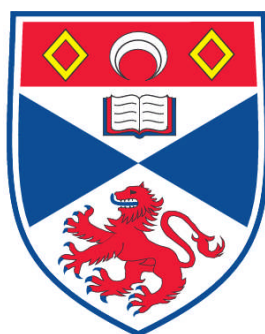# EVOLVING COMPLEX SYSTEMS FROM SIMPLE MOLECULES

Jan Sadownik

**A Thesis Submitted for the Degree of PhD
at the
University of St. Andrews**
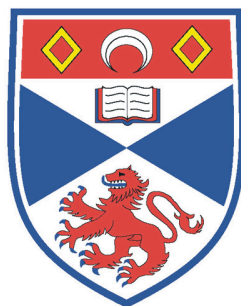
2009

Evolving Complex Systems from Simple Molecules

Jan Sadownik

A Thesis Presented for the Degree of

Doctor of Philosophy

in the

School of Chemistry

University of St Andrews

St Andrews

December 2008

I, Jan Sadownik, hereby certify that this thesis has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

date ……….…   signature of candidate ……………………

I was admitted as a research student in January 2005 and as a candidate for the degree of Doctor of Philosophy in January 2005; the higher study for which this is a record was carried out in the University of St Andrews between 2005 and 2008.

date …………..   signature of candidate ……………………

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Doctor of Philosophy in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

date …………...   signature of supervisor …………………...

In submitting this thesis to the University of St Andrews I wish access to it to be subject to the following conditions: for a period of 1 year from the date of submission, the thesis shall be withheld from use.
I understand, however, that the title and abstract of the thesis will be published during this period of restricted access; and that after the expiry of this period the thesis will be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright in the work not being affected thereby, and a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permission that may be required in order to allow such access and migration.

date ……………   signature of candidate ………………………

Synopsis

Until very recently, synthetic chemistry has focussed on the creation of chemical entities with desirable properties through the programmed application of isolated chemical reactions, either individually or in a cascade that afford a target compound selectively. By contrast, biological systems operate using a plethora of complex interconnected signaling and metabolic networks with multiple checkpoint controls and feedback loops allowing biological systems to adapt and respond rapidly to external stimuli. Systems chemistry attempts to capture the complexity and emergent phenomena prevalent in the life sciences within a wholly synthetic chemical framework. In this approach, complex phenomena are expressed by a group of synthetic chemical entities designed to interact and react with many partners within the ensemble in programmed ways. In this manner, it should be possible to create synthetic chemical systems whose properties are not simply the linear sum of the attributes of the individual components.

Chapter 1 discusses the role of complex networks in various aspects of chemistry-related research from the origin of life to nanotechnology. Further, it introduces the concept of Systems chemistry, giving various examples of dynamic covalent networks, self-replicating systems and molecular logic gates, showing the applications of complex system research.

Chapter 2 discusses the components of replicator design. Further, it introduces a network based on recognition mediated reactions that is implemented by length-segregation of the substrates and displays properties of self-sorting.

Chapter 3 presents a fully addressable chemical system based on auto- and cross-catalytic properties of product templates. The system is described by Boolean logic operations with different template inputs giving different template outputs.

Chapter 4 introduces a dynamic network which fate is determined by a single recognition event. The replicator is capable of exploiting and dominating the exchanging pool of reagents in order to amplify its own formation at the expense of other species through the non-linear kinetics inherent in minimal replication.

Chapter 5 focuses on the development of complex dynamic systems from structurally simple molecules. The new approach allows creating multicomponent networks with many reaction pathways operating simultaneously from readily available substrates.

Acknowledgements

I would like to thank my supervisor Prof Douglas Philp for all the help and guidance he has given me over the last four years.

I would like to thank all the members of the Philp group that I had the pleasure to work with. Firstly, a big thanks goes to Annick Vidonne as we started on the same day and together we shall finish. Not forgetting Dr Simon Turega, Dr Evan Wood, Dr Vicente del Amo, Peter Donnelly, Craig Robertson, Jürgen Huck, Izzaty Hassan and all the project students.

Special thanks goes to the technical staff at the University of St Andrews. Thanks to Melanja Smith and Tomas Lebl for NMR spectroscopy assistance. Caroline Horsburgh for mass spectrometry, Prof Alexandra Slawin for X-ray crystallography and Sylvia Williamson for elemental analysis.

I would like to thank my parents for never doubting in me and helping me through these four years away.

Finally I would like to thank my wife Marta for supporting and helping me at home, and my son Emil who arrived just six months ago and turned out to be an amazing person and a never ending source of inspiration (as well as sleepless nights). To him I dedicate this work.

List of abbreviations

3-D	three-dimensional
Ac	acyl group
Boc	tertbutoxycarbonyl group
CA	carbonic anhydrase
d	doublet
DCC	dynamic covalent chemistry
DCL	dynamic combinatorial library
DCM	dichloromethane
dd	doublet of doublets
DMF	dimethylformamide
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNP	1,5-dioxynaphthalene
EDC	1-ethyl-3-(3'-dimethylaminopropyl)carbodiimide
EDTA	ethylene diamine tetra-acetic acid
EF	enhancement factor
Et	ethyl group
g	gram(s)
h	hour(s)
HMTA	hexamethtlenetetraamine
HNA	hexitol nucleic acid
HOBT	1-hydroxybenzotrialzol
HOMO	highest occupied molecular orbital
HPLC	high performance liquid chromatography
K_a	association constant
K_d	dissociation constant
kJ	kilojoule(s)
LOMO	lowest occupied molecular orbital
LUMO	lowest unoccupied molecular orbital
M	molar
m	multiplet

m.p.	melting point
MALDI-TOF	matrix-assisted laser desorption/ionisation-time of flight mass spectrometry
Me	methyl group
MHz	megahertz
min	minute(s)
ml	millilitre(s)
mM	millimolar
mm	millimetre(s)
mmol	millimole(s)
NAD^+	nicotinamide adenine dinucleotide
NASA	National Aeronautics and Space Administration
nm	nanometre(s)
NMR	nuclear magnetic resonance
o	ortho
p	para
PET	photoinduced electron transfer
Ph	phenyl group
PNA	peptide nucleic acid
ppm	part(s) per million
p-RNA	pyranosyl ribonucleic acid
PTSA	*p*-toluenesulphonic acid
r.t.	room temperature
RNA	ribonucleic acids
s	second(s) / singlet
t	time / triplet
TFA	trifluoroacetic acid
THF	tetrahydrofuran
TLC	thin layer chromatography
TMS	tetramethylsilane
TNA	threofuranosyl nucleic acid
TTF	tetrathiafulvalene
UV-Vis	ultraviolet-visible

Table of contents

1. Introduction

1.1 Preamble

The boundaries between the sciences are fading. Modern chemistry can only progress in the company of biology, physics, mathematics and engineering. Research is becoming interdisciplinary, feeding from many sources, in order to achieve new goals. With more and more knowledge available at hand the trend is to look at the broader picture, outside the constraints of one discipline. Such approach can be applied at many levels: looking at chemistry in biology, looking at reactions in reaction networks, looking at compounds in intractable mixtures.

Systems chemistry is an emerging area that attempts to incorporate these principles in chemical research. Complex networks are often associated with sociological and biological phenomena, but with the rise of modern analytical techniques they start to find their way into experimental chemistry. The ideas of chemical networks are, however, as old as life itself, being the focus of many prebiotic theories. Most researchers agree the living form emerged from a simpler chemical system, but how this transfer occurred is still a subject to an intense ongoing debate. Some answers could be found if life was to be synthesised *a priori* in a laboratory. This feat has not yet been accomplished, but with the development of nanotechnology and molecular biology one might expect very interesting results in the near future. The one thing for sure is that a complex chemical system will lie at the heart of such assembly.

Systems chemistry incorporates many aspects of modern chemical research that could be relevant to the development of the ideas above. Bridging the gap between dynamic combinatorial chemistry, self-replication and molecular logic could help find answers to questions ranging from the origin of life to nanofabrication of molecular computers. In order to achieve these goals, it has to be realised that complex system research is possible even with very simple molecules and the results could be highly rewarding.

1.2 Complex networks

The study of complex networks[1-3] is relevant to practically all fields of science, from neurobiology to statistical physics. A network is an interconnected group or system. It consists of a series of nodes, which represent single entities, such as people,

computers, molecules or proteins. The connections between the nodes could represent acquaintance, sharing of information or influencing behaviour. The properties of nodes and edges define the structure of the network (Figure 1). Nodes can be of the same or different nature, some nodes can have more edges than others and depending on what properties define a node some have a greater weight than other members of the network. The connections can also be of different types and weights with some nodes more strongly connected than other ones. A specific type of connection is a directed edge, when one node can influence another but is not influenced by it in return. Graphs composed of directed edges are themselves called directed graphs, which can be either cyclic, meaning they contain closed loops of edges, or acyclic, meaning they do not.
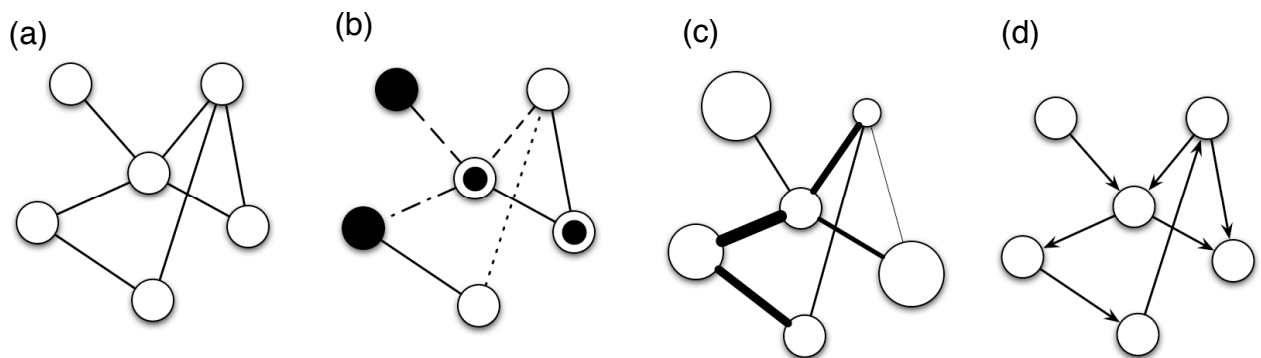
Figure 1. Examples of various types of network architectures. (a) An undirected network with only a single type of node and a single type of edge. (b) A network with a number of discrete node and edge types. (c) A network with varying node and edge weights. (d) A directed network in which each edge has a direction.

A significant problem with networks is data presentation. Even the most complex physical structure can be presented with modern three-dimensional imaging techniques as for example crystal structures of proteins. The most common way to depict networks is a wiring diagram, which at a higher complexity turns into an intricate tangle (Figure 2). Unfortunately, the higher the complexity of the system the more interesting are its properties. The subtle long-distance connections across the entire system are especially difficult to depict. The problem being that a change to any single part of the network will have an effect on the whole system.
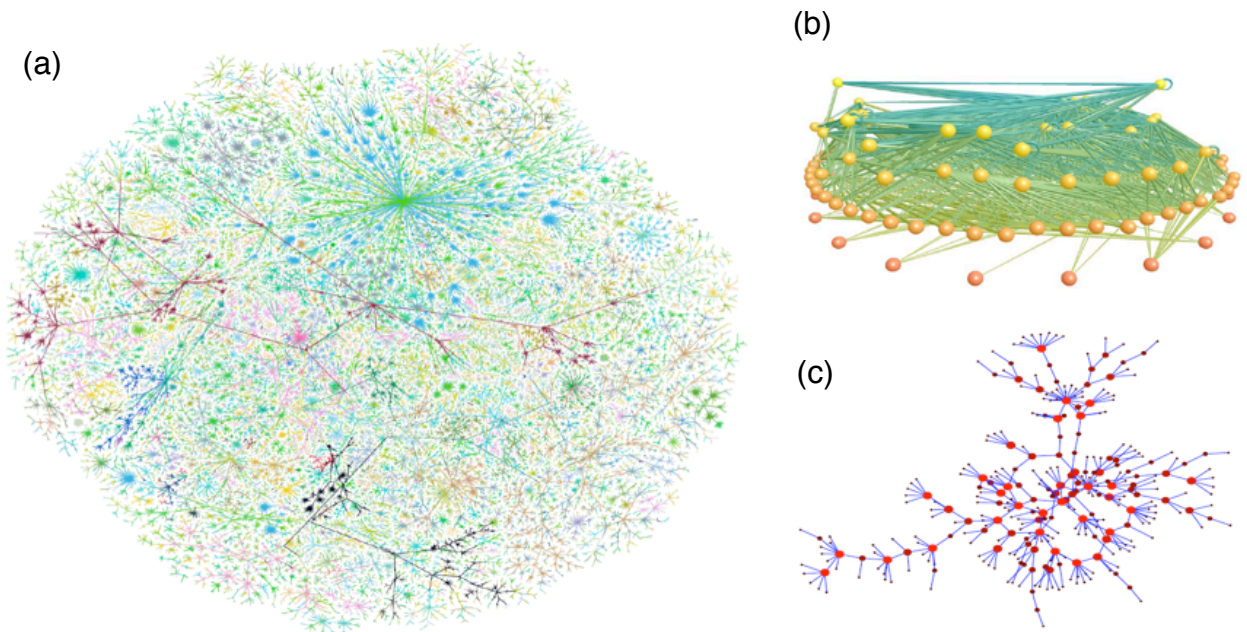
Figure 2. Examples of studied networks. (a) A visualisation of the network structure of the Internet. (b) A food web of predator-prey interactions between species in a freshwater lake. (c) A social network of sexual contacts. Figure taken from ref. 2.

Some other features of networks that make them inherently difficult to understand include network evolution, which makes the diagram change over time. Again the problem is too many interdependent variables that would have to be plotted against time.

The most basic feature[4] of any network is its architecture. A large number of nodes represented on a two-dimensional plane can be connected in a number of ways. The two extreme scenarios[5] are either a regular network when neighbouring nodes are connected in an orderly fashion or a totally random set of connections. However, like with any assumptions the real life networks lie somewhere in between. The social or biological networks are never fully ordered or fully random. There are several models that try to define the transition state between random and ordered. One of them is a small-world[6] approach, in which not all nodes are neighbours, but most can be reached from every other by a small number of hops or steps (Figure 3).
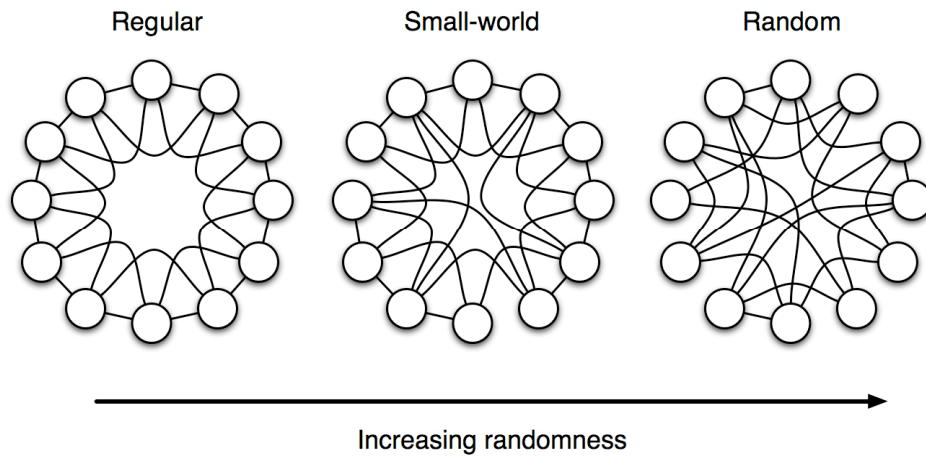
Figure 3. Schematic representations of the regular, small-world and random networks.

This feature reflects the famous experiment[7] by Milgram in the 1960s, which became the basis for the six-degrees of separation theory, currently widely proven by members of internet social networks *e.g.* Facebook. In theory, every person in a social network is connected to another person by maximum of six other people (often referred to as six hand-shakes). This property exists to some extent as the result of clustering which can be found in many networks. If node A is connected to node B and node B to node C, then there is a heightened probability that A will also be connected to C. In the language of social networks, the friend of your friend is also likely to be your friend. This situation leads to some nodes having many more connections than others, which is usually a feature of naturally occurring networks and is defined as a scale-free[8] property (Figure 4a). In such a system, the number of connections per node ranges from a very large number for a small number of nodes and relatively few for the rest. Such system displays the small-world properties as well. These characteristics lead to the formation of a community structure[9] in networks (Figure 4b), in which network nodes are joined together in tightly knit groups, between which there are only looser connections. The presence of communities in networks is a signature of the hierarchical nature of complex systems[10]. Grouping nodes into communities is also one of the ways to simplify the representation of the network and many ways to detect communities in a network are being studied[11-12].

Figure 4. Complex network models. (a) A scale-free network with highly connected nodes (black). (b) A network with community structure of highly interlinked modules connected to each other by a few links.
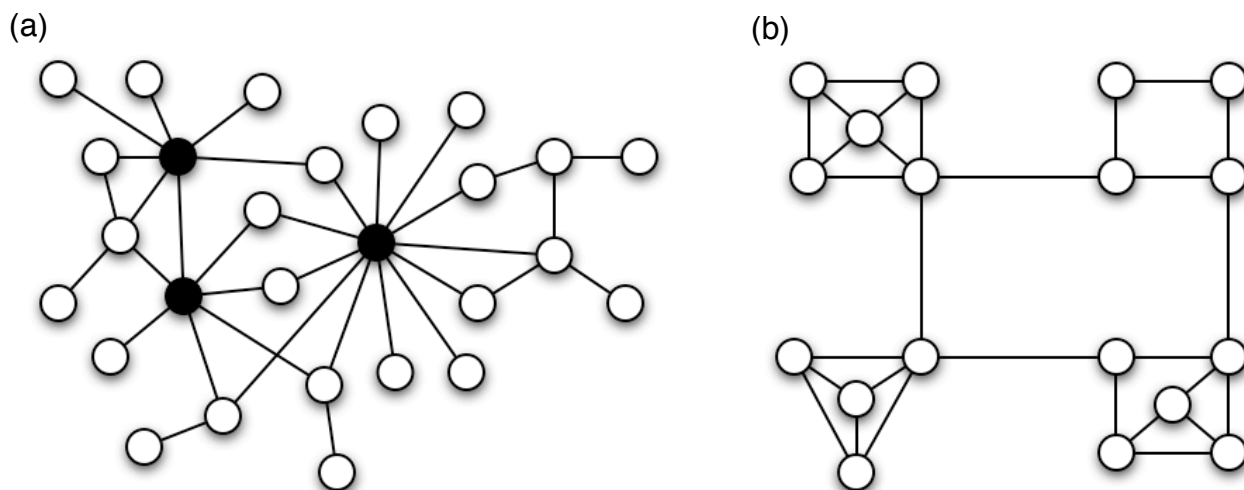
A network of interacting proteins in living systems is sometimes called an 'interactome'. The interactome possesses a scale-free topology with some proteins interacting with a large number of partners. It has been shown[13] that by a random deletion of proteins in a model yeast system, the removal of the 'hub' proteins is much more likely to lead to cell death than a removal of a peripheral protein.

1.3 The origin of life

Life is one of the most complex networks studied by man. The drive to understand even the basic aspects of life is still a great challenge and comprehension of it as a whole will most likely prove impossible. The emergence of life on Earth is a question that has intrigued and challenged humans since the beginnings of science. Understanding the innate complexity of living systems poses a series of important questions, which will only be answered by a concerted effort involving chemistry, biophysics, molecular biology, genetics, evolutionary research, geology, atmospheric science and astronomy. All of these disciplines work at a different length scales, ranging from single molecules to whole planet systems. Chemists are privileged to be working at the very bottom of this ladder, at the most basic level that is closest to life's origin. Studying simple molecules that are capable of organising themselves into more complex systems may bring us closer to understanding how non-living matter transformed[14-18] into living organisms (Figure 5). Since life is based on organic compounds, the first issue is how did organic compounds appear on the early Earth.

However, this problem is not entirely relevant to a modern chemist. A major task is to investigate how first cells emerged from these organic molecules and to define the minimal requirements for a self-sustained chemical system capable of undergoing Darwinian evolution, which is the definition of life proposed[19] by the Exobiology Program within NASA. This definition was used[20] originally by Horowitz and Miller and is considered as a general working definition by Luisi in his discussion[21] on the various definitions of life.

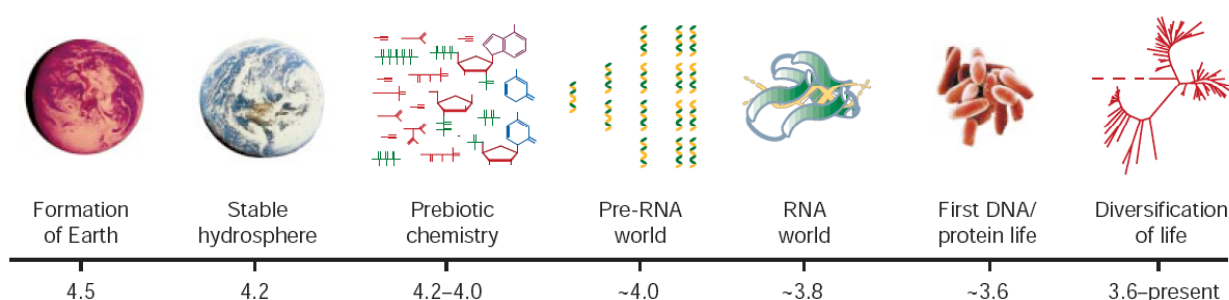| Formation of Earth	Stable hydrosphere	Prebiotic chemistry	Pre-RNA world	RNA world	First DNA/ protein life	Diversification of life
4.5	4.2	4.2–4.0	~4.0	~3.8	~3.6	3.6–present

Figure 5. Timeline of events pertaining to the early Earth, with approximate dates in billions of years before the present. Figure taken from ref. 47.

The search for the origin of life as a scientific research field, supported by solid empirical experiment, dates back to the beginning of the 20th century with a small tract published[22] by Oparin. The progress in the field is driven by firstly understanding the general principles of contemporary biological processes and then using these observations to provide increasingly satisfactory explanations of life's origin. Chemical research in this area, however, has to be broadened, as we should not only define life in its cellular form and how it came to be, but following Eschenmoser and Kisakürek, we should consider[23] that "the aim of an experimental aetiological chemistry is not primarily to delineate the pathway along which our (natural) life on Earth could have originated, but to provide decisive experimental evidence, through the realisation of model systems that life can arise as a result of the organisation of the organic matter". This statement brings us towards the idea of artificial, engineered life. Of course, the first attempts focus on creation of life in its minimal form. Two different approaches exist. The first approach[24-25] strips down an existing organism, leaving behind only the necessary components needed to fulfil the definition of life. This approach, however, still leaves hundreds of genes and thousands of different proteins and other molecules inside a cellular compartment making it impossible to

6

monitor all of the basic life processes at molecular level. A second approach[26-27] is to gather the minimal amount of components needed for life inside a preassembled compartment. This approach is more appealing to chemists as it relates to chemical synthesis of complex, functional mixtures.

1.4 The RNA World

Before we can discuss the attempts to synthesise life, we have to identify the target for this challenge. The target brings us back to the scenario of the early life on Earth. One of the most popular theories is the so-called RNA World[28-29.] The idea was coined after the discovery that some RNA molecules, called ribozymes, possess[30-32] catalytic properties and, thus, could have been[33] the basis of life before DNA. For the discovery[34] of the enzymatic cleavage of RNA by RNA, Cech and Altman were awarded the Nobel Prize in 1989. Ribozymes[35] lack distinct functional groups, but exhibit a strong potential to bind metal ions. It is postulated that in many cases the presence of the specifically positioned metal ion is responsible for their catalytic activity.

The hammerhead[36] ribozyme (Figure 6), at ~40 nucleotides, is the smallest of the naturally occurring ribozymes, and mediates replication within circular virus-like RNAs that infect plants. Recent experiments show that the hammerhead motif is the most efficient self-cleaving sequence that can be isolated from randomised pools of RNA, suggesting that it may have arisen multiple times during the evolution of functional RNA molecules[37]. Consisting of three short helices connected at a conserved sequence junction, the hammerhead catalyses site-specific cleavage of one of its own phosphodiester bonds *via* nucleophilic attack of the adjacent 28-oxygen at the scissile phosphate. Divalent metal ions have been identified in the structure of the hammerhead, but their role in the catalysis is still unclear. The same phosphodiester transfer reaction is also catalysed by the hepatitis delta virus ribozyme as well as the hairpin ribozyme, but ribozymes obtained through *in vitro* selection techniques were found to catalyse a number of other reactions. These ribozymes were identified to form[38] a nucleotide from a base plus a sugar, synthesise[39-40] amide bonds, form[41] Michael adducts such as those involved in the methylation of uridine monophosphate to give thymidine monophosphate, and form[42] acyl-coenzyme A, which is found in many protein enzymes.
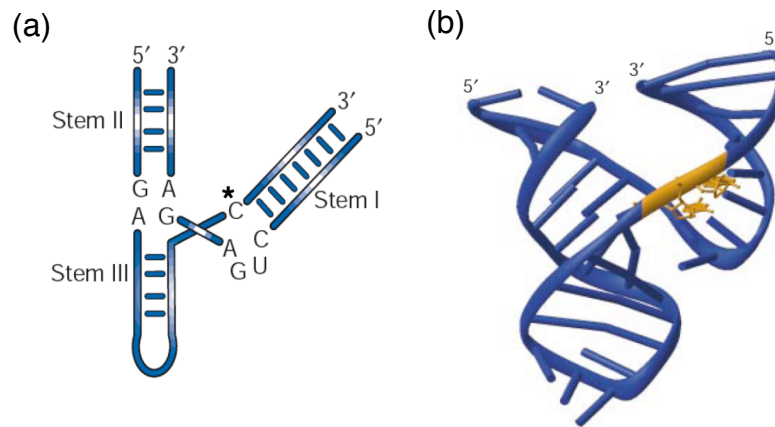
Figure 6. (a) Secondary structure of the hammerhead ribozyme. Nucleotides important for catalytic activity are indicated; the cleavage site is indicated by a star. (b) Crystal structure of the hammerhead ribozyme. Figure adapted from ref. 30.

In the RNA World, RNA served as both the coding material and an enzyme-like catalyst, probably assisted later on by metal ions, amino acids and other small molecule cofactors. This scenario[43-44] allows for replication without protein-based enzymes. In the RNA World the first stage of evolution occurred when RNA molecules, by performing their catalytic activities, assembled themselves from a nucleotide soup. Next, they evolved in a self-replicating fashion and by recombination and mutation gained new catalytic functions. The hypothesis[45] states that as cellular metabolism became more sophisticated there was an increasing demand for biocatalysts providing the drive for the transition to protein-based enzymes. Protein-based enzymes were templated by RNA and carried the same catalytic properties as RNA only performed more efficiently, which led to the overtaking of the ribozyme function by more complex proteins. Like all theories about the origin of life, this one also has its drawbacks[46]. There is good evidence to suggest an RNA-based life in the early stages of the development of Earth, but it is hard to believe that there were no evolutionary processes, which operated before the emergence of RNA. The problem of a spontaneous synthesis of RNA on early Earth is the same that every chemist struggles with in the laboratory everyday. It is the problem of mixtures and how to obtain the one single desired product out of them. Each of the components of RNA would have been surrounded by several closely related analogues, which could assemble in almost any combination. The question asked[47] is how RNA emerged from the 'primordial soup' and how it was possible to devise a self-replicating mechanism that would operate independently from all other byproducts[48] present in the early Earth medium.

Pentose sugars such as ribose are a necessary component of nucleic acids, but it is still unclear how ribose could have been specifically selected from the mixture of sugars produced by the formose reaction, especially that ribose constitutes for less than 1% of the reaction products. For example, the prebiotic synthesis of sugars from formaldehyde can be biased by starting from glycoaldehyde phosphate, leading[49] to ribose 2,4-diphosphate as the predominant pentose sugar. Eschenmoser conducted[50-52] an extensive study on RNA alternatives (Figure 7). The target was to synthesise structures based on Watson-Crick base pairs[53], but with different sugar backbones and study their properties compared to naturally occurring RNA. It became apparent that base pairing is not unique just for RNA and is not specific for the double helix type of conformation. Also, RNA does not display the strongest strand association, so it cannot be the feature that made ribose backboned polymers the basis of modern life. For example pyranosyl-RNA displays higher base-pairing strength than natural furanosyl-RNA. The research disqualified hexose-nucleic acids as potential natural alternatives for pentose-based structures as the bulk of the monomer unit disables the base-pairing.

| PNA | p-RNA | TNA | HNA |
| peptide nucleic acid | pyranosyl-RNA | threofuranosyl nucleic acid | hexitol nucleic acid |

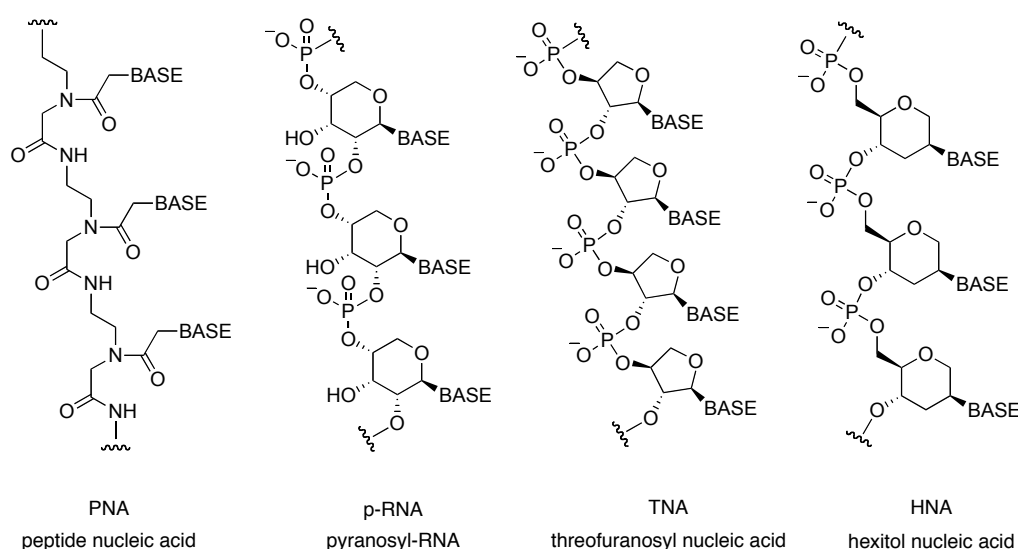**Figure 7.** Examples of nucleic acid analogues, which form stable base-pair duplexes.

Since the publication[54] by Oro purines and pyrimidines are usually considered[55] products of hydrogen cyanide condensation (Figure 8), but an entirely plausible prebiotic route to polynucleotides has not yet been proposed. The main problem seems to be that ribonucleotides just seem too complex to envisage a possible, direct

way how such molecules could have emerged form a 'primordial soup'. All of the subunits would have to accumulate in sufficient amounts in the same time and space. For example, prebiotic synthesis of purines and pyrimidines are clearly incompatible[56] with those proposed for the synthesis of ribose. Prebiotic synthesis of nucleic acids is improbable[57] when considering a spontaneous assembly of all the necessary components from a mixture of random clutter without any influence of catalysts.
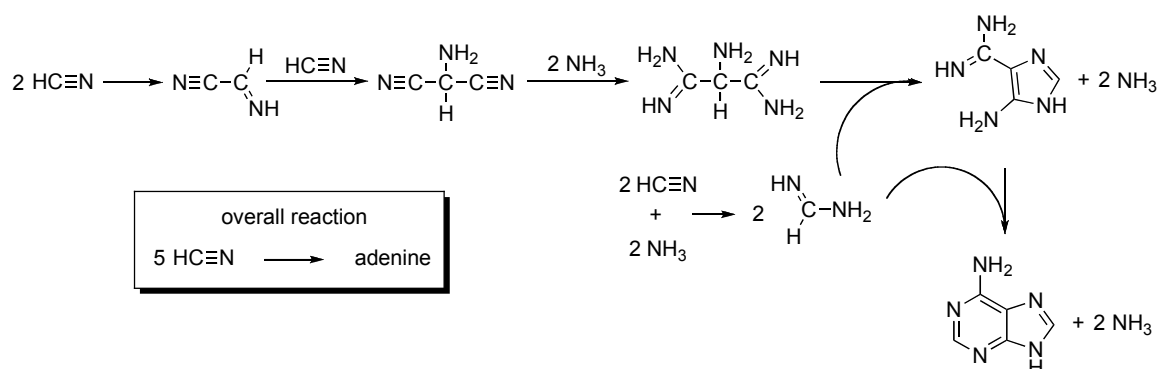

Figure 8. Mechanism of formation of adenine from hydrogen cyanide.

One of the popular theories considers the involvement[58] of natural clays as catalysts for the polymerisation reaction and some theories consider[59] clays as pre-RNA inorganic genetic material. Ferris conducted[60-61] an extensive survey on polymerisation reactions of activated RNA monomers. Using montmorillonite clay it was possible to synthesise 40-mer nucleotides. The size is comparable to the hammerhead ribozyme and so the oligomer is long enough to carry catalytic properties. The chemistry involved an activated RNA monomer that was allowed to adsorb onto the clay and after one to three days a whole array of polynucleotides was obtained. In addition, the linking regiochemistry is predominantly the correct 5' to 3' linkage and not the incorrect 5' to 2' linkage, which characterises the major product in the absence of any template.

This synthetic route is very attractive partly because of the ready availability of the catalyst, however, the clay used for the experiments was significantly different from the native mineral found in Nature. The purification of montmorillonite for laboratory use greatly enhances its catalytic properties making the polymerisation process much more efficient than it would have been in the early Earth environment. The second issue is the availability of activated monomers in the prebiotic scenario, as nucleotides do not condense in aqueous solution without being converted to a form

containing a high-energy bond.

In the polynucleotides responsible for the structure of nucleic acids all ribose is in the D-ribose configuration. This asymmetry indicates that at some point chirally pure nucleic acid chains had to be synthesised. Selecting ribose over other sugars is statistically very improbable and selecting only one of its enantiomers is even less plausible. Orgel and coworkers gave[62] an insight into a possible mechanism of such 'symmetry breaking'. Orgel studied template-directed synthesis of poly(G) by poly(C). When poly(D-C) was used as a template for enantiomerically pure D-guanosine 5'-phospho-2-methylimidazole (D-2-MeImpG) the reaction outcome were 3',5'-linked oligo(G)s ranging from dimers up to 30-mers. Far less efficient process was observed when the same template was used to direct the formation from a racemic mixture of monomers. No product greater than an octamer was observed. First of all the (L-2-MeImpG) was incorporated less readily than the D-form and secondly once it was incorporated into the chain it acted as an inhibitor for further elongation. The cross-inhibition results when the incorporation of the D-monomer forces the base into a *syn* conformation changing the orientation of the sugar-phosphate component in relation to the template disabling the addition of the next monomer. This behaviour could explain the favourable synthesis of only one of the enantiomers of the nucleic acids.

Even though life's origin dates back billions of years, the ongoing debate about the issues of the transformation of organic matter into life is still raging supported by modern experimental chemistry.

1.5 Simple molecules and the metabolism-first scenario

The ideas concerning the origin of life can be divided into two major trends. The vision of RNA with its self-replicating properties as being central to the emergence of life falls into the 'replication first'[63] category. There is, however, an alternative approach. It is the concept that small molecule interactions were[64] central to the origin of life. Simple organic molecules are more likely to have created a primitive metabolism hence this idea is termed 'metabolism-first'[65] scenario. Dyson using a purely hypothetical model suggested[66] that complexity is the real key to the origin of life. Given sufficient complexity, Dyson argues that the system must 'crystallise' into a set of mutually catalytic molecules that possess the property of 'closure'. Closure indicates the formation of a cycle, rather than an infinite linear process.

The sequence of chemical reactions where any product is the same as a substrate of some preceding step is often termed[67] a hypercycle. The compounds involved in a hypercycle are usually connected by cross catalysis and some or all of them display autocatalytic behaviour. The cycle as a whole can also be viewed as being catalytic since all of the components are utilised and subsequently returned to the system with every turn of the cycle. The hypercycle can be found in the basic functions of life such as the citric acid cycle or the replication of single-stranded RNA. Hypercycles are a principle of natural self-organisation allowing an integration and coherent evolution of a set of functionally coupled self-replicating entities. Hypercyclic organisation[68] is a minimal requirement to guarantee the evolution of a translational apparatus. Primitive entities capable of replication would be most likely ineffective and error-prone. The simplest way to enhance their performance is to assemble the self-replicators into a hypercycle (Figure 9). The replicators would not only catalyse their own formation but also the next species in the cycle.
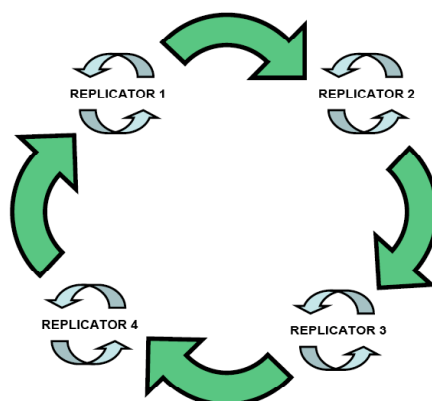

Figure 9. Schematic representation of a hypercycle of replicators.

Through such network no replicator could exclude another. The aid given by one entity to the next would ultimately be fed back to itself through the cycle. This catalytic relationship would lead to the promotion of the best replicators and evolution. The encapsulation of the hypercycle into compartments would lead to the development of cells and competition between them.

This idea was extended[69] by Shapiro who claimed that for a mixture of simple organic compounds to evolve in the direction of life it would be necessary for the network to develop processes of self-organisation. The hierarchy in the mixture would be established with accumulation of certain components of the mixture. This

enhancement of concentration is established at the expense of other members of the mixture or by synthesis from external substrates. Since the products of given reactions in the cycle are also catalysts for other pathways, a rise in abundance of one of the members in the mixture can influence another part of the cycle. Certainly to drive the catalytic cycles of the system an external source of free energy is necessary to counter the overall negative entropy of the process. Although all the reactions in the metabolism scheme are reversible, it is possible for an irreversible 'driver' reaction that is coupled to an energy source transmitting energy into the cycle. Even though in this theory metabolism developed before the emergence of molecules capable of storing genetic information, it is still possible for the mixture to carry heredity information that would be represented by the identity and concentration of its components. All these properties can be accumulated in a term 'compositional genome'[70]. The system, because of its hypercyclic topology (Figure 10) is able to adapt to the changes in the environment. If by some natural 'disaster' one of the paths is blocked, the system is capable of finding a way around the blockage. Evolution of the system proceeds through such circumstances, where new pathways are grafted[71] onto the system and new catalytic interactions can form as a response to the outside stimuli. Grafting refers to a mechanism of evolution in which two previously independent autocatalytic systems symbiotically interact in a manner that increases the growth rate of both and leads to their mutual interdependence.
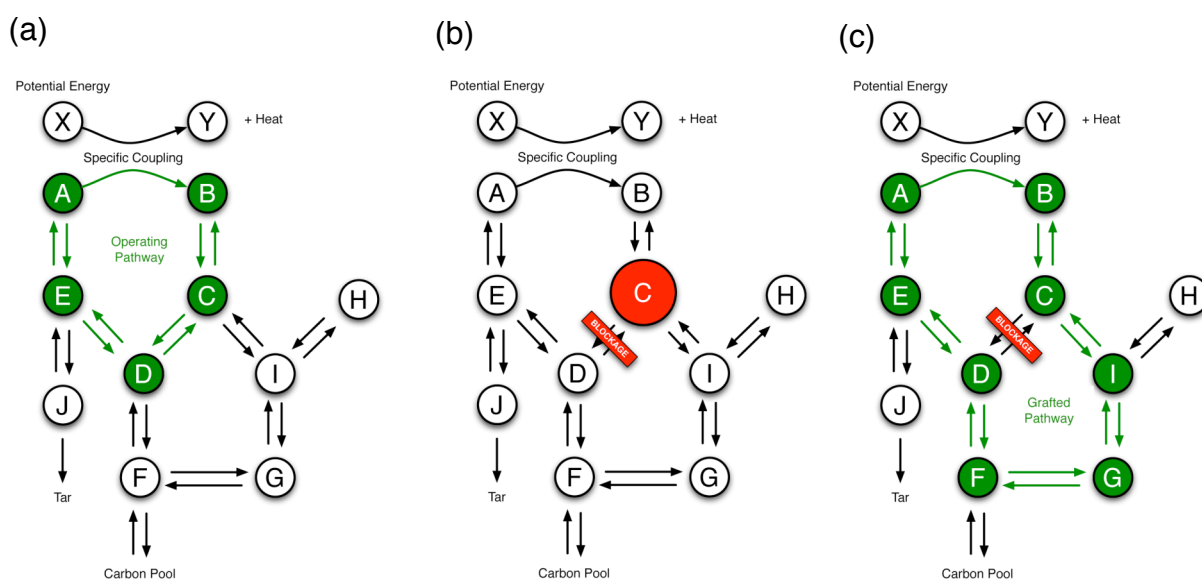

Figure 10. (a) A hypothetical system of simple organic molecules with a hypercyclic operating pathway. (b) A blockage of the C-D connection by a 'natural disaster' results in an accumulation of C. (c) The accumulation of C induces a new grafted pathway in order to bypass the blockage.

Through such behaviour, evolution could have started in a world containing no instructional and catalytic polymers and subsequently lead to the formation of biomolecules that are the basis of life as we know today. Of course, we probably will never be able to comprehend and understand the entire network of interconnected metabolic pathways that formed the prebiotic system, although the research into complex systems is becoming an important part of many scientific disciplines.

1.6 Attempts in synthesising life

Cellular life[72] cannot be described in terms of only RNA or any other information-carrying polymer nor as metabolism or as compartment alone. Cellular life emerges from the coupling among these three components. A container is necessary in order to at least confine reactions in a limited space, where interactions are more likely to occur[73]. It also provides a spatial location able to effectively facilitate division of function between the cellular processes. Moreover, in modern cells, the membrane is an active cell component[74], channelling nutrients in and waste products out of the cell by means of specialised transport catalysts. A metabolism provides[75] the source of non-equilibrium and a means of energy storage required in order to build and maintain cellular components. It is also required to allow cell growth[76] to occur and eventually replication into two different but similar copies.

The main component of a basic synthetic protocell[77-78] would be an RNA molecule that can act both as a template for the storage and transmission of genetic information, and as an RNA polymerase that can replicate its own sequence. This process has to take place in a compartment, because by keeping molecules that are closely related together and disabling free mixing, advantageous mutations can lead to preferential replication. This difference would lead to the formation of compartments containing more efficient mutants than the original molecules giving them advantage needed for evolution. The research[79-81] towards synthesising an RNA strand that would meet such requirements is developing quickly, however, it is a goal still to be reached.

Modern cells base their compartmentalisation on lipid membranes and it would only be sensible to try to incorporate such assemblies into artificial cells. It has been proven such vesicles can self-assemble, which is a big advantage in synthetic protocols. The vesicle component of a protocell must possess rather unusual

properties, including spontaneous growth and division. The specific membrane structures proposed for the purpose of life synthesis are micelles, which are bilayer vesicles[82]. Such micelles have been shown[83-85] to be able to self-replicate once their growth becomes thermodynamically unstable when they reach a certain size.

Szostak and coworkers proved[86] experimentally that vesicles could grow and replicate (Figure 11). Vesicles formed by fatty acids in water undergo[87] spontaneous assembly and growth. Szostak used myristoleate **1** vesicles adding dilute fatty acid micelles to preformed vesicles slowly, over a 4 hour period. The reason behind using high dilution was that vesicle growth could be enhanced and the formation of new vesicles reduced by eliminating the transiently high micelle concentration present immediately after micelle addition to vesicles. Examination of an initial population of myristoleate vesicles, prepared by extrusion through 100 nm pore filters, revealed an average diameter of approximately 90 nm. Slow addition of one equivalent of myristoleate micelles resulted in a steady increase in average vesicle size, up to a final mean diameter of 130 nm, consistent with quantitative incorporation of the new myristoleate into the preformed vesicles and leading to a doubling of surface area. The division of the vesicles was induced by extrusion through small-pore polycarbonate membrane filters. 100 nm myristoleate vesicles, encapsulating 20 mM calcein dye, were grown to 140 nm and then extruded through the pores to a final mean size of 88 nm. The loss of the dye from inside of the vesicle was only slightly greater than that demanded by the geometric constraints due to reduction in size. This experiment proved vesicles could divide when subjected to mechanical stress, with retention of the encapsulated materials.
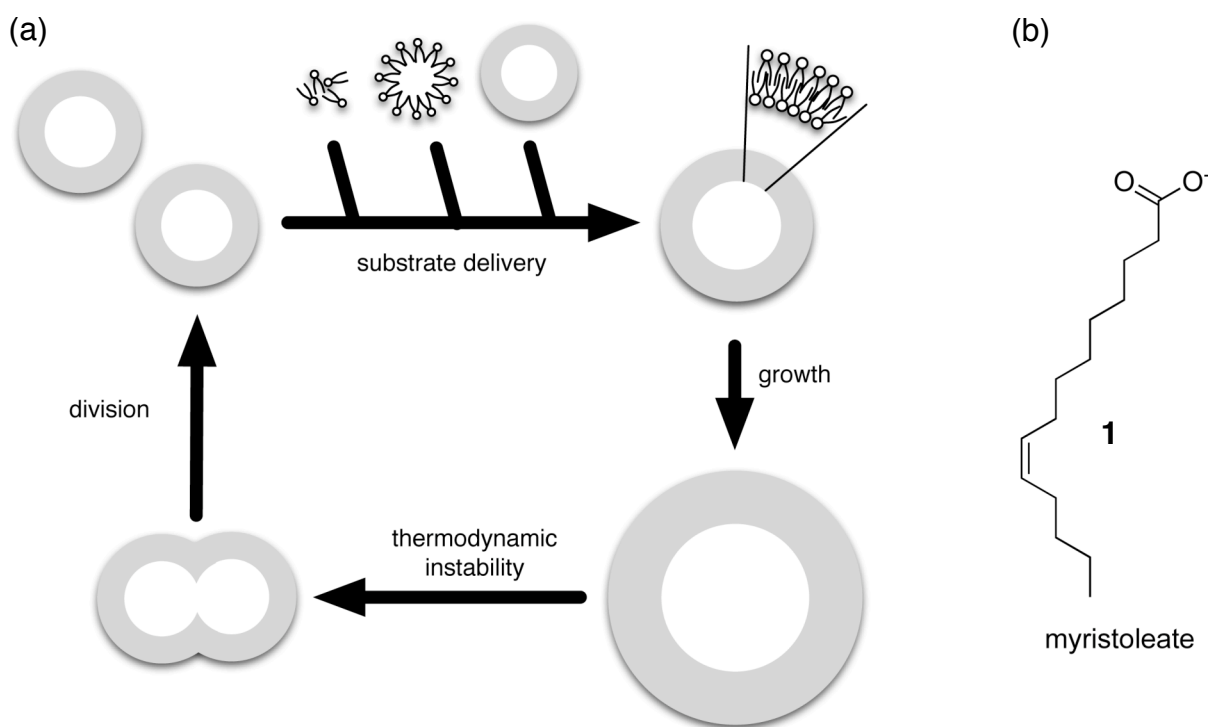
Figure 11. (a) Modes of vesicle growth and division. Self-replicating membrane vesicles can grow when supplied with appropriate substrates and may divide either spontaneously or under the influence of external forces. (b) Structure of fatty acid cation, myristoleate **1** used by Szostak to create self-replicating vesicles.

The membrane used for creating an artificial cell should also allow for small molecules to pass freely, while keeping the macromolecules encapsulated allowing for a primitive feed. A further requirement for such system to be termed living is the necessity for both replicating cycles, the ribozyme cycle and the membrane cycle, to operate independently. Once this characteristic is achieved, the protocells can start competing[88] with each other and evolving in a way that neither of its components is capable on its own. This feature is one of the characteristics of complex interacting systems that are often targets of modern synthetic chemistry.

1.7 Nanotechnology

From the examples above we can clearly see that the search for the origin of life leads us to investigate how organic matter is able to organise itself and through that organisation develop new properties. This idea is also the theme[89] for nanotechnology, whose development is often related[90] to the origin of life research. According[91] to Drexler the goal of nanotechnology is a self-replicating, programmable architecture capable of atom-by-atom manipulation of both biological and

non-biological matter. Since Nature was able to create life by such an approach, nanotechnology is believed to be a very powerful strategy for the creation of machines and devices. The idea stirs up both enthusiasm and fear, since nanotechnology is expected, for good and bad, to strongly influence[92] the future of mankind.

As in the search for the minimal life, there are two different approaches to miniaturisation: the top-down and the bottom-up approach. The top-down approach is the current technology of the electronics industry, which is able to manipulate increasingly smaller pieces of matter by lithography and related techniques. It is now apparent that the top-down approach is subject to limitations[93-94] for dimensions smaller than 50 nm. This gap is still very large on the scale of atoms and molecules. Therefore, quoting[95] Feynman, "there is still plenty of room at the bottom" for further miniaturisation. The answer to closing the gap is the bottom-up approach, analogous to creating a protocell from the basic components necessary for life, the idea is to use atoms and molecules to build up to nanostructures. The ideal paradigm for the bottom-up approach is self-assembly. In order for structures to self-assemble, shape, proportions and properties of the final, desired structure are encoded in the molecules used. Nature offers[96] countless examples of functional self-assembly that stimulate scientists to design artificial, non-living systems.

The term nanotechnology is used to describe[97] phenomena in many areas of chemistry. Organic chemists contribute mostly in the area of molecular devices[89-101]. This term describes organising organic molecules in complex functional manner for the synthesis[102-103] of molecular motors and machines. A major goal of this research is to harness molecular motion, which use is widely spread in Nature. The targets are mechanically interlocked molecules like catenanes[104] and rotaxanes[105] (Figure 12), which allow for utilising the translational or rotational motion of their components.
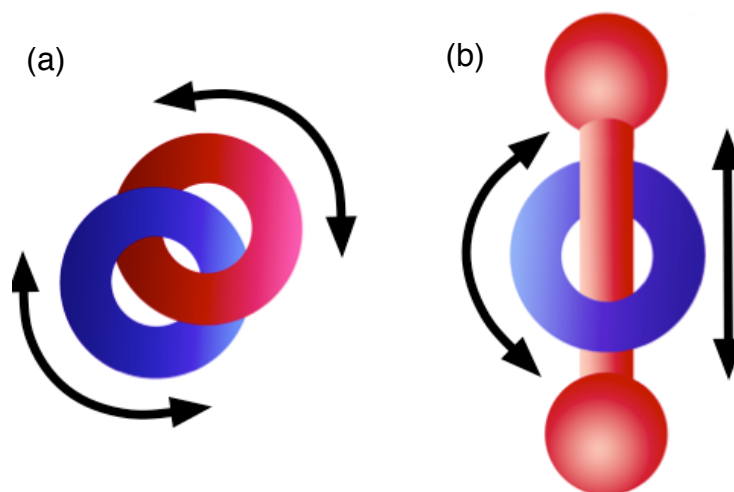
Figure 12. Schematic representation of (a) [2]catenane and (b) [2]rotaxane. Arrows show possible large-amplitude modes of movement for one component relative to another.

The rotaxane has proven to have the biggest potential when it came to displaying function. The nanodevices based on rotaxane topology include molecular shuttles[106-108], elevators[109] and muscles[110-111].

The ability to manipulate single molecules by modern techniques is outstanding. The tools to do so are again borrowed from Nature. The assembly is guided in most cases either by donor-acceptor interactions between π-systems, hydrogen bonds or metal-ligand interactions.

Stoddart's pioneering work on the assembly[112-113] of mechanically interlocked molecules such as rotaxanes and catenanes was based on bipyridinium herbicide paraquat as the electron acceptor. Two bipyridinium motifs were incorporated into a macrocycle in order to create a π-electron deficient host. Guests that can be bound into the rigid rectangular cavity include catechol and hydroquinone ethers[114], 1,5-dioxynaphthalene derivatives[115], aromatic aminoacids[116] and similar π-electron rich structures. This methodology led Stoddart and coworkers to develop[117-119] an astounding array of molecular machines and supramolecular structures and devices such as switches. A rotaxane by definition consists of a macrocycle threaded onto a dumbbell shaped linear molecule. The bulky stoppers at the ends of the thread prevent the macrocycle from dissociating from it creating a mechanical bond. Even though in order to disassemble the rotaxane it is necessary to break a covalent bond it is possible for the macrocycle to shift freely along the thread. It is possible to design the thread in such a way that there would be two recognition sites for the macrocycle making it possible that in a given set of conditions the ring will sit either on one

station or the other. In this particular case (Figure 13) the two sites are a benzidine and a biphenol ring systems. The solution of the rotaxane **2a** in acetonitrile when analysed with ^1H NMR spectroscopy shows the macrocycle is associated preferentially with the benzidine unit in an 84:16 ratio. This selectivity is due to a higher affinity of the tetracationic cyclophane to the benzidine. The addition of CF_3COOD to the solution protonates the benzidine, which results in a repulsion of the positively charged macrocycle from the now positively charged benzidine. The analysis of the acidified solution reveals an exclusive population of rotaxanes with the biphenol group inside the macrocycle cavity. The addition of pyridine to the solution deprotonates the benzidine and reverts the system to the original equilibrium.
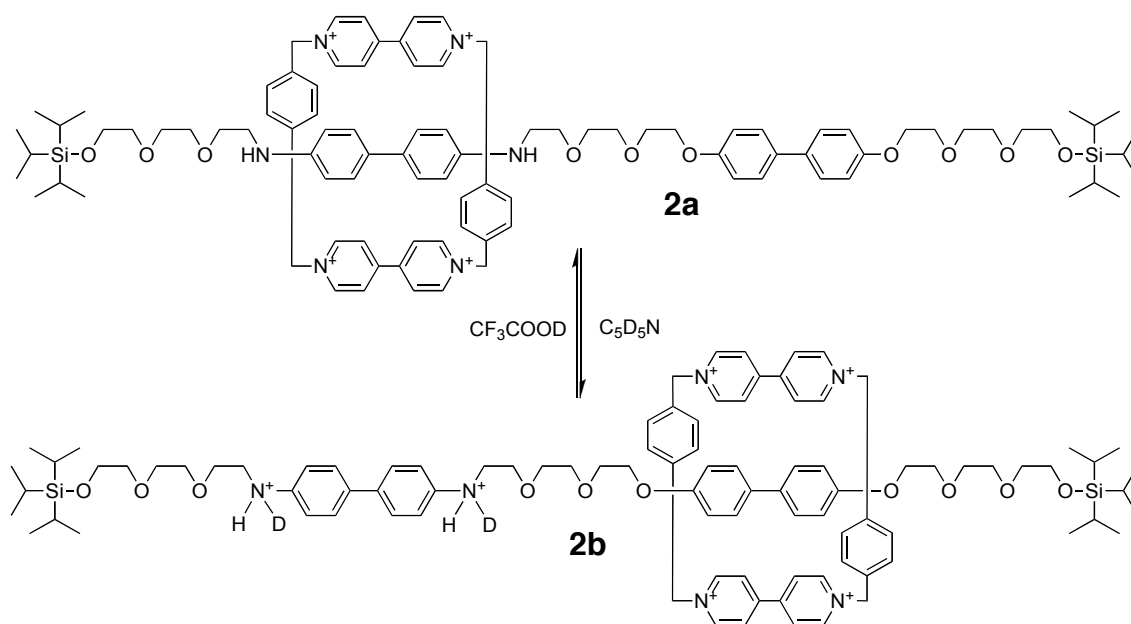

Figure 13. The shuttling of the macrocycle along the dumbbell-shaped thread is controlled by protonating\deprotonating the benzidine unit.

Leigh synthesised[120] one of the first catenanes **5**, self-assembled from commercially available substrates **3** and **4** (Figure 14). The condensation of eight molecules into an interlocked structure was driven purely by hydrogen bond recognition giving a 20% yield in one step.
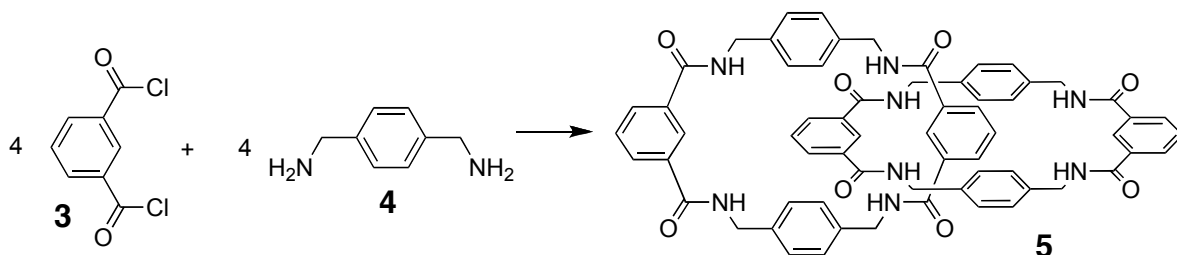

Figure 14. One step, eight-molecule condensation to give the [2]catenane **5**.

Catenanes can be used as devices when it is possible to control their rotational motion. For example, the behaviour of a [2]catenane **6** (Figure 15) with a similar structure to the one described above can be altered with change of solvent. In halogenated solvents such as CDCl$_3$, the two macrocycles interact through hydrogen bonds. Each of the identical rings adopts a different conformation, one effectively acting as the host and the other as the guest. In a hydrogen-bond-disrupting solvent such as [D$_6$]DMSO, the preferred conformation has the amides exposed on the surface where they can interact with the solvent, while the hydrophobic alkyl chains are shielded inside the molecule to avoid the polar solvent. Such system can be considered as a solvent sensitive switch.
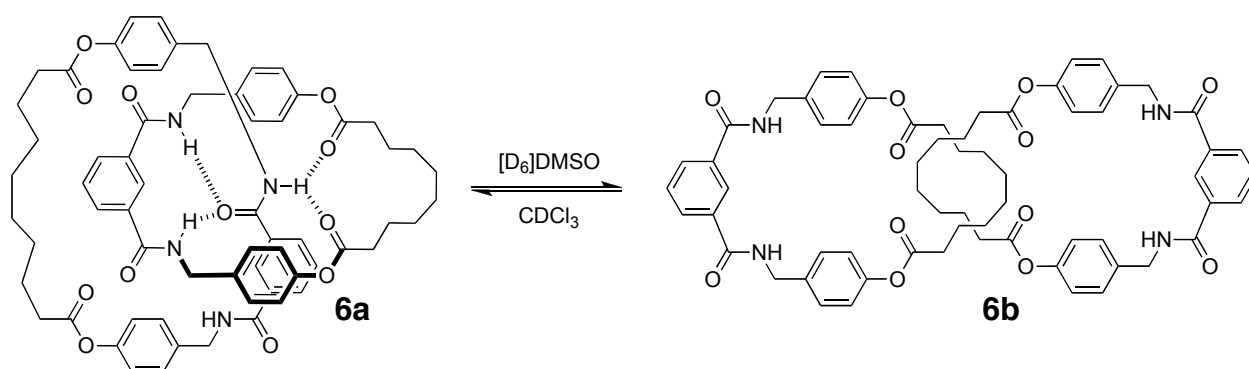

Figure 15. Translational isomerism in an amphiphilic benzylic amide [2]catenane **6**.

Some of the most outstanding examples of self-assembled structures come from the laboratory of Fujita. Using multicomponent metal-ligand self-assembly it was possible[121,122] to create discreet 3-D hollow structures such as cages, cones, capsules and boxes. The edges of the structures are designed rigid organic compounds while the corners are metal atoms with specific coordination angles that govern the overall geometry. The cages have various shapes with either triangular or square sides. The structures obtained with this methodology are of exceptional size with cavities large enough to act as molecular flasks to encapsulate guest molecules and regulate or promote specific reactions (Figure 16).
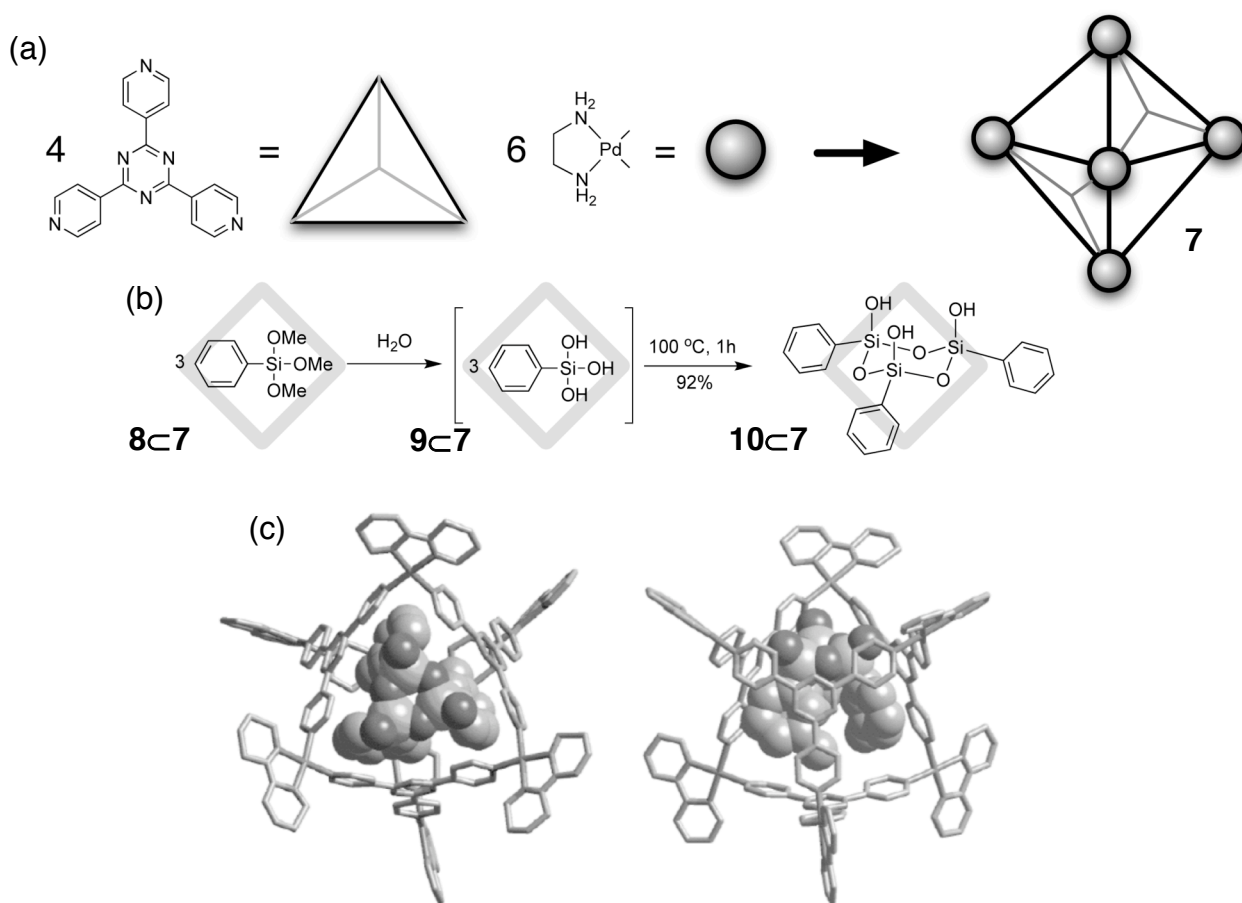
Figure 16. A self-assembled coordination cage **7**. (a) Synthesis scheme. (b) Synthesis of a siloxane trimer **10** inside the coordination cage **7**. (c) Crystal structure of the siloxane trimer **10** trapped inside the cage **7**. Figure adapted from ref. 125.

A coordination cage provides an isolated nanospace for molecules, where otherwise labile species can be considerably stabilised in the protection of the cage. The products are trapped most effectively and stabilised in a cage if they are prepared *in situ* from smaller components coming through the openings of the cage. The authors succeeded[123-124] in preparing a labile siloxane trimer **10** as a stable form in the self-assembled cage cavity **7**. When the condensation reaction of phenyltrialoxysilane **9** was carried out in the presence of the cage **7**, cyclic trimer **10** was formed within the cavity in 92% isolated yield. In this reaction, three molecules per cage were hydrolysed and subsequently condensed to form a trimer.

The substrates **8** for the reaction are small enough to enter into the cage but the product **10** however, being larger, cannot escape the cage. Fujita compares this synthesis to assembling a ship inside a bottle. Other reactions performed inside this type of cages include[125] photochemical oxidation of alkanes and [2+2] intermolecular photodimerisation of olefins.

These examples show how, from a mixture of simple compounds under the right conditions, very complex structures can spontaneously emerge. Over the last 20 years, creating more and more complex structures has been the target of supramolecular chemistry. The idea of the 'chemistry beyond the molecule'[126] focuses on the interactions that involve non-covalent bonds and recognition and try to encode the properties into the assembly, again trying to mimic Nature. The plethora of examples shows how chemists embrace multi-component complex systems. This idea led to combining these efforts under a common umbrella, leading to the emergence of systems chemistry.

1.8 Systems chemistry

The fields of sociology and biology have given rise[127] to the greatest weight of publications concerning complex networks. Chemistry is often omitted in the field of complexity and chemists themselves have a history of considering complex mixtures as an unwanted necessity of their research. Synthetic chemists focused mostly on devising single transformations that produce a desired compound with given properties selectively. In such approaches mixtures of compounds are undesired and need to be eliminated with various purification techniques. This situation is no longer the case, since an insight[128] into complex networks has been made possible by the development of advanced analytical methods. The definition of complexity given[129] by Whitesides is that such a system is characterised by a high sensitivity to initial conditions and any small changes in the environment. It also must consist of a large number of independent interacting components and multiple pathways through which these components might evolve.

The term 'systems chemistry' was first used[130] in 2005 by von Kiedrowski in his paper describing the kinetic and computational analysis of a nearly exponential organic replicator. In this paper, von Kiedrowski claims new research "could open a door to a field that may be termed systems chemistry, namely, the design of prespecified dynamic behaviour". In the same year, a systems chemistry workshop was held[131] in Italy where a more detailed definition was given describing systems chemistry as "a new field of chemistry seen as the offspring of prebiotic and supramolecular chemistry on the one hand and theoretical biology and complex systems research on the other". Moreover, systems chemistry deals with complex dynamic phenomena

and incorporating them into ordered systems with emergent nonlinear properties. The property of the system results from the interactions of its components and is not just a simple sum of the properties of the mixture members in isolation. It should also be possible to derive from the information collected from the system the individual properties of the components simultaneously, providing one can find a way to deconvolute such data.

Systems chemistry rather than a totally new area of chemistry is a title under which a few already quite well established branches of chemistry can cooperate. Otto[132] focuses mostly on the role of dynamic combinatorial chemistry (DCC) as being central to developing complex systems, while von Kiedrowski[130] strongly emphasises the role of self-replicating autocatalytic cycles, both of which will be discussed further. Other areas of development include molecular logic, oscillating reactions, self-assembly and self-sorting processes. It seems that the mixture of different approaches to systems chemistry is as complex as the systems it sets out to describe. In my opinion, systems chemistry should focus on the emergent properties of systems of distinctly interconnected species. The systems should be designed in such a way as to direct the behaviour of the system with simple chemical inputs.

1.9 Dynamic combinatorial chemistry

Dynamic combinatorial chemistry[133-135] is defined as combinatorial chemistry under thermodynamic control. The objects of investigation in DCC are dynamic combinatorial libraries (DCLs), which are currently considered a starting point for the development of systems chemistry. A DCL is a mixture of compounds that continuously interchange their subunits, resulting in a distribution of products, which is thermodynamically controlled. A change of stability of one member will influence other members, so the final equilibrium distribution is derived from the sum of the stabilities of all species in a library. This property means that DCLs are addressable, making them an interesting target for complex system research.

The first application that stood behind the development of DCC was drug discovery[136]. This area has emerged from combinatorial chemistry[137-140] in which a large static population of different compounds was synthesised using divergent methods. The aim was to produce a library with as high structural diversity as possible and subsequently screen the properties of the library members for

bioactivity or specific catalysis. Generation by synthetic methods of such large libraries has its limitations, so it was only logical to find a new approach to introduce diversity and complexity in a mixture.

Lehn discusses[141] the transition from static libraries to dynamic ones introducing a term virtual libraries. According to him, since all combinations of interchanging compounds are possible even if some are not present in the pool, it makes the assembly virtual. The term however did not come into general use and such systems are commonly described as dynamic combinatorial libraries.

The first necessary condition to create a DCL is to find an appropriate reversible connection process (Figure 17). The bond formed and broken can be both of covalent and non-covalent nature. Dynamic covalent chemistry[142] is a term often used to describe the first case. The reaction used in a DCL has to meet a number of requirements. Firstly, it has to be reversible on a reasonable time scale in the given experimental conditions. Also it cannot interfere with other processes that the library is subjected to. An interesting property that could be designed into the library is the possibility of 'freezing' the exchange. If the reversible processes are stopped it would be possible to analyse, separate and handle the library members without fear of their decomposition.
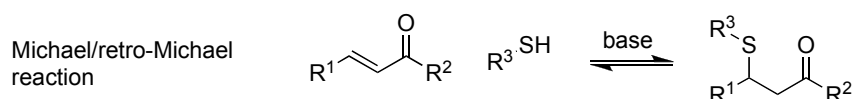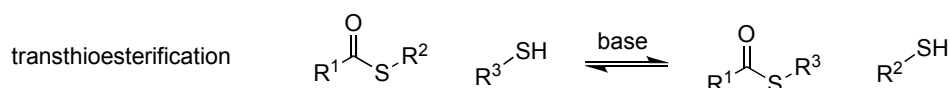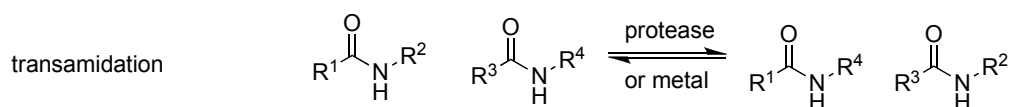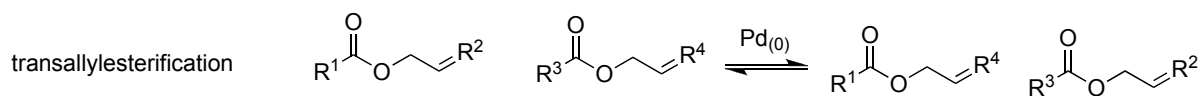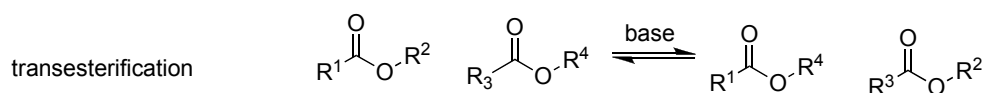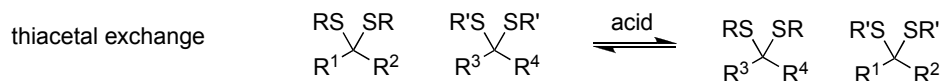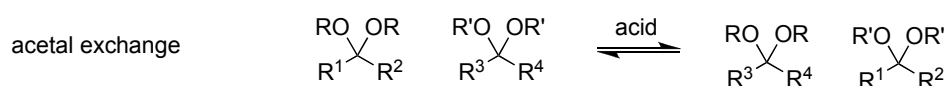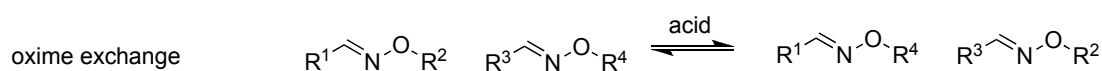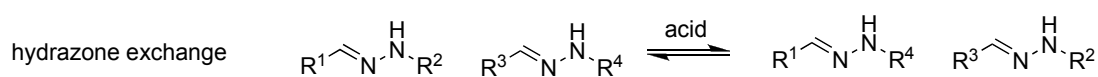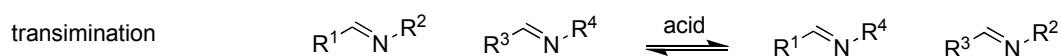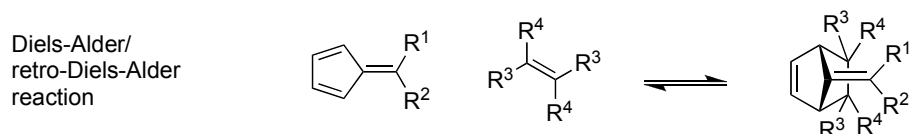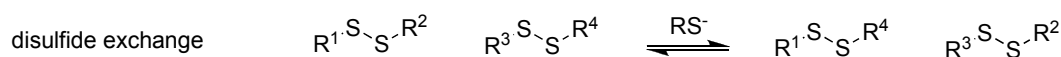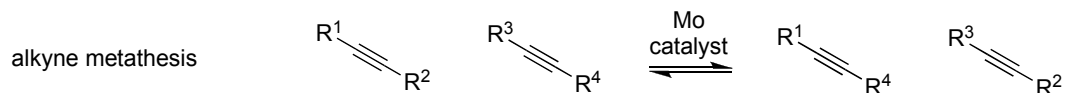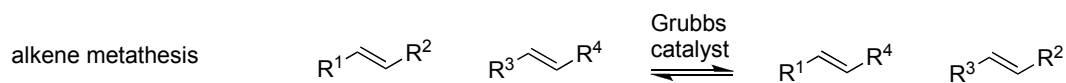
Figure 17. Examples of reversible reactions used in dynamic covalent chemistry.

25

The chemistry of the reversible covalent bond has many examples. The kinetics of the formation of the covalent bond is generally slower than that of the non-covalent bond since covalent bonds are much stronger. In order to form an exchanging library it is often necessary to make use of catalysts that enable the process to occur on a reasonable time scale. Sometimes the removal of the catalyst is a convenient way to 'freeze' the reaction. One of the properties of the bond is its symmetry (Figure 18). Library design will change depending if bond-forming reactive groups are the same or different. The symmetry of the bond dictates whether the building blocks will self-oligomerise or only form alternate sequences. The symmetrical bond forms larger libraries, but the unsymmetrical bond offers greater control over the structure of the library members.

(a)   (b)

$$R_1-SH + R_2-SH \rightleftharpoons R_1-S^{S-R_2}$$

$$R_1\!\!-\!\!\overset{O}{\diagup} + R_2-NH_2 \rightleftharpoons R_1\!\!-\!\!\overset{N-R_2}{\diagup}$$
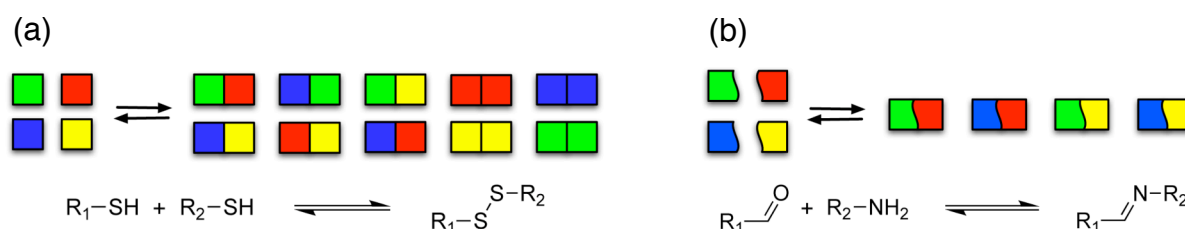
Figure 18. Schemes of dynamic combinatorial libraries created using (a) a symmetrical bond and (b) unsymmetrical bond, with reaction examples.

Examples of symmetrical bonds include disulfide bridges and alkene double bonds formed in the metathesis reaction. The disulfide exchange[143] reaction is of considerable interest since it plays an important role in the folding[144] of proteins. The mechanism consists of one thiolate anion displacing another one from the disulfate. Thiols can be readily oxidised to disulfides in the presence of air. Exchange between disulfides[145] can be easily initiated and is relatively quick in the presence of a catalytic amount of thiol under neutral or mildly basic conditions. The advantage of using neutral conditions is that the exchange is compatible with the presence of biomolecules.

A whole family of exchange reactions involves the carbonyl group. The carbonyl bonds are non-symmetrical and relatively labile, being broken reversibly when an electronegative atom is adjacent to the reactive group. Esterification, for example, is a reversible reaction and when synthesising esters the equilibrium is usually shifted towards the condensation products by introducing an excess of one of the substrates or by removing water from the reaction mixture. Achieving efficient ester exchange

when none of the components is in excess is more difficult and often has to be catalysed by acids or bases. This approach leads on the other hand to unwanted hydrolysis, which influences the distribution of compounds in the library. In order to avoid these undesired pathways another approach[146] uses enzymes such as esterases to promote exchange and this idea has also been applied to form exchanging pools of compounds using[147] amides. The amide bond is very stable so the exchange conditions not involving biological catalysts are quite harsh (high temperature, organometallic catalysts) making the applications limited. Other carbonyl based exchange reactions include aldol exchange[148], thioester exchange[149] and Michael addition[150] and work is carried out to utilise these processes in the creation of DCLs.

The bond that has some of the most practical applications[151] in DCC is the C=N bond. It is formed in the synthesis of imines when condensing an amine with an aldehyde. The substitution patterns and electronic properties of the substrates for imine formation play an important role in the stability of the product. Like with esters, water removal is most of the times necessary to push the equilibrium towards the adducts. The most stable and thus most extensively used in DCC are aromatic imines. Acids promote the exchange and often are used in dynamic libraries. It is possible to work with imines in aqueous solutions although the ease of hydrolysis of these compounds does cause problems. In order to overcome this issue, many systems work in water free or biphasic conditions. The rate of hydrolysis is also increased by electron-withdrawing substituents while the use of electron-donating groups stabilises the adduct. A useful property of the imine bond is that it can be 'fixed' by reducing the imine to a secondary amine by various reducing agents.

A similar class of compounds that base on the reversible C=N bond are hydrazones. Hydrazones are more stable than imines thanks to their mesomeric nature. This property decreases the electrophilicity of the molecule slowing down hydrolysis but also making it less prone to exchange. The exchange[152] is possible only at low pH or at high temperatures. The use of the acyl or other electron-withdrawing group on the amine subunit is also a common way to mediate exchange and allow for the use of hydrazones in DCC.

Non-covalent bond formation and breaking has also been used to create DCLs. The two most common types of bonds are metal-ligand coordination[153-154] and hydrogen bonding[155-156] motifs.

1.10 Examples of dynamic covalent systems

DCLs are addressable systems, since a change of stability of one member will influence other members, so the final equilibrium distribution is derived from the sum of the stabilities of all species in a library. This property is used for selecting a desired component from the library. If one component engages in some kind of non-covalent interaction with a host molecule present in the mixture its concentration will increase over other library members. Non-covalent host-guest interactions are the domain of supramolecular chemistry and in a great degree it is thanks to a rapid development of this area of chemistry that DCC came to be. Actually, supramolecular chemistry is also dynamic by its nature[157]. The liability of the non-covalent interaction connecting a supramolecular assembly allows for a constant association and dissociation of the building blocks. The self-assembly process is also controlled by thermodynamics selecting the most stable structure. Supramolecular chemists spent decades developing new receptors for a wide variety of guests. Many outstanding, elegantly designed molecules have been synthesised improving our understanding of non-covalent interactions. However, designing, synthesising, evaluating and improving receptors by traditional synthetic techniques is a cumbersome process. DCC is a very attractive approach to development[158] of synthetic receptors and sensors, creation[159-162] of supramolecular assemblies as well as in the search[163-166] for ligands for biomacromolecules.

Sometimes a dynamic approach reveals features that are not apparent under kinetic control. Rowan and Sanders used[167] a transesterification reaction to prove that from a library of monomers a macrocycle of a certain size was favoured over other products. The monomers **11** were derivatives of cinchona alkaloids each equipped with a methyl ester group on one end and a hydroxyl group at the other. Both cinchonidine and quinine derived monomers under thermodynamic conditions form cyclic trimers **13** selectively (Figure 19). When the reaction was conducted under kinetic control a much broader spectrum of products was reported. The trimer **13** was still the most abundant product but in a much smaller selectivity, 37% compared to over 90% under thermodynamic control, and also a moderate quantity (23%) of the tetramer was observed. Different experiments were conducted using oligomers instead of monomers as starting material and introducing competing substrates for the esterification reaction, however under DCC conditions the predisposition of the

cinchona alkaloid derivatives to form the most stable cyclic trimers was always predominant.

The authors conclude that the substrate is predisposed[168] to stabilise the trimer **13**. Predisposition has to be distinguished from preorganisation. The latter refers to a favourable conformation of monomers that holds the reactive groups in close proximity favouring one pathway over alternatives. Preorganisation is often accomplished by associating the reactive species on a template and since it enhances the rate of reaction between the building blocks it has a kinetic effect on the system. Predisposition according to the authors is a strong conformational or structural preference expressed by the building block once incorporated into a larger assembly, giving rise to a particular product without the use of a template.
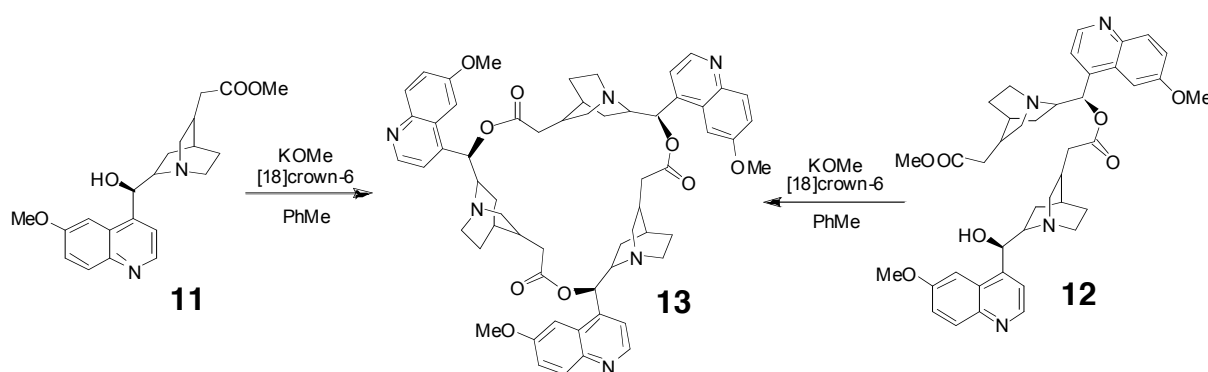
Figure 19. Formation of a cyclic trimer **13** as the dominant product using a monomer **11** or a dimer **12** as starting material.

A similar behaviour was reported[169] by Philp and Comina in a mixture of bis(borazaaromatic) compounds **14** that undergo reversible dehydration to form linear or cyclic oligomers (Figure 20). Upon addition of 4Å molecular sieves the slow process of dehydration was monitored by MALDI-TOF mass spectrometry. Not surprisingly initially the linear and cyclic dimers are the predominant species in the mixture. As the dehydration progresses higher oligomers can be observed, but after 4 days the higher molecular weight species diminish, to leave exclusively the cyclic dimer **15** in solution after 12 days. The predisposition of the specific structure to assemble exclusively into dimers is a good example of achieving selectivity in a dynamic mixture without using a template.

Figure 20. Exclusive formation of cyclic dimer **15** upon slow dehydration of a bis(boraza aromatic) compound **14**.

Introducing a guest template into a dynamic library of potential receptors will result in a distribution favouring the structure that is most thermodynamically stabilised by the s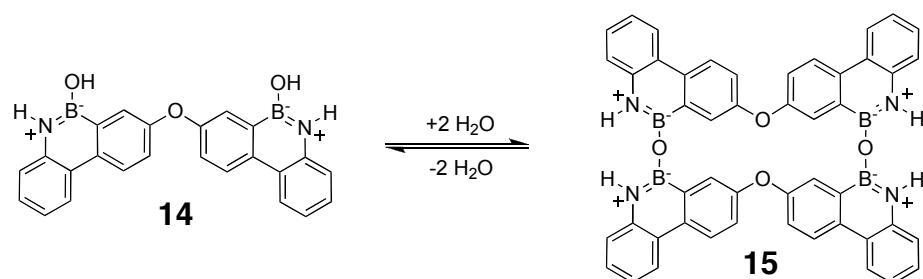pecific template. In other words the subunits will assemble themselves around the template automatically adopting the best combination at a highest concentration. Such process is described[170] as molding when the target substrate is used as an *exo*-receptor. Casting, on the other hand, utilises a target receptor to direct the assembly of the optimal substrate from the library using the target as an *endo*-receptor (Figure 21).
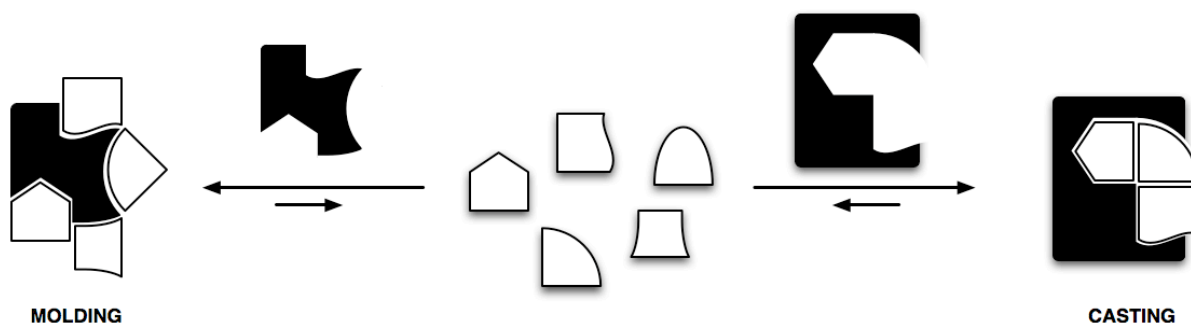
Figure 21. Assembling of a structure from DCL components upon addition of a receptor either by molding or casting.

Otto and Kubik presented[171] one of the examples of optimising a neutral receptor that binds inorganic ions. A sandwich type complex of an ion such as a halide or a sulfate and two cyclic hexapeptides was reported. If two of the cyclopeptides were to be linked by a covalent bond the complex would be even more stable. The problem was choosing the proper linker. A dynamic combinatorial approach using reversible disulfide chemistry was chosen to solve this issue (Figure 22). A range of dithiol derivative spacers **17-22** with different lengths and geometries were used to form a dynamic library of potential receptors with a compound where two peptide rings were linked by a simple disulfide bond **16**. The addition of an anionic guest amplified the host with the highest affinity.
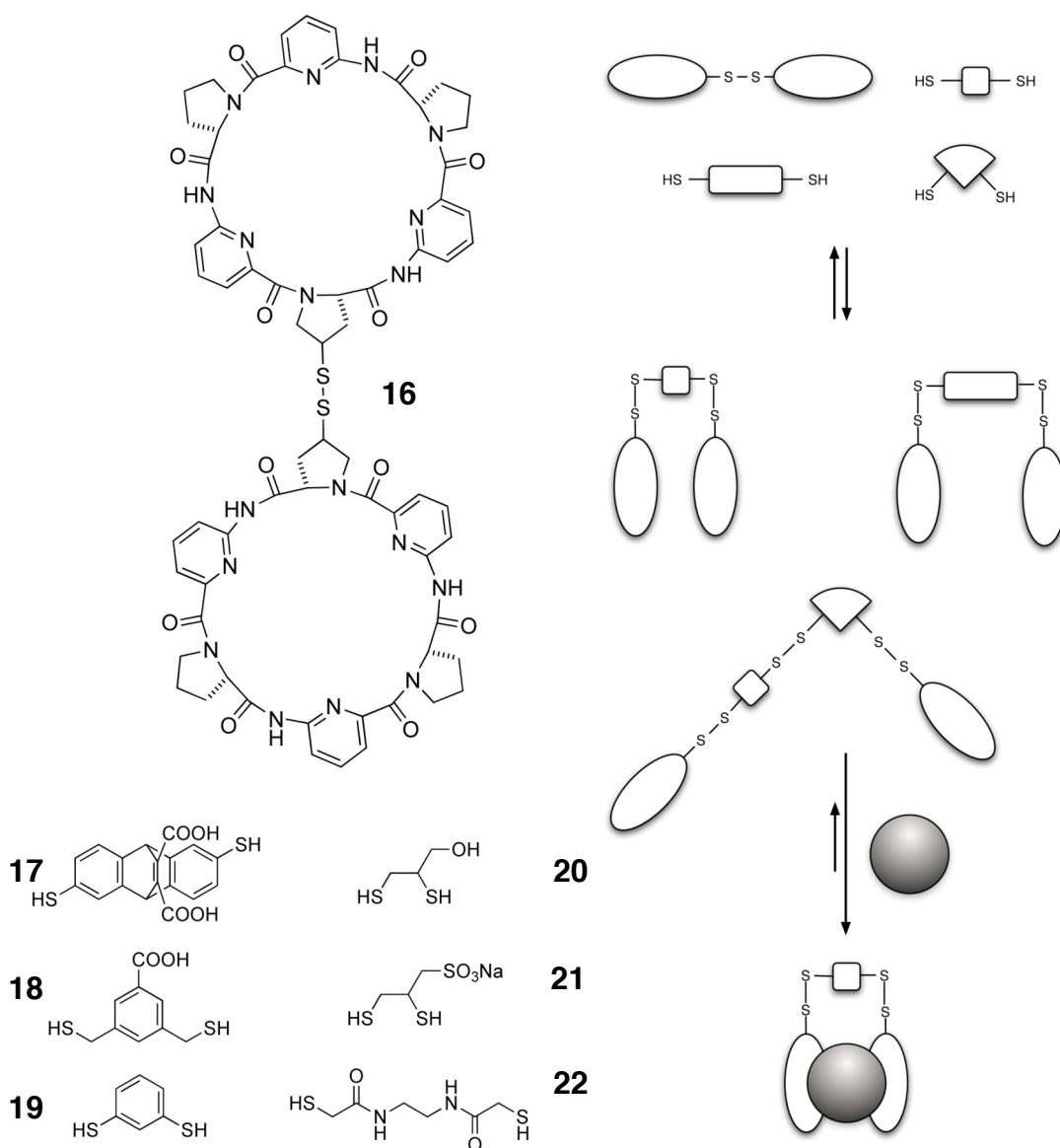
Figure 22. Design of a DCL for the optimisation of an ion receptor. An array of disulfide linkers **17-22** formed structures with two peptide macrocycles. The addition of an ionic guest (grey sphere) selected the most stable sandwich-type complex.

Imines were one of the first compounds that were used to create a DCL in search for inhibitors of biological macromolecules. Huc and Lehn used[172] an imine based library in the presence of carbonic anhydrase (Figure 23). The parent compounds for the generation of the library were four amines **23-26** and three aldehydes **27-29**. From the interchanging mixture of imines one compound **30** was amplified twofold over the other ones. In order to analyse the mixture the imines had to be reduced to the corresponding amines using sodium cyanoborohydride. It has been observed[173] that the imine analogues used in the DCL are good models of the biological activity of the amines.
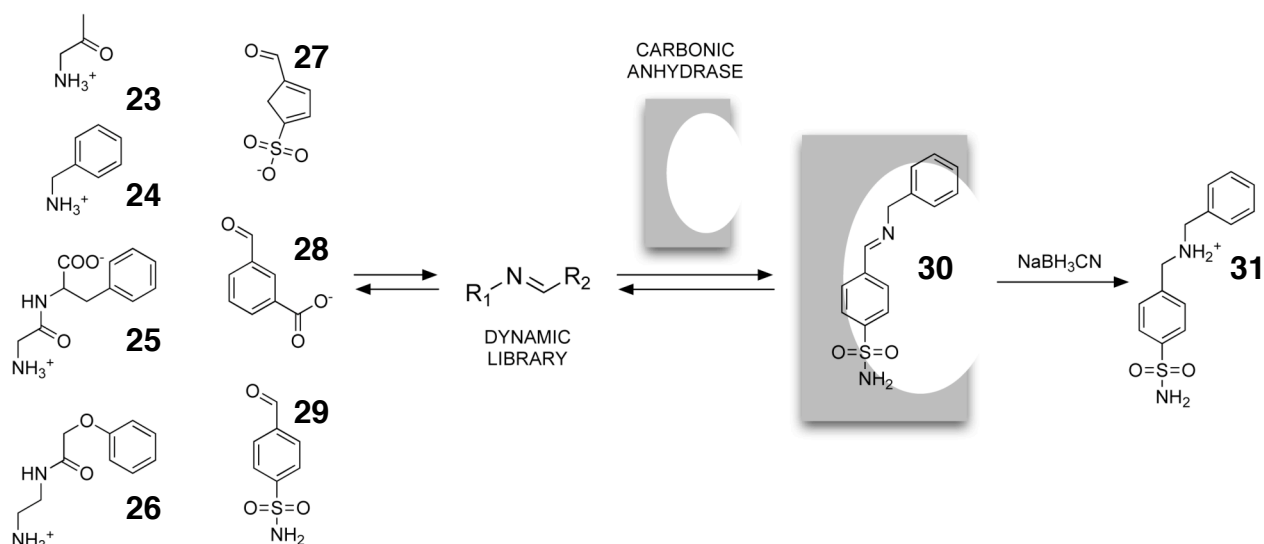
Figure 23. Selection of the best inhibitor for carbonic anhydrase using an imine-based DCL with subsequent reduction of the imine **30** to an amine **31**.

Metal cation receptors have also been a target of imine-based dynamic combinatorial chemistry. Gotor[174] used pyridinedicarboxaldehyde **32** with a homochiral diamine **33**. Different macrocycles were amplified depending on the size of the metal ion. Ba^{2+} induced the amplification of the smaller macrocycle **34** consisting of two amines and two aldehydes, while the bigger Cd^{2+} gave the [3+3] adduct **35** in almost quantitative yield (Figure 24). The imine compounds were also reduced to the stable amines, however the transformation resulted in a decrease in the binding affinity of the complex.
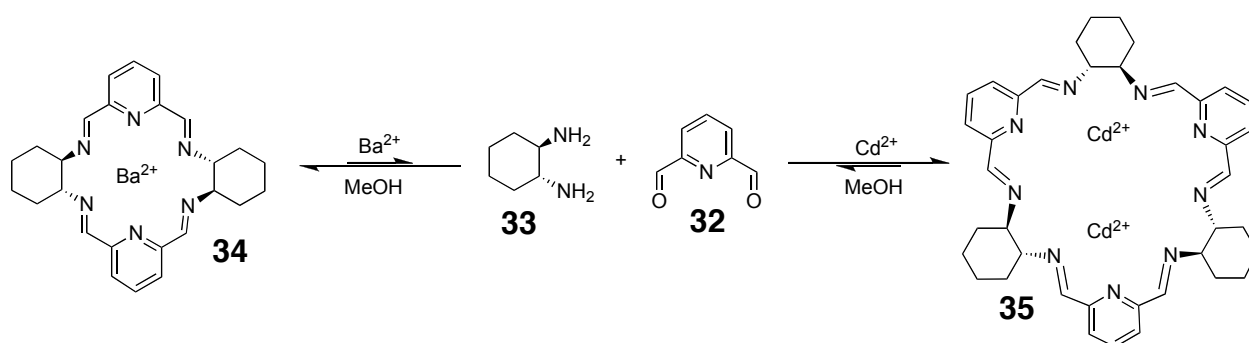
Figure 24. Amplification of different macrocycles from an imine-based DCL using different metal ions.

Imines have been used extensively by Stoddart[175] in research where DCC meets supramolecular chemistry. One of the most formidable examples of utilising the reversible imine bond is the formation of the molecular borromean rings **39** (Figure 25). This elegant assembly consists of three interwoven macrocycles **38**

arranged in such a way that the breaking of one of the rings enables the disassembly of the other two. The imine based structure consists of six dialdehydes **36** and six diamines **37** and the assembly is templated by six Zn^{2+} ions. Again a way of fixing the macrocycles is a reduction of the imines to the amines, in the case of the borromean rings also a removal of the zinc ions with EDTA is possible to give a stable mechanically interlocked structure.
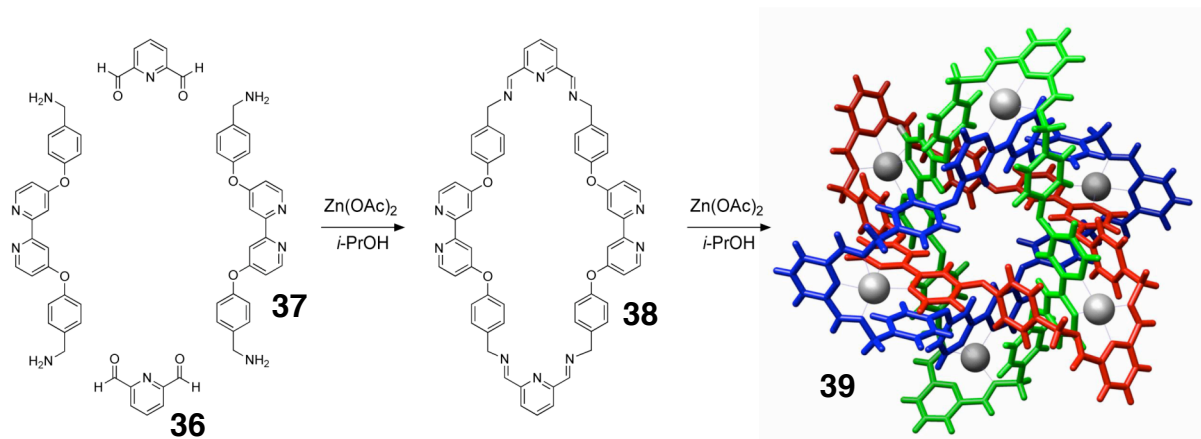
Figure 25. The synthesis and chemical structure of macrocycle **38** and the X-ray crystal structure of the molecular borromean rings assembly **39**. Figure adapted from ref. 151.

Sanders and coworkers amplified[176] a supramolecular structure from a hydrazone-based DCL. A library of peptide-hydrazone building blocks **40** when allowed to exchange in the presence of an acetylcholine template, form a [2]-catenane **41** as the most stable structure (Figure 26). Without the template the mixture consists of cyclic oligomers from a simple dimer up to a hexamer. Upon the introduction of the template and letting the mixture equilibrate the catenane **41** accounts for 70% of the mixture after 44 days. The interactions between the catenane receptor and the guest are not completely elucidated however it is sure that it is a very specific directional binding in order to create such an unexpected three dimensional structure.
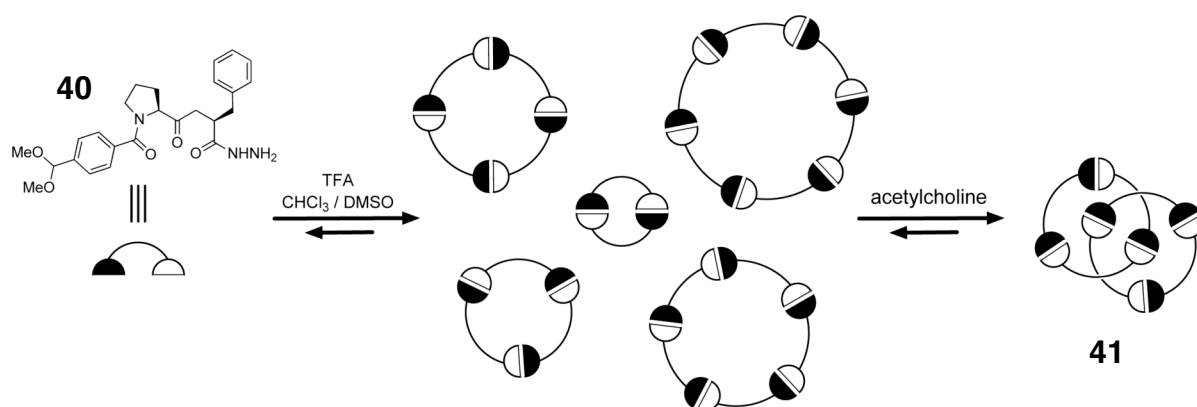
Figure 26. Amplifying a [2]catenane **41** from a hydrazone-based DCL of oligomeric macrocycles using acetylcholine as the template.

Imine and hydrazone exchanges are the most prominent members of the family of exchange reactions based on nitrogen nucleophiles. Both the amine and the hydrazone however are not particularly good nucleophiles and the exchange mainly proceeds *via* the hydrolytic mechanism. Also neither of the bonds presents any utility in synthetic applications. The only useful reaction of the imine is its reduction to the secondary amine.

A logical extension of the methodology is the use of hydroxylamines. Hydroxylamines are excellent nucleophiles and react rapidly with aldehydes in non-polar solvents to form the corresponding nitrone. Unlike hydrazones, diaryl nitrones are capable of participating in a number of irreversible synthetic transformations directly, the most attractive of which are dipolar cycloadditions. A description of nitrone exchange in DCC has appeared in the patent literature[177]. The reaction is, however, restricted to aqueous buffers and the generation of unreactive nitrones mostly for the development in drug discovery. Philp has reported[178] that nitrones are, indeed, capable of undergoing exchange under mild conditions in chloroform. This dynamic exchange is demonstrated and exploited within the context of the selection of a receptor for a dicarboxylic acid from a mixture of nitrones. The small dynamic library consists of four nitrones two of them bearing a single aminopyridine recognition site **42** and **43**, one incapable of recognition **45** and one with two aminopyridine recognition sites **46**. The template is a glutaric acid derivative **44** and thus should amplify the concentration of the complementary nitrone **46**. After 48 h a selectivity of 4:1 was achieved for the side of the equilibrium containing the targeted nitrone **46** (Figure 27). Further exchange processes involving nitrones and their incorporation into kinetic processes will be discussed in the results section.
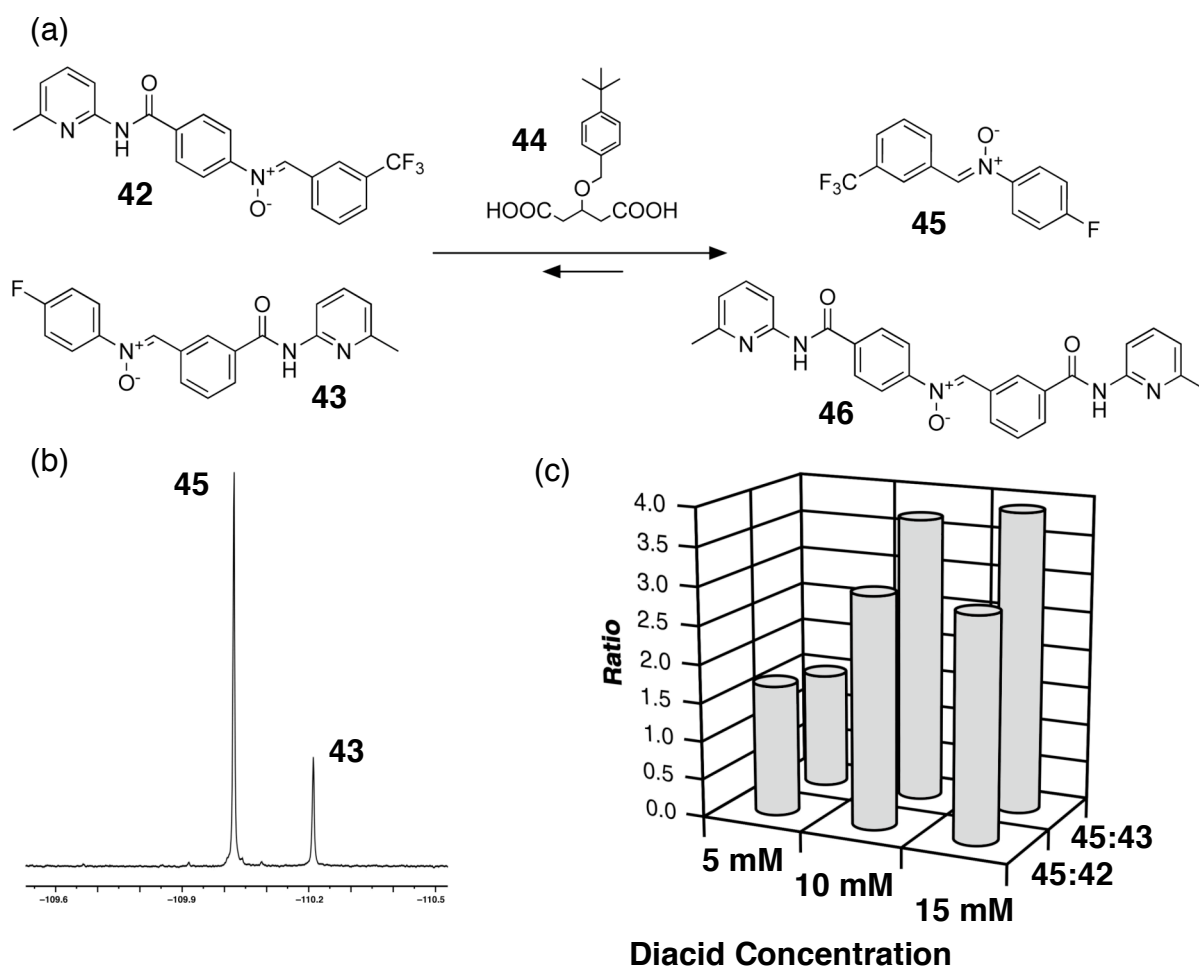
Figure 27. A nitrone exchange reaction. (a) Addition of a diacidic guest **44** stabilises the nitrone with two amidopyridine recognition sites **46** pushing the equilibrium to the right side of the reaction. (b) Partial 282.3 MHz ^{19}F NMR spectrum displaying peaks of **45** and **43**. (c) The graph shows the ratio between **45** and **43** as well as **45** and **42** upon addition to the exchange reaction of 5, 10 and 15 mM of the diacid template **44**.

Nitschke focused his DCC research on expanding the scope of the libraries by incorporating multiple, orthogonal exchange processes into a single system. The described[179] system consists of an imine library formed from compounds **47-50**, which collapses, in a self-sorting fashion upon addition of Fe^{2+} and Cu^+ salts (Figure 28) to form two complexes **51** and **52**. The complexes are formed from specific library components while thermodynamic equilibrium eliminates all other products.
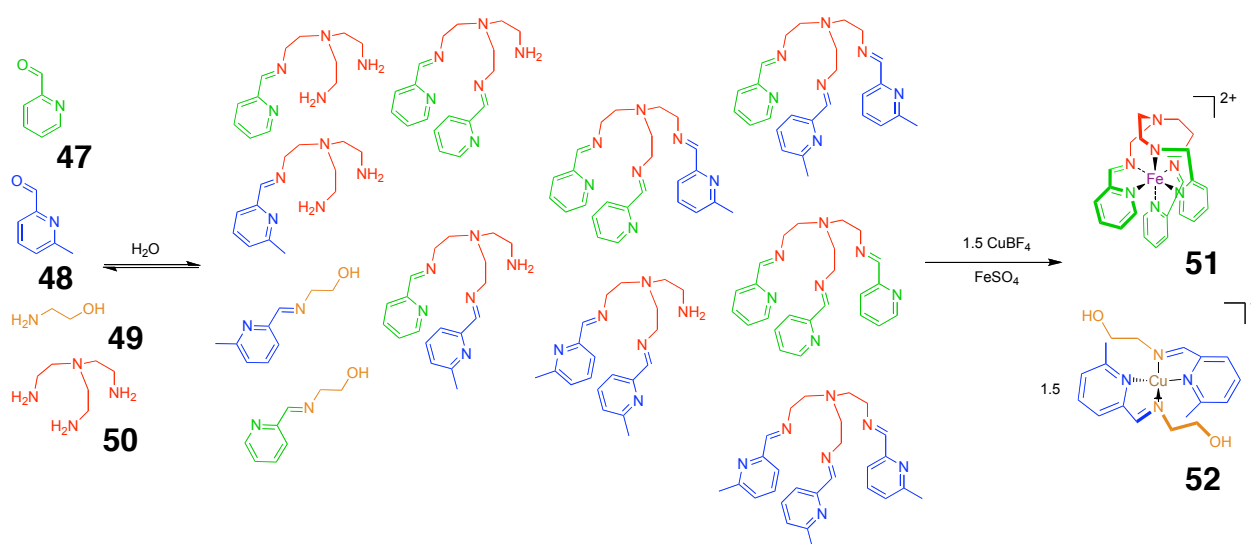
Figure 28. Self-sorting of an imine-based DCL upon addition of Fe^{2+} and Cu^+ salts.

This effect was studied more closely on individual ions in smaller libraries to reveal a subtle interplay of steric and electronic factors of both covalent and coordinative nature. A system based on these interactions was even further developed[180] by incorporating a third, apart from imine and metal-ligand, reversible bonding process (Figure 30). The amine units of the DCL were linked by a disulfide bridge, which allowed for further exchange processes to take place. The reference for the system was an exchange between two disulfides: one with two amine groups **53** and one with two unreactive methoxy groups **54** (Figure 29). This exchange gives a mixture that favours the hetero disulfide **55** in a ratio of 2:1.
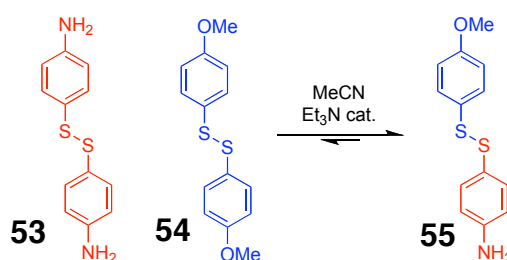
Figure 29. A disulfide exchange reaction used by Nitschke as a reference for the more complex DCL system (Figure 30) giving a measurable ratio of the homo to hetero-disulfide.

The way to bias this equilibrium was to add a pyridine aldehyde **47**, which is capable of forming an imine with the amine group of disulfides **53** and **55** and Fe^{+2} ions that will assemble the components into a complex. The result is a completely different DCL with a new composition and a new set of components.
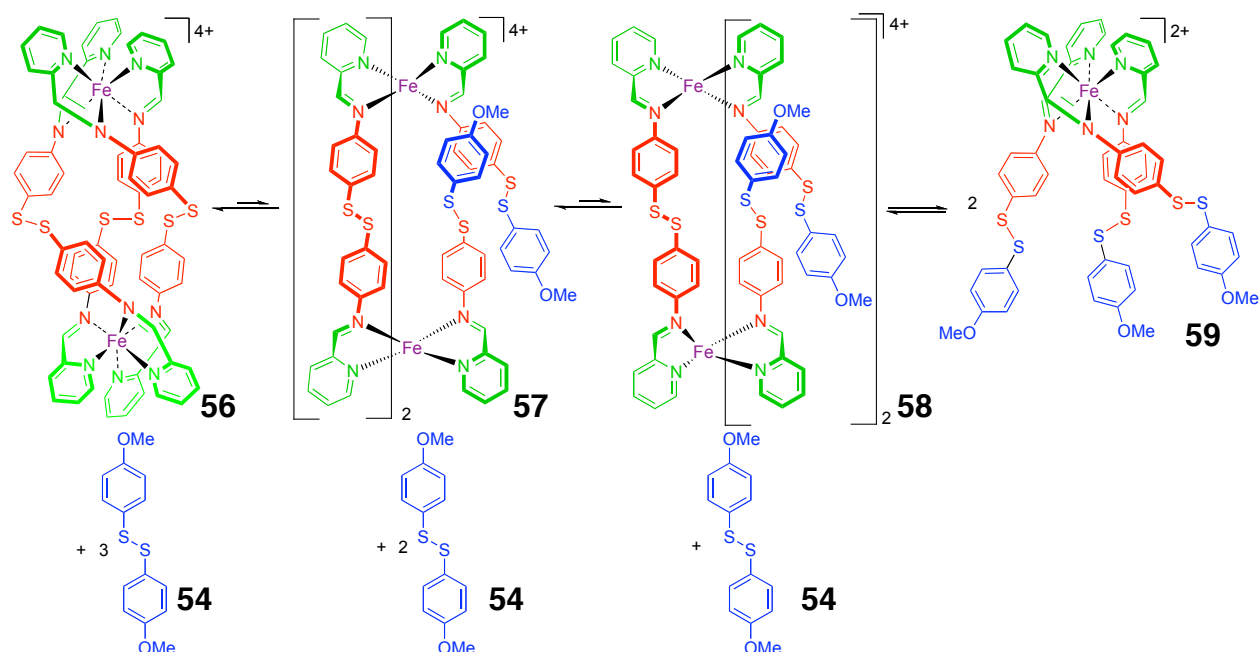
Figure 30. A DCL system consisting of three orthogonal reversible processes: imine exchange, disulfide exchange and metal-ligand exchange.

The DCL strongly favours the homo **53** disulfide as it is incorporated into the stable metal complex **56** changing the equilibrium position, which compared to the reference experiment, is shifted by a factor of 179. Due to different processes that govern the dynamic system it is possible to address them selectively and orthogonally. Transmetallation was performed on the system replacing the metal centres as well as adding of competing ligands. All these stimuli resulted in a response of the library and change of its composition.

In the above examples DCLs were used to identify and amplify new receptors. If such compounds would be equipped with appropriate signalling units they could be used as sensors for the presence of the specific host. There is also another approach where a whole DCL is used as a sensor. As mentioned before all components of a DCL are interconnected and their distribution is dependant on the environment. The idea is very appealing considering the adaptive nature of a DCL and a multitude of its components that could be tuned for a large number of analytes. If the members of the library would be of different colours then the read-out of the sensor would be a fingerprint UV-Vis spectrum over its whole range. Buryak and Severin created[181] a DCL based on metal-ligand reversible chemistry (Figure 31) The authors mixed metal salts of Cu and Ni with three available dyes: Arsenazo I **60**, Glycine Crezol Red **61** and Methylcalcein Blue **62** in an aqueous buffer. The mixture consisted of various combinations of metal-die complexes each with its characteristic UV-Vis absorption.

Any change to the composition of the library would result in a change of the collective spectrum. The sensor library was used to distinguish between various dipeptides. The dipeptides would associate with the metal ions releasing some dies disturbing the library equilibrium. One example is a differentiating between two isomers His-Ala and Ala-His. The UV-Vis spectra recorded upon addition of each of the dipeptides to the library are significantly different and could be used to identify these compounds. This approach was also used for a more challenging examples of more related peptides as Gly-Ala, Val-Phe, Ala-Phe, Phe-Ala, D-Phe-Ala and also proved successful.
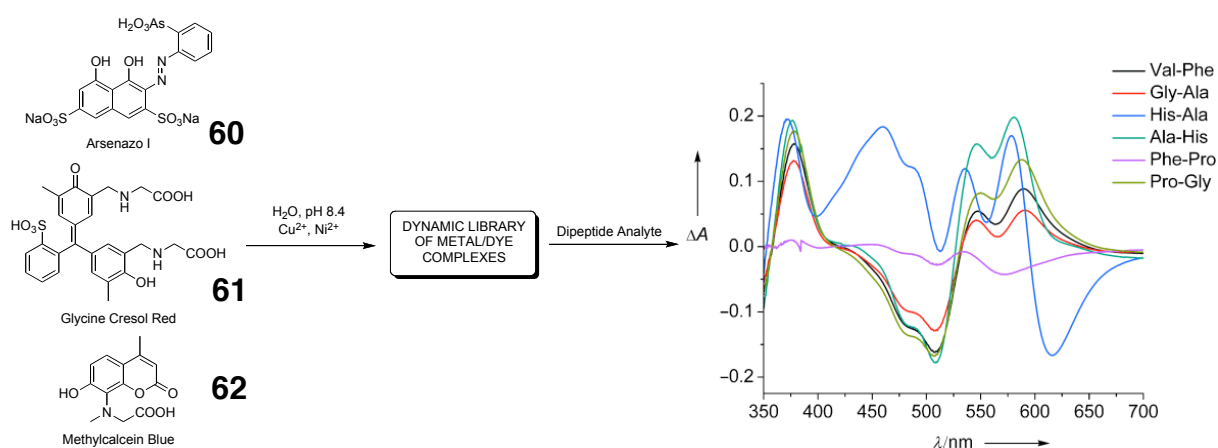
Figure 31. Formation of a DCL based on metal-ligand exchange between metal ions and organic dyes **60**-**62**. The library was utilised as a sensor for detecting dipeptides giving characteristic UV/Vis spectra for specific analytes. Figure adapted from ref. 181.

So far, all demonstrated examples of DCLs were based on receptor assisted combinatorial chemistry. In such strategy a receptor is present in the library during its synthesis and amplifies the formation of the most associating component. This kind of amplification is however inherently linear since the selectivity that is engendered within the DCL is directly proportional to, and limited by, the amount of template added and to the difference in affinity of the template for the target compared to the other components of the mixture. In order to increase selectivity in DCLs and break out of the purely thermodynamic regime, elements of kinetic irreversible chemistry are being introduced into dynamic combinatorial libraries. The term pseudo-dynamic combinatorial library was coined by Gleason, Kazlauskas and coworkers to describe[182] their system involving irreversible reactions. Instead of a thermodynamic equilibrium the library bases on an irreversible creation and destruction of its

components. The recognition step is still thermodynamic and the library members associate with the receptor like in a traditional DCL. The amplification is achieved kinetically when the most tightly bound members of the library are being destroyed slowest of all and unbound compounds are destroyed the fastest. The experimental system utilised[183] sulfonamide-dipeptide inhibitors with carbonic anhydrase (CA) as the receptor and a protease as a destructive enzyme. The reaction vessel was separated by a membrane to keep the two enzymes separate, but letting the library members to pass freely (Figure 32).
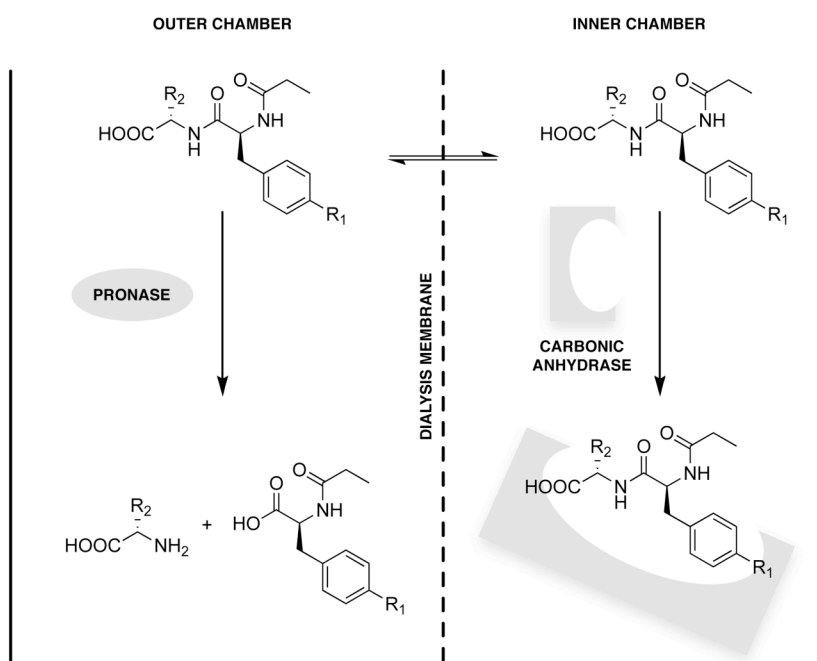
Figure 32. Design for a DCL utilising selective destruction. Dipeptide components can diffuse across the dialysis membrane between the two chambers. The pronase in the outer chamber cleaves the dipeptides to their corresponding constituents. The carbonic anhydrase in the inner chamber prevents the strongest binding blocks to diffuse back and thus be destroyed selecting the inhibitor with the highest affinity.

When a known CA inhibitor and a non-binding dipeptide were introduced into the system a selectivity of over 20:1 was achieved after 12 h. The more interesting setup consisted of two competing inhibitors. The ratio of the binding constants equalled 2.1:1, but in a pseudo-dynamic scenario a selectivity of 3.8:1 was achieved clearly beyond the thermodynamic limitations. One of the problems of pseudo-dynamic approach is finding a proper kinetic destruction component that would have a broad spectrum of activity and would not bias the selectivity. Another problem is that instead of one component being amplified during the experiment most of the library is

destroyed, which is contradictive with the approach of a synthetic chemist. A more constructive approach is to develop a DCL that performs in a creative and not destructive manner and to amplify a compound kinetically by utilising it to make a more complex and useful structure.

Ramström investigated[184] the coupling of a DCL to an irreversible secondary reaction. In what he termed a dynamic combinatorial resolution process the thermodynamically selected library member undergoes an irreversible reaction and the product dissociates from the template allowing for further recognition events. Ramström demonstrated this principle using a reversible nitroaldol reaction of a nitroalkane **68** with 5 aldehydes **63-67** forming a new carbon-carbon bond in the presence of a base (Figure 33). The template for the selection was an enzyme lipase PS-C I, which also catalyses a transesterification reaction of nitroalcohols. The acyl donor for the reaction was *p*-chlorophenyl acetate. The best substrate **69** for the enzymatic reaction was selected from the library and the transesterification product consisted for more than half of the mixture of the esters and the total yield being more than 80%. Another advantage of using a kinetic enzymatic process was the asymmetric discrimination leading to a 99% enantiomeric excess of the major product.
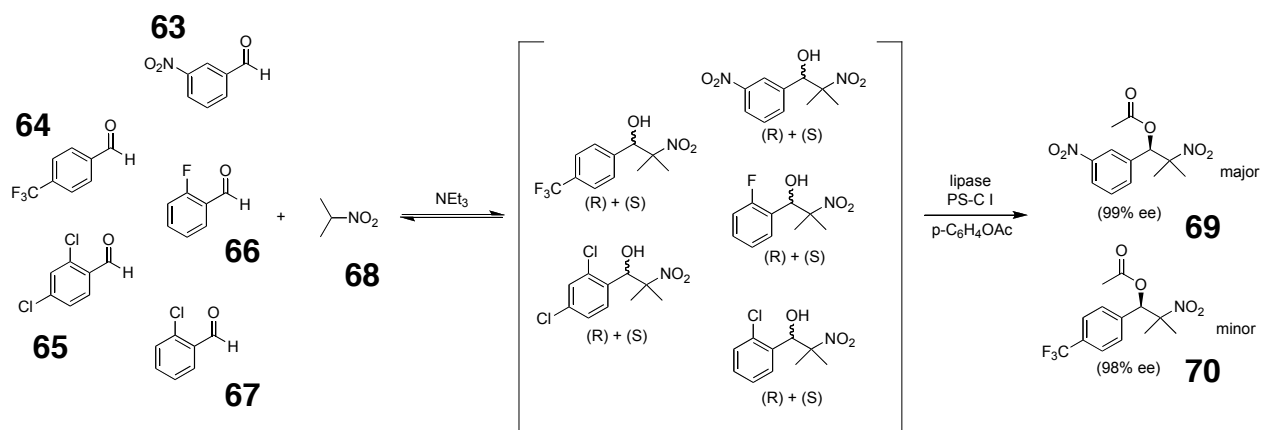
Figure 33. Generation of a nitroaldol-based DCL for lipase-mediated asymmetric resolution.

Ramström created[185] a DCL based on the same connection/disconnection process, but in order to resolve the library used an intramolecular cyclisation reaction (Figure 34). The dynamic mixture consists of nitroalcohols and their substrates: nitroethane **71** and differently substituted benzaldehydes. The nitroalcohol **72** with a CN substituent at the 2 position on the aldehyde ring is capable of undergoing a 5-*exo*-dig irreversible cyclisation to form an iminolactone **73**. After 24 h the whole

nitroethane **71** is incorporated into the cyclisation product **73** and the initially formed nitroalcohols disappear from the mixture.
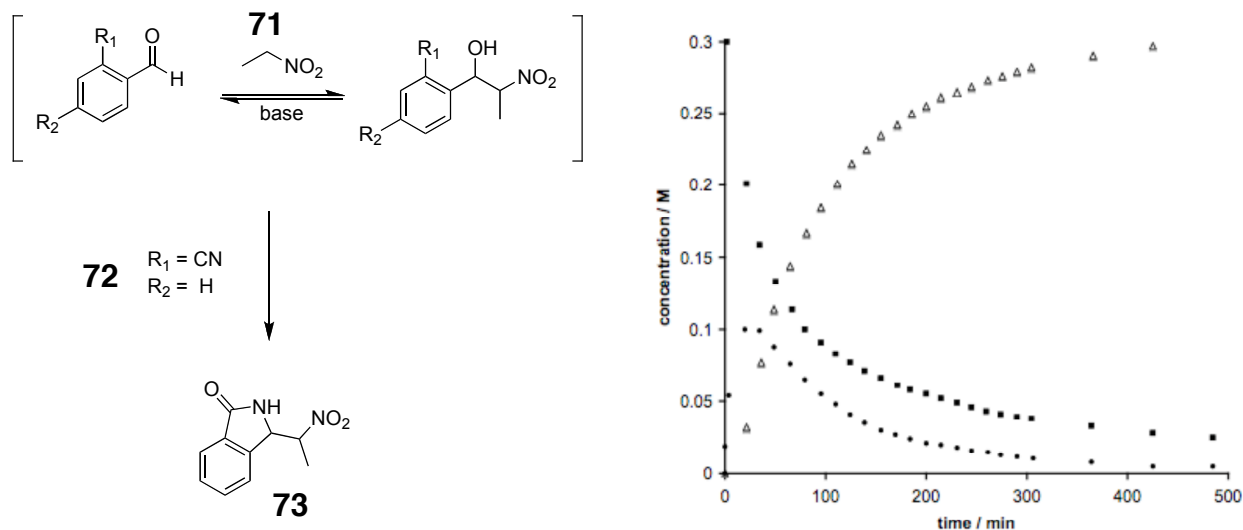
Figure 34. Selection from a nitroaldol-based DCL is achieved by intramolecular cyclisation reaction. The graph follows the reaction composition over time: the formation of the cyclised product **73** (triangles), the corresponding aldehyde (squares) and the nitroalcohol intermediate (circles).

This nitroaldol system was expanded[186] further, when the stereochemistry of the product was considered. The two diastereoisomeric forms of the iminolactone **77** are in equilibrium in the presence of base. However one of the diastereomeric forms displays a much greater affinity for crystallisation. The created system consists of a dynamic library of nitroalcohols formed from aldehydes **74-76** and nitroethane **71** from which a dynamic pair of diastereomers is selected kinetically and further separated by crystallisation (Figure 35).
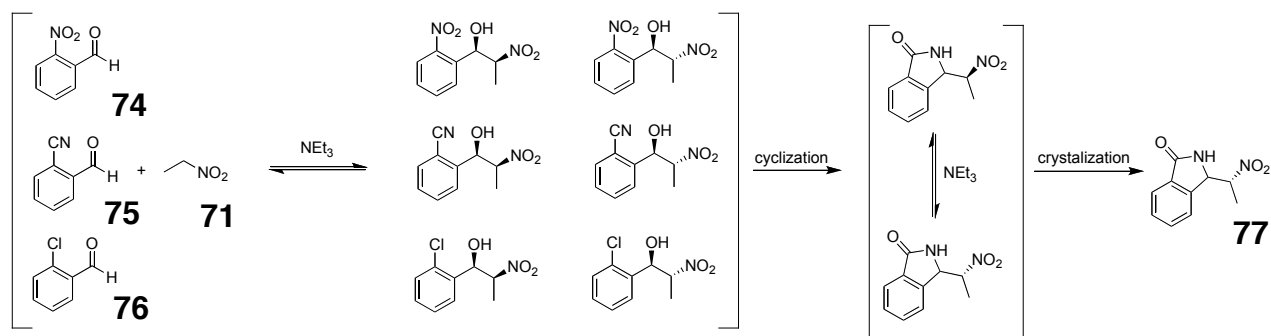
Figure 35. Intramolecular cyclisation is used to select a given structure from a nitroaldol-based DCL with a subsequent separation of diastereoisomers by crystallisation.

Philp and coworkers used[187] an intramolecular hydrogen bond to influence the distribution in a DCL. The authors used a reversible Diels-Alder reaction to create a DCL from a family of acid maleimides with different linker lengths **78-80** (Figure 36). The reactive partner for the maleimides **81** carries a complementary amidopyridine recognition site that is capable of associating with the carboxylic acid. The recognition can play a dual role: on one hand could amplify the rate of the reaction, upon the association of the substrates making the reaction pseudo-intramolecular. On the other hand the recognition could be conserved in the product stabilising the cycloadduct shifting the equilibrium of the reversible reaction.
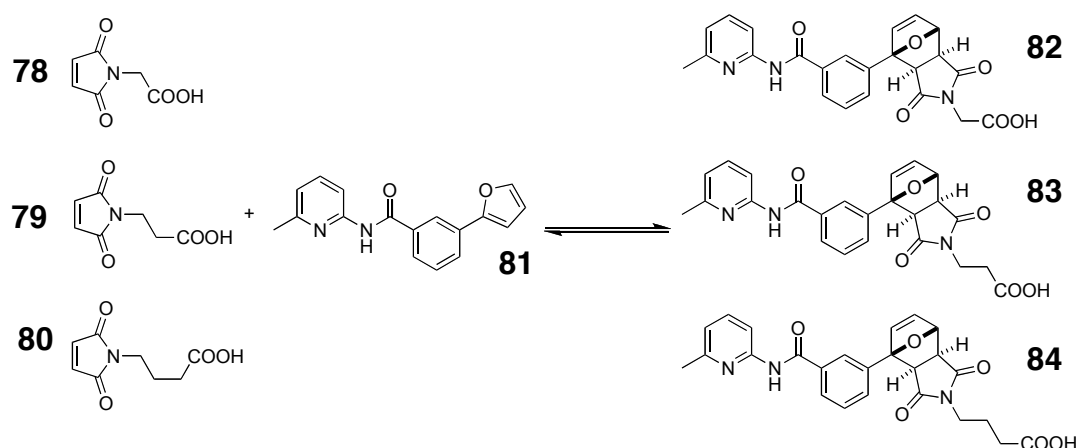
Figure 36. A Diels-Alder reaction-based DCL forms three cycloadducts **82-84** capable of intramolecular recognition using a double hydrogen-bond motif between a carboxylic acid and amidopyridine.

The maleimide with two-carbon atom spacer **81** has been shown[188] to have the best geometry to thermodynamically stabilise the cycloadduct **85** it forms. However it is the reaction of the maleimide with a one carbon atom spacer **78** that has the highest initial rate, suggesting easiest access to the reaction transition state, but it lacks the flexibility to stabilise the product. The reaction of the maleimide with the longest spacer **80** is disfavoured both thermodynamically and kinetically. A DCL prepared with the four building blocks **78-81** will display a different selectivity over time of the experiment: in the beginning cycloadduct **82** will dominate the mixture but over time it is maleimide **79** that forms the thermodynamically most stable product **83** (Figure 37).
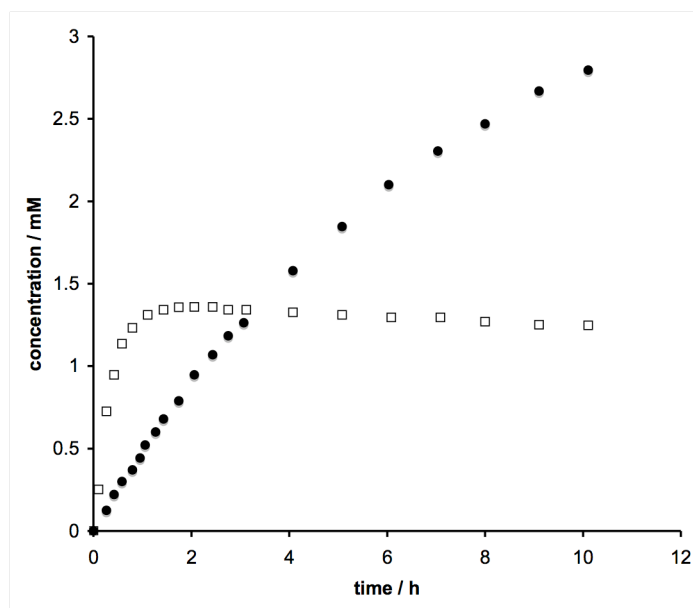
Figure 37. The graph follows the concentration of cycloadduct **82** (squares) and **83** (filled circles) in a Diels-Alder reaction-based DCL. Cycloadduct **82** dominates the mixture at the start of the experiment with the highest initial rate. The thermodynamically stabilised product **83** is dominant in the later stages of the reaction with the concentration of **82** falling.

DCC is developing and more and more experimental examples appear in the literature. However, with the growth of the size of DCLs there starts to be a problem with the analysis of such complicated networks. That is why in parallel to experimental work a lot of computational modelling of DCLs is carried out. Severin has conducted[189] a broad numerical simulation on different types of small dynamic libraries. General trends and characteristics were defined for the consideration in the design of DCLs. The conclusions[190] are quite surprising as from these studies it is apparent that the assembly with the highest affinity is not necessarily the one amplified the most. Two DCL models were investigated: type A where a single building block forms a library of differently sized oligomers and type B with two or more building blocks form oligomers of a specific number of components (Figure 38).
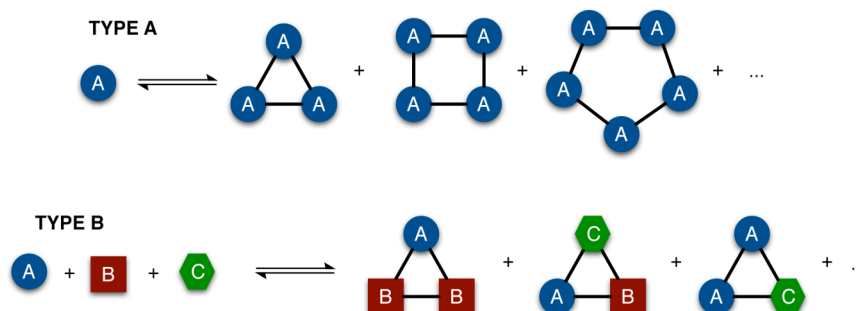
Figure 38. Two types of theoretically investigated DCL types. Type A consists of a single species capable of forming cyclic oligomers. Type B consists of different building blocks that form a structure with only a specific number of components.

The results for DCL type A show the tendency for the formation of oligomers of relatively small aggregation number. For DCL type B it is apparent that the composition of the aggregate reflects the composition of the library. These preferences are mainly observed for DCLs with an excess of target and high association constants. In order to prevent the DCL from selecting a structure that is not necessarily the best binder there are a couple of considerations to follow. According to the simulations it is advantageous to work at relatively low concentrations of the target and to design the DCL in such a way that the dominating species in the mixture are the building blocks and not the aggregates.

Another set of simulations, this time for large DCLs, was conducted[191] by Sanders and coworkers. The aim of the calculation was again to assess the correlation between host-guest binding and host amplification. The correlation was either good or bad depending on the distribution of the association constants between the library members (Figure 39). The factor that prevented the highest binding compound to be amplified was competition for building blocks.
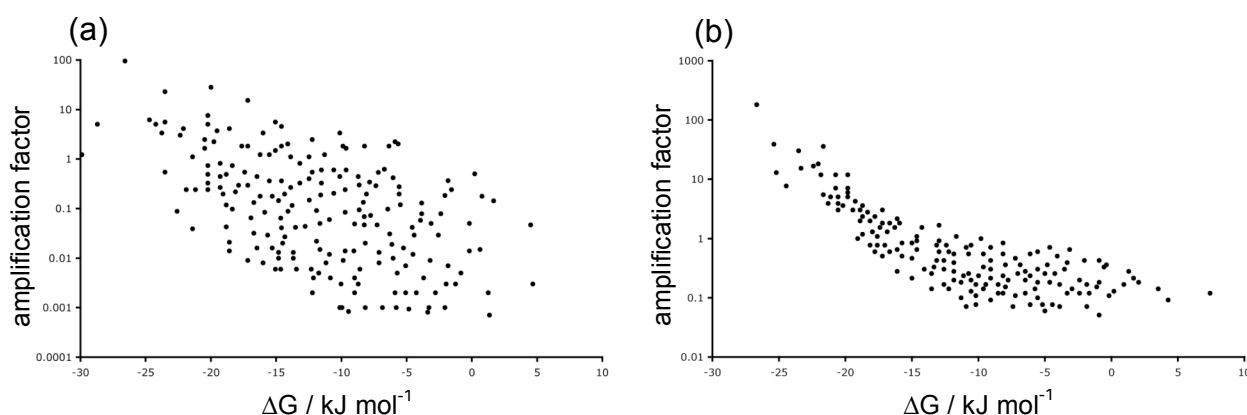
Figure 39.   The relationship between amplification and free energy of binding for all of the hosts in two randomly generated DCLs, that differ only in the way the binding constants are distributed over various hosts, with (a) bad and (b) good correlation.

If another member of the library with a lower affinity for the target is present in the library at a relatively high concentration and at the same time uses the same building blocks as the highest binding compound it is possible that it will prevail in the library. Again, the answer to this problem is to work at substochiometric amounts of template. The limited amount of template, however, reduces the overall selectivity. The point stated again, is that the design and utilisation of DCLs is, as for all complex systems, a challenging task and a fine balancing act.

Another look into emergent phenomena[192] of large DCLs was taken by Otto and

coworkers. While simulating large number of DCLs they have noticed that at certain conditions unexpected patterns appear in correlation between an amplification factor and template affinity of DCL components (Figure 40).

These patterns emerge from the inherent properties of DCLs some of which were already mentioned. First of all in order to observe the patterns a small aggregate, most likely a dimer with a considerable affinity for the template, has to be present at a high concentration in the library
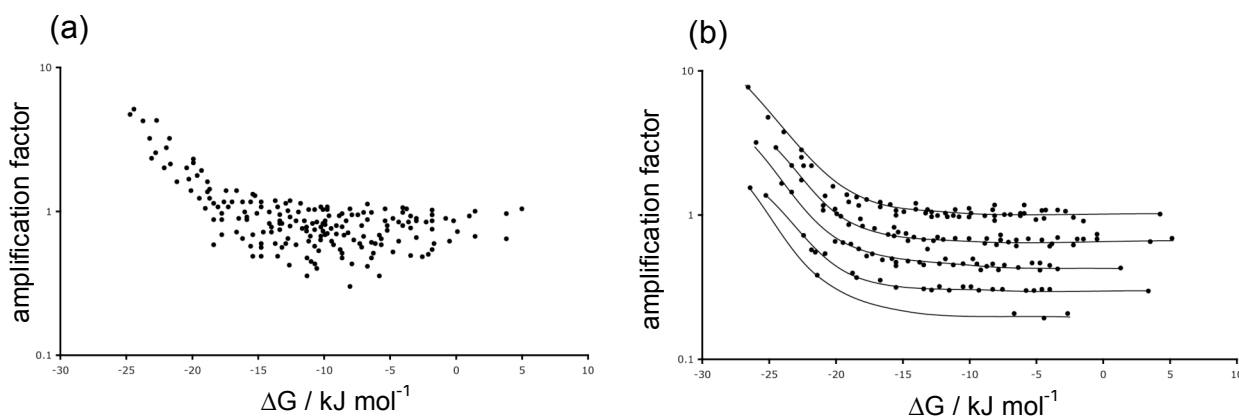
Figure 40. (a) Typical correlation between the amplification factor and the affinity for the template for a randomly generated DCL. (b) Emergence of a pattern in the same library with different set of binding constants.

It seems more favourable to create a lot of copies of the dimer than a higher oligomer, even if it is a better binder for the template. Since the two building blocks necessary to form the given dimer are mostly incorporated in this species, any other library member that incorporates these building blocks will be penalised and their concentration upon addition of the template will not be amplified as much as the components that do not compete with the dimer for building blocks. This example shows how one specie with given properties can influence the whole library not only by influencing the thermodynamic equilibrium by binding the template but also in more indirect ways through the competition for building blocks.

1.11 Self-replicating systems

There is no question that there would be no life without replication. The emergence of replication from simpler systems is one of the central concepts in the origin of life research. The ability to replicate is also an appealing synthetic technique which

guarantees nonlinear amplification of production that is practically error-free. This feature could be used to amplify a specific component from a mixture. As the processes within dynamic combinatorial libraries, self-replication bases strongly on recognition. The precursor of self-replication is templated synthesis[193], the difference being that in replication the product of the reaction is the template for its own formation. The dependence of the rate of self-replication on the amount of template in the system makes the process addressable what in result can be interpreted as chemical logic behaviour. These properties make self-replication a desirable feature incorporated into complex systems.

The development[194-196] of self-replication methodology bases on the minimal model of self-replication (Figure 41). This model contains three distinct reaction channels. The first is the uncatalysed bimolecular reaction between reagents **A** and **B** to afford the template **T**. A key requirement of the model is that **A** and **B** bear complementary recognition sites and partner reactive sites separated by a spacer. One consequence of the presence of the recognition sites is that **A** and **B** can associate with each other to form a binary complex, [**A·B**]. The formation of this complex opens a second reaction channel – the binary or [**A·B**] complex channel – in which **A** and **B** are preorganised with respect to each other and the reaction between them is pseudointramolecular. The product of this reaction channel forms a closed template **T**$_{inactive}$ in which the recognition used to assemble the binary complex lives on in the template, thus, although rate acceleration is achieved by this mechanism, this template is catalytically inert. The third reaction channel available to the system is the autocatalytic cycle. In this channel, **A** and **B** bind reversibly to the open template **T** to form a catalytic ternary complex [**A·B·T**]. In a manner similar to the [**A·B**] complex, the reaction between **A** and **B** is also rendered pseudointramolecular. Bond formation occurs between **A** and **B** to give the product duplex [**T·T**], which then dissociates to return two molecules of **T** to the start of the autocatalytic cycle. Thus, assuming the open template **T** presents its recognition sites in the correct orientation, it can act as a template for its own formation, transmitting molecular information through the formation of identical copies of itself.
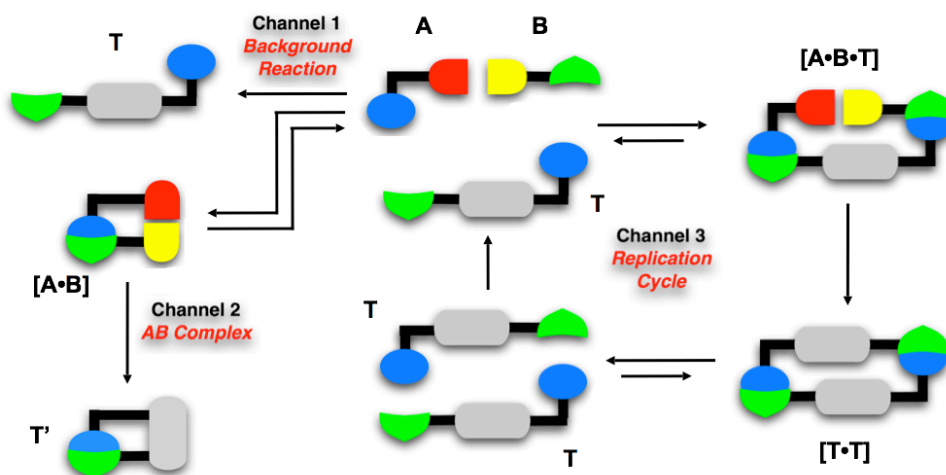
Figure 41. The minimal model of self-replication. Reagents **A** and **B** can react through three pathways – an uncatalysed bimolecular reaction, a recognition-mediated pseudounimolecular pathway mediated by a binary complex [**A·B**] and a recognition-mediated pseudounimolecular autocatalytic cycle mediated by a ternary complex [**A·B·T**]

In a perfect scenario one autocatalytic template would give two, two would give four, four would give eight, rendering the product growth exponential. Such acceleration is still to be achieved in a minimal model. Most systems are characterised by parabolic[197] growth of varying efficiency (Figure 42).
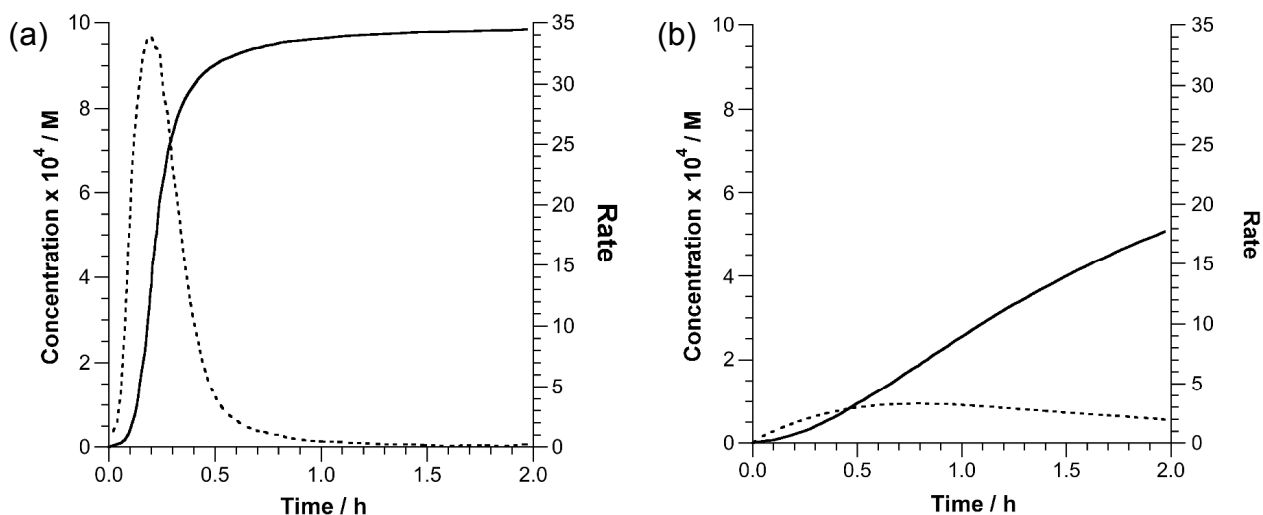
Figure 42. Simulated concentration *vs.* time (solid line) and rate *vs.* time (dashed line) profiles for replicators displaying (a) exponential and (b) parabolic growth.

One of the problems with the efficiency of self-replicators is selectively channelling the reaction through the autocatalytic pathway disabling other channels. Another problem in the design of an efficient self-replicator is a fine interplay of the binding strength. The binding has to be strong enough to readily assemble the ternary complex [**A·B·T**], on the other hand the product duplex [**T·T**] has to be able to

47

dissociate in order to return the template to the beginning of the autocatalytic cycle. These obstacles make self-replication research a challenging field, but the possibilities of the methodology are very significant both to the origin of life and systems chemistry.

Since Nature is undoubtedly scientist's greatest inspiration, the first replicating systems were created by modifying the existing biological replicating machinery. The replicating[198] entity that is central to all life is DNA. Although the concept of the DNA's complementary strands where one is the template for the synthesis of the other is quite simple, the whole process of replication requires complex enzymatic machinery. The goal of replication research is to enable replication in its minimal form without the need for enzymatic activity. One of the first examples that use nucleotide recognition properties to promote replication was presented[199] by von Kiedrowski in 1986 which was published[200] in an improved form in 1991. In the experiment two trinucleotides **85** and **86** underwent condensation on a hexanucleotide complementary template **87**, which at the same time was the product of the condensation (Figure 43). The substrates were trideoxynucleotide 3-phospahte with the 5'-terminus protected as a methyl ester with the base sequence CCG **85** and a second trideoxynucleotide 3-phosphate bearing a complementary CGG sequence with the 3'-terminus protected with an *o*-chlorophenyl group **86**. The reaction was activated with EDC and monitored by HPLC.

Figure 43. A nucleotide-based self-replicating system consisting of two trinucleotide units which ligation is templated by a hexanucleotide complementary product template.

This example was the first system where a characteristic sigmoidal shape profile for the time course of the product formation was observed (Figure 44). The shape of the curve signifies that the rate of reaction is slower at the beginning due to a low concentration of template. Once there is a sufficient amount of the product the reaction reaches its maximum autocatalytic rate and later slows down again with the depletion of starting materials.

Figure 44. Concentration *vs.* time profile of the reaction between trinucleotides **85** and **86** without the addition of preformed template **87** (squares) and with 0.32 mM of the template **87** added at the start of the reaction (circles).

As in the search for the origin of life where DNA is used, RNA and its catalytic properties is also considered as an analogue. Joyce reported[201] a system involving a self-replicating ligase ribozyme (Figure 45). As in the RNA World the reaction used was an RNA-catalysed condensation of RNA fragments designed in such a way as to produce an exact copy. When the preformed template **90** was added to the system Joyce observed an initial increase of rate of reaction between fragments **88** and **89**, but at later stages of the reaction the rate was equal to that of the untemplated reaction. Joyce explained that even though the system was designed to form identical copies of the starting template it seems that the ribozymes of the next generation lost their catalytic properties and had greater difficulty forming a productive complex with the two substrates. A more probable explanation is that the templates displayed too strong binding in order to dissociate and allow for the catalytic turnover.

Figure 45. Ribozyme-based self-replicating system. (a) Structure of the ternary complex utilised in the self-replicating cycle (b). (c) Graph following the reaction between the ribozyme fragments **88** and **89** without preformed template **90** (circles) and with 0.4 μM of ribozyme template **90**.

The system was also expanded[202] to perform in a cross-catalytic manner, where two ribozymes direct each other's synthesis from a total of four building blocks. The cross-catalysis pathway was demonstrated to be significantly more efficient and so dominated other processes in the system. Joyce remarked that the systems developed so far are not capable of Darwinian evolution and that an incorporation of self-replication into more complex system is necessary in order to fulfil this goal.

One of the first synthetic small-molecule self-replicating system was developed[203-204] by Rebek. The reaction used to join the building blocks **91** and **92** was a simple amide formation and the recognition still held a resemblance to the base pairing observed in Nature, utilising adenosine and Kemp's triacid (Figure 46). Rebek's search for an efficient replicator has shown that the design of the spacer between the recognition and the reactive site is very important. The first attempts with a short and flexible spacer resulted in the building blocks to react mainly *via* the AB complex

pathway. The first system that Rebek claimed to be self-replicating utilised a naphthalene spacer. The reaction profile did not show the sigmoidal profile, but there was a strong template effect when preformed template **93** was added at the start of the reaction. The lack of the characteristic shape was explained by a considerable amount of the AB complex pathway in the system. The replication mechanism proposed by Rebek was a subject of a controversial discussion and many authors proposed alternative explanations to the rate acceleration. Menger suggested[205] that the observed phenomenon is nothing more than amide catalysis of aminolisis caused by stabilising the zwitterionic intermediate through hydrogen bonding. Rebek refuted[206] these accusations and furthermore Reinhoudt performed[207] a detailed kinetic analysis proving a significant contribution of the ternary complex in the reaction mechanism. The bottom line in the whole argument is that Rebek's system was relatively inefficient and despite further modifications[208] has mostly historic significance.
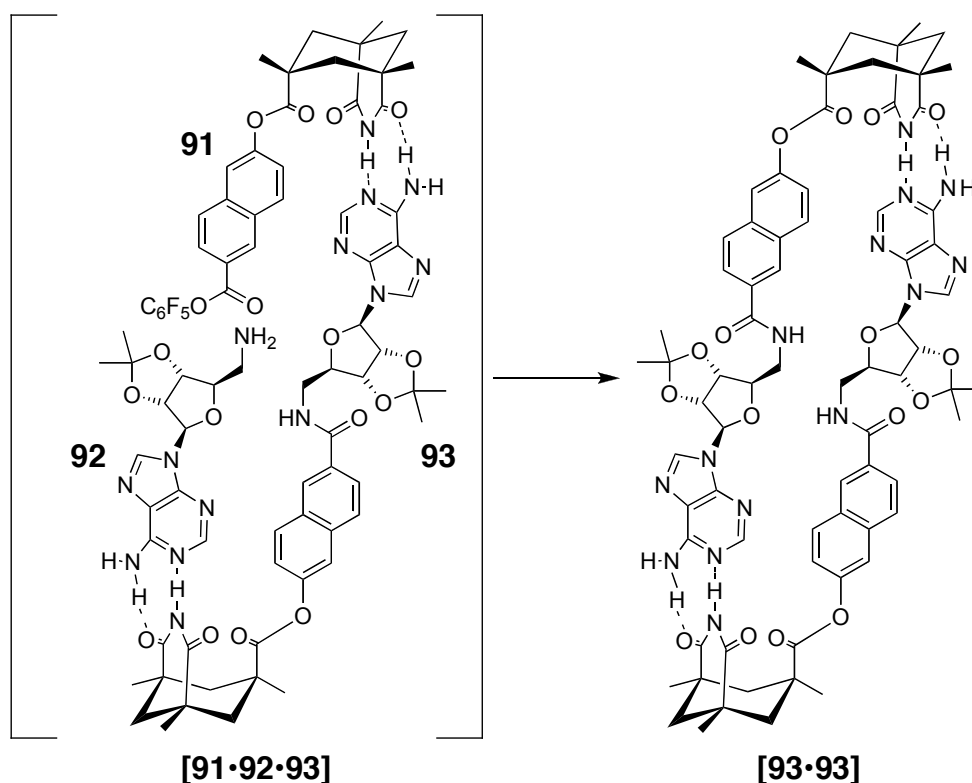

Figure 46. Formation of a template duplex **[93·93]** from the ternary complex **[91·92·93]** by amide formation utilising recognition motif between Kemp's triacid and adenosine.

Xu and Giuseppone introduced[209] elements of dynamic combinatorial chemistry into Rebek's system (Figure 47). A reversible imine synthesis reaction was used instead of an irreversible amide bond formation. Building blocks based on Rebek's structures

94-98 were used with either an aldehyde or an amine functionality to create a dynamic combinatorial library.
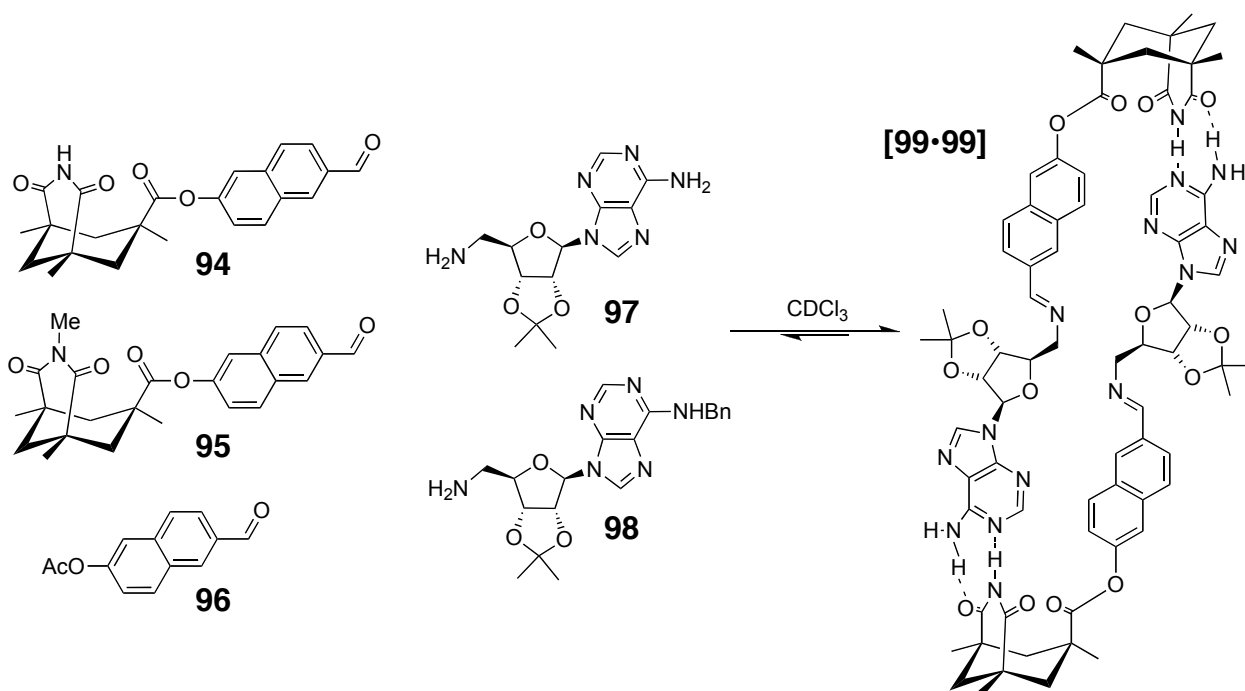

Figure 47. An imine DCL based in the self-replicating system developed by Rebek. The imine **99** with complementary recognition sites stabilises itself by forming a duplex **[99·99]**.

Aldehyde **94** and amine **97** were equipped with complementary deprotected recognition sites between Kemp's triacid and adenosine. The imine formed from these building blocks **99** is capable of forming a duplex (responsible for replication in Rebek's system) thus thermodynamically stabilising this structure over other possible combinations from the library. The expression of the duplex **[99·99]** was increased more than 83% compared to the control DCL without any recognition present. The authors tried to prove that the selection proceeds *via* self-replication, but since both the covalent linkage and the recognition are reversible processes the self-stabilisation of the duplex is purely thermodynamic. Self-replication is by definition a kinetic process and therefore cannot be used in the case presented above. The authors tried to overcome this problem by describing the system behaviour as self-duplication, however without providing a solid definition for this term.

Von Kiedrowski realised that in order to fully control the processes in the system it has to be as simple as possible. Firstly, the author stepped back from using nucelobase recognition and secondly used a simple bond forming step between an aldehyde and an amine. The recognition motif was the amidinium–carboxylate salt bridge (Figure 48).
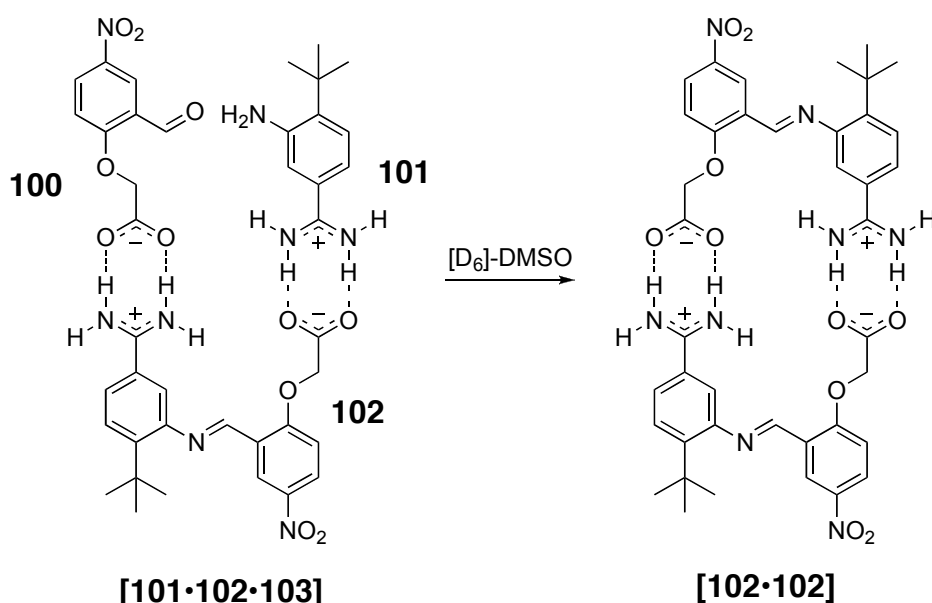
Figure 48. Formation of a template duplex **[102·102]** from the ternary complex **[101·102·103]** using imine formation and utilising the recognition motif between amidinium and carboxylate ions.

The main proof of the self-replicating mechanism is again the template effect, when there was more preformed imine **102** in solution the faster was the initial rate of the reaction between **100** and **101**. The above attempt based on the assumption that simple, specifically designed synthetic systems are a necessary tool for the understanding of more complex phenomena.

Wang and Sutherland also did not use base pairing as recognition in their carefully designed[210] self-replicating system (Figure 49). The recognition based on a three hydrogen bond motif between pyridone **103** and naphthyridine **104** derivatives. The connecting reaction was a [4+2] Diels-Alder cycloaddition. The process was studied using NMR spectroscopy and the rate profile obtained clearly depicted sigmoidal character, which was indicative of autocatalysis. The addition of as little as 0.05 equivalents of the preformed template **105** resulted in a disappearance of the lag period, which confirmed that the reaction proceeded through a recognition-mediated pathway. Kinetically this is a successful system, but there remain some unanswered questions regarding the stereochemistry of the process. The authors provide no evidence if only one of the isomers acts as a template or is crosscatalysis between the diastereoisomers possible.
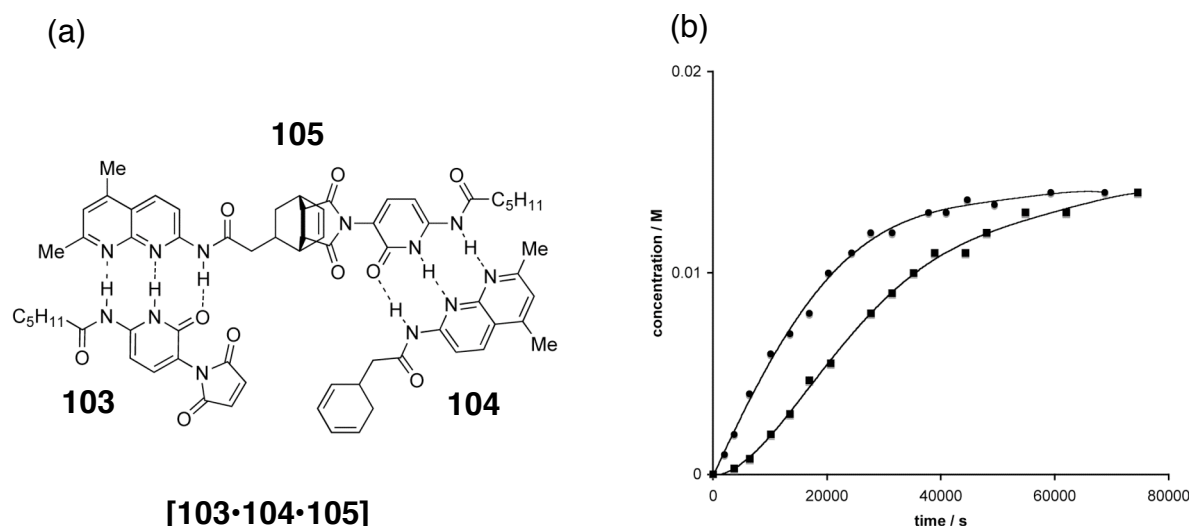
Figure 49. (a) Structure of the ternary **[103·104·105]** complex of a self-replicating system based on the Diels-Alder reaction and utilising triple hydrogen bond motif recognition. (b) The graph follows the formation of the cycloadduct **105** without (squares) and with 10 mol% of preformed template **105** added at the start of the experiment (circles).

Von Kiedrowski conducted[130] a mechanistic and a stereochemical study on a variant of the Wang and Sutherland system. The authors have shown that only one of the four potential diastereomers **106** is capable of both autocatalysis and cross catalysis. The reaction of the racemic mixture showed all the characteristics of a self-replicator and NMR spectroscopy as well as X-ray crystallography proved the formation of only one diastereoisomer. The enantiomers of the diene were separated and each of them was characterised separately for its catalytic behaviour. Both of the enantiomers of **106** display template effect when either of the enantiopure templates is added to the reactions indicating crosscatalysis between them (Figure 50).
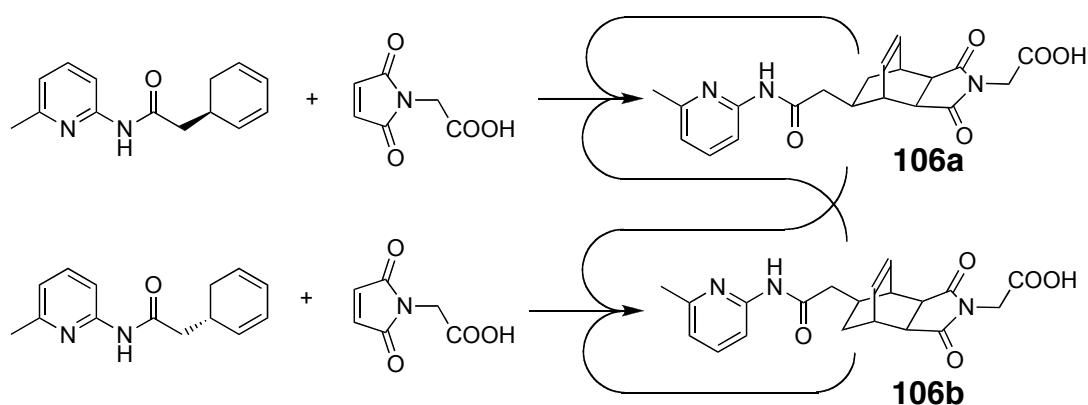

Figure 50. The products **106** both display auto- and cross- catalytic template effects.

Philp and coworkers contributed to the study of synthetic self-replicating systems. Their research focused not only on the self-replicating pathway, but also on the AB complex reaction and general considerations of the recognition mediated reactions.

One of the first self-replicating systems that emerged[211] from our laboratory employed the same recognition motif as above and a cycloaddition reaction between a maleimide **78** and an azide **107** (Figure 51). This moderately efficient replicating system displays both recognition-mediated rate acceleration and a strong template effect.
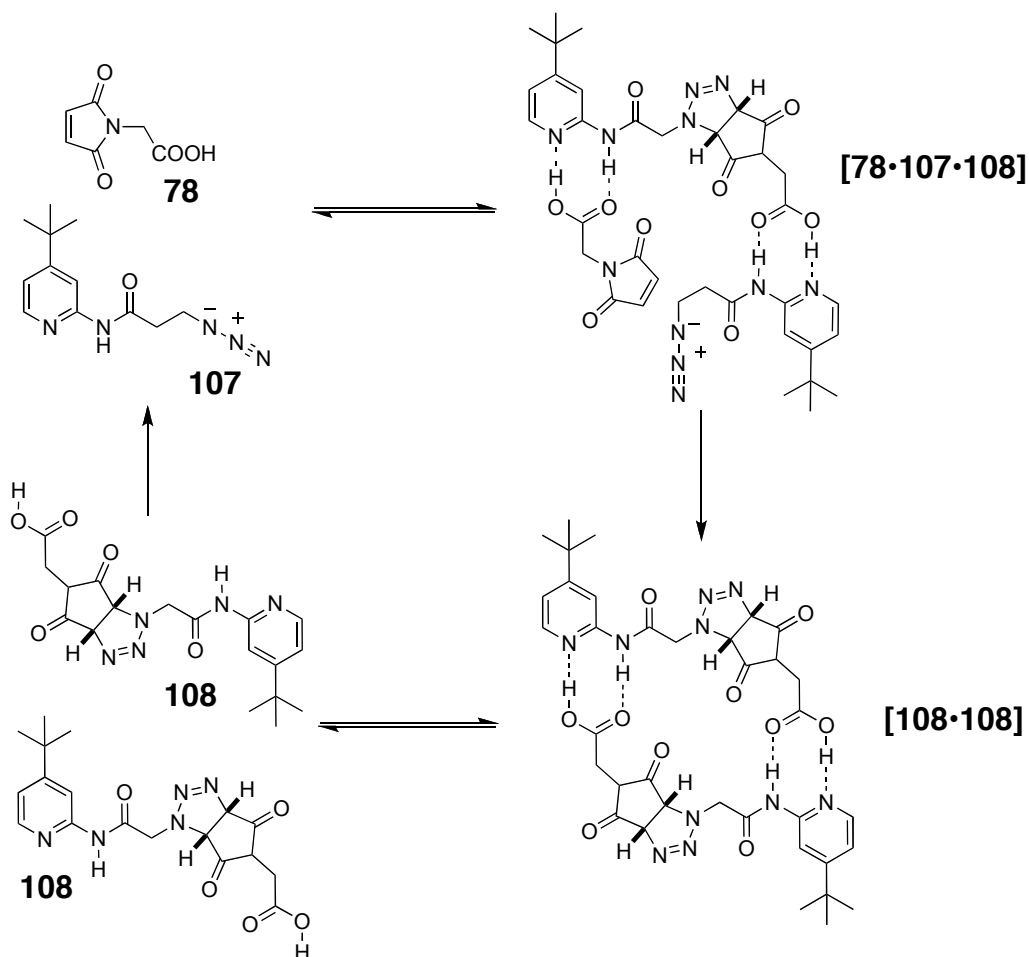

Figure 51. Synthetic self-replicating system based on azide-maleimide cycloaddition utilising carboxylic acid and amidopyridine hydrogen bond recognition motif.

The low efficiency of the system was explained by the product duplex **[108·108]** being very stable, disabling the release of free template **108** into solution. Philp and coworkers set out to modify the system in order to overcome these problems. The chosen recognition motif was preserved in the future systems, but the biggest challenge was to design the appropriate geometry of the template to achieve the highest efficiency. Molecular modelling techniques were an often-used tool in the group research.

This effort led to the development[212] of a system based on a reaction between a maleimide **78** and a nitrone **109** (Figure 52). The interesting aspect of this reaction is

the different possible diastereoisomeric product outcome. The hope was that, thanks to the template directed replication, a stereochemical information transfer would be possible to the next generation of molecules. This assumption proved correct since the *trans* diastereomer of the isoxazolidine product **110** was amplified much more than the *cis* when comparing the recognition enabled reaction to the control. The formation of *trans* product displayed the characteristic sigmoidal shape and showed a template effect. Neither of these could be observed for the *cis* product. The slight enhancement in the formation of the *cis* isoxazolidine is explained by the involvement of the AB complex pathway.
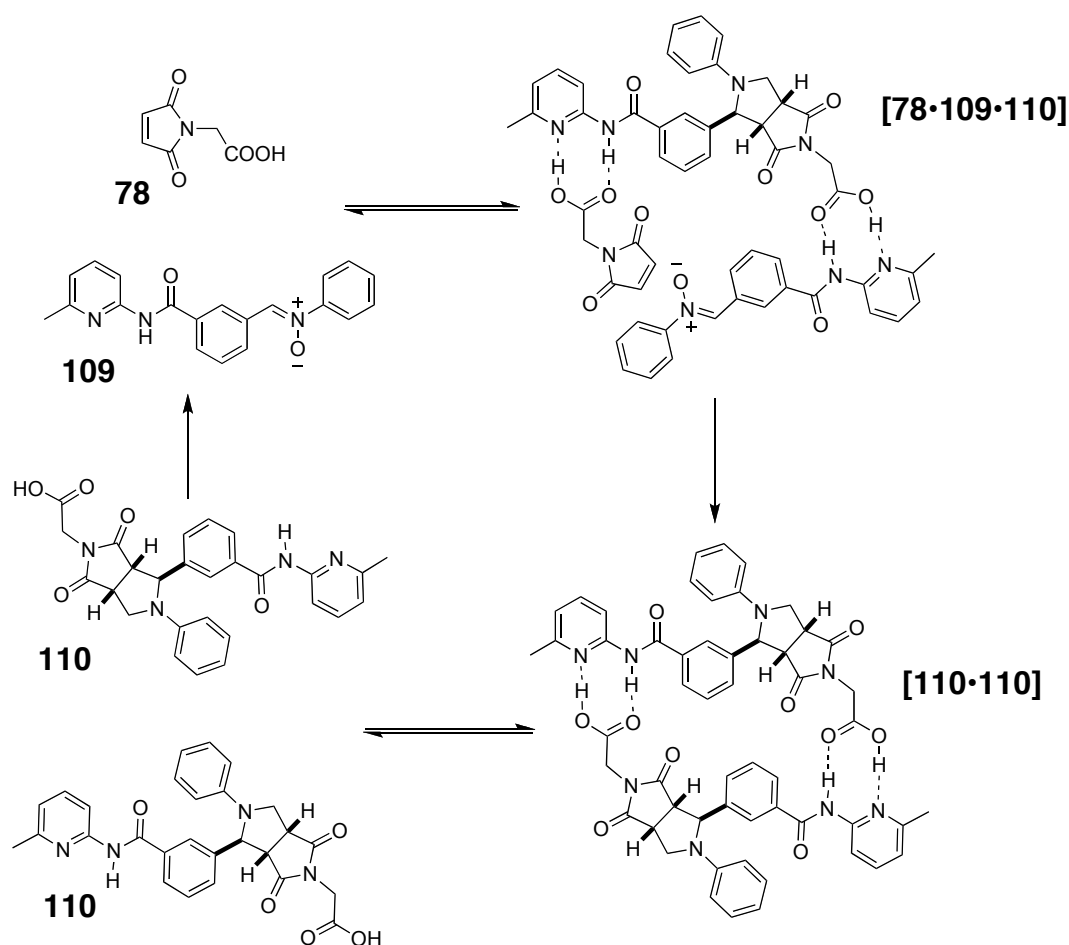

Figure 52. A synthetic self-replicating system based on 1,3-dipolar cycloaddition reaction between a nitrone **109** and a maleimide **78** utilising carboxylic acid-amidopyridine recognition motif.

A related efficient synthetic system (Figure 53) was developed[213] by Philp and Kassianidis. The nitrone and the maleimide can also react giving two diastereoisomers (*cis* and *trans*). Electronic structure calculations revealed that only the *trans* isomer is in a form of an open template that is capable of associating the starting materials in a correct configuration for self-replicating to occur. The *cis* isomer on the other hand displays closed template geometry. This observation

implied that one of the isomers would be amplified *via* a self-replicating pathway and the other *via* the AB pathway.

When a control reaction was run with the recognition disabled the ratio of the isomers was 3:1 in favour of the *trans* and the conversion was 9%. If the reaction was run in the same conditions with the recognition enabled the selectivity rose dramatically in favour of the *trans* to 115:1 and the conversion was greater than 85%. A characteristic sigmoidal shape of the profile of formation was clearly observed, which disappeared after doping the reaction with 10 mol% of preformed template. The presence of template in solution from the start of the reaction allows for maximum autocatalytic rate to be achieved immediately so no lag period is observed.
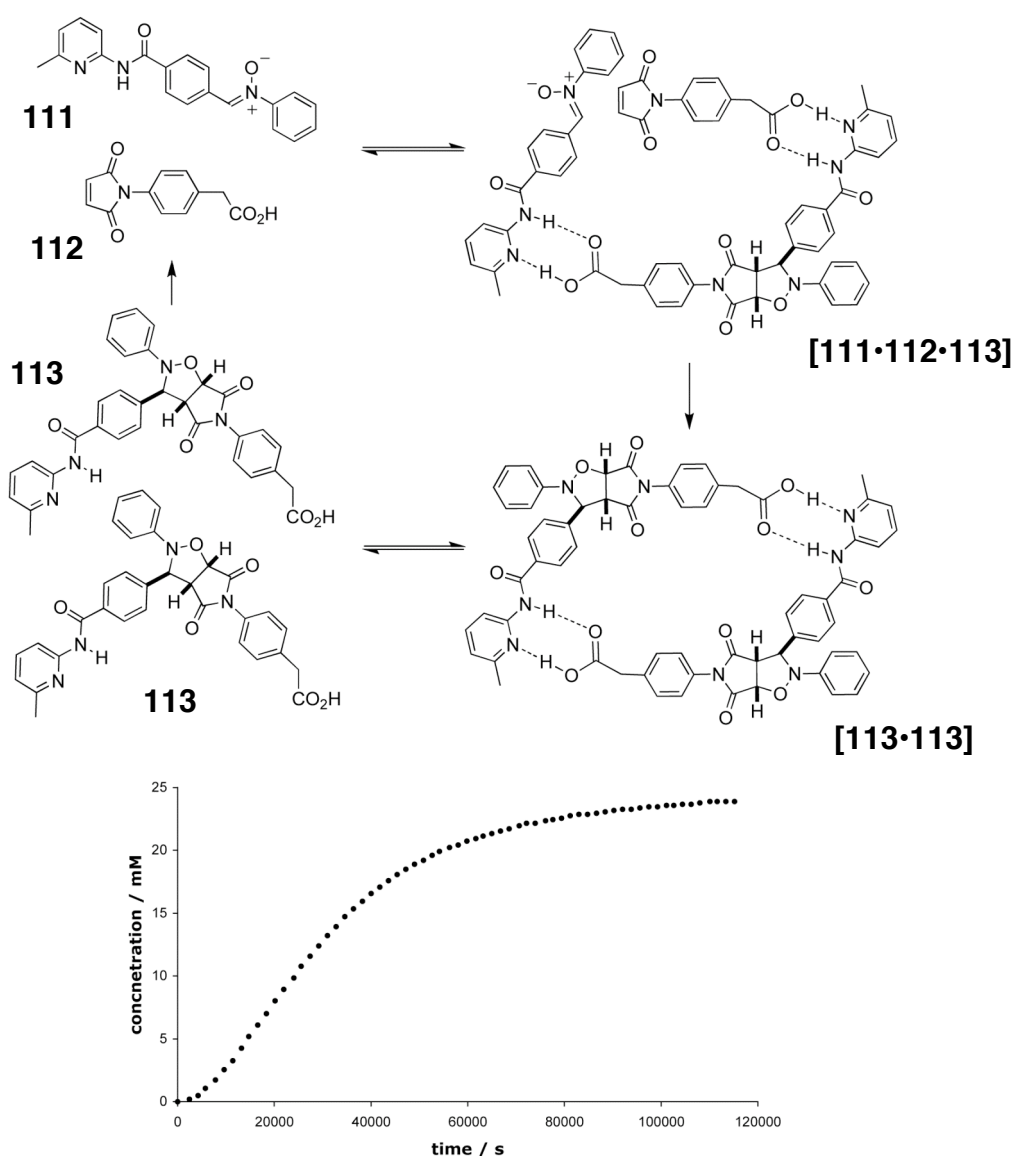

Figure 53. A self-replicating cycle of the *trans* diastereoisomer **113**. The graph follows the formation of *trans*-**113** over the course of the reaction in CDCl$_3$ at –10 ºC displaying a characteristic sigmoidal shape of reaction profile.

58

This system has become the base for further research in our group and was used for the incorporation of minimal self-replicating systems into more complex networks and will be mentioned numerously in the results and discussion section of this work.

A different approach to self-replication balances between natural and synthetic systems. Work developed[214] by Ghadiri employs the use of polypeptides as self-replicating entities (Figure 54). Unlike nucleic acids polypeptides are not linearly self-complementary. In order to utilise this kind of structures for self-replication one has to make use of their secondary structure. Ghadiri based his system on a simple coiled coil folding motif in order to achieve the interactions necessary to create a stable template. The peptide used was a 32-residue polypeptide similar to the yeast transcription factor GCN4.
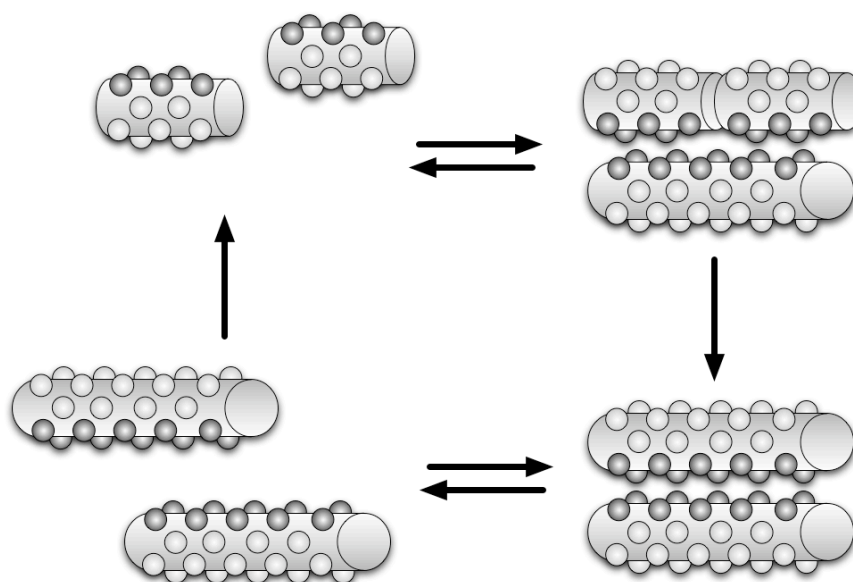

Figure 54. A peptide-based self-replicating system. The cylinders represent recognition-capable coiled coil secondary polypeptide structures.

The replication had to be performed at specific salt concentration and pH so that the secondary structure would be conserved. Kent ligation (Figure 55) was used[215] for fragment coupling. The reaction proceeds *via* an initial trans-thioesterification between C-terminal thioester of electrophile polypeptide and the N-terminal cysteine sulfhydryl group of nucleophile polypeptide. The resulting intermediate rapidly rearranges *via* intramolecular S to N acyl transfer to give the copy of the template.
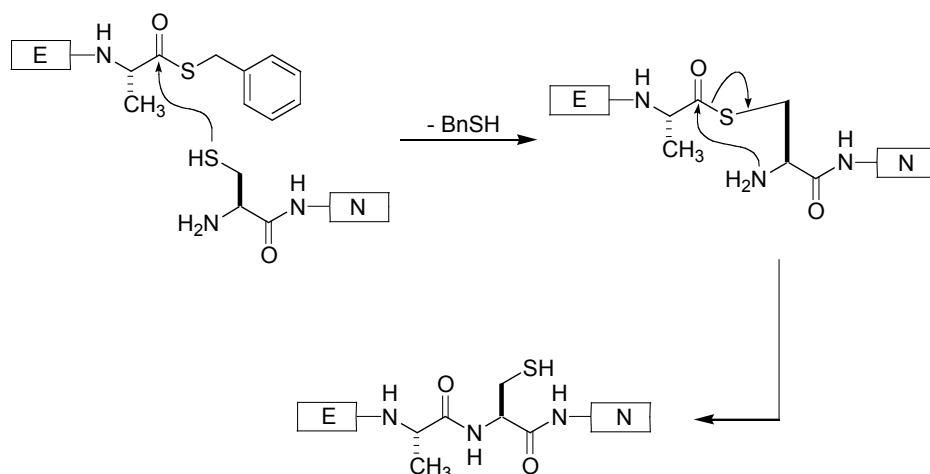
Figure 55. A scheme of Kent legation used as the connecting reaction in peptide-based self-replicating systems.

Chmielewski also developed peptide based self-replicating systems focusing on the improvement of replication efficiency. The main idea was to destabilise the template duplex. One of the approaches[216] was to shorten the polypeptide length weakening the interactions allowing for a better turnover of the cycle. Another way to achieve the same goal was to change the geometry of the template. Introduction of a proline kink in the polypeptide coil proved to enhance replication efficiency (Figure 56).
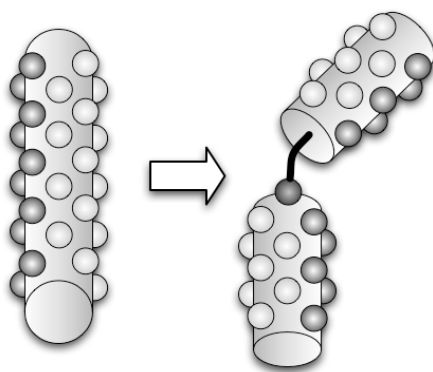

Figure 56. The effect of proline substitution in the replicating peptide-based template. The resulting 'kink' destabilises the template duplex enhancing replication.

It was Ghadiri who started incorporating the self-replicating peptides into more complex systems introducing cross-catalysis and hypercyclic networks. These systems will be described in the next section.

1.12 Logic operations in chemical systems

Logic is an inseparable part of chemistry. The design of every synthesis is based on logical assumptions that a chosen reaction will give the desired product from a set of substrates. Chemists try to choose a reaction that will give the highest selectivity and yield enhancing the chance for success. Usually it is the product of such reaction that is of scientific interest and not the logic itself, this approach however is changing. Molecular logic[217-218] has become an area of interest in the last years and found its way into systems chemistry. In order to create an evolving complex system some degree of logic has to be engendered into the design. The rational basis for such approach is the obvious fact[219] that in living organisms all data is stored, transferred and processed by chemical means. It has already been discussed how single molecules and assemblies as well as mixtures respond to inputs such as presence of analytes or changes in the environment. The output that can be read as the change of properties of a given sensor is of logical nature. The idea of a chemical computer has been around for a number of years although only recently the development of supramolecular chemistry and analytical methodology made rapid advances in this area possible.

The first comparison of chemical logic devices that needs to be made is to the modern semiconductor based electronics. It has to be stressed that molecular logic does not necessarily have to follow the semiconductor paradigm. One issue is the nature of addressing the device. In the semiconductor scenario the input is an electrical signal send along a wire. An electrochemical setup can be addressed this way but usually in a chemical counterpart the input is a molecule or atom hitting the device by diffusion in a solution. Another often used input is fluorescent light which can be transmitted from a distance simply through the medium. Photochemistry is very appealing for the design[220] of chemical logic gates, firstly for the ease of analysis and secondly for the broad knowledge of the components that feature either photochromic or quenching properties. It is a promising idea as modern data transfer is largely based on optical signals transferred *via* optic fibres over long distances.

The nature of the input and output raises the question about connectivity of the devices. Often the two signals are of a different nature making it difficult to arrange linearly single molecular devices in order for them to perform complex operations. Finally, there is an issue about the stability of the molecular devices that so far is

miniscule compared to the longevity of the silicone based electronics and incorporation of the chemical structures into efficient large-scale manufacturing processes.

The synthesis of complex structures with multiple modules each with a specific function is now possible opening the route for the development of molecular logic. The main synthetic targets are molecular switches[221-222] which convert input stimuli into output signals. This function is usually incorporated into a single molecule[223] or a supramolecular assembly (e.g. switchable rotaxanes[224]) that will change its state under appropriate conditions that indicate an input and the properties of the new state will be read as output. The switches themselves cannot be described as systems, they are, however components of systems to be and this possibility has to be included in their design. The binary logic performed by these chemical entities[225] reflects the Boolean functions[226] of traditional logic gates used in electronics (Figure 57). The operations can be described by simple ones and zeros. A one is assigned to the system in which we can observe a change in conditions or properties, a zero describes no input or no change in observable parameters of the system. The ones and zeros can be incorporated into truth tables, which are specific to every Boolean function.
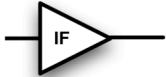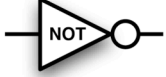
GATE	INPUTS			
	0,0	1,0	0,1	1,1
IF	0	1		
NOT	1	0		
OR	0	1	1	1
AND	0	0	0	1
NOR	1	0	0	0
NAND	1	1	1	0
XOR	0	1	1	0
XNOR	1	0	0	1

Figure 57. Summary of the two one-input logic gates and six two-input logic gates, showing their truth tables and symbols.

De Silva and coworkers gave[227] an example of a single molecule photochemistry-based AND logic gate (Figure 58). The specific molecule **114** comprises three functional units separated by spacers. At one end of the molecule is an aromatic anthracenyl fluorophore ring system that would normally fluoresce blue when it is exposed to ultraviolet light. In this system, however, the fluorescence is quenched by photoinduced electron transfer (PET), in which an electron is transferred to the anthracenyl group from either the nitrogen or oxygen atoms in the other parts of the molecule. However, both of these groups can be specifically addressed: the amine group can act as a receptor for hydrogen ions and the crown ether can be used to capture sodium cations. If only one of the two receptor sites is occupied PET will still

occur from the other, and still no fluorescence will be observed. If, however, both receptor sites are filled, that is both ion inputs are present, PET is prevented because the electrons are involved in ion binding and strong fluorescence is observed from the anthracenyl. The overall result is AND logic, with output of one (fluorescence) only occurring when both inputs are present. The nature of the recognition sites and the fluorophore can be modified to create similar systems that respond to different inputs and give different outputs.
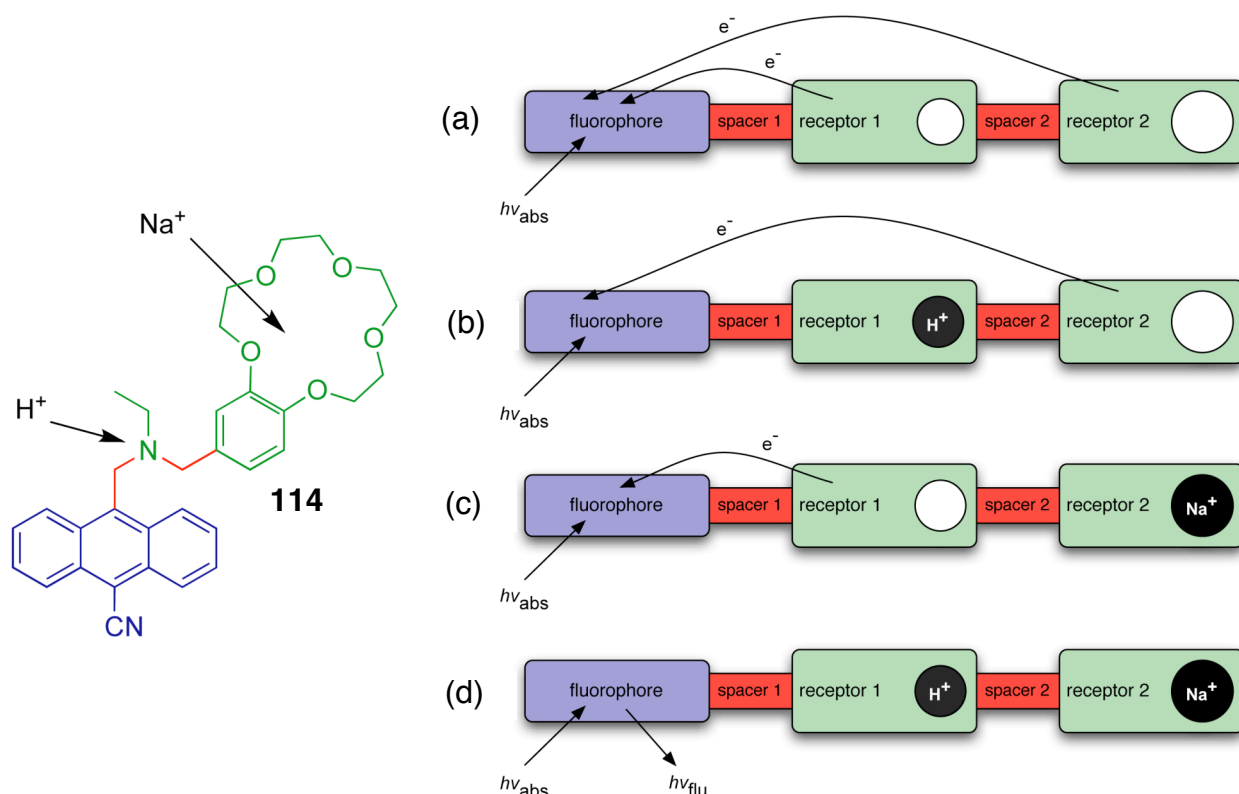

Figure 58. A molecular logic gate based on photoinduced electron transfer. Two chemical receptors (green): H^+ binding amine and Na^+ binding benzocrown ether are connected by hydrocarbon spacers (red) to each other and cyanoanthracene dye (purple). The dye is capable of absorbing and emitting light. (a) The system has two possible pathways of transferring an electron to the excited fluorophore - one from each of the unoccupied receptors - resulting in quenching of the fluorescence. (b,c) Occupation of only one of the receptors still results in quenching of the fluorescence. (d) Fluorescence results only if both receptors are occupied.

Mechanically interlocked molecules are perfect candidates to perform switching operations through their translational motion. The most prominent development in the integration of [2]-rotaxanes into working memory devices was done[228] by Dichtel, Heath and Stoddart (Figure 59). The molecule **115** can exist in two distinct states without changing its composition by reversibly shifting the macrocycle from one recognition site to another. These bistable rotaxanes have been integrated into

nanowire crossbar devices opening a route to memory devices of unprecedented density.
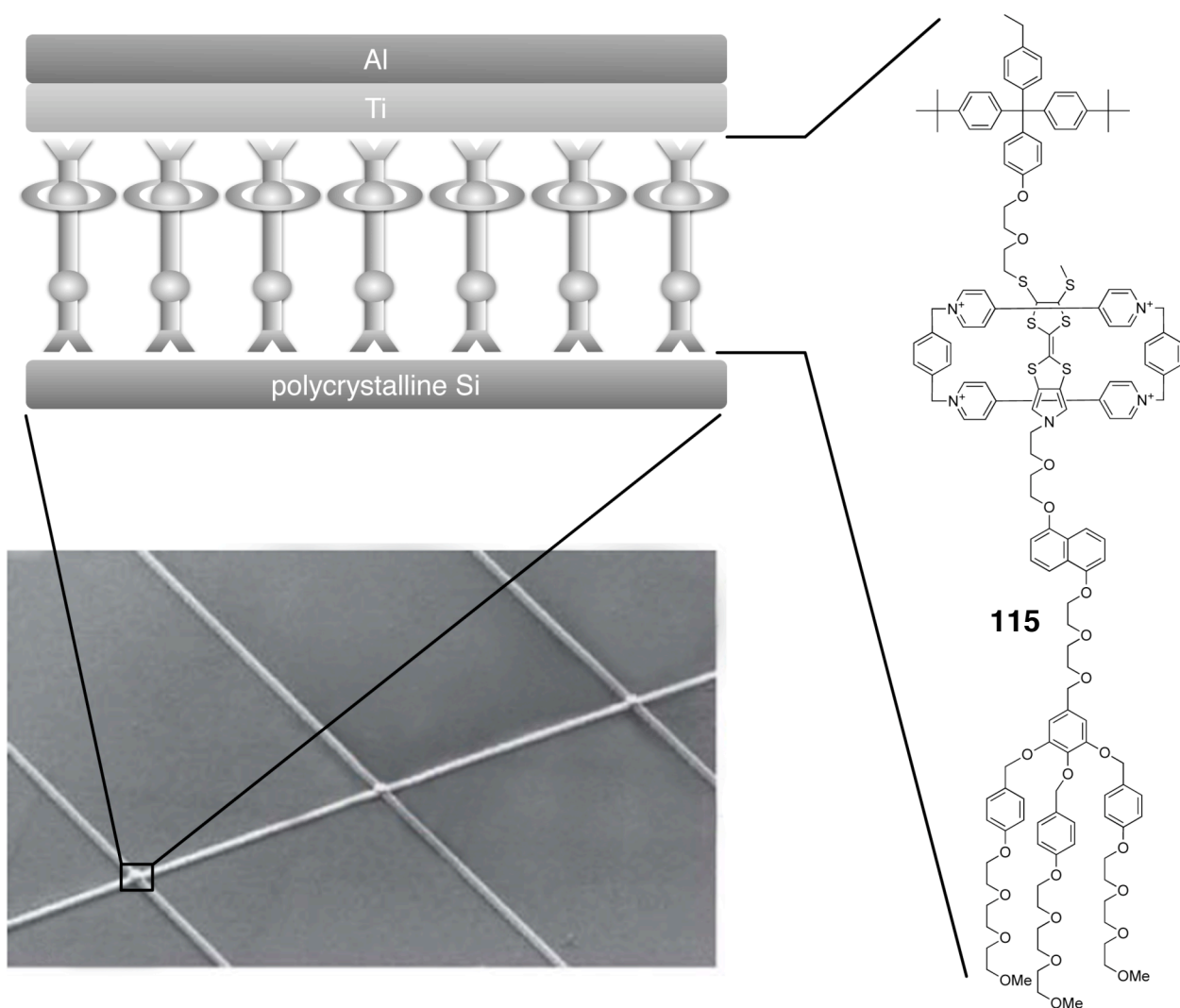
Figure 59. A bistable [2]rotaxane **115** was is incorporated into a crossbar nanowire device. The scanning electron microscope image shows the crossing point being 7000 nm^2 incorporating *ca.* 500 molecules of **115** between the wires. Figure adapted from ref. 228.

The macrocycle in this particular rotaxane is the previously described cyclobis(paraquat-p-phenylene) with its two π-electron-deficient bipyridinium units and the thread containing π-election-rich units: tetrathiafulvalene (TTF) and 1,5-dioxynaphthalene (DNP). In the ground state, when the input is zero, the ring preferably binds to the TTF unit in a 9:1 ratio. This ratio can be influenced by electrochemically oxidising the TTF to a cation. The coulombic repulsion between the positive charges of TTF and paraquat make this position less favourable. When the oxidation is considered an input then the output of the switch is the movement of the macrocycle to the DNP station. The necessary step for the incorporation of such

molecules into electronic devices is the possibility of adsorbing them onto a surface without jeopardizing their switching properties. It proved possible and a crossbar device incorporating the bistable rotaxane **115** was constructed. Depending on which recognition site was occupied by the macrocycle the two states reflected the high and low conductivity, this property was used to write, read and erase a program code. This first demonstration of a working ultradense memory indicates that these devices may be successfully scaled to junctions containing a very small number of molecules. Although the rotaxane-based memory is an outstanding achievement, a straightforward swap of voltage based bulk semiconductors for the 'soft' chemical counterparts is still impossible. The architecture of the described device is still similar to the semiconductor used generally in electronics, it is only the nature of the switching that is chemical. The next step is to devise a new approach where a chemical system as a whole would act as logic device.

The first efforts in this area reach for the well characterised biomolecules abundant in Nature. A whole array of DNA[229-230] and related ribozyme[231] based logic systems have been developed. These systems use the recognition properties of the biological information carriers. The logic bases on the fact that a certain DNA fragment will hybridise with its complementary partner and not another random oligonucleotide.

The idea of using DNA for computing was first crafted[232] by Adleman in 1994. His brilliant insight and initial implementation of DNA computation focused on a version of the 'travelling salesman' problem (a Hamiltonian Path problem) — the sequences of individual DNA strands represent cities and the recognition between them codes the paths between cities. When strands hybridise, it represents the joining of two cities by a path. When an 'answer' of the correct length assembles, it represents the solution to the problem of how to pass through each city exactly once (Figure 60).
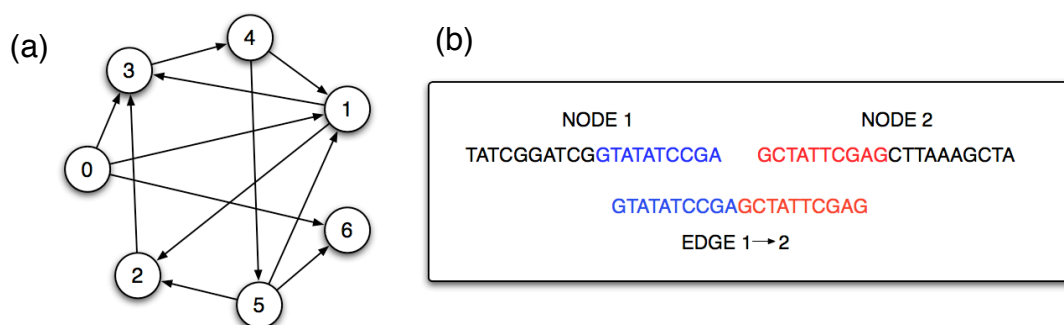

Figure 60. (a) A directed graph with a unique Hamiltonian path from 0 to 6. (b) Encoding the graph in DNA. For each node a random 20-mer oligonucleotide is generated for the path from node 1 to 2 another 20-mer is generated utilising 3' 10-mer of node 1 and 5' 10-mer of node 2.

Even though the experiment was a success not all possible paths emerged from the DNA interactions. A silicone based computer would solve this problem fully in less than a nanosecond. There is however a foreseeable advantage of DNA based computing and that is parallelism not possible in conventional linear semiconductor logic. Unfortunately to make use of this property and outperform classical computers it would take an incredible amount of oligonucleotide strands, and the more DNA fragments the higher possibility of error.

More advantages could come from combining the DNA logic with biological properties of these molecules. A strand carrying an input sequence could recognise a complementary part of the more complex specie triggering a change of biological properties. Such a response through, for example, a change of conformation can lead to engagement of a biochemical pathway or overexpression of a chosen metabolite. Such systems apart from being focused on performing logic for the use in molecular computing are also designed for therapeutic purposes. An idea that as an input the system would interpret disease markers from a cell, analyse them and as an output deliver a therapeutic agent is highly attractive. This function indicated that such systems would have to operate in solution under physiological conditions, putting them farther away from conventional electronics.

An extensive study into deoxyribozyme systems with a future focus on therapeutic purposes has been undertaken[233] by Stojanovic and coworkers. Their long-term goal is to construct macromolecular systems able to enter specific cell types and therein sense multiple molecular markers of diseases. The signals could be analysed to result in a simple binary output, for example, cell death or cell survival. The target molecules for this task are oligonucleotides, which are substrates for deoxyribozyme logic gates consisting of RNA based catalysts[234]. A precise protocol to construct and test a complete set of ribozyme-based logic gates whose self-cleavage activity is modulated by the presence of oligonucleotide units was described[235] by Penchovsky and Breaker. Again, as in the RNA World scenario, when it comes to complex system operation RNA seems to be the molecule of choice. Oligonucleotides are also quite well understood regarding their recognition properties, as well as stability and modes of intercellular delivery. In order to construct a logic gate, a ribozyme of a certain sequence has to be engineered in such a way that it is either activated or deactivated in the presence of a given oligonucleotide (Figure 61). A very useful feature of such systems is that the input and the output are of the same nature. The oligonucleotide

triggers cleavage that produces a different oligonucleotide, which can be further utilised in the next logic operation. Such function allows incorporating a chain of logic gates into more complex systems.
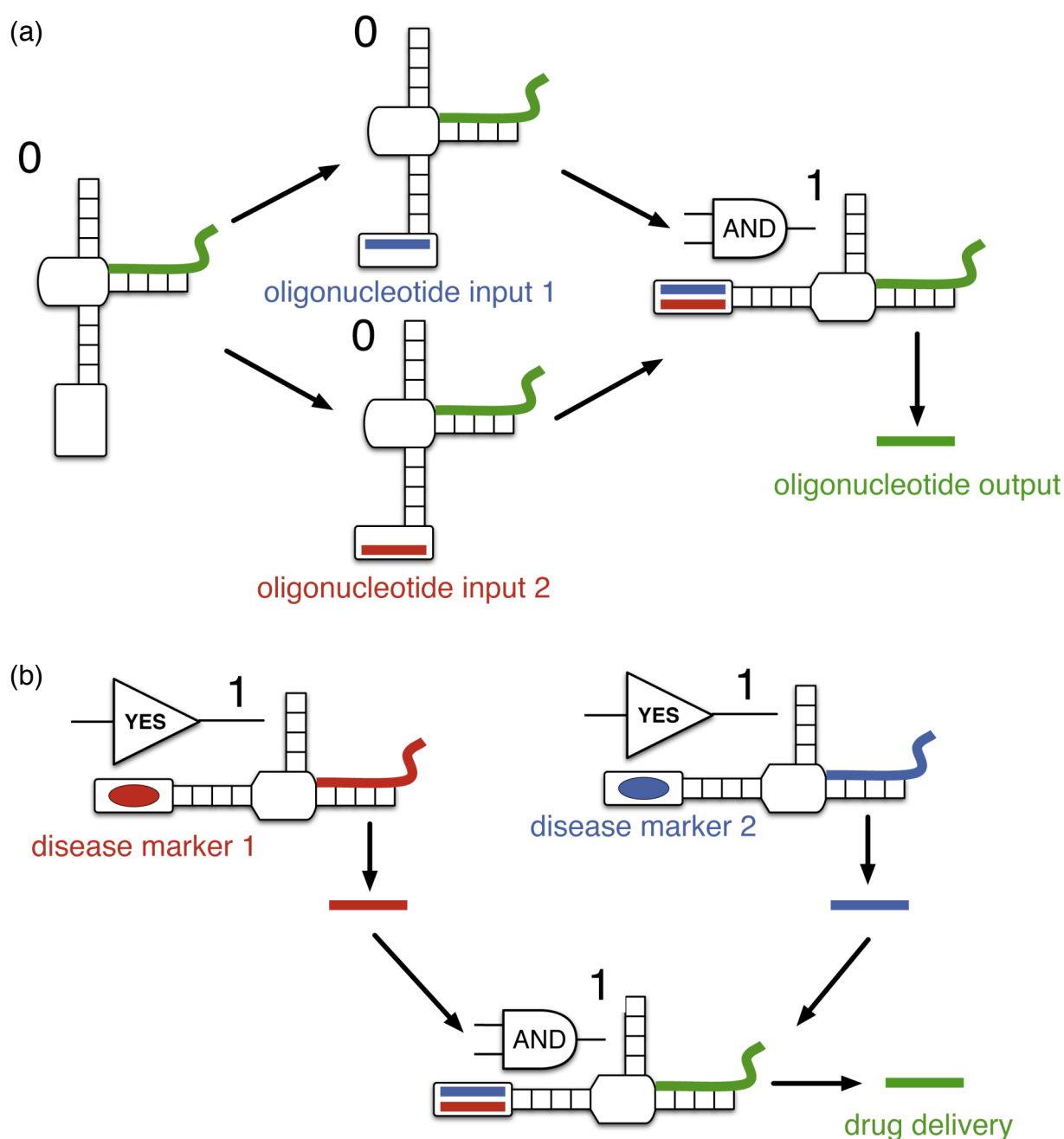
Figure 61. (a) Schematic representation of a ribozyme-based AND logic gate. The AND gate recognises two oligonucleotides 1 (blue) and 2 (red) and has four possible states. Three states exhibit no self-cleavage activity (0), while the state with both oligonucleotides bound displays self-cleavage (1) releasing an output oligonucleotide. (b) Operation of a therapeutic automaton implementing AND logic and analyzing two disease markers. The YES gates recognise two protein disease markers (blue and red ellipses) and become active (1) releasing specific oligonucleotides (blue and red). The oligonucleotides activate the AND gate which releases another oligonucleotide (green) that is capable of activating a downstream drug delivery element.

A more complex system could also incorporate[236] enzymes such as ligases into ribozyme circuits. The ligases could interact with the system to synthesise larger input/output oligonucleotides, opening pathways for new logic units. Deoxyribozyme-based logic was also implemented[237-238] by Stojanovic in an impressive molecular automaton that is capable of simulating a game of tic-tac-toe against a human opponent. An automaton is a device that converts information from one form into another according to a definite procedure. The automaton developed by Stojanovic incorporates 23 molecular gates and one active catalyst in nine wells that correspond to the possible moves on a tic-tac-toe board. Every human move is coded as an input of specific oligonucleotide and the automaton after analysis replies by a change of fluorescence in a given cell. The whole system is designed not to able to lose.

Enzymes themselves are a promising target[239] for the construction of logic gates. The specific substrates or inhibitors of enzymes could be used as inputs to activate or deactivate the enzyme. An example of such approach comprises of a biocatalytic logic gate system created by a combination of two enzymes—horseradish peroxidase and glucose dehydrogenase in conjunction with two chemicals that act as activating signals, namely hydrogen peroxide (input A) and glucose (input B) (Figure 62).
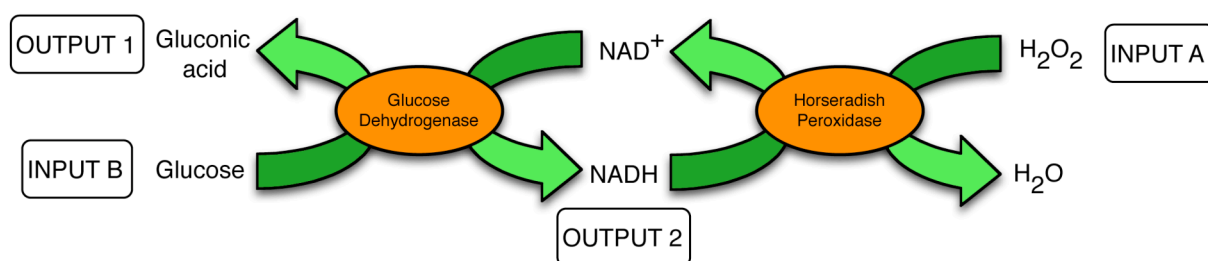
Figure 62. Enzyme-based logic gate constructed with glucose dehydrogenase and horseradish peroxidase. The two enzymes are connected with a NAD^+/NADH reduction/oxidation cycle. The inputs for the system are either hydrogen peroxide or glucose. The monitored outputs are gluconic acid and NADH.

The two enzymes are connected by a NAD^+/NADH reduction/oxidation cycle. The output is a change in absorbance caused by either gluconic acid or NADH and depending on the conditions and which output is considered this system can mimic the AND, NOR and InhibAND logic operations.

Ghadiri combined[240] the recognition properties of DNA with the activity of enzymes. A semisynthetic enzyme was equipped with a DNA recognition fragment that becomes a complementary site for the inhibitor (Figure 63). Such setup allows for

programmable regulation of the enzyme activity. Simply adding the complementary inhibitor turns off the activity. This process can be reversed by adding a strand of DNA that will bind to the enzyme and displace the inhibitor or bind more preferentially to the inhibitor making it impossible to inhibit the enzyme. The two modes of activation can be treated as two inputs of a logic gate making it possible to mimic the AND, OR and NOR logic gate operations. The output is an enhanced production through the enzyme catalysed reaction.
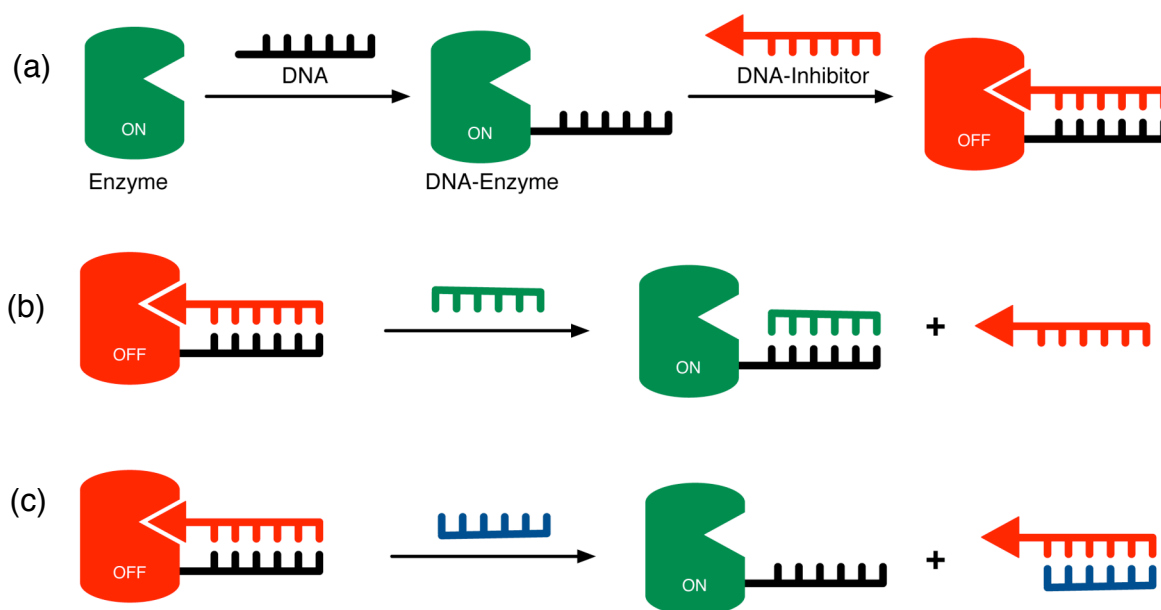

Figure 63. (a) Synthesis of a DNA-Enzyme and inhibition of enzyme activity by a complementary DNA fragment coupled to an inhibitor. (b) The enzyme activity can be restored by displacing the DNA-Inhibitor with an inactive DNA strand the exhibits higher potency for the DNA-Enzyme or c) by a DNA strand binding to the DNA-Inhibitor.

One of the most prominent systems that combine replication, complex behaviour and logic comes again from the work of Ghadiri. The system is based on the self-replicating peptide that has already been discussed in the self-replication section. Before introducing Boolean logic a directed molecular network had to be designed[241]. The design base was a sequence-dependent auto and cross catalysed peptide fragment condensation reaction in an aqueous solution. The system was created using an array of 81 sequence similar 32-residue coiled-coil peptides. Firstly calculated graph architecture was elucidated uncovering 25 nodes joined by 53 vector edges (Figure 64).
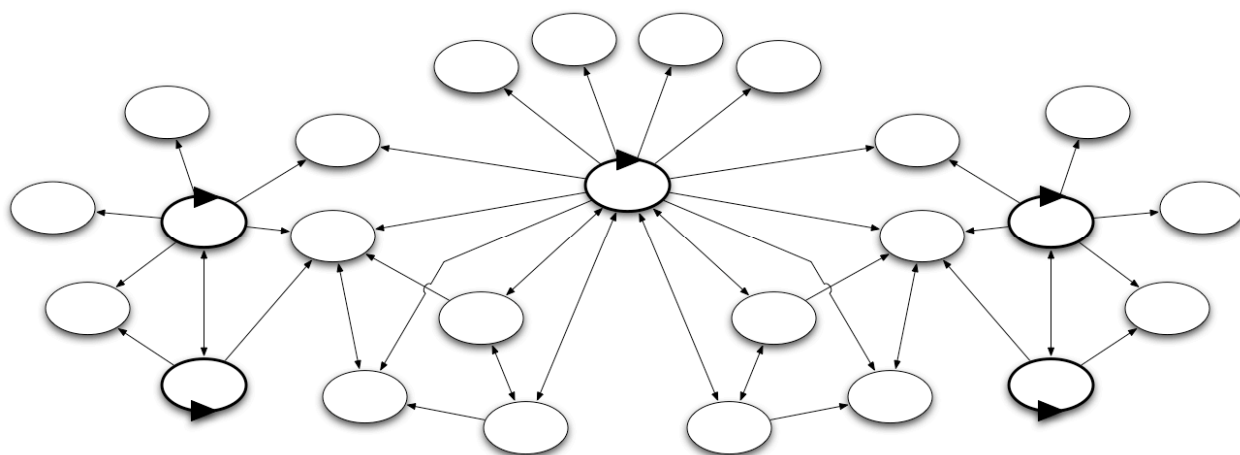
Figure 64. Calculated network architecture illustrating formation of a self-organised peptide network composed of 25 nodes joined by 53 vector edges.

To validate the theoretical calculations nine major nodes were selected, synthesised and their activity tested experimentally (Figure 65). Three of the nodes displayed strong autocatalytic property and a total of 20 cross-catalytic interactions were characterised. Not all pathways were observed in the system and one of the reasons for that is the strong engagement in autocatalysis of some of the nodes, which does not leave any template for the weaker cross-catalytic pathways. When the autocatalytic behaviour was inhibited the weaker pathways were observed as well.
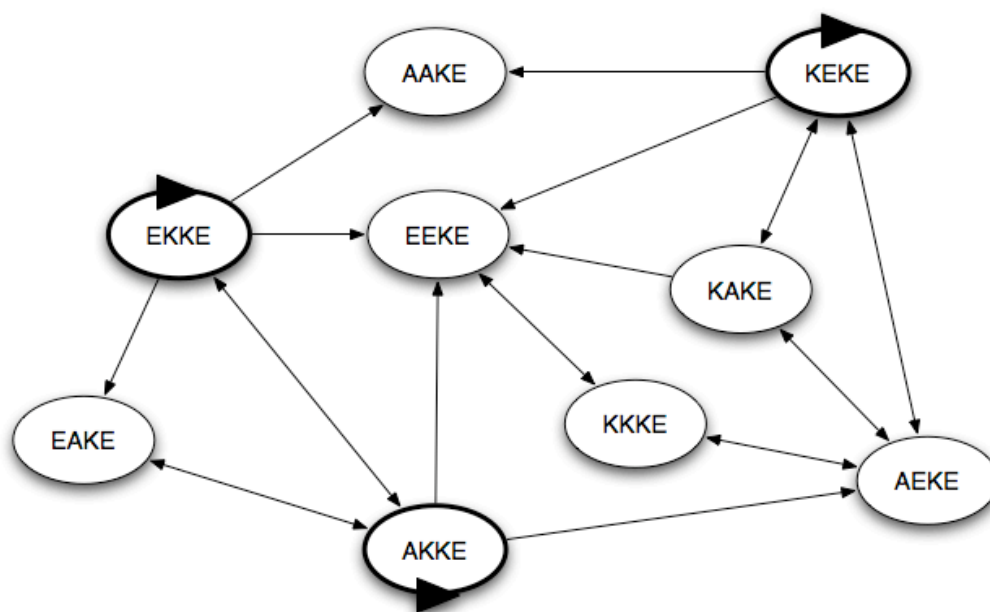

Figure 65. Experimentally derived network architecture representing part of the calculated network (Figure 64). Two sets of experiments were needed to reveal both the strong and weak catalytic pathways.

The system was also seeded with all 9 polypeptide templates (nodes) to see how an initial rate enhancement of the processes utilising the particular templates would be

71

affected. As predicted the enhancement was observed for just certain nodes that are connected to the seeded node and in particular the autocatalytic cycles were clearly enhanced. The study proved that a synthetic chemical system may posses the features of much larger biological system. The connections were characterised by the rate of formation of each of the nodes and the rates were significantly different in the system than in isolation. The possibility of templating the system opens a route to use such assemblies to perform logical operations, where the input would be a specific template. Such behaviour was described[242] by Ashkenasy and Ghadiri using a peptide network to mimic the OR, NOR and NOTIF logic gate operations. A fragment of the network consisting of 5 nodes with well established connections was extracted from the bigger system described above. By carefully choosing which nodes act as inputs and which as outputs various logic operations could be performed on this cut down system.

Philp and coworkers devised a simpler synthetic system with engendered auto and cross catalytic properties that perform logic operations (Figure 66). The system bases on reactions between nitrone **109** and a maleimides **80** and **81** Both reactions have been proven[212] to display self-replicating properties, although the reaction of maleimide **81** leading to the formation of template **116** it is less efficient. Having two self-replicating systems with very similar template structures opened a possibility for cross catalysis. Investigation proved that template **116** is capable of cross catalyzing the system based on template **110** even more efficiently than **110** can catalyse its own formation. The original template **110** however does not cross catalyse the new system based on **116**, but even inhibits its autocatalytic properties. This surprising result has been explained using electronic structure calculations that uncovered a subtle unfavourable conformational strain that disables the template effect. Since the two templates display different catalytic properties, they could be used as inputs for a chemical logic gate. The replicators are of similar efficiency so in the absence of any added template no selectivity was observed, which was interpreted as an output of zero. Addition of the template **116**, **110**, or both, resulted in the amplification of the template **110** over **116** (output equals one) either through auto or cross catalysis. Such behaviour can be interpreted as an OR logic gate.
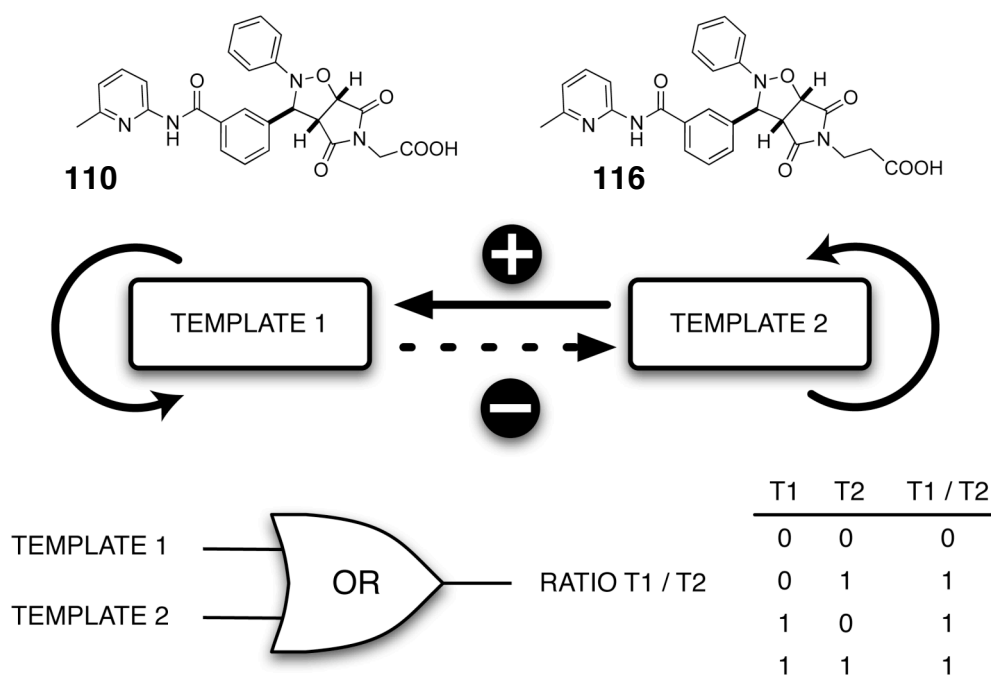
Figure 66. An OR logic gate system comprising of two self-replicating templates **110** and **116**. Since Template 2 **116** cross-catalyses Template 1 **110** (+) and Template 1 cross-inhibits Template 2 (-) addition of any or both templates results in a higher formation of Template 1. If the ratio of Template 1 to Template 2 is more than 1 then the output of the system is positive.

Engendering logic into whole systems instead of just single molecular switches can prove much more versatile allowing combinations of different inputs and outputs depending on what operation we want the system to perform. The above examples truly encompass the ideas of system chemistry, combining complex mixtures of interdependent species and logical, responsive operations.

Defining systems chemistry is not an easy task. The origins of the idea are clearly rooted in the search for early life and the chemistry associated with it. The next step was embracing the complexity idea by chemists, which led to a new approach to mixtures and their properties. The development of advanced analytical techniques made this insight possible, and the incredible examples of molecular machines and nanostructures proved that the bottom up approach to very subtle chemical problems is possible. Dynamic combinatorial chemistry has become a tool to create model systems with complex emergent properties. Ways of amplifying specific structures or properties from mixtures are developed both based on thermodynamics and molecular recognition as well as kinetics and self-replication. Finally incorporating logic into chemical structures and whole systems is seen as the proof of ultimate

control. The future applications range from nanotechnology, intelligent drugs, molecular computers, to maybe one day answering the ultimate question of life.

1.13 Aims and objectives

Complex systems are associated with many aspects of research. From the examples described in the previous Sections, it is apparent that chemistry is embracing the idea of complex systems and networks utilising it to search for the origin of life, synthesise life itself or form sophisticated nanostructures. From this basis a new area of systems chemistry emerged that attempts to capture the complexity and emergent phenomena of synthetic chemical systems. Systems chemistry is still not well defined and sometimes may seem to deal with phenomena hard to explain and understand. In order to develop this area we have to provide research that reflects the ideas of systems chemistry using simple, easy to analyse components that give clear results. The aim of this work is to show it is possible to combine self-replication, dynamic covalent chemistry and chemical logic operations in a minimal complexity scenario. The building blocks for the networks have to be facile to synthesise and easy to monitor. The emergent properties of the systems, on the other hand, display phenomena such as self-sorting, logic operations and kinetic selection from a dynamic combinatorial library. This goal is achieved using just two types of chemical reactions and one type of recognition motif incorporated into a reaction which product displays self-replicating properties. The complexity does not lie in the molecules themselves, but rather in careful design and the insight into subtle interconnections that are responsible for the emergent properties of the systems.

2. Length segregation in chemical networks

2.1 Integrating length segregation in self-replication design

The first aim of the project was to identify a replicating system that would be efficient, robust, and easily modified in order to incorporate it into a more complex network. The design of the replicator was based on the systems previously developed in our laboratory. The experience in artificial replication research within the group ensured that a useful system could be identified, which would exhibit high efficiency, synthetic simplicity and reproducibility of results.

A minimal model of self-replication (Figure 67) consists of three reaction channels. Two of these pathways display recognition-mediated enhancement, but only one pathway is autocatalytic – the replicator. In order to observe the highest selectivity the reaction has to be designed in such a way as to direct all starting material through a single channel. The structure of the substrates and the conditions of the procedure should be chosen in such a way, as to eliminate undesired pathways.
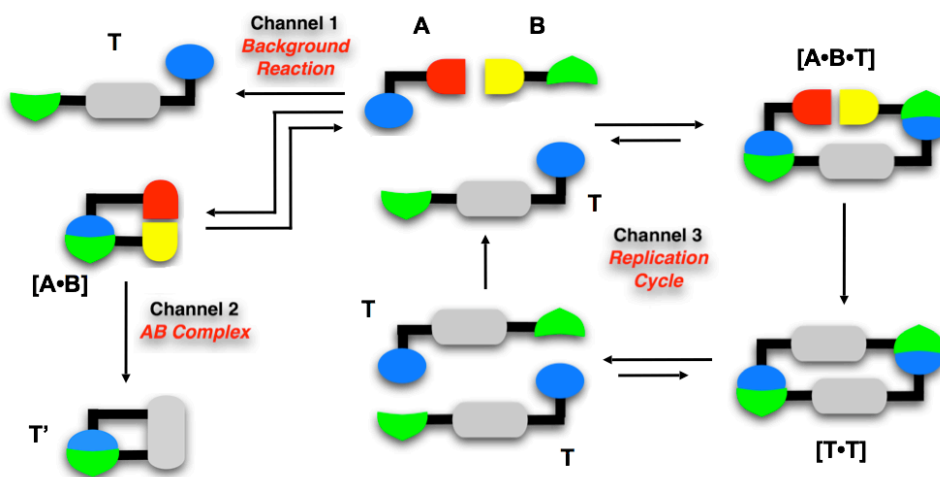
Figure 67. The minimal model of self-replication. Reagents A and B can react through three pathways – an uncatalysed bimolecular background reaction, a recognition-mediated pseudounimolecular pathway mediated by a binary complex [A•B] and a recognition-mediated pseudounimolecular autocatalytic cycle mediated by a ternary complex [A•B•T]

The bimolecular pathway, unmediated by recognition, is slow and unselective. This pathway is always present to some extent in the system, since it bases on random collision of reacting molecules. The involvement of this reaction can only be influenced by changes in concentration and temperature. The lowering of either of

the parameters will slow down the bimolecular channel to a much greater extent than any of the recognition-mediated pathways.

The most complex and thus most difficult to design is the replicator. In order to direct the reaction *via* the replicating cycle the chosen conditions must suppress the bimolecular pathway but still allow for efficient autocatalytic turnover of the replicator. The elimination of the AB complex channel can be achieved by careful design of the building blocks. In the AB complex the building blocks upon recognition bring the reactive sites into close proximity facilitating the reaction. An increase in rate is observed since the reaction is now pseudounimolecular. The simplest way to disable this reaction pathway is to use substrates with sufficiently different lengths of linkers connecting the recognition and the reactive sites. Upon recognition the reactive sites will be too far away to stabilise the transition state (Figure 68a).
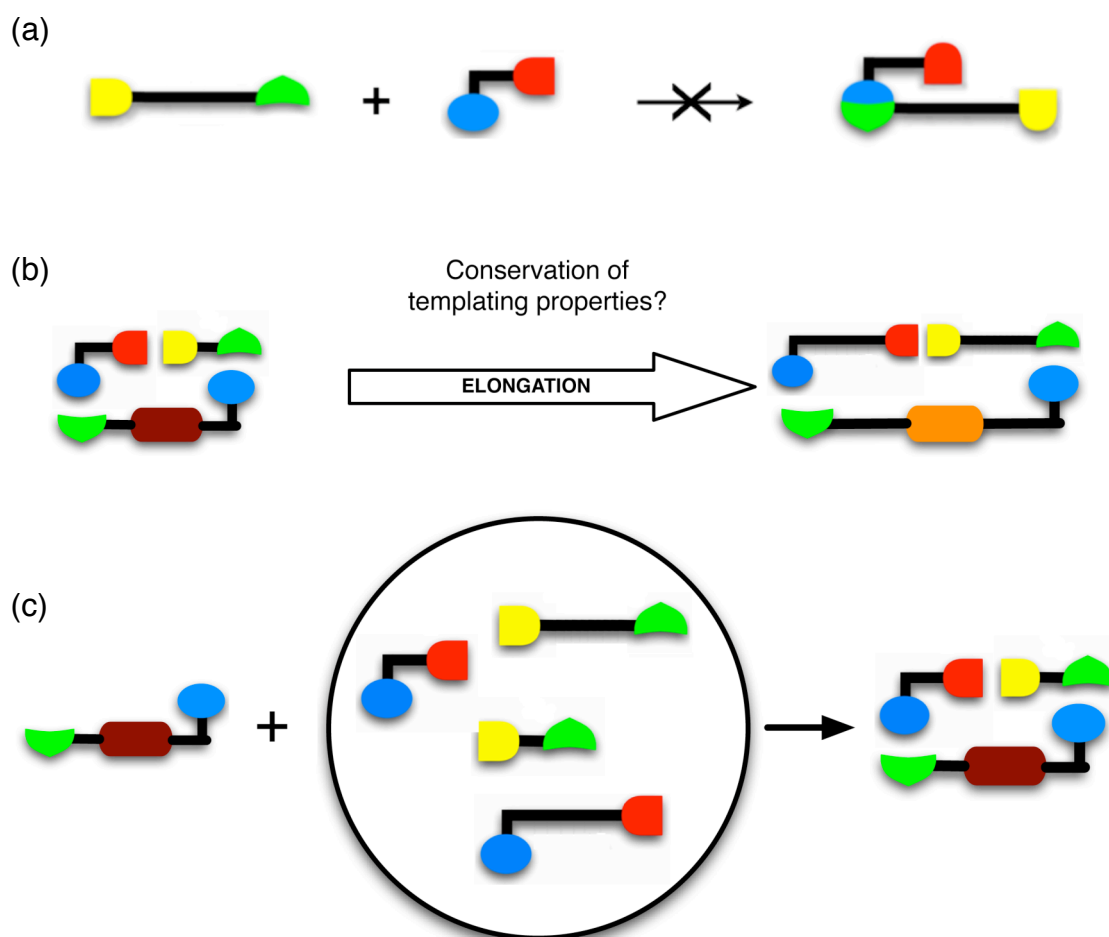

Figure 68. Introduction of length segregation into a minimal self-replicating systems in order to: (a) eliminate the AB complex pathway; (b) investigate if upon elongation of building blocks the templating properties of short analogues will be conserved; (c) create a system utilising components of different lengths that could be selectively addressed by templates of different lengths.

Eliminating the AB pathway is only one reason for changing the length of the studied molecules. Another question is, if changing the length of the building blocks of a known, efficient replicator will alter its properties (Figure 68b). Since the linker is kept straight and both building blocks are extended by the same length, theoretically the templating properties should be conserved.

Building blocks of different lengths will produce templates of different lengths. This property can be used to induce selectivity in the system. Templates of different lengths will only catalyse the reaction of the building blocks that fit the given templates inducing selective autocatalysis (Figure 68c). Cross-catalysis is possible when different building blocks form templates of the same length. These relationships could be used to create connections in a network of compounds with templating properties.

2.2 Components and properties of efficient self-replicating systems

The design of the building blocks is the key to achieving efficient selectivity. The building blocks **A** and **B** consist of three features: the recognition site, the reactive site and the linker between the two functional groups. The recognition motif based on hydrogen bonds between a carboxylic acid and amidopyridine has been studied extensively in our group. The association through the two hydrogen bonds has been studied on model **117** and **118** compounds and the association constant was determined[243] in CDCl$_3$ at 273 K to be 1000 M^{-1} (Figure 69).

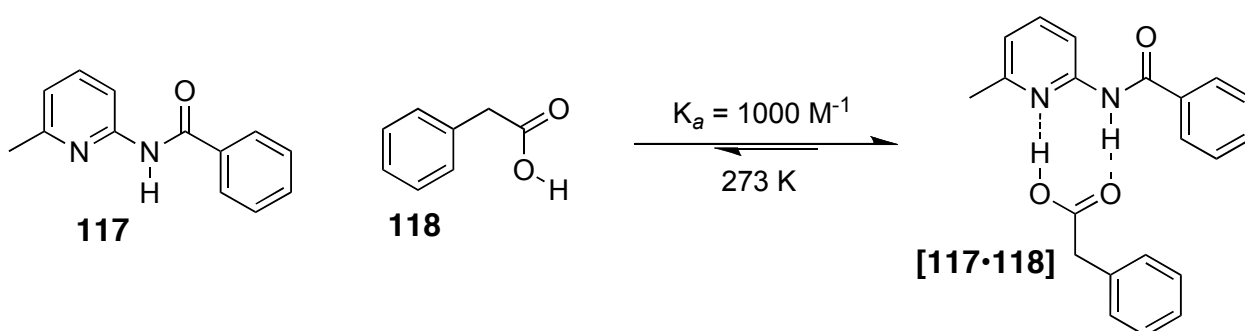**Figure 69**.   The association between amidopyridine **117** and carboxylic acid **118**.

A maleimide and a nitrone group are the reactive sites used in the design. The nitrone acts as a dipole and the maleimide acts as a dipolarohile. These properties allow the compounds to undergo a 1,3-dipolar cycloaddition (Figure 70). This particular reaction gives rise to two diastereoisomers described as *cis* and *trans*. The

compound in which all hydrogen atoms in the bicyclic isoxazolidine ring system are on the same face is rendered the *cis*. In the other isomer the nitrone-derived hydrogen atom is *trans* to the two hydrogen atoms originally located on the maleimide group. The formation of the isomers depends on the orientation of the building blocks in the transition state. If the reaction between the nitrone and the maleimide does not involve recognition events, than the ratio between the *trans* and *cis* isomer is usually close to 3:1, independently of the substituents on the aromatic rings adjacent to the reactive sites. The higher abundance of the *trans* isomer is a result of the energy difference of the two transition states. This ratio will be characteristic for the control reactions (without recognition) in all the described systems as well as the bimolecular reaction channel of the minimal model of self-replication.

If the reactive sites were connected to the complementary recognition motifs then different geometry of the diastereoisomers would result in different recognition properties.
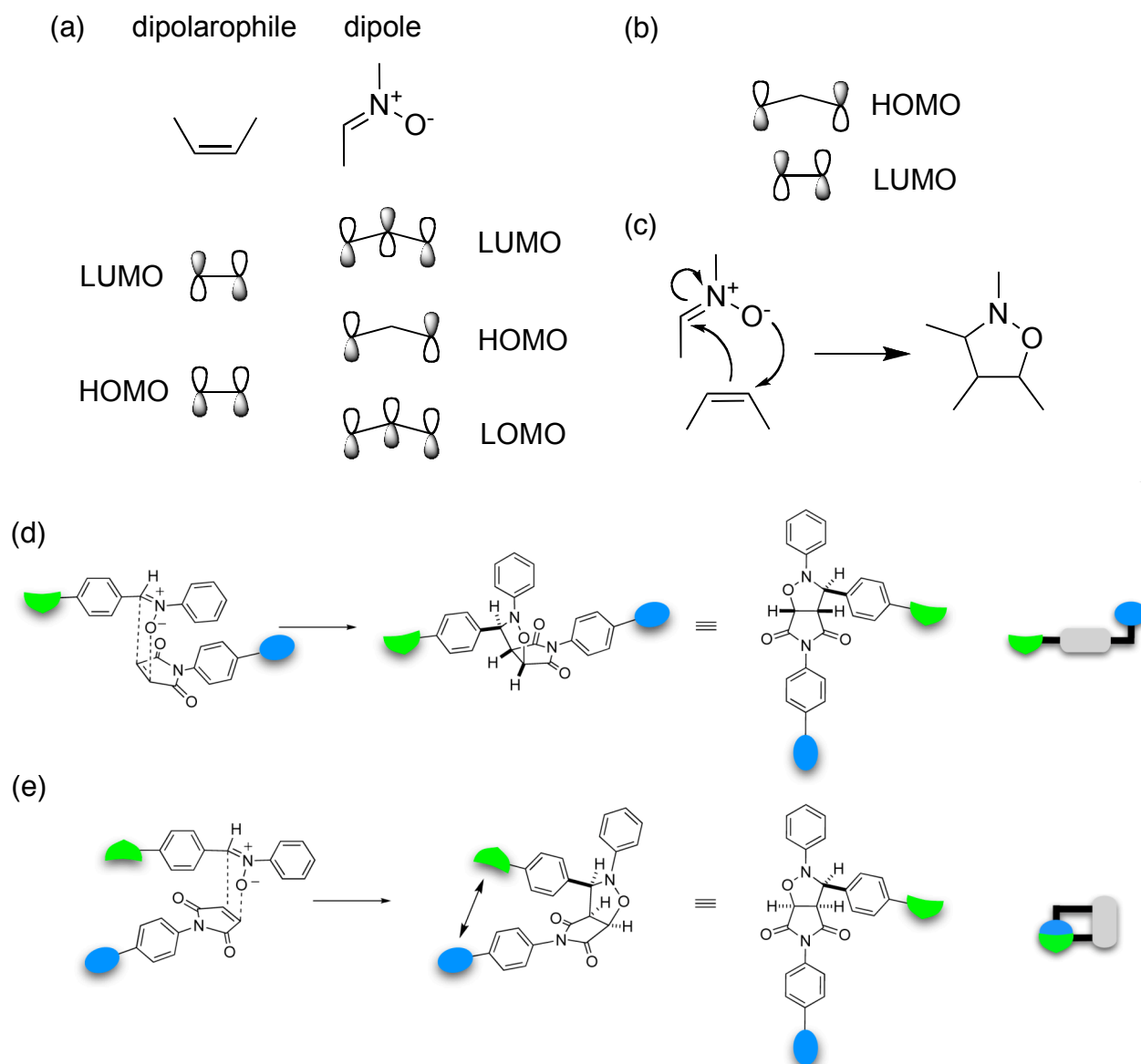
Figure 70. (a) Key molecular orbitals of the dipolarophile and the dipole. (b) Type I interaction between the highest occupied orbital of the dipole and the lowest unoccupied orbital of the dipolarophile is responsible for the cycloaddition reaction (c). 1,3-dipolar cycloaddition between a nitrone and a maleimide leads to the formation of two diastereoisomers. Transition state, structure and cartoon representation of (d) *trans* diastereoisomer and (e) *cis* diastereoisomer.

In the *trans* transition state, the recognition sites of the molecules are held far apart and, hence, we should not expect this product to display intramolecular recognition. In the *cis* transition state the molecules position themselves in such a way that the recognition sites are close together. If the geometry of the building blocks allows, the recognition could stabilise the *cis* transition state, leading to an accelerated formation of this isomer. Such scenario is present in the AB complex pathway of the minimal model of self-replication. When the reaction of choice in the self-replicating system is

the described above 1,3-dipolar cycloaddition then we can expect that the AB complex channel is responsible for the observed enhancement of the formation of the *cis* isomer. On the other hand the *trans* cycloadduct should assume a geometry with the free complementary recognition sites at the both ends of the molecule. Such property describes the template molecule **T** of the self-replication pathway. In the systems described, only the *trans* diastereoisomer displays self-replicating properties. Comparing the rates of formation of the two isomers, in the control reactions and the recognition-mediated reactions, is a good indication about the efficiency of the replicating pathway (*trans* selectivity) as well as the involvement of the AB pathway (*cis* selectivity).

The final component of the building blocks is the molecular linker that connects the recognition site and the reactive site. In order to control the geometry of the substrate the linker has to be kept rigid. The linkers can be modified around two parameters: length and angle. Incorporating aromatic rings or ethylene bonds in order to extend the linker can modify the length. The angle changes with different substitution patterns on the benzene rings. The angle at which the reactive site and the recognition motif are relatively to each other is one of the major factors determining if the reaction will follow the AB complex or the replicating pathway. If the bond angles in the product structure allow for intramolecular recognition of the complementary recognition sites than the resulting 'closed' configuration will be inactive for autocatalysis.

An efficient self-replicating system is expected to display very characteristic properties. In order to assess these properties a control reaction is needed. Since recognition is necessary for the replication to occur the chosen control reaction has the recognition site blocked. A simple substitution of the carboxylic acid group by the corresponding methyl ester is sufficient to block the formation of the hydrogen bonds responsible for the formation of the active complexes. The observed rate and selectivity without recognition present will be termed the background or the control reaction. As mentioned previously the 1,3-dipolar cycloaddition used to create the described systems is a rather slow and unselective reaction when no recognition-mediated events take place. This property makes it an ideal candidate for improvement. Knowing the properties of the control reaction it is possible to evaluate the efficiency and the pathway *via* which the recognition-mediated reaction proceeds. The enhancement in the formation of one of the diastereoisomers when compared

with the background reaction will indicate if the dominant pathway is the self-replicating cycle or the AB complex channel. The shape of the reaction profile is a very important factor that describes the system. If the profile of the formation of the *trans* isomer is sigmoidal, with a distinct lag period at the start of the experiment, then the reaction is autocatalytic and in this case most likely self-replicating. The reaction rate relies on the amount of template in solution and since the template is at the same time the product of the reaction, then the maximum rate is reached after a certain time into the experiment. Since there is no template in solution at the start of the experiment the product can only by formed through the slow bimolecular pathway. The template is also an instructional species passing onto its copy stereochemical information (the *trans* isomer only catalyses the formation of the *trans* isomer). Without template the system is much more error-prone, therefore the lag period is an important time to some extent determining the selectivity of the reaction.

If an enhancement of the *cis* isomer is observed the reaction profile is not sigmoidal and the rate of the reaction is highest when the concentration of the substrates is highest at t = 0. This enhancement is indicative of the AB complex pathway since the geometry of the *cis* isomer allows the formation of a folded self-complementary structure. The comparison of emergent properties of an AB complex pathway and a self-replicating pathway are summarised in Figure 71.
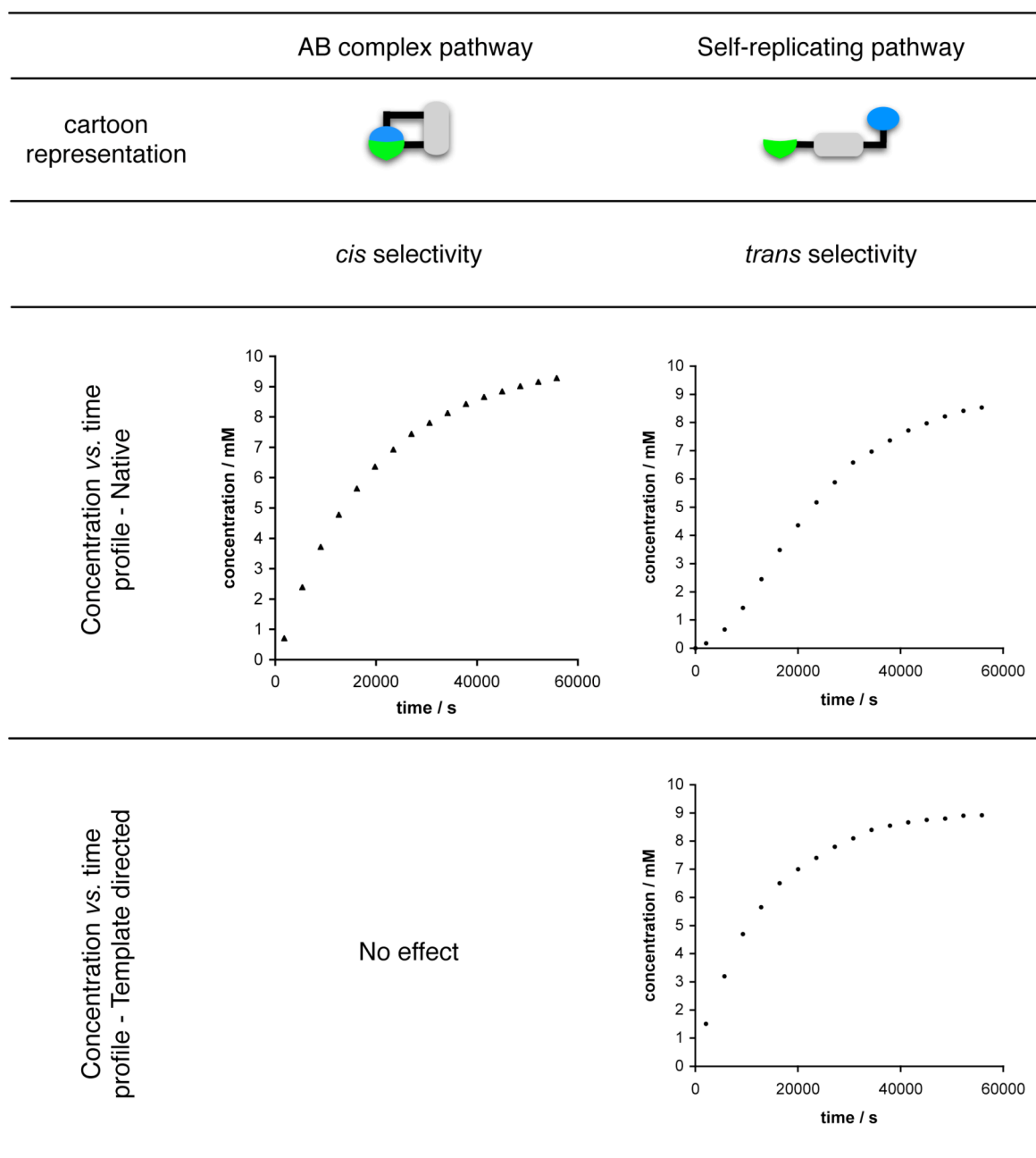
Figure 71. Comparison of the general behaviour of 1,3-dipolar cycloaddition reactions following the AB complex pathway and the self-replicating pathway.

These enhancements of rate of the reaction are observed when the control ester substrate is replaced with the analogous carboxylic acid substrate. One of the explanations of this behaviour could be just a simple acid catalysis and not the effect of recognition-mediated processes. In order to disprove this theory the previously developed replicating systems were allowed to react in the presence of another carboxylic acid. If the reaction would be acid catalysed than its rate would be even higher, however, if it were the recognition that was responsible for the enhancement,

then the efficiency of the system would be suppressed since the acid would compete for the amidopyridine recognition sites and act as an inhibitor. The latter case was always the result (Figure 72), proving that the observed properties of the reaction are the effect of recognition.
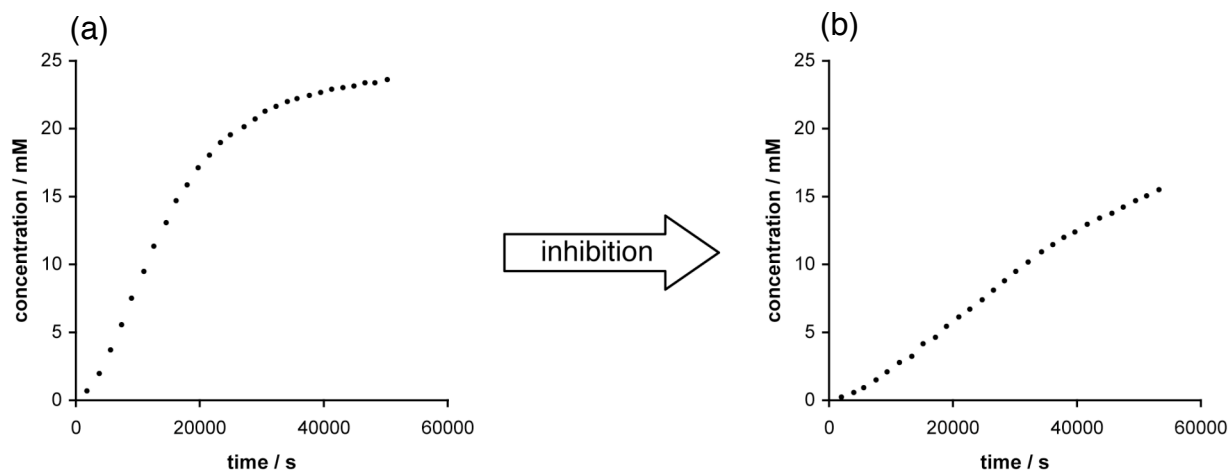

Figure 72. Concentration *vs.* time profiles of (a) an efficient replicator based on a reaction between a nitrone and a maleimide utilising aminopyridine-carboxylic acid recognition and (b) the same reaction doped with 30 mol% of benzoic acid that acts as recognition inhibitor.

The systems described in this work are based on the replicators thoroughly studied[244-246] previously in our group, hence the experiments with the competitive inhibitor were not performed for the newly designed building blocks.

A distinctive proof that the system is indeed self-replicating can be obtained by performing an appropriate doping experiment. The previously mentioned lag period of the replicator is caused by an insufficient amount of the template in solution for the reaction to assume its maximum autocatalytic rate. If the experiment were to be conducted with a sufficient amount of the preformed template (product) in solution together with the starting building blocks the lag period should disappear. The rate of the reaction should be highest from the start of the experiment and comparable to the maximum rate of the formation of the products in the undoped experiment. Such behaviour is only observed for the self-replicating pathway. If the reaction proceeds *via* an AB complex channel the addition of the template to such system will not affect the rate of the product formation. An efficient self-replicator will display an outstanding enhancement of the rate of formation of the *trans* isomer when compared with the control reaction and a strong template effect when doped with preformed product resulting in a disappearance of the lag period.

All these observations are a result of numerous self-replicating systems developed in our laboratory over the previous years. Investigating different reactions, building block geometries and substitution patterns led to a firm understanding of the mechanism of replication and to discovering some very efficient replicators. The concrete experimental foundation left by my predecessors allowed for the expansion of the replicator methodology into more complex networks.

2.3 Synthesis and properties of a replicator

One previously mentioned but underdeveloped system in our group was used to explore if length segregation of the building blocks would eliminate the AB pathway. Accordingly, the nitrone and the maleimides were prepared from appropriate starting materials (Figure 73). 2-Amino-6-methylpyridine **119** was condensed with chloromethylbenzoyl chloride **120** to afford the 4-chloromethyl-*N*-picoline **121** in almost quantitative yield. Treatment of the chloromethyl substituted amide **121** with hexamethylene tetramine and hydrochloric acid under Sommelet reaction conditions[247] afforded the aldehyde equipped with the recognition motif **122**. The tolylhydroxylamine **123** was prepared by Zn-catalysed reduction of the 1-methyl-4-nitrobenzene in the presence of excess of ammonium chloride and was used without further purification. The aldehyde and the hydroxylamine were dissolved in ethanol and kept in the dark for three days resulting in the precipitation of the desired nitrone **124**, which was separated by filtration.
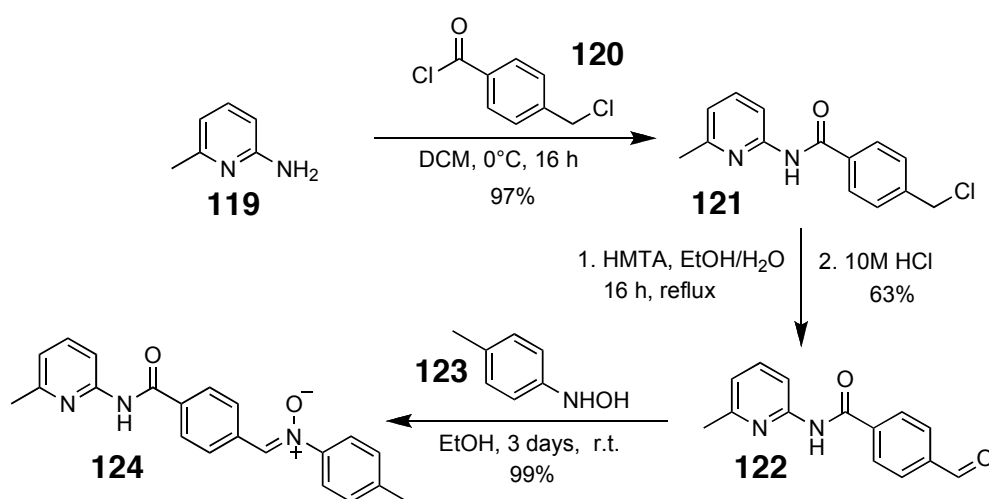

Figure 73. Synthesis of nitrone **124**.

In order to synthesise the maleimide building blocks (Figure 74) 3-aminobenzoic acid **125** and maleic anhydride **126** were dissolved in THF resulting in the diacid

intermediate **127** precipitating from solution. The solid was dissolved in acetic anhydride and heated with the addition of sodium acetate yielding the desired acid maleimide **128**. In order to study the reaction with the recognition site blocked a methyl ester of **128** was prepared. Acid maleimide **128** was dissolved in DMSO with methyl iodide and caesium carbonate affording the desired ester **129**.
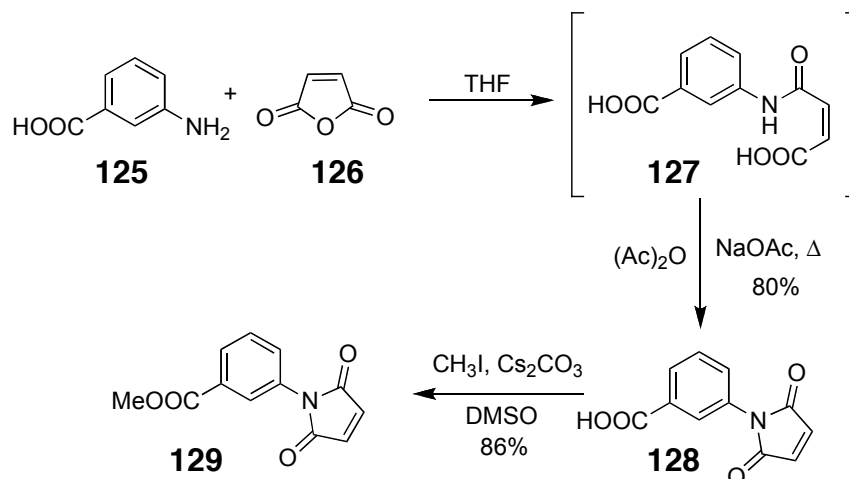

Figure 74. Synthesis of maleimides **128** and **129**.

The 1,3-dipolar cycloaddition reaction (Figure 75) was performed in CDCl$_3$ in order to monitor the reaction with ^{1}H NMR spectroscopy. Two diastereoisomeric products were identified in the reaction mixture. The cycloadduct **trans-131** was separated by crystallisation and characterised.

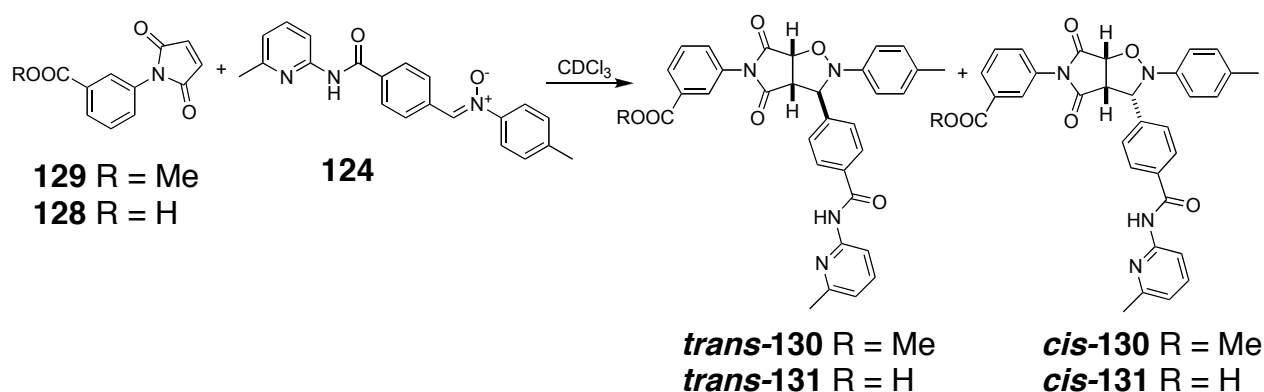**Figure 75.** Products of 1,3-dipolar cycloaddition between maleimides **128** and **129** with nitrone **124**.

A kinetic analysis of the system was undertaken. In a typical experiment, a sample suitable for analysis by NMR spectroscopy was prepared in a 5 mm NMR tube (Wilmad 528PP) by mixing appropriate volumes of stock solutions in CDCl$_3$ such that the total volume was 0.7 ml and the concentration of each component equalled

10 mM. The NMR tube was transferred to an NMR spectrometer, regulated at 273 K, and 500.1 MHz ^{1}H NMR spectra were acquired automatically every 1800 s over a period of 16 hours. Analysis and deconvolution of each of the ^{1}H NMR spectra recorded during this time was performed using Bruker Topspin software (Version 2.0 pl 3, Bruker Biospin, Germany, 2006). The data acquired from deconvolution was used to determine the concentrations of the mixture components at different time points. This procedure allows plotting concentration *vs.* time profiles for the products of the reaction.

Firstly, a control experiment (Figure 76a) was run using the methyl ester of the maleimide **129**. ([**129**] = [**124**] = 10 mM). The ester group disables the recognition rendering the reaction bimolecular. This approach allows assessing the rate and selectivity of the reaction when no recognition is present.
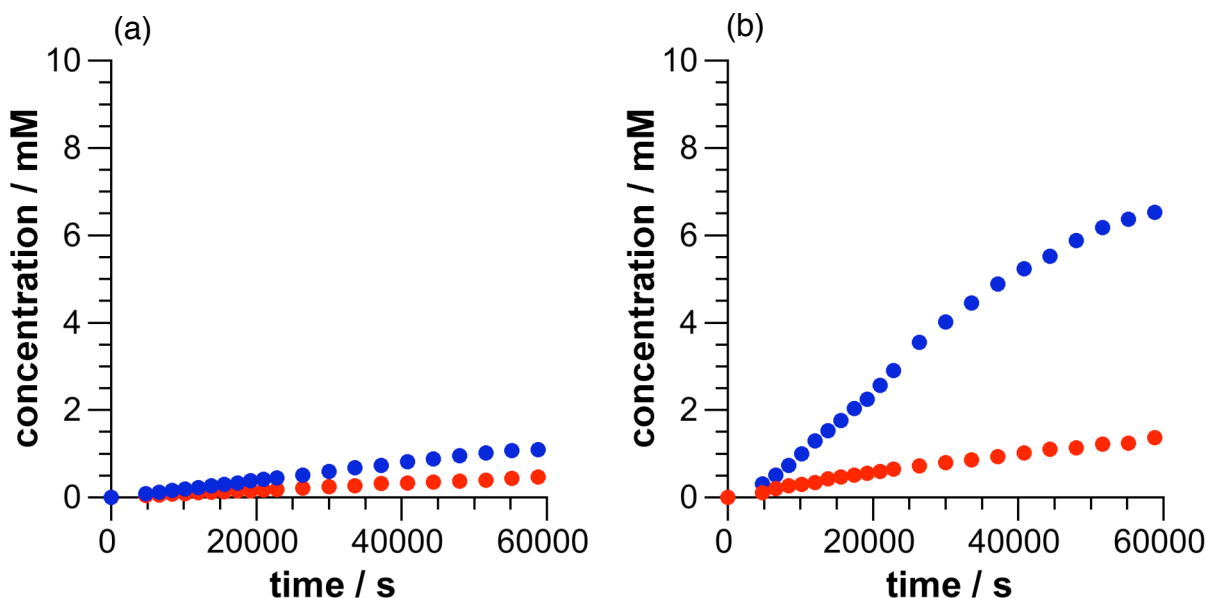

Figure 76. Concentration *vs.* time plots of (a) control reaction between ester maleimide **129** and nitrone **124** ([**129**] = [**124**] = 10 mM) following the formation of products: ***trans*-130** (●) and ***cis*-130** (●). (b) Recognition-enabled reaction between maleimide **128** and nitrone **124** ([**128**] = [**124**] = 10 mM) following the formation of products: ***trans*-131** (●) and ***cis*-131** (●).

The control reaction is inherently slow reaching a conversion of 15% and the selectivity is 2.3:1 in favour of the ***trans*-130** after 16 h. The reaction was performed in the same conditions with recognition deprotected using the acid maleimide **128** as a substrate ([**128**] = [**124**] = 10 mM). The concentration *vs.* time plot (Figure 76b) reveals a significant enhancement in the production of the ***trans*-131** cycloadduct. The reaction profile clearly shows a sigmoidal shape suggesting that the ***trans*-131**

product acts as a template in a self-replicating process. The *cis*-**131** isomer is also slightly amplified reaching 1.3 mM in the recognition experiment compared to 0.5 mM of *cis*-**130** in the control. Since the closed template configuration of this product makes it possible to be formed *via* the AB complex, one can suspect the involvement of this pathway in the recognition-mediated reaction.

2.4 Achieving length segregation in synthesis

The reaction clearly displays replicating behaviour for the *trans*-**131** isomer and some AB complex pathway enhancement for the *cis*-**131** isomer. Blocking the AB complex channel could direct more of the starting material through the autocatalytic cycle and result in higher replicator efficiency and selectivity. Changing the length of one of the building blocks will result in a spatial separation of the reactive sites upon recognition of the building blocks making the AB complex inactive. In order to thoroughly analyse this approach elongated analogues of both the nitrone and the maleimide were synthesised.

The single phenyl linker was exchanged for a longer diphenylethyne group (Figure 77).
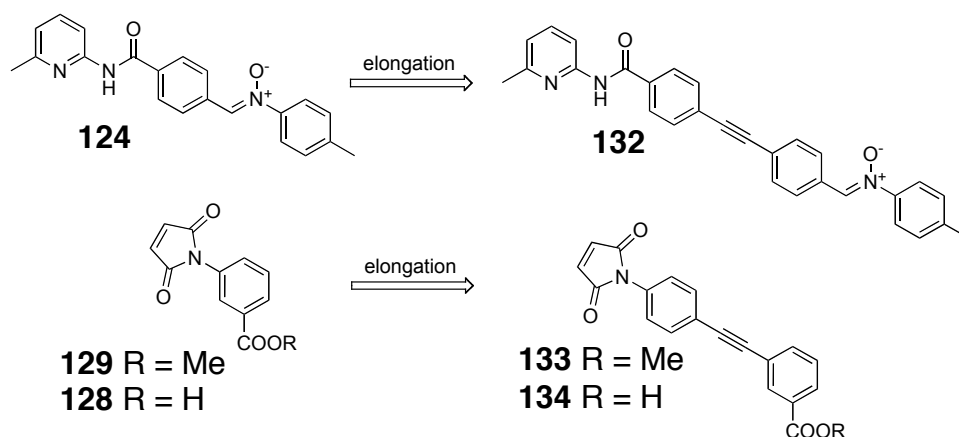
Figure 77. Synthetic design of elongated versions of compounds **124**, **128** and **129**.

The long analogues of the nitrone and the maleimides were prepared from appropriate starting materials. 4-ethynylaniline **135** was protected with a Boc group to form **136** and the 3-bromobenzoic acid **137** was converted to the methyl ester **138**. The resulting compounds were reacted under Sonogashira conditions[248] to afford the functionalised diphenylethyne **139** (Figure 78).

Figure 78. Synthesis of diphenylethyne **139**.

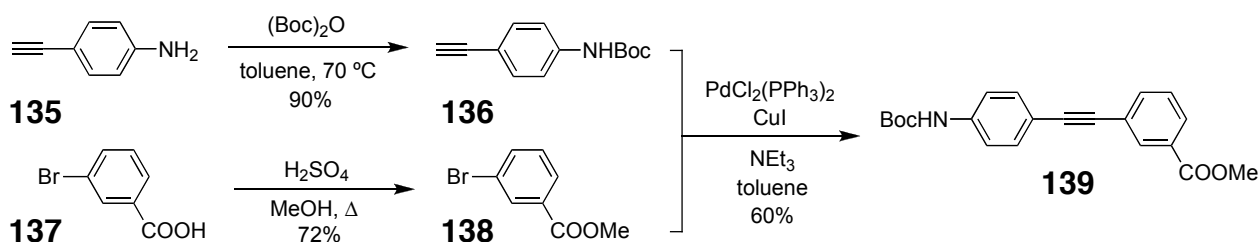The ester was hydrolysed with lithium hydroxide and the Boc group was removed with TFA affording the aminoacid intermediate **141**. The maleimide was generated by the addition of maleic anhydride **126** and subsequent cyclisation affording the elongated acid maleimide **134** (Figure 79).
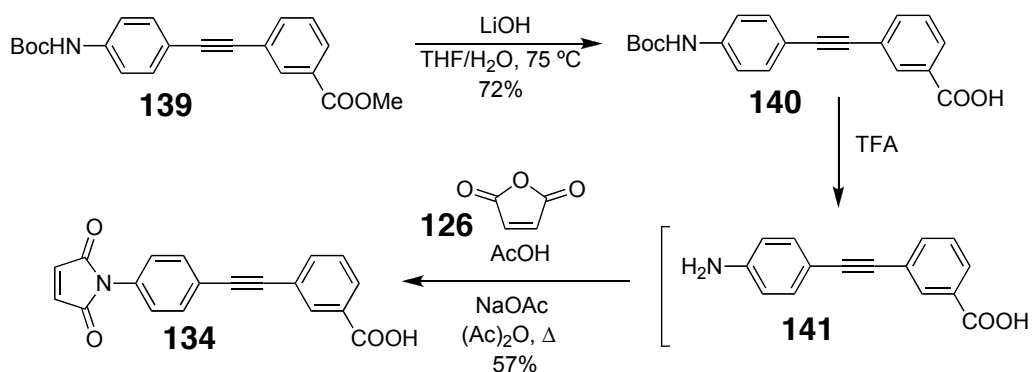

Figure 79. Synthesis of elongated acid maleimide **134**.

In order to synthesise the ester control compound **133** only the Boc group was removed from compound **139** using TFA. The methyl ester group was conserved through further synthetic steps yielding the desired ester maleimide **133** (Figure 80).
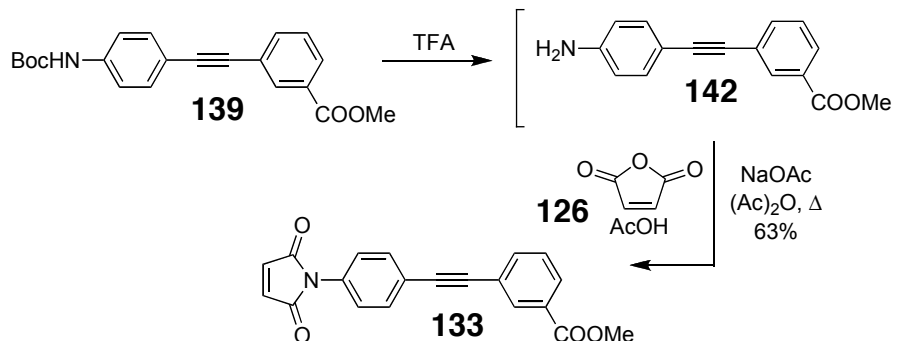

Figure 80. Synthesis of elongated ester maleimide **133**.

The first step in the synthesis of the elongated nitrone **132** was a condensation of 2-amino-6-methylpyridine **119** and 4-iodobenzoyl chloride **143** to afford **144**. This iodo-substituted amide **144** was condensed with 4-ethynylbenzaldehyde **145** in a Pd-catalysed Sonogashira reaction to afford compound **146** bearing a recognition site and an aldehyde group separated by a diphenylethyne linker. The aldehyde reacted readily with the freshly prepared tolylhydroxylamine **123** to afford the desired nitrone **132** (Figure 81).
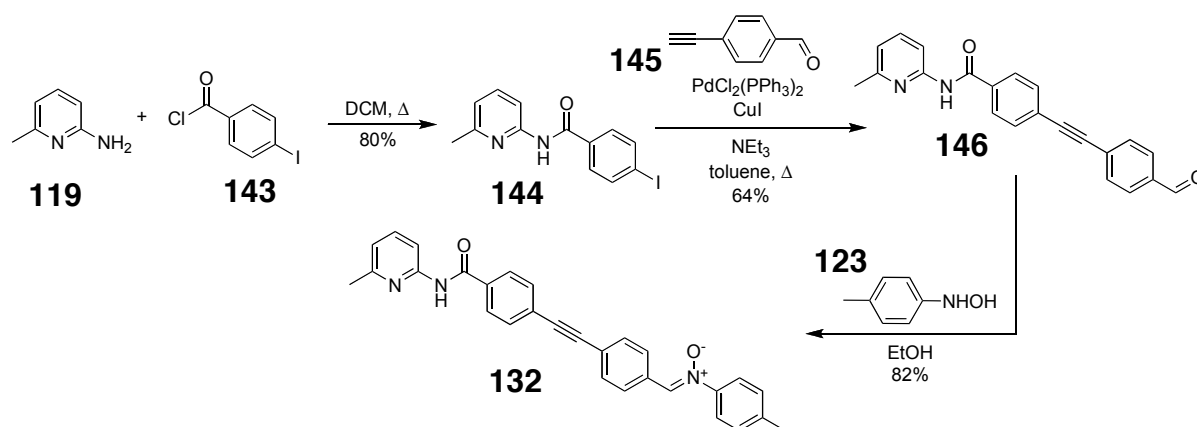

Figure 81. Synthesis of elongated nitrone **132**.

2.5 Substrate length and replication efficiency

The modified compounds together with the shorter original building blocks open a possibility of new reacting pairs with substrates of different length: long nitrone **132** - short maleimide **128** and short nitrone **124** - long maleimide **134**. The difference in length of the reacting partners should eliminate any enhancement in the formation of the *cis* product when comparing the control experiment to the recognition-enabled experiment. All 1,3-dipolar cycloaddition (Figure 82) reactions were performed in CDCl$_3$ at 273 K with the starting concentrations of 10 mM. The reactions were monitored using 500 MHz ^{1}H NMR spectroscopy and some species in the reaction mixtures are present at concentrations below the detection level of this analytical technique. This inconvenience results in some data points missing in the concentration *vs.* time plots describing the kinetic experiments.
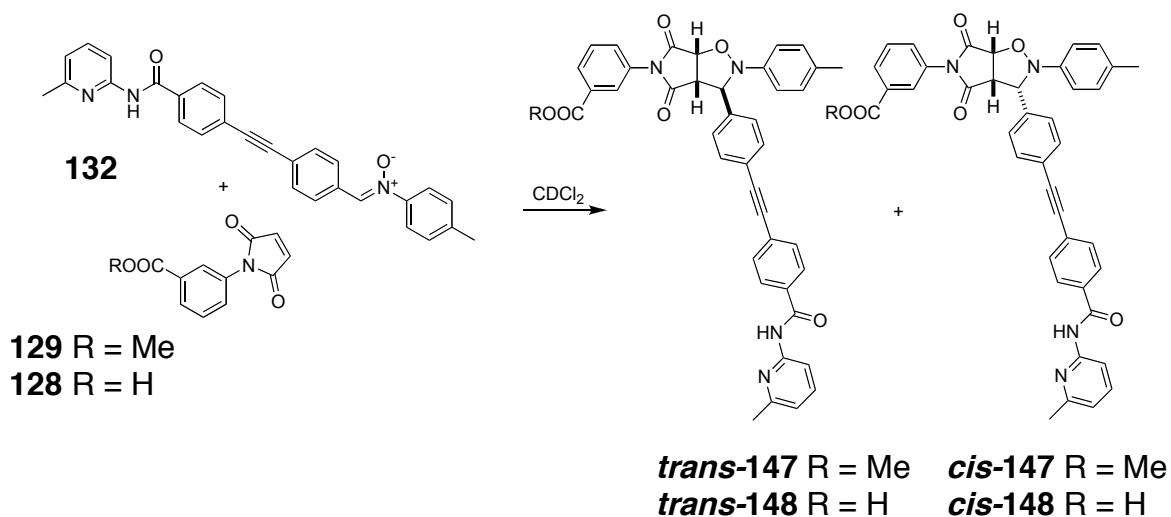
Figure 82. Products of 1,3-dipolar cycloaddition between maleimide **128** and **129** with nitrone **132**.

Figure 83. Concentration *vs.* time plots of (a) control reaction between ester maleimide **129** and nitrone **132** ([**129**] = [**132**] = 10 mM) following the formation of products: *trans*-**147** (●) and *cis*-**147** (●). (b) Recognition-enabled reaction between maleimide **128** and nitrone **132** ([**128**] = [**132**] = 10 mM) following the formation of products: *trans*-**148** (●) and *cis*-**148** (●).

The control reaction (Figure 83a) between the short ester maleimide **129** and the long nitrone **132** ([**129**] = [**132**] = 10 mM) displays poor conversion of 14% and a selectivity of 2:1 in favour of *trans*-**147** after 16 h. When comparing the reaction profile for the formation of the *cis*-**147** cycloadduct in the control and *cis*-**148** in the recognition-enabled reaction ([**128**] = [**132**] = 10 mM) (Figure 83b) there is no difference between them. Since the AB complex channel is responsible for the recognition-mediated enhancement of the *cis* isomer, such observation suggests this channel is disabled. The formation of the *trans*-**148** product is faster than the control

90

reaching 2.7 mM after 16 h compared to 0.9 mM of **trans-147** for the control. The efficiency of the self-replication of the length-mismatched pair is, disappointingly, poorer than in the native short maleimide **128** - short nitrone **124** system (Figure 76). Changing the length of the nitrone did eliminate the AB complex channel, but even though the geometry of the nitrone was preserved the replicating properties of the new template **trans-148** were compromised.

The reaction of long maleimide **134** and short nitrone **124** (Figure 84) was also investigated in hope of it displaying better reactivity *via* the self-replicating channel.

Figure 84. Products of 1,3-dipolar cycloaddition between maleimide **134** and **133** with nitrone **124**.

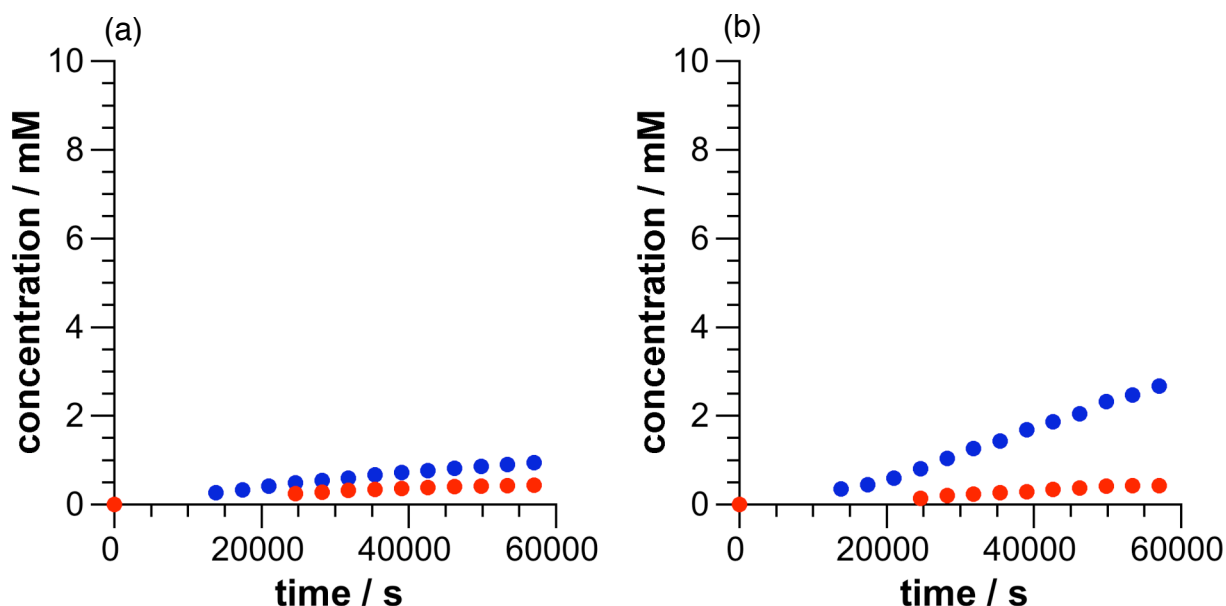**Figure 85.** Concentration *vs.* time plots of (a) control reaction between ester maleimide **133** and nitrone **124** ([**133**] = [**124**] = 10 mM) following the formation of products: **trans-149** (●) and **cis-149** (●). (b) Recognition-enabled reaction between maleimide **134** and nitrone **124** ([**134**] = [**124**] = 10 mM) following the formation of products: **trans-150** (●) and **cis-150** (●). Missing points at the early stages of the experiment are a result of low concentration of species in the reaction mixture.

The kinetics of the control ([**133**] = [**124**] = 10 mM) (Figure 85a) and the recognition-mediated reaction ([**134**] = [**124**] = 10 mM) (Figure 85b) were monitored by 500 MHz ^{1}H NMR spectroscopy. Analysis of the data for the long maleimide **134** - short nitrone **124** reaction, shows no evidence for any AB complex pathway in the recognition-mediated reaction – there is no enhancement in the formation of the ***cis*-150** isomer upon the introduction of recognition. The self-replicating activity is also greatly compromised with only a slightest enhancement. The concentration of ***trans*-149** after 16 h is 1.4 mM for the control and 1.7 mM for ***trans*-150** when recognition is present. The above results suggest that changing the length of one of the components of the reaction has an effect on the geometry of the open template necessary for self-replication. This observation is quite surprising, as the only change is the length of linker. A rigid linker should not compromise the templating property of the cycloadduct. However, if a longer linker is seen as a longer lever, than any small changes in geometry of the complex at one end of the molecule will have a larger effect on the other end. It is easy to envisage that since the complementary recognition sites are now further apart a precise assembly of a ternary complex needed for replication may be more difficult then when all components are relatively shorter.

An insight into how length of linker influences the replicating properties of the reaction can be gained by investigating the reactive pair of the two elongated derivatives: the long maleimide **134** - long nitrone **132** reaction (Figure 86).
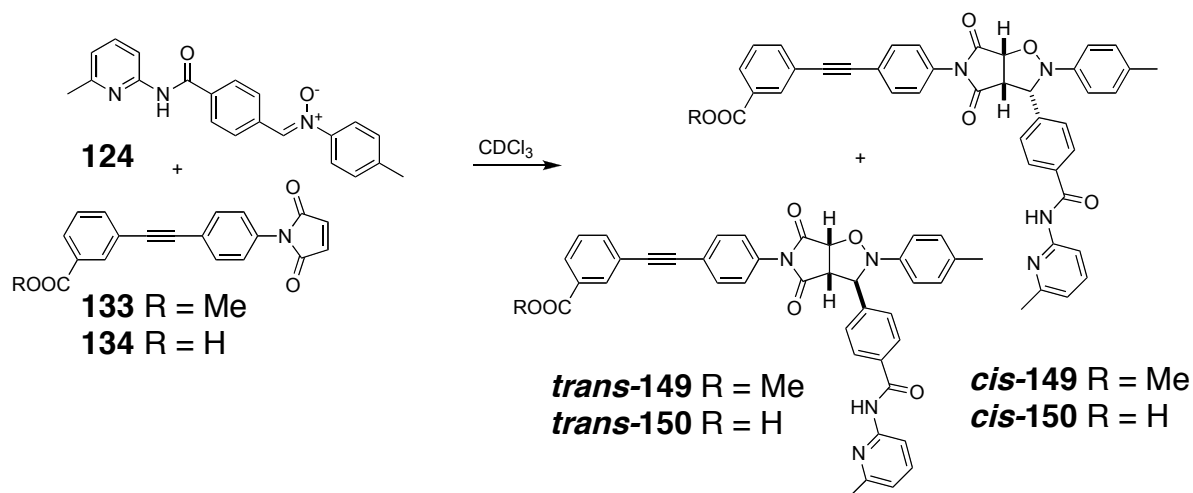

Figure 86. Products of 1,3-dipolar cycloaddition between maleimide **134** and **133** with nitrone **132**.

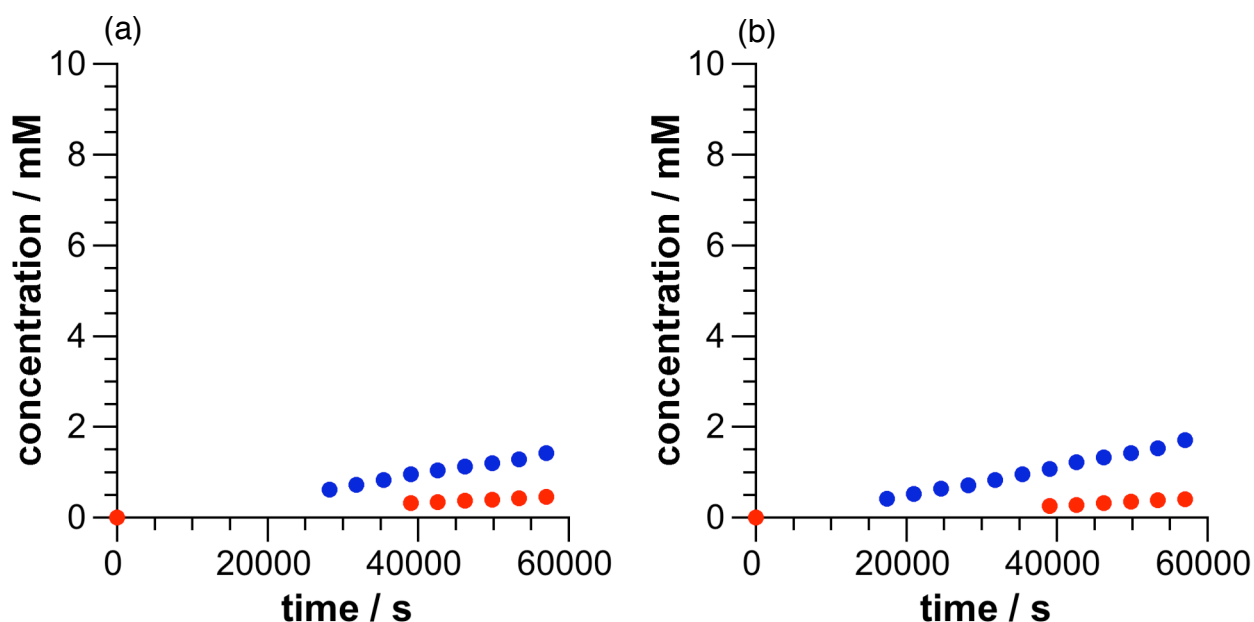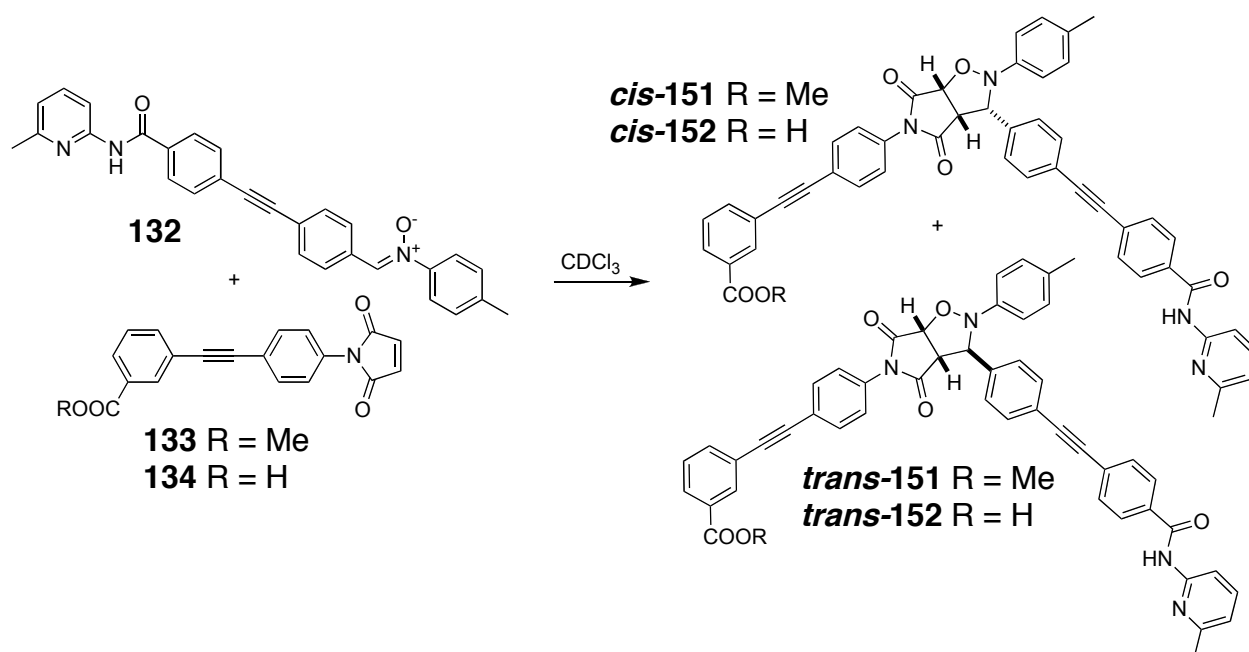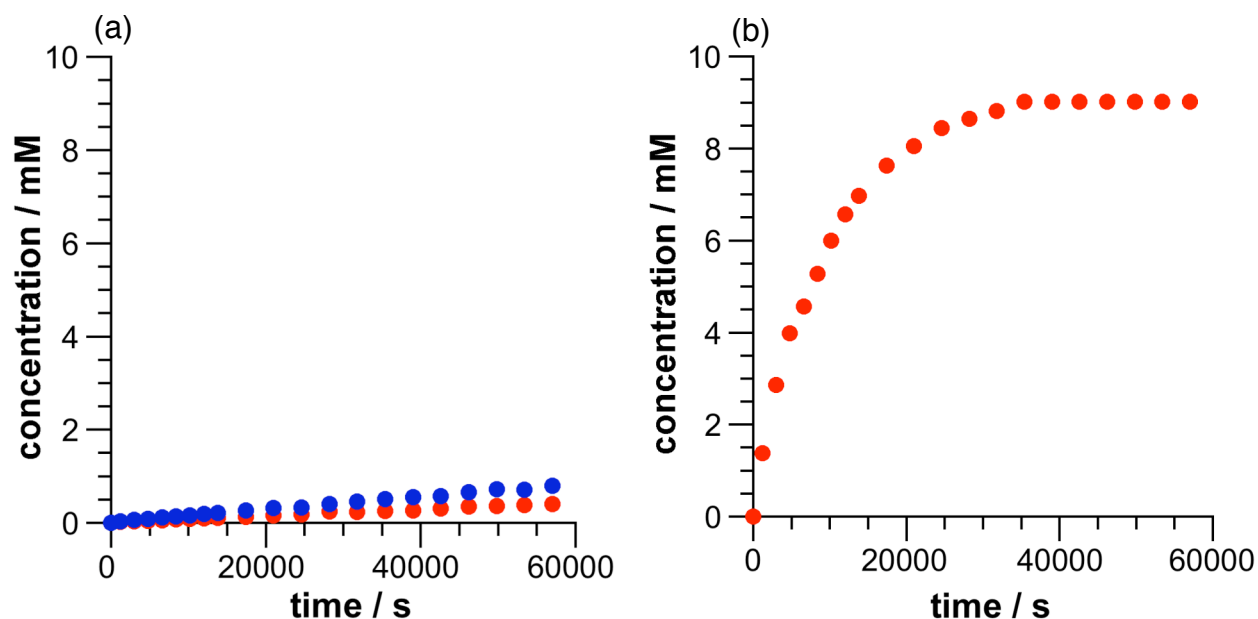**Figure 87.** Concentration *vs.* time plots of (a) control reaction between ester maleimide **133** and nitrone **132** ([**133**] = [**132**] = 10 mM) following the formation of products: ***trans*-151** (●) and ***cis*-151** (●). (b) Recognition-enabled reaction between maleimide **134** and nitrone **132** ([**134**] = [**124**] = 10 mM) following the specific formation of product ***cis*-152** (●).

The control reaction ([**133**] = [**132**] = 10 mM) (Figure 87a) is not any different than the previous examples with the conversion reaching 12% and selectivity 2:1 in favour of the ***trans*-151** after 16 h. The recognition-enabled reaction ([**134**] = [**124**] = 10 mM) (Figure 87b) shows, however, a completely inverted selectivity. The only product formed is the ***cis*-152** diastereoisomer. This selectivity and the great enhancement in formation of this particular cycloadduct indicate the whole of reaction proceeds *via* the AB complex channel with the conversion reaching 90% after 16 h. The longer linkers in the substrates seem to allow for the folding of the molecules upon recognition and a subsequent pseudo-intramolecular reaction. The NMR spectroscopy data proves that the hydrogen bonds are conserved in the product, forming a closed template. It seems the bigger distance between the reactive site and a recognition site allows for a higher degree of flexibility when assembling the reactive complex. The more compact AB complex in this case is a more favourable structure than the extensively long open template, hence the specificity of the reaction towards the ***cis*-152** isomer.

The discussed building blocks: maleimides **128** and **134** and their esters **129** and **134** together with nitrones **124** and **132** form eight reactive combinations four control

reactions and four recognition-mediated reactions with different properties (Figure 88).
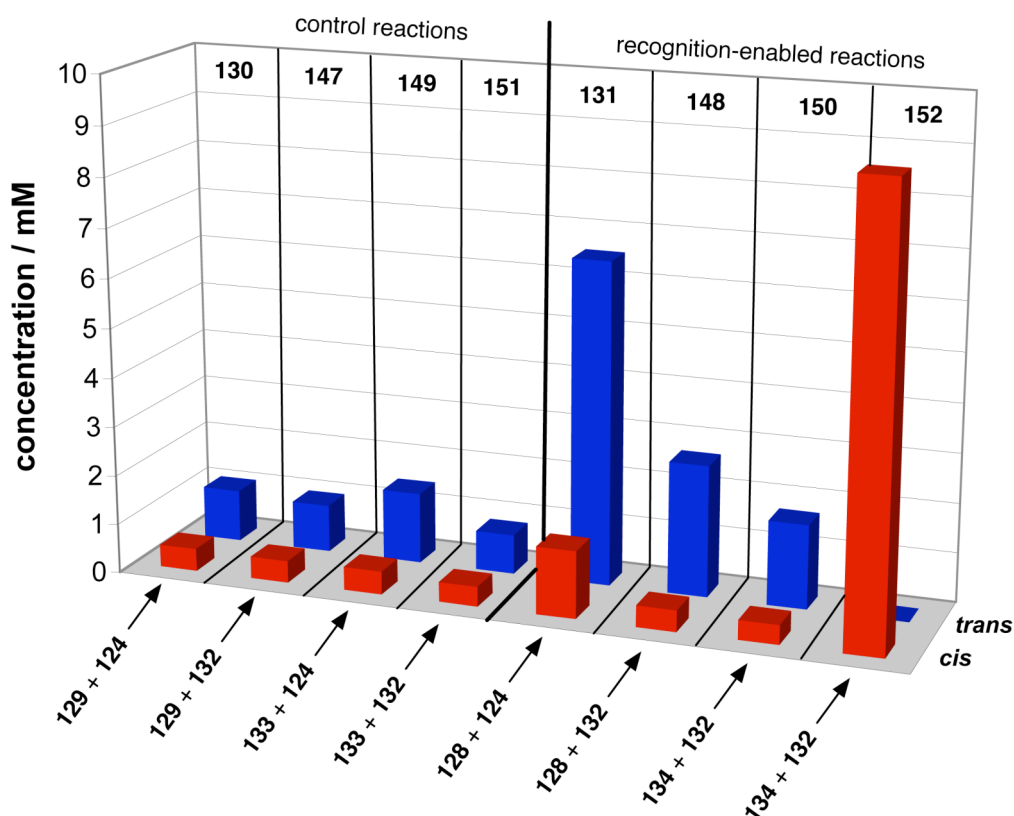
Figure 88. Concentrations of all cycloadducts derived from maleimides **129** and **133** and nitrones **124** and **132** (control) and maleimides **128** and **132** and nitrones **124** and **132** (recognition) after 16 h into the reaction. The *trans* isomers are represented by blue bars (■) and the *cis* isomers by red bars (■).

2.6 Length segregation in a reaction network

The four recognition enabled building blocks of different lengths (**128**, **134**, **124**, **132**) capable of four reactive combinations that generate eight products through different recognition-mediated pathways would create an interesting network if all four substrates would be allowed to react simultaneously (Figure 89). Within the reactive pairs we can identify a fairly efficient replicator ***trans*-131** formed from the short reacting pair and a very efficient and selective AB complex reaction generating ***cis*-152** from the long reactive pair. These pathways are expected to dominate the network, since the two length-mismatched pairs show only slight response to the introduction of recognition.
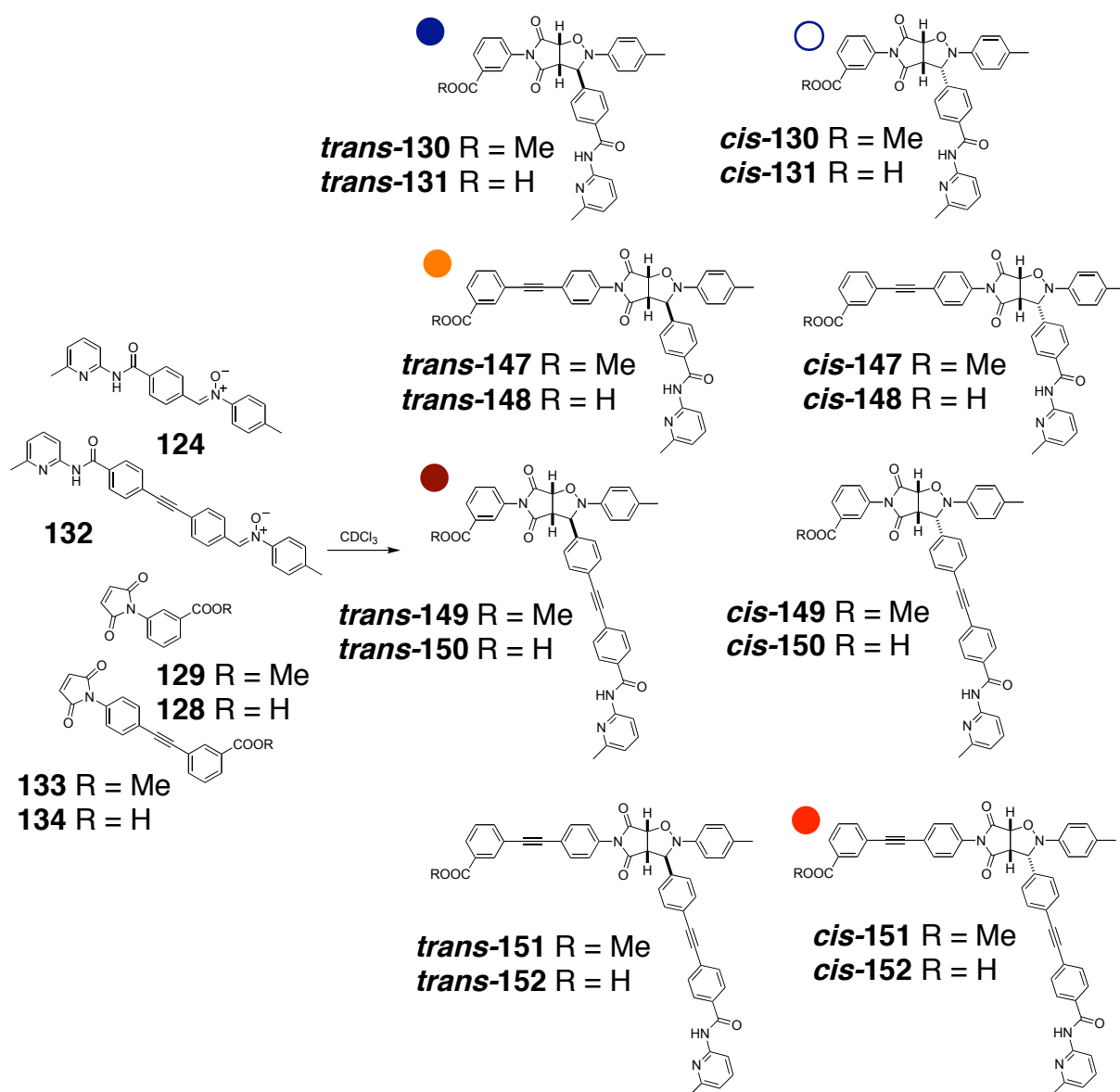
Figure 89. All possible system components generated from the mixture of four building blocks **124**, **132**, **129**, **128**, **133** and **134**. The products are colour-coded for future reference with replicator *trans*-**131** blue and product of the AB complex pathway *cis*-**152** red.

A control reaction was performed using the two ester maleimides **129** and **133** and the two nitrones **124** and **132** each at 10 mM in CDCl$_3$ at 273 K in order to assess how the network will evolve without recognition present (Figure 90a).

All four *trans* cycloadducts and all *cis* cycloadducts follow the same reaction path with no selectivity between them. The native ratio between the *cis* and *trans* products is conserved reaching 2.4:1 in favour of the *trans*. The overall yield is 14% with none of the products exceeding the concentration of 1 mM after 16 h.

When the four substrates were reacted under same conditions ([**128**] = [**134**] = [**124**] = [**132**] = 10 mM) only with the recognition deprotected, a completely different scenario unravelled (Figure 90b).

Figure 90. Concentration *vs.* time plots of (a) control reaction between ester maleimides **129** and **133** and nitrones **124** and **132** ([**129**] = [**133**] = [**124**] = [**132**] = 10 mM) following the formation of products: ***trans*-130,147,149,151** (●) and ***cis*-130,147,149,151** (●). (b) Recognition-enabled reaction between maleimides **128** and **134** and nitrones **124** and *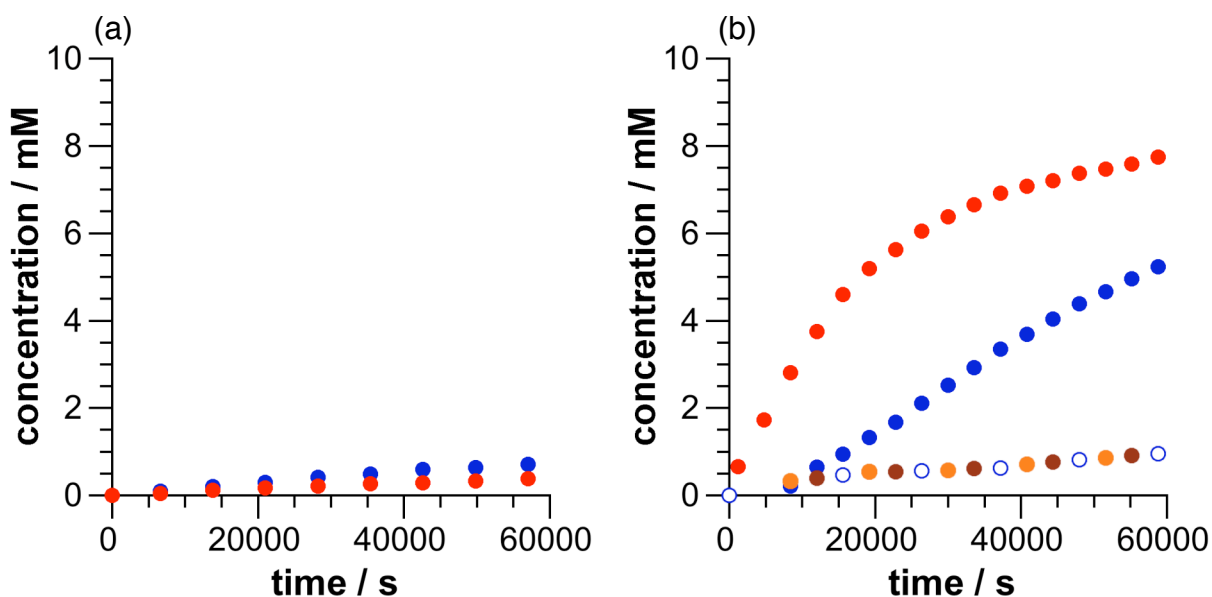*132** ([**128**] = [**134**] = [**124**] = [**132**] = 10 mM) following the formation of products ***cis*-152** (●), ***trans*-131** (●), ***cis*-131** (○), ***trans*-148** (●) and ***trans*-150** (●).

As predicted, the dominating reaction is the AB complex pathway involving the long components **134** and **132**. The second most efficient pathway is the short component replicator ***trans*-131**.

The two dominating reactions do not compete directly for the building blocks since the replicator only incorporates the short derivatives and only the elongated components are necessary to form ***cis*-152** *via* he AB complex pathway. The concentration of the dominating ***cis*-152** cycloadduct after 16 h is 7.6 mM (9 mM in isolation) and the self-replicating ***trans*-131** product reaches 5.2 mM (6.5 mM in isolation). Both reactions are less efficient in the network than in isolation. The reason for this situation is to some extent the competition with the length-mismatched reactions, as well as more possible binding interactions inhibiting the formation of the reactive complexes.

The AB complex channel is usually more efficient than the replicator when operating side by side. For the formation of the AB complex only two molecules need to recognise each other and no template is needed while the self-replicator bases on a ternary complex and the presence of template in solution. This situation makes the

AB complex pathway operate better especially at low concentrations in the beginning of the experiment giving it a clear advantage.

Other components observed above the limit of detection in the system were the two length-mismatched *trans* cycloadducts (***trans*-148** and ***trans*-150**) and the ***cis*-131** product of the reaction between the short maleimide **128** and short nitrone **124**. All these products followed practically the same reaction profile reaching the concentration of 1 mM each after 16 h. The short component ***cis*-131** product is able to compete with the length-mismatched *trans* products because there is a degree of the AB complex pathway in its formation. The mismatched products are disfavoured since long components are necessary for their formation and these are being rapidly incorporated in the formation of the long component ***cis*-152** cycloadduct and the short components are utilised by the replicator. The overall yield reached by the system is 78% after 16 h. What is clearly observed in the system is that from eight possible products two are selected, each through a different reaction pathway. In order to quantify the changes in the system upon introduction of recognition an enhancement factor was introduced as the log of the ratio between the concentrations of a given product in the recognition-enabled reaction and the control reaction after 16 h (Figure 91).

$$\text{enhancement factor} = \log \frac{[\text{product}]_{\text{recognition}}}{[\text{product}]_{\text{control}}}$$
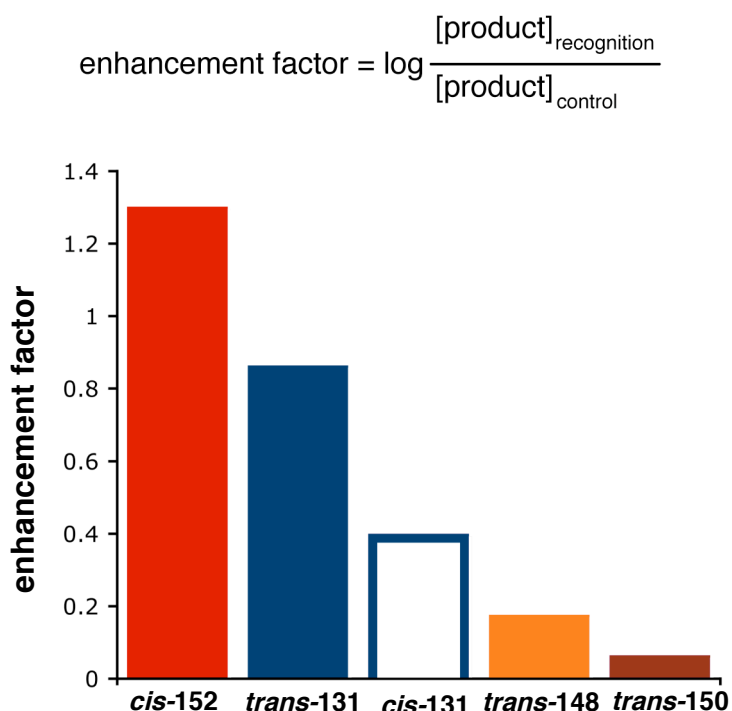

Figure 91. Equation describing the amplification factor and the amplification factors for compounds detected in the system derived from maleimides **128** and **134** and nitrones **124** and **132**.

Without recognition present the system displayed no selectivity and poor conversion providing no differentiating factor between the components (Figure 90a). The introduction of recognition in the system resulted in a self-sorting of the network components from the most efficient AB complex pathway to the very weak long maleimide **134** - short nitrone **124** replicator ***trans*-150**. A simple recognition event unveiled the properties engendered in the design of the building blocks in an intractable mixture.

2.7 Addressing the reaction network

The AB complex pathway is clearly faster than the self-replicator. However, a unique property of the autocatalytic reaction is that doping the reaction with preformed template can enhance its performance. The effect of template doping was first tested on the isolated reaction between the short maleimide **128** and the short nitrone **124**. The native experiment was repeated in the same conditions only with 10 mol% of template ***trans*-131** added together with the maleimide **128** and the nitrone **124** ([**128**] = [**124**] = 10 mM, [***trans*-131**] = 1 mM) (Figure 92).
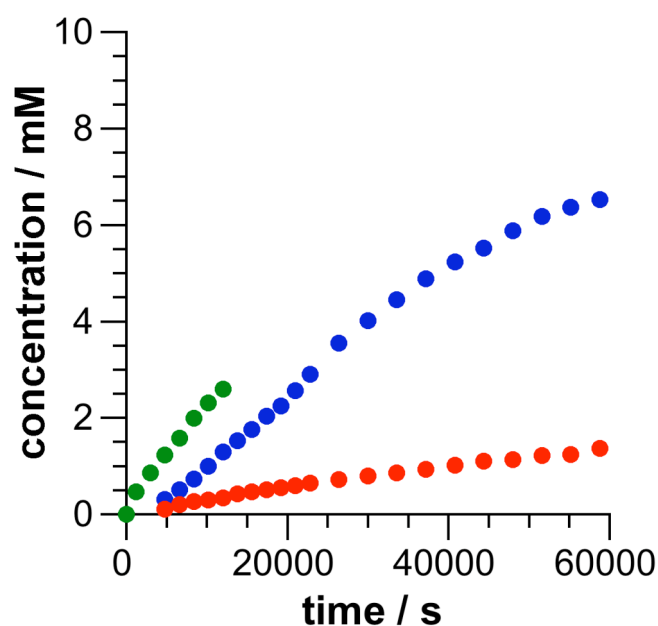
Figure 92. Concentration *vs.* time plot of recognition-enabled reaction between maleimide **128** and nitrone **124** doped with 10 mol% of preformed template ***trans*-131** ([**128**] = [**124**] = 10 mM, [***trans*-131**] = 1 mM) following the specific formation of product: ***trans*-131** (●) overlaid on a the plot of the undoped reaction between maleimide **128** and nitrone **124** ([**128**] = [**124**] = 10 mM) following the formation of products: ***trans*-131** (●) and ***cis*-131** (●).

Precipitate was observed in the sample after 4 h making further data collection impossible. Fortunately, the biggest difference between the doped and native experiment is, in fact, observed at the start of the experiment. The native reaction exhibits a lag period, which is caused by an insufficient amount of template *trans*-**131** in solution for the reaction to assume its maximum rate. In the doped experiment, since the template is already in solution from the start of the reaction, the maximum rate is achieved immediately. The addition of the template also should increase the selectivity of the system in favour of the replicating isomer *trans*-**131**. The experiment with all four building blocks and added preformed short-component *trans* cycloadduct ([**128**] = [**134**] = [**124**] = [**132**] = 10 mM, [*trans*-**131**] = 1 mM) was to prove if the template effect was significant enough to even the chances in the competition between the replicator and the AB complex channel (Figure 93).
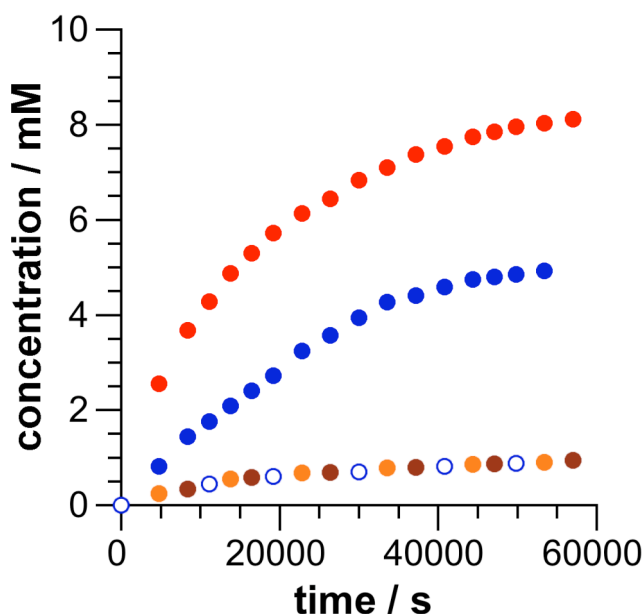
Figure 93. Concentration *vs.* time plot of doped recognition-enabled reaction between maleimides **128** and **134** and nitrones **124** and **132** ([**128**] = [**134**] = [**124**] = [**132**] = 10 mM, [*trans*-**131**] = 1 mM) following the formation of products *cis*-**152** (●), *trans*-**131** (●), *cis*-**131** (○), *trans*-**148** (●) and *trans*-**150** (●).

The disappearance of the lag period for the short component reaction is apparent in the four-substrate system doped with *trans*-**131**. Even though the formation of the self-replicating template is faster at the start of the experiment *trans*-**131** product reaches practically the same concentration after 16 h as in the undoped experiment (4.9 mM in the undoped experiment and 5.2 mM when doped with preformed template). In both experiments the total conversion of the system is the same

reaching 78%. Interestingly, the dominating AB complex pathway seems to reach higher concentration of product *cis*-**152** than in the undoped reaction: 8.1 mM compared to 7.7 mM. The template *trans*-**131** cannot have any direct recognition effect on the long component *cis*-**152** product. The explanation could lay in the dependence of the reactions for the same building blocks. Since the two dominating pathways do not compete directly for building blocks as one uses only the short components and the other only the long. In the doped experiment the short components **128** and **124** will be incorporated much more rapidly in the early stages of the evolution of the system than in the undoped reaction. The lack of short building blocks disfavours the length-mismatched reactions, which need both the short and the long substrates. Both of the length-mismatched reaction products *trans*-**148** and *trans*-**150** reach a conversion that is lower by 0.1 mM after 16 h compared to the undoped reaction. Since these reactions are slightly suppressed more long building blocks **134** and **132** are available for the AB complex pathway allowing for a higher yield of *cis*-**152** *via* this particular channel.
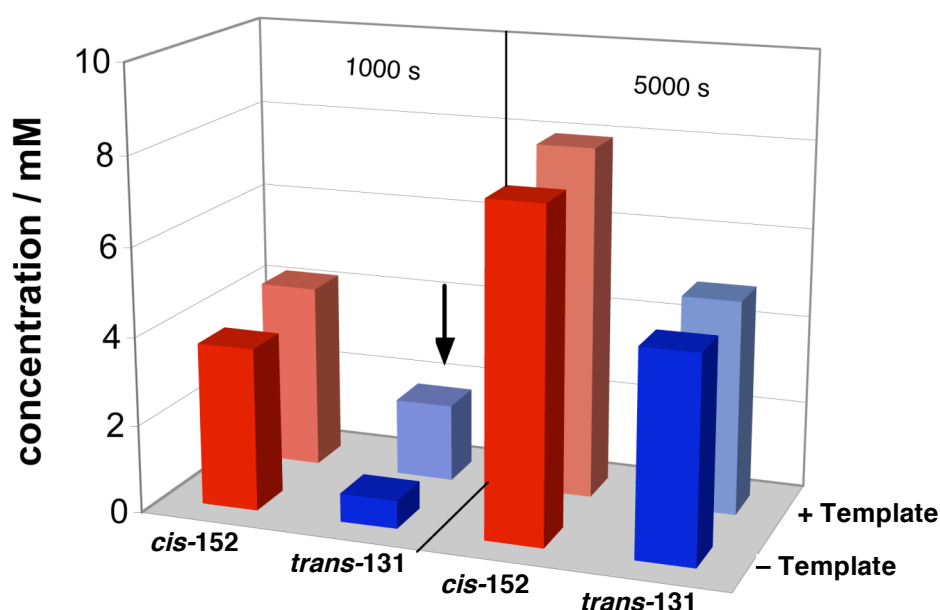
Figure 94. Comparison of the concentration of the two dominating products *cis*-**152** (■) and *trans*-**131** (■) in the system derived from maleimides **128** and **134** and nitrones **124** and **132** at 1000 and 5000 s. The doping of the system with preformed *trans*-**131** (+ Template) enhances the formation of this cycloadduct (arrow) most at the early stages of the experiment compared to the native, undoped reaction (– Template).

The comparison of the two dominating reaction channels reveals the biggest difference in concentration of the self-replicator *trans*-**131** between the native and the doped experiment is at the start of the experiment (Figure 94). The fact, that seeding

of the system with template **_trans_-131** mainly influences the initial rate of the replicator, explains this observation. Adding preformed template has a bigger influence on how early the maximum rate is reached, than on the rate itself.

The system, considering its structural simplicity, shows some very interesting properties. Recognition-mediated sorting is a phenomenon proving how a simple recognition motif can uncover the hidden properties of reacting molecules directing the evolution of the network. The selectivity in the mixture is achieved solely by kinetic properties of the reactions with two independent and mechanistically different pathways amplifying specific compounds. The template effect of the replicator allows for selectively addressing that single pathway without disrupting the performance of other reactions. All the processes in the system are based on the same type of reaction and the same type of recognition. Considering the similarity of all the components all the emergent properties of the network are the effect of length segregation introduced into the building blocks. It is possible to identify some distinct connections that are important in the design of such kind of networks. One of them is the direct template effect that can be used to selectively promote replicators. The other more subtle effect is the competition for building blocks, which is responsible for the promotion of the most efficient pathways in the system and elimination of the weakest. A given pathway has to incorporate the highest amount of components necessary for its 'survival' as early in the reaction as possible in order to achieve domination over the competition. The limitation in developing the described system is that the dominant AB complex pathway cannot be directly addressed, since the closed template configuration of **_cis_-152** is catalytically inert. The next step is to design a system consisting of only replicators of comparable efficiencies that could be selectively addressed with their unique templates. Length segregation is a promising methodology to achieve this goal, as templates of different length will selectively promote the replicators that utilise only the components that fit the given template.

3. Logic operations in a network of replicators

3.1 Length segregation of an efficient replicator

In order to develop a system of length-segregated replicators an efficient reaction which product displays self-replicating properties had to be first identified. Hopefully, the replicating properties would be conserved with the elongation of the building blocks. If all combinations of substrates with different lengths display templating properties such scenario would allow forming a fully addressable replicator-based system.

One of the most efficient replicating systems developed in our laboratory consists of the nitrone **124** as in the previous example, but a differently substituted acid maleimide **112**. The length segregation of the building blocks would hopefully give new reactive pairs with similar self-replicating activity. The maleimide was prepared from a commercially available 4-aminophenylacetic acid **153**, which was reacted with maleic anhydride **126** in acetic acid. In order to form the ester **154**, a reaction developed[249] by Madhusudhan was applied. The carboxylic acid **112** was injected into a suspension of silica gel supported $NaHSO_4$ in methanol affording the desired compound **154** (Figure 95).
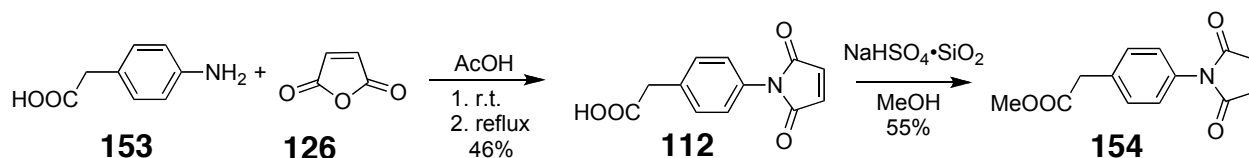

Figure 95. Synthesis of maleimides **112** and **154**.

The new maleimides **112** and **154** were reacted with the nitrone **124** in $CDCl_3$ to give rise to a new pair of cycloadducts (Figure 96).
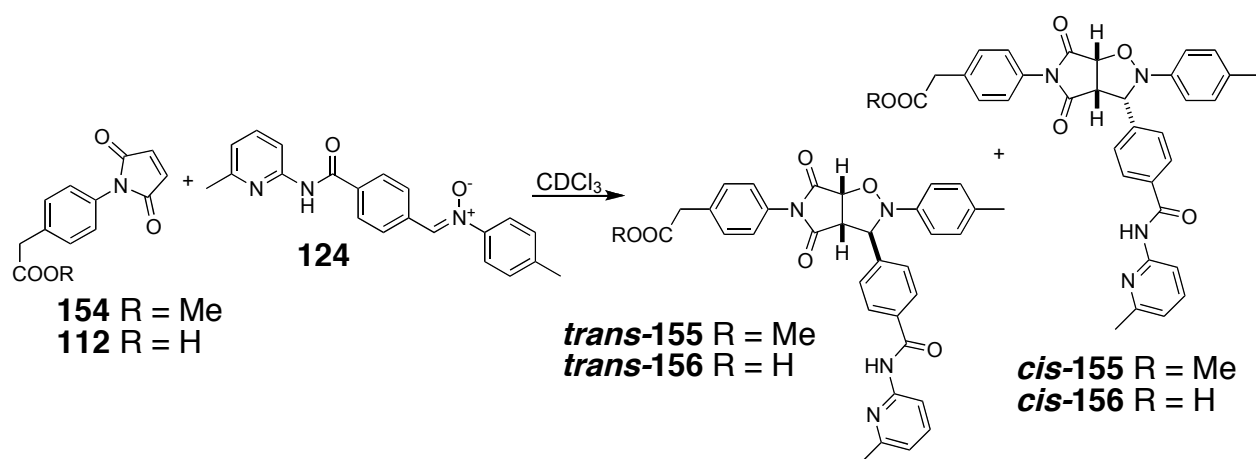
Figure 96. Products of 1,3-dipolar cycloaddition between maleimide **154** and **112** with nitrone **124**.

The control reaction without recognition present (Figure 97a) results in a slow formation of both diastereoisomers of **155** in 3.3:1 ratio in favour of the ***trans*-155** and a yield of less than 9% after 16 h. When recognition is present (Figure 97b) the reaction is rendered *trans* specific as no ***cis*-155** cycloadduct is detected in the reaction mixture. The concentration *vs.* time profile has a characteristic sigmoidal profile of the formation of ***trans*-156** with a slight lag period and reaches a conversion of 85% after 16 h.
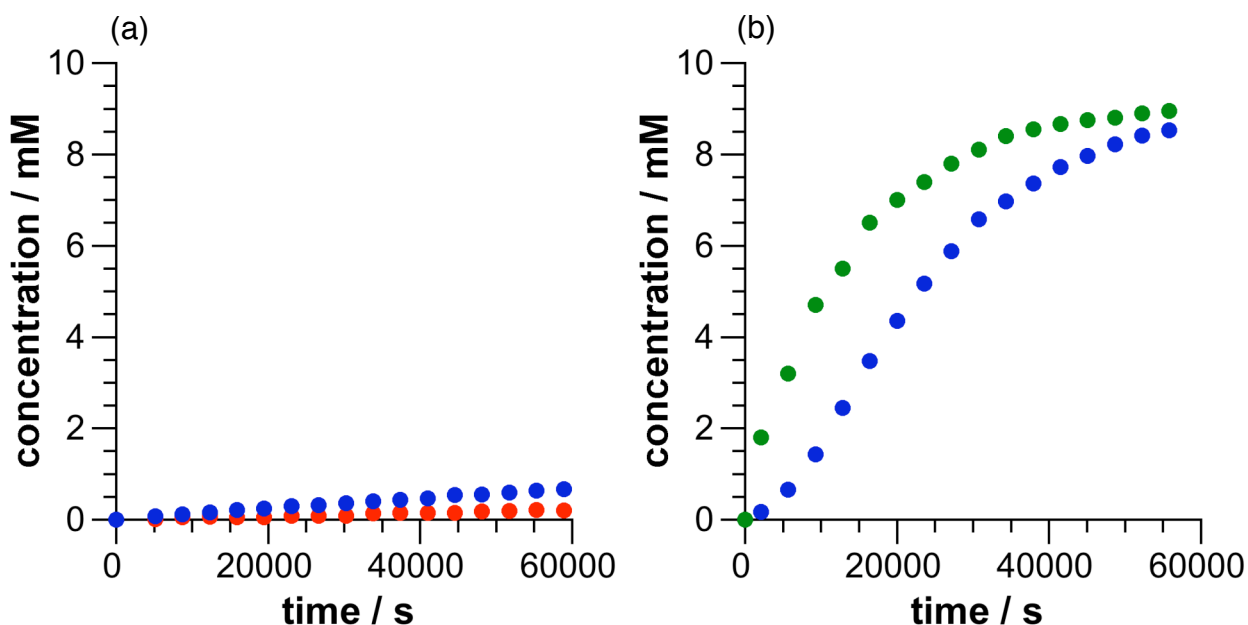
Figure 97. Concentration *vs.* time plots of (a) control reaction between ester maleimide **154** and nitrone **124** ([**154**] = [**124**] = 10 mM) following the formation of products: ***trans*-155** (●) and ***cis*-155** (●). (b) Recognition-enabled reaction between maleimide **112** and nitrone **124** ([**134**] = [**124**] = 10 mM) following the specific formation of product ***trans*-156** (●). The graph is overlaid with a concentration *vs.* time plot of the same reaction doped with 10 mol% of preformed ***trans*-156** ([**134**] = [**124**] = 10 mM, [***trans*-156**] = 1 mM) following the formation of ***trans*-156** (●).

The self-replicator **trans-156** shows a strong template effect. Upon addition of 10 mol% of preformed **trans-156** cycloadduct at the start of the experiment the lag period disappears and the reaction assumes its maximum autocatalytic rate from t = 0. The overall yield of the templated reaction reaches 90%. This data confirms that the reacting compounds **112** and **124** form an efficient replicating system.

The building blocks of the system were modified in order to achieve length segregation (Figure 98).
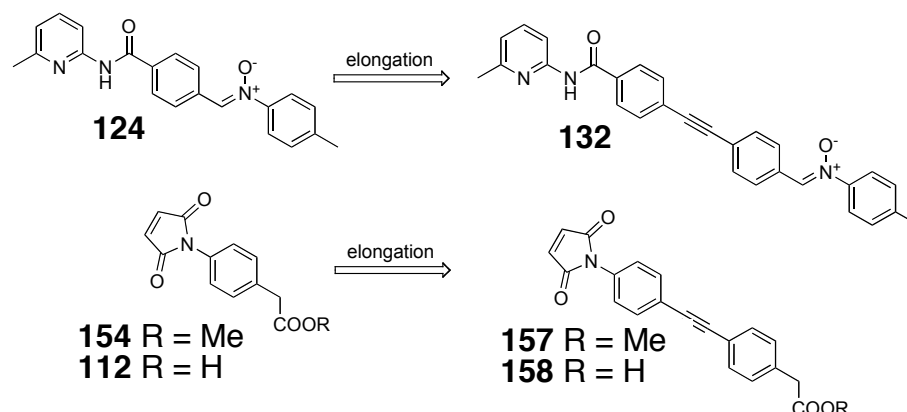
Figure 98. Synthetic design of elongated versions of compounds **124**, **154** and **112**.

The synthesis of the long nitrone **132** was already described (Figure 81). The long maleimide acid **158** and ester **157** were synthesised from appropriate starting materials. 4-Bromophenylacetic acid **159** was converted to a methyl ester **160** by standard acid catalysed reaction with methanol. The ester **160** was subsequently reacted with 4-ethynylaniline **135** under Sonogashira conditions. Addition of maleic anhydride **126** and cyclisation in acetic acid resulted in the formation of the control ester compound **157** (Figure 99).

Figure 99. Synthesis of elongated ester maleimide **157**.

In order to form the maleimide compound with the recognition site deprotected **158**, ester **161** was deprotected with lithium hydroxide to give the amino acid **163**. The approach to the formation of the maleimide group had to be modified since the previously used harsh conditions of cyclisation in acetic acid did not show satisfactory results. The formation of the elongated maleimide **158** was finally accomplished using EDC and HOBT in DMF. The conditions usually used for the formation of amides allow performing the reaction at room temperature. The yield was quite poor, but supplied sufficient amount of material to perform the necessary experiments (Figure 100).

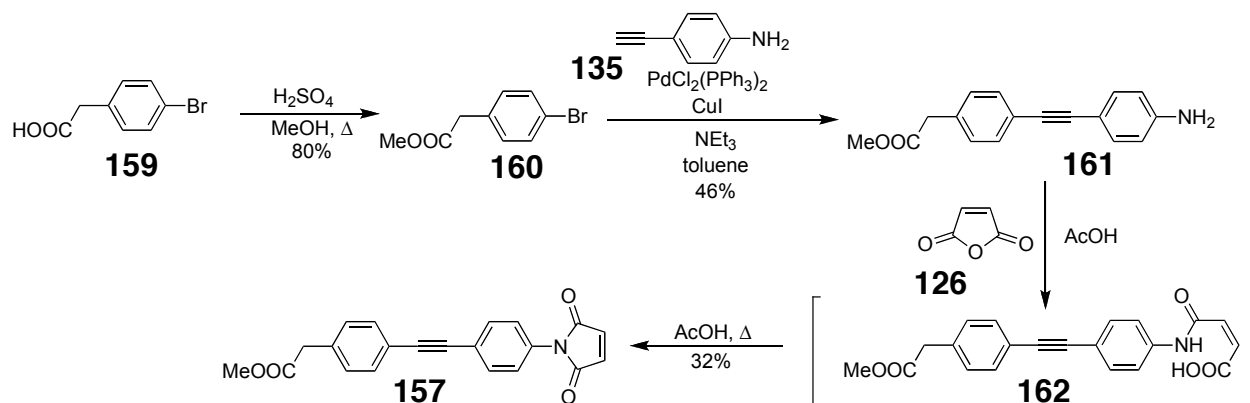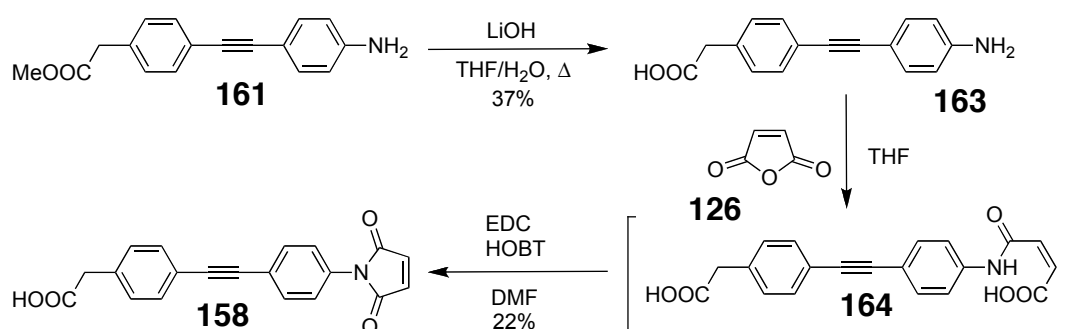**Figure 100.** Synthesis of elongated acid maleimide **158**.

The long components synthesised above together with the short building blocks are able to form four reactive pairs without recognition present and four pairs with recognition. The new reactions had to be assessed for their efficiency and selectivity in order to determine which reaction channel is dominant.

3.2 Kinetic investigation of length segregated reactive pairs

The reactions of the long components: maleimides **157** and **158** and nitrone **132,** were first investigated in order to assess their properties (Figure 101). In the previous system the long substrates **134** and **132** instead of forming a replicating system displayed AB complex pathway reactivity. The new long maleimide **158** has a different geometry and hopefully this change will bias the reaction outcome towards the replicating channel.

Figure 101. Products of 1,3-dipolar cycloaddition between maleimide **157** and **158** with nitrone **132**.

Figure 102. Concentration *vs.* time plots of (a) control reaction between ester maleimide **157** and nitrone **132** ([**157**] = [**132**] = 10 mM) following the formation of products: ***trans*-165** (●) and ***cis*-165** (●). (b) Recognition-enabled reaction between maleimide **158** and nitrone **132** ([**158**] = [**132**] = 10 mM) following the formation of products ***trans*-166** (●) and ***cis*-166** (●). The graph is overlaid with a concentration *vs.* time plot of the same reaction doped with 12 mol% of preformed ***trans*-166** ([**158**] = [**132**] = 10 mM, [***trans*-166**] = 1.2 mM) following the formation of ***trans*-166** (●).

The product of the reaction proved to be self-replicating. The control ([**157**] = [**132**] = 10 mM) (Figure 102a) is slow when no recognition is present and the final ratio of ***trans*-165** to ***cis*-165** is 3.9:1 with an overall conversion of 12%. When the reaction is run in the same conditions with the acid recognition site deprotected ([**158**] = [**132**]

106

= 10 mM) (Figure 102b) there is strong evidence of replicating activity of the *trans*-166 isomer. The profile has a characteristic sigmoidal shape, although it is clearly not as efficient as the short - short counterpart *trans*-156. There is an enhancement of the *cis*-166 isomer (0.3 mM in the control and 0.9 mM in the recognition-enabled reaction after 16 h) suggesting some AB complex pathway involvement. It is possible to investigate the response of the system to addition of preformed template. When 12 mol% of cycloadduct *trans*-166, derived from the two long components, is added to the solution with the starting materials ([**158**] = [**132**] = 10 mM, [*trans*-166] = 1.2 mM), the reaction has a significantly different concentration-time profile. There is no lag period, with the reaction assuming its maximum rate from t = 0, and a higher conversion is observed reaching 84% compared to 70% in the undoped reaction after 16 h. In this case the self-replicating properties are conserved with the elongation of the building blocks. This observation opens a possibility of the two replicators operating alongside each other in the same reaction medium. Since both the long-component and short-component systems respond to their particular templates each of the replicators could be addressed selectively within the system.

When both long and both short substrates are in solution more reactive pairs are possible than just between the components of the same length. The two pairs of the mismatched reactive partners also had to be assessed for their reactivity (Figure 103).
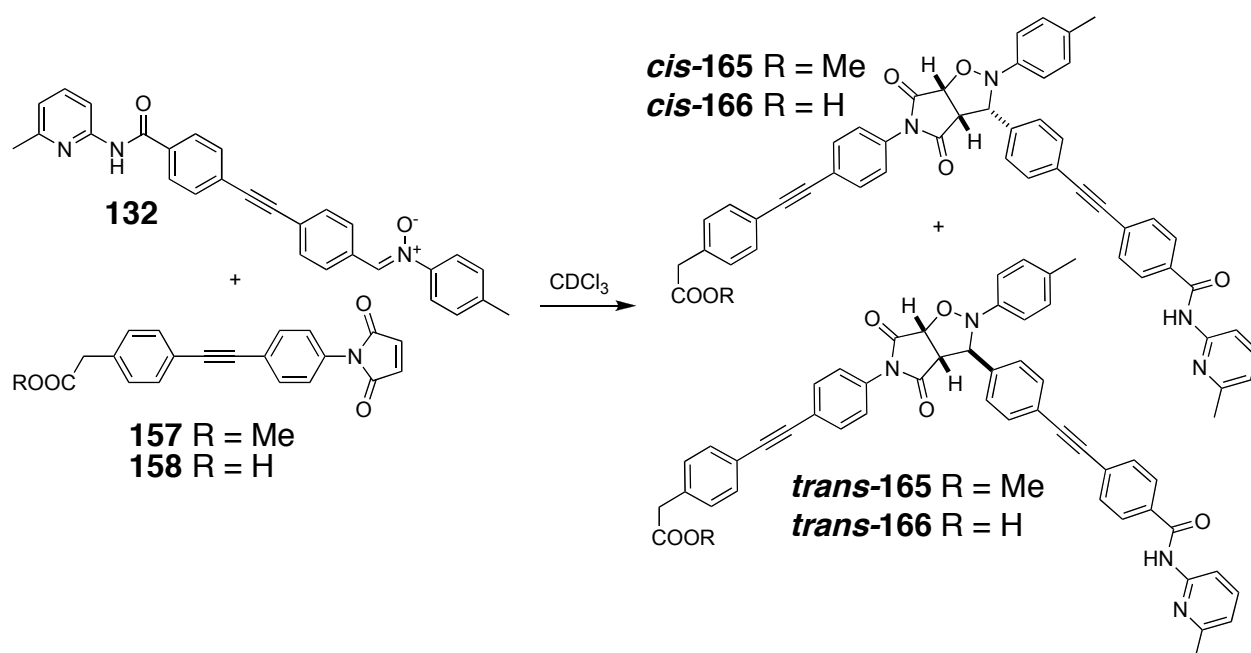

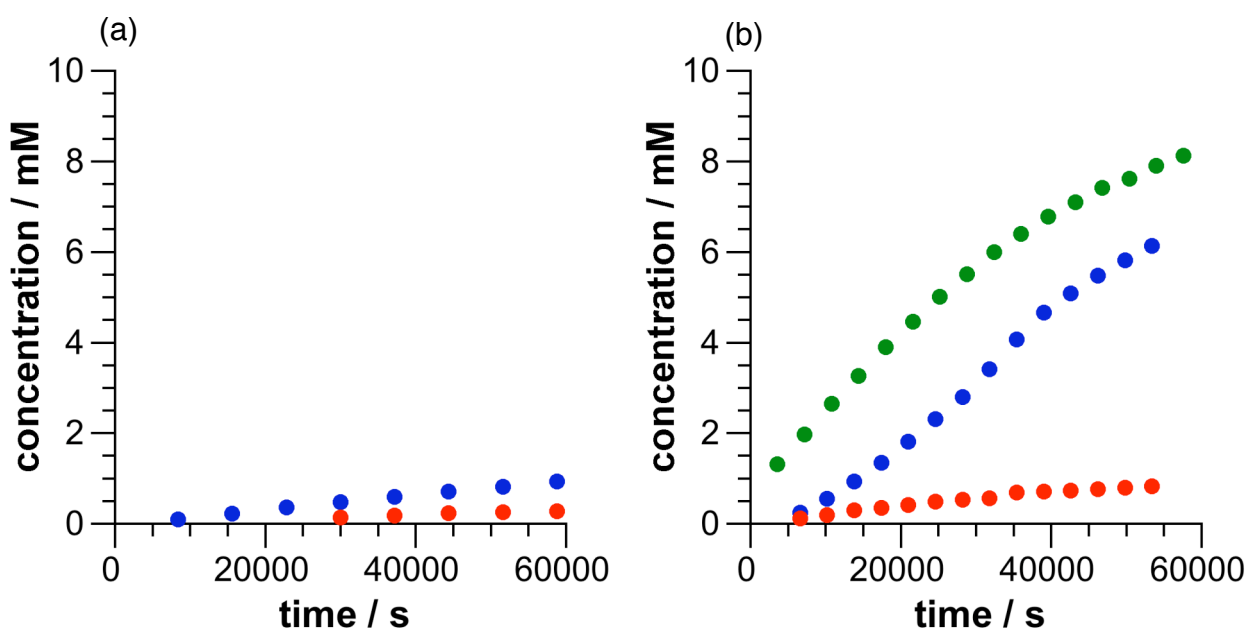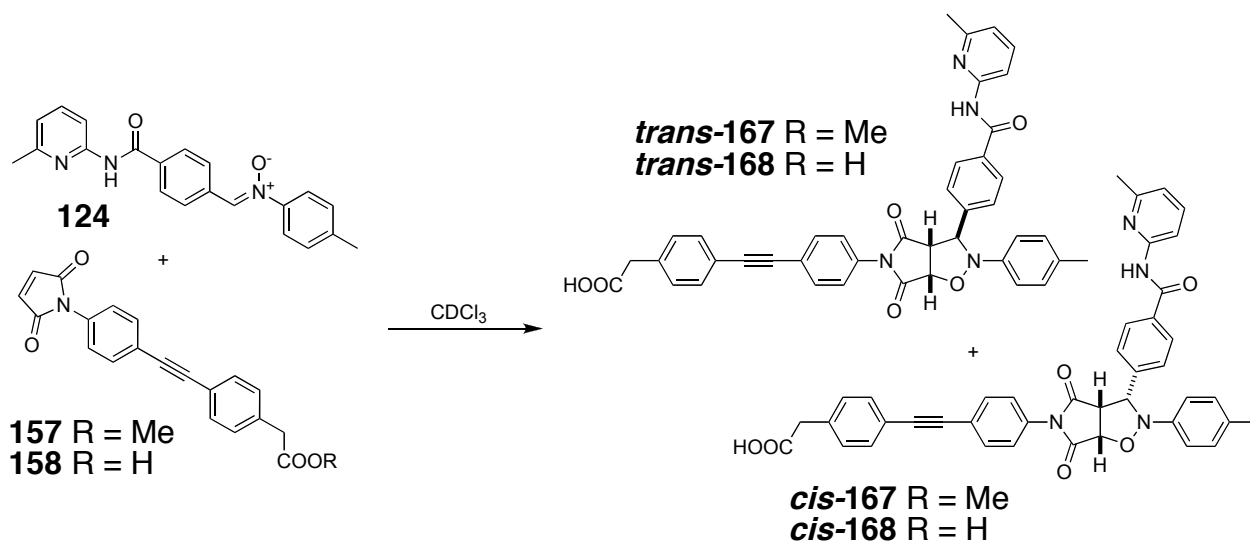**Figure 103.** Products of 1,3-dipolar cycloaddition between maleimide **157** and **158** with nitrone **124**.
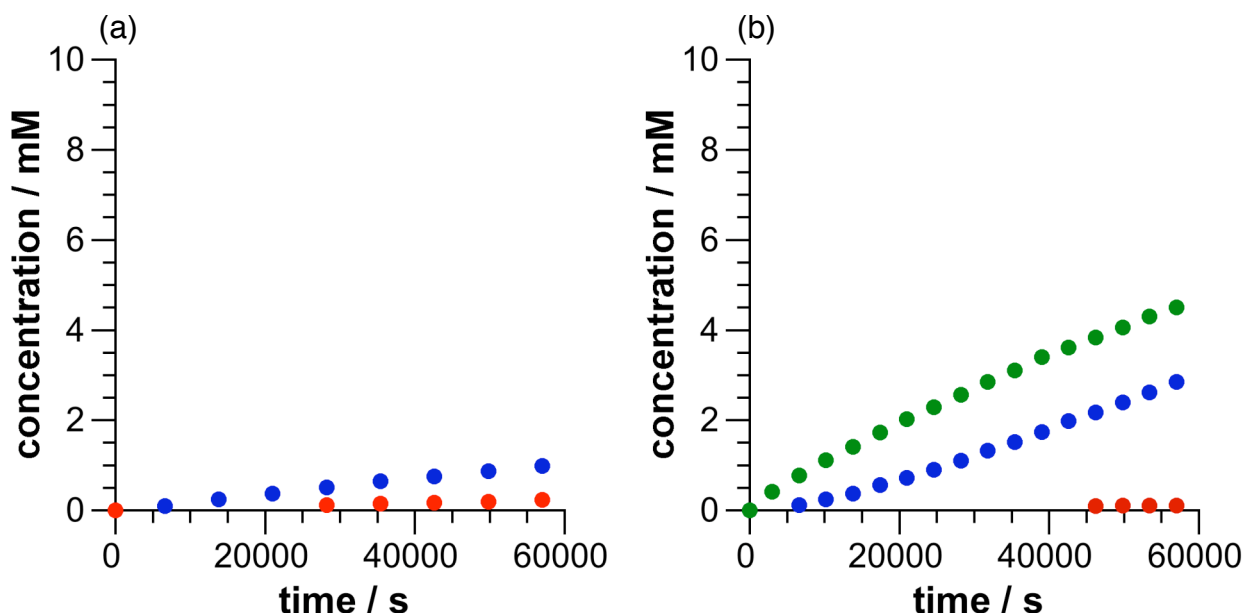
Figure 104. Concentration *vs.* time plots of (a) control reaction between ester maleimide **157** and nitrone **124** ([**157**] = [**124**] = 10 mM) following the formation of products: *trans*-**168** (●) and *cis*-**168** (●). (b) Recognition-enabled reaction between maleimide **158** and nitrone **124** ([**158**] = [**124**] = 10 mM) following the formation of products *trans*-**168** (●) and *cis*-**168** (●). The graph is overlaid with a concentration *vs.* time plot of the same reaction doped with 50 mol% of preformed *trans*-**168** ([**158**] = [**132**] = 10 mM, [*trans*-**166**] = 5 mM) following the formation of *trans*-**168** (●).

The reaction between the short nitrone **124** and the long maleimides: the ester **157** and the acid **158** was run in CDCl₃ at 273 K with the starting concentrations of the substrates at 10 mM. The control 1,3-dipolar cycloaddition reaction ([**157**] = [**124**] = 10 mM) was no different when compared to all previous reactions with the recognition blocked.

The reaction of the long acid maleimide **158** and the short nitrone **124** with recognition deprotected ([**158**] = [**124**] = 10 mM) shows an enhancement of formation of the *trans*-**168** isomer, although the lag period is very long and the conversion reaches only 2.9 mM. This enhancement is sufficient enough to suppress the formation of the *cis* isomer, which concentration reaches only about 0.1 mM. The addition of the preformed template *trans*-**168** with the substrates enhances the formation of the *trans*-**168** isomer even more. The reaction with 50 mol% of template *trans*-**168** is faster and concentration of the enhanced compound after 16 h reaches 4.5 mM. In this system some replicating activity is apparent, but the reaction is slower from the length-matched replicators *trans*-**156** and *trans*-**166**.

The other length-mismatched reactive pair is the short maleimide **112** and the long nitrone **132** (Figure 105).
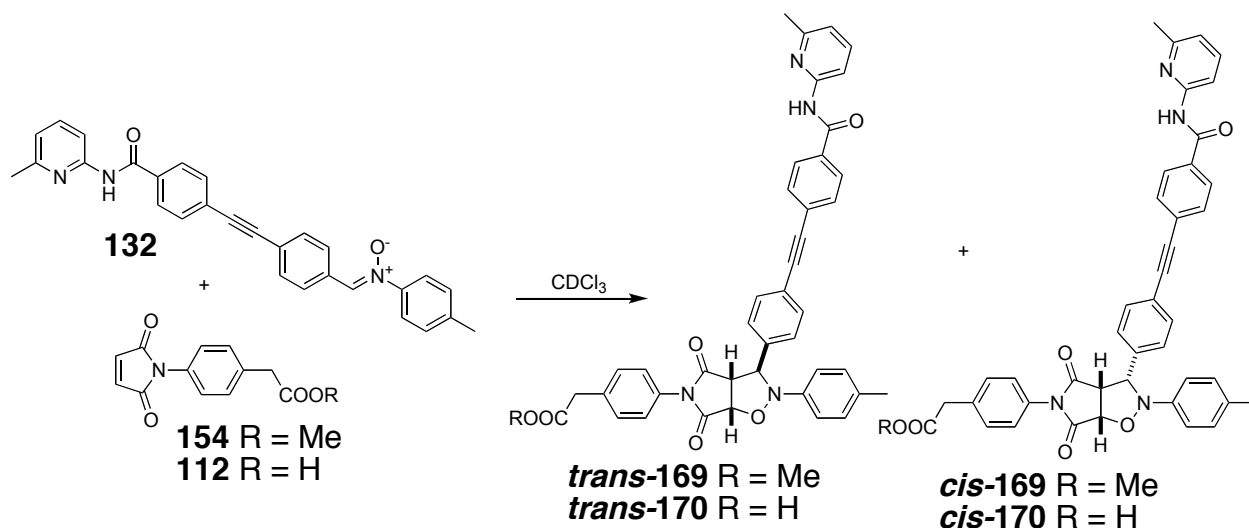
Figure 105. Products of 1,3-dipolar cycloaddition between maleimide **154** and **112** with nitrone **132**.
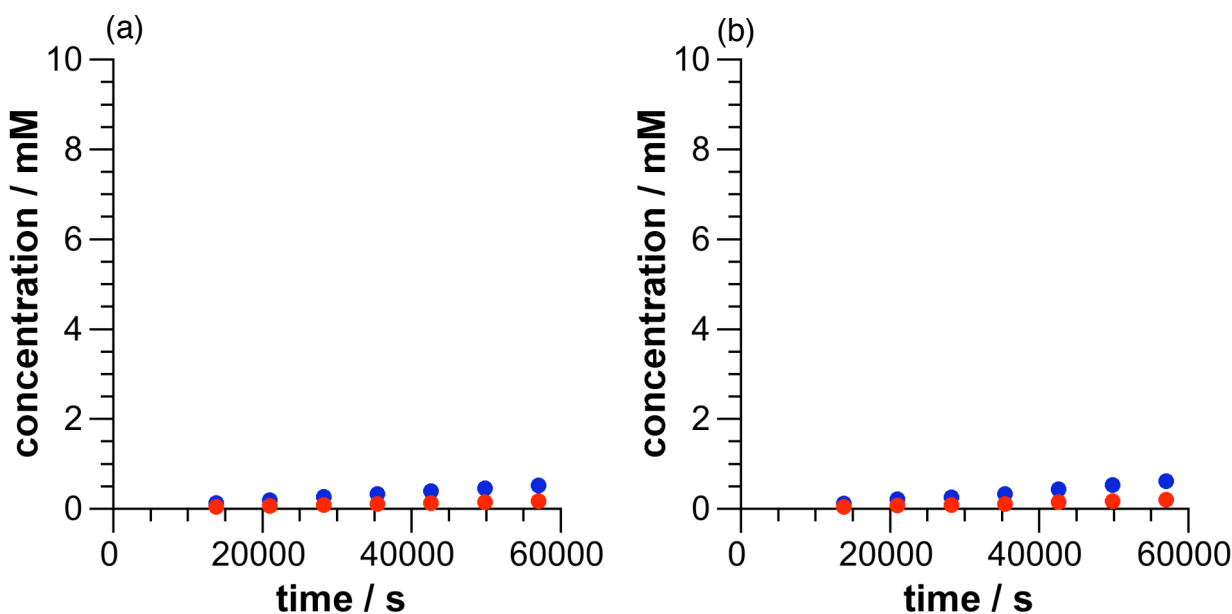
Figure 106. Concentration *vs.* time plots of (a) control reaction between ester maleimide **155** and nitrone **132** ([**155**] = [**132**] = 10 mM) following the formation of products: ***trans*-169** (●) and ***cis*-169** (●). (b) Recognition-enabled reaction between maleimide **112** and nitrone **132** ([**112**] = [**132**] = 10 mM) following the formation of products ***trans*-170** (●) and ***cis*-170**.

The reactions of the long nitrone **132** do not show any change of activity when comparing the control ([**155**] = [**132**] = 10 mM) (Figure 106a) and the recognition reaction ([**112**] = [**132**] = 10 mM) (Figure 106b). This result is surprising and yet unexplained. The first suspicion was that the system exhibits a very long lag period, but allowing the reaction to run for up to 3 days did not show any enhancement of the formation of any of the isomers. The system does not respond to the addition of the

template ***trans*-170**, which is also very insoluble once isolated from solution, making it difficult to dope the system with a sufficient amount.

Replicators base their reactivity on selective autocatalysis, with the template being the product of the reaction it catalyses. There is another possibility of a templated process, where the template cross-catalyses the formation of a different product. In order for such a process to exist the two cross-catalytic templates have to have a different, but complementary structure.

3.3 Cross-catalysis in a replicator network

There is no possibility of any cross-catalytic relationship between the long-component ***trans*-166** and short-component ***trans*-156** template. The products of the length-mismatched pairs ***trans*-168** and ***trans*-170** are of the same length so a possibility of cross-catalysis exists between the two products. Molecular modelling was employed to assess this possibility (Figure 107).

Figure 107. Modelling of the four trans isomer products derived from maleimides **112** and **158** and nitrones **124** and **132**.

The results make it apparent that the long maleimide - short nitrone ***trans*-168** and the short maleimide - long nitrone ***trans*-170** templates are of the same length but their cross-catalytic properties were to be proven experimentally (Figure 108).
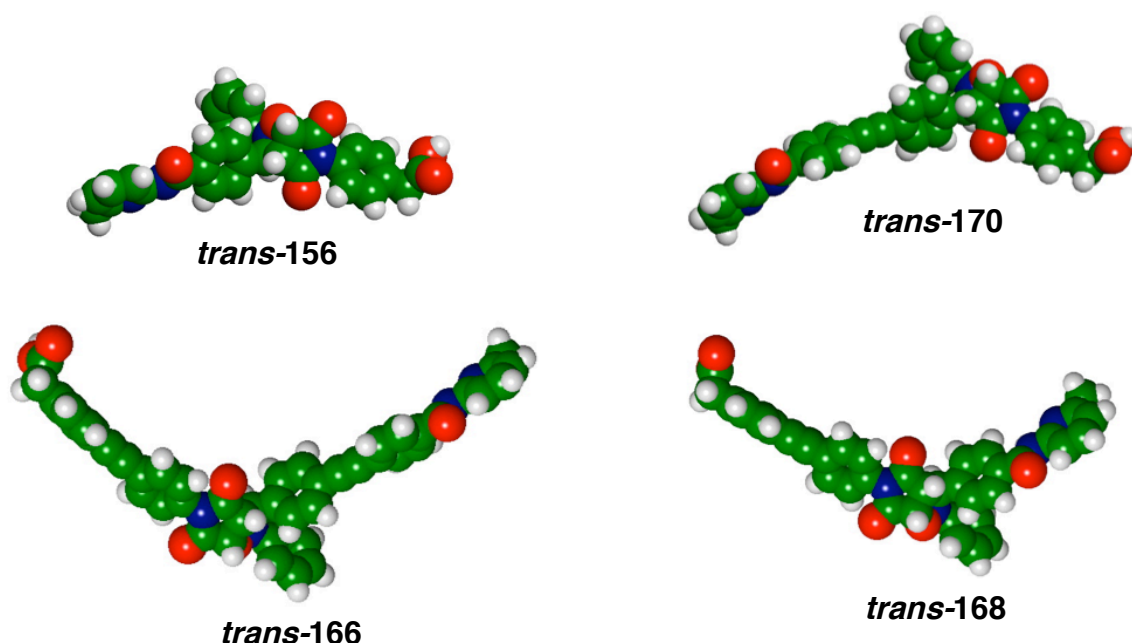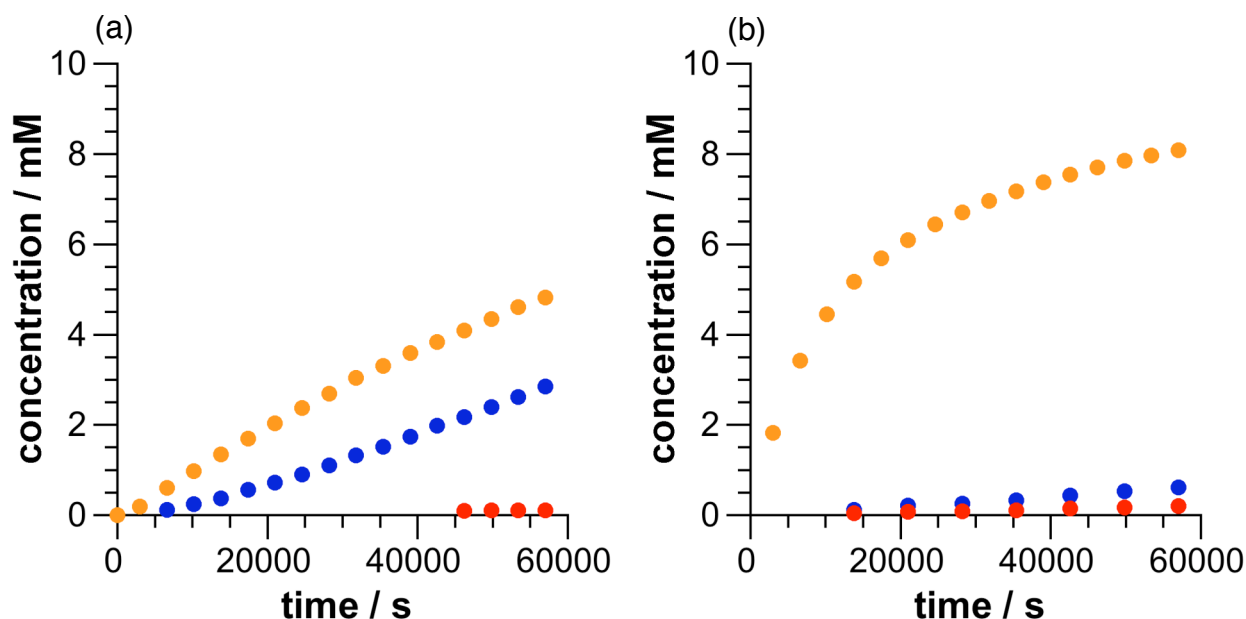
Figure 108. Concentration *vs.* time plots of (a) Recognition-enabled reaction between maleimide **158** and nitrone **124** doped with 2.5 mol% of *trans*-**170** ([**158**] = [**124**] = 10 mM, [*trans*-**170**] = 0.25 mM) following the formation of products *trans*-**168** (🟠) overlaid on a the plot of the undoped reaction between maleimide **158** and nitrone **124** ([**158**] = [**124**] = 10 mM) following the formation of products: *trans*-**168** (🔵) and *cis*-**168** (🔴). (b) Recognition-enabled reaction between maleimide **112** and nitrone **132** doped with 50 mol% of *trans*-**168** ([**112**] = [**132**] = 10 mM, [*trans*-**168**] = 5 mM) following the formation of products *trans*-**170** (🟠) overlaid on a the plot of the undoped reaction between maleimide **112** and nitrone **132** ([**112**] = [**132**] = 10 mM) following the formation of products: *trans*-**170** (🔵) and *cis*-**170** (🔴).

When the reaction between the long maleimide **158** and a short nitrone **124** is seeded with the template formed from the short maleimide **112** and the long nitrone **132** ([**157**] = [**124**] = 10 mM, [*trans*-**170**] = 0.25 mM) the effect is comparable to the one observed when an autocatalytic template *trans*-**168** is added (Figure 108a). Surprisingly, the rates are similar even though only 2.5 mol% of the cross-catalytic template *trans*-**170** is added and the previous experiment 50 mol% of the autocatalytic *trans*-**168** template was used as a dopant. This result indicates that, for efficient templates, only small amounts need be added since the system becomes saturated at higher loadings. In this case the cross-catalysis is just as efficient as the autocatalysis. The other pair, namely the short maleimide **112** and the long nitrone **132**, shows a more astounding cross-catalytic effect (Figure 108b). This system did not show any recognition-mediated activity on its own, however, when doped with 50 mol% of the preformed template *trans*-**168** ([**112**] = [**132**] = 10 mM, [*trans*-**168**] = 5 mM) the acceleration of the reaction is dramatic. This result is very encouraging when thinking of using the studied compounds to a form a more complex system,

ince there should an interesting connection between the two length-mismatched products **trans-168** and **trans-170** *via* their cross-catalytic properties. Also the long nitrone **132** - short maleimide **112** system that in the first instance seemed inactive, in the presence of its cross-catalytic partner is expected to show activity in the network.

3.4 Replicator-based reaction network

Having established the properties of the reactive pairs in isolation (Figure 109) it is possible to envisage the behaviour of the system as a whole, when all four building blocks **112**, **158**, **124** and **132** are present in solution.
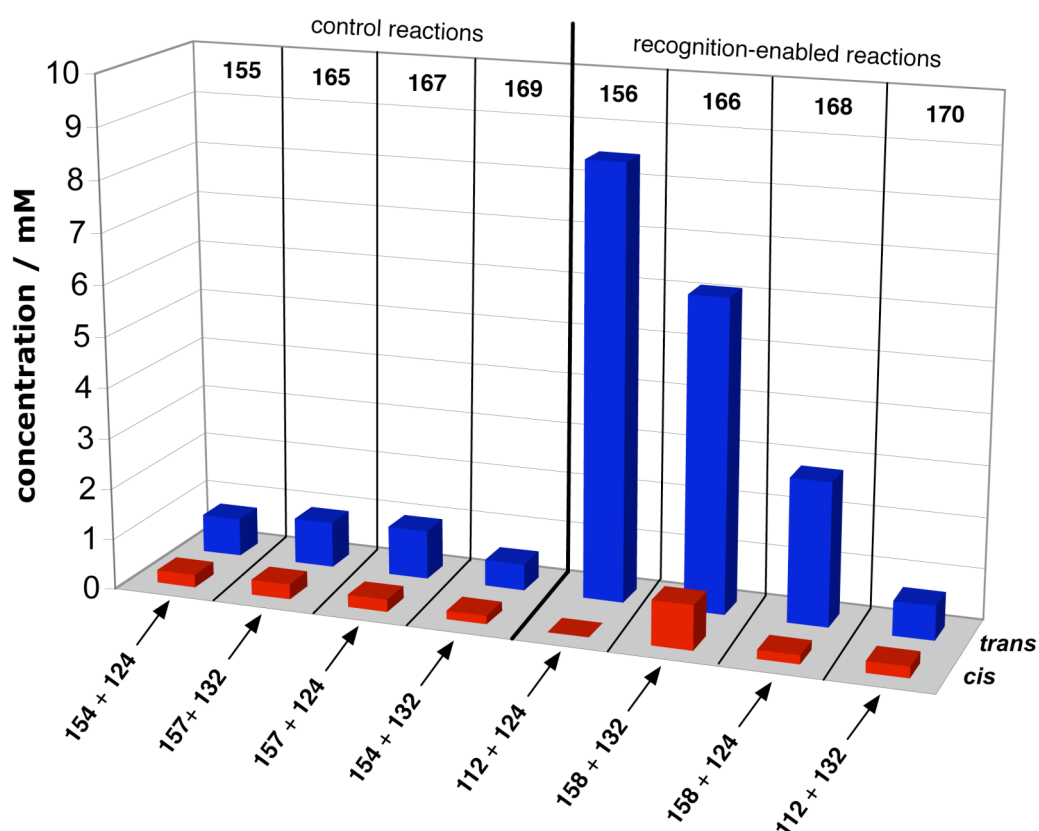
Figure 109. Concentrations of all cycloadducts derived from maleimides **154** and **157** and nitrones **124** and **132** (control) and maleimides **112** and **158** and nitrones **124** and **132** (recognition) after 16 h into the reaction. The *trans* isomers are represented by blue bars ( ) and the *cis* isomers by red bars ( ).

The short-component and the long-component pairs are expected to be involved in their autocatalytic cycles utilising templates **trans-156** and **trans-166**. The behaviour of the length-mismatched pairs would be more complex, since only **trans-168** is a replicator and there is a strong cross-catalytic connection between the two reactions. Competition for building blocks is going to be another important aspect governing the behaviour of the system. Although there is no template-directed relationship between

the length-matched and length-mismatched pairs, all systems use a limited number of common building blocks. For example if the short - short system efficiently incorporates the short components, both of the length-mismatched systems will suffer from lower concentration of starting material.

In the network each of the compounds with templating properties will be considered a node with three nodes being autocatalytic (Figure 110). There will be two different kinds of connections between the nodes: strong, direct template effect and a weaker building block competition effect. Three nodes are autocatalytic (*trans*-156, *trans*-166 and *trans*-168) and will respond positively to the addition of their specific template. The length-mismatched nodes are connected by a positive cross-catalytic edge. The input into one of these nodes will transfer the enhancement to the other one. The length-matched and the length-mismatched nodes are connected by negative edges that base on building block competition. The enhancement of one node will result in the suppression of the node through this type of connection. If the formation of a given node is suppressed then it could be expected that other nodes depending on the same building blocks will be enhanced since more starting material could be channelled to form that product.
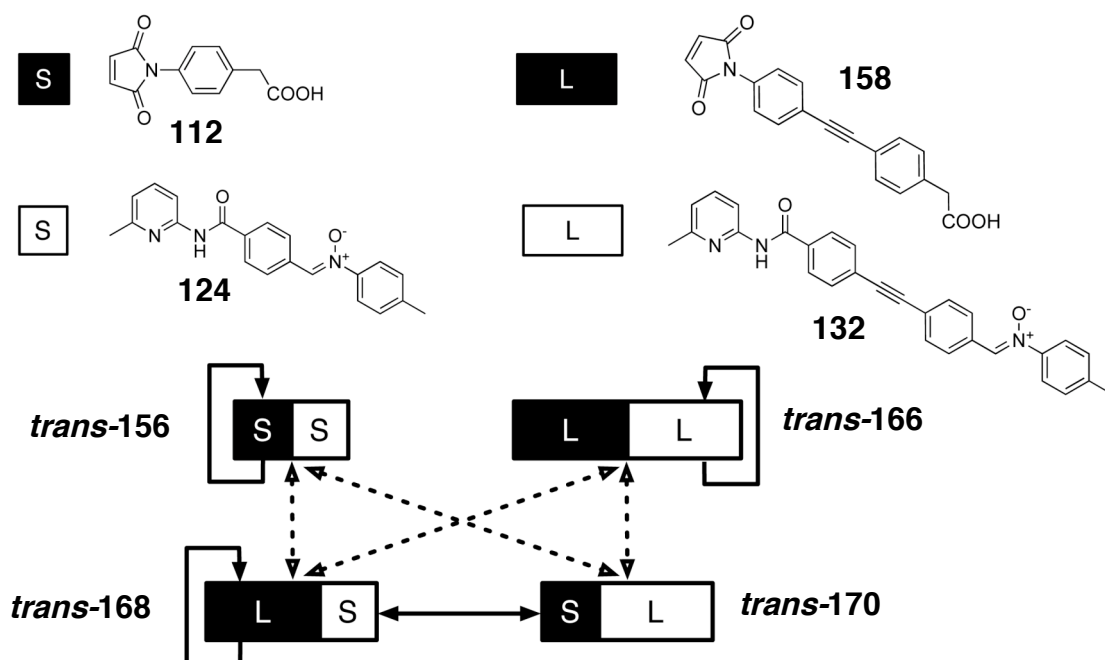
Figure 110. Schematic of the designed length segregated system. The building blocks **112, 158**, **124** and **132** can form four nodes. Three of the nodes SS, LL and LS possess self-replicating properties and positively respond to the addition of the product template (solid arrows). Nodes LS and SL are connected by a cross catalytic relationship (solid arrow): one node is enhanced in the presence of the other. The length-matched nodes SS and LL are connected with the length-mismatched nodes LS and SL by negative 'building block competition' edges (dashed arrows): enhancement of one node will lead to the suppression of the node connected through such edge.

3.5 One-input molecular logic in a reaction network

The system is expected to show properties of molecular logic. The addition of one of the four preformed templates is to be considered an input. The output of the system will be an overexpression of particular products (Figure 111). Which node is enhanced depends on the input and the architecture of the network.
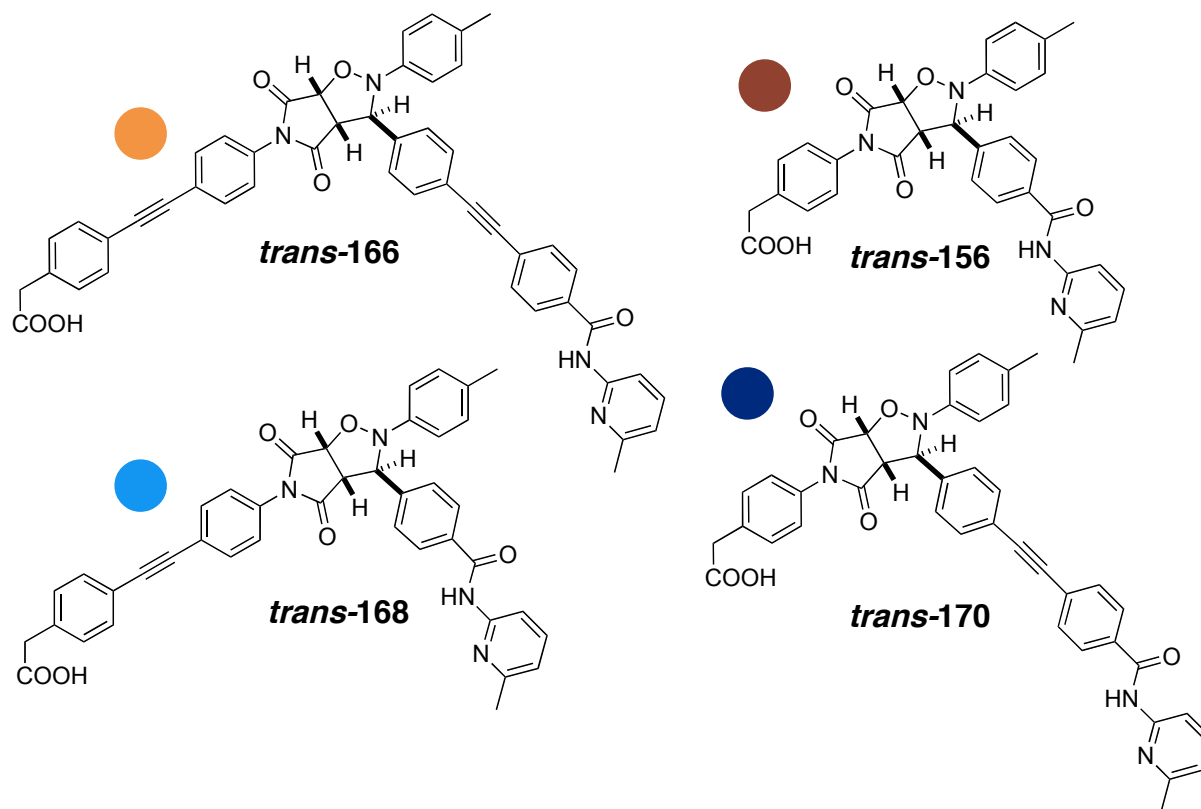
Figure 111. Structures of the four cycloadducts that describe the nodes of the chemical system. The structures are colour coded for future reference.

Since the *trans* isomers act as open templates for auto and cross-catalysis only these diastereoisomers will be used to describe the behaviour of the system. The *cis* isomers are present in the system, but only to a very limited extent and do not influence the network.

Firstly, the ground state of the system had to be identified as the coexistence of the nodes within the system without any added template. The four building blocks were allowed to react in CDCl$_3$ at 273 K with starting concentrations of 10 mM ([**112**] = [**158**] = [**124**] = [**132**] = 10 mM) (Figure 112).

Figure 112. Concentration *vs.* time plot of a recognition-enabled reaction between maleimides **112** and **158** and nitrones **124** and **132** ([**112**] = [**158**] = [**124**] = [**132**] =10 mM) following the formation of products *trans*-**156** (●), *trans*-**166** (●), *trans*-**168** (●) and *trans*-**170** (●).

The four *trans* cycloadducts *trans*-**156**, *trans*-**166**, *trans*-**168** and *trans*-**170** are formed with distinct sigmoidal-shaped concentration *vs.* time profiles. The shape of the reactive pathway suggests some kind of template-directed process for all the products. The short-component reaction forming *trans*-**156** was identified as the most efficient replicator in isolation and still is in the four-component experiment. The second best replicator in the isolated scenario was the long-component system *trans*-**166**, however, in the network it performs the worst. This inconsistency is easily explained, when the cross-catalytic relationship between the length-mismatched products *trans*-**168** and *trans*-**170** is realised. The long maleimide - short nitrone *trans*-**168** reaction is both autocatalytic and cross-catalytic making it more efficient in the network than on its own. The short maleimide - long nitrone *trans*-**170** reaction even though is not autocatalytic shows a dramatic enhancement of rate in the presence of the other length-mismatched template *trans*-**168**. The practically identical reaction profile of the formation of these two cycloadducts shows how strong is the connection between the two nodes.

The next set of experiments was to determine how the system would respond to the addition of different preformed templates *trans*-**156**, *trans*-**166**, *trans*-**168** and *trans*-**170**. In order to describe the system an enhancement factor was defined as the log of the ratio between the concentrations the products product in the doped reaction and the undoped reaction (Figure 112) after 16 h. If the formation of a given

cycloadduct is enhanced the amplification factor is positive, if the concentration is lower than in the ground state the enhancement factor is negative.

$$\text{enhancement factor} = \log \frac{[\text{product}]_{\text{doped}}}{[\text{product}]_{\text{undoped}}}$$

In order to test the logic response of the system, 30 mol% of the short component template was added together with the four building blocks ([**112**] = [**158**] = [**124**] = [**132**] = 10 mM, [*trans*-**156**] = 3 mM) (Figure 113). The doped replicator is suspected to dominate in the network, since the short maleimide - short nitrone reaction displays a strong template effect. The reactions are monitored using 500 MHz ^{1}H NMR spectroscopy. The use of this analytical technique is very convenient although there are limitations. The reactions are run at 10 mM concentrations of starting materials so the products, especially at the early stages of the experiment, often fall below the level of detection. Also, since the four cycloadducts are of similar structure it is common for their corresponding signals to overlap in the NMR spectrum. These drawbacks result in some data points missing in the concentration *vs.* time plots describing the experiments.
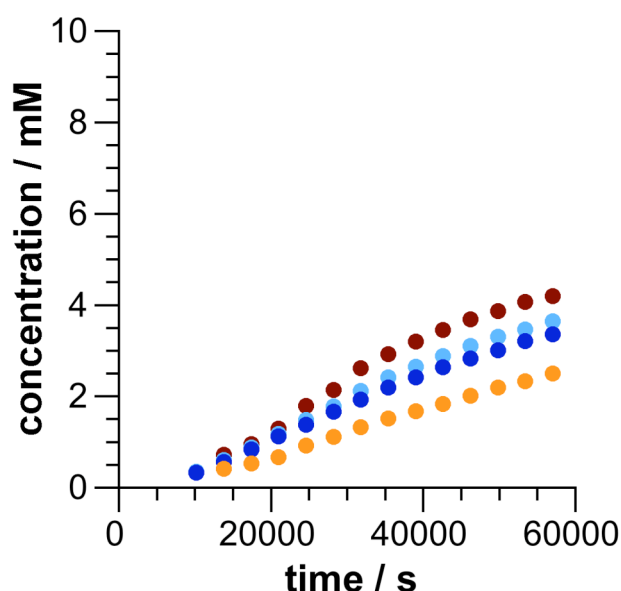

Figure 113. (a) Concentration *vs.* time plot of a recognition-enabled reaction between maleimides **112** and **158** and nitrones **124** and **132** doped with 30 mol% of preformed *trans*-**156** ([**112**] = [**158**] = [**124**] = [**132**] = 10 mM, [*trans*-**156**] = 3 mM) following the formation of products *trans*-**156** (●), *trans*-**166** (●), *trans*-**168** (●) and *trans*-**170** (●). (b) Enhancement factors for products *trans*-**156** (■), *trans*-**166** (■), *trans*-**168** (■) and *trans*-**170** (■).

The short component replicator ***trans*-156** is definitely the dominant reaction pathway in the system. Since the short component pool (compounds **112** and **124**) is rapidly depleted, the two mismatched products ***trans*-168** and ***trans*-170** are suppressed, as they need the short building blocks as well. Surprisingly the conversion of the long component product ***trans*-166** is enhanced even though this system is not directly doped with template. This enhancement is explained by a higher abundance of the long building blocks in solution. Since there are no short substrates for the length-mismatched reactions the long components are not used by these subsystems and are left behind to react with each other.

The addition of the template could be considered as an input and the overexpression of specific compounds as the output. In this particular system we have four possible inputs and four possible outputs. Using binary logic the addition of a template will be considered an input of 1. The enhancement of any of the template will be considered an output of 1 and no change or suppression an output of 0. This assignment allows for the creation of truth tables for our system (Figure 114).

trans	156	166	168	170
IN	1	0	0	0
OUT	1	1	0	0

Figure 114. A truth table and a graphic scheme for the system doped with ***trans*-156**. The input of 1 for ***trans*-156** generates an output of 1 for ***trans*-156** and ***trans*-166**.

In order to evaluate the logic behaviour of the system, other inputs had to be tested. The next template to be added was the long-component ***trans*-166** cycloadduct ([**112**] = [**158**] = [**124**] = [**132**] = 10 mM, [***trans*-166**] = 1.5 mM). Again, the doped product was predicted to dominate the reaction mixture (figure 115).
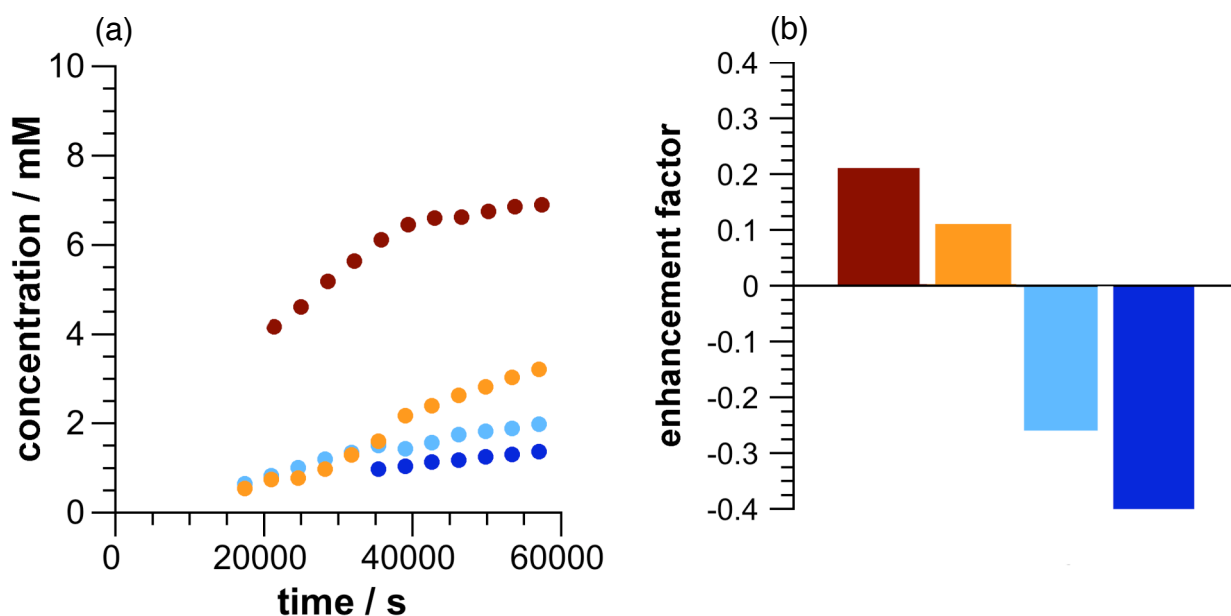
Figure 115. (a) Concentration *vs.* time plot of a recognition-enabled reaction between maleimides **112** and **158** and nitrones **124** and **132** doped with 15 mol% of preformed ***trans*-166** ([**112**] = [**158**] = [**124**] = [**132**] = 10 mM, [***trans*-166**] = 1.5 mM) following the formation of products ***trans*-156** (●), ***trans*-166** (●), ***trans*-168** (●) and ***trans*-170** (●). (b) Enhancement factors for products ***trans*-156** (■), ***trans*-166** (■), ***trans*-168** (■) and ***trans*-170** (■).

As a result of solubility issues only 15 mol% of the long component template ***trans*-166** was added to the system, but still this product dominates the mixture. The long cycloadduct ***trans*-166** reaches the highest concentration of 5.6 mM after 16 h, but its advantage over the next best reaction is not as large as that of the short product ***trans*-156** in the previous example (Figure 113). First of all, the amount of added template ***trans*-166** is half that as ***trans*-156** added in the first experiment and secondly, the long-component replicator is not as efficient in isolation as the short-component one. For the same reasons the length-mismatched products ***trans*-168** and ***trans*-170** are suppressed to a lesser degree than in the experiment with the short component template as the input. The input of the long component template ***trans*-166** generates the same output as the short component template ***trans*-156**. The addition of any of the length-matched templates will enhance their own formation and the formation of the other length-matched product even though there is no direct connection between the two subsystems. The influence is generated across the network through the subtle relationships that base on the competition for building blocks (Figure 116).
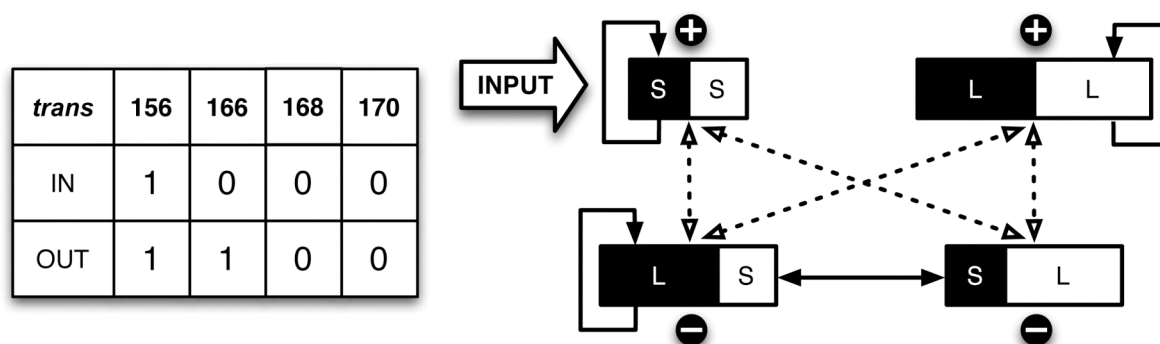
Figure 116. A truth table and a graphic scheme for the system doped with ***trans*-166**. The input of 1 for ***trans*-166** generates an output of 1 for ***trans*-156** and ***trans*-166**.

The next experiment was to show what influence on the system the long maleimide - short nitrone template ***trans*-168** would have. This compound is the only template that exhibits both auto- and cross-catalytic properties. The expected outcome of the experiment is an enhancement of both of the length-mismatched products ***trans*-168** and ***trans*-170** given their cross-catalytic properties (Figure 117).
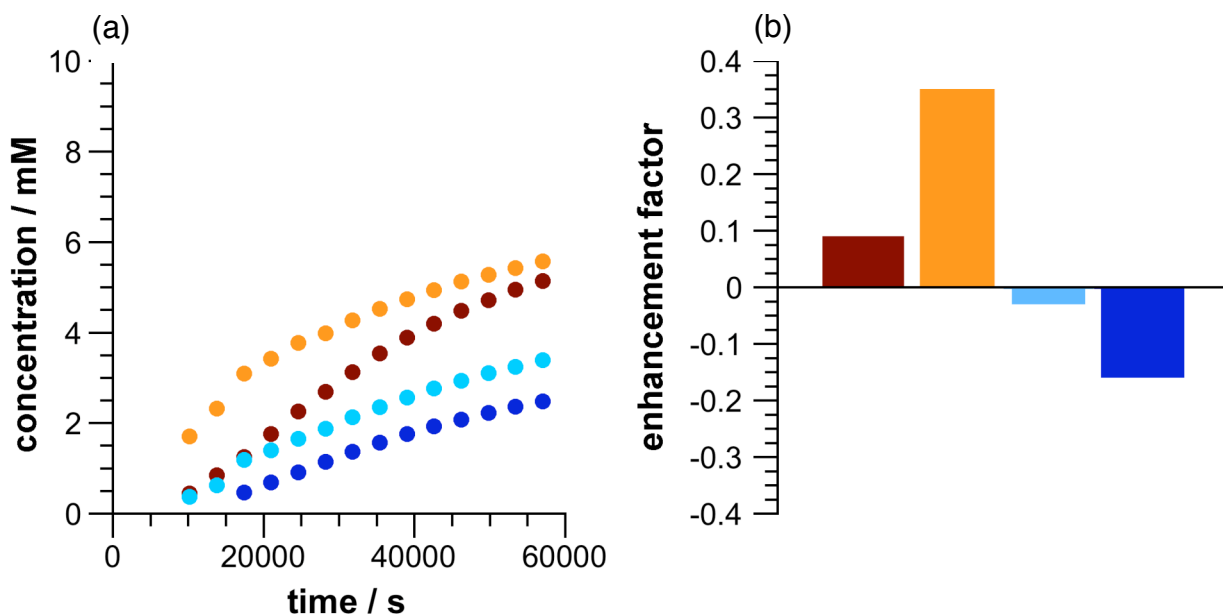

Figure 117. (a) Concentration *vs.* time plot of a recognition-enabled reaction between maleimides **112** and **158** and nitrones **124** and **132** doped with 30 mol% of preformed ***trans*-168** ([**112**] = [**158**] = [**124**] = [**132**] = 10 mM, [***trans*-168**] = 3 mM) following the formation of products ***trans*-156** (●), ***trans*-166** (●), ***trans*-168** (●) and ***trans*-170** (●). (b) Enhancement factors for products ***trans*-156** (■), ***trans*-166** (■), ***trans*-168** (■) and ***trans*-170** (■).

In this experiment 30 mol% of template ***trans*-168** was added together with the four building blocks **168** ([**112**] = [**158**] = [**124**] = [**132**] = 10 mM, [***trans*-168**] = 3 mM) and the evolution of the system was monitored by ^{1}H NMR spectroscopy. This time we observe the disappearance of the lag period for both of the length-mismatched

products ***trans*-168** and ***trans*-170**, since the added template catalyses the formation of both cycloadducts. All short and long substrates are channelled to form the two templated products, resulting in a dramatic suppression of both of the length-matched cycloadducts ***trans*-156** and ***trans*-166**. The cross-catalysis is such an efficient process that even the short component subsystem ***trans*-156**, which is one of the most efficient replicators reported to date, loses in the competition for building blocks. The truth table for the specific input reveals that the addition of one length-mismatched template enhances to formation of both length-mismatched products (Figure 118).
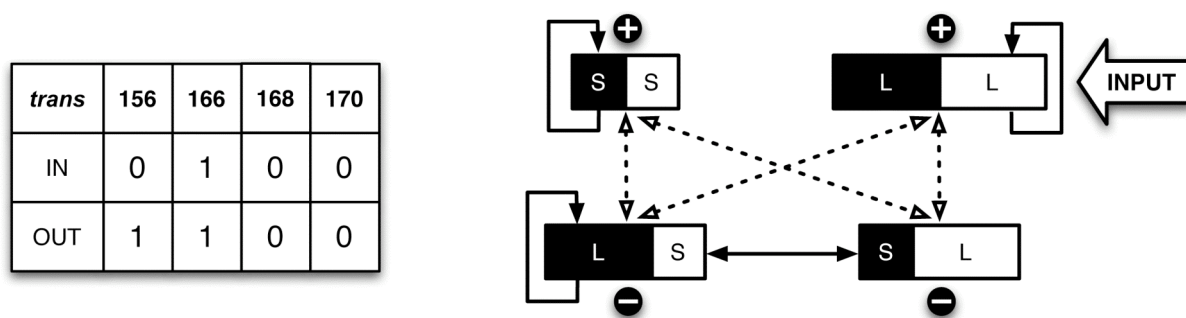

Figure 118. A truth table and a graphic scheme for the system doped with ***trans*-168**. The input of 1 for ***trans*-168** generates an output of 1 for ***trans*-168** and ***trans*-170**.

A problem arose when trying to evaluate the influence of the short maleimide - long nitrone template ***trans*-170** on the system. As mentioned previously this particular compound is difficult to redissolve once isolated from solution and only 4 mol% could be added with the other building blocks ([**112**] = [**158**] = [**124**] = [**132**] = 10 mM, [***trans*-170**] = 0.4 mM). This drawback may result in a smaller response than in the previously described scenarios (Figure 119).
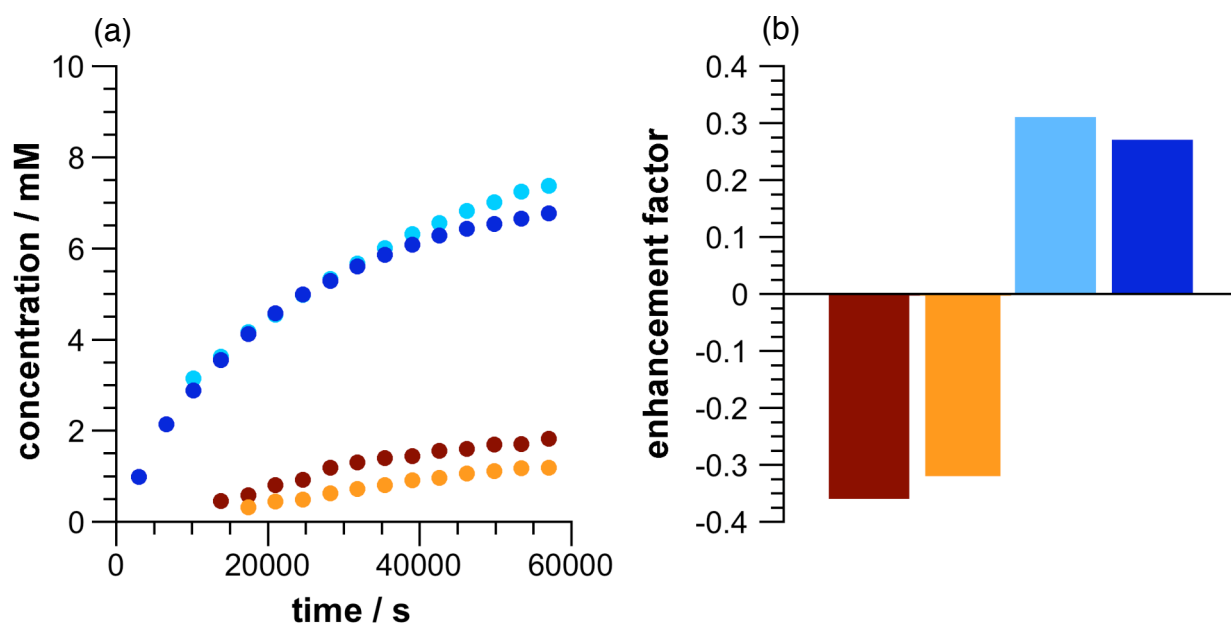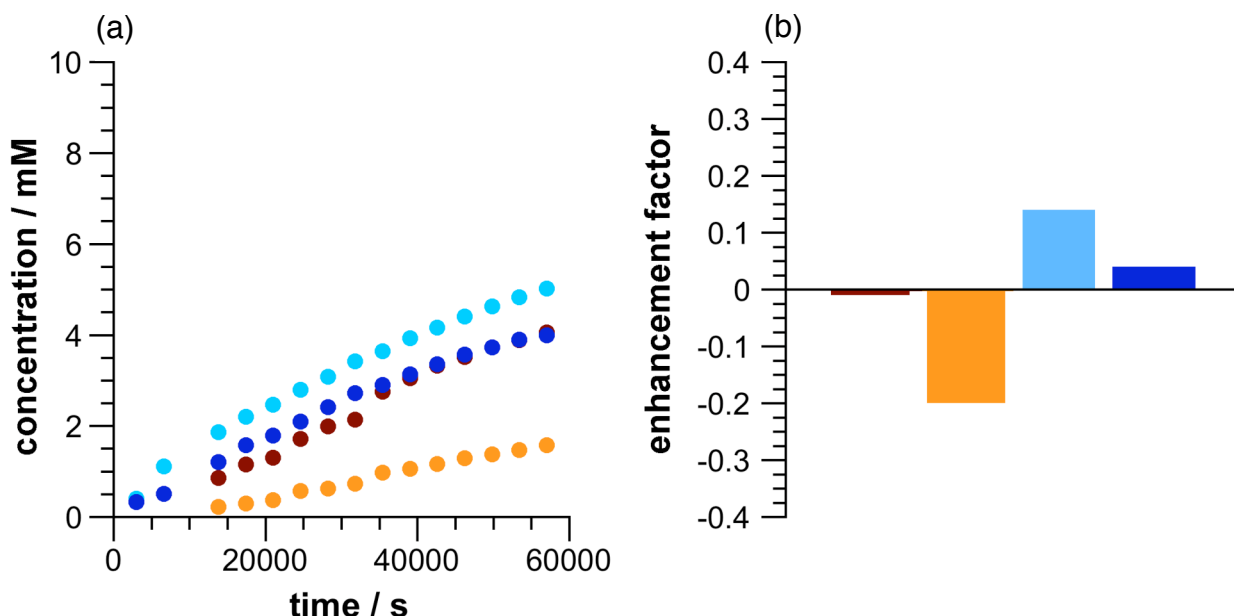
Figure 119. (a) Concentration *vs.* time plot of a recognition-enabled reaction between maleimides **112** and **158** and nitrones **124** and **132** doped with 4 mol% of preformed *trans*-**170** ([**112**] = [**158**] = [**124**] = [**132**] = 10 mM, [*trans*-**170**] = 0.4 mM) following the formation of products *trans*-**156** (●), *trans*-**166** (●), *trans*-**168** (●) and *trans*-**170** (●). (b) Enhancement factors for products *trans*-**156** (■), *trans*-**166** (■), *trans*-**168** (■) and *trans*-**170** (■).

In this experiment, the templated subsystem does not show the highest enhancement. This result is explained by the lack of autocatalytic properties of the short maleimide - long nitrone compound *trans*-**170** and a small loading of template. The long maleimide - short nitrone product *trans*-**168** exhibits the most efficient formation since it is cross-catalysed by the added template *trans*-**170** and subsequently engages into its own autocatalytic cycle at the same time cross-catalysing over to the other length-mismatched product *trans*-**170**. The slight enhancement of the doped product does not result from direct template effect, but from the cross-catalysis. This process is much less efficient resulting in just a very small suppression of the efficient short-component replicator *trans*-**156**. Analysing the truth table (Figure 120) reveals that the input of either of the length-mismatched templates results in the enhancement of both of products linked by their cross-catalytic properties.

trans	156	166	168	170
IN	0	0	0	1
OUT	0	0	1	1

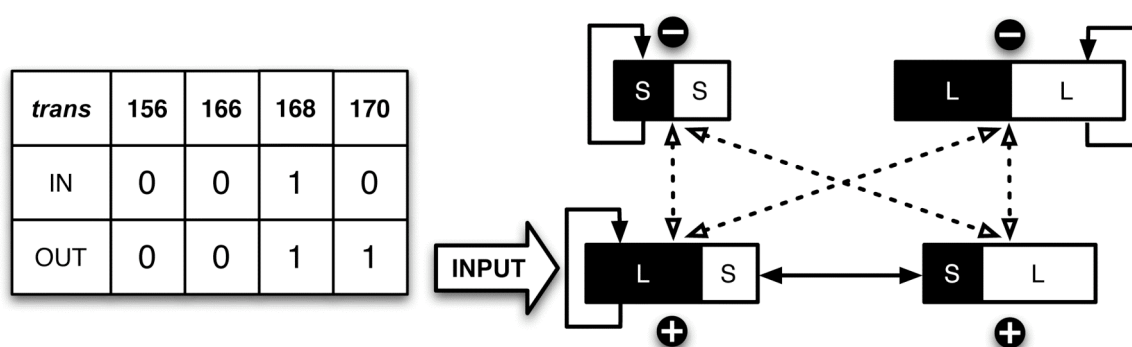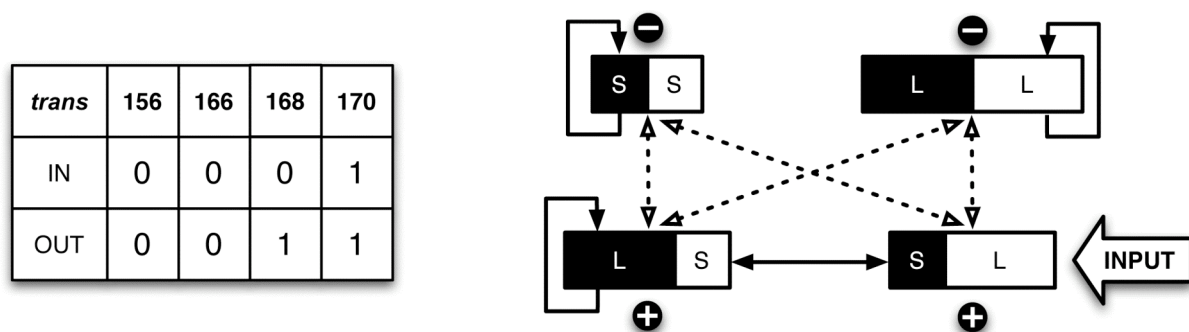**Figure 120.** A truth table and a graphic scheme for the system doped with *trans*-170. The input of 1 for *trans*-170 generates an output of 1 for *trans*-168 and *trans*-170.

When there is only a single input, the response of the system is very logical. Since there are four possible inputs and four possible outputs the behaviour of the network cannot be described as a single logic gate. The way to picture the system is as a circuit built of two bifurcated OR logic gates, with a separate gate for the length-matched and the length-mismatched subsystems (Figure 121).
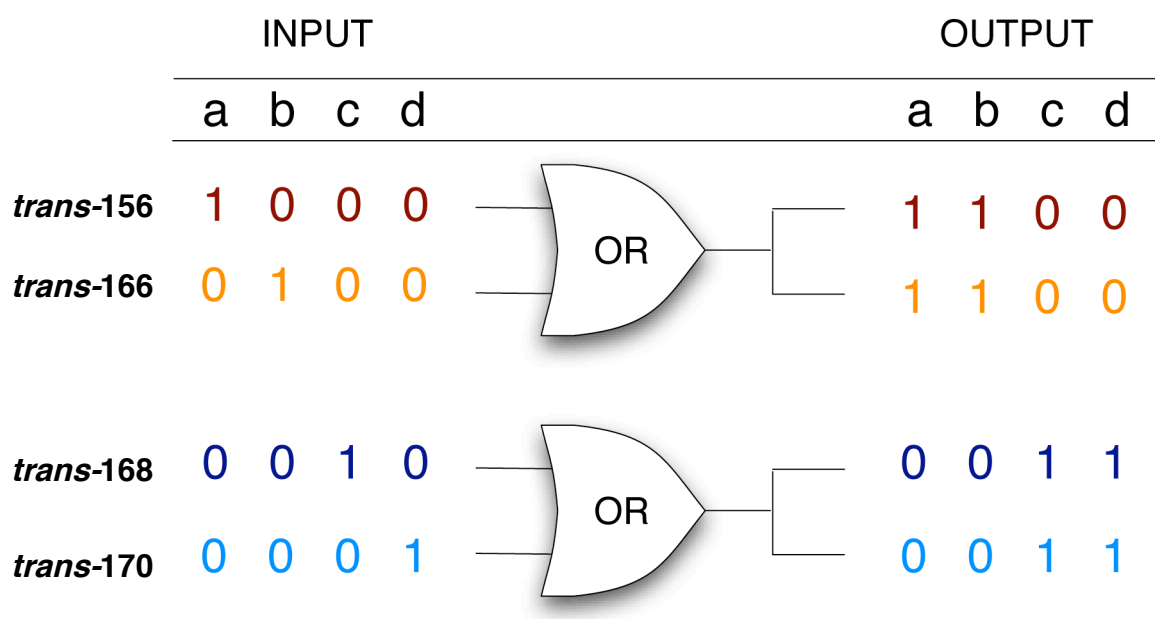
Figure 121. Truth table describing the system as a circuit made of two bifurcated OR logic gates. The columns represent the input and output in particular experiments described in: (a) Figure 113, (b) Figure 115, (c) Figure 117, (d) Figure 119.

The ground state of the system (Figure 112) relies on all reactions being recognition-mediated: either auto or cross-catalytic. In such situation the template effect will be a strong influence deciding the outcome of the system. The experiments with single template input consistently prove the structure of the network displaying the template properties as well as the connections through building block competition.

3.6 Two-input molecular logic in a reaction network

The response of the system is suspected to become more complicated to explain when there is more than one input. The input by definition is quantified, either a one or a zero, and none should be stronger than other. The predictions in the described system, with more than one input, are less accurate because the effects of different templates are of different weights. A number of experiments were conducted with two template inputs, however, problems with solubility resulted in different concentrations of templates added together with the starting materials. Some interesting results and trends were still observed.

The most obvious response of the system is expected when both the short-component *trans*-**156** and the long-component *trans*-**166** templates where added. Since there is no template relationship and no building block competition between them, both of the length-matched subsystems should be amplified by the expense of the length-mismatched products (Figure 122).

Figure 122. (a) Concentration *vs.* time plot of a recognition-enabled reaction between maleimides **112** and **158** and nitrones **124** and **132** doped with 15 mol% of preformed *trans*-**156** and 30 mol% of *trans*-**166** ([**112**] = [**158**] = [**124**] = [**132**] = 10 mM, [*trans*-**156**] = 1.5 mM, [*trans*-**166**] = 3 mM) following the formation of products *trans*-**156** (●), *trans*-**166** (●), *trans*-**168** (●) and *trans*-**170** (●). (b) Enhancement factors for products *trans*-**156** (■), *trans*-**166** (■), *trans*-**168** (■) and *trans*-**170** (■).

Upon addition of 15 mol% of the short-component template *trans*-**156** and 30 mol% of the long-component template *trans*-**166** to the four component system ([**112**] =

123

[**158**] = [**124**] = [**132**] = 10 mM, [*trans*-**156**] = 1.5 mM, [*trans*-**166**] = 3 mM) the disappearance of the lag period for both length-matched products is observed. Both products reach practically the same concentration, but it is the long maleimide - long nitrone template *trans*-**166** that is amplified more, mostly due twice the amount of added template. The output confirms the predictions that since each length-matched template added on its own catalyses both length-matched product, then the addition of both templates will influence the same selectivity in the system.

A more interesting result is expected when the input consists of two templates that give different outputs. Templating one of the length-matched replicators should result in a strong enhancement of the chosen subsystem, however the cross-catalytic relationship between the length-mismatched products establishes their strong position in the network.

In this experiment 30 mol% of the short component *trans*-**156** template and 20 mol% of the soluble long maleimide - short nitrone *trans*-**168** template were added with the starting materials ([**112**] = [**158**] = [**124**] = [**132**] = 10 mM, [*trans*-**156**] = 3 mM, [*trans*-**168**] = 2 mM) (Figure 123).
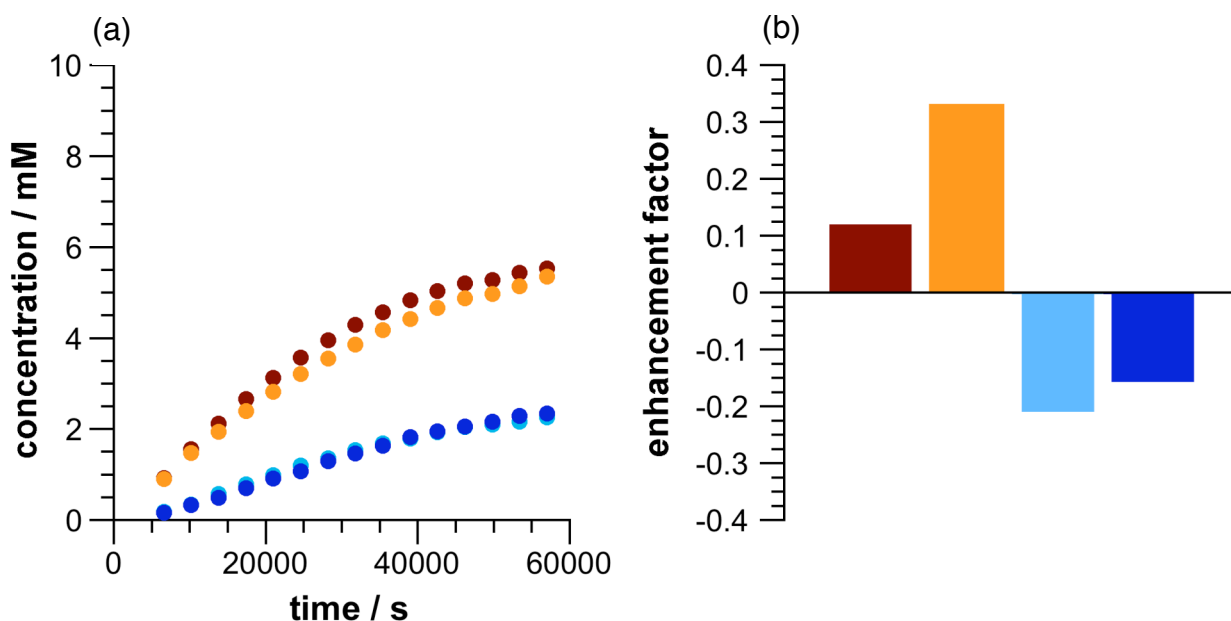

Figure 123. (a) Concentration *vs.* time plot of a recognition-enabled reaction between maleimides **112** and **158** and nitrones **124** and **132** doped with 30 mol% of preformed *trans*-**156** and 20 mol% of *trans*-**168** ([**112**] = [**158**] = [**124**] = [**132**] = 10 mM, [*trans*-**156**] = 3 mM, [*trans*-**168**] = 2 mM) following the formation of products *trans*-**156** (●), *trans*-**166** (●), *trans*-**168** (●) and *trans*-**170** (●). (b) Enhancement factors for products *trans*-**156** (■), *trans*-**166** (■), *trans*-**168** (■) and *trans*-**170** (■).

Since the added length-mismatched cycloadduct *trans*-168 templates both length-mismatched products, the only untemplated process is the reaction of the long components to form *trans*-166. The addition of just the short-component template *trans*-156 on its own (Figure 113) resulted in a slight enhancement of the long-component product *trans*-166, but in this example this fine relationship is broken and the long-component product *trans*-166 is suppressed to the expense of all other templated processes.

The system was doped with 30 mol% of the long-component template *trans*-166 and 40 mol% of the long maleimide - short nitrone template *trans*-168 ([112] = [158] = [124] = [132] = 10 mM, [*trans*-166] = 3 mM, [*trans*-168] = 4 mM) (Figure 124).
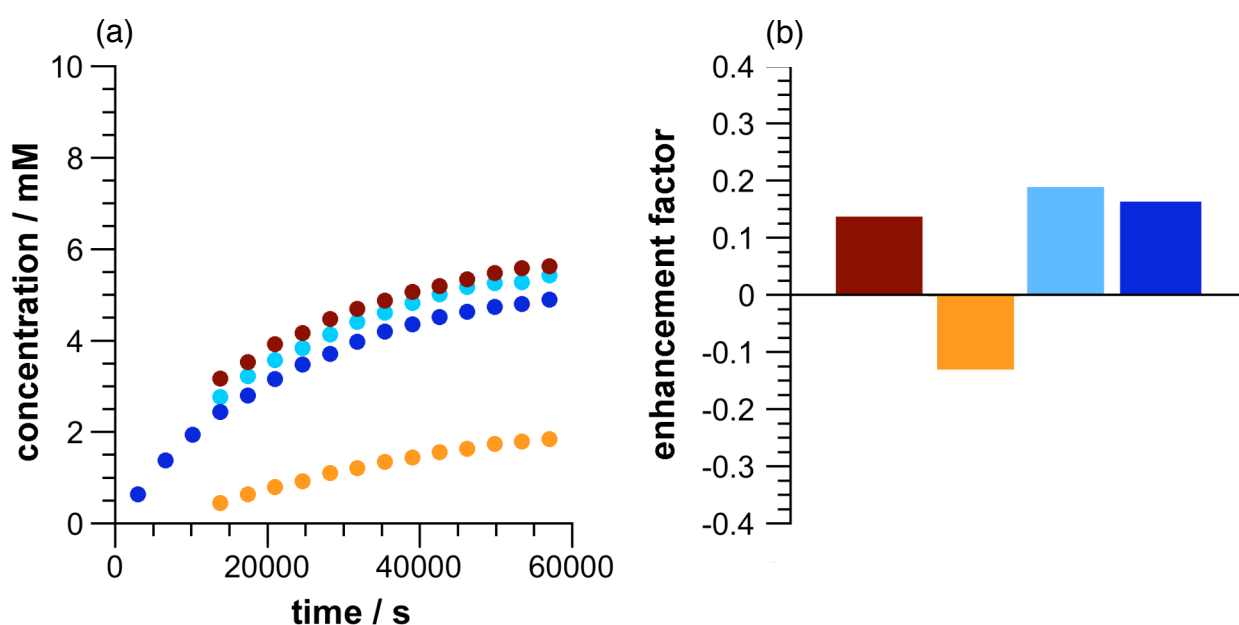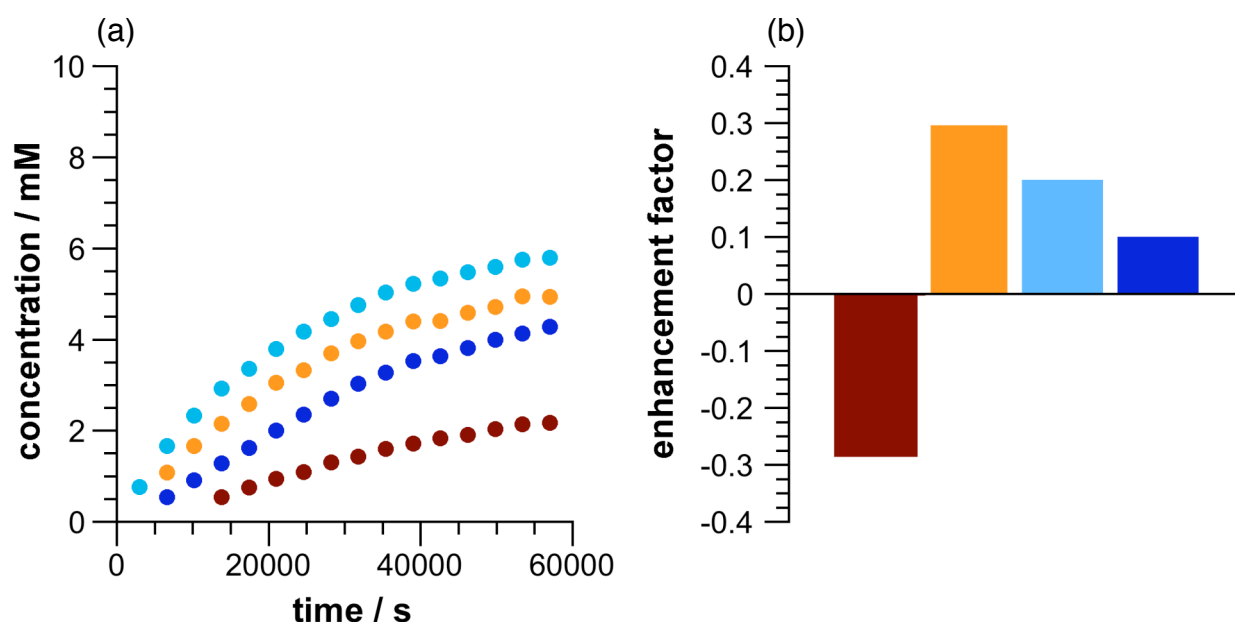

Figure 124. (a) Concentration *vs.* time plot of a recognition-enabled reaction between maleimides **112** and **158** and nitrones **124** and **132** doped with 30 mol% of preformed *trans*-166 and 40 mol% of *trans*-168 ([112] = [158] = [124] = [132] = 10 mM, [*trans*-166] = 3 mM, [*trans*-168] = 4 mM) following the formation of products *trans*-156 (●), *trans*-166 (●), *trans*-168 (●) and *trans*-170 (●). (b) Enhancement factors for products *trans*-156 (■), *trans*-166 (■), *trans*-168 (■) and *trans*-170 (■).

The situation is similar to the one described in the previous example with the untemplated subsystem suffering the greatest suppression, only this time it is the short-component replicator *trans*-156. The addition of the long maleimide - short nitrone template *trans*-168 always results in the enhancement of both of the length-mismatched products. Short maleimide - long nitrone *trans*-170 template was not tested in a two-input scenario because of the limited solubility of this compound.

In a two-input scenario the template effects are so strong that the more subtle

connections, based on competition for building blocks, are no longer detected in the network. If we employ the logic gate methodology to picture the system with two possible inputs the circuit will consist of two YES gates and an OR gate. The part of the graph responsible for the length-mismatched subsystems does not change. The length-matched inputs, however, display just a simple YES logic with the input being equal to the output. The two length-matched systems now operate independently of each other.
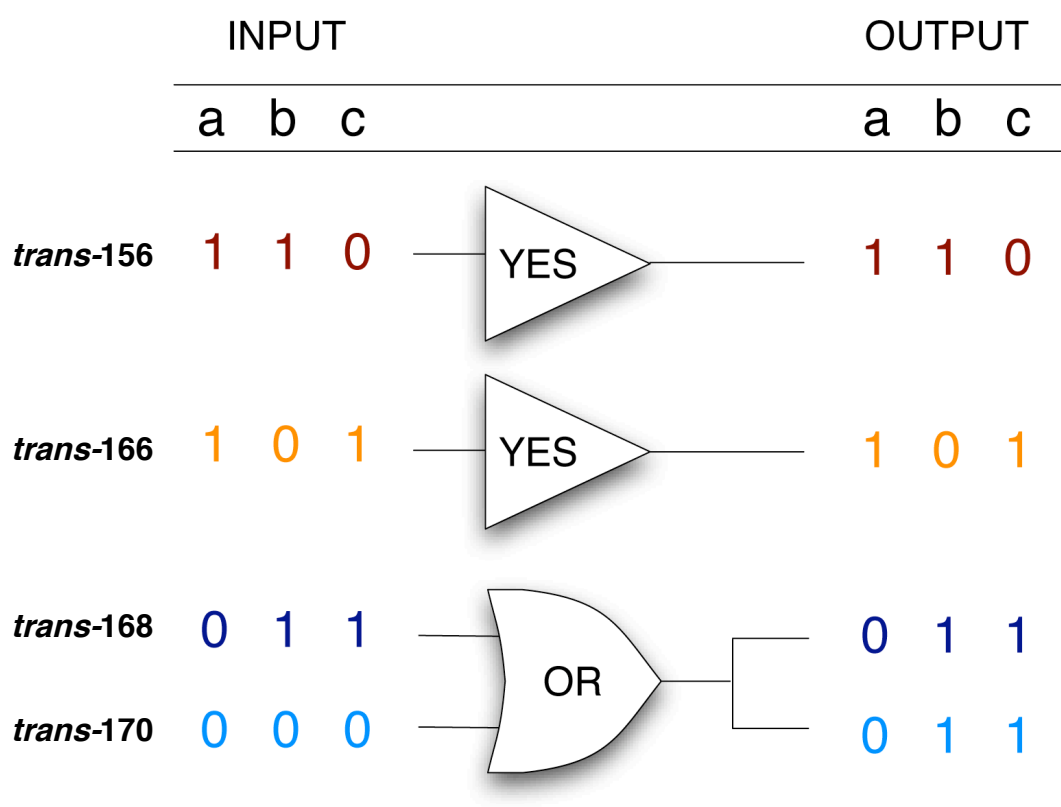
Figure 125. Truth table describing the system as a circuit made of two YES logic gates and a bifurcated OR gate. The columns represent the input and output in particular experiments described in: (a) Figure 122, (b) Figure 123, (c) Figure 124.

3.7 Conclusions

The network, although synthetically simple, displays a surprisingly complex behaviour. A straightforward idea of elongating the building blocks of an efficient replicator led to the formation of this elegant system. It has to be understood that the reaction and the recognition is the same for all the building blocks, the only difference is the length of the molecules. This small alteration opened up a plethora of new properties and possible reactive combinations. The investigation of all possible reactions led to identifying the node properties and the possible connections between

126

them. The network is fully addressable as all nodes participate in some kind of template-mediated process either auto or cross-catalytic. This property allows for the behaviour of the network to be explained in terms of logic operations. Many 'chemical logic gates' use the change of conditions or properties as inputs and output. Also, one type of input (*e.g.* ion recognition) generates a different type of output (*e.g.* fluorescence). The replicator system is different. First of all the input and the output is an actual chemical entity, a compound of given properties. Since the input is actually one of the possible products of the network the composition of the mixture does not change as no new compounds are being introduced and the relative abundance of products determines the output. Such property is only possible when using self-replication as the mean of amplifying specific species from the mixture, since replication is self-sustaining making this system unique in its operation. The input and the output are of the same nature, hence it would be possible to envisage using the output of the network as in input for a next similar system forming a simple circuit.

The systems described in this Chapter possess a high degree of complexity mostly due to the replicating nature of the templates used in the base of their function. All the network components are supplied into the system pool in their defined form, carrying specific properties that then reflect the properties of the network. The starting point for the system is static, all components are in place for the kinetic evolution, which resolves the mixture displaying the properties engendered in the substrates. A more challenging, as well as intriguing, scenario would be provided if the system would exhibit thermodynamically driven combinatorial properties. Coupling a kinetic irreversible reaction, such as a replicator, to a dynamic reagent pool would open possibilities to create more flexible and robust systems.

4. Coupling recognition-mediated reactions to dynamic covalent systems

4.1 Nitrone-imine exchange reaction

The chemistry of reversible covalent bond formation – dynamic covalent chemistry – allows for the generation of networks of interconverting compounds, termed dynamic combinatorial libraries (DCLs). Since the system operates under thermodynamic control, the distribution of library members is governed by their relative free energies. Processes that are capable of altering the free energy relationships within the dynamic network can influence the distribution of library members – the final equilibrium distribution being derived from the sum of the stabilities of all species in a library. DCLs are, in fact, chemical systems with complex network of interactions between all their members.

Selection procedures within DCLs based on thermodynamic control can be limited, since selectivity is based on the differences in binding constants. This property makes thermodynamic selection inherently linear. In order to achieve better selectivity, higher concentration of the template has to be used. On the other hand working with high concentrations of template could lead to undesired, counterintuitive results (see Introduction). The construction of DCLs with large number of components is limited to low concentrations making the thermodynamic selection process less efficient. The answer to achieving high selectivity is to employ irreversible processes into the DCL. The irreversibility of a given process introduces a kinetic aspect into the library. Kinetic amplification using a fast, selective reaction can lead to a very efficient resolution of a DCL upon breaking the purely thermodynamic regime.

Recognition-mediated reactions can be utilised as a tool for selection and amplification of target molecules within DCLs. One of the possibilities of selecting a target compound within a DCL is to employ an irreversible chemical reaction in a constructive manner, using recognition to direct reaction selectively to one library member, and, hence, influence the composition of the library. In order for this approach to be successful, the covalent bond that is made or broken in order to create the DCL must, in itself, be reactive, making it possible to couple the DCL to a recognition-mediated irreversible chemical reaction. Recent self-replicating systems

studied within the group based on the reaction of nitrones. Working extensively with nitrone-based systems led to the discovery that nitrones are capable of undergoing dynamic exchange in non-polar solvents such as chloroform. A key feature of nitrone exchange, when compared to dynamic covalent reactions based on other nitrogen nucleophiles, is the ability of nitrones to participate in irreversible chemical reactions, such as 1,3-dipolar cycloadditions.

The exchange reaction between nitrones is an interesting addition to the family of nitrogen nucleophile-based reversible processes, however, it is not ideal to create a DCL for the purpose of selectively selecting a compound using an irreversible reaction. In a DCL composed solely of nitrones all members of the library would display similar reactivity. Another approach is to introduce a different type of compound that would exchange with nitrones, but displayed orthogonal reactivity. Nitrones possess structural similarity to another group of compounds widely used for the creation of DCLs, namely, imines. If nitrones were to exchange with imines (Figure 126) the number of possible reactive species in the library would be reduced by half, already inducing some selectivity in the system.
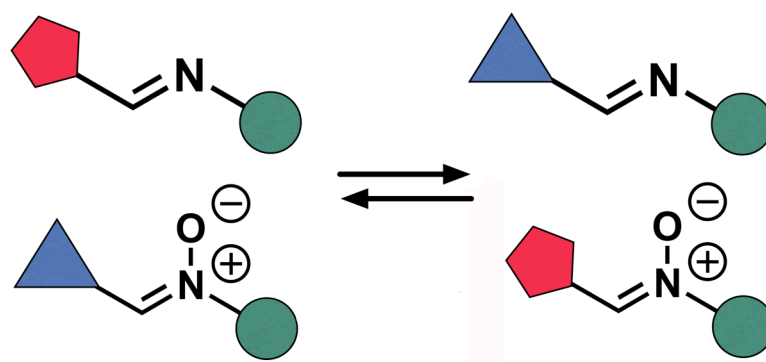
Figure 126. Graphic scheme of imine-nitrone exchange.

4.2 Synthesis of the dynamic combinatorial library

A small dynamic library was developed, based on the exchange reaction between imines and nitrones, using compounds that later would be utilised to couple the DCL to an irreversible process. The compounds necessary to carry out the exchange experiments were easily synthesised from the appropriate starting materials. The imines were formed by a simple condensation of an aldehyde and an aniline (Figure 127).
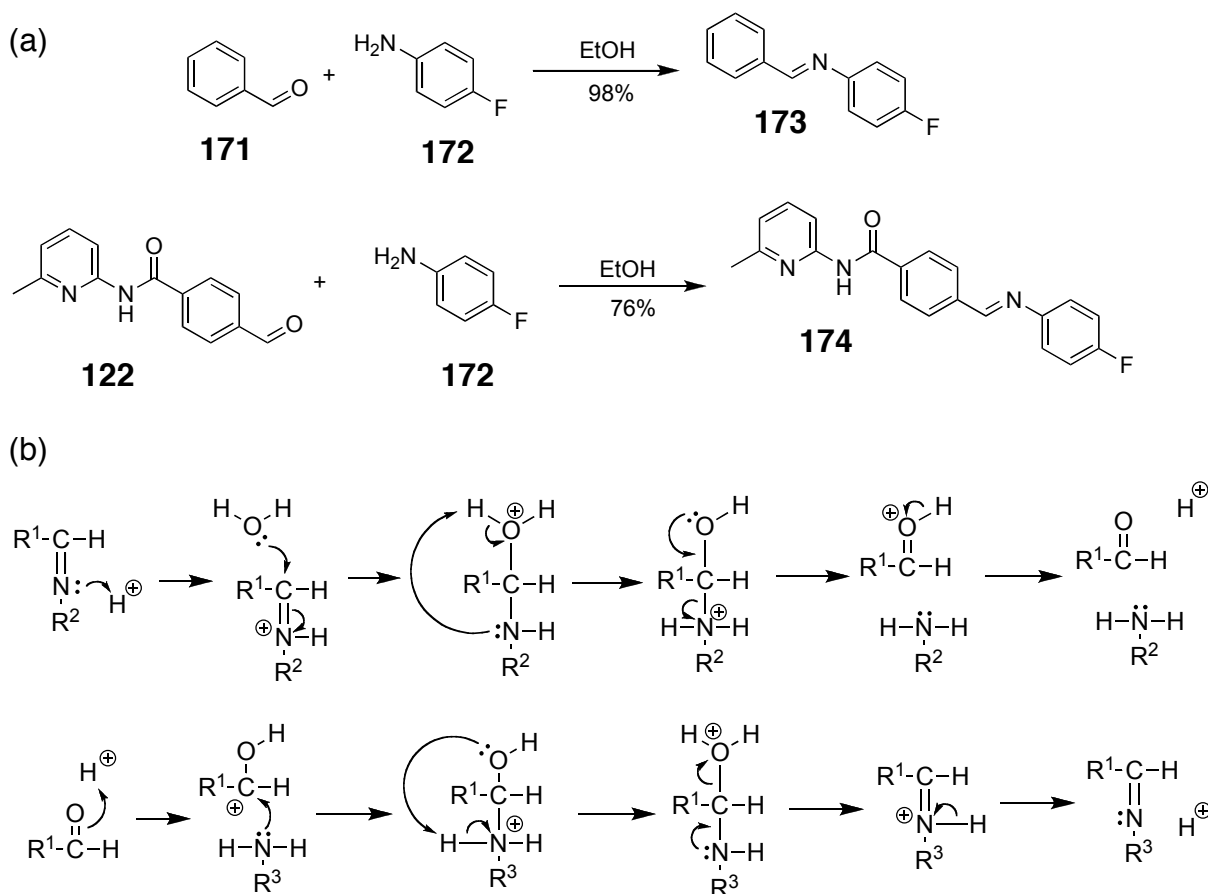
Figure 127. (a) Synthesis of imines **173** and **174**. (b) Mechanism of imine hydrolysis and subsequent imine formation responsible for imine exchange in conditions where both acid and water are present.

Imine **173** is the condensation product of benzaldehyde **171** and 4-fluoroaniline **172** reacted in ethanol giving a high yield of product separated by filtration from the reaction mixture. The imine bearing a recognition site **174** was synthesised in an analogous reaction using aldehyde **122** as the substrate.

The synthesis of the nitrone compounds (Figure 128) was conducted using the same aldehydes **171** and **122** with 4-fluorophenylhydroxylamine **175** instead of the aniline **172**. The hydroxylamine **175** was derived from 4-fluoronitrobenzene reduced with bismuth chloride and potassium borohydride and used immediately in fear of decomposition.
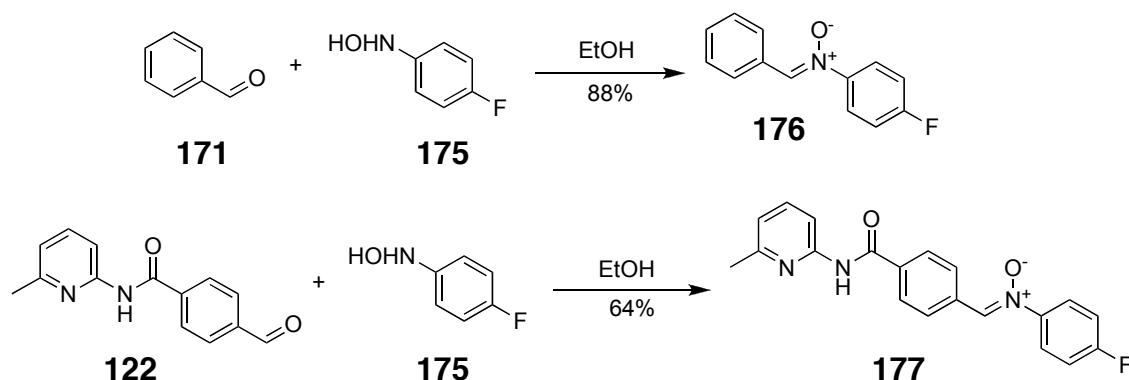
Figure 128. Synthesis of nitrones **176** and **177**.

The next task was to choose the appropriate conditions for the exchange to perform on a reasonable time scale and give repeatable results. The exchange reactions are sensitive to temperature, concentration of starting materials and the amount of acid and water in solution. The conditions of the exchange had to be similar to the ones used for the studies of the recognition-mediated reactions, since the goal was to have both types of processes perform in one system. The chosen starting material concentration was 20 mM and the temperature 273 K. The choice of solvent proved to be a greater challenge. As the previous work was performed in chloroform, this solvent was the first choice. The problem with chloroform is, however, that every batch of this hygroscopic solvent includes a different amount of water and traces of acid. It would be possible to carefully dry and deacidify the chloroform, but both water and acid promote the exchange, and removing them completely from the solution made the exchange too slow to observe any significant changes during the planned time of the experiment (16 h). A solvent system was required that would include both acid and water, but controlled and equal amounts every time the experiment is run. In order to achieve this goal deuterated DCM was chosen as an alternative solvent, being less hygroscopic then chloroform and without acid contamination. In order to introduce defined amounts of both water and acid into the solvent, the sample to be used in the experiment was saturated with *p*-tuloenesulphonic acid monohydrate (PTSA). The amount of acid and water dissolved in DCM allowed for the exchange to occur at the planned time scale, and importantly, proved not to interfere with the recognition processes and other reactions we were hoping to incorporate into our system. In order to prepare the solvent, CD_2Cl_2 (2 ml) was stirred rapidly with 100 mg of PTSA monohydrate in a volumetric flask at room temperature for 20 min. The suspension was then filtered and the solvent used immediately.

131

The construction of the dynamic library components consist of two aldehydes, one of which **171** bears an amidopyridine recognition site. The presence of 4-fluoroaniline permits the formation of the unreactive imines **173** and **174** in the library and the presence of 4-fluorophenylhydroxylamine, in turn, permits the formation of the reactive nitrones **176** and **177**. Therefore, at equilibrium, the DCL contains two imines and two nitrones, compounds **173, 174, 176** and **177** (Figure 129) and the components necessary to synthesise these compounds. NMR spectroscopy was chosen as the analytical technique to monitor the dynamic systems. With the growing complexity of the mixtures ^{1}H NMR spectroscopy proved to be insufficient, as many of the network components have similar structural motifs. Analysis of such mixtures results in overlapping of the characteristic ^{1}H NMR signals in the spectrum making it impossible to distinguish the library components from each other. In order to overcome this problem, all major compounds were equipped with a fluorine tag making it possible to monitor the system using ^{19}F NMR spectroscopy. This approach allows for clear monitoring of the evolution of the networks and all the individual components.
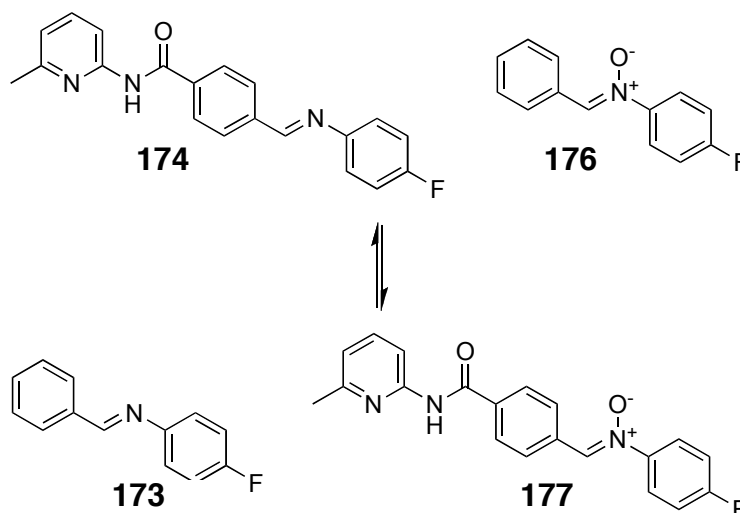
Figure 129. The exchange reaction between the synthesised imines **173** and **174** and nitrones **176** and **177**.

The starting point of the exchange reaction consists of one imine and one nitrone. The first experiment was to prove if starting from any pair of building blocks either **174** and **176** or **173** and **177** would lead to the same equilibrium point. When reacting the recognition imine **174** with a simple nitrone **176** and analyzing the mixture after 24 h five compounds were identified in the mixture: two nitrones **176** and **177**, two imines **173** and **174** and 4-fluoroaniline **172**.
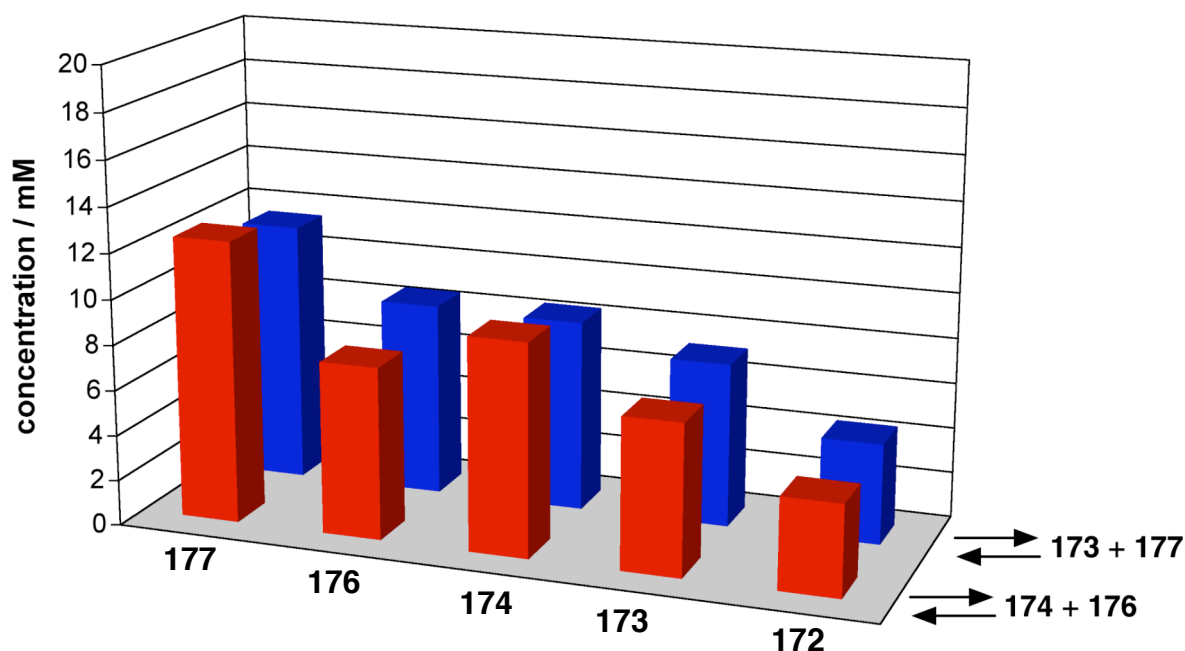
Figure 130. Concentration of the five compounds **172**, **173**, **174**, **176** and **177** detected in the DCL at equilibrium. The starting point for the exchange was either [**174**] = [**176**] = 20 mM; red bars ( ) or [**173**] = [**177**] = 20 mM; blue bars ( ).

The differences in concentrations of the library members are caused by their relative hydrolytic stability. The nitrones are more stable than the imines with the recognition nitrone **177** being the most abundant compound in the mixture. The presence of aniline **172** can be explained by the hydrolysis of the imines **173** and **174**, the hydroxylamine **175**, on the other hand, is not detected, being practically fully incorporated into the nitrone species. Whether the starting point for the exchange is the recognition nitrone **177** and the simple imine **176** or the recognition imine **174** and the simple nitrone **176** the concentrations of library members are practically identical, thus allowing us to establish the equilibrium point for the studied exchange.

There is a number of possible exchange processes in the library. Both imines and nitrones can revert to their starting materials, but it is expected for the imines to hydrolyse much more readily. The hydrolysed constituents can recombine to form new species, but there is a possibility of a direct reaction between the imine and the nitrone.

The exchange process in DCM saturated with PTSA starting with the recognition imine **174** and the simple nitrone **176** was monitored using ^{19}F NMR spectroscopy over a period of 16 hours at 273 K in order to study the evolution of the system with time (Figure 131).
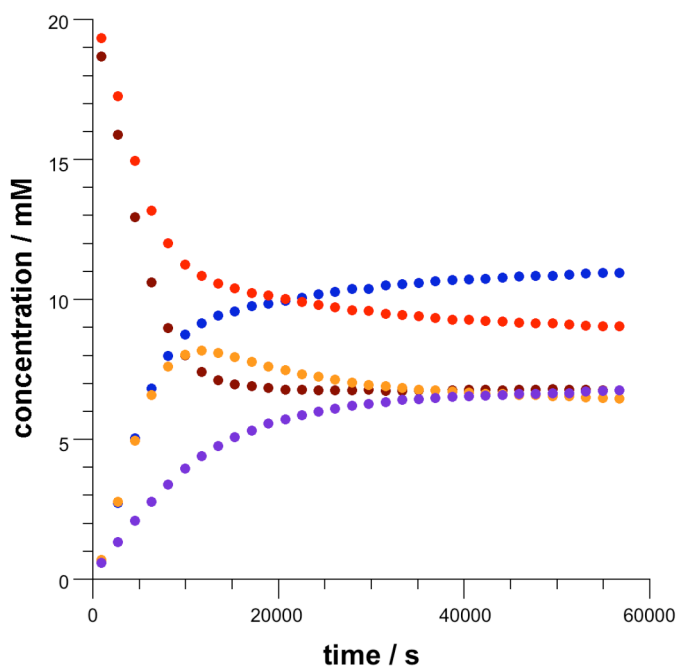
Figure 131. Concentration *vs.* time plot of the exchange reaction between imine **174** and nitrone **176** ([**174**] = [**176**] = 20 mM) following the depletion of substrates: imine **174** (●) and nitrone **176** (●) and the formation of products nitrone **177** (●), imine **173** (●), and aniline **172** (●).

As expected the concentration of the starting materials **174** and **176** goes down and subsequently the formation of the new imine/nitrone **173**/**177** pair is observed. It is worth noting that the final equilibrium point of the system actually consists of two separate equilibria. The nitrones, being more hydrolytically stable, are present at higher concentrations forming the first equilibrium. The imines together with the aniline form a lower concentration equilibrium that includes the reversible decomposition of the imine species, hence the observed amount of the aniline **172**. The formation of the simple imine **173** displays an interesting profile with the concentration rapidly reaching 8 mM in the first 1000 s and than falling due to hydrolysis to below 7 mM at equilibrium. The equilibrium concentrations of the main library components are: [**177**] = 11.0 mM, [**176**] = 9.1 mM, [**174**] = 6.8 mM, [**173**] = 6.5 mM.

4.3 Coupling a DCL to an AB complex pathway reaction

The library has been specifically designed in order to incorporate recognition-mediated irreversible reactions as means of influencing its composition. There are two distinct features of the compounds in the library – the reactive site and the recognition site. The four compounds feature the four possible combinations of the

134

two features. Imine **174** bears only the recognition, while nitrone **176** is the carrier of the reactivity. Between the products of the exchange reaction imine **173** does not posses any of the characteristic properties while nitrone **177** bears both the recognition and the nitrone reactive site. This library member is capable of undergoing an irreversible 1,3-dipolar cycloaddition reaction with a maleimide that is directed by a recognition event and therefore is the target for amplification. Recognition and reactivity are demonstrated to be essential for achieving selectivity within the library and this combination is shown to have a profound effect on the final composition of the library.

Material can be transferred irreversibly from this exchange pool to a pool of products through the reaction of either nitrone **176** or nitrone **177** with a maleimide added together with the exchange reaction substrates. It has already been demonstrated that it is possible to achieve significant rate accelerations in cycloaddition reactions through the formation of a reactive AB complex between the 1,3-dipole and the 2π components. Hence, it was necessary to identify the most appropriate structure for a dipolarophile that could recognise **177** and, hence, participate in a 1,3-dipolar cycloaddition reaction through a recognition-mediated pathway such as the AB complex channel. Electronic structure calculations were used to design maleimide **179**.
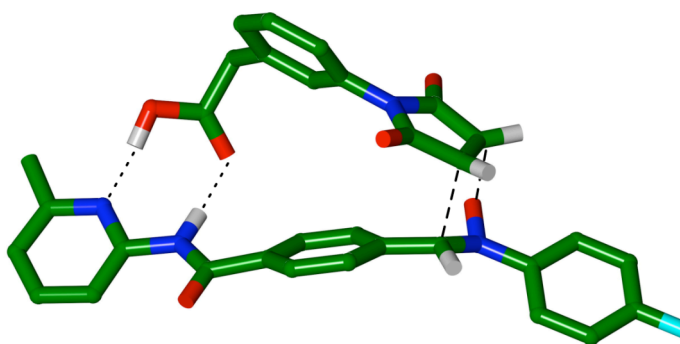
Figure 132. Calculated transition state (B3LYP/6-31G+(d,p)) for the reaction between nitrone **177** and maleimide **179** within the [**177·179**] binary complex. This transition state leads to *cis*-**182**. C atoms are coloured green, O atoms are coloured red, N atoms are coloured blue, H atoms are coloured white and the F atom is coloured cyan. Most H atoms have been omitted for clarity. Hydrogen bonds are represented by dotted lines and partial covalent bonds by dashed lines.

The calculations indicated that maleimide **179** was capable of binding to nitrone **177** and this complex was capable of accessing the transition state leading to *cis*-**182**.

Since the AB complex channel is the chosen path of amplification, thus it is expected the *cis* isomer being the dominating product of the reaction. Maleimides **179** and **180** were synthesised from commercially available compounds in an analogous manner as maleimide **112** (Figure 133).
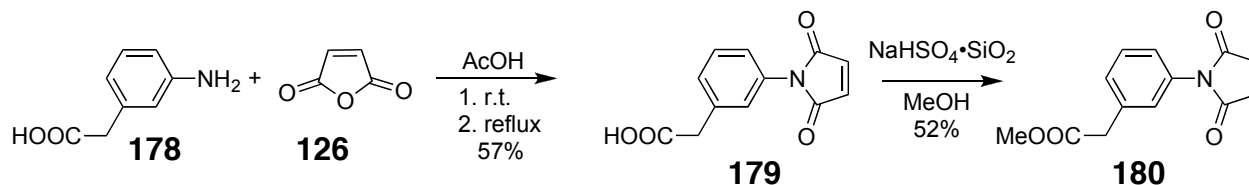
Figure 133. Synthesis of maleimides **179** and **180**.

In order to verify experimentally the predicted behaviour of the binary complex, the reactions of recognition-enabled nitrone **177** with both the control maleimide **180** and the recognition-enabled maleimide **179** were assessed (Figure 134). These reactions were performed in CD_2Cl_2 saturated with PTSA at 273 K with the initial concentrations of the substrates equal to 20 mM and were monitored by 500 MHz ^{1}H NMR spectroscopy. After 60000 s, the total yield of the control reaction ([**177**] = [**180**] = 20 mM) was 15% and the ***trans*-181**:***cis*-181** ratio was 3:1. When the same experiment was conducted with the recognition motif enabled ([**177**] = [**179**] = 20 mM) the conversion after 60000 s rose to 95% and the diastereoisomeric ratio was 35:1 in favour of ***cis*-182** (Figure 135).
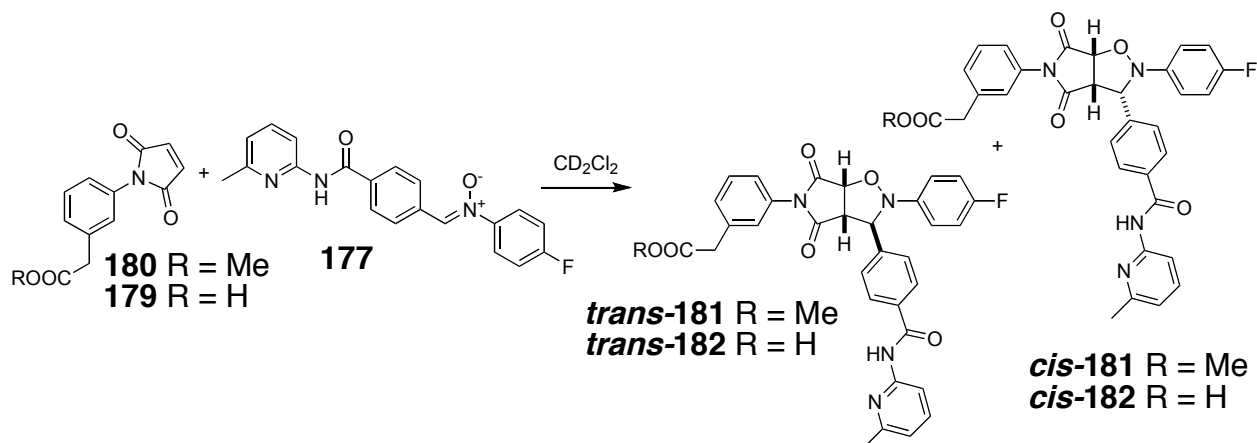
Figure 134. Products of 1,3-dipolar cycloaddition between maleimide **179** and **180** with nitrone **177**.
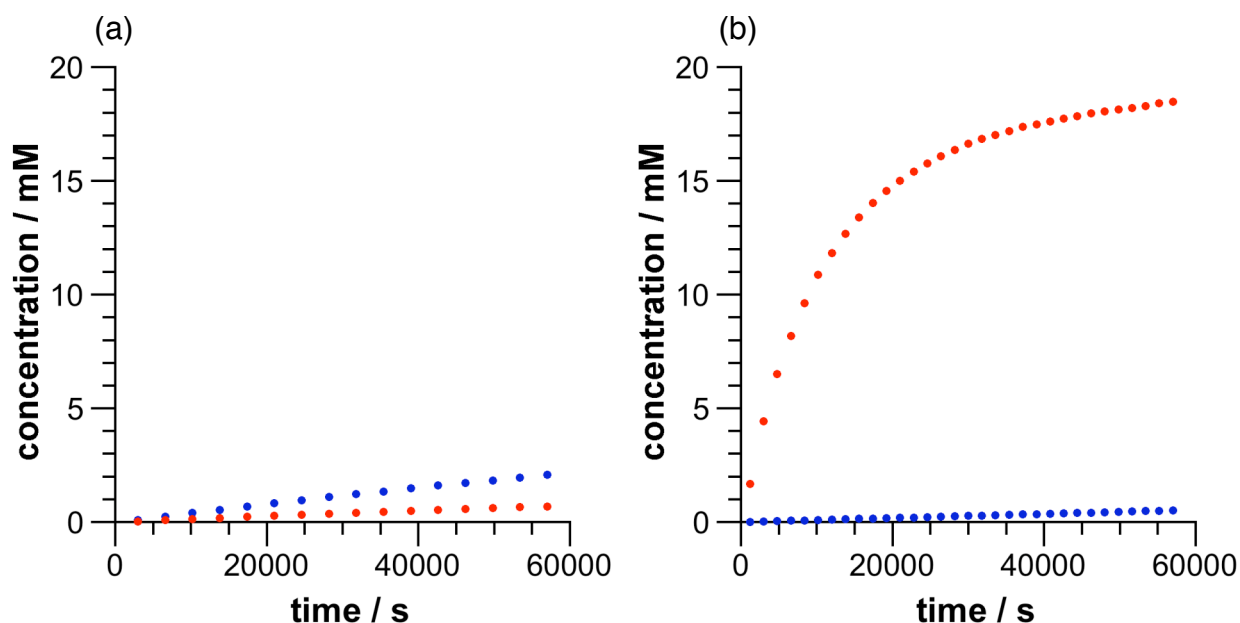
Figure 135. Concentration *vs.* time plots of (a) control reaction between ester maleimide **180** and nitrone **177** ([**180**] = [**177**] = 20 mM) following the formation of products: ***trans*-181** (●) and ***cis*-181** (●). (b) Recognition-enabled reaction between maleimide **179** and nitrone **177** ([**179**] = [**177**] = 20 mM) following the formation of products ***trans*-182** (●) and ***cis*-182** (●).

Having demonstrated that the mixture of imines and nitrones is indeed dynamic, and the reaction chosen for the amplification is fast and selective it was now possible to attempt the coupling of this exchange process to the cycloaddition reaction (Figure 136).
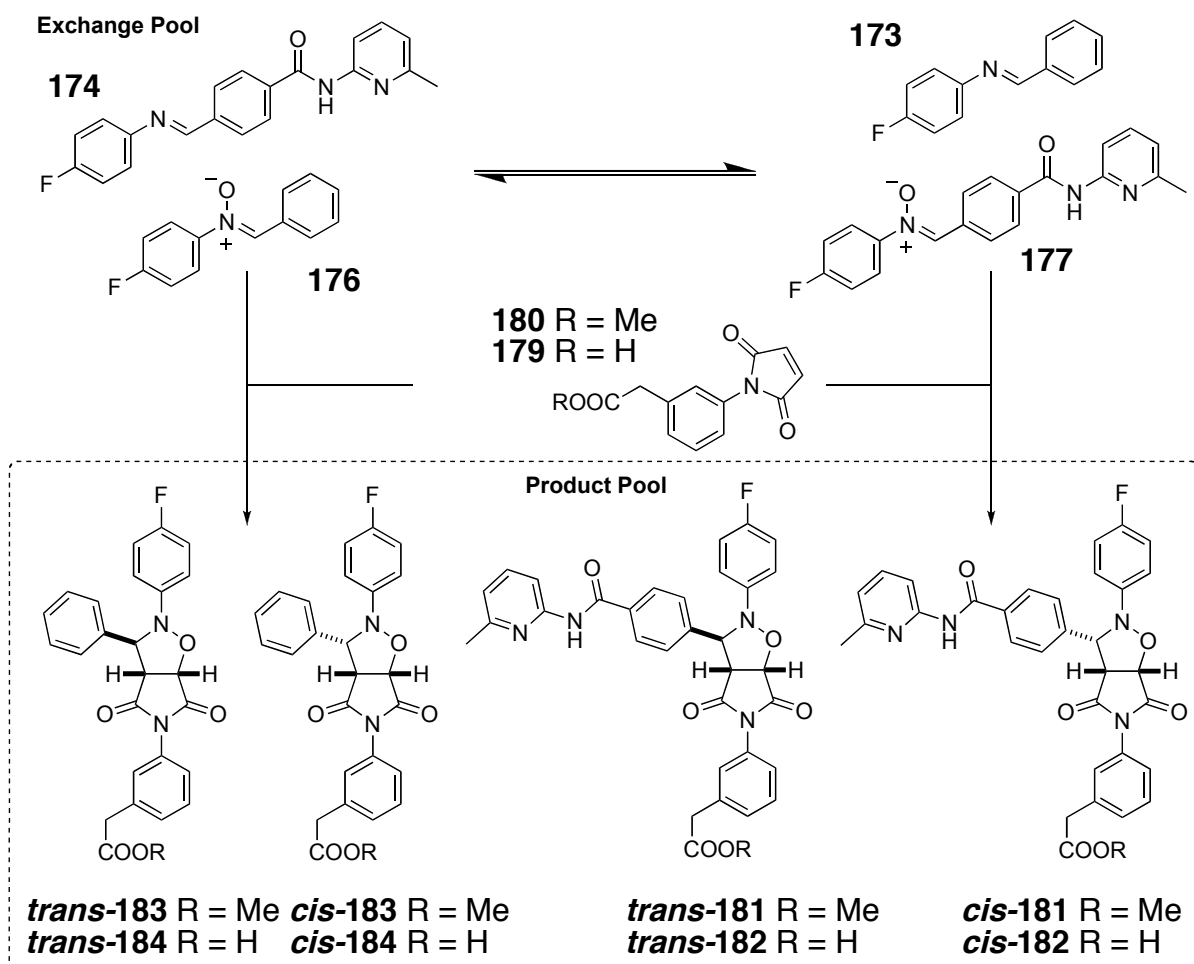
Figure 136. A pool of compounds containing imines **173** and **174** and nitrones **176** and **177** can exchange freely in CD$_2$Cl$_2$ saturated with *p*-toluenesulfonic acid monohydrate at 273 K. Material can be transferred irreversibly to a pool of products, present in the same solution, that cannot be interconverted or returned to the exchange pool, through reaction of nitrones **176** or **177** with an appropriate maleimide (**179** or **180**).

The control experiment with the recognition blocked was to determine if simply permitting an irreversible reaction within the exchange pool generated any selectivity in the system. To this end, a mixture of imine **174** and nitrone **176** in CD$_2$Cl$_2$ saturated with PTSA ([**174**] = [**176**] = 20 mM) was prepared and maleimide **180** was added immediately as the dipolarophile ([**180**] = 20 mM). Since the rate of exchange is sensitive to the amount of acid and in the recognition-mediated reaction an acid maleimide **179** will be used, an equivalent of phenylacetic acid was also added to the reaction mixture. The composition of this mixture was then allowed to evolve at 273 K for 16 hours and monitored by ^{19}F NMR spectroscopy every 30 minutes during this period. The concentrations of each of the species containing a fluorine tag present were then determined for each time point. The compounds were identified by their characteristic ^{19}F chemical shift in the spectrum (Figure 137).

138
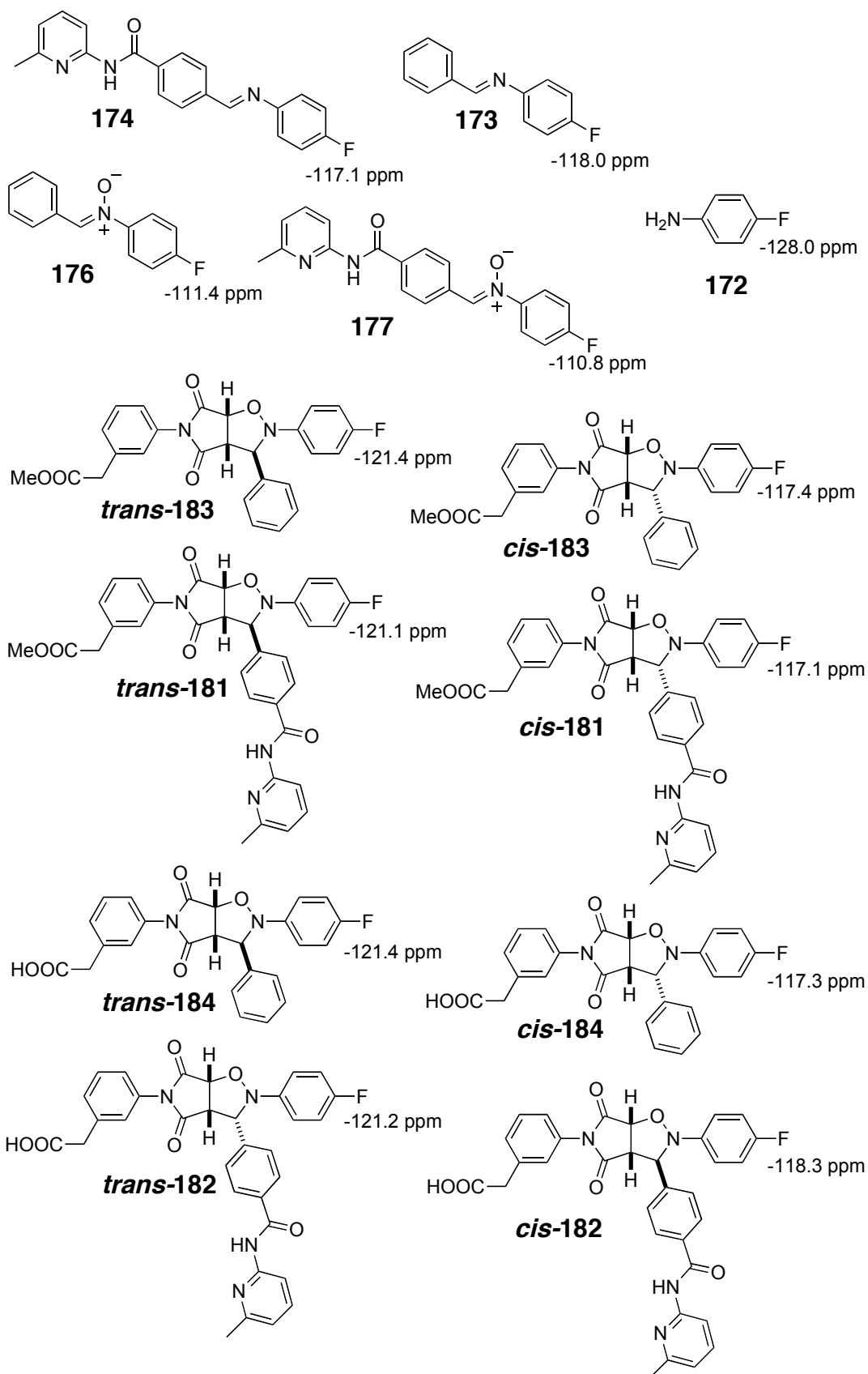
Figure 137. 470.4 MHz ^{19}F NMR chemical shifts, recorded in CD_2Cl_2 saturated with PTSA at 273 K, for all of the fluorine-containing species observed in the exchange experiments.

In the product pool of the control reaction, after 16 h, *trans*-183 and *cis*-183 – the products of reaction between nitrone 176 and maleimide 180 – are present at a total concentration of 3 mM ([*trans*-183]:[*cis*-183] = 2:1). After the same time, *trans*-181 and *cis*-181 – the products of reaction between nitrone 177 and maleimide 180 – are present at a total concentration of 1.4 mM ([*trans*-181]:[*cis*-181] = 2.5:1). Thus, the overall conversion for all cycloaddition reactions within the library is only 22% (Figure 138). Within the exchange pool, after 16 h, compounds 173, 174, 176 and 177 are all present in almost equal amounts ([177] = 7.6 mM, [176]= 7.1 mM, [174] = 8.1 mM, [173] = 8.5 mM). This composition is close to that observed in the exchange experiments where the maleimide is absent.

Figure 138. Concentration *vs.* time plots of (a) the evolution of the exchange pool generated from imine 174 and nitrone 176 ([174] = [176] = 20 mM) following the depletion of substrates: imine 174 (●) and nitrone 176 (●) and the formation of products nitrone 177 (●), imine 173 (●), and aniline 172 (●). (b) Evolution of the product pool generated from the exchange pool by the reaction of nitrones 176 and 177 with maleimide 180 following formation of products *trans*-183 (●), *cis*-183 (○), *trans*-181 (●) and *cis*-181 (○).

There are significant differences in the reaction profiles between the exchange without any maleimide added (Figure 131) and the described control experiment (Figure 138). The exchange is slower in the latter case with both imines and nitrones following the same profile and the different equilibria for the two classes of compounds are not observed with all compounds in the exchange pool reaching practically the same concentration. The imines still hydrolyse more readily and the

presence of 4-fluoroaniline **172** proves this assumption. Both nitrones **176** and **177** are to some extent converted into their corresponding cycloadducts when reacting with the maleimide **180** explaining their lower concentration. Interestingly the concentration of all cycloadducts compared to 4-fluoroaniline is practically the same throughout the experiment suggesting the rate of the imine hydrolysis and nitrone - maleimide reaction is equal explaining why the nitrone and the imine concentration follow the same reaction profiles. The overall decrease in rate of exchange is more difficult to explain and this issue will be addressed later. Still the rate of the 1,3-dipolar cycloaddition reactions between **176** and **180** and **177** and **180** are much slower than the rate of exchange. Therefore, these irreversible reactions have little influence on the exchange process as nitrones **176** and **177** are depleted at similar rates. In other words both sides of the exchange reaction are influenced in the same manner so the equilibrium is not driven to any particular side of the exchange. Additionally, both cycloaddition reactions are rather unselective resulting in poor overall selectivity in the product pool. Thus, despite the fact that nitrone **177** is not present in the starting exchange pool, after 16 hours, the ratio of cycloadducts arising from nitrone **176** to those arising from nitrone **177** is only 2:1 and there is rather poor diastereoselectivity in the product pool in general. It is clear from these results that simply coupling exchange to the irreversible cycloaddition reaction generates no selectivity in either the exchange pool or in the product pool.

Next, we attempted to determine the effect that a recognition process can have on chemical reactivity within the context of the exchanging library. The anticipation was that, by use of maleimide **179**, which possesses a carboxylic acid recognition site complementary with the amidopyridine recognition site present in nitrone **177**, it would be possible to exploit the more than a 100-fold acceleration in the rate of reaction between nitrone **177** and maleimide **179** to drive the exchange process. Accordingly, a mixture of imine **174**, nitrone **176** and maleimide **179** in CD_2Cl_2 saturated with PTSA ([**174**] = [**176**] = [**179**] = 20 mM) was prepared. Once again, the composition of the mixture was allowed to evolve at 273 K for 16 hours and the coupled exchange and reaction processes were monitored by ^{19}F NMR spectroscopy as described previously (Figure 139).
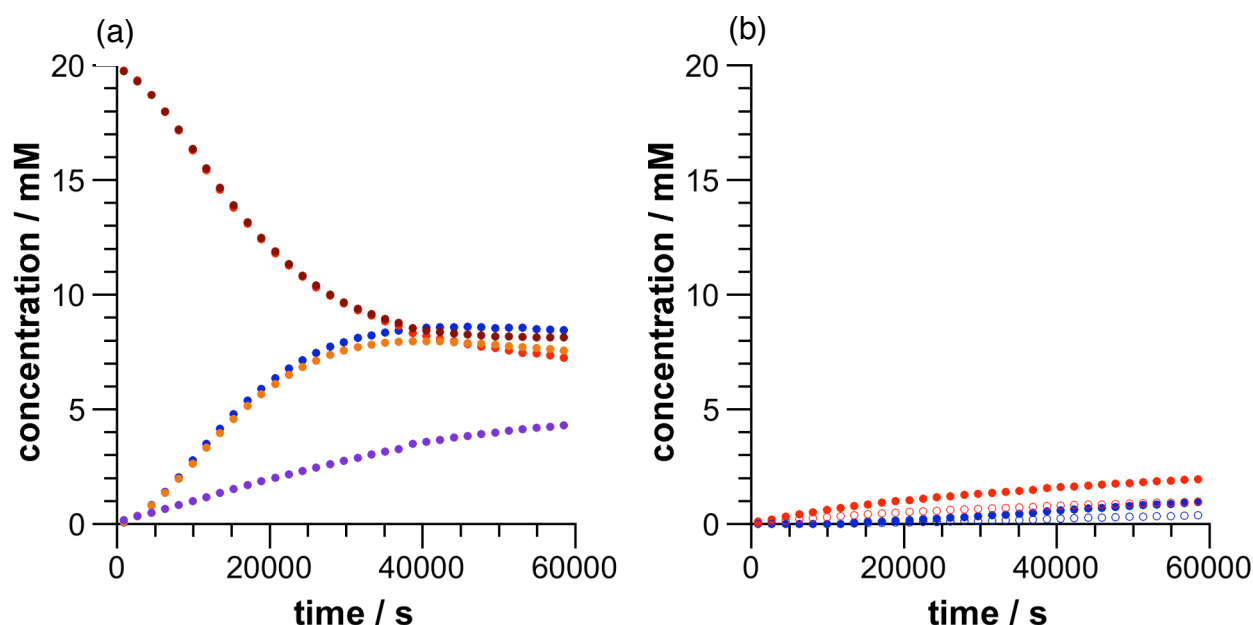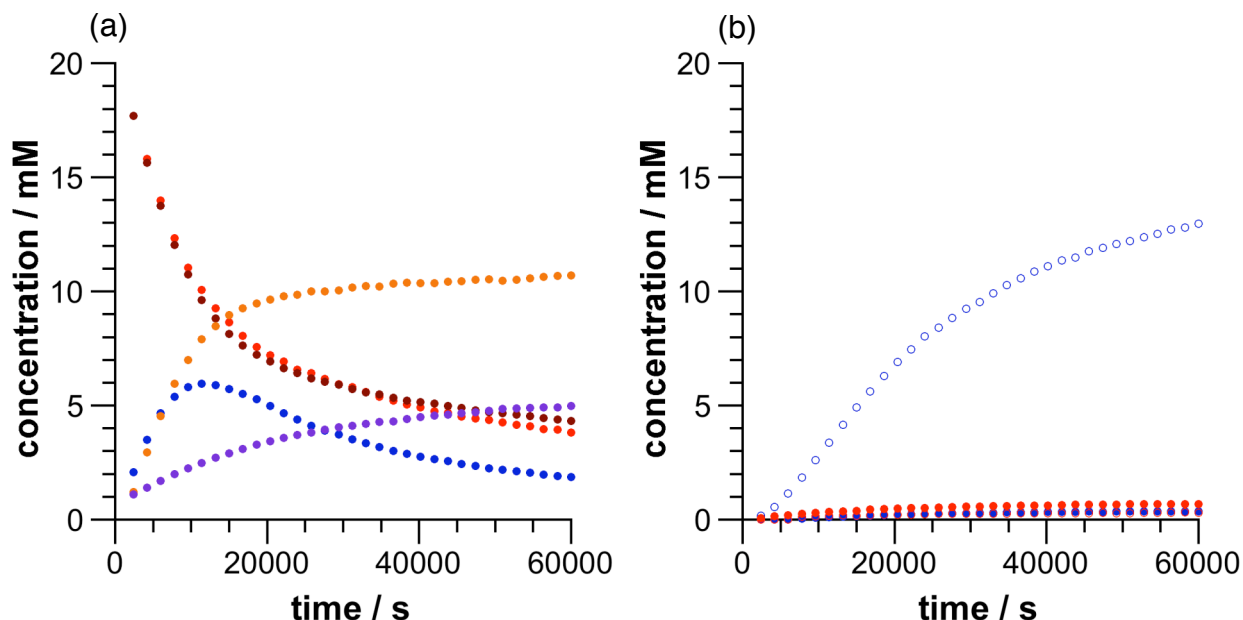
Figure 139. Concentration *vs.* time plots of (a) the evolution of the exchange pool generated from imine **174** and nitrone **176** ([**174**] = [**176**] = 20 mM) following the depletion of substrates: imine **174** (●) and nitrone **176** (●) and the formation of products nitrone **177** (●), imine **173** (●), and aniline **172** (●). (b) Evolution of the product pool generated from the exchange pool by the reaction of nitrones **176** and **177** with maleimide **179** following formation of products *trans*-**184** (●), *cis*-**184** (○), *trans*-**182** (●) and *cis*-**182** (○).

When comparing the ^{19}F NMR spectra of the control and the recognition-enabled reaction after 16 h it is apparent that in the latter case one compound emerges above all other in the mixture (Figure 140).

142
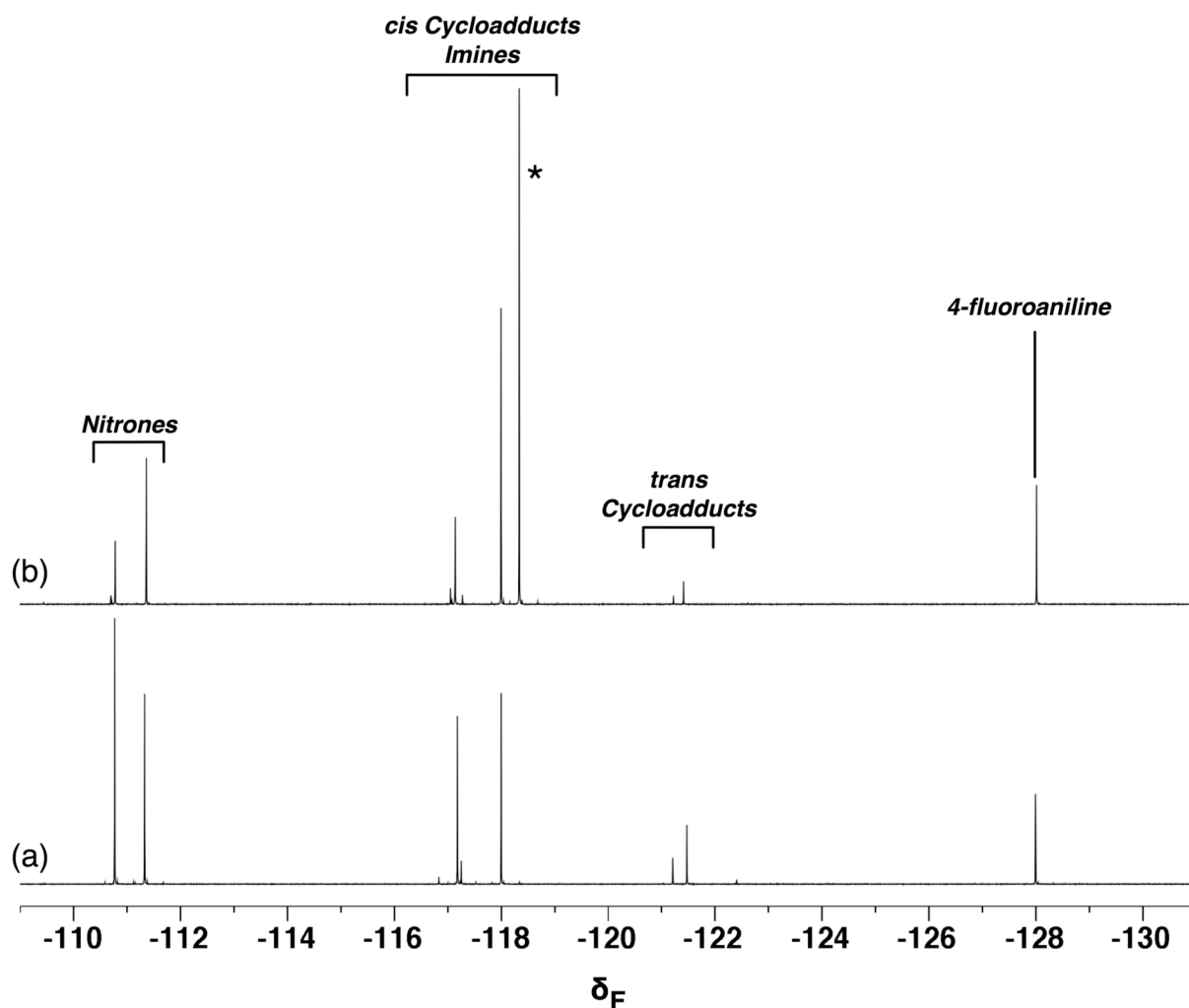
Figure 140. Partial 470.4 MHz ¹⁹F NMR spectra, recorded in CD₂Cl₂ saturated with PTSA at 273 K, of (a) the control exchange experiment, starting from **174**, **176** and **180**, after 16 hours and (b) the recognition-mediated exchange experiment, starting from **174**, **176** and **179**, after 16 hours. The starred resonance is that arising from *cis*-**182**.

In the product pool, after 16 h, *trans*-**184** and *cis*-**184** – the products of reaction between nitrone **176** and maleimide **179** – are present at a total concentration of 1.1 mM ([*trans*-**184**]:[*cis*-**184**] = 2:1). After the same time, *trans*-**182** and *cis*-**182** – the products of reaction between nitrone **177** and maleimide **179** – are present at a total concentration of 13.3 mM ([*trans*-**182**]:[*cis*-**182**] = 1:43). Thus, the overall conversion for all cycloaddition reactions within the library is now 72% and cycloadduct *cis*-**182** constitutes for more than 90% of the total cycloadduct in the product pool. The effects of the introduction of the recognition process into the system are equally significant in the exchange pool. After 16 h, compounds **173**, **174**, **176** and **177** are present in dramatically different concentrations compared to the same exchange process in the presence of the control maleimide **180**. In particular,

the concentration of the two nitrones **176** and **177** are depressed significantly ([**176**] = 3.8 mM, [**177**] = 1.9 mM). In this case, the rate of the 1,3-dipolar cycloaddition reaction between nitrone **176** and maleimide **179** is still slower than the rate of exchange since it is not recognition mediated. However, the reaction between nitrone **177** and **179** is comparable in rate to the exchange processes. Thus, once exchange generates a concentration of recognition nitrone **177** close to the K_d (~ 2 mM) of the reactive complex, rapid reaction to form *cis*-**182** will occur through this complex, thereby removing nitrone **177** from the exchange pool. This rapid depletion of one component of the exchange pool drives the equilibration of compounds **173**, **174**, **176** and **177** to regenerate the depleted species. It is clear from the results that coupling exchange to the irreversible recognition-mediated cycloaddition reaction generates significant selectivity in both the exchange pool and in the product pool.

The introduction of a carboxylic acid recognition site on maleimide **179** accelerates the reaction between **179** and one member of the exchange pool, nitrone **177**, by a factor of 125. In order to understand the broader effect of this single, recognition-mediated reaction process within the manifold of exchanging species, it is instructive to make an overall comparison between the compositions of the exchange and product pools after 16 h in the absence and in the presence of recognition. For the purposes of comparison, we define the enhancement factor, EF, for a given species as being the log of ratio of the concentration of that species after 16 h in the experiment where recognition is active to that of the control experiment (Figure 141).

Within the exchange pool, only imine **173** is enhanced (EF = 0.15). This small enhancement arises because neither of the compounds required to form imine **173** – 4-fluoroaniline **172** and benzaldehyde **171** – are necessary for the recognition-mediated reaction. By contrast, imine **174** and nitrones **176** and **177** are all suppressed (EF(**174**) = –0.27; EF(**176**) = –0.27; EF(**177**) = –0.66) since all three compounds contain at least one component required for the recognition-mediated reaction. In the reaction pool, the formation of one cycloadduct, *cis*-**182**, is clearly enhanced and the other three cycloadducts are suppressed. The enhancement factor for cycloadduct *cis*-**182** is 1.43 and those of the other three cycloadducts, *trans*-**182**, *trans*-**184** and *cis*-**184**, are –0.42, –0.46 and –0.57 respectively. These changes generate selectivity for *cis*-**182** of 19:1 over the next most abundant cycloadduct, namely *trans*-**184**.

$$\text{enhancement factor} = \log \frac{[\text{product}]_{\text{recognition}}}{[\text{product}]_{\text{control}}}$$
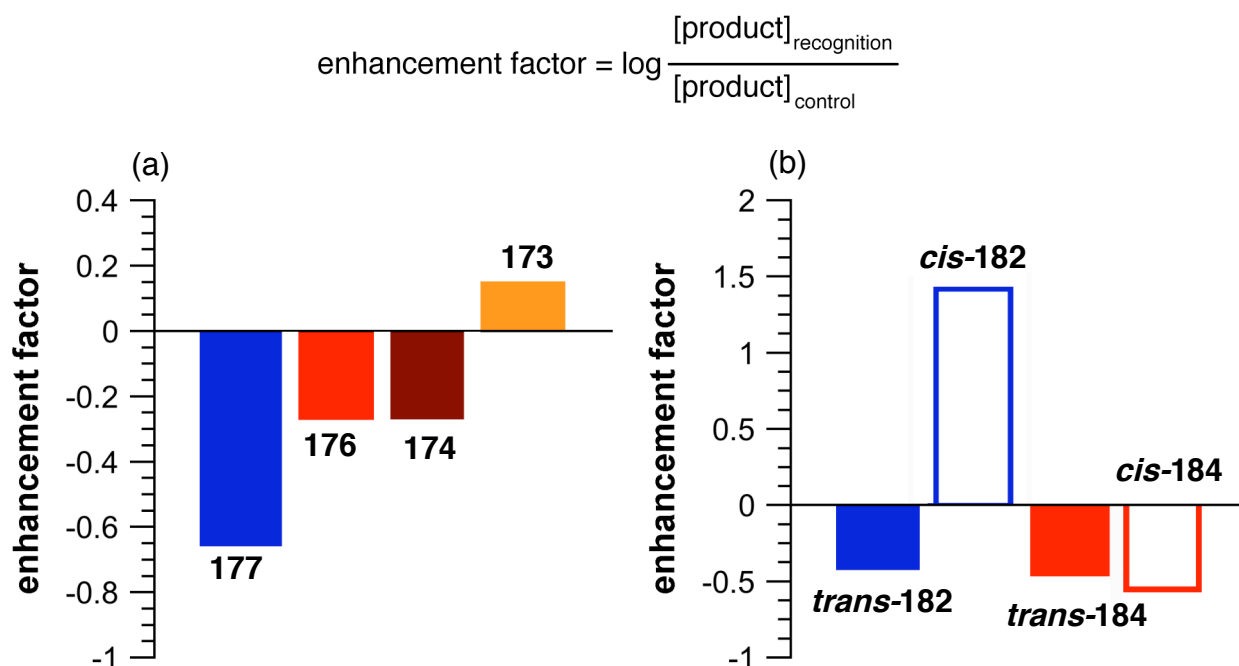
Figure 141. Enhancement factors for compounds in (a) the exchange pool: nitrones **177** and **176** and imines **174** and **173** (b) the product pool: cycloadducts **_trans_-182**, **_cis_-182**, **_trans_-184** and **_cis_-184**.

Three different situations involving the exchange reaction were described: the exchange on its own, the control reaction using maleimide **180** and the recognition-mediated experiment using maleimide **179**. The concentration _vs._ time plots (Figures 131, 138 and 139) for the three cases differ significantly. What is most striking is that the exchange process seems slowest in the case when the exchange is coupled to a slow irreversible reaction without recognition present (Figure 138). In order to explain this observation, it is necessary to answer the question what drives the exchange. It seems that any imbalance that is generated between the library members 'kickstarts' the exchange allowing it to reach the equilibrium faster. Lets first focus on the two examples where the exchange is fast: the exchange on its own (Figure 131) and the recognition-mediated coupled exchange (Figure 139). In the first case the difference between the imine and nitrone components is that the imines hydrolyse much more readily than nitrones. Upon the hydrolysis the concentration of the imine falls, compared to the nitrone, creating the imbalance necessary to drive the exchange. In the second case the exchange is coupled to a fast irreversible process. Since only one of the nitrones is rapidly removed from the exchange pool that in itself is enough to drive the rate of exchange. The scenario is different when we consider the exchange coupled to a slow irreversible process (Figure 138). In this case both

145

nitrones react at the same rate as well as both imines hydrolyse at similar rates. Incidentally, the rate of the cycloaddition and the imine hydrolysis is also of similar magnitude, so both types of building blocks at both sides of the exchange reaction are influenced practically in the same manner. In such scenario there is no imbalance to drive the exchange. It is clear when looking at the concentration *vs.* time plot as both the imine and the nitrone follow the same reactions profiles. The exchange of course occurs but at a lower rate.

The fate of an exchanging pool of compounds can be influenced profoundly by a single, appropriately designed, recognition event. In the system presented here, one of two reactive members of an exchanging pool of compounds can be targeted and transformed rapidly by a recognition-mediated irreversible cycloaddition reaction, altering dramatically the final composition of the system. The amplification is purely kinetic. Comparing to a classic thermodynamic amplification from a DCL, in this case no template is needed. The differentiating factor is the maleimide substrate that is designed to utilise the recognition event for kinetic amplification. The target nitrone displays the only proper combination of recognition and reactivity in order to be metabolised forming the dominating cycloadduct. In such approach it is not exactly a library member that is selected from a mixture, but its product, a more complex structure. The reaction is capable of influencing the exchange pool as it removes a selected component from the thermodynamically controlled exchange regime. The effects demonstrated here rely on exploiting the AB complex pathway as means of amplifying the desired compound. A more challenging and interesting approach is to resolve the library using self-replication.

4.4 Coupling the DCL to a self-replicator

The challenge of coupling a nitrone-based replicator to the previously described dynamic library based on nitrone-imine exchange carries new implications. The selectivity in the dynamic system based on the AB complex pathway described in the previous section was astounding and probably will not be matched by the replicator. The components were well designed and their engendered properties gave the expected result, so why modify the system towards replication methodology? Analysing the properties of the substrates of the system based on the AB complex pathway makes the outcome quite apparent. Self-replication is a much more

interesting and subtle mechanism. First of all the replicator is not linear, its properties become apparent with the evolution of the reaction. The templating properties that make the replicator a selfish catalyst are engendered in the structure of the components, but it is the mechanism that is responsible for the enhancement of the specific product. These properties make the information embedded in the replicator-based system more abstract. The other property that makes the replicator unique is the possibility of addressing the system with preformed template. This feature allows for creating systems that will respond to the addition of a small instructional input. Since the system is dynamic, such input will influence the whole of the network, reconfiguring it in order to accommodate the imported data.

The research into self-replicating systems and dynamic libraries makes it possible to demonstrate that a synthetic replicator, by virtue of its autocatalytic properties, is capable of exploiting a network of reactions within a dynamic library to amplify its own formation at the expense of other species. This methodology is very interesting as a way of resolving a dynamic combinatorial library. Also, the idea of a replicator evolving from a complex mixture is common in the research concerning the origin of life. Designing a structurally and analytically simple system that combines DCC with self-replication is an important addition to modern systems chemistry. In this approach, complex dynamic phenomena are expressed by a group of synthetic chemical entities designed to interact and react with many partners within the ensemble in programmed ways. In this manner, it should be possible to create synthetic chemical systems whose properties are not simply the linear sum of the attributes of the individual components. These new system-level properties emerge through the interactions of chemical networks assembled from the many predesigned components.

The construction of the library is analogous to the one described previously. The only change is the substitution pattern on the substrate maleimide **112** responsible for the transfer of the selected component from the exchange pool to the products pool. This time one of the cycloadducts formed in the product pool – **trans-188** – is capable of catalyzing its own formation, *i.e.* it is capable of self-replication. The reaction is analogous to the replicating cycle described in Section 2.3.
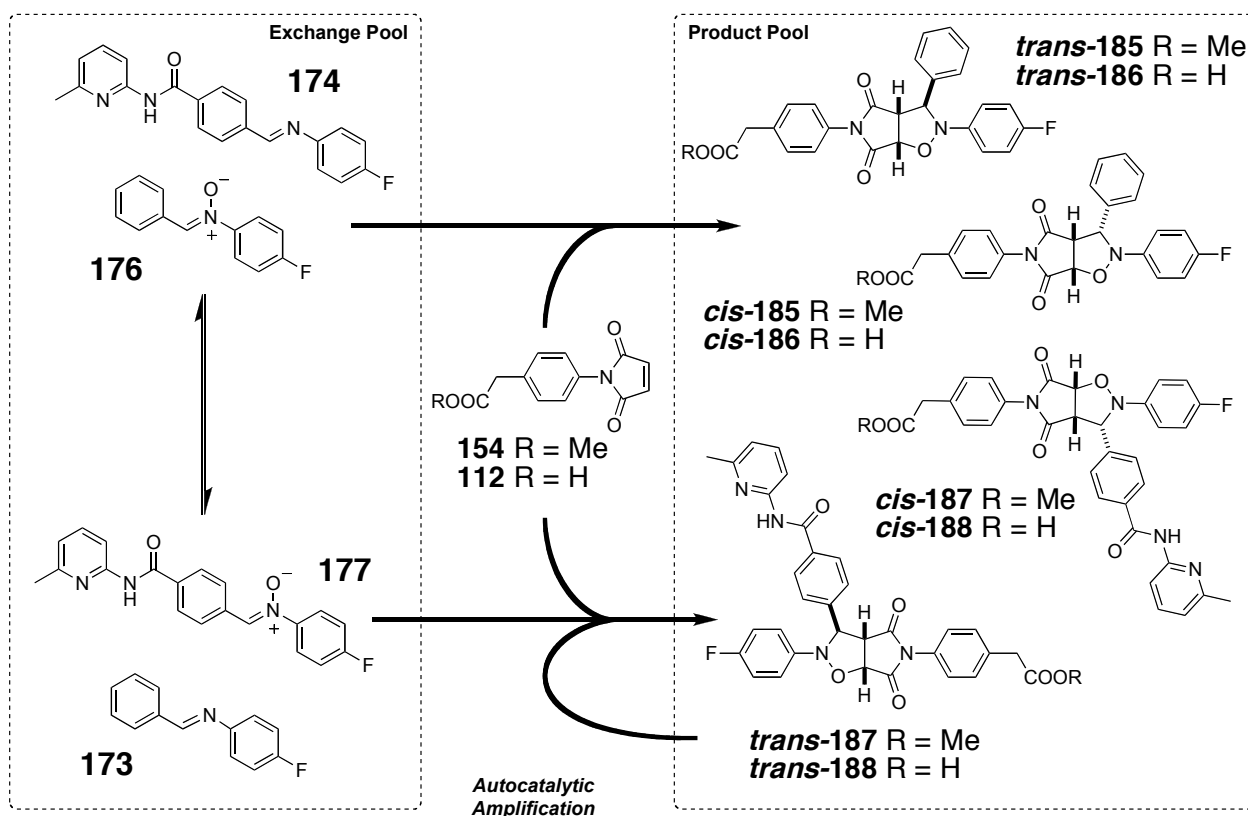
Figure 142. A pool of compounds containing imines **173** and **174** and nitrones **176** and **177** can exchange freely in CD$_2$Cl$_2$ saturated with *p*-toluenesulfonic acid monohydrate at 273 K. Material can be transferred irreversibly to a pool of products, present in the same solution, that cannot be interconverted or returned to the exchange pool, through reaction of nitrones **176** or **177** with an appropriate maleimide (**154** or **112**). When maleimide **112** is used as the dipolarophile, replicator *trans*-**188** is formed in the product pool and this species can act as a catalyst for its own formation.

The replication process relies on the ability of *trans*-**188** to act as a template for its own formation through the recognition and binding of nitrone **177** and maleimide **112** – the components required to form *trans*-**188**. The replicator was investigated in isolation applying the exchange conditions (CD$_2$Cl$_2$ saturated with PTSA) (Figure 143).

148
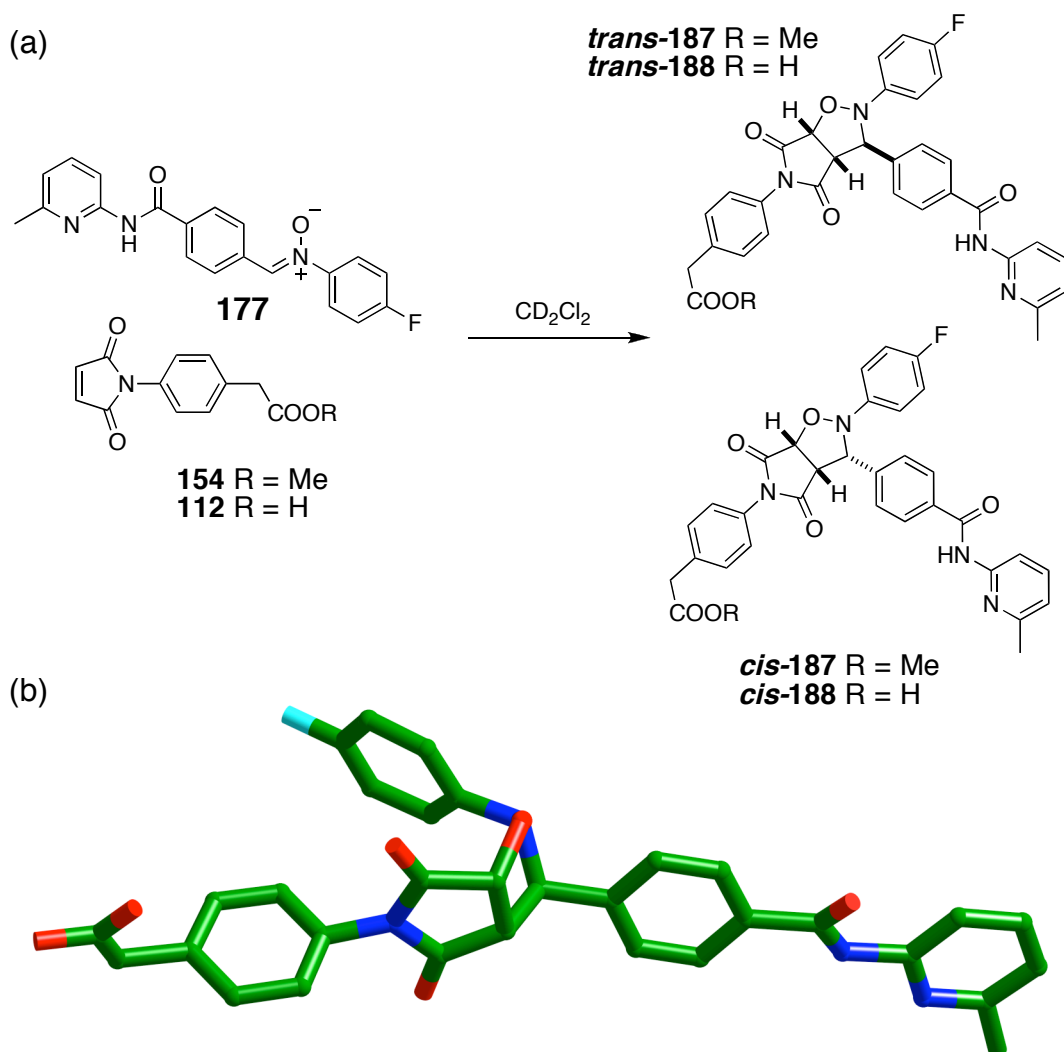
Figure 143. (a) Products of 1,3-dipolar cycloaddition between maleimide **154** and **112** with nitrone **177**. (b) Solid state structure of **trans-188**, determined by single crystal X-ray diffraction. C atoms are coloured green, O atoms are coloured red, N atoms are coloured blue, H atoms are coloured white and the F atom is coloured cyan. H atoms have been omitted for clarity.

The control reaction (Figure 144a) reaches a conversion of 15% with a ratio of 2.4:1 in favour of the **trans-187** cycloadduct after 60000 s. When recognition is introduced the reaction shows different properties (Figure 144b). The formed catalytic ternary complex, accelerates the cycloaddition reaction between the nitrone **177** and the maleimide **112** by more than 100 fold and the stereochemistry of the **trans-188** template is transcribed faithfully – the **trans-188**:**cis-188** ratio is more than 90:1 with a total conversion of 83%. Upon addition of 10 mol% of preformed template **trans-188** the lag period disappears and the reaction starts at its maximum rate. The reaction profile was not recorded for the full time period of the experiment because of precipitation.
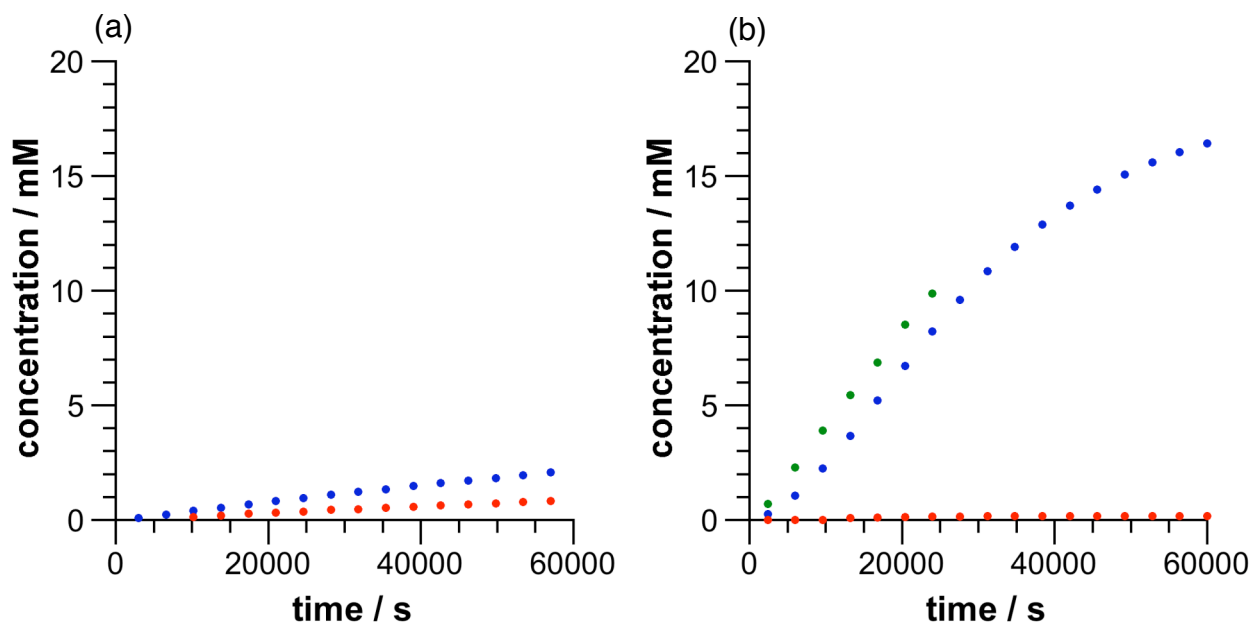
Figure 144. Concentration *vs.* time plots of (a) control reaction between ester maleimide **154** and nitrone **177** ([**154**] = [**177**] = 20 mM) following the formation of products: *trans*-**187** (●) and *cis*-**187** (●). (b) Recognition-enabled reaction between maleimide **112** and nitrone **177** ([**112**] = [**177**] = 20 mM) following the formation of products *trans*-**188** (●) and *cis*-**188** (●). The graph is overlaid with a concentration *vs.* time plot of the same reaction doped with 10 mol% of preformed *trans*-**188** ([**112**] = [**177**] = 20 mM, [*trans*-**188**] = 2 mM) following the formation of *trans*-**188** (●).

The compounds in the system were identified by their characteristic ^{19}F chemical shifts. (Figure 145).
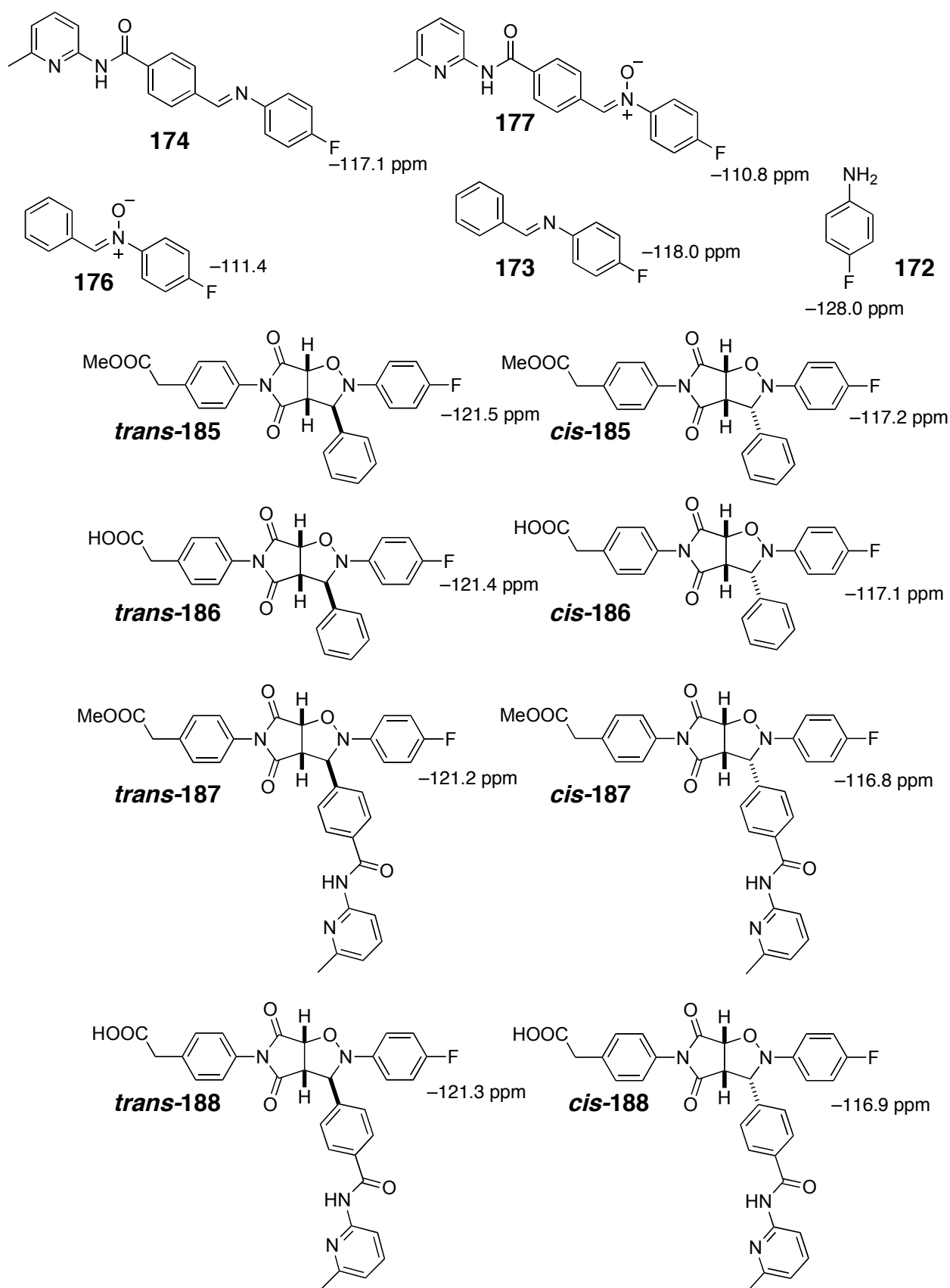
Figure 145. 470.4 MHz ^{19}F NMR chemical shifts, recorded in CD_2Cl_2 saturated with PTSA at 273 K, for all of the fluorine-containing species observed in the exchange experiments.

Again, in the exchange pool, only imine **174** and nitrone **176** are present initially. Since our replicating template **trans-188** is formed from nitrone **177**, some exchange must occur before this replicator can be formed. It was anticipated that, by use of maleimide **112**, which possesses a carboxylic acid recognition site complementary with the amidopyridine recognition site present in nitrone **177**, it would be possible to exploit the more than 100-fold acceleration in the rate of reaction between nitrone **177** and maleimide **112** engendered for the formation of the autocatalytic template **trans-188** within the catalytic ternary complex. This replication process would, in turn, drive the exchange towards the formation of nitrone **177**. Additionally, it was envisaged that the autocatalytic behaviour of **trans-188** would express itself progressively, resulting in this species becoming the dominant one in the product pool at the end of the experiment. Accordingly, a mixture of imine **174**, nitrone **176** and maleimide **112** in CD_2Cl_2 saturated with PTSA ([**174**] = [**176**] = [**112**] = 20 mM) was prepared. Once again, the composition of the mixture was allowed to evolve at 273 K for 16 hours and the coupled exchange and reaction processes were monitored by ^{19}F NMR spectroscopy as described previously (Figure 146).
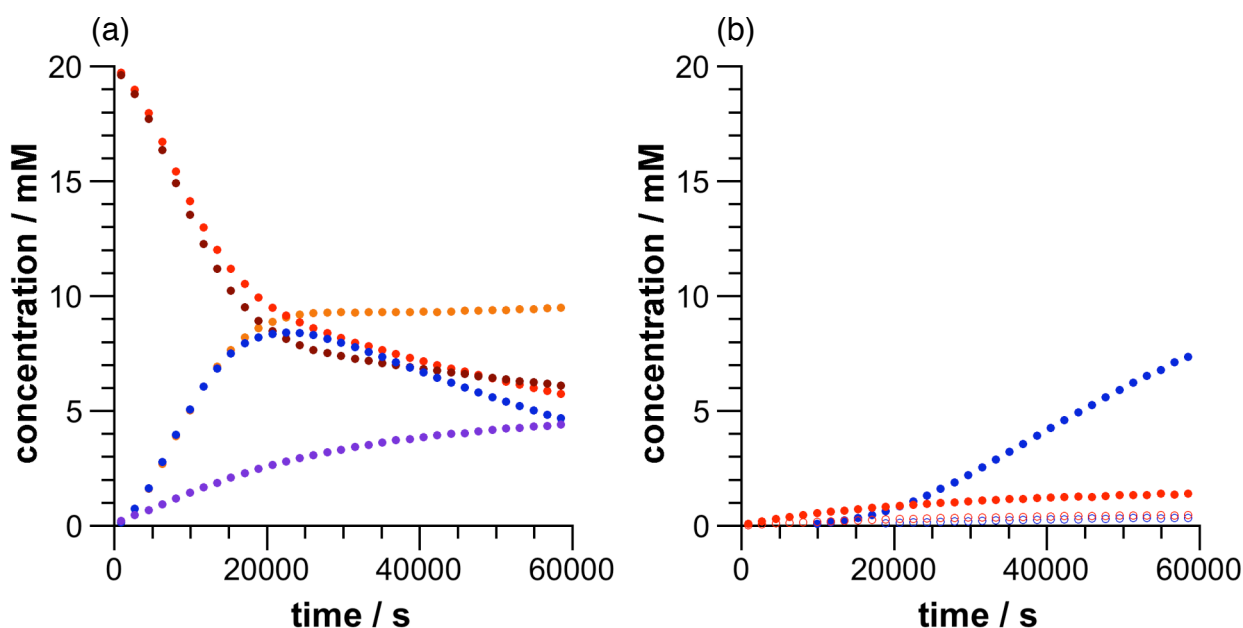
Figure 146. Concentration *vs.* time plots of a) the evolution of the exchange pool generated from imine **174** and nitrone **176** ([**174**] = [**176**] = 20 mM) following the depletion of substrates: imine **174** (●) and nitrone **176** (●) and the formation of products nitrone **177** (●), imine **173** (●), and aniline **172** (●). (b) Evolution of the product pool generated from the exchange pool by the reaction of nitrones **176** and **177** with maleimide **112** following formation of products **trans-186** (●), **cis-186** (○), **trans-188** (●) and **cis-188** (○).

In the product pool, after 16 hours, *trans*-**186** and *cis*-**186** – the products of reaction between nitrone **176** and maleimide **112** – are present at a total concentration of 1.9 mM ([*trans*-**186**]:[*cis*-**186**] = 3:1). After the same time, *trans*-**188** and *cis*-**188** – the products of reaction between nitrone **177** and maleimide **112** – are present at a total concentration of 7.7 mM ([*trans*-**188**]:[*cis*-**188**] = 21:1). Thus, the overall conversion for all cycloaddition reactions within the library is now 48% and cycloadduct *trans*-**188** constitutes almost 80% of the total cycloadduct in the product pool. The effects of the introduction of replication into the system are equally significant in the exchange pool. After 16 hours, compounds **173**, **174**, **176** and **177** are present in dramatically different concentrations compared to the same exchange process in the presence of the control maleimide **154**. The reaction profile of the control reaction is the same as one described in Figure 138 where the same DCL was treated with maleimide **179**, and thus will not again be described in detail. In the recognition-enabled reaction, the rate of the 1,3-dipolar cycloaddition reaction between **176** and **112** is still slower than the rate of exchange. However, the reaction between nitrone **177** and **112** has a comparable maximum rate to the exchange processes. Thus, once exchange generates a concentration of nitrone **177** close to the K_d (~ 2 mM) for the carboxylic acid·amidopyridine complex, reaction to form *trans*-**188** will start to occur through the autocatalytic process mediated by the ternary complex, thereby removing nitrone **177** from the exchange pool rapidly. The depletion of **177** from the exchange pool drives the exchange equilibria involving compounds **173**, **174**, **176** and **177**, regenerating the species removed by the autocatalytic reaction. As the concentration of *trans*-**188** increases, the effect of autocatalysis is to increase the rate of the depletion of **177** until the formation of **177** through exchange becomes limiting overall. Once again an enhancement factor was calculated for the species in the exchange and the product pools (Figure 147).
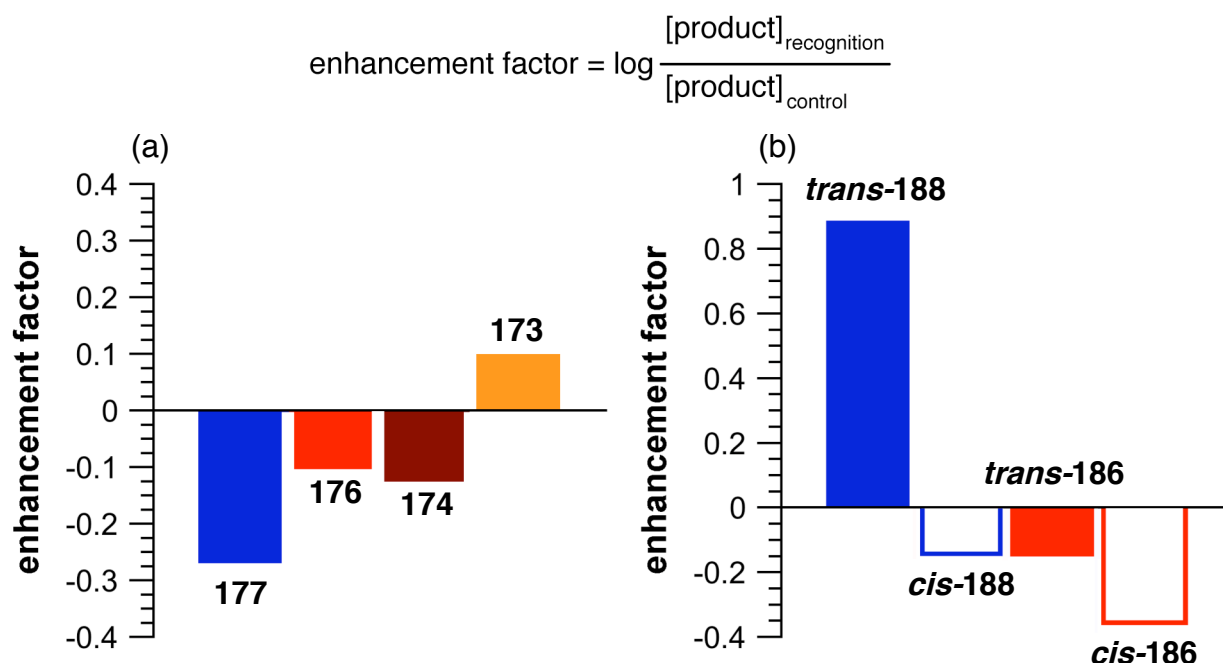
$$\text{enhancement factor} = \log \frac{[\text{product}]_{\text{recognition}}}{[\text{product}]_{\text{control}}}$$

Figure 147. Enhancement factors for compounds in (a) the exchange pool: nitrones **177** and **176** and imines **174** and **173** (b) the product pool: cycloadducts **trans-188**, **cis-188**, **trans-186** and **cis-186**.

In the exchange pool only imine **173** is enhanced, as it does not use any of the components necessary for reaction nor recognition. In the product pool the replicator **trans-188** is enhanced while all other cycloadducts are suppressed.

4.5 Templating a dynamic reaction network

The key feature of a self-replicator is its ability to template its own formation. Experiments in which the reaction mixture is seeded with a small amount of a replicating template are usually used to demonstrate replicating behaviour and should result in a significant enhancement in the formation of the replicator. One might view the addition of template **trans-188** to the exchange pool as an informational input, instructing our dynamic system to synthesise **trans-188**. Therefore, we prepared a mixture of imine **174**, nitrone **176**, maleimide **112** and replicator **trans-188** in CD_2Cl_2 saturated with PTSA ([**174**] = [**176**] = [**112**] = 20 mM; [**trans-188**] = 2 mM). This mixture was allowed to evolve at 273 K for 16 hours and the coupled exchange and reaction processes were monitored ^{19}F NMR spectroscopy as described previously (Fig 148).
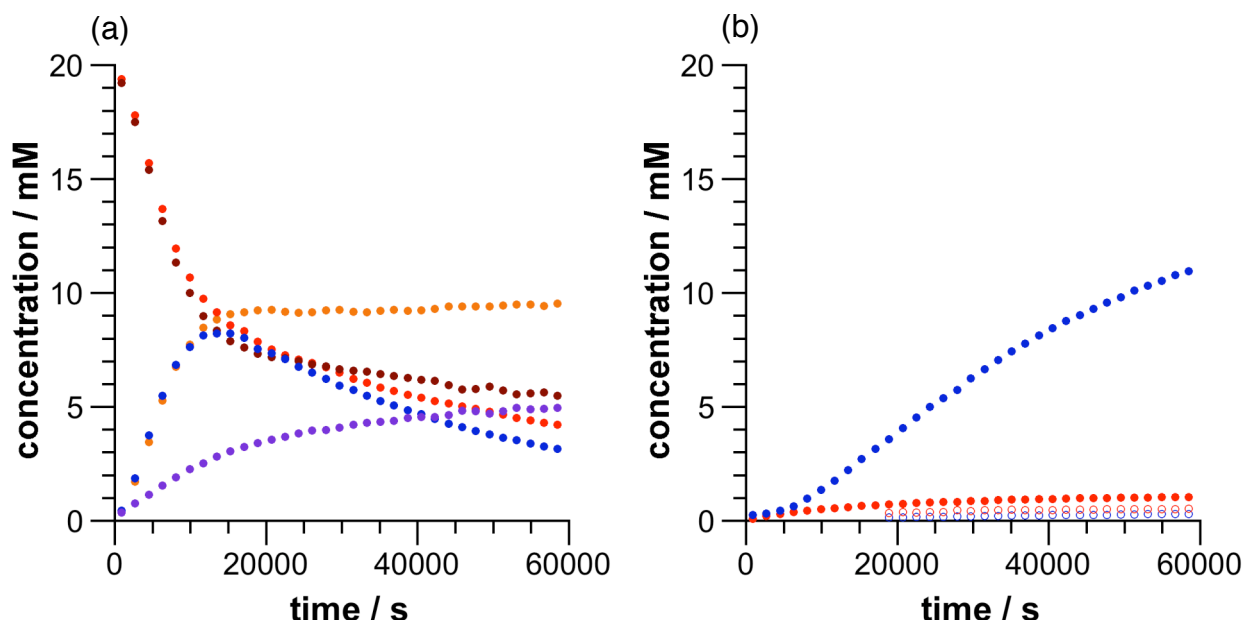
Figure 148. Concentration *vs.* time plots of (a) the evolution of the exchange pool generated from imine **174** and nitrone **176** doped with 10 mol% of preformed ***trans*-188** ([**174**] = [**176**] = 20 mM, [***trans*-188**] = 2 mM) following the depletion of substrates: imine **174** (●) and nitrone **176** (●) and the formation of products nitrone **177** (●), imine **173** (●), and aniline **172** (●). (b) Evolution of the product pool generated from the exchange pool by the reaction of nitrones **176** and **177** with maleimide **112** following formation of products ***trans*-186** (●), ***cis*-186** (○), ***trans*-188** (●) and ***cis*-188** (○).

In the product pool, after 16 h, ***trans*-186** and ***cis*-186** – the products of reaction between nitrone **176** and maleimide **112** – are present at a total concentration of 1.5 mM ([***trans*-186**]:[***cis*-186**] = 2:1). After the same time, ***trans*-188** and ***cis*-188** – the products of reaction between nitrone **177** and maleimide **112** – are present at a total concentration of 11.3 mM ([***trans*-188**]:[***cis*-188**] = 38:1). Thus, the overall conversion for all cycloaddition reactions within the library is now 64% and cycloadduct ***trans*-188** constitutes 88% of the total cycloadduct in the product pool.

The differences in product distribution between the control, recognition-enabled reaction and the template-doped reaction can be easily observed by comparing the ^{19}F NMR spectra of the three cases after 16 h (Figure 149).
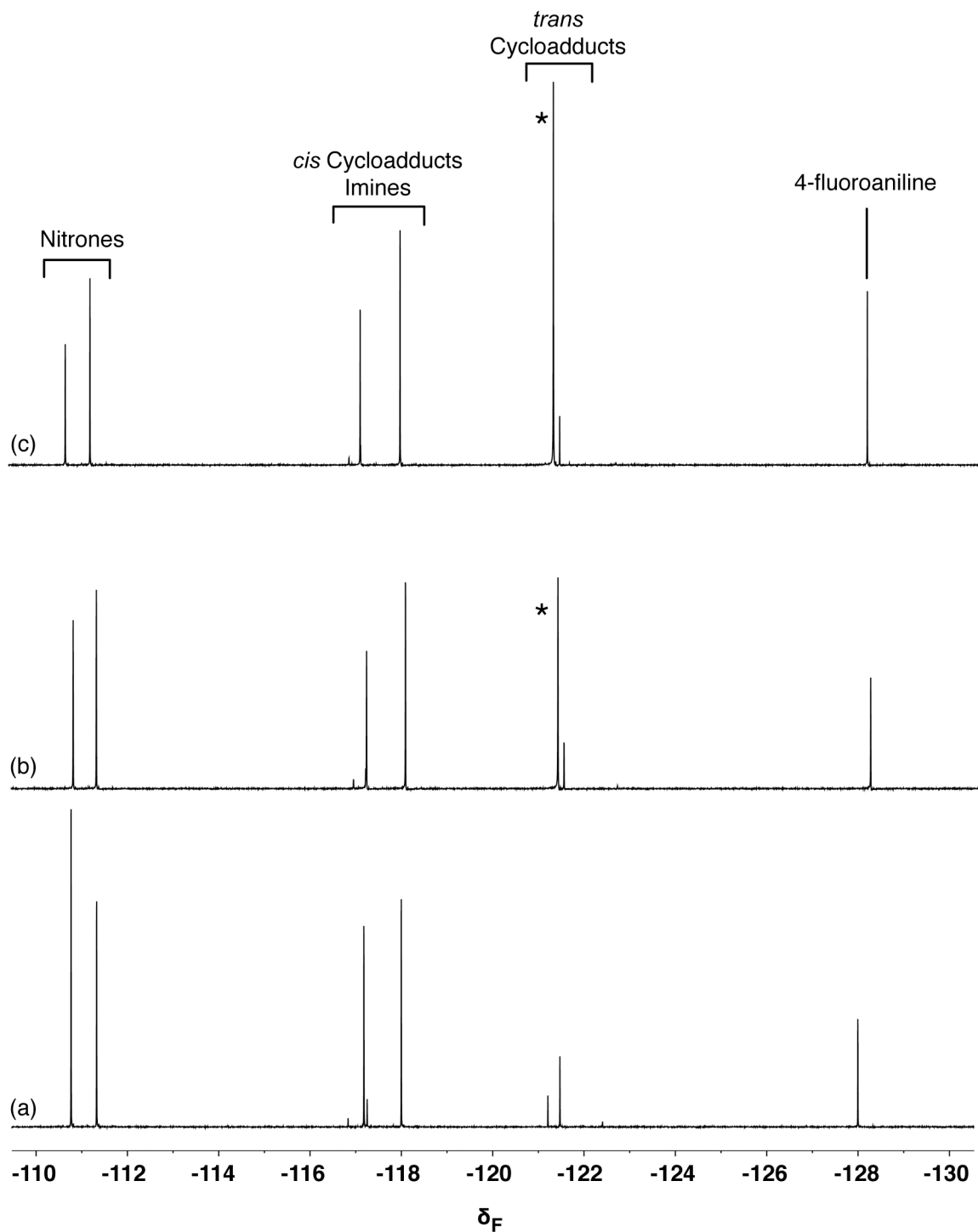
Figure 149. Partial 470.4 MHz ^{19}F NMR spectra, recorded in CD_2Cl_2 saturated with PTSA at 273 K, of a) the control exchange experiment, starting from **174**, **176** and **154** after 16 h, b) the recognition-mediated exchange experiment, starting from **174**, **176** and **112** after 16 h and c) the recognition-mediated exchange experiment, starting from **174**, **176** and **112** and 10 mol% of *trans*-**188** after 16 h. The starred resonance is that arising from *trans*-**188**.

4.6 Analysis of network system-level properties

It is instructive to take a system-level view (Figure 150) of the experiments described above. One can view the exchange pool as containing a finite amount of a resource (the hydroxylamine), which can be converted by the exchange processes into two useable forms – nitrones **176** and **177**. These nitrones can then be converted irreversibly (metabolised) through the cycloaddition reaction with maleimide **112** forming four possible products. The total concentration of all cycloadducts that can be formed is therefore equal to the amount of resource available (20 mM).
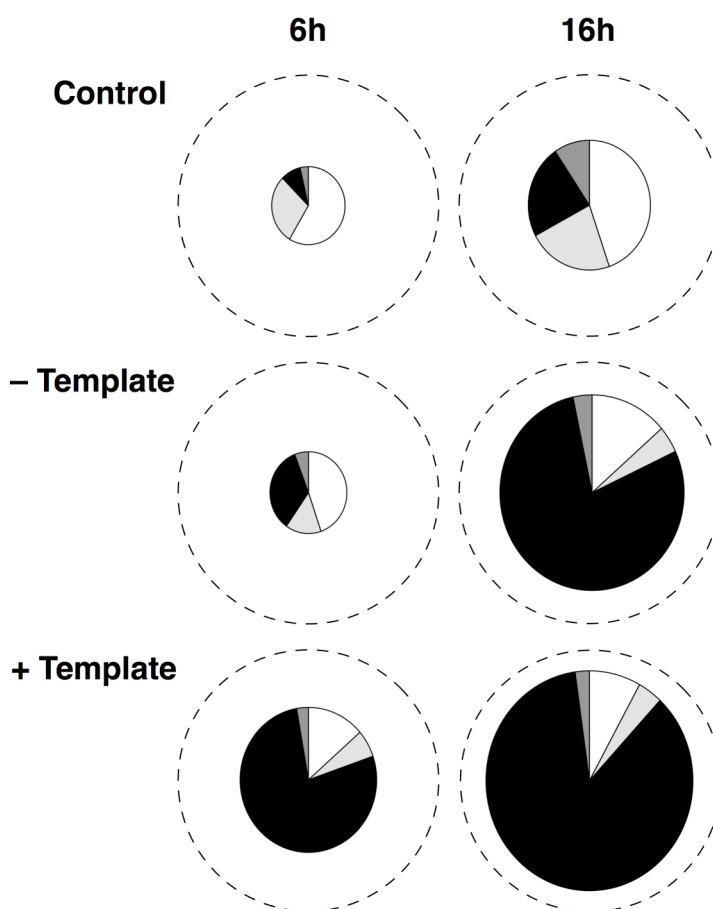
Figure 150. Composition of the product pool after 6 hours and 16 hours. The dashed circle represents 100% conversion to cycloadducts. The area of the pie charts indicates the actual conversion. The shading of the pie chart wedges indicates the composition of the product pool – white = **trans-186**, light grey **cis-186**, dark grey = **cis-188**, black = **trans-188**. Starting concentrations – Control: [**174**] = [**176**] = [**154**] = 20 mM. – Template: [**174**] = [**176**] = [**112**] = 20 mM. +Template: [**174**] = [**176**] = [**112**] = 20 mM; [**trans-188**] = 2 mM. All experiments were performed in CD_2Cl_2 saturated with p-toluenesulfonic acid monohydrate at 273 K.

In the control experiment, there is no overall controlling influence within the system and exploitation of the resource through the reactions of nitrone **176** and **177** is slow (conversion 21% after 16 hours). Since there is no recognition-mediated events in the system, both reactions are bimolecular and the ratio of the products reflects the

availability of the substrates in the mixture and the native selectivity engendered in the 1,3-dipolar cycloaddition reaction. When we unblock the recognition and allow the replicator to emerge within the system it takes some time for its effect to become evident, since the properties of the replicator are non-linear. After 6 hours, the total conversion is very similar to the control experiment (9% and 8%, respectively), but, crucially, the composition of the mixture is not – the replicator, *trans*-188, now makes up 34% of the product pool as opposed to only 9% in the control experiment. From this point onwards, the replicator dominates the system having enough template in solution to assume its maximum autocatalytic rate in the given conditions. Of the 8.7 mM of hydroxylamine converted to cycloadducts in the next 10 hours, 7.7 mM is converted by the replicator and by 16 hours *trans*-188 constitutes 79% of the product pool. The rise of the replicator was accomplished at the expense of other species in the library.

When the exchanging pool is seeded with the replicator the effects are even more dramatic. After 6 hours, the total conversion is much higher (29%), and the composition of the mixture is dominated by the replicator – *trans*-188 already makes up 77% of the product pool. The replicator domination is proportional to the amount of the template product in solution. The more template present the higher chances of the replicator to be the fittest species in the evolution of the network. Adding preformed template at the start of the experiment dramatically enhances this effect. In the next 10 hours, a further 7.6 mM of hydroxylamine is converted to cycloadducts – 92% of this conversion is the formation of *trans*-188 catalysed by itself. The replicator has achieved domination. After 16 hours *trans*-188 constitutes 86% of the product pool and a total of 64% of the initial hydroxylamine resource has been converted. An explanation for the different behaviour of the system in the presence of the instructional template can be found by examining the rates of formation of *trans*-188 in the different scenarios (Figure 151).
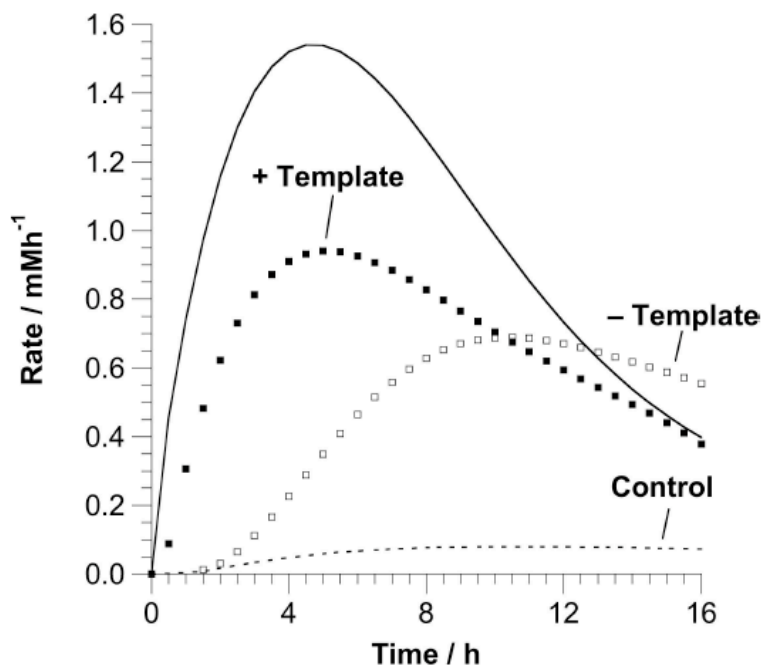
Figure 151. Rate vs. time profiles in the absence (−Template) and in the presence (+Template) of *trans*-**188** (10 mol%) added at the start of the experiment. Data for the control experiment and for replicator *trans*-**188** operating in isolation within a non-dynamic experiment (solid line) are shown for comparison purposes.

In the absence of added replicator, the maximum rate of formation of *trans*-**188** (0.69 mMh^{-1}) is achieved 10 hours into the experiment. By contrast, in the presence of 10 mol% added *trans*-**188**, the maximum rate for replicator formation achieved is higher (0.94 mMh^{-1}) and occurs earlier (5 hours). Therefore, the effect of the small amount of added template is to engender selectivity principally in the early phases of the experiment by ensuring rapid and selective consumption of nitrone **177** through the intermediacy of the recognition processes, which assemble the catalytic ternary complex. The earlier the replicator reaches its full potential the higher the domination over other pathways.

In summary, it has been demonstrated that a replicating template is capable of exploiting and dominating an exchanging pool of reagents in order to amplify its own formation. The structural complexity and information content of the replicator *trans*-**188** are low as its molecular weight is only 580 Da, possesses only three stereocentres and the system relies on a single recognition motif for its function. Therefore, in comparison to nucleic acids, it can be considered to be structurally and informationally simple, but it is still capable of driving the network of exchange

reactions by virtue of the non-linear kinetics inherent in minimal replication. Thus, despite the fact that, at the start of all of the experiments, the concentration of nitrone **177**, which is required to form the replicator, is zero, replicator *trans*-**188** is, in all cases, the dominant species found in the product pool. The results lend weight to hypotheses that a primitive metabolism incorporating a system of autocatalysed reaction cycles might be able to select and amplify replicators leading to molecular evolution. The template-doped reaction is of special interest as the addition of *trans*-**188** can be described as an instructional input saying 'make this'. The dynamic system will adapt in order to create the seeded compound efficiently. In the described system only one such process is possible, but it could be the basis for more complex systems that will adapt to give different outputs with different instructional templates.

4.7 Competition between a replicator and an AB complex pathway in a dynamic scenario

After demonstrating that the dynamic combinatorial library can be influenced using the amplification engendered in the AB complex pathway and the self-replicating cycle. It is possible to envisage both these processes competing for staring material in a single system. The experiment is easily designed using the already utilised library, but this time using both the *meta* **179** and *para* **112** substituted maleimides together. Both maleimides compete for the recognition nitrone **177** formed in the exchange process. Since there are two maleimides present in solution a total of eight products can be formed: four cis isomers (*cis*-**182**, *cis*-**184**, *cis*-**186**, *cis*-**188**) and four trans isomers (*trans*-**182**, *trans*-**184**, *trans*-**186**, *trans*-**188**) (Figure 152). It is expected for two products to be amplified: *trans*-**188** and *cis*-**182**.
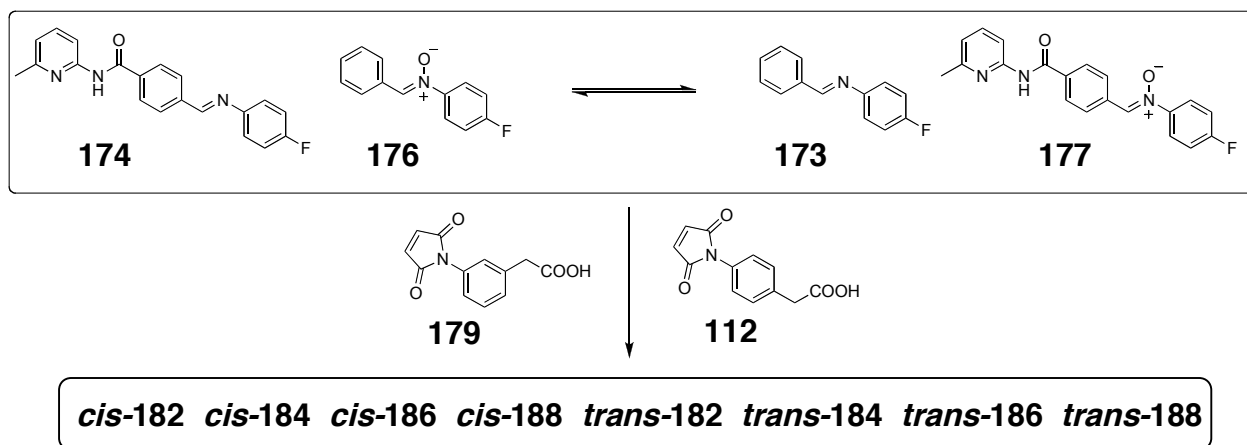
Figure 152. Addition of maleimides **179** and **112** to the dynamic combinatorial library created from imine **174** and nitrone **176** results in the formation of eight possible cycloadducts.

Basing our knowledge on previous experiments it is expected that the AB complex pathway will be faster than the self-replicator in the early stage of the reaction, therefore *cis*-**182** product should be amplified more than *trans*-**188**.

When the imine-nitrone library is subjected to both maleimides **179** and **112** it is possible to distinguish all 13 fluorine-containing compounds in solution using ^{19}F NMR spectroscopy (Figure 153).
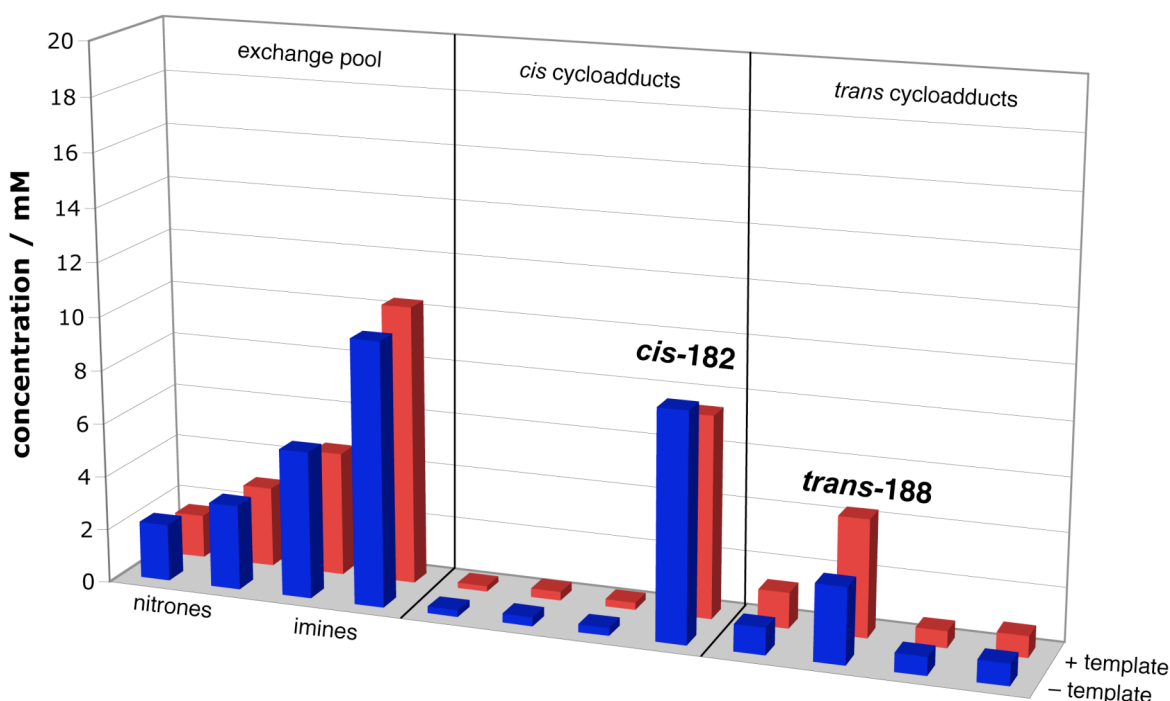
Figure 153. Distribution of compounds in a DCL generated from imine **174** and nitrone **176** together with maleimides **179** and **112** ([**174**] = [**176**] = [**179**] = [**112**] = 20 mM) after 16 h (-template, ■). When the same experiment is repeated with 10 mol% of template *trans*-**188** the formation of *trans*-**188** is enhanced (+ template, ■). In both cases *cis*-**182** is the dominating cycloadduct in the mixture.

161

In the product pool **cis-182** is most abundant reaching a concentration of 8.5 mM and is 28 times higher than the next *cis* product. When comparing the trans cycloadducts it is apparent that the self-replicating **trans-188** has the highest concentration and is the second most abundant product reaching 2.8 mM. The selectivity among the *trans* isomers is not as good as among the *cis* series with the dominating **trans-188** being only 3.3 times higher than the next *trans* product.

It is possible to make use of the templating properties of the replicator in order to enhance its performance. In the next experiment aside the two maleimides **179** and **112** 10 mol% of preformed template **trans-188** was added to the DCL. The hope was that such approach would even the chances in the competition for the hydroxylamine resource between the AB complex pathway and the replicator. An increase in the concentration of **trans-188** at the expense of **cis-182** was observed. The concentration of **trans-188** replicator after 16 h reaches 4.3 mM, compared to 2.8 mM in the undoped reaction and the dominating **cis-182** product reaches 7.5 mM which is 1 mM less then in the previous experiment. This experiment proves that the AB complex pathway being a more simple mechanism of utilising recognition as means of amplifying the rate of reaction dominates the replicator in a competition scenario. Since the recognition nitrone is being fed into the solution *via* the exchange the cycloaddition reactions have to perform at relatively low concentrations. This limitation favours the AB complex pathway, which is more efficient at lower concentrations than the replicator, which bases on the reactive ternary complex. The system and its emergent properties are another example of self-sorting. The amplification from a dynamic combinatorial library is performed in an ordered fashion through two distinct pathways. The compounds have engendered in their design the recognition-mediated reactivity and to what extent it can be utilised to promote their expression form the mixture.

4.8 Conclusions

The systems described here represent one of the first experimental notifications that combine the fundamental ideas behind the emerging area of systems chemistry in a small system. Firstly, the networks are based on dynamic combinatorial chemistry, but on top of that possess supramolecular recognition features, nonlinear amplification and self-replication. It is now possible to determine what are the minimal

requirements for such a system to operate. Utilising synthetically simple building blocks relying on a single recognition event to determine the fate of the system allows for a detailed kinetic study of how the systems evolve with time showing exquisite selectivity. Coupling of a thermodynamic process to a kinetic amplification technique is a promising approach to achieving high selectivity from a DCL. The significant response of the system to a small input of instructional template is also encouraging. This property suggests that it should be possible to develop more complex dynamic recognition-mediated reaction networks, relying on multiple recognition events, such as a combination of auto- and crosscatalytic replicators, to generate and express more complex programmed responses to template inputs through recognition-mediated processes. Systems chemistry attempts to capture the complexity and emergent phenomena prevalent in the life sciences within a wholly synthetic chemical framework. The unique coupling between thermodynamic and kinetic phenomena shown in a minimum complexity scenario brings us closer to understanding the workings of non-linear biological processes.

5. Dynamic reaction networks from simple molecules

5.1 Synthesis of a DCL from structurally simple substrates

The exchange experiments described in the previous Chapter relied on nitrones and imines as starting material. Both of these types of compounds are synthesised *via* reversible reactions of aldehydes with either hydroxylamine or amine. It can be envisaged that the dynamic library is created from the simplest possible building blocks: appropriate aldehydes, amines and hydroxylamines. When the idea of the kinetic amplification is employed in such scenario the outcome of the system should be similar to when using already preformed imines and nitrones for the generation of a DCL. The advantage of such approach is that the starting point for the exchange reaction consists of a mixture of very basic, mostly commercially available compounds. The products amplified from the mixture, on the other hand, posses a high degree of complexity.

The components of the library were identified retrosythetically from the imines and nitrones used in the previous Chapter and consist of: aldehyde **122**, benzaldehyde **171**, 4-fluoroanilne **172** and 4-fluorophenylhydroxylamine **175**.

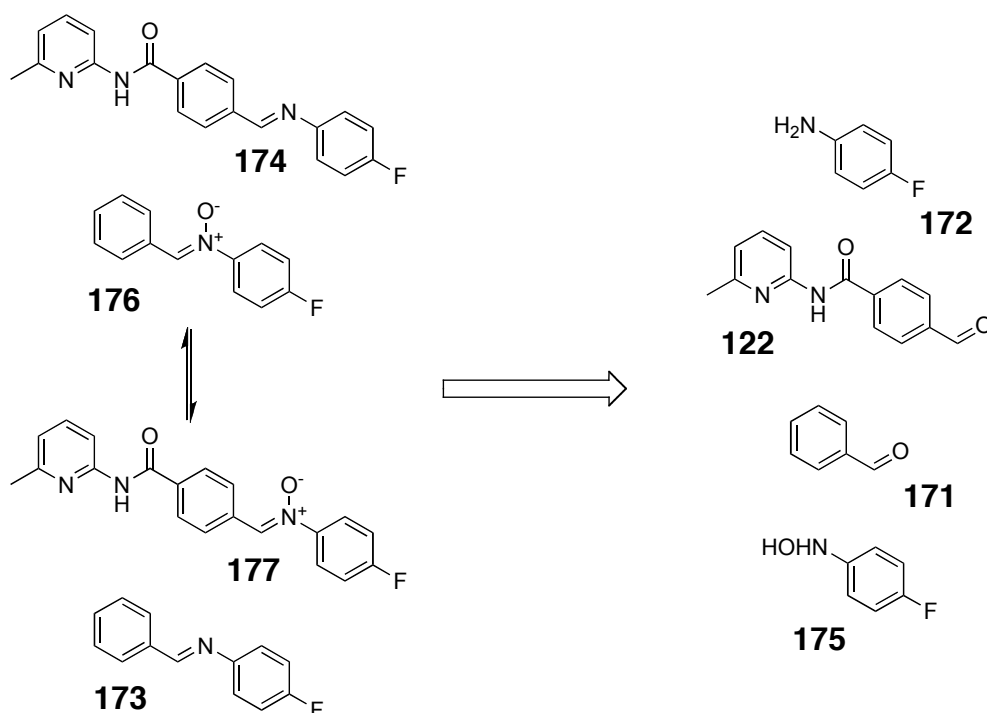

Figure 154. Retrosynthetical analysis of the described DCL in order to identify structurally simpler starting materials **122**, **171**, **172** and **175** from imines **173** and **174** and nitrones **176** and **177**.

The first task was to establish if starting from just aldehydes and nitrogen nucleophile species would lead to the same distribution of library members as the previously studied exchange reactions using nitrones and imines. In order to determine the equilibrium point aldehydes **122** and **171** were dissolved together with 4-fluoroaniline **172** and 4-fluorophenylhydroxylamine **175** ([**122**] = [**171**] = [**172**] = [**175**] = 20 mM) in CD_2Cl_2 saturated with PTSA and left at 273 K until no further changes to the concentrations could be observed (24 h).

Four compounds were used to construct the DCL compared to only two (imine and nitrone) in the previous Chapter. However, at equilibrium five compounds are detected (two nitrones **176** and **177**, two imines **173** and **174** and aniline **172**) at practically the same concentrations.
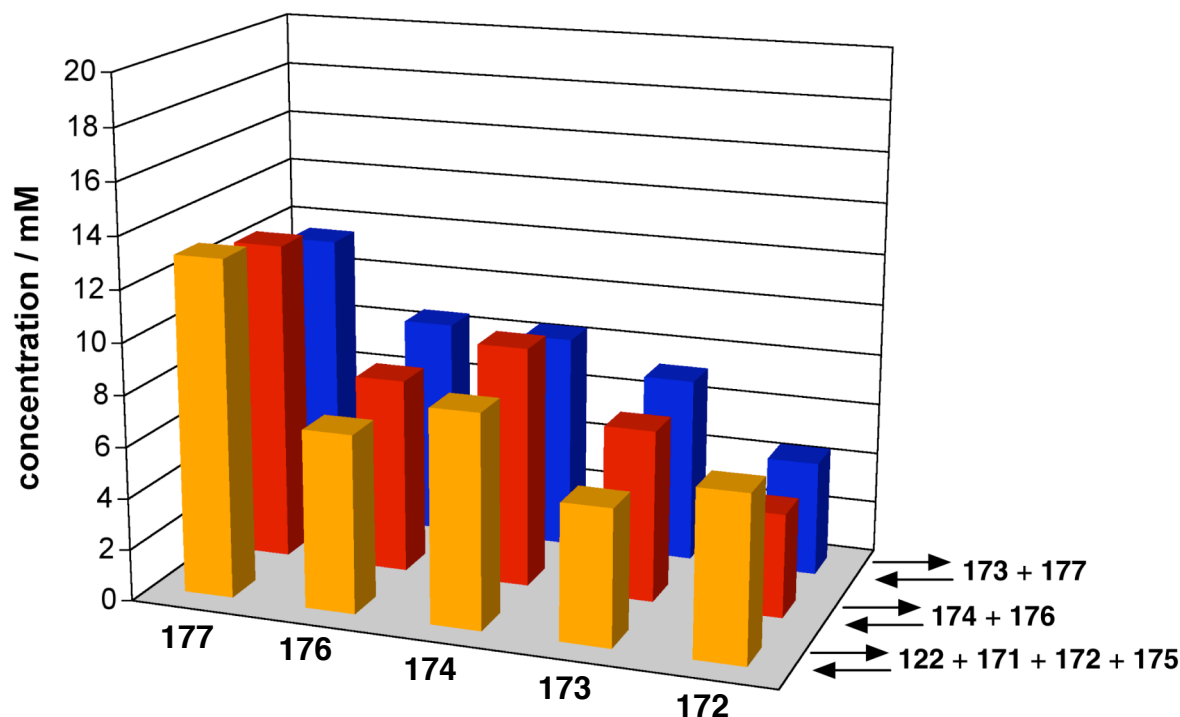
Figure 155. Concentration of the five compounds **172**, **173**, **174**, **176** and **177** detected in the DCL at equilibrium. The starting point for the exchange was either [**173**] = [**177**] = 20 mM; blue bars (■), [**174**] = [**176**] = 20 mM; red bars (■), or [**122**] = [**171**] = [**172**] = [**175**] = 20 mM; orange bars (■). Compare to Figure 130.

The equilibrium point relies on thermodynamics and does not change with different starting materials, however, the route to the equilibrium depends on the initial conditions. The nature and rates of the processes evolving within the system will influence the outcome when irreversible processes are introduced into the DCL.

5.2 Coupling the simple-substrate DCL to a self-replicator

In the examples described previously, at the start of the exchange experiment, one nitrone **176** was at maximum concentration while the second nitrone **177** had to be generated by the exchange process. In the case when we use the simpler starting materials, none of the nitrones are present at the start and both have to be synthesised *in situ* from their corresponding building blocks. The hydroxylamine unit **175** is common for both nitrones, so a competition for this building block is going to be an issue in the system. The nitrone with the higher rate of formation will have an advantage in the subsequent cycloaddition reaction (Figure 156).
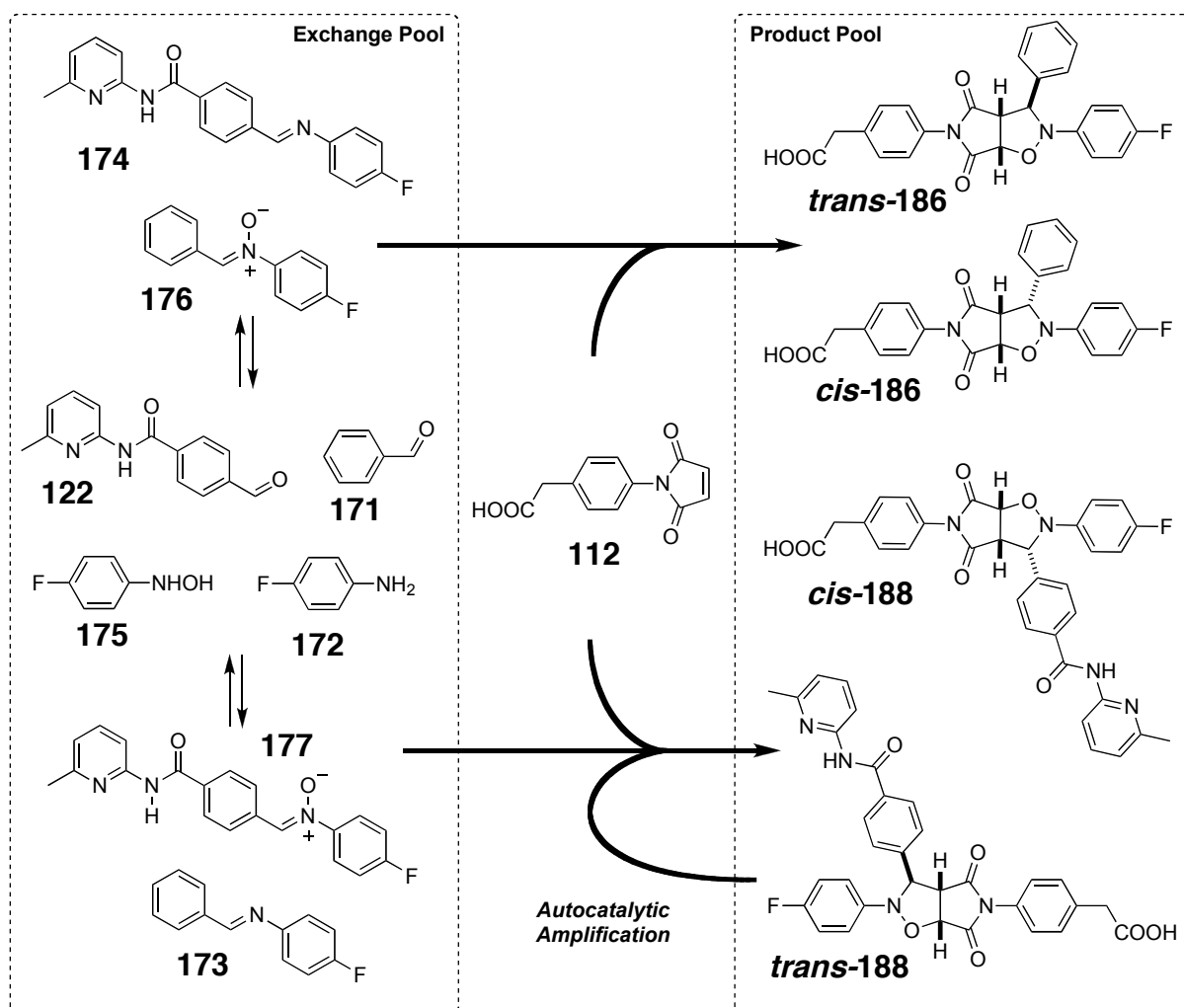
Figure 156. A pool of compounds containing aldehydes **122** and **171**, aniline **172** and hydroxylamine **175** can combine to form imines **173** and **174** and nitrones **176** and **177**. All compounds can exchange freely in CD$_2$Cl$_2$ saturated with *p*-toluenesulfonic acid monohydrate at 273 K. Material can be transferred irreversibly to a pool of products, present in the same solution, that cannot be interconverted or returned to the exchange pool, through reaction of nitrones **176** or **177** with an appropriate maleimide **112**.

The structurally simple compounds were dissolved in CD$_2$Cl$_2$ saturated with PTSA together with maleimide **112** responsible for the formation of the replicating template

with nitrone **177** ([**122**] = [**171**] = [**172**] = [**175**] = [**112**] **=** 20 mM). The system was monitored using ^{19}F NMR spectroscopy at 273 K over a period of 16 h (Figure 157).
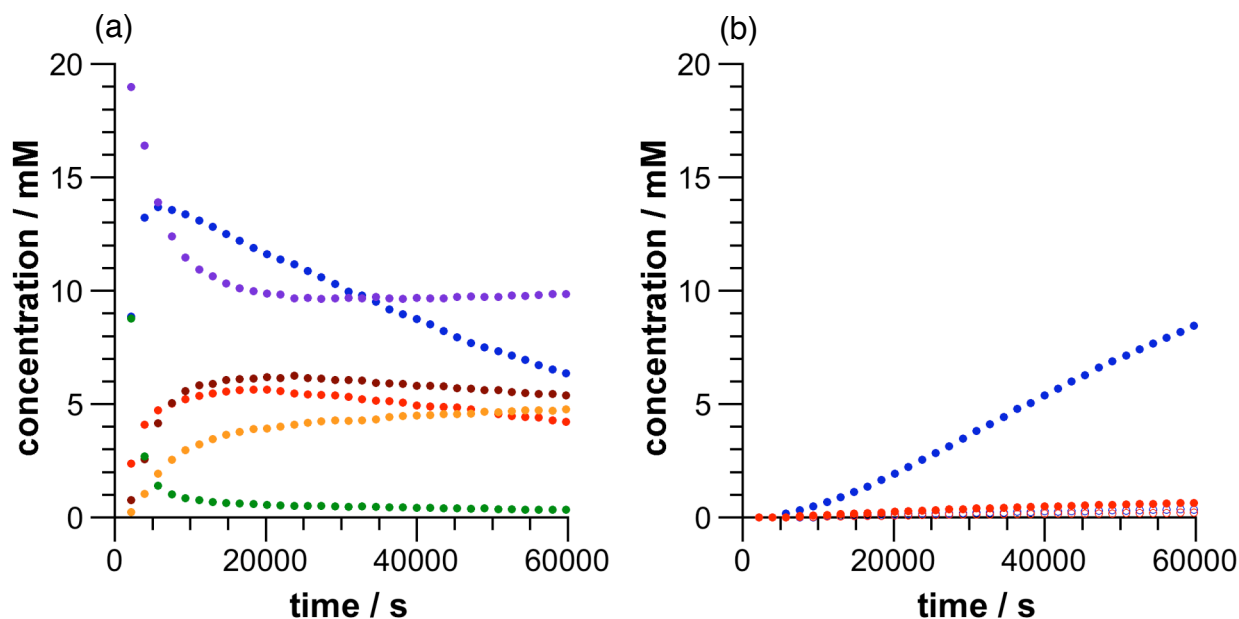

Figure 157. Concentration *vs.* time plots of (a) the evolution of the exchange pool generated from aldehydes **122** and **171**, aniline **172** and hydroxylamine **175** ([**122**] = [**171**] = [**172**] = [**175**] = 20 mM) following the depletion of substrates: aniline **172** (●) and hydroxylamine **175** (●) and the formation of products nitrones **177** (●) and **176** (●) and imines **173** (●) and **174** (●). (b) Evolution of the product pool generated from the exchange pool by the reaction of nitrones **176** and **177** with maleimide **112** following formation of products *trans*-**186** (●), *cis*-**186** (○), *trans*-**188** (●) and *cis*-**188** (○).

In this scenario the only compounds with a starting concentration of 20 mM that bear a fluorine tag are the aniline **172** and the hydroxylamine **175**. As expected the hydroxylamine being much more nucleophilic than the aniline is depleted more rapidly forming the two nitrones **176** and **177**. Unfortunately the nitrones are being formed at considerably different rates. The recognition nitrone **177** is synthesised faster reaching a concentration of 14 mM after 5000 s. In the same time nitrone **176** reaches only 5 mM. At this point already 95% of the hydroxylamine **175** is consumed. This disproportion increases the selectivity in the exchange pool in favour of the recognition nitrone. The selectivity will be transmitted further to the product pool promoting the replicator *trans*-**188**, which is already favoured, as recognition-mediated reaction is faster than the cycloaddition of nitrone **176** that bears no recognition. The difference in rates is a result of different reactivity of the two aldehydes, with the recognition aldehyde **122** being more reactive towards the hydroxylamine **175** then benzaldehyde **171**.

5.3 Single component reactivity modification and DCL properties

The goal of the experiment was to achieve selection by employing solely the recognition events, such as replication, for resolving the library and the concentration advantage of the recognition nitrone **177** can be considered as 'loading the dice' in favour of the replicator. In order to minimise the effect of the higher reactivity of the recognition aldehyde a different competing non-recognition aldehyde had to be chosen. Instead of using just a simple benzaldehyde **171**, 4-trifluoromethyl benzaldehyde **189** was chosen (Figure 158). The effect of the CF_3 group was to activate the aldehyde reacting group towards the condensation reaction with the hydroxylamine. The above experiment was repeated using 4-trifluoromethyl benzaldehyde **189** instead of benzaldehyde **171** (Figure 159).
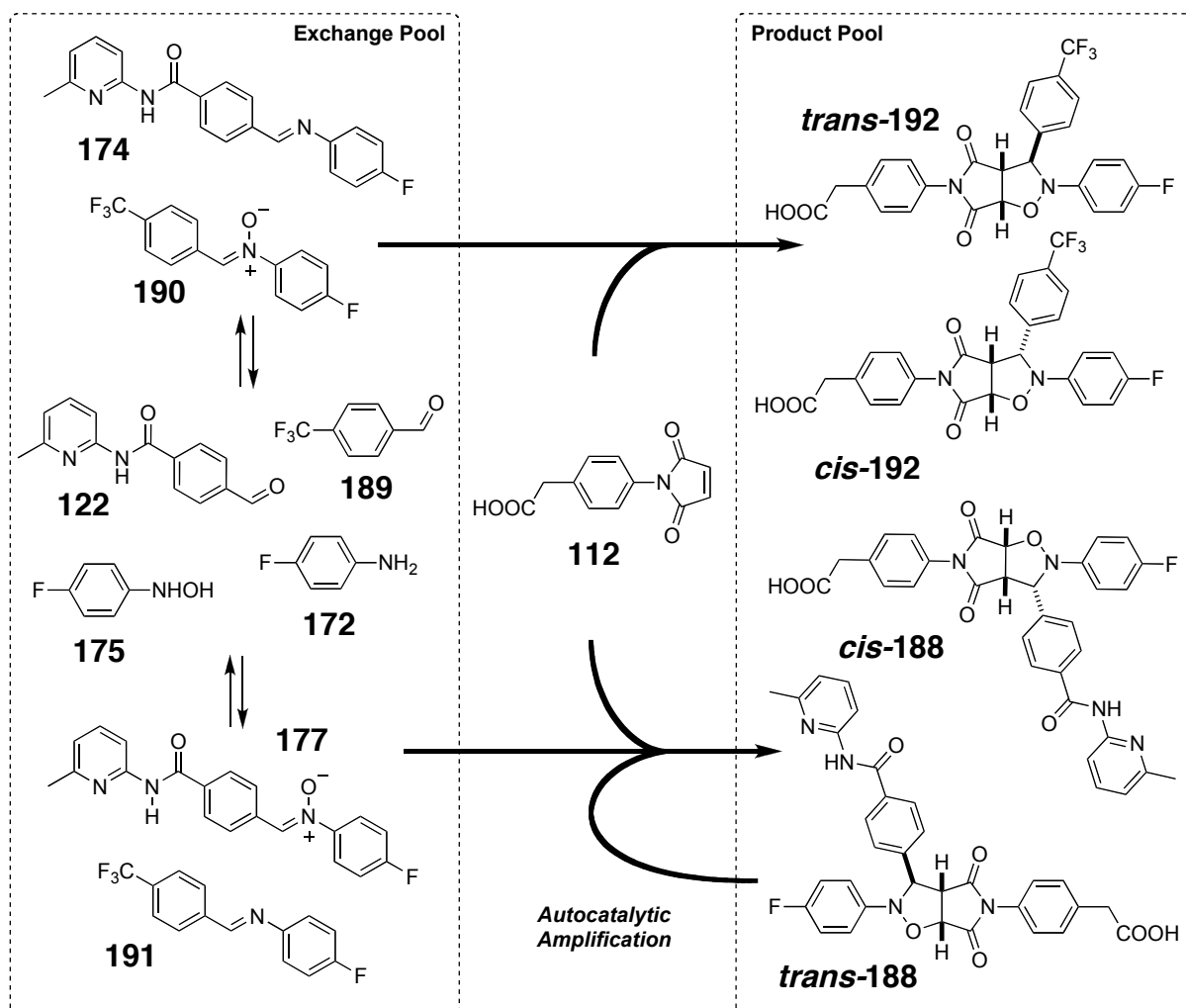
Figure 158. A pool of compounds containing aldehydes **122** and **189**, aniline **172** and hydroxylamine **175** can combine to form imines **174** and **191** and nitrones **177** and **190**. All compounds can exchange freely in CD_2Cl_2 saturated with *p*-toluenesulfonic acid monohydrate at 273 K. Material can be transferred irreversibly to a pool of products, present in the same solution, that cannot be interconverted or returned to the exchange pool, through reaction of nitrones **177** or **190** with an appropriate maleimide **112**.

168

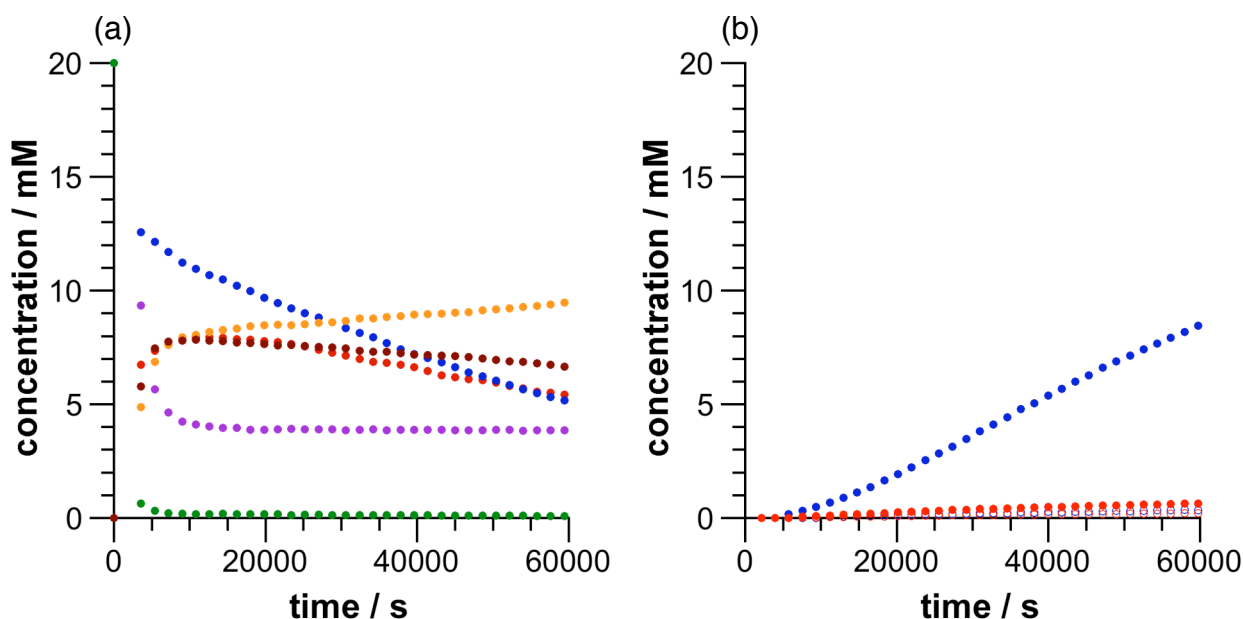**Figure 159.** Concentration *vs.* time plots of (a) the evolution of the exchange pool generated from aldehydes **122** and **189**, aniline **172** and hydroxylamine **175** ([**122**] = [**189**] = [**172**] = [**175**] = 20 mM) following the depletion of substrates: aniline **172** (⬤) and hydroxylamine **175** (⬤) and the formation of products nitrones **177** (⬤) and **190** (⬤) and imines **191** (⬤) and **174** (⬤). (b) Evolution of the product pool generated from the exchange pool by the reaction of nitrones **190** and **177** with maleimide **112** following formation of products *trans*-**192** (⬤), *cis*-**192** (◯), *trans*-**188** (⬤) and *cis*-**188** (◯).

Both nitrones are formed very quickly with the recognition nitrone **177** dominating, but the difference is smaller than when using benzaldehyde. At 5000 s the recognition nitrone **177** reaches the concentration of 12 mM while nitrone **190** is at 7.8 mM. Again, at this point almost all of the hydroxylamine **175** (99%) is consumed. The difference in concentration of the nitrones at this early stage is 4.2 mM and compared with 9 mM in the previous scenario is an improvement. The recognition nitrone **177** is still dominating, but to a lesser degree. The formation of all components in this kind of system depends on their relative rates of formation as well as on the rate of exchange between them. Another apparent difference between the two systems (one using benzaldehyde **171** and the other using p-trifluoromethyl benzaldehyde **189**) is an improved stability of the imine that utilises the aldehyde **189**. Nitrone **190** would probably reach higher concentration if not for the higher rate of the formation of imine **191** that uses the same aldehyde. The higher presence of imines in the system using CF$_3$ aldehyde **189** is clearly noticed when we compare the amount of aniline **172** at equilibrium. The amount of aniline **172** falls to 10 mM in the first case (Figure 157a) and to 4 mM in the latter (Figure 159a). This difference indicates there is 6 mM more

of imine species in solution compared to the first example. This result shows how a slight change of reactivity of one component can influence the distribution of all library members. The main property of such systems is the interconnectedness of all species and the change of properties of any component will change the emergent properties of the system as a whole. Interestingly, the differences in the exchange pool do not influence the distribution in the products pool. The dominating replicator *trans*-**188** in both cases reaches the concentration of about 8 mM. In both scenarios the concentration of the recognition nitrone **177** is practically the same explaining the similar rate of formation of *trans*-**188**. The selectivity over the other *trans* cycloadduct is better when using benzaldehyde **171** to create the DCL and reaches 13:1 after 16 h compared to 10:1 when using the CF$_3$ aldehyde **189**. This difference is easily explained by a higher concentration of nitrone **190** in the second case promoting the bimolecular reaction generating cycloadduct *trans*-**192**. The selectivity is also higher than in the scenario where the simple nitrone **176** and recognition imine **174** were the starting materials for the DCL. In this case the bimolecular reaction performed better especially at the start of the experiment as the concentration of the substrates for the cycloadduct *trans*-**186** was 20 mM, while in the described system the maximum concentration of nitrone **191** is 8 mM.

Overall the performance of the recognition-mediated reactions in the dynamic scenario is poorer than in a static reaction where just two substrates are used. First of all in the dynamic systems the recognition-mediated reactions have to perform at lower concentrations. Another explanation for such behaviour is the presence of competitive unreactive binders for the carboxylic acid recognition site. The recognition imine **174**, as well as the recognition aldehyde **122**, engages in formation of hydrogen bonds with the carboxylic acid group of the maleimide **112**. These compounds might be considered inhibitors for the formation of the reactive ternary complex necessary in the process of self-replication. In such mixture the performance of the replicator is limited.

It has been demonstrated that one can use self-replication to achieve selection from a DCL constructed from structurally simple building blocks. The most important aspect of this kind of approach is the comparison of the simplicity of the starting material compared to the complexity of the enhanced product. It is possible to observe all the complexities of a replicator coupled to a DCL in a simple to prepare experiment.

5.4 Coupling the simple-substrate DCL to an AB complex reaction

A nitrone-imine DCL generated by using structurally simple compounds can also be coupled with amplification based on the AB complex pathway. To the starting materials (aldehydes **122** and **189**, 4-fluoroaniline **172** and hydroxylamine **175**) *meta* acid maleimide **179** was added ([**122**] = [**189**] = [**172**] = [**175**] = [**179**] = 20 mM) (Figure 160). Again the target compound is the recognition nitrone **177** that will react with the maleimide **179** through the reactive AB complex as described in Section 2.4. It is expected for the reaction to be much faster especially at the start of the experiment, so a more rapid depletion of compounds that use the same building blocks as the enhanced product will be observed (Figure 161).
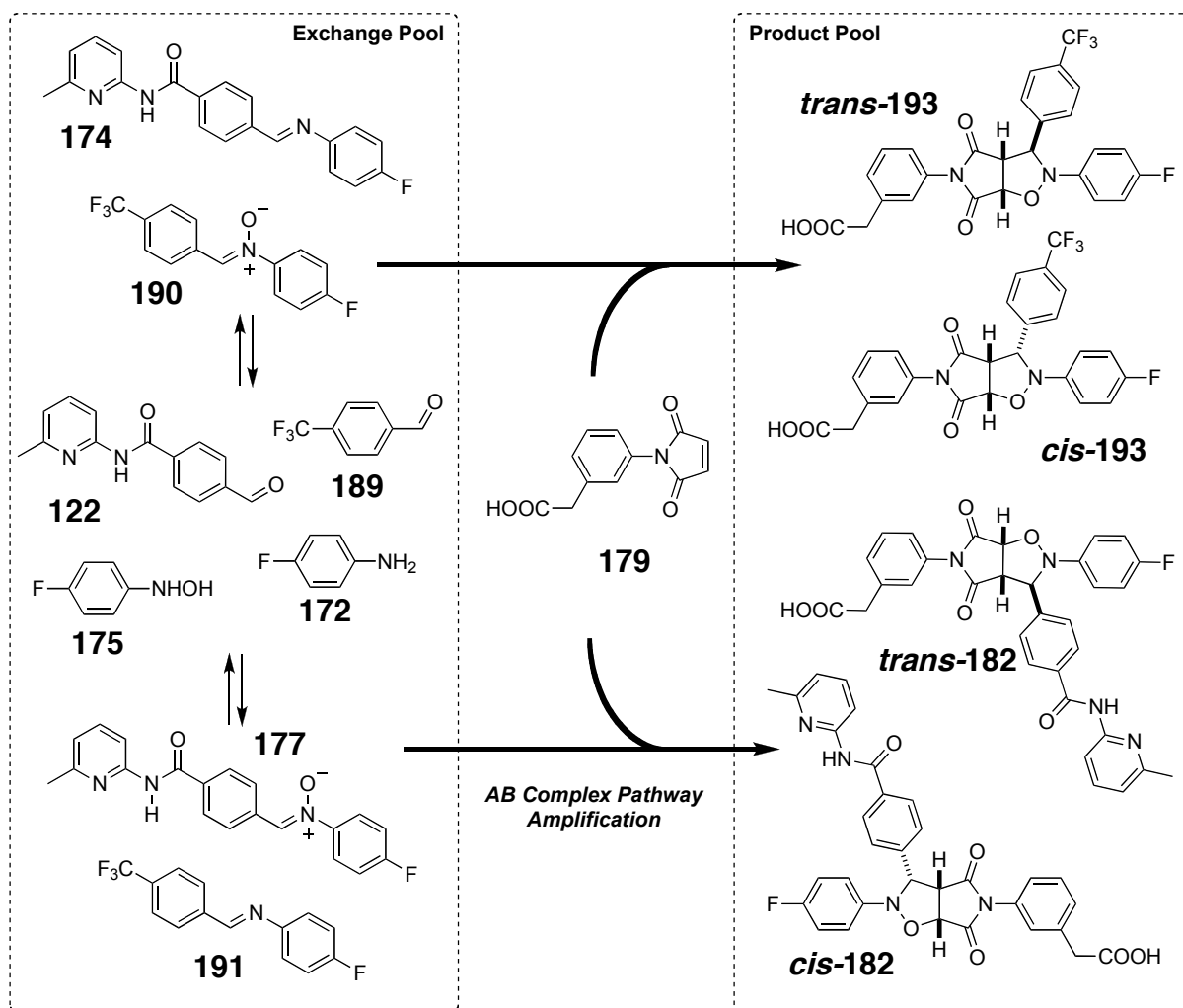
Figure 160. A pool of compounds containing aldehydes **122** and **189**, aniline **172** and hydroxylamine **175** can combine to form imines **174** and **191** and nitrones **177** and **190**. All compounds can exchange freely in CD$_2$Cl$_2$ saturated with *p*-toluenesulfonic acid monohydrate at 273 K. Material can be transferred irreversibly to a pool of products, present in the same solution, that cannot be interconverted or returned to the exchange pool, through reaction of nitrones **177** or **190** with an appropriate maleimide **179**.

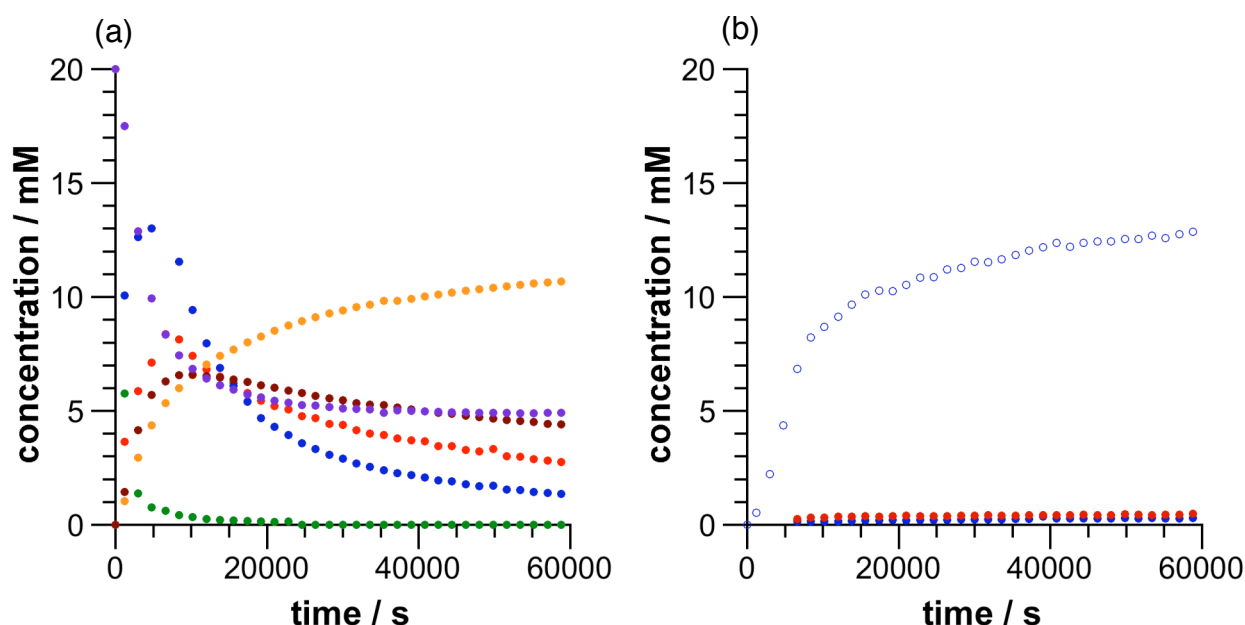**Figure 161.** Concentration *vs.* time plots of (a) the evolution of the exchange pool generated from aldehydes **122** and **189**, aniline **172** and hydroxylamine **175** ([**122**] = [**189**] = [**172**] = [**175**] = 20 mM) following the depletion of substrates: aniline **172** (●) and hydroxylamine **175** (●) and the formation of products nitrones **177** (●) and **190** (●) and imines **191** (●) and **174** (●). (b) Evolution of the product pool generated from the exchange pool by the reaction of nitrones **190** and **177** with maleimide **179** following formation of products *trans*-**193** (●), *cis*-**193** (○), *trans*-**182** (●) and *cis*-**182** (○).

The reaction amplified *via* the AB complex pathway is very fast and only a very small lag period is observed. Once the recognition nitrone **177** reaches its maximum concentration of 13 mM at 5000 s the formation of the ***cis*-182** is already at its highest rate. From that point the rate of the AB complex reaction slows down, as the concentration of nitrone **177** drops and the rate of exchange becomes the limiting factor. The ***cis*-182** cycloadduct dominates the product pool reaching a selectivity of 26:1 over the next most abundant product ***trans*-193**. In this case the cycloadduct ***cis*-182** reaches the highest concentration of all compounds in the mixture including imine **191**.

5.5 Opening new reaction pathways by geometrical modifications of DCL components

Using just basic substrates greatly simplifies the setup of the DCL allowing for a ready formation of many interesting combinations and easy screening of the mixture using ^{19}F NMR spectroscopy. So far, the differentiating agent in the systems was the substitution pattern of the maleimide, either *para* **112** or *meta* **179**. The difference in

geometry of the maleimide channelled the reaction either *via* the self-replicating or the AB complex pathway. It is possible to introduce a similar structural motif in the aldehyde component with the recognition group *meta* or *para* to the formyl group. The *meta* substituted aldehyde was synthesised from commercially available substrates (Figure 162). 2-Amino-6-methylpyridine **119** was condensed with 3-chloromethylbenzoyl chloride **194** to afford the 3-chloromethyl-*N*-picoline **195** in almost quantitative yield. Treatment of the chloromethyl substituted amide **195** with hexamethylene tetramine and hydrochloric acid afforded the *meta* substituted aldehyde **196**.
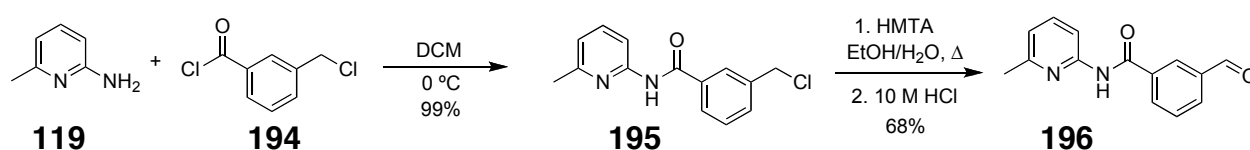
Figure 162. Synthesis of aldehyde **196**.

Aldehyde **196** will lead to the formation of nitrones of different geometry which together with the *meta* and *para* maleimides **179** and **112** will form new possible reactive pairs. So far, only the reactions of the *para* nitrone **177** were investigated. The *meta* nitrone **197** will display different reactivity (Figure 163).

Figure 163. Properties of different reactive pairs formed using nitrones **177** and **197** and maleimides **112** and **179**.

173

All experiments were conducted with the starting concentration of the substrates equal to 20 mM in CD_2Cl_2 saturated with PTSA at 273 K and monitored with ^{19}F NMR spectroscopy after 16 h. The possible combinations of library components led to formation of systems with increasing complexity from 4 up to 12 cycloadducts derived from the exchanging pool of substrates.
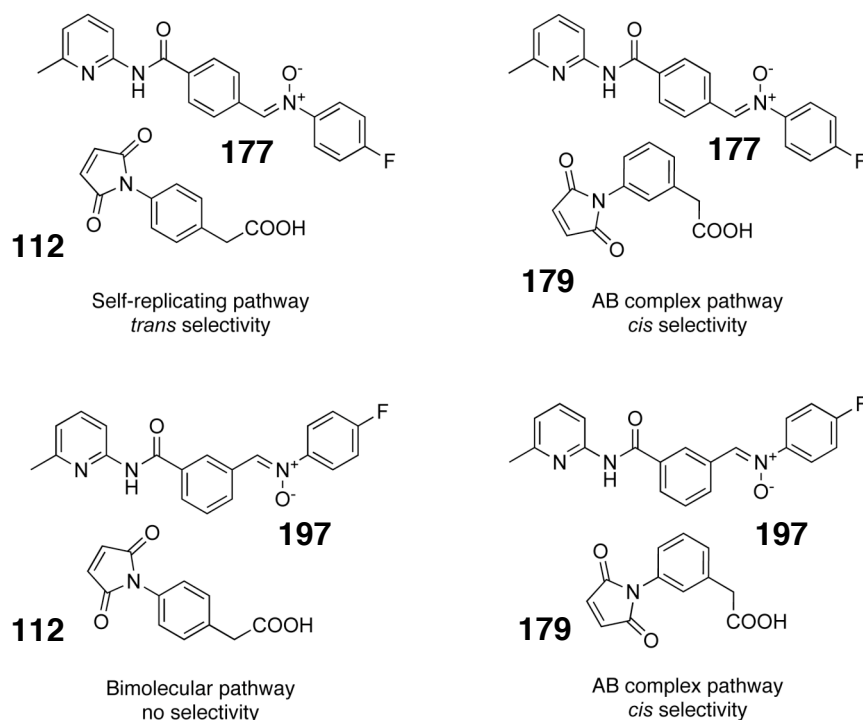
Constructing the system using *meta* aldehyde **196**, 4-trifluoromethyl benzaldehyde **189**, 4-fluoroaniline **172**, 4-fluorophenylhydroxylamine **175** and the *para* substituted maleimide **112** ([**196**] = [**189**] = [**172**] = [**175**] = [**112**] = 20 mM) establishes the nitrone-imine exchange system with four possible cycloaddition products (Figure 164). The reaction between the *meta* nitrone **197** and the *para* maleimide **112** does not display any enhancement at 273 K neither of the ***cis*-198** cycloadduct nor the ***trans*-198** cycloadducts (Figure 165). This lack of activity is a result of unfavourable relative geometry of the substrates that upon recognition does not stabilise the transition state of neither the AB complex nor the ternary complex of the replicating cycle.
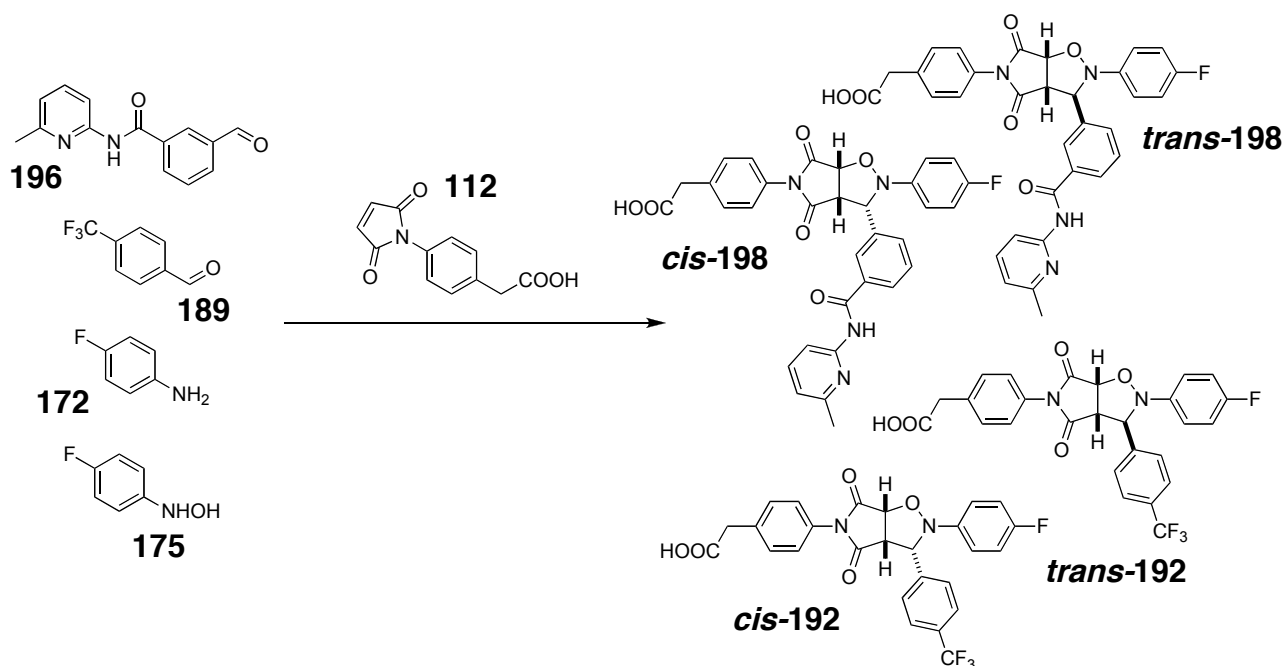

Figure 164. Cycloadducts derived from a DCL created from compounds **196**, **189**, **172** and **175** when treated with maleimide **112**.

Figure 165. Concentrations of the *cis* cycloadducts (■) and *trans* cycloadducts (■) detected in the reaction mixture of a DCL created from compounds **196**, **189**, **172** and **175** treated with maleimide **112** ([**196**] = [**189**] = [**172**] = [**175**] = [**112**] = 20 mM).

The ratios between the *cis* and *trans* cycloadducts derived from the exchange pool (1:2.1 and 1:2.4) suggest that the reactions of both nitrones **197** and **190** with maleimide **112** are bimolecular. The cycloadducts derived from aldehyde **196** are present in slightly higher concentrations since the recognition nitrone **197** is synthesised *in situ* faster than nitrone **190**. The recognition does not play a role in the enhancement of the formation of cycloadducts **198**, since native *trans* to *cis* ratio for a 1,3-dipolar cycloaddition is conserved. Overall there is no significant selectivity in the product pool.

A different distribution of the products is expected when the same exchanging pool is treated with maleimide **179** (Figure 166). The geometry of the complex between the *meta* nitrone **197** and the *meta* maleimide **179** allows for a pseudo-intramolecular reaction opening the AB complex channel. This property is expected to enhance the formation of the *cis* diastereoisomer in the mixture (Figure 167).

Figure 166. Cycloadducts derived from a DCL created from compounds **196**, **189**, **172** and **175** when treated with maleimide **179**.
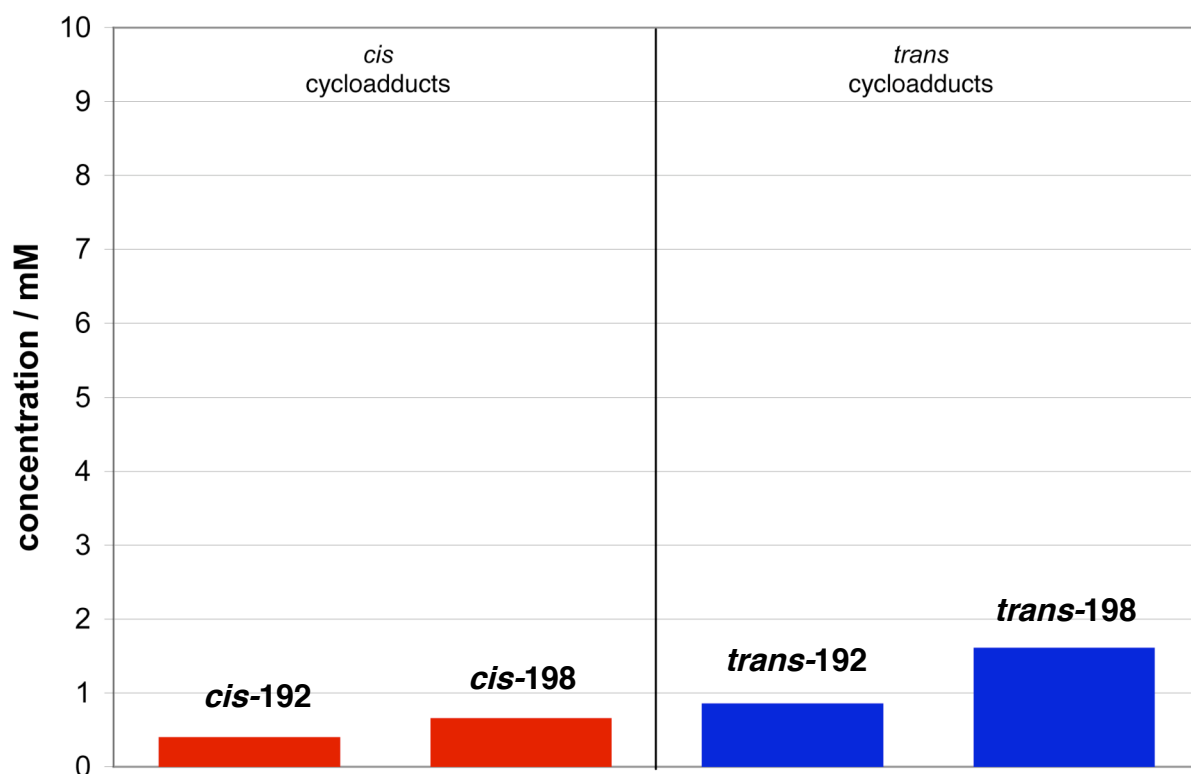

Figure 167. Concentrations of the *cis* cycloadducts (■) and *trans* cycloadducts (■) detected in the reaction mixture of a DCL created from compounds **196**, **189**, **172** and **175** treated with maleimide **179** ([**196**] = [**189**] = [**172**] = [**175**] = [**179**] = 20 mM).

The enhancement of the recognition **cis-199** product is apparent with the compound reaching a concentration of 6.1 mM achieving a 23:1 selectivity over the **cis-193** cycloadduct after 16 h. When comparing this result to the previously described

AB complex pathway system using the *para* aldehyde **112** and the *meta* maleimide **179** (Figure 161) this system is much less efficient. The **cis-182** product in the previous experiment reached a concentration of 13 mM after the same time with a more than 120:1 selectivity over the **cis-193** product. This result indicates that the *meta* nitrone **197** and the *meta* maleimide **179** complex is far from assuming the perfect geometry to create a reactive AB complex pathway. If the two AB complex pathways were to compete with each other in one network it would be expected for **cis-182** to dominate over **cis-199**.

The use of structurally simple substrates for the formation of the networks allows for an easy setup of an experiment to prove the above assumption. When both recognition aldehydes (**122** and **196**) are introduced together with 4-trifluoromethyl benzaldehyde **189**, 4-fluoroaniline **172**, 4-fluorophenylhydroxylamine **175** ([**122**] = [**196**] = [**189**] = [**172**] = [**175**] = [**179**] = 20 mM) (Figure 168) the exchange pool consists of six compounds (three nitrones and three imines) as well as six compounds will be formed in the products pool including **cis-182** and **cis-199** (Figure 169).
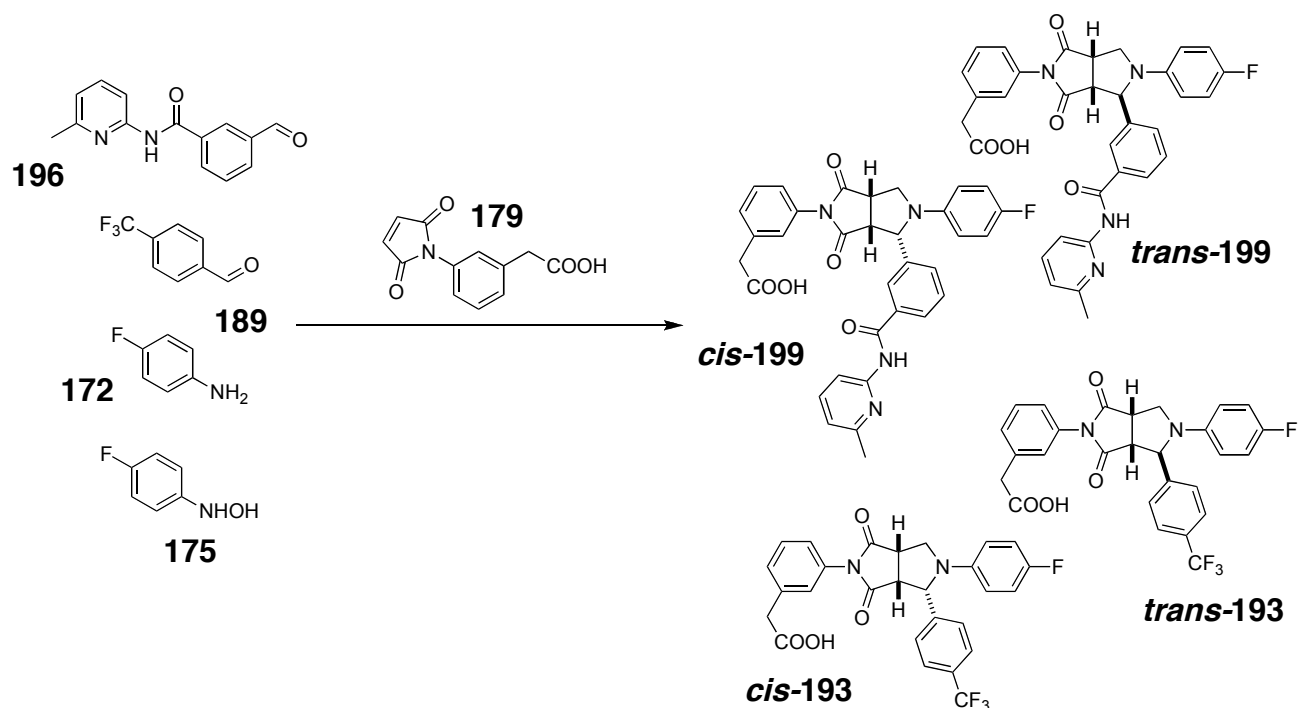

Figure 168. Cycloadducts derived from a DCL created from compounds **122**, **196**, **189**, **172** and **175** when treated with maleimide **179**.
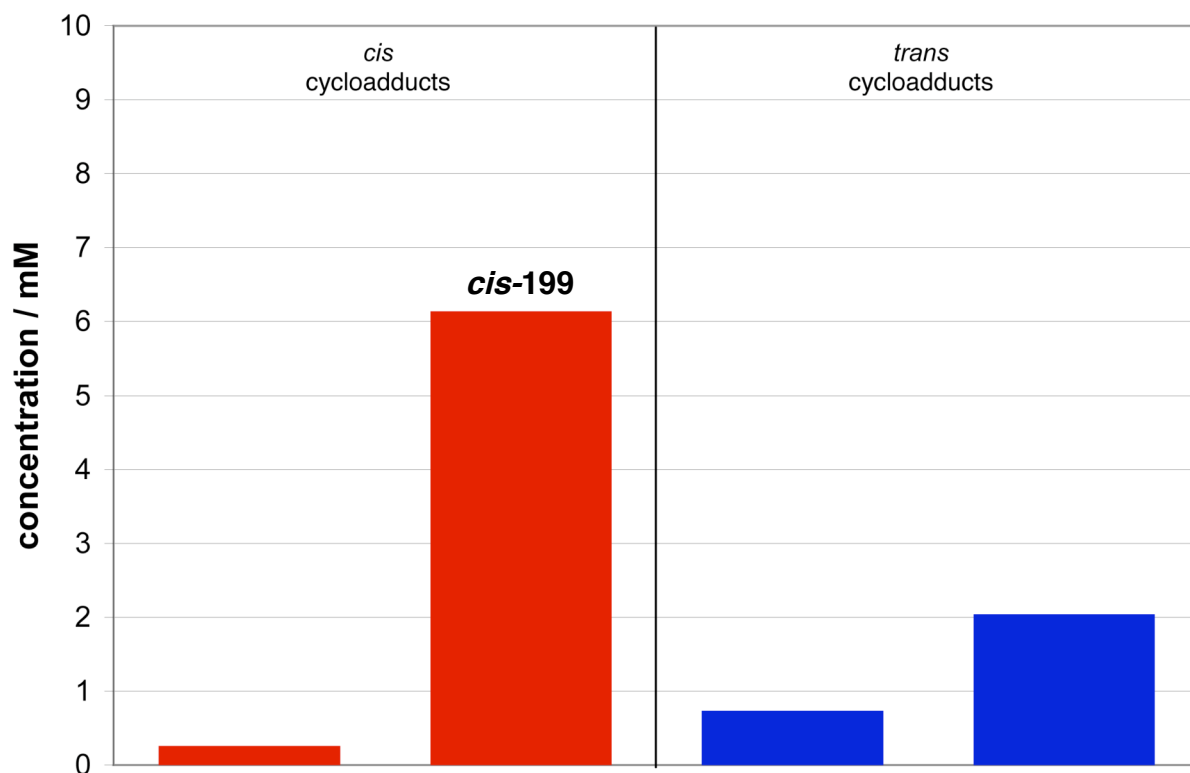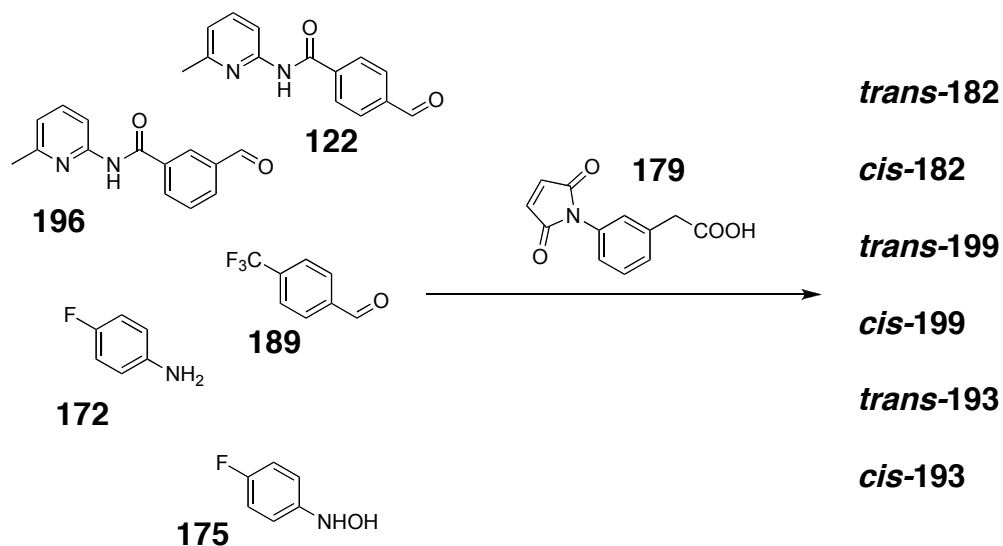
Figure 169. Concentrations of the *cis* cycloadducts (■) and *trans* cycloadducts (■) detected in the reaction mixture of a DCL created from compounds **112**, **196**, **189**, **172** and **175** treated with maleimide **179** ([**112**] = [**196**] = [**189**] = [**172**] = [**175**] = [**179**] = 20 mM).

Since both the *meta* nitrone **197** and the *para* nitrone **177** react with the *meta* maleimide **179** *via* the AB complex pathway a competition between the formation of two *cis* cycloadducts **cis-182** and **cis-199** is observed. The reaction of *para* nitrone **177** with the *meta* maleimide **179** is more efficient than the corresponding recognition-mediated reaction of the *meta* nitrone **197** with the same maleimide. This difference is observed in the final concentrations in the product pool. **Cis-182** reaches a concentration of 8.8 mM and the other enhanced cis product **cis-199** 2.6 mM. None of the other compounds in the product pool exceed the concentration of 1 mM. The properties of the two AB complex pathway reactions were transferred faithfully into a dynamic scenario and the more efficient one as predicted dominated the network.

When the same DCL is treated with the *para* maleimide **112** (Figure 169) a different scenario is expected to unfold. The *meta* nitrone **197** is unreactive towards the *para* maleimide **112** and the *para* nitrone **177** - *para* maleimide **112** reaction is the previously described efficient replicator. The recognition-mediated reaction forming **trans-188** should have no competition in the system (Figure 170).

Figure 169. Cycloadducts derived from a DCL created from compounds **122**, **196**, **189**, **172** and **175** when treated with maleimide **122**.
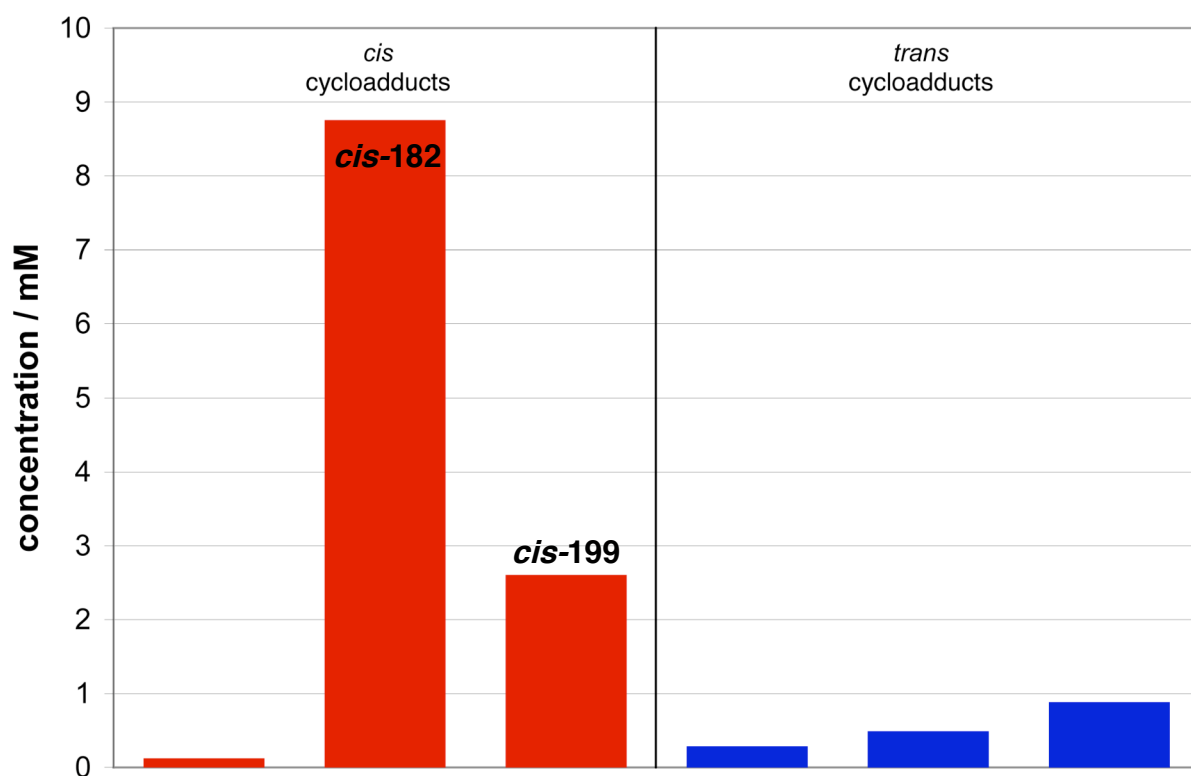

Figure 170. Concentrations of the *cis* cycloadducts (■) and *trans* cycloadducts (■) detected in the reaction mixture of a DCL created from compounds **112**, **196**, **189**, **172** and **175** treated with maleimide **122** ([**112**] = [**196**] = [**189**] = [**172**] = [**175**] = [**122**] = 20 mM).

There is no detectable selectivity between the *cis* cycloadducts, however, the replicating product ***trans*-188** clearly dominates the *trans* isomers reaching a concentration of 4.6 mM. The selectivity over the next most abundant *trans* compound is 4:1.

Three recognition-mediated enhancement pathways have been identified among the reactive combinations: two AB complex reactions forming ***cis*-182** and ***cis*-199** and

one replicator ***trans*-188**. It is possible to increase the complexity of the system by using both recognition aldehydes **122** and **196** and both maleimides **112** and **179** (Figure 171). In such scenario there will be six compounds in the exchange pool, but twelve compounds in the product pool including all three enhancement pathways (Figure 172).
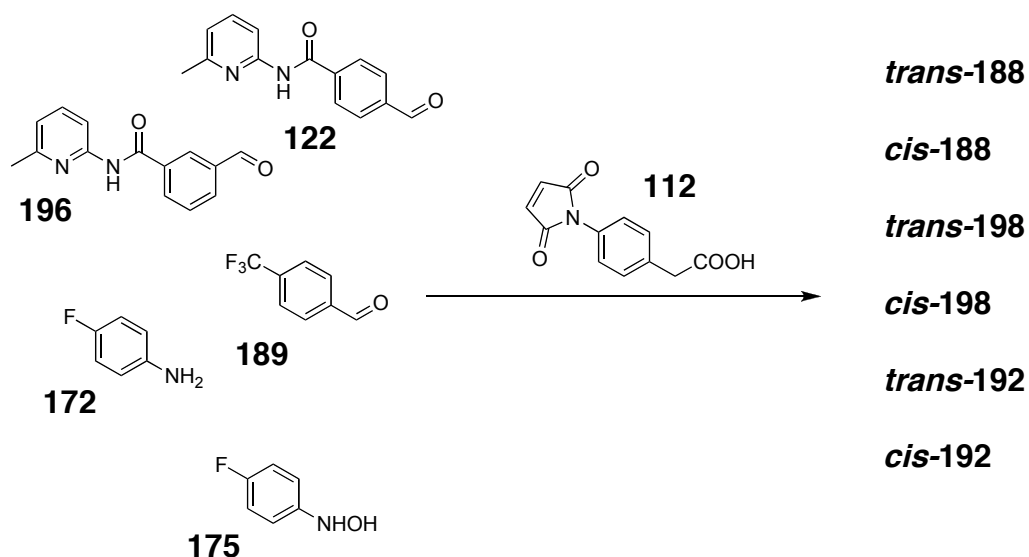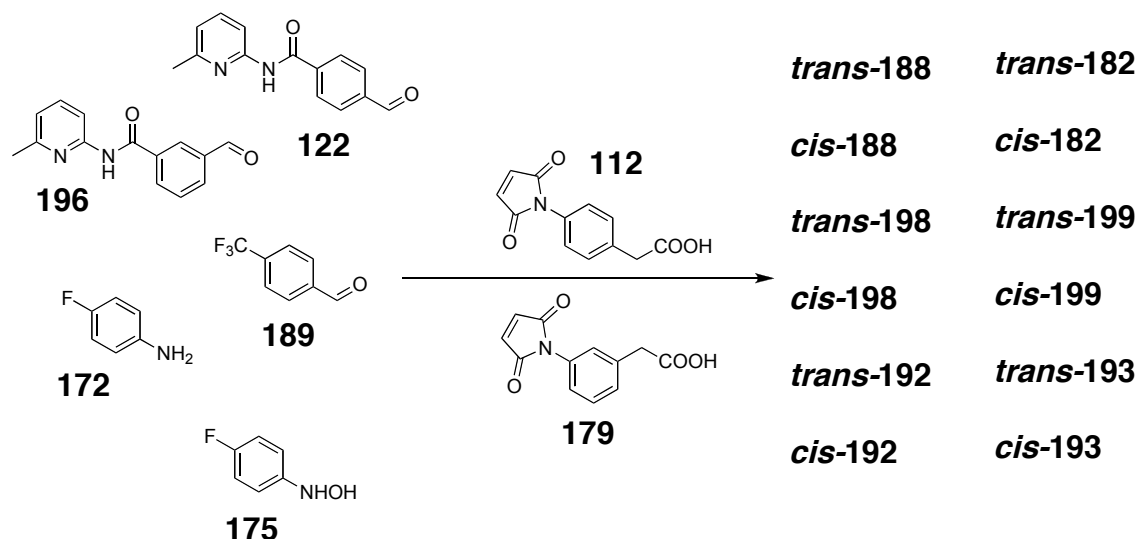

Figure 171. Cycloadducts derived from a DCL created from compounds **122**, **196**, **189**, **172** and **175** when treated with maleimides **122** and **179**.
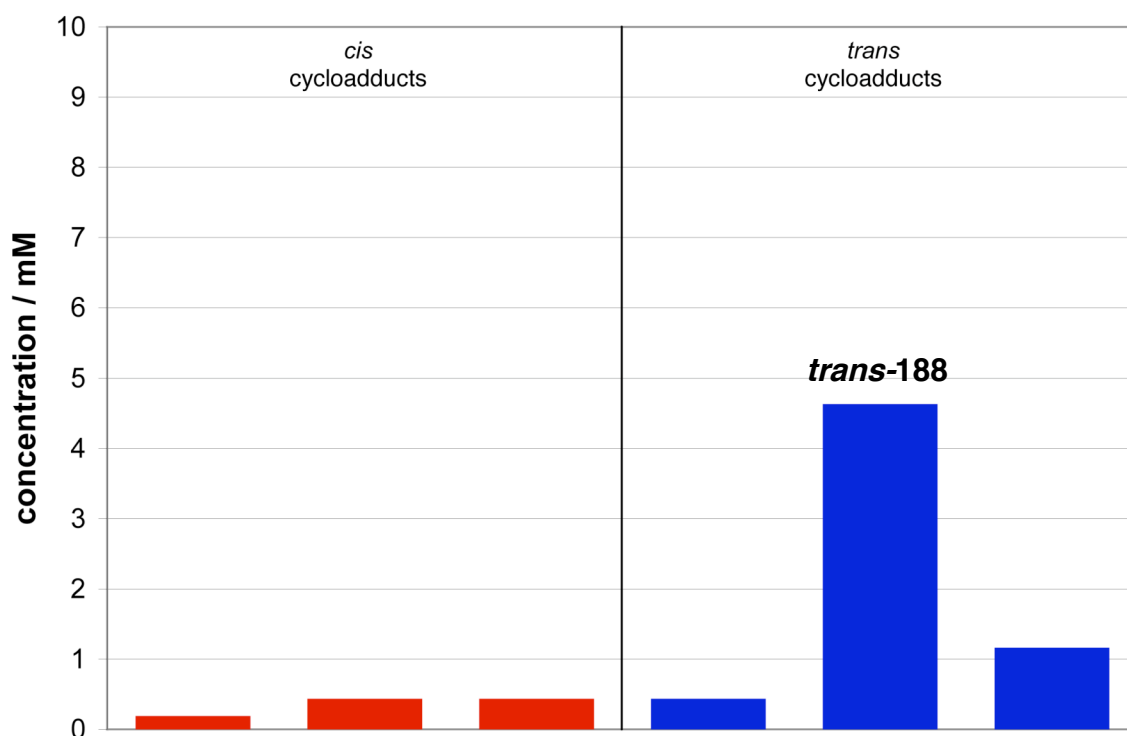

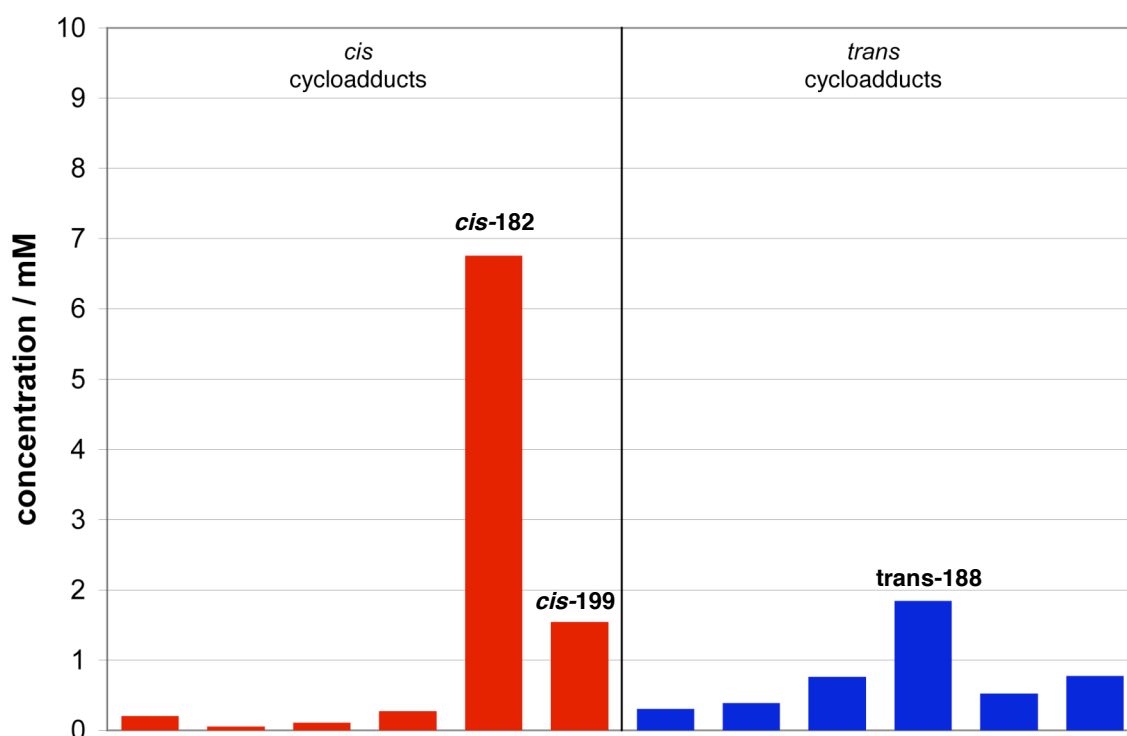**Figure 172.** Concentrations of the *cis* cycloadducts (■) and *trans* cycloadducts (■) detected in the reaction mixture of a DCL created from compounds **112**, **196**, **189**, **172** and **175** treated with maleimide **122** and **179** ([**112**] = [**196**] = [**189**] = [**172**] = [**175**] = [**122**] = [**179**] = 20 mM).

All enhancement pathways compete for the available hydroxylamine resource. Systems where the AB complex pathway competed with a replicator were already described previously and from those results it is expected for the most efficient AB complex product to be the one selected over the others. The most abundant product *cis*-**182** is derived from *para* nitrone **177** and *meta* maleimide **179** reaching 6.8 mM. There is also a slight increase of *cis*-**198** product that stands out over the other four remaining *cis* cycloadducts reaching the concentration of 1.5 mM, while none of the other *cis* products exceed 0.3 mM. Among the *trans* products the replicator is only slightly dominating since it loses the competition for the hydroxylamine resource with the AB complex pathway products. The selectivity of *trans*-**188** at a concentration of 1.8 mM is only 2.4:1 over the next most abundant *trans* cycloadduct.

These experiments shed light on how properties of individual reactions can be carried over into a more complex dynamic scenario. Simple geometric modifications of the building blocks led to enhancement of different products form the exchanging networks. The behaviour of the systems could be predicted knowing the reactivity of the components. The approach using simple building blocks to construct the DCLs is a very straightforward method of investigating how different kinetic processes perform in a dynamic scenario. The idea of using structurally similar building blocks in different combinations, with each combination giving different emergent properties is very interesting in scope of systems chemistry development. The possible combinations are countless and the methodology opens scope for further investigations.

5.6 Selecting a replicator from a 35-member DCL

The final goal of the project was to prove that the developed methodology is effective in a considerably larger DCL. The idea was to use 5 aldehydes, among which only one **122** bears a recognition site, 4 amines, and a hydroxylamine **175**. This set of building blocks will lead to the generation of 20 imines and 5 nitrones that upon addition of the maleimide can be metabolised into 10 cycloadducts. The total number of species in the mixture is 46 (35 in the exchanging pool) out of which only one *trans*-**188** is capable of replication (Figure 173). The starting concentrations of all

compounds were 20 mM and the reaction was run in CD$_2$Cl$_2$ saturated with PTSA at 273 K.
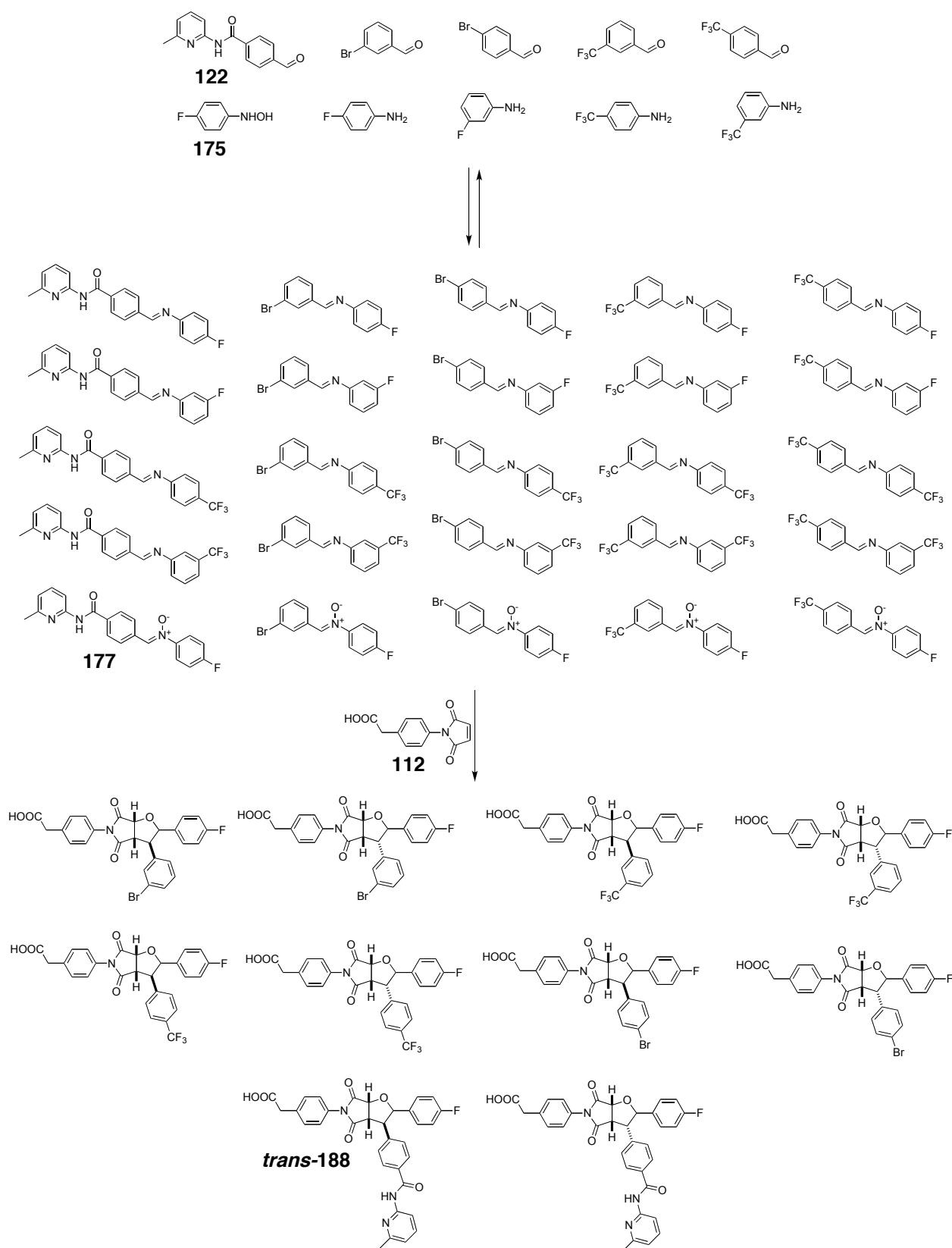
Figure 173. Cycloadducts derived from a DCL created from 5 aldehydes, 4 amines and hydroxylamine **175** treated with maleimide **112**.

The problem for the replicator **trans**-188 in such scenario is that the distribution of the hydroxylamine resource will be much wider than in the previous examples, as there will be five reactive nitrones in solution at one time. The reactions in the system will have to perform at lower concentrations. Such situation will on one hand slow down the bimolecular background reactions, but also effect the replicator. The resolution of the library may take longer, but high selectivity is expected (Figure 174).
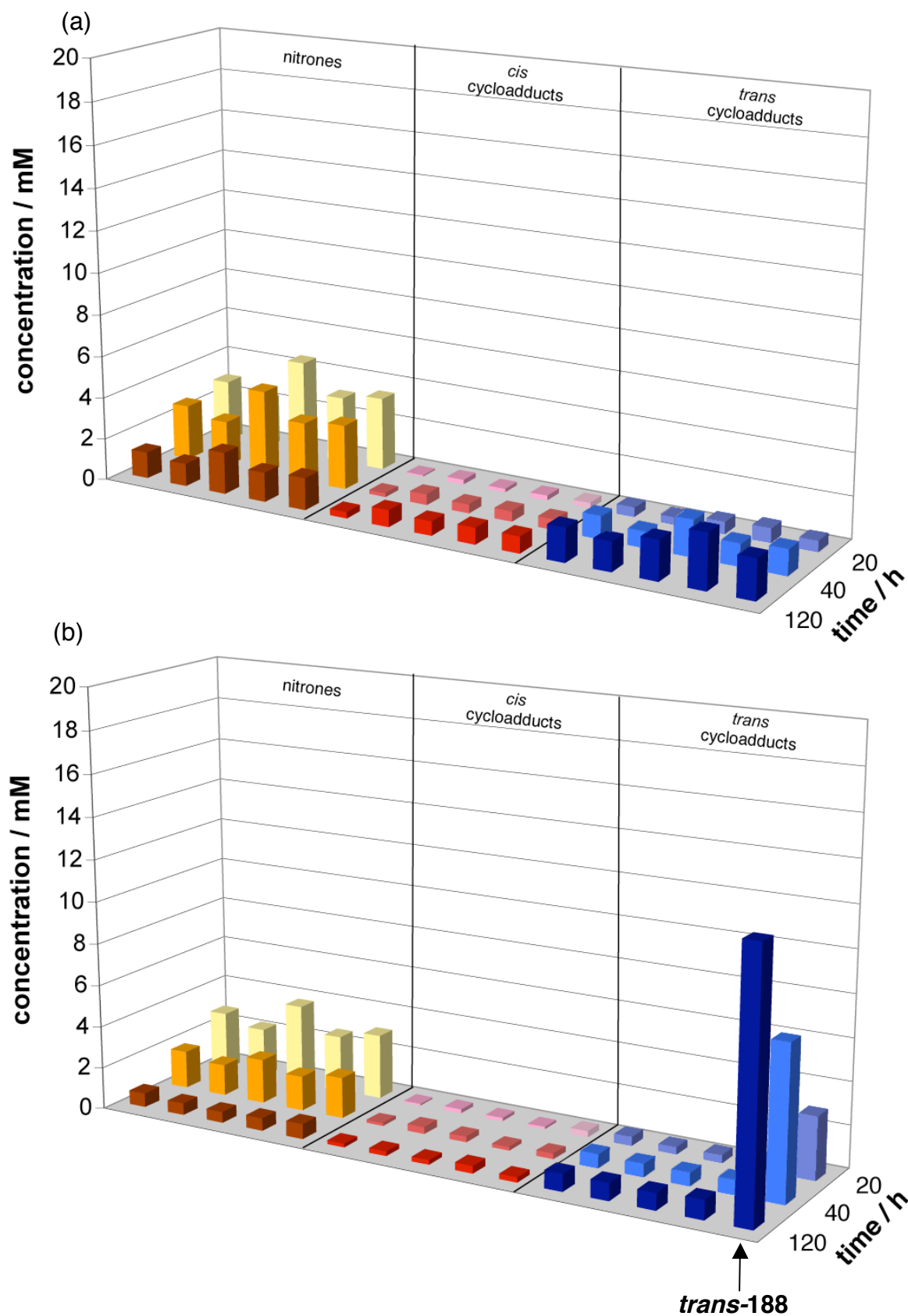
Figure 174. Concentrations of compounds containing the hydroxylamine unit: nitrones (■), *cis* cycloadducts (■) and *trans* cycloadducts (■) detected after 20, 40 and 120 h in the reaction mixture of a DCL created from 5 aldehydes, 4 amines and hydroxylamine **175** (Figure 173) treated with (a) ester maleimide and (b) acid maleimide **112**.

184

Since not all library components include a fluorine tag, for clarity, only the distribution of hydroxylamine-containing species is monitored: the nitrones and the cycloadducts. First, a control reaction was run using the ester maleimide **156** in order to observe how the DCL behaves when no recognition is present. The library was monitored after 16, 40 and 120 hours. After 16 h in the control reaction nitrones are the most abundant hydroxylamine-containing species with the concentration ranging from 2 to 4 mM. The highest concentration among the cycloadducts is only 0.7 mM. After 120 hours 56% of the nitrones is metabolised into the corresponding cycloadducts with the most abundant cycloadduct at 2.6 mM and the highest nitrone concentration at 2 mM. Overall there is no significant selectivity in the system, with the slight differences resulting from relative reactivities and stabilities of the differently substituted nitrones.

The situation changes when recognition is present within the system upon the addition of the acid maleimide **112**. Already, after 16 h there is selectivity in the products pool with the replicating product *trans*-**188** at 3 mM having a 7:1 advantage over the next *trans* cycloadduct. However the concentration of the nitrone species still ranges between 2.3 mM and 4.6 mM indicating there is more than 14 mM of unmetabolised hydroxylamine in the system. The exchange process keeps regenerating the recognition nitrone **177** so it is still possible for more of the replicating product to be formed. After 40 h the concentration of the replicator *trans*-**188** reaches 7.2 mM and at that point is the dominating hydroxylamine-containing species in the mixture. After 120 h the selected compound *trans*-**188** is at 12.3 mM with none of the other compounds utilising the hydroxylamine unit exceeding the concentration of 1 mM.

The experiment was repeated using the *meta* substituted maleimide **179** to metabolise the nitrones into the cycloadducts. As demonstrated previously the recognition nitrone **177** with maleimide **179** form a reactive AB complex. The expected compound to be amplified is *cis*-**182**. The ratio between the amplified product and other cycloadducts should be higher than in the previous example since the AB complex channel is more efficient at lower concentrations then the replicator. The system was left to evolve at 273 K for 120 h and the concentration of hydroxylamine-containing species was determined using ^{19}F NMR spectroscopy (Figure 175).
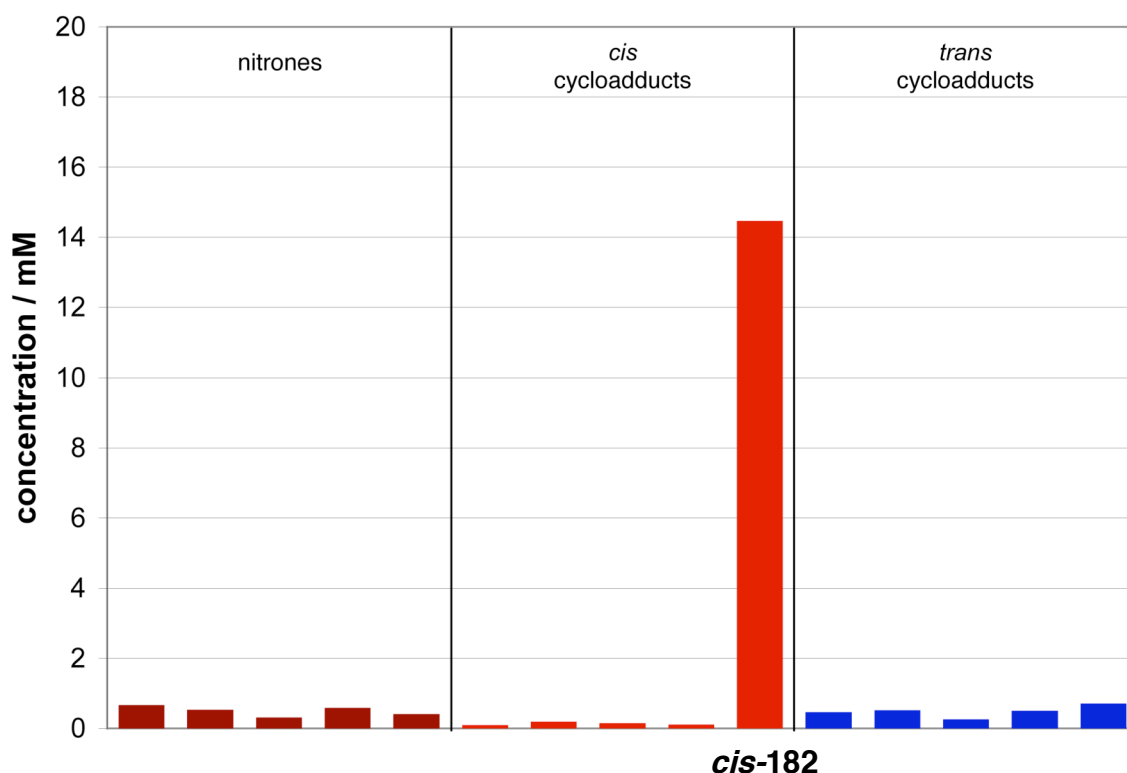
Figure 175. Concentrations of compounds containing the hydroxylamine unit: nitrones (▉), *cis* cycloadducts (▉) and *trans* cycloadducts (▉) detected after 120 h in the reaction mixture of a DCL created from 5 aldehydes, 4 amines and hydroxylamine **175** treated with acid maleimide **179**.

The product of the AB complex pathway **cis-182** clearly dominates over all other monitored compounds. The ratio between **cis-182** and the next most abundant cycloadduct is more then 20:1. The amplified product **cis-182** reaches a concentration of 14.5 mM with none of the other compounds utilising the hydroxylamine exceeding 1 mM.

Monitoring the systems over a longer period of time revealed the fate of the hydroxylamine resource. Not only selectivity was achieved in the product pool, but also among all hydroxylamine-containing species. The recognition-mediated processes were able to dominate a complex dynamic mixture performing at low concentrations All bimolecular pathways were greatly suppressed due to low concentration of substrates and most show minimal growth of product over the period of the experiment. The recognition nitrone **177** was siphoned from the dynamic pool *via* the irreversible cycloaddition reaction driving the exchange processes towards its own formation. The regeneration of the recognition nitrone allowed for a great majority of the hydroxylamine resource to be channelled through a single reaction pathway. This experiment definitely proves how a single recognition event can

influence a selection of a species from a complex mixture and that the developed methodology can be used in bigger libraries opening scope for further research in this field.

6. Future work and general conclusions

6.1 Solubility issues

There are several approaches for future work on the systems described in the previous Chapters. A first approach would modify the existing systems in order to overcome the problems with their analysis in order to study their properties in more detail. The systems described in Chapters 2 and 3 were analysed using ^{1}H NMR spectroscopy. The problems associated with this technique included overlapping of signals and difficulty with interpreting the spectra at low concentrations. The ^{19}F NMR spectroscopy used in Chapters 4 and 5 provided much improved signal dispersion and therefore precision. In more complex systems developed in the future, fluorine tagging of the components should allow for more precise analysis.

One of the major problems of the system described in Section 3.2 that prohibits precise analysis of all possible reactive combinations is poor solubility of the template consisting of the short maleimide **112** and the long nitrone **132**. Solubility has often been an issue in the study of replicating systems in our laboratory. Recently, some progress has been made into synthesis of soluble replicator building blocks. The major strategy consists of adding long hydrocarbon chain substituents to the known replicators that showed problems with solubility (Figure 176). Compound **200** could be an alternative to maleimide **112** allowing for the corresponding cycloadducts to be more soluble. If such approach would be applied to the system described in Section 3.2 it could be possible to perform the experiments at higher concentrations and with equal amounts of input template. This scenario would allow gaining better insight into the workings of the network and getting a full spectrum of its input/output properties.
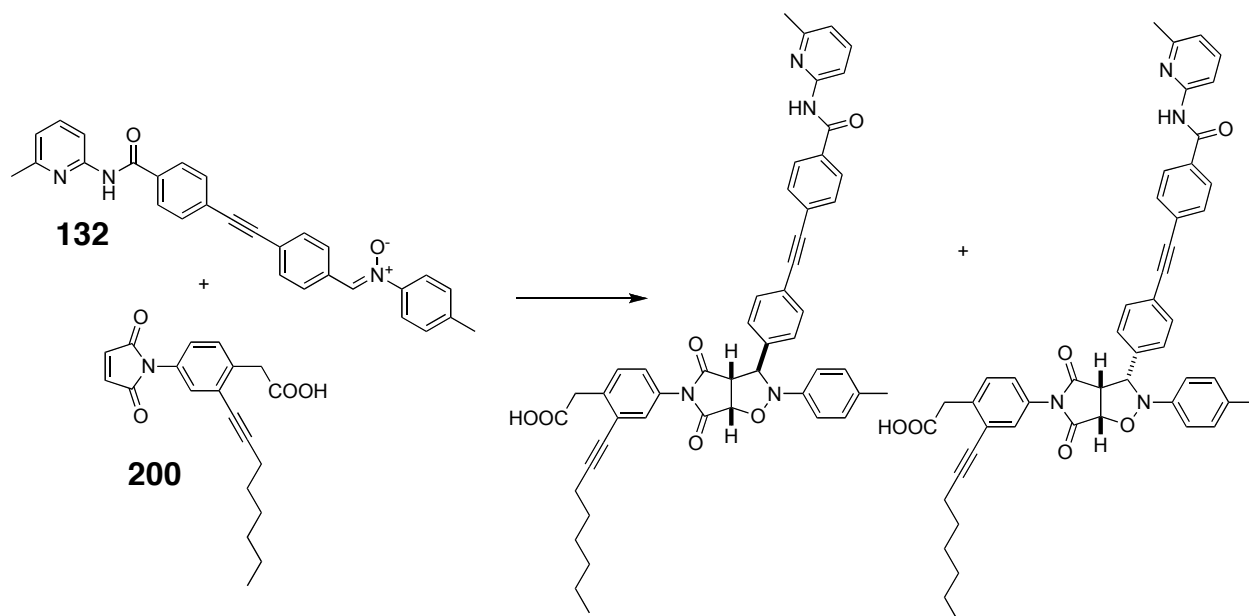
Figure 176. Maleimide **200** is functionalised with a long hydrocarbon chain in order to improve the solubility of the products of the reaction with nitrone **132**.

6.2 Expanding the logic

Logic gates are only of limited interest if it is not possible to connect them into circuits. One of the outstanding properties of replication-based logic gates described in Chapter 3 is that the nature of the input and the output are the same. It could be envisaged that the mixture, which is formed by the chemical logic gate, is used as an input for the next reaction. The input mixture would consist of the four templates of different concentrations. If such mixture was added to a fresh solution of starting materials the replicators that are at higher concentrations should template their own formation more efficiently than the ones at lower concentrations. In other words the bias in the input mixture towards certain replicators should be transferred and enhanced in the next logic operation experiment. If this experiment were repeated a several times the difference between the enhanced and the suppressed products would grow with every iteration.

The static systems described in Chapter 3 could be equipped with a dynamic element as the systems in Chapter 4. If the building blocks, apart from being length-segregated, could also exchange freely the recognition and the reactive sites, a new dimension would be introduced into the chemical logic gate. In such scenario all four templates would not only have to compete for the building blocks of appropriate length, but also for the available hydroxylamine resource in the system. The dynamic libraries described in this thesis consist of only one replicator competing with

bimolecular reactions in the dynamic mixture. Length segregation would create a scenario, where from a single DCL, the addition of one template would enhance one replicator and the addition of a different template a different replicator (Figure 177). The addition of a template would determine which replicator incorporates most of the available hydroxylamine resource and is selected from the mixture over other products. Such system could also be described by Boolean logic.
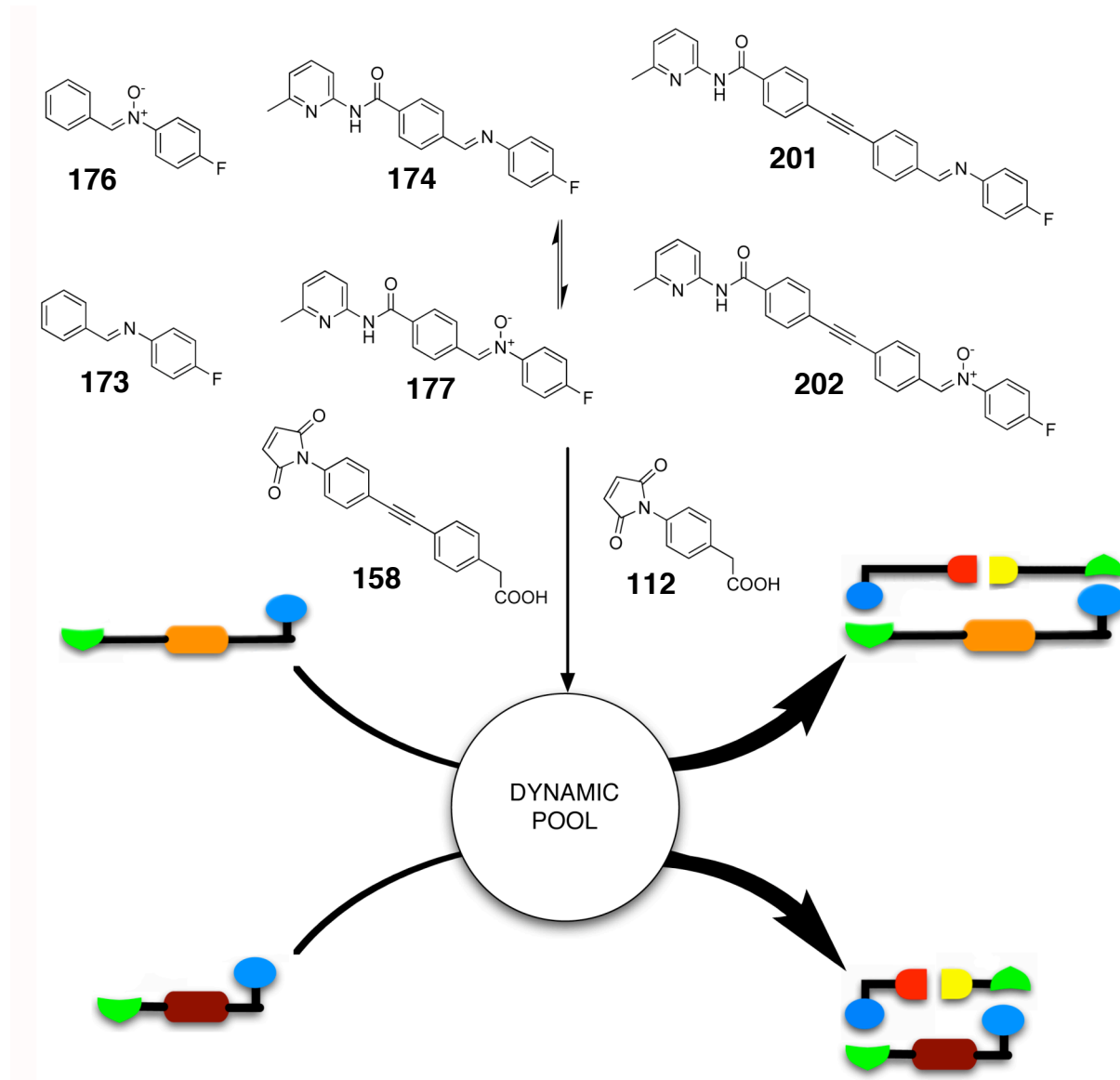
Figure 177. Design of a length segregated logic system based on a dynamic combinatorial library. Addition of a template of a specific length will lead to the reconfiguration of the exchanging pool in order to amplify that template.

6.3 Selective enhancement from a DCL

Another approach to developing a system that would respond to the addition of a specific template is using the geometry of the molecules as the differentiating factor. The reactivity of nitrones **177** and **197** was already discussed in Chapters 4 and 5,

but from previous studies it is known that these compounds together with maleimide **78** (Figure 178) form replicators of similar efficiencies. If the two nitrones were to be formed in a dynamic pool together with maleimide **78**, a possibility opens to template replicator **204** or **205** selectively. Nitrones **177** and **197** would be generated *via* the exchange reaction from the corresponding imines **174** and **203**. The addition of template **204** to the dynamic library will result in promoting that replicator and the hydroxylamine resource will be channelled to form this particular product. The addition of template **205** will change the distribution of library components in order to promote the formation of **205**.
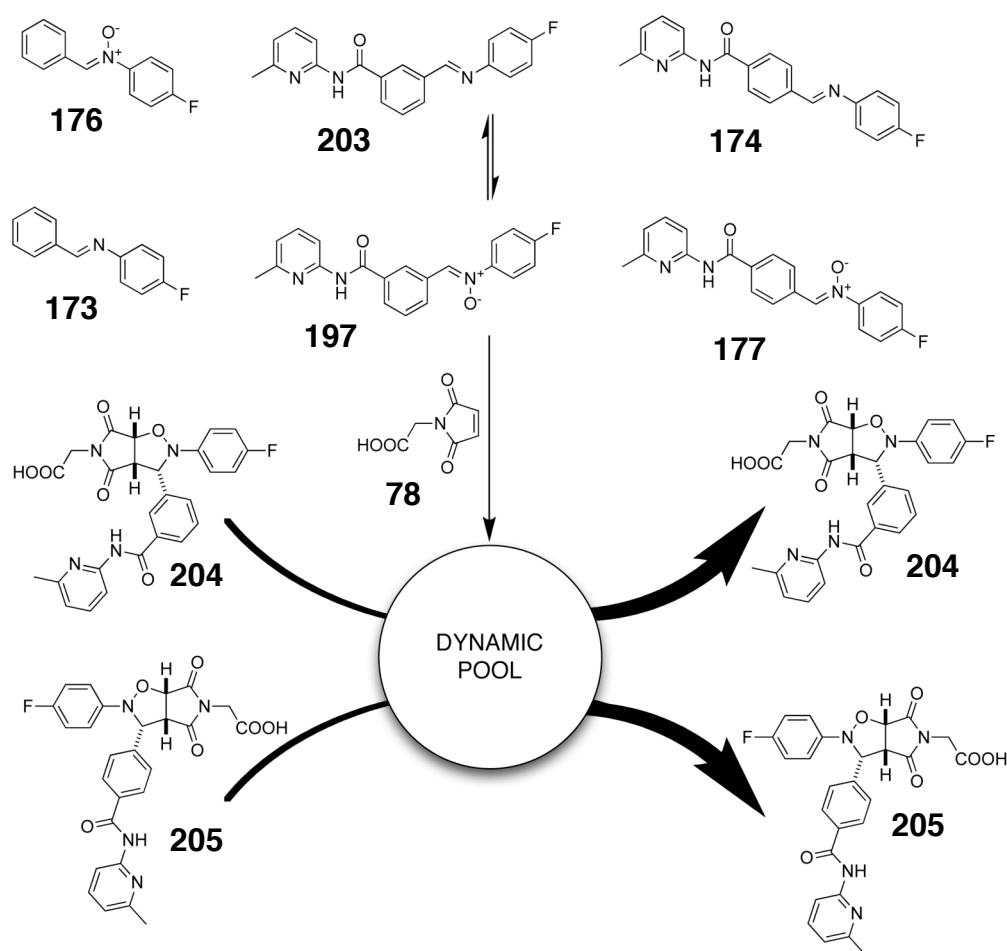
Figure 178. Design of a dynamic combinatorial library that generates two recognition-bearing nitrones with different geometries: **177** and **197**. Both nitrones react through the self-replicating pathway with maleimide **78**. Addition of a specific template **204** or **205** should enhance the templated replicator over the other.

The challenge is to design a dynamic mixture consisting of simple building blocks that when seeded with a proper template input would dynamically rearrange to selectively produce that compound. The library would interpret an input as an instruction describing the kind of species to produce. The mixture could be a

resource of undefined building blocks for a predesigned assembly. The addition of an instructional template would trigger the mixture to produce a specific component of the assembly. If a different building block would be needed at any time, another instructional template would trigger the mixture to start producing the new component. Using the methodology described in the previous chapters it is possible to envisage such system bringing us closer to developing complex assemblies from a dynamic pool of undefined components in a controlled manner.

6.4 General conclusions

The research described in this thesis reflects the ideas of systems chemistry. The focus is set on the study of how reactions occur in different scenarios and not necessarily, on the properties of the products, unless the product influences the reaction. The target is not synthesis and isolation of a single, structurally challenging compound, but rather a creation and exploitation of mixtures of increasing complexity. The overall goal is to keep the system building blocks as simple as possible while, on the other hand, expanding the networks created from these building blocks. Thanks to such approach the systems have a considerable degree of complexity in their emergent properties, but it is still possible to monitor practically every aspect of the mixture with available analytical techniques.

The properties of all systems are engendered in the recognition properties of the building blocks. The minimal model of self-replication based on two building blocks is on its own a complex, non-linear system. It is possible to incorporate self-replicating methodology into designing more complex networks of unique emergent properties. Increasing the number of building blocks from two to even as little as four, opens the possibility of many more recognition-mediated pathways to operate simultaneously in one network. Such scenario is the case in the systems where the recognition properties are governed by the length of the molecules. The uniquely designed systems show a clear response to a simple chemical input (template), which in fact is a very small alteration of the starting conditions. The output of the system can be interpreted logically and it is possible to depict clearly the network connecting the system components. Another specific feature of the systems described in this thesis is the ability to address selectively a certain pathway in a mixture of compounds which all bear the same reactive site and the same recognition site.

With the introduction of a dynamic component to the described systems, a new degree of complexity is reached, but structurally, the building blocks are still simple to synthesise and analyse. With the growing complexity of the system fluorine tagging was introduced making it possible for very precise and clear analysis using ^{19}F NMR spectroscopy. In the dynamic combinatorial systems the ability to selectively address the system comes with the passing of the recognition and the reactive site between the DCL members. The mode of selection from the described dynamic libraries is a unique coupling of thermodynamics and kinetics. The library is, by its nature, a thermodynamic system while the replicator introduces an irreversible, kinetic factor. This combination allows for very high selectivities and a large degree of control over the fate of the library. The next step was the realisation that such complex systems can be prepared from basic organic compound practically eliminating any synthesis of starting materials. This approach, together with fluorine tagging, allows preparing an array of many combinations and easy screening of the mixtures. It is worth noting that the more complex the described networks become the simpler the starting building blocks are. This approach proves that complex system research is possible using simple synthetic compounds and many theories arising from theoretical insight could now be proved in a laboratory. Supramolecular chemistry is often referred to as 'chemistry beyond the molecule'. It is possible to term Systems chemistry as the chemistry beyond the reaction, as the focus should be reset from perfecting a single chemical reaction to give a better yield and selectivity. With today's analytical techniques it is possible to gain insight into entire networks of reactions. As described in this work, complex system research can be an easily accessible area of chemistry with a large scope for future development.

7. Experimental

7.1 General procedures

Chemicals and solvents were purchased from Sigma-Aldrich, Alfa Aesar, Fisher Scientific or VWR and were used as received unless otherwise stated. CH_2Cl_2, MeCN and hexane were distilled over calcium hydride under a nitrogen atmosphere. Et_2O and tetrahydrofuran were dried by heating under reflux with sodium-benzophenone ketyl under a nitrogen atmosphere and collected by distillation. Thin-layer chromatography (TLC) was performed on aluminum plates coated with Merck Kieselgel 60 F_{254}. Developed plates were dried and analysed under a UV lamp (366 nm), where necessary stained with iodine, $KMnO_4$ and ninhydrin solution to aid visualisation and identification. Melting points were recorded using an Electrothermal 9200 melting point apparatus and are uncorrected. Column chromatography was performed on ICN EcoChrom Silica 32-63 μm or Apollo Scientific Silica Gel 40-63 μm. Microanalysis (CHN) was carried out at the University of St. Andrews. Chemicals and samples used prior to kinetic experiments were weighed using a Sartorius balance (MC1 Analytic AC 120S) with an accuracy of ± 0.1 mg.

7.2 General NMR spectroscopy procedures

^{1}H Nuclear magnetic resonance (NMR) spectra were recorded on a Bruker Avance 300 (300.1 MHz), Bruker Avance II 400 (400.1 MHz) and a Varian UNITYplus (500.1 MHz) spectrometer using the deuterated solvent as the lock and the residual solvent as the internal reference in all cases. ^{13}C NMR spectra were acquired using the PENDANT or DEPTQ sequences on Bruker Avance 300 (75.5 MHz) or a Bruker Avance II 400 (100.6 MHz) spectrometer. ^{19}F NMR spectra were recorded using a Bruker Avance II 400 (376.5 MHz) and Bruker Avance 500 (470.4 MHz) spectrometer and were referenced to CCl_3F = 0 ppm. All ^{1}H and ^{19}F coupling constants are quoted in Hz and all spectra were recorded at 298 K, unless otherwise stated. The chemical shift information for each resonance signal is given in parts per million (ppm). The symbols s, d, t, m are used in the assignment of ^{1}H, ^{13}C and ^{19}F spectra donate singlet, doublet, triplet and multiplet respectively. All ^{1}H, ^{13}C and ^{19}F spectra were

analyzed using Bruker Topspin (Version 2.0 pl 3, Bruker Biospin, Germany, 2006) or iNMR (Version 2.3.6, Giuseppe Balacco, 2007) software.

7.3 ^{1}H NMR spectroscopy kinetic measurements and deconvolution of data

500 MHz ^{1}H NMR spectroscopy was used for kinetic analysis of the systems described in Chapters 2 and 3. Masses of reagents were measured using a Sartorius BP 211D balance (± 0.01 mg). Stock solutions of the appropriate starting materials were made using a fresh batch of CDCl$_3$ and equilibrated in a thermostatically-controlled water bath at 273 K. In a typical experiment, an NMR sample was prepared in a 5 mm NMR tube (Wilmad 528PP) by mixing appropriate volumes of stock solutions such that the total volume was 0.7 ml and the concentration of each component was 10 mM. A polyethylene pressure cap was then applied to the top of the tube to prevent solvent evaporation. The NMR tube was transferred to an NMR spectrometer (Varian UNITYplus) regulated at 273 K, and 500.1 MHz ^{1}H NMR spectra were acquired automatically every 1800 s over a period of 16 hours. Analysis and deconvolution of each of the ^{1}H NMR spectra recorded during this time was performed using Bruker Topspin software (Version 2.0 pl 3, Bruker Biospin, Germany, 2006).

The chemical reaction monitored in the described systems was a 1,3-dipolar cycloaddition between a maleimide and a nitrone, giving rise to two cycloadduct products. The percentage completion of the reaction was monitored through the disappearance of the maleimide protons signal and the appearance of the cycloadduct proton signals (Figure 179).
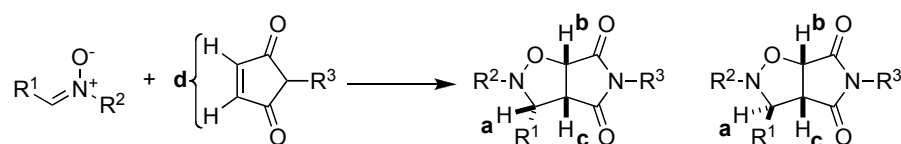
Figure 179. Scheme of the studied 1,3-dipolar cycloaddition reaction between a nitrone and a maleimide. Protons which signals were used to follow the reaction by ^{1}H NMR spectroscopy are highlighted: **a**, **b**, **c** for the cycloadducts and **d** for the maleimide.

Two diastereoisomeric products of the reaction between a nitrone and a maleimide are easily distinguishable in the ^{1}H NMR spectrum. The signals of the isoxazolidine protons give a different splitting pattern for the two cycloadducts (Figure 180). Using

molecular modelling and the appropriate crystal structures it is possible to measure the dihedral angles of the protons in the fused ring system. In the *trans* isomer protons (a) and (b) form a dihedral angle of *ca.* 90°. The Karplus equation[250] predicts no coupling between these protons, thus proton (a) appears as a singlet in the ^{1}H NMR spectrum. Protons (b) and (c) form a dihedral angle of 20° and both display a coupling constant of 8 Hz, which is also in accordance with the Karplus equation. In the *cis* isomer all protons in the isoxazolidine ring form dihedral angles of around 20° and as expected the coupling constant is 8 Hz for the observed dublets.

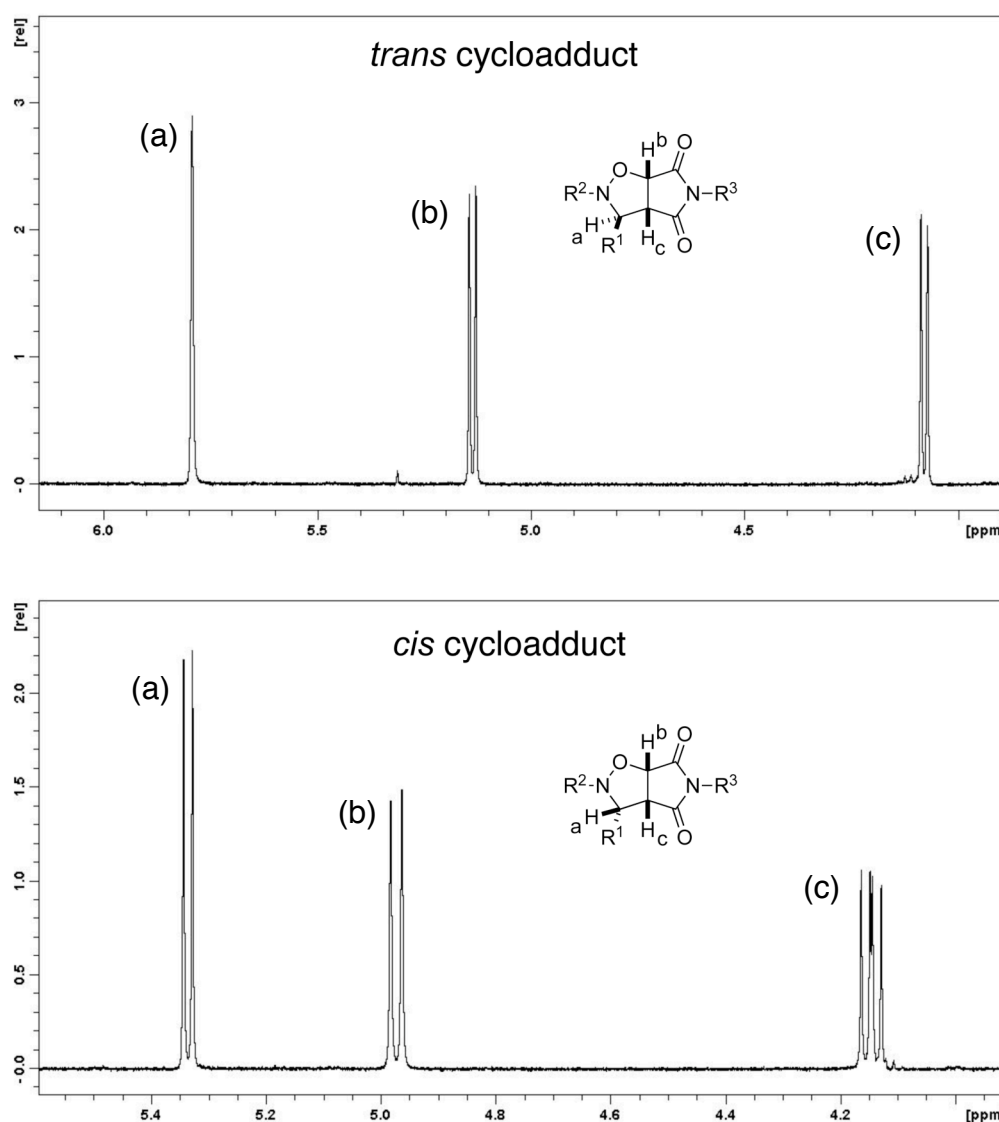**Figure 180.** Partial 500 MHz ^{1}H NMR spectra featuring characteristic splitting patterns for the proton signals arising of the isoxazolidine ring for the *trans* and *cis* cycloadducts. For the *trans* cycloadduct the signal for proton (a) is singlet and for (b) and (c) a doublet. For the *cis* isomer the signals for protons (a) and (b) are doublets and for proton (c) a doublet of doublets.

196

In order to extract product concentrations from the ^{1}H NMR spectrum the area of the maleimide protons signal and one of the signals of the corresponding cycloadduct proton signals was calculated through deconvolution of the spectrum. The area of the maleimide protons signal was divided by two and added to the area of the cycloadduct proton signal. The obtained number signifies the area of a proton signal of a species at a concentration of 10 mM and was used as a reference to calculate the concentration of all species detectable in the spectrum. This procedure was repeated for all spectra recorded at different time points. The time at which each spectrum was recorded was extracted from the log file, which accompanies the NMR spectroscopy data. Knowing the concentration of the products at each time point we can plot a reaction profile for each reaction using graphing software KaleidaGraph (Synergy Software, Reading, PA, USA, version 4.1, 2008).

7.4 Examples of ^{1}H NMR spectra in system analysis

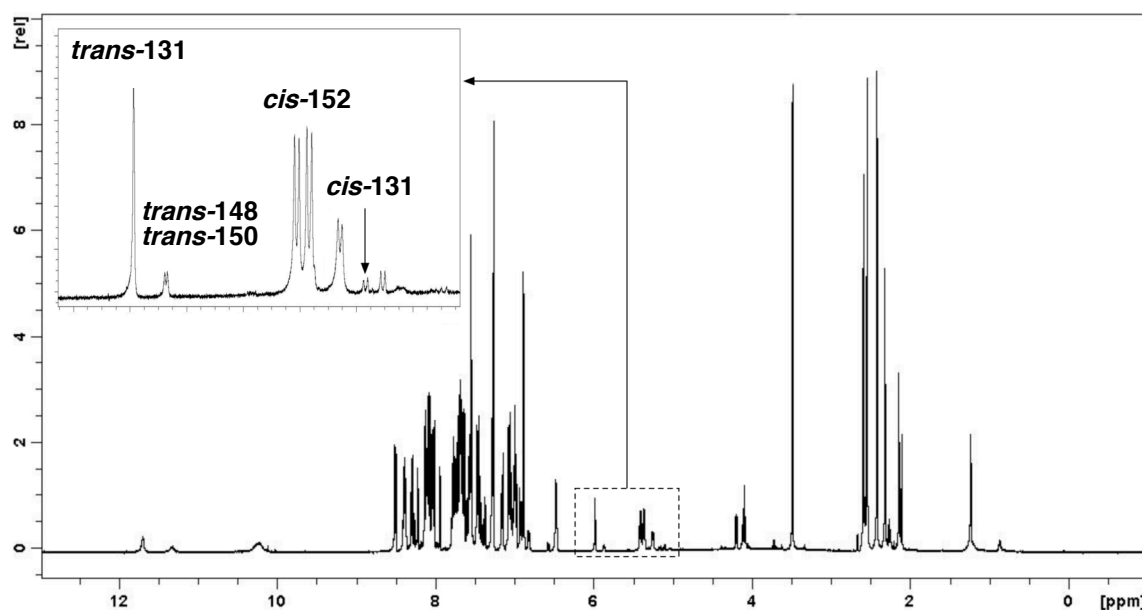**Figure 181.** 500.1 MHz ^{1}H NMR spectrum recorded after 8 h a reaction between maleimides **128** and **134** and nitrones **124** and **132** ([**128**] = [**134**] = [**124**] = [**132**] = 10 mM, [***trans*-131**] = 1 mM) used to derive the concentration *vs.* time plot in Figure 93. The inset shows the signals used to identify the products of the reaction.

Figure 182. 500.1 MHz ^{1}H NMR spectrum recorded after 16 h of a reaction between maleimides **112** and **158** and nitrones **124** and **132** ([**112**] = [**158**] = [**124**] = [**132**] =10 mM) used to derive the concentration *vs.* time plot in Figure 112. The inset shows the signals used to identify the products of the reaction.

Figure 183. 500.1 MHz ^{1}H NMR spectrum recorded after 16 h of a reaction between maleimides **112** and **158** and nitrones **124** and **132** doped with 30 mol% of preformed ***trans*-156** ([**112**] = [**158**] = [**124**] = [**132**] =10 mM, [***trans*-156**] = 3 mM) used to derive the concentration *vs.* time plot in Figure 113. The inset shows the signals used to identify the products of the reaction.
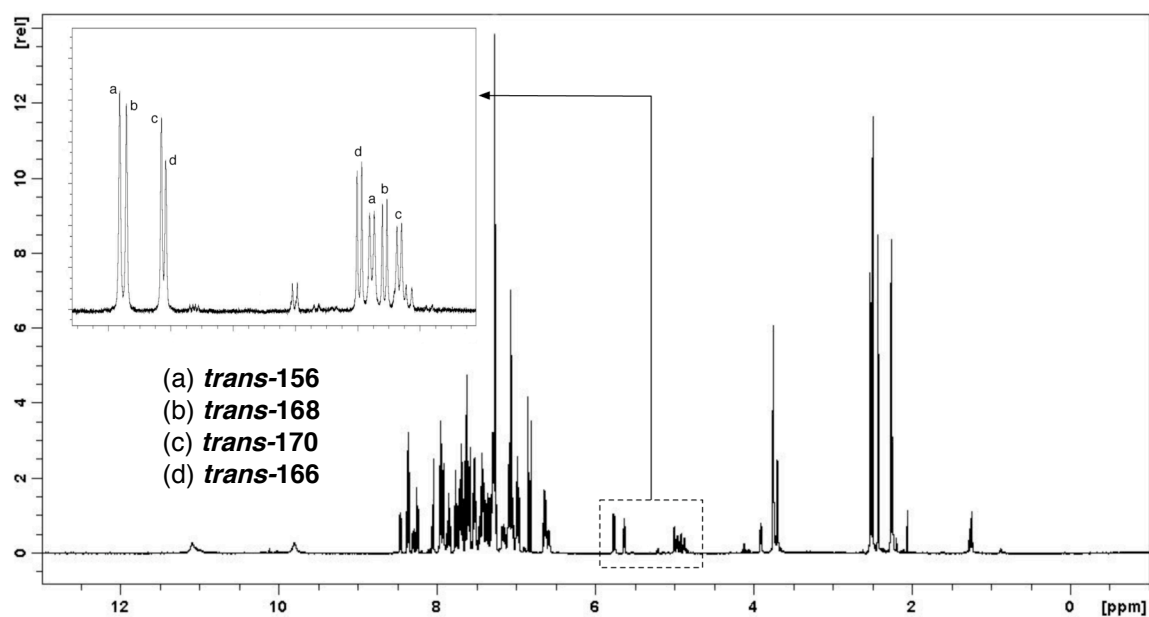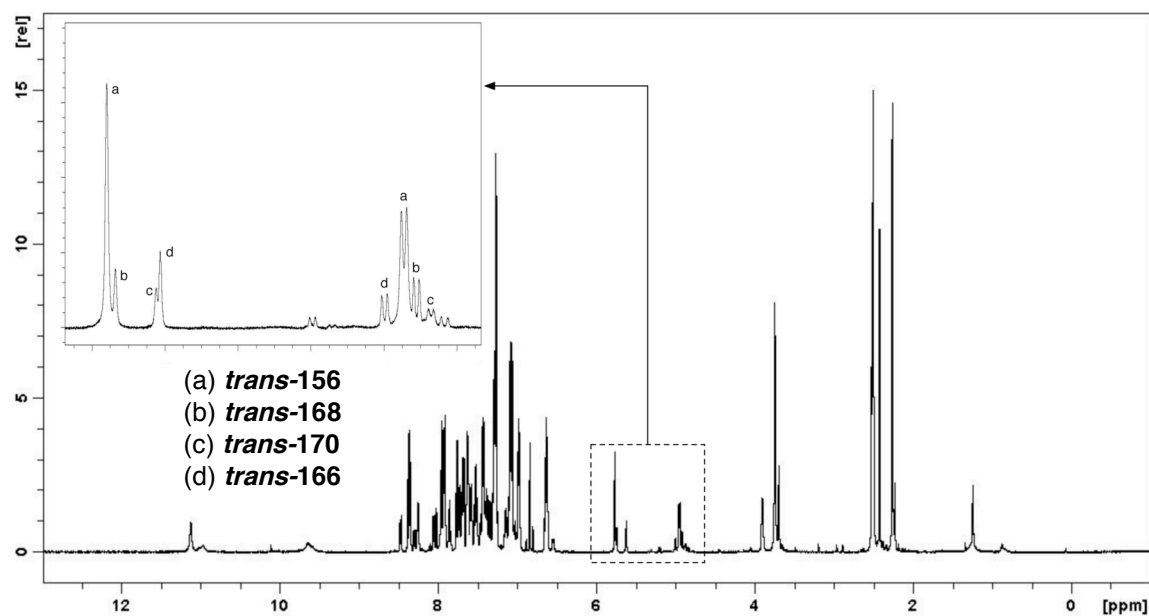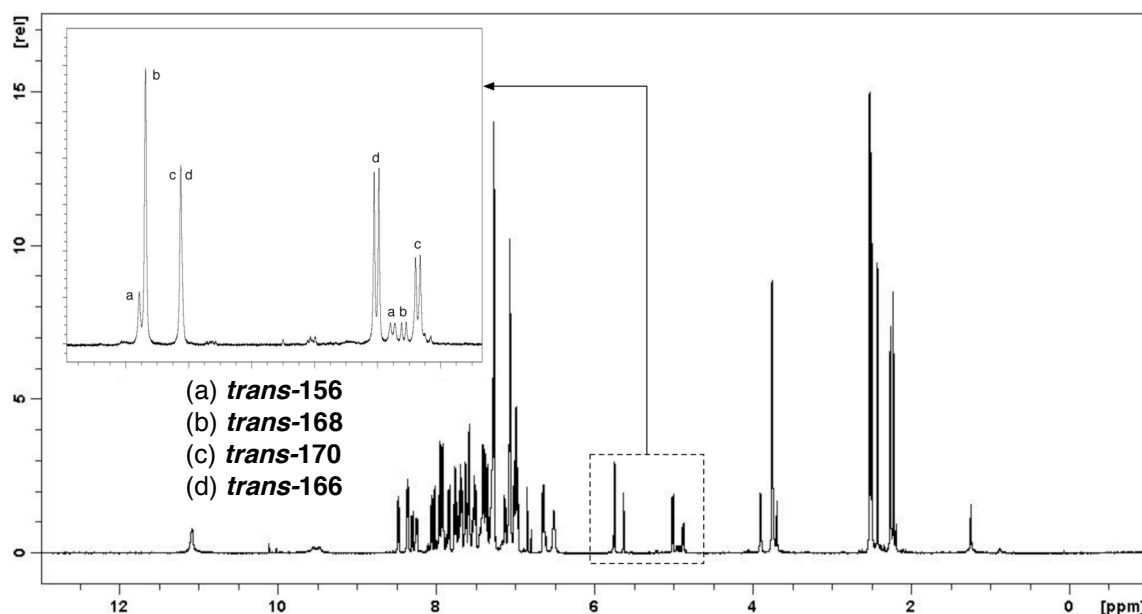
Figure 184. 500.1 MHz ^{1}H NMR spectrum recorded after 16 h of a reaction between maleimides **112** and **158** and nitrones **124** and **132** doped with 30 mol% of preformed *trans*-**168** ([**112**] = [**158**] = [**124**] = [**132**] =10 mM, [*trans*-**168**] = 3 mM) used to derive the concentration *vs.* time plot in Figure 117. The inset shows the signals used to identify the products of the reaction.

7.5 ^{19}F NMR spectroscopy kinetic measurements and deconvolution of data

470.4 MHz ^{19}F NMR spectroscopy was used for kinetic analysis of the systems described in Chapters 4 and 5. In order to prepare the solvent for the imine-nitrone exchange reactions CD$_2$Cl$_2$ (2 ml) was stirred rapidly with 100 mg of PTSA monohydrate in a volumetric flask at room temperature for 20 min. The suspension was filtered and the solvent used immediately. Masses of reagents were measured using a Sartorius BP 211D balance (± 0.01 mg). Stock solutions of the appropriate starting materials were made using freshly prepared solvent and equilibrated in a thermostatically-controlled water bath at 273 K. In a typical experiment, an NMR sample was prepared in a 5 mm NMR tube (Wilmad 528PP) by mixing appropriate volumes of stock solutions such that the total volume was 0.7 ml and the concentration of each component was 20 mM. A polyethylene pressure cap was then applied to the top of the tube to prevent solvent evaporation. The NMR tube was transferred to an NMR spectrometer Bruker Avance 500 regulated at 273 K, and 470.4 MHz ^{19}F{^{1}H} NMR were acquired automatically every 1800 s over a period of 16 hours. Analysis and deconvolution of each of the ^{19}F NMR spectra recorded during this time was performed using Bruker Topspin software (Version 2.0 pl 3, Bruker Biospin, Germany, 2006). The percentage completion of the reaction was

monitored through the disappearance of the nitrone fluorine signals and the appearance of the cycloadduct fluorine signals. In order to extract product concentrations from the ^{19}F NMR spectrum the areas of all fluorine signals was calculated through deconvolution of the spectrum. The area of the nitrone fluorine signals and the area of all cycloadduct fluorine signals were summed. The obtained number signifies the area of a proton signal of a species at a concentration of 20 mM and was used as a reference to calculate the concentration of other species detectable in the spectrum. This procedure was repeated for all spectra recorded at different time points. The time at which each spectrum was recorded was extracted from the log file, which accompanies the NMR spectroscopy data. Knowing the concentration of the products at each time point we can plot a reaction profile for each reaction using graphing software KaleidaGraph (Synergy Software, Reading, PA, USA, version 4.1, 2008).

7.6 Examples of ^{19}F NMR spectra in system analysis
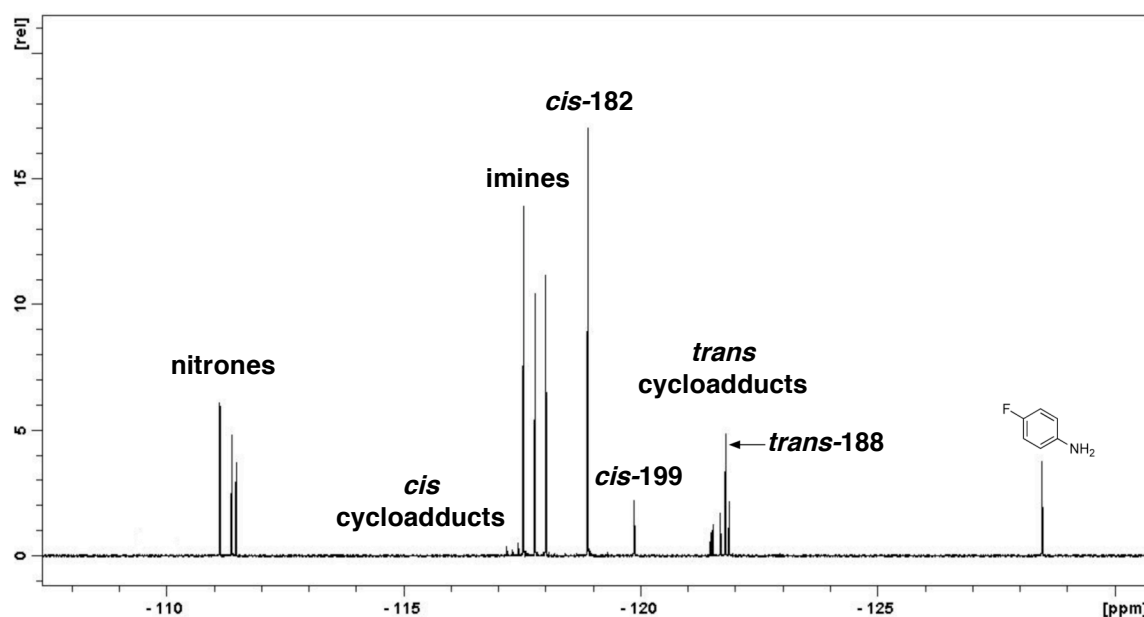
Figure 185. Partial 470.4 MHz ^{19}F NMR spectrum recorded after 16 h of a reaction mixture of a DCL created from compounds **112**, **196**, **189**, **172** and **175** treated with maleimide **122** and **179** ([**112**] = [**196**] = [**189**] = [**172**] = [**175**] = [**122**] = [**179**] = 20 mM) used to derive the plot in Figure 172.

Figure 186. Partial 470.4 MHz ^{19}F NMR spectrum recorded after 16 h of a reaction mixture of a DCL created from 5 aldehydes, 4 amines and hydroxylamine **175** (Figure 173) treated with acid maleimide **112** used to derive the plot in Figure 174b. The inset shows the signals of the five *trans* cycloadduct products with ***trans*-188** dominating.
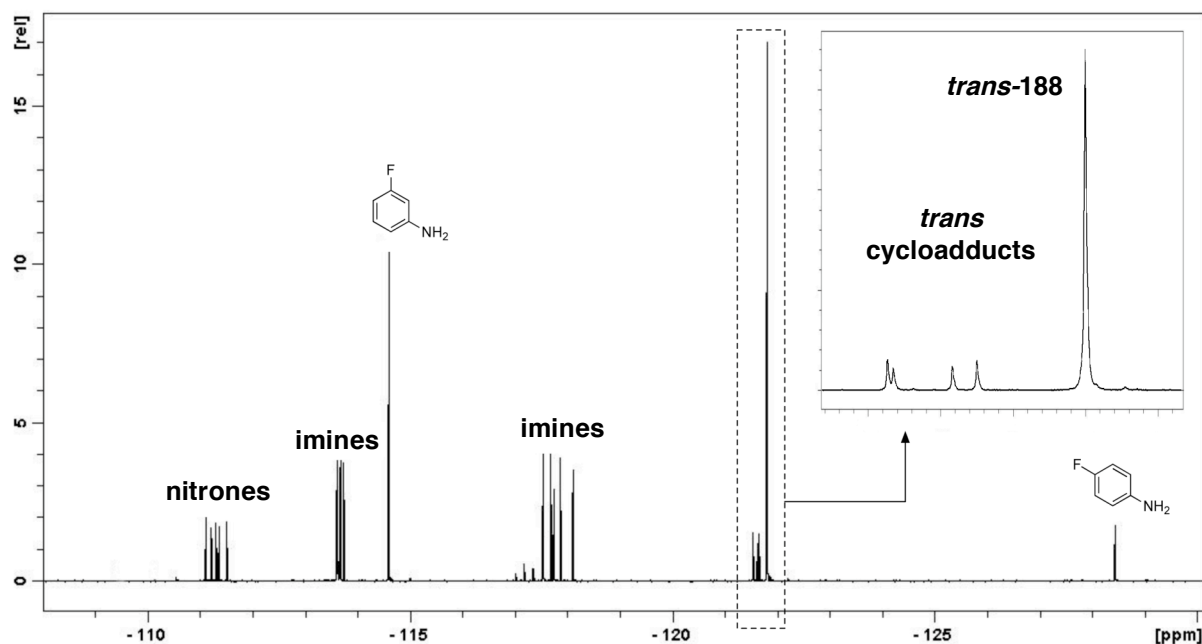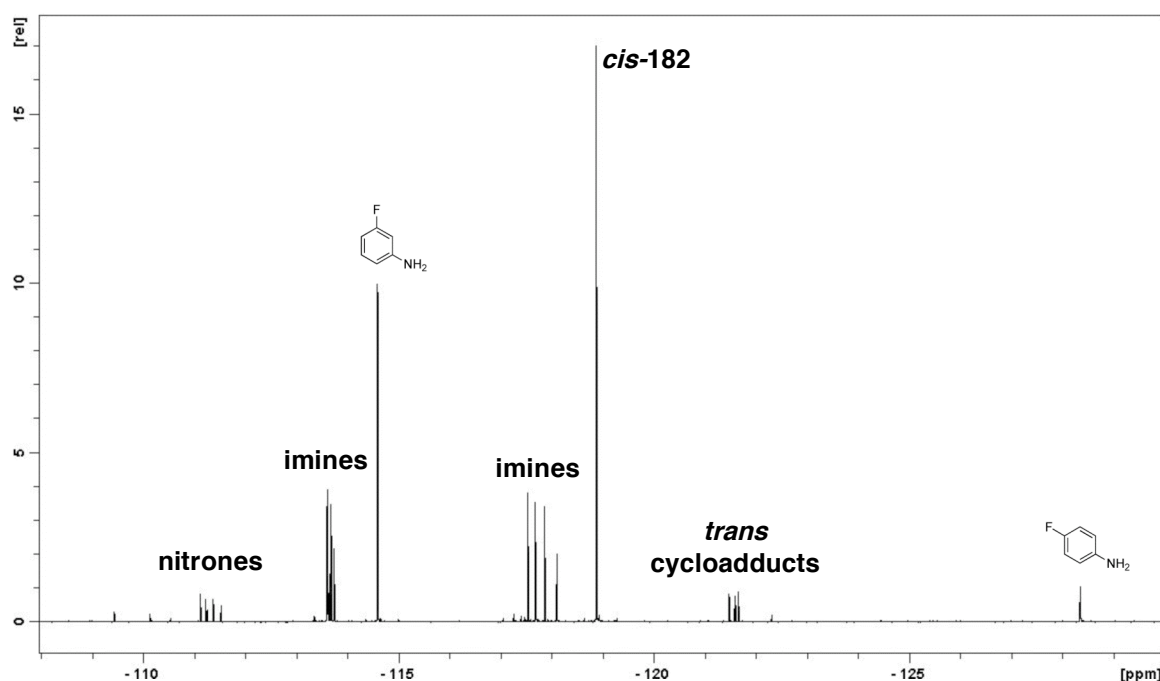

Figure 187. Partial 470.4 MHz ^{19}F NMR spectrum recorded after 16 h of a reaction mixture of a DCL created from 5 aldehydes, 4 amines and hydroxylamine **175** (Figure 173) treated with acid maleimide **179** used to derive the plot in Figure 175.

7.7 Synthetic procedures

The synthetic procedures are arranged with the structure and the number of the compound above the description of the procedure and spectroscopic data.
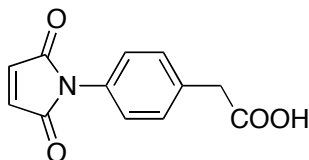
112

2-(4-(2,5-dioxo-2,5-dihydro-1*H*-pyrrol-1-yl)phenyl)acetic acid[213]

4-aminophenylacetic acid **153** (5.00 g, 33.1 mmol) was dissolved in glacial acetic acid (200 ml) and maleic anhydride **126** (3.24 g, 33.1 mmol) was added. The mixture was stirred at room temperature for 12 h followed by heating to reflux for further 5 h. The solvent was removed *in vacuo* and the residue was purified on a silica gel column (DCM : ethyl acetate 5%). This afforded a yellow solid, which was recrystallised from DCM (3.53 g, 46%). M.p. = 152.1-153.2 °C (lit.[213] 151.7-153.1 °C); ^{1}H NMR (300.1 MHz, CDCl$_3$, 25°C): δ = 7.41 (d, ^3J(H,H) = 9 Hz, 2H; ArH), 7.32 (d, ^3J(H,H) = 9 Hz, 2H; ArH), 6.85 (s, 2H; 2xCH), 3.60 (s, 2H; CH$_2$); ^{13}C NMR (75.5 MHz, CDCl$_3$, 25°C): δ = 177.8 (CO), 169.9 (CO), 134.6 (CH), 133.5 (ArC), 130.8 (ArCH), 130.6 (ArC), 126.5 (ArCH), 41.1 (CH$_2$); MS (EI): *m/z* (%): 231 (10) [M$^+$], 186 (100), 214 (15); HRMS (EI): calculated for C$_{12}$H$_9$NO$_4$ [M$^+$] 231.0532, found 231.0526.
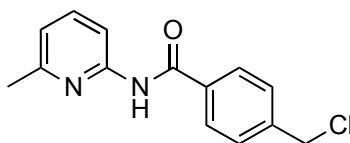
121

4-Chloromethyl-*N*-(6-methylpyridin-2-yl)benzamide[213]

4-(Chloromethyl)benzoyl chloride **120** (5.00 g, 27.0 mmol) was dissolved in DCM (50 ml) and the solution was added dropwise to 6-methyl-2-aminopyridine **119** (8.60 g, 80.0 mmol) and the mixture was stirred at room temperature for 16 h. The resultant precipitate was filtered and the solvent was removed *in vacuo*. The residue was purified by silica gel flash column chromatography (3:1 hexane : ethyl acetate) to

afford the desired compound as a colourless solid (5.50 g, 80%). M.p. = 126.1-127.3 °C; ^{1}H NMR (300.1 MHz, CDCl$_3$, 25°C): δ = 8.65 (s, 1H; NH), 8.18 (d, ^3J(H,H) = 8 Hz, 1H; ArH), 7.91 (d, ^3J(H,H) = 8.5 Hz, 2H; ArH), 7.64 (dd, ^3J(H,H) = 8 Hz and 7.5 Hz, 1H; ArH), 7.50 (d, ^3J(H,H) = 8.5 Hz, 2H; ArH), 6.93 (d, ^3J(H,H) = 7.5 Hz, 1H; ArH), 4.62 (s, 2H; CH$_2$) 2.45 (s, 3H; CH$_3$); ^{13}C NMR (75.5 MHz, CDCl$_3$ 25°C): δ = 165.0 (CO), 157.4 (ArC), 151.1 (ArC), 141.9 (ArC) 139.2 (ArCH), 134.7 (ArC), 129.3 (ArCH), 128.1 (ArCH), 120.0 (ArCH), 111.4 (ArCH), 45.7 (CH$_2$), 24.3 (CH$_3$); MS (EI+): m/z (%): 260 (18) [M$^+$], 153 (100), EA: Found C 64.49 H 5.03 N 10.74, C$_{14}$H$_{13}$N$_2$OCl requires C 64.43 H 5.15 N 10.67.
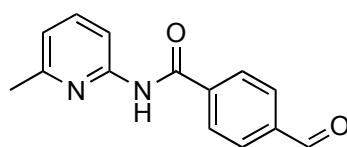
122

N-(6-methylpyridin-2-yl)-3-formylbenzamide[213]

Hexamethylene tetramine (1.90 g, 13.8 mmol) was added to a solution of 4-chloromethyl-*N*-(6-methylpyridin-2-yl)benzamide **121** (1.20 g, 4.8 mmol) in ethanol/water (1:1, 5 ml) and heated to reflux for 4 h. After that time 12 M HCl (2 ml) was added and the reaction mixture was refluxed for further 30 min. The solution was left to cool to room temperature and diluted with water (20 ml). The reaction mixture was extracted with DCM (2 x 100 ml). The combined organic fractions were dried over MgSO$_4$, filtered and the solvent was removed *in vacuo* affording the desired compound as a colourless solid (0.95 g, 85%). M.p. = 106.7-107.9 °C (lit.[213] 108.5-109.6 °C); ^{1}H NMR (300.1 MHz, CDCl$_3$, 25°C): δ = 10.00 (s, 1H; CHO), 8.80 (s, 1H; NH), 8.14 (d, ^3J(H,H) = 8 Hz, 1H; ArH), 8.03 (d, ^3J(H,H) = 8.3 Hz, 2H; ArH), 7.92 (d, ^3J(H,H) = 8.3 Hz, 2H; ArH), 7.62 (dd, ^3J(H,H) = 8.2 Hz and 7.4 Hz, 1H; ArH), 6.89 (d, ^3J(H,H) = 7.4 Hz, 1H; ArH), 2.45 (s, 3H; CH$_3$); ^{13}C NMR (75.5 MHz, CDCl$_3$, 25°C): δ = 191.8 (CO), 164.8 (CO), 156.3 (ArC), 140.6 (ArCH), 139.3 (ArC) 139.1 (ArC), 130.4 (ArCH), 128.6 (ArCH), 120.3 (ArCH), 112.1 (ArCH), 23.6 (CH$_3$); MS (ES+): m/z (%): 240 (7) [M$^+$], 239 (100); HRMS (EI): calculated for C$_{14}$H$_{12}$N$_2$O$_2$ [M+H$^+$] 241.0977, found 241.0977.
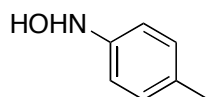
123

N-p-tolylhydroxylamine[251]

Ammonium chloride (1.90 g, 36.0 mmol) and 4-nitrotoluene (4.30 g, 31.1 mmol) were dispersed in water with ethanol mixture (2:5, 60 ml). While stirring rapidly at room temperature, zinc dust (9.60 g, 146.0 mmol) was added portion-wise, this suspension was left to stir for 15 minutes. The slurry formed was then filtered through celite with warm water and left to cool to room temperature. The reaction mixture was extracted using diethyl ether (3 x 250 ml). The organic fractions were combined, dried over $MgSO_4$ and filtered. The solvent was removed *in vacuo* to afford the crude product as a yellow solid, which was purified by recrystallisation from diethyl ether/hexane (1:4) to afford the desired product as a colourless solid (0.95 g, 25%). M.p. = 91.1-92.0 °C (lit.[252] 93.5-94.0 °C). ^{1}H NMR (300.1 MHz, $CDCl_3$, 25°C): δ = 7.11 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 6.95 (d, ^3J(H,H) = 8 Hz, 2H; ArH) 6.06 (br s, 2H; NHOH), 2.32 (s, 3H; CH_3). ^{13}C NMR (75.5 MHz, $CDCl_3$, 25°C): δ = 147.4 (ArC), 132.5 (ArCH), 129.9 (ArCH), 115.9 (ArC), 21.1 (CH_3).
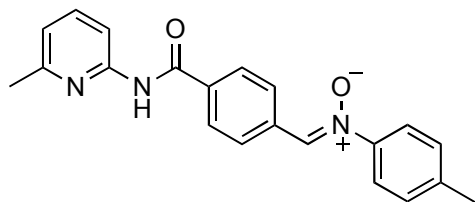
124

(*Z*)-4-methyl-*N*-(4-(6-methylpyridin-2-ylcarbamoyl)benzylidine)aniline oxide

N-(6-methylpyridin-2-yl)-3-formylbenzamide **122** (0.59 g, 2.4 mmol) and *N-p*-tolylhydroxylamine **123** (0.30 g, 2.4 mmol) were dissolved in ethanol (10 ml). The solution was left to stand for 24 h in the dark at room temperature. The resulting precipitate was filtered and washed with hexane affording the desired compound as a colourless solid (0.75 g, 90%). M.p. = 195.3-196.5 °C; ^{1}H NMR (300.1 MHz, $CDCl_3$, 25°C): δ = 8.67 (s, 1H; NH), 8.44 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 8.15 (d, ^3J(H,H) = 8 Hz, 1H; ArH), 7.98 (d, ^3J(H,H) = 8.7 Hz, 2H; ArH), 7.93 (s, 1H; CH), 7.62 (d, ^3J(H,H) = 8.7 Hz, 2H; ArH), 7.61 (dd, ^3J(H,H) = 8.4 Hz and 7.5 Hz, 1H; ArH), 7.23 (d, ^3J(H,H)

= 8 Hz, 2H; ArH), 6.90 (d, ^3J(H,H) = 7.5 Hz, 1H; ArH), 2.44 (s, 3H, CH$_3$), 2.37 (s, 3H, CH$_3$); ^{13}C NMR (75.5 MHz, CDCl$_3$, 25°C): δ = 165.1 (CO), 157.2 (ArC), 151.1 (ArC), 147.1 (ArC), 141.1 (ArC) 139.3 (ArCH), 135.8 (ArC), 134.4 (ArC), 133.5 (CH), 130.1 (ArCH), 129.4 (ArCH), 128.0 (ArCH), 121.9 (ArCH), 120.0 (ArCH), 111.6 (ArCH), 24.3 (CH$_3$), 21.6 (CH$_3$); MS (ES+): m/z (%): 368.1 (100) [M+Na$^+$]; HRMS (ES+): calculated for C$_{21}$H$_{19}$N$_3$O$_2$Na [M+Na$^+$] 368.1375, found 368.1377.
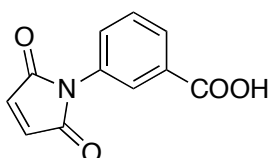
128

3-(2,5-dioxo-2,5-dihydro-1*H*-pyrrol-1-yl)benzoic acid

3-aminobenzoic acid **125** (3.00 g, 21.9 mmol) and maleic anhydride **126** (2.2 g, 21.90 mmol) were stirred in THF (50 ml) at room temperature until a precipitate was formed. After filtration it was redissolved in acetic anhydride (50 ml) and sodium acetate was added (0.36 g, 4.4 mmol). The solution was stirred at 50°C for 4 h. After that time an excess of water was added and the solution was left stirring at room temperature for 16 h. The formed precipitate was filtered and purified on a silica gel column (DCM : MeOH 5%) affording the desired compound as a colourless solid (3.8 g, 80%). M.p. = 231.8-232.6 °C (lit.[253] 232 °C); ^{1}H NMR (300.1 MHz, [D$_6$]DMSO, 25°C): δ = 7.88 7.92 (m, 2H; ArH), 7.66-7.58 (m, 2H; ArH), 7.19 (s, 2H; 2xCH); ^{13}C NMR (75.5 MHz, [D$_6$]DMSO, 25°C): δ = 169.7 (CO), 166.6 (CO), 134.7 (CH), 131.9 (ArC), 131.6 (ArC), 130.8 (ArCH), 129.2 (ArCH), 128.3 (ArCH), 127.3 (ArCH). MS (ES+): m/z (%): 240 (100) [M+Na$^+$]; HRMS (ES+): calculated for C$_{11}$H$_7$NO$_4$Na [M+Na$^+$] 240.0272, found 240.0273.
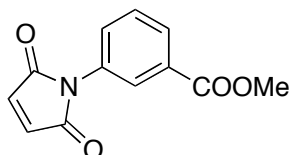
129

Methyl 3-(2,5-dioxo-2,5-dihydro-1*H*-pyrrol-1-yl)benzoate

Methyl iodide (0.46 g, 3.2 mmol) was added to a stirring solution of 3-(2,5-dioxo-2,5-

dihydro-1*H*-pyrrol-1-yl)benzoic acid **128** (0.35 g, 1.6 mmol) and caesium carbonate (0.26 g, 1.6 mmol) in DMSO (15 ml). The mixture was left stirring in the dark under nitrogen for 16 h. The mixture was diluted with water and extracted with DCM. The combined extracts were dried over MgSO$_4$. The solvent was removed and the residue was purified using silica gel column chromatography (DCM) affording the desired product (0.32 g, 86%). M.p. = 129.6-129.8 °C; ^{1}H NMR (300.1 MHz, CDCl$_3$, 25°C): δ = 8.01-7.97 (m, 2H; ArH), 7.50-7.48 (m, 2H; ArH), 6.81 (s, 2H; 2xCH), 3.86 (s, 3H; CH$_3$); ^{13}C NMR (100.6 MHz, [D$_6$]DMSO, 25°C): δ = 169.7 (CO), 165.5 (CO), 134.7 (2xCH), 132.0 (ArC), 131.3 (ArCH), 130.3 (ArC), 129.4 (ArCH), 128.1 (ArCH), 127.1 (ArCH), 52.3 (CH$_3$); MS (ES+): *m/z* (%): 254 (100) [M+Na$^+$]; HRMS (ES+): calculated for C$_{12}$H$_9$NO$_4$Na [M+Na$^+$] 254.0434, found 254.0429.
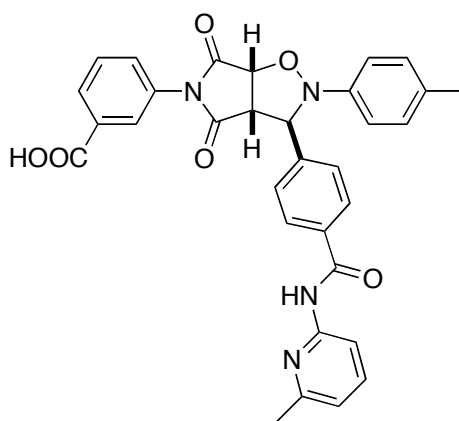
***trans*-131**

3-(3-(4-(6-methylpyridin-2-ylcarbamoyl)phenyl)-4,6-dioxo-2-*p*-tolyldihydro-2*H*-pyrrolo[3,4-*d*]isoxazol-5(3*H*,6*H*,6a*H*)-yl)benzoic acid

3-(2,5-dioxo-2,5-dihydro-1*H*-pyrrol-1-yl)benzoic **128** acid (0.10 g, 0.5 mmol) and (*Z*)-4-methyl-*N*-(4-(6-methylpyridin-2-ylcarbamoyl)benzylidene)aniline oxide **124** (0.16 g, 0.5 mmol) were dissolved in chloroform (5 ml) and left in room temperature for 24 h. The solvent was removed *in vacuo* and the crude residue was recrystallised form ethanol yielding the product as a white solid (0.15 g, 58%). M.p. = 181.2-182.8 °C; ^{1}H NMR (400.1 MHz, [D$_6$]DMSO, 25°C): δ = 10.74 (s, 1H; NH), 8.07 (d, ^{3}*J*(H,H) = 8 Hz, 2H; ArH), 8.07 (d, ^{3}*J*(H,H) = 8 Hz, 1H; ArH), 7.91 (d, ^{3}*J*(H,H) = 8 Hz, 1H; ArH), 7.74-7.66 (m, 3H; ArH), 7.52 (t, ^{3}*J*(H,H) = 8 Hz, 1H; ArH), 7.21 (s, 1H; ArH), 7.14-7.12 (m, 2H; ArH), 7.04-7.01 (m, 3H; ArH), 6.89 (d, ^{3}*J*(H,H) = 8 Hz, 1H; ArH), 5.96 (s, 1H; CH), 5.41 (d, ^{3}*J*(H,H) = 7 Hz, 1H; CH), 4.14 (d, ^{3}*J*(H,H) = 7 Hz, 1H; CH),

2.45 (s, 3H; CH$_3$), 2.19 (s, 3H; CH$_3$); ^{13}C NMR (100.6 MHz, [D$_6$]DMSO, 25°C): δ = 174.2 (CO), 173.1 (CO), 166.2 (CO), 165.4 (CO), 156.5 (ArC), 151.4 (ArC), 146.5 (ArC), 142.9 (ArC), 138.4 (ArCH), 133.4 (ArC), 131.8 (ArC), 131.6 (ArC), 131.5 (ArC),130.6 (ArCH), 129.4 (ArCH), 129.3 (ArCH), 129.1 (ArCH), 128.1 (ArCH), 127.2 (ArCH), 127.0 (ArCH), 119.1 (ArCH), 114.3 (ArCH), 111.7 (ArCH), 77.85 (CH), 68.3 (CH), 56.5 (CH), 23.5 (CH$_3$), 20.0 (CH$_3$); MS (ES–): m/z (%): 561.29 (100) [M–H$^+$]; HRMS (ES–): calculated for C$_{32}$H$_{25}$N$_4$O$_6$ [M–H$^+$] 561.1769, found 561.1774.
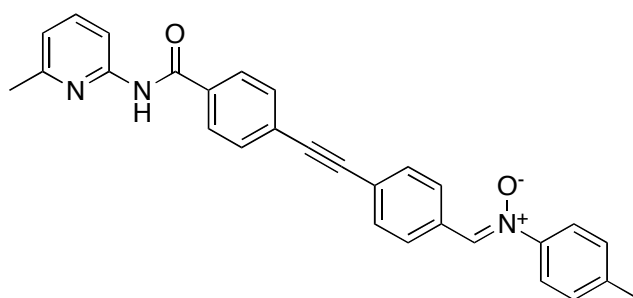
132

(Z)-4-methyl-N-(4-((4-(6-methylpyridin-2-ylcarbamoyl)phenyl)ethynyl) benzylidene)aniline oxide

4-((4-formylphenyl)ethynyl)-N-(6-methylpyridin-2-yl)benzamide **146** (1.00 g, 2.9 mmol) and N-p-tolylhydroxylamine **123** (0.38 g, 2.9 mmol) were dissolved in minimal amount of ethanol and left in a freezer for 16 h. After that time a precipitate was observed and filtered affording the desired product (1.07 g, 82%). M.p. = 233.5-233.8 °C; ^{1}H NMR (300.1 MHz, CDCl$_3$, 25°C): δ = 8.54 (s, 1H; NH), 8.32 (d, ^{3}J(H,H) = 8 Hz, 2H; ArH), 8.12 (d, ^{3}J(H,H) = 8 Hz, 1H; ArH), 7.88-7.86 (m, 3H; ArH), 7.62-7.56 (m, 7H; ArH), 7.21 (d, ^{3}J(H,H) = 8 Hz, 2H; ArH), 6.88 (d, ^{3}J(H,H) = 8 Hz, 1H; ArH), 2.41 (s, 3H; CH$_3$), 2.36 (s, 3H; CH$_3$); ^{13}C NMR (75.5 MHz, CDCl$_3$, 25°C): δ = 164.8 (CO), 156.8 (ArC), 150.6 (ArC), 146.8 (ArC), 140.5 (ArC), 139.0 (ArCH), 133.8 (ArC), 133.3 (ArCH), 132.0 (ArCH), 131.9 (ArCH), 130.9 (ArC), 129.7 (ArCH), 128.7 (ArCH), 127.3 (ArCH), 127.0 (ArC), 124.7 (ArC), 121.5 (ArCH), 119.6 (ArCH), 111.0 (CH), 92.2 (C), 90.8 (C), 23.9 (CH$_3$), 21.2 (CH$_3$); MS (ES+): m/z (%): 468.22 (100) [M+Na$^+$]; HRMS (ES+): calculated for C$_{29}$H$_{24}$N$_3$O$_2$ [M+H$^+$] 446.1859, found 446.1869.
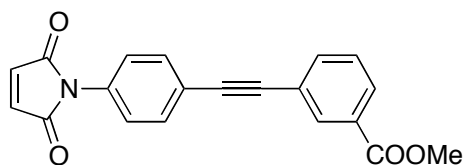
133

Methyl 3-((4-(2,5-dioxo-2,5-dihydro-1*H*-pyrrol-1-yl)phenyl)ethynyl)benzoate

Methyl 3-((4-(*tert*-butoxycarbonylamino)phenyl)ethynyl)benzoate **139** (0.50 g, 1.4 mmol) was treated with TFA (5 ml) in order to remove the Boc protecting group and after stirring for 15 min the solvent was evaporated. The residue was redissolved in acetic acid (20 ml) and maleic anhydride **126** (0.14 g, 1.4 mmol) was added and the mixture was stirred for 16 h. A precipitate was observed which was filtered and dissolved in acetic anhydride (5 ml). Sodium acetate (0.02 g, 0.3 mmol) was added and the mixture was refluxed for 4 h. After cooling, ice cold water (30 ml) was added and the mixture was stirred for 16 h. After that time the precipitate was observed and filtered, affording the desired product (0.30 g, 63%). M.p. = 142.7-143.0 °C; ^{1}H NMR (400.1 MHz, [D$_6$]DMSO, 25°C): δ = 8.09 (s, 1H; ArH); 7.98 (d, ^3J(H,H) = 8 Hz, 1H; ArH); 7.83 (d, ^3J(H,H) = 8 Hz, 1H; ArH); 7.70 (d, ^3J(H,H) = 8 Hz, 2H; ArH); 7.59 (t, ^3J(H,H) = 8 Hz, 1H; ArH); 7.42 (d, ^3J(H,H) = 8 Hz, 2H; ArH); 7.18 (s, 2H; 2xCH), 3.87 (s, 3H; CH$_3$); ^{13}C NMR (100.6 MHz, [D$_6$]DMSO, 25°C): δ =169.6 (CO), 165.4 (CO), 135.8 (ArCH), 134.7 (CH), 132.0 (ArC), 131.9 (ArCH), 131.8 (ArCH), 130.2 (ArC), 129.5 (ArCH), 129.3 (ArCH), 126.6 (ArCH), 122.6 (ArC), 120.8 (ArC), 89.6 (C), 88.6 (C), 52.4 (CH$_3$); MS (ES–): *m*/*z* (%): 362 (100) [M–H$^+$+MeOH].
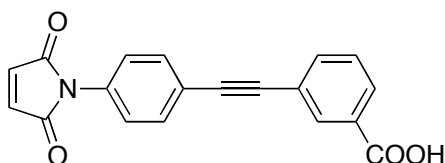
134

## 3-((4-(2,5-dioxo-2,5-dihydro-1*H*-pyrrol-1-yl)phenyl)ethynyl)benzoic acid

3-((4-(*tert*-butoxycarbonylamino)phenyl)ethynyl)benzoic acid **140** (0.20 g, 0.6 mmol) was dissolved in TFA (5 ml) and stirred for 15 min to remove the Boc protecting group. The solvent was removed and replaced with acetic acid (10 ml). To a stirring solution maleic anhydride **126** (0.06 g, 0.8 mmol) was added and the solution was left

in room temperature for 16 h. A precipitate was observed, filtered and dissolved in acetic anhydride (5 ml). Sodium acetate (0.01 g, 0.1 mmol) was added to the solution and the reaction was refluxed for 4 h. After that time the solution was cooled and water (20 ml) was added to the solution. The mixture was stirred in room temperature for 16 h. A precipitate was observed, filtered and dissolved in TFA in order to remove any mixed anhydride that may have been formed. The solvent was removed and the residue was recrystallised from ethyl acetate affording the desired product (0.10 g, 57%). M.p. = 259.6-259.7 °C; (300.1 MHz, [D_6]DMSO, 25°C): δ = 8.08 (s, 1H; ArH), 7.97 (d, ^3J(H,H) = 8 Hz, 1H; ArH), 7.81 (d, ^3J(H,H) = 8 Hz, 1H; ArH), 7.71 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 7.58 (t, ^3J(H,H) = 8 Hz, 1H; ArH), 7.43 (d, ^3J(H,H) = 8 Hz, 1H; ArH), 7.18 (s, 2H; 2xCH); ^{13}C NMR (75.5 MHz, [D_6]DMSO, 25°C): δ = 169.6 (CO), 166.5 (CO), 135.4 (ArCH), 134.7 (CH), 132.0 (ArCH), 131.9 (ArCH), 131.3 (ArC), 129.6 (ArCH), 129.3 (ArCH), 128.6 (ArC), 126.7 (ArCH), 122.4 (ArC), 123.0 (ArC), 89.4 (C), 88.9 (C); MS (ES-): m/z (%): 316 (100) [M–H⁺], 272 (90); HRMS (ES-): calculated for $C_{19}H_{10}NO_4$ [M–H⁺] 316.0607, found 316.0610.
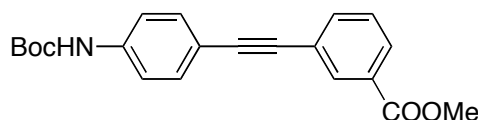
139

Methyl 3-((4-(*tert*-butoxycarbonylamino)phenyl)ethynyl)benzoate

A mixture of toluene (20 ml) and triethylamine (20 ml) was degassed with argon for 15 min. After that time *tert*-butyl 4-ethynylphenylcarbamate **136** (2.00 g, 9.22 mmol) and 3-bromobenzoic acid methyl ester **138** (1.96 g, 9.2 mmol) were added. The mixture was degassed further for 15 min and dichlorobis(triphenylphosphine) palladium(II) (0.36 g, 0.5 mmol) and copper iodide (0.09 g, 0.5 mmol) were added. The mixture was stirred under argon for 16 h at 85 °C. The solvent was removed *in vacuo* and the residue was dissolved in DCM and purified on a silica gel chromatography column (DCM) affording the desired product (1.94 g, 60%). M.p. = 171.9-172.2 °C; ^{1}H NMR (300.1 MHz, CDCl$_3$, 25°C): δ = 8.12 (s, 1H; ArH), 7.91 (d, ^3J(H,H) = 8 Hz, 1H; ArH), 7.61 (d, ^3J(H,H) = 8 Hz, 1H; ArH), 7.42-7.28 (m, 5H; ArH), 6.49 (s, 1H; NH), 3.86 (s, 3H; CH$_3$), 1.46 (s, 9H; 3xCH$_3$); ^{13}C NMR (100.6 MHz, [D_6]DMSO, 25°C): δ = 165.5 (CO), 152.5 (CO), 140.3 (ArC), 135.5 (ArCH), 132.1

(ArCH), 131.5 (ArCH), 130.1 (ArC), 129.3 (ArCH), 128.8 (ArCH), 123.2 (ArC), 117.9 (ArCH), 114.8 (ArC), 90.7 (C), 87.0 (C), 79.5 (C), 52.3 (CH$_3$), 28.0 (CH$_3$); MS (ES–): m/z (%): 350 (100) [M–H$^+$]; HRMS (ES–): calculated for C$_{21}$H$_{20}$NO$_4$ [M–H$^+$] 350.1397, found 350.1392.
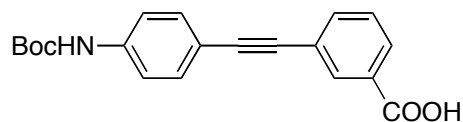
140

3-((4-(*tert*-butoxycarbonylamino)phenyl)ethynyl)benzoic acid

Methyl 3-((4-(*tert*-butoxycarbonylamino)phenyl)ethynyl)benzoate **139** (0.30 g, 0.8 mmol) was dissolved in THF (4 ml) and a solution of lithium hydroxide (0.12 g, 5.0 mmol) in water (2 ml) was added. The mixture was stirred for 16 h at 75 °C. After that time the solvent was evaporated and the residue was redissolved in water acidified with 1M HCl and extracted wit ethyl acetate. The organic fraction was dried with MgSO$_4$ and evaporated affording the desired product (0.21 g, 72%). M.p. = 203 °C dec.; ^{1}H NMR (300.1 MHz, [D$_6$]DMSO, 25°C): δ = 9.53 (s, 1H; NH), 7.98 (s, 1H; ArH), 7.90 (d, ^3J(H,H) = 8 Hz, 1H; ArH), 7.71 (d, ^3J(H,H) = 8 Hz, 1H; ArH), 7.54-7.42 (m, 5H; ArH), 1.43 (s, 9H; 3xCH$_3$); ^{13}C NMR (75.5 MHz, [D$_6$]DMSO, 25°C): δ = 166.7 (CO), 152.6 (CO), 140.1 (ArC), 135.1 (ArCH), 132.1 (ArCH), 131.7 (ArCH), 131.3 (ArC), 129.3 (ArCH), 129.0 (ArCH), 122.9 (ArC), 118.0 (ArCH), 115.0 (ArC), 90.4 (C), 87.2 (C), 79.7 (C), 28.0 (CH$_3$); MS (ES–): m/z (%): 336 (100) [M–H$^+$]; HRMS (ES–): calculated for C$_{20}$H$_{18}$NO$_4$ [M–H$^+$] 336.1238, found 336.1236.
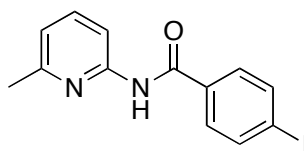
144

4-iodo-*N*-(6-methylpyridin-2-yl)benzamide

2-amino-6-methylpyridine **119** (4.32 g, 40.0 mmol) was dissolved in dry DCM (100 ml) and added dropwise to a solution of 4-iodobenzoyl chloride **143** (4.25 g, 16.0 mmol) in dry DCM (50 ml) under a positive pressure of nitrogen. The reaction mixture

was stirred at reflux for 12 h and allowed to cool. The solution was washed with water (50 ml) and 10% sodium bicarbonate (50 ml). The aqueous layer was further washed with DCM (2 x 50 ml). The organic layers were combined and dried over MgSO$_4$. The mixture was filtered and the solvent was removed *in vacuo* to afford the desired product as a colourless solid (4.32 g, 80%). M.p. = 120.1-122.3 °C; ^{1}H NMR (300.1 MHz, CDCl$_3$, 25 °C): δ = 8.66 (s, 1H; NH), 8.14 (d, ^3J(H,H) = 8 Hz, 1H; ArH), 7.80 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 7.63 (t, ^3J(H,H) = 8 Hz, 1H; ArH), 7.61 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 6.91 (d, ^3J(H,H) = 8 Hz, 1H; ArH), 2.43 (s, 3H; CH$_3$); ^{13}C NMR (75.5 MHz, CDCl$_3$, 25°C): δ = 165.0 (CO), 157.0 (ArC), 150.6 (ArC), 138.8 (ArC), 138.0 (ArCH), 136.2 (ArCH), 133.8 (ArC), 128.8 (ArCH), 119.7 (ArCH), 111.1 (ArCH), 99.4 (ArC), 24.0 (CH$_3$); MS (EI): m/z (%): 338 (23) [M$^+$], 310 (57), 231 (82), 108 (100); HRMS (EI): calculated for C$_{13}$H$_{11}$IN$_2$ONa [M+Na$^+$] 360.9824, found 360.9814.
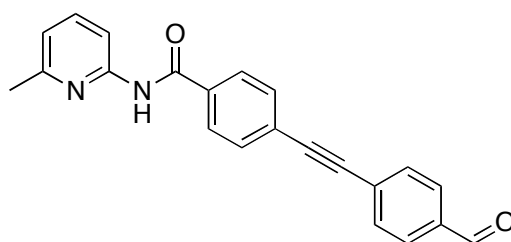
146

4-((4-formylphenyl)ethynyl)-*N*-(6-methylpyridin-2-yl)benzamide

A mixture of toluene (20 ml) and triethylamine (20 ml) was degassed with argon for 15 min. After that time 4-Iodo-N-(6-methyl-pyridin-2-yl)-benzamide **144** (2.00 g, 5.9 mmol) and 4-ethynyl-benzaldehyde **145** (0.77 g, 5.9 mmol) were added. The mixture was degassed further for 15 min and bis(triphenylphosphine)palladium dichloride (0.23 g, 0.3 mmol) and copper iodide (0.06 g, 0.3 mmol) were added. The mixture was stirred under argon for 16 h at 85 °C. The solvent was removed and the residue was dissolved in DCM and purified on a silica gel chromatography column (DCM) affording the desired product (1.29 g, 64%). M.p. = 197.8-198.1 °C; ^{1}H NMR (300.1 MHz, CDCl$_3$, 25 °C): δ = 9.97 (s, 1H; CHO), 8.43 (s, 1H; NH), 8.11 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 7.90-7.81 (m, 4H; ArH), 7.65-7.57 (m, 5H; ArH), 6.89 (d, ^3J(H,H) = 8 Hz, 1H; ArH), 2.42 (s, 3H; CH$_3$); ^{13}C NMR (75.5 MHz, CDCl$_3$, 25°C): δ = 191.4 (CHO), 164.9 (CO), 156.9 (ArC), 150.5 (ArC), 138.9 (ArCH), 135.8 (ArC), 134.2 (ArC), 132.3 (ArCH), 132.1 (ArCH), 129.6 (ArCH), 128.9 (ArC), 127.3 (ArCH), 126.5

(ArC), 119.7 (ArCH), 111.0 (ArCH), 92.2 (C), 91.1 (C), 24.0 (CH$_3$); MS (CI+): *m/z* (%): 341 (100) [M+H$^+$], 312 (25), 233 (20); HRMS (CI+): calculated for C$_{22}$H$_{17}$N$_2$O$_2$ [M+H$^+$] 341.1304, found 341.1290.
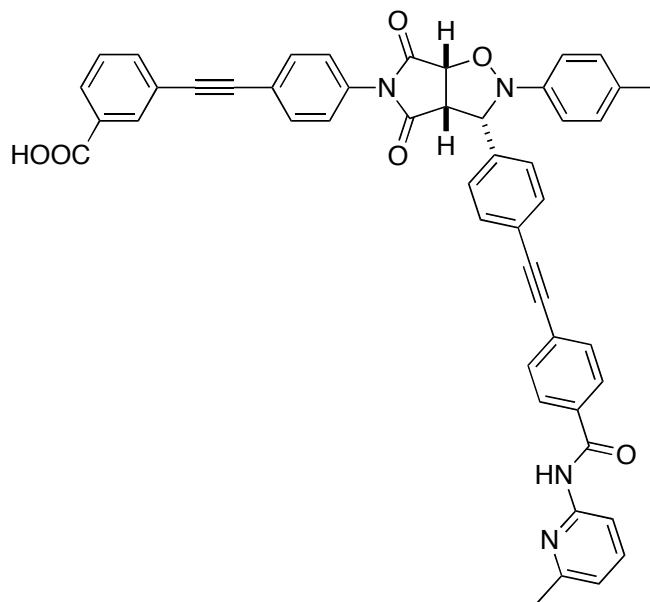
cis-152

3-((4-(3-(4-((4-(6-methylpyridin-2-ylcarbamoyl)phenyl)ethynyl)phenyl)-4,6-dioxo-2-*p*-tolyldihydro-2*H*-pyrrolo[3,4-*d*]isoxazol-5(3*H*,6*H*,6a*H*)-yl)phenyl)ethynyl)benzoic acid

3-((4-(2,5-dioxo-2,5-dihydro-1*H*-pyrrol-1-yl)phenyl)ethynyl)benzoic acid (0.10 g, 0.3 mmol) **134** and (*Z*)-4-methyl-*N*-(4-((4-(6-methylpyridin-2-ylcarbamoyl)phenyl)ethynyl) benzylidene)aniline oxide **132** (0.14 g, 0.3 mmol) were dissolved in chloroform (5 ml) and left at room temperature for 16 h. The solvent was removed *in vacuo* and the residue was recrystallised from ethanol yielding the product as a yellow solid (0.19 g, 80%). M.p. = 202.3-203.7 °C; ^{1}H NMR (400.1 MHz, [D$_6$]DMSO, 25°C): δ = 10.80 (s, 1H; NH), 8.07-7.95 (m, 5H; ArH), 7.81 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 7.74-7.69 (m, 3H; ArH), 7.64 (d, ^3J(H,H) = 8 Hz, 1H; ArH), 7.59-7.55 (m, 3H; ArH), 7.48 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 7.24 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 7.09 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 7.04-7.01 (m, 3H; ArH), 5.39 (d, ^3J(H,H) = 8 Hz, 1H; CH), 4.98 (d, ^3J(H,H) = 9 Hz, 1H; CH), 4.23 (dd, 1H, ^3J(H,H) = 9 Hz and 8 Hz, 1H; CH), 2.44 (s, 3H; CH$_3$), 1.46 (s, 3H; CH$_3$); ^{13}C NMR (100.6 MHz, [D$_6$]DMSO, 25°C): δ = 174.2 (CO), 171.9 (CO), 166.5 (CO), 156.1 (CO), 156.6 (ArC), 151.4 (ArC), 144.3 (ArC), 138.4 (ArCH), 136.3 (ArC), 135.4 (ArCH), 134.8 (ArC), 133.9 (ArC), 132.1 (ArCH), 132.0 (ArCH), 131.7 (ArCH),

131.4 (ArC), 131.2 (ArCH), 129.6 (ArCH), 129.3 (ArCH), 129.2 (ArCH), 128.4 (ArCH), 128.1 (ArCH), 127.4 (ArC), 126.8 (ArCH), 125.4 (ArC), 122.3 (ArC), 122.0 (ArC), 121.7 (ArC), 120.4 (ArCH), 119.1 (ArCH), 111.7 (ArCH), 91.2 (C), 89.3 (C), 89.2 (C), 89.1 (C), 77.2 (CH), 70.3 (CH), 54.9 (CH), 23.5 (CH$_3$), 20.3 (CH$_3$); MS (ES–): *m/z* (%): 761 (100) [M–H$^+$], 762 (15) [M$^+$]; HRMS (ES–): calculated for C$_{48}$H$_{33}$N$_4$O$_6$ [M–H$^+$] 761.2399, found 761.2400.
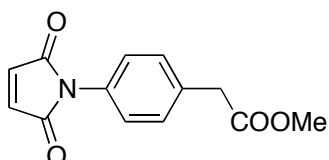
154

Methyl 2-(4-(2,5-dioxo-2,5-dihydro-1H-pyrrol-1-yl)phenyl)acetate[213]

2-(4-(2,5-dioxo-2,5-dihydro-1H-pyrrol-1-yl)phenyl)acetic acid **112** (1.00 g, 4.3 mmol) was added to a solution of silica supported sodium hydrogen sulfate (100 mg/mmol) in methanol (20 ml) and left stirring under nitrogen for 30 h. The suspension was filtered and the solution was evaporated *in vacuo* to yield a yellow residue, which was purified on a silica gel chromatography column (DCM) affording the desired product as a yellow solid (0.58 g, 55%). M.p. = 86.1-87.2 °C (lit.[213] 86.7-87.6 °C); ^{1}H NMR (300.1 MHz, CDCl$_3$, 25°C): δ = 7.32 (d, ^3J(H,H) = 9 Hz, 2H; ArH), 7.24 (d, ^3J(H,H) = 9 Hz, 2H; ArH), 6.79 (s, 2H; 2xCH), 3.63 (s, 3H; CH$_3$), 3.59 (s, 2H; CH$_2$); ^{13}C NMR (75.5 MHz, CDCl$_3$, 25°C): δ = 171.5 (CO), 169.9 (CO), 134.6 (CH), 134.2 (ArC), 130.5 (ArCH), 126.5 (ArCH), 52.6 (CH$_3$), 41.2 (CH$_2$); MS (EI): m/z (%): 245 (40) [M$^+$], 186 (100); HRMS (EI): calculated for C$_{13}$H$_{11}$NO$_4$ [M$^+$] 245.0688, found 245.0694.
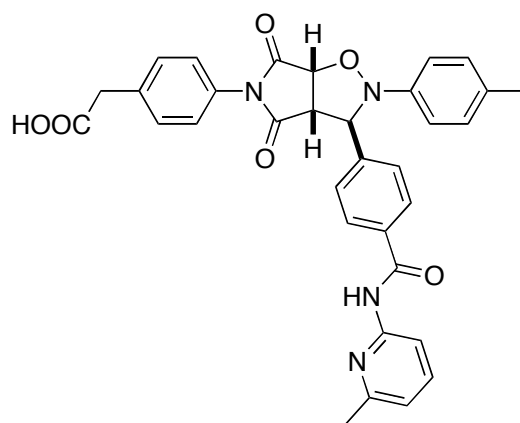
trans-156

2-(4-(3-(4-(6-methylpyridin-2-ylcarbamoyl)phenyl)-4,6-dioxo-2-*p*-tolyldihydro-2*H*-pyrrolo[3,4-*d*]isoxazol-5(3*H*,6*H*,6a*H*)-yl)phenyl)acetic acid

2-(4-(2,5-dioxo-2,5-dihydro-1*H*-pyrrol-1-yl)phenyl)acetic acid **112** (0.09 g, 0.4 mmol) and (*Z*)-4-methyl-*N*-(4-(6-methylpyridin-2-ylcarbamoyl)benzylidene)aniline oxide **124** (0.13 g, 0.4 mmol) were dissolved in chloroform (15 ml). The solution was left in the dark at 0°C. The resulting precipitate was filtered and washed with chloroform to afford the desired product as a colourless solid (0.18 g, 80%). M.p. = 194 °C dec.; ^{1}H NMR (300.1 MHz, CDCl$_3$, 25°C): δ = 11.09 (s, 1H; NH), 8.38 (d, ^3J(H,H) = 8 Hz, 1H; ArH), 8.94 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 7.77 (dd, ^3J(H,H) = 8 Hz and 7 Hz, 1H; ArH), 7.44 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 7.32 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 7.14-7.07 (m, 4H; ArH), 6.99 (d, ^3J(H,H) = 7 Hz, 1H; ArH), 6.98 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 5.78 (s, 1H; CH), 4.98 (d, ^3J(H,H) = 7 Hz, 1H; CH), 3.92 (d, ^3J(H,H) = 7 Hz, 1H; CH), 2.54 (s, 2H; CH$_2$), 2.54 (s, 3H; CH$_3$), 2.28 (s, 3H; CH$_3$); ^{13}C NMR (75.5 MHz, CDCl$_3$, 25°C): δ = 176.6 (CO), 174.8 (CO), 173.4 (CO), 166.5 (ArC), 155.7 (ArC), 151.7 (ArC), 146.4 (ArC), 142.6 (ArC), 140.4 (ArCH), 135.4 (ArC), 133.6 (ArC), 132.7 (ArC), 130.5 (ArCH), 129.9 (ArCH), 128.8 (ArCH), 126.5 (ArCH), 126.1 (ArCH), 119.9 (ArCH), 114.6 (ArCH), 113.2 (ArCH), 77.0 (CH), 68.8 (CH), 57.1 (CH), 40.2 (CH$_2$), 22.0 (CH$_3$), 20.5 (CH$_3$); MS (ES+): *m/z* (%): 577 (5) [M+H$^+$], 599 (100); EA: Found C 68.3 H 4.8 N 9.5, C$_{33}$H$_{28}$N$_4$O$_6$ requires C 68.7 H 4.9 N 9.7.
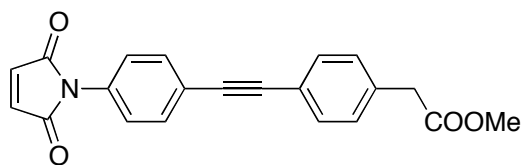
157

Methyl2-(4-((4-(2,5-dioxo-2,5-dihydro-1*H*-pyrrol-1-yl)phenyl)ethynyl)phenyl) acetate

Methyl 2-(4-((4-aminophenyl)ethynyl)phenyl)ethanoate **161** (2.00 g, 7.5 mmol) and maleic anhydride **126** (0.75 g, 7.5 mmol) were dissolved in glacial acetic acid (45 ml). The solution was stirred at room temperature for 12 h, followed by heating to reflux for 5 h. The solvent was removed *in vacuo*. The residue was purified on silica gel column (hexane : ethyl acetate, 3:1) affording the desired product as off-yellow solid (0.83 g, 32%). M.p. = 129.2-130.1 °C; ^{1}H NMR (300.1 MHz, CDCl$_3$, 25°C): δ = 7.53 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 7.42 ((d, ^3J(H,H) = 8 Hz, 2H; ArH), 7.29 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 7.20 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 6.78 (s, 2H; 2xCH), 3.63 (s, 3H; CH$_3$), 3.57 (s, 2H; CH$_2$); ^{13}C NMR (75.5 MHz, CDCl$_3$, 25°C): δ = 171.6 (CO), 169.2 (CO), 134.4 (ArC), 134.3 (ArCH), 132.3 (ArCH), 131.9 (ArCH), 131.1 (ArC), 129.4 (CH), 125.6 (ArCH), 122.9 (ArC), 121.9 (ArC), 90.1 (C), 88.7 (C), 52.2 (CH$_2$), 41.1 (CH$_3$). MS (CI+): m/z (%): 346.10 (100) [M+H$^+$], 345 (20) [M$^+$], 348 (30), 286 (25); HRMS (CI+): calculated for C$_{21}$H$_{16}$NO$_4$ 346.1071 [M$^+$+H], found 346.1079.
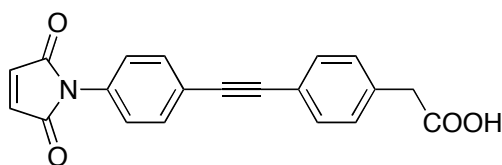
158

2-(4-((4-(2,5-dioxo-2,5-dihydro-1*H*-pyrrol-1-yl)phenyl)ethynyl)phenyl)acetic acid

2-(4-((4-aminophenyl)ethynyl)phenyl)ethanoic acid **163** (1.00 g, 4.0 mmol), and maleic anhydride **126** (0.39 g, 4.0 mmol) were dissolved in THF (20 ml) and stirred at room temperature for 5 h until a yellow precipitate was observed. The precipitate was filtered, redissolved in DMF (60 ml) and cooled to 0 °C. *N*-hydroxybenzotriazole (0.54 g, 4.0 mmol) was added first followed by the addition of 1-(3-dimethylaminopropyl)-3-ethyl carbodiimide hydrochloride (0.77 g, 4.0 mmol). The reaction mixture was

allowed to warm to room temperature and was stirred for 12 h. Water was added to the solution (50 ml) and the mixture was extracted with ethyl acetate (50 ml). The organic layer was washed with ice water (50 ml), brine (50 ml) and 1 M phosphoric acid (50 ml). The organic layer was dried over $MgSO_4$, filtered and the solvent was removed *in vacuo*. The residue was purified on silica gel column (DCM : 5% methanol). The product was obtained as a yellow solid (0.32 g, 22%). M.p. = 206.8-207.6 °C; ^{1}H NMR (300.1 MHz, [D_6]DMSO, 25°C): δ = 12.44 (s, 1H; COOH), 7.67 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 7.53 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 7.42 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 7.33 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 7.21 (s, 2H; 2xCH), 3.64 (s, 2H; CH_2), 3.56 (s, 3H; CH_3); ^{13}C NMR (75.5 MHz, [D_6]DMSO, 25°C): δ = 172.2 (CO), 169.6 (CO), 136.1 (ArC), 134.8 (ArCH), 131.8 (ArCH), 131.7 (ArC), 131.3 (CH), 129.9 (ArCH), 126.6 (ArCH), 121.4 (ArC), 120.3 (ArC), 89.9 (C), 88.5 (C), 40.4 (CH_2); MS (CI+): m/z (%): 332 (100) [M+H$^+$] , 331 (20) [M$^+$], 286 (20); HRMS (CI+): calculated for $C_{20}H_{14}NO_4$ 332.0920 [M+H$^+$], found 332.0923.
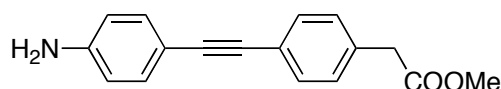
161

Methyl 2-(4-((4-aminophenyl)ethynyl)phenyl)acetate

A mixture of 60 ml of triethylamine and 20 ml of THF was degasses with argon. 4-ethynylaniline **135** (5.00 g, 42.7 mmol) and methyl 2-(4-bromophenyl)acetate **160** (9.70 g, 42.7 mmol) were added and the solution was further degassed with argon. Bis(triphenylphosphine)palladium dichloride (1.50 g, 2.1 mmol) and copper iodide (0.40 g, 2.1 mmol) were added to the solution. The reaction mixture was stirred at 80 °C for 16 h. The solvent was removed *in vacuo* and the residue was purified on silica gel column (DCM). The product was obtained as a colourless solid (5.21 g, 46%). M.p. = 95.8-96.6 °C; ^{1}H NMR (300.1 MHz, CDCl$_3$, 25°C): δ = 7.49 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 7.37 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 7.28 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 6.67 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 3.87 (s, 2H; NH_2), 3.74 (s, 3H; CH_3), 3.67 (s, 2H; CH_2); ^{13}C NMR (75.5 MHz, CDCl$_3$, 25°C): δ = 171.9 (CO), 146.7 (ArC), 133.5 (ArC), 133.0 (ArCH), 131.5 (ArCH), 129.3 (ArCH), 122.8 (ArC), 114.8 (ArCH), 122.6 (ArC), 90.3 (C), 87.1 (C), 52.2 (CH_2), 41.1 (CH_3); MS (ES+): m/z (%): 288 (100) [M+Na$^+$]; HRMS (ES+): calculated for $C_{17}H_{16}NO_2$ 266.1177 [M+H$^+$], found 266.1181.
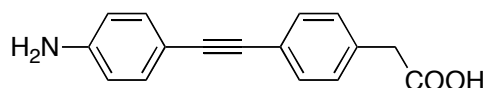
163

2-(4-((4-aminophenyl)ethynyl)phenyl)acetic acid

Methyl 2-(4-((4-aminophenyl)ethynyl)phenyl)ethanoate **161** (2.00 g, 7.5 mmol) was dissolved in 30 ml of THF. Lithium hydroxide (1.10 g, 45.2 mmol) was dissolved in 15 ml of water. The solutions were combined and stirred rapidly at 70 °C for 16 h. The solution was cooled and water (30 ml) was added. The solution was washed with ethyl acetate (50 ml) and acidified with 1 M hydrochloric acid until a precipitate was observed. The precipitate was filtered and dissolved in hot 10 % sodium bicarbonate (20 ml). 1 M hydrochloric acid was added dropwise to the solution until a faint precipitate was observed. At this point the slurry was extracted with ethyl acetate (2 x 30 ml). The organic layer was dried with $MgSO_4$, filtered and the solvent was removed *in vacuo*. The product was obtained as a colourless solid (0.72 g, 37%). M.p. = 166 °C dec.; ^{1}H NMR (300.1 MHz, [D$_6$]DMSO, 25°C): δ = 7.39 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 7.26 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 7.19 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 6.56 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 5.56 (s, 2H; NH$_2$), 3.59 (s, 2H; CH$_2$); ^{13}C NMR (75.5 MHz, [D$_6$]DMSO, 25°C): δ = 174.4 (CO), 149.4 (ArC), 134.7 (ArC), 132.5 (ArCH), 130.7 (ArCH), 129.7 (ArCH), 121.7 (ArC), 113.6 (ArCH), 108.1 (ArC), 91.1 (C), 86.3 (C), 40.4 (CH$_2$); MS (CI+): m/z (%): 252 (100) [M+H$^+$], 251 (90) [M$^+$], 206 (50); HRMS (CI+): calculated for C$_{16}$H$_{14}$NO$_2$ 252.1030 [M+H$^+$], found 252.1025.
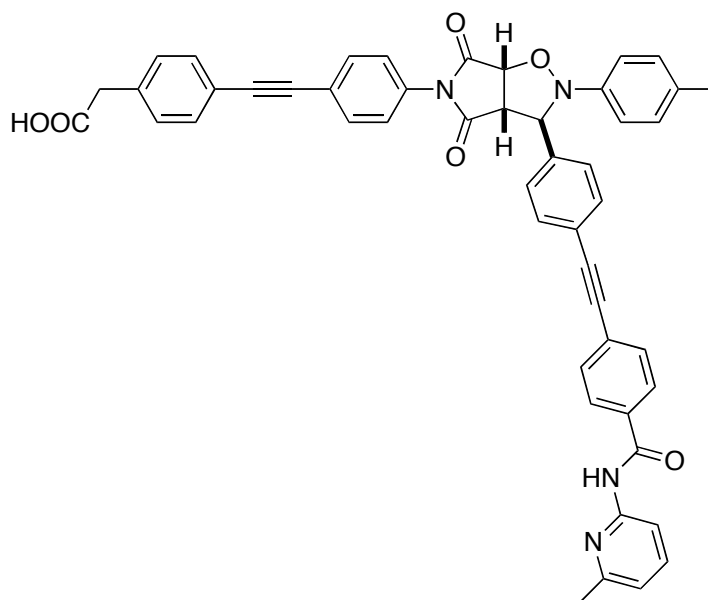
trans-166

2-(4-((4-(3-(4-((4-(6-methylpyridin-2-ylcarbamoyl)phenyl)ethynyl)phenyl)-4,6-dioxo-2-*p*-tolyldihydro-2*H*-pyrrolo[3,4-*d*]isoxazol-5(3*H*,6*H*,6a*H*)-yl)phenyl) ethynyl)phenyl) acetic acid

2-(4-((4-(2,5-dioxo-2,5-dihydro-1*H*-pyrrol-1-yl)phenyl)ethynyl)phenyl)acetic acid **158** (0.13 g, 0.4 mmol) and (*Z*)-4-methyl-*N*-(4-((4-(6-methylpyridin-2-ylcarbamoyl)phenyl) ethynyl)benzylidene)aniline oxide **132** (0.18 g, 0.4 mmol) were dissolved in chloroform (15 ml). The solution was left stirring in the dark at room temperature. The resulting precipitate was filtered and recrystallised from methanol to afford the desired product as a colourless solid (0.20 g, 63%). M.p. = 228.6-229.3 °C; (400.1 MHz, [D$_6$]DMSO, 25°C): δ = 12.41 (s, 1H; COOH), 10.82 (s, 1H; NH), 8.06 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 7.99 (d, ^3J(H,H) = 8 Hz, 1H; ArH), 7.72 (t, ^3J(H,H) = 8 Hz, 1H; ArH), 7.68-7.62 (m, 5H; ArH), 7.56 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 7.51 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 7.32 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 7.11-7.02 (m, 6H; ArH), 6.69 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 5.89 (s, 1H; CH), 5.41 (d, ^3J(H,H) = 7 Hz, 1H; CH), 4.14 (d, ^3J(H,H) = 7 Hz, 1H; CH), 3.62 (s, 2H; CH$_2$), 2.45 (s, 3H; CH$_3$), 2.22 (s, 3H; CH$_3$); ^{13}C NMR (100.6 MHz, [D$_6$]DMSO, 25°C): δ = 174.3 (CO), 174.1 (CO), 173.0 (CO), 172.3 (CO), 165.2 (ArC), 156.6 (ArC), 151.4 (ArC), 146.3 (ArC), 140.1 (ArC), 138.4 (ArCH), 136.2 (ArC), 133.9 (ArC), 131.7 (ArCH), 131.4 (ArC), 131.3 (ArCH), 131.2 (ArCH), 129.9 (ArCH), 129.4 (ArCH), 128.4 (ArCH), 127.7 (ArCH), 126.6 (ArCH), 125.5 (ArC), 122.6 (ArC), 121.3 (ArC), 120.4 (ArCH), 120.1 (ArC), 119.1 (ArCH), 114.6 (ArCH), 111.7 (ArCH), 91.2 (C), 90.3 (C), 89.0 (C), 88.2 (C), 77.65 (CH), 68.3 (CH), 56.3

(CH), 40.4 (CH$_2$), 23.5 (CH$_3$), 20.0 (CH$_3$); MS (ES+): *m/z* (%): 777 (100) [M+H$^+$], 779 (10); HRMS (ES+): calculated for C$_{49}$H$_{37}$N$_4$O$_6$ [M+H$^+$] 777.2720, found 777.2713.

trans-168

2-(4-((4-(3-(4-(6-methylpyridin-2-ylcarbamoyl)phenyl)-4,6-dioxo-2-*p*-tolyldihydro-2*H*-pyrrolo[3,4-*d*]isoxazol-5(3*H*,6*H*,6a*H*)-yl)phenyl)ethynyl)phenyl)acetic acid

2-(4-((4-(2,5-dioxo-2,5-dihydro-1*H*-pyrrol-1-yl)phenyl)ethynyl)phenyl)acetic acid **158** (0.13 g, 0.4 mmol) and (*Z*)-4-methyl-*N*-(4-(6-methylpyridin-2-ylcarbamoyl) benzylidene) aniline oxide **124** (0.13 g, 0.4 mmol) were dissolved in chloroform (15 ml). The solution was left in the dark at room temperature. The resulting precipitate was filtered and recrystallised from ethyl acetate to afford the desired product as a colourless solid (0.20 g, 73%). M.p. = 214.2-215.5 °C; ^{1}H NMR (400.1 MHz, [D$_6$]DMSO, 25°C): δ = 2.44 (s, 1H; COOH), 10.74 (s, 1H; NH), 8.08 (d, ^{3}*J*(H,H) = 8 Hz, 2H; ArH), 8.02 (d, ^{3}*J*(H,H) = 8 Hz, 1H; ArH), 7.73-7.66 (m, 3H; ArH), 7.56 (d, ^{3}*J*(H,H) = 8 Hz, 2H; ArH), 7.51 (d, ^{3}*J*(H,H) = 8 Hz, 2H; ArH), 7.32 (d, ^{3}*J*(H,H) = 8 Hz, 2H; ArH), 7.14-7.01 (m, 5H; ArH), 6.68 (d, ^{3}*J*(H,H) = 8 Hz, 2H; ArH), 5.95 (s, 1H; CH), 5.42 (d, ^{3}*J*(H,H) = 7 Hz, 1H; CH), 4.16 (d, ^{3}*J*(H,H) = 7 Hz, 1H; CH), 3.63 (s, 2H; CH$_2$), 2.45 (s, 3H; CH$_3$), 2.22 (s, 3H; CH$_3$); ^{13}C NMR (100.6 MHz, [D$_6$]DMSO, 25°C): δ = 174.1 (CO), 173.0 (CO), 172.3 (CO), 165.4 (CO), 156.6 (ArC), 151.5 (ArC), 146.5 (ArC), 142.8 (ArC), 138.3 (ArCH), 136.2 (ArC), 133.5 (ArC), 131.7 (ArC), 131.6 (ArCH), 131.4 (ArC), 131.3 (ArCH), 129.9 (ArCH), 129.4 (ArCH), 128.3 (ArCH), 127.0 (ArCH), 126.6 (ArCH), 122.6 (ArC), 120.1 (ArC), 119.1 (ArCH), 114.5 (ArCH), 111.7 (ArCH), 90.3 (C), 88.2 (C), 77.7 (CH), 68.3 (CH), 56.4 (CH), 40.4 (CH$_2$), 23.5 (CH$_3$), 20.0 (CH$_3$); MS (ES+): *m/z* (%): 677 (100) [M+H$^+$], 678 (25), 346 (10); HRMS (ES$^+$): calculated for C$_{41}$H$_{33}$N$_4$O$_6$ [M+H$^+$] 677.2396, found 677.2400.
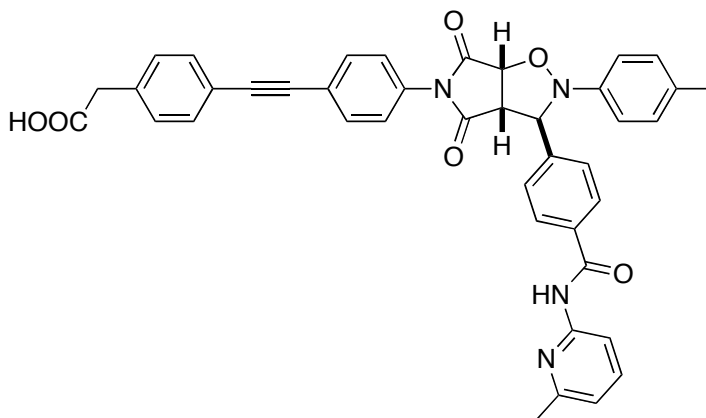
***trans*-170**

2-(4-(3-(4-((4-(6-methylpyridin-2-ylcarbamoyl)phenyl)ethynyl)phenyl)-4,6-dioxo-2-*p*-tolyldihydro-2*H*-pyrrolo[3,4-*d*]isoxazol-5(3*H*,6*H*,6a*H*)-yl)phenyl)acetic acid

2-(4-(2,5-dioxo-2,5-dihydro-1*H*-pyrrol-1-yl)phenyl)acetic acid **112** (0.09 g, 0.4 mmol) and(*Z*)-4-methyl-*N*-(4-((4-(6-methylpyridin-2-ylcarbamoyl)phenyl)ethynyl)benzylidene) aniline oxide **132** (0.18 g, 0.4 mmol) 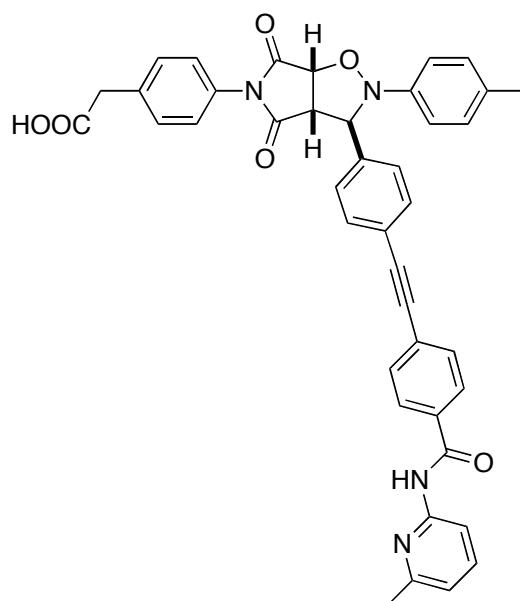and 2-(4-((4-(3-(4-(6-methylpyridin-2-ylcarbamoyl)phenyl)-4,6-dioxo-2-*p*-tolyldihydro-2*H*-pyrrolo[3,4-*d*]isoxazol-5(3*H*,6*H*,6a*H*)-yl)phenyl)ethynyl) phenyl)acetic acid **168** (0.02 g, 0.03 mmol) were dissolved in chloroform (15 ml). The solution was left stirring in the dark at room temperature. The resulting precipitate was filtered and recrystallised from methanol to afford the desired product as a colourless solid (0.18 g, 66%). M.p. = 228.8-228.9 °C; (400.1 MHz, [D$_6$]DMSO, 25°C): δ = 12.42 (s, 1H; COOH), 10.81 (s, 1H; NH), 8.07 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 8.01 (d, ^3J(H,H) = 8 Hz, 1H; ArH), 7.72 (t, ^3J(H,H) = 8 Hz, 1H; ArH), 7.68-7.62 (m, 6H; ArH), 7.25 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 7.12-7.01 (m, 5H; ArH), 6.53 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 5.89 (s, 1H; CH), 5.38 (d, ^3J(H,H) = 7 Hz, 1H; CH), 4.13 (d, ^3J(H,H) = 7 Hz, 1H; CH), 3.59 (s, 2H; CH$_2$), 2.45 (s, 3H; CH$_3$), 2.22 (s, 3H; CH$_3$); ^{13}C NMR (100.6 MHz, [D$_6$]DMSO, 25°C): δ = 174.4 (CO), 173.2 (CO), 172.4 (CO), 165.2 (CO), 156.6 (ArC), 151.4 (ArC), 146.5 (ArC), 140.2 (ArC), 138.4 (ArCH), 135.7 (ArC), 133.9 (ArC), 131.7 (ArCH), 131.5 (ArC), 131.2 (ArCH), 129.9 (ArC), 129.8 (ArCH), 129.4 (ArCH), 128.4 (ArCH), 127.6 (ArCH), 126.2 (ArCH), 125.5 (ArC), 121.2 (ArC), 119.1 (ArCH), 114.5 (ArCH), 111.7 (ArCH), 91.25 (C), 89.0 (C),

77.7 (CH), 68.3 (CH), 56.3 (CH), 40.1 (CH$_2$), 23.5 (CH$_3$), 20.0 (CH$_3$); MS (ES+): *m/z* (%): 677 (20) [M+H$^+$], 446 (100); HRMS (ES$^+$): calculated for C$_{41}$H$_{33}$N$_4$O$_6$ [M+H$^+$] 677.2397, found 677.2400.
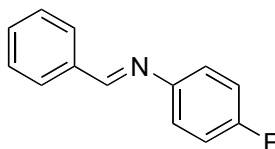
173

(*E*)-*N*-benzylidene-4-fluoroaniline[254]

Benzaldehyde **171** (0.50 g, 4.7 mmol) and 4-fluoroaniline **172** (0.52 g, 4.7 mmol) were dissolved in minimal amount of ethanol and left in the absence of light for 16 h. The resulting precipitate was filtered affording the desired compound as a colourless crystalline solid (0.92 g, 98%). M.p. = 56.8-57.2 °C (lit.[254] 56.8 °C); ^{1}H NMR (400.1 MHz, CDCl$_3$, 25°C): δ = 8.47 (s, 1H; CH), 7.95-7.92 (m, 2H; ArH), 7.50-7.48 (m, 3H; ArH), 7.25 (d, ^{3}*J*(H,H) = 9 Hz, 2H; ArH) 7.12 (d, ^{3}*J*(H,H) = 9 Hz, 2H; ArH); ^{13}C NMR (100.6 MHz, CDCl$_3$, 25°C): δ = 162.9 (ArC), 160.2 (CH), 159.7 (ArC), 136.1 (ArC), 131.5 (ArCH), 128.9 (ArCH), 128.8 (ArCH), 122.3 (d, ^{3}*J*(C,F) = 8 Hz; ArCH), 115.9 (d, ^{2}*J*(C,F) = 23 Hz; ArCH); ^{19}F NMR (376.5 MHz, CDCl$_3$, 25°C): δ = −117.8; MS (ES+): *m/z* (%): 200 (100) [M+H$^+$]; HRMS (ES+): calculated for C$_{13}$H$_{11}$NF [M+H$^+$] 200.0874 found 200.0876.
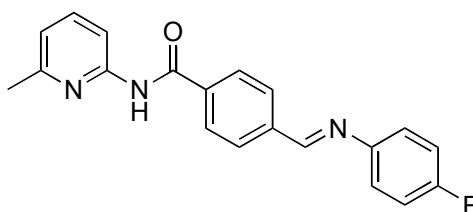
174

(*E*)-4-((4-fluorophenylimino)methyl)-*N*-(6-methylpyridin-2-yl)benzamide

N-(6-methyl-2 pyridyl)-4-formyl benzamide **122** (0.50 g, 2.1 mmol) and 4-fluoroaniline **172** (0.23 g, 2.1 mmol) were dissolved in minimal amount of ethanol and left in the absence of light for 16 h. The resulting precipitate was filtered affording the desired compound as a yellow solid (0.52 g, 76 %). M.p. = 154.8-155.4 °C; ^{1}H NMR

(400.1 MHz, CDCl$_3$, 25°C): δ = 8.62 (s, 1H; NH), 8.54 (s, 1H; CH), 8.22 (d, ^3J(H,H) = 8 Hz, 1H; ArH); 8.07-8.02 (m, 4H; ArH), 7.69 (t, ^3J(H,H) = 8 Hz, 1H; ArH), 7.29-7.25 (m, 2H; ArH), 7.16-7.10 (m, 2H; ArH), 6.90 (d, ^3J(H,H) = 8 Hz, 1H; ArH), 2.50 (s, 3H; CH$_3$); ^{13}C NMR (100.6 MHz, CDCl$_3$, 25°C): δ = 164.9 (CO), 163.2 (ArC), 160.0 (ArC), 158.6 (CH), 157.0 (ArC), 150.6 (ArC), 139.6 (ArC), 138.9 (ArCH),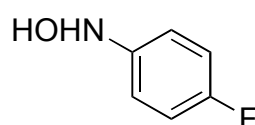 136.6 (ArC), 129.1 (ArCH), 127.7 (ArCH), 122.5 (d, ^3J(C,F) = 8 Hz; ArCH), 119.7 (ArCH), 116.0, (d, ^2J(C,F) = 23 Hz; ArCH) 111.1 (ArCH), 24.0 (CH$_3$); ^{19}F NMR (376.5 MHz, CDCl$_3$, 25°C): δ = −116.7 (ArF); MS (ES+): m/z (%): 356 (100) [M+Na$^+$]; HRMS (ES+) calculated for C$_{20}$H$_{16}$N$_3$OFNa [M+Na$^+$] 356.1174, found 356.1175.

175

N-(4-fluorophenyl)hydroxylamine[255]

4-fluoronitrobenzene (2.00 g, 14.0 mmol) was dissolved in ethanol/water mixture (20 ml, 1:1) and placed in a water bath. Bismuth chloride (0.90 g, 2.8 mmol) was added followed by potassium borohydride (1.10 g, 21.0 mmol) over a 5 min period. The solution was allowed to stir for 10 min before the excess borohydride was neutralised to pH 7 with 0.5 M HCl. The product was extracted into ether (3 x 100 ml). The combined organic leyers were dried over magnesium sulfate, filtered and the solvent was removed *in vacuo* to yield a crude product the was purified by recrystallisation in ethyl acetate/hexane affording the desired product as a brown powder (1.34 g, 74%). M.p. = ^{1}H NMR (400.1 MHz, CDCl$_3$, 25°C): δ = 6.93-6.86 (m, 4H; ArH), 6.05 (br, 2H; NHOH); ^{13}C NMR (100.6 MHz, CDCl$_3$, 25°C): δ = 158.7 (d, ^1J(C,F) = 240 Hz; ArC), 145.5 (d, ^4J(C,F) = 2 Hz; ArC), 116.4 (d, ^3J(C,F) = 8 Hz; ArCH), 115.6 (d, ^2J(C,F) = 23 Hz; ArCH); ^{19}F NMR (376.5 MHz, CDCl$_3$, 25°C): δ = −121.8.
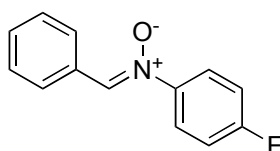
176

(*E*)-*N*-benzylidene-4-fluoroaniline oxide[256]

Benzaldehyde **171** (0.50 g, 4.7 mmol) and *N*-(4-fluorophenyl)hydroxylamine **175** (0.60 g, 4.7 mmol) were dissolved in minimal amount of ethanol and left in the absence of light for 16 h. The resulting precipitate was filtered affording the desired compound as a colourless crystalline solid (0.88 g, 88 %). M.p. = 169.5-170.3 °C (lit.[257] 172 °C); ^{1}H NMR (400.1 MHz, CDCl$_3$, 25°C): δ = 8.31-8.21 (m, 2H; ArH), 7.80 (s, 1H; CH), 7.71 (d, ^{3}J(H,H) = 9 Hz, 2H; ArH), 7.41-7.39 (m, 3H; ArH), 7.17 (d, ^{3}J(H,H) = 9 Hz, 2H; ArH); ^{13}C NMR (100.6 MHz, CDCl$_3$, 25°C): δ = 164.8(ArC), 161.4 (ArC), 134.6 (CH), 131.1 (ArCH), 130.6 (ArC), 129.1 (ArCH), 128.7 (ArCH), 123.7 (d, ^{3}J(C,F) = 9 Hz; ArCH), 116.1 (d, ^{2}J(C,F) = 23 Hz; ArCH); ^{19}F NMR (376.5 MHz, CDCl$_3$, 25°C): δ = −110.4; MS (ES+): *m/z* (%): 234 (100) [M+Na$^+$], 453 (20); HRMS (ES+): calculated for C$_{13}$H$_{10}$NOFNa [M+Na$^+$] 238.0647, found 238.0644.
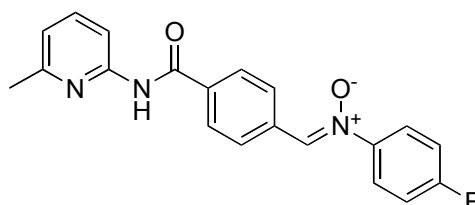
177

(*Z*)-4-fluoro-*N*-(4-(6-methylpyridin-2-ylcarbamoyl)benzylidine)aniline oxide

N-(6-methyl-2-pyridyl)-4-formyl benzamide **122** (0.57 g, 2.4 mmol) and *N*-(4-fluorophenyl)hydroxylamine **175** (0.30 g, 2.4 mmol) were dissolved together in ethanol (10 ml) and left in the absence of light to react for 16 h. The resulting precipitate was filtered and washed with cold ethanol to yield the desired product as a colourless solid (0.54 g, 64%). M.p. = 217.2-218.3 °C; ^{1}H NMR (400.1 MHz, CDCl$_3$, 25°C): δ = 8.70 (s, 1H; NH), 8.37 (d, ^{3}J(H,H) = 8 Hz, 2H; ArH), 8.12 (d, ^{3}J(H,H) = 8 Hz, 1H; ArH); 7.96 (d, ^{3}J(H,H) = 8 Hz, 2H; ArH), 7.54 (s, 1H; CH), 7.72 (d, ^{3}J(H,H) = 8 Hz, 2H; ArH), 7.58 (t, ^{3}J(H,H) = 8 Hz, 1H; ArH), 7.08 (d, ^{3}J(H,H) = 8 Hz, 2H; ArH), 6.87 (d, ^{3}J(H,H) = 8 Hz, 1H; ArH), 2.36 (s, 3H; CH$_3$); ^{13}C NMR (100.6 MHz, CDCl$_3$,

25°C): δ = 165.2 (CO), 165.0 (ArC), 161.6 (ArC), 157.0 (ArC), 150.7 (ArC), 138.8 (ArCH), 135.8 (ArC), 133.7 (ArC), 133.3 (CH), 129.0 (ArC),127.6 (ArC), 123.7 (d, ^3J(C,F) = 9 Hz; ArCH), 119.7 (ArCH), 116.2 (d, ^2J(C,F) = 23 Hz; ArCH), 111.1 (ArCH), 24.0 (CH$_3$); ^{19}F NMR (376.5 MHz, CDCl$_3$, 25°C): δ = −109.6 (ArF); MS (ES+): m/z (%): 372 (100) [M+Na$^+$]; HRMS (ES+): calculated for C$_{20}$H$_{16}$N$_3$O$_2$FNa [M+Na$^+$] 372.1124, found 372.1114.
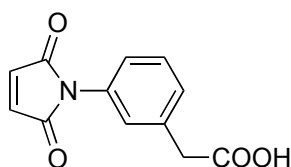
179

2-(3-(2,5-dioxo-2,5-dihydro-1*H*-pyrrol-1-yl)phenyl)acetic acid

3-aminophenylacetic acid **178** (5.00 g, 33.1 mmol) and maleic anhydride **126** (3.20 g, 32.7 mmol) was dissolved in glacial acetic acid (200 ml). The solution was stirred at room temperature under nitrogen for 12 h, followed by heating to reflux for further 5 h. The solvent was removed *in vacuo* and the residue was purified on silica gel chromatography column (hexane:ethyl acetate, 3:1) affording the desired product as a yellow solid which was recrystallised from DCM (3.53 g, 46%). M.p. = 114.3-115.0 °C; ^{1}H NMR (300.1 MHz, [D$_6$]DMSO, 25°C): δ = 12.46 (s, 1H; COOH), 7.46-7.41 (m, 1H; ArH), 7.32-7.28 (m, 1H; ArH), 7.24-7.20 (m, 2H; ArH), 7.16 (s, 2H; 2xCH), 3.63 (s, 2H; CH$_2$); ^{13}C NMR (75.5 MHz, [D$_6$]DMSO, 25°C): δ = 172.8 (CO), 170.3 (CO), 136.3 (ArC), 135.0 (CH), 131.8 (ArC), 129.3 (ArCH), 129.1(ArCH), 128.0 (ArCH), 125.5 (ArCH), 40.7 (CH$_2$); MS (CI+): *m/z* (%): 231 (15) [M$^+$], 186 (100), 214 (15); HRMS (CI+): calculated for C$_{12}$H$_9$NO$_4$ [M$^+$] 231.0532, found 231.0540.
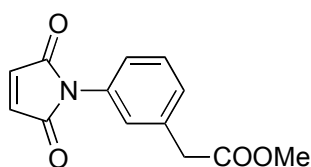
180

Methyl 2-(3-(2,5-dioxo-2,5-dihydro-1H-pyrrol-1-yl)phenyl)acetate

2-(3-(2,5-dioxo-2,5-dihydro-1*H*-pyrrol-1-yl)phenyl)acetic acid **179** (1.00 g, 4.3 mmol)

was added to a solution of NaHSO$_4$·SiO$_2$ (100 mg/mmol) in methanol (20 ml) and left to stir under nitrogen atmosphere for 30 h. The suspension was filtered to separate the catalyst and filtrate was evaporated under reduced pressure to yield a yellow residue. The crude product was purified by silica gel flash column chromatography (DCM) affording a yellow solid (0.54 g, 51%). M.p. = 58.3-59.0 °C; ^{1}H NMR (300.1 MHz, CDCl$_3$, 25°C): δ = 7.42 (t, ^3J(H,H) = 8 Hz, 1H; ArH), 7.27-7.38 (m, 3H; ArH), 6.82 (s, 2H; 2xCH), 3.69 (s, 3H; CH$_3$), 3.68 (s, 2H; CH$_2$); ^{13}C NMR (75.5 MHz, CDCl$_3$, 25°C): δ = 171.9 (CO), 169.8 (CO), 135.5 (ArC), 134.6 (CH), 131.8 (ArC), 129.7 (ArCH), 129.3 (ArCH), 127.3 (ArCH), 125.2 (ArCH), 52.6 (CH$_3$), 41.3 (CH$_2$); MS (ES+): m/z (%): 268 (100) [M+Na$^+$]; HRMS (CI+): calculated for C$_{13}$H$_{12}$NO$_4$ [M+H$^+$] 246.0766, found 246.0771.
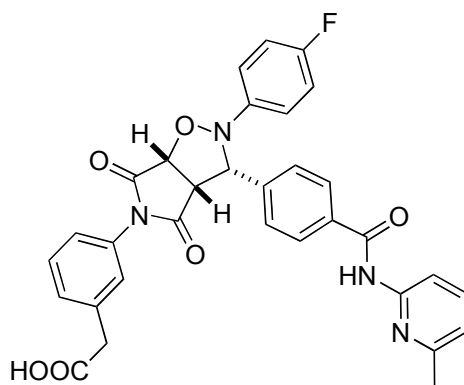
cis-182

2-(3-(2-(4-fluorophenyl)-3-(4-(6-methylpyridin-2-ylcarbamoyl)phenyl)-4,6-dioxodihydro-2*H*-pyrrolo[3,4-*d*]isoxazol-5(3*H*,6*H*,6a*H*)-yl)phenyl)acetic acid

2-(3-(2,5-dioxo-2,5-dihydro-1*H*-pyrrol-1-yl)phenyl)acetic acid **179** (0.13 g, 0.6 mmol) and (*Z*)-4-fluoro-*N*-(4-(6-methylpyridin-2-ylcarbamoyl)benzylidene)aniline oxide **177** (0.20 g, 0.6 mmol) were dissolved in chloroform (10 ml) and left at 0 °C for 16 h. The resulting precipitate was filtered and recrystallised from ethanol affording the desired product as a colourless solid (0.30 g, 90%). M.p. = 214.0-214.9 °C; ^{1}H NMR (400.1 MHz, CDCl$_3$, 25°C): δ = 11.23 (s, 1H; NH), 8.18 (d, ^3J(H,H) = 8 Hz, 1H; ArH), 7.94 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 7.66 (t, ^3J(H,H) = 8 Hz, 1H; ArH), 7.55 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 7.42 (s, 1H; ArH), 7.29 (t, ^3J(H,H) = 8 Hz, 1H; ArH), 7.15 (d, ^3J(H,H) = 8 Hz, 1H; ArH), 7.08 (d, ^3J(H,H) = 8 Hz, 1H; ArH), 7.02 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 6.89-6.83 (m, 3H), 5.21 (d, ^3J(H,H) = 8 Hz, 1H; CH), 4.83 (d, ^3J(H,H) = 10 Hz, 1H; CH), 4.03 (dd, ^3J(H,H) = 8 Hz and 10 Hz, 1H; CH), 3.70 (s, 2H; CH$_2$), 2.40 (s, 3H; CH$_3$); ^{13}C

NMR (100.6 MHz, CDCl$_3$, 25°C): δ = 176.2 (CO), 173.8 (CO), 170.7 (CO), 167.0 (CO), 161.2 (ArC), 158.8 (ArC), 151.6 (ArC), 142.9 (ArC), 140.4 (ArCH), 140.4 (ArCH), 138.5 (ArC), 135.5 (ArC), 135.0 (ArC), 131.2 (ArC), 129.8 (ArCH), 128.9 (ArCH), 127.0 (ArCH), 126.4 (ArCH), 123.9 (ArCH), 120.7 (d, ^3J(C,F) = 9 Hz; ArCH), 119.8 (ArCH), 115.8 (d, ^2J(C,F) = 23 Hz; ArCH), 113.1 (ArCH), 77.2 (CH), 70.6 (CH), 54.6 (CH), 38.8 (CH$_2$), 22.1(CH$_3$); ^{19}F NMR (376.5 MHz, CDCl$_3$, 25°C): δ = −117.7 (ArF); MS (ES+): *m/z* (%): 603 (100) [M+Na$^+$]; HRMS (ES+): calculated for C$_{32}$H$_{25}$N$_4$O$_6$NaF [M+Na$^+$] 603.1643, found 603.1656.
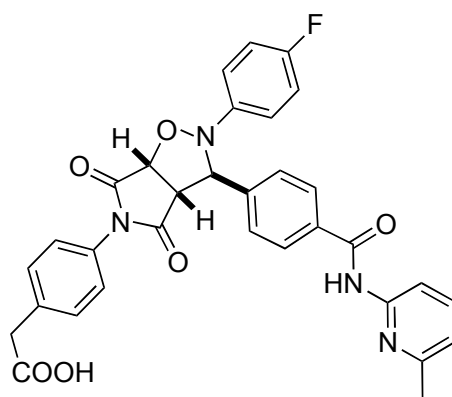

***trans*-188**

2-(4-(2-(4-fluorophenyl)-3-(4-(6-methylpyridin-2-ylcarbamoyl)phenyl)-4,6-dioxodihydro-2*H*-pyrrolo[3,4-*d*]isoxazol-5(3*H*,6*H*,6a*H*)-yl)phenyl)acetic acid

2-(4-(2,5-dioxo-2,5-dihydro-1*H*-pyrrol-1-yl)phenyl)acetic acid **112** (0.13 g, 0.6 mmol) and (*Z*)-4-fluoro-*N*-(4-(6-methylpyridin-2-ylcarbamoyl)benzylidine)aniline oxide **177** (0.20 g, 0.6 mmol) were dissolved in chloroform (10 ml) and left at 0°C for 16 h. The resulting precipitate was filtered and recrystallised from ethanol affording the desired product as a colourless solid (0.30 g, 90%). M.p. = 203.0-205.9 °C; ^{1}H NMR (400.1 MHz, [D$_6$]DMSO, 25°C): δ = 12.42 (s, 1H; COOH), 10.72 (s, 1H; NH), 8.05 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 8.00 (d, ^3J(H,H) = 8 Hz, 1H; ArH), 7.71 (t, ^3J(H,H)=8 Hz, 1H; ArH), 7.64 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 7.29 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 7.29 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 7.23 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 7.10 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 7.02 (d, ^3J(H,H) = 8 Hz, 1H; ArH), 6.68 (d, ^3J(H,H) = 8 Hz, 2H; ArH), 5.94 (s, 1H; CH), 5.44 (d, ^3J(H,H) = 7 Hz, 1H; CH), 4.16 (d, ^3J(H,H) = 7 Hz, 1H; CH), 3.60 (s, 2H; CH$_2$), 3.33 (s, 3H; CH$_3$); ^{13}C NMR (100.6 MHz, [D$_6$]DMSO, 25°C): δ = 174.3 (CO), 173.2 (CO), 172.4 (CO), 165.4 (CO), 159.1 (ArC), 156.7 (ArC), 156.5

(ArC), 151.5 (ArC), 145.0 (ArC), 142.5 (ArC), 138.4 (ArCH), 135.8 (ArC), 133.5 (ArC), 129.9 (ArCH), 128.3 (ArCH), 127.1 (ArCH), 125.9 (ArCH), 119.0 (ArCH), 116.3 (d, ^3J(C,F) = 8 Hz; ArCH), 115.6 (d, ^2J(C,F) = 23 Hz; ArCH), 111.6 (ArCH), 77.7 (CH), 68.4 (CH), 56.3 (CH), 40.0 (CH$_2$), 23.5 (CH$_3$); ^{19}F NMR (376.5 MHz, CDCl$_3$, 25°C): δ = −121.6 (ArF); MS (ES+): *m/z* (%): 581 (100) [M+H$^+$]; HRMS (ES+): calculated for C$_{35}$H$_{24}$N$_5$O$_3$F [M+H$^+$] 581.1862, found 581.1863.
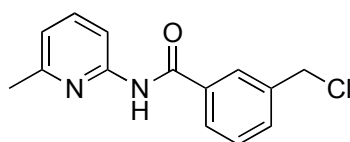
195

3-(chloromethyl)-*N*-(6-methylpyridin-2-yl)benzamide

2-amino-6-picoline **119** (5.15 g, 48.0 mmol) dissolved in DCM (50 ml) was added dropwise to 3-(chloromethyl)benzoyl chloride **194** (3.00 g, 16.0 mmol) at 0 °C. The solution was allowed to warm to room temperature and was left stirring for 16 h. The solvent was removed *in vacuo* and the residue was purified by silica gel flash column chromatography (hexane : ethyl acetate, 3:2) to afford the desired compound as a colourless solid (4.05 g, 98%). M.p. = 66.2-67.5 °C; ^{1}H NMR (300.1 MHz, CDCl$_3$, 25°C): δ = 8.53 (s, 1H; NH), 8.17 (d, ^3J(H,H) = 8 Hz, 1H; ArCH), 7.94 (s, 1H; ArCH), 7.86 (d, ^3J(H,H) = 8 Hz, 1H; ArCH), 7.64 (dd, ^3J(H,H) = 8 Hz and 7 Hz, 1H; ArCH), 7.54 (d, ^3J(H,H) = 8 Hz, 1H; ArCH), 7.51 (dd, ^3J(H,H) = 8 Hz and 8 Hz, 1H; ArCH), 6.93 (d, ^3J(H,H) = 7 Hz, 1H; ArCH), 4.63 (s, 2H; CH$_2$), 2.47 (s, 3H; CH$_3$); ^{13}C NMR (75.5 MHz, CDCl$_3$, 25°C): δ = 165.2 (CO), 156.9 (ArC), 150.8 (ArC), 138.9 (ArCH), 138.3 (ArC), 134.9 (ArC), 132.2 (ArCH), 129.3 (ArCH), 127.5 (ArCH), 127.2 (ArCH), 119.6 (ArCH), 111.2 (ArCH), 45.5 (CH$_2$), 23.9 (CH$_3$); MS (EI): *m/z* (%): 262 (15) [M$^+$(^{37}Cl)], 260 [M$^+$(^{35}Cl)], 232 (39), 231 (100), 153 (97), 125 (97), 89 (62); HRMS (EI): calculated for C$_{14}$H$_{13}$N$_2$OCl [M$^+$] 260.0716, found 260.0719.
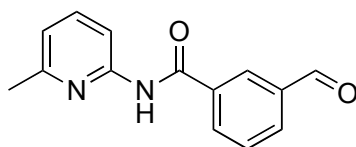
196

3-formyl-*N*-(6-methylpyridin-2-yl)benzamide

3-(chloromethyl)-*N*-(6-methylpyridin-2-yl)benzamide **195** (3.90 g, 15.0 mmol) was dissolved with hexamethylene tetramine (6.45 g, 45.0 mmol) in ethanol/water (18 ml, 1:1) and heated to reflux for 16 h. After this time 12 M HCl (3 ml) was added and the reaction mixture was refluxed for further 30 min. After cooling to room temperature the mixture was diluted with water (100 ml) and extracted with DCM (300 ml). The combined organic extracts were washed with water and dried over $MgSO_4$ and the solvent was removed *in vacuo*. The residue was purified by silica gel flash column chromatography (hexane : ethyl acetate, 3:2) to afford the desired compound as a colourless solid (2.20 g, 61%). M.p. = 126.4-128.7 °C; ^{1}H NMR (300.1 MHz, CDCl$_3$, 25°C): δ = 10.10 (s, 1H; CHO), 8.70 (s, 1H; NH), 8.43 (s, 1H; ArCH), 8.24-8.17 (m, 2H; ArCH), 8.10-8.07 (m, 1H; ArCH), 7.73-7.65 (m, 2H; ArCH), 6.97 (d, ^3J(H,H) = 8 Hz, 1H; ArCH), 2.49 (s, 3H; CH$_3$); ^{13}C NMR (75.5 MHz, CDCl$_3$, 25°C): δ = 191.5 (CHO), 164.3 (CO), 156.8 (ArC), 150.4 (ArC), 139.0 (ArCH), 136.5 (ArC), 135.2 (ArC), 133.0 (ArCH), 132.6 (ArCH), 130.6 (ArCH), 128.3 (ArCH), 119.7 (ArCH), 111.1 (ArCH), 23.7 (CH$_3$); MS (CI): *m/z* (%): 241 (100) [M+H$^+$], 211 (5); HRMS (EI): calculated for C$_{14}$H$_{12}$N$_2$O$_2$ [M$^+$] 240.0899, found 240.0888.

8. References

1. S.H. Strogatz, *Nature* **2001**, *410*, 268-276.

2. M. E. J. Newman, *SIAM Review* **2003**, *45*, 167-256.

3. T. Vicsek, *Nature* **2002**, *418*, 131.

4. D. Bray, *Science* **2003**, *301*, 1864-1865,

5. D. J. Watts, S. H. Strogatz, *Nature* **1998**, *393*, 440-442.

6. M. Kochen (Ed.), *The Small World* **1989** Ablex, Norwood, NJ.

7. J. Travers, S. Milgram, *Sociometry* **1969**, *32*, 425-443.

8. A.-L. Barabasi, R. Albert, *Science* **1999**, *286*, 509-511.

9. M. Girvan, M. E. J. Newman, *Proc. Natl. Acad. Sci. U.S.A.* **2002,** *99*, 7821–7826.

10. E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, A.-L. Barabasi, *Science* **2002**, *297*, 1551-1555.

11. G. Palla, I. Derenyi, I. Farkas, T. Vicsek, *Nature* **2005**, *435*, 814-818.

12. M. Rosvall, C. T. Bergstrom, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 1118-1123.

13. J.-D. J. Han, N. Bertin, T. Hao, D. S. Goldberg, G. F. Berriz, L. V. Zhang, D. Dupuy, A. J. M. Walhout, M. E. Cusick, F. P. Roth, M. Vidal, *Nature* **2004**, *430*, 88-93.

14. A. Eschenmoser, *Tetrahedron* **2007**, *63*, 12821-12844.

15. P.A. Lindahl, *Orig. Life Evol. Biosph.* **2004**, *34*, 371-389.

16. B. M. Rode, D. Fitz, T. Jakschitz, *Chem. Biodiv.* **2007**, *4*, 2674-2702.

17. I. Fry, *The Emergence of Life on Earth: A Scientific Overview* **2000**, Rutgers University Press, New Brunswick, NJ.

18. S. Kauffman, *Orig. Life Evol. Biosph.* **2007**, *37*, 315-322.

19. G. F. Joyce, in D. W. Deamer and G. R. Fleischaker (eds.), the foreword of 'Origins of Life: The Central Concepts' **1994**, Jones and Bartlett, Boston. The original reference is a 1992 internal NASA document entitled 'Exobiology: Discipline Science Plan'.

20. N. Horowitz, S. Miller, in L. Zechmeister (ed.), *Progress in the Chemistry of Natural Products,* **1962**, *20*, 4230-4590.

21. P. L. Luisi, *Orig. Life. Evol. Biosph.* **1998**, *28*, 613-622.

22. A. I. Oparin, *The Origin of Life* [Proiskhozdenie zhizny] **1924**, Moskovshii Rabochii, Moscow. English translation in: J. D. Bernal, *The Origin of Life* **1967**, Weidenfeld and Nicolson, London.

23. A. Eschenmoser, M. V. Kisakürek, *Helv. Chim. Acta.* **1996**, *79*, 1249-1259.

24. C. A. Hutchison, S. N. Peterson, S. R. Gill, R. T. Cline, O. White, C. M. Fraser, H. O. Smith, J. C. Venter*, Science* **1999**, *286*, 2165-2169.

25. P. L. Luisi, T. Oberholzer, A. Lazcano, *Helv. Chim. Acta* **2002**, *85*, 1759-1777.

26. D. Deamer, *Trends Biotechnol.* **2005,** *23*, 336-338.

27. S. Rasmussen, L. Chen, M. Nilsson, S. Abe, *Artif. Life* **2003**, *9*, 269-316.

28. R. F. Gesteland, T. R. Cech and J. F. Atkins (Eds), *The RNA World*, 2nd ed. **1999**, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

29. D. P. Bartel, P. J. Unrau *Trends Biochem. Sci.* **1999**, *24*, M9-M13.

30. F. H. Westheimer, *Nature* **1986**, *319*, 534-536.

31. K. Kruger, *Cell* **1982**, *31*, 147-157.

32. A. Luptak, J. W. Szostak, *Ribozymes and RNA Catalysis* **2008**, 123-133.

33. J. Demongeota, A. Moreirab, *J. Theor. Biol.* **2007**, *249*, 314-324.

34. T. R. Cech, *Bioscience Reports* **2004**, *24*, 363-385.

35. J. A. Doudna, T. R. Cech, *Nature* **2002**, *418*, 222-228.

36. W. G. Scott, *Ribozymes and RNA Catalysis* **2008**, 48-65.

37. K. Salehi-Ashtiani, J. W. Szostak, *Nature* **2001**, *414*, 82-84.

38. P. J. Unrau, D. P. Bartel, *Nature* **1998**, *395*, 260-263.

39. P. A. Lohse, J. W. Szostak, *Nature* **1996**, *381*, 442-444.

40. T. W. Wiegand, R. C. Janssen, B. E. Eaton, *Chem. Biol.* **1997**,*4,* 675-683.

41. G. Sengle, A. Eisenfuhr, P. S. Arora, J. S. Nowick, M. Famulok, *Chem. Biol.* **2001**, *8*, 459-473.

42. V. R. Jadhav, M. Yarus, *Biochemistry* **2002**, *41*, 723-729.

43. W. Gilbert, *Nature* **1986**, *319*, 618- 618.

44. G. F. Joyce, *Nature* **1989**, *338*, 217-224.

45. L. E. Orgel, *Orig. Life Evol. Biosph.* **2003**, *33*, 211-218.

46. P.-A. Monnard, *Orig. Life Evol. Biosph.* **2007**, *37*, 387-390.

47. G. F. Joyce, *Nature* **2002**, *418*, 214-221.

48. A. W. Schwartz, *Chem. Biodiv.* **2007**, *4*, 656-664.

49. D. Mller, S. Pitsch, A. Kittaka, E. Wagner, C. Wintner, A. Eschenmoser, *Helv. Chim. Acta* **1990**, *73*, 1410-1468.

50. A. Eschenmoser, *Science* **1999**, *284*, 2118-2124.

51. M. Beier, F. Reck, T. Wagner, R. Krishnmurthy, A. Eschenmoser, *Science* **1999**, *283*, 699-703.

52. M. Ferencic, G. Reddy, X. Wu, S. Guntha, J. Nandy, R. Krishnmurthy, A. Eschenmoser, *Chem. Biodivers.* **2004**, *1*, 939-979.

53. J. D. Watson, F. H. C. Crick, *Nature* **1953**, *171*, 964-967.

54. J. Oro, *Nature* **1961**, *191*, 1193-1194.

55. J. P. Ferris, P. C. Joshi, E. H. Edelson, J. G. Lawless, *J. Mol. Evol* **1978**, *11*, 293-311.

56. G. C. Mills, D. Kenyon, *Origins & Design* **1996**, *17*, 1-7.

57. R. Shapiro, *Orig. Life* **1984**, *14*, 565-570.

58. G. Ertem, *Orig. Life Evol. Biosph.* **2004**, *34*, 549-570.

59. A. G. Cairns-Smith, *Chem. Eur. J.* **2008**, *14*, 3830-3839.

60. J. P. Ferris, *Orig. Life Evol. Biosph.* **2002**, *32*, 311-332.

61. J. P. Ferris, *Phil. Trans. R. Soc. B* **2006**, *361*, 1777-1786.

62. G. F. Joyce, A. W. Schwartz, S. L. Miller, L. E. Orgel, *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 4398-4402.

63. A. Ross, *Orig. Life Evol. Biosph.* **2004**, *34*, 307-321.

64. R. Shapiro, *Q. Rev. Biol.* **2006**, *81*, 105-125.

65. F. AL Anet, *Curr. Opin. Chem. Biol.* **2004**, *8*, 654-659.

66. F. Dyson, *Origins of Life. Second Edition*. **1999**, Cambridge University Press, Cambridge.

67. M. Eiger, P. Schuster, *The Hypercycle: A Principle of Natural Self-Organization* **1979**, Springer-Verlag, NY.

68. J. M. Smith, E. Szathmáry, *The Major Transitions in Evolution* **1995**, Oxford University Press, Oxford.

69. R. Shapiro, *Sci. Am.* **2007**, June, 47-53.

70. D. Segre, D. Lancet, *EMBO Reports* **2000**, *1*, 217-222.

71. G. Wächtershäuser, *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 200-204.

72. R. V. Sole, A. Munteanu, C. Rodriguez-Caso, J. Macia, *Phil. Trans. R. Soc. B* **2007**, *362*, 1727-1739.

73. D. Deamer, J. Dworkin, S. Sandford, M. Bernstein, L. Allamandola, *Astrobiology* **2002**, *2*, 371-381.

74. P. Pohorille, K. Schweighofer, M. A. Wilson, *Astrobiology* **2005**, *5*, 1-17.

75. E. Smith, H. J. Morowitz, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*,168-173.

76. I. A. Chen, *Science* **2006**, *314*, 1558-1559.

77. E. Szathmary, M. Santos, C. Fernando, *Top. Curr. Chem.* **2005**, *259*, 167-211.

78. J. W. Szostak, D. P. Bartel, P. L. Luisi, *Nature* **2001**, *409*, 387-390.

79. E. H. Ekland, J. W. Szostak, D. P. Bartel, *Science* **1995**, *269*, 364-370.

80. E. H. Ekland, D. P. Bartel, *Nucleic Acids Res.* **1995**, *23*, 3231-3238.

81. N. H. Bergman,W. K. Johnston, D. P. Bartel, *Biochemistry* **2000***, 39*, 3115-3123.

82. J. D. Swalen, D. L. Allara, J. D. Andrade, E. A. Chandross, S. Garoff, J. Israelachvili, T. J. McCarthy, R. Murray, R. F. Pease, J. F. Rabolt, K. J. Wynne, H. Yu, *Langmuir* **1987**, *3*, 932-950.

83. P. A. Bachmann, P. L. Luisi, J. Lang, *Nature* **1992**, *357*, 57-59.

84. H. Fellermann, R. V. Solé, *Phil. Trans. R. Soc. B* **2007**, *362*, 1803-1811.

85. M. M. Hanczyc, J.W. Szostak, *Curr. Opin. Chem. Biol.* **2004**, *8*, 660-664.

86. M. M. Hanczyc, S. M. Fujikawa, J. W. Szostak, *Science* **2003**, *302*, 618-622.

87. P. Walde, R. Wick, M. Fresta, A. Mangone, P. L. Luisi, *J. Am. Chem. Soc.* **1994**, *116*, 11649-11654.

88. I. A. Chen, R. W. Roberts, J. W. Szostak, *Science* **2004**, *305*, 1474-1476.

89. V. Balzani, *Small* **2005**, *1*, 278-283.

90. S. J. Sowerby, N. G. Holm, G. B. Petersen, *BioSystems* **2001**, *61*, 69-78.

91. K. E. Drexler, *Engines of Creation* **1987**, Anchor Books, New York, NY.

92. W. S. Bainbridge, *J. Nanopart. Res.* **2002**, *4*, 561-570.

93. *International Technology Roadmap for Semiconductors (ITRS)* **2007**, available at http://www.itrs.net (accessed May 2008).

94. S. E. Thompson, S. Parthasarathy, *Materials Today* **2006**, *9*, 20-25.

95. R. P. Feynman, *Eng. Sci.* **1960**, *23*, 22-26.

96. G.M. Whitesides, B. Grzybowski, *Science* **2002**, *295*, 2418-2421.

97. G. M. Whitesides, Small **2005**, *2*, 172-179.

98. P. L. Anelli, P. R. Ashton, R. Ballardini, V. Balzani, M. Delgado, M. T. Gandolfi, T. T. Goodnow, A. E. Kaifer, D. Philp, M. Pietraszkiewicz, L. Prodi, M. V. Reddington, A. M. Z. Slawin, N. Spencer, J. F. Stoddart, C. Vicent, D. J. Williams, *J. Am. Chem. Soc.* **1992**, *114 ,* 193-218.

99. R. E. Gillard, F. M. Raymo, J. F. Stoddart, *Chem. Eur. J.* **1997**, *3*, 1933-1940.

100. S. J. Cantrill, A. R. Pease, J. F. Stoddart, *J. Chem. Soc. Dalton Trans.* **2000**, *21*, 3715-3734.

101. J. F. Stoddart, H. M. Colquhoun, *Tetrahedron* **2008**, 64, 8231-8263.

102. E. R. Kay, D. A. Leigh, F. Zerbetto, *Angew. Chem. Int. Ed.* **2007**, *46*, 72-191.

103. V. Balzani, A. Credi, F M Raymo, J. F. Stoddart, *Angew. Chem. Int. Ed.* **2000**, *39*, 3348-3391.

104. M. S. Vickers, P. D. Beer, *Chem. Soc. Rev.* **2007**, *36*, 211-225.

105. S. J. Loeb, *Chem. Soc. Rev.* **2007**, *36*, 226-235.

106. D. S. Marlin, D. G. Cabrera, D. A. Leigh, A. M. Z. Slawin*, Angew. Chem. Int. Ed.* **2006**, *45*, 1385-1390.

107. H. Murakami, A. Kawabuchi, R. Matsumoto, T. Ido, N. Nakashima, *J. Am. Chem. Soc.* **2005**, *127*, 15891-15899.

108. A. Mateo-Alonso, G. Fioravanti, M. Marcaccio, F. Paolucci, G. M. A. Rahman, C. Ehli, D. M. Guldi, M. Prato, *Chem. Commun.* **2007**, 1945-1948.

109. J. D. Badjic, V. Balzani A. Credi, S. Silvi, J. F. Stoddart, *Science* **2004**, *303*, 1845-1849.

110. Y. Liu, A. H. Flood, P. A. Bonvallet, S. A. Vignon, B. Northrop, H.-R. Tseng, J. Jeppesen, T. J. Huang, B. Brough, M. Baller, S. Magonov, S. Solares, W. A. Goddard III, C.-M. Ho, J. F. Stoddart, *J. Am. Chem. Soc.* **2005**, *127*, 9745-9759.

111. M. C. Jimenez, C. O. Dietrich-Buchecker, J. P. Sauvage, *Angew. Chem. Int. Ed.* **2000**, *39*, 3284-3287.

112. P. R. Ashton, D. Philp, N. Spencer, J. F. Stoddart, *Chem. Commun.* **1991**, 1677-1679.

113. D. B. Amabilino, P. R. Ashton, M. S. Tolley, J. F. Stoddart, D. J. Williams, *Angew. Chem. Int. Ed. Engl.* **1993,** *32*, 1297-1301.

114. P. R. Ashton, B. Odell. M. V Reddington. A. M. Z. Slawin, J. F. Stoddart, D. J. Williams, *Angew. Chem. Int. Ed. Engl.* **1988**, *27*, 1550-1556.

115. M. V. Reddington, A. M. Z. Slawin, N. Spencer, J. F. Stoddart, C. Vicent. D. J. Williams, *Chem. Commun.* **1991**, 630-632.

116. T. T. Goodnow, M. V. Reddington, J. F. Stoddart, A. E. Kaifer, *J. Am. Chem. Soc.* **1991**, *113*, 4335-4341.

117. P. R. Ashton, V. Baldoni, V. Balzani, A. Credi, H. D. A. Hoffmann, M.-V. Martínez-Díaz, F. M. Raymo, J. F. Stoddart, M. Venturi, *Chem. Eur. J.* **2001**, *7*, 3482-3493.

118. P. R. Ashton, S. E. Boyd, A. Brindle, S. J. Langford, S. Menzer, L. Pe, J. A. Preece, F. M. Raymo, N. Spencer, J. F. Stoddart, A. J. P. White, D. J. Williams, *New J. Chem.* **1999**, *23*, 587-602.

119. D. Philp, J. F. Stoddart, *Angew.Chem. Int. Ed. Engl.* **1996**, *35*, 1154-1196.

120. A. G. Johnston, D. A. Leigh, R. J. Pritchard, M. D. Deegan, *Angew. Chem. Int. Ed. Engl.* **1995**, *34*, 1209-1212.

121. W.-Y. Sun, M. Yoshizawa, T. Kusukawa, M. Fujita, *Curr. Opin. Chem. Biol.* **2002**, *6*, 757-764.

122. V. Maurizot, M. Yoshizawa, M. Kawano, M. Fujita, *Dalton Trans.* **2006**, 2750-2756.

123. M. Yoshizawa, T. Kusukawa, M. Fujita, K. Yamaguchi, *J. Am. Chem. Soc.* **2000**, *122*, 6311-6312.

124. M. Yoshizawa, T. Kusukawa, M. Fujita, S. Sakamoto, K. Yamaguchi, *J. Am. Chem. Soc.* **2001**, *123*, 10454-10459.

125. M. Yoshizawa, M. Fujita, *Pure Appl. Chem.* **2005**, *77*, 1107-1112.

126. G. R. Desiraju, *Nature* **2001**, *412*, 397-400.

127. E. de Silva, M. P. H. Stumpf, *J. R. Soc. Interface* **2005**, *2*, 419-430.

128. Z. Dadon, N. Wagner, G. Ashkenasy, *Angew. Chem. Int. Ed.* **2008**, *47*, 6128-6136.

129. G. M. Whitesides, R. F. Ismagilov, *Science* **1999**, *284*, 89-92.

130. M. Kindermann, I. Stahl, M. Reimold, W. M. Pankau, G. von Kiedrowski, *Angew. Chem. Int. Ed.* **2005**, *44*, 6750-6755.

131. J. Stankiewicz, L. H. Eckardt, *Angew. Chem. Int. Ed.* **2006**, *45*, 342-344.

132. R. F. Ludlow, S. Otto, *Chem. Soc. Rev.* **2008**, *37*, 101-108.

133. P. T. Corbett, J. Leclaire, L. Vial, K. R. West, J.-L. Wietor, J. K. M. Sanders, S. Otto, *Chem. Rev.* **2006**, *106*, 3652-3711.

134. S. Ladame, *Org. Biomol. Chem.* **2008**, *6*, 219-226.

135. S. Otto, R. L. E. Furlan, J. K. M. Sanders, *Curr. Opin. Chem. Biol.* **2002**, *6*, 321-327.

136. J.-M. Lehn, A. V. Eliseev, *Science* **2001**, *291*, 2331-2332.

137. G. Lowe*, Chem. Soc. Rev.* **1995**, *24*, 309-317.

138. F. Balkenhohl, C. von dem Bussche-Hünnefeld, A. Lansky, C. Zechel*, Angew. Chem. Int. Ed. Engl.* **1996**, *35*, 2288-2337.

139. J. W. Szostak (Ed.), *Chem. Rev.* **1997**, *97*, special issue.

140. K. T. Chapman, G. F.Joyce, W. C. Still (Eds), *Curr. Opin. Chem. Biol.* **1997**, *1*, special issue.

141. J.-M. Lehn, *Chem. Eur. J.* **1999**, *5*, 2455-2463.

142. S. J. Rowan, S. J. Cantrill, G .R. L. Cousins, J. K. M. Sanders, J. F. Stoddart, *Angew. Chem. Int. Ed.* **2002**, *41*, 898-952.

143. S. Otto, R. L. E. Furlan, J. K. M. Sanders, *J. Am. Chem. Soc.* **2000**, *122*, 12063-12064.

144. H. F. Gilbert, *J. Biol. Chem.* **1997**, *272*, 29399-29402.

145. A. L. Kieran, A. D. Bond, A. M. Belenguer, J. K. M. Sanders, *Chem. Commun.* **2003**, 2674-2675.

146. C.H. Wong, G. M. Whitesides, *Enzymes in Synthetc Organic Chemistry* **1995**, Pergamon, Oxford.

147. S. Toledano, R. J. Williams, V. Jayawarna, R. V. Ulijn, *J. Am. Chem. Soc.* **2006**, *128*, 1070-1071.

148. R. J. Lins, S. L. Flitsch, N. J. Turner, E. Irving, S. A. Brown, *Tetrahedron* **2004**, *60*, 771-780.

149. R. Larsson, Z. Pei, O. Ramström*, Angew. Chem. Int. Ed.* **2004**, *43*, 3716-3718.

150. B. Shi, M. F. Greaney, *Chem. Commun.* **2005**, 886–888.

151. C. D. Meyer, C. S. Joiner, J. F. Stoddart, *Chem. Soc. Rev.* **2007**, *36*, 1705-1723.

152. S. L. Roberts, R. L. E. Furlan, S. Otto, J. K. M. Sanders, *Org. Biomol. Chem.* **2003**, *1*, 1625-1633.

153. I. Saur, K. Severin, *Chem. Commun.* **2005**, 1471-1473.

154. M. T. Stone, J. S. Moore, *J. Am. Chem. Soc.* **2005**, *127*, 5928-5935.

155. F. Cardullo, M. Crego Callama, B. H. M. Snellink-Ruel, J. L. Weidmann, A. Bielejewska, R. Fokkens, N. M. M. Nibbering, P. Timmerman, D. N. Reinhoudt, *Chem. Commun.* **2000**, 367-368.

156. A. T. ten Cate, P. Y. W. Dankers, R. P. Sijbesma, E. W. Meier, *J. Org. Chem.* **2005**, *70*, 5799-5803.

157. J.-M. Lehn, *Chem. Soc. Rev.* **2007**, *36*, 151-160.

158. S. Otto, K. Severin, *Top. Curr. Chem.* **2007**, *277*, 267-288.

159. B. Shi, R. Stevenson, D. J. Campopiano, M. F. Greaney, *J. Am. Chem. Soc.* **2006**, *128*, 8459-8467.

160. L. Milanesi, C. A. Hunter, S. E. Sedelinkova, J. P. Waltho, *Chem. Eur. J.* **2006**, *12*, 1081-1087.

161. O. Ramström, S. Lohmann, T. Bunyapaiboonsri, J.-M. Lehn, *Chem. Eur. J.* **2004**, *10*, 1711-1715.

162. M. Hochgürtel, R. Biesinger, H. Kroth, D. Piecha, M. W. Hofmann, S. Krause, O. Schaaf, C. Nicolau, A. V. Eliseev, *J. Med. Chem.* **2003**, *46*, 356-358.

163. R. F. Ludlow, J. Liu, H. Li, S. L. Roberts, J. K. M. Sanders, S. Otto, *Angew. Chem. Int. Ed.* **2007**, *46*, 5762-5764.

164. P. C. Haussmann, S. I. Khan, J. F. Stoddaart, *J. Org. Chem.* **2007**, *72*, 6708-6713.

165. J. M. C. A. Kerckhoffs, M. A. Mateos-Timoneda, D. N. Reinhoudt, M. Crego-Calama, *Chem. Eur. J.* **2007**, *13*, 2377-2385.

166. E. Stulz, S. M. Scott, A. D. Bond, S. J. Teat, J. K. M. Sanders, *Chem. Eur. J.* **2003**, *9*, 6039-6048.

167. S. J. Rowan, J. K. M. Sanders, *J. Org. Chem.* **1998**, *63*, 1536-1546.

168. S. J. Rowan, D. G. Hamilton, P. A. Brady, J. K. M. Sanders, *J. Am. Chem. Soc.* **1997**, *119*, 2578-2579.

169. P. J. Comina, D. Philp, B. M. Kariuki, K. D. M. Harris, *Chem. Commun.* **1999**, 2279-2280.

170. L. M. Greig, D. Philp, *Chem. Soc. Rev.* **2001**, *30*, 287-302.

171. S. Otto, S. Kubik, *J. Am. Chem. Soc.* **2003**, *125*, 7804-7805.

172. I. Huc, J.-M. Lehn, *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 2106-2110.

173. S. Gerber-Lemaire, F. Popowycz, E. Rodriguez-Garcia, A. T. C. Asenjo, I. Robina, P. Vogel, *ChemBioChem* **2002**, *3*, 466-470.

174. A. Gonzalez-Alvarez, I. Alfonso, F. Lopez-Ortiz, A. Aguirre, S. Garcia-Granada, V. Gotor, *Eur. J. Org. Chem.* **2004**, 1117-1127.

175. K. S. Chichak, S. J. Cantrill, A. R. Pease, S.-H. Chiu, G. W. V. Cave, J. L. Atwood, J. F. Stoddart, *Science* **2004**, *304*, 1308-1312.

176. R. T. S. Lam, A. Belenguer, S. L. Roberts, C. Naumann, T. Jarrosson, S. Otto, J. K. M. Sanders, *Science* **2005**, *308*, 667-669.

177. M. Hochgürtel, *Eur. Pat. Appl.* **2005**, EP 1496038 A1.

178. S. M. Turega, C. Lorenz, J. W. Sadownik, D. Philp, *Chem. Commun.* **2008**, 4076-4078.

179. D. Schultz, J. R. Nitschke, *Angew. Chem. Int. Ed.* **2006**, *45*, 2453-2456.

180. R. J. Sarma, S. Otto, J. R. Nitschke, *Chem. Eur. J.* **2007**, *13*, 9542-9546.

181. A. Buryak, K. Severin, *Angew. Chem. Int. Ed.* **2005**, *44*, 7935-7938.

182. J. D. Cheeseman, A. D. Corbett, J. L. Gleason, R. J. Kazlauskas, *Chem. Eur. J.* **2005**, *11*, 1708-1716.

183. J. D. Cheeseman, A. D. Corbett, R. Shu, J. Croteau, J. L. Gleason, R. J. Kazlauskas, *J. Am. Chem. Soc.* **2002**, *124*, 5692-5701.

184. P. Vongvilai, M. Angelin, R. Larsson, O. Ramström, *Angew. Chem. Int. Ed.* **2007**, *46*, 948-950.

185. M. Angelin, P. Vongvilai, A. Fischer, O. Ramström, *Chem. Commun.* **2008**, 768-770.

186. M. Angelin, A. Fischer, O. Ramström, *J. Org. Chem.* **2008**, *73*, 3593-3595.

187. R. Bennes, D. Philp, *Org. Lett.* **2006**, *8*, 3651-3654.

188. R. Bennes, D. Philp, N. Spencer, B. M. Kariuki, K. D. M. Harris, *Org. Lett.* **1999**, *1*, 1087-1090.

189. K. Severin, *Chem. Eur. J.* **2004**, *10*, 2565-258.

190. B. de Bruin, P. Hauwert, J. N. H. Reek, *Angew. Chem. Int. Ed.* **2006**, *45*, 2660-2663.

191. P. T. Corbett, S. Otto, J. K. M. Sanders, *Chem. Eur. J.* **2004**, *10*, 3139-3143.

192. P. T. Corbett, J. K. M. Sanders, S. Otto, *Angew. Chem. Int. Ed.* **2007**, *46*, 8858-8861.

193. C. A. Schalley, F. Vögtle, K. H. Dötz (Eds.), *Top. Curr. Chem.* **2005**, *248 and 249,* special editions.

194. A. Robertson, A. J. Sinclair, D. Philp, *Chem. Soc. Rev.* **2000**, *29*, 141-152.

195. N. Paul, G. F. Joyce, *Curr. Opin. Chem. Biol.* **2004**, *8*, 634-639.

196. V. Patzke, G. von Kiedrowski, *ARKIVOC* **2007**, *5*, 293-310.

197. G. von Kiedrowski, *Bioorg. Chem. Front.* **1993**, *3*, 113-146.

198. A. Kornberg, *J. Biol. Chem.* **1988**, *263*, 1-4.

199. G. von Kiedrowski, *Angew. Chem. Int. Ed. Engl.* **1986**, *25,* 932-935.

200. G. von Kiedrowski, B. Wlotzka, J. Helbing, M. Matzen, S. Jordan, *Angew. Chem. Int. Ed. Engl.* **1991**, *30,* 423-426.

201. N. Paul, G. F. Joyce, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12733-12740.

202. D.-E. Kim, G. F. Joyce, *Chem. Biol.* **2004**, *11*, 1505-1512.

203. T. Tjivikua, P. Ballester, J. Rebek, Jr. *J. Am. Chem. Soc.* **1990**, *112*, 1249-1250.

204. J. S. Nowick, Q. Feng, T. Tjivikua, P. Ballester, J. Rebek, Jr., *J. Am. Chem. Soc.* **1991**, *113*, 8831-8839.

205. F. M. Menger, A. V. Eliseev, N. A. Khanjin, *J. Am. Chem. Soc.* **1994**, *116*, 3613-3614.

206. M. M. Conn, E. A. Wintner, J. Rebek, Jr., *J. Am. Chem. Soc.* **1994**, *116*, 8823-8824.

207. D. N. Reinhoudt, D. M. Rudkevich, F. de Jong, *J. Am. Chem. Soc.* **1996**, *118*, 6880-6889.

208. V. Rotello, J.-I. Hong, J. Rebek, Jr., *J. Am. Chem. Soc.* **1991**, *113*, 9422-9423.

209. S. Xu, N. Giuseppone, *J. Am. Chem. Soc.* **2008**, *130*, 1826-1827.

210. B. Wang, I. O. Sutherland, *Chem. Commun.* **1997**, 1495-1496.

211. J. M. Quayle, A. M. Z. Slawin, D. Philp, *Tetrahedron Lett.* **2002**, *43*, 7229-7233.

212. V. C. Allen, D. Philp, N. Spencer, *Org. Lett.* **2001**, *3*, 777-780.

213. E. Kassianidis, D. Philp, *Angew. Chem. Int. Ed.* **2006**, *45*, 6344-6348.

214. D. H. Lee, J. R. Granja, J. A. Martinez, K. Severin, M. R. Ghadiri, *Nature* **1996**, *382*, 525-528.

215. A. J. Kennan, V. Haridas, K. Severin, D. H. Lee, M. R. Ghadiri, *J. Am. Chem. Soc.* **2001**, *123*, 1797-1803.

216. X. Li, J. Chmielewski, *Org. Biomol. Chem.* **2003**, *1*, 901-904.

217. A. Credi, *Angew. Chem. Int. Ed.* **2007**, *46*, 5472-5475.

218. A. P. de Silva, S. Uchiyama, *Nature Nanotech.* **2007**, *2*, 399-409.

219. D. S. Goodsell, *Bionanotechnology: Lessons from Nature* **2004**, Wiley, Hoboken.

220. F. M. Raymo, S. Giordani, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 4941-4944.

221. B. L. Feringa, *J. Org. Chem.* **2007**, *72*, 6635-6652.

222. D. Gust, T. A. Moore, A. L. Moore, *Chem. Commun.* **2006**, 1169-1178.

223. M. W. Holman, D. M. Adams, *ChemPhysChem* **2004**, *5*, 1831-1836.

224. H. Tian, Q.-C. Wang, *Chem. Soc. Rev.* **2006**, *35*, 361-374.

225. A. P. de Silva, N. D. McClenaghan, *Chem. Eur. J.* **2004**, *10*, 574-586.

226. J. R. Gregg, *Ones and Zeros: Understanding Boolean Algebra, Digital Circuits and the Logic of Sets*, **1998**, Wiley, NY.

227. A. P. de Silva, H. Q. N. Gunaratne, C. P. McCoy, *Nature* **1993**, *364*, 42-44.

228. W. R. Dichtel, J. R. Heath, J. F. Stoddart, *Phil. Trans. R. Soc. A* **2007**, *365*, 1607-1625.

229. J. C. Cox, A. D. Ellington, *Curr. Biol.* **2001**, *11*, R336.

230. J. C. Cox, D. S. Cohen, A. D. Ellington, *Trends. Biotechnol.* **1999**, *17*, 151-154.

231. M. N. Stojanovic, D. Stefanovic, *J. Am. Chem. Soc.* **2003**, *125*, 6673-6676.

232. L. M. Adleman, *Science* **1994**, *266*, 1021-1024.

233. M. N. Stojanovic, T. E. Mitchell, D. Stefanovic, *J. Am. Chem. Soc.*, **2002**, *124*, 3555-3561.

234. Y. Li, R. R. Breaker, *Curr. Opin. Struct. Biol.* **1999**, *9*, 315-323.

235. R. Penchovsky, R. R. Breaker, *Nature Biotechnol.* **2005**, *23*, 1424-1433.

236. M. N. Stojanovic, S. Semova, D. Kolpashchikov, J. Macdonald, C. Morgan, D. Stefanovic, *J. Am. Chem. Soc.* **2005**, *127*, 6914-6915.

237. M. N. Stojanovic, D. Stefanovic, *Nature Biotechnol.* **2003**, *21*, 1069-1074.

238. J. J. Tabor, A. D. Ellington, *Nature Biotechnol.* **2003**, *21*, 1013-1015.

239. R. Baron, O. Lioubashevski, E. Katz, T. Niazov, I. Willner, *Angew. Chem. Int. Ed.* **2006**, *45*, 1572-1576.

240. N. C. Gianneschi, M. R. Ghadiri, *Angew. Chem. Int. Ed.* **2007**, *46*, 3955-3958.

241. G. Ashkenasy, R. Jagasia, M. Yadav, M. R. Ghadiri, *Proc. Natl. Acad. Sci. U.S.A.* **2004***, 101*, 10872-10877.

242. G. Ashkenasy, M. R. Ghadiri, *J. Am. Chem. Soc.* **2004**, *126*, 11140-11141.

243. E. A. Wood, PhD Thesis, University of St Andrews, UK, **2008**.

244. V. C. Allen, PhD Thesis, University of St Andrews, UK, **2000**.

245. J. M. Quayle, PhD Thesis, University of St Andrews, UK, **2002**.

246. E. Kassianidis, PhD Thesis, University of St Andrews, UK, **2005**.

247. M. Sommelet, *Compt. Rend.* **1913**, *157*, 852-854.

248. K. Sonogashira, Y. Tohda, N. Hagihara, *Tetrahedron Lett.* **1975**, *16*, 4467-4470.

249. B. Das, B. Venkataiah, P. Madhusudhan, *Synlett.* **2000**, *1*, 59-60.

250 M. Karplus, *J. Am. Chem. Soc.* **1963**, *85*, 2870-2871.

251. T. B. Patrick, J. A. Schield, D. G. Kirchner, *J. Org. Chem.* **1974**, *39*, 1758-1761.

252. E. Bamberger, *Chem Ber.* **1895**, *28*, 245-251.

253. J. Liu, *J. Chin. Chem. Soc.* **1975**, *22*, 243-247.

254. A. Roe, J. A. Montgomery, *J. Am. Chem. Soc.* **1953**, *75*, 910-912.

255. F. J. Weigert, W. A. Sheppard *J. Org. Chem.*, **1976**, *41*, 4006-4012.

256. R. E. Banks, R. A. DuBoisson, R. G. Pritchard, A. E. Tipping, *Acta Cryst.* **1995**, *C51*, 1470-1472.

257. M. le Guyader, *Bull. Soc. Chim. Fr.* **1966**, 1848-1858.