Imperial College London

Development of a framework for designing nucleic acid-based, out-of-equilibrium catalytic reaction networks.

Ismael Mullor Ruiz

Thesis submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy

Department of Bioengineering

April 2021

Declaration of originality

I hereby declare that the work presented in this thesis has not been submitted for any other degree or professional qualification, and that it is the result of my own independent work.

Ismael Mullor Ruiz (Candidate)

Declaration of Copyright

Ismael Mullor Ruiz.

Abstract

DNA nanotechnology and Toehold Mediated Strand Displacement (TMSD) offer the possibility of building systems that display complex algorithmic behaviours. However, up to the present point most systems based in this technology act as one-shot systems that relax into equilibrium when they operate and lose any chance of responsiveness and adaptation to changing environmental outputs.

This situation is starkly opposed to how living systems react to changes in their environment and implement finely-tuned responses. Cells perform this feat via biochemical transduction networks, which have previously been described as distributed computational systems operating out of thermodynamic equilibrium. While transduction networks are highly complex, their fundamental building motif is well known: the push-pull network. In this motif, a substrate is switched between two states via two fuel-consuming catalysts. This out-of-equilibrium operation allows push-pull systems to propagate information robustly in signal-transduction cascades presenting while complex dynamics.

Understanding the main operational constraints of push-pull systems is a fundamental question in systems biology and a requirement for their proper use in synthetic biology. Moreover, building DNA-based analogues of push-pull systems would allow us to explore these fundamental questions and provides an ideal engineering platform for implementing non-equilibrium information processing systems in biological and nanotechnology contexts.

While emulating the behaviour of transduction network in TMSD systems is possible in principle, there are several performance issues that hinder this possibility. In order to overcome these limitations, we propose the Active Circuits of Duplex Catalysts (ACDC) Framework to meet this challenge. In ACDC, all species are DNA duplexes that interact directly via four-way strand exchange. The present thesis demonstrates that the ACDC Framework can successfully implement all the prerequisite to build extended catalytic reaction networks. Additionally, we discuss functional and formal limitations of the Framework as well as the role of thermodynamic drives in overcoming them.

Publications associated with this research

The experimental work on the implementation of a hidden thermodynamic drive inside a seesaw gate reaction motif was featured in:

-Haley, N. E., Ouldridge, T. E., Mullor Ruiz, I., Geraldini, A., Louis, A. A., Bath, J., & Turberfield, A. J. (2020). Design of hidden thermodynamic driving for non-equilibrium systems via mismatch elimination during DNA strand displacement. *Nature communications*, *11*(1), 1-11.

The discussion of the formal structure, nomenclature, and formal limitations of the ACDC Framework as well as the proposal for troubleshooting these formal limitations were featured in:

-Lankinen A., Mullor Ruiz I. & Ouldridge T. E. (2020), Implementing non-equilibrium networks with active circuits of duplex catalysts. *26th International Conference on DNA Computing and Molecular Programming (DNA 26)*, Publisher: Schloss Dagstuhl--Leibniz-Zentrum, Pages: 1-25.

Acknowledgements

The present thesis is the result of many of people that directly and indirectly have made it possible to me to be at this situation. And for them, the recognition for what they have given to make this possible must be given. These years have been a trip with a lot of personal changes besides my growth as a scientist, and there's a lot to recognize.

The first people who should be recognized are evidently my supervisors, Dr Thomas E. Ouldridge and Prof. Guy-Bart V. Stan. When we first met almost 5 years ago, I had very clear that the path I wanted to follow in my life was in science -and more precisely it was in nucleic acid nanotechnology. But I never expected to be given an opportunity as good as the one I received. And I really feel honoured of having had the privilege of working with them and being challenged by problems and fields that were tangential at most to my then-interests- but have now become fields I want to dive in. They gave me a challenge after a time in my life in which I was told I was not good for what I felt it was my passion, as well as new ideas into which dive and get lost. And for all that, I thank them.

Evidently one of this Ph.D could have not been done without any type of material support, and for that I want to thank the funding bodies of my supervisors. Hence all of them at all the levels must be acknowledged. Beginning with the Department of Bioengineering at Imperial College London as well as external bodies the Royal Society for the Research Fellowship that allowed Dr Ouldridge to settle his group in the Department.

But no funding support would have resulted in anything with a proper work environment, and needless to say, the people in both the Ouldridge and the Stan labs have been the best travel partners I could find in this endeavour. Beginning with Dr Wooli Bae, whose continuous support and expertise on FRET was fundamental to design how to trace the different components of the networks we built. Alongside him the list is enormous – five years can make a lot when it comes to meeting people. However, alongside him, Javi, Alicia, Jordy, Jenny, Abishenk, Rory, Franc, Rakesh, Aditya, Tomislav, Tom Addenbrooke and all the master students that came and went through the Ouldridge group should be acknowledged. For giving a precious feedback that helped me unclutter my message when I wasn't delivering it properly for my audience as well as for helping me when I felt the clumsiest and for having patience with me at my worst. And the same should be said about the amazing people of the Stan Lab. For all that, Alice, Albert, Andreas, Marios, Nicolas, Ezster, and all those that have come and go, thanks for being there and making Imperial and the lab the amazing place it is. And for the people of the labs I am not associated with (Rodrigo, Joaquin, Lara, Davina, Gediz, Tanel, Cameron, Tigran, Amrit, Angeles, Silvia Mercedes and many others). You also made this time here amazing.

Our external collaborators have been also a crucial part of the development of the project providing with guidance and feedback in crucial moments of the thesis. For all this, I would also like to thank Prof Andrew Tuberfield and Prof. Ard Louis from University of Oxford

as well as Dr. Natalie E. Haley for having allowed me to collaborate in the research about hidden thermodynamic drives that they were carrying out alongside my supervisor Dr Ouldridge as well as thanking Dr. Andrew Phillips from Microsoft Research Cambridge with suggestions on adjusting reaction parameters that vastly improved the system's functionality. Although more informal, the discussions provided by Dr. Christian Cuba Samaniego from UCLA also resulted crucial in order to understand where the relevance of punctual results really resided.

In addition to all the people that I met in this wonderful 5 years, Prof. Manuel Monleon and Dr Carlos Marti-Gastaldo, my former supervisors, should also be thanked. Without their previous support, I would not have had the chance to shine and find my path, even when I was riddle with doubt and anxiety.

But if my previous supervisors are acknowledged, my supervisees should be as well, since alongside them I learnt what entails to mentor and guide. It goes specially for Yash Shukla and Antti Lankinen, whose brilliance never ceased to amaze me. I am pretty sure that without them, (specially in Antti's case) this thesis would not have been the same. But it also goes for all the group projects I handled in one way or another. For all the people of iGEM 2018, Shiv, Lidia, Albi, Thomas, Yutong, Siwat, Josh, Dielza, Luis and Will- the BIOMOD 2019 - Albi again, Hansa, Georg, Sophie, Shabzs, Edward, Alisdair, Kevin) and the little help I tried to give to the iGEM team in this last turbulent year (Olive, Raymond, Gabby, Gabriel, Hia, Maria. Elosie and the others). I feel honoured of watching what you achieved and have been a little part of it.

And last, but not least important, my deepest thanks go for my family and friends who were not physically there during the process, but still supported me at the distance. Mom, Dad, Pascual, Maria, Adolfo, Menacho, Rous, Pepelu, the Alexes (Ferri and Damian Serrano), Orphen, Salva, Rael, David, Elena, Sergio, Edu, Becca, AJ, Ester, all my extended family... You cannot imagine how much I thank you for being there, and for having supported me even in my worst state of mind when the distance (in all the possible senses) between us was at its worse. I might have faced hardship throughout these years but I thank you for being there, having grown alongside me and make me find the way back when I needed it most.

I am leaving lots of people out in this humble attempt at recognition - unintentionally and otherwise. And certainly, to try sum up things, as the Bard said, all the world's a stage and men and women are merely players. So, for all those who played a role in this play, either big or small, heroic or otherwise. thanks for doing so while you starred in the plays of your own. Every single action leads us to where we are, and certainly, I could not be happier with who I am right now and what I achieved and the stories I carry with me after this experience.

Ismael Mullor Ruiz

Table of contents

List of figures

List of tables

Chapter 1: Introduction to nucleic acid nanotechnology and current state of the art.

1.1.- Nucleic Acids: Brief history and biological function.

Nucleic acids are a family of biomolecules discovered by Swiss chemist Friedrich Miescher in 1868 who isolated them from white blood cell samples under the name "nuclein"[1] and characterized their slightly acidic behaviour. Chemically speaking, nucleic acids can be defined as aperiodic polymers formed by monomers called nucleotides. At the same time, each of these monomers is composed of three subunits: a phosphate group, a sugar molecule (Ribose or its derivate 2-Deoxyribose) and a nucleobase molecule which can be either a purine (Adenine, Guanine) or a pyrimidine (Thymine, Uracil, Cytosine). Their importance to biological studies was proved crucial by Oswald Avery, Colin MacLeod and Maclyn McCarthy's discovery of one of these molecules (Deoxyribonucleic Acid –abridged DNA-) being the physical basis for storage of the genetic information.[2] This discovery, along with contemporaneous discoveries regarding biomolecular crystallography and the establishment of the dependency of biological function on structure, raised the question of the structure of the DNA molecule. Such knowledge would shed some light on the heritability mechanisms of genetic information. This structure was first described by James Watson and Francis Crick[3] in 1954, in a work that would earn both of them along with Maurice Wilkins the Nobel Prize in 1963: **the double helix** (displayed in Figure 1).

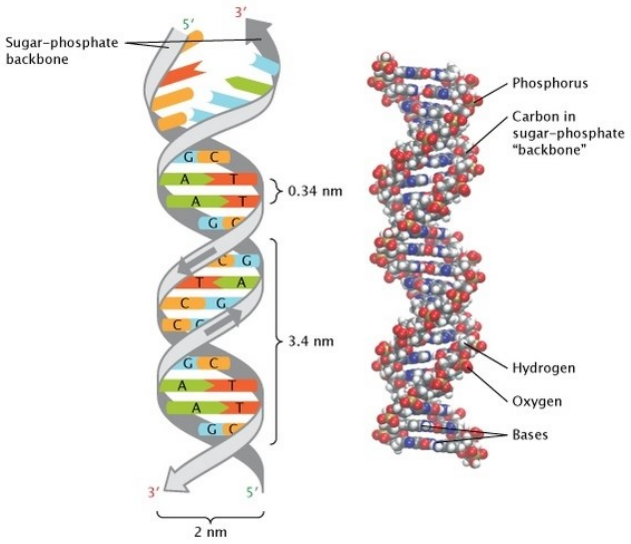

Figure 1. Structure of the B-DNA double helix. Adapted from Pray (2008)[4]. © Nature Education.

In the aforementioned double helix structure, two DNA molecules are placed one opposite each other in an antiparallel manner. The sugar molecules and the phosphate groups of the nucleotides placed on the outside form the backbone. The phosphate group bonds the nucleotides of the chain through a covalent bond between the 5' hydroxyl group of the 2'-Deoxyribose of one nucleotide and the 3' hydroxyl group of the adjacent nucleotide. Perpendicularly to the backbone are placed the nucleobases, which form specific parings Adenine–Thymine and Guanine-Cytosine via hydrogen bonds (henceforth called Watson-Crick pairings). The most common, stable and widely considered DNA structure (B-DNA), the double helix, is a dextrogyrous one with the presence of two differently sized grooves (minor and major groove) with a total diameter of 2.37 nm and a thread pitch of 3.54 nm (approximately equivalent to 10.4 basepairs). Other possible structures include A-DNA or Z-DNA, formed by changes in the hydration state or the ionic force of the medium in which the molecule is immersed. In these structures, the thread pitch, the diameter and even the sense of the pitch change.

In addition to these structures, we can also find configurations that do not rely on Watson-Crick pairings but rather on Hoogsteen[5,6] pairings like DNA triplexes[7] and G-Quadruplexes[8]. However, none of these structures will be taken into account into the project discussion since B-DNA is the most stable configuration of the DNA molecule in typical physiological conditions. This circumstance does not deny that Hoogsteen pairings are relevant biologically or technologically: there is a growing body of literature that highlights that G-quadruplex intermediates have relevance in tasks as important as maintenance of promoters and telomeres[9], and there are potential applications of triple helixes in pH and electrochemically-based sensors[10]. However, such endeavours, as we previously pointed, are out of the scope of the present thesis. As mentioned before, B-DNA and Watson-Crick pairings are the most stable and hence the ones that are easier to occur and predict, a property of crucial relevance for our purposes and that allows us to design and predict DNA assemblies with computer-assisted design tools such as NUPACK[11] or CaDNAno[12].

The discovery of the double helix structure of DNA by Watson and Crick in 1954 is historically considered as the starting point of molecular biology, since the model built did not only take into account all the observed data as the X-Ray diffraction structures obtained by Rosalind Franklin, Maurice Wilkins and Raymond Gosling[13], or Edwin Chargaff's empirically obtained nucleotides ratios[14], but yielded a deeper understanding of the processes that involved DNA. Subsequent breakthroughs began with the discovery of the replication mechanisms that could explain heritability of genetic information, as the authors noted on their foundational work. This work, along with the discovery of all the protein machinery that made DNA replication possible, as well as the translation of the DNA information into functional proteins would later culminate in the so-called Molecular Biology Central Dogma[15]. In this work, Crick gave the first description of how genetic information flows from DNA to proteins through RNA transcription inside living cells (even though it had to be revised after the discovery of retroviruses[16]). This understanding paved the way for the upcoming genetic engineering revolution and the subsequent surge of synthetic biology as a field on its own

with the addition of engineering standards and design practices[17]. But, besides this acknowledged paradigm shift, Watson and Crick's double helix work would later be revealed to be foundational for another different scientific subfield: **nucleic acid nanotechnology.**

1.2. - Nucleic Acid Nanotechnology: a short introduction.

1.2.1. - Origins of the field: Structural DNA nanotechnology and directed self-assembly.

The term nanotechnology refers to the design, manipulation, and control of matter on the dimensional scale of the very few nanometres, often involving the manipulation of single molecules or a few atoms with precision. The term was first coined by Norio Taniguchi in 1974 referring to thin film and ion beam lithography processes for semiconductor devices that allowed the controlled and reproducible building and miniaturisation of the devices to the aforementioned dimensional scale[18]. Despite this, hints of this possibility as well as some of the possible implications were already given by Richard Feynman on his seminal 1959 talk "*There's plenty of room at the bottom*"[19]. Although nanotechnology was born and focused on solid state physics, electronics, and semiconductor devices design, it quickly began to grow and include concepts, experimental approaches and goals from other fields such as colloid science, surface science, coordination chemistry, supramolecular chemistry, or molecular magnetism, only to mention a few.

In the early 1980s, DNA chemistry became one of these tools used for atomically-precise nanotechnology with Nadrian Seeman's foundational work in structural **DNA nanotechnology** [20]. Here, the Watson-Crick base-pairing present in the double helix was exploited in order to build DNA junctions that could be connected with each other, thus producing programmable regular two- and three-dimensional structures such as regular shapes and lattices[21]. These structures were obtained via rational design by exploiting the high specificity and predictability that Watson-Crick base pairings confer to the molecular binding as well as to the structure formation.

Initially, these programmable structures were proposed as a tool for assisted crystallisation of proteins, but the realisation of prescribed structures with DNA wasn't as straightforward as Seeman initially devised and the first full realisation of periodic 3D lattice DNA crystals from a preconfigured structural motif was achieved in 2009[22], almost 30 years after Seeman first came up with the concept. However, during the development of the technology, several concepts and tools of crucial importance were tested and spawned subfields of their own. In this vein, Seeman and collaborators developed motifs such as the **Double Crossover (DX**) [23] (depicted in Figure 2) inspired by the occurrence of 4-stranded intermediates in Holiday Junctions during meiotic DNA recombination. Using this motif as the basic subunit, Seeman demonstrated that extending the end of the motif, several subunits could assemble to form 2D lattices[24], assembly of tiles, or tensegrity-driven[25] trigonal subunits that could further self-assemble in periodic structures in 2D and 3D. DX-motif-based structures would later also become the basis for simple devices that exhibited mechanical

actuation[26]. These devices, all by themselves would later spawn a whole new field, dynamic DNA nanotechnology, which will be the focus of the following subsection.

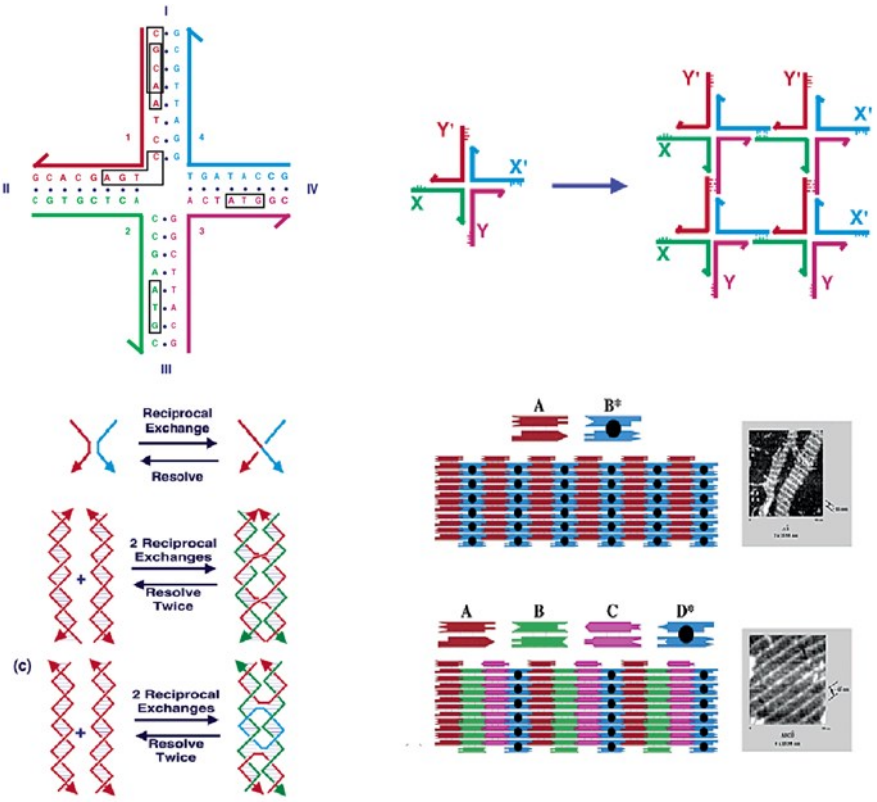

Figure 2 (a) Structure of a 4-way junction with its sequence domain. (b) Formation of a periodic two-dimensional lattice based on the 4-way junction motif from (a). (c) Formation of the double crossed (DX) motif both from antiparallel double strands of from parallel double strands. (d) Assembly of DX tiles of two and four different types on a surface. Adapted from Seeman (2003)[27].

In parallel with DX-based systems, other paradigms of directed DNA self-assembly arose years later and extended the possibilities of directed assembly of DNA. Amongst these directed assembly strategies, there are three that have gained the attention of the scientific community particularly. The first of these strategies is DNA origami.

DNA origami was developed by Paul K Rothemud at Caltech in 2006[28] and presented a novelty when compared with Seeman's assemblies: instead of assembling many subunits composed of the same type of building bricks and building with them a periodic structure, DNA origami works with two different types of building subunits as can be seen on Figure 3. The first type of subunit is the large backbone, normally comprised by a single strand of the natural bacterial phage genome. The other type of subunit are short oligonucleotides, called staples, that are designed to direct the bending of the backbone (with the help of some thermal energy that avoids the folding of the structure into an undesired local energy minimum instead of the programmed desired global minimum) into a discrete, defined structure.

The origami structures can be either 2D or 3D[29] and the methodology allows for larger structures than those that could be achieved back in the day by DX tiles assembly. These origami can stand on their own, but their design versatility allows the inclusion of unpaired edges in the structure that can interact with other well-defined edges giving place to a defined macromolecular meta-assembly[30] if the edges of the origami are designed for that purpose since the staples allow the edges of the origami to be blunt and behave as finite independent objects.

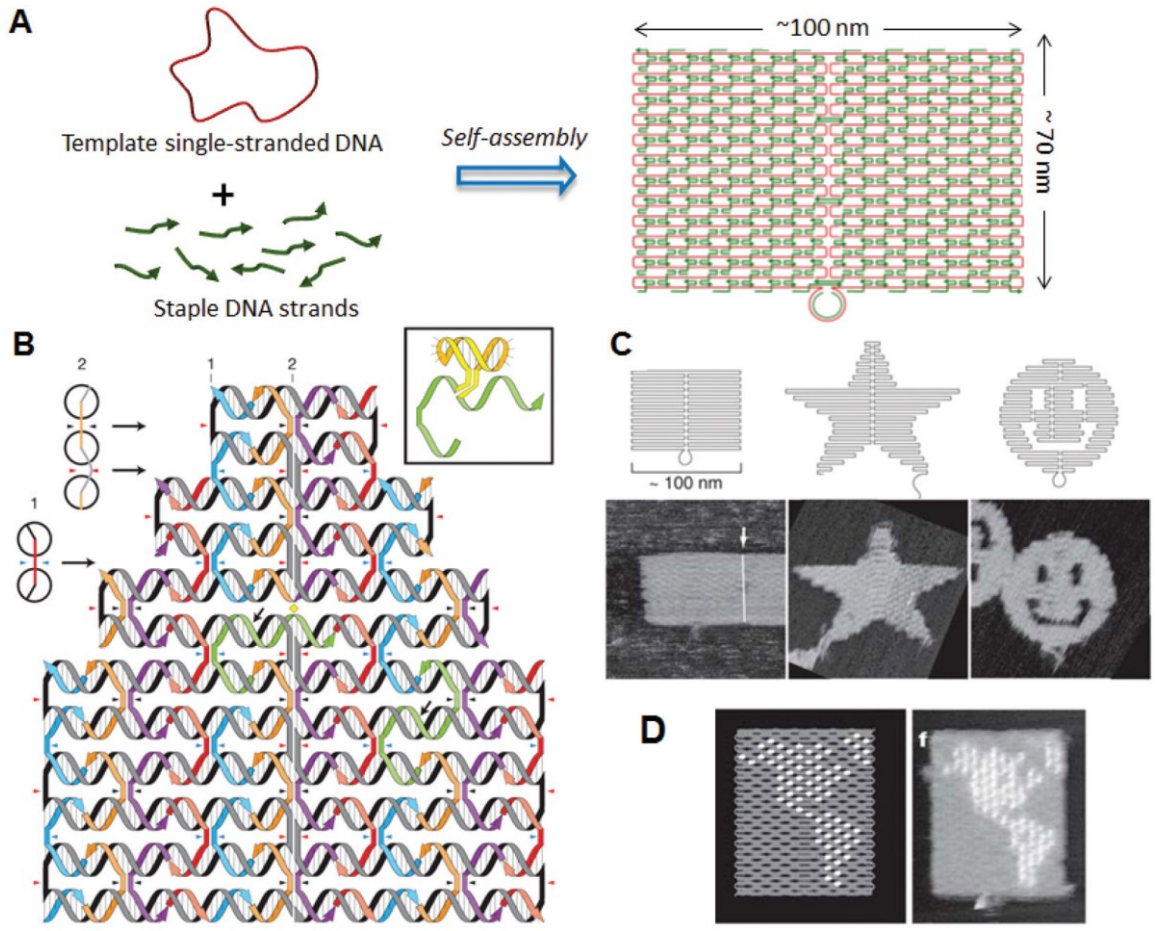

Figure 3. (A) Quick schematics of the formation of a rectangular DNA origami. (B) Detailed schematic showing in double-helixes how the staples are located throughout the origami and how do they orient themselves throughout the process. (C) Example of different discrete 2D forms built with the origami technique obtained via AFM. (D) Example of the implementation of different heights via staple elongation. The elongated staples appear as brighter points in the AFM image. Adapted from Endo et al (2013)[31].

DNA blocks, on the other hand, are a much more recent development from Peng Yin's lab at the Wyss institute in Harvard[32]. The DNA blocks strategy is based on the design of short oligonucleotide molecules (in principle each block is 32 nucleotides long, but modifications in length can be done depending on shape particularities as well as if a particular brick is part of the edge of the prescribed structure). In this working framework, the Watson-Crick pairings are designed in such a way that the sequence can be divided in four specific interaction domains, with the possibility of model the single-stranded bricks as a LEGO-like brick as depicted in Figure 4. Again, due to Watson-Crick pairing specificity, the possible interactions

between the bricks on a brick set can be defined only by their edges' sequence. Hence, on a single run in a single pot, all the given bricks of a set can assemble non –hierarchically in a prescribed 3D structure which, while having nanometric precision in their positioning, can reach a size in the scale of microns.

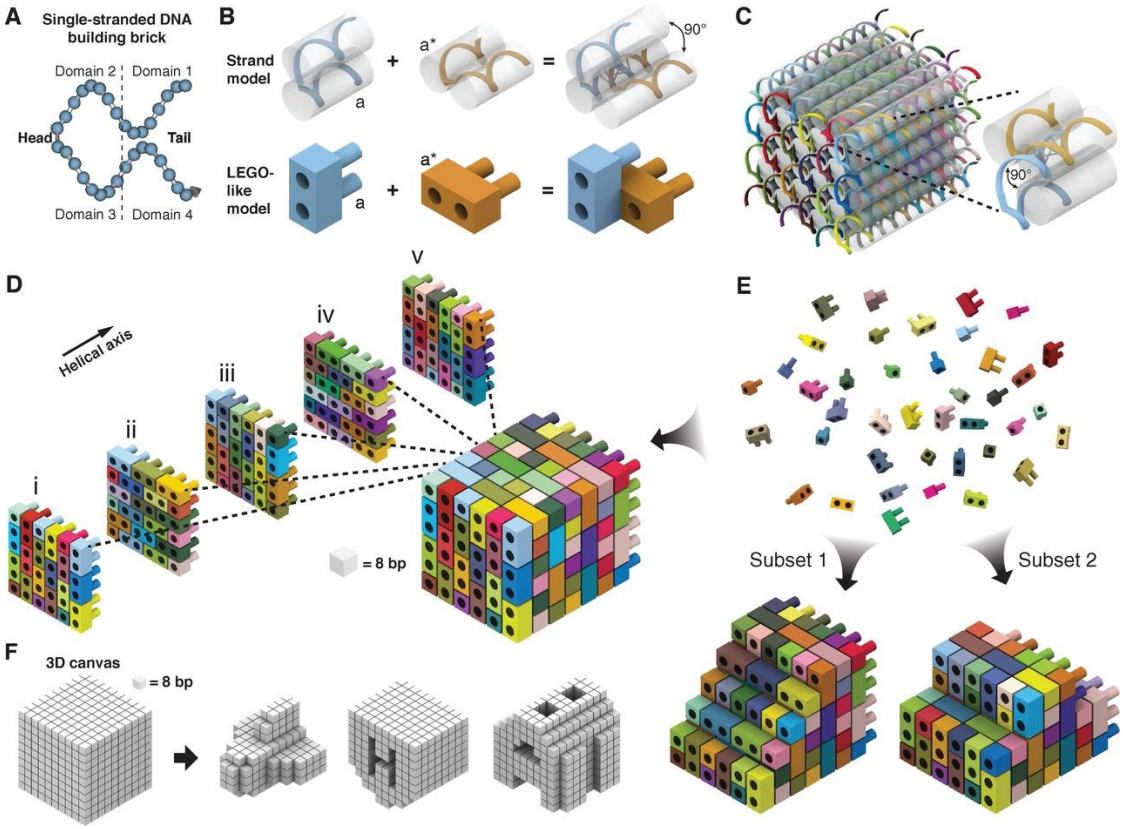

Figure 4. (A) Beads representation of a single DNA brick unit. Each nucleotide is represented by a ball in this abstraction in which the head and the tail of the brick are highlighted. (B) Schematic showing in double-helixes as well as with the LEGO-like representation how two DNA bricks get assembled one with each other. (C) Example of a discrete 3D polyhedron shape built with DNA bricks represented with DNA helixes bundles (D) Bricks representation of the assembly layer by layer of the polyhedron at (C). (E) Schematic representation of the dependency between the assembled shape and the set of bricks available in the system. (F) Demonstration of how to program discrete shapes in DNA bricks. Adapted from Ke (2014)[32].

The last relevant structural DNA nanotechnology framework developed after Ned Seeman's foundational work **is DNA wireframe[33].** As, the names says, in this framework (described in Figure 5), DNA can form finite structures, but instead of building them in bulk as in DNA origami or DNA bricks, DNA constitutes only the edges of the structure. As with DNA origami, the design process of the structure is completely computer assisted, allowing us to implement an automatized pipeline in which the structures can go directly from graph definition to DNA sequence realisation after calculations that optimize the tensions to which each edge is submitted in the structure as well as minimize the number of strands required.

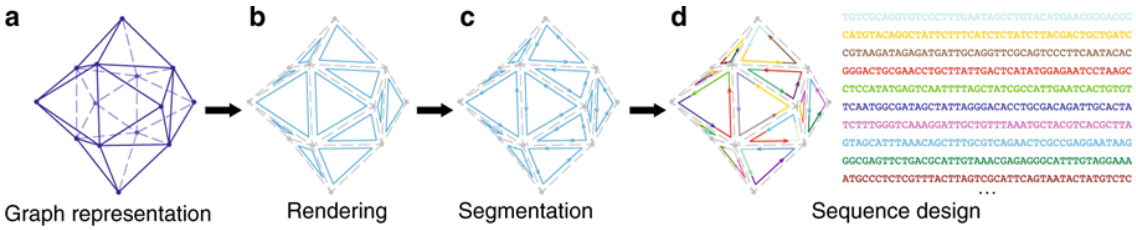

Figure 5. Quick summary on the basic working steps of an automatized pipeline for sequence design for a specific DNA Wireframe structure- Adapted from Wang et al (2019)[34].

One important notion that emerged during the development of directed self-assembly systems is that the interactions of each subunit can be defined by the Watson-Crick pairings it can form and hence, that it can be programmed. This programmability allows a designer to embed an algorithm in the assembly process, and use the building of a given structure as the output of a computation. This notion of algorithmic self-assembly was first demonstrated by Erik Winfree and collaborators for the DX tile assembly[35] of a fractal Sierpinsky triangle (as depicted in Figure 6). The demonstration of the possibility of algorithmic self-assembly, alongside Leonard Adelman's[36] 1994 paper on the solution of the 5-nodes Hamiltonian using programmed DNA interactions are widely considered to be the founding works of the **molecular programming** field. Albeit their differences (Adelman's system was based on PCR, hence requiring enzymes and a heat source external to the system) both works showed the computational power that directed Watson-Crick interactions could provide. Algorithmic self-assembled systems have continued to be a topic of great interest for the research community till the present day[37].

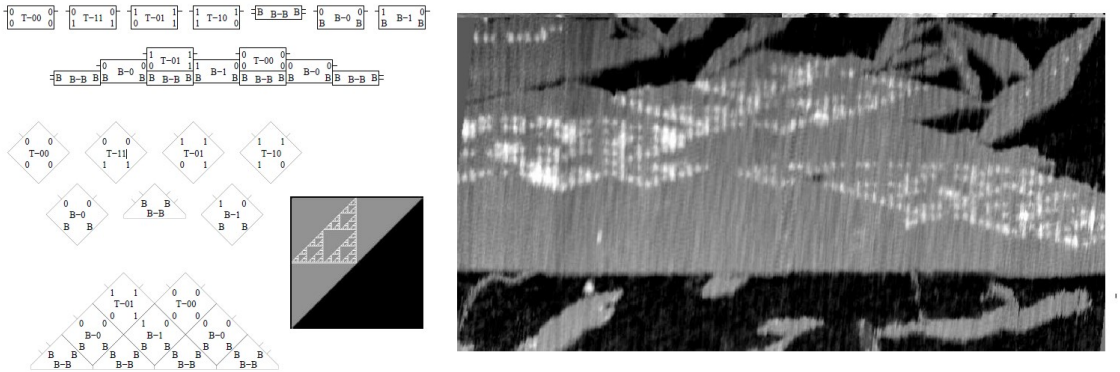

Figure 6. (L) Schematics of the design of DNA DX tile design for the algorithmic assembly of fractal Sierpinsky triangle. (R) AFM images of the actual algorithmic assembly of the preprogramed pattern. Adapted from Rothemud et al (1999)[35].

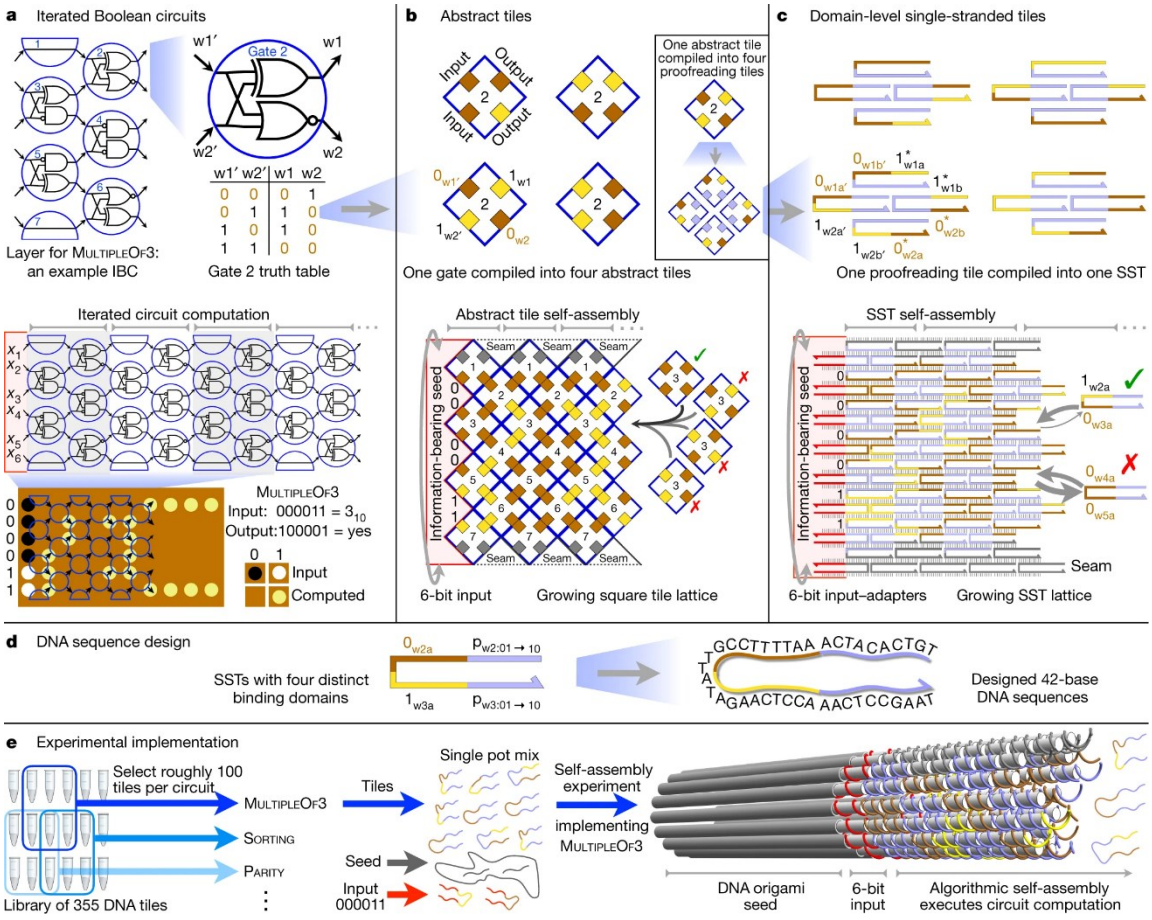

Figure 7. (A) Representation of the logic gates abstraction of a universal computing architecture based in arbitrarily extensive interactions between 2-Inputs-2-outputs logic gates. (B) Implementation of the formalism described in (A) in a tiles assembly system. In this framework all the tiles have 4 possible interaction: two corresponding to the inputs of the logic gate and the other two corresponding to the outputs of the given gate. Using the possible permutations of these inputs and outputs, different functions with different outputs corresponding to the assembled structure. (C) Formal implementation of the theoretical 2D tiles of (B) in single-stranded DNA tiles. (D). Domain and sequence design of the tiles. Adapted from Woods (2019)[37].

Many articles speculating about the possibilities of algorithmic DNA reactions and their massive parallelization appeared at the turn of the Millennium. However, a huge number of drawbacks kept these possibilities from coming into practical fruition. One of these drawbacks, particularly relevant to algorithmic self-assembly, lies in the fact that self-assembled systems execute their algorithm by being driven to a state of thermodynamic equilibrium. This means that once the process has been completed, there is no way of rebooting the system or reprogramming it rather than rebuilding it from scratch, hence posing a serious problem in terms of continuous operation and performance. These drawbacks and how are they faced will become an important part of the discussion in the present chapter and the thesis as a whole.

1.2.2. - Dynamic DNA nanotechnology: actuators, sensors and computers.

As we have seen in the previous section, the origins of DNA nanotechnology relied on the design of self-assembled static structures that converged into an equilibrium state. But

this was not the only way in which nucleic acids demonstrated potential for technological advances at the nanoscale. In the late nineties and at the beginning of this Millennium, two foundational works demonstrated that DNA could not only be a static building material, but could also be used to build active components, thereby kick-starting the field of dynamic DNA nanotechnology[24]. The first of these foundational works was published in 1999, again by Seeman's group. It consisted –as seen in Figure 8- in a simple actuator that through a change of conformation between the B and Z conformations of DNA activates or inactivates a FRET pair of fluorophores[26].

The working mechanism of the actuator required changes in the salinity conditions of the buffer in which the device is working, which, despite being a characteristic desirable in technologies such as sensors and being a feature that has been researched and desired, could be an eventual operative limitation for any device meant to **operate autonomously** without continuous external changes of the overall conditions of any kind.

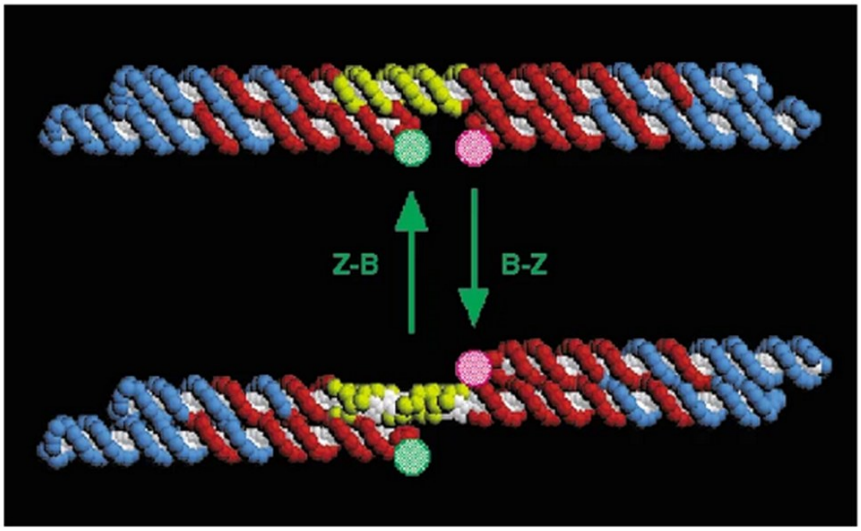

Figure 8. Salinity triggered transition between B and Z-DNA of a DX junction-based actuator. In this structure the changes in structure triggered by the changes in salinity result in a change of the FRET signal detected in bulk-. Adapted from Mao et al (1999)[26].

However, shortly after the appearance of these rudimentary actuators, Bernard Yurke, Andrew Tuberfield and collaborators introduced the first DNA-fuelled DNA nanomachine[38]. This nanomachine consisted of a pair of tweezers whose functionality was provided by the directed hybridization of DNA molecules that acted as a molecular fuel –as seen in Figure 9-. This notion of the use of DNA molecules as fuels for autonomous machines would be later exploited and expanded by the DNA walkers –like the one depicted in Figure 10- designed initially by John Reif and developed by John Reif, Andrew Tuberfield[39,40] and collaborators to be later carried on by many others[41,42,43]. In a similar manner to how myosin proteins can walk alongside an actin filament, a DNA walker can hybridize and go along a DNA single filament using a fuel molecule at each step producing as a result net displacement as well as an inert duplex.

This imitation of natural biomolecular systems with programmable nucleic acid systems[44], seen not only in the aforementioned DNA walkers, but also in machines like the

DNA Origami-based pores inserted in lipid bilayers[44], led Ned Seeman to coin the term ***kleptobiology.*** With this neologism, Seeman referred to a design strategy in which the underlying working mechanisms of a given system are transplanted into another biomolecular substrate that does not have such function in nature. But the most important development that these rudimentary nanomachines made possible wasn't in the *kleptobiology* concept, but rather in the molecular mechanism that made possible the use of molecular fuels: the selective hybridization of specific molecules only when certain domain sequences called **toeholds** were exposed.

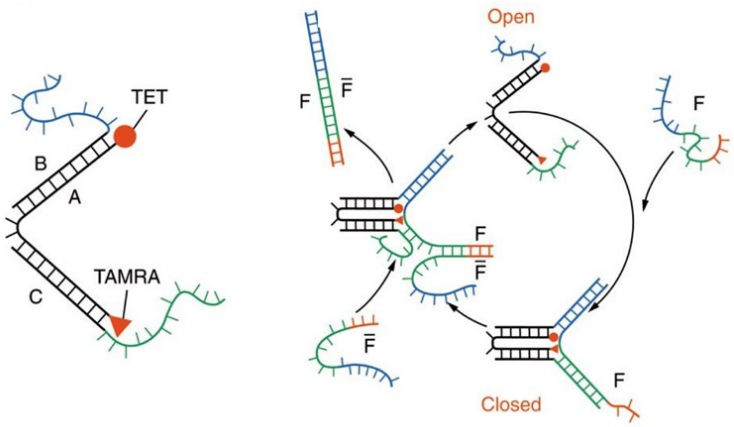

Figure 9. Autonomous DNA tweezers. The tweezers can be in two different states (open and closed and the transition between the two states involves the hybridisation of either a fuel or an ant-fuel strand to the tweezer when its either open or closed respectively. The hybridisation of the corresponding strands is mediated through the exposure of the corresponding toehold. The transition between the two states is monitored through the fluorescent signal of the TAMRA dye attached to one side of the tweezers-. Adapted from Yurke et al (2000)[38].

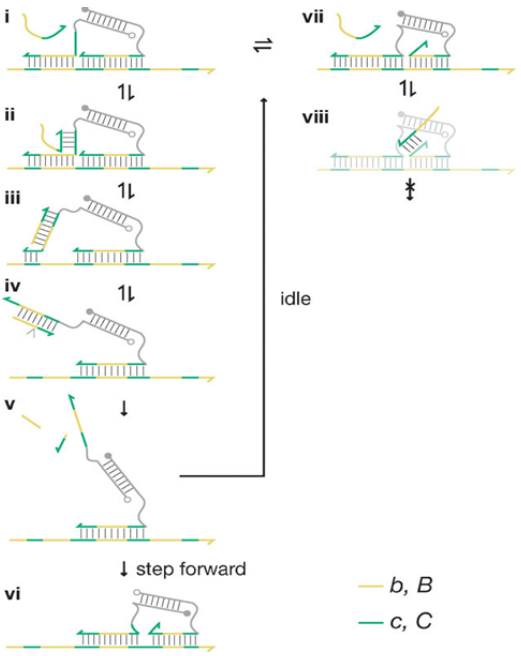

Figure 10. Illustration of a DNA-powered walker. This walker features 2 feed that recognize the domain structure to which they are attached in the single- stranded lane. The process of the binding and unbinding of the feet from the track is in principle random, but adding a mechanism that after the unbinding of a foot destroys the incumbent fuel (i.e. enzymes) allows the walker to produce net motion. Adapted from Bath et al (2009)[27].

The introduction of the concept of toehold became one of major relevance in the field of dynamic DNA nanotechnology field and its importance cannot be overstated. As seen previously with the assembly of static structures, the presence of an exposed, given sequence in the system determines the possibility of a given reaction via Watson-Crick pairing. However, with the concept of toehold it was introduced the possibility of **toehold exchange** (thus meaning that, as a result of the toehold mediated reaction, another toehold became exposed hence allowing molecular programmers to implement the dynamical multi-stepped molecular machinery described previously).

However, the introduction of toeholds DNA nanotechnology not only brought the possibility of designing complex autonomous machinery, but also the chance to fine tune the speed of process since the variation of length of the toehold allows us to have a control the reaction rate of six orders of magnitude [45]. This exquisite degree of control given by toeholds also extends to the possibility of determining thermodynamic endpoints of a given system when no toehold is exposed after reaction and to the chance of implement them in different strand configurations as seen in Figure 11 c and d, opening more design possibilities.

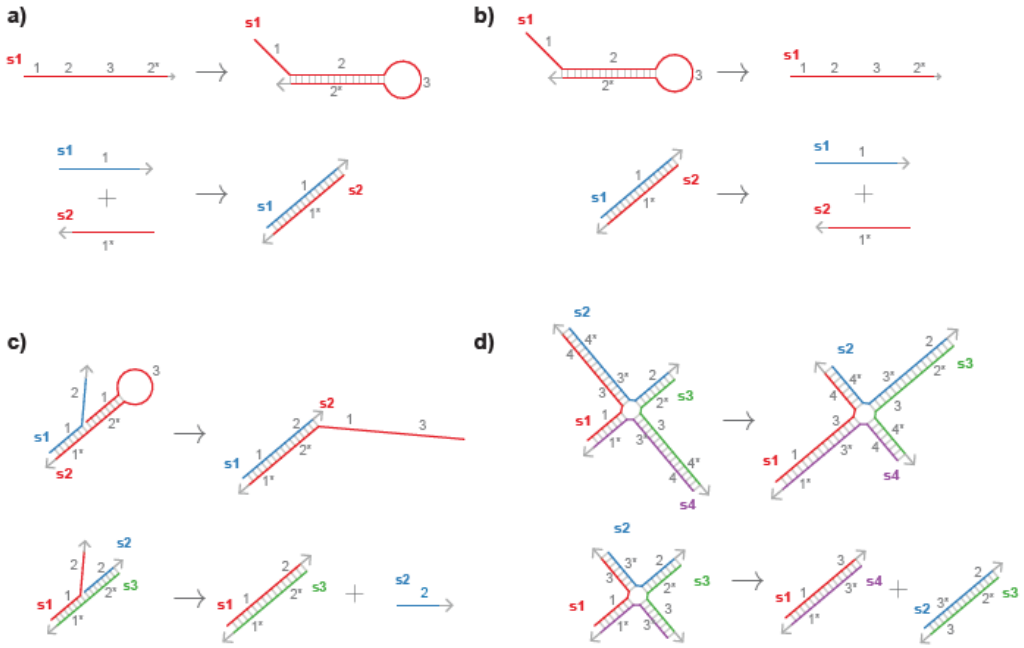

Figure 11. (a) Examples of hybridization in both a single strand forming a hairpin and between two strands forming a duplex. (b) Spontaneous dissociation of the structures formed in (a) giving place to their fundamental constituents. (c) Toehold-mediated strand displacement reaction of the opening of the hairpin and the displacement of s2. (d) Toehold mediated 4-way branch migration resulting on the formation of a 4-way junction that gets resolved into 2 duplexes. Adapted from Grun et al. (2015)[46].

It must be said, however, that despite toeholds being fundamental for dynamic DNA nanotechnology systems (particularly for computing systems), and as it can be deduced from Figure 11, there currently isn't a single toehold recognition-based framework, but rather many with variations that affect performance. **Hybridisation chain reactions[47] (HCR)** developed by the group of Niles Pierce constitutes the first paradigm for DNA computing

reaction, in which the addition of an input strand allowed floating DNA hairpins to assemble sequentially a series of hairpins into a preprogramed linear structure, as seen in Figure 12.

The change in the opening of the hairpins and the appearance of the structure can be associated to a change in a fluorescence signal that can be quantified to determine how much of the input has reacted. The technique, besides, being a pioneering framework which is still being substantially developed, has found applications in research on embryonic development in which the system is designed to use a specific mRNA as initiation input for the chain reaction[48]. The capability of interfacing our designed biomolecular systems with those in living systems is a non-trivial point. However, it is this very fact that allows us to exploit them to create systems interesting for applications in biomedicine or synthetic biology.

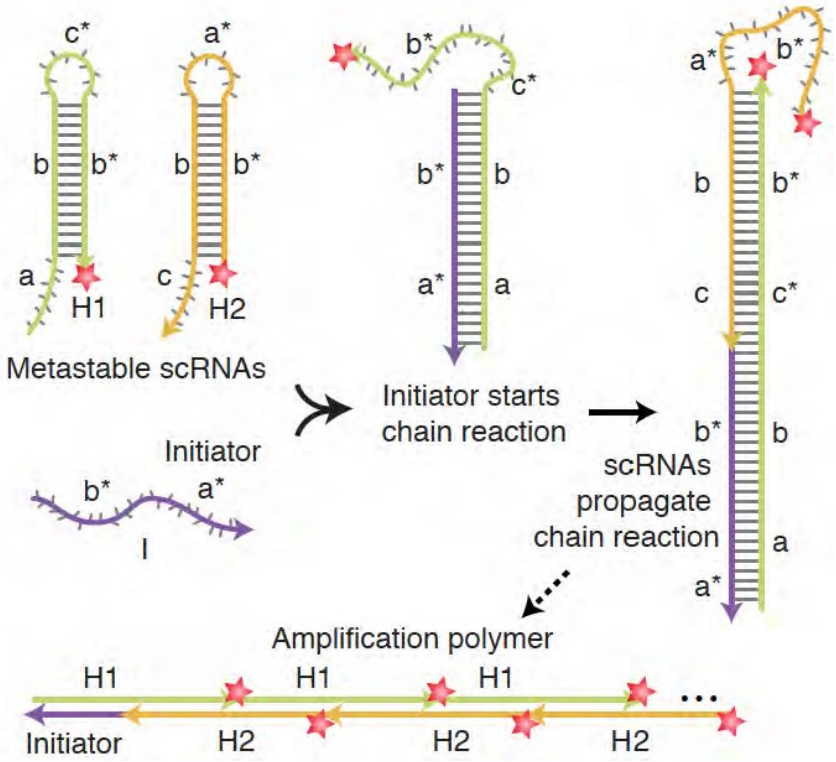

Figure 12. Diagram of an HCR amplification-based system. The formation of the polymer can only happen if the initiator is present on the system-. Adapted from Choi et al (2010)[48].

Importantly, as stated before, HCR is not the only nor the dominant framework in dynamic DNA nanotechnology. Currently this honour sits with **DNA strand displacement (DSD)** reaction networks[45] which will become the focus of the present project and will be explained later at further lengths. DSD reaction networks were developed shortly after HCR in 2006. Once DSD was established, various paradigms followed such as the **PEN** (Primer-Endonuclease-Nickase)[49] **Toolbox** developed by Yannick Rondelez, which includes the action of enzymes as components of the system. The PEN Toolbox uses the aforementioned enzymes to mediate the toehold exposure and strand displacement reactions and has been used to demonstrate the possibility of creating spatial control of the reactions via control of reaction-diffusion reactions in a microfluidic environment. Other notable enzyme-mediated

systems include the **PER (Primer Exchange Reactions) Toolbox** developed by Peng Yin and collaborators at Harvard[50]. The PER Toolbox has the capability to register and produce different outputs depending on the temporal sequence of inputs.

However, despite their apparent improved functionality over DNA-only systems, enzyme-mediated systems have the drawback of a lesser controllability since enzymes specificity and reaction rate are harder to control than in DNA-only systems and much harder to tune given the higher complexity of enzymatic machinery. Hence, the focus of our project will be on the working framework that has been proved to offer the best compromise between versatility, performance and controllability: **DNA Strand Displacement reactions**.

1.2.2.1.- DNA Strand Displacement (DSD) Reactions: fundamentals, implementation on Chemical Reaction Networks (CRN) and limitations.

As said before, the predominant framework for DNA computation is DNA Strand Displacement reactions (DSD). This technology, initially developed by Erik Winfree, Georg Seelig, David Soloveichik, David Zhang, Andrew Turberfield and Bernard Yurke[51], has become a hot topic in DNA nanotechnology since its inception. It has at its basis a reaction called **toehold-mediated strand displacement (TMSD)**. At the core of this basic process (displayed in Figure 13) are two basic components: a single-stranded DNA input and a duplex in which the corresponding output is hybridised with another strand (the gate) that has a little fragment remaining single stranded (the toehold). The input is totally complementary to the gate strand, and so its binding to the gate is thermodynamically favoured over the output. Toehold-mediated strand displacement takes place in a two-step process. During the first step, the union of the input fragment to the toehold complementary fragment takes place, while in the second step the output strand gets released through an exchange of base pairs.

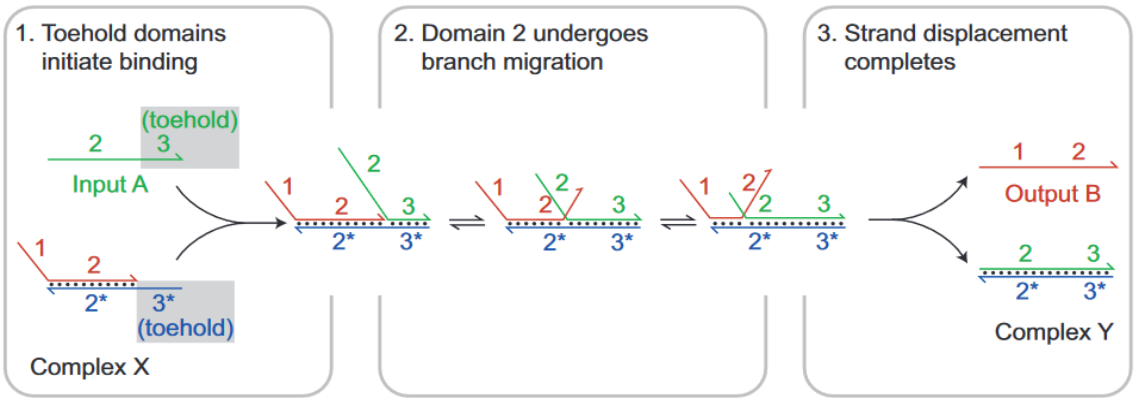

Figure 13. Diagram of a toehold-mediated strand displacement reaction. In this reaction, the toehold sequence is marked in red. The output of the reaction can hybridise with another toehold of another reaction thus allowing to create programmable cascades. Adapted from Zhang, Seelig (2011)[52].

It must be remarked that the biophysics of this process are deeply understood in terms of the different free energy landscapes of all the reaction network species[45]. Its principles are robust and stable enough to be able to be transplanted into other similar toehold-based reaction pathways, such as 4-way branch migrations[53], as well as systems based on other nucleic acids such as RNA[54]. In this case, it must be noted that the design space for the RNA molecule is even broader, due to structural and thermodynamic differences with DNA. This enzyme-free reaction network approach is remarkably versatile and has become predominant in the field in comparison to other architectures like those proposed in the foundational work by Alderman or the enzyme-based approaches because of the aforementioned versatility and the high degree of control available over all the components in the system.

However, the real power of DSD systems lies on their scalability and integration into large reaction networks. Using cascades of DSD reactions, basic Boolean logic gate functions can be implemented. Furthermore, scaling up these networks to more extensive systems via sequential exposure of different toeholds and with some tweaks such as the seesaw motif (a strand with a double toehold that allows to implement catalytic inputs via fuel consumption in DSD systems)[55] displacement has been used to demonstrate complex devices such as a calculator able to compute molecularly the square root of an integer number with 4 decimals[56] or a neural network capable of pattern recognition[57]. For all these reasons and due to the fact that DSD reaction networks can form the basis for the construction of programmable and direct interfaces with and within living organisms, DSD-based computational systems are a field of growing interest. The development of DSD reaction networks interfaces is a field of interest for intelligent drug delivery systems[58] or, when translated into RNA circuits, for the creation of modular and orthogonal control systems for synthetic biology circuits[59] that impose a lesser metabolic burden to host organism than heterologous protein-based control networks[60]. It can be argued that the algorithmic assembly systems discussed in the previous section are also suitable for the tasks required for such applications. However, DSD systems have functional subunits that are much easier to build, program and scale up than any other programmable molecular assembly method. These characteristics justify the choice of designs based DSD reaction networks.

The aforementioned applications are characterized by the fact that all these systems require embedding of different types of chemical species in the same system in order to perform the computation. Hence, the proper exploitation of DSD networks requires a formal framework able to conceptualize a given algorithm as a set of chemical reactions. During the years following the experimental development of DSD networks, theoretical computer scientists like Erik Winfree, Luca Cardelli and David Soloveichik applied a pre-existing theoretical framework known as **chemical reaction networks (CRNs)** theory and mapped experimental DSD networks into CRNs.

Within this framework any possible theoretical chemical reaction could be designed and their behaviour predicted by using the dynamics and the result of the set of reactions as the function we want to implement. CRNs have become the core framework for many

important designs of biomolecular controllers in the field of control engineering applied to synthetic biology. CRNs are theoretical constructs, and as such, they are formal representations of a desired behaviour in a physical system. This implies that the experimental realisation of a CRN can be done on a series of different substrates and a considerable number of CRNs have been translated into physical protein systems. However, for all the reasons cited previously regarding the control of physical parameters as well as ease of design, nucleic acids can easily be considered the best candidates for the realisation of a given theoretical CRN. Despite the advantages, there are still issues posed by current DSD implementations as detailed hereafter.

The problems that DSD faces when used as means to realise a given CRN can, in principle, be divided into two categories: those related to the actual realisation of the CRN and those related to the performance of the experimental system. The realisation problems of a theoretical CRN are related to the fact that a same given CRN can be approximated -if the dynamics and concentrations of the physical system allow to do so and stochastic effects are negligible[61] with a different set of species that contain at least the core elements of the theoretical CRN and their formal interactions, but not necessarily sharing the same physical interactions[59] This means that a given CRN might need intermediate reactions just to be formally realized – hence resulting in possible unexpected problems related to the ancillary experimental reactions that were not captured by the CRN formalism. Deciding on the best realisation is not a trivial task and many efforts in the last years have been directed towards developing verification methods that allow minimization of the number of side reactions[62,63].

Performance problems for the experimental realisations of CRN begin with the fact that CRN theory is typically focused on deterministic behaviours while biomolecular systems are inherently stochastic. This means that the approximation of the DSD network via CRN theory will only hold while working in bulk with a very large number of molecules. But even when the system can be treated as a deterministic one, it does not eliminate the chance of performance troubles. Complex algorithms usually require a large set of different chemical species (i.e. the square root calculator[56] required more than 130 different DNA species) which can result in a higher risk of unwanted leak reactions or a higher burden if encoded and expressed in a living cell. Finally, an operational issue that must be addressed is that, as for the algorithmic self-assembled systems described before, most DSD experimental implementations of a given CRN are designed to be driven to a thermodynamic equilibrium state when executed, thus resulting in the inability to be reconfigured, rebooted or operated continuously without external inputs[64,65,] except for certain given circuits[66].

Of all these troubles, only the rebooting of the circuits has been addressed recently with time renewable circuits as those cited in Reference 65, although this has only been demonstrated for single gate elements rather than for large scale, complex circuits. But it must be noted that this regeneration steps require external action on the device. This means that are not capable of autonomously adapt to the changes in the environment.

However, when observing, natural, information-processing biomolecular circuits in transduction networks, one can witness an outstanding adaptation and reconfiguration capability as well the capability of performing complex tasks like sensing and inferencing. This seems to indicate that natural systems are defined by a series of characteristics that DSD systems currently do not properly capture. Hence, the final section of this present chapter will give an overview of the desired features that define these systems.

1.3.- Biological transduction networks and circuits: from biology to kleptobiology.

Cells are remarkable when studied as engineering systems. Feats that are essential for the cell survival like detecting and moving towards increasing concentration gradient of nutrients as well as regulating gene expression according to environmental changes are tasks that require the integration of several simultaneous inputs as well as the activation of actuation systems in real time[67]. Living systems are capable of performing such tasks reliably in an incredibly noisy environment, which can be seen as an outstanding feat by itself. These information processing capabilities are directly tied to particular properties of the biophysics of living systems.

At a first glance, it could seem that cellular signal processing systems operate in a way that is totally alien to how man-made systems work. However, the lack of similitude becomes only an appearance after some closer examination. Instead of relying on silicon circuits, signal transduction networks in living systems are formed by catalysts (normally kinases and phosphatases) that are able to act as switches with multiple states over a given substrate that normally is also an enzyme, thus allowing to the building of layered catalytic systems as seen in Figure 14.

In this layered architecture, catalysts have opposing effects over the activation state of a given substrate and information can be encoded in the relative concentrations of the catalysts present in the medium or their variations[68]. Moreover, the fact that the network system is catalytic in nature confers the system some interesting features. This way, since catalysts are not consumed and can be kept in a given state via fuel consumption, the network can retain the information associated in case no perturbation is made effective upon the system. However, with varying external inputs in the network via sensing (as seen in Figure 14), differences in catalysts arise leading to a defined distinct cellular response that activates or deactivates a given cellular process, continuously responding to these external variations.

It does not escape our attention that the use of terms as layers in a network system as well as a working mechanism that entails summation of different weights with variable opposing values in a converging point of a network is tremendously familiar to those who work with artificial neural networks in silico. And certainly, should be, since artificial neural networks do really capture many of the features on how signal processing works in cells and hence, begin to yield enough understanding to be able to replicate it.

This analogy between signal transduction networks and artificial neural networks might seem fashionable given how ubiquitous the formalism has become with the boom in machine learning methods and applications for practically any conceivable research field. However, the idea is quite an established one that has been reported at least 25 years ago by Dennis Bray[69] and it's a mapping that is still being explored[70]. It must be noted that since artificial neural networks are only a model there are some non-trivial features of the biological transduction networks –particularly those related to their catalytic nature- are not captured properly. However, this idea holds much more heuristic power than could be thought initially.

It is precisely by establishing working parallelisms (both enzymatic and artificial networks operate with layered circuits and analogic signal processing) that we can see with greater clarity that signal transduction networks do possess a functional fundamental unit akin to the perceptron[71] in an artificial neural networks: the **push-pull network motif** (also known in other sources of bibliography as the **catalytic pair motif**).

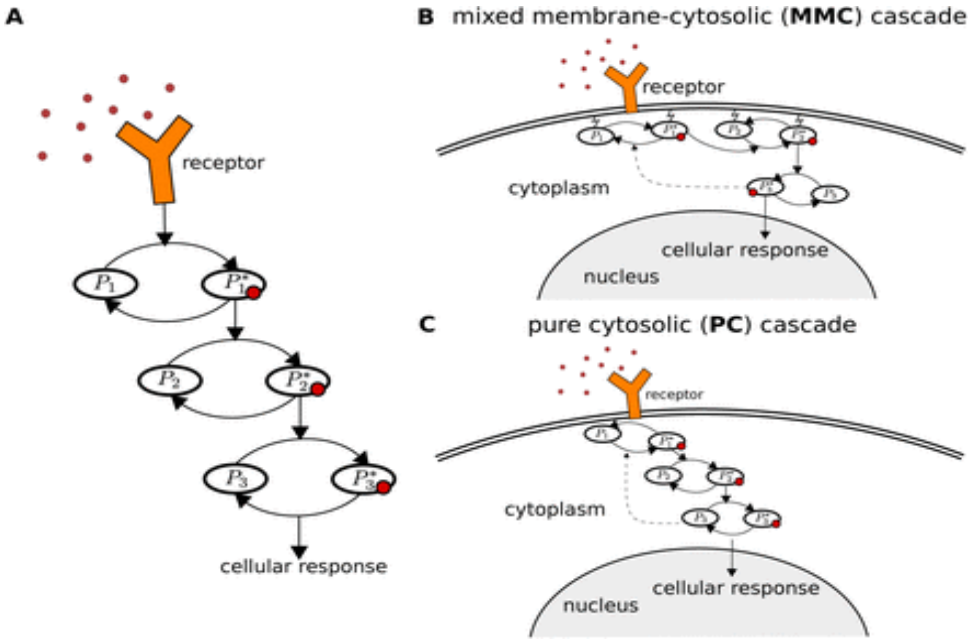

Figure 14. (A) Schematic figure a general transduction cascade. (B) Schematic figure of a mixed membrane-cytosolic. (C) (A) Schematic figure of a purely cytosolic transduction cascade. Adapted from Giese et al (2018)[72].

1.3.1- Push-pull networks: description and fundamental properties of the motif

As specified in the previous section the push-pull network motif is considered the fundamental motif of signal transduction inside cells[73]. This motif is comprised by a substrate protein that can be switched between, at least, two states corresponding to an active and an inactive state as we can see in Figure 15.

These activation and inactivation processes do not occur spontaneously but rather is mediated by the action of two opposing catalysts that act on the substrate. Both catalysts act on the same molecule and in a cycle that, depending on the ratios of the opposing catalysts, will activate more or less substrate and the amount of active substrate will vary as the ratios

between the catalysts do in real time. This active substrate will then be able to act on a different substrate in another catalytic layer. This basic structure suggests that the aforementioned analogy between push-pulls and artificial neurons is indeed much more direct than the previous subsection of this chapter suggested. But the interest that push-pull networks have as systems goes beyond the similarities they might have with any man-made computation architecture.

Push-pull networks are considered to be the fundamental unit of biochemical transduction networks. And as such, they have been studied and characterized extensively. This focus, in turn, has yielded a very deep understanding on the functionalities of catalytic pairs and how these relate to the fact that the operating units are catalysts. Hence, we have observed that the catalytic nature of the motif allows it to amplify the signal downstream[74] the substrate via cascading, split the signal activating more than one substrate or performing time-integration of the signal through cascading action conferring the system extra robustness to stochasticity.

The catalytic nature of its constituent elements not only allows the motif to perform all the functions previously detailed, but also allows it to act as a regulating insulator motif[75] between different circuit modules that avoids noise propagation and retroactivity. These functions can be performed by a single catalytic pair, but more interesting properties arise when these network motifs organize themselves in more extended networks and cascades.

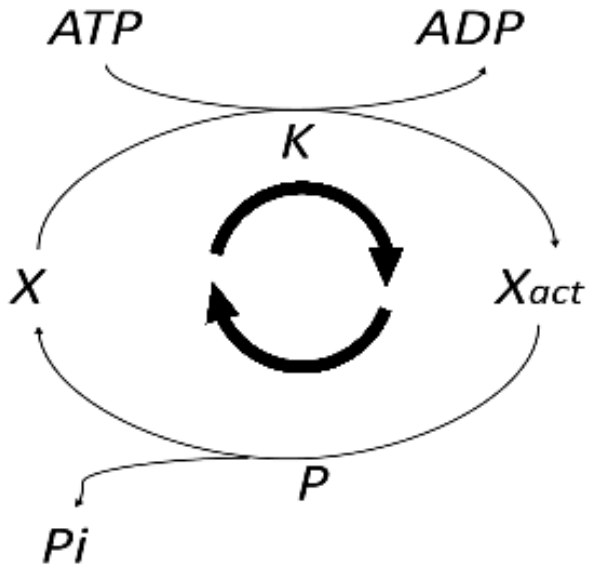

Figure 15. Schematic figure of the basic push-pull motif. On it it is featured the substrate X (for simplification it is assumed that only has two activation states) and two catalysts acting on it : a kinase (K) that drive the substrate to its active state consuming ATP as a chemical fuel involving a process of phosphorilation and a Phosphatatase (P) that deactivates the substrate dephosporilating it. In the motif the two reactions compete one against each other and the outcome is determined mostly by the ratios between catalysts.

As we have seen before with the different experimental implementations of chemical reaction networks, the physical substrate upon which a formal computing paradigm is implemented matters to its performance and potential exploitable features. With respect to

what enzymatic transduction cascades add to network computing implementations there is more than meets the eye.

In push-pull networks substrates do not necessarily limit themselves to two active/inactive states. Proteins in nature are allosteric[76], meaning that they can have several activation sites and some of them are only accessible to their target enzymes once a previous modification has been added. This property adds more complexity to the circuits that can be built with proteins in two ways.

The first extra layer of complexity that allosterism provides to push-pull networks is that the necessity of a previous input in order to have activation of a given type implies that not all inputs are equal to the substrate and hierarchies of inputs can be established, resulting in more complex behaviours.

The second way in which allosterism adds extra complexity lies in the fact that different activation states result in different reaction kinetics. This change on the kinetics can happen very fast and can change the reaction activation profile very quickly, switching from an analogic-like signal to a step-like digital answer. This implies that depending on the conditions one type of signal processing can be favoured over another or they can be combined to exploit the different strengths of both types of computing simultaneously in the same network.

This ability to exploit simultaneously both analogic and digital computations has been proposed as one of the unique properties of living systems as computational substrates that could result in *cellular supremacy* – this is, making them potentially more suitable for certain types of algorithms. This heterotic analogic/digital computing is not the only factor that has been proposed as fundamental for *cellular supremacy*[77] – others like concurrency, task distribution and high parallelisation have been proposed – but it is fascinating in the fact that the very architecture of transduction networks seems to allow this transition from analogic to digital computing through various mechanisms.

It has been proposed theoretically that when the catalysts of a push-pull network are working in saturation, small oscillations in the amount of catalyst available could yield this type of response[78]. However, without saturation or allosterism, a step-like activation of a substrate can be achieved if its activating catalysts is upregulated by another activating catalyst[76].

1.3.2.- Extended push-pull-based networks: interest and non-trivial properties.

As we have seen before, push-pull networks are characterized by an outstanding performance in terms of information processing. With an apparently simple architecture this motif can perform a great variety of tasks related with the process of external signals inside the cell. However, as we had stated in the previous subsection, push-pull networks are the basic functional subunit of biological transduction networks. Again, the mappings between our biochemical network motif and an idealized artificial network become relevant due to the

fact that their real power and versatility arises when both neurons and push-pull organize in more extended networks.

Hence, in nature we can observe how push pull networks can extend themselves in two different ways:

- **Horizontally:** In horizontal network extensions we can observe either that the protein substrate of the network is activated by more than one catalyst in the same direction or the given catalytic pair acting simultaneously in more than one target protein in the same layer. More elements are added but all are in the same layer and subsequently, they keep their roles as substrates or catalysts.

- **Vertically**: In vertical extension (or cascading) creating cascades of push-pull networks in which the substrate of a push-pull acts as a catalyst on a subsequent layer.

This way, through the integration of the fundamental motif in extended networks, the signal processing network can gain extra features. The first and most obvious possibility lies in the capability of integrating several different inputs in the same substrate using different catalysts (which is known as **signal joining**). In addition to this we also have the possibility of acting on different targets in the same layer simultaneously which is called **signal splitting**. Signal splitting allows the network to perform a synchronized multiple response to a given input, since signal branching allows a catalyst molecule to act on different targets and then cascading can lead to each of these targets to act on different subcircuits and get the signal amplified on each layer.

Another way with which extended networks improve performance of their single push-pulls is by the way of adding **functional redundancy** that provides extra robustness to the system in case any of the acting catalysts gets inhibited or compromised. This redundancy can be obtained by the signal joining mechanism we have described previously having two different catalysts pushing the same substrate in the same direction in the same catalytic layer. This means that a target molecule can be switched by many different catalysts on the same layer. In addition to the extra robustness, functional redundancy can affect the activation speed of a given substrate: in principle, having more net catalyst pushing in one direction should accelerate the process. However, increasing indefinitely the amount of catalyst will not increase indefinitely the reaction speed due to the limited availability of the substrate to this catalyst excess.

Additionally, the architecture in cascade of the push-pull network-based systems has been demonstrated to have exceptional properties with regards its robustness: theoretical studies have shown that signal transduction cascades work near the Wiener-Kolmogorov limit acting not only as non-lineal signal amplifiers, but also as noise suppressors, being able to perform both tasks[74] at the same time.

This can be better seen imagining a situation in which a great punctual signal perturbation (or environmental noise) happens in the higher end of a cascade. If the cascade consisted of a single catalyst element acting on an on/off target, that punctual perturbation

would have been transmitted into the substrate and that substrate would have been activated resulting, for example, in an accidental activation of a gene that should have not been activated.

However, if we put an opposing catalyst element to the same target and we put another push-pull downstream the situation changes drastically. If the amount of inactivating catalysts of the first push-pull element of the cascade outnumbers drastically the amount of activating ones affected by the perturbation, this punctual perturbation will not be transmitted to the following step, thus providing implicitly a thresholding function to the system. However, if the perturbation is stronger and much more prolonged in time, due to an input to which the cell should respond, eventually the activated substrate of the first layer will be able to act as a catalyst on the second one amplifying the signal non-linearly with respect to the amount of activated substrate.

In addition to all the advantages mentioned before, it must be noted that theoretically the organisation of the push-pull networks in cascades would also allow them to implement oscillatory behaviours[79] when implementing a feedback loop motif. However, these oscillatory behaviours are rarely observed inside signal transduction cascades and it has been discussed that their absence even when such loops are present allows the cascades to maintain their robustness to temporal signal delays and that they achieve through loading effects on the catalysts[80].

1.3.3.-Push-Pulls as non-trivial out-of-equilibrium systems.

The last notion that explains the outstanding performance of living systems is the fact that living systems are out-of-equilibrium systems. Cells are open systems that can freely exchange matter and energy with the environment. This feature, combined with the production of entropy in the surroundings, makes it possible for cells to create complex circuits with proteins. Such cellular circuits can transmit and process information performing functions such as information copying and data integration from the surroundings. These circuits are sustained in time in a quasi-steady state through the consumption of energy via chemical fuels[81].

The fact that the state of a circuit given certain inputs is not an equilibrium one and is maintained by consumption of an energetic fuel might look trivial, but it allows the circuit to implement transient responses only active as long as there is fuel available to the circuit. Again, since the active input-output configuration of the circuit is out of equilibrium, it means that, as we described at the beginning of the third section of this chapter, the space of chemical species reacting and interacting is not closed and at any given moment another interacting species can transiently come and alter the input-output configuration of the circuit in real time. This alteration of the circuit configuration can happen in either direction, either vertically inside a given cascade adding new layers or in terms of introducing new downstream targets in another layer that might activate a parallel circuit. This property is responsible for the great malleability and adaptability of biochemical circuits.

In this sense, push-pull networks constitute one of the simplest systems in which the out-of-equilibrium features that define living systems are present and hence, the interest to understand them and which are the fundamental limits that define them. Among these properties and features, the capability of producing a copy of a given state into another biomolecular is one of the most interesting ones. Information transfer between different substrates is one of the defining features of life. In fact, the central dogma of molecular biology mentioned previously in the first section of this chapter describes precisely how information flows from its original storage substrate (DNA) into either another storage copy or functional substrates. Both cases are examples of out-of-equilibrium processes since they act autonomously and require an energy cost to be fully functional[82].

As we have said before, two-step push-pull cascades are a limit case of autonomous information copying[83] and so far they have demonstrated theoretically to be extremely energetically efficient (albeit still far from working at Landauer's limit[84]). Determining whether they can operate closer to this limit without compromising any of the functionalities is a question that remains open to discussion today in the field of thermodynamics. It must be noted that the relevance of the insight provided by the study of push-pull networks such is not limited to biomolecular systems but rather to the whole field of computational thermodynamics. As stated in the previous paragraph push-pull cascades are a minimal copier that works using an extremely low energy expense. But the catalysts employed by the cells in order to execute the task work use chemical fuels in a physical irreversible process. In theory is possible to achieve reliable logical irreversible computation without the need for a physical irreversible process[85].

Biomolecular systems - DNA nanotechnology-based most precisely- are an extremely powerful tool to explore the plausibility and trade-offs of these fundamental assumptions[86] since they allow us to design systems in which the free energy landscapes are programmable and completely predictable. In our present case, the main feature that we search to implement in our reaction system is making the enthalpic free energy difference between the two sides of the reactions that constitute our system as low as possible without compromising the system's functionality.

1.4- Thesis overview

In summary, the aim of the present thesis is to explore and determine whether these interesting properties of natural transduction systems (mainly the architecture based on push-pull network motifs that work out of thermodynamic equilibrium and can be organised in multiple layers with flexibility in their response) can be transplanted into DNA strand displacement systems and exploited to create more robust biomolecular computing systems. The motivation on building this stems from two different motivations. The first motivation lies in the need to explore the fundamental limitations that copying in computing devices face no matter which type of substrate are they implemented, either silicon on nucleotides chains. Then the second objective of the project is building a framework that can bring all the features

of natural, biomolecular out-of-equilibrium catalyst-based networks into man-made DNA nanotechnology-based systems.

For such endeavour, we will proceed to introduce the basic framework of our project in the second chapter of this present thesis: the **Active Circuits of Duplex Catalysts framework** (abridged as **ACDC Framework**). This chapter will delve into discussions about how current catalytic DNA systems fail to capture wholly the features that define catalysts in natural systems and how our system design choices account for these features as well as for ensuring the orthogonality of every potential catalyst species in the system.

The third chapter of the thesis will give the experimental evidence for the step-by step realisation of the fundamental push-pull network that defines biochemical transduction network. From the reversible single step reactions to the catalytic reaction to finally, the full assembly of the out-of-equilibrium network motif. This chapter will bring on the discussion on some of the limits that these systems seem to face and how do they relate to the performance in real-life systems.

Chapter 4 will deal with the construction of extended networks, demonstrating that the ACDC Framework allows us to put different target species for a given set of catalysts as well as implementing cascading systems. However, we will discuss how despite in principle being able to fully realize extended networks with the aforementioned framework, how do practical limitations arise.

Finally, Chapter 5 will explore the implementation of hidden thermodynamic drives in the systems as means to overcome troubles faced in previous episodes.

Chapter 2: Active Circuits of Duplex Catalysts (ACDC) Framework: Design considerations and feasibility.

2.1.- Motivation of design choices.

The objective of our work is to design and test DNA reaction motifs that improve present CRN implementations. We do this by drawing inspiration from natural biomolecular information processing networks. In order to reach this objective, there are four distinct properties which we have considered as the most relevant with regards to CRN implementation and functionality.

The first property we want to implement in our network motifs is to make these motifs continuously responsive in time. This means that activation and deactivation processes in our networks must be reversible and should occur in real time. This time-responsiveness property requires our network motifs to work out of thermodynamic equilibrium instead of being one-shot implementations that become inert once they reach equilibrium.

The second property we want to implement to make motifs that can be integrated into larger composable networks. This allows us to implement complex functions and behaviours using basic network motifs.

The third property is to make all the major functional elements of the motif catalytic. Hence, we will be using the term Catalytic Control Network – abridged CCN – to refer to the networks we want to build with our system. The main interest in the use of catalytic elements lies in the fact that these catalytic elements do not get used or consumed during the information processing allows to create out-of-equilibrium systems with the desired properties previously described through the consumption of chemical fuels while preserving the information they have embedded.

The first interesting informational property that we must highlight is that catalytic elements can maintain a given state via fuel consumption. However, since the state of interest is an out-of-equilibrium one, this means that an opposing catalyst can act in the opposite direction against the initial one, thus enabling the reversibility property we first listed.

A rough schematic diagram that summarizes the desired properties for our CRN DNA implementation is featured in Figure 15. In this figure, we can see an element X being activated and deactivated by other two catalytic elements while X itself is able to interact with and activate another layer of catalytic elements.

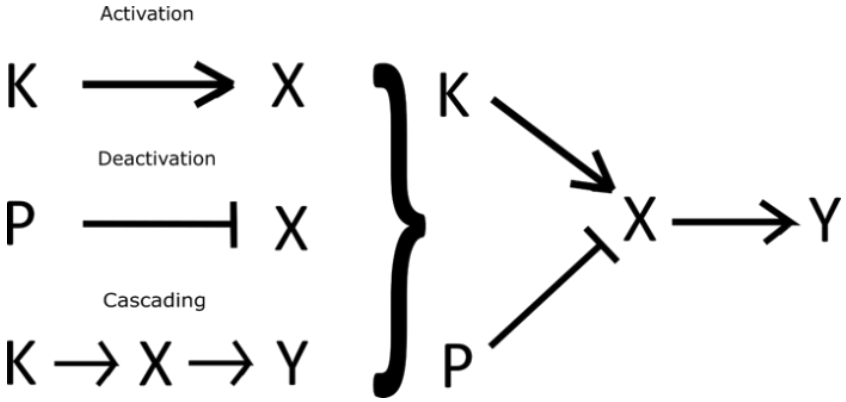

Figure 15. Formal summary of the desired reaction functionalities of our proposed system.

The last property we are looking to implement is to enable spatial localisation of the network elements of our system. Again, this is another property observed in natural catalytic networks. In these networks, the different elements that are grouped in scaffolds that create functional modules in which diffusion effects are minimized speeding reaction rates and lowering unwanted side activation reactions. However, scaffolding of catalytic elements also has regulatory roles in type of response of transduction networks and its precise role is currently under discussion[87-89]. Hence having the possibility of implementing such spatial localisation in a simpler DNA-based system can help us answer some of these questions.

While none of these properties are totally alien to DNA nanotechnology, not all the possible physical implementations of CRNs allow all of them simultaneously. Moreover, the third desired property for our reaction scheme (making all the major elements of the network catalytic in nature in order to add improved information processing properties) is an even less trivial task that requires a deep discussion in order to determine the adequacy of reaction geometries for this property.

2.1.1-Problems with current implementations of catalytic motifs in DSD: catalysis vs pseudo-catalysis

By definition, catalysts are molecular element that when present, – normally in small concentrations – interact physically directly with the reactants of a given chemical reaction to speed up the reaction by providing a reaction pathway with a lower activation energy barrier[90] without being consumed in the process. Catalysts, despite their action, do not alter the equilibrium point of the reaction upon which they act. Their effect is limited to the kinetics of the reaction and they are recycled during the chemical reaction process.

Life is heavily dependent upon catalytic processes. Enzymes, the catalysts upon which the whole of biochemistry relies, are incredibly complex molecules that allow biochemical reactions to be feasible in a timescale compatible with life. Without enzymes, the biochemical reactions at the core of the metabolism of living cells would be either thermodynamically unfavoured or, if they still tended to occur, they would not do so not in timescales compatible with cellular operations.

The role of enzymes in metabolism is not limited to enable metabolic processes but also includes the spatial and temporal regulation processes in the different contexts in which living systems thrive and operate, since the transduction networks that control the cell response are also formed by enzymes. These network structures are the inspiration from which we want to draw for our desired nucleic acid systems. and many of the features that make this inspiration such a remarkable paradigm in information processing (possibility of complex dynamics, time-integration of signal, strong memory correlations between inputs and outputs) are directly derived from the catalytic nature of their action.

Implementations of catalytic motifs in nucleic acid strand displacement motifs are not a novelty in the field. Back in 2011, Winfree and Qian developed the seesaw gate motif[55] showing a powerful example of a catalytic reaction motif which proved it could be extended into larger networks by building a 4-bit square root calculator and a 4-neuron neural network based on this seesaw motif.

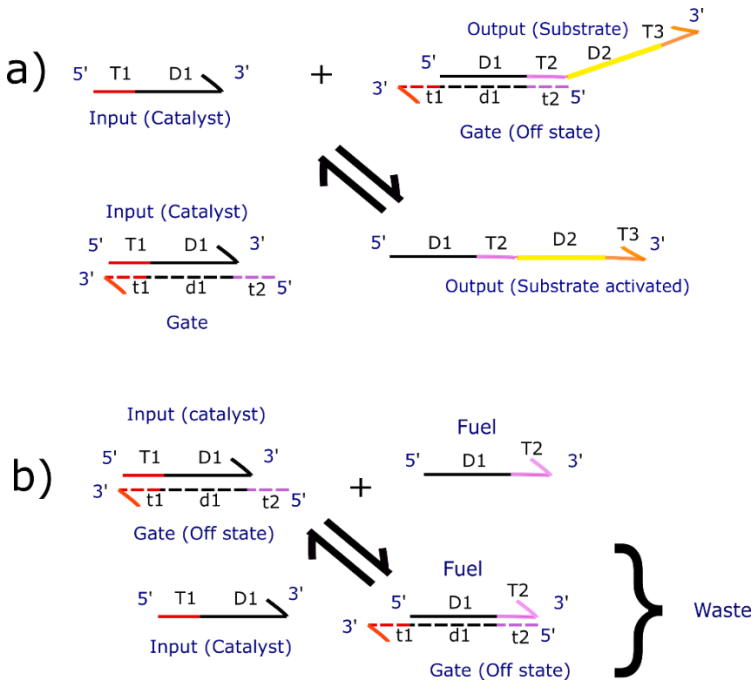

Figure 16. Implementation of the CRN of the catalytic activation in a seesaw gate. Step a) features the activation step of substrate X whereas step b) features the regeneration of the catalytic input via fuel consumption.

In the seesaw gate motif depicted in the Figure 16, the catalytic element is the input of the system. This single-stranded DNA input triggers a strand displacement reaction with the Gate-Output complex via toehold recognition. This Gate-Output complex is formally equivalent to a switched-off substrate X and releases a single stranded DNA that plays the role of the active form of the substrate X. Later on, this gate-input complex interacts with a single-stranded fuel DNA that kicks out the input, thus generating a Gate-Fuel complex (formally equivalent to a waste molecule). As we have seen, this reaction motif can be extended.

However, this approach has issues that limit its adequacy as a basis for implementing our desired CCNs which will be described below.

While catalytic activation is feasible in seesaw gates, the deactivation reaction faces implementation problems. In this case, the activity of the motif is catalytic and we can perform catalytic activation of a substrate. However, implementing the deactivation in this setup entails binding back the Output to the Gate. This reaction is the exact opposite of the activation, but realizing this operation involves a role exchange of the Fuel and the Output molecule from the original reaction. Moreover, the very architecture of the seesaw gate entails that there is only one possible catalyst per each substrate and that the role it plays depends only on the concentrations of free Fuel and Output available. This circumstance violates the condition we initially imposed in our design about making deactivation of an element through the action of a catalyst different to the one that activates the substrate. Hence it becomes apparent that seesaw gate motif functionality falls short for all the purposes we want to implement in our reaction scheme.

Due to these limitations, the search for alternatives to the basic reaction scheme becomes mandatory. For example, DNA reactions mediated by multi-stranded gate polymers[91,92] reaction geometries designed independently by Erik Winfree, Lulu Qian, David Soloveichik on one side and Luca Cardelli on the other could be an alternative. Cardelli's implementation scheme is illustrated in the Figure 17. This reaction geometry allows the implementation of any arbitrary CRN. Theoretically, the implementation could be as easy as defining a trimolecular reaction in which, a catalyst species C is defined both in the set of input and output species of the reaction while the different states of X as well as the chemical fuel and the waste are defined in reactants and products. The result is the CRN pictured below in (1) with the implementation displayed in Figure 17.

$$X + C + F \rightarrow Xact + C + W \quad (1).$$

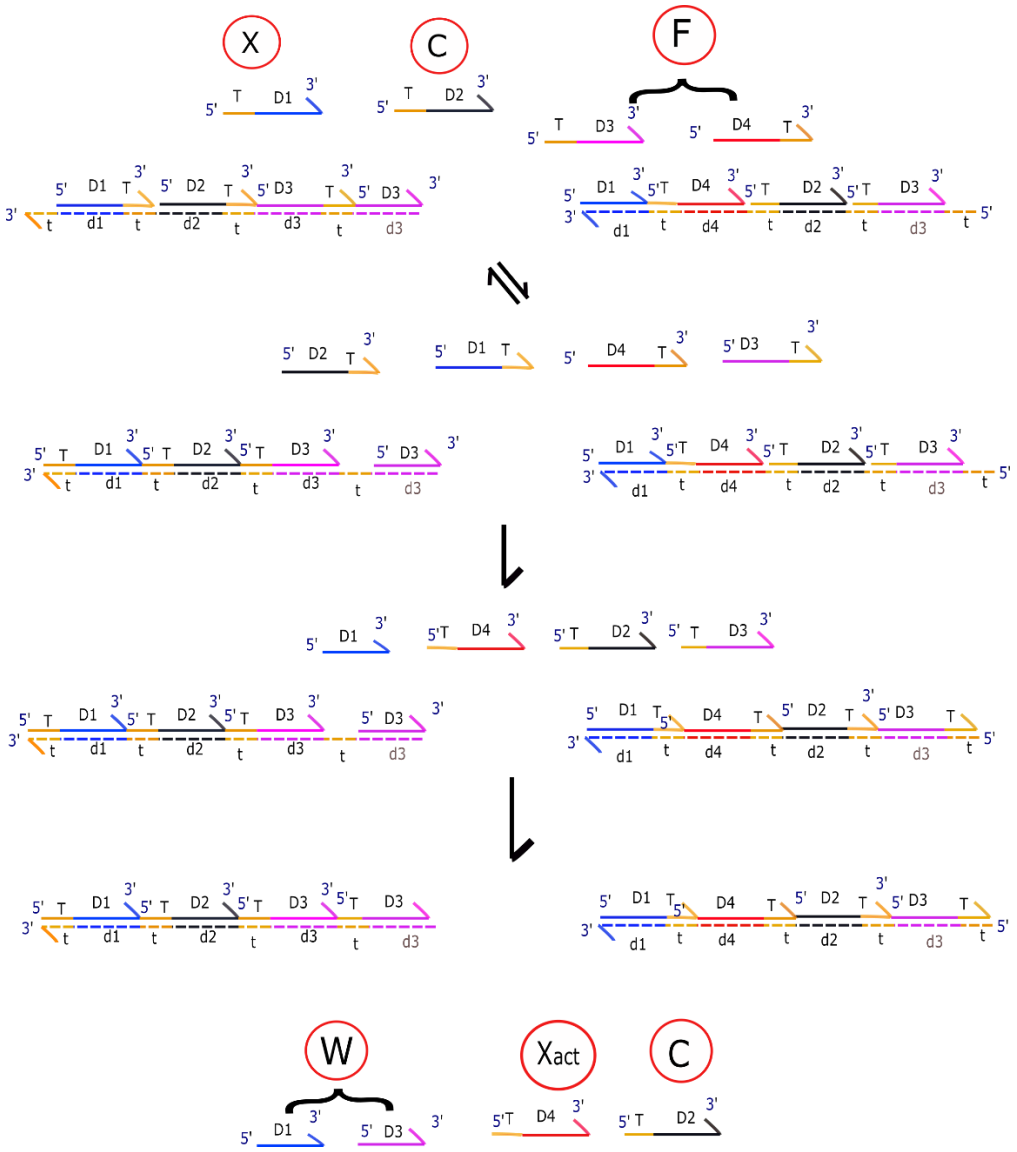

Figure 17. Implementation of the CRN of the catalytic turnover X + F +C → X_{act} + W + C in a setup based on the use of multi-stranded gate polymers (specifically, the design created by Luca Cardelli for Cardelli (2010)[90]). The implementation is a pseudo-catalytic since there is no real recycling of the catalyst C in the setup, and requires multi-stranded complexes to be prepared beforehand.

Although this reaction geometry apparently allows us to implement the catalytic turnover reaction and being Turing complete should allow us to implement the deactivation reaction, both Cardelli's and Winfree-Qian-Soloveichik implementations still present some issues that are common to both and make them be far from the most desirable implementation of our desired CRN.

First, this implementation is a complex implementation of our theoretical chemical reaction network as it requires ancillary species such as the multi-stranded complex gate and the intermediate single-stranded species attached to it to work. The complexity of these multistranded complexes might be problematic in terms of modelling and predicting its behaviour as well as for assemble the complexes experimentally. Since complex gates are

difficult to anneal in the correct stoichiometry, more prone to potential leak reactions than simpler architectures and more difficult to spatially localize.

The second problem that this implementation poses to the functionality of our desired system is that its working is pseudo-catalytic. Instead of physically recycling the same strand of the initial catalytic cycle, the catalyst C is consumed in the first polymer and then released in the second one, thus making the core of the catalyst not physically conserved along the process. In principle, this mechanism would work in bulk solution without major problems. However, if we want to run the reaction in a spatially-localized configuration, similar to signal transduction network in cells, the catalyst strands that are initially bound to the scaffold would end up being used. Then, the reaction will stop after consuming all the localized catalysts species.

Thirdly, in this implementation, it is not the pseudo-catalyst the species that reacts directly with all the species of the chemical reaction network. All direct physical interactions are mediated by the scaffold strand of the polymer in the ancillary gate. This mechanism can lead to the effective sequestration of the substrate or catalysts in case one of the complex gates or the ancillary species is not in the same concentration with the one acting as an input. Such sequestration can be attenuated through the use of very high concentrations of the ancillary strands present in the gates acting as a reaction buffer. But this would slow down the reaction due to the competition between the reaction of our major species with the first gate and the reaction of the ancillary species with the gate.

Hence, summarizing this section, all present implementations of a catalytic turnover reaction in strand displacement systems present different problems by design. Seesaw gates offer the possibility of implementing real catalytic action. However, each defined substrate can only have a single catalyst defined for itself which makes deactivation reactions essentially equal to activation reactions with switched roles of fuels and substrates. Complex polymer implementations are pseudo-catalytic, excessively complex and, as a consequence of these two problems, cannot be properly localized spatially and are prone to sequestration issues. Hence, in order to make our purposed reaction scheme, we need a novel reaction geometry that allows us to solve these problems.

2.2- Introduction to our framework: Active Circuits of Duplex Catalysts Framework (ACDC Framework).

As we have seen in the previous section, current catalytic motifs are prone to multiple faults in their design that can limit their performance and their suitability to build CCNs. Hence examining the limitations we currently face, we can find ways of overcoming them in the simplest, most efficient way.

In all previous designs, we observed that interactions between the different species were mediated by another strand that indicated that the single stranded species they were bound at the moment – whether it was the substrate, the catalyst of the fuel - were inactive. In order to tackle these problems associated with the previous, we must find a reaction

geometry that allows us to keep a functional molecular core conserved throughout the process rather than resort to use dual-rail logic to emulate the catalytic reaction.

Moreover, the need for binding between the catalysts and substrates implies that the activation state of the substrate cannot be encoded by a bound/unbound state of the substrate. Hence, the second feature we need to implement in our system is that all the possible states of a given species must be explicitly declared and encoded somewhere else rather the chemical species themselves.

This means that the active or inactive state of a given chemical species cannot represented by a particular strand (the null or inactive state must exist) and that the state of the species must be encoded in a strand of its own while keeping the identity core of the interacting species coded in another strand.

The need for states explicitly declared in separate strands imposes a new constraint in the system: given that the identity of the chemical species and the activation or binding state are coded in different strands, our reacting chemical species must be duplexes for all the types of species of our system. This use of duplexes does not only allow us to declare explicitly the state of each species of the system, but also to place the function in strands that do not define the identity of the chemical species. This way, we can avoid the functional overlap between fuels and substrates that appeared in the seesaw gate implementation or in the faulty polymer implementation. It must be noted that the idea of multiple states of a same variable encoded in a strand is not completely alien to DSD reaction schemes, since many approximated implementations of digital computing in DSD rely on the concept of dual-rail logic [93] in which the null values of variables is encoded not by the absence of a given species, but by the use of an inert chemical species. However, whereas regular implementations of dual-rail logic still implemented the state and the identity in the same strands, our proposed duplex implementation performs this logic by encoding the state and the identity in different strands.

We have new constraints and conditions that arise in the duplex scheme. Both catalysts and substrates must be duplexes and the interaction between both types of species must be direct. Hence, regular TMSD must be discarded as some of the interacting species are single-stranded. Fortunately, such mechanism exists in the shape of the 4-way branch migration in which the formation of a 4-stranded intermediate triggers a strand exchange process between the reacting duplexes.

Such a mechanism is inspired by the Holiday Junctions used by eukaryotic cells for chromosomic recombination. And again, like the case of dual-rail logic, the use of 4-way junctions to implement reaction mechanisms is not an alien topic for DNA nanotechnology. Previous theoretical work by Winfree and Qian[92] depicted in Figure 18 already pointed to the feasibility of using DNA duplex species for DNA computing. The use of duplexes and associative toeholds for spatially localized reaction systems has remained a source of inspiration for several molecular programmable systems of all sorts [93-97].

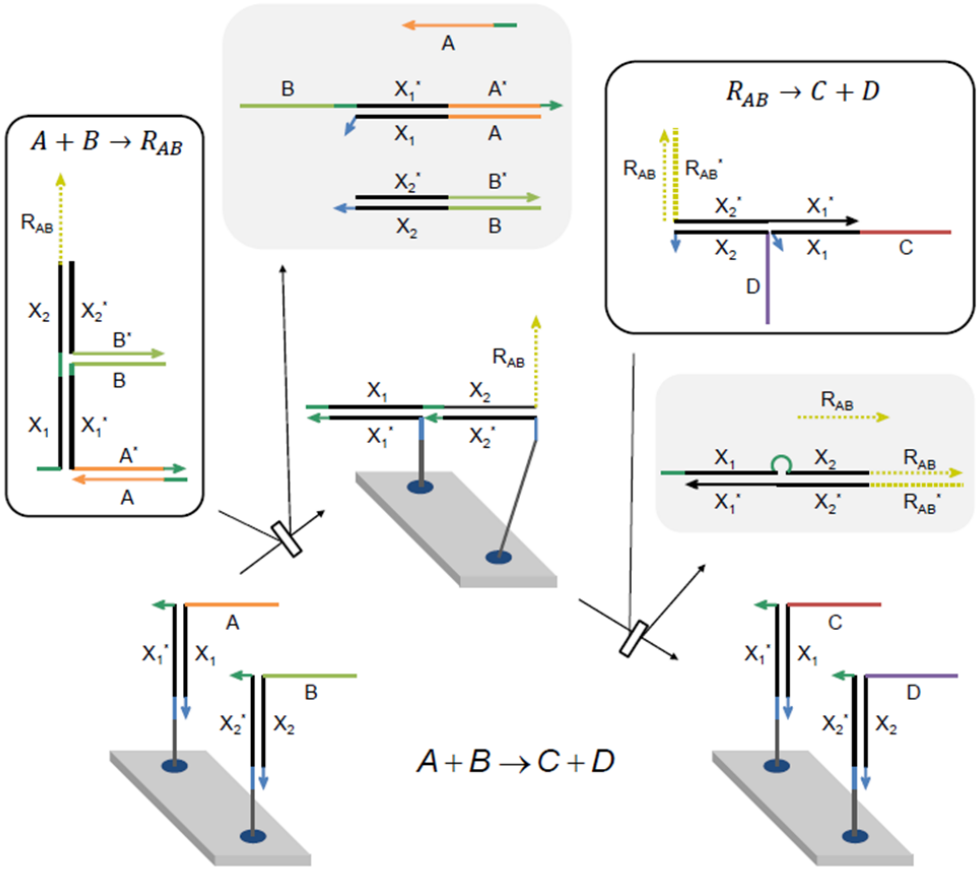

Figure 18. Implementation of an A + B→ C +D CNR in a spatially localized duplex-based 4-way branch migration system. Adapted from Winfree, Qian (2010)[92].

Even without the spatial localisation feature, the use of 4-way branch migration schemes has been of high interest for the DNA nanotechnology community. This mechanism has been studied in the past[98] and used in practical applications such as directed hierarchical assembly,[97] as well as in implementing low leakage catalytic reaction motifs[98] in which there is real contact between the catalyst and the substrate species.

However, it must be noted that the motifs described in the work of Hughes and Kotani[98] and depicted in figure 18 are not suitable to the construction of out-of-equilibrium systems because its reaction products are inert duplexes that are in equilibrium. Moreover, the catalysts of this formalism are single-stranded. This design feature limits the number of possible catalysts that can trigger a given conversion to only one, limiting the network design space in a similar manner to what we have seen with the seesaw gate. In addition to this circumstance, the species employed in this system can have different number of arms, adding unwanted complexity to the system.

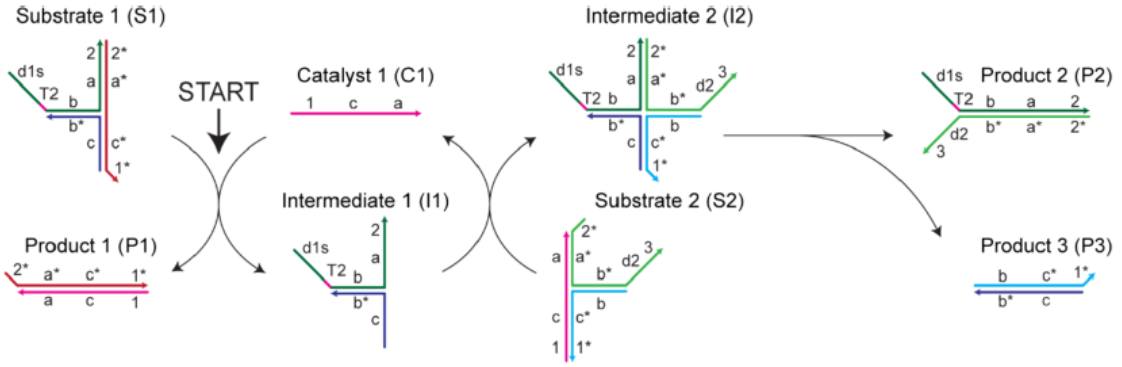

Figure 19. Implementation of a catalytic turnover reaction based in multiarmed complexes. Adapted from Hughes,Kotani (2017)[98].

The possibility of having more than one catalyst acting on the same substrate is mandatory to create the out-of-equilibrium, switchable multistate systems we intend to design for CCNs. Without it, the push-pull motif, which we called in our previous chapter as the minimal functional subunit of biological transduction networks, cannot exist without a functional overlap that results in performance troubles. If different catalysts act on the same substrate, it implies that they share a functional domain that recognizes specifically this substrate. Such a common interaction domain cannot be the one that defines the identity of the catalyst.

So, summarizing all the conditions required for building our robust catalytic motif, we need a reaction motif in which:

- The reacting species are duplexes
- The species react via 4-way branch migration
- The identity and the activation state of the different chemical species are in different strands.
- The strands of the target and the catalyst species must bind each other to initiate the process.
- There must be more than one possible catalyst interacting with a given species

Taking all these conditions into consideration and taking inspiration by the localized reaction scheme described by Winfree and Qian and depicted in Figure 19 we have designed a reaction framework that accomplishes all of them and allows us to implement real robust catalytic reactions for out-of-equilibrium systems: the "**Active Circuits of Duplex Catalysts**" framework (abridged as **ACDC Framework** from the present moment onwards).

2.2.1- General description of the ACDC Framework

2.2.1.1- Species in the ACDC Framework

The ACDC Framework is based on the existence of 2 main types of duplexes that accomplish all the roles in a catalytic process. The first type of chemical species are the **major**

species, whose general structure is described in Figure 20. All major species in ACDC are either catalysts or substrates and all of them are anatomically identical. Each major species is a duplex consisting of two strands, called the identity strand and the state strand. The identity strand is the preserved molecular core while the state strand specifies if the species is activated. The two strands that constitute a major species at a given time are bound by 3 different functional domains. In the centre of the strand is placed the core domain. This domain is common to all the strands and has the role of providing stability to the binding of the strands.

Flanking the core domain are found two different hidden toeholds that are not exposed in the case of the major species but will be functional in other species of the system. Next to the hidden toeholds, major species also feature unbound exposed toeholds both upstream (in our convention, corresponding with the toehold with the 5' extreme of the identity strand) and downstream (in our design convention, corresponding with the toehold with the 3' extreme of the identity strand). These toeholds are the responsible for the recognition and reaction triggering between major species. It must be noted that the same duplex species can have different roles in different reactions being the target in one of them and the catalyst on another one – which is a required condition to implement vertical cascading and multi-layered circuits on any framework.

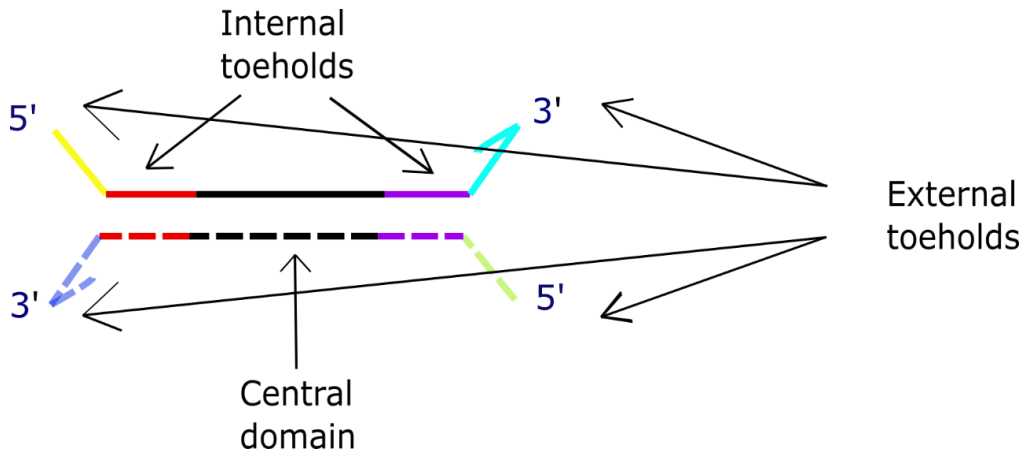

Figure 20. General structure of a major chemical species in the ACDC Framework. It must be noted that in the framework only substrate and catalyst species have this given structure.

The other type of species on which the ACDC Framework are called **ancillary species** (which are represented in Figure 21). The strands that compose ancillary species are the same that constitute the main species of the system, but their roles and the way they are bound differ from the major species. Ancillary species act as the fuel or waste species that are consumed and produced through a catalytic reaction of our CCNs. Reaction intermediates formed by the interaction of a substrate and a given catalyst are also part of the ancillary species of the ACDC Framework.

In major species, the identity strand and the state strand of a given species are bound by the central domain and the two internal toeholds of the strand. On the other hand, in the ancillary species, the two strands are bound by the central domain as well as the two toeholds

- one internal and one external. It must be remarked that the number of base pairs that stabilize the formation of the duplexes is the same for all the species of a given reaction or CCN.

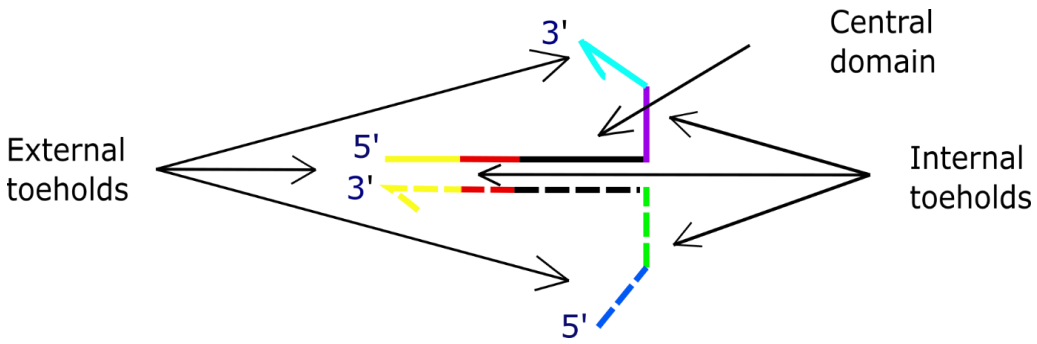

Figure 21. General structure of an ancillary chemical species in the ACDC Framework. Again, in the ACDC Framework, only the species acting as chemical fuels (as well as catalytic reaction waste) and the reaction intermediates between a given substrate and a catalyst have this duplex geometry.

Full catalytic reactions in the ACDC Framework occur in a two-step reversible process. The first step (as depicted on Figure 22) is triggered when two major species (one catalyst and one substrate) recognize each other via their complementary external toeholds, as depicted in Figure 21, to form a 4-way junction. In this junction, the two former duplexes undergo a strand exchange process between the regions of the junction that are complementary to each other. The strand exchange takes places until the toeholds exchanged are not complementary and the junction dissociates to release two ancillary species: a reaction intermediate and a waste molecule. The reaction intermediate is formed with the identity strands of both substrate and catalyst. As the number of base pairs remains constant throughout the process, this single step is reversible in nature and the resulting ancillary species of the step reaction can rebind and turn back into the original major species. The overall direction of the reaction is dictated by the entropic component of the reaction'sf free energy. Thus, the concentration imbalances between the species present in the medium will be used to drive a reaction.

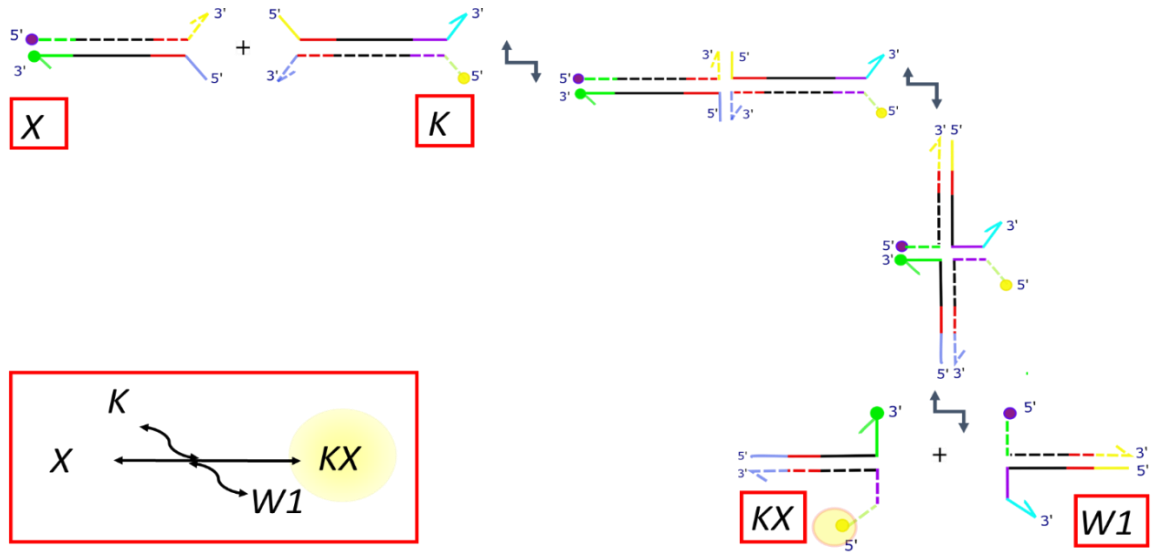

Figure 22. First reaction step of a catalytic turnover reaction in the ACDC Framework. In this first step two major species (a substrate X and a catalyst K) recognize each other and trigger a 4-way branch migration reaction that can give place to two ancillary species: a waste molecule and an intermediate formed by the two identity strands of both the substrate and the catalyst. In principle all the reaction steps in the framework are reversible. It must be noted that in this depiction of the reaction step, the downstream toeholds of the substrate X as well as one of the upstream toeholds of K are being substituted by FRET species that allow reaction tracing.

The second step of a catalytic reaction in the ACDC Framework is the reaction of the ancillary reaction intermediate with a fuel molecule as depicted in Figure 23. The fuel molecule is virtually identical to the waste molecule produced in the first step except for one of the unbound external toeholds – the one that is in the downstream interface of the state strand of the initial substrate. This means that the fuel molecule has the same exposed inner toeholds that the waste molecule created in the first reaction step has.

As a consequence, the fuel and the reaction intermediate can recognize each other, form a 4-way junction and take part in a strand exchange. The result of this strand exchange is the regeneration of the catalyst species and the generation of a substrate species that is identical to the original one except for one external toehold that determines the activation state. Again, this reaction is entropically driven by concentration imbalances - as it happened with the first reaction step - with the fuel in this case the one that is in overabundance driving the reaction to completion.

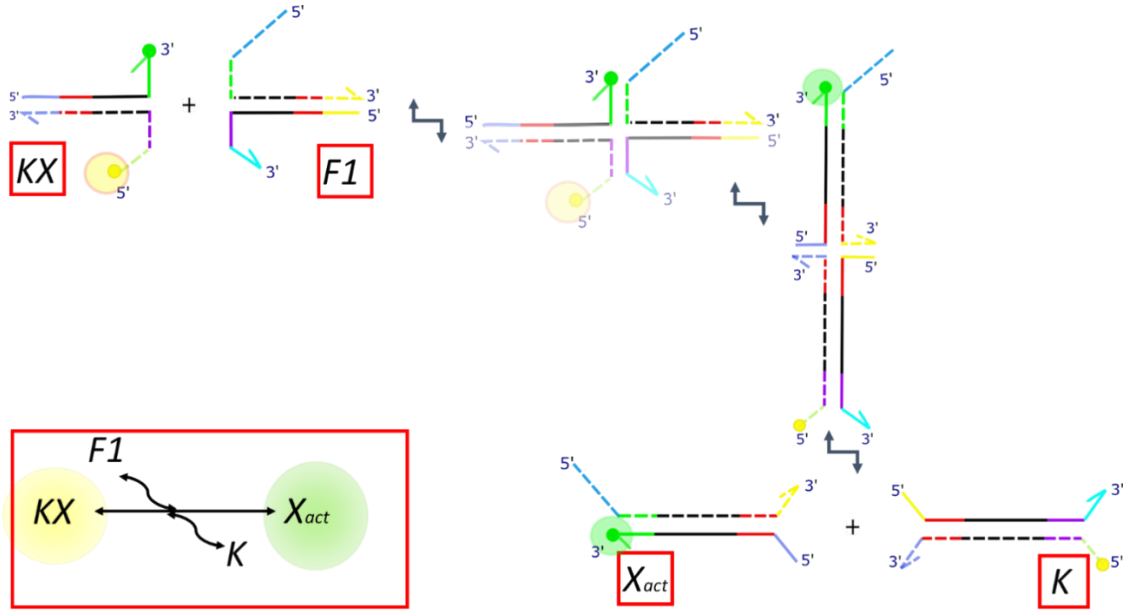

Figure 23. Second reaction step of a single catalytic turnover reaction in the ACDC Framework. In this second step the intermediate species from the initial step reacts with another ancillary species (a fuel) regenerating the catalyst and exchanging a toehold from the initial substrate species X leading to a different activation state. Again, some of the toeholds have been substituted by FRET elements to trace the reaction and the reaction is essentially reversible.

The proposed mechanism for a catalytic reaction in the ACDC Framework allows us to implement catalytic reactions and to implement more than a single catalyst per substrate as opposed to previous catalytic TMSD implementations like the seesaw gate. We are able to do so creating another catalyst species (that we will dub P) that shares the same downstream interaction domains than the catalyst K but a totally different set of domains -internal and external- upstream. Hence the new catalyst is able to recognize the same substrate, but generates and interacts with a new set of ancillary species.

Moreover, the ACDC Framework is able to separate the identity and the activation state of the chemical species as stated in the discussion of the previous section. It can also define a different activation state with the change of a single toehold downstream, since the 4 way branch migration process require the presence of both toeholds to be triggered effectively. Hence, removing one of the toeholds results in a reaction deactivation while an exchange that adds a functional toehold needed for an interaction results in an activation process. These features suggest that we can define different activation states beyond activation or inactivation using different lengths of toeholds that bind with a different rate. This in turn suggests that tuning of the rates might be achievable inside the ACDC Framework

In short, defining the ACDC Framework and defining the duplex species mechanism allows us to envision a way in which push-pull-based CCN can be implemented into nucleic acid reaction systems. However, the envisioning of a working mechanism does not directly imply the realisability of the aforementioned mechanism. The design process of a given CCN and the potential shortcomings that can be faced in the process as well as the solutions implemented will be the subject of the following subsection.

2.2.1.2 Design of ACDC-based systems

The pipeline requires the definition of the nucleic acid structures of the species we want to design, followed by the definition of the domain structure of the strands we want to define in our system as well as the way that such strands are going to be combined in every single species.

So, the steps required to design a CCN in the ACDC Framework consist of the design of every single reaction of activation/deactivation that constitute the desired CCN. Once the reactions have been defined, all the species taking part in it, both main and ancillary, must be identified with all the subsequent interactions. The knowledge of the internal and external toeholds required by the aforementioned reactions comes next as a consequence of the interactions required. Finally, once the domain structure of the strands as well as their role in each species is defined, comes the last step consisting of the assignment of the sequences corresponding to each domain.

This last step of the design pipeline is assisted through the use of the NUPACK web tool[11]. NUPACK allows us to perform the sequence assignation of the domains ensuring the orthogonality of the reactions minimizing the risk of unwanted interactions at the same time that the overall free energy difference that drives the reactions that form the system is kept under a minimum. The whole design pipeline going from the highest abstraction level to the sequence is summarized in Figure 24.

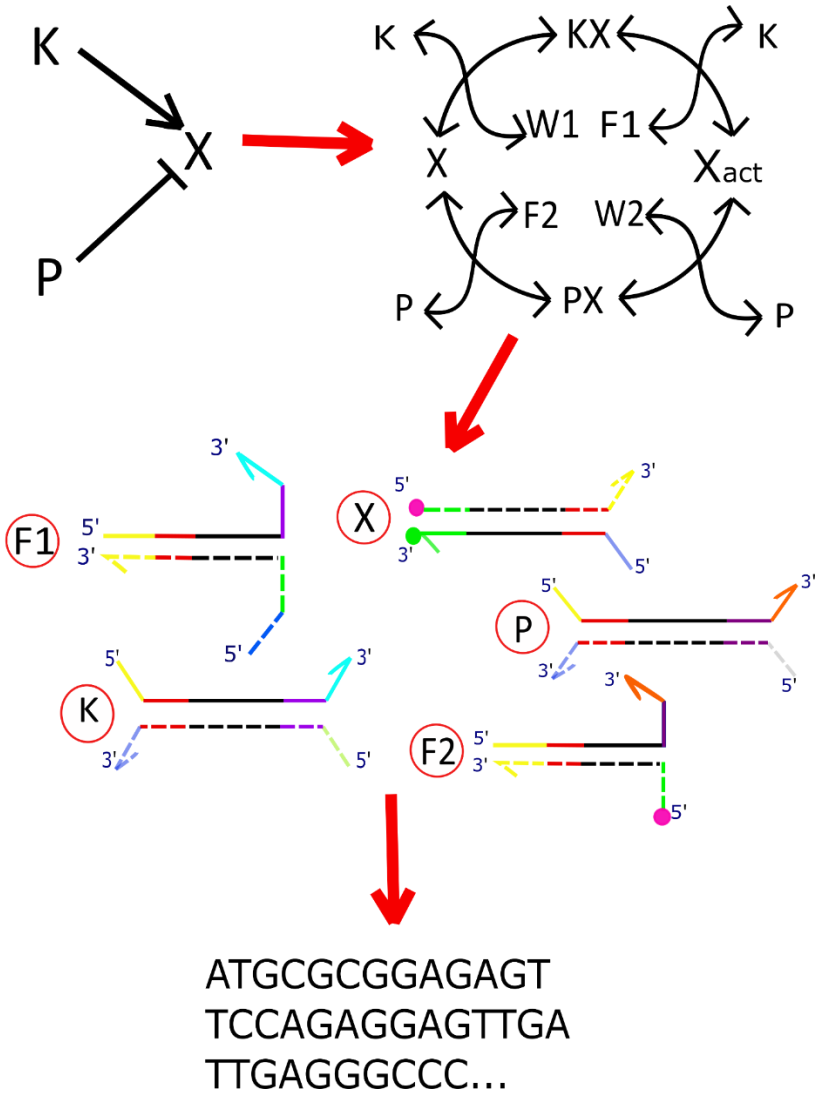

Figure 24. Summary of the original design pipeline for the design of a CCN in the ACDC Framework.

Initially, all the ACDC systems envisioned for the Chapters 2 to 4 of the present thesis were designed manually at the level of domains, before feeding the required domain structure into NUPACK to generate sequences. By encoding the structures in NUPACK -and threading structures in those domains with the sequence constraints however it must be noted that the NUPACK analysis did not assess for domain leakiness or repetition. The list of domains and strands designed for all the tested networks is available at the supplementary information of the chapter.

However, compilers that are able to verify all the possible reactions of a defined CCN and have been developed recently in our group[99]. This compiler has proved itself a useful tool since it has allowed me to explore design limitations emerging on extended ACDC reaction networks that will be treated in Chapter 5 of the present thesis.

Despite having these powerful tools, the ability to design a given network in the ACDC is not a direct proof of its feasibility. In order to make the proposed CCNs functionally feasible, we need to demonstrate that the motifs are actually functional.

2.2.1.3. Feasibility tests of the ACDC Framework

In order to assess whether our push-pull-based CCN are feasible or not, the first step we need to assess all the assumptions taken in the basic design of our species and identify the potential risks that might compromise the catalytic action of the network. Then we need to design a way to observe whether these risks occur and find potential solutions in order to make the reaction scheme suitable for building our desired CCNs Hence, observing the fundamentals of an ACDC-based are three fundamental risks derived from the use of duplexes as the reacting species in our system.

The first risk is derived from the very fact of using duplexes, as all the species must be prepared by annealing the constituent strands with absolute stoichiometric precision. If any of the strands were in a slightly higher concentration than intended, it would result in the excess strand interacting with the duplexes in a regular TMSD reaction and triggering the activation or deactivation of the substrate species as seen in Figure 25. Unfortunately, the chances of such concentration imbalances are almost certain due to errors in the DNA strands synthesis as well as in experimental manipulation. Hence, our experimental protocol must implement a procedure to render these leak reactions irrelevant.

Then, even with the risk of leak reactions eliminated, new problems arise due to the use of 4-way junctions. As we saw in the initial chapter, many of the original implementations of structural DNA nanotechnology envisioned by Ned Seeman had 4-way junction-based subunits as their constituents. Hence it could be expected that the 4-stranded intermediates could be stable enough to be kept in a little fraction in equilibrium despite the fact that the DNA duplexes are a more stable configuration. The formation of these stable 4-stranded species, as depicted in Figure 26 would result in sequestration and undesired distortions of the reaction dynamics, making the system harder to model and analyse. Given this reason, the reaction conditions must be adjusted to supress 4-stranded species held together by 4-way junctions.

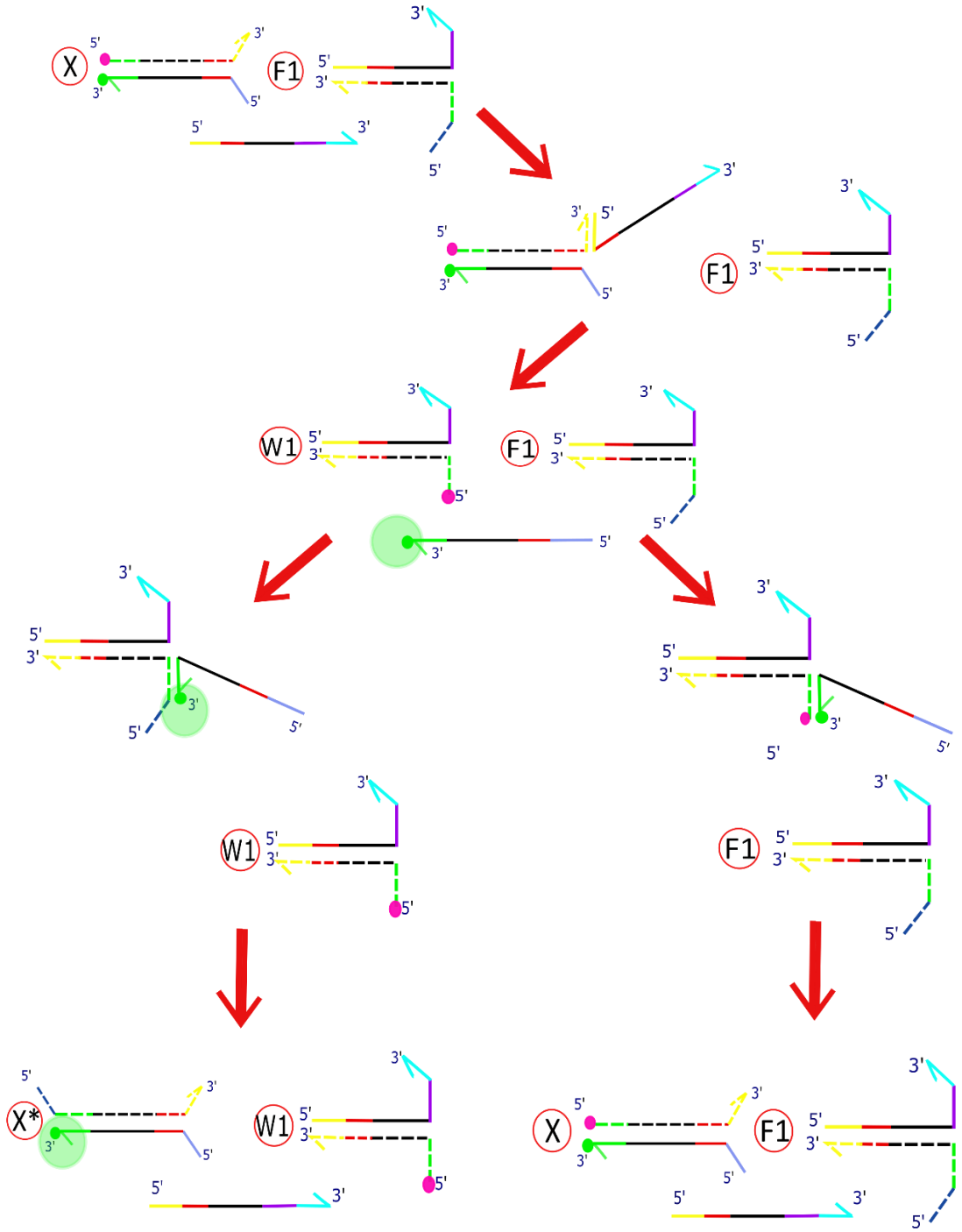

Figure 25. Representation of a leak reaction caused by the accidental excess of the fuel identity strand in the medium. This type of reaction was the first type of inherent risk identified in our duplex-based system that should be tested and addressed.

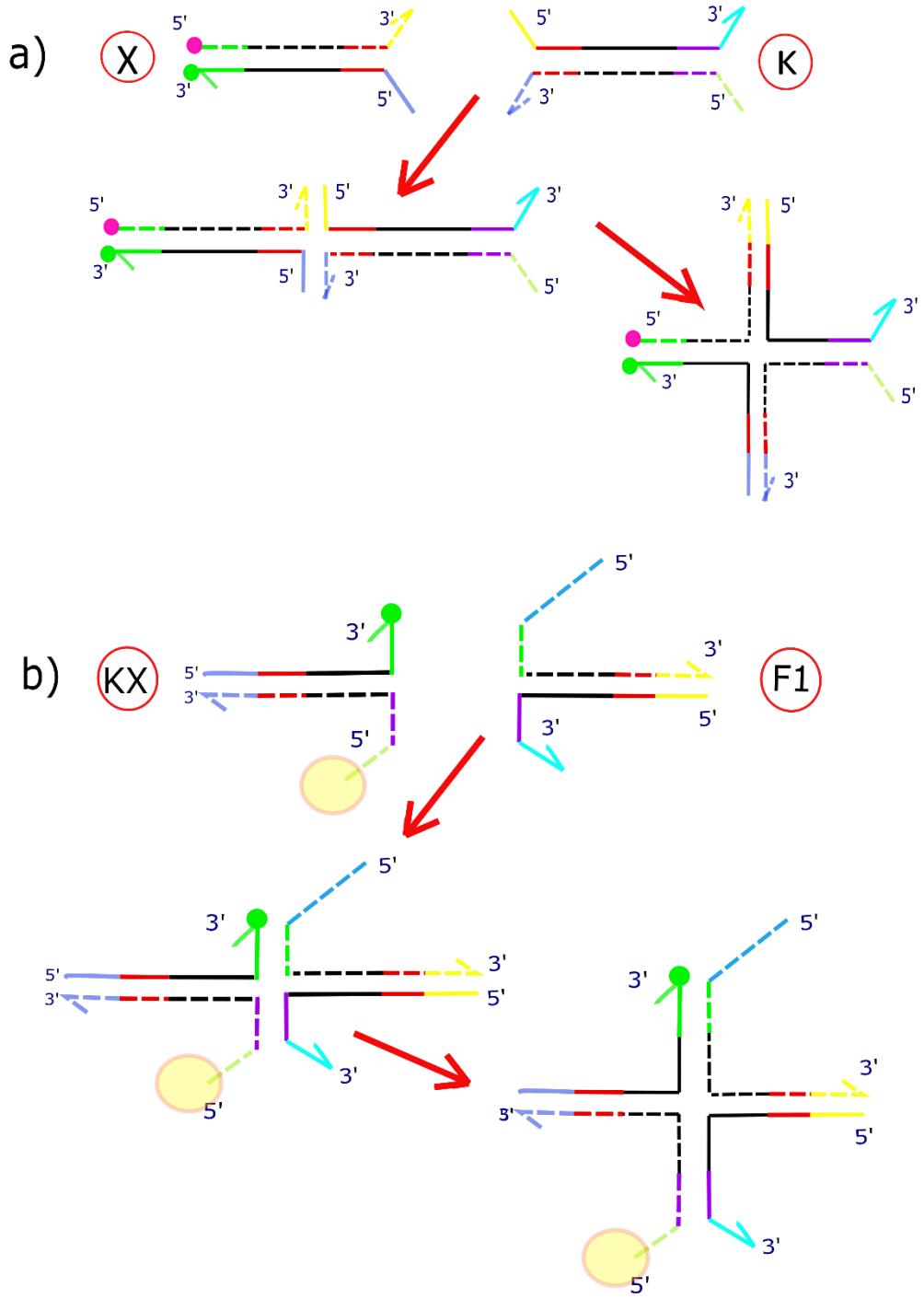

Figure 26. Theoretical formation of stable 4-stranded species. (A) Formation of a 4-stranded intermediate in the first step of a catalytic activation reaction in the ACDC Framework between a catalyst and a substrate. (B) Formation of a 4-stranded intermediate in the second reaction step of catalytic activation reaction in the ACDC Framework between a reaction intermediate and its fuel.

Finally, the last identified inherent risk for ACDC-based CCN is the possibility of leaky single toehold reactions as described in Figure 26. The presence of slow leaky reactions with

a single toehold was described originally by Dabby in 2013[52] and they can compromise the catalytic activity of our system in two different ways, which are depicted in Figure 27.

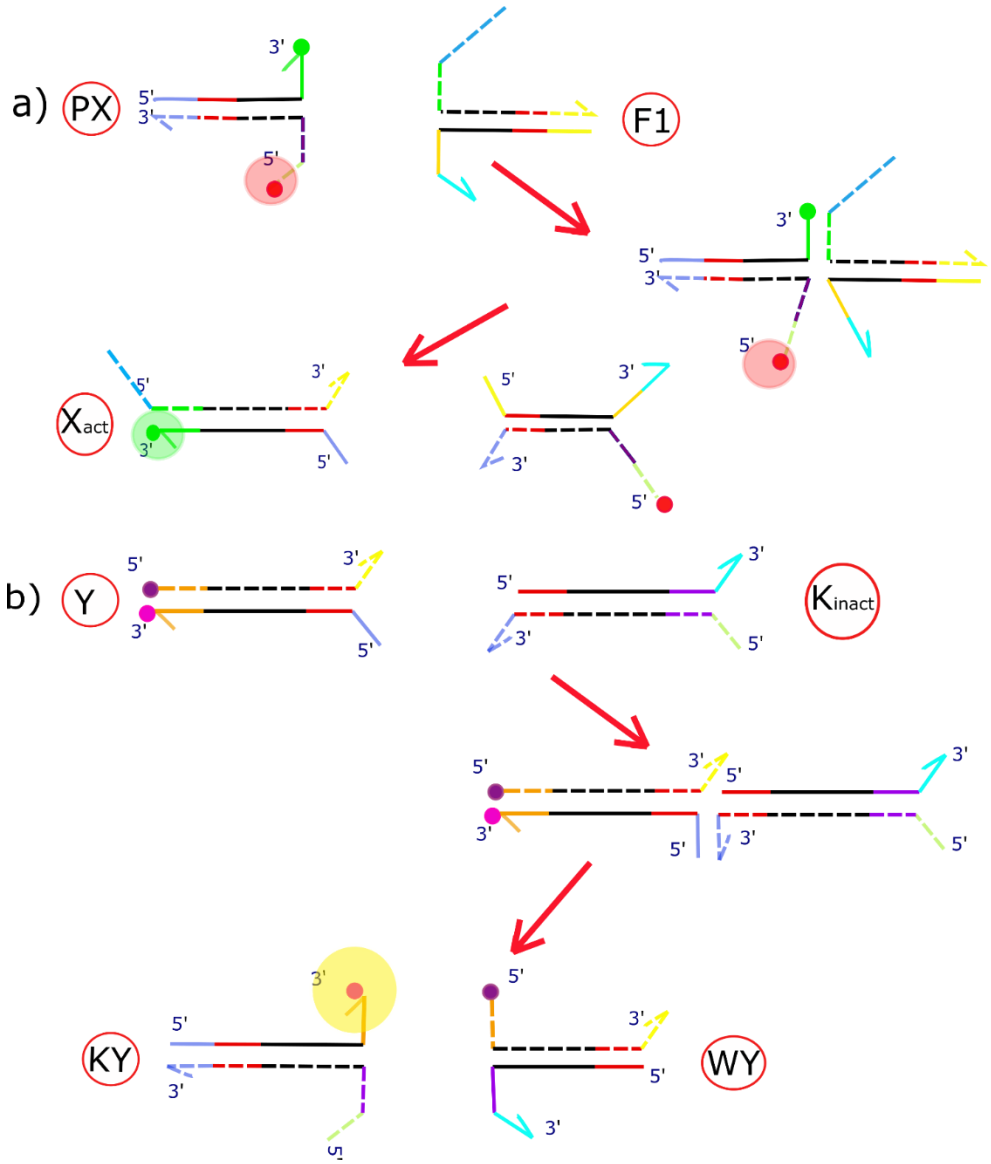

Figure 27. Theoretical triggering of single toehold mediated strand exchange. This potential problem can occur in the ACDC Framework in the shape of two different reactions: (A) reaction between an intermediate (PX in the figure) and a non-corresponding ancillary species (F1) . This type of reaction is known as a crosstalk reaction. (B) Reaction between an inactive catalyst (K_{inact}) and a given substrate (Y). This reaction is known as an uncontrolled cascading reaction.

First, the single toehold reactions between duplexes could occur as different catalysts that act on the same substrate share interacting inner toeholds for the state strand. This would mean that different substrate-catalyst intermediates could interact with fuels that are not designed for. This would result in design problems for two different reasons. First, these reactions would destroy orthogonality between catalysts as seen in Figure 27a and it can also act a source of mutual retroactivity between catalysts. The latter would destroy modularity and impose loading effects on our system[100].

It could be argued that having a universal fuel could be a solution that simplifies the design of the system while making it more akin to how do transduction networks function in vivo with all the catalytic proteins using ATP or GTP as their fuel. But this solution would extend the aforementioned mutual retroactivity problems to all the designed CCN.

The second way that this single toehold reaction compromises the catalytic control of a CCN is when this reaction occurs between an inactive catalyst and a substrate via single complementary external toeholds as seen in Figure 27b. While the rate would vary between the two activation states, the presence of non-negligible leaks would severely limit the possibility of exerting catalytic control over a cascading system rendering it useless. Hence, proper feasibility controls must include probing conditions in which these single toehold reactions cannot happen.

Although these reactions are thermodynamically unfavoured due to the formation of an unmatched toehold, this thermodynamic barrier could not be enough if one of the reactants is in overabundance making the existence of these leaks non-trivial. We therefore establish in the following subsections whether any of the aforementioned challenges are real pitfalls and propose modifications to address any problems.

2.2.1.3.1 Single strand leak resilience protocol

Aware of how troublesome the effects of the aforementioned concentration imbalances could be, we developed a methodology to render any possible strand excess inert. This methodology is based on mixing one of the strands that constitutes the duplex with an intentional concentration excess of a 100% more than the other strand of the duplex. The excess strand is chosen for each case so that it lowers the signal baseline in the Alexa 488 channel. The excess strand is then rendered inert through the addition of an equal concentration of a “blocker” strand designed as in Figure 28.

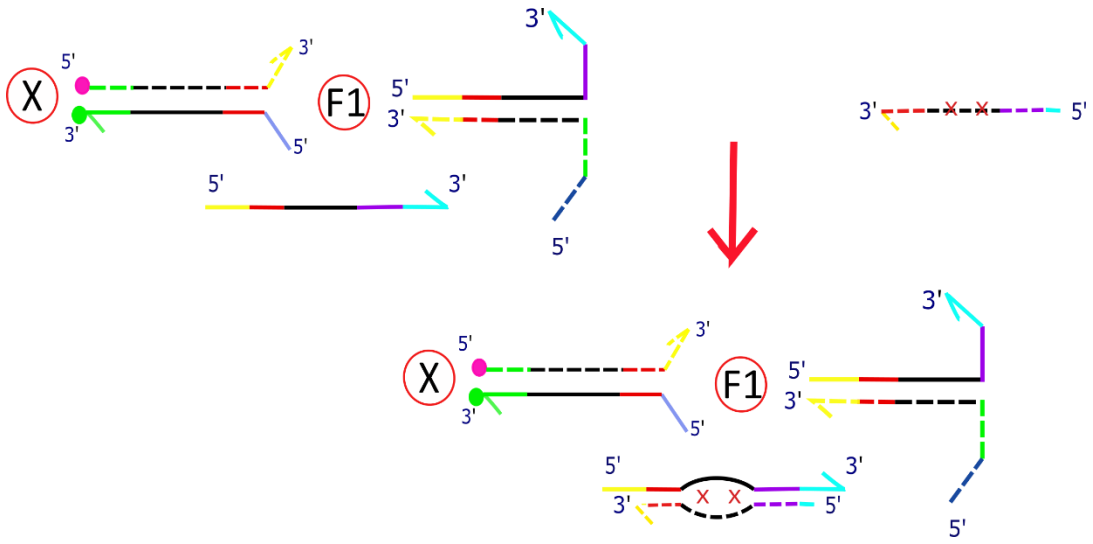

Figure 28. Reaction mechanism for the excess strand blocking strands. Based on this mechanism we can neutralize excesses of strands even in very large ratios and assure that stoichiometry is effectively 1:1. In the present example the state strand for X_{act} is floating freely and without the blocker could trigger the reactions depicted previously in Figure 26. However, in this case there is also a blocker strand (in dashed lines) with the mismatches binding to it and avoiding the invasion process.

This blocker strand is designed to be completely complementary to the intentional excess strand except in the terminal 5' and 3' binding domains in which they bind to only 1 base-pair. Also, two mismatches are introduced to the central domain. These design tweaks are added to disfavour as much as possible the displacement of the already-bound strand added in excess of while still allowing the blocker to bind to the excess.

In order to test the efficacy of the blocker strands, we designed two experiments to assess their action. The first experiment consisting of the observation of the fluorescence dynamics of the substrate X in a total concentration of 30 nM when placed in presence of a 1 µM excess of F1 without any catalyst present. In the experiment, the excess of fuel was prepared following three different methodologies.

In the first one, the Fuel was prepared in a 1:1 stochiometric ratio between the two constituent strands coming from a 30 µM prediluted stock and left to anneal at room temperature during 2 hours with an equivalent volume of TAE buffer Na^+ 1M to adjust the final concentration to 10 µM.

In the second variation of the fuel preparation methodology, the times were kept, but the ratio between the strands was 2:1 in favour to the acting invading strand- as seen in Figure 25. This variation was implemented as a positive control for the leak and activation of the substrate X and was done mixing the 30 µM stock of the strand that was not in excess with the equivalent volume of a 60 µM stock of the excess strand that again was left to anneal for 2 hours and after which got added a volume of Na^+ 1 M buffer to adjust the final concentration to 10 µM F1 +10 µM excess strand.

The third variation on the preparation of the substrates followed the same procedure as in the second variation, but the step of adjusting the volume was substituted by the addition of an equivalent volume of 60 µM stock of the blocker strand excess to the mix instead of the TAE Buffer Na^+ 1 M buffer and left to anneal for 2 hours, so the final mix of that fuel has the excess of invading strand blocked by the blocker. Substrates and fuels were placed together alongside controls and were monitored via fluorescence tracing. The samples were excited at a 465 nm wavelength and monitored in the emission wavelength of the Alexa 488 fluorophore of the substrate X (530 nm) during 9 hours following a timestep of 32 seconds and a total of 1000 measurement steps at 25 Celsius on a ClarioSTAR plate reader using 70 flashes per reading. The results of this experiment are depicted in Figure 29.

The second experiment was a variation of this experiment in which the fluorescence tracing is done for X_{act} and W1 in excess. Following the same procedure in the preparation of the substrate and the waste, in this series of experiments we intended to observe was a decrease in the fluorescence. This decrease is expected to be given by the invasion of the excess strand to the substrate X_{act} and the posterior binding of the released labelled strand to the waste molecule in the surroundings similarly to how it did it in the case of X and F1. W1 had three preparation variations that were analogous to those of the experiment of Figure 29. The fluorescence tracing conditions (excitation, emission, timestep, timespan and temperature) were kept from those experiments as well. Results of these experiments are depicted on Figure 30

We observed that, as we wanted, no leak reaction occurred when the blockers were used whereas leaks happened when fuels were used in the absence of blockers. The leak reactions demonstrate not only the use of the blocker strands worked as intended but also when no common interaction domain between duplexes is exposed, the reaction steps cannot happen. This also confirmed the property that reaction steps can only happen with duplexes that share the same geometry. However, it must be reported that all the reactions that involved an excess of W1 – which had a quencher - had a decrease in the baseline fluorescence level. Experiments with the species X_{act} (The activated form of the substrate X that appears depicted at the bottom of figure and the strand with the quencher bound to its corresponding quencher confirmed that the decrease was due to environmental quenching caused by the great excess of quencher in the environment as seen in Figure 31.

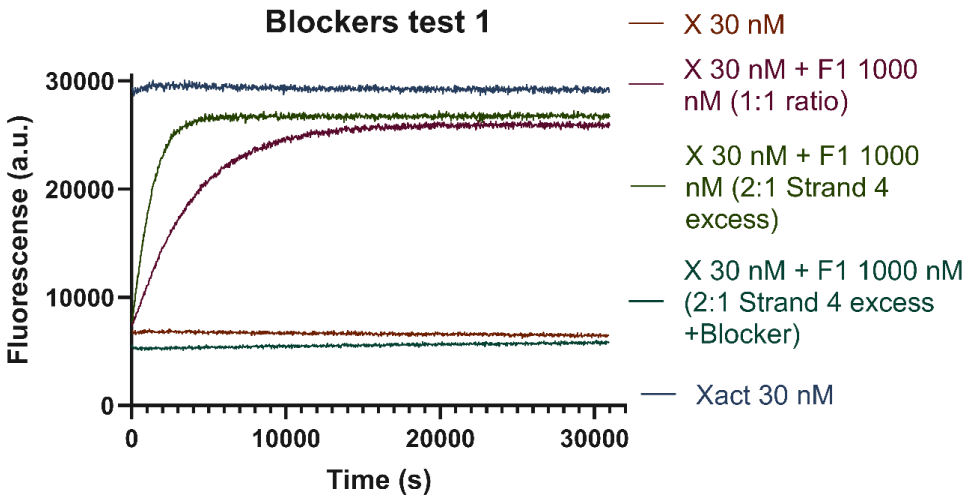

Figure 29. Fluorescence tracing for the test of the blockers. A depicted in Figure 24 substrate X was challenged with a large F1 excess, triggering leaky activation as seen there. As we see activation happens with both the 1:1 stoichiometry and the intended unblocked excess of strand 2, but not when the excess is blocked as we intended and depicted in Figure 27.

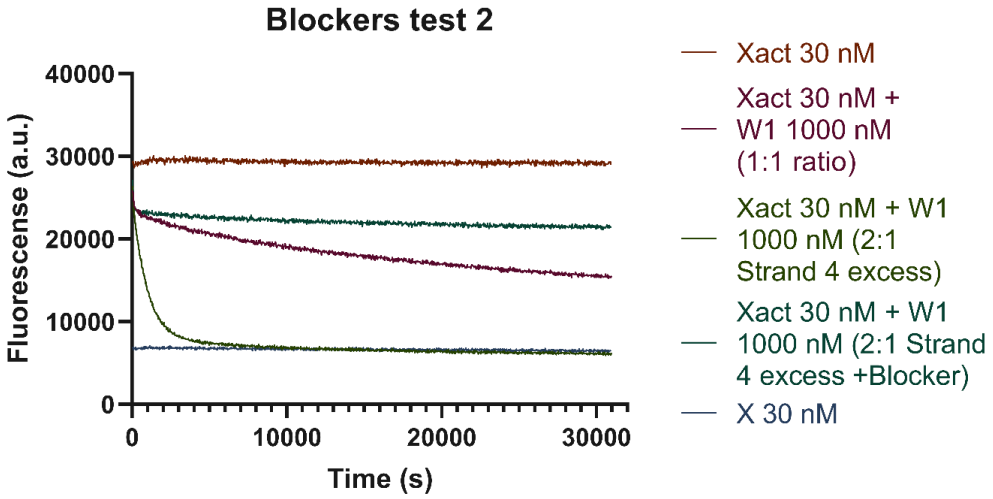

Figure 30 Fluorescence tracing for the test of the blockers working mechanism. A substrate X* was challenged with a large W1 excess triggering leaky activation. As we see deactivation clearly happens with both the 1:1 stoichiometry and the intended unblocked excess. The blocked excess seems to give a quick fluorescence drop, but this is due to environmental quenching by the very large quantity of duplexes with an Iowa Black FQ quencher. The working mechanisms underlying this test are analogous to those depicted in Figures 24, 27 and 28, but for leaky activation.

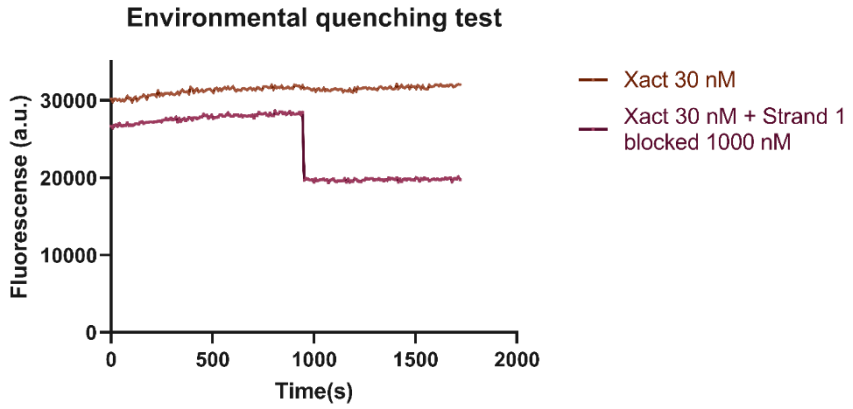

Figure 31. Test of the environmental quenching of big excesses of Iowa Black FQ. X_{act} and X_{act} + Strand 1 (bearing the Iowa Black FQ quencher) neutralized by its corresponding blocker. Blocked strand 1 causes a sudden drop-off due to environmental quenching. In this case, no possible 4-way branch migration is possible in the system, hence all the drop in the fluorescence in this case will be due only to the excess of quencher acting in the environment of the fluorophore.

2.2.1.3.2.- 4-way junction stability.

The second type of potential initial drawback we tested was checking how stable are the 4-stranded intermediates formed at the beginning of each reaction step. We needed to assess this possibility because if such intermediates were stable, there would be a non-negligible chance of them hampering the reactions and affecting the dynamics of the system.

In order to assess this possibility, we performed experiments with modified versions of the K and X species as well as the ancillary species KX and W1. These DNA species had two types of modifications that were requested for these experiments. The first modification consisted on the extension of the bound domain of these species adding an extra 10 basepair domain. This new extra domain was different between the two interacting species and could not be exchanged in the formation of the 4-way junction, thus leading to the formation of a stable 4-stranded junction. Moreover, the interacting exposed domains of the duplexes had a moiety attached to allow us to monitor the formation of the junction via fluorescence decrease. More precisely the variations for X and KX, as depicted in Figure 32, had an Alexa 488 fluorophore adjacent to the upstream interaction domain while the K and the W1 analogues had an Iowa Black FQ quencher. This way, as seen in Figure 32, once the corresponding duplexes (K and X for one side, and KX and W1 for another) interact with each other, the fluorescence decreases.

The duplexes were assembled using the protocol settled for between duplexes. Some of these K and W duplexes also included an unbound extra single-stranded domain to ensure the formation of the stable 4 strand junction as a positive control as seen on Figure 32 b. Sequences for these strands are available in the supplementary info. All the reactions were monitored for an excitation of 485 nm and an emission of 530 nm using a timestep of 52 seconds on a ClarioSTAR plate reader using 70 flashes per reading.

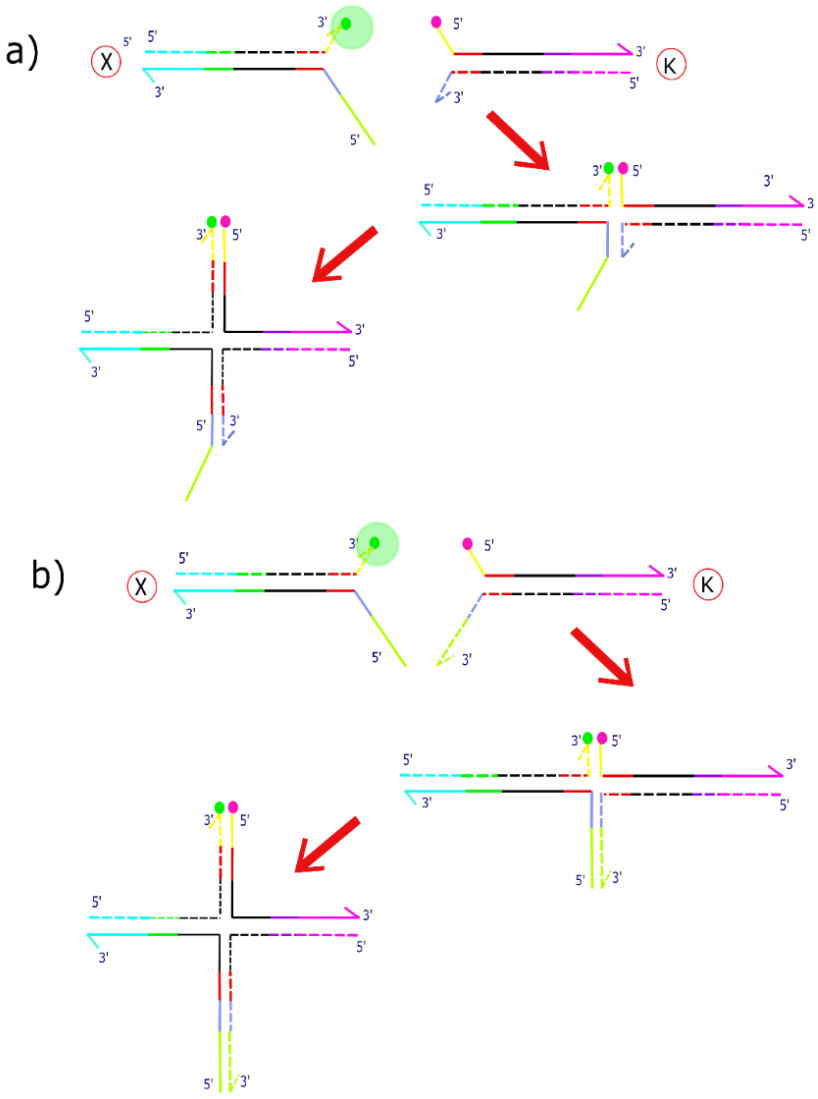

Figure 32. Reaction mechanism for the monitoring of the formation of stable with (a) and without the stabilizing extra domain (b) for the union of two major species (equivalent in sequence to K and X) modified. Analogous modifications were done to their equivalents in the resulting ancillary species (KX and W1) for their corresponding test.

When testing the stable junction formation, we observed the formation of some stable junction with the regular double toeholds for our standard systems at our initial conditions. In fact, in some cases (such as Figure 33) that featured a high number of GC pairing the junction was as stable as the positive control. So, in order to limit the formation of these unwanted metastable intermediates, we reduced the buffer condition to 250 nM of Na^+ and increased the temperature to 37 °C – which is actually closer to physiological condition. With these changes, the formation of the 4-stranded intermediates was limited as seen in Figures 35 and 36.

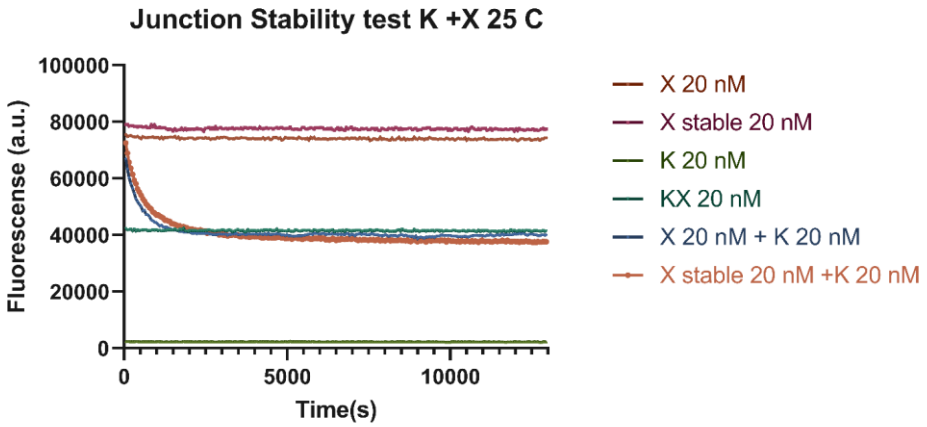

Figure 33. Formation of stable 4-stranded intermediates at 25 degrees Celsius for the reaction of the modified major species K and X. As observed, even without the addition of the extra stabilizing domain in the junction, the union of the duplexes is strong enough to keep the 4-stranded junction bound.

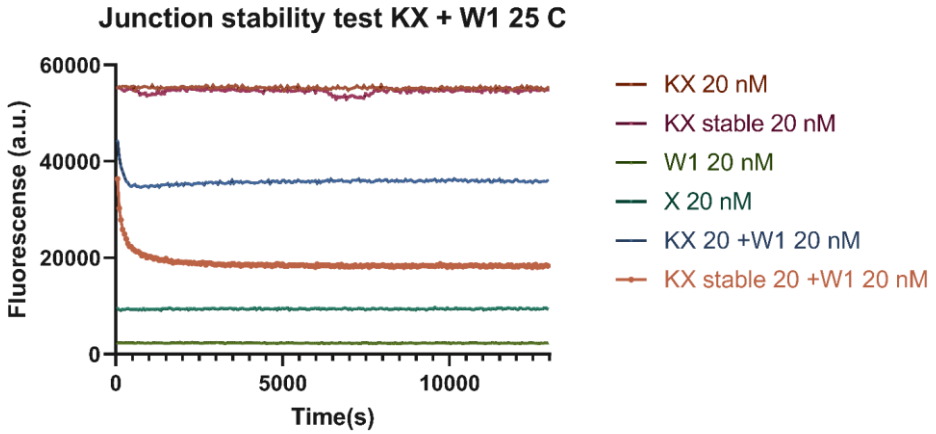

Figure 34. Formation of stable 4-stranded intermediates at 25 degrees Celsius for the reaction of the modified ancillary species KX and W1. As observed, even without the addition of the extra stabilizing domain in the junction, the union of the duplexes is strong enough to keep a substantial fraction of the 4-stranded junction bound.

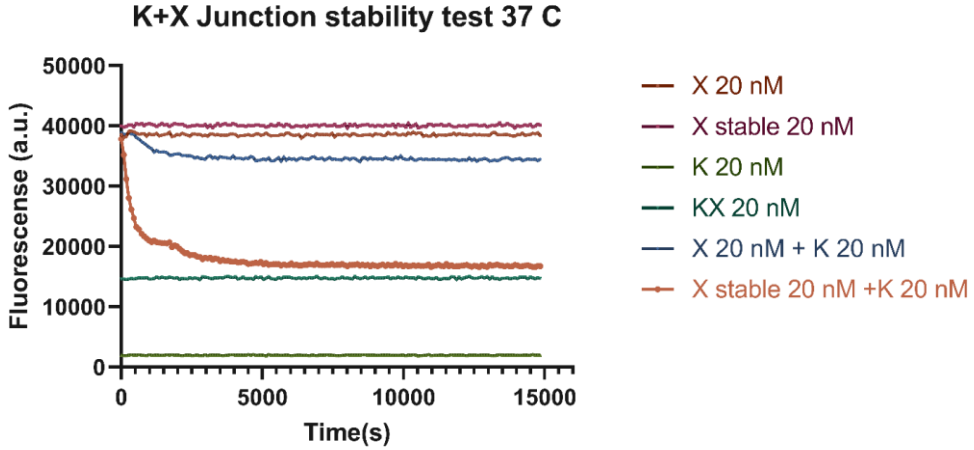

Figure 35. Formation of stable 4-stranded intermediates at 37 degrees Celsius for the reaction of the modified major species K and X. As observed, in this case the stabilizing domain is still working as previously observed but now the regular junction is much less stable giving less than a 20% of the prior fluorescence decrease.

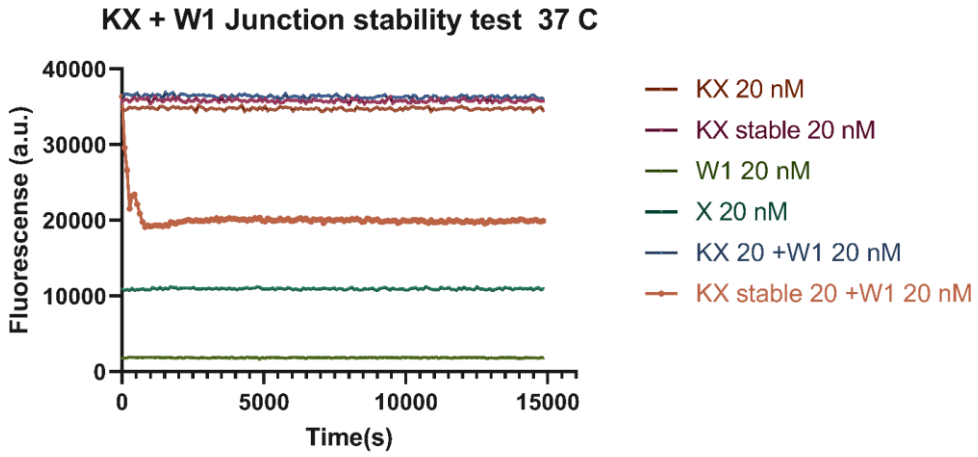

Figure 36. Formation of stable 4-stranded intermediates at 37 degrees Celsius for the reaction of the modified ancillary species. As observed, in this case, the 4- way junction is not at all stable in the new temperature conditions.

2.2.1.3.3.- Fuel orthogonality.

These experiments consisted of the observation of the second step of an ACDC reaction using fuels that are not meant to work with the intermediate duplex. In this case, the duplexes have same geometry, but they share only one complementary interacting domains instead of two in an exposed available interface. This way we prepared the ancillary species KX and PX as well as the fuels F1, F2, and the waste W1, W2 following the protocol for preparing species with blocker strands established in section 2.2.1.3.1 of the present chapter. Information about the corresponding excess strands is given in the supplementary. These complexes were added as controls (as well as the substrate in the states X and X_{act} as controls for deactivation or activation) a 20 nM concentration. Then, in separated well each intermediate complex was mixed with the fuel or the waste corresponding to the other catalyst (this is F2 or W2 for KX and F1 and W1 for PX) in a 1:1 ratio and we monitored the changes in fluorescence that might indicate that the ancillary species were reacting with each other using the conditions of excitation and emission (465 nm and 530 with a gain of 2000 and 70 readings per well) used previously in the blocker strand tests. The measurement equipment for these experiments was again the ClarioSTAR plate reader. The measurements were taken during using a 31 seconds timestep for 3 hours and 20 minutes at 37 Celsius and 250 nM Na^+.

As seen on Figure 37, no reaction is observed in any of the 4 cases studied, confirming that reactions can be only triggered with the presence of the two correct pairs of complementary available interacting domains in the specific interfaces. These experiments are a further proof of the catalytic power of ACDC since, given that no crosstalk is possible in the system, the driving force for each ACDC reaction will only be the amount of fuel for each type of catalyst available. With this experiment, we confirm that our catalysts are orthogonal ensuring that their behaviour lacks undesired crosstalk.

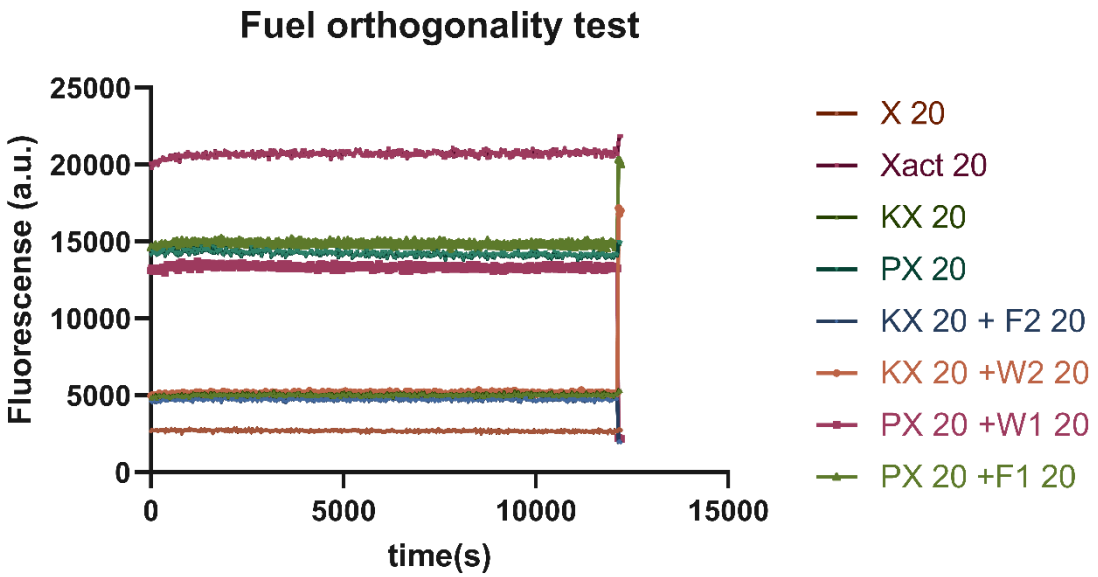

Figure 37. Test for the orthogonality of fuels. As we can observe, no change in fluoresce was observed in either of the two intermediates (KX and PX) react with any of the ancillary species that correspond with the wrong catalyst (F2 and W2 in the case of KX and F1 and W1 in the case of PX). The increase/decrease in the endpoint is due to the addition of totally complementary strands to the identity strand of the substrate X that have either an Iowa Black FQ quencher (thus causing the signal to drop to the level of X) or not (making in this case to increase to the level of X_{act}). The high increase and decrease in the signal at the end of the measurement is due to the addition of a strand fully complementary to the state strand of X -with or without a quencher- in order to observe the full activation or deactivation depending on the case.

2.2.1.3.4. Cascade leaky activity control.

In these experiments, we tried to trigger a reaction between both active/inactive forms of a substrate Y and a catalyst K in its inactive state without any fluorophore that might do. The interacting exposed downstream domain of the state strand of the catalyst (Strand 4) was removed completely so to avoid interactions. Moreover, again we have two duplexes of the same geometry reacting with only one of interacting domains being complementary but in this case are major species (a substrate and a catalyst).

The substrate Y in both its active and inactive versions and the inactive version of catalyst K were prepared using the protocols settled in sections 2.2.1.3.1 and 2.2.1.3.2 (information about excess strands is again available in the supplementary information). The equipment in which the fluorescence was measured, again was the ClarioSTAR plate reader. However, while temperature salinity and gain conditions of the measurement remained the same for as in section 2.2.1.3.3. the excitation and emission wavelengths of the measurement were changed. This change was due to the fact that the fluorophore of the substrate Y was Cy3 instead of Alexa 488 as it was in this case of substrate X. In this case the excitation wavelength was 530 nm while the emission wavelength was 585 nm. In this particular case the timestep was of 18 seconds and the measurement was taken during 18,000 seconds in order to ensure we could really see any slow leak if it happened.

The first thing we were able to assure -as can be seen in Figure 38- is that both states of Y and KY had different fluorescence signals at the same concentration which allowed us to

make significant comparisons. As we saw in the fuel orthogonality tests, no reaction is triggered in any case by the inert catalyst. It also means the state of activation of the catalyst can be in principle regulated in the ACDC-based systems. This opens the possibility for fine-tuned control of different layers of circuits.

However, it must be noted that as an additional design consideration, when studying a domain that codes for an inactive catalyst state, it's convenient not to place any fluorophores in the aforementioned toehold. This fact follows since the fluorophores can give extra stability to the union that can be enough to overcome the energy barriers that make unfavourable the reaction (specially in short oligos where the impact on the melting temperature can be very noticeable[101]).

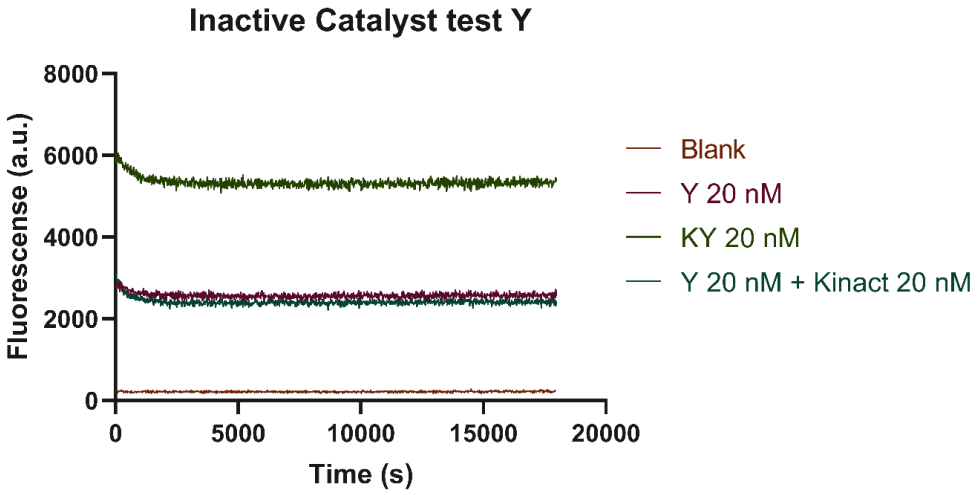

Figure 38. Test for the leaky reaction of an inactive catalyst K acting with its substrate Y. As it can be seen, no change in fluorescence was triggered when the catalyst was added compared to the catalyst-fee control. Hence the on/off state of a given catalyst is robust, ensuring that in principle cascading can be properly controlled.

2.3 General experimental design.

Based on our preliminary studies, we established the following protocols that will govern all experiments reported unless otherwise indicated. The first step of the process consisted in the obtention of the DNA strands using the design cycle described previously. With our pipeline, we were able to design seven DNA strands that when combined a system with a Target species X with two active states and two Catalysts (K and P) with their corresponding pairs of Fuels (1 Fuel species per Catalyst per Target State) that allows us to implement in principle the full push-pull motif that we want to test as our network unit.

The species obtained initially were later used as an input in another design cycle in which we incorporated Target species Y for our initial catalysts K and P (adding only 1 more strand to the initial set of seven). A second design cycle was performed in order to design and a set of Catalysts (KK and PK) with their corresponding set of Fuels each that had the Catalyst K of the original network their substrate species. This implied that in this new extended

network K also formed intermediate species with each catalyst upon interacting with them (CKK in the case of the KK catalyst and CPK in the case of the PK catalyst).

This addition of these strands (five extra to the set previously built) allows us to build a two-step transduction cascade. The following table will detail the Strands that form our duplexes following the output number that NUPACK as assigned them on the cycle designed. The code for NUPACK as well as the resulting sequences is included in the supplementary information. The total list of species for this extended system is available in tables 1 and 2

Species	**Identity Strand**	**State Strand**
X (Substrate)	Strand 3	Strand 1
X_{act} (Substrate)	Strand 3	Strand 2
Y (Substrate)	Strand 13	Strand 1
Y_{act} (Substrate)	Strand 13	Strand 2
K (Substrate/Catalyst)	Strand 5	Strand 4
K_{inact} (Substrate/Catalyst)	Strand 5	Strand 14
P (Catalyst)	Strand 7	Strand 6
KK (Catalyst)	Strand 19	Strand 18
PK (Catalyst)	Strand 17	Strand 16

Table 1: List of major species and their constituent strands obtained in the design cycle previously specified.

Species		
KX (Intermediate)	Strand 3	Strand 5
PX (Intermediate)	Strand 3	Strand 7
KY (Intermediate)	Strand 13	Strand 5
PY (Intermediate)	Strand 13	Strand 7
CKK (Intermediate)	Strand 5	Strand 19
CPK (Intermediate)	Strand 5	Strand 17
F1 (Fuel)	Strand 4	Strand 2
W1 (Waste)	Strand 4	Strand 1
F2 (Fuel)	Strand 6	Strand 1
W2 (Waste)	Strand 6	Strand 2
FKK (Fuel)	Strand 18	Strand 4
WKK (Waste)	Strand 18	Strand 14
FPK (Fuel)	Strand 16	Strand 14
WPK (Waste)	Strand 16	Strand 4

Table 2: List of ancillary species and their constituent strands obtained in the design cycle previously specified. It must be noted that ancillary species in our setup don't have defined identity and state strands.

This initial strands designs obtained were later modified in order to allow the fluorescent labelling of the different substrate and Intermediate species for the reactions. The fluorescent dyes were chosen from selected bibliography[102] by their capability to generate different FRET signals when they are one in the proximity of a given donor dye that is the same to all of them.

This common donor dye was placed in the Identity strand of the Target species X substituting the non-interacting upstream domain while the acceptors of the FRET were placed in the end of the downstream domains of the Identity strands of the catalyst. The inactive state Strand of the Target X was also modified to include a quencher substituting the interaction domain adjacent to the FRET donor in order to block any signal from this fluorophore when the Target is on its inactive state. This substitution again was possible due to the domain removed not taking part on any reaction on the designed system so far. The fluorophores chosen for each strand as well as the following:

Substrate X Identity Strand: Alexa 488.

Catalyst K Identity Strand: Cy3.

Catalyst P Identity Strand: Cy5 .

Target X Inactive State Strand: Iowa Black FQ.

The common excitation wavelength chosen for the possible 3 wavelengths was 465 nm (20 nm from the donor maximum excitation wavelength) in order to minimize the excitation bleed-through over Cy3, whereas the reading wavelength chosen for every possible combination of acceptor with the Alexa 488 fluorophore were the following, trying in every moment to minimize the reading window to avoid bleed-throughs from other fluorophores.

-Alexa 488/Nothing: 520/ 15 nm width.

-Alexa 488/Cy3: 600/20 nm width.

-Alexa 488/Cy5: 690/30 nm width.

On the other side, the identity strand of species Y was virtually identical in its interaction domains with species X, but it changed the Alexa 488 donor fluorophore for a Cy3 donor excited at 530 nm and measured at 585 nm. In all the experiments that involved Y the identity strands of the catalysts K and P were modified, changing the acceptor fluorophore for the Iowa Black FQ Quencher because in these experiments the focus of the characterisation of the system lies in seeing the net effect of the signal split rather than observing the evolution in Substrates and Catalysts.

All the strands with the extra specifications were ordered from IDT with an HPLC purification step and then prepared at their LabReady formulation (TE Buffer 1x at pH 8.0 and 100˚µM concentration of DNA) that after arrival were aliquoted and ready to use in the concentrations required for each experiment.

2.3.1. General experimental procedure

Once we obtained our DNA, we designed a general experimental procedure that worked as the basis for all the experimental work. This procedure was the one that was followed in all our experiments, unless when changes were noted. The general procedure to prepare our duplex species begins with preparing 30 and 60 µM dilutions of the strands that will not and will be in excess, respectively, out of the 100 µM stocks ordered from IDT. These strand dilutions are prepared following the concentrations detailed below:

-Non-excess strands (Per each 10 µL prepared): 3 µL of DNA 100 µM stock + 4.5 µL TAE Buffer 1x (Sigma Aldrich) + 2.5 µL of TAE buffer 1x Na^+ 1 M.

-Excess strands - including blockers - (Per each 10 µL prepared): 6 µL of DNA 100 µM stock + 1.5 µL TAE Buffer 1x (Sigma Aldrich) + 2.5 µL of TAE buffer 1x Na^+ 1 M.

In all the previous preparations, TAE 1X was prepared from a 25X concentrated stock (Sigma Aldrich). The 1x Na^+ 1 M stocks used for preparation were prepared adding 2.93 grams of NaCl (Sigma Aldrich) in 50 ml of an aliquot of TAE Buffer 1x. TAE Buffer 1x 250 mM for dilution of the stocks of duplexes as well as final experiments was also prepared diluting 0.98 g of NaCl in 50 ml of TAE Buffer 1x. Excesses strands with their corresponding blockers are specified in the supplementary info of the chapter

The general procedure to obtain 10 µM stocks of our duplex species is to mix 10 µL of the non-excess strand dilution with 10 µL of the excess strand dilution to later homogenize through serial re-pipetting and vortexing. After this, the duplexes are left to anneal for two hours at room temperature and added with 10 µL of the corresponding blocker strand, re-homogenize and left for two hours prior to the reaction to let the blockers act. When required, the 10 µM concentration of duplexes can be diluted into working stocks easier to handle in terms of volume (1 µM for the Target species and 100 nM for the Catalyst species are the general orders of magnitude, but this can vary depending on the specific experiment and will be noted when relevant). Timesteps of the measurements were adjusted for the specifications of the type of experiments as well as the gain of every emission fluorophore pair. For the single step experiments as well as the orthogonality tests for our design the measurement time step was 42 seconds. Variations of these timesteps depending on the type of experiment will be detailed throughout the thesis

After preparation, corresponding reaction mixtures were placed in the corresponding wells on 96-well plate with clear bottom (Corning) and adjusted to a final volume of 200 µl in all cases. The reactions tracing was measured in a ClarioSTAR plate reader (BMG Labtech) using the following specifications for each reading wavelength (in all the cases the measurement of the optics are taken from the bottom of the plate). All experiments were performed at 37 Celsius, as determined per the junction stability tests.

Alexa 488/Nothing: 520/15 nm (70 Flashes per second, 2000 gain)

Alexa 488/Cy3: 600/20 nm (70 Flashes per second, 2000 gain).

Alexa 488/Cy5: 690/30 nm (70 Flashes per second, 2500 gain).

Cy3/Nothing : 585/20 (70 Flashes per second, 2000 gain).

Chapter 3: Experimental implementation of a single catalytic reaction in the ACDC Framework. Results and considerations.

3.1-Single step reactions

3.1.1.- Initial considerations

The results obtained in the last section of chapter 2 allow us to consider that in principle, reversible CCNs are realisable within the ACDC Framework. Hence, we can test the single reaction steps for a given catalytic reaction of the push-pull cycle as well as the whole catalytic reactions themselves.

In order to perform single-step experiments, we proceed to assemble our species to a 10 nM stock using the protocol detailed in Section 2.3.1 of Chapter 2. The strands used to form the duplex species were the modified versions with the fluorophores attached, substituting some of the non-required external domains of the strands. With these modifications we can make fluorescence tracings of the substrate species X, not only detecting the active and the inactive state as major species, but also the ancillary intermediates KX and KP formed with the catalysts K and P. The general outline of the species that take part in each catalytic reaction can be found in Figures 39, 40, 41 and 42, respectively.

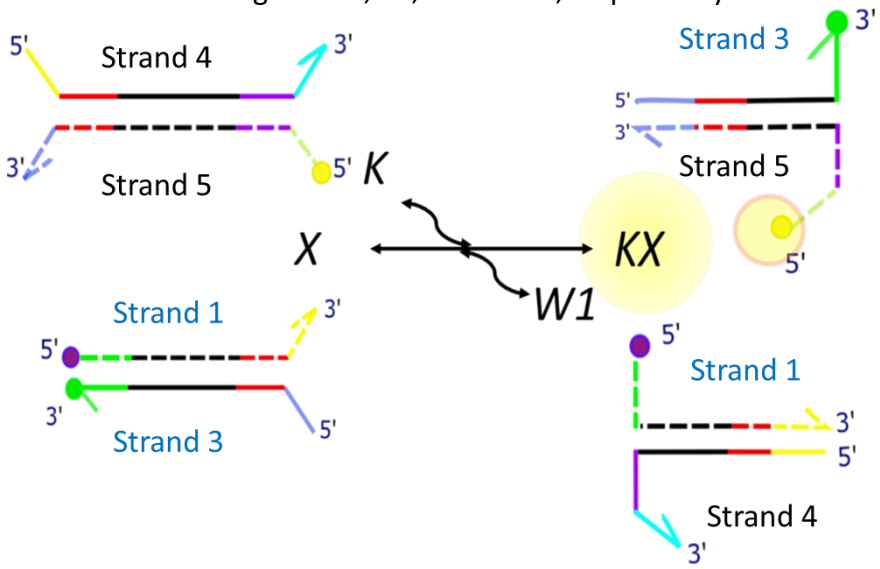

Figure 39. Diagram of the first step of the push reaction K +X ↔ KX +W1. The common central domain is depicted in black, while the toeholds are depicted in colour. Identity strands of both the substrate X and the catalyst K have fluorophores attached (Alexa 488 -green- in the case of the X identity strand and Cy3-yellow- in the case of K identity strand). The state strand of X that codes for it inactive state bears a quencher (Iowa Black FQ).

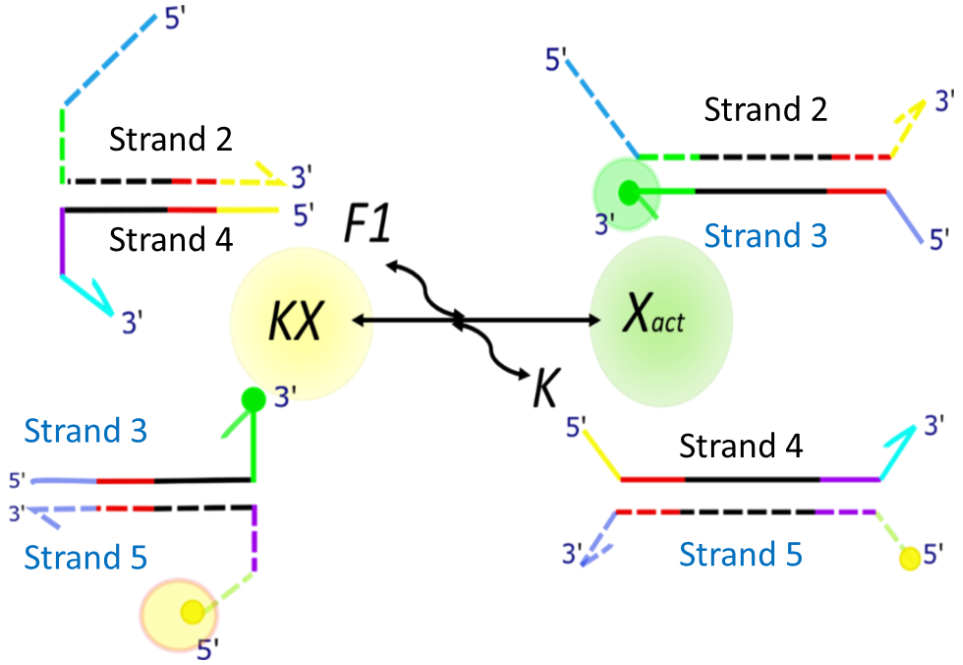

Figure 40. Diagram of the second step of the push reaction Kx +X ↔ X_{act} +K. The common central domain is depicted in black, while the toeholds are depicted in colour. Identity strands of both the substrate X and the catalyst K have fluorophores attached (Alexa 488 -green- in the case of the X identity strand and Cy3-yellow- in the case of K identity strand).

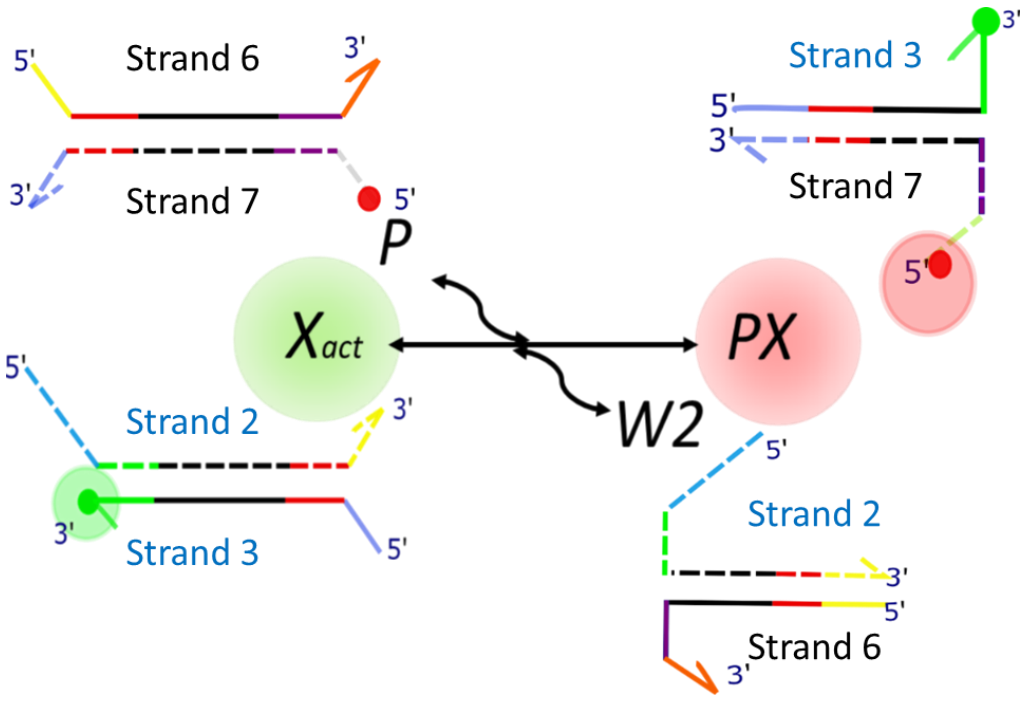

Figure 41 Diagram of the first step of the push reaction X_{act} +P ↔ PX + W2. The common central domain is depicted in black, while the toeholds are depicted in colour. Identity strands of both the substrate X and the catalyst K have fluorophores attached (Alexa 488 -green- in the case of the X identity strand and Cy5-red- in the case of K identity strand).

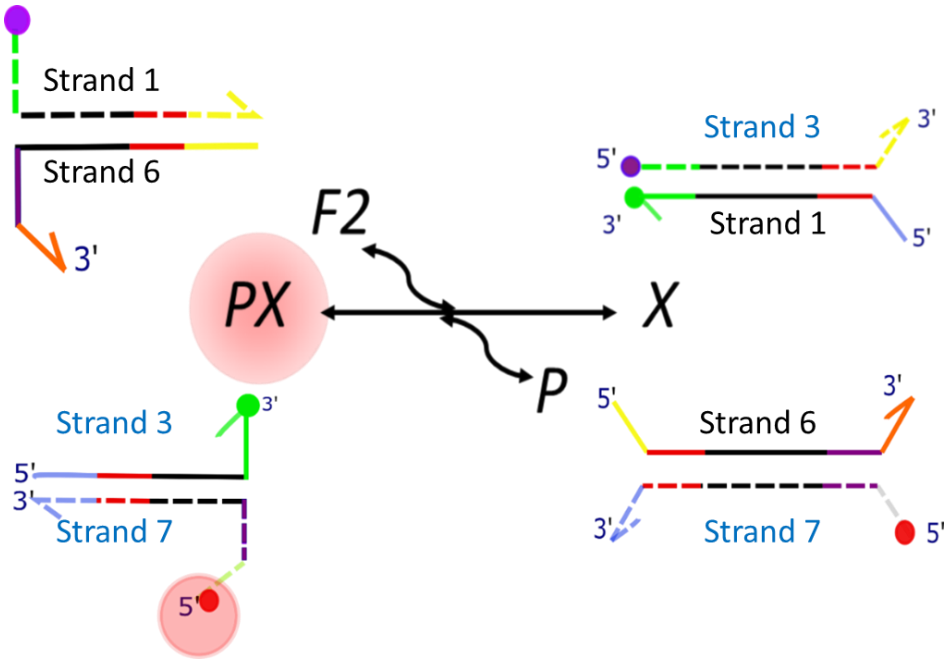

Figure 42. Diagram of the second step of the push reaction PX +F2 ↔ X +P. The common central domain is depicted in black, while the toeholds are depicted in colour. Identity strands of both the substrate X and the catalyst K have fluorophores attached (Alexa 488 -green- in the case of the X identity strand and Cy5-red- in the case of K identity strand). The state strand that encodes the inactive state of X bears a quencher (Iowa Black FQ) that suppresses all the emission in all wavelengths.

The two main features that we want to test in this round of experiments are the reaction rates for the reaction steps as well as the intended reversible nature of the reaction steps. Such features were tested in experiments in which a fixed concentration of a given substrate or reaction intermediate species reacts with a variable amount of the catalyst or the appropriate ancillary fuel/waste (see Figures 43, 44, 45 and 46). When looking at both the reaction of the substrate with the catalyst as well as the backwards reaction between the products, we can see if, as we intended, the reaction is reversible.

In each case fluorescence tracing was performed in two different fluorescence channels, one corresponding to the signal given by the FRET donor placed in the identity strand of the identity strand of the substrate X (Alexa 488, which as per the general protocol we described in Section 2.3.1 we excite at of 465 nm and for which we have settled the readout wavelength at 520 nm) and the other one corresponding to the signal of the FRET acceptor for the identity strand of the catalyst that is used in each (Cy3 for the FRET of the activation of the substrate using the catalyst K – for which we have a read and Cy3 for the FRET of the deactivation of the substrate using the catalyst P). In all cases the measurement time step employed was of 42 seconds with 430 total measurements taken for each experiment, giving 5 hours of measurements for each reaction.

3.1.2.- Experimental results

As stated in subsection 3.1.1, each single reaction was observed in both possible directions and its temporal evolution was observed via changes in the fluorescence. e

observed – as seen in the Figures 43 to 50 – that the reactions occurred in both directions as intended, e.g when we had X and K, initially, we observed how the green fluorescence emission as well as the FRET signal corresponding to the formation of KX increased; and when we began with KX and added W1 we observed the opposite reaction happening. However, we also observed that, despite this reversibility, some small biases that favoured one reaction direction to another – even in reactions that in principle, should relax to the same equilibrium endpoint.

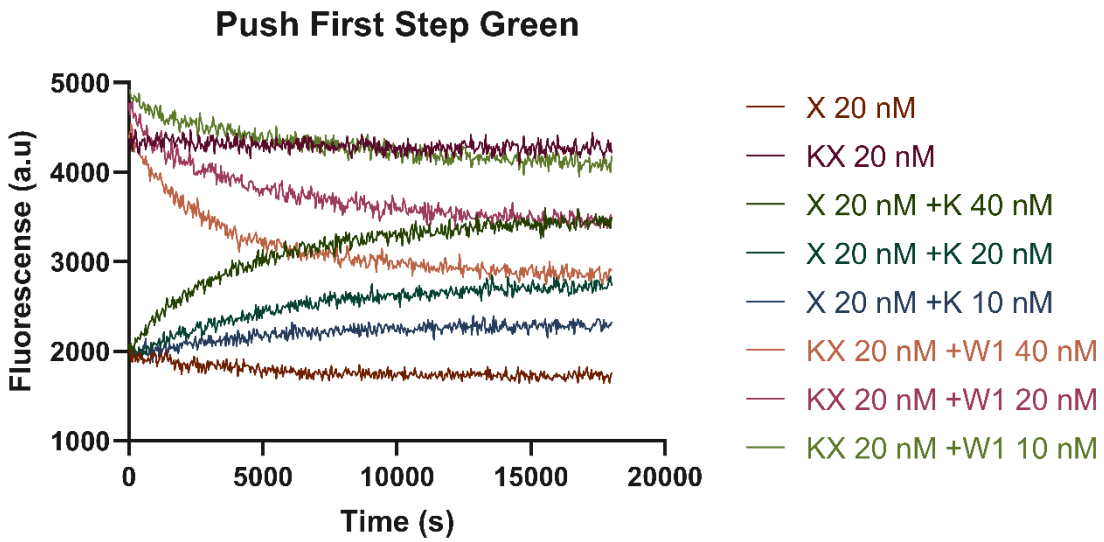

Figure 43. Fluorescence tracing of the first reaction step of the push reaction when observed on the 520 nm fluorescence channel corresponding to the emission of the Alexa 488 fluorophore. In the present graph, the lower fluorescence level corresponds with X 20 nM while the higher levels correspond to KX 20 nM. The rising curves correspond with the X + K reacting with each other to form KX and W1 while the falling ones correspond with the reaction of KX and W1 to form back K and X.

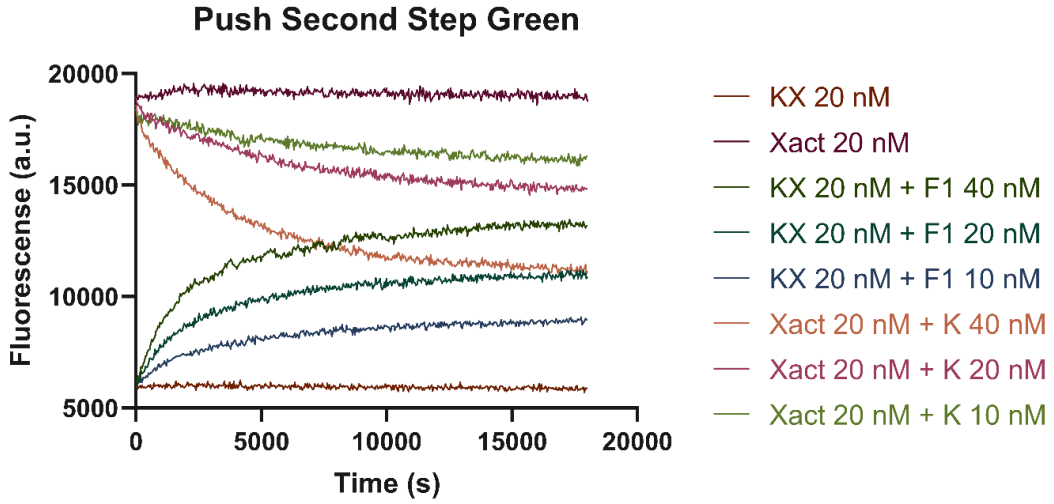

Figure 44. Fluorescence tracing of the second reaction step of the push reaction when observed on the 520 nm fluorescence channel corresponding to the emission of the Alexa 488 fluorophore. In the present graph, the lower fluorescence level corresponds with KX 20 nM while the higher levels correspond to Xact 20 nM. The rising curves correspond with the KX + F1 reacting with each other to form Xact and K while the falling ones correspond with the reaction of K and Xact to form back KX and F1.

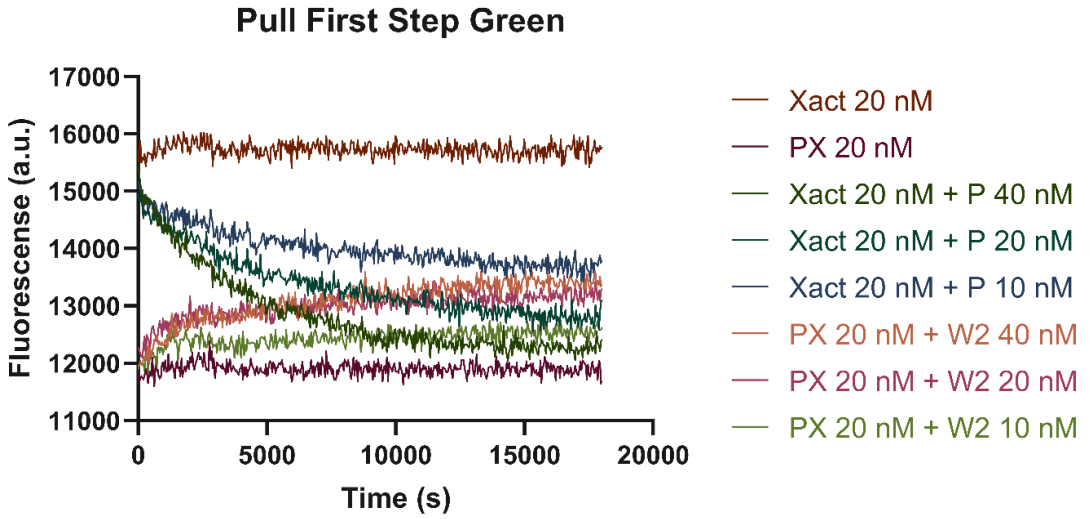

Figure 45. Fluorescence tracing of the first reaction step of the pull reaction when observed on the 520 nm fluorescence channel corresponding to the emission of the Alexa 488 fluorophore. In the present graph, the lower fluorescence level corresponds with PX 20 nM while the higher levels correspond to Xact 20 nM. The falling curves correspond with the Xact + P reacting with each other to form PX and W2 while the rising ones correspond with the reaction of PX and W2 to form back P and Xact.

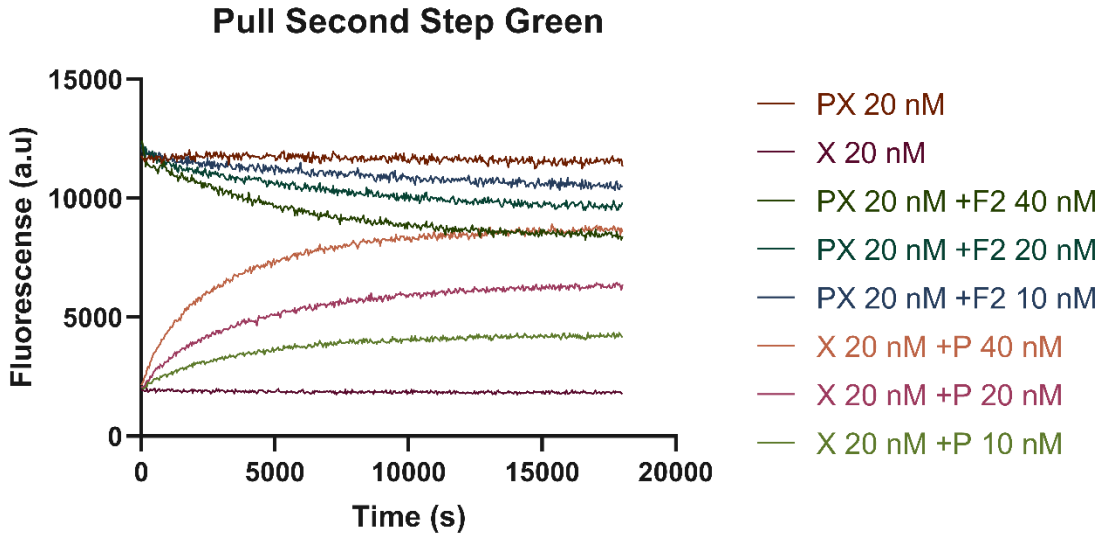

Figure 46. Fluorescence tracing of the second reaction step of the pull reaction when observed on the 520 nm fluorescence channel corresponding to the emission of the Alexa 488 fluorophore. In the present graph, the lower fluorescence level corresponds with X 20 nM while the higher levels correspond to PX 20 nM. The falling curves correspond with the PX + F2 reacting with each other to form X and P while the rising ones correspond with the reaction of X and P to form back PX and F2.

However, when analysing the single step reactions in the FRET channel, we observed shifted backgrounds in certain sets of reactions that made the results harder to interpret. More precisely, we observed this shift in reactions that involved the addition of different amounts of the catalyst K when observed on the channel corresponding to the FRET signal emitted by the KX intermediate. This shift on the baseline was lineally dependent on the quantity of K added as we can observe in Figures 46 and 47, which led us to conclude that

such background effect was due to direct excitation of the Cy3 fluorophore present on Strand 5 by the wavelength supposed to excite only the Alexa 488 donor.

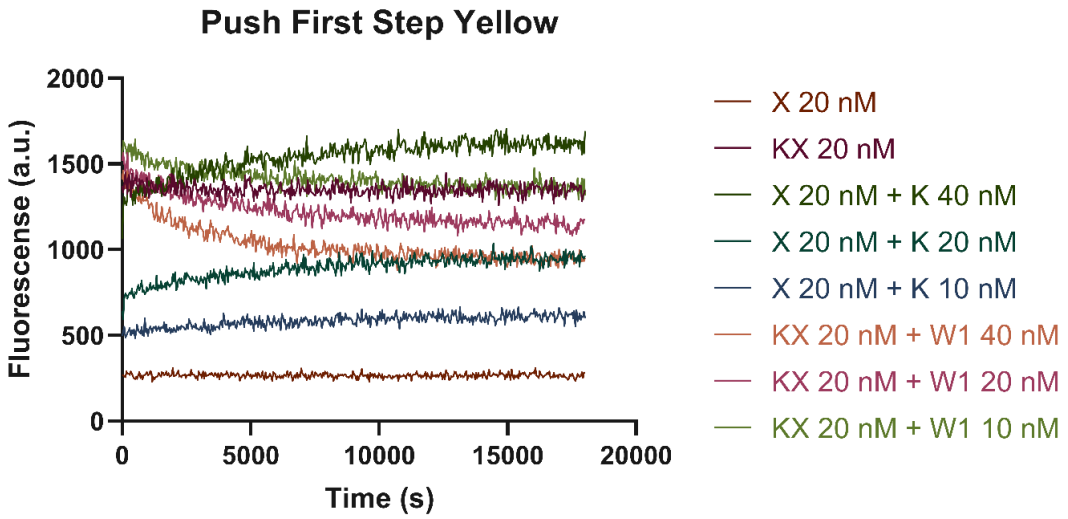

Figure 47. Fluorescence tracing of the first reaction step of the push reaction when observed on the 600 nm fluorescence channel corresponding to the emission of the Alexa 488-Cy3 FRET pair at 600 nm. In the present graph, the lower fluorescence level corresponds with X 20 nM while the higher levels correspond to KX 20 nM. The rising curves correspond with the X + K reacting with each other to form KX and W1 while the falling ones correspond with the reaction of KX and W1 to form back K and X. Effects of the indirect activation of the Cy3 fluorophore can be noted in the rising K +X curves.

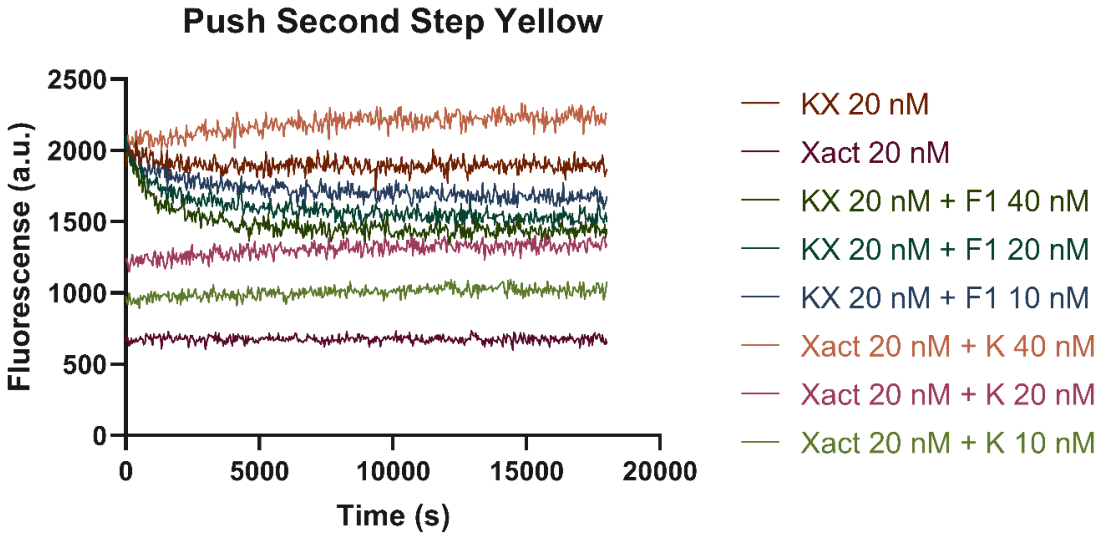

Figure 48. Fluorescence tracing of the second reaction step of the push reaction when observed on the 600 nm fluorescence channel corresponding to the emission of the Alexa 488-Cy3 FRET pair at 600 nm. In the present graph, the lower fluorescence level corresponds with Xact 20 nM while the higher levels correspond to KX 20 nM. The rising curves correspond with the X + K reacting with each other to form KX and F1 while the falling ones correspond with the reaction of KX and F1 to form back K and X. Effects of the indirect activation of the Cy3 fluorophore can be noted in the rising K +X curves.

In principle this can be accounted for by means of including a control with a known quantity of Strand 5 blocked, as well as just K on its own. Knowing the average increase per nanomole added in the background, we can properly assess this when studying fluorescence tracing and hence make comparable assessments of the evolution of the reaction.

This direct excitation problem, however, was not observed on the pull catalyst experiments, as we can see in Figures 49 and 50 due to the fact that there is no direct excitation of the red fluorophore in this FRET pair. In these figures we can observe that, in the case of Figure 49 the tracing profile was the inverse of that seen on Figure 45, while it was directly analogous between Figures 46 and 50. In both cases, the signal was noisier and more compressed than in the green channel due to the fact that FRET efficiency is never complete and non-radiative quenching phenomena occur during the process.

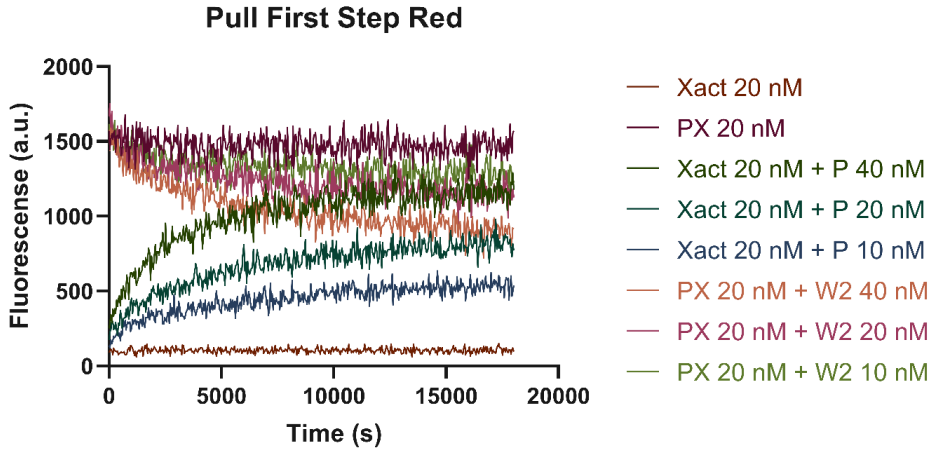

Figure 49. Fluorescence tracing of the first reaction step of the pull reaction when observed on the 600 nm fluorescence channel corresponding to the emission of the Alexa 488-Cy5 FRET pair at 690 nm. In the present graph, the lower fluorescence level corresponds with Xact 20 nM while the higher levels correspond to PX 20 nM. The rising curves correspond with the Xact + K reacting with each other to form PX and W2 while the falling ones correspond with the reaction of PX and W2 to form back P and X_{act}.

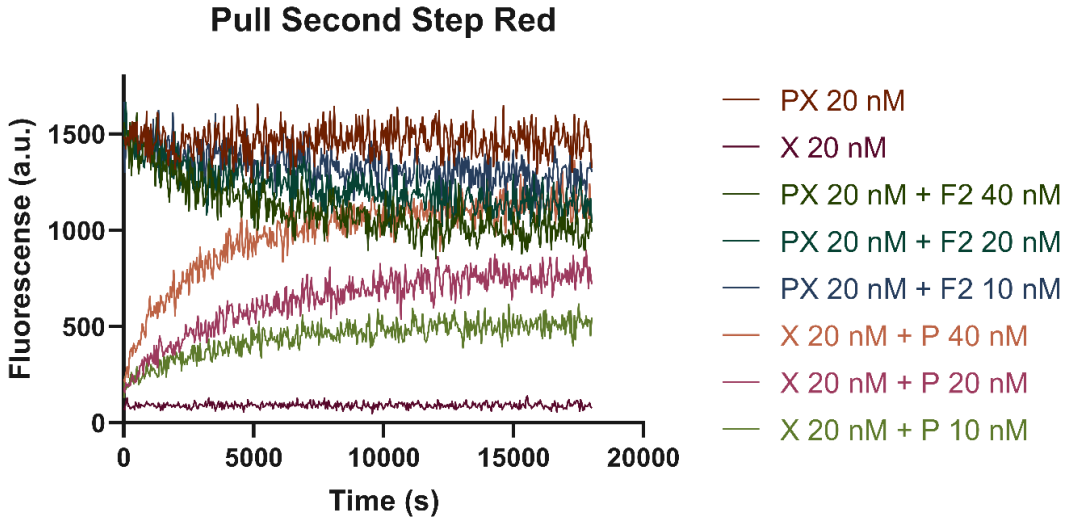

Figure 50. Fluorescence tracing of the second reaction step of the pull reaction when observed on the 600 nm fluorescence channel corresponding to the emission of the Alexa 488-Cy3 FRET pair at 690 nm. In the present graph, the lower fluorescence level corresponds with X 20 nM while the higher levels correspond to KX 20 nM. The rising curves correspond with the X + KP reacting with each other to form PX and W2 while the falling ones correspond with the reaction of PX and F2 to form P and X.

It might seem that these apparent reaction endpoint asymmetries might pose a risk to the working of the system since they could impose a small bias on the reaction direction. However, such risk can be discarded with a reasonable certainty if we are able to demonstrate that the reaction rates for both directions are comparable. If such is the case, that would entail that, as we intended from the very beginning, the catalytic process can be driven purely by concentration imbalances and hence be entropically driven. This desired equal range on reaction rates in both directions also entails that the enthalpic free energy differences between the different states that comprise a reaction are small and can be overcome by the entropic part of the free energy by providing relatively small concentration imbalances.

The fact that both possible reaction directions occur is a very strong indicator that our designed reactions don't have an impossible-to-overcome bias, since we have observed that both reaction directions for each given substep are kinetically accessible. However, rigorous modelling of the system and fitting of the experimental data previously shown can help substantiate this case and explore some of the causes of this apparent imbalance.

3.1.3.- Data fitting.

As we stated in the previous subsection of the chapter, rigorous treatment of our data will allow us to draw more substantiated conclusions about our system. First, we should consider a model that properly captures the phenomena we are studying. For the sake of simplicity, we're going to consider that the mathematical model we build for the first reaction step in the forward direction can be applied for the rest of the reaction steps of the system and directions of reaction, since their behaviour is meant to be the same given that all the single step reactions have the same topology and the toeholds that mediate them are the same length in both directions.

Hence, we are considering the single step reaction (1)

$$K + X \leftrightarrow KX + W \tag{1}$$

In this case, we are monitoring the increase or decrease in the fluorescence due to the formation or disassembly of the KX intermediate. Hence, we consider the initial ODE (2)

$$\frac{d[KX]}{dt} = \gamma_1[K][X] - \gamma_2[KX][W1], \tag{2}$$

with γ_1 and γ_2 being the forwards and backwards reaction rates. This particular case of the ODE describes the temporal evolution of the single step reaction and can be reformulated to express all terms in function of only the variable [KX] by making some assumptions. The first assumption consists of establishing the conservation laws for K (3) and X (4):

$$[\mathrm{K}] = [\mathrm{K}]_{\mathrm{total}} - [\mathrm{KX}] \tag{3}$$

and

$$[X] = [X]_{total} - [KX] \tag{4}$$

In which the terms $[K]_{total}$ and $[X]_{total}$ are constants that are determined by the initial conditions of the experiment.

The second simplifying assumption is expressed via equation (5), which reflects the fact that the concentrations of [KX] and [W1] are equal since they have the same initial concentration and are produced as a pair in the reaction. We can make this assumption due to the fact that the model is only taking into account a single step of the catalytic turnover. However, this assumption does not work in case we wanted to model the whole catalytic turnover process.

$$[KX] = [W1] \tag{5}$$

Using (3), (4) and (5), we can turn Equation (2) into:

$$\frac{d[KX]}{dt} = \gamma_1([K]_{total} - [KX])([X]_{total} - [KX]) - \gamma_2[KX]^2 \tag{6}$$

Then expanding and rearranging we arrive at Equation (7):

$$\frac{d[KX]}{dt} = (\gamma_1 - \gamma_2)[KX]^2 - \gamma_1([K]_{total} + [X]_{total})[KX] + \gamma_1([K]_{total}[X]_{total}). \tag{7}$$

This ODE can be solved analytically after a rearrangement step in which we group d[KX] with the [KX] terms and extract $\frac{1}{(\gamma_1-\gamma_2)}$ as a common term. Doing so, we get Equation 8:

$$dt = \frac{1}{(\gamma_1 - \gamma_2)}\left[\frac{d[KX]}{[KX]^2 - \frac{\gamma_1}{(\gamma_1 - \gamma_2)}([K]_{total} + [X]_{total})[KX] + \frac{\gamma_1}{(\gamma_1 - \gamma_2)}([K]_{total}[X]_{total})}\right] \tag{8}$$

The polynomial on the denominator has the shape of a quadratic equation of the form, $ax^2 + bx + c$ with $x = \frac{-b \pm \sqrt{b^2-4ac}}{2a}$. Solving this quadratic equation allows us to decompose the fraction into two simple integrable fractions of the type $\frac{A}{\alpha[KX]+\beta} + \frac{B}{\delta[KX]+\varepsilon}$.

Figuring out the values of the coefficients with the help of the quadratic equation and subsequently integrating each fraction separately results in t being equal to a sum of logarithms. Further rearrangement to place [KX] as the output of the integrated function yields equation (9)

$$[KX] = \frac{([K]_{total}[X]_{total})}{(\gamma_1 - \gamma_2)} \frac{e^{-2H(\gamma_1-\gamma_2)} - 1}{J + H - (J - H)e^{-2H(\gamma_1-\gamma_2)}} \tag{9}$$

where

$$J = \frac{\gamma_1}{2(\gamma_1-\gamma_2)}([K]_{total} + [X]_{total}) \tag{10}$$

and

$$H = \sqrt{J^2 - \frac{\gamma_1}{(\gamma_1-\gamma_2)}([K]_{total}[X]_{total})} \tag{11}$$

With this model already built and with the collaboration of my supervisor Dr Thomas E. Ouldridge, we attempted to fit the experimental data traces to the trajectories predicted by the model via mean square differences minimization using the MATLAB function *fminsearch*.

However, the fitting step requires the conversion of the concentrations predicted by the model into the corresponding fluorescence signal. We perform this conversion using the equation 12

$$F = F_{\text{background}} + \alpha_A[A] + \alpha_B[B] + \alpha_C[C] + \alpha_D[D]. \tag{12}$$

Here A, B, C and D are the different species taking part in a reaction of the type

$$A + B \rightleftharpoons C + D \tag{13}$$

while the α coefficients (α_A, α_B, etc.) are the associated multiplier factors for each species at a given wavelength of interest.

The signal $F_{background}$ is composed of a signal F_{blank} that is taken from the blank buffer control and the signal corresponding to the excess of labelled blocked inert strands (which only was noticeable in the case of KX and K) following the relationship

$$F_{\text{background}} = F_{\text{blank}} + \alpha_{\text{blocked}}[\text{blocked}], \tag{14}$$

where

$$\alpha_{\text{blocked}} = \frac{\varphi_{\text{blocked}} - \varphi_{\text{blank}}}{[\text{blocked control}]} \tag{15}$$

in which the [blocked control] concentrations are 20nM in all experiments and where the φ factors, φ_{blocked} and φ_{blank}, are the average fluorescence over time of the blocked strands and the buffer, respectively.

The multipliers for the different species taking part in the general reaction were determined using equation (16) where, following our previously introduced notation, $\varphi_{\text{X control}}$ is the average fluorescence over time of the X 20 nM control.

$$\alpha_{\text{X}}^{\text{default}} = \frac{\varphi_{\text{X control}} - \varphi_{\text{blank}}}{[\text{X control}]} \tag{16}$$

The only exceptions were for the case of K and KX. Here, the control volume contained an equal concentration of blocked fluorescent strand; a correction was calculated as stated in equation (17) for K:

$$\alpha_{\text{K}}^{\text{default}} = \frac{\varphi_{\text{K control}} - \varphi_{\text{blank}} - \alpha_{\text{blocked}}[\text{K control}]}{[\text{K control}]} \tag{17}$$

However, due to experimental variations caused by time delays between injection on the sample and the onset of the measurement that tried to be minimized, the initial fluorescence signal might have noticeable lower or bigger signal than the controls employed. Hence, the default α multipliers had to be changed and re-scaled by using the expression used in equation (18) (except again, in the cases of K and KX in which the fluorescent strand 5 bound to the blocker strand was present and the modification of the multiplier was of the shape of Equation (19)).

$$\alpha_X^i = \alpha_{\text{X}}^{\text{default}} \frac{\varphi_{X\text{ initial}}^i - \varphi_{\text{blank initial}}^i}{\varphi_{X\text{ control initial}}^i - \varphi_{\text{blank initial}}^i}. \tag{18}$$

$$\alpha_K^i = \alpha_{\text{K}}^{\text{default}} \frac{\varphi_{K\text{ initial}}^i - \alpha_{\text{blocked initial}}[\text{K control}] - \varphi_{\text{blank initial}}^i}{\varphi_{K\text{ control initial}}^i - \alpha_{\text{blocked initial}}[\text{K control}] - \varphi_{\text{blank initial}}^i} \tag{19}$$

The initial model was modified for each direction and step and the fit of the experimental data was done for all cases in all fluorescence channels. The graphs for all the fittings are found in Figures 51 to 74 while the values of the reaction rates calculated by the fitting are found in Table 1. For the first third part of all the trajectories, the squared error was

multiplied by 3 and then summed in order to give these parts of the trajectories – where most of the dynamics happen – a larger weight for the fitting.

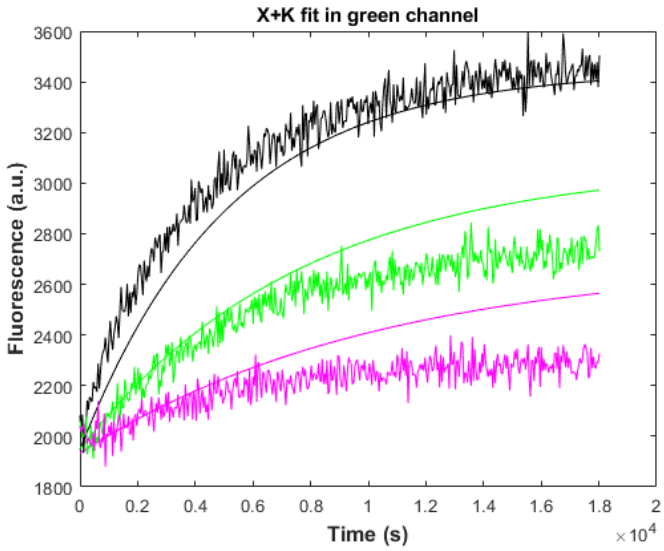

Figure 51. Fitting of the first reaction step of the push in the forwards direction with the data observed in the green channel. Time units in the X axis are seconds. In the figure, the solid lines correspond to the fittings while the rugged lines correspond to the actual experimental fluorescence tracings. The black lines stand for the X 20 nM + K 40 nM reaction, while the green curves stand for the X 20 nM + K 20 nM reaction and the magenta one for the X 20 nM + K 10 nM reaction. The fitting was done only for the forward reaction rate (γ_1 in our general model).

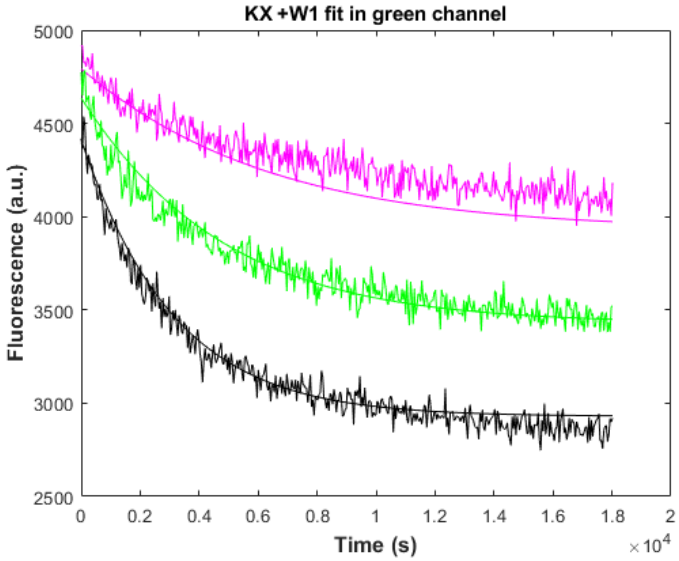

Figure 52. Fitting of the first reaction step of the push in the backwards direction with the data observed in the green channel. Time units in the X axis are seconds. In the figure, the solid lines correspond to the fittings while the rugged lines correspond to the actual experimental fluorescence tracings. The black lines stand for the KX 20 nM + W1 40 nM reaction, while the green curves stand for the KX 20 nM + W1 20 nM reaction and the magenta one for the KX 20 nM + W1 10 nM reaction. The fitting was done only for the backwards reaction rate (γ_2 in our general model)

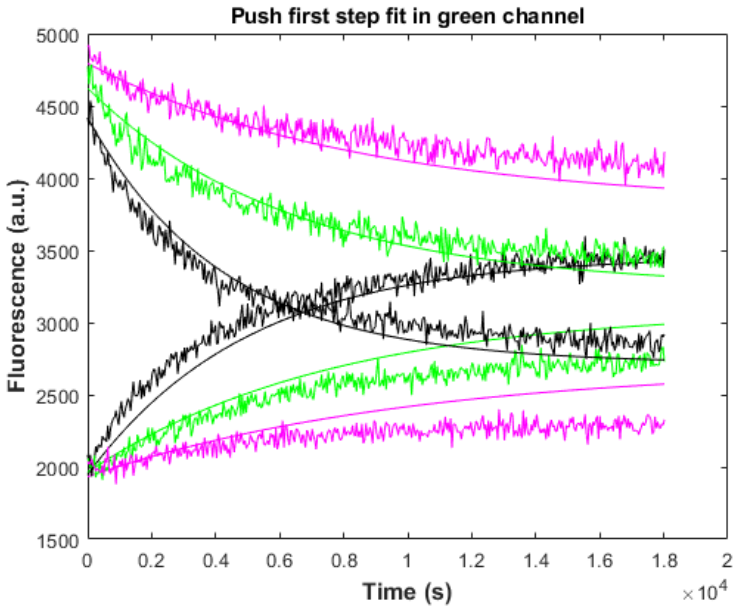

Figure 53. Fitting of the first reaction step of the push in both directions with the data observed in the green channel. Time units in the X axis are seconds. In the figure, the solid lines correspond to the fittings while the rugged lines correspond to the actual experimental fluorescence tracings. Rising curves correspond with the forward reaction K+X, while the decreasing curves correspond with the backwards reaction KX + W1. Colour codes are the same than those from figure 50 and 51. The fitting was done for both reaction rates.

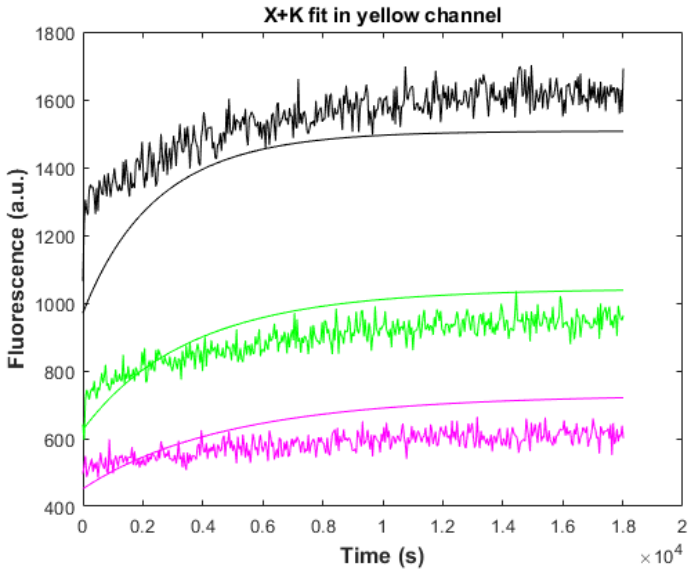

Figure 54. Fitting of the first reaction step of the push in the forward direction with the data observed in the yellow channel. Time units on the X axis are seconds. In the figure, the solid lines correspond to the fittings while the rugged lines correspond to the actual experimental fluorescence tracings. The black lines stand for the X 20 nM + K 40 nM reaction, while the green curves stand for the X 20 nM + K 20 nM reaction and the magenta one for the X 20 nM + K 10 nM reaction. The fitting was done only for the forward reaction rate.

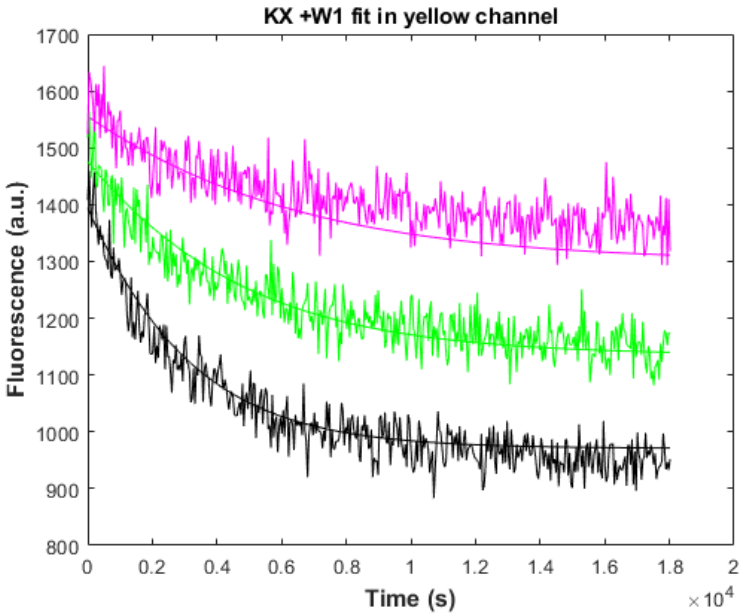

Figure 55. Fitting of the first reaction step of the push in the backward direction with the data observed in the yellow channel. Time units on the X axis are seconds. In the figure, the solid lines correspond to the fittings while the rugged lines correspond to the actual experimental fluorescence tracings. The black lines stand for the KX 20 nM + W1 40 nM reaction, while the green curves stand for the KX 20 nM + W1 20 nM reaction and the magenta one for the KX 20 nM + W1 10 nM reaction. The fitting was done only for the backwards reaction rate.

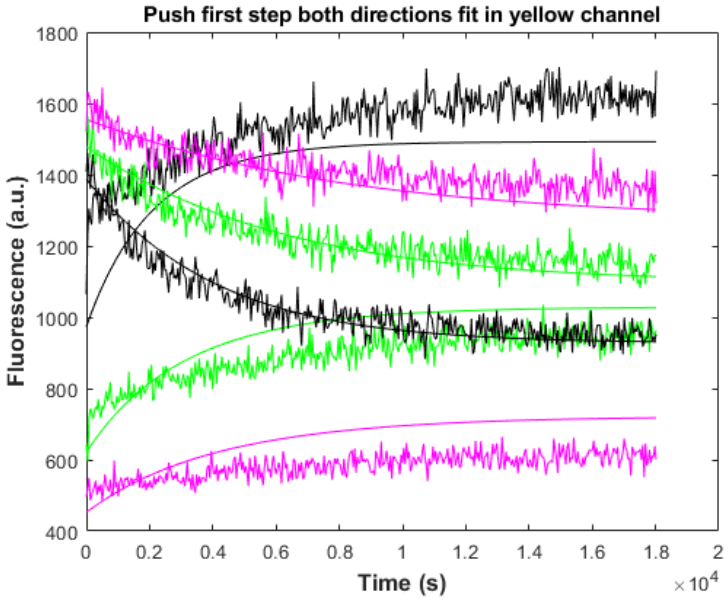

Figure 56. Fitting of the first reaction step of the push in both directions with the data observed in the yellow channel. Time units on the X axis are seconds. In the figure, the solid lines correspond to the fittings while the rugged lines correspond to the actual experimental fluorescence tracings. Rising curves correspond with the forward reaction K+X, while the decreasing curves correspond with the backwards reaction KX + W1. Colour codes are the same than those from figure 53 and 54 The fitting was done for both reaction rates.

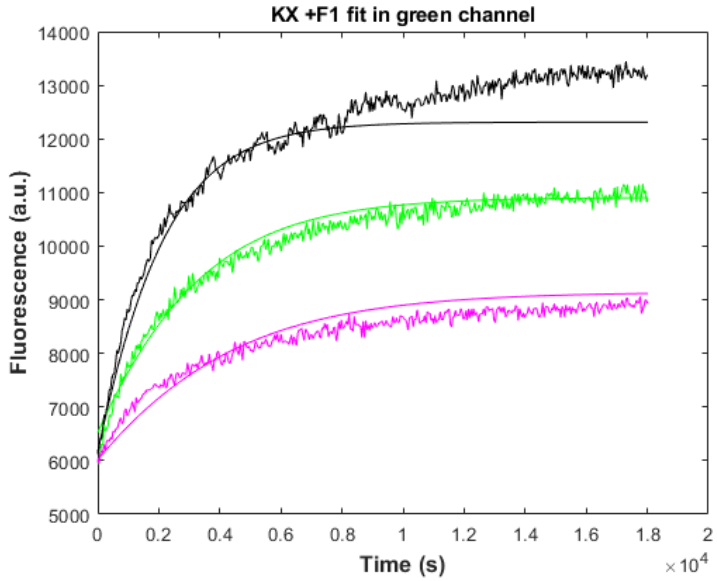

Figure 57. Fitting of the second reaction step of the push in the forward direction with the data observed in the green channel. Time units on the X axis are seconds. In the figure, the solid lines correspond to the fittings while the rugged lines correspond to the actual experimental fluorescence tracings. The black lines stand for the KX 20 nM + F1 40 nM reaction, while the green curves stand for the KX 20 nM + F1 20 nM reaction and the magenta one for the KX 20 nM + F1 10 nM reaction. The fitting was done only for the forward reaction rate.

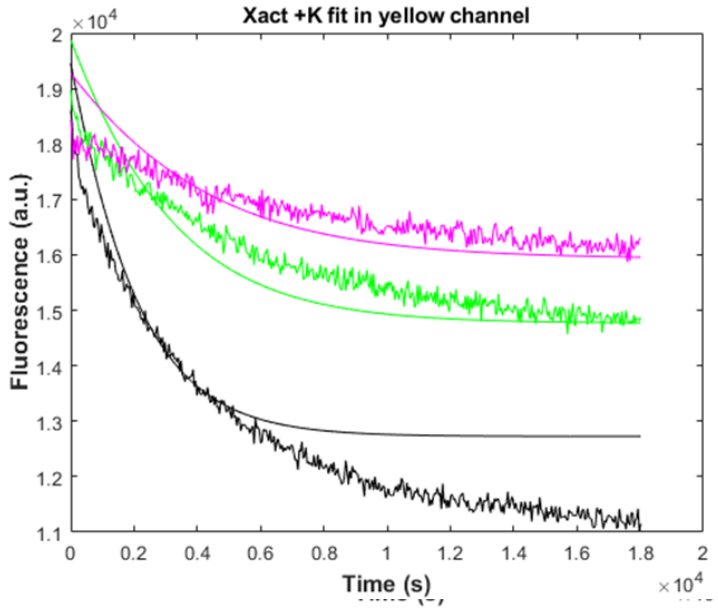

Figure 58. Fitting of the second reaction step of the push in the backwards direction with the data observed in the green channel. Time units on the X axis are seconds. In the figure, the solid lines correspond t0 the fittings while the rugged lines correspond to the actual experimental fluorescence tracings. The black lines stand for the Xact 20 nM + K 40 nM reaction, while the green curves stand for the Xact 20 nM + K 20 nM reaction and the magenta one for the Xact 20 nM + K 10 nM reaction. The fitting was done only for the backward reaction rate.

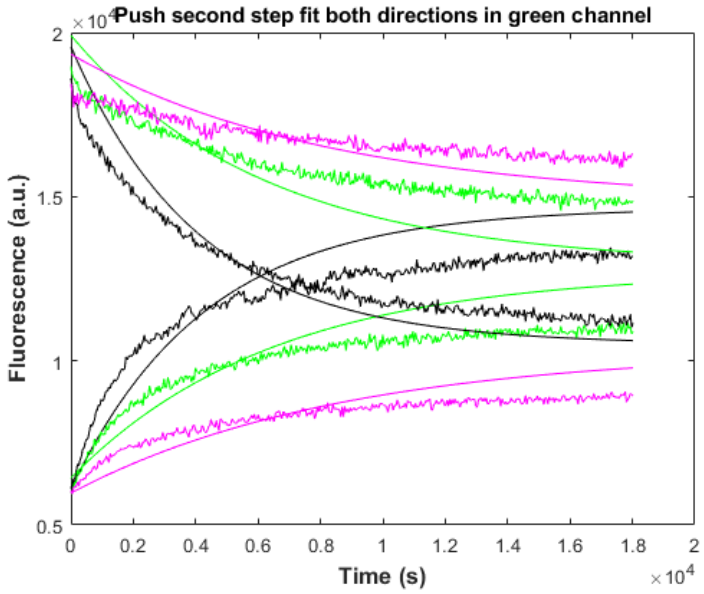

Figure 59. Fitting of the second reaction step of the push in both directions with the data observed in the green channel. Time units on the X axis are seconds. In the figure, the solid lines correspond to the fittings while the rugged lines correspond to the actual experimental fluorescence tracings. Rising curves correspond with the forward reaction KX +F1, while the decreasing curves correspond with the backwards reaction Xact + K. Colour codes are the same than those from figure 50 and 51 The fitting was done for both reaction rates.

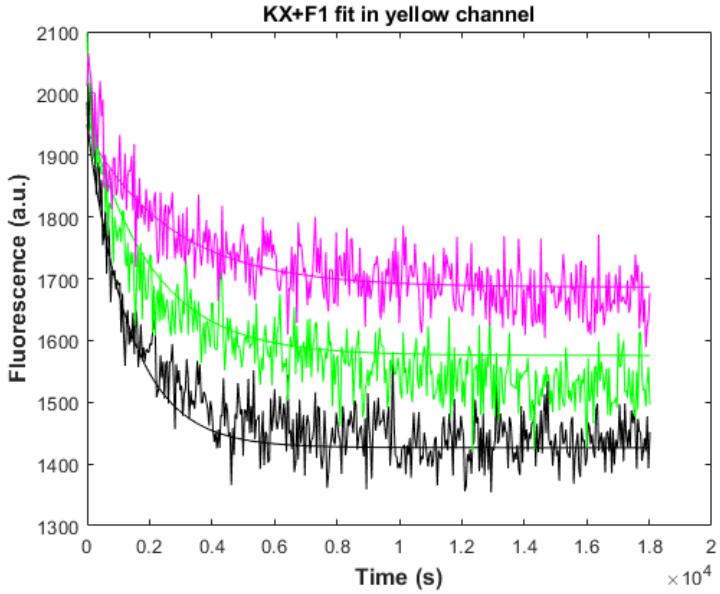

Figure 60. Fitting of the second reaction step of the push in the forward direction with the data observed in the yellow channel. Time units on the X axis are seconds. In the figure, the solid lines correspond to the fittings while the rugged lines correspond to the actual experimental fluorescence tracings. The black lines stand for the KX 20 nM + F1 40 nM reaction, while the green curves stand for the KX 20 nM + F1 20 nM reaction and the magenta one for the KX 20 nM + F1 10 nM reaction. The fitting was done only for the forward reaction rate.

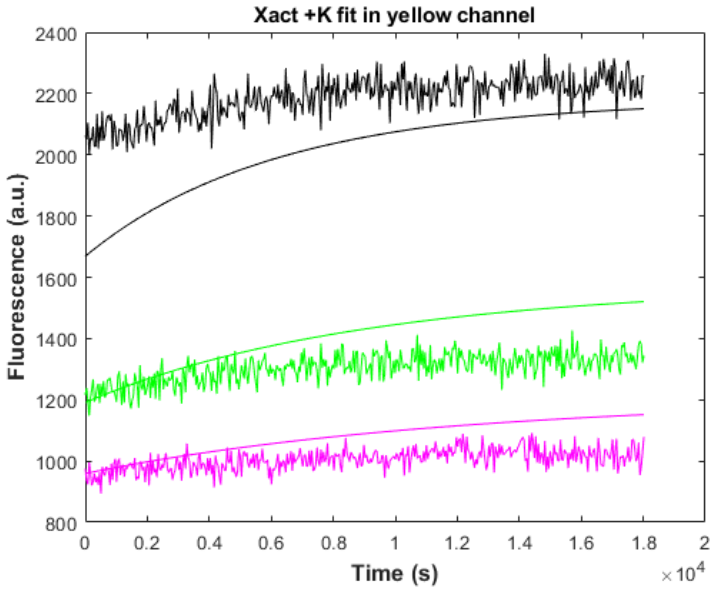

Figure 61. Fitting of the second reaction step of the push in the backwards direction with the data observed in the yellow channel. Time units on the X axis are seconds. In the figure, the solid lines correspond to the fittings while the rugged lines correspond to the actual experimental fluorescence tracings. The black lines stand for the Xact 20 nM + K 40 nM reaction, while the green curves stand for the Xact 20 nM + K 20 nM reaction and the magenta one for the Xact 20 nM + K 10 nM reaction. The fitting was done only for the backward reaction rate.

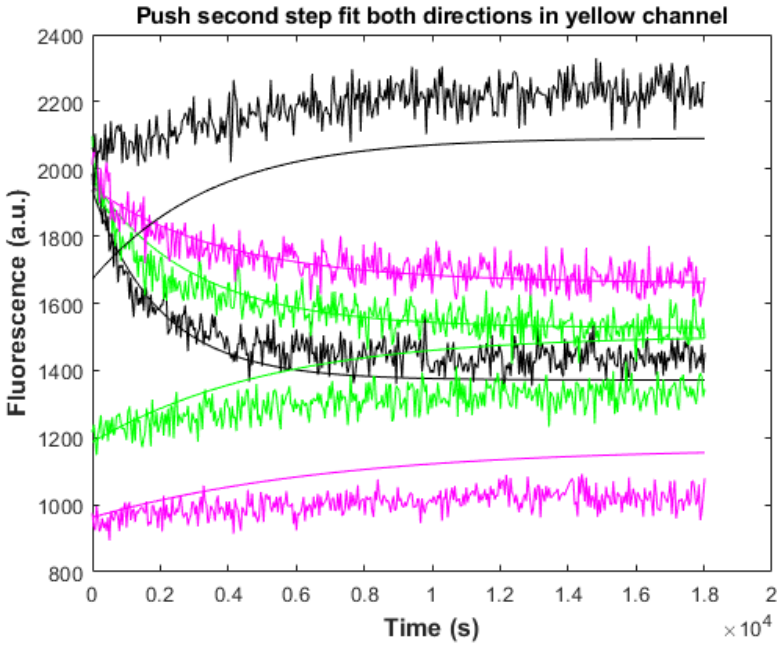

Figure 62. Fitting of the second reaction step of the push in both direction with the data observed in the yellow channel. Time units on the X axis are seconds. In the figure, the solid lines correspond to the fittings while the rugged lines correspond to the actual experimental fluorescence tracings. Rising curves correspond with the backward reaction K +Xact, while the decreasing curves correspond with the forward reaction KX+ F1. Colour codes are the same than those from Figures 58 and 59 The fitting was done for both reaction rates.

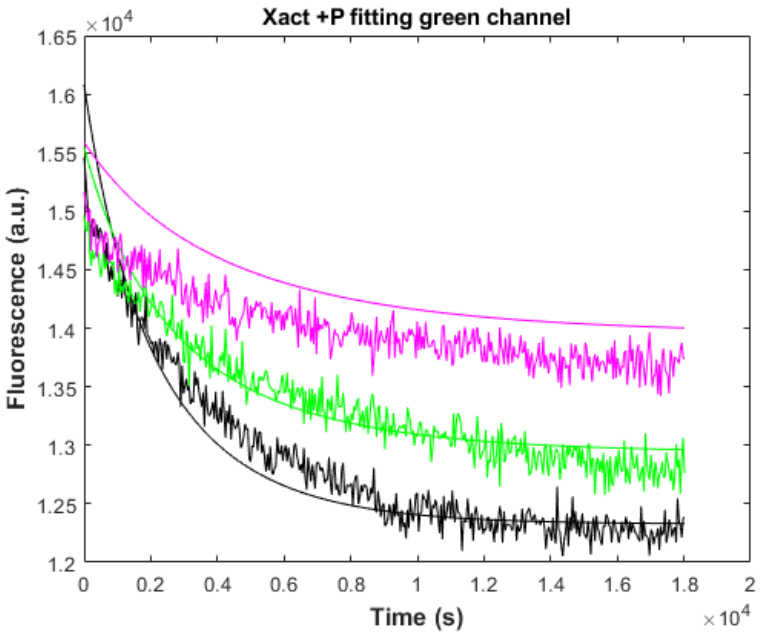

Figure 63. Fitting of the first reaction step of the pull in the forwards direction with the data observed in the green channel. Time units on the X axis are seconds. Units on the Y axis are M x 10^{-8}. In the figure, the solid lines correspond to the fittings while the rugged lines correspond to the actual experimental fluorescence tracings. The black lines stand for the Xact 20 nM + P 40 nM reaction, while the green curves stand for the Xact 20 nM + P 20 nM reaction and the magenta one for the Xact 20 nM + P 10 nM reaction. The fitting was done only for the forward reaction rate.

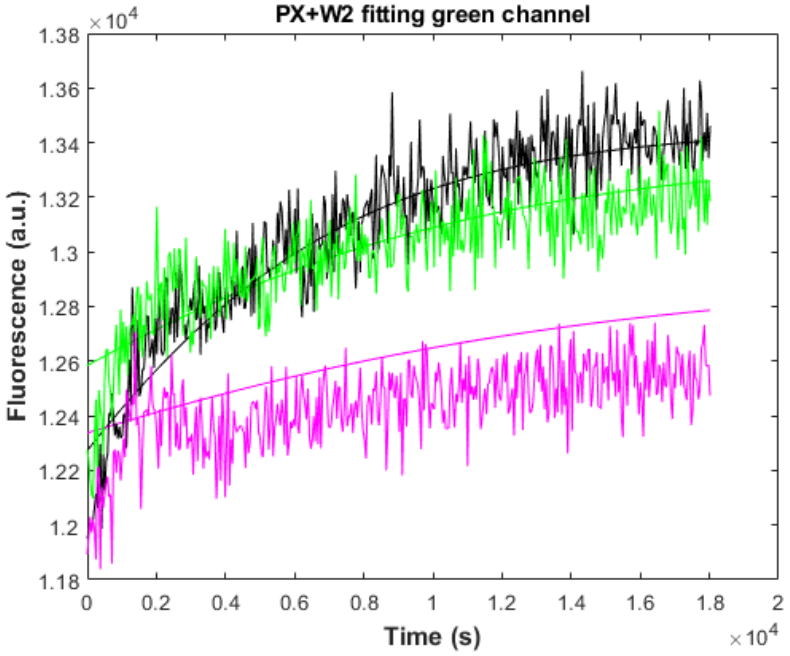

Figure 64. Fitting of the first reaction step of the pull in the backwards direction with the data observed in the green channel. Time units on the X axis are seconds. Units on the Y axis are M x 10^{-8}. In the figure, the solid lines correspond to the fittings while the rugged lines correspond to the actual experimental fluorescence tracings. The black lines stand for the PX 20 nM + W2 40 nM reaction, while the green curves stand for the PX 20 nM + W2 20 nM reaction and the magenta one for the PX 20 nM + W2 10 nM reaction. The fitting was done only for the backward reaction rate.

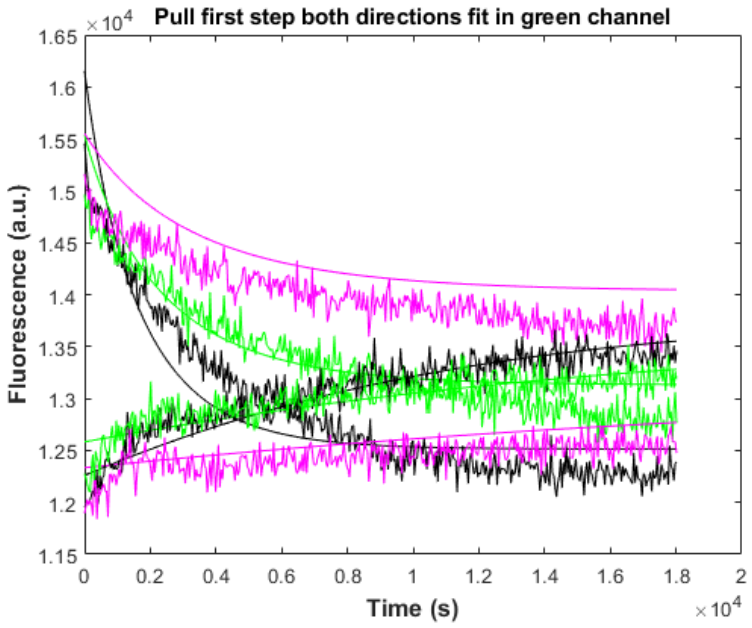

Figure 65. Fitting of the first reaction step of the pull in both directions with the data observed in the green channel. Time units on the X axis are seconds. Units on the Y axis are M x 10^{-8}. In the figure, the solid lines correspond to the fittings while the rugged lines correspond to the actual experimental fluorescence tracings. Rising curves correspond to the backward reaction PX +W2, while the decreasing curves correspond to the backwards reaction Xact + K. Colour codes are the same than those from figure 61 and 62 The fitting was done for both reaction rates.

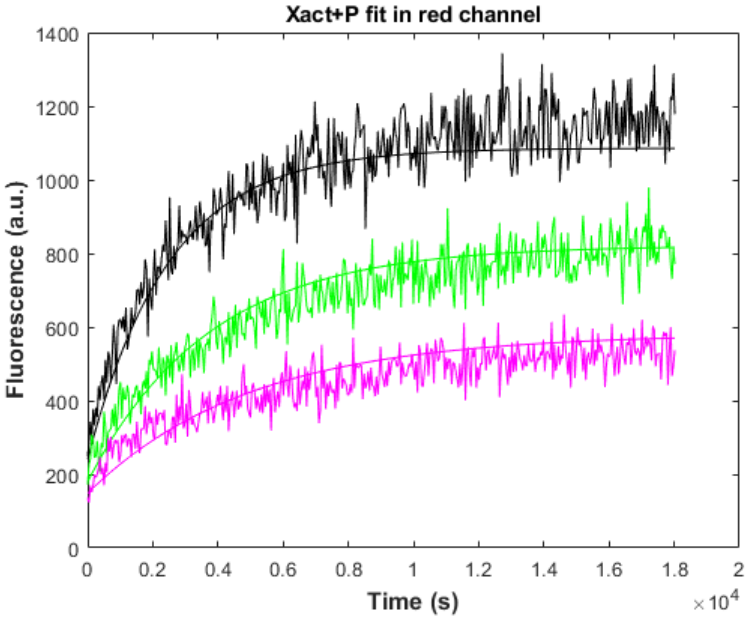

Figure 66. Fitting of the first reaction step of the pull in the forward direction with the data observed in the red channel. Time units on the X axis are seconds. In the figure, the solid lines correspond to the fittings while the rugged lines correspond to the actual experimental fluorescence tracings. The black lines stand for the Xact 20 nM + P 40 nM reaction, while the green curves stand for the Xact 20 nM + P 20 nM reaction and the magenta one for the Xact 20 nM + P 10 nM reaction. The fitting was done only for the forward reaction rate.

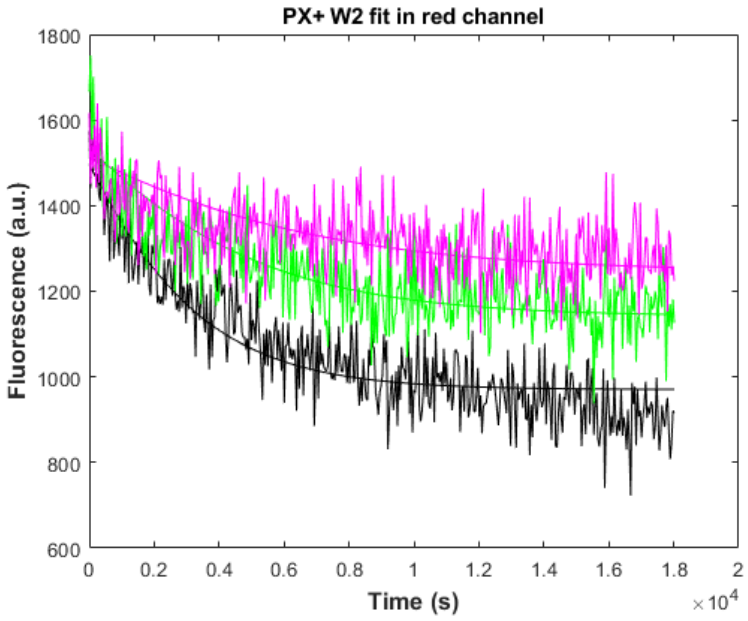

Figure 67. Fitting of the first reaction step of the pull in the backward direction with the data observed in the red channel. Time units on the X axis are seconds. In the figure, the solid lines correspond to the fittings while the rugged lines correspond to the actual experimental fluorescence tracings. The black lines stand for the PX 20 nM + W2 40 nM reaction, while the green curves stand for the PX 20 nM + W2 20 nM reaction and the magenta one for the PX 20 nM + W2 10 nM reaction. The fitting was done only for the backward reaction rate.

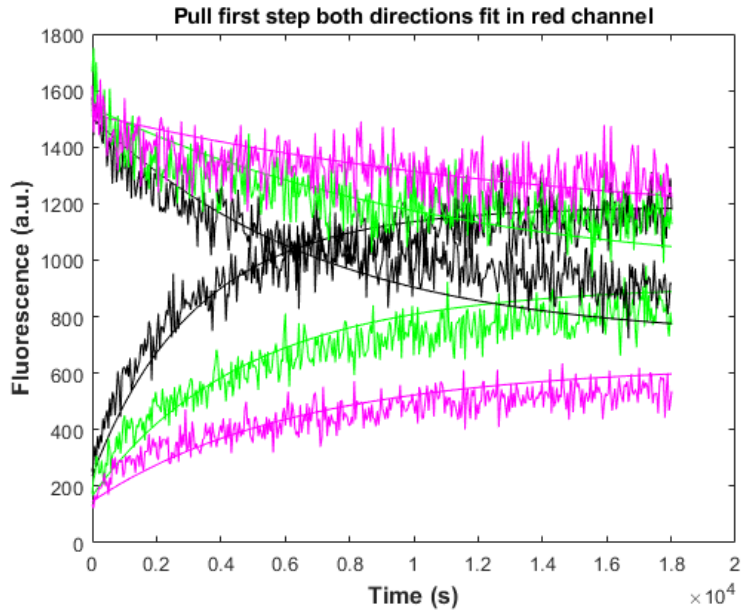

Figure 68. Fitting of the first reaction step of the pull in both directions with the data observed in the red channel. Time units on the X axis are seconds. In the figure, the solid lines correspond with the fittings while the rugged lines correspond to the actual experimental fluorescence tracings. Rising curves correspond to the forward reaction Xact +P, while the decreasing curves correspond to the backwards reaction PX t + W2. Colour codes are the same than those from figure 64 and 65 The fitting was done for both reaction rates.

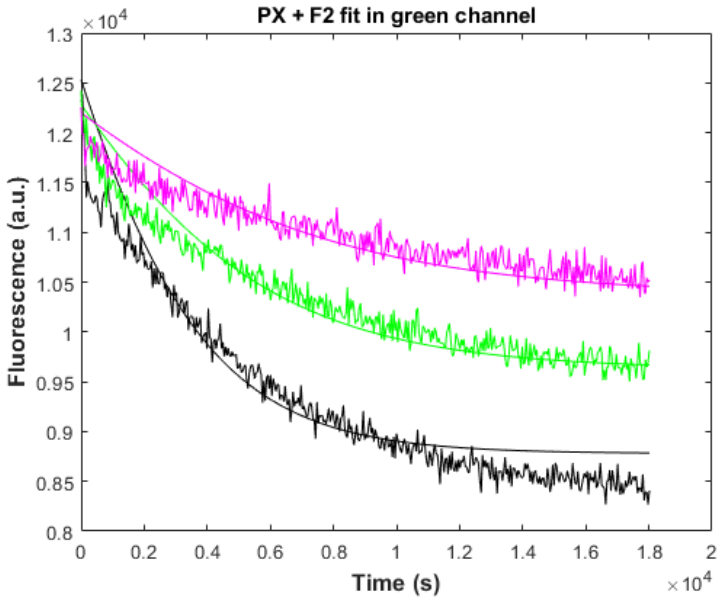

Figure 69. Fitting of the second reaction step of the pull in the forward direction with the data observed in the green channel. Time units on the X axis are seconds. In the figure, the solid lines correspond to the fittings while the rugged lines correspond to the actual experimental fluorescence tracings. The black lines stand for the PX 20 nM + F2 40 nM reaction, while the green curves stand for the PX 20 nM + F2 20 nM reaction and the magenta one for the PX 20 nM + F2 10 nM reaction. The fitting was done only for the forward reaction rate.

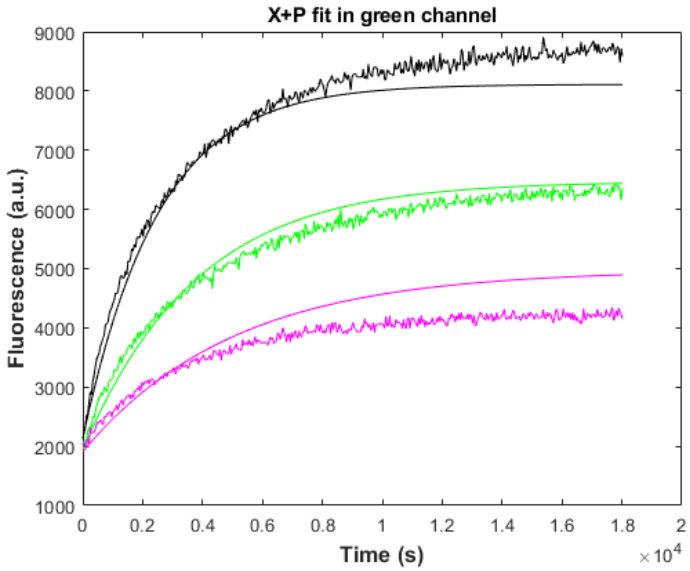

Figure 70. Fitting of the second reaction step of the pull in the backwards direction with the data observed in the green channel. Time units on the X axis are seconds. In the figure, the solid lines correspond to the fittings while the cracked lines correspond to the actual experimental fluorescence tracings. The black lines stand for the X 20 nM + P 40 nM reaction, while the green curves stand for the X 20 nM + P 20 nM reaction and the magenta one for the X 20 nM + P 10 nM reaction. The fitting was done only for the backwards reaction rate.

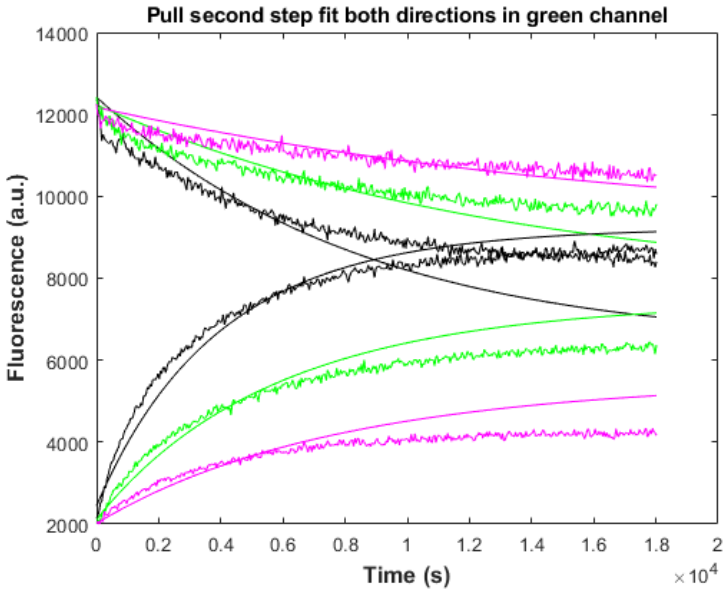

Figure 71. Fitting of the second reaction step of the pull in both directions with the data observed in the green channel. Time units on the X axis are seconds. In the figure, the solid lines correspond to the fittings while the rugged lines correspond to the actual experimental fluorescence tracings. Rising curves correspond to the backwards reaction X +P, while the decreasing curves correspond to the backwards reaction PX t + F2. Colour codes are the same than those from figure 67 and 68 The fitting was done for both reaction rates.

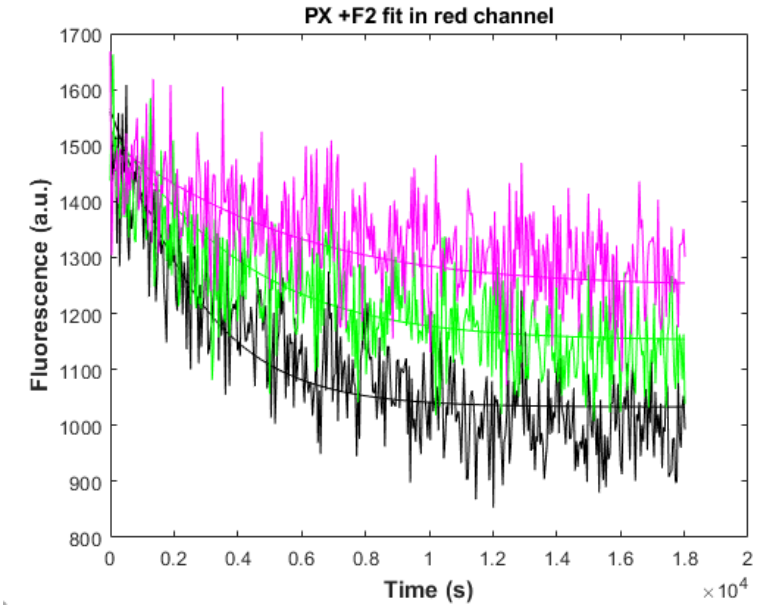

Figure 72. Fitting of the second reaction step of the pull in the forward direction with the data observed in the green channel. Time units on the X axis are seconds. In the figure, the solid lines correspond to the fittings while the cracked lines correspond to the actual experimental fluorescence tracings. The black lines stand for the PX 20 nM + F2 40 nM reaction, while the green curves stand for the PX 20 nM + F2 20 nM reaction and the magenta one for the PX 20 nM + F2 10 nM reaction. The fitting was done only for the forward reaction rate.

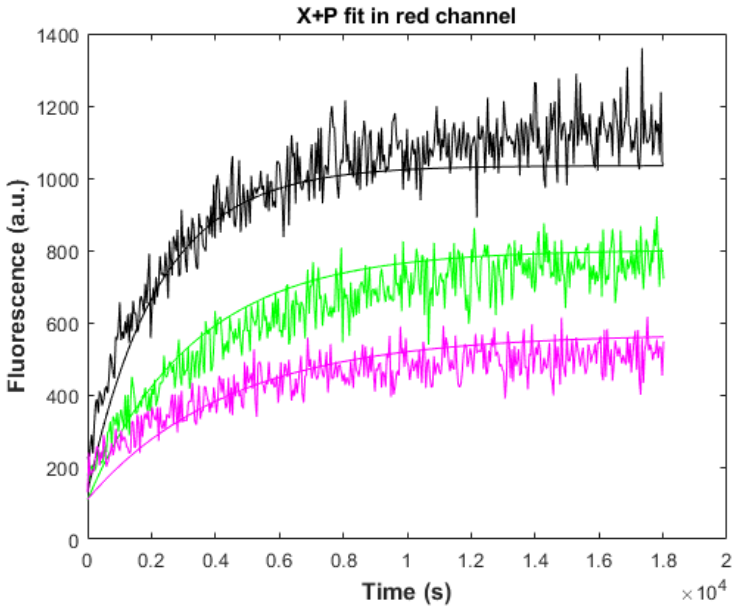

Figure 73. Fitting of the second reaction step of the pull in the backwards direction with the data observed in the red channel. Time units in the X axis is seconds. In the figure, the solid lines correspond to the fittings while the rugged lines correspond to the actual experimental fluorescence tracings. The black lines stand for the X 20 nM + P 40 nM reaction, while the green curves stand for the X 20 nM + P 20 nM reaction and the magenta one for the X 20 nM + P 10 nM reaction. The fitting was done only for the backwards reaction rate.

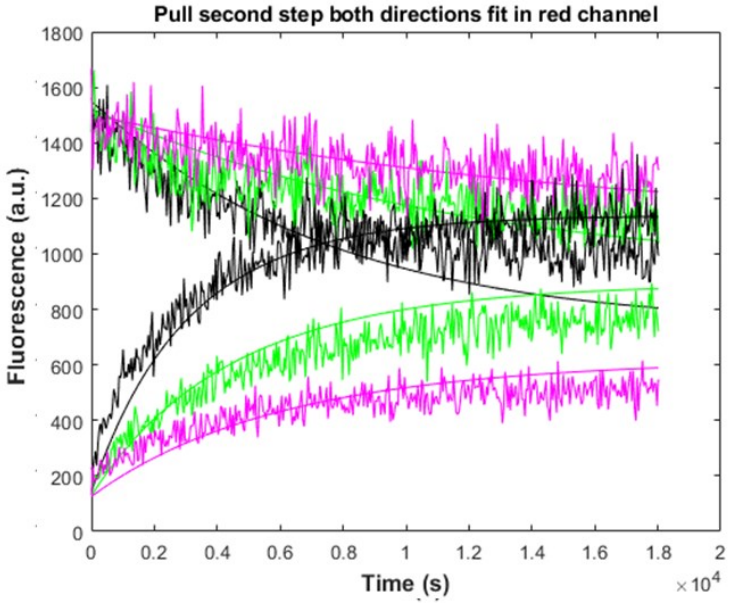

Figure 74. Fitting of the second reaction step of the pull in both directions with the data observed in the red channel. Time units in the X axis is seconds. In the figure, the solid lines correspond to the fittings while the rugged lines correspond to the actual experimental fluorescence tracings. Rising curves correspond to the backwards reaction X +P, while the decreasing curves correspond to the backwards reaction PX t + F2. Colour codes are the same than those from figure 61 and 62 The fitting was done for both reaction rates.

Reaction	G channel forwards only	G channel backwards only	G channel combined	FRET channel forwards only	FRET channel backwards only	FRET channel combined
X+K	3.3709x10^3	8.8473x10^3	3.4026x10^3	6.0519x10^3	1.4346x10^4	7.3358x10^3
KX + W1	3.9526x10^3	5.6890x10^3	3.8277x10^3	3.1396x10^3	6.0333x10^3	4.4185x10^3
KX+F1	9.1859x10^3	2.5651x10^4	3.8983x10^3	1.2289x10^4	1.6286x10^3	7.8915x10^3
X*+K	3.0493x10^4	9.1148x10^3	4.0160x10^3	1.2289x10^4	2.4826x10^3	5.1351x10^3
X* + P	5.7566x10^3	2.9380x10^4	8.0685x10^3	6.6902x10^3	4.0310x10^4	4.6121x10^3
PX +W2	8.3932x10^2	3.3081x10^3	1.9062x10^3	6.5487x10^3	6.5982x10^3	2.5913x10^3
PX + F1	6.2598x10^3	8.5499x10^3	2.1636x10^3	7.5625x10^3	6.4086x10^3	2.4665x10^3
X + P	5.1112x10^4	6.8873x10^3	3.8441x10^3	5.7625x10^4	7.3315x10^3	4.8690x10^3

Table 3. Numerical results of the fitting with the analytical model. All the quantities in the table have $M^{=1}$ s^{-1} units. Marked in yellow are the rates considered for each reaction when the reactants were those were present on the reaction. Marked in bold squares are the fittings that can be deemed as reliable for each reaction.

In the first place, we note that this model allowed us to find a reasonable fit for all reactions for, at least, one of the fluorescence channels observed. As we can observe on Table 1, all the rates obtained by the fitting are between the orders of magnitude of 10^3 and 10^4 M^{-1} s^{-1}. While it was already known that 4-way branch migration is a generally a slower process than the 3-way branch migration reactions that most TMSD systems use, in accordance with the characterization done by Nadine Dabby in her Ph.D thesis[52], these reaction rates correspond to those expected when using 4 nucleotides double toehold. However, it should be noted that Dabby's data on the characterisation of the reaction rates of a 4-way branch migration reaction were done using duplexes that were designed to maximize their base pairing after the reaction, using two duplexes with no toeholds, and hence incorporating a noticeable Gibbs free energy difference, that in turn had an effect in the chemical potential and by extension, on the kinetics of the system. This might partially explain why despite having toeholds longer by 1 base-pair, the reaction rates are in the same order of magnitude.

The second remarkable feature that we can see in our fitting is that regardless of the reaction it is trying to fit, the model seems to impose a bias in the reaction direction. In all cases, when considering a given reaction step in a given direction, the model seemed to give a lower reaction rate for the reaction that had the actual reactants present in the experiment rather than its reverse. This in principle would entail that in fact, there exists a bias towards a

given side of the reaction. But the fact that the reaction direction bias varies with whether direction is being experimentally tested is inconsistent with this explanation.

Moreover, according to the thermodynamics of the system if the mix of strands was the same initially (as it was in those reactions fitted by the green lines in the fitting figures) they would be supposed to converge in the same equilibrium endpoint. However, in most of these reactions such wasn't the case and stayed away of each other when in equilibrium – except in the case of the fitting of the first step of the pull in the green channel.

Guesses about the purity of the strands or the possibility that stable 4-stranded sequestering intermediates might be occurring in the reaction were launched. However, the first one was quite hard to assess given that the purity range that the supplier provided was quite difficult to test.

On the other hand, we tried to consider a reaction model with an intermediate:

$$K + X \quad \leftrightarrow \quad I \quad \leftrightarrow \quad KX + W$$

(19).

In this setup, instead of a single ODE, we were dealing with an ODE system:

$$\frac{d[KX]}{dt} = \gamma_3[I] - \gamma_4[KX][W1],$$

$$\frac{d[I]}{dt} = \gamma_1[X][K] - \gamma_2[I],$$

in which many of the simplifications and conservation laws that we initially applied in our previous model do not hold. This forced us to solve the ODE system of the model numerically instead of analytically. We used the Runge-Kutta method to solve the model numerically to fit the four reaction rates, alongside the fluorescence of I of the model and the offset times to our data. This new fitting did not yield any better results than those already obtained with the analytical model we built. The convergence of the reactions was not met either under these conditions. Hence, the presence of a long-lived reaction intermediate did not seem to have any strong evidence for or against from what could be concluded from our models.

Possible answers to these anomalies might come from closer examination of the reaction mechanics. It must be accounted that no detailed analysis of the reaction dynamics (either theoretically or in a single molecule setting) was performed and many a priori assumptions taken come from Dabby's analysis and might not hold in our specific case in which the reactions are designed to be reversible and have a enthalpic free energy difference of roughly about K_BT between the two sides. Moreover, chemical potential equations establish indirect relationships between reaction rates and the Gibbs free energy of the reaction[103,104]. The nature of the driving force of our designed reaction potentially entails the possibility of the rate not being constant throughout the reaction and becoming infinitely

slow as the reaction progresses. However, this possibility hasn't been properly probed in the context of this thesis and should be taken as only an educated guess.

Besides this apparent limitation on explaining fully what is happening in our reactions, all the reactions we tested were proved to be reversible with both reaction directions being kinetically accessible. Moreover, neither great differences of thermodynamic stability between all the states of the system were observed. Hence, we proceeded to test the next step of the system, which consisted in testing the full catalytic reactions both for the activation and the deactivation of the substrate.

3.2. - Full catalytic reactions

3.2.1.- Catalytic activation and catalytic deactivation

Despite the issues regarding the reaction endpoint asymmetries in the single-step reactions, we still can, in principle, consider our purposed system as apt for implementing the catalytic reactions we envisioned as the basis for the CCNs we want to build in our system.

This initial consideration can be done since, as we have seen in our reaction fittings, both sides of each step are reachable, and the reaction rates for each direction are of the same order of magnitude. However, the functionality of the full catalytic reaction must be assessed before we attempt to build larger networks. In order to see this catalytic action properly assessed we must play with three variables of the system that will tell us if the activation process is catalytic.

The first variable that we must play with is the amount of catalyst. Because catalysts, by their very nature, get recycled after the activation process, they can activate amounts of substrate larger than those of the catalyst present in a process called multiple turnover. Hence, we must observe that small quantities of the catalyst are able to activate a quantity of the substrate larger than that of the catalyst present in the system.

However, the next two variables to assess the catalytic nature of our reaction systems are related to the fact that our intended ACDC-based reactions are not only catalytic reactions, but are also fuel-driven out-of-equilibrium reactions. Due to this, the endpoint of the reaction is determined by the amount of fuel available to the catalyst, while the amount of catalyst only affects to the speed of the reaction. Hence reactions with the same amount of fuel but with different amounts of catalysts should have the same endpoint, whereas keeping the amount of catalyst the same while varying the quantity of fuel should lead to changes in the reaction endpoint without changes on the reaction rate.

Hence, we designed two experiments focusing on the catalytic activation of X into X_{act} by K, using F1 as the fuel molecule of the process as seen in Figure 75. These experiments were done to see the effects of the following variations: one using a constant amount of fuel (100 nM) to covert 20 nM of the substrate X using 5 and 10 nM of the catalyst and another one through which we observed the conversion of 20 nM of the substrate X using a fixed

amount of catalyst (5 nM) but different concentrations of the fuel (50, 100 and 200 nM respectively). We set a timestep of 88 seconds for a total of 1000 measuring steps in order to trace the dynamics into full completion.

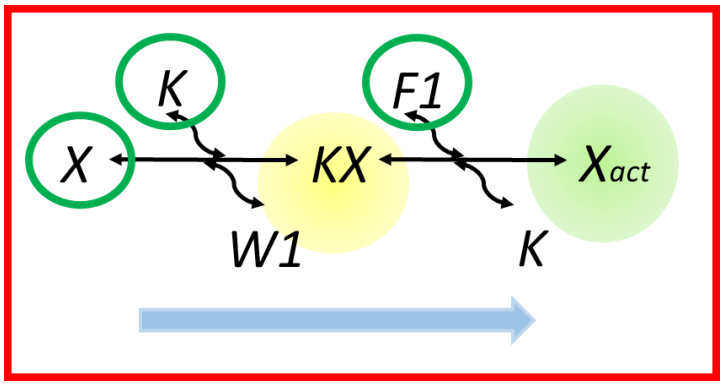

Figure 75. Schematics for the catalytic push reaction assembling the two single reaction steps. The reaction consists of the catalytic activation of X into Xact using K as a catalyst and F1 as the fuel. Species initially present in the system are circled in green in the figure. Reaction direction indicated by the blue arrow.

In the first experiment as depicted in Figure 76, we can see that the endpoint reached at the end of the reaction is roughly the same. We also observe that the reaction rate is twice as fast in the case of the reaction with [K]= 10 nM. However, in both cases, we observe that the total amount of substrate X converted is much bigger than the amount of catalyst. More precisely, the total turnover observed is around 75% of the total turnover. However, despite having a large supply of fuel, we observe that the catalytic turnover of the system is not complete. The reasons behind this behaviour will be explored in future sections of this chapter.

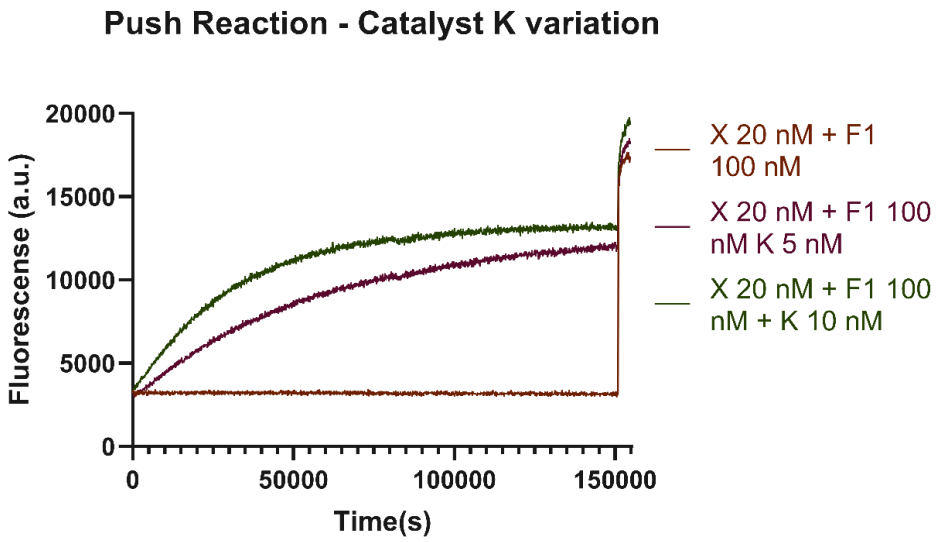

Figure 76. Fluorescence tracings for the catalytic push varying the amount of catalyst K while keeping the fuel constant. The reaction consists on the activation of X into X_{act} using K as the catalyst and F1 as a fuel. The huge increase in the fluorescence observed at the end of the kinetics corresponds with the addition of an excess of a DNA complementary to Strand 3.3 in order to get the full signal release of the experiment. We can observe that the turnover is roughly the same in both cases, but the rate varies with the amount of catalyst.

On the other hand, when observing the effects of the fuel variation, we observe that, as expected, the timescale of the reaction dynamics is similar in all three cases, while the endpoint varies as we see on Figure 77.

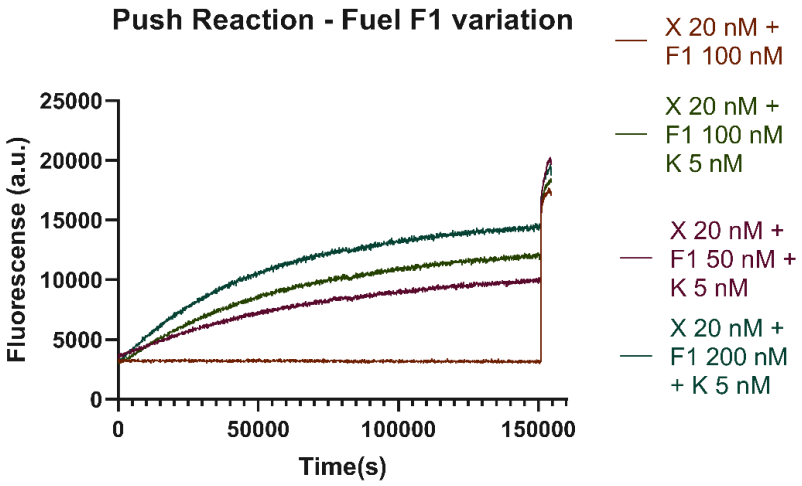

Figure 77. Fluorescence tracings for the catalytic push varying the amount of chemical fuel while keeping the catalyst constant. As we see the reaction rate is roughly the same in all three cases but the endpoint varies with the amount of fuel. The huge increase in the fluorescence observed at the end of the kinetics corresponds with the addition of an excess of a DNA complementary to Strand 3.3 in order to get the full signal release of the experiment.

In summary, we can assert with trust that our intended catalytic activation is working, albeit the reaction speed is one order of magnitude slower than the seesaw catalytic mechanism devised by Winfree and Qian[55]. However, the reason why ACDC became the design choice for our proposed system was not the speed of reaction but the possibility of implementing multiple catalysts per substrate in order to allow catalytic deactivation and the construction of the push-pull motif.

With ACDC, we can implement a new catalyst with its own supply of fuel – hence making it orthogonal to the one we have already implemented in the push. Following an analogous reasoning to that followed previously, we can justify that catalytic deactivation can also be achieved within our framework. As we saw in the case of the activation with the K catalyst, we can observe that the deactivation process can be achieved using the P catalyst (as seen in Figure 78). The catalytic nature of the deactivation process can be seen in Figures 79 and 80. Hence we can expect that the push-pull motif and more extended catalytic networks can me realized using the ACDC Framework.

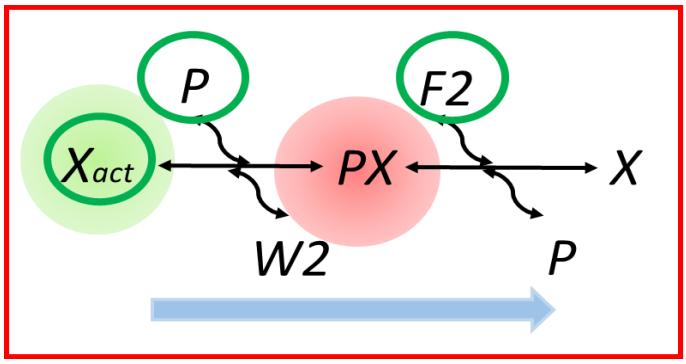

Figure 78. Schematics for the catalytic push reaction assembling the two single reaction steps, reaction assembling the two single reaction steps. The reaction consists of the catalytic deactivation of Xact into X using P as a catalyst and F2 as the fuel. Species initially present in the system are circled in green in the figure. Reaction direction indicated by the blue arrow.

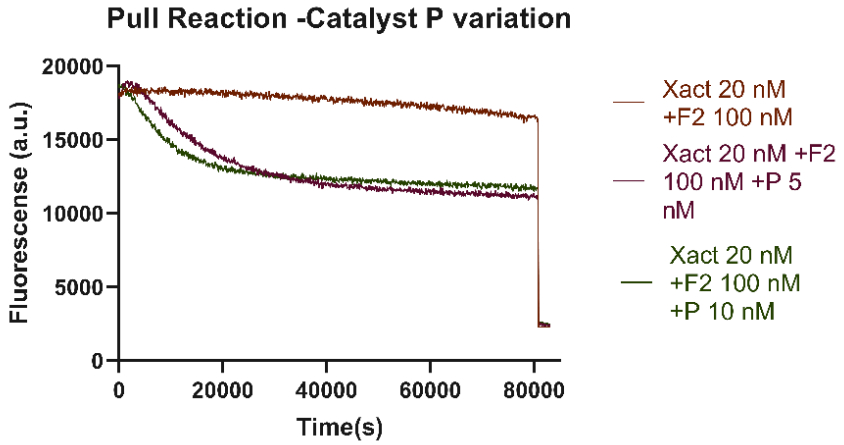

Figure 79. Fluorescence tracings for the catalytic pull varying the amount of catalyst P while keeping the fuel constant. The reaction is slower with 5 nm of catalyst (red curve) compared to 10 nM of catalyst (green curve), but both curve reaches same endpoint with the same amount of fuel. The green curve reaction rate is 1.8x times faster than the red one, which is to be expected for the amount of catalyst added. The huge drop in the fluorescence observed at the end of the kinetics corresponds with the addition of an excess of a DNA complementary to Strand 3.3 with an Iowa Black FQ quencher in order to get the full substrate deactivation on the experiment.

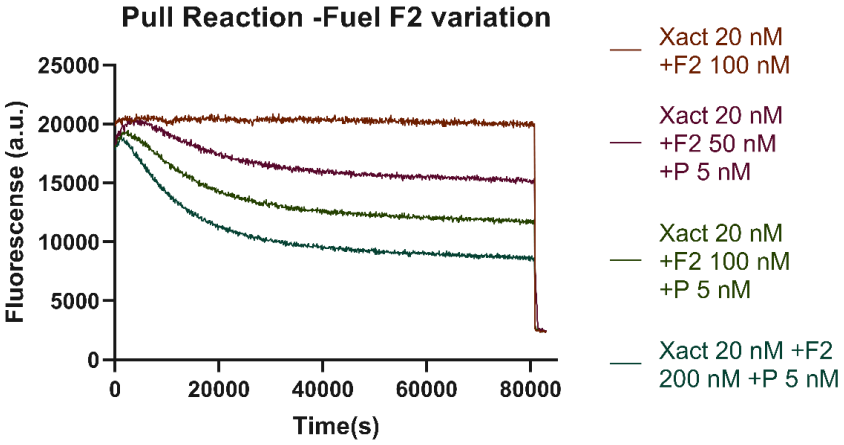

Figure 80. Fluorescence tracings for the catalytic pull varying the amount of fuel while keeping the catalyst P constant. In all three different F2 concentrations the reaction rate is the same as expected for having the same amount of catalyst. However, the endpoints are different, as they are determined by the amount of fuel. The huge drop in the fluorescence observed at the end of the kinetics corresponds with the addition of an excess of a DNA complementary to Strand 3.3 with an Iowa Black FQ quencher in order to get the full substrate deactivation on the experiment.

The activating and deactivating nature of the catalysts is not intrinsic to the catalysts themselves, but rather comes from the fuel supply they use in order to perform the catalytic turnover. This fact could be regarded as trivial. However, as we will discuss in the next subsections of the chapter, this entails profound implications to the performance of the system as well as for CCNs in general.

3.2.2.- Reversibility and loading effects: Two sides of a same coin

We have observed that each one of the ACDC reactions that constitute the push-pull cycle works as intended and that its function is catalytic as we initially intended. However, the system we have designed, as we have stated in previous instances throughout this dissertation, is an entropy-driven one in which the reaction steps do not have any enthalpy loss that drives them irreversibly towards one direction. This means that the reaction directions are determined by the type of fuel available in excess in the system rather than by any intrinsic preferential activity of the catalyst in one direction or another.

Consequently, we can use the waste molecules produced in the first reaction step of our "canonical" reactions as our excess fuels to perform the opposite reaction of that which we initially intended. We can see in Figures 81 to 83 how the catalyst K can perform a deactivation process using W1 as the fuel molecule while Figures 84 to 86 illustrate how the P catalyst can perform an activation process using W2 as the fuel. And in both cases the process is still catalytic in nature.

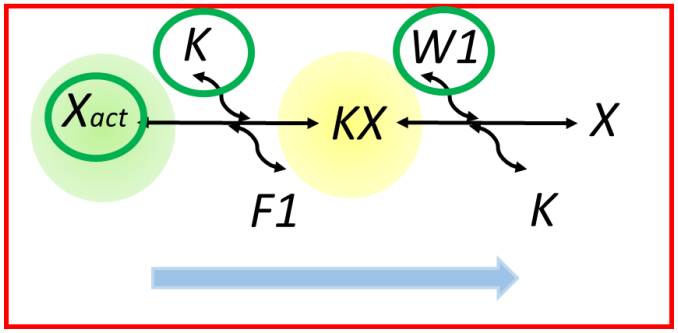

Figure 81. Schematics for the catalytic reverse push reaction – also named "pull with K", assembling the two single reaction steps using W1 as the driving fuel of the process. The reaction consists of the catalytic deactivation of X_{act} into X using K as a catalyst and W1 as the fuel. Species initially present in the system are circled in green in the figure. Reaction direction indicated by the blue arrow.

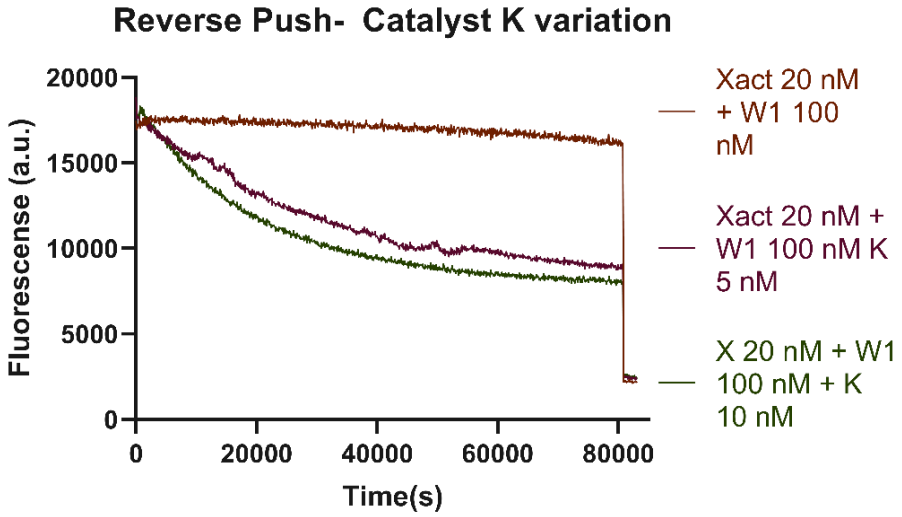

Figure 82. Fluorescence tracings for the catalytic reverse push reaction – also named "pull with K", assembling the two single reaction steps using W1 as the driving fuel of the process. The present experiment varies the amount of catalyst K while keeping constant the amount of fuel in each experiment. As we s is to be expected for the amount of catalyst added. The huge drop in the fluorescence observed at the end of the kinetics corresponds with the addition of an excess of a DNA complementary to Strand 3.3 with in order to get the full substrate deactivation on the experiment.

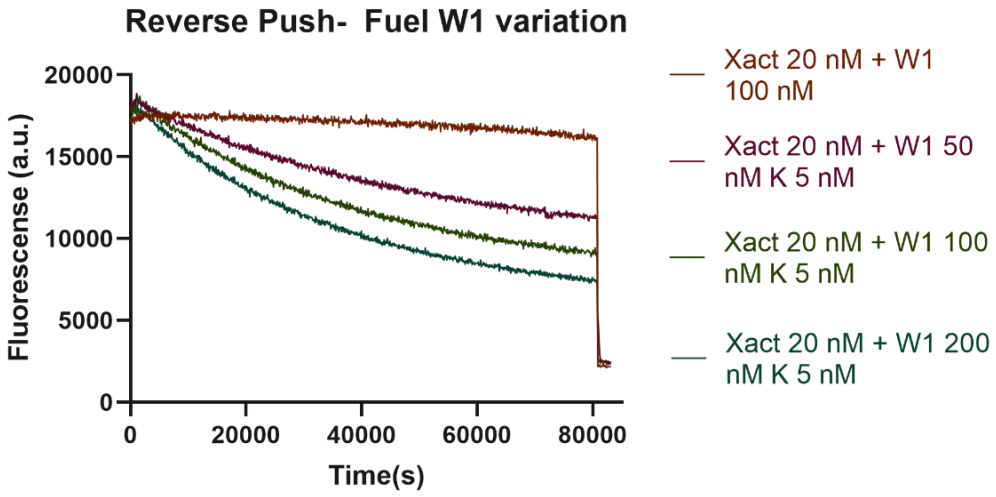

Figure 83. Fluorescence tracings for the catalytic reverse push reaction – also named "pull with K" assembling the two single reaction steps using W1 as the driving fuel of the process. The present experiment varies the amount of W1 while keeping constant the amount of catalyst in each experiment. Again, we see different amounts of fuel resulting in different endpoints. The huge drop in the fluorescence observed at the end of the kinetics corresponds with the addition of an excess of a DNA complementary to Strand 3.3 with in order to get the full substrate deactivation on the experiment.

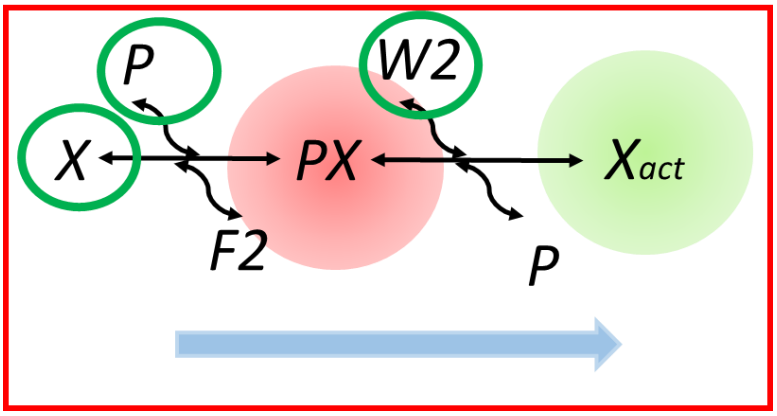

Figure 84. Schematics for the catalytic reverse push reaction – also named "push with P" assembling the two single reaction steps using W1 as the driving fuel of the process. The reaction consists of the catalytic activation of X into X_{act} using P as a catalyst and W2 as the fuel. Species initially present in the system are circled in green in the figure. Reaction direction indicated by the blue arrow.

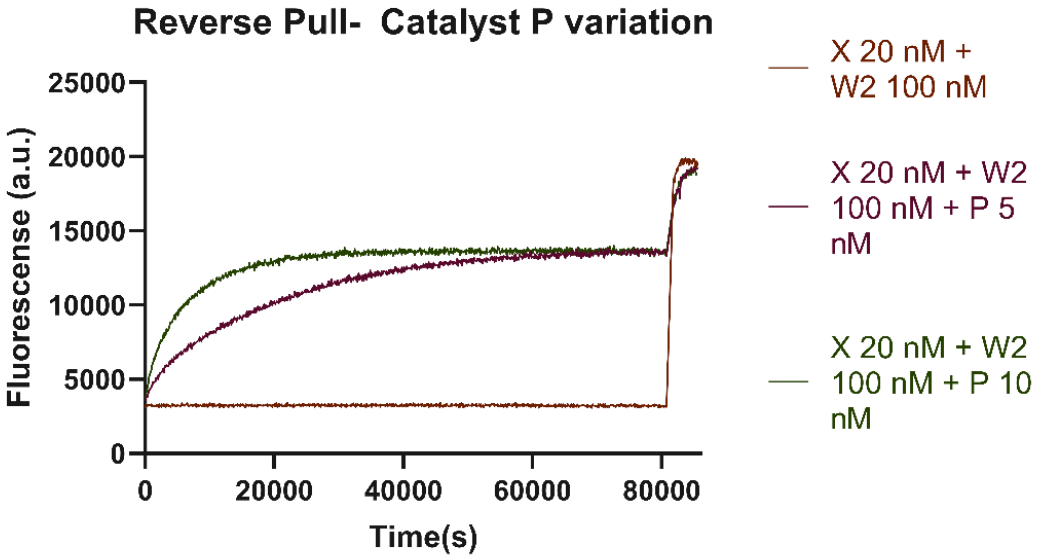

Figure 85. Fluorescence tracings for the catalytic reverse pull reaction – also named "push with P" assembling the two single reaction steps using W1 as the driving fuel of the process. The present experiment varies the amount of catalyst P while keeping constant the amount of fuel in each experiment. Again, we see, even more clearly than in previous cases, that the variation on the amount of catalysts affects the reaction of the system, but not its endpoint. The huge increase in the fluorescence observed at the end of the kinetics corresponds with the addition of an excess of a DNA complementary to Strand 3.3 in order to get the full signal release of the experiment.

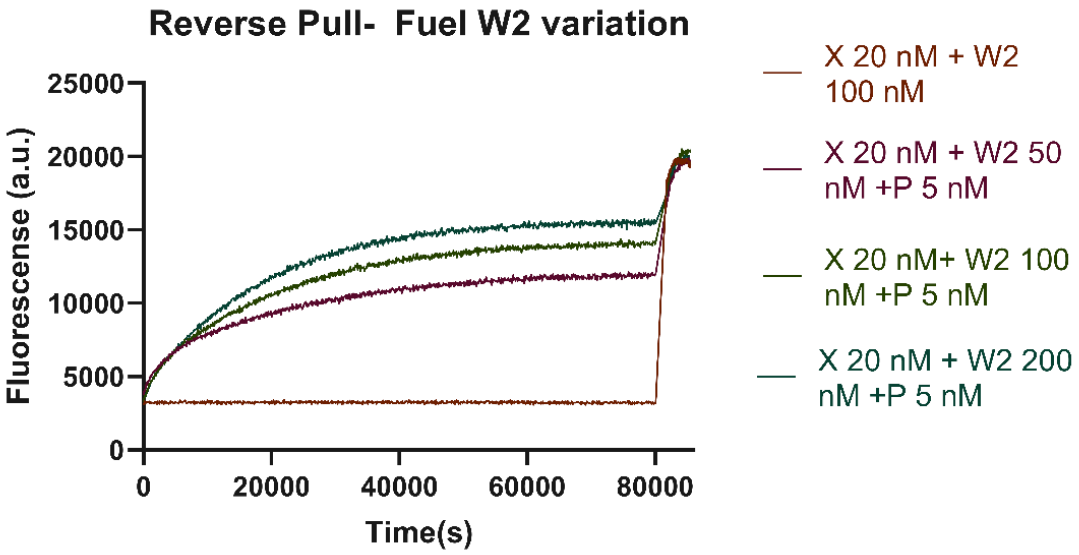

Figure 86. Fluorescence tracings for the catalytic reverse pull reaction – also named "push with K" assembling the two single reaction steps using W2 as the driving fuel of the process. The present experiment varies the amount of W2 while keeping constant the amount of catalyst in each experiment. Again, we see the variation in the fuel supply driving the system to different endpoints. The huge increase in the fluorescence observed at the end of the kinetics corresponds with the addition of an excess of a DNA complementary to Strand 3.3 in order to get the full signal release of the experiment.

This reversibility of the reaction has deeper implications on the working of the system. The fact that the activation and the deactivation can be done with any of the two catalysts we designed (which is to be expected given that the recognition domains are identical in both of them) implies that catalysts do not distinguish between the active and the inactive form of the substrate they interact with. This means that even when having a preferential reaction direction marked by the fuels, the product can come back and bind to the catalyst, thereby sequestering it and acting as an effective inhibitor as we see in Figure 87. The reaction chosen to illustrate this effect is at every moment the activation of X using the P catalyst. This reaction was chosen because the P catalyst does not present any type of indirect excitation at any channel, thus making all the results easier to directly interpret with regards the point we want to make.

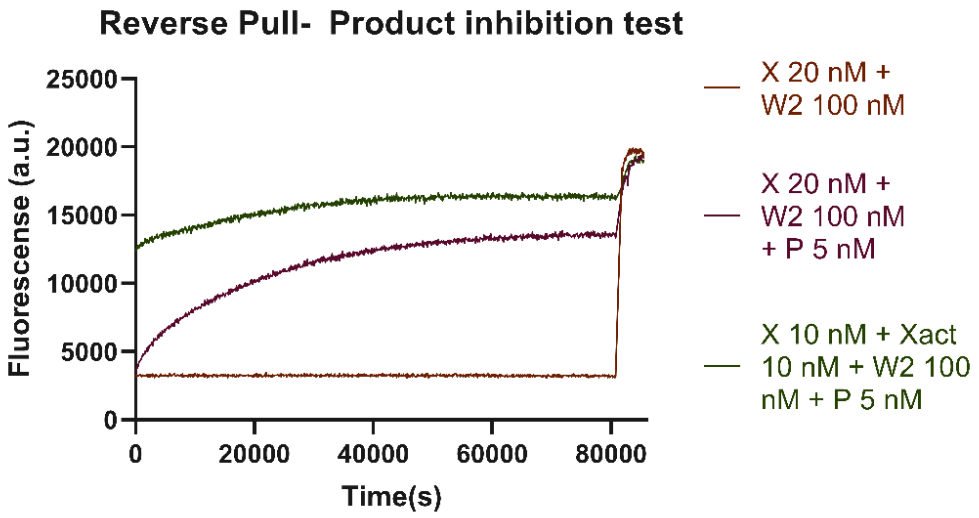

Figure 87. Fluorescence tracings for the catalytic reverse pull reaction – also named "push with P" assembling the two single reaction steps using W2 as the driving fuel of the process. The present experiment varies the amount of active substrate X while keeping constant the amount of fuel and catalyst on each experiment. The aim of the experiment consists on proving that there is product inhibition in our systems. As we see, in the conditions observed on the original experiment beginning with no activated substrate, the push leads to turnover two thirds of the substrate. The net conversion of X into Xact is lower and slower than that observed in the previous case. The huge increase in the fluorescence observed at the end of the kinetics corresponds with the addition of an excess of a DNA complementary to Strand 3.3 in order to get the full signal release of the experiment.

This product inhibition should not be totally unexpected in our design since it's a feature observed in protein-based transduction networks in nature. Previous research by Bluthgen[80] and collaborators has proposed that this product inhibition alongside sequestration on the catalysts has a role on softening time delays in activation of different parts of a given transduction network as well as help to avoid behaviours such as time sustained oscillation in negative feedback loop transduction networks. Taking the product inhibition as a feature would be one that would allow us to enhance the robustness to perturbations of the circuits in a similar manner to that of the buffering DNA circuits designed by the lab of Rebecca Schulman[105]. However, this product inhibition can also be seen as a form of retroactivity that limits the effective activation range of our system. Moreover, the presence of this product inhibition might be, alongside the unexplained dampening of our reaction dynamics seen in the single step experiments why the fuel effectivity does not scale lineally with the concentration.

It's not hard to conclude that implementing intrinsic kinetic discrimination in the binding to different states of the substrate could be a feature of interest to design ACDC systems with expanded functionality. However, attaining this kinetic discrimination has deep physical implications related to the system performance. Determining that the substrate to which the catalyst is bound is the one that we intend can be mapped into a readout operation in which the enzyme acts like the reader and the substrate state (active or inactive) codes for the two states of a bit (0 or 1). But this reading operation does not come without a cost.

The link between information and thermodynamics is one that has been thoroughly explored by physicists in the last two-hundred years since James Clerk Maxwell's foundational

thought experiment about a microscopic Demon capable of reducing the entropy of a system through the measurement of the system[106]. In the decades following the initial formulation of this experiment several progresses were made, but the two most notable were the reformulation of the problem by Leo Szilard with his theoretical engine that showed that the information of a system – understood as the degrees of freedom required to accurately describe it – has an energy and an entropy associated[107] with it, as well as the determination by Rolf Landauer in 1961 that any computational operation has a lower energy bound of $KTln2$[108], where K is the Boltzmann constant and T the system's temperature.

Further experimental and theoretical implementations of Maxwell's Demon and Szilard engine[109-111] showed that no measurement could be exploited without paying an entropic cost higher than that of the work required in the measurement. It must be remarked that such tests were proved in different physical substrates, including biochemical catalytic systems in which the components of a reaction in our ACDC Framework can be easily identified and mapped[112].

Hence, if we want to implement any kind of kinetic discrimination that improves the functionality of the systems designed with the ACDC Framework, the implementation of a mechanism that allows to pay the energetic cost of the measurement becomes mandatory. It must also be added that the notion of paying energetic costs in order to ensure the fidelity of a readout operation in biomolecular systems is not a new one, given that the concept of kinetic proofreading introduced by John Hopfield[113] in the models of replication is founded in little free energy differences between the addition of a correct or an incorrect monomer to a copying monomer.

3.2.3,- Sequestration in our catalysts

In addition to the problem that reversibility poses to the functionality of our catalytic reactions, we must add that the very fact of having the single step reactions as reversible is also a source of trouble in our reactions.

In our setup, the catalyst-substrate complexes are as intrinsically stable as any other species of the system. This thermodynamic stability in turn leads to a sequestration situation that is harder to regulate and tune, even when fuel is in excess because fuel turnover can become a bottleneck (as we observe on Figure 88 in which doubling up the amount of fuel available in the system has diminishing returns in reducing the amount of the intermediate PX). This situation results in catalyst sequestration and imposes another loading effect in terms of the catalyst in addition to that caused by product inhibition. In fact, product inhibition does also contribute to this loading effect on the catalysts.

On the other hand, the presence of this sequestration allows for the existence of saturation regimes -as we observe in Figure 89- if the substrate is more abundant.

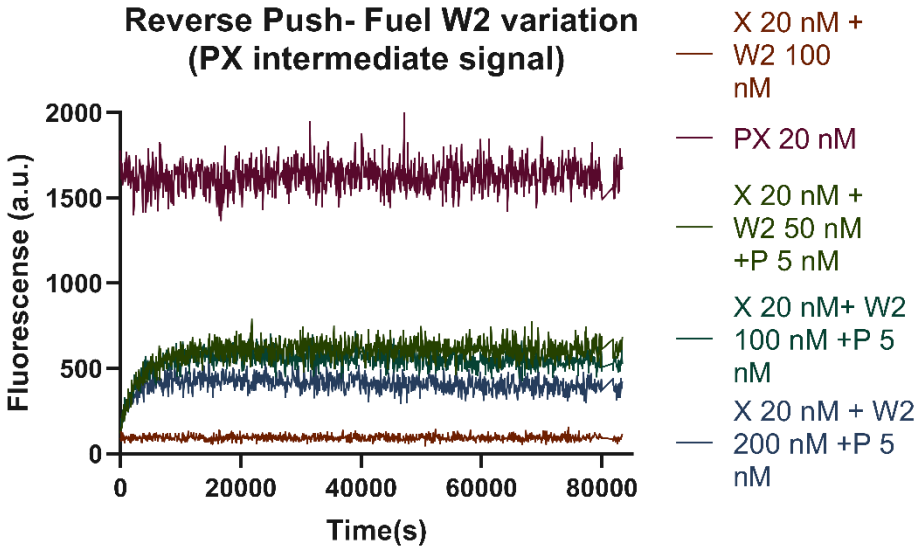

Figure 88. Fluorescence tracings for the presence of the intermediate PX in the catalytic reverse pull reaction – also named "push with P" assembling the two single reaction steps using W2 as the driving fuel of the process. The present experiment varies the amount of chemical fuel while keeping constant the amount of catalyst constant in each experiment, showing the bottleneck that fuel consumption rate poses (even doubling the fuel present in the medium from 100 to 200 nM doesn't result in a significantly lower sequestration).

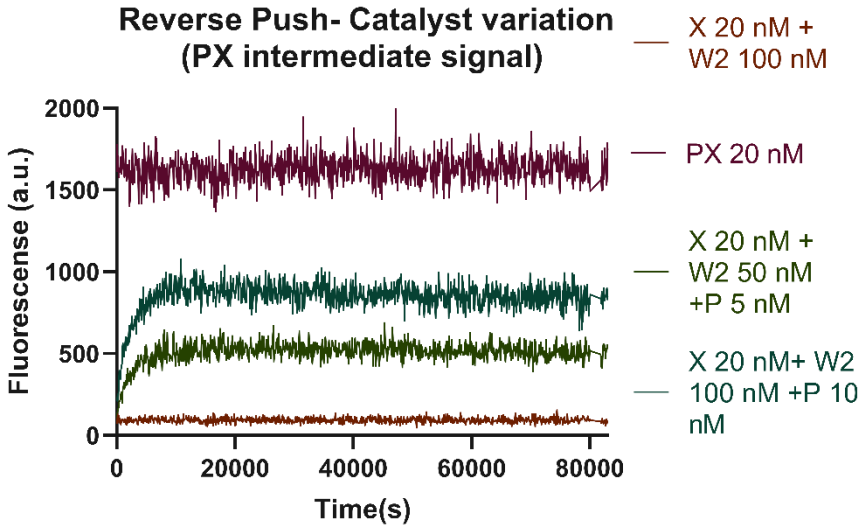

Figure 89. Fluorescence tracings for for the presence of the intermediate PX in the catalytic reverse pull reaction – also named "push with P" assembling the two single reaction steps using W2 as the driving fuel of the process. The present experiment varies the amount of catalyst P while keeping constant the amount of fuel on each experiment. In this case the reaction is observed on the red channel, allowing us to observe the formation of the intermediate during the catalytic reaction. And as we can observe saturation during the catalytic process reaches near 100% of the catalyst in the tested conditions when compared with the intermediate control of PX 20 nM.

While already noticeable in single catalytic reactions, the performance characteristics identified throughout this chapter will become even more relevant when studying the assembly of these catalytic reactions into larger reaction motifs and networks, as we will see in Chapter 4.

Chapter 4: Experimental implementation of extended reaction networks in the ACDC Framework: results and considerations.

In Chapter 3, we confirmed that catalytic reactions can be implemented using the ACDC Framework and that a given activation reaction can be reversed by using a second catalyst opposing the first one. In principle, this should ensure that both catalysts can be working simultaneously and, as a consequence, that a push-pull (as depicted in Figure 90) motif can be built. Hence, the next logical step that we will follow lies in observing two catalysts working simultaneously in the same reaction setup.

However, a single push-pull network, while being the minimal unit for out-of-equilibrium CCNs, is functionally limited. Hence, we must explore the possibility of connecting different ACDC reactions in different extended configurations. So, the second half of the present chapter will explore the basic operations that can allow us to implement more functionally complex networks.

More precisely, what we intend to explore are three operations:

1. The possibility of splitting the action of a given catalyst between different substrates simultaneously. This network would allow any circuit to implement concurrency and to activate simultaneously two parts of the same circuit.

2. Having two separate catalysts acting in the same direction on the same substrate. This will allow us to add functional redundancy over the actuation of a given substrate as well enabling the construction of multi-input systems.

3. With these basic two operations, we can have control on what happens on a given catalytic layer of a CCN. But as we saw in the first chapter, transduction networks are organized in cascades that allow cells to implement signal amplification between layers. Cascading also allows turning the dynamics of the system from regular Michaelis-Menten-like responses into sigmoidal step-like responses and helping mitigate the noise inherent in low copy, biomolecular systems. Hence, we need to show that ACDC-based systems can implement cascading action in order to ultimately demonstrate that they can be used to build extended networks.

4.1, - Push-Pull Network

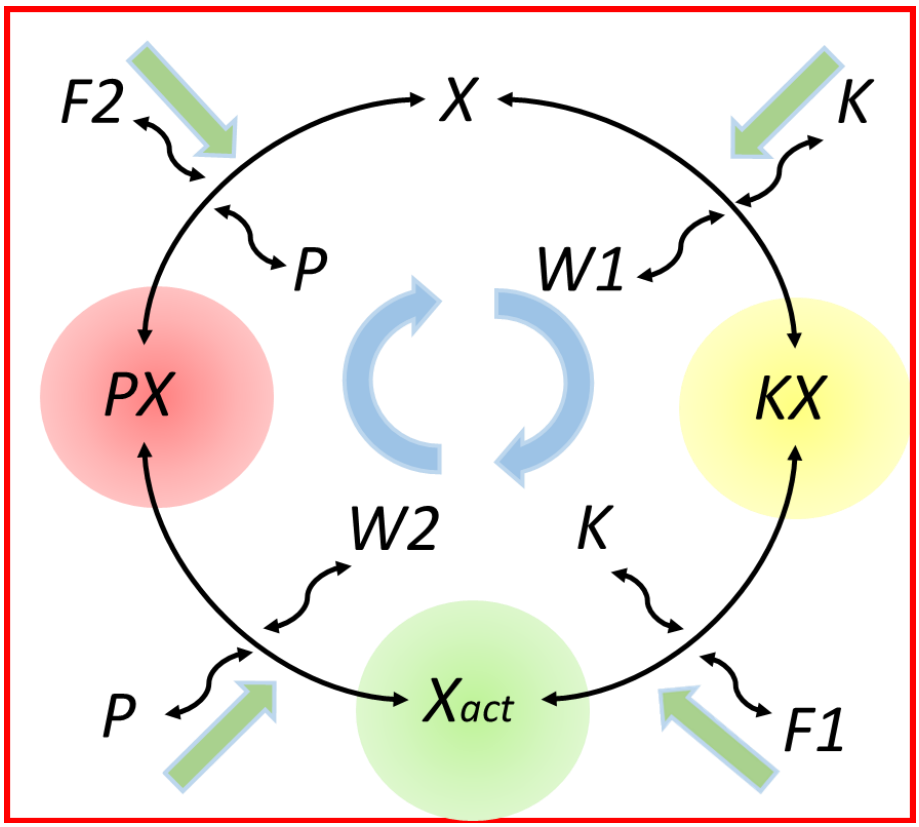

Figure 90. Reaction diagram with all the reversible reaction components of an ACDC-based push-pull network. The species meant to produce fluorescent signal are circled with a translucid colour circle corresponding to their wavelength emission: Green (520 nm corresponding to the emission of the Alexa 488 fluorophore) in the case of X_{act}, Yellow (600 nm corresponding to the non-overlapping emission chosen for the Alexa 488-Cy3 FRET pair) in the case of the intermediate KX, and red (690 nm corresponding to the emission of the Alexa 488-Cy5 FRET pair) in the case of the intermediate PX. The cycle works clockwise with K consuming F1 and producing W1 to switch X into X_{act} on its first half, whereas on its second half P turns X_{act} back into X by consuming F2 and producing W2. The inner blue circular arrows depict the general directionality of the cycle. The green translucent arrows depict the consumption direction of reactants for each substep.

As we mentioned in our initial chapter, push-pull networks are considered to be the fundamental functional subunit of transduction networks. They are functionally interesting because they work in an out-of-equilibrium quasi-steady state that is decoupled from the equilibrium endpoint. This means that the output level of the network can be modified through variations of the catalysts' ratios. Hence our interest when studying these systems lies in verifying that our DNA ACDC-based analogue system behaves analogously to natural push-pull (protein-based) networks upon which it is based.

In order to make this verification, we must do two types of experiments: one that proves that the system can be reversed by sequentially pushing and pulling the system and another series of experiments in which we can observe the transition of the system from the out-of-equilibrium steady state determined by the catalysts to the system's equilibrium endpoint.

4.1.1.- Sequential addition of the catalysts

In these experiments we began with all the substrate at a concentration of 20 nM in the inactive state as well as with a supply of 500 nM for the fuels of both the K and the P catalysts. The experimental conditions for the fluorescence tracings used were a variation of those used previously in section 2.3 of Chapter 2 and the whole Chapter 3. While the excitation conditions as well as the photodetector gain conditions used were the same as in the previous chapter, the timestep was changed to one of 88 seconds. Under these conditions, we were able to observe that after 12 hours, using 10 nM of the catalyst K, we were able to reach a turnover of approximately 85% of the available substrate before plateauing (see Figure 91). The percentage was established using as a reference for a 100% activation the final triggering level after the addition of a complementary strand in excess and using the signal of the control without the catalysts (the brown line in Figure 91) as the 0% activation control. This is summarised in the following equation used to calculate the turnover:

$$\% = 100\, x(Activation - Negative\ Control)/(Positive\ Control - Negative\ Control)$$

After the system reached a plateau, the opposing catalyst P was added to the mix at a concentration of 10 nM. And as expected, the addition of the catalyst caused a drop on the fluorescence, confirming that the catalytic deactivation process took place. More precisely, we observed that the signal of the activated substrate dropped until it reached a 50/50 ratio of active to inactive substrate.

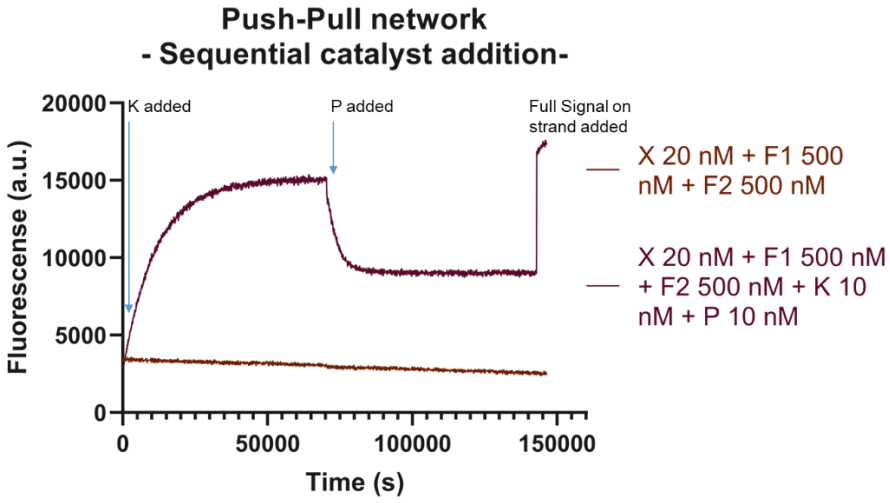

Figure 91. Fluorescence tracing of the sequential addition of the catalysts for the push-pull network. Fluorescence is observed in the emission channel corresponding the emission wavelength of the Alexa 488 fluorophore (Green). Each catalyst was present at total concentration of 10 nM, while the substrate X was present at a 20 nM concentration and the fuels for each catalyst were at a 500 nM excess to ensure a high drive as well as sustained maintenance of the resulting steady state. At the end of the experiment a strand complementary to strand 3.3 was added in excess to provide the full activation reference.

As counterintuitive as this might look, this result is the one to expect if both catalysts consume fuel at the same rate and are present at the same quantity in the system. When

these conditions are met, the system reaches an out-of-equilibrium quasi-steady state in which a balance exists between the antagonistic actions of the catalysts. This state is kept as long as the fuel supply allows for the maintenance of this state.

However, we must not discard that what we have reached in the present experiment is actually an out-of-equilibrium steady state instead of an equilibrium. Hence in the next experiments, we shall demonstrate that our catalytic system is able to reach an out-of-equilibrium steady state in which we can vary the activation level prescribed testing variations of the catalysts ratios Moreover, in order to demonstrate that the equilibrium state and the quasi-steady state of the system are different regimes, we will also monitor the transition into equilibrium once the fuel is already consumed.

4.1.2.- Demonstration of the out-of-equilibrium quasi-steady state regime of the push-pull motif.

Out-of-equilibrium systems are fundamentally driven kinetically. This means that variations of the catalysts' concentrations are the main determinant of the degree of activation or deactivation of the out-of-equilibrium regime. This means that all the systems with the same fuel ratios will converge to the same equilibrium point regardless on which catalysts ratios have triggered the reactions – which in turn will have different out-of-equilibrium states.

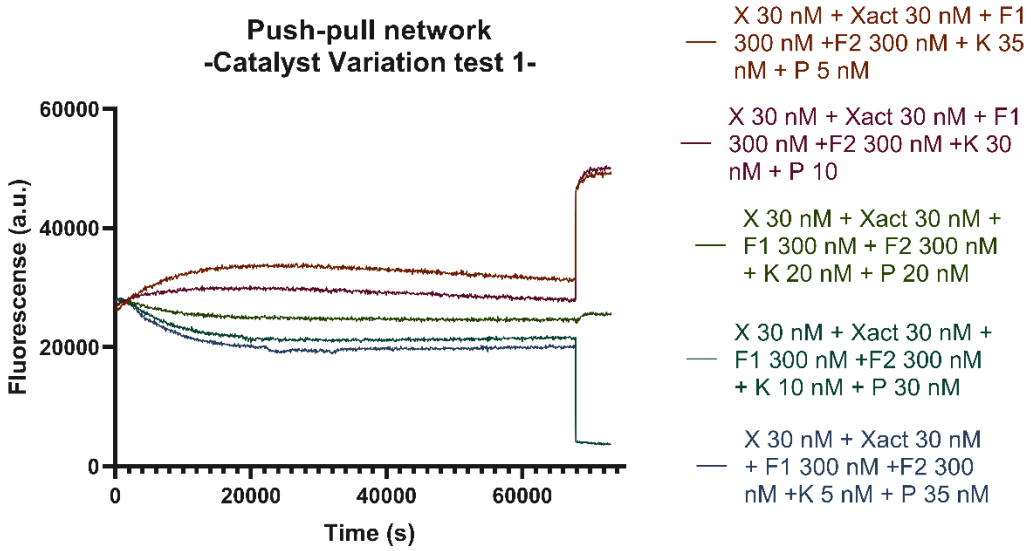

Figure 92. Fluorescence tracing of the simultaneous addition of the catalysts of the push-pull network. This test was designed to probe the effects of varying the ratio of catalysts in the system. Fluorescence is observed in the emission channel corresponding the emission wavelength of the Alexa 488 fluorophore (Green). At the end of this experiment, in order to get a picture of the dynamic range of the experiment series, two types of strands were added. First a strand complementary to strand 3.3 was added in excess to the K 35 nM P 5 nM and K 30 nM P 10 nM experiments to get the signal equivalent to the top activation. On the other hand the K 10 nM P 30 nM and the K 5 nM P 35 nM had an Iowa Black FQ-conjugated version of the aforementioned strand excess added to give the full deactivation signal equivalent.

As we see in Figure 92, different catalyst ratios are driving the system to different maximums in their out-of-equilibrium steady states. This indicates that, in principle, the push-pull cycle is

working as intended. Despite this difference of activation due to the catalysts, we also observe that even when the system is much more biased towards a given direction, either if it is activation or deactivation, it does not go to the full extent of the system. Many causes previously described on Chapter 3 could explain this trouble with the system's dynamic range:

- The lack of an enthalpic drive on the system inherent to the design's reversibility
-The product inhibition observed in section 3.2.2. of Chapter 3
-The unexplained thermodynamic anomaly observed at the fittings of the single steps.

While we can observe some eventual decay in the signal, it could be argued that we are still not observing relaxation into equilibrium. However, when halving the amount of fuel from 300 nM for each catalyst to 150 nM (see Figure 93) we can see that the endpoint of all different five P/K ratios that we used initially is the same, thus confirming that a relaxation to the same equilibrium endpoint happens. The data in Figure 93 also confirms that larger supplies of fuels, even when present in the same ratios, lead to higher drives with broader dynamic ranges that are sustained for longer time.

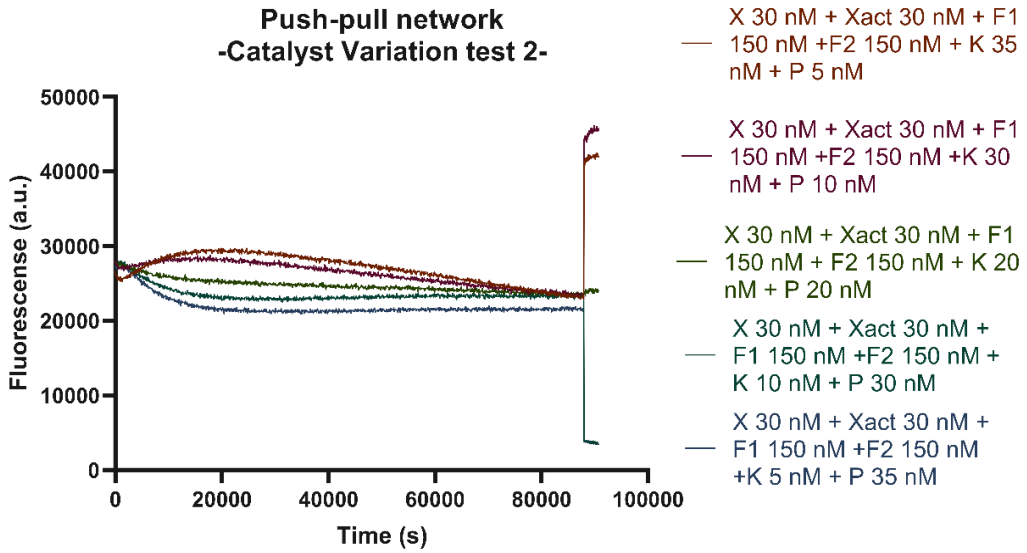

Figure 93. Fluorescence tracing of the simultaneous addition of the catalysts of the push-pull network observed on the green channel halving the fuel supply from that used at Figure 92. This test was designed to probe the effects of varying the ratio of catalysts in the system. The amounts of catalyst added are indicated in the legend of the figure, while the common conditions for all the experiments were [X] = [X_{act}] = 30 nM and [F1] = [F2] = 150 nM. At the end of this experiment, in order to get a picture of the dynamic range of the experiment series, two types of strands were added. First a strand complementary to strand 3.3 was added in excess to the K 35 nM P 5 nM and K 30 nM P 10 nM experiment to get the signal equivalent to the top activation. On the other hand the K 10 nM P 30 nM and the K 5 nM P 35 nM had an Iowa Black FQ-conjugated version of the aforementioned strand excess added to give the full deactivation signal equivalent.

On the other side, and as another consequence of this logic, using the same ratio of catalysts while using different ratios in the fuel, will result in different equilibrium points. As predicted and seen in Figure 94, beginning in the same initial point with the different fuels leads to divergences on the endpoint over long periods of time.

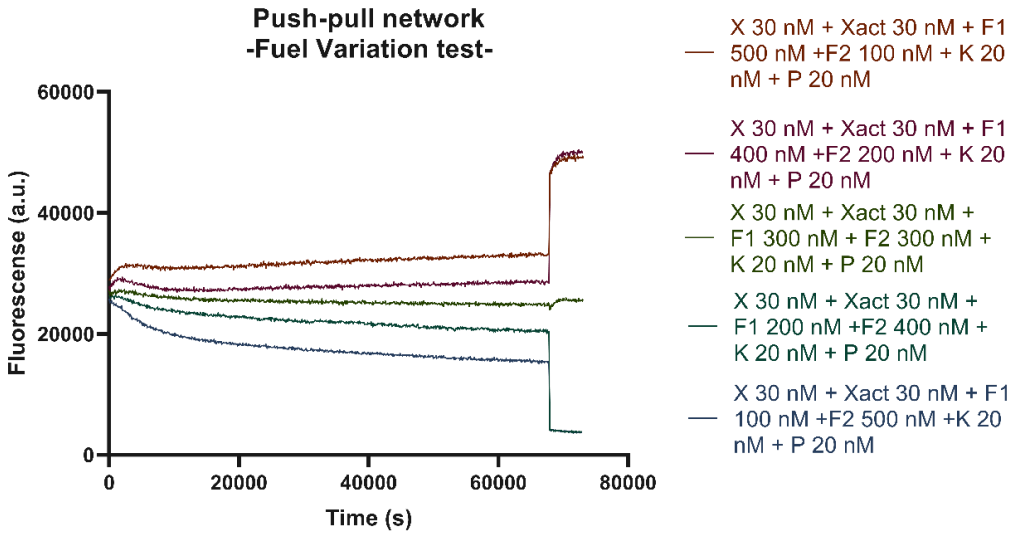

Figure 94. Fluorescence tracing of the simultaneous addition of the catalysts for the push-pull network during a fuel variation experiment. The traces are observed in the emission channel corresponding the emission wavelength of the Alexa 488 fluorophore (Green). First a strand complementary to strand 3.3 was added in excess to the F1 500 nM F2 100 nM and F1 400 nM F2 200 nM experiments to get the signal equivalent to the top activation. On the other hand the F1 200 nM F2 200 nM and the F1 100 nM F2 500 nM had an Iowa Black FQ-conjugated version of the aforementioned strand excess added to give the full deactivation signal equivalent.

In conclusion, we can confidently assert that our push-pull system is working out-of-equilibrium system. We effectively can observe how the steady state and the equilibrium state are decoupled one from another and we can exert control over its activation and deactivation by changing catalysts ratios as well as fuel supplies.

On the other side, we can observe limitations in the activation and deactivation range of the system. These limitations are caused by various causes such as the product inhibition caused by the agnosticism of the catalysts to the activation state of the substrate X derived as well as the fact that the system dominant drive is an entropic one. Solutions for these problems will be addressed in the next chapter of the present thesis.

4.1.3.- Possibility of transient action?

As we observed in this system, we have two different regimes in a push-pull network dominated by two different forces: the out-of-equilibrium quasi steady state dominated kinetically by the catalyst action, and the equilibrium dominated by the supply of fuel in either direction. These regimes are dominated differently and are decoupled temporally one from another. This feature implies that the system could in principle have two opposing tendencies on each of the states and be driven in a transition from one to another.

As a consequence, this leads to the possibility of creating transient activation of a substrate using a push pull network that is kinetically driven to the activation of the substrate, but thermodynamically tends to the inactivation. Hence, in order to test how feasible the construction of these transiently activated modules is, we modified the concentrations and initial conditions of our push-pull module. In these experiments, while keeping the same FRET

conditions as in Section 4.1.2 we changed the initial conditions of the system. Hence, the the experiment began with no activated substrate X with a clear kinetic bias towards activation by having 3.5 times more K than P, and biased toward deactivation in equilibrium by having a 5x to 10x excess of F2 with respect to F1. The results of these tests are displayed in Figure 94.

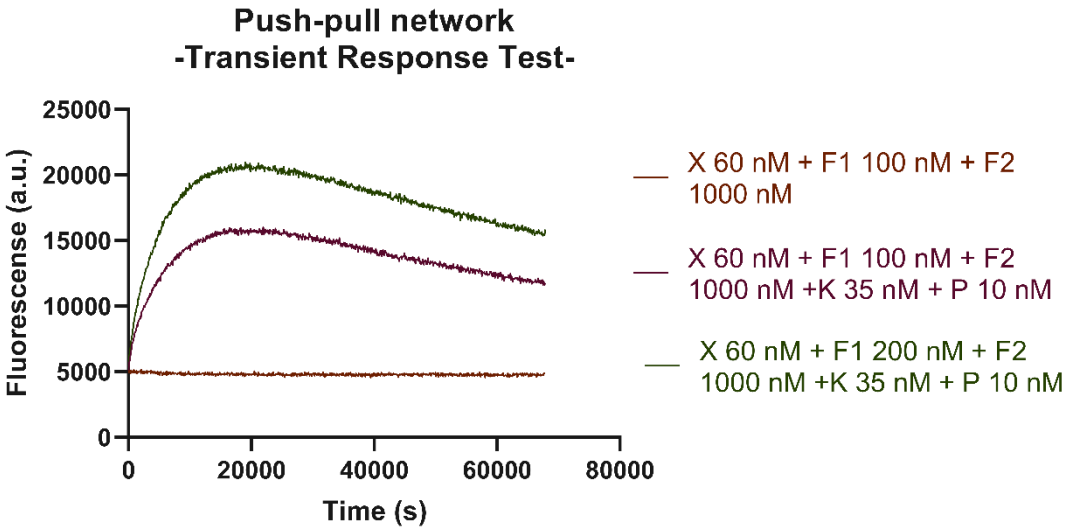

Figure 95. Fluorescence tracing of the transient response action test of the push-pull network. Fluorescence is observed in the emission channel corresponding the emission wavelength of the Alexa 488 fluorophore (Green). In the present figure we intend to display how the decoupling of the steady state and the equilibrium state on the system leads to the possibility of implementing transient, pulse-like activation. However, the deactivation dynamics become too slow to create sharp pulses.

As we see in Figure 95, the implemented system does not activate to the fullest range despite having more than enough fuel molecules to consume in the process. Additionally, the deactivation process is dampened due to the causes listed in section 4.1.2.

4.1.4.- Possibility of zero-order ultrasensitivity? General lessons for circuit design.

One of the most interesting possibilities that the push-pull motif presents is the possibility of achieving zero-order ultrasensitivity with this motif. This behaviour was first predicted theoretically for the mitogen-activated kinase cascades by Koshland Jr and Goldbeter in their seminal 1981 paper[114]. In this paper, they predicted that in a given single push-pull unit that works with its catalysts at saturations, very small variations on the concentration of any of the catalysts would result on pronounced activation or deactivation of their substrate. Koshland and Goldbeter argued this could explain partially why living systems are able to give fast adaptable responses to external outputs.

Over the years – given the importance of transduction pathways in cell regulation processes both in bacteria[115] and eukaryotic cells (normally being based on the MAPK cascade system)[116-117] – the interest in understanding the role of this action mechanism in transduction pathways grew in the scientific community and so did the impact of the original work by Koshland Jr and Goldbeter. Moreover, in recent years the synthetic biology community has displayed a great interest in the understanding of push-pull networks as fundamental units of

transduction as well as in exploiting this behaviour in artificial feedback controllers that ideally could display fast analog-to-digital response [73,118].

In principle, when examining the responses of our own ACDC-based push-pull network in Figures 93 and 94, we can observe that we have a limited dynamic range in our system. Moreover, in addition to how broad the dynamic range of the system can become by increasing the fuel available we also observe that (very) skewed ratios towards activation or deactivation do not seem to result in much greater activation or deactivation. This situation seems to indicate that our system doesn't have an ultrasensitive response of the zero-order type. However, the reasons why our ACDC-based push-pull system doesn't display the behaviour the theory predicted are not trivial and are related to some of the functional issues described throughout this thesis. Goldbeter and Koshland's case, while extremely relevant to the field and literature, relied on theoretical assumptions that made the zero-order ultrasensitivity a very specific case instead of a general functional feature of transduction systems.

The first claim is that Koshland and Goldbeter assumed was that the enzymes in a push-pull network work in a saturated regime with the substrates being much more abundant. However, in nature, the most common case is that the catalysts and substrates of the transduction networks network are present in very similar concentrations. As noted by Bluntgen[90], when this is the case, the downstream product tends to get sequestered by the catalysts, destroying any zero-order ultrasensitive behaviour. However, it has also been argued that, in nature, kinases and phosphatases are not necessarily diffusing freely, but rather are scaffolded in protein complexes. The spatial confinement effect that the scaffolding confers to transduction cascades has been linked to different regulatory effects[119-121], which theoretically includes facilitating zero-order ultrasensitivity by enabling substrate excess concentration locally.

It must be noted that beyond the effects to zero-order ultrasensitivity, sequestration in intermediates poses problems in other aspects of the design. Alongside compromising zero-order ultrasensitivity, sequestration causes loading effects that also compromise modularity[100] and dampen the dynamic behaviour of a given circuit. Moreover, the existence of this internal loading effect even in kinase-phosphatase-based push-pulls constitutes a counterargument to the use of push-pull networks as insulator motifs to reduce loading effects and restore modularity, as proposed by authors such as Eduardo Sontag or Domitilla Del Vecchio[122] – or at least constitutes such counterargument if we do not consider any measurement to reduce internal loading effects in the catalysts that form the push-pull motif itself.

The second reason that might compromise the plausibility of zero-order ultrasensitive behaviours is the presence of product inhibition in the system. The first account of this effect was described by Ortega[123] and collaborators in 2002. In this work they described how, when presuming a non-negligible chance of product inhibition, zero-order ultrasensitive behaviours broke – which can be a specific case in some natural, enzyme-based systems but is the case more generally for an ACDC-based push-pull network. However, subsequent theoretical works

both for catalytic pairs as well as for bifunctional catalytic systems [124,125] explored this claim and proposed that different zero-order ultrasensitivity regimes could be reached by limiting the product inhibition the catalysts suffered.

Importantly, despite being phenomena initially described for kinase/phosphatases-based systems, it is likely that any transduction system based on the ACDC framework will suffer from the sequestration and product inhibition issues described in this section to an even greater degree than its protein analogue network since the reaction intermediates and the products of the catalytic turnover in ACDC are equally energetically favoured – thus favouring sequestration – and the fact that catalysts don't have inbuilt any mechanism to distinguish different states of a substrate

However, it must be noted that, while the conditions for the setting of a zero-order ultrasensitive response are very much disputed in the bibliography, it is not the only way that ultrasensitive behaviours can be implemented in transduction networks. As noted by Ferrell and collaborators[78], other mechanisms such as cascading and allosterism can result in a switch-like behaviour in which the dynamics of the system transition from being described by a Michaelis-Menten dynamics to a Hill-type dynamics. This typically results in fast, analog-to-digital signal processing switch.

In conclusion despite the failure to implement an ultrasensitive behaviour in a single push-pull subunit, the rigorous study of the possible causes that might enable these behaviours can teach us lessons related to circuits design and functionality. Following this approach, we can learn about when the desired behaviour is plausible and find possible alternative mechanisms to implement the behaviour-which in many cases entails the construction of extended networks.

While the vast majority of the arguments about network functionality have been drawn from examples based on natural protein-based systems, all these arguments still hold when applied to nucleic-acid systems. From a functional point of view, the nucleic-acid analogues we have engineered can be considered as extreme cases of properties that are hardly tuneable in protein-based systems – like product inhibition – hence demonstrating their usefulness as a design exploration tool for biochemical control networks.

4.2.- Extending networks in the same catalytic layer

As stated in subsection 4.1, while push-pull networks are considered the basic motif of transduction networks *in vivo*, not all the properties that make these networks such remarkable information processing systems can be reduced to the catalytic nature of their components or to the out-of-equilibrium nature of the push-pull motif. These properties rather rely on the ways in which these push-pull motifs integrate into larger networks.

In order to integrate push-pulls into larger networks, we must first identify the basic realisable operations that allow us to do such integration. In the present iteration of the ACDC

framework the basic realisable operations that enable the design of larger complex networks are: signal branching, signal integration and cascading.

The signal branching operation consists of the simultaneous action of a given catalyst on many substrates. This operation is functionally relevant because it allows us to perform different parallel operations in a given network and theoretically give our CNNs the ability to operate concurrently and execute different parts of a biomolecular "algorithm" in parallel. and is performed by a motif we have dubbed the "split" network.

Signal integration, on the other hand, consists of the simultaneous action of two given catalysts on the same substrate and in the same direction. The relevance of this operation lies in the fact that it allows us to introduce functional redundancy in a network by making more than one catalyst capable of performing a given catalytic task, which results in increased functional robustness. Moreover, retaking the analogy between transduction elements and the elements of a perceptron established by Dennis Bray,[69] this operation allows us to sum different inputs into a node. This operation is performed by the "join" network motif.

The third and final basic realisable operation we need to implement and demonstrate in our ACDC systems is cascading. With this operation, we have the activation of a substrate in one catalytic layer that subsequently can act as a catalyst in a deeper catalytic layer. This operation is relevant from the functional point of view because, as we have commented before, it allows us to implement non-linear amplification of the signal, as well as being one of the mechanisms used by transduction networks to implement ultrasensitive responses in transduction systems. This operation is implemented through the "cascade" network.

4.2.1, - Split network.

In order to implement the split network, we designed a new substrate Y (Depicted in Figure 96) which has a new identity strand called Strand 13. Strand 13 is analogue to Strand 3, in which they both share the same external upstream domains that interact with the catalysts when present in a major species as well as those inner toeholds that mediate the fuel consumption step. However, Strand 13 and Strand 3 differ in the downstream domain corresponding to the fluorophore used to report the catalytic activation. This means that these two strands share the same state strands and can be used as depicted in the reaction diagram of Figure 97.

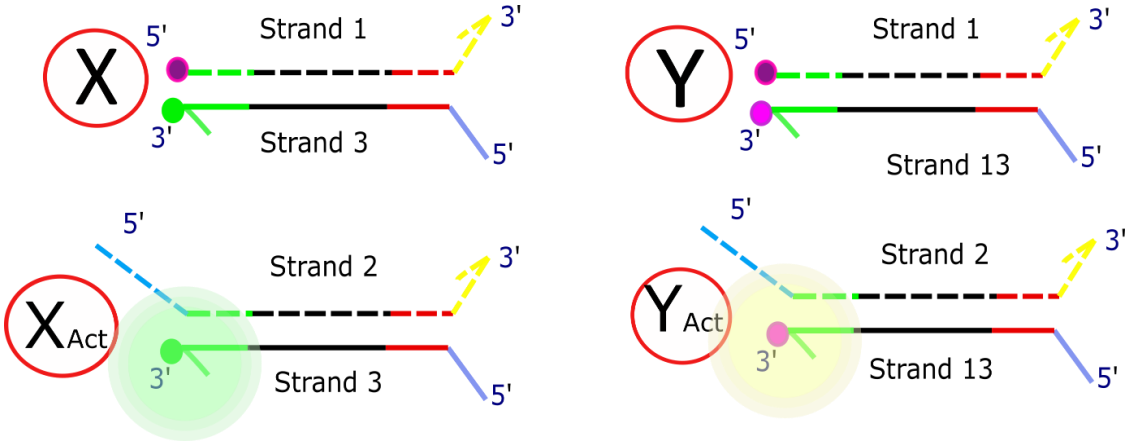

Figure 96. Strands and domains diagrams of the two substrates X and Y that are used on the split network. As we can see, their only difference is in the downstream domain -which is also translated in different fluorescence for each substrate.

In the case of Strand 13, we chose Cy3 as the reporting fluorophore, given that it was already proven to be compatible with the Iowa Black FQ quencher used in the OFF-state strand, Strand 1. But instead of making it work as the acceptor of a FRET pair, as we used it as a normal fluorescence probe with excitation at a 530 nm as noted per the general protocol described in section 2.3.1 of Chapter 2. The catalyst strands of our system were also modified in this experimental setup. Since in this case, we were not interested in the visualisation of the intermediates, we decided to use an analogue to Strands 5 (the identity strand of the catalyst K) in which the Cy5 fluorophore is replaced by an Iowa Black FQ quencher so the only contribution to the fluorescence increases comes from the active substrates X and Y. The timestep chosen for this experiment was 88 seconds in order to stay consistent with the previous experiments. The gain for both channels was set to 2000 with 70 flashes per measurement. When observing simultaneously the channels corresponding to the activation of X and Y, the first thing we noted was that contrary to what happened when doing FRET for which there was indirect excitation when using the same excitation radiation, in this case we did not encounter such problem.

Studying the activation in the green channel (Figure 98) we observed no signal when the only substrate present in the medium was Y. On the contrary, when X was present, either alone or alongside Y, we observed catalytic activation of X. This activation was notably slower when X and Y were present in the system thus indicating that the two substrates were competing for the catalyst and the fuel present in the medium and slowing each other's activation due to the mutual retroactivity that the substrate exerted. However, this retroactivity effect seemed to be limited to the speed of the reaction, rather than to the endpoint. On the other hand, observing the reactions in the channel corresponding to the signal of Cy3 in Figure 99, the image of the process we obtained was a mirror image to that obtained in the green channel switching X for Y. Activation of Y was observed when it was present either on its own or when competing with X the for the catalyst with the same decrease in rate. This situation confirms that the catalysts are really agnostic to the substrates in this iteration of the split network. This is a result we could expect since the domains interacting with the catalysts and fuels of both substrates are the same and hence.

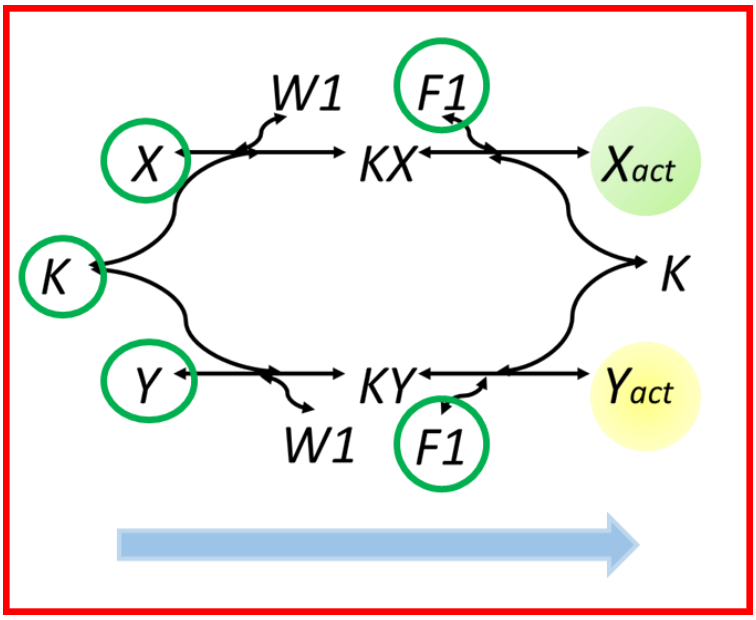

Figure 97. Reaction diagram with all the reversible reaction components of an ACDC-based split network. The species meant to produce a fluorescent signal are circled with a translucid colour circle corresponding to their wavelength emission: Green (520 nm corresponding to the emission of the Alexa 488 fluorophore) in the case of X_{act} and Yellow (580 nm corresponding to the non-overlapping emission chosen for the Cy3 fluorophore) in the case of the Y_{act} substrate. Circled in green are the species that are initially in the reaction while the reaction direction is marked by the translucid blue arrow.

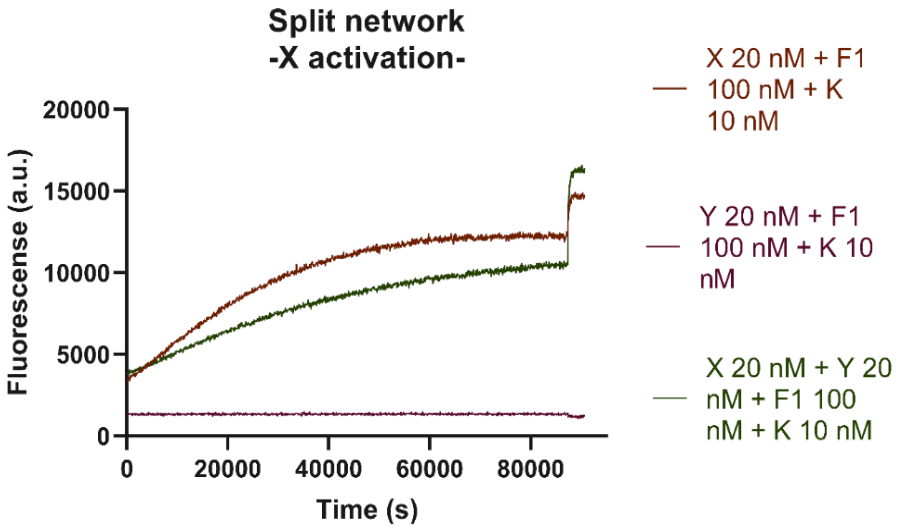

Figure 98. Fluorescence tracing for the activation of X by the catalyst K when it's on its own and forming part of the split network. Fluorescence is observed in the emission channel corresponding the emission wavelength of the Alexa 488 fluorophore (Green). We can observe the activation of X both when it is alone or competing with Y. In the latter case we can see how the two substrates compete for the same catalyst. The fluorescence jump in the end corresponds to the addition of a strand completely complementary to the identity strands of X and Y in order to release the full activation signal

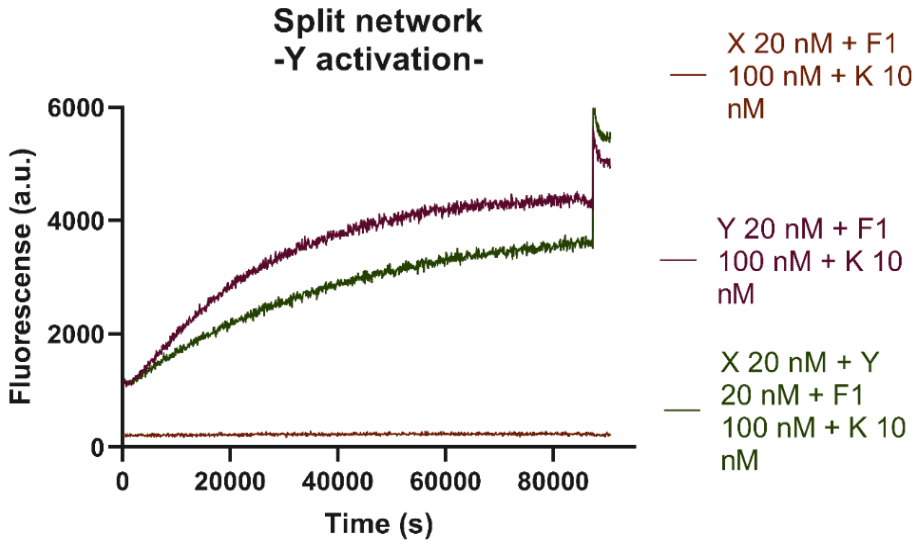

Figure 99. Fluorescence tracing of the activation of Y by the catalyst K when it's on its own and forming part of the split network. Fluorescence is observed in the emission channel corresponding the emission wavelength of the Cy3 fluorophore (Yellow). We can observe the activation of Y both when it is alone or competing with X. In the latter case we can see how the two substrates compete for the same catalyst. The fluorescence jump in the end corresponds to the addition of a strand completely complementary to the identity strands of X and Y in order to release the full activation signal

4.2.2. - Join network

In order to implement our proposed join network (Figure 100), we exploited the fact that our catalytic turnover reactions are intrinsically reversible and that the fuels and waste molecules of each reaction accomplish that role only due to their presence in excess in the medium. Hence, we can have our two catalysts from our push-pull system K and P working in the same direction if, as we did in our experiment, we use an excess of W2 in P and we do a reverse pull. No further modifications of our catalyst strands were necessary. We used the same DNA strands and experimental conditions to those we used for the catalytic turnover experiments from Chapter 3 (gain of 2000 on our BMG ClarioSTAR for the 465 nm excitation of the Alexa 488 fluorophore acting and a timestep between measurements of 88 seconds). Again, given that we only expected to have a single substrate to activate, for assembling K we resorted to the Cy3 labeled Strand 5.3 that we used previously on the FRET experiments for the single steps, the catalytic turnover experiments and the push-pull experiments.

In our test, depicted in Figure 101 we compared fluorescence outputs in three setups: (1) using the catalyst K at a 10 nM concentration with 100 nM of F1 as the driving force, (2) using the catalyst P at a 10 nM concentration with 100 nM of W2 and the real join network, and (3) using both K and P at a 5 nM concentration using 50 nM of their corresponding fuels as the pushing force. When comparing the kinetics and endpoints of each of these three cases as depicted in Figure 101, we saw no noticeable difference, thus confirming that the join action is possible and that the catalysts actions can be directly grouped if they all go in the same direction – which also has implications in terms of simplifying the modelling of the dynamics of the system.

Again, this observation reinforces the point observed in the previous subsection about the catalyst being agnostic to their chemical substrates as long as the interaction domains between these two distinct species remain the same.

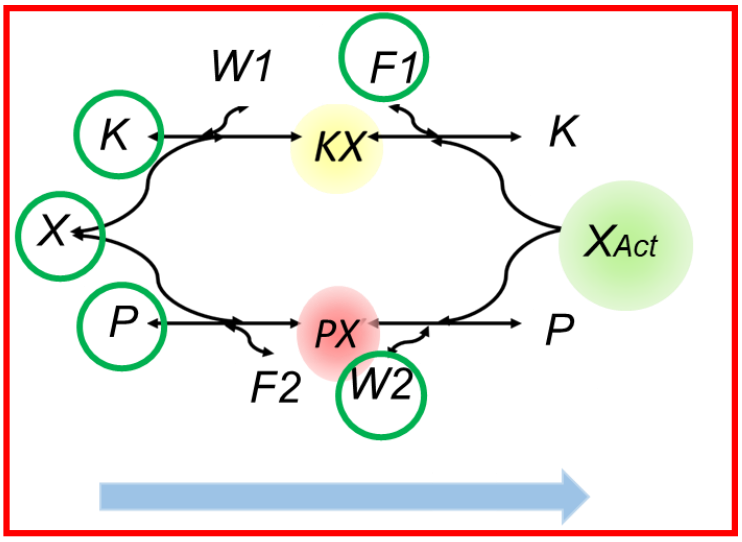

Figure 100. Reaction diagram with all the reversible reaction components of an ACDC-based join network. In the present experiment we are meant to follow the conversion of X into X_{act} measuring the emission in the green channel (520 nm corresponding to the emission of the Alexa 488 fluorophore) although we can also follow the dynamics of the catalysts. Surrounded by a green circle are all the reactants initially present in the reaction whereas the translucid blue arrow indicates the direction of the reaction

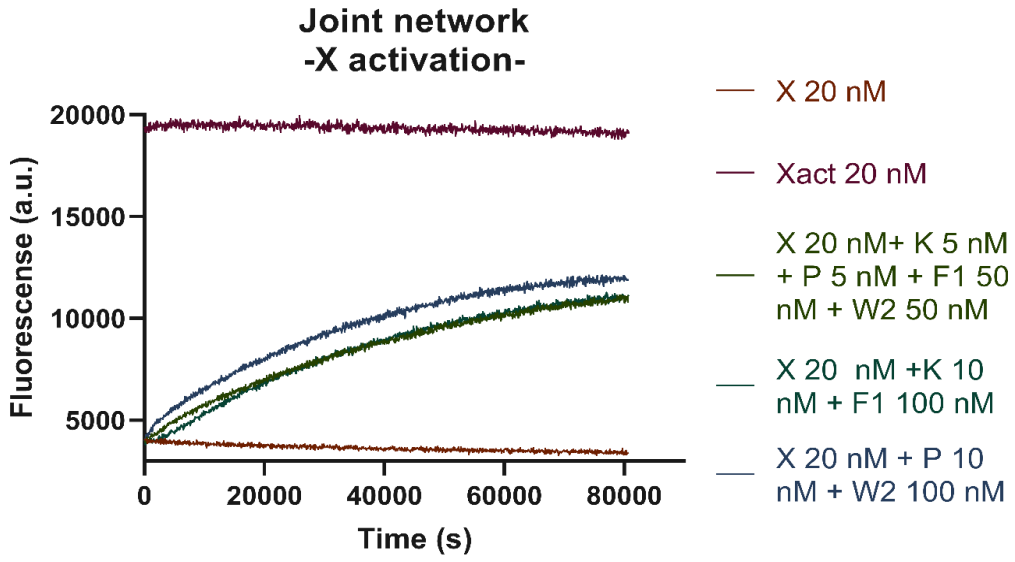

Figure 101. Fluorescence tracing in the case of a simultaneous addition of the catalysts for the join network compared with the activation by same net quantity of catalyst and fuel of a single type. The traces are observed in the emission channel corresponding the emission wavelength of the Alexa 488 fluorophore (Green). In this case we opted for a well with the fully activated substrate Xact as the control for the full dynamic range since it could be equivalent given that there was no environmental quenching due to the low dilution of the strands with a quencher. We can observe in this case that the rate is roughly equivalent for all three cases, thus meaning that effectively, the two catalysts and fuels are acting with an efficiency comparable to that of a single catalyst and the equivalent net amount of fuel.

4.3.- Extending networks into different layers: cascading networks.

In order to test our cascading system (Figure 102), we established a variant of Strand 5 (Strand 5.6) labelled with Cy3 in the downstream interaction domain E*, as well as an analogue of Strand 4 (Strand 14), which is acting as an inactive state strand that has the downstream interaction domain G switched to an Iowa Black FQ quencher. This way K is able to have two states (one catalytically inactive and the active state we have been using up so far) that we can distinguish via their signature fluorescence. Analogously to X, we designed two catalysts KK and PK that can oppose each other in the activation or deactivation of K. The experimental setup employed was recycled from the split network experiments mentioned on section 4.2 of the present chapter.

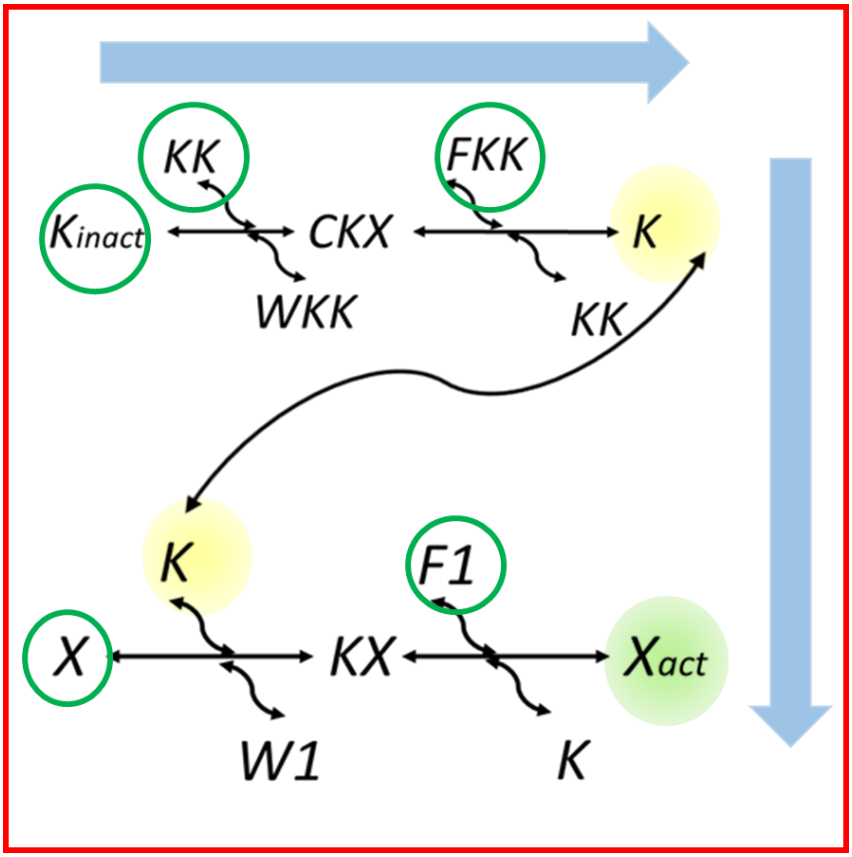

Figure 102. Reaction diagram with all the reversible reaction components of an ACDC-based cascade network. The species meant to give out a fluorescent signal are circled with a translucid colour circle corresponding to their wavelength emission: Green (520 nm corresponding to the emission of the Alexa 488 fluorophore) in the case of X_{act}. Yellow (580 nm corresponding to the non-overlapping emission chosen for the Cy3 fluorophore) in the case of the K substrate/catalyst.

The first round of experiments consisted in testing the effect of the cascade of different concentration variations of the upstream elements in the first catalytic layer of the system beginning with the substrate $K_{inact,}$ continuing with the Fuel FKK and ending with the variation

of the catalyst KK. As a general observation applicable to all the experiments, we observed that the inactive catalyst/substrate K_{inact} displayed little to no leak action when in presence of X and F1, which by itself stands as a testament to the robustness of our framework in the present working conditions.

When observing the behaviour of the cascade while varying K_{inact}, we observed that the same amount of KK and FKK were able to give proportionally the same activation in the case of [K_{inact}] = 10 nM and [K_{inact}] = 20 nM (around a 30% of the total K_{inact} present). While proportionally being the same, the process in the case of the 20 nM yielded twice as much catalyst and the reaction rate was initially faster. On the other hand, in the [K_{inact}] = 5 nM experiment, the excess of KK to the substrate led to a sequestration problem that was limiting the activation of K as seen in Figure 103.

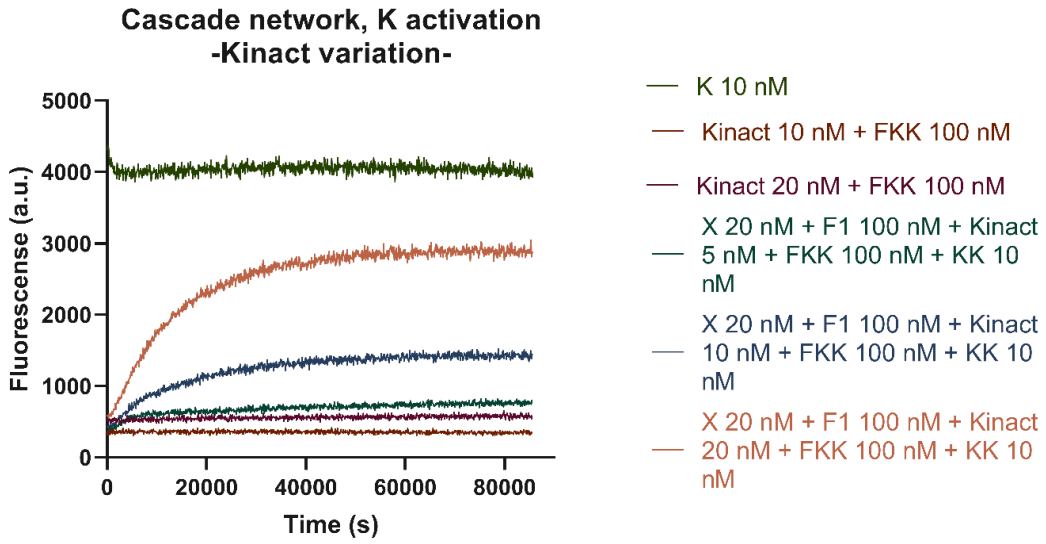

Figure 103. Fluorescence tracing for the addition of the upstream catalyst for the cascade network. Fluorescence is observed in the emission channel corresponding the emission wavelength of the Cy3 fluorophore (Yellow). In this experiment the aim is to study the effect of the variation of the amount of K_{inact} on the dynamics of the system. We observe that a greater presence of substrate K_{inact} leads to a faster initial activation rate. However the proportion of the K_{inact} activated to the total available is the same for a given fuel supply. We also observe that having the catalyst KK in excess to its substrate Kinact leads to sequestration.

Looking at the activation of the cascade network from the point of view of the downstream substrate X, we observe that, they all tend to converge to the same endpoint – which is expected given that the amount F1 available in all cases is the same. We also observe that with more K, reaction became faster and we were able to observe a moderate sigmoidal effect in the [K_{inact}] = 20 nM experiment – albeit dampened by the sequesterering effects that are inherent to our reversible reaction, as we see in Figure 104.

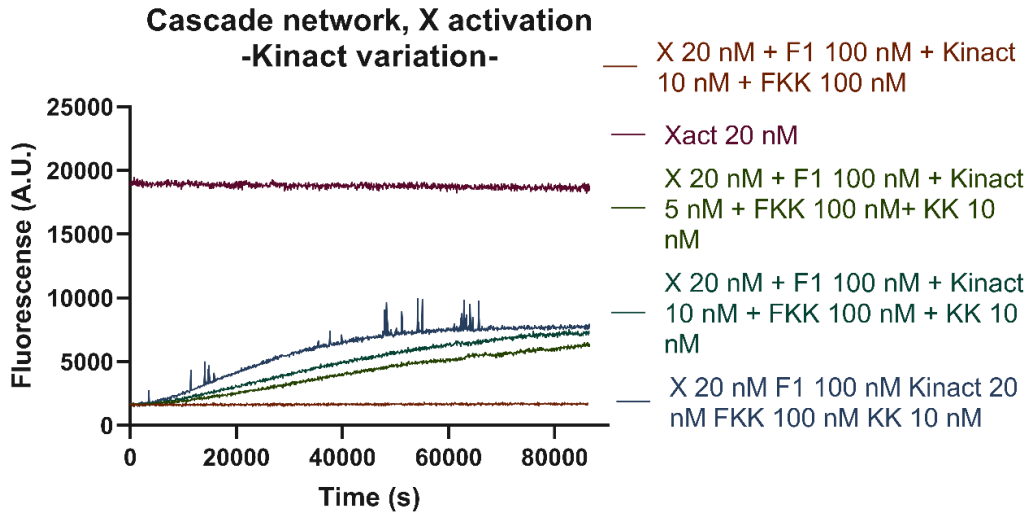

Figure 104. Fluorescence tracing for the addition of the upstream catalyst to the cascade network. Fluorescence is observed in the emission channel the Alexa 488 fluorophore (Green). In this experiment the aim is studying the effect of the variation of the amount of K_{inact} on the dynamics of the system. We observe that the endpoints go the same level in all three cases. The higher availability of K in the system leads to a faster response.

On the other hand, as seen in Figure 105, variations on the fuel FKK had a very noticeable effect on the endpoints of how much K_{inact} was turned on in the first layer. This was totally expected and is coherent with the catalytic turnover experiments of the Chapter 3 in which fuel variations keeping the catalyst fixed varied the reaction endpoint but not the speed of the reaction. Similarly, the amount of K turned on by a given fuel supply does not scale up linearly for the very same reasons -mainly product inhibition and the loading effects present on the system

However when we observed the effect of the variation of FKK in the activation of X as depicted on Figure 106, we observe that the greater activation of K in the upper part of the cascade that results of having a bigger fuel supply makes the activation of X faster. However, while the effect in the activation is notable, it is not as visible as in the case of the K_{inact} variation previously studied. Moreover, in any case of this particular experiment we observe any of the activation curves plateauing.

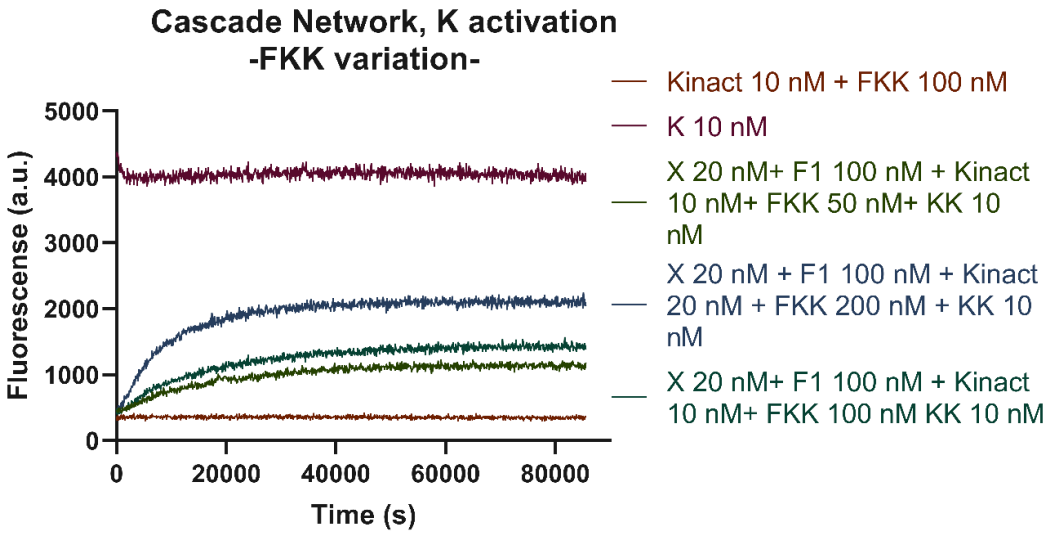

Figure 105. Fluorescence tracing for the addition of the upstream catalyst in the cascade network. Fluorescence is observed in the emission channel corresponding the emission wavelength of the Cy3 fluorophore (Yellow). In this experiment the aim is to study the effect of the variation of the amount of FKK in the dynamics of the system. The behaviour is the expectable to a fuel variation experiment in which the rate is the same in all cases, but the endpoint varies with the amount of fuel.

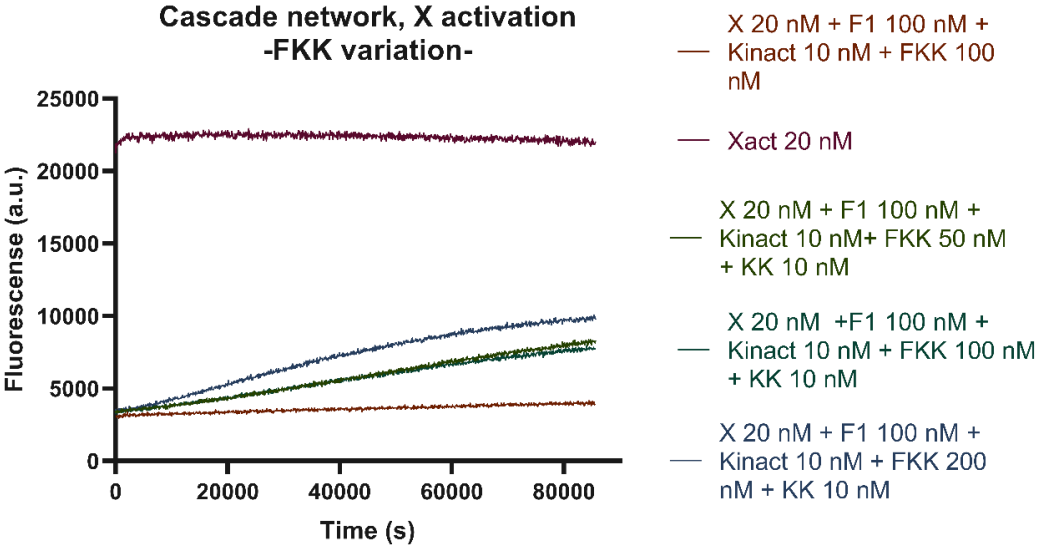

Figure 106. Fluorescence tracing for the addition of the upstream catalyst in the cascade network. Fluorescence is observed in the emission channel the Alexa 488 fluorophore (Green). In this experiment the aim is to study the effect of the variation of the amount of FKK in the dynamics of the system. We observe dynamics correlated to those observed in Figure 103 and as expected, greater amounts of K activated result in a faster activation of X. However, the variation of FKK is less impactful on the activation of X than varying K_{inact} and we don't seem to observe any of the curves plateauing.

Finally, we studied the effect of varying the concentration of the upstream catalyst KK in the dynamics of the full cascade. As we observed in Figure 107 that the greatest turnover is produced with the lower catalyst condition.

This might seem counterintuitive at first. However, we need to remember that in our ACDC Framework reaction intermediates of a catalytic turnover are as stable as the reaction

products. Hence a catalyst excess is not going to favour a faster turnover but rather to make the substrate spend more time being sequestered (and as we saw in section 3.2.2 of Chapter 3, increasing the amount of fuel on the system to limit the quantity of intermediate present is limited by the fuel turnover rate of the catalyst).

This sequestration upstream in turn results in a lower availability of K to activate the second layer of the cascade as seen on Figure 108. As we saw in the K_{inact} variation experiment, in the sequestered conditions, less K is being released in the medium, and it is released slower. Hence the conditions with the least sequestering ([KK] = 5 nM) leads to the shaper, more sigmoidal-like dynamics.

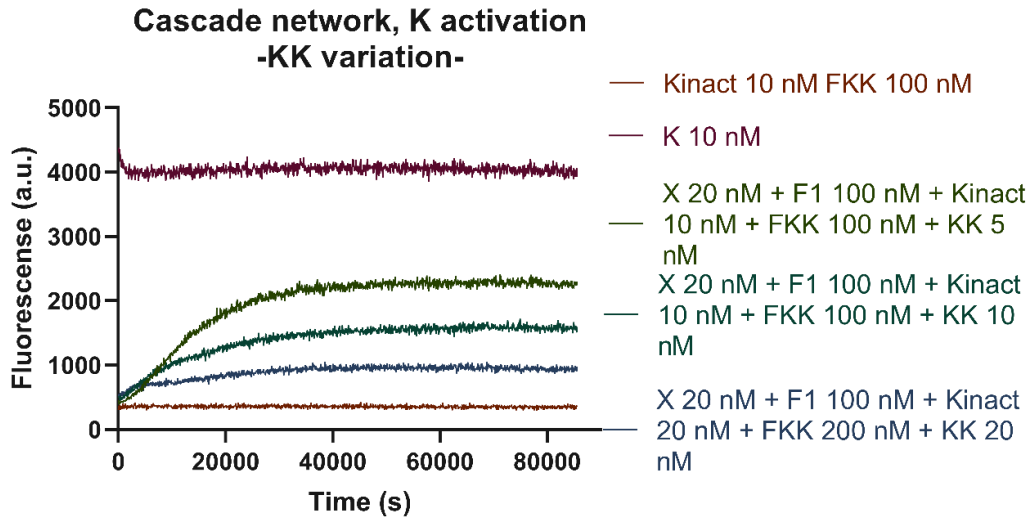

Figure 107. Fluorescence tracing for the addition of the upstream catalyst in the cascade network. Fluorescence is observed in the emission channel corresponding the emission wavelength of the Cy3 fluorophore (Yellow). In this experiment the aim is to study the effect of the variation of the amount of KK on the dynamics of the system. We again observe that having the catalyst KK in excess to its substrate K_{inact} leads to sequestration and limits the activation of K.

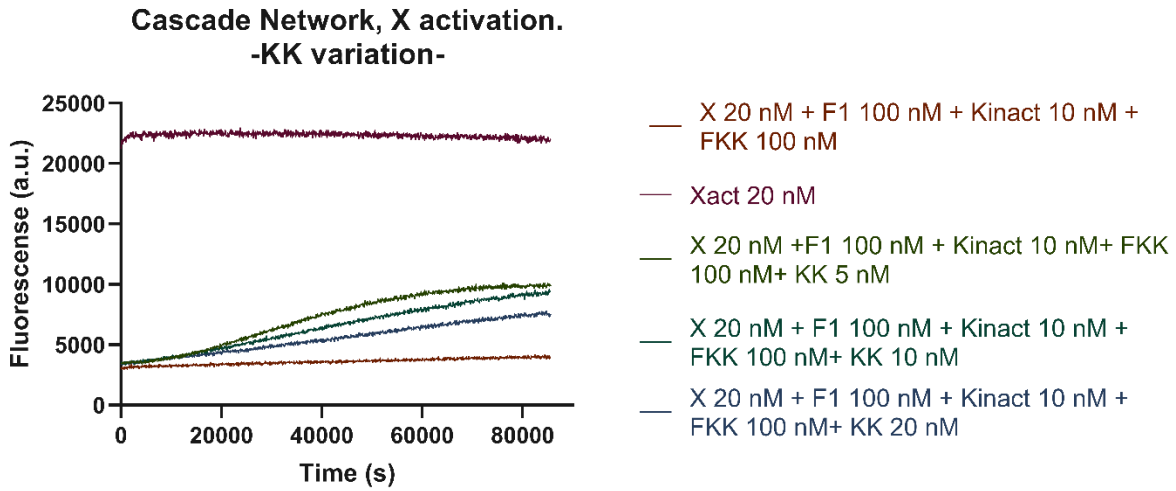

Figure 108. Fluorescence tracing for the addition of the upstream catalyst to the cascade network. Fluorescence is observed in the emission channel corresponding the emission wavelength of the Alexa 488 fluorophore (Green). In this experiment the aim is to study the effect of the variation of the amount of KK on the dynamics of the system. Again, we can observe that the low turnvover of K upstream limits the reaction speed of the activation of X. However, the faster cases seem to converge in the same endpoint, which is expected given the F1 supply is the same in all three cases.

We conclude that it is possible to build cascade system using the ACDC Framework, evn though ACDC-based cascade systems suffer of sequestration issues that affect the activation speed of the reaction if catalysts are in excess with regards the substrates they act upon.

On an ending note for this chapter, we must note that, as we described at the beginning of this subsection, we designed two different catalysts that can act simultaneously on K_{inact}/K. Hence, we can demonstrate that push pulls can be cascaded and make them act as functional subunits in the cascades (Figure 109). Furthermore, we can tune the input and output of each subunit by varying the catalysts. This molecular motif was proposed by McBride, Shah and Del Vecchio[100] as a molecular insulator motif – although simple push-pull motifs can perform such insulation task – which, when implemented between different parts of a biomolecular circuit in which one exerts a loading effect, can allow the designer to restore the modularity between the parts. This consideration can be traced back to the fact that traditionally, push-pulls have been considered insulator motifs. However, as we have seen in the section about zero-order ultrasensitivity, single push-pulls suffer internal loading effects that might make them not suitable to such modularity restoration tasks.

In order to test the functionality of the push-pull cascade, the working conditions, of the fluorophores' activation were kept from the cascade network experiment. However, the timestep was modified to a 147 seconds one. This modification was done due to the fact that the catalysts amounts on each layer have been halved to keep sequestration low. With this setup, we programmed the push-pull cascade to reach the quasi-steady state of the first layer at $[K_{inact}] = [K]$. In principle the basic desired features of the network such as the leakless controlled activation are present. Moreover, when analyzing the first catalytic layer of the network we observe that, as per Figure 110, the activated amount of K is the one we prescribed.

The second layer of the cascade (Figure 111) is also meant to give a roughly equal ratio of X to X_{act}. However, when observing the dynamics of the second layer, we observe that only a 25% of X is turned into X_{act} and this is happening at a rate that is half of that of the first layers. This effect might be due to the fact that in the initial conditions of the system, the catalyst activation is not synchronized. Hence, at time = 0 K is not active while P is active and able to bind to X. In this situation, P is able to sequester in PX a sizeable quantity of the available fraction of X and add another source of retroactivity.

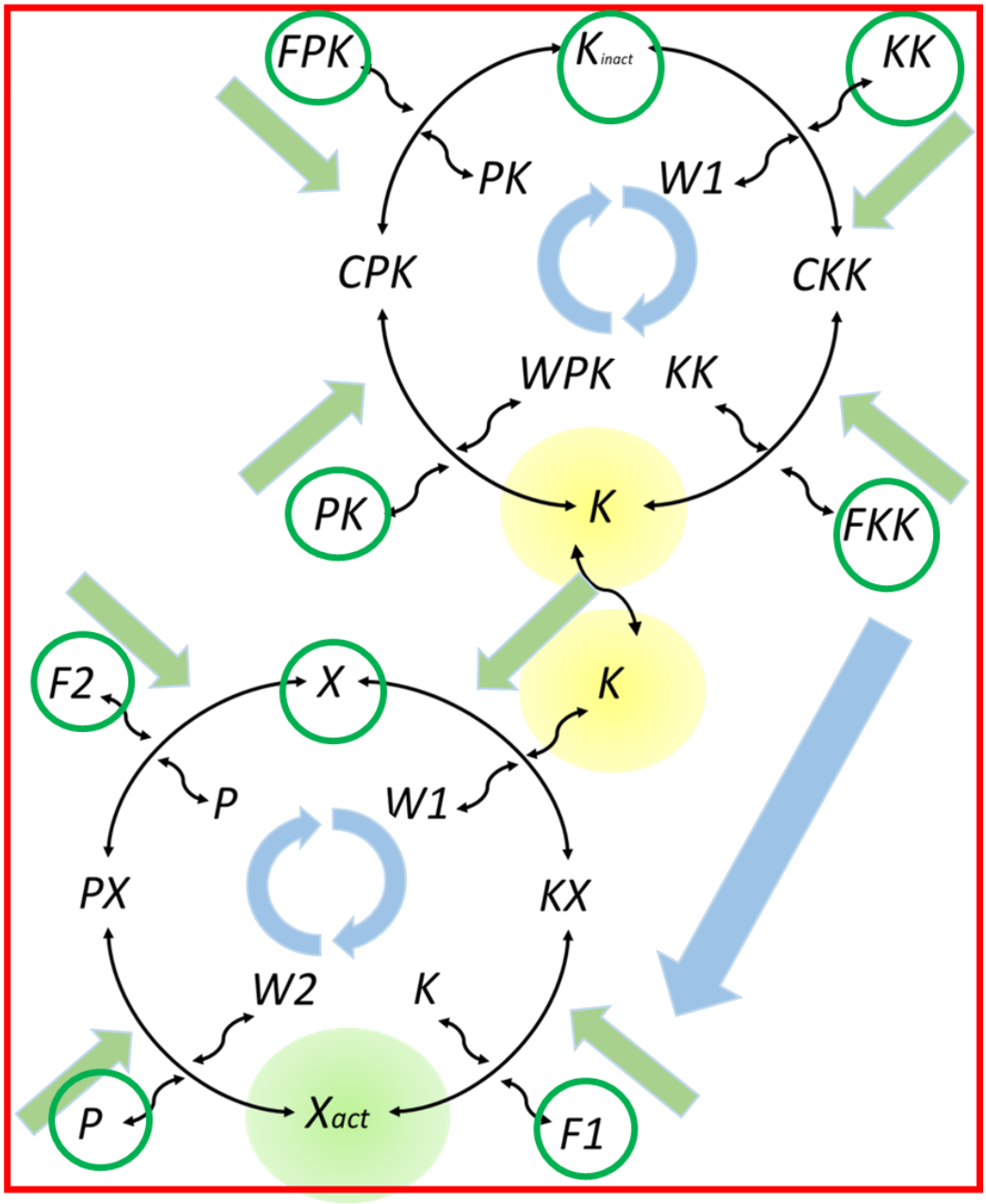

Figure 109. Reaction diagram with all the reversible reaction components of an ACDC-based push-pull cascade network. The species meant to give away any fluorescent signal are circled with a translucid colour circle corresponding to their wavelength emission: Green (520 nm corresponding to the emission of the Alexa 488 fluorophore) in the case of X_{act}, Yellow (580 nm corresponding to the non-overlapping emission chosen for the Cy3 fluorophore) in the case of the K substrate/catalyst. In the present case, none of the catalysts other than K have a fluorophore attached to them but rather an Iowa Black FQ quencher since, again, like in the regular cascading network, we are only interested in the signals resulting from the activation of K and X. The blue arrows indicate the directionality of the reactions, whereas green circles indicate which species are initially present in the reaction. As per the convention of Figure 90, the green arrows indicate in which directions the different species of each push-pull are being incorporated

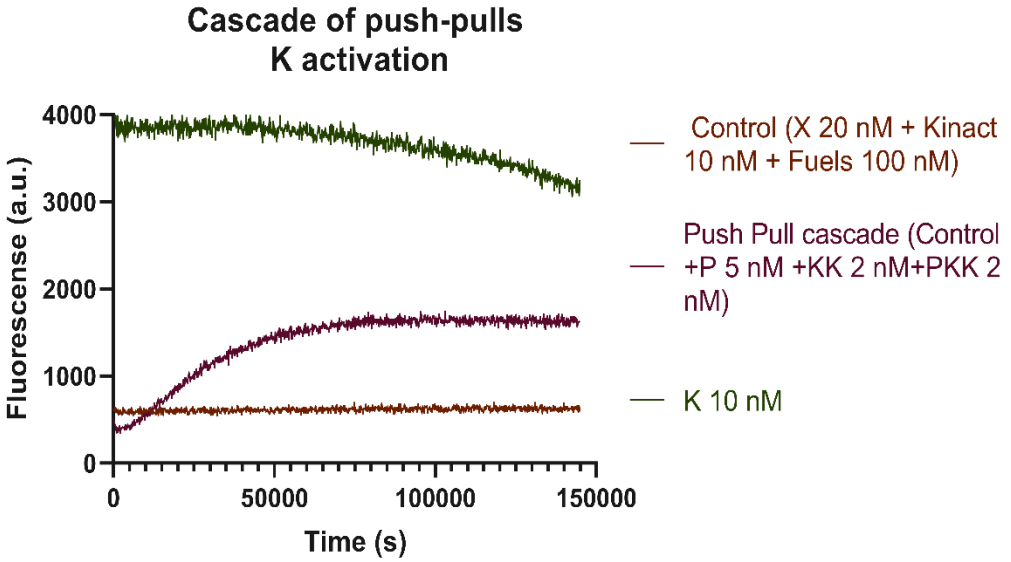

Figure 110. Fluorescence tracing for the addition of the upstream catalysts to the push-pull cascade network. Fluorescence is observed in the emission channel corresponding the emission wavelength of the Cy3 fluorophore (Yellow). In this channel, we are able to observe how the upstream push-pull goes to the prescribed out-of-equilibrium and equilibrium states.

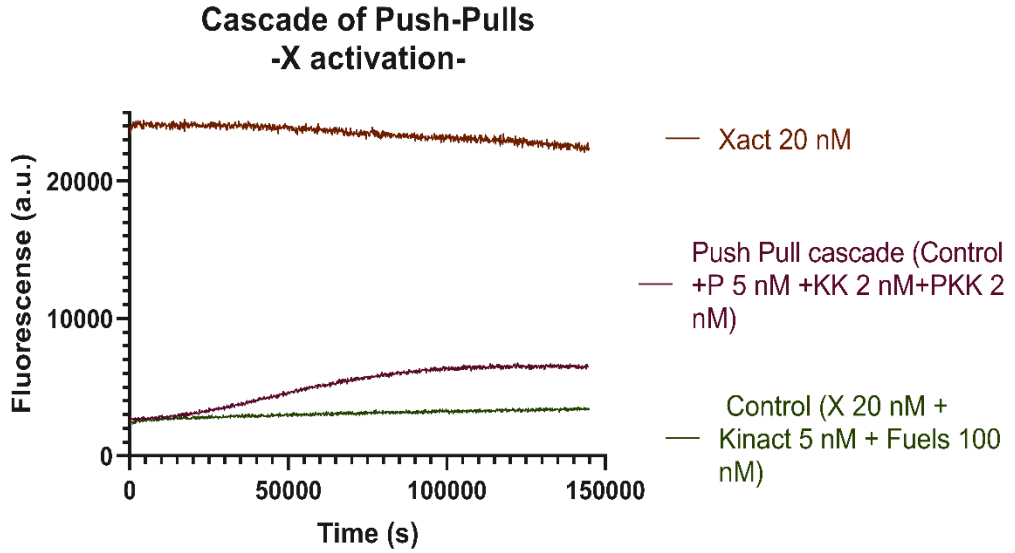

Figure 111. Fluorescence tracing for the addition of the upstream catalyst to the push-pull cascade network. Fluorescence is observed in the emission channel corresponding the emission wavelength of the Alexa 488 fluorophore (Green). Despite having prescribed a 50% activation state as the upstream push-pull network from which its catalyst K comes from, retroactivity of the P catalyst alongside sequestration leads to dampened dynamics both in terms of reaction rate and endpoint.

In conclusion, we can assert that the ACDC Framework allows us to build the catalytic reactions required to construct a push-pull motif – which is widely considered to be the fundamental functional subunit upon which transduction networks are based. Moreover, we have demonstrated that these nucleic acid-based subunits really perform out-of-equilibrium similarly to their protein counterparts and can be integrated in larger, extended network

configurations. However, sequestration and product inhibition can hinder the functionality of these networks compromising properties such as the possibility of zero-order ultrasensitivity or the chance of doing controlled, transient activation. Hence, in the next and final chapter of the thesis we will explore a strategy to potentially mitigate these problems.

Chapter 5: Functional and formal limitations of the ACDC Framework. Troubleshooting via mismatch implementations.

5.1. - Summary of the limitations of the system.

As we have seen in the previous chapter, the ACDC Framework allows us to build the push-pull motif that constitutes the minimal functional unit of transduction networks and successfully implement the properties that make these out-of-equilibrium systems so remarkable for synthetic biology and DNA nanotechnology. In addition to this, we have also observed how the catalytic reactions that form the push-pull motif can be incorporated in extended networks with added functionality via the signal splitting, signal joining and cascading operations. However, despite these successes in designing and implementing a working framework for CCNs that emulates functionalities of transduction networks, some major drawbacks have been found and highlighted throughout the previous chapters of the present thesis. The drawbacks found fall inside two different categories related to the effect that these drawbacks cause.

The first category of limitations that we have identified are the **functional limitations.** When we talk about functional limitations, we talk about those that limit how a network built using the ACDC Framework will perform with regards the desired behaviour. These limitations in problems in aspects of the network such the reaction rate, the dynamic range of the activation or the responsiveness of the network.

On the other hand, the other type of drawbacks found in the ACDC Framework are those we have decided to name as **formal limitations.** These formal limitations are much more fundamental in their nature and they relate to the constraints inherent to the reaction mechanism the framework is based on and how these constraints limit which network motifs can be realized inside the ACDC Framework. The present section of this chapter will detail both types of limitations and their effects in the system design

5.1.1, – Functional limitations of the ACDC Framework.

The first functional limitation we observed on our system was the lack of a significant enthalpic free energy component in its single step reactions. Although we designed the system to work only with an entropic drive in order to minimize the energetic expense of the operation of the system, this strategy proved to be of limited use. This situation resulted in the fact that all reaction steps were essentially reversible. While this situation might not seem

detrimental in principle, it entails that reaction steps in the opposite direction of those we have designed are present and interfering significantly with the performance of the CCN. As a consequence, this lack of intrinsic enthalpic free energy results on the need to add of increasingly high reservoirs of fuel with diminishing results. This in turn made complete turnover in either direction hard to obtain.

Related to this reaction step reversibility problem appeared the second of the functional limitations of our system functional limitations in the shape of product inhibition. Due to the fact that all the reaction steps that form a catalytic reaction in the ACDC framework are reversible, we observed that the second reaction step could be engineered to go in either direction. This means that, after a catalytic turnover reaction, the product could bind back to the catalyst and trap it in an intermediate complex. This inhibition in turn limits the availability of the catalyst and makes it less effective the further the system was pushed in either direction given.

The third limitation inbuilt in our system's functionality was related with the stability of the intermediate complex that forms after a catalyst and a substrate bind with each other. As we observe in Figures 88 and 89 in Chapter 3, the signal of the complex PX was almost the highest possible, thus indicating that in practical terms the catalyst was mostly bound to the substrate and was rarely free in the system. The existence of these long-lived complexes in turn results in a loading effect on the catalysts that at the same time causes dampening of the dynamics of the system. These loading effect in turn can cause mutual retroactivity between two different substrates if they share a catalyst and a fuel supply. In addition to this, the occurrence of these loading effects is also helped and exacerbated by the possibility of product inhibition in the catalytic reaction. Moreover, if the catalyst was not working on a saturation regime and it was more abundant than the catalyst, this sequestration problem would become even more prevalent.

In addition to these problems, we also must add the unexplained anomaly on the endpoint of the single step experiments. In this anomaly we found that experimental setups that were meant to converge to the same endpoint (due to having the same strand composition in different initial configurations of duplexes) did not get to converge. This unexplained anomaly imposes limitations on the dynamics of the system at the single step level that might translate into – and maybe cause – some of the previous issues.

However, in addition to these functional issues, as we described at the beginning of the chapter, some fundamental formal problems exist within the ACDC Framework and will be the object of the discussion of the next subsection of the chapter.

5.1.1. - Formal logical limitations of the ACDC Framework.

As we stated in the previous subsection, besides the functional issues that affect the speed or the dynamic range of any prescribed network in the ACDC Framework, there are issues that affect which networks can be realized due to the very mechanism that the framework is based on.

To begin with, the most basic limitation we have found is the number of different toeholds our systems can use. As noted by Johnson and Qian[126], the specificity of recognition of different species in 4-way branch migration-based systems comes from the toeholds rather than from the displacing domains. With a toehold length of 5 nucleotides we have a total of 5^4 = 625 different combinations of bases to make toeholds. These 625 sequences give place to a total of 312 possible interactions in our system. However, the size of the network cannot be deducted directly from this estimation. In the first place, as we saw in chapter two, for two species to interact we need to have two different complementary double toeholds to trigger an interaction between two species. The fact that toeholds act in combinations adds greater freedom in terms of potential design space. But on the other hand, there is a non-negligible possibility of cross interactions between toeholds that are not completely complementary. This poses a risk in terms of possible leak or crosstalk. In conclusion, we can conclude that we cannot build an arbitrarily large network of any kind.

Even when taking this limitation into account, even within the size constraints of the system, there are certain network motifs that cannot be inside the ACDC Framework. Moreover, the realisability limitations, as proved by the work by Antti Lankinen[99] and the author and codirector of the present thesis are inherent to the 4-way strand exchange mechanism the whole ACDC Framework is based on.

The limitation that we face in that regard is that autocatalytic motifs in which a given species can activate itself are forbidden by the framework due to the fact that, as stated in the aforementioned paper by Lankinen and the author of this thesis[99], its realisation would require at some point to have three or more complementary toeholds binding together, which breaks the foundation of the reversible 4-way branch migration reaction upon which the system is based. If, in a given reaction step, any species inside the framework had three or more complementary toeholds bound, that step would be maximizing the number of basepairs and thus creating a free energy well in the system that acts as a sink that can't be driven out of equilibrium.

More worryingly, however, multi-step cascades and hence loop architectures are also severely limited inside the ACDC Framework. As observed in the work by Lankinen[99] and co-authors, the ACDC framework is limited when it comes to implement networks with more than 4 distinct catalytic layers (which in turn limits the possibility of implementing functional loops of any kind).

This limitation arises from the fact that the activation-deactivation processes within the ACDC Framework are based on toehold exchange. These toehold exchange processes entail that the same strands have different roles and thus toeholds are present in different catalytic layers and are recycled from one layer to another. Due to toehold recycling, any cascade with more than 4 layers would face a situation in which ancillary species from the first and the fourth layer would potentially bind one to each other by stable 10 nucleotide duplexes, as depicted in Figure 112.

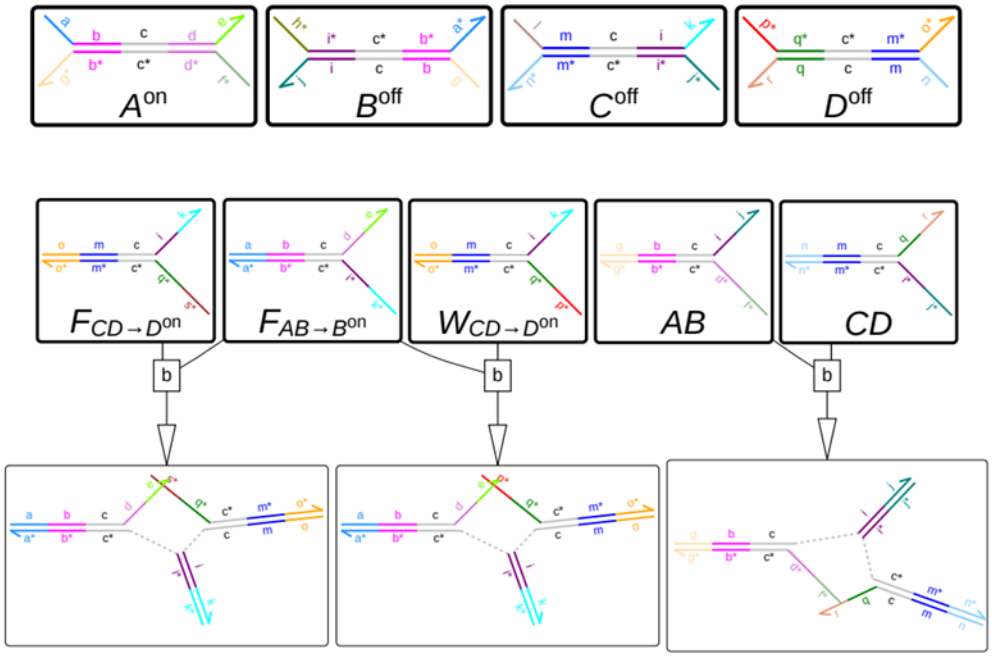

Figure 112. Example of dimerization via toehold binding in a theoretical 4-layer catalytic cascade network. On the first row of the figure, we can find the major species that constitute the catalysts and substrates. On the second row, we can find examples of the unwanted dimerization of ancillary species – whether they are fuels/waste or reaction intermediates via the formation of the unwanted 10 basepairs duplexes – that occur as a consequence of the recycling of the toeholds throughout the network. Adapted from Lankinen et al (2020)[99].

However, the situation is not necessarily unsolvable and in the next subsection of the chapter, we will provide the basic idea for the introduction of a hidden thermodynamic drive in DNA strand displacement systems, and show how this basic idea can translate into the ACDC framework, potentially mitigating both types of functional issues.

5.2.- Introducing the concept of destabilization and thermodynamic drive via mismatches.

As we saw in the first chapter of this thesis, almost all enzyme-free dynamic DNA nanotechnology systems are based on toehold-mediated strand displacement. In these systems, the toehold triggers the displacement process and the length of the toehold is the dominant factor governing the reaction rate of the process.

However, this approach has problems when in terms of the implementation of finely tuned control over the thermodynamics and kinetics of a given step. In principle, it is possible to give a thermodynamic bias to a toehold-mediated strand displacement reaction by ensuring the products have more basepairs than the initial reactants. While this approach is valid, it creates problems when the input molecules have to be recycled, as it should be for catalysis-based reaction frameworks such as are the seesaw gates designed by Winfree and Qian [55] or our ACDC Framework.

We will take the seesaw gate as the example in which we will illustrate the subtleties we want to discuss (see Figure 113). Implementing an extra drive on the first step of the reaction via basepair maximization of the binding of invading strand (A in Figure 111) with the gate molecule D would result in a higher release of the signal molecule B. However, this extra drive would result into the trapping of the input on the gate molecule. In principle, adding an even higher drive on the second step making the fuel bind not only to the toehold and the displacer domain, but also to part of the toehold the input has previously used to bind to the gate, would look like a feasible solution. However, this addition of extra binding basepairs to bias the reaction also means that the fuel molecule C can invade via a small toehold the gate-output complex BD directly in the absence of A, which results in a loss of catalytic control via a leak reaction between the fuel and the gate-output complex. However, there are solutions that allow us to decouple the implementation of a thermodynamic drive from the binding of a DNA strand to a toehold.

As seen in research by Andrew Tuberfield's group [127-] for which the author provided experimental results-- the implementation and subsequent repair of mismatches can be a way by which the addition of a thermodynamic drive can be achieved without leak-enhancing toeholds.

According to the model developed by John Santalucia[128, 129], the presence of a mismatch (i.e. the presence of a basepair in a duplex structure that does not correspond with a Watson-Crick pair) disrupts and destabilizes the double helix structure. The effect of this destabilization is dependent of the position of the mismatch inside the duplex. If the mismatch is placed at the near-end of the DNA duplex, the surrounding basepairs can fray to accommodate and compensate for the disruption. However, if the mismatch is placed deeper insider the duplex, the destabilizing effect is maximized.

Sufficiently long duplexes with a mismatch are kinetically stable and don't spontaneously dissociate despite their higher free energy when compared with the mismatch-less versions of the same duplex. Hence, we decided that placing one of these mismatches in the centre of a DNA duplex and repairing this mismatch in one of the steps of a catalytic turnover (more precisely on the regeneration of the catalyst A) drives the system to the formation of a stable duplex CD. The repair of such mismatches during a strand displacement reaction step is favoured thermodynamically, thus making the mismatch-bearing state of the duplexes a metastable state. This metastability implementation strategy has noticeable effects on the thermodynamics of the reaction, providing a clear direction for the free energy of the reaction, while at the same time keeping the leak reaction rate low due to the lack of an accessible toehold to the fuel to invade in absence of the input.

This metastability and its associated mismatch repair strategy – dubbed hidden thermodynamic drive – was successfully implemented on basic circuits such as a pulse generator or the seesaw gate depicted in Figure 113. As expected, in the latter, we were able to observe a very noticeable increase on the turnover per molecule of fuel as well as an increase in the reaction rates without causing any new leaks between the gate-output complex in the process.

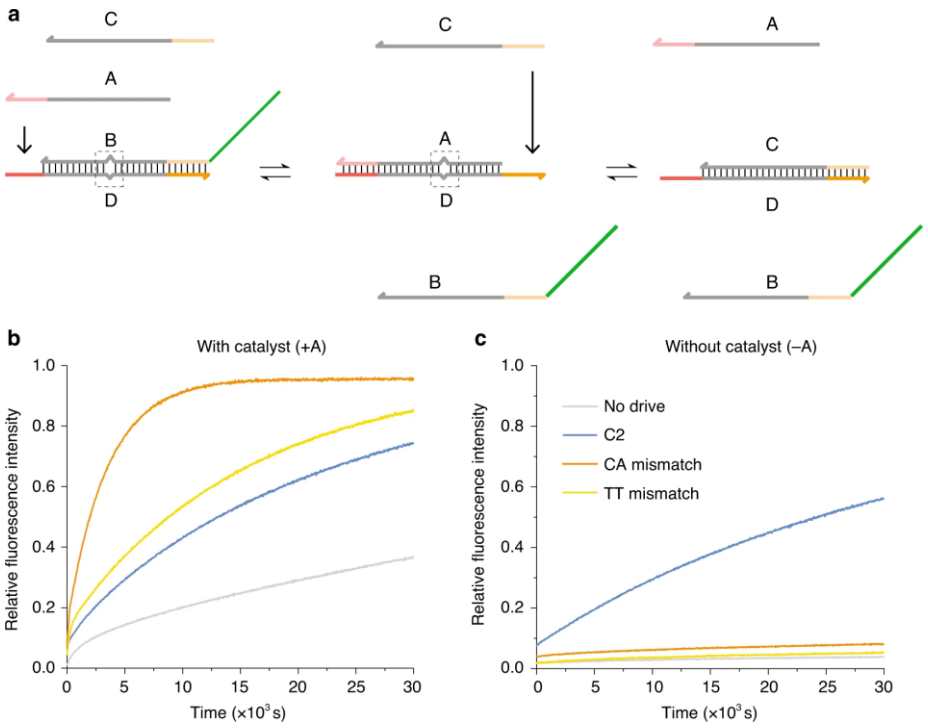

Figure 113. Implementation of a hidden thermodynamic drive in the seesaw gate motif. (a) Displays the diagram of the seesaw gate with the hidden thermodynamic drive. The system begins with the gate-output complex BD in which the catalytic input A binds to the toehold and triggers a strand displacement reaction that keeps the mismatch present on the initial BD duplex. The AD duplex gets invaded by the catalytic fuel C, which triggers the recycling of the input A while repairing the mismatch initially present. Once released the strand B interacts with a fluorophore reporter system initially present in the reaction, thereby triggering emission of the fluorescence signal we measured. Dynamics of the B strand and the reporter were probed independently and the release reaction of the reporter was fast enough to consider the release dynamics as instantaneous when in presence of B. (b) Signal triggering of the motif with and without the hidden thermodynamic drive as well as with as well as with a control with two extra basepairs in the 3' extreme of the fuel molecule. We can observe that without drive of any kind the turnover is limited, while the addition of a thermodynamic drives – especially CA mismatches – increases the turnover of the system. (c) Leak reactions triggered by the different fuels without the action of a catalyst on the reaction. No significant leak reaction – with the subsequent conservation of the catalytic control – was observed on the mismatch-based systems as opposed to the C2 fuel. Figure adapted from Haley et al (2020). Results obtained experimentally by the author of the present thesis. Sequences and experiment design are available in the supplementary information.

Having proved the usefulness of mismatches as a tool to bias the thermodynamics of the system, we will now describe how mismatches offer potential solutions to the problems found in our system.

5.2.1- Application of mismatch repairing strategies to the ACDC Framework.

As we described in section 5.1, the types of problems found in the ACDC Framework belong to two different categories: functionality problems and formal logical problem. The first type of problem comprises two basic but interrelated phenomena occurring during catalytic turnover: the fact that the ancillary catalyst-substrate complex is stable, leading to

the appearance of sequestration and a loading effect over the catalyst, and the fact that catalysts are not able to distinguish between the different activation states of a given substrate, hence leading to the appearance of product inhibition – which in turn makes the sequestration problem even more prevalent.

The first of these functional drawbacks can be solved through the implementation of a thermodynamic drive. Adding this drive, we can create a free energy difference between the two sides of the reaction that effectively that drives the system towards the regeneration of the catalyst. This in turn results in a shortening of the lifetime of the ancillary intermediate and alleviates the loading effect on the catalyst, driving further the reaction, as well as, potentially, eliminating or alleviating all the loading effects that occur when the substrate is not saturating the catalysts as we saw during the cascading experiments of Chapter 4.

The first consideration to take when designing a thermodynamic drive is considering in which step of the reaction it should be implemented and in which molecular species it should be placed. However, the characteristics we want to keep in our reactions give constraints on these placement choices.

The first constraint is that the reactions in the ACDC Framework are catalytic. Hence, the catalyst must be recovered and be usable after the reaction. Hence this conservation of functionality of the catalyst discards the catalyst as a possibility in which to implement the hidden thermodynamic drive.

The second characteristic that constrains the placement of the drive is that, if we want to implement the push-pull motif, we need to be able to implement reactions that can switch effectively the two activation states of a substrate. Hence having the reactants and products with a clear energetic bias is detrimental to this switching between states since it would skew one of the reactions of the push-pull cycle greatly. This situation leads us to the conclusion that the only species in which the hidden drive can be placed without compromising functionality is in the fuel molecule. This establishes that the mismatch should be repaired in the second step of the catalytic turnover, analogously to the implementation of the seesaw gate at Figure 113.

However as important as the choice of the reaction and the species is the choice of mismatch placement inside the duplex. In principle, the highest thermodynamic drive is achieved by placing the mismatch inside the central stabilizing domain common to all the species. However, in this case we hypothesize that the optimal place to put the mismatch is 4 nucleotides inside the bound external domain of the fuel, thereby placing the mismatch in the state strand for the substrate as seen in Figure 113. This design choice is justified for two main reasons. First, with this exact placement, we can ensure that the core of the catalyst is conserved without any alterations of the catalytic turnover process, which was one of the main initial motivations for the development of the ACDC Framework.

With this specific mismatch placement, the catalyst gets recycled while the substrate gains two new toeholds. These new toeholds are placed both in the upstream and the

downstream interaction domains. While the latter toehold exchange was already featured in the mismatch-less version of the framework, the former wasn't. This variation in this toehold makes the re-binding of the catalyst to the product harder than it was in the original setup. As illustrated in Figure 114, it reduces the tendency of the product to bind back to the catalyst, hence solving the product inhibition problem.

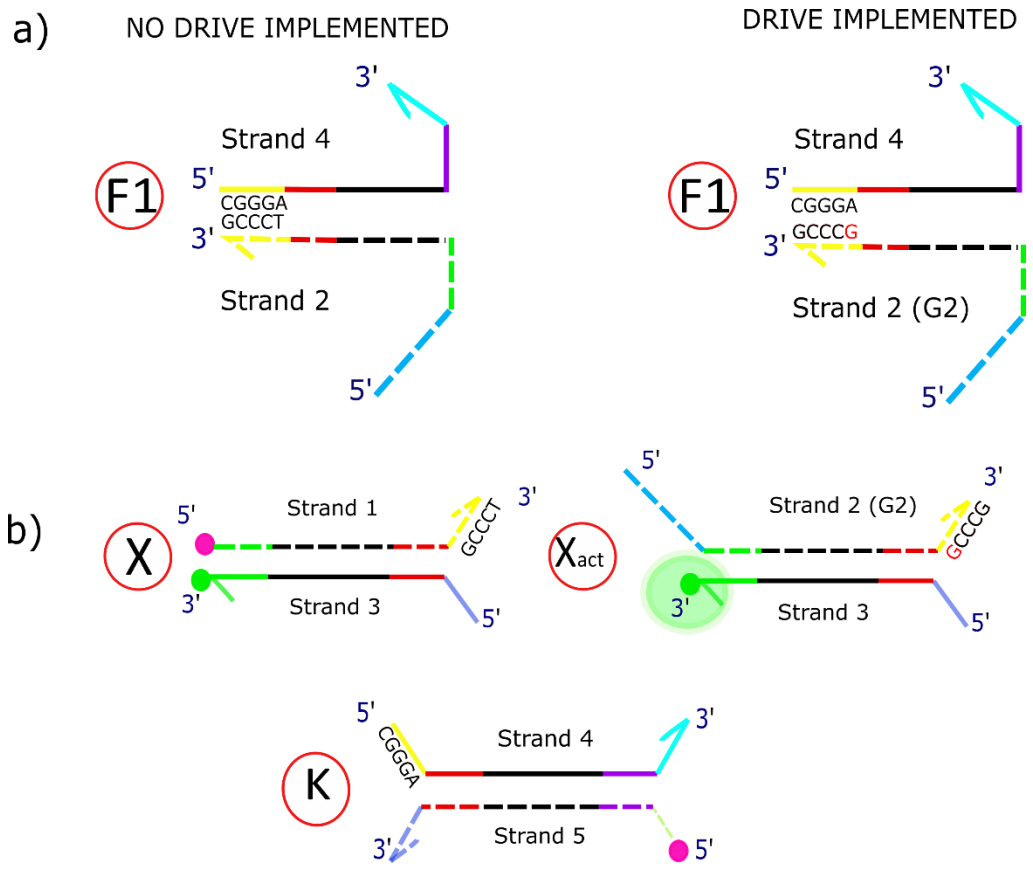

Figure 114. a) Diagram of the implementation of the hidden thermodynamic drive on the fuel F1 of the catalytic activation push. With the help of a mismatch hidden 4 nucleotides inside the duplex (in the yellow domains of the diagram) we are able to implement a de-destabilisation in the duplex that gets repaired once the fuel is consumed and the Strand 2 (2G) that was part of F1 becomes the state strand of X_{act}. This way the reaction gains an enthalpic thermodynamic drive. b) In addition to the thermodynamic drive, the turnover of X into X_{act} using the fuel with the drive results in the exposure of the toehold that was bound with the mismatch. The fact that X_{act} binding back to K would entail reforming a mismatch acts as a safeguard against product inhibition in the system.

It must be noted that, as per the aforementioned work by Lankinen, Ouldridge and the present author[99] our implementation strategy alongside the fuel orthogonality between catalysts allows us to keep specificity between the substrate activation state and the catalyst for the two catalysts that act upon a substrate in a push pull motif.

To test the adequacy of the strategy, we implemented the hidden thermodynamic drive in a single-step reaction, as well as in a full-catalytic turnover reaction and a cascade

network (as seen in Figures 115 to 117). In all this cases, the implementation of the drive on at least one of the reaction steps, makes the reaction irreversible.

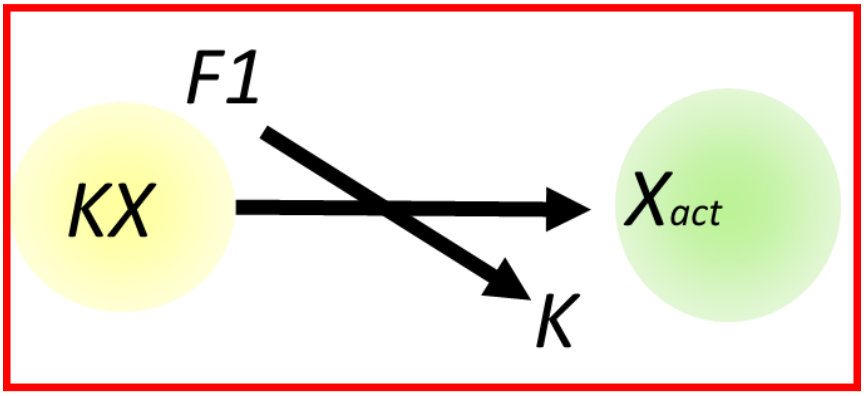

Figure 115. Diagram of the effectively irreversible second step of the push reaction with the thermodynamic drive. The reaction is marked as irreversible with the mismatch strategy, hence the arrows in the schematic representation follow only one direction and are marked with a bolder weight than the original diagrams depicted in Chapters 3 and 4.

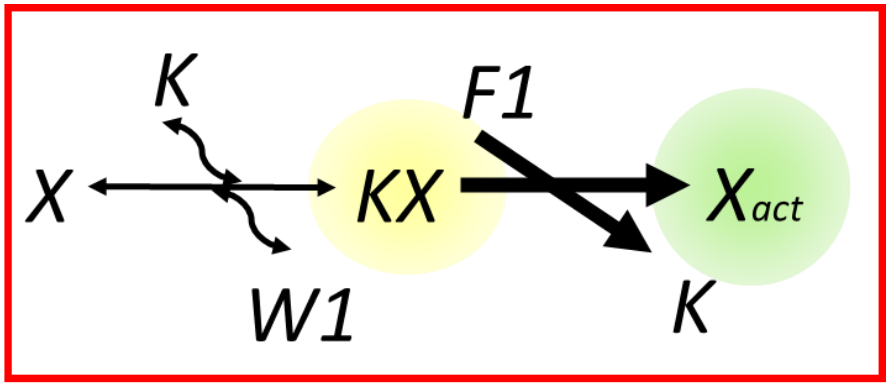

Figure 116. Diagram of the full implementation of the Push reaction with the thermodynamic drive. In this variation of the Push reaction there is a thermodynamic drive implemented in the second step of the reaction that biases the process towards the production of X_{act} as well as avoids the binding back of the product X_{act} to the catalyst as explained in Figure 112.

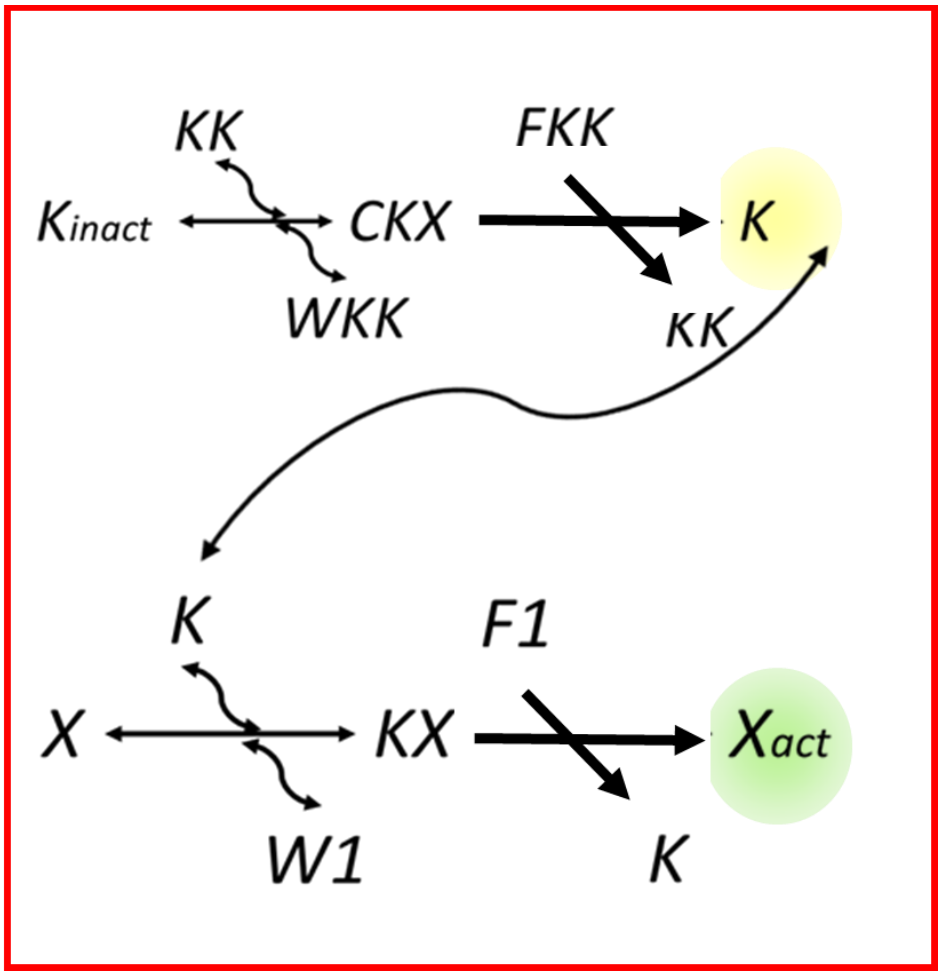

Figure 117. Diagram of the full implementation of the cascade network with the thermodynamic drive implemented in the two reactions of the cascade. In the present case, the experiments with the drive in the two levels of the cascade will have an emphasis on whether the drive could be used to overcome the sequestration problems that K suffers when KK is in excess.

The second type of limitations faced by the original design of the ACDC Framework were those we classified as fundamental formal limitations of the framework. Due to being based on toehold exchanges and recycling toeholds that are recycled in different catalytic layers, any network entailing any feedback loop as well as more than 4 catalytic layers cannot be realized due to the potential dimerization of the ancillary species.

However, in the same paper in which we discussed how these fundamental limitations arose, we discussed how, in principle, two types of mismatches placed on these ancillary species can act as a potential solution to these issues, thereby expanding the scope of the types of CCNs that can be designed within the ACDC Framework. These types of mismatches are either placed on the inner edge of the outer toehold (in the first type,) or in the outer edge of the inner toehold (on the second type). The first type of mismatch becomes exposed which remains hidden. Both types are required to implement the safeguard mechanism against the unwanted dimerization. However due to constraints, the experiments performed so far could only focus in testing the effectiveness of the first type. Despite this we can reasonably infer

that the results over the effectiveness of the first type of mismatch can give an educated guess on whether second type of mismatches would also destabilize unwanted complexes.

5.3.- Results

5.3.1.- Solving functional problems.

The result of implementing the mismatch on the aforementioned second step of the catalytic turnover for the push can be observed in Figures 118, which was obtained following the exact same protocol for the drive-less single-step experiments carried out previously in Chapter 3. We can observe that the implementation of the mismatch within the reaction works as we intended since we observe a clear thermodynamic bias making the reaction go in the forward direction, but not into the backward direction as intended. Moreover, the bias allowed us to see that in the reaction steps in which the fuel was in excess with respect the ancillary complex the turnover was 100%.

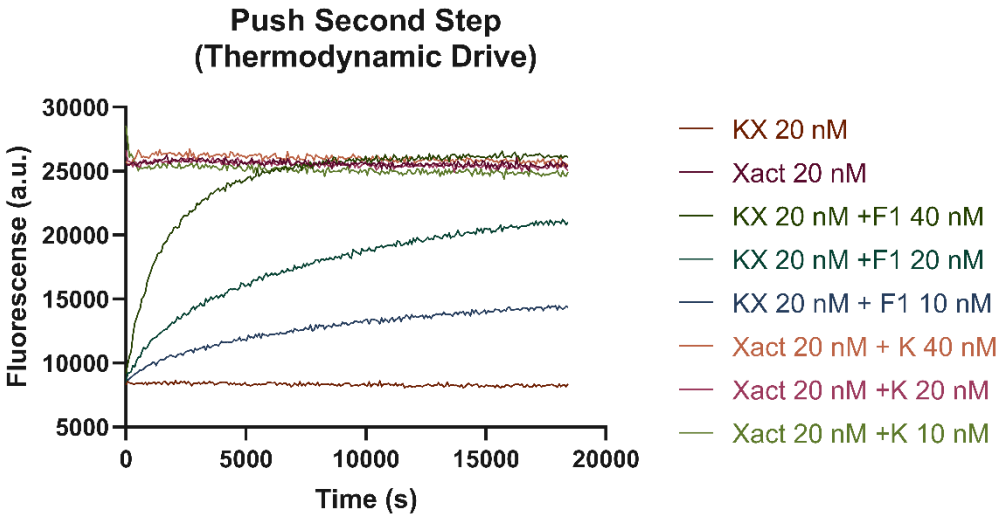

Figure 118. Fluorescence tracing for the second reaction step of the push reaction when observed on the 520 nm fluorescence channel corresponding to the emission of the Alexa 488 fluorophore. In the present graph, the lower fluorescence level corresponds to KX 20 nM, while the higher levels correspond to Xact 20 nM. The rising curves correspond to KX + F1 reacting with each other to form Xact and K while the slightly decreasing ones correspond to the reaction of K and Xact to form back KX and F1. We can observe a clear bias towards the intended forward reaction KX + F1, while we observe no noticeable reaction between Xact and the catalyst K.

When studying how the implementation of the hidden thermodynamic drive affected the full catalytic turnover reactions overall (Figures 119 and 112), we observed that in all the cases the turnover was complete after 24 hours and that the reactions did not plateau in any case. Moreover, more changes to the overall dynamic were observed when considering catalyst variation and the fuel variation experiments. When analysing the results of the catalyst variation experiment as seen in Figure 119, we observed that the reactions in both cases go to full completion – in contrast to the reaction without the drive – and that the rate of the reaction that had [K] = 10 nM was twice as big as when [K] = 5 nM as expected. However, when looking at the fuel variation experiments (Figure 120), we observed that doubling up of

the fuel (with the same quantity of catalyst) had no effect, thereby implying that with the thermodynamic drive implemented, big fuel excesses are not needed to drive the system.

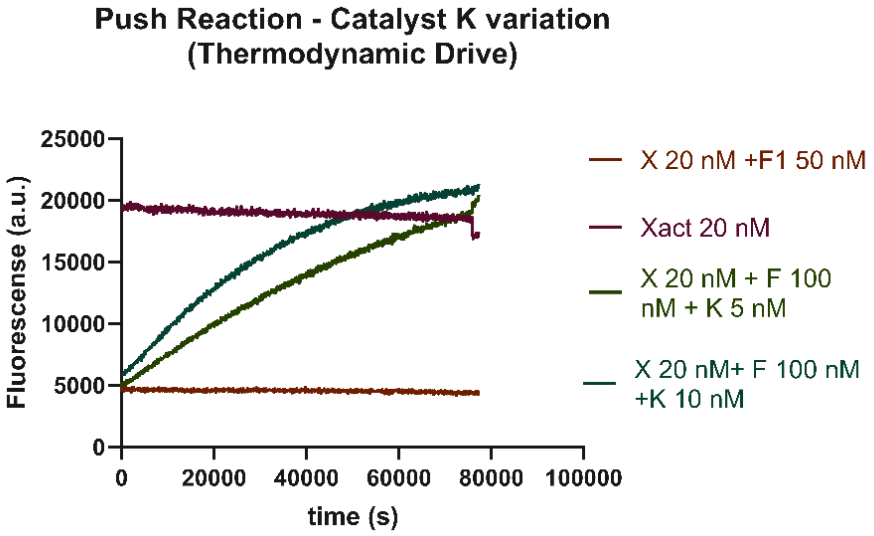

Figure 119. Fluorescence tracings for the catalytic push with the thermodynamic drive. In these experiments we vary the amount of catalyst K while keeping the amount of fuel equal between measurements. Tracings observed on the Alexa 488 fluorophore (green). It can be observed that the reaction speed differences between the two tracings are lower than in the case without the thermodynamic drive but still inside the range of a factor of 2 between [K]= 5 nM and [K] = 10 nM. After 24 hours the reactions were brought to full completion by adding an excess of a strand completely complementary to Strand 3 to fully release fluorescence.

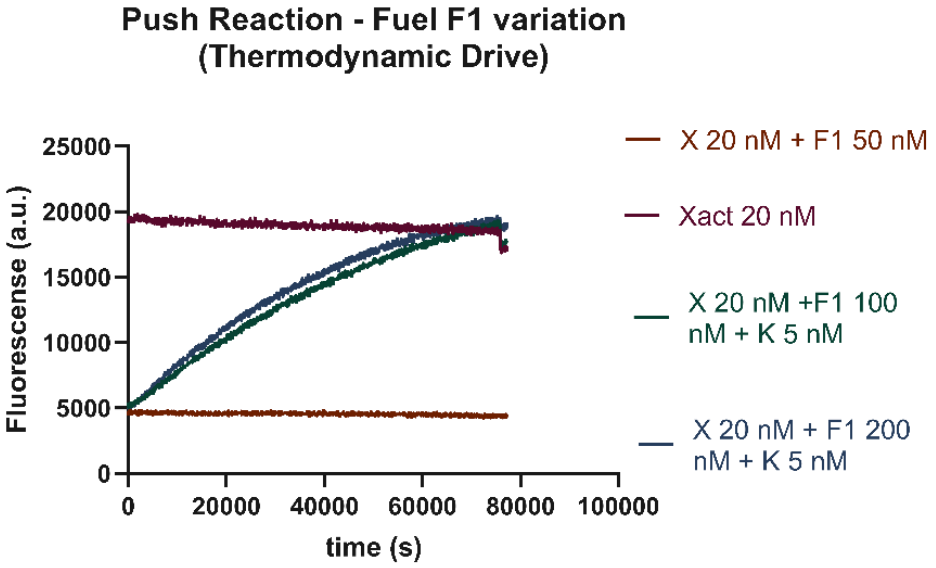

Figure 120. Fluorescence tracings for the catalytic push with the thermodynamic drive. In these experiments we vary the amount of F1 while keeping the amount of K equal between measurements. Tracings observed on the Alexa 488 fluorophore (green). After 24 hours, the reactions were brought to full completion by adding an excess of a strand completely complementary to Strand 3 to release full fluorescence, but the fluorescence did not change when compared with the level reached by the reaction turnover.

The next step consisted in the evaluation of the mismatches as a strategy to prevent product inhibition. As we see in Figure 122, we were able to observe that when we implemented the drive in a reaction that began with half of the substrate X activated, the reaction also reached the endpoint as it did when it began with no substrate in the active state. This situation alongside the results on the single-step reaction tests depicted on Figure 118. Moreover, the fact that the slope of the fluorescence tracing for [X]= [X_{act}] = 10 nM is half of the reaction with no X_{act} initially seems to further indicate that the unwanted product inhibition is being supressed. This is what we would expect from having only half the amount substrate present as in our case.

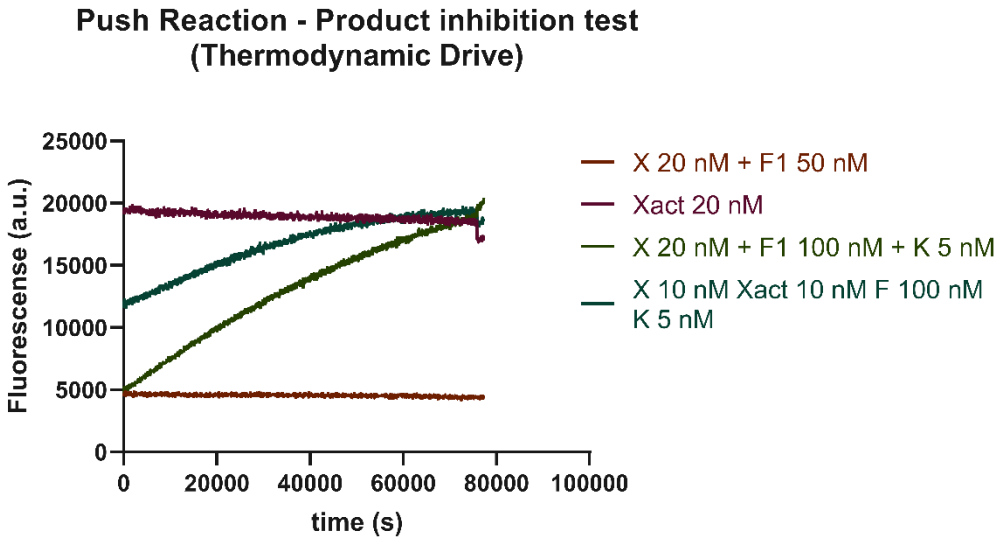

Figure 121. Fluorescence tracings for the push reaction with a thermodynamic drive using different starting points to probe the existence of product inhibition and its associated retroactivity on the system. As observed, the endpoint and the reaction rate are identical in the two circumstances, thus proving the problems associated with this issue can be sorted out with this particular implementation of the hidden thermodynamic drive. After 24 hours the reactions were brought to full completion by adding an excess of a strand completely complementary to Strand 3 to release full fluorescence.

We also observed that the thermodynamic drive was successfully implemented into both catalytic turnover reactions of the cascade network we designed in Chapter 4. Recycling the same protocol setup used on the cascade results of that chapter both in terms of timestep and excitation wavelength, we observed that the range that each of the catalytic layers could have was almost full when compared to the controls in the experiments (Figures 122 to 127). Moreover, the turnover shape in the case of the cascaded second layer is sharper. In addition, when observing the experiments varying KK (Figures 126 and 127) we still can observe that the excess of the upstream catalyst is not causing such a pronounced dampening effect on the dynamics of the system when compared to the experiment without the thermodynamic drive designed and studied in Chapter 4. This indicates that the sequestration of K by KK has lessened in the new system. In turn, on the second layer, the behaviour of K is roughly the same.

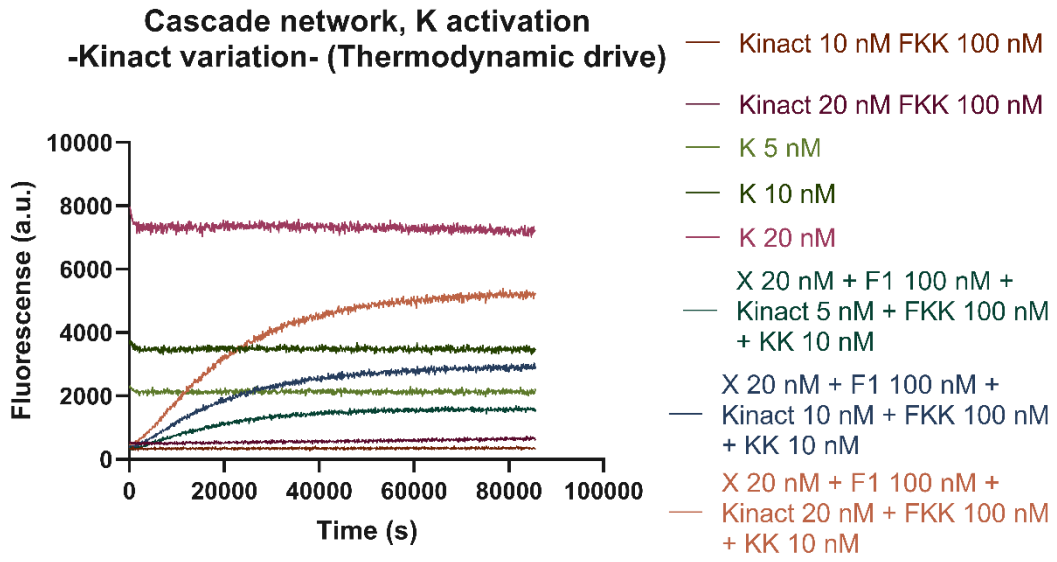

Figure 122. Fluorescence tracing for the addition of the upstream catalyst of the cascade network observed in the emission channel corresponding the emission wavelength of the Cy3 fluorophore (Yellow). In this experiment the aim is studying the effect of the variation of the amount of K_{inact} in the dynamics of the activation of K in the cascade when a hidden thermodynamic drive is implemented in the catalytic reactions. We can observe a greater presence of K activated than that that we observed in the drive-less versions of this experiments seen in figure 101 of Chapter 4, with the total activated proportion of K going between an 80 and a 90% of the maximum expected as per the reaction control.

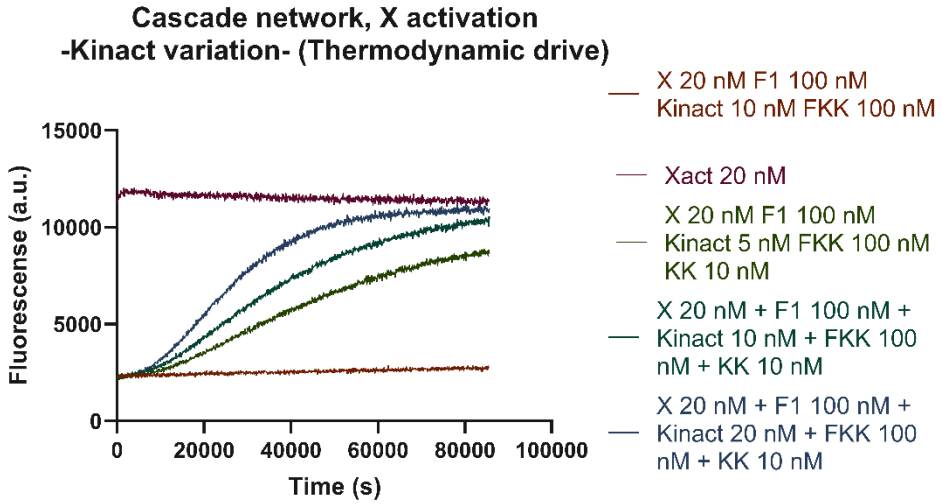

Figure 123. Fluorescence tracing for the addition of the upstream catalyst of the cascade network observed in the emission channel corresponding the emission wavelength of the Alexa 488 fluorophore (Green). In this experiment the aim is to study the effect of the variation of the amount of K_{inact} in the dynamics of the activation of X when a hidden thermodynamic drive is implemented in the catalytic reactions. We observe that a greater presence of K_{inact} leads to a faster reaction and that the reaction seems to reach completion.

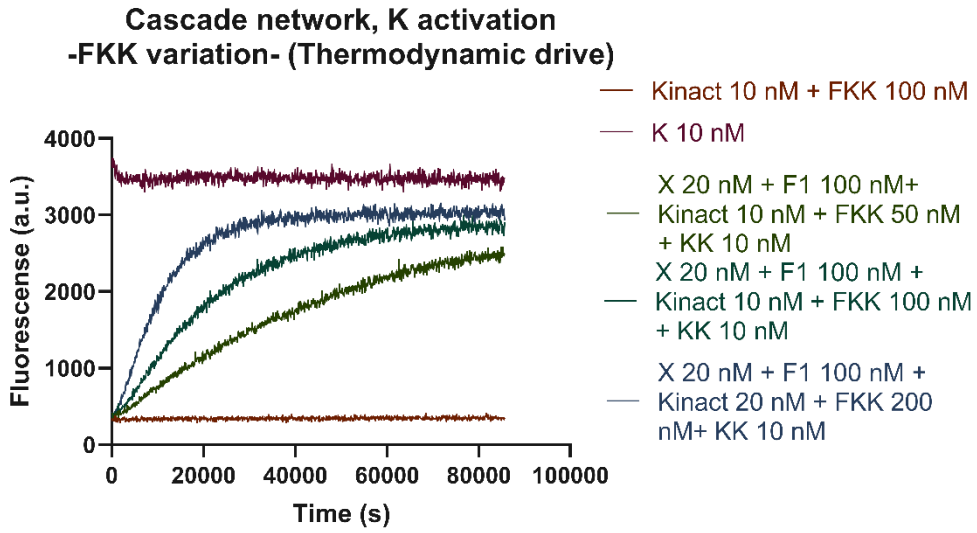

Figure 124. Fluorescence tracing for the addition of the upstream catalyst of the cascade network observed in the emission channel corresponding the emission wavelength of the Cy3 fluorophore (Yellow). In this experiment the aim is to study the effect of the variation of the amount of FKK in the dynamics of the activation of K in the cascade. We can observe that while the drive in all cases seems to enable the activation of all Kinact, the more abundant FKK is the faster the reaction seems to become – which contradicts the data observed on Figure 120. No explanation of this phenomenon has been found, although we speculate that in this case the binding of fuel to the intermediate might not be as strong as in the case of Figure 120, hence making the step more concentration-driven despite the presence of the enthalpic drive. However, the endpoint of the reaction seems to be independent from the amount of fuel, which supports the hypothesis that the thermodynamics of the system is dominated by the enthalpic contribution.

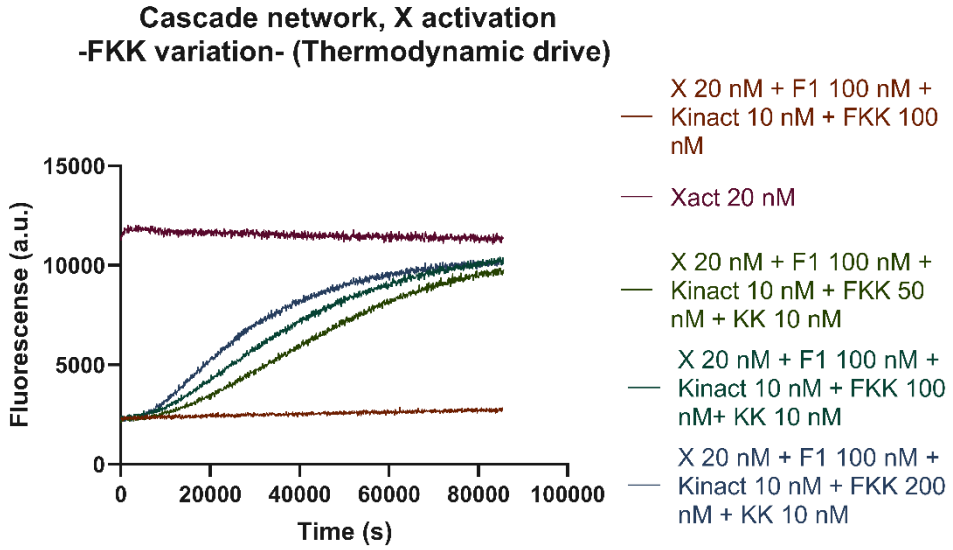

Figure 125. Fluorescence tracing for the addition of the upstream catalyst of the cascade network observed in the emission channel corresponding the emission wavelength of the Alexa 488 fluorophore (Green). In this experiment the aim is to study the effect of the variation of the amount of FKK on the activation of Xact. In this case we observe that all reactions reach the same endpoint but with different rates. These differences are caused by the faster activation that the greater presence of FKK causes.

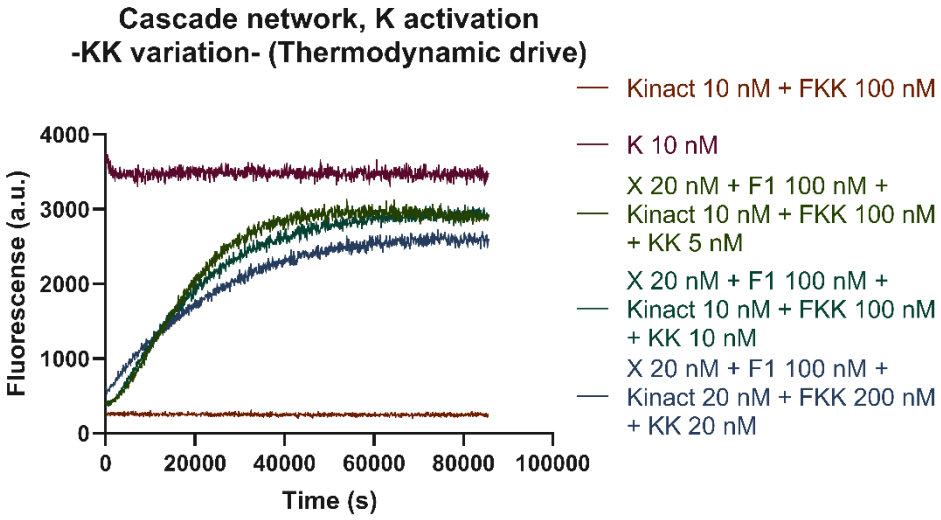

Figure 126. Fluorescence tracing for the addition of the upstream catalyst of the cascade network observed in the emission channel corresponding the emission wavelength of the Cy3 fluorophore (Yellow). In this experiment the aim is to study the effect of the variation of the amount of KK in the dynamics of the activation of the K. We observe that the rates are still faster with lesser quantities of KK injected in the system. Despite this, the differences are much less pronounced, and the drive seems to be avoiding the sequestration and the binding back of K into KK. This seems to suggest that the fuel consumption rate is the limiting factor of the activation in this case – which is coherent with the results presented in the two previous figures.

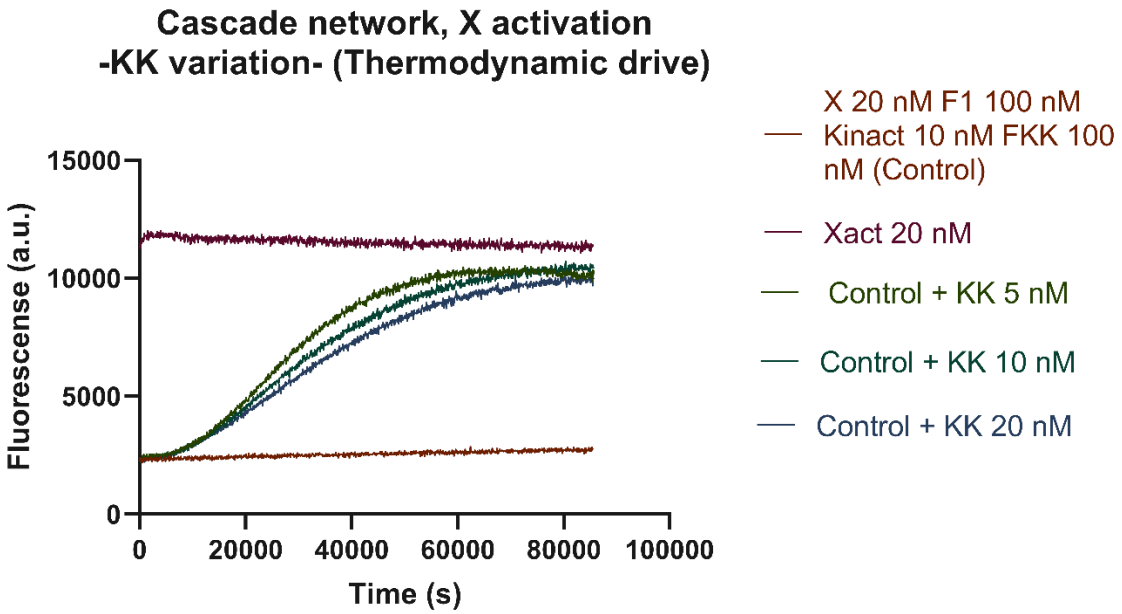

Figure 127. Fluorescence tracing for the addition of the upstream catalyst of the cascade network observed in the emission channel corresponding the emission wavelength of the Alexa 488 fluorophore (Green). In this experiment the aim is to study the effect of the variation of the amount of KK on the activation of X at the base of the cascade. In this case, since the drive has mitigated the sequestration effect that KK exerts over K when it is in excess.

5.3.2.-Application to fundamental formal problems

As we mentioned before, in the theoretical paper by Lankinen and collaborators[99] , we stated that the formal limitations faced by an ACDC-based CCN could be mitigated via the use of mismatches that destabilize complexes formed between ancillary species. In order to test this, we tested the dimerization in three setups. The first setup consisted in testing the binding of two duplexes designed only to bind via the formation a 10 basepair duplex. These duplexes as seen in Figure 128 were designed by placing an Alexa 488 fluorophore and an Iowa Black FQ quencher together when the two duplexes were bound, hence leading to a traceable reduction of the fluorescence of the system. The second setup was a variation of the first one in which a mismatch was placed in the inner side of one of the toeholds that mediate the hybridisation in the duplex with the fluorophore (which we dubbed Duplex 1 in all the configurations). The third and last setup was a variation of the first one that included an extra 10 basepairs complementary domain in the binding of the exposed toeholds. This configuration is designed as a positive control for the binding between the two duplexes in order to compare how strong the binding of the two previous configurations is when compared to the third – which is designed to give a strong binding between the duplexes and work as a positive control.

Using the same Alexa 488 excitation and emission wavelengths and gains specified in Chapter 2 but with a 42 second timestep, two types of experiments were performed with these duplexes. In the first one, Duplex 2 (the one with the Iowa Black FQ quencher) was added to the Duplex 1 in an equimolar ratio. In the second type of experiment, however, Duplex 2 was added in a tenfold excess in regards to Duplex 1 to test if high concentrations of ancillary species – as are the fuels in our reactions – might force this undesired dimerization. As we can see in Figures 129 and 130 the use of the mismatches avoids such unwanted dimerization from happening even when one of the ancillary species is at a tenfold excess as would be the case with the fuels and the intermediate complexes. Hence, we can take these results as a first indicator that the ACDC Framework allows us to build more complex motifs than those depicted previously in this thesis.

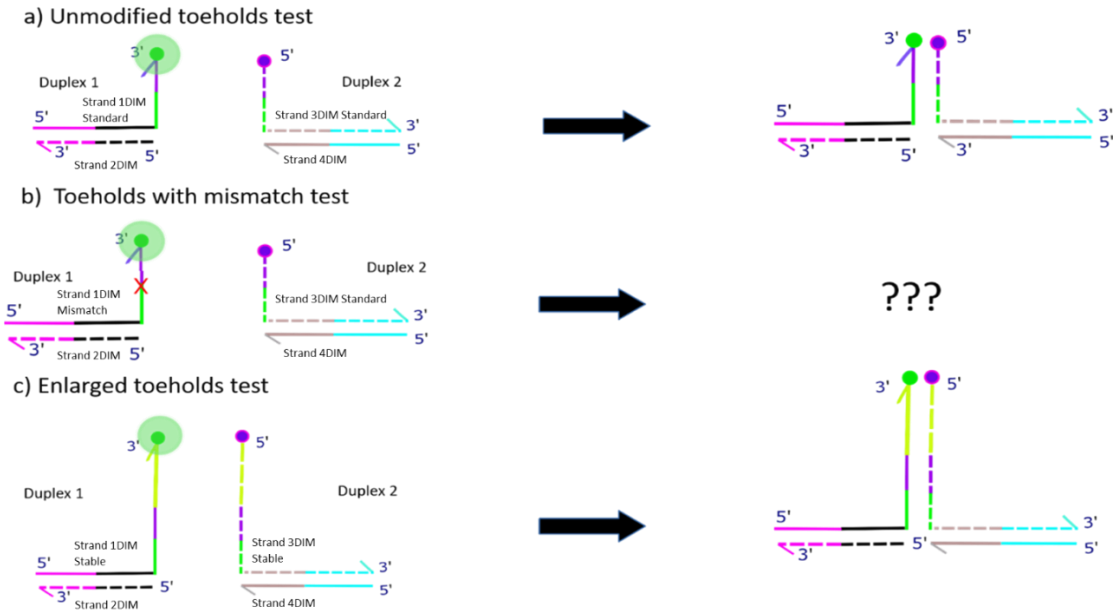

Figure 128. Diagram of the implementation of the anti-dimerization test for the design of extended networks. These three tests search to see the formation of a dimer via the union of a two 5 basepair exposed toeholds like those present in the ancillary species of an ACDC reaction. The duplexes are designed in such a way that if a dimer forms, we can expect a drop in fluorescence. In the unmodified toeholds test, the exposed toeholds are totally complementary, whereas in the test in b) a mismatch in the inner part of the outer toehold (the first type of mismatch described by Lankinen[99]) is introduced to destabilize the union. The third test is a variant of the first one in which the duplexes include an additional complementary 10 basepair domain is. With this domain, the binding of the duplexes is guaranteed thus acting as a positive binding control.

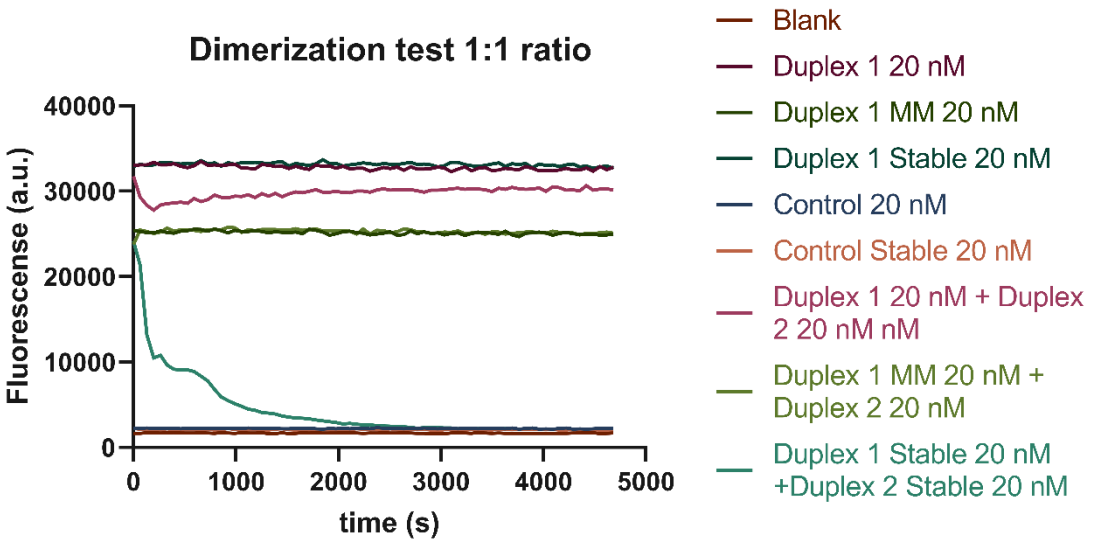

Figure 129. Fluorescence tracings for the dimerization test using a 1:1 ratio between the duplexes. As expected, the mismatch avoided the occurrence of dimers, while the stable dimers bound completely. Dimerization between the duplexes in the standard setup still could occur, albeit in a lower proportion. However, the ratio between the different ancillary species is not necessarily 1:1, hence explaining the need to extend the test to a situation where one of the dimers is more abundant than the other.

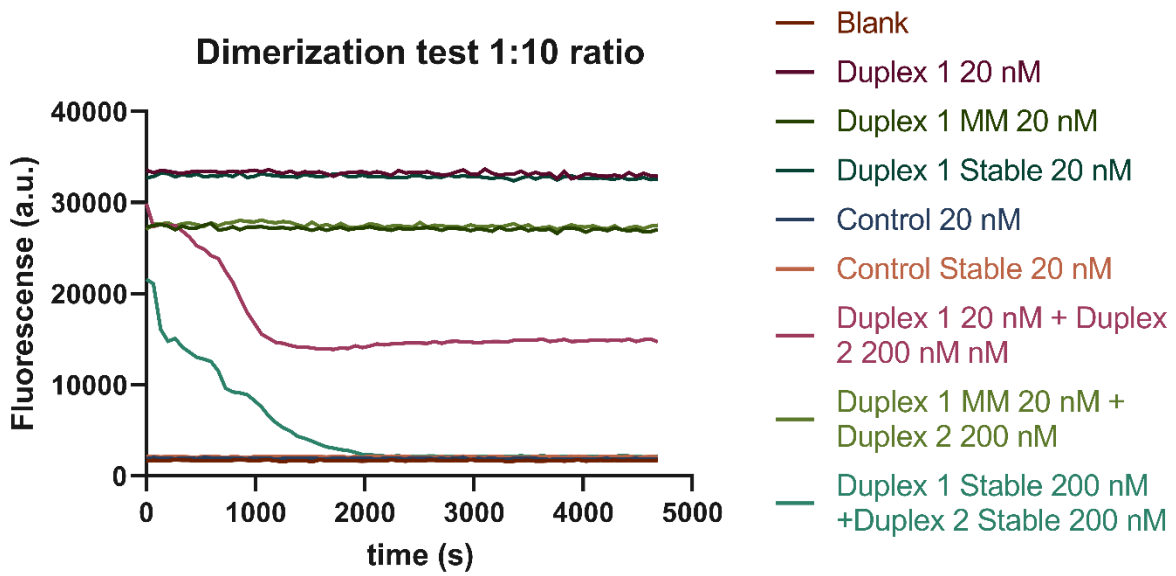

Figure 130. Fluorescence tracings for the dimerization test using a 1:10 ratio between the duplexes. As expected, the mismatch avoided the occurrence of dimers, while the stable dimers bound completely. Dimerization between the duplexes in the standard setup, is much more pronounced. However, as seen in this very figure, the presence of the mismatch avoids dimerization even in such big excesses.

In conclusion, in the light of the results shown in the present chapter, we conclude that the implementation of mismatches is a promising strategy to help attenuating the functional and formal limitations that the ACDC Framework presented in its first iteration. However, further testing needs to be done to assess how these mismatches are implemented in full push-pull motifs. Moreover, while the test to verify whether the implementation of a mismatch in an ancillary species shows that the mechanism works, at least physically, still remains to be seen a full implementation of the strategy designed described Lankinen's paper[99] to make feasible the design of loop networks inside the ACDC Framework.

Conclusions

The construction of time-responsive circuits has been an elusive matter to molecular programmers for decades. DNA nanotechnology and molecular programming has demonstrated that nucleic acids are a powerful platform upon which we can embed complex algorithms. However, the design of nucleic acid circuits capable of changing their behaviour in real time in response to variations in their environment – akin to what biological transduction networks do in nature – has so far been overlooked. While many applications of DNA nanotechnology such as biosensors do not rely on time-responsiveness to be properly functional, the use of nucleic acid circuits in synthetic biology requires time-responsiveness since adaptation and response to environmental changes is fundamental for life.

Given that actual signal processing systems in living systems can have such properties, it became apparent that the best way to tackle this problem is to borrow from nature and biology. In the case of the present work we went further than just borrowing a programmable, information-rich building block like DNA; we also considered the fundamental properties that make signal processing in biology so remarkable. Time-responsiveness is a property derived from the fact that transduction networks are systems operating out of thermodynamic equilibrium. In this out-of- equilibrium state catalysts act as the information carrier molecules transmitting signals to other catalysts, forming large extended processing networks. These networks keep a responsive state through continuous fuel consumption and can vary their output through changes in the catalysts. However, examination of existing reaction formalism implementations revealed their shortcomings in terms of functionality, that made the implementation of our desired reaction networks unattainable or impractical via such reaction formalisms.

Due to these limitations, we designed the ACDC Framework. Using DNA duplexes and 4-way junctions as the basis of our framework, we designed ACDC to be a tool that allowed us to circumvent the design problems of previous DNA strand displacement formalisms via two fundamental features. The first one was the definition of major species (substrates and catalysts) that conserved a molecular core throughout the whole catalytic reaction. The second one was the ability to keep the activation state and the identity of the substrate in two different strands of the duplex. Using ACDC, as demonstrated via our experimental results, we have been able to build proper catalytic reactions and integrate these into functional push-pull motifs that operate out of equilibrium. In addition to this, we also demonstrated that the basic operations that would allow us to build extended reaction networks are feasible via the ACDC Framework.

The ACDC Framework also allowed us to explore fundamental questions pertaining to the thermodynamics of computation. By way of nearing to zero the enthalpic free energy component of each of the chemical reactions that composed a catalytic ACDC-based reaction,

we were able to buildsystems that work out of thermodynamic equilibrium like the natural transduction systems while operating a lower energy cost than phosphorylation-based protein networks via only fuel imbalances. Moreover it must be remarked that the ACDC framework could allow us to create a functional version of a Szilard engine with a few experimental setup modifications [112].

However, while this type of low energy-consuming circuits proved to be functional, it also became evident that the design of reversible reactions entailed a series of functional drawbacks such as product inhibition, loading effects, unwanted sequestration, or the decreasing effectiveness of the fuels that power the chemical turnover. Additionally, the recycling of toeholds in our reaction formalism indicated that, without any kind of energy barrier to or bias prevent unwanted interactions, multi-layered loop motifs are not practically feasible. Hence, it is not hard to conclude that while enthalpic costs might not be necessary to build simple systems, they might be necessary as means to enhance their performance and make them more robust.

Working in collaboration with Andrew Tuberfield's group, we were demonstrated that implementing a hidden thermodynamic drive in our system can solve many of the functional problems previously enumerated. Moreover, we were able to provide the basis of a functional barrier mechanism based on mismatches that might make loop motifs feasible inside the ACDC Framework. Further research on how the barrier mechanism translates into real circuits as well as how the thermodynamic drives affect the behaviour of a push-pull could not be included in the present thesis due time and logistic constraints.

In addition to what was achieved in the present thesis, there are several potential directions for future research that are hinted by the present results. While being functional, present ACDC implementations operate at a timescale that might be too slow to make them useful as controllers in relevant biological processes. Hence, finding ways of speeding up the 4-way branch migrations reactions is a research direction that remains to be addressed. Moreover, biological implementations of ACDC-based circuits require addressing issues beyond the speed of the reactions. While the work of Dr Wooli Bae has shown that regular TMSD reactions can be implemented in genetically encoded RNA circuits in reaction buffer[130], it still remains to be seen how well ACDC species translate into this scheme. On the other hand, while spatial clustering of catalysts – a feature of biological transduction networks – is theoretically possible using the ACDC Framework, its experimental realisation and validation still remains an open problem.

References

1.- Miescher-Rüsch, F. (1871). Ueber die chemische Zusammensetzung der Eiterzellen. *Medicinisch-chemische Untersuchungen*. **4**: 441–460.

2.- McCarty, M., & Avery, O. T. (1946). Studies on the chemical nature of the substance inducing transformation of pneumococcal types. *Journal of Experimental Medicine*, *83*(2), 89-96.

3.- Watson, J. D., & Crick, F. H. (1953). A structure for deoxyribose nucleic acid. *Nature*, *171*(4356), 737-738.

4.- Pray, L. (2008). Discovery of DNA structure and function: Watson and Crick. *Nature Education*, *1*(1), 100.

5.- Hoogsteen, K. (1963). The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine. *Acta Crystallographica*, *16*(9), 907-916.

6.- Zhou, H., Hintze, B. J., Kimsey, I. J., Sathyamoorthy, B., Yang, S., Richardson, J. S., & Al-Hashimi, H. M. (2015). New insights into Hoogsteen base pairs in DNA duplexes from a structure-based survey. *Nucleic acids research*, gkv241.

7.- Frank-Kamenetskii, M. D., & Mirkin, S. M. (1995). Triplex DNA structures. *Annual review of biochemistry*, *64*(1), 65-95.

8.- Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K., & Neidle, S. (2006). Quadruplex DNA: sequence, topology and structure. *Nucleic acids research*, *34*(19), 5402-5415.

9.- Rhodes, D., & Lipps, H. J. (2015). G-quadruplexes and their regulatory roles in biology. *Nucleic acids research*, *43(18),* 8627-863

10.- Hu, Y., Cecconello, A., Idili, A., Ricci, F., & Willner, I. (2017). Triplex DNA nanostructures: From basic properties to applications. *Angewandte Chemie International Edition, 56(48),* 15210-15233.

11.- Zadeh, J. N., Steenberg, C. D., Bois, J. S., Wolfe, B. R., Pierce, M. B., Khan, A. R., ... & Pierce, N. A. (2011). NUPACK: analysis and design of nucleic acid systems. *Journal of computational chemistry, 32(1),* 170-173.

12.- Douglas, S. M., Marblestone, A. H., Teerapittayanon, S., Vazquez, A., Church, G. M., & Shih, W. M. (2009). Rapid prototyping of 3D DNA-origami shapes with caDNAno. *Nucleic acids research*, *37*(15), 5001-5006.

13.- Franklin, R. E., & Gosling, R. G. (1953). The structure of sodium thymonucleate fibres. I. The influence of water content. *Acta Crystallographica*, *6*(8-9), 673-677.

14.- Chargaff, E., Zamenhof, S., & Green, C. (1950). Human desoxypentose nucleic acid: Composition of human desoxypentose nucleic acid. *Nature*, *165*(4202), 756-757.

15.- Crick, F. (1970). Central dogma of molecular biology. *Nature*, *227*(5258), 561-563.

16.- Varmus, H. E. (1988). Retroviruses. *Science*, *240*(4858), 1427-1436.

17 .-Arkin, A. (2008). Setting the standard in synthetic biology. *Nature biotechnology*, *26*(7), 771-774

18.- Taniguchi, N. (1974). On the basic concept of nano-technology Proceedings of the International Conference on Production Engineering Tokyo Part II. Japan Society of Precision Engineering.

19.- Feynman, R. P. (1960). There's plenty of room at the bottom. *Engineering and science*, *23*(5), 22-36.

20.- Seeman, N. C. (1982). Nucleic acid junctions and lattices. *Journal of theoretical biology*, *99*(2), 237-247.

21.- Zhang, F., Nangreave, J., Liu, Y., & Yan, H. (2014). Structural DNA nanotechnology: state of the art and future perspective. *Journal of the American Chemical Society*, *136*(32), 11198-11211.

22.- Zheng, J., Birktoft, J. J., Chen, Y., Wang, T., Sha, R., Constantinou, P. E., ... & Seeman, N. C. (2009). From molecular to macroscopic via the rational design of a self-assembled 3D DNA crystal. *Nature*, *461*(7260), 74-77.

23.- Fu, T.-J. & Seeman, N. C. DNA double-crossover molecules. *Biochemistry* **32**, 3211–3220 (1993)

24.- Winfree, E., Liu, F., Wenzler, L. A., & Seeman, N. C. (1998). Design and self-assembly of two-dimensional DNA crystals. *Nature*, *394*(6693), 539-544.

25.-Seeman, N. C., Wang, H., Yang, X., Liu, F., Mao, C., Sun, W., ... & Wong, M. H. (1998). New motifs in DNA nanotechnology. *Nanotechnology*, *9*(3), 257.

26.- Mao, C., Sun, W., Shen, Z., & Seeman, N. C. (1999). A nanomechanical device based on the B–Z transition of DNA. *Nature*, *397*(6715), 144-146.

27.- Seeman, N. C. (2003). Biochemistry and structural DNA nanotechnology: an evolving symbiotic relationship. *Biochemistry*, *42*(24), 7259-7269.

28.- Rothemund, P. W. (2006). Folding DNA to create nanoscale shapes and patterns. *Nature*, *440*(7082), 297-302.

29.- Douglas, S. M., Dietz, H., Liedl, T., Högberg, B., Graf, F., & Shih, W. M. (2009). Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature*, *459*(7245), 414-418.

30.- Tikhomirov, G., Petersen, P., & Qian, L. (2017). Fractal assembly of micrometre-scale DNA origami arrays with arbitrary patterns. *Nature*, *552*(7683), 67-71

31.- Endo, M., Yang, Y., & Sugiyama, H. (2013). DNA origami technology for biomaterials applications. *Biomaterials Science*, *1*(4), 347-360

32.- Ke, Y., Ong, L. L., Sun, W., Song, J., Dong, M., Shih, W. M., & Yin, P. (2014). DNA brick crystals with prescribed depths. *Nature chemistry*, *6*(11), 994-1002.

33.- Zhang, F., Jiang, S., Wu, S., Li, Y., Mao, C., Liu, Y., & Yan, H. (2015). Complex wireframe DNA origami nanostructures with multi-arm junction vertices. *Nature nanotechnology*, *10*(9), 779.

34-- Wang, W., Chen, S., An, B., Huang, K., Bai, T., Xu, M., ... & Wei, B. (2019). Complex wireframe DNA nanostructures from simple building blocks. *Nature communications*, *10*(1), 1-8.

35.- Rothemund, P. W., Papadakis, N., & Winfree, E. (2004). Algorithmic self-assembly of DNA Sierpinski triangles. *PLoS Biol*, *2*(12), e424.

36- Adleman, L. M. (1994). Molecular computation of solutions to combinatorial problems. *Science*. 266 (5187), 1021–1024

37.- Woods, D., Doty, D., Myhrvold, C., Hui, J., Zhou, F., Yin, P., & Winfree, E. (2019). Diverse and robust molecular algorithms using reprogrammable DNA self-assembly. *Nature*, *567*(7748), 366-372.

38.- Yurke, B., Turberfield, A. J., Mills, A. P., Simmel, F. C., & Neumann, J. L. (2000). A DNA-fuelled molecular machine made of DNA. *Nature*, *406*(6796), 605-608.

39.- Reif, J. H. (2003). The design of autonomous DNA nano-mechanical devices: Walking and rolling DNA. *Natural Computing*, *2*(4), 439.

40..- Yin, Peng; Yan, Hao; Daniel, Xiaoju G.; Turberfield, Andrew J.; Reif, John H. (2004). "A Unidirectional DNA Walker Moving Autonomously Along a Linear Track". *Angewandte Chemie International Edition*. 43 (37): 4906–491141.- Shin, J. S., & Pierce, N. A. (2004). A synthetic DNA walker for molecular transport. *Journal of the American Chemical Society*, *126*(35), 10834-10835.

42.- Chen, J., Luo, Z., Sun, C., Huang, Z., Zhou, C., Yin, S., ... & Li, Y. (2019). Research Progress of DNA Walker and Its Recent Applications in Biosensor. *TrAC Trends in Analytical Chemistry*, 115626.

43 .- Bath, J., Green, S. J., Allen, K. E., & Turberfield, A. J. (2009). Mechanism for a directional, processive, and reversible DNA motor. *Small*, *5*(13), 1513-1516.

44.- Burns, J. R., Stulz, E., & Howorka, S. (2013). Self-assembled DNA nanopores that span lipid bilayers. *Nano letters*, *13*(6), 2351-2356.

45.- Zhang, D. Y., & Winfree, E. (2009). Control of DNA strand displacement kinetics using toehold exchange. *Journal of the American Chemical Society*, *131*(47), 17303-17314.

46.- Grun, C., Sarma, K., Wolfe, B., Shin, S. W., & Winfree, E. (2015). A domain-level DNA strand displacement reaction enumerator allowing arbitrary non-pseudoknotted secondary structures. *arXiv preprint arXiv:1505.03738*.

47.- Dirks, R. M., & Pierce, N. A. (2004). Triggered amplification by hybridization chain reaction. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(43), 15275-15278.

48.- Choi, H. M., Chang, J. Y., Trinh, L. A., Padilla, J. E., Fraser, S. E., & Pierce, N. A. (2010). Programmable in situ amplification for multiplexed imaging of mRNA expression. *Nature biotechnology*, *28*(11), 1208.

49.- Fujii, T., & Rondelez, Y. (2013). Predator–prey molecular ecosystems. *ACS nano*, *7*(1), 27-34.

50.- Kishi, J. Y., Schaus, T. E., Gopalkrishnan, N., Xuan, F., & Yin, P. (2018). Programmable autonomous synthesis of single-stranded DNA. *Nature chemistry*, *10*(2), 155.

51.- Seelig, G., Soloveichik, D., Zhang, D. Y., & Winfree, E. (2006). Enzyme-free nucleic acid logic circuits. *Science*, *314*(5805), 1585-1588.

52.- Zhang, D. Y., & Seelig, G. (2011). Dynamic DNA nanotechnology using strand-displacement reactions. *Nature chemistry*, *3*(2), 103.

53.- Dabby, N. L. (2013). *Synthetic molecular machines for active self-assembly: prototype algorithms, designs, and experimental study*. California Institute of Technology. PhD thesis,

54.- Qiu, M., Khisamutdinov, E., Zhao, Z., Pan, C., Choi, J. W., Leontis, N. B., & Guo, P. (2013). RNA nanotechnology for computer design and in vivo computation. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *371*(2000), 20120310

55.- Qian, L., & Winfree, E. (2011). A simple DNA gate motif for synthesizing large-scale circuits. *Journal of the Royal Society Interface*, *8*(62), 1281-1297.

56.- Qian, L., & Winfree, E. (2011). Scaling up digital circuit computation with DNA strand displacement cascades. *Science*, *332*(6034), 1196-1201.

57.- Qian, L., Winfree, E., & Bruck, J. (2011). Neural network computation with DNA strand displacement cascades. *Nature*, *475*(7356), 368-372.

58.- Douglas, S. M., Bachelet, I., & Church, G. M. (2012). A logic-gated nanorobot for targeted transport of molecular payloads. *Science*, *335*(6070), 831-834.

59.- Na, D., Yoo, S. M., Chung, H., Park, H., Park, J. H., & Lee, S. Y. (2013). Metabolic engineering of Escherichia coli using synthetic small regulatory RNAs. *Nature biotechnology*, *31*(2), 170-174.

60.- Mellin, J. R., & Cossart, P. (2015). Unexpected versatility in bacterial riboswitches. *Trends in Genetics*, *31*(3), 150-156.

61.- Plesa, T., Zygalakis, K. C., Anderson, D. F., & Erban, R. (2018). Noise control for molecular computing. *Journal of The Royal Society Interface*, *15*(144), 20180199.

62.-Johnson, R., Dong, Q., & Winfree, E. (2019). Verifying chemical reaction network implementations: a bisimulation approach. *Theoretical Computer Science*, *765*, 3-46.

63.-Shin, S. W., Thachuk, C., & Winfree, E. (2019). Verifying chemical reaction network implementations: A pathway decomposition approach. *Theoretical Computer Science*, *765*, 67-96.

64.- Genot, A. J., Bath, J., & Turberfield, A. J. (2011). Reversible logic circuits made of DNA. *Journal of the American Chemical Society*, *133*(50), 20080-20083.

65.- Garg, S., Shah, S., Bui, H., Song, T., Mokhtar, R., & Reif, J. (2018). Renewable Time-Responsive DNA Circuits. *Small*, *14*(33), 1801470.

66.- Srinivas, N., Parkin, J., Seelig, G., Winfree, E., & Soloveichik, D. (2017). Enzyme-free nucleic acid dynamical systems. *Science*, *358*(6369), eaal2052.

67.- Tiger, C. F., Krause, F., Cedersund, G., Palmér, R., Klipp, E., Hohmann, S., ... & Krantz, M. (2012). A framework for mapping, visualisation and automatic model creation of signal-transduction networks. *Molecular systems biology*, *8*(1).

68.- Purvis, J. E., & Lahav, G. (2013). Encoding and decoding cellular information through signaling dynamics. *Cell*, *152*(5), 945-956.

69.- Bray, D. (1995). Protein molecules as computational elements in living cells. *Nature*, *376*(6538), 307-312.

70.- Poole, W., Ortiz-Munoz, A., Behera, A., Jones, N. S., Ouldridge, T. E., Winfree, E., & Gopalkrishnan, M. (2017, September). Chemical boltzmann machines. In *International Conference on DNA-Based Computers* (pp. 210-231). Springer, Cham.

71.- Block, H. D. (1962). The perceptron: A model for brain functioning. *i. Reviews of Modern Physics*, 34(1), 123.

72.- Giese, W., Milicic, G., Schröder, A., & Klipp, E. (2018). Spatial modeling of the membrane-cytosolic interface in protein kinase signal transduction. *PLoS computational biology*, *14*(4), e1006075.

73.- Bashor, C. J., & Collins, J. J. (2018). Understanding biological regulation through synthetic biology. *Annual review of biophysics*, *47*, 399-423.

74.- Hinczewski, M., & Thirumalai, D. (2014). Cellular signaling networks function as generalized Wiener-Kolmogorov filters to suppress noise. *Physical Review X*, *4*(4), 041017.

75.- Deshpande, A., & Ouldridge, T. E. (2017). High rates of fuel consumption are not required by insulating motifs to suppress retroactivity in biochemical circuits. *Engineering Biology*, *1*(2), 86-99.

76.- Ferrell, J. E., & Ha, S. H. (2014). Ultrasensitivity part II: multisite phosphorylation, stoichiometric inhibitors, and positive feedback. *Trends in biochemical sciences*, *39*(11), 556-569.

77.- Grozinger, L., Amos, M., Gorochowski, T. E., Carbonell, P., Oyarzún, D. A., Stoof, R., ... & Goñi-Moreno, A. (2019). Pathways to cellular supremacy in biocomputing. *Nature communications*, *10*(1), 1-11

78.- Ferrell Jr, J. E., & Ha, S. H. (2014). Ultrasensitivity part I: Michaelian responses and zero-order ultrasensitivity. *Trends in biochemical sciences*, *39*(10), 496-503

79.- Qu, Z., & Vondriska, T. M. (2009). The effects of cascade length, kinetics and feedback loops on biological signal transduction dynamics in a simplified cascade model. *Physical biology*, *6*(1), 016007

80.- Blüthgen, N., Bruggeman, F. J., Legewie, S., Herzel, H., Westerhoff, H. V., & Kholodenko, B. N. (2006). Effects of sequestration on signal transduction cascades. *The FEBS journal*, *273*(5), 895-906.

81.-McGrath, T., Jones, N. S., ten Wolde, P. R., & Ouldridge, T. E. (2017). Biochemical machines for the interconversion of mutual information and work. *Physical review letters*, *118*(2), 028101

82.- Ouldridge, T. E., & ten Wolde, P. R. (2017). Fundamental costs in the production and destruction of persistent polymer copies. *Physical review letters*, *118*(15), 158103.

83.-Ouldridge, T. E., Govern, C. C., & ten Wolde, P. R. (2017). Thermodynamics of computational copying in biochemical systems. *Physical Review X*, *7*(2), 021004

84.- Mehta, P., Lang, A. H., & Schwab, D. J. (2016). Landauer in the age of synthetic biology: energy consumption and information processing in biochemical networks. *Journal of Statistical Physics*, *162*(5), 1153-1166.

85.-Bennett, C. H. (1973). Logical reversibility of computation. *IBM journal of Research and Development*, *17*(6), 525-532.

86.- Ouldridge, T. E. (2018). The importance of thermodynamics for molecular systems, and the importance of molecular systems for thermodynamics. *Natural computing*, *17*(1), 3-29.

87.- Sun, J., Yi, M., Yang, L., Wei, W., Ding, Y., & Jia, Y. (2014). Enhancement of tunability of MAPK cascade due to coexistence of processive and distributive phosphorylation mechanisms. *Biophysical journal*, *106*(5), 1215-1226.

88.- Takahashi, S., & Pryciak, P. M. (2008). Membrane localization of scaffold proteins promotes graded signaling in the yeast MAP kinase cascade. *Current Biology*, *18*(16), 1184-1191.

89.- Locasale, J. W., Shaw, A. S., & Chakraborty, A. K. (2007). Scaffold proteins confer diverse regulatory properties to protein kinase cascades. *Proceedings of the National Academy of Sciences*, *104*(33), 13307-13312.

90.- Cardelli, L. (2010). Two-domain DNA strand displacement. *arXiv preprint arXiv:1006.2993*.

91.- Qian, L., Soloveichik, D., & Winfree, E. (2010, June). Efficient Turing-universal computation with DNA polymers. In *International Workshop on DNA-Based Computers* (pp. 123-140). Springer, Berlin, Heidelberg.

92.- Qian, L., & Winfree, E. (2014, September). Parallel and scalable computation and spatial dynamics with DNA-based chemical reaction networks on a surface. In *International Workshop on DNA-Based Computers* (pp. 114-131). Springer, Cham.

93.- Teichmann, M., Kopperger, E., & Simmel, F. C. (2014). Robustness of localized DNA strand displacement cascades. *ACS nano*, *8*(8), 8487-8496.

94.- Mullor Ruiz, I., Arbona, J. M., Lad, A., Mendoza, O., Aimé, J. P., & Elezgaray, J. (2015). Connecting localized DNA strand displacement reactions. *Nanoscale*, *7*(30), 12970-12978.

95.- Bui, H., Shah, S., Mokhtar, R., Song, T., Garg, S., & Reif, J. (2018). Localized DNA hybridization chain reactions on DNA origami. *ACS nano*, *12*(2), 1146-1155.

96.- Chatterjee, G., Dalchau, N., Muscat, R. A., Phillips, A., & Seelig, G. (2017). A spatially localized architecture for fast and modular DNA computing. *Nature nanotechnology*, *12*(9), 920.

97.- Lin, T., Yan, J., Ong, L. L., Robaszewski, J., Lu, H. D., Mi, Y., ... & Wei, B. (2018). Hierarchical Assembly of DNA Nanostructures Based on Four-Way Toehold-Mediated Strand Displacement. *Nano letters*, *18*(8), 4791-4795.

98.- Kotani, S., & Hughes, W. L. (2017). Multi-arm junctions for dynamic DNA nanotechnology. *Journal of the American Chemical Society*, *139*(18), 6363-6368.

99.- Lankinen A., Mullor Ruiz I., & Ouldridge T. E. (2020), Implementing non-equilibrium networks with active circuits of duplex catalysts. *26th International Conference on DNA Computing and Molecular Programming (DNA 26)*, Publisher: Schloss Dagstuhl--Leibniz-Zentrum, Pages: 1-25

100.- McBride, C., Shah, R., & Del Vecchio, D. (2019). The effect of loads in molecular communications. *Proceedings of the IEEE*, *107*(7), 1369-1386.

101.- Moreira, B. G., You, Y., & Owczarzy, R. (2015). Cy3 and Cy5 dyes attached to oligonucleotide terminus stabilize DNA duplexes: predictive thermodynamic model. *Biophysical chemistry*, *198*, 36-44.

102.- Lee, J., Lee, S., Ragunathan, K., Joo, C., Ha, T., & Hohng, S. (2010). Single-molecule four-color FRET. *Angewandte Chemie International Edition*, *49*(51), 9922-9925.

103.- Pekař, M. Thermodynamics and Reaction Rates. In Thermodynamics—Interaction Studies—Solids, Liquids and Gases; Moreno-Pirajan, J.C., Ed.; InTech: Rijeka, Croatia, 2011; pp. 673–694.

104.- Pekař, M. (2018). Thermodynamic analysis of chemically reacting mixtures—Comparison of first and second order models. *Frontiers in chemistry*, *6*, 35.

105.- Scalise, D., Dutta, N., & Schulman, R. (2018). DNA strand buffers. *Journal of the American Chemical Society*, *140*(38), 12069-12076.

106.- Maxwell JC. (1879). The Sorting Demon of Maxwell. *Nature 20*, 126

107.- Szilard, L. (1929). Über die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen. *Zeitschrift für Physik*, *53*(11), 840-856.

108.- Landauer, R. (1961). Irreversibility and heat generation in the computing process. *IBM journal of research and development*, *5*(3), 183-191.

109.- Serreli, V., Lee, C. F., Kay, E. R., & Leigh, D. A. (2007). A molecular information ratchet. *Nature*, *445*(7127), 523-527.

110.- Koski, J. V., Maisi, V. F., Pekola, J. P., & Averin, D. V. (2014). Experimental realization of a Szilard engine with a single electron. *Proceedings of the National Academy of Sciences*, *111*(38), 13786-13789.

111.- Bengtsson, J., Tengstrand, M. N., Wacker, A., Samuelsson, P., Ueda, M., Linke, H., & Reimann, S. M. (2018). Quantum Szilard engine with attractively interacting bosons. *Physical review letters*, *120*(10), 100601.

112.- Brittain, R. A., Jones, N. S., & Ouldridge, T. E. (2019). Biochemical Szilard engines for memory-limited inference. *New Journal of Physics*, *21*(6), 063022.

113.- Hopfield, J. J. (1974). Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proceedings of the National Academy of Sciences*, *71*(10), 4135-4139.

114.- Goldbeter, A., & Koshland, D. E. (1981). An amplified sensitivity arising from covalent modification in biological systems. *Proceedings of the National Academy of Sciences*, *78*(11), 6840-6844.

115.- Stock, J. B., Stock, A. M., & Mottonen, J. M. (1990). Signal transduction in bacteria. *Nature*, *344*(6265), 395-400.

116.- Burotto, M., Chiou, V. L., Lee, J. M., & Kohn, E. C. (2014). The MAPK pathway across different malignancies: a new perspective. *Cancer*, *120*(22), 3446-3456.

117.- Munoz, L., & Ammit, A. J. (2010). Targeting p38 MAPK pathway for the treatment of Alzheimer's disease. *Neuropharmacology*, *58*(3), 561-568.

118.-Samaniego, C. C., & Franco, E. (2021). Ultrasensitive molecular controllers for quasi-integral feedback. *Cell Systems*.

119.- Takahashi, S., & Pryciak, P. M. (2008). Membrane localization of scaffold proteins promotes graded signaling in the yeast MAP kinase cascade. *Current Biology*, *18*(16), 1184-1191.

120.- Locasale, J. W., Shaw, A. S., & Chakraborty, A. K. (2007). Scaffold proteins confer diverse regulatory properties to protein kinase cascades. *Proceedings of the National Academy of Sciences*, *104*(33), 13307-13312.

121.- Feng, S., Ollivier, J. F., & Soyer, O. S. (2016). Enzyme sequestration as a tuning point in controlling response dynamics of signalling networks. *PLoS computational biology*, *12*(5), e1004918.

122.- Del Vecchio, D., Ninfa, A. J., & Sontag, E. D. (2008). Modular cell biology: retroactivity and insulation. *Molecular systems biology*, *4*(1), 161.

123.- Ortega, F., Acerenza, L., Westerhoff, H. V., Mas, F., & Cascante, M. (2002). Product dependence and bifunctionality compromise the ultrasensitivity of signal transduction cascades. *Proceedings of the National Academy of Sciences*, *99*(3), 1170-1175.

124.- Straube, R. (2013). Sensitivity and robustness in covalent modification cycles with a bifunctional converter enzyme. *Biophysical journal*, *105*(8), 1925-1933**.**

125- Xu, Y., & Gunawardena, J. (2012). Realistic enzymology for post-translational modification: zero-order ultrasensitivity revisited. *Journal of theoretical biology*, *311*, 139-152.

126.- Johnson, R., Dong, Q., & Winfree, E. (2019). Verifying chemical reaction network implementations: a bisimulation approach. *Theoretical Computer Science*, *765*, 3-46.

127.- Haley, N. E., Ouldridge, T. E.,Mullor Ruiz, I., Geraldini, A., Louis, A. A., Bath, J., & Turberfield, A. J. (2020). Design of hidden thermodynamic driving for non-equilibrium systems via mismatch elimination during DNA strand displacement. *Nature communications*, *11*(1), 1-11.

128.- SantaLucia, J., Allawi, H. T., & Seneviratne, P. A. (1996). Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry*, *35*(11), 3555-3562.

129.- SantaLucia, J. (1998). A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proceedings of the National Academy of Sciences*, *95*(4), 1460-1465.

130.- Bae, W., Stan, G. B. V., & Ouldridge, T. E. (2020). In situ Generation of RNA Complexes for Synthetic Molecular Strand-Displacement Circuits in Autonomous Systems. *Nano Letters*.

Appendices

Chapter 2

NUPACK code for the initial design of the push-pull

```
#Design of the strands for the DNA-based push-pull motif#

material = dna
temperature[C] = 25.0
trials = 10

# target structures #

structure X =
.....(((((((((((((((((((((((.....+.....)))))))))))))))))))))))).....
structure K =
.....(((((((((((((((((((((((.....+.....)))))))))))))))))))))))).....
structure KX =
..........((((((((((((((((((((((((+))))))))))))))))))))))))..........
structure FKXXact =
..........((((((((((((((((((((((((+))))))))))))))))))))))))..........
structure FKXactX =
..........((((((((((((((((((((((((+))))))))))))))))))))))))..........
structure Xact =
.....(((((((((((((((((((((((.....+.....)))))))))))))))))))))))).....
structure P =
.....(((((((((((((((((((((((.....+.....)))))))))))))))))))))))).....
structure PX =
..........((((((((((((((((((((((((+))))))))))))))))))))))))..........
structure FPXactX =
..........((((((((((((((((((((((((+))))))))))))))))))))))))..........
structure FPXXact =
..........((((((((((((((((((((((((+))))))))))))))))))))))))..........



# sequence domains # #Domain P in the original input is renamed Z in
order to avoid trouble with the P construct (x also gave some trouble #

domain C = N15
domain B = N5
domain D = N5
domain I = N5
domain J = N5
domain M = N5
domain A = N5
domain H = N5
domain F = N5
domain G = N5
domain E = N5
domain L = N5
domain N = N5
domain O = N5
```

```
domain Q = N10

# strands # #strand 14 is the same as strand 4#

strand s1 = F* B* C* D* G*
strand s2 = H* B* C* D* G*
strand s3 = E D C B A
strand s4 = G D C I L
strand s5 = J* I* C* D* E*
strand s6 = G D C M O
strand s7 = N* M* C* D* E*


# thread strands onto target structures


X.seq = s1 s3
K.seq = s4 s5
KX.seq = s5 s3
FKXXact.seq = s2 s4
FKXactX.seq = s1 s4
Xact.seq =  s2 s3
P.seq = s6 s7
PX.seq = s7 s3
FPXactX.seq = s1 s6
FPXXact.seq = s2 s6



# prevent sequence patterns #

prevent = AAAA, CCCC, GGGG, UUUU, KKKKKK, MMMMMM, RRRRRR, SSSSSS,
WWWWWW, YYYYYY
```

NUPACK code for the push pull cascade

```
#Design of the strands for the DNA-based push-pull cascade#

material = dna
temperature[C] = 25.0
trials = 10

# target structures #

structure X =
.....(((((((((((((((((((((((.....+.....)))))))))))))))))))))).....
structure K =
.....(((((((((((((((((((((((.....+.....)))))))))))))))))))))).....
structure KX =
..........(((((((((((((((((((((((+)))))))))))))))))))))))..........
structure FKXXact =
..........(((((((((((((((((((((((+)))))))))))))))))))))))..........
structure FKXactX =
..........(((((((((((((((((((((((+)))))))))))))))))))))))..........
structure Xact =
.....(((((((((((((((((((((((.....+.....)))))))))))))))))))))).....
```

```
structure P =
.....(((((((((((((((((((((((.....+.....)))))))))))))))))))))).....
structure PX =
..........(((((((((((((((((((((((+)))))))))))))))))))))).........
structure FPXactX =
..........(((((((((((((((((((((((+)))))))))))))))))))))).........
structure FPXXact =
..........(((((((((((((((((((((((+)))))))))))))))))))))).........
structure Rep = .....(((((((((((((+)))))))))))))
structure Y =
.....(((((((((((((((((((((((.....+.....)))))))))))))))))))))).....
structure Yact=
.....(((((((((((((((((((((((.....+.....)))))))))))))))))))))).....
structure KY =
..........(((((((((((((((((((((((+)))))))))))))))))))))).........
structure PY =
..........(((((((((((((((((((((((+)))))))))))))))))))))).........
structure KK =
.....(((((((((((((((((((((((.....+.....)))))))))))))))))))))).....
structure PK =
.....(((((((((((((((((((((((.....+.....)))))))))))))))))))))).....
structure CKK =
..........(((((((((((((((((((((((+)))))))))))))))))))))).........
structure CKP =
..........(((((((((((((((((((((((+)))))))))))))))))))))).........
structure FKK =
..........(((((((((((((((((((((((+)))))))))))))))))))))).........
structure FKP =
..........(((((((((((((((((((((((+)))))))))))))))))))))).........



# sequence domains # #Domain P in the original input is renamed Z in
order to avoid trouble with the P construct (x also gave some trouble #

domain C = TCTGCTTAAGCCGTG
domain B = TCTGC
domain D = CGTAG
domain I = AGATT
domain J = TGATG
domain M = ATAAC
domain A = ACTTA
domain H = GGTCG
domain F = AATAT
domain G = CGGGA
domain E = GCCTT
domain L = ACGCT
domain N = GTAAG
domain O = TCCAC
domain Z = CGATT
domain Q = GTGTATTTGG
domain R = GAACA
domain S = ATCTT
domain T = TATAA
domain U = N5
domain V = AGTTT
domain W = TTGGA
```

```
# strands # #strand 14 is the same as strand 4#

strand s1 = F* B* C* D* G*
strand s2 = H* B* C* D* G*
strand s3 = E D C B A
strand s4 = G D C I L
strand s5 = J* I* C* D* E*
strand s6 = G D C M O
strand s7 = N* M* C* D* E*
strand s8 = Q* H*
strand s9 = A* H Q
strand s13 = E D C B Z
strand s16 = L* I* C* U* V*
strand s17 = W U C I J
strand s18 = L* I* C* R* S*
strand s19 = T R C I J


# thread strands onto target structures


X.seq = s1 s3
K.seq = s4 s5
KX.seq = s5 s3
FKXXact.seq = s2 s4
FKXactX.seq = s1 s4
Xact.seq =  s2 s3
P.seq = s6 s7
PX.seq = s7 s3
FPXactX.seq = s1 s6
FPXXact.seq = s2 s6
Rep.seq = s9 s8
Y.seq = s1 s13
Yact.seq = s2 s13
KY.seq = s5 s13
PY.seq = s7 s13
KK.seq = s19 s18
PK.seq = s17 s16
CKK.seq = s19 s5
CKP.seq = s17 s5
FKK.seq = s4 s18
FKP.seq = s4 s16



# prevent sequence patterns #

prevent = AAAA, CCCC, GGGG, UUUU, KKKKKK, MMMMMM, RRRRRR, SSSSSS,
WWWWWW, YYYYYY
```

List of excess strands used to prepare every species designed in this chapter with the previous NUPACK codes (conserved for the rest of experiments)

- Substrate X: (Strand 1)
- Substrate X_{act}: (Strand 2)
- Substrate Y Inactive: (Strand 1)

- Substrate Y_{act}: (Strand 2)
- Intermediate species KX: (Strand 5)
- Intermediate species PX: (Strand 7)
- Intermediate species KY: (Strand 5)
- Intermediate species PY: (Strand 7)
- Catalyst species K: (Strand 5)
- Catalyst species P: (Strand 7)
- Fuel F1: (Strand 4)
- Waste W1: (Strand 4)
- Fuel F2: (Strand 2)
- Waste W2: (Strand 2)
- Substrate species $K_{inactive}$: (Strand 14)
- Intermediate CKK: (Strand 19)
- Intermediate sp CPK: (Strand 17)
- Catalyst KK: (Strand 19)
- Catalyst PK: (Strand 16)
- Fuel FKK : (Strand 18)
- Waste WKK : (Strand 18)
- Fuel FPK : (Strand 16)
- Fuel WPK : (Strand 16)

List of domains and strands with modifications used in this chapter

All sequences are indicated in the 5' to 3' sense.

The general notation in these appendices is that the asterisk domains are the complementary of those with the same letter without it.

The general notation with the blockers is that the the strands with a ,b subindex should bind to the excess of those with whom they share a number (i.e. Strand 1.b should block Strand 1 excesses)

Brackets in a domain indicate when a domain present initially in the NUPACK output design has been capped or removed for functional purposes of the experiment.

Leak reaction tests

Domains

A: ACTTA

B: TCTGC

C: TCTGCTTAAGCCGTG

C[2mm]:TCTcCTTAAcCCGTG

D: CGTAG

E: GCCTT

F: AATAT

G: CGGGA

A*: TAAGT

B*: GCAGA

C*: CACGGCTTAAGCAGA

C*[2mm]: CACcGCTTAAcCAGA

D*: CTACG

E*: AAGGC

F*: ATATT

G*: TCCCG

H*: CGACC

Q*: CCAAATACAC

Strands

Strand 1.2

/Iowa Black FQ/ F*[truncated] B* C* D* G*

/5IABkFQ/[ATA]TTGCAGACACGGCTTAAGCAGACTACGTCCCG

Strand 2.1

Q* H* B* C* D* G*

CCAAATACACCGACCGCAGACACGGCTTAAGCAGACTACGTCCCG

Strand 3.3

E D C B A[truncated] /Alexa 488/

GCCTTCGTAGTCTGCTTAAGCCGTGTCTGC[ACTTA]C/Alexa 488/

Strand 4.2

G D C I L[truncated]

CGGGACGTAGTCTGCTTAAGCCGTGAGATTAC[GCT]

Strand 1.2b

G[truncated] D C[2mm] B F[truncated]

[CGGG]ACGTAGTCTcCTTAAcCCGTGTCTGCA[ATAT]

Strand 2.1b

G[truncated] D C[2mm] B H[truncated]

[CGGG]ACGTAGTCTcCTTAAcCCGTGTCTGCG[GTCG]

Strand 4.2b

L*[truncated] I* C*[2mm] D* G*[truncated]

[AGCG]TAATCTCACcGCTTAAcCAGACTACGT[CCCG]

Junction Stability tests

Domains

B: TCTGC

C :TCTGCTTAAGCCGTG

D: CGTAG

E: GCCTT

G: CGGGA

I: AGATT

Θ1: GGTCAATGTA

Θ2: CATATTAAAT

Φ1: ATGTTGAGGC

Φ2: ATGTTGAATT

Σ1: AATAATCT

Σ2: GCCCGAAT

B*: GCAGA

C*: CACGGCTTAAGCAGA

D*: CTACG

E*: AAGGC

G*: TCCCG

I*: AATCT

Θ1*: TACATTGACC

Θ2*: ATTTAATATG

Φ1*: GCCTCAACAT

Φ2*: AATTCAACAT

Σ1*: AGATTATT

Σ2*: TGCCATTC

Strands

Strand 1_7GC

Θ1* C* D* G* /Alexa 488/

TACATTGACC CACGGCTTAAGCAGA CTACG TCCCG

Strand 2_7GC

E D C Θ1

GCCTT CGTAG TCTGCTTAAGCCGTG GGTCAATGTA

Strand 2_Stable7GC

Σ1 E D C Θ1

AATAATCT GCCTT CGTAG TCTGCTTAAGCCGTG GGTCAATGTA

Strand 3_7GC

Θ1* C* D* E* Σ1 *:

TACATTGACC CACGGCTTAAGCAGA CTACG AAGGC AGATTATT

Strand4 _7GC

/Iowa Black FQ/ G D C Θ1

/Iowa Black FQ/ CGGGA CGTAG TCTGCTTAAGCCGTG ATGTTGAGGC

Strand1_4GC

Θ2 C B

ATTTAATATG CACGGCTTAAGCAGA TCTGC

Strand1_Stable4GC
Θ2 C B Σ1

ATTTAATATG CACGGCTTAAGCAGA TCTGC GCCCGAA

Strand2_4GC

Σ1* B* C* Φ2*:

TGCCATTC GCAGA CACGGCTTAAGCAGA AATTCAACAT

Strand3_4GC

Φ2 C I /Iowa Black FQ/

ATGTTGAATT CACGGCTTAAGCAGA AGATT /Iowa Black FQ/

Strand3_4GC

/Alexa488/ I*C* Θ2*

/Alexa 488/ AATCT CACGGCTTAAGCAGA ATTTAATATG

The interacting duplexes were Duplex 12 + Duplex 34 in the test with 7 GC (the one analogous to the interaction of K +X) and Duplex 14 + Duplex 23 in the test (the one analogous to the KX + F interaction).

Catalyst orthogonality tests

Domains

A: ACTTA

B: TCTGC

C: TCTGCTTAAGCCGTG

C[2mm]:TCTcCTTAAcCCGTG

D: CGTAG

E: GCCTT

F: AATAT

G: CGGGA

H: GGTCG

I: AGATT

J: TGATG

L: ACGCT

M: ATAAC

N: GTAAG

O: TCCAC

A*: TAAGT

B*: GCAGA

C*: CACGGCTTAAGCAGA

C*[2mm]: CACcGCTTAAcCAGA

D*: CTACG

E*: AAGGC

F*: ATATT

G*: TCCCG

H*: CGACC

I*: AATCT

J*: CATCA

L*: AGCGT

M*: GTTAT

N*: CTTAC

O*: GTGGA

Q*: CCAAATACAC

Strands

Strand 1.2

/Iowa Black FQ/ F*[truncated] B* C* D* G*

/5IABkFQ/[ATA]TTGCAGACACGGCTTAAGCAGACTACGTCCCG

Strand 2.1

Q* H* B* C* D* G*

CCAAATACACCGACCGCAGACACGGCTTAAGCAGACTACGTCCCG

Strand 3.3

E D C B A[truncated] /Alexa 488/

GCCTTCGTAGTCTGCTTAAGCCGTGTCTGC[ACTTA]C/Alexa 488/

Strand 4.2

G D C I L[truncated]

CGGGACGTAGTCTGCTTAAGCCGTGAGATTAC[GCT]

Strand 5.3

Cy3 J*[truncated] I* C* D* E*

/Cy3/[CAT]CAAATCTCACGGCTTAAGCAGACTACGAAGGC

Strand 6.2

G D C M O[truncated]

CGGGACGTAGTCTGCTTAAGCCGTGATAACTC[CAC]

Strand 7.2

Al647 N*[truncated] M* C* D* E*

/Cy5/[CTT]ACGTTATCACGGCTTAAGCAGACTACGAAGGC

Strand 1.2b

G[truncated] D C[2mm] B F[truncated]

[CGGG]ACGTAGTCTcCTTAAcCCGTGTCTGCA[ATAT]

Strand 2.1b

G[truncated] D C[2mm] B H[truncated]

[CGGG]ACGTAGTCTcCTTAAcCCGTGTCTGCG[GTCG]

Strand 4.2b

L*[truncated] I* C*[2mm] D* G*[truncated]

[AGCG]TAATCTCACcGCTTAAcCAGACTACGT[CCCG]

Strand 5.3b

E[truncated] D C[2mm] I J[truncated]

[GCCT]TCGTAGTCTcCTTAAcCCGTGAGATTT[GATG]

Strand 6.2b

O*[truncated] M* C*[2mm] D* G*[truncated]

[GTGG]AGTTATCACcGCTTAAcCAGACTACGT[CCCG]

Strand 7.2b

E[truncated] D C[2mm] M N[truncated]

[GCCT]TCGTAGTCTcCTTAAcCCGTGATAACG[TAAG]

ON displacer

Q* H* B* C* D* E*

CCAAATACACCGACCGCAGACACGGCTTAAGCAGACTACGAAGGC

OFF displacer

/ Iowa Black FQ/ F*(trunc) B* C* D* E*

/5IABkFQ/[ATA]TT CAGACACGGCTTAAGCAGACTACGAAGGC

Catalyst inactive tests

Domains

B: TCTGC

C: TCTGCTTAAGCCGTG

C[2mm]:TCTcCTTAAcCCGTG

D: CGTAG

E: GCCTT

F: AATAT

G: CGGGA

H: GGTCG

I: AGATT

J: TGATG

L: ACGCT

P: CGATT

B*: GCAGA

C*: CACGGCTTAAGCAGA

C*[2mm]: CACcGCTTAAcCAGA

D*: CTACG

E*: AAGGC

F*: ATATT

G*: TCCCG

H*: CGACC

I*: AATCT

J*: CATCA

L*: AGCGT

P*: AATCG

Q*: CCAAATACAC

Strands

Strand 1.2

/Iowa Black FQ/ F*[truncated] B* C* D* G*

/5IABkFQ/[ATA]TTGCAGACACGGCTTAAGCAGACTACGTCCCG

Strand 13.1

E D C B P[truncated] Cy3

GCCTTCGTAGTCTGCTTAAGCCGTGTCTGCCG[ATT]/3Cy3Sp/

Strand 14.1 UNALBELED

G[removed completely] D C I L

[CGGGA]CGTAGTCTGCTTAAGCCGTGAGATTACGCT

Strand 5.3_ UNLABELED

J*[truncated] I* C* D* E*

[CAT]CAAATCTCACGGCTTAAGCAGACTACGAAGGC

Strand 1.2b

G[truncated] D C[2mm] B F[truncated]

[CGGG]ACGTAGTCTcCTTAAcCCGTGTCTGCA[ATAT]

Strand 5.3b

E[truncated] D C[2mm] I J[truncated]

[GCCT]TCGTAGTCTcCTTAAcCCGTGAGATTT[GATG]

Chapter 3

List of domains and strands with modifications used in this chapter

Domains

A: ACTTA

B: TCTGC

C: TCTGCTTAAGCCGTG

C[2mm]:TCTcCTTAAcCCGTG

D: CGTAG

E: GCCTT

F: AATAT

G: CGGGA

H: GGTCG

I: AGATT

J: TGATG

L: ACGCT

M: ATAAC

N: GTAAG

O: TCCAC

A*: TAAGT

B*: GCAGA

C*: CACGGCTTAAGCAGA

C*[2mm]: CACcGCTTAAcCAGA

D*: CTACG

E*: AAGGC

F*: ATATT

G*: TCCCG

H*: CGACC

I*: AATCT

J*: CATCA

L*: AGCGT

M*: GTTAT

N*: CTTAC

O*: GTGGA

Q*: CCAAATACAC

Strands

Strand 1.2 -

/Iowa Black FQ/ F*[truncated] B* C* D* G*

/5IABkFQ/[ATA]TTGCAGACACGGCTTAAGCAGACTACGTCCCG

Strand 2.1 -

Q* H* B* C* D* G*

CCAAATACACCGACCGCAGACACGGCTTAAGCAGACTACGTCCCG

Strand 3.3

E D C B A[truncated] /Alexa 488/

GCCTTCGTAGTCTGCTTAAGCCGTGTCTGC[ACTTA]C/Alexa 488/

Strand 4.2

G D C I L[truncated]

CGGGACGTAGTCTGCTTAAGCCGTGAGATTAC[GCT]

Strand 5.3

Cy3 J*[truncated] I* C* D* E*

/Cy3/[CAT]CAAATCTCACGGCTTAAGCAGACTACGAAGGC

Strand 6.2

G D C M O[truncated]

CGGGACGTAGTCTGCTTAAGCCGTGATAACTC[CAC]

Strand 7.2

Al647 N*[truncated] M* C* D* E*

/Cy5/[CTT]ACGTTATCACGGCTTAAGCAGACTACGAAGGC

Strand 1.2b

G[truncated] D C[2mm] B F[truncated]

[CGGG]ACGTAGTCTcCTTAAcCCGTGTCTGCA[ATAT]

Strand 2.1b

G[truncated] D C[2mm] B H[truncated]

[CGGG]ACGTAGTCTcCTTAAcCCGTGTCTGCG[GTCG]

Strand 4.2b

L*[truncated] I* C*[2mm] D* G*[truncated]

[AGCG]TAATCTCACcGCTTAAcCAGACTACGT[CCCG]

Strand 5.3b

E[truncated] D C[2mm] I J[truncated]

[GCCT]TCGTAGTCTcCTTAAcCCGTGAGATTT[GATG]

Strand 6.2b

O*[truncated] M* C*[2mm] D* G*[truncated]

[GTGG]AGTTATCACcGCTTAAcCAGACTACGT[CCCG]

Strand 7.2b

E[truncated] D C[2mm] M N[truncated]

[GCCT]TCGTAGTCTcCTTAAcCCGTGATAACG[TAAG]

ON displacer

Q* H* B* C* D* E*

CCAAATACACCGACCGCAGACACGGCTTAAGCAGACTACGAAGGC

OFF displacer

/ Iowa Black FQ/ F*(trunc) B* C* D* E*

/5IABkFQ/[ATA]TT CAGACACGGCTTAAGCAGACTACGAAGGC

Chapter 4

List of domains and strands with modifications used in this chapter

Single Push pull and join network experiments.

Domains

A: ACTTA

B: TCTGC

C: TCTGCTTAAGCCGTG

C[2mm]:TCTcCTTAAcCCGTG

D: CGTAG

E: GCCTT

F: AATAT

G: CGGGA

H: GGTCG

I: AGATT

J: TGATG

L: ACGCT

M: ATAAC

N: GTAAG

O: TCCAC

A*: TAAGT

B*: GCAGA

C*: CACGGCTTAAGCAGA

C*[2mm]: CACcGCTTAAcCAGA

D*: CTACG

E*: AAGGC

F*: ATATT

G*: TCCCG

H*: CGACC

I*: AATCT

J*: CATCA

L*: AGCGT

M*: GTTAT

N*: CTTAC

O*: GTGGA

Q*: CCAAATACAC

Strands

Strand 1.2

/Iowa Black FQ/ F*[truncated] B* C* D* G*

/5IABkFQ/[ATA]TTGCAGACACGGCTTAAGCAGACTACGTCCCG

Strand 2.1

Q* H* B* C* D* G*

CCAAATACACCGACCGCAGACACGGCTTAAGCAGACTACGTCCCG

Strand 3.3

E D C B A[truncated] /Alexa 488/

GCCTTCGTAGTCTGCTTAAGCCGTGTCTGC[ACTTA]C/Alexa 488/

Strand 4.2

G D C I L[truncated]

CGGGACGTAGTCTGCTTAAGCCGTGAGATTAC[GCT]

Strand 5.3

Cy3 J*[truncated] I* C* D* E*

/Cy3/[CAT]CAAATCTCACGGCTTAAGCAGACTACGAAGGC

Strand 6.2

G D C M O[truncated]

CGGGACGTAGTCTGCTTAAGCCGTGATAACTC[CAC]

Strand 7.2

Al647 N*[truncated] M* C* D* E*

/Cy5/[CTT]ACGTTATCACGGCTTAAGCAGACTACGAAGGC

Strand 1.2b

G[truncated] D C[2mm] B F[truncated]

[CGGG]ACGTAGTCTcCTTAAcCCGTGTCTGCA[ATAT]

Strand 2.1b

G[truncated] D C[2mm] B H[truncated]

[CGGG]ACGTAGTCTcCTTAAcCCGTGTCTGCG[GTCG]

Strand 4.2b

L*[truncated] I* C*[2mm] D* G*[truncated]

[AGCG]TAATCTCACcGCTTAAcCAGACTACGT[CCCG]

Strand 5.3b

E[truncated] D C[2mm] I J[truncated]

[GCCT]TCGTAGTCTcCTTAAcCCGTGAGATTT[GATG]

Strand 6.2b

O*[truncated] M* C*[2mm] D* G*[truncated]

[GTGG]AGTTATCACcGCTTAAcCAGACTACGT[CCCG]

Strand 7.2b

E[truncated] D C[2mm] M N[truncated]

[GCCT]TCGTAGTCTcCTTAAcCCGTGATAACG[TAAG]

ON displacer

Q* H* B* C* D* E*

CCAAATACACCGACCGCAGACACGGCTTAAGCAGACTACGAAGGC

OFF displacer

/ Iowa Black FQ/ F*(trunc) B* C* D* E*

/5IABkFQ/[ATA]TT CAGACACGGCTTAAGCAGACTACGAAGGC

Split network experiments

Domains

A: ACTTA

B: TCTGC

C: TCTGCTTAAGCCGTG

C[2mm]:TCTcCTTAAcCCGTG

D: CGTAG

E: GCCTT

F: AATAT

G: CGGGA

H: GGTCG

I: AGATT

J: TGATG

L: ACGCT

P: CGATT

A*: TAAGT

B*: GCAGA

C*: CACGGCTTAAGCAGA

C*[2mm]: CACcGCTTAAcCAGA

D*: CTACG

E*: AAGGC

F*: ATATT

G*: TCCCG

H*: CGACC

I*: AATCT

J*: CATCA

L*: AGCGT

P*: AATCG

Q*: CCAAATACAC

Strands

Strand 1.2

/Iowa Black FQ/ F*[truncated] B* C* D* G*

/5IABkFQ/[ATA]TTGCAGACACGGCTTAAGCAGACTACGTCCCG

Strand 2.1

Q* H* B* C* D* G*

CCAAATACACCGACCGCAGACACGGCTTAAGCAGACTACGTCCCG

Strand 3.3

E D C B A[truncated] /Alexa 488/

GCCTTCGTAGTCTGCTTAAGCCGTGTCTGC[ACTTA]C/Alexa 488/

Strand 4.2

G D C I L[truncated]

CGGGACGTAGTCTGCTTAAGCCGTGAGATTAC[GCT]

Strand 5.3

Cy3 J*[truncated] I* C* D* E*

/Cy3/[CAT]CAAATCTCACGGCTTAAGCAGACTACGAAGGC

Strand 13.1

E D C B P[truncated] Cy3

GCCTTCGTAGTCTGCTTAAGCCGTGTCTGCCG[ATT]/3Cy3Sp/

Strand 1.2b

G[truncated] D C[2mm] B F[truncated]

[CGGG]ACGTAGTCTcCTTAAcCCGTGTCTGCA[ATAT]

Strand 2.1b

G[truncated] D C[2mm] B H[truncated]

[CGGG]ACGTAGTCTcCTTAAcCCGTGTCTGCG[GTCG]

Strand 4.2b

L*[truncated] I* C*[2mm] D* G*[truncated]

[AGCG]TAATCTCACcGCTTAAcCAGACTACGT[CCCG

Strand 5.3b

E[truncated] D C[2mm] I J[truncated]

[GCCT]TCGTAGTCTcCTTAAcCCGTGAGATTT[GATG]

Cascade and cascade of push-pulls experiments

Domains

A: ACTTA

B: TCTGC

C: TCTGCTTAAGCCGTG

C[2mm]:TCTcCTTAAcCCGTG

D: CGTAG

E: GCCTT

F: AATAT

G: CGGGA

H: GGTCG

I: AGATT

J: TGATG

L: ACGCT

M: ATAAC

N: GTAAG

O: TCCAC

P: CGATT (substitute of the downstream domain of X in Y)

R: GAACA

S: ATCTT

T: TATAA

U: TTCCA

V: AGTTT

W: GATTG

A*: TAAGT

B*: GCAGA

C*: CACGGCTTAAGCAGA

C*[2mm]: CACcGCTTAAcCAGA

D*: CTACG

E*: AAGGC

F*: ATATT

G*: TCCCG

H*: CGACC

I*: AATCT

J*: CATCA

L*: AGCGT

M*: GTTAT

N*: CTTAC

O*: GTGGA

P*: AATCG

Q*: CCAAATACAC

R*: TGTTC

S*: AAGAT

T*: TTATA

U*: TGGAA

V*: AAACT

W*: CAATC

Strands

Strand 1.2 -

/Iowa Black FQ/ F*[truncated] B* C* D* G*

/5IABkFQ/[ATA]TTGCAGACACGGCTTAAGCAGACTACGTCCCG

Strand 2.1 -

Q* H* B* C* D* G*

CCAAATACACCGACCGCAGACACGGCTTAAGCAGACTACGTCCCG

Strand 3.3

E D C B A[truncated] /Alexa 488/

GCCTTCGTAGTCTGCTTAAGCCGTGTCTGC[ACTTA]C/Alexa 488/

Strand 6.2

G D C M O[truncated]

CGGGACGTAGTCTGCTTAAGCCGTGATAACTC[CAC]

Strand 7.2 QUENCHED

Iowa Black FQ N*[truncated] M* C* D* E*

/Iowa Black FQ/[CTT]ACGTTATCACGGCTTAAGCAGACTACGAAGGC

Strand 4.3

G D C I L

CGGGA CGTAG TCTGCTTAAGCCGTG AGATT ACGCT

Strand 5.6

/Cy3/ J* I* C* D* E*

CATCAAATCTCACGGCTTAAGCAGACTACGAAGGC

Strand 14.3

G[removed completely] D C I L /5IABkFQ/

[CGGGA]CGTAGTCTGCTTAAGCCGTGAGATTACGCT /5IABkFQ/

Strand 18.1

L* I* C* R* S*[truncated]

AGCGTAATCTCACGGCTTAAGCAGATGTTCA[AGAT]

Strand 19.1 QUENCHED

/Iowa Black FQ/T[truncated] R C I J

/Iowa Black FQ/ [TATA]AGAACATCTGCTTAAGCCGTGAGATTTGATG

Strand 16.1

L* I* C* U* V*[truncated]

AGCGTAATCTCACGGCTTAAGCAGATGGAAA[AACT]

Strand 17.1 QUENCHED

/Iowa Black FQ/W[truncated] U C I J

[GATT]GTTCCATCTGCTTAAGCCGTGAGATTTGATG

Strand 1.2b

G[truncated] D C[2mm] B F[truncated]

[CGGG]ACGTAGTCTcCTTAAcCCGTGTCTGCA[ATAT]

Strand 2.1b

G[truncated] D C[2mm] B H[truncated]

[CGGG]ACGTAGTCTcCTTAAcCCGTGTCTGCG[GTCG]

Strand 4.2b

L*[truncated] I* C*[2mm] D* G*[truncated]

[AGCG]TAATCTCACcGCTTAAcCAGACTACGT[CCCG]

Strand 6.2b

O*[truncated] M* C*[2mm] D* G*[truncated]

[GTGG]AGTTATCACcGCTTAAcCAGACTACGT[CCCG]

Strand 7.2b

E[truncated] D C[2mm] M N[truncated]

[GCCT]TCGTAGTCTcCTTAAcCCGTGATAACG[TAAG]

Strand 18.1b

S [truncated] R C[2mm] I L [truncated]

[ATCT]TGAACATCTcCTTAAcCCGTGAGATTA[CGCT]

Strand 19.1b

J*[truncated] I* C*[2mm] R* T*[truncated]

[CATCA]AATCTCACcGCTTAAcCAGATGTTCT[TATA]

Strand 16.1b

V [truncated] U C [2mm] I L[truncated]

[AGTT]TTTCCATCTcCTTAAcCCGTGAGATTA[CGCT]

Strand 17.1b

J*[truncated] I* C*[2mm] U* W* [truncated]

[CATC]AAATCTCACcGCTTAAcCAGATTGGAC[AATC]

Chapter 5

Experimental protocol for the results of Figure 113

Experimental design

Species preparation

Oligonucleotides were purchased from Integrated DNA Technologies (IDT) at an initial concentration of 100 µM in the LabReady formulation provided by the supplier (pH 8.0 TE Buffer). All the strands were HPLC-purified.

Catalyst B strands were set to an intermediate concentration of 1 µM diluting 1 µL of the initial stock provided by IDT in 99 µL of TAE buffer 1x Na^+ 1M, then shaking vigorously in the vortex and letting to rest at room temperature. The final working concentration for the real time tracing and leak reaction experiments was 50 nM and was obtained by diluting 5 µl of the intermediate concentration in 95 µL of TAE buffer 1x Na^+ 1 M. The final concentration for the endpoint experiments was 100 nM and was prepared by adding 10 µl of the intermediate concentration in 95 µL of TAE buffer 1x Na^+ 1 M. The working concentration for the Fuel strand was set at 10 µM by diluting 4 µl from the stock provided by IDT at 36 µL of TAE buffer Na^+ 1 M, then shaking vigorously in the vortex and letting to rest at room temperature. Gate-Output complexes were prepared by adding both strands at 1.2:1 ratio the strands (1.2 µL of gate strand per each microliter of output at 98 µL of TAE Buffer 1x Na^+ 1 M) to TAE Buffer 1x Na^x 1M. In the case of the TT mismatch complexes in which the ratio was 1.2:1.1 because titration showed that the concentration of the Output-TT mismatch strand was actually lower than specified and titration proved that those ratios yielded comparable input release results to those of the other two Gate-Output. The Reporter-Lock complexes were prepared diluting the strands in 1:1.3 ratios (5 µL of the Reporter strand mixed with in 188.5 uL of TAE Buffer 1 x full quenching of the strand checked by titration). Both the Gate-Output and the Reporter-Lock complexes were annealed by being left at room temperature (25°C).

Fluorimetry traces protocol

The experiments were performed in a CLARIOstar Plate Reader (BMG Labtech) using as our support 96 wells black polystyrene plates with the transparent bottom (SIGMA ALDRICH). The optics settings of the experiments was using bottom optics measuring all the time on a fixed point of the well. Shaking the plate during 5 seconds before the first measurement to ensure homogeneity in the wells was also included in the protocol. The excitation and reading wavelengths used for measuring the fluorescence of Cy3 were those provided by default for the fluorophore (520/30 nm for the excitation wavelength and 590/20 nm for the reading wavelength). The gain set for the experiments was 1600 and temperature

was held at 25 °C throughout the whole experiments by the spectrofluorimeter's internal temperature control. The measurements were performed each 30 seconds. Samples were not illuminated between the measurements and no appreciable photobleaching effect was observed for these experiments. All different conditions of the same types of experiments were performed simultaneously in order to improve internal consistency between readings.

For the real time dynamics and leak reaction experiments, the experiments began with the addition of the Gate-Output complexes and the Reporter-Lock complexes (8 and 20 µl of the corresponding working concentration dilutions described in the previous section) in TAE buffer 1x Na^+ 1M previously added to the corresponding well (168 µL for the leak experiments and 148 µl for the real time dynamics of the system experiments). After this step the content of the well was homogenised taking 148 µl of the well carefully and adding it back to the well twice, changing the tip of the pipette each time to avoid cross-contamination of the wells. After this mixing we measured the signal for 5 minutes in order to see if the baseline given is coherent with the level expected for the quenched reporter. After this step, the fuel strands were incorporated by the addition of 4 µL of the working concentration dilution prepared previously, again mixed and homogenized following the same steps as with the addition of the Gate-Output and the Reporter-Lock complexes to the well. Then 20 uL of the working concentration of the corresponding input strands were added to the corresponding well using a 10-100 µL multichannel pipette (Eppendorf) adding up to a final 200 µL volume on each well, then homogenized twice with taking and re-adding 20 µl from the corresponding wells with the multichannel pipette. The experiments were left to run for 30.000 s after which an excess of input (4 uL of standard Input strand 100 µM) was added in order to check the endpoint of the experiment and allow for the normalisation of data.

Sequences for the experiment of Figure 113

(all sequences are in 5' to 3' sense)

Gate :

GAGTGG GCGAAGGGTTCTTGGAGGAG AGGGTC GG

Output_no_mm:

GTCCCTATTACCCTTGCC GACCCT CTCCTCCAAGAACCCTTCGC

Catalytic input_no_mm :

CT CTCCTCCAAGAACCCTTCGC CCACTC

Fuel_NoExtraBp:

CC GACCCT CTCCTCCAAGAACCCTTCGC

Fuel_Extra2:

CC GACCCT CTCCTCCAAGAACCCTTCGC CC

Output_CAmm:

GTCCCTATTACCCTTGCC GACCCT CTCCTCCAAAAACCCTTCGC

Input_CAmm:

CT CTCCTCCAAAAACCCTTCGC CCACTC

Output_TTmm:

GTCCCTATTACCCTTGCC GACCCT CTCCTCCATGAACCCTTCGC

Input_TTmm:

CT CTCCTCCATGAACCCTTCGC CCACTC

Reporter:

AGGGTC GGCAAGGGTAATAGGGAC ttt /3Cy3Sp/

Lockreporter1_hh:

/5IABkFQ/ ttt GTCCCTATTACCCTTGCC

Domains and strands for the implementation of the drive in the ACDC catalytic reactions

Domains

A: ACTTA

B: TCTGC

C: TCTGCTTAAGCCGTG

C[2mm]:TCTcCTTAAcCCGTG

D: CGTAG

E: GCCTT

F: AATAT

G: CGGGA

H: GGTCG

I: AGATT

J: TGATG

L: ACGCT

M: ATAAC

R: GAACA

S: ATCTT

T: TATAA

B*: GCAGA

C*: CACGGCTTAAGCAGA

C*[2mm]: CACcGCTTAAcCAGA

D*: CTACG

E*: AAGGC

F*: ATATT

G*: TCCCG

G2*: ACCCG

H*: CGACC

I*: AATCT

J*: CATCA

L*: AGCGT

Q*: CCAAATACAC

R*: TGTTC

S*: AAGAT

T*: TTATA

Strands

Strand 1.2 - length 32

/Iowa Black FQ/ F*[truncated] B* C* D* G*

/5IABkFQ/[ATA]TTGCAGACACGGCTTAAGCAGACTACGTCCCG

Strand 2.1 - length 45

Q* H* B* C* D* G*

CCAAATACACCGACCGCAGACACGGCTTAAGCAGACTACGTCCCG

Strand 3.3 length 32

E D C B A[truncated] /Alexa 488/

GCCTTCGTAGTCTGCTTAAGCCGTGTCTGC[ACTTA]C/Alexa 488/

Strand 4.3

G D C I L

CGGGA CGTAG TCTGCTTAAGCCGTG AGATT ACGCT

Strand 5.6

/Cy3/ J* I* C* D* E*

CATCAAATCTCACGGCTTAAGCAGACTACGAAGGC

Strand 14.3

G[removed completely] D C I L /5IABkFQ/

[CGGGA]CGTAGTCTGCTTAAGCCGTGAGATTACGCT /5IABkFQ/

Strand 18.1

L* I* C* R* S*[truncated]

AGCGTAATCTCACGGCTTAAGCAGATGTTCA[AGAT]

Strand 19.1 QUENCHED

/Iowa Black FQ/T[truncated] R C I J

/Iowa Black FQ/ [TATA]AGAACATCTGCTTAAGCCGTGAGATTTGATG

Strand 1.2b

G[truncated] D C[2mm] B F[truncated]

[CGGG]ACGTAGTCTcCTTAAcCCGTGTCTGCA[ATAT]

Strand 2.1b

G[truncated] D C[2mm] B H[truncated]

[CGGG]ACGTAGTCTcCTTAAcCCGTGTCTGCG[GTCG]

Strand 4.2b

L*[truncated] I* C*[2mm] D* G*[truncated]

[AGCG]TAATCTCACcGCTTAAcCAGACTACGT[CCCG]

Strand 18.1b

S [truncated] R C[2mm] I L [truncated]

[ATCT]TGAACATCTcCTTAAcCCGTGAGATTA[CGCT]

Strand 19.1b

J*[truncated] I* C*[2mm] R* T*[truncated]

[CATCA]AATCTCACcGCTTAAcCAGATGTTCT[TATA]

Strand 2.1 G2*

Q* H* B* C* D* G2*

CCAAATACACCGACCGCAGACACGGCTTAAGCAGACTACG ACCCG

Strand 16.1 L2*

L2* I* C* U* V*[truncated]

AGCGA AATCTCACGGCTTAAGCAGATGGAAA[AACT

Domains and strands for the dimerization test

As stated in the body of the thesis, the dimers tested are dimers formed by a any variant of 1 and 2 aganst dimers formed by any variant of 3 and 4.

Domains

C: TCTGCTTAAGCCGTG

C[2mm]:TCTcCTTAAcCCGTG

Θ1: GGTCAATGTA

Θ2: CATATTAAAT

Φ1: ATGTTGAGGC

Φ1mm: ATGTTCAGGC

Σ1: AATAATCTCC

C*: CACGGCTTAAGCAGA

C*[2mm]: CACcGCTTAAcCAGA

Θ1*: TACATTGACC

Θ2*: ATTTAATATG

Φ1*: GCCTCAACAT

Φ2*: AATTCAACAT

Σ1*: AGATTATT

Strands

Strand 1DIM Standard
C Θ1 Φ1 tt /Alexa 488/
TCTGCTTAAGCCGTG GGTCAATGTA ATGTTGAGGC TT /Alexa 488/

Strand 1DIM Mismatch
C Θ1 Φ1mm tt /Alexa 488/
TCTGCTTAAGCCGTG GGTCAATGTA ATGTTCAGGC TT /Alexa 488/

Strand 1DIM Stable
C Θ1 Φ1 Σ1 tt /Alexa 488/
TCTGCTTAAGCCGTG GGTCAATGTA ATGTTGAGGC AATAATCTCC TT/Alexa 488/

Strand 2DIM
Θ1* C*
TACATTGACC CACGGCTTAAGCAGA

Strand 2DIM.b
C 2 mm Θ1 1mm
[T] TCTcCTTAAcCCGTG ggtcaatgt[a]

Strand 3DIM standard
/IowaBlack FQ/ TT Φ1* Θ2* C*

/IowaBlack FQ/ TT GCCTCAACAT ATTTAATATG CACGGCTTAAGCAGA

Strand 3DIM stable

/IowaBlack FQ/ TT Σ1* Φ1* Θ2* C*

/IowaBlack FQ/ TT GGAGATTATT GCCTCAACAT ATTTAATATG CACGGCTTAAGCAGA

Strand 4DIM

C Θ2

TCTGCTTAAGCCGTG CATATTAAAT

Strand 4DIM.b

Θ2* C*2mm

[A]TTTAATATG CACCGCTTAACCAG[A]