Leonard J. Prins
Dipartimento di Scienze Chimiche
Università di Padova
leonard.prins@unipd.it
www.chimica.unipd.it/leonard.prins/pubblica/

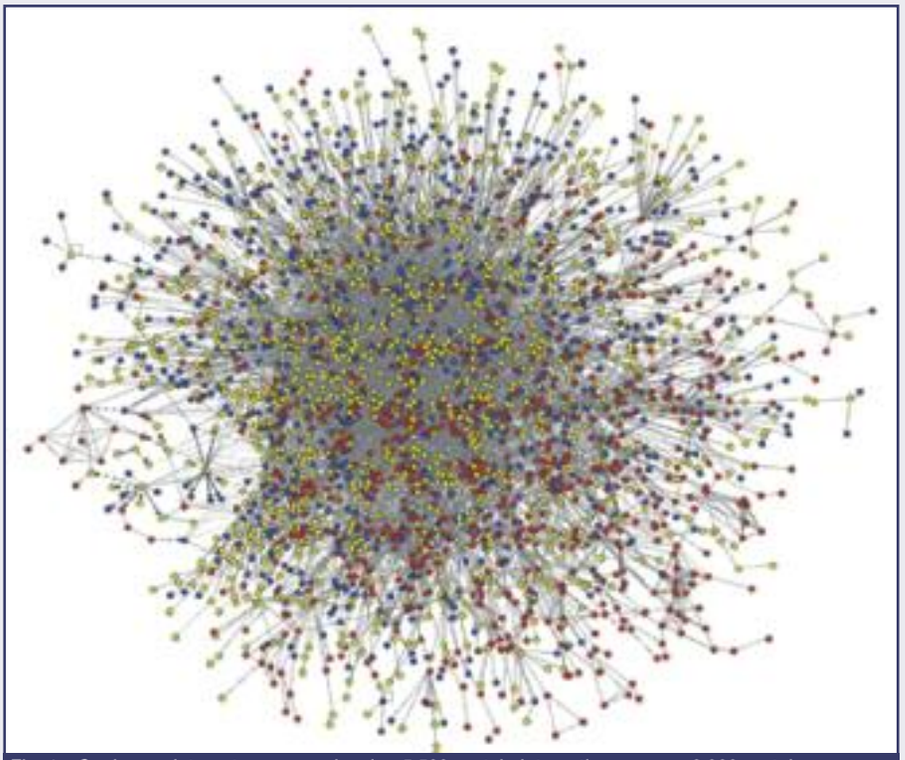

Fig. 1 - *C. elegans* interactome map showing 5,500 protein interactions among 3,000 proteins

SYSTEMS CHEMISTRY: THE MAGIC OF MIXTURES

Systems chemistry is chemistry in search for complexity. This review describes current progress in the development of artificial molecular networks able of self-replication and Darwinian evolution. These properties are network characteristics, since a single pure component is unable of these tasks. The unprecedented potential of complex networks for the development of receptors and catalysts will be discussed.

Chemists don't like mixtures. This has an historical background, because chemistry has all about to do with the correlation of molecular property to molecular structure (for example binding affinity or catalytic activity). This requires molecules as pure as possible, to avoid unpleasant situations in which an observed effect originating from an impurity is mistakingly attributed to the principal component. Consequently, the actual synthesis of a new compound is considered a job half-way done, and purification is often the tedious and time-consuming part. This reductionistic approach has allowed chemistry to rise to great heights in understanding the origin of molecular properties. However, there is a current awareness that this approach is also posing limits. Looking at Nature, which has always been the main source of inspiration, it can be observed that function arises from complex networks of molecules, rather than from a single molecule [1]. Examples are metabolic and signalling pathways, feedback loops and control mechanisms, but also entire systems as the cell and the organism itself. Biologists extensively study such networks in a field called systems biology, which now constitutes one of the major pillars of the proteomics era [2]. Just to give an impression, in Fig. 1 a map of over 5,500 protein interactions among 3,000 proteins in *C. elegans* is reproduced from a special section in *Nature* in 2009 dedicated to systems biology [3].

Considering the strength, achievements and properties of natural systems it is surprising that only now chemists are starting to get interested in molecular networks [4, 5]. The emergence of systems chemistry as a research area is demonstrated by the opening of new research centres in Groningen (Centre for System Chemistry) and Eindhoven (Institute for Complex Molecular Systems) both in the Netherlands, the launch of a new journal entitled *Systems Chemistry,* a European network in the form of a Cost Action and a series of ESF high-level research conferences. What exactly is systems chemistry? This was the starting question posed by many speakers during the opening symposium of the Groningen Centre for System Chemistry. It is interesting to note that many of those were actually *involved* in systems chemistry, another sign of the embryonic stage of this research area. Systems chemistry can be defined as the chemistry that targets the collective properties of a molecular network which cannot be traced back to the single components. The aim of this review is to provide illustrative examples which show the potential of systems chemistry.

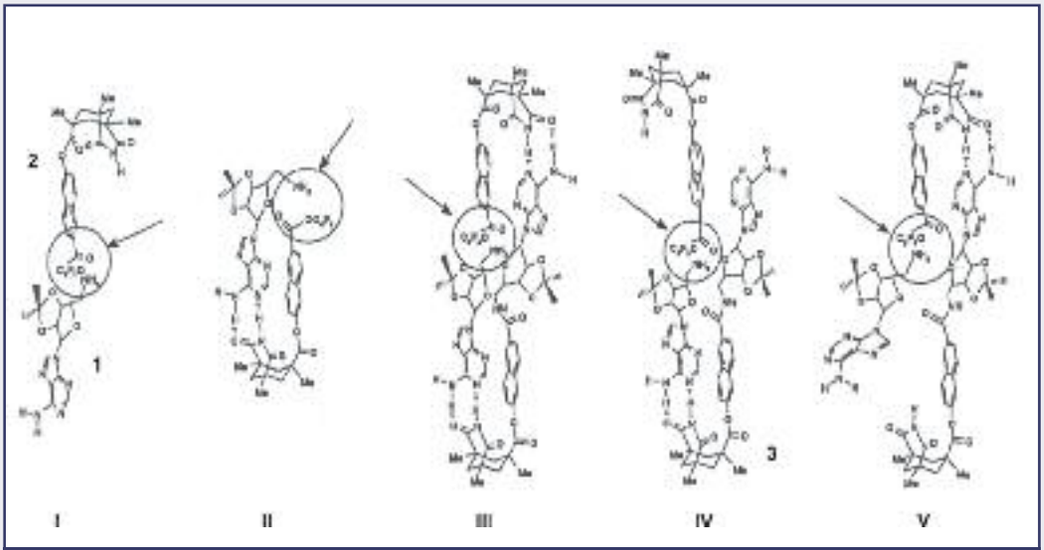

Fig. 2 - Relevant reaction pathways in Rebek's replicating system (I : background, II: preassociative, III: termolecular, IV: activated bimolecular, V: bimolecular). Arrows indicate bond formation

Self-replicating molecules

It has long fascinated scientists, but not only, to understand how a collection of molecules gives rise to something that is recognised as being alive [6]. Jack Szostak, the 2009 winner of the Nobel prize in Medicine, last year wrote a commentary in Nature entitled 'Systems chemistry on early Earth' [7] discussing the recent discoveries that activated pyrimidine ribonucleotides could be synthesized under prebiotically plausible conditions [8] *and* that an RNA enzyme could self-replicate [9]. These results sustain the theory that RNA may have played a crucial role in early days of Earth. A long-standing research goal has been the development of an artificial chemical system that bears the fundamental characteristics of life: the ability to self-replicate, to evolve through Darwinian evolution and to sustain itself [10].

The discovery by von Kiedrowski in 1986 that a hexadeoxynucleotide templated its own formation from the complementary activated trinucleotides sparked a quest for self-replicating molecules [11]. Here, the self-replicating molecules proposed by Rebek *et al.* will be discussed in detail, not only because it became the topic of a controversy with Menger, but also because it illustrates how quickly complexity is generated. The system proposed by Rebek *et al.* in 1991 was based on amide formation between amino adenosine **1** and the Kemp's triacid based activated ester **2** (Fig. 2) [12, 13]. It was observed that the addition of the product **3** to the initial reaction mixture increased the initial rate of formation of **3** up to roughly 40%. In self-replication studies this is generally taken as the best evidence of self-replication, *i.e.* the product **3** acts as a template for its own formation. In fact, NMR titrations indicated that **3** dimerizes with an association constant K_{ass} of 630 M^{-1}.

As in all self-replicating systems, the autocatalytic reaction has to compete with other pathways leading to the same product. For the given system Rebek included the second order background reaction (I in Fig. 2) and formation of **3** via intramolecular attack of the amine in the binary complex **1•2** (II in Fig. 2). In the latter case this results in the initial formation of a *cis*-amide which subsequently converts into *trans* to provide the stretched isomer of **3**. Being one of the first examples of a truly artificial system and being promoted as a primitive sign of life, this contribution aroused great interest, but also scepticism. For Menger *et al.* there was no reason to ascribe the observed catalytic effect to self-replication, since also primary amides were shown to catalyse the formation of **3** (which contains an internal amide bond) [14, 15]. The following dispute was ended in 1996 by a detailed kinetic analysis performed by Reinhoudt *et al*. [16] Limiting themselves to complexes composed of at most four species, a set of 16 thermodynamic equilibria was defined that accounted for the distribution of the building blocks. Five possible reactions (templated and not) that lead to formation of product **3** were considered (Fig. 1). To cut a long story short, it was concluded that the self-replication as defined by Rebek indeed contributed to catalysis with a rate enhancement factor of 6.8. Nonetheless, the presence of competing bimolecular reactions, as those indicated by Menger, significantly complicated the overall kinetic picture. Since then a plethora of self-replicating molecules has been reported and the concept itself is no longer under discussion. The take-home message here is that even a system that appears relatively simple, because composed of few components, can already express a level of complexity that generally acts as an excellent repellent for most chemists. However, analytical tools are now on such a sophisticated level that the analysis of complex systems is no longer a

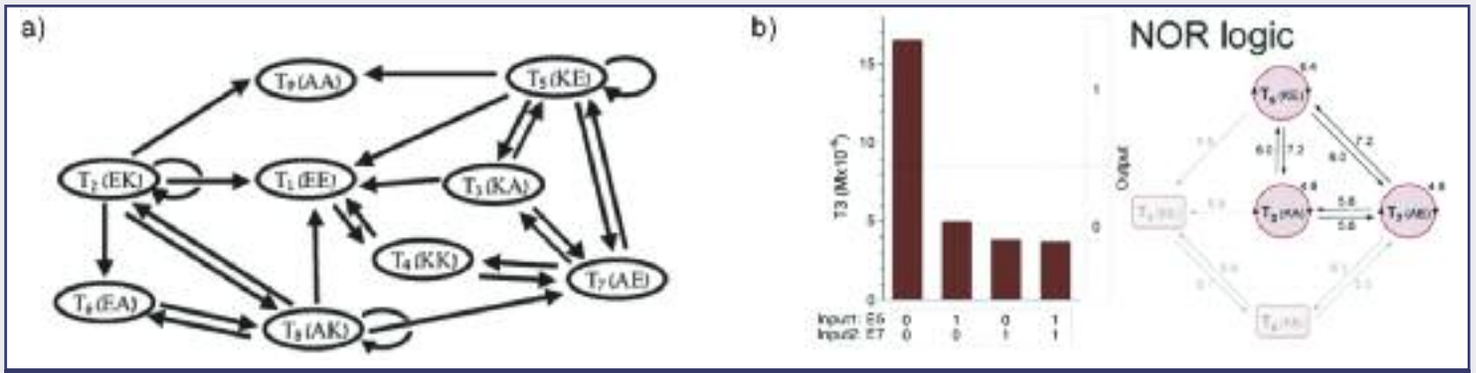

Fig. 3 - a) Experimentally derived network architecture as developed by Ghadiri *et al.* Each node represents a product (T_{1-9}) and each arrow indicates a template-assisted ligation pathway (of the peptide fragments N and E_1-$_9$) pointing from the template towards the product; b) example of the expression of a NOR logic gate expressed by a peptide sub-network (highlighted). The amount of product T_3 formed (output) is given as a function of the absence (o) or presence (1) of different combinations of input peptides E_5 and/or E_7

limiting factor. For instance, Ghadiri *et al.* created a complex network of auto and cross catalytic reactions between peptides (Fig. 3a) [17]. This network extended on their seminal contributions on peptide replicators based on the ligation of electro- and nucleophilic peptide fragments templated by the product through the formation of coiled-coil superstructures [18, 19]. A substoichiometic amount of nucleophilic peptide, **N**, was reacted with a mixture of nine electrophilic peptide fragments, $\mathbf{E_{1-9}}$, and the formation of products $\mathbf{T_{1-9}}$ was followed in time under native and denaturing conditions. Network connectivity was assessed by repeating the experiment several times seeding the mixture with either one of the templates $\mathbf{T_{1-9}}$. Importantly, the obtained network was in good agreement with the expected network based on theoretical calculations. This shows that even complex systems can be rationally designed to a large extent. Similar to natural systems, Ghadiri and Ashkenasy showed that it is possible to attribute function to small sub-networks within a larger framework [20]. For example, within the peptide network depicted in Fig. 3b, the rate of $\mathbf{T_3}$ production can be negatively affected by the addition of either $\mathbf{E_5}$ or $\mathbf{E_7}$ that suppress the autocatalytic cycle of $\mathbf{T_3}$ because of an enhancement of the more favourable $\mathbf{T_3 \rightarrow T_5}$ and/or $\mathbf{T_3 \rightarrow T_7}$ pathways. This is the equivalent of the Boolean logic function NOR (that produces a value *true* if and only if both operands are false, or, in this case, if both input peptides $\mathbf{E_5}$ and $\mathbf{E_7}$ are absent). The interconnectivity within a molecular network allows many other logic gates to be defined evaluating the relative cross catalytic and autocatalytic rates of subsets of products [21]. As such, systems of this type are attracting interest as alternative computing systems [22].

Darwinian evolution in molecular systems

One of the marvels of natural systems is their ability to interact with their environment. Systems, which can be entire organisms, but also smaller subunits, continuously respond, react, and adapt to external stimuli. Interaction with the environment is also the driving force for Darwinian evolution, which can be simply interpreted as mutation and 'survival of the fittest'. The independent discovery in 1997 by the groups of Lehn [23] and Sanders [24] that artificial systems can display similar behaviour was an enormous breakthrough and led to the establishment of the field of dynamic combinatorial chemistry [25]. This area deals with combinatorial libraries, but opposed to conventional ones, the library components are formed through the reversible connection of molecular subunits (Fig. 4) [26]. These are either noncovalent bonds (*e.g.* hydrogen or coordinative bonds) or, more frequently, dynamic covalent bonds (*e.g.* imines or disulfides). The reversibility of these bonds allows exchange between the library components possible and makes that the whole system is under thermodynamic control. Consequently, the library composition is determined by the relative thermodynamic stabilities of the library components. This is highly exciting, because it means that the system has the possibility to adapt to external stimuli able to alter the thermodynamic landscape.

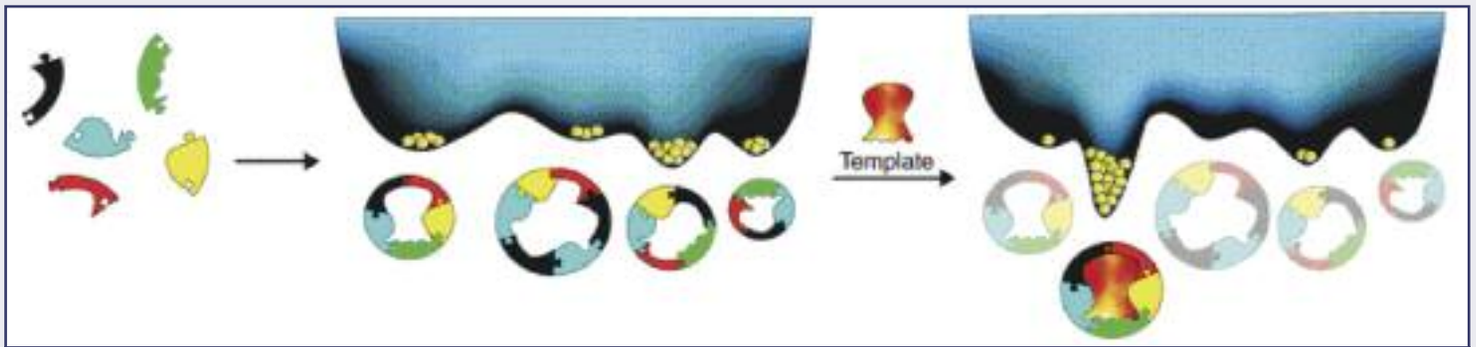

Fig. 4 - Concept of dynamic combinatorial chemistry. Addition of a template changes the free energy landscape and causes a redistribution of the dynamic library favouring the best fitted receptor. Image taken from [26]

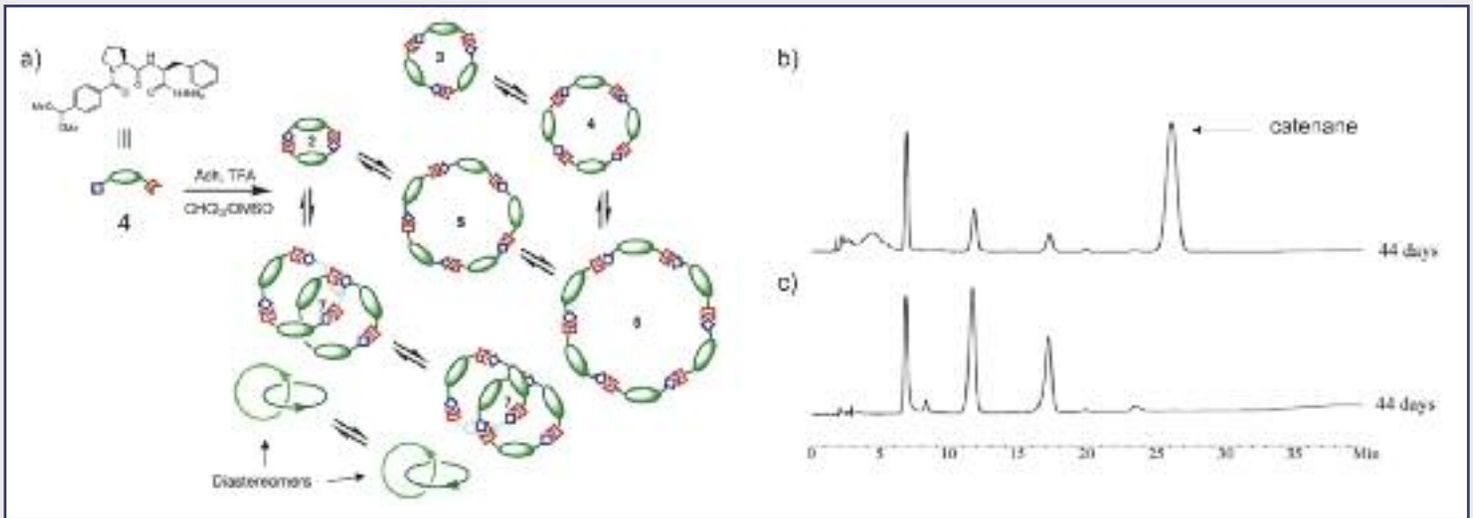

Fig. 5 - a) Formation of a dynamic combinatorial library of macrocycles and catenanes from building block 4. HPLC chromatograms of the DCL in the presence (b) and absence (c) of acetylcholine

Imagine a library of potential receptors for a substrate. Addition of the substrate to the dynamic library will make it probe the various components and form the thermodynamically most stable complex with the best receptor present. This will act as a thermodynamic sink and the dynamic library will ideally respond by increasing the concentration of that receptor. The advantage is clear. Since this is a process of spontaneous self-selection by the target, it implies that the difficult burden of rationally designing receptors (or other species) has been relieved. Rather, the challenge has become to design dynamic libraries that cover a portion of chemical space as large as possible. The ability of a dynamic library to respond spontaneously can give rise to surprising outcomes that no one would have ever been to anticipate. This is illustrated with one of the most striking examples that has been reported so far.

The group of Sanders and Otto prepared a dynamic library of macrocycles of different size upon addition of trifluoroacetic acid to a solution of peptide building block **4** (Fig. 5a) [27]. TFA liberates the aldehyde group and also catalyzes the formation and exchange of hydrazones with the hydrazide unit. HPLC analysis revealed that initially linear intermediates were formed that rapidly converted into a series of macrocycles (up to at least the hexamer). Thermodynamic equilibrium was reached in 3 days indicated by the absence of any further changes in time in the chromatogram (Fig. 5c). The surprise came when the experiment was repeated in the presence of the neurotransmitter acetylcholine ACh. Also here, initially a series of linear and cyclic oligomers was formed, but within one hour a new peak appeared that was never observed in the non-templated experiment. Over a long period of 44 days this peak continued to grow in intensity reaching a maximum level of 70% of the material in the library (based on HPLC) corresponding to an amplification of more than three orders of magnitude (Fig. 5b). Isolation and characterization of this peak by a combination of NMR and MS revealed that this product was a [2]-catenane composed of two interlocked cyclic trimers. Remarkably, in the absence of ACh the ^{1}H NMR spectrum of the catenane is ill-defined indicating the presence of slowly interconverting isomers. Addition of the substrate sharpens the spectrum indicating the formation of a well-defined complex. Isothermal calorimetry revealed an impressive binding constant of 1.4×10^7 M^{-1} in a 95:5 mixture of $CHCl_3$:DMSO.

Dynamic combinatorial chemistry has been most frequently used for the discovery of receptors using the substrate as trigger. However, systems that respond to various other type of stimuli (*e.g.* light, pH, pressure) have also been reported and dynamic combinatorial chemistry has been used to develop responsive materials, sensors, etc. [25] Recently, we have started a project that aims at the development of catalysts by means of dynamic combinatorial chemistry, which has remained one of the least studied applications [28]. The concept is to use as trigger a compound which is a stable analogue of the transition state (TSA) of a

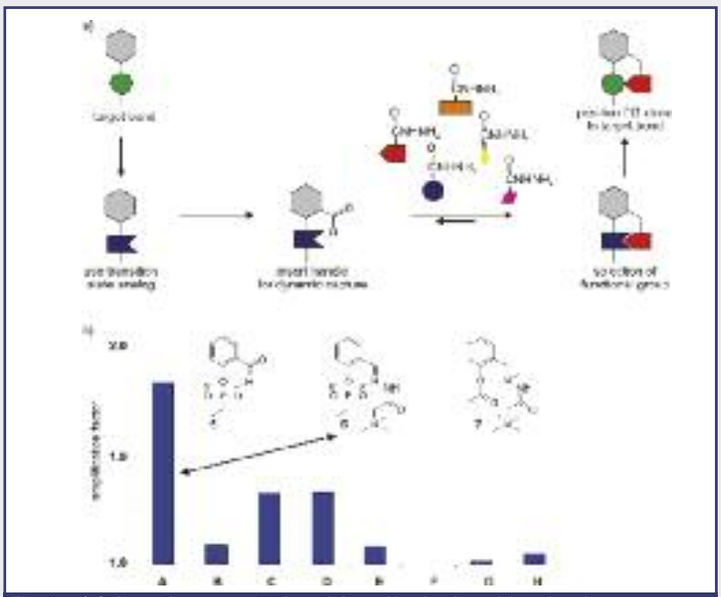

Fig. 6 - a) Schematic representation of the self-selection of functional groups that assist in the intramolecular cleavage of a target bond; b) Observed amplification factors for a hydrazide library identifying the strongest interaction in hydrazone 6 between the ammonium and phosphonate group [29]

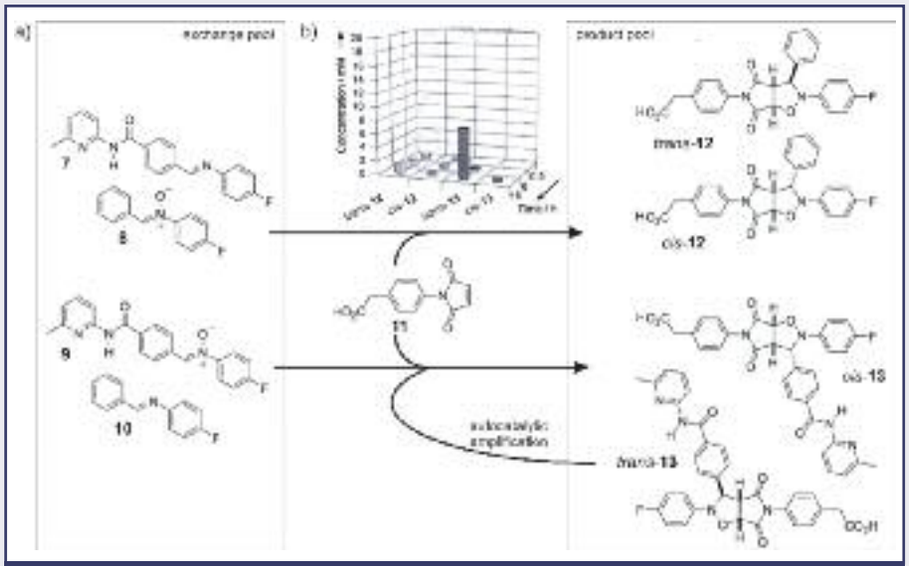

Fig. 7 - a) Self-amplification of a replicator from a dynamic imine/nitrone library; b) Distribution in the product pool after reaction with maleimide 11

reaction (Fig. 6a). The dynamic library will respond by increasing the concentration of the library member that most strongly interacts with the TSA. Added to the reaction under scrutiny, the isolated member will act as a catalyst by lowering the energy of the transition state. As a proof-of-principle we have focused on the self-selection of functional groups that can assist intramolecularly in the cleavage of a flanking ester bond [29]. Evidently this does not yield a true catalyst, but our choice to start from here was driven by the knowledge that intramolecular recognition events are much easier to detect [30]. So, our target molecule is 2-ethylphosphonoxybenzaldehyde **5** containing the phosphonate group as TSA for the basic cleavage of an ester moiety and an aldehyde for reversible hydrazone formation (Fig. 6b). Exposure of this target to a library of functional hydrazides yields a mixture of the corresponding hydrazones some of which stabilized through intramolecular interactions between the phosphonate and the functional group of the hydrazide unit. The occurrence of intramolecular interactions emerges immediately from a comparison of the library distribution with the distribution of a reference library in which the target is absent (using 2-methoxybenzaldehyde) [31]. We have developed new analytical tools based on fast ^{1}H-^{13}C HSQC NMR spectroscopy (with Damien Jeannerat from the University of Geneva) [32] and UV/Vis-spectroscopy [33] to rapidly determine the mixture composition. Comparison of the relative hydrazone concentrations in both libraries revealed the strongest amplification for hydrazone **6** equipped with an ammonium group. Positioning of that ammonium close to the ester moiety in functionalized phenylacetate **7** indeed resulted in a moderately enhanced cleavage rate [29]. A series of control experiments supported the hypothesis that this was indeed due to transition state stabilization. Evidently, the road from here to true catalyst development is still long, but these initial results are a promising start. The adaptability of a dynamic combinatorial library is a unique systems property and creates unprecedented possibilities for the discovery of receptors, catalysts, materials among others.

Smart mixtures

It is the aim of systems chemistry to exploit concepts such as self-replication and Darwinian evolution and to combine them. Its potential emerges clearly from the following example that was recently reported by the group of Philp [34]. They took a small four component dynamic library of 2 imines and 2 nitrones able to interconvert (Fig. 7a). At thermodynamic equilibrium a 1.0:1.4:1.7:1.0 composition was observed for compounds **7**, **8**, **9**, and **10**, respectively. These components were chosen because of two characteristics. Two components of the library (**7** and **9**) contain an amido pyridine unit that can complex a carboxylic acid through the formation of two hydrogen bonds [35]. Two components (**8** and, importantly, again **9**) are nitrones able to undergo a 1,3-dipolar cycloaddition with maleimides [36]. The addition of maleimide **11**, containing a carboxylic acid, to this dynamic misture gives a spectacular result. The maleimide reacts with both nitrones **7** and **9** and causes an irreversible transfer of material to a product pool forming two pairs of diastereomeric cycloadducts: *cis*- and *trans*-**12** and *cis*- and *trans*-**13**, respectively. The special effect is caused by *trans*-**13** which is able to catalyse its own formation through the ternary complex [**9**•**11**•*trans*-**13**]. So, whereas *trans*-**13** is initially formed through the bimolecular reaction between **9** and **11**, the autocatalytic pathway takes over progressively. As a consequence, nitrone **9** is depleted from the library at a much higher rate compared to **7**. Being dynamic, the library responds by shifting its composition towards the formation of more nitrone **8**. Both the process of self-replication and evolution work in the same direction and highly efficiently drive the system towards the formation of *trans*-**13**. After 16 hours and an overall conversion of 48%, *trans*-**13** constitutes almost 80% of the product pool (Fig. 7b). In contrast, for a reference maleimide lacking the carboxylic acid recognition module a conversion of only 21% was obtained in the same time interval with a rather uniform distribution of the four diastereomeric products.

This example demonstrates the significant advances that have been made studying and manipulating artificial systems. Nonetheless, a crucial difference still exists with natural systems. Currently, changes in artificial systems are driven by a thermodynamic need to reach a more stable state. In contrast, natural (living !) systems do not operate at all under thermodynamic control [37]. Rather an off-equilibrium state is maintained to create energy differences which can be turned into action when needed. Obviously, the maintenance of this state requires the continuous need for external energy input. Development of artificial systems with this property would mean an enormous breakthrough in the field of systems chemistry.

Conclusions

Chemists are now in the position to face the challenges of mixtures and are starting to appreciate the accompanying complexity and potential. Compared to biologists, the approach taken by chemists

is fundamentally different. Whether biologists tackle molecular networks in a top-down approach focusing in on specific parts of an organism, cells, or cycles, chemists have the unique capability of constructing molecular networks in a bottom-up approach starting from the molecular constituents.
It is worth remembering that a similar dual approach on the borderline of chemistry and physics has lead towards the rise of nanotechnology. As illustrated by the examples in this review, systems chemistry can give conceptually new input in many fields, such as recognition, sensing, and catalysis and is obviously intimately related to the philosophical question of how life originated on Earth.

Acknowledgements: The author is grateful to prof. Paolo Scrimin for stimulating discussions and to Cristian Guarise, Giovanni Zaupa, Giulio Gasparini, Marta Dal Molin, Matteo Graziani and Anna Sansone for their collaboration on various projects. Financial support from the University of Padova (CPDA054893) and the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC Starting Grant agreement n° 239898 is acknowledged.

Leonard Prins (1974) obtained his PhD degree in 2001 from the University of Twente, Netherlands, under the supervision of David Reinhoudt. After postdoctoral periods with Peter Dervan (Caltech) and Giulia Licini (Padova), he is currently an assistant professor at the University of Padova, Italy. He has been awarded the H.J. Backerprize 2001 by the Organic Chemistry Division of the Royal Dutch Chemical Society, the European Young Chemist Award (silver) 2008 by EuCheMs and the 'Ciamician' medal 2009 by the Organic Chemistry Division of the Italian Chemical Society. He is a recipient of an ERC starting grant 2009. His current research interests are the application of dynamic chemistry for receptor and catalyst discovery and the development of multivalent enzyme-like catalysts.

References

[1] H. Kitano, *Science,* 2002, **295**, 1662.
[2] See the Institute for Systems Biology; www.systemsbiology.org
[3] See the special section dedicated to systems biology in *Nature,* 2009, **460**, 415.
[4] J. Stankiewicz, L.H. Eckardt, *Angew. Chem. Int. Ed.,* 2006, **45**, 342.
[5] R.F. Ludlow, S. Otto, *Chem. Soc. Rev.* 2008, **37**, 101.
[6] It is not the intention of this review to embark on the origin-of-life discussion. For an excellent starting point from a chemists' perspective see: P.L. Luisi, The Emergence of Life - From Chemical Origins to Synthetic Biology, Cambridge University Press, New York, 2006.
[7] J.W. Szostak, *Nature,* 2009, **459**, 171.
[8] M.W. Powner *et al., Nature,* 2009, **459**, 239.
[9] T.A. Lincoln, G.F. Joyce, *Science,* 2009, **323**, 1229.
[10] L.E. Orgel, *Crit. Rev. Biochem. Mol. Biol.,* 2004, **39**, 99.
[11] G. Von Kiedrowski, *Angew. Chem. Int. Ed. Engl.,* 1986, **25**, 932.
[12] T. Tjivikua *et al., J. Am. Chem. Soc.,* 1990, **112**, 1249.
[13] J.S. Nowick *et al., J. Am. Chem. Soc.,* 1991, **113**, 8831.
[14] F.M. Menger *et al., J. Am. Chem. Soc.,* 1994, **116**, 3613.
[15] F.M. Menger *et al., J. Org. Chem.,* 1995, **60**, 2870.
[16] D.N. Reinhoudt *et al., J. Am. Chem. Soc.,* 1996, **118**, 6880.
[17] G. Ashkenasy *et al., Proc. Natl. Acad. Sci. USA,* 2004, **101**, 10872.
[18] D.H. Lee *et al., Nature,* 1996, **382**, 525.
[19] A. Saghatelian *et al., Nature,* 2001, **409**, 797.
[20] G. Ashkenasy, M.R. Ghadiri, *J. Am. Chem. Soc.,* 2004, **126**, 11140.
[21] N. Wagner, G. Ashkenasy, *Chem. Eur. J.,* 2009, **15**, 1765.
[22] Z. Ezziane, *Nanotechnology,* 2006, **17**, R27.
[23] I. Huc, J.M. Lehn, *Proc. Natl. Acad. Sci. USA,* 1997, **94**, 2106.
[24] P.A. Brady, J.K.M. Sanders, *J. Chem. Soc. Perkin Trans. 1,* 1997, 3237.
[25] P.T. Corbett *et al., Chem. Rev.,* 2006, **106**, 3652.
[26] S. Otto *et al., Curr. Opin. Chem. Biol.,* 2002, **6**, 321.
[27] R.T.S. Lam *et al., Science,* 2005, **308**, 667.
[28] B. Brisig, *et al., Angew. Chem. Int. Ed.,* 2003, **42**, 1270.
[29] G. Gasparini *et al., Angew. Chem. Int. Ed.,* 2008, **47**, 2475.
[30] L.J. Prins, P. Scrimin, *Angew. Chem. Int. Ed.,* 2009, **48**, 2288.
[31] G. Gasparini *et al., Chem. Commun.,* 2007, 1340.
[32] G. Gasparini *et al., Chem. Commun.,* 2008, 3034.
[33] G. Gasparini *et al., Angew. Chem. Int. Ed.,* 2009, **48**, 4546.
[34] J.W. Sadownik, D. Philp, *Angew. Chem. Int. Ed.,* 2008, **47**, 9965.
[35] R.M. Bennes, D. Philp, *Org. Lett.,* 2006, **8**, 3651.
[36] E. Kassianidis, D. Philp, *Angew. Chem. Int. Ed.,* 2006, **45**, 6344.
[37] A. Pross, *Chem. Eur. J.,* 2009, **15**, 8374.

RIASSUNTO

Chimica dei Sistemi: la magia delle miscele

La Chimica dei Sistemi è la chimica che si occupa della complessità. Questo lavoro prende in esame i progressi più recenti compiuti nello sviluppo di network molecolari sintetici in grado di autoreplicarsi ed evolvere in maniera darwiniana. Si tratta di proprietà caratteristiche del sistema di molecole che non sono presenti nei singoli componenti isolati. Vengono pure prese in esame le potenzialità, largamente inesplorate, di network complessi per lo sviluppo di recettori molecolari e catalizzatori.