Implementing Non-Equilibrium Networks with Active Circuits of Duplex Catalysts

Antti Lankinen
Department of Bioengineering, Imperial College London, UK
antti.lankinen15@imperial.ac.uk

Ismael Mullor Ruiz
Department of Bioengineering and Imperial College Centre for Synthetic Biology, Imperial College London, UK
i.mullor-ruiz16@imperial.ac.uk

Thomas E. Ouldridge
Department of Bioengineering and Imperial College Centre for Synthetic Biology, Imperial College London, UK
t.ouldridge@imperial.ac.uk

Abstract

DNA strand displacement (DSD) reactions have been used to construct chemical reaction networks in which species act catalytically at the level of the overall stoichiometry of reactions. These effective catalytic reactions are typically realised through one or more of the following: many-stranded gate complexes to coordinate the catalysis, indirect interaction between the catalyst and its substrate, and the recovery of a distinct "catalyst" strand from the one that triggered the reaction. These facts make emulation of the out-of-equilibrium catalytic circuitry of living cells more difficult. Here, we propose a new framework for constructing catalytic DSD networks: Active Circuits of Duplex Catalysts (ACDC). ACDC components are all double-stranded complexes, with reactions occurring through 4-way strand exchange. Catalysts directly bind to their substrates, and the "identity" strand of the catalyst recovered at the end of a reaction is the same molecule as the one that initiated it. We analyse the capability of the framework to implement catalytic circuits analogous to phosphorylation networks in living cells. We also propose two methods of systematically introducing mismatches within DNA strands to avoid leak reactions and introduce driving through net base pair formation. We then combine these results into a compiler to automate the process of designing DNA strands that realise any catalytic network allowed by our framework.

2012 ACM Subject Classification Hardware → Biology-related information processing

Keywords and phrases DNA strand displacement, Catalysis, Information-processing networks

Digital Object Identifier 10.4230/LIPIcs.DNA.2020.7

Supplementary Material A compiler to generate optimal sequences for each strand in any allowed catalytic network is available at `https://zenodo.org/record/3948343`.

1 Introduction

DNA is an attractive engineering material due to the high specificity of Watson-Crick base pairing and well-characterised thermodynamics of DNA hybridisation [13, 40], which give DNA the most predictable and programmable interactions of any natural or synthetic molecule [43]. DNA computing involves exploiting these properties to assemble computational devices made of DNA. The computational circuits are typically realised using DNA strand displacement (DSD) reactions, in which sections of DNA strands called *domains* with partial or full complementarity hybridise, displacing one or more previously hybridised strands in the process [55]. DSD is initiated by the binding of short complementary sequences called *toeholds*. It is helpful to divide DSD reactions into a few common reaction steps, including:


© Antti Lankinen, Ismael Mullor Ruiz, and Thomas E. Ouldridge;
licensed under Creative Commons License CC-BY



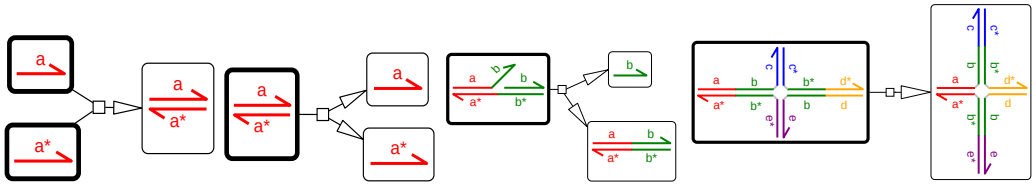
binding, unbinding, and three- or four-way strand displacement and branch migration, shown in Figure 1. DSD is an attractive scheme for computation as it can be used as a medium in which to realise chemical reaction networks (CRNs) [44], which provide an abstraction of systems exhibiting mass-action chemical kinetics and have been shown to be Turing complete [27]. DSD is then Turing complete as well [34,52]. DSD has been used to construct, for example, logic circuits [35,42], artificial neural networks [9,17,38], dynamical systems [46], catalytic networks [8,36,56], and other computational devices [1,53]. To facilitate testing and realisation of DSD systems, domain-level design tools [23,45] as well as domain-to-sequence translation [54] software have been introduced.

While DNA nanotechnology is concerned with using DNA as a non-biological material, a key goal of DNA nanotechnology is the imitation and augmentation of cellular systems. It is therefore worth considering how these natural systems typically perform computation and information processing. One ubiquitous biological paradigm for signal propagation and processing is the catalytic activation network, as exemplified by kinases [20, 28, 29]. Kinases are catalysts that modify substrates by phosphorylation and consume ATP in the process. These substrates can be, for example, transcription factors, but can also be kinases themselves that are either activated or deactivated by phosphorylation. The opposite function, dephosphorylation, is performed by phosphatases [4]. The emergent catalytic network then performs information propagation or computation by converting species, kinases and phosphatases, between their active and passive states. Kinase cascades are featured in many key biological functions, such as cellular growth, adhesion, and differentiation [28, 51] and long-term potentiation [47].

Most catalytic networks - including many with a simple topology and a constant steady state, such as a a single kinase and phosphatase species competing to activate/deactivate a substrate - operate out of equilibrium and consume fuel even in their steady state. This fuel-consuming, non-equilibrium behaviour is vital in allowing them to perform functions such as signal splitting, amplification, time integration and insulation [5, 12, 18, 30, 31]. Moreover, since the key molecular species are recovered rather than consumed by reactions, catalytic networks can operate continuously, responding to stimuli as they change over time - unlike many architectures for DSD-based computation and information processing that operate by allowing the key components to be consumed [1, 9, 38]. This ability to operate continuously is invaluable in autonomous environments such as living cells.

In this work, we propose a minimal mechanism for implementing reaction networks of molecules that exist in catalytically active and inactive states, a simple abstraction of natural kinase networks. In these catalytic activation networks, we implement arbitrary activation reactions of the form $A^{\text{on}} + B^{\text{off}} + \sum_i F_i \rightarrow A^{\text{on}} + B^{\text{on}} + \sum_i W_i,\ i \in \{1, 2, 3...\}$. Here, the catalyst A in its active state A^{on} drives B between its inactive and active states (B^{off}, B^{on}) by the conversion of one or more fuel molecules $\{F_i\}$ into waste $\{W_i\}$. Equivalent deactivation reactions in which an active catalyst deactivates a substrate are also considered.

The rest of this paper is organised as follows. In Section 2, we propose and motivate the concept of a direct bimolecular catalytic reaction and consider the necessary conditions for DSD species that are able to perform such reactions. Section 3 introduces a novel DSD framework to implement these reactions, and its computational properties are analysed in Section 4. Based on these findings, we propose a systematic method of introducing mismatched base pairs within species in our framework to improve its function in Section 5. We combine our findings and propositions into a software to automate the sequence-level design of any CRN that is realisable within our framework, and detail this software in Section 6. In Section 7, we discuss our framework, findings, and future work. We conclude the paper in Section 8.



(a) Bind. **(b)** Unbind. **(c)** Displace (3-way). **(d)** Branch migrate (4-way).

Figure 1 Basic reaction steps in the DSD formalism, as represented by Visual DSD [23]. Each domain is represented by a letter and a colour. "*" denotes the Watson-Crick complement. The barbed end of a strand indicates the 3' end.

2 Direct Action of Molecular Catalysts

Catalytic processes are those in which a species is both a reactant and a product. Such processes cannot result from an individual elementary reaction of binding, unbinding or unimolecular state change; catalysis is therefore only defined at the level of the overall stoichiometry of a series of elementary reactions. As a corollary, the same overall stoichiometry can result from many different combinations of elementary steps.

In kinase cascades, functional changes in substrates are a result of direct binding of the catalyst to the substrate. Moreover, the essential products of the reaction (the activated substrate and recovered catalyst) are the same molecules that initially bound to each other - albeit with some modification of certain residues, or turnover of small molecules such as ATP or ADP to which they are bound. Motivated by these facts, we propose the following definition for a direct bimolecular catalytic activation reaction.

▶ **Definition 1** (Direct bimolecular catalytic activation). *Consider the (non-elementary) process*

$$A^{\text{on}} + B^{\text{off}} + \sum_i F_i \rightarrow A^{\text{on}} + B^{\text{on}} + \sum_i W_i, \ i \in \{1, 2, 3...\},$$

where A *in active state* A^{on} *catalyses the conversion of* B *from inactive* B^{off} *to active* B^{on}*, using one or more ancillary fuels* $\{F_i\}$ *and producing waste* $\{W_i\}$*. The overall reaction is a direct bimolecular catalytic activation reaction if and only if:*

1. *The reaction is initialised with the interaction of* A *in state* A^{on} *and* B *in state* B^{off}*.*
2. A *and* B *molecules have molecular cores that are retained in the products, rather than the input molecules being consumed and distinct outputs released.*

Deactivation reactions have an equivalent form, but convert B *from* B^{on} *to* B^{off}*. If the same overall reaction stoichiometry is implemented by any set of reactions and species that violate conditions 1 and 2, we label the process a* pseudocatalytic bimolecular activation reaction.

Direct bimolecular catalytic (de)activation reactions have some important functional properties. The first is that, if the first step of the reaction requires the presence of A and B, nothing can happen unless both molecules are present. In pseudocatalytic implementations, as we discuss below, it is possible to produce activated B^{on} or sequester A even if no B molecules in state B^{off} are present, violating the logic of activation-based networks. The second is that the persistence of a molecular core of both the substrate and the catalyst allows either or both to be localised on a surface or scaffold, as is observed for some kinase cascades in living cells [14, 41, 50] and is often proposed for DNA-based systems [6, 7, 37, 39, 48].

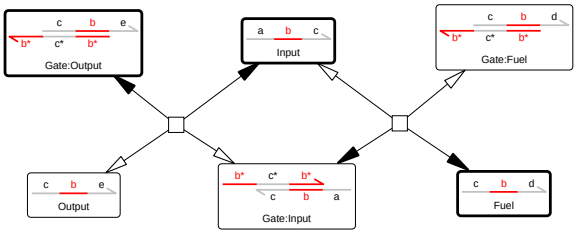


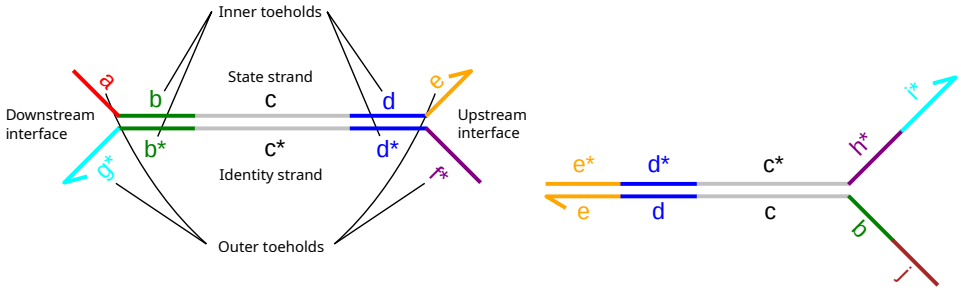
Figure 2 Catalytic reaction using a seesaw gate [19, 36]. Reactants are shown in bold boxes; the input acts pseudocatalytically to "convert" the fuel into an output, with ancillary gate complexes consumed and produced. Each compound reaction is illustrated by a small square, and consists of sequential bind, displace, and unbind reactions. All reactions are reversible; open arrows indicate reactions proceeding forwards, and closed arrows by reactions proceeding backwards.

A number of DNA computing frameworks have been developed to implement reactions of the stoichiometry of Definition 1. The simplest, illustrated in Figure 2 (a), involves a two-step seesaw gate [19, 36]. An input molecule (representing A in state A^{on} in Definition 1) binds to a gate-output complex (F), releasing the output (B in state B^{on}). The input is then displaced by a molecule conventionally described as the fuel, but fulfilling the role of B in state B^{off} from Definition 1 in the context of catalysis, recovering A in state A^{on} and generating a waste duplex (W). Although the A molecule recovered at the end of the process is the same one that initiated the process, the strands representing B^{off} and B^{on} molecules are distinct and the reaction is not initiated by the binding of A and B; it is therefore pseudocatalytic.

This pseudocatalysis can have important consequences. If a small quantity of input the strand representing A^{on} is added to a solution containing the gate-output complex F but no strand representing B^{off}, a large fraction of the A^{on} strand will be sequestered and a corresponding amount of the B^{on} strand produced. This sequestration of A and production of activated B from nothing violates the logic of ideal catalytic activation networks.

More complex strategies to implement reactions of the stoichiometry of Definition 1 using DSD exist [8, 34]. These approaches rely on the catalyst and substrate (A^{on} and B^{off} from Definition 1) interacting with a gate, rather than binding to each other, and the recovered catalyst and product are separate strands - the reactions are therefore pseudocatalytic. In certain limits, these strategies can approximate a mass-action dependence of reaction rates on the concentrations of A^{on} and B^{off} [8, 33], providing a better approximation to the logic of ideal catalytic activation circuits than the simple seesaw motif. The price, however, is the need to construct large multi-stranded gate complexes to facilitate the reaction; the complexity of these motifs is a major barrier to implementing such systems in autonomous setting such as living cells. Moreover, localising catalysts and substrates to a scaffold or surface remains challenging when the molecules themselves are not recovered.

We now consider how to design minimal DSD-based units that implement direct bimolecular catalytic (de)activation in catalytic activation networks. If the core of the substrate species B must be retained in both B^{off} and B^{on}, B^{off} and B^{on} cannot simply be two strands with a slightly different sequence. Instead, B^{off} and B^{on} must either be distinct complexes of strands, in which at least one strand is common, have different secondary structure within a single strand, or both. To avoid complexities in balancing the thermodynamics of hairpin loop formation with bimolecular association, and suppressing the kinetics of unimolecular rearrangement, we do not pursue the possibility of engineering metastable secondary structure



(a) Major species. **(b)** Ancillary species.

Figure 3 (a) Structure of major species in the ACDC system (substrates or catalysts), illustrating upstream and downstream interfaces, and inner and outer toeholds. The long central domain forms a stable binding duplex. (b) Structure of ancillary species (fuel, waste or substrate-catalyst complex).

within a strand. At least one of the states B^{off} and B^{on} of B must therefore consist of at least two strands. Moreover, since each activation state of each species must be a viable substrate in an arbitrary catalytic (de)activation network, the simplest approach that allows for a generic catalytic mechanism is to implement all substrate/catalyst species as two-stranded complexes.

3 ACDC: A Duplex-Based Catalytic DSD Framework

We introduce the Active Circuits of Duplex Catalysts (ACDC) scheme to implement catalytic activation networks through direct bimolecular catalytic (de)activation. Each reaction has three inputs: a substrate, a catalyst, and a single fuel complex. The outputs are a modified substrate, the recovered catalyst and a waste complex. The domain-level structures of these species are shown in Figure 3.

Substrates and catalysts – hereafter referred to as *major species* – are structurally identical. Each consists of two strands, each of which has one central long domain ($\sim$ 20 nucleotides (nt)) and two toeholds ($\sim$ 5 nt) on each side of the long domain. In major species, these strands are called the *identity strand* and the *state strand*. The identity strand is the preserved molecular core; the state strand specifies the activation state of a major species at a particular time (specifically, through the domain at its 5′ end - labelled “a” in Fig. 3).

The two strands in a major species are bound by three central domains; the *outer toeholds* at either end of the strands are *available* (unbound). Major species thus contain two *interfaces* at either end of the molecule, both displaying two available toeholds, one on each constituent strand. The *inner toeholds*, which are bound in major species, are described as *hidden*. We call the interface at the 5’ end of the state strand and the 3’ end of the identity strand the *downstream* interface and the interface with the 3’ end of the state strand and 5’ end of the identity strand the *upstream* interface.

All other two-stranded species in ACDC, including fuel and waste species, are described as *ancillary species*. They have a distinct structure from major species (Figure 3). Ancillary species also consist of two strands of five domains, but are bound by the central long domain and two shorter flanking toeholds (one outer toehold and one inner toehold) on one side. They therefore possess just one interface of available toeholds, but this interface presents two contiguous available toeholds on each strand.

The catalytic reaction of a single ACDC unit proceeds as shown in Figure 4. The downstream interface of the catalyst A in state A^{on} and upstream interface of the substrate B in state B^{off} bind together through recognition of all four available toeholds in the relevant interfaces. The resultant complex undergoes a 4-way branch migration, with the base pairs between the state and identity strand of the substrate and catalyst being exchanged for base pairs between the two state strands and the two identity strands. After the exchange of a hidden toehold and the central binding domain, the 4-stranded complex is held together by only two inner toeholds on either side of a 4-way junction. Dissociation by spontaneous detachment of these toeholds creates two ancillary product species, a waste $W_{AB\to B^{\text{on}}}$ and an intermediate complex AB. The sequence of these three reactions is called the *2r-4* reaction [21].

The fuel $F_{AB\to B^{\text{on}}}$ is identical to the waste, except for a single toehold. This toehold corresponds to the outer toehold of the state strand of B from the downstream interface. $F_{AB\to B^{\text{on}}}$ and AB can undergo another *2r-4* reaction, producing B in state B^{on} (equivalent to B^{off}, but with a single domain changed in the downstream interface) and recovering the catalyst. With the downstream interface of substrate B changed from that of B^{off} into that of B^{on}, the substrate has been activated and could act as a catalyst to another reaction, provided that an appropriate downstream substrate and fuel were present. An equivalent catalytic process could trigger another reaction converting B from B^{on} to B^{off}, *deactivating* B, analogous to dephosphorylation by a phosphatase. Note that the domain structures of ancillary species participating in an ACDC catalytic unit are unambiguously specified by the major species involved, since the individual strands are the same.

The ACDC mechanism borrows significantly from Qian and Winfree's design for surface-bound reaction networks [37]. In particular, that proposed framework also includes double-stranded species with identity and state strands, and exploits 4-way strand exchange reactions. However, the mechanistic details are more complex; species do not directly bind to each other, and interactions are mediated by multi-stranded gate complexes.

The basic ACDC unit in Figure 4 satisfies the conditions of Definition 1 for direct bimolecular catalytic activation, since the reaction is initiated by the binding of A in state A^{on} and B in state B^{off}, and the identity strands in the major species are retained throughout. In this case, a single fuel molecule is consumed and a single waste produced by a single catalytic conversion. ACDC relies on the experimentally-verified mechanism of toehold-mediated 4-way branch migration [10, 22, 25, 49]. The number of base pairs and complexes is unchanged by each *2r-4* reaction, and therefore a bias for clockwise activation cycles (as opposed to anticlockwise deactivation) would require a large excess of fuel complexes $F_{AB\to B^{\text{on}}}$ relative to waste $W_{AB\to B^{\text{on}}}$. In addition, for a single catalytic cycle to operate as intended, the following assumptions must hold:

▶ **Assumption 2** (Stability of complexes). *It is assumed that strands bound together by long domains are stable and will not spontaneously dissociate. It is also assumed that if two strands are bound by a pair of complementary domains, any adjacent pairs of complementary domains that could bind to form a contiguous duplex are not available.*

▶ **Assumption 3** (Detachment of products). *It is assumed that 4-stranded complexes bound together by two pairs of toehold domains either side of a junction can dissociate into duplexes.*

▶ **Assumption 4** (Need for two complementary toeholds to trigger branch migration). *It is assumed that if a 4-stranded complex is formed by the binding of a single pair of toehold domains, it will dissociate into product duplexes, rather than undergo branch migration.*

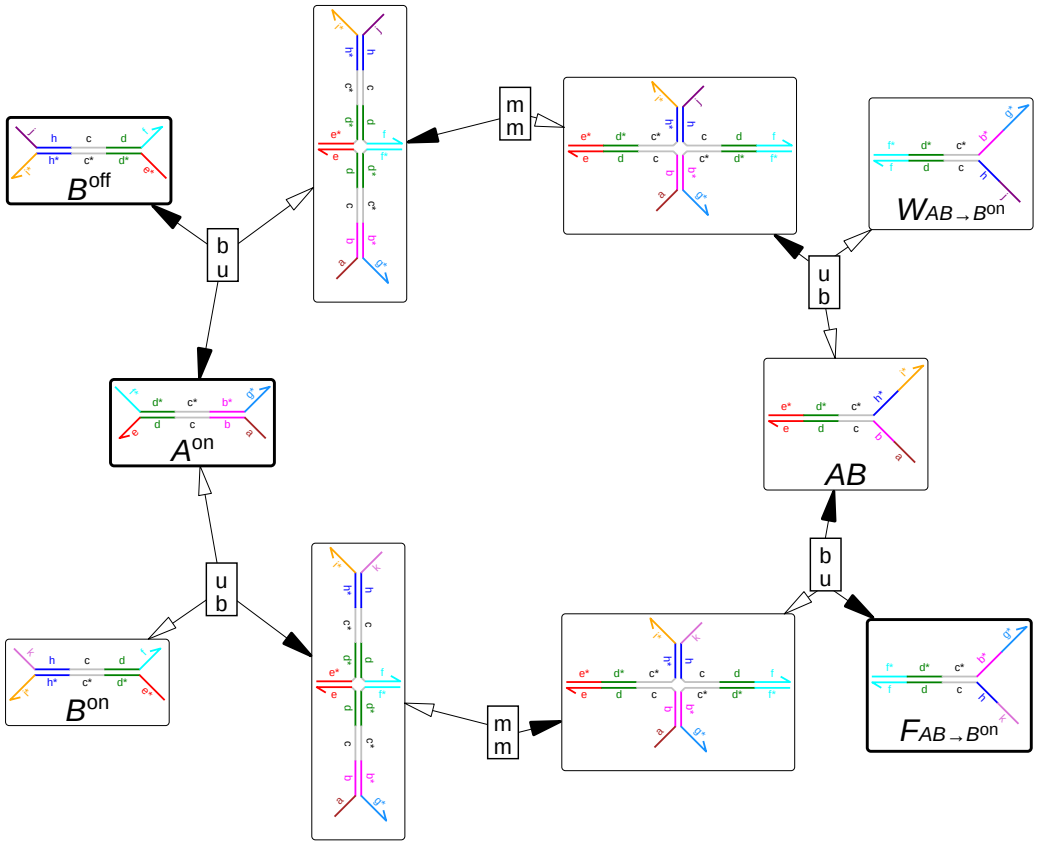


Figure 4 A basic ACDC reaction unit $A^{on} + B^{off} + F_{AB \to B^{on}} \to A^{on} + B^{on} + W_{AB \to B^{on}}$, as represented by Visual DSD [23]. Inputs to the reaction are shown in bold, and each small box corresponding to a reaction step is labelled with b/u (bind/unbind) or m (migrate). Imbalances in the concentration of fuel and waste drive the reaction clockwise (the direction indicated by open arrows).

Assumption 2 ensures that the system keeps its duplex-based structure, and that toeholds are well hidden in complexes when required. Assumption 3 is necessary to avoid all species being sequestered into 4-stranded complexes. Note that the assumption is not that detachment must happen extremely quickly, since such 4-stranded complexes need to be metastable enough to initiate branch migration with reasonable frequency. It is equivalent to the need for single toeholds to detach in 3-way toehold exchange reactions [36]. In practice, toehold length and conditions such as temperature could be tuned to optimize the relative propensity for branch migration and detachment. Given a reasonable balance between branch migration and detachment, Assumption 4 – which enables the switching of B from B^{off} and B^{on} to have a downstream effect – is also likely to be satisfied.

4 Domain-based constraints in ACDC Networks

Larger catalytic activation networks can be constructed from the basic ACDC units of Figure 4, since the substrate B in its activated state B^{on} can itself act as a catalyst. To describe these networks, let us now formalise the notation so that roman letters A, B, C etc. represent the nodes of the catalytic network, and italic symbols $A^{on}, B^{off}, F_{AB \to B^{on}}, AB$ etc.

represent the actual double-stranded molecular species in solution, carrying both identity and state information where appropriate. In this formalism, let A $\rightarrow$ B be a shorthand for the reaction $A^{\text{on}} + B^{\text{off}} + F_{AB \rightarrow B^{\text{on}}} \rightarrow A^{\text{on}} + B^{\text{on}} + W_{AB \rightarrow B^{\text{on}}}$ and C $\dashv$ B a shorthand for the reaction $C^{\text{on}} + B^{\text{on}} + F_{CB \rightarrow B^{\text{off}}} \rightarrow C^{\text{on}} + B^{\text{off}} + W_{CB \rightarrow B^{\text{off}}}$. Then, any potential catalytic activation network can be represented as a weighted directed graph, where nodes represent catalyst/substrate in the network and edges represent activation (edge weight 1) or deactivation (edge weight -1). Is it possible to realise any such graph using ACDC?

▶ **Assumption 5** (Toehold orthogonality). *We assume that there are sufficiently many toehold domain sequences that cross-talk between non-complementary domains is negligible.*

Since ACDC components share a long central domain, specificity is entirely driven through toehold recognition. As noted by Johnson, [21], there is a finite number of orthogonal short toehold domains that limits the size of the connected network that can be constructed. We assume that the network of interest does not violate this limit. We instead ask the realisability question at the level of domains.

▶ **Definition 6** (Realisability). *A catalytic activation network is realisable using the ACDC framework if a domain structure for the major species, which implies the domain structure of the ancillary species, can be specified such that:*

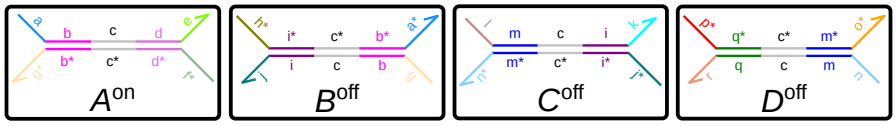
1. *All network edges* A $\rightarrow$ B (A $\dashv$ B) *are realised through basic ACDC units as illustrated in Figure 4. These units implement the overall reaction*

$$A^{\text{on}} + B^{\text{off}}(B^{\text{on}}) + F_{AB \rightarrow B^{\text{on}}(B^{\text{off}})} \rightarrow A^{\text{on}} + B^{\text{on}}(B^{\text{off}}) + W_{AB \rightarrow B^{\text{on}}(B^{\text{off}})},$$

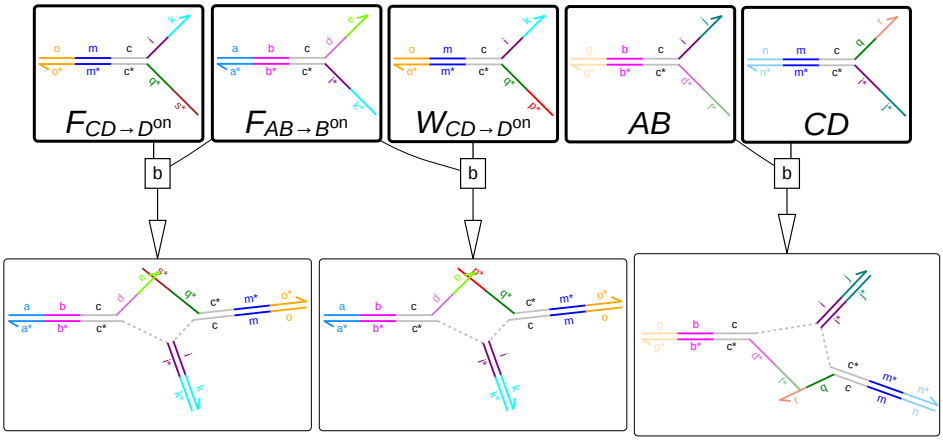
   where the bracketed terms apply to deactivation reactions.
2. *Other than the pairs of species that undergo reactions implied by condition 1, no pairs of species exist for which: (a) it is possible to exchange a pair of strands between the species and retain three contiguous bound domains in both the resultant complexes; and (b) the two species are able to bind via two available pairs of complementary toeholds. If (a) and (b) are both satisfied, the pair of species could undergo 2r-4 reactions as illustrated in Figure 4.*
3. *No two strands can form an uninterrupted duplex of four bound domains or more.*
4. *No two species (including all wastes, fuels and catalyst-substrate complexes) possess two available toehold pairs that could form a contiguous complementary duplex as shown in Figure 5(b).*

Condition 2 rules out reactions that respect the architecture of ACDC, but which involve reactants that are not intended to interact. Condition 3 rules out strand exchange reactions that allow an increase in the number of bound domains, which would sequester additional toeholds and violate the ACDC architecture (it is assumed that strand exchange reactions that would reduce the number of bound domains can be neglected). Condition 4 rules out the formation of 4-stranded complexes that can only dissociate by disrupting an uninterrupted two-toehold duplex. Contiguous duplexes of this kind are potentially stable, even if they cannot undergo strand exchange, and would potentially sequester components.

▶ **Lemma 7** (Realisability with activation implies realisability with deactivation). *If a catalytic activation network with purely activation reactions is realisable using the basic ACDC formalism, it is also realisable using the basic ACDC formalism if any subset of those reactions are converted to deactivation.*



(a) Major species.



(b) Ancillary species and unwanted reactions.

Figure 5 Major species and a subset of ancillary species from an implementation of A → B → C → D using the ACDC formalism. Three unwanted reactions, as identified in Lemma 10, occur between the shown ancillary species.

Proof. A deactivation reaction is simply an activation reaction with the role of the fuel and waste reversed. Therefore a domain structure specification that realises a given network with activation reactions also realises all networks of the same structure. ◀

4.1 Realisability of Motifs in the ACDC formalism

Since there are infinitely many networks, we restrict our analysis to a set of motifs (generalised versions of the minimal examples depicted in Figure 6), establishing whether these motifs can be realised in isolation. The *split*, *integrate*, *cascade*, *self-activation*, *bidirectional edge*, *feedback loop* (FBL), and *feedforward loop* (FFL) are chosen because of their importance in biology and synthetic biology [2, 15, 16]. The proofs of theorems not explicitly given in this section are provided in Appendix B.

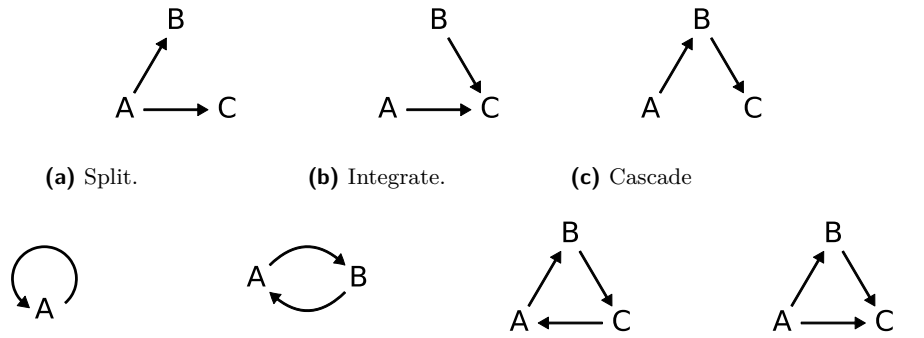
4.1.1 Motifs Without Loops

Theorems 8 and 9 establish that arbitrarily complex split and integrate motifs, constructed using ACDC in accordance with Definition 1, are realisable as per Definition 6.

▶ **Theorem 8** (Split motifs are realisable). *Consider the N reactions*

$$\mathrm{A} \to \mathrm{B}_1 \quad \mathrm{A} \to \mathrm{B}_2 \quad \dots \quad \mathrm{A} \to \mathrm{B}_N,$$

in which all B_i are distinct from A*. This network is realisable for any $N \geq 1$.*



(a) Split. **(b)** Integrate. **(c)** Cascade

(d) Auto-activation loop. **(e)** Bidirectional edge. **(f)** Feedback loop. **(g)** Feedforward loop.

Figure 6 Minimal example motifs of interest in a catalytic activation network.

▶ **Theorem 9** (Integrate motifs are realisable). *Consider the N reactions*

$$\mathrm{A}_1 \to \mathrm{B} \quad \mathrm{A}_2 \to \mathrm{B} \quad \dots \quad \mathrm{A}_N \to \mathrm{B},$$

in which all A_i *are distinct from* B. *This network is realisable for any* $N \geq 1$.

Although all networks consist of simply combining split and integrate motifs for each node, proving that all split and integrate motifs are realisable in isolation does not prove that any network assembled from them is realisable. We therefore explore other simple motifs. For example, consider the *cascade* motif (a 3-component example is illustrated in Figure 6).

▶ **Lemma 10** (The ancillary species of a catalyst's upstream reactions and substrate's downstream reactions cause leak reactions). *Consider a reaction* $\mathrm{B} \to \mathrm{C}$, *and further assume that* $\mathrm{A} \to \mathrm{B}$ *and* $\mathrm{C} \to \mathrm{D}$ *for a species* A *and a species* D. *Then AB and CD, and $F_{AB \to B^{\mathrm{on}}}$ and $F_{CD \to D^{\mathrm{on}}}/W_{CD \to D^{\mathrm{on}}}$ possess two available toehold pairs that could form a contiguous complementary duplex. No other violations of realisability occur.*

This failure is illustrated in Figure 5. The essence of the problem is that both the inner and outer toehold domains from the downstream end of B^{on} are available in AB and $F_{AB \to B^{\mathrm{on}}}$, and the inner and outer toehold domains from the upstream end of C are available in CD, $F_{CD \to D^{\mathrm{on}}}$ and $W_{CD \to D^{\mathrm{on}}}$. Since the downstream end of B^{on} is complementary to the upstream end of C^{off}, the result is that the species can bind to each other strongly.

▶ **Theorem 11** (Cascades with at most 3 components are realisable; longer cascades are not realisable). *Consider the set of N reactions* $\mathrm{A}_1 \to \mathrm{A}_2$, $\mathrm{A}_2 \to \mathrm{A}_3 \dots \mathrm{A}_{N-1} \to \mathrm{A}_N$, *in which all* A_i *are distinct. This network is realisable if and only if* $N \leq 3$.

Proof. A direct consequence of Lemma 10 and Definition 6. ◀

▶ **Theorem 12** (Long cascades are non-realisable due to a particular type of leak reaction only). *Consider the set of reactions* $\mathrm{A}_1 \to \mathrm{A}_2$, $\mathrm{A}_2 \to \mathrm{A}_3 \dots \mathrm{A}_{N-1} \to \mathrm{A}_N$ *for* $N > 3$, *in which all* A_i *are distinct. This network would be realisable if reactions between ancillary species A_iA_{i+1} and $A_{i+2}A_{i+3}$, and $F_{A_iA_{i+1} \to A_{i+1}^{\mathrm{on}}}$ and $F_{A_{i+2}A_{i+3} \to A_{i+3}^{\mathrm{on}}}/W_{A_{i+2}A_{i+3} \to A_{i+3}^{\mathrm{on}}}$, were absent.*

The result of Theorem 11 is discouraging, since cascades are a major feature of kinase networks [20, 29]. Nonetheless, we will continue the analysis of remaining motifs, and present a potential solution in Section 5.

4.1.2 Motifs With Loops

A network possesses a loop if it is possible to traverse a path that begins and ends at the same node without using the same edge twice. For the purposes of this classification, a given (directed) edge can be traversed in either direction. Loops are common components of natural networks, providing the possibility of oscillation, bistability and filtering [2, 11].

▶ **Theorem 13** (Loops of odd length are not realisable). *Consider a system of reactions* $\mathrm{A}_1 \leftrightarrow \mathrm{A}_2 \leftrightarrow \mathrm{A}_3 \ldots \mathrm{A}_{N-1} \leftrightarrow \mathrm{A}_1$*, where* $\leftrightarrow$ *indicates a catalytic activation in either direction. This network, a* directionless loop*, is not realizable if* N *is odd, unless the long central domain is self-complementary.*

Proof. ACDC circuits require that the long central domain alternates between a sequence and its complement in the identity strands of catalysts and their substrates. If N is odd, then the sequence must be self-complementary for this alternation to happen. ◀

Introducing a self-complementary central domain is a strategy that risks a competition between duplexes and single-stranded hairpins. We do not consider it further.

▶ **Theorem 14** (Self interactions and bidirectional edges are not realisable). *Consider a system of reactions* $\mathrm{A}_1 \rightarrow \mathrm{A}_2 \rightarrow \mathrm{A}_3 \ldots \mathrm{A}_{N-1} \rightarrow \mathrm{A}_1$*. This network is not realisable if* $N \leq 2$*.*

The ACDC system is not inherently suited to auto-activation or bidirectional interactions. These motifs require complementarity between both the downstream and upstream toeholds of either a single species, or two species. Strands in the system therefore violate condition 3 of Definition 6 and will tend to hybridise to form fully complementary duplexes.

An isolated feedback loop is a network of size N with a single directed path around the network. A simple example of length 3 is shown in Fig. 6(f).

▶ **Theorem 15** (Feedback loops are not realisable). *Consider the feedback loop* $\mathrm{A}_1 \rightarrow \mathrm{A}_2 \rightarrow \mathrm{A}_3 \ldots \mathrm{A}_{N-1} \rightarrow \mathrm{A}_1$*. Such a system is not realisable for any* N*.*

Proof. A direct consequence of Theorems 11, 13, and 14. ◀

As a consequence of Theorems 13 and 14, any realisable feedback loop must have $N \geq 4$. However, a feedback loop of this length faces the same issues as a cascade: formation of stable, undesired products between ancillary species. As with cascades, the problem is essentially local, due to interactions between ancillary species in reaction n and reaction $n + 2$.

▶ **Theorem 16** (Long feedback loops with an even number of units are non-realisable due to a particular type of leak reaction only). *Consider the feedback loop*

$$\mathrm{A}_1 \rightarrow \mathrm{A}_2 \quad \mathrm{A}_2 \rightarrow \mathrm{A}_3 \quad \ldots \quad \mathrm{A}_{N-1} \rightarrow \mathrm{A}_N \quad \mathrm{A}_N \rightarrow \mathrm{A}_1$$

For N *even,* $N \geq 4$*, this network would be realisable if reactions between ancillary species* A_iA_{i+1} *and* $A_{i+2}A_{i+3}$*, and* $F_{A_iA_{i+1}\rightarrow A_{i+1}^{\mathrm{on}}}$ *and* $F_{A_{i+2}A_{i+3}\rightarrow A_{i+3}^{\mathrm{on}}}/W_{A_{i+2}A_{i+3}\rightarrow A_{i+3}^{\mathrm{on}}}$*, were absent. Here, the index* j *in* A_j *should be interpreted modularly:* $\mathrm{A}_j = \mathrm{A}_{j-N}$ *for* $j > N$*.*

An isolated feedforward loop is a network of size N with two directed paths from one node i to another node j. Every other node appears exactly once in one of these paths. An example with path lengths of 1 and 2 is shown in Figure 6.

▶ **Theorem 17** (The relative lengths of paths are constrained in feedforward loops). *Consider the generalised feedforward loop*

$$\begin{array}{llll} \mathrm{A} \to \mathrm{B}_1 & \mathrm{B}_1 \to \mathrm{B}_2 & \dots & \mathrm{B}_{N-1} \to \mathrm{B}_N \quad \mathrm{B}_N \to \mathrm{D} \\ \mathrm{A} \to \mathrm{C}_1 & \mathrm{C}_1 \to \mathrm{C}_2 & \dots & \mathrm{C}_{M-1} \to \mathrm{C}_M \quad \mathrm{C}_M \to \mathrm{D} \end{array}$$

For such a network to be realisable, it is necessary that $N \geq 1$, $M \geq 1$, and $N - M$ is even.

Proof. The claim about $N - M$ having to be even follows from Theorem 13.

Assume for contradiction that a FFL with $N = 0$ and $M \geq 2$ and even is realisable. Since A activates C_1, and both A and C_M activate D, it must be that C_M can also perform a branch migration with C_1, which is an unwanted reaction violating condition 2 of Definition 6. ◀

Since each path in a feedforward loop is a cascade, Theorems 11 and 17 imply that only feedforward loops with a single intermediate in each branch are realisable.

▶ **Theorem 18** (Realisability of feedforward loops). *Consider the generalised feedforward loop*

$$\begin{array}{llll} \mathrm{A} \to \mathrm{B}_1 & \mathrm{B}_1 \to \mathrm{B}_2 & \dots & \mathrm{B}_{N-1} \to \mathrm{B}_N \quad \mathrm{B}_N \to \mathrm{D} \\ \mathrm{A} \to \mathrm{C}_1 & \mathrm{C}_1 \to \mathrm{C}_2 & \dots & \mathrm{C}_{M-1} \to \mathrm{C}_M \quad \mathrm{C}_M \to \mathrm{D} \end{array}$$

Such a system is realisable if and only if $N = 1$ and $M = 1$.

Proof. As a consequence of Theorems 8, 9, 11, and 17, all other FFLs are not realisable. The realisability of the FFL with N=1 and M=1 can be verified by inspection. ◀

Typically, feedforward loops use branches of different lengths to achieve a complex response to a signal over time [2,11]. Such networks are not realisable. Indeed, our analysis of various motifs has revealed that the majority are not realisable. Broadly speaking, there are a number of small motifs (e.g. auto-activation, bi-directional reactions, feedforward loops with no intermediates in one branch) that cannot be achieved because the major species themselves interact directly. In addition, loops of odd total length are not realisable due to the nature of complementary base pairs. However, most motifs are ruled out because of a single type of interaction, between the ancillary species in one reaction and the ancillary species in another reaction that occurs two steps downstream. In Section 5, we propose a strategy to overcome this last problem, massively increasing the scope of the ACDC framework.

5 Overcoming the Cascade Leak Reaction and Introducing Hidden Thermodynamic Drive

The most severe limitation of the ACDC system detailed in Section 3 is expressed by Theorem 11. Long cascades, and loops incorporating cascades, are non-realisable due to interactions between ancillary species of a given reaction, and ancillary species of a reaction separated by two catalytic steps (Theorem 12).

▶ **Assumption 19** (Mismatches destabilise complexes held together by two contiguous toehold domains). *We assume that a single mismatched C-C or G-G base pair, positioned adjacent to the interface of two toehold domains, is sufficiently destabilizing that an unwanted complex formed only by the binding of these toehold domains no longer precludes realisability.*

The basic design of the ACDC motif assumes that toehold binding is relatively weak; two toehold domains on either side of a junction must be able to dissociate by Assumption 2. Individual C-C or G-G mismatches are known to be highly destabilising [40], and should similarly allow for two contiguous domains to detach. Given Assumption 19, the challenge is then to systematically introduce mismatches so that all interactions between ancillary species identified in Theorem 12 are compromised by a mismatch, without compromising intended circuit activity. Our full scheme is visualised in Figure 7.

▶ **Definition 20** (Mismatches proposed to destabilize unintended complexes)**.** *We propose the following mismatches.*

1. *We propose that the upstream interface of every major species is made distinct for active and inactive states. Specifically, we introduce a G base at the inner edge of the outer toehold domain of the state strand of the inactive species, and a C base in the same position for the active species. Catalysts that (de)activate that species possess a C(G) in the complementary position of their downstream interface.*
2. *We introduce a C-C mismatch at the outer edge of the inner toehold domain at the downstream interface of each major species. This mismatch is eliminated in the formation of waste complexes, and retained in the substrate-catalyst complexes.*

▶ **Assumption 21** (Mismatches cannot cause leak reactions)**.** *We assume that the sequence constraints introduced by mismatch inclusion do not violate Assumption 5, and that the destabilisation of duplexes does not violate Assumption 2.*

In practice, mismatches will likely result in some increase in the rate of interactions between otherwise hidden toeholds; we assume that these rates remain negligible.

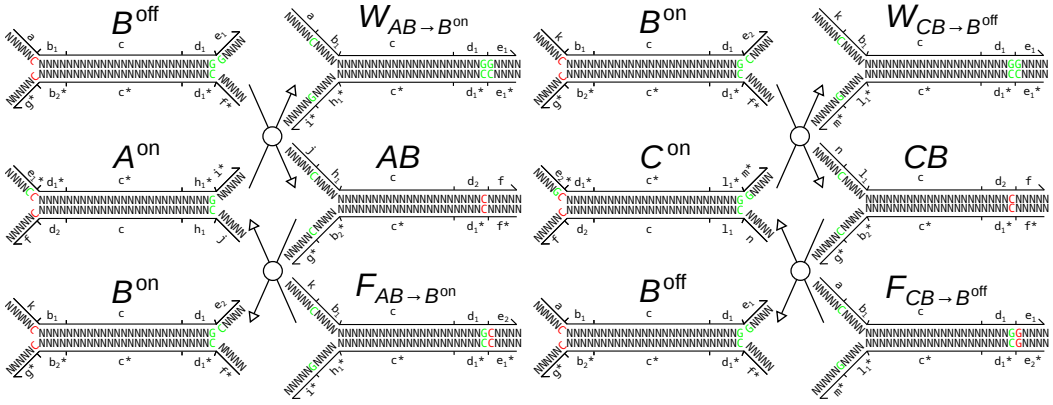
▶ **Theorem 22** (Mismatches successfully destabilize unintended complexes)**.** *The scheme proposed in Definition 20 satisfies the following:*

1. *All motifs that are realisable in the mismatch-free ACDC design remain realisable in the mismatch-based scheme.*
2. *Cascades of arbitrary length N with at most the first and last reactions deactivating are realisable;*
3. *Feedback loops with N even and $N \geq 6$ in which all reactions are activating are realisable;*
4. *Feedforward loops with $N \geq 1$, $M \geq 1$, $N - M$ even, in which at most the first and last reactions are deactivating in each branch, are realisable.*

The proof of Theorem 22 is given in Appendix B.

Note that the introduction of mismatches proposed in Definition 20 invalidates Lemma 7, since the downstream domains of activating and deactivating catalysts are now distinct. Indeed, the described strategy only eliminates unwanted sequestration in cascades in which the intermediate steps are activating. Nonetheless, it makes complex networks in which - for example - deactivating catalysts are always active realisable. Networks of this kind are common in biology [20, 29].

The first type of mismatch in Definition 20 ensures that there is always a C-C mismatch between the upstream toeholds of the state strand of A_{i+2}^{on} and the downstream toeholds of the state strand of A_{i+1}^{on} in the cascade $\mathrm{A}_i \to / \dashv \mathrm{A}_{i+1} \to \mathrm{A}_{i+2} \to / \dashv \mathrm{A}_{i+3}$, weakening the unwanted binding between the fuel and waste species identified in Theorem 12. Here $\to / \dashv$ indicates activation or deactivation. The second type of mismatch in Definition 20 ensures that the upstream toeholds of the identity strand of A_{i+2} are no longer fully complementary to the downstream toeholds of A_{i+1} in the cascade $\mathrm{A}_i \to / \dashv \mathrm{A}_{i+1} \to / \dashv \mathrm{A}_{i+2} \to / \dashv \mathrm{A}_{i+3}$, weakening the unwanted binding between ancillary species $A_i A_{i+1}$ and $A_{i+2} A_{i+3}$.



(a) A → B. **(b)** C ⊣ B.

Figure 7 Illustration of the proposed mismatch schemes for reactions A → B and C ⊣ B, assuming toeholds of length 5 nucleotides and central domains of length 17 nucleotides. Specific mismatched bases are highlighted in red, and the same bases are highlighted in green when not part of a mismatch. The domains are separated with ticks on each species, and upstream interfaces of the major species are shown on the right of each diagram.

Having proposed these mismatches, it is important to determine that they would not compromise the intended reactions. The first type of mismatch in Definition 20 is not present in any complex that must form during the operation of the network; only in the initially-prepared fuel and if a (de)activating catalyst binds to an (in)active substrate. It therefore presents no issues for intended reactions.

The second type of mismatch in Definition 20 is more subtle. When a catalyst A^{on} interacts with its substrate B^{off}, a mismatch at the very end of the catalyst duplex is converted into a mismatch within the stem of of the catalyst-substrate complex AB. Since mismatches are known to be more destabilizing in duplex interiors [32, 40], this conversion represents a local barrier to branch migration. The thermodynamic favourability of the full *2r-4* reaction $A^{\text{on}} + B^{\text{off}} \to AB + W_{AB \to B^{\text{on}}}$ (or the equivalent step in a deactivation reaction) is marginal, as the mismatch at the downstream end of B^{off} counters this barrier. We assume that the local barriers introduced would not prohibit the intended reactions - indeed, conventional 3-way strand displacement is able to proceed through unmitigated C-C mismatch formation, albeit with a significant effect on kinetics [26]. In this case, any penalty is likely to be far weaker.

The second step of the catalytic turnover, $AB + F_{AB \to B^{\text{on}}} \to A^{\text{on}} + B^{\text{on}}$ (or the equivalent in a deactivation) is thermodynamically favourable (two internal mismatches are converted into exterior mismatches) and without local barriers, although one of the toeholds is effectively shortened to 4 base pairs. The overall catalytic (de)activation cycle effectively eliminates a single C-C (G-G) mismatch initially present in the fuel. The reaction as a whole is therefore driven forwards by the free energy of base-pairing via "hidden thermodynamic driving" [19]; products are more stable than reactants without consumption of initially available toeholds. In this sense, the mismatches proposed in Definition 20 will improve the efficacy of the ACDC motif, as the concentration excess of fuel relative to waste required to drive the reaction in the desired direction would be reduced.

6 A Compiler for ACDC Networks

To construct an ACDC network that implements a given graph, three things need to be done: (1) verification that the network is realisable; (2) enumerating all domains on all species given the graph topology; and (3) compile sequences for each domain and thus for each strand present in the system. We have created an ACDC compiler with this functionality [24]. While compilers for DSD systems that could be potentially be extended to accommodate our framework exist [3, 46], we decided to make our own since our framework has unique requirements about verifying the feasibility of a given CRN and introducing mismatches within domains.

The first part is done, at least at the level of each cascade and loop present, by analysing the properties of a given graph. For every pair of nodes i, j, all directed simple paths are computed. We search for paths of length $N \geq 3$ that containing edge weights of -1 anywhere other than at the first or last edge; these cascades are not rendered realisable by our mismatch scheme, per Theorem 22. Moreover, if there exists more than 1 path between the nodes, then either a FFL (at least two paths from i to j or from j to i) or a FBL (at least one path from i to j and from j to i) exists in the graph. Furthermore, if there exists more than 1 path between the nodes after transforming the graph to an undirected form, there exists a "directionless loop" (Theorem 13) in the graph. The realisability of the loop(s) can be verified from the lengths of the paths according to Theorems 13 and 22.

If a given graph is found to be realisable, then domains are assigned for each strand of each species, such that all complementarities and mismatches required by the topology are satisfied. This ask can be achieved by local analysis of the network topology.

Finally, a NUPACK [54] script is generated to generate optimal sequences for each strand. The required mismatches are hard-coded into the domain definitions in the script. The software is available at `https://zenodo.org/record/3948343`.

7 Discussion

We have introduced the ACDC scheme for constructing DNA-based networks that perform direct catalysis, analysed its shortcomings, and subsequently proposed practical improvements. As of now, we have focused only on the realisability of ACDC implementations for some graphs, not their dynamical behaviour. Three natural directions for further theoretical investigation are: (1) proving the realisability of arbitrary networks; (2) implementing additional hidden thermodynamic driving so that both *2r-4* substeps of a catalytic reaction are thermodynamically downhill; and (3) automated design of ACDC networks to perform some desired transfer function between input concentrations $x_i(t),\ i = 1..N$ and output concentrations $y_j(t),\ j = 1..M$. With regard to the first, we conjecture that all violations of realisability in arbitrary networks are attributable to the causes identified in Section 4.

Equally important, however, is experimentally testing the ACDC motif. Whilst 4-way branch migration has been used in several contexts [10, 22, 25, 49], the toehold exchange mechanism proposed here is relatively untested. It is also important to establish that the mismatches function as intended, limiting sequestration reactions and providing strong overall thermodynamic driving without causing excessive local barriers that frustrate the necessary reactions. A final consideration is the possibility of leak reactions involving non-complementary toeholds that we have assumed to be negligible. It remains to be established that unintended reactions will occur at a negligible rate, particularly in the context of species containing mismatches. This research is ongoing within the group.

A key property of ACDC is the two recognition interfaces within each species and the inherent symmetry in the species that follows. While this symmetry is a design feature that allows both substrate-like and catalyst-like behaviour for a single species, it also has a drawback that domains that are essential for some reaction to occur are also present in reactions where they only act as identity placeholders (downstream interface of a catalyst and an upstream interface of a substrate) that do not interact with any other domain. Consider the reaction in Figure 4; the identity of the "placeholder domains" a, b, g, h, i, j, k that aren't involved in the initial binding and migration reactions could be swapped to arbitrary domains that aren't complementary with d, e, f or each other in only one species and the reaction could still occur (assuming the correct fuel species is generated based on the substrate and catalyst). However, this may not be possible if A and B are part of some larger computational network where the placeholder domain identities are important. Another drawback of the symmetry is the limitation of loop lengths to even numbers, characterised in Theorem 13. An obvious potential mitigation to this problem is to make the central domain its own complement, although this choice risks the formation of self-complementary hairpins.

The weaknesses of the ACDC motif invite the exploration of other possible designs of catalytic activation networks that operate via direct bimolecular catalysis. It is an open question as to whether the shortcomings of ACDC can be mitigated without a substantial increase in complexity or abandoning the mechanism of direct catalytic action.

8 Conclusion

We have established the concept of a direct catalytic reaction and discussed why previous work on catalytic DNA computing does not fulfil this definition. We have then proposed a framework, ACDC, for implementing non-equilibrium catalytic (de)activation networks using direct catalytic activation, analogous to systems seen in living cells. ACDC is simple in the sense that all species contain only two strands - an important consideration in the context of implementing DSD circuitry in a broad range of contexts.

We have analysed the framework's expressiveness by exploring the implementation of seven network motifs with ACDC. The basic design is highly limited by the inherent symmetry of components, prohibiting long cascades and most feedforward and feedback loops. However, we propose that systematic placement of mismatches can obviate these difficulties in many contexts. Moreover, we argue that these initially-present mismatches can contribute a "hidden thermodynamic driving" [19] to the ACDC motifs, increasing the robustness of the design to subtleties in DNA thermodynamics and reducing the concentration imbalances of fuels required to drive the reactions forward. We present a compiler for the sequence design of ACDC-based networks that implements these findings [24].

References

1 Leonard Adleman. Molecular computation of solutions to combinatorial problems. *Science*, 266(5187):1021–1024, November 1994. doi:10.1126/science.7973651.

2 Uri Alon. *An introduction to systems biology: design principles of biological circuits, Second Edition*. CRC Press LLC, Boca Raton, UNITED STATES, 2019.

3 Stefan Badelt, Seung Woo Shin, Robert F. Johnson, Qing Dong, Chris Thachuk, and Erik Winfree. A general-purpose CRN-to-DSD compiler with formal verification, optimization, and simulation capabilities. In *DNA computing and molecular programming*, Lecture notes in computer science, pages 232–248. Springer, Cham, September 2017. doi:10.1007/978-3-319-66799-7_15.

4 David Barford, Amit K. Das, and Marie-Pierre Egloff. The structure and mechanism of protein phosphatases: insights into catalysis and regulation. *Annual Review of Biophysics and Biomolecular Structure*, 27(1):133–164, June 1998. Publisher: Annual Reviews. doi:10.1146/annurev.biophys.27.1.133.

5 John P. Barton and Eduardo D. Sontag. The energy costs of insulators in biochemical networks. *Biophysical Journal*, 104(6):1380–1390, March 2013. doi:10.1016/j.bpj.2013.01.056.

6 Hieu Bui, Shalin Shah, Reem Mokhtar, Tianqi Song, Sudhanshu Garg, and John Reif. Localized DNA hybridization chain reactions on DNA origami. *ACS Nano*, 12(2):1146–1155, February 2018. Publisher: American Chemical Society. doi:10.1021/acsnano.7b06699.

7 Gourab Chatterjee, Neil Dalchau, Richard A. Muscat, Andrew Phillips, and Georg Seelig. A spatially localized architecture for fast and modular DNA computing. *Nature Nanotechnology*, 12(9):920–927, September 2017. Number: 9 Publisher: Nature Publishing Group. doi:10.1038/nnano.2017.127.

8 Yuan-Jyue Chen, Neil Dalchau, Niranjan Srinivas, Andrew Phillips, Luca Cardelli, David Soloveichik, and Georg Seelig. Programmable chemical controllers made from DNA. *Nature Nanotechnology*, 8(10):755–762, October 2013. doi:10.1038/nnano.2013.189.

9 Kevin M. Cherry and Lulu Qian. Scaling up molecular pattern recognition with DNA-based winner-take-all neural networks. *Nature*, 559(7714):370–376, July 2018. doi:10.1038/s41586-018-0289-6.

10 Nadine L. Dabby. *Synthetic molecular machines for active self-assembly: prototype algorithms, designs, and experimental study*. PhD thesis, California Institute of Technology, Pasadena, California, 2013. URL: https://pdfs.semanticscholar.org/e668/440cdb786ea7c2d0d6ae306c5aefef1208f6.pdf.

11 Wiet de Ronde and Pieter Rein ten Wolde. Multiplexing oscillatory biochemical signals. *Physical Biology*, 11(2):026004, April 2014. doi:10.1088/1478-3975/11/2/026004.

12 Abhishek Deshpande and Thomas E. Ouldridge. High rates of fuel consumption are not required by insulating motifs to suppress retroactivity in biochemical circuits. *Engineering Biology*, 1(2):86–99, December 2017. Publisher: IET Digital Library. doi:10.1049/enb.2017.0017.

13 Robert M. Dirks, Justin S. Bois, Joseph M. Schaeffer, Erik Winfree, and Niles A. Pierce. Thermodynamic analysis of interacting nucleic acid strands. *SIAM Review*, 49(1):65–88, January 2007. Publisher: Society for Industrial and Applied Mathematics. doi:10.1137/060651100.

14 Elaine A. Elion. Ste5: a meeting place for MAP kinases and their associates. *Trends in Cell Biology*, 5(8):322–327, August 1995. doi:10.1016/S0962-8924(00)89055-8.

15 Michael B. Elowitz and Stanislas Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403(6767):335–338, January 2000. Number: 6767 Publisher: Nature Publishing Group. doi:10.1038/35002125.

16 Timothy S. Gardner, Charles R. Cantor, and James J. Collins. Construction of a genetic toggle switch in Escherichia coli. *Nature*, 403(6767):339–342, January 2000. Number: 6767 Publisher: Nature Publishing Group. doi:10.1038/35002131.

17 Anthony J. Genot, Teruo Fujii, and Yannick Rondelez. Scaling down DNA circuits with competitive neural networks. *Journal of The Royal Society Interface*, 10(85):20130212, August 2013. Publisher: Royal Society. doi:10.1098/rsif.2013.0212.

18 Christopher C. Govern and Pieter Rein ten Wolde. Energy dissipation and noise correlations in biochemical sensing. *Physical Review Letters*, 113(25):258102, December 2014. Publisher: American Physical Society. doi:10.1103/PhysRevLett.113.258102.

19 Natalie E. C. Haley, Thomas E. Ouldridge, Ismael Mullor Ruiz, Alessandro Geraldini, Ard A. Louis, Jonathan Bath, and Andrew J. Turberfield. Design of hidden thermodynamic driving for non-equilibrium systems via mismatch elimination during DNA strand displacement. *Nature Communications*, 11(1):2562, May 2020. Number: 1 Publisher: Nature Publishing Group. doi:10.1038/s41467-020-16353-y.

20 Ira Herskowitz. MAP kinase pathways in yeast: for mating and more. *Cell*, 80(2):187–197, January 1995. doi:10.1016/0092-8674(95)90402-6.

21 Robert F. Johnson. Impossibility of sufficiently simple chemical reaction network implementations in DNA strand displacement. In Ian McQuillan and Shinnosuke Seki, editors, *Unconventional computation and natural computation*, Lecture notes in computer science, pages 136–149. Springer International Publishing, 2019. doi:10.1007/978-3-030-19311-9_12.

22 Shohei Kotani and William L. Hughes. Multi-arm junctions for dynamic DNA nanotechnology. *Journal of the American Chemical Society*, 139(18):6363–6368, May 2017. doi:10.1021/jacs.7b00530.

23 Matthew R. Lakin, Simon Youssef, Filippo Polo, Stephen Emmott, and Andrew Phillips. Visual DSD: a design and analysis tool for DNA strand displacement systems. *Bioinformatics*, 27(22):3211–3213, November 2011. doi:10.1093/bioinformatics/btr543.

24 Antti Lankinen. ACDC compiler, July 2020. URL: https://zenodo.org/record/3948343.

25 Tong Lin, Jun Yan, Luvena L. Ong, Joanna Robaszewski, Hoang D. Lu, Yongli Mi, Peng Yin, and Bryan Wei. Hierarchical assembly of DNA nanostructures based on four-way toehold-mediated strand displacement. *Nano Letters*, 18(8):4791–4795, August 2018. Publisher: American Chemical Society. doi:10.1021/acs.nanolett.8b01355.

26 Robert R. F. Machinek, Thomas E. Ouldridge, Natalie E. C. Haley, Jonathan Bath, and Andrew J. Turberfield. Programmable energy landscapes for kinetic control of DNA strand displacement. *Nature Communications*, 5(1):1–9, November 2014. Number: 1 Publisher: Nature Publishing Group. doi:10.1038/ncomms6324.

27 Marcelo O. Magnasco. Chemical kinetics is Turing universal. *Physical Review Letters*, 78(6):1190–1193, February 1997. doi:10.1103/PhysRevLett.78.1190.

28 G. Manning, D. B. Whyte, R. Martinez, T. Hunter, and S. Sudarsanam. The protein kinase complement of the human genome. *Science*, 298(5600):1912–1934, December 2002. Publisher: American Association for the Advancement of Science Section: Review. doi:10.1126/science.1075762.

29 Christopher J. Marshall. MAP kinase kinase kinase, MAP kinase kinase and MAP kinase. *Current Opinion in Genetics & Development*, 4(1):82–89, February 1994. doi:10.1016/0959-437X(94)90095-7.

30 Pankaj Mehta, Alex H. Lang, and David J. Schwab. Landauer in the agea of synthetic biology: energy consumption and information processing in biochemical networks. *Journal of Statistical Physics*, 162(5):1153–1166, March 2016. doi:10.1007/s10955-015-1431-6.

31 Thomas E. Ouldridge, Christopher C. Govern, and Pieter Rein ten Wolde. Thermodynamics of computational copying in biochemical systems. *Physical Review X*, 7(2):021004, April 2017. Publisher: American Physical Society. doi:10.1103/PhysRevX.7.021004.

32 Thomas E. Ouldridge, Ard A. Louis, and Jonathan P. K. Doye. Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model. *The Journal of Chemical Physics*, 134(8):085101, February 2011. Publisher: American Institute of Physics. doi:10.1063/1.3552946.

33 Tomislav Plesa. Stochastic approximation of high-molecular by bi-molecular reactions. *arXiv:1811.02766 [math, q-bio]*, November 2018. arXiv: 1811.02766. URL: http://arxiv.org/abs/1811.02766.

34 Lulu Qian, David Soloveichik, and Erik Winfree. Efficient Turing-universal computation with DNA polymers. In Yasubumi Sakakibara and Yongli Mi, editors, *DNA computing and molecular programming*, Lecture notes in computer science, pages 123–140, Berlin, Heidelberg, 2011. Springer. doi:10.1007/978-3-642-18305-8_12.

35 Lulu Qian and Erik Winfree. Scaling up digital circuit computation with DNA strand displacement cascades. *Science*, 332(6034):1196–1201, June 2011. doi:10.1126/science.1200520.

36 Lulu Qian and Erik Winfree. A simple DNA gate motif for synthesizing large-scale circuits. *Journal of the Royal Society Interface*, 8(62):1281–1297, September 2011. doi:10.1098/rsif.2010.0729.

37 Lulu Qian and Erik Winfree. Parallel and scalable computation and spatial dynamics with DNA-based chemical reaction networks on a surface. In Satoshi Murata and Satoshi Kobayashi, editors, *DNA computing and molecular programming*, Lecture notes in computer science, pages 114–131, Cham, 2014. Springer International Publishing. doi:10.1007/978-3-319-11295-4_8.

38 Lulu Qian, Erik Winfree, and Jehoshua Bruck. Neural network computation with DNA strand displacement cascades. *Nature*, 475(7356):368–372, July 2011. doi:10.1038/nature10262.

39 Ismael Mullor Ruiz, Jean-Michel Arbona, Amitkumar Lad, Oscar Mendoza, Jean-Pierre Aimé, and Juan Elezgaray. Connecting localized DNA strand displacement reactions. *Nanoscale*, 7(30):12970–12978, July 2015. Publisher: The Royal Society of Chemistry. doi:10.1039/C5NR02434J.

40 John SantaLucia and Donald Hicks. The thermodynamics of DNA structural motifs. *Annual Review of Biophysics and Biomolecular Structure*, 33(1):415–440, 2004. _eprint: https://doi.org/10.1146/annurev.biophys.32.110601.141800. doi:10.1146/annurev.biophys.32.110601.141800.

41 Hans J. Schaeffer, Andrew D. Catling, Scott T. Eblen, Lara S. Collier, Anke Krauss, and Michael J. Weber. MP1: a MEK binding partner that enhances enzymatic activation of the MAP kinase cascade. *Science*, 281(5383):1668–1671, September 1998. Publisher: American Association for the Advancement of Science Section: Report. doi:10.1126/science.281.5383.1668.

42 Georg Seelig, David Soloveichik, David Yu Zhang, and Erik Winfree. Enzyme-free nucleic acid logic circuits. *Science*, 314(5805):1585–1588, December 2006. Publisher: American Association for the Advancement of Science Section: Report. doi:10.1126/science.1132493.

43 Nadrian C. Seeman and Hanadi F. Sleiman. DNA nanotechnology. *Nature Reviews Materials*, 3(1):1–23, November 2017. doi:10.1038/natrevmats.2017.68.

44 David Soloveichik, Georg Seelig, and Erik Winfree. DNA as a universal substrate for chemical kinetics. *Proceedings of the National Academy of Sciences*, 107(12):5393–5398, March 2010. doi:10.1073/pnas.0909380107.

45 Carlo Spaccasassi, Matthew R. Lakin, and Andrew Phillips. A logic programming language for computational nucleic acid devices. *ACS synthetic biology*, 8(7):1530–1547, July 2019. doi:10.1021/acssynbio.8b00229.

46 Niranjan Srinivas, James Parkin, Georg Seelig, Erik Winfree, and David Soloveichik. Enzyme-free nucleic acid dynamical systems. *Science*, 358(6369), December 2017. doi:10.1126/science.aal2052.

47 J. David Sweatt. The neuronal MAP kinase cascade: a biochemical signal integration system subserving synaptic plasticity and memory. *Journal of Neurochemistry*, 76(1):1–10, 2001. doi:10.1046/j.1471-4159.2001.00054.x.

48 Mario Teichmann, Enzo Kopperger, and Friedrich C. Simmel. Robustness of localized DNA strand displacement cascades. *ACS Nano*, 8(8):8487–8496, August 2014. Publisher: American Chemical Society. doi:10.1021/nn503073p.

49 Suvir Venkataraman, Robert M. Dirks, Paul W. K. Rothemund, Erik Winfree, and Niles A. Pierce. An autonomous polymerization motor powered by DNA hybridization. *Nature Nanotechnology*, 2(8):490–494, August 2007. Number: 8 Publisher: Nature Publishing Group. doi:10.1038/nnano.2007.225.

50 Alan J. Whitmarsh, Julie Cavanagh, Cathy Tournier, Jun Yasuda, and Roger J. Davis. A mammalian scaffold complex that selectively mediates MAP kinase activation. *Science*, 281(5383):1671–1674, September 1998. Publisher: American Association for the Advancement of Science Section: Report. doi:10.1126/science.281.5383.1671.

51 Christian Widmann, Spencer Gibson, Matthew B. Jarpe, and Gary L. Johnson. Mitogen-activated protein kinase: conservation of a three-kinase module from yeast to human. *Physiolo-*

gical Reviews, 79(1):143–180, January 1999. Publisher: American Physiological Society. doi:10.1152/physrev.1999.79.1.143.

52 Wataru Yahiro and Masami Hagiya. Implementation of Turing machine using DNA strand displacement. In Carlos Martín-Vide, Takaaki Mizuki, and Miguel A. Vega-Rodríguez, editors, *Theory and Practice of Natural Computing*, Lecture notes in computer science, pages 161–172, Cham, 2016. Springer International Publishing. doi:10.1007/978-3-319-49001-4_13.

53 Peng Yin, Harry M. T. Choi, Colby R. Calvert, and Niles A. Pierce. Programming biomolecular self-assembly pathways. *Nature*, 451(7176):318–322, January 2008. Number: 7176 Publisher: Nature Publishing Group. doi:10.1038/nature06451.

54 Joseph N. Zadeh, Conrad D. Steenberg, Justin S. Bois, Brian R. Wolfe, Marshall B. Pierce, Asif R. Khan, Robert M. Dirks, and Niles A. Pierce. NUPACK: Analysis and design of nucleic acid systems. *Journal of Computational Chemistry*, 32(1):170–173, 2011. doi:10.1002/jcc.21596.

55 David Yu Zhang and Georg Seelig. Dynamic DNA nanotechnology using strand-displacement reactions. *Nature Chemistry*, 3(2):103–113, February 2011. doi:10.1038/nchem.957.

56 David Yu Zhang, Andrew J. Turberfield, Bernard Yurke, and Erik Winfree. Engineering entropy-driven reactions and networks catalyzed by DNA. *Science*, 318(5853):1121–1125, November 2007. Publisher: American Association for the Advancement of Science Section: Report. doi:10.1126/science.1148532.

A Notation For ACDC Species and Reactions

$[a\ b]$ denotes a strand consisting of domains a and b. Logical not is denoted by $\neg$ and logical and by $\wedge$.

▶ **Definition 23.** *(ACDC major species structure). Each major species in an ACDC network consists of two strands, each of which have one long domain and four toehold domains. The two strands are called state strand and identity strand based on the fact that one strand decodes the state of the species and other the identity. A major species X has the following domains (note the use of H for "inner" to avoid confusion with "identity"):*

- $SH5(X)$*: the inner toehold domain on the 5' side (downstream end) of the state strand.*
- $SO5(X)$*: the outer toehold domain on the 5' side (downstream end) of the state strand.*
- $SH3(X)$*: the inner toehold domain on the 3' side (upstream end) of the state strand.*
- $SO3(X)$*: the outer toehold domain on the 3' side (upstream end) of the state strand.*
- $IH5(X)$*: the inner toehold domain on the 5' side (upstream end) of the identity strand.*
- $IO5(X)$*: the outer toehold domain on the 5' side (upstream end) of the identity strand.*
- $IH3(X)$*: the inner toehold domain on the 3' side (downstream end) of the identity strand.*
- $IO3(X)$*: the outer toehold domain on the 3' side (downstream end) of the identity strand.*
- $SL(X)$*: the long domain on the state strand.*
- $IL(X)$*: the long domain on the identity strand.*

▶ **Definition 24.** *(Subset and logical operations for ACDC species). The following operations will be useful in the analysis of ACDC networks:*

- *Complementarity* $\diamond$ *:* $x \diamond y$ *is true for sequences* x, y *iff* $x = y^*$ *(and* $x^* = y$*).*
- *Complementarity with mismatch* $\square$ *:* $x \square y$ *is true for sequences* x, y *iff* $x = y^*$ *(and* $x^* = y$*) except for a single centrally-placed C-C or G-G mismatch.* $x \square y$ *is distinct from* $\neg x \diamond y$*, for which it is assumed that interactions between* x *and* y *are negligible.*
- $5'$ *(downstream end) state toehold sequence* $S5(X) := [SO5(X)\ SH5(X)]$.
- $3'$ *(upstream end) state toehold sequence* $S3(X) := [SH3(X)\ SO3(X)]$.
- $5'$ *(upstream end) identity toehold sequence* $I5(X) := [IO5(X)\ IH5(X)]$.
- $3'$ *(downstream end) identity toehold sequence* $I3(X) := [IH3(X)\ IO3(X)]$.

▶ **Definition 25.** *(Major species). A major species X must satisfy*

$$\neg\big(SO5(X)\diamond IO3(X)\big)\wedge\big(SH5(X)\diamond IH3(X)\big)\wedge \\ \big(SL(X)\diamond IL(X)\big)\wedge \\ \big(SH3(X)\diamond IH5(X)\big)\wedge\neg\big(SO3(X)\diamond IO5(X)\big).$$

▶ **Definition 26.** *(Domain complementarities in an ACDC reaction without mismatches). An ACDC reaction* $\mathrm{A}\rightarrow\mathrm{B}$ *or* $\mathrm{A}\dashv\mathrm{B}$ *implies*

$$\begin{aligned} &S5(A^{\mathrm{on}})\diamond S3(B^{\mathrm{off}})=S3(B^{\mathrm{on}}) &&\wedge\\ &IL(A^{\mathrm{on}})=IL(A^{\mathrm{off}})\diamond IL(B^{\mathrm{off}})=IL(B^{\mathrm{on}}) &&\wedge\\ &I3(A^{\mathrm{on}})=I3(A^{\mathrm{off}})\diamond I5(B^{\mathrm{off}})=I5(B^{\mathrm{on}}). \end{aligned}$$

Domains not constrained by these requirements are non-complementary. We emphasize that the domains of ancillary species involved in $\mathrm{A}\rightarrow\mathrm{B}$ *are determined unambiguously by the domains of the relevant major species.*

▶ **Definition 27.** *(Domain complementarities in ACDC reactions with mismatches). An ACDC reaction* $\mathrm{A}\rightarrow/\dashv\mathrm{B}$ *with mismatches placed as per Definition 20 implies*

$$\begin{aligned} &S5(A^{\mathrm{on}})\diamond S3(B^{\mathrm{off}})\ /\ S5(A^{\mathrm{on}})\square S3(B^{\mathrm{off}}) &&\wedge\\ &S5(A^{\mathrm{on}})\square S3(B^{\mathrm{on}})\ /\ S5(A^{\mathrm{on}})\diamond S3(B^{\mathrm{on}}) &&\wedge\\ &IL(A^{\mathrm{on}})=IL(A^{\mathrm{off}})\diamond IL(B^{\mathrm{off}})=IL(B^{\mathrm{on}}) &&\wedge\\ &I3(A^{\mathrm{on}})=I3(A^{\mathrm{off}})\square I5(B^{\mathrm{off}})=I5(B^{\mathrm{on}}). \end{aligned}$$

Domains not constrained by these requirements are non-complementary.

B Proofs of Theorems and Lemmas 8 - 22

▶ **Theorem 8** (Split motifs are realisable)**.** *Consider the N reactions* $\mathrm{A}\rightarrow\mathrm{B}_1,\mathrm{A}\rightarrow\mathrm{B}_2,\dots\mathrm{A}\rightarrow\mathrm{B}_N$*, in which all* B_i *are distinct from* A*. Such a network is realisable for any* $N\geq 1$*.*

Proof. By induction. Assume that the split motif is realisable for a given $N=M>0$. If so, a valid domain level implementation exists for $N=M$. Now consider the species $B^{\mathrm{off}}_{M+1}, B^{\mathrm{on}}_{M+1}, AB_{M+1}W_{AB_{M+1}\rightarrow B^{\mathrm{on}}_{M+1}}, F_{AB_{M+1}\rightarrow B^{\mathrm{on}}_{M+1}}$ *related* to a putative additional node B_{M+1}. Let these species be identical to those of B_1, except with the domains that function as the downstream end $(SO5, SH5, IO3, IH3)$ in $B^{\mathrm{off}}_{M+1}, B^{\mathrm{on}}_{M+1}$ changed to have no complementarity with any domains in the existing valid implementation for $N=M$. This assignment is possible by Assumption 5. Since the upstream domains $SO3, SH3, IO5, IH5$ of $B^{\mathrm{on}}_{M+1}, B^{\mathrm{off}}_{M+1}$ are identical to those of $B^{\mathrm{on}}_1, B^{\mathrm{off}}_1$, Definition 26 implies $\mathrm{A}\rightarrow\mathrm{B}$ as required by condition 1 of Definition 6. Moreover, since the species related to node B_{M+1} are identical to those of the existing node B_1, except for the downstream domains with no complementarity to the rest of the network, no new violations of conditions 2-4 of Definition 6 can occur due to interactions between the species related to B_{M+1} and A^{on} or those related to B_i for $1<i\leq M$. By considering the species defined in Figure 4 for B_1, and replacing domains h,i,j,k with $h_{M+1},i_{M+1},j_{M+1},k_{M+1}$ to define B_{M+1}, it is straightforward to establish that no violations of conditions 2-4 of Definition 6 occur between the species related to B_{M+1} and B_1. Therefore if a split motif of size $N=M$ is realisable, a split motif of size $N=M+1$ is realisable. Given the valid implementation for $N=1$ in Figure 4, split motifs of arbitrary $N>0$ are realisable. ◀

▶ **Theorem 9** (Integrate motifs are realisable). *Consider the N reactions $\mathrm{A}_1 \to \mathrm{B}, \mathrm{A}_2 \to \mathrm{B}, \ldots \mathrm{A}_N \to \mathrm{B}$, in which all A_i are distinct from* B. *This network is realisable for any $N \geq 1$.*

Proof. The proof is identical to that of Theorem 8 with the direction of catalysis interchanged. ◀

▶ **Lemma 10** (The ancillary species of a catalyst's upstream reactions and substrate's downstream reactions cause leak reactions). *Consider a reaction* $\mathrm{B} \to \mathrm{C}$, *and further assume that* $\mathrm{A} \to \mathrm{B}$ *and* $\mathrm{C} \to \mathrm{D}$ *for a species* A *and a species* D. *Then AB and CD, and $F_{AB \to B^{\mathrm{on}}}$ and $F_{CD \to D^{\mathrm{on}}}/W_{CD \to D^{\mathrm{on}}}$ possess two available toehold pairs that could form a contiguous complementary duplex. No other violations of realisability occur.*

Proof. Consider the following major species:

$$
\begin{aligned}
A^{\mathrm{on}} &:= [a \;\; b \;\; c \;\; d \;\; e\;] \\
&\quad\;\; [g^* \; b^* \; c^* \; d^* \; f^*] \\
B^{\mathrm{off}} &:= [h^* \; i^* \; c^* \; b^* \; a^*] \\
&\quad\;\; [j \;\; i \;\; c \;\; b \;\; g\;] \\
C^{\mathrm{off}} &:= [l \;\; m \;\; c \;\; i \;\; k\;] \\
&\quad\;\; [n^* \; m^* \; c^* \; i^* \; j^*] \\
D^{\mathrm{off}} &:= [p^* \; q^* \; c^* \; m^* \; o^*] \\
&\quad\;\; [r \;\; q \;\; c \;\; m \;\; n\;]
\end{aligned}
\qquad\qquad
\begin{aligned}
B^{\mathrm{on}} &:= [k^* \; i^* \; c^* \; b^* \; a^*] \\
&\quad\;\; [j \;\; i \;\; c \;\; b \;\; g\;] \\
C^{\mathrm{on}} &:= [o \;\; m \;\; c \;\; i \;\; k\;] \\
&\quad\;\; [n^* \; m^* \; c^* \; i^* \; j^*] \\
D^{\mathrm{on}} &:= [s^* \; q^* \; c^* \; m^* \; o^*] \\
&\quad\;\; [r \;\; q \;\; c \;\; m \;\; n\;]
\end{aligned}
$$

where the top (bottom) strand of each species is the state (identity) strand in 5'-3' (3'-5') direction. These species and the accordingly generated ancillary species implement the cascade $\mathrm{A} \to \mathrm{B} \to \mathrm{C} \to \mathrm{D}$. Conditions 1-3 of Definition 6 are satisfied.

To establish whether condition 4 of Definition 6 is necessarily violated, consider the unbound domains on the ancillary species in the system $\mathrm{A} \to \mathrm{B} \to \mathrm{C} \to \mathrm{D}$:

- $I5(A^{\mathrm{off}})$, $I3(B^{\mathrm{off}})$ in AB
- $S3(A^{\mathrm{on}})$, $S5(B^{\mathrm{on}})$ in $F_{AB \to B^{\mathrm{on}}}$
- $S3(A^{\mathrm{on}})$, $S5(B^{\mathrm{off}})$ in $W_{AB \to B^{\mathrm{on}}}$
- $I5(B^{\mathrm{off}})$, $I3(C^{\mathrm{off}})$ in BC
- $S3(B^{\mathrm{on}})$, $S5(C^{\mathrm{on}})$ in $F_{BC \to C^{\mathrm{on}}}$
- $S3(B^{\mathrm{on}})$, $S5(C^{\mathrm{off}})$ in $W_{BC \to C^{\mathrm{on}}}$
- $I5(C^{\mathrm{off}})$, $I3(D^{\mathrm{off}})$ in CD
- $S3(C^{\mathrm{on}})$, $S5(D^{\mathrm{on}})$ in $F_{CD \to D^{\mathrm{on}}}$
- $S3(C^{\mathrm{on}})$, $S5(D^{\mathrm{off}})$ in $W_{CD \to D^{\mathrm{on}}}$.

Definition 26 requires that $I3(B^{\mathrm{off}}) \diamond I5(C^{\mathrm{off}})$, $S5(B^{\mathrm{on}}) \diamond S3(C^{\mathrm{on}})$. These constraints are manifested in the example above as $[j, i]$ being present in the identity strand of $B^{\mathrm{off}}/B^{\mathrm{on}}$ and $[i^*, j^*]$ in the identity strand of $C^{\mathrm{off}}/C^{\mathrm{on}}$, and $[i, k]$ being present in the state strand of $C^{\mathrm{off}}/C^{\mathrm{on}}$ and $[k^*, i^*]$ in the state strand of B^{on}. Consequently AB and BC can bind by the two contiguous toehold domains $I3(B^{\mathrm{off}})$, $I5(C^{\mathrm{off}})$, and $F_{AB \to B^{\mathrm{on}}}$ can bind with $F_{CD \to D^{\mathrm{on}}}$ and $W_{CD \to D^{\mathrm{on}}}$ by the two contiguous toehold domains in $S5(B^{\mathrm{on}})$, $S3(C^{\mathrm{on}})$. No other violations of condition 4 occur in the proposed implementation. ◀

▶ **Theorem 12** (Long cascades are non-realisable due to a particular type of leak reaction only). *Consider the set of reactions $\mathrm{A}_1 \to \mathrm{A}_2$, $\mathrm{A}_2 \to \mathrm{A}_3 \ldots \mathrm{A}_{N-1} \to \mathrm{A}_N$ for $N > 3$, in which all A_i are distinct. This network would be realisable if reactions between ancillary species $A_i A_{i+1}$ and $A_{i+2} A_{i+3}$, and $F_{A_i A_{i+1} \to A^{\mathrm{on}}_{i+1}}$ and $F_{A_{i+2} A_{i+3} \to A^{\mathrm{on}}_{i+3}}/W_{A_{i+2} A_{i+3} \to A^{\mathrm{on}}_{i+3}}$, were absent.*

Proof. By induction. Assume that an implementation of a cascade of length $N > 3$ exists in which: (a) for any toehold domain x present in the downstream [upstream] end of A_M^{off} or A_M^{on}, $S5(A_M^{\text{off}}), S5(A_M^{\text{on}}), I3(A_M^{\text{off}}) = I3(A_M^{\text{on}})$ $\left[S3(A_M^{\text{off}}) = S3(A_M^{\text{on}}), I5(A_M^{\text{off}}) = I5(A_M^{\text{on}})\right]$, the presence of x and x^* in major species is restricted to the downstream [upstream] end of A_M^{off} and A_M^{on} and the upstream [downstream] end of A_{M+1}^{off} and A_{M+1}^{on} [A_{M-1}^{off} and A_{M-1}^{on}], $S3(A_{M+1}^{\text{off}}) = S3(A_{M+1}^{\text{on}}), I5(A_{M+1}^{\text{off}}) = I5(A_{M+1}^{\text{on}})$ $\left[S5(A_{M-1}^{\text{off}}), S5(A_{M-1}^{\text{on}}), I3(A_{M-1}^{\text{off}}) = I3(A_{M-1}^{\text{on}})\right]$; and (b) the only violations of realisability are those stated in this theorem. Lemma 10 gives an implementation for $N = 4$ satisfying these conditions.

Let us consider adding a new layer A_{N+1} to the cascade. The toeholds $S3(A_{N+1}^{\text{off}}) = S3(A_{N+1}^{\text{on}}), I5(A_{N+1}^{\text{off}}) = I5(A_{N+1}^{\text{on}})$ are complements of $S5(A_N^{\text{on}}), I3(A_N^{\text{on}})$, respectively, and the toeholds $S5(A_{N+1}^{\text{off}}), S5(A_{N+1}^{\text{on}}), I3(A_{N+1}^{\text{off}}) = I3(A_{N+1}^{\text{on}})$ can be orthogonal to all other toeholds by Assumption 5. This choice preserves assumption (a) above for the $N+1$-layer cascade. Definition 26 indicates that when the implied ancillary species are included, $\mathrm{A}_N \to \mathrm{A}_{N+1}$ as required by condition 1 of Definition 6. Moreover, the only toeholds in the new species are either non-complementary to the rest of the network, or taken from the upstream and downstream ends of A_N^{off} and A_N^{on}. By (a), these toeholds are only present in major species of nodes A_{N-1}, A_N and A_{N+1} and the ancillary species associated with them. To identify violations of conditions 2-4 of Definition 6, it is therefore sufficient to consider the isolated 4-level cascade $\mathrm{A}_{N-2} \to \mathrm{A}_{N-1} \to \mathrm{A}_N \to \mathrm{A}_{N+1}$ only. This analysis proceeds exactly as in Lemma 10; the proposed $N+1$-layer cascade therefore preserves assumption (b) as well as (a). Given that a domain-level implementation satisfying assumption (a) and (b) is given in Lemma 10 for $N = 4$, we therefore conclude that an implementation satisfying (a) and (b) can be constructed for arbitrary $N > 3$. Consequently there are no restrictions on realisability of cascades for $N > 3$ other than those stated in the theorem. ◀

▶ **Theorem 14** (Self interactions and bidirectional edges are not realisable). *Consider a system of reactions* $\mathrm{A}_1 \to \mathrm{A}_2 \to \mathrm{A}_3 \ldots \mathrm{A}_{N-1} \to \mathrm{A}_1$. *This network is not realisable if* $N \leq 2$.

Proof. The result for $N = 1$ is a direct consequence of Theorem 13. For $N = 2$, consider the set of reactions: $\mathrm{A} \to \mathrm{B}$, $\mathrm{B} \to \mathrm{A}$. By Definition 26, $\mathrm{A} \to \mathrm{B}$ implies $I3(A^{\text{off}}) \diamond I5(B^{\text{off}})$ and $IL(A^{\text{off}}) \diamond IL(B^{\text{off}})$. In addition, $\mathrm{B} \to \mathrm{A}$ implies $I5(A^{\text{off}}) \diamond I3(B^{\text{off}})$. The identity strands of A and B are then fully complementary, violating condition 3 of Definition 6. ◀

▶ **Theorem 16** (Long feedback loops with an even number of units are non-realisable due to a particular type of leak reaction only). *Consider the feedback loop* $\mathrm{A}_1 \to \mathrm{A}_2, \mathrm{A}_2 \to \mathrm{A}_3, \ldots \mathrm{A}_{N-1} \to \mathrm{A}_N, \mathrm{A}_N \to \mathrm{A}_1$. *For* N *even,* $N \geq 4$, *this network would be realisable if reactions between ancillary species* A_iA_{i+1} *and* $A_{i+2}A_{i+3}$, *and* $F_{A_iA_{i+1} \to A_{i+1}^{\text{on}}}$ *and* $F_{A_{i+2}A_{i+3} \to A_{i+3}^{\text{on}}}/W_{A_{i+2}A_{i+3} \to A_{i+3}^{\text{on}}}$, *were absent. Here, the index* j *in* A_j *should be interpreted modularly:* $\mathrm{A}_j = \mathrm{A}_{j-N}$ *for* $j > N$.

Proof. For N even, $N \geq 4$, a loop obeying condition 1 of Definition 6 can be constructed from the cascades identified in the proof of Theorem 12 by setting the otherwise unconstrained toeholds $S5(A_N^{\text{on}}), I3(A_N^{\text{on}})$ to $S5(A_N^{\text{on}}) \diamond S3(A_1^{\text{off}}) = S3(A_1^{\text{on}}), I3(A_N^{\text{on}}) \diamond I5(A_1^{\text{off}}) = I5(A_1^{\text{off}})$. To identify the violations of realisability that arise from conditions 2-4 of Definition 6, let us first consider a cascade without the $\mathrm{A}_N \to \mathrm{A}_1$ reaction. The only violations of realisability are those identified in Theorem 12: between A_iA_{i+1} and $A_{i+2}A_{i+3}$, and $F_{A_iA_{i+1} \to A_{i+1}^{\text{on}}}$ and $F_{A_{i+2}A_{i+3} \to A_{i+3}^{\text{on}}}/W_{A_{i+2}A_{i+3} \to A_{i+3}^{\text{on}}}$, without interpreting the index modularly. Now we consider the additional effect of requiring $\mathrm{A}_N \to \mathrm{A}_1$. The only domains that must be changed are $S5(A_N^{\text{on}})$ and $I3(A_N^{\text{on}})$. These domains and their complements are only present in the species of $\mathrm{A}_{N-1} \to \mathrm{A}_N$, $\mathrm{A}_N \to \mathrm{A}_1$, $\mathrm{A}_1 \to \mathrm{A}_2$, and so it is sufficient to consider only this cascade to identify additional violations of realisability. By Lemma 10, the resultant violations of realisability are exactly those stated in the theorem. ◀

▶ **Theorem 22** (Mismatches successfully destabilize unintended complexes)**.** *The scheme proposed in Definition 20 satisfies the following:*

1. *All motifs that are realisable in the mismatch-free ACDC design remain realisable in the mismatch-based scheme.*
2. *Cascades of arbitrary length N with at most the first and last reactions deactivating are realisable;*
3. *Feedback loops with N even and $N \geq 6$ in which all reactions are activating are realisable;*
4. *Feedforward loops with $N \geq 1$, $M \geq 1$, $N - M$ even, in which at most the first and last reactions are deactivating in each branch, are realisable.*

Proof. Consider the first claim. For any network in which it is possible to select domains that satisfy Definition 25 and Definition 26, it is trivial to convert those domains to satisfy 25 and 27 by introducing the mismatches in major species, and adjusting ancillary species compensate. By Assumption 19, these changes do not introduce new violations of realisability.

Now consider the second claim. By the first claim and the construction in Theorem 12, it is sufficient to consider whether the sequestration reactions identified in Lemma 10 for an $N = 4$ cascade occur in the mismatch-based scheme of Definition 20. First, consider the unbound domains in the ancillary species in the system $A \to / \dashv B \to C \to / \dashv D$, with mismatches placed as per Definition 20:

- $I5(A^{\text{off}})$, $I3(B^{\text{off}})$ in AB
- $S3(A^{\text{on}})$, $S5(B^{\text{on}})$ in $F_{AB\to B^{\text{on}}}/W_{AB\to B^{\text{off}}}$
- $S3(A^{\text{on}})$, $S5(B^{\text{off}})$ in $W_{AB\to B^{\text{on}}}/F_{AB\to B^{\text{off}}}$
- $I5(B^{\text{off}})$, $I3(C^{\text{off}})$ in BC
- $S3(B^{\text{on}})$, $S5(C^{\text{on}})$ in $F_{BC\to C^{\text{on}}}$
- $S3(B^{\text{on}})$, $S5(C^{\text{off}})$ in $W_{BC\to C^{\text{on}}}$
- $I5(C^{\text{off}})$, $I3(D^{\text{off}})$ in CD
- $S3(C^{\text{on}})$, $S5(D^{\text{on}})$ in $F_{CD\to D^{\text{on}}}/W_{CD\to D^{\text{off}}}$
- $S3(C^{\text{on}})$, $S5(D^{\text{off}})$ in $W_{CD\to D^{\text{on}}}/F_{CD\to D^{\text{off}}}$.

By Definition 27, the reaction B $\to$ C implies $I3(B^{\text{off}})\Box I5(C^{\text{off}})$, $S5(B^{\text{on}})\Box S3(C^{\text{on}})$. Moreover, $\neg S5(B^{\text{off}}) \diamond S3(C^{\text{on}})$. By Assumption 19, none of the violations of realisability that would otherwise occur due to binding of AB and CD; $F_{AB\to B^{\text{on}}}\ W_{AB\to B^{\text{off}}}$ and $F_{CD\to D^{\text{on}}}\ W_{CD\to D^{\text{off}}}$; and $F_{AB\to B^{\text{on}}}\ W_{AB\to B^{\text{off}}}$ and $W_{CD\to D^{\text{on}}}\ F_{CD\to D^{\text{off}}}$ characterised by Lemma 10, occur. Note that if B $\dashv$ C in the above network, Definition 27 implies $S5(B^{\text{on}}) \diamond S3(C^{\text{on}})$, meaning that sequestration reactions still occur between ancillary species. Cascades with deactivation reactions as intermediate steps are therefore not realisable.

Now consider the third claim. By the construction in Theorem 16 and the first claim of this Theorem, it is sufficient to consider only the sequestration reactions listed in Theorem 16. Further, since the only difference between a feedback loop with exclusively activating interactions and an activating cascade with N species is that $\mathrm{A}_N \to \mathrm{A}_1$, by the second claim of this Theorem we need only consider changes in realisability due to the introduction of $\mathrm{A}_N \to \mathrm{A}_1$ to a cascade. For $N \geq 6$, imposing $I3(A_N^{\text{off}})\Box I5(A_1^{\text{off}})$, $S5(A_N^{\text{on}})\Box S(3)A_1^{\text{on}}$, as required by $\mathrm{A}_N \to \mathrm{A}_1$, does not create new realisability violations for a cascade of length N with exclusively activating reactions. The ancillary species of the reactions $\mathrm{A}_{N-2} \to \mathrm{A}_{N-1}, \mathrm{A}_{N-1} \to \mathrm{A}_N, \mathrm{A}_N \to \mathrm{A}_1, \mathrm{A}_1 \to \mathrm{A}_2, \mathrm{A}_2 \to \mathrm{A}_3$ can only form complexes held together by two contiguous toehold domains with a central mismatch, and thus do not violate realisability by Assumption 19.

The above argument does not apply to FBLs of length $N = 4$, which remain non-realisable. In that case, adding the reaction $A_N \rightarrow A_1$ allows complexes of ancillary species bound by two separate sets of contiguous toehold domains, each with a central mismatch, either side of a 4-way junction. The short periodicity of an $N = 4$ loop means that the unwanted interaction identified in Lemma 10 happens twice for each pair of ancillary species. We do not assume in Assumption 19 that such a structure will dissociate. We also note that feedback loops with any deactivating reactions remain non-realisable, since each reaction $A_i \rightarrow A_{i+1}$ is effectively an intermediate reaction between $A_{i-1} \rightarrow A_i$ and $A_{i+1} \rightarrow A_{i+2}$.

Finally we turn to the fourth claim. By the first claim of this Theorem, and Theorem 17, it is sufficient to consider only the potential unwanted sequestration reactions between ancillary species identified in Theorem 17 for each feed-forward branch. The proof is then identical to that of the second claim of this Theorem. ◀