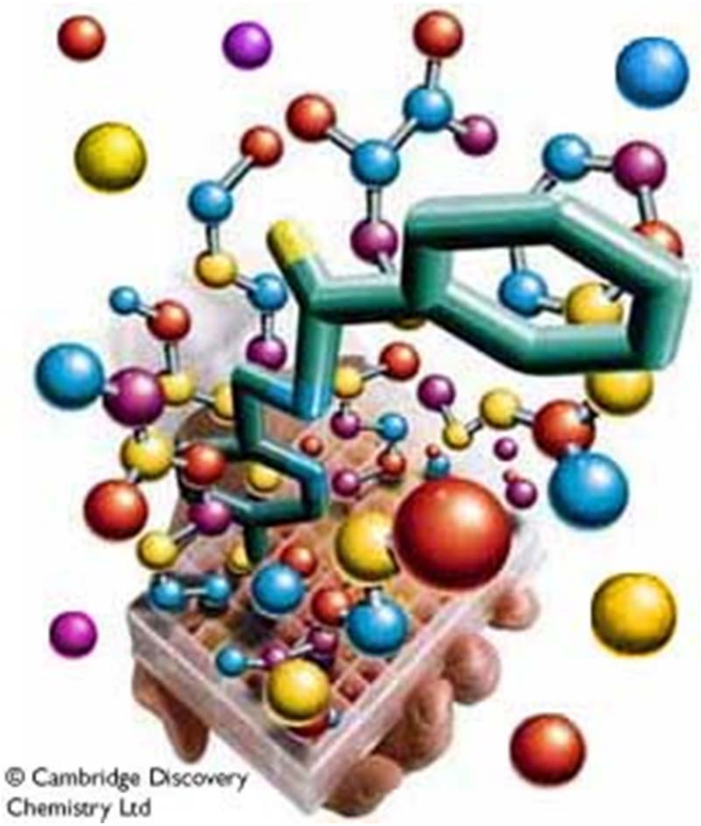
# Combinatorial Chemistry



The basic principle of combinatorial chemistry is to prepare and screen a large number of different molecules at the same time

# Combinatorial Chemistry
searching for a needle in a haystack

# History of Combinatorial Chemistry

1963 Solid Phase Peptide Synthesis

Merrifield (Nobel Prize 1984)

1982 Split-mix-synthesis

Furka

Multipin Synthesis 1984

Geysen

'Tea Bag' Synthesis 1984

Houghten

1993 Binary Encoding

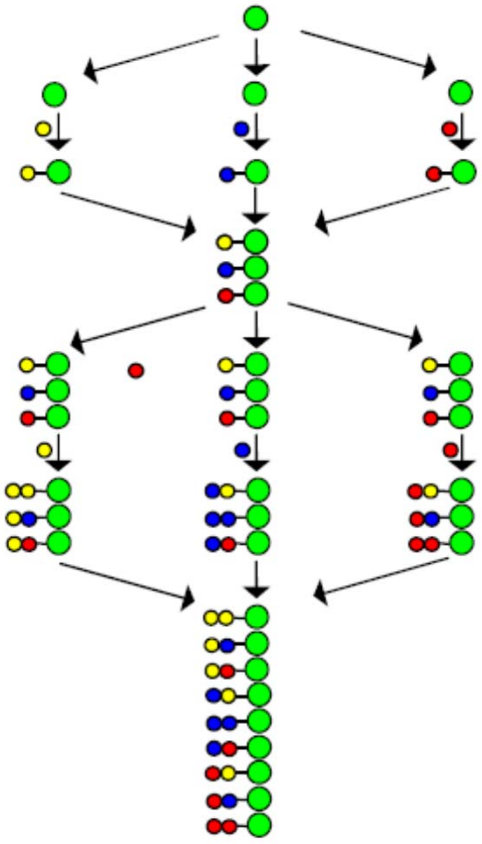
Clark W. Still

Schreiber

Diversity oriented synthesis 2000

# Split-mix concept

Synthesis of all dipeptides using three amino acids X, Y, Z ($3^2 = 9$ compounds)
One compound per bead (one bead/one component libraries)



*a*

Only 6 coupling steps are needed to make 9 compounds.

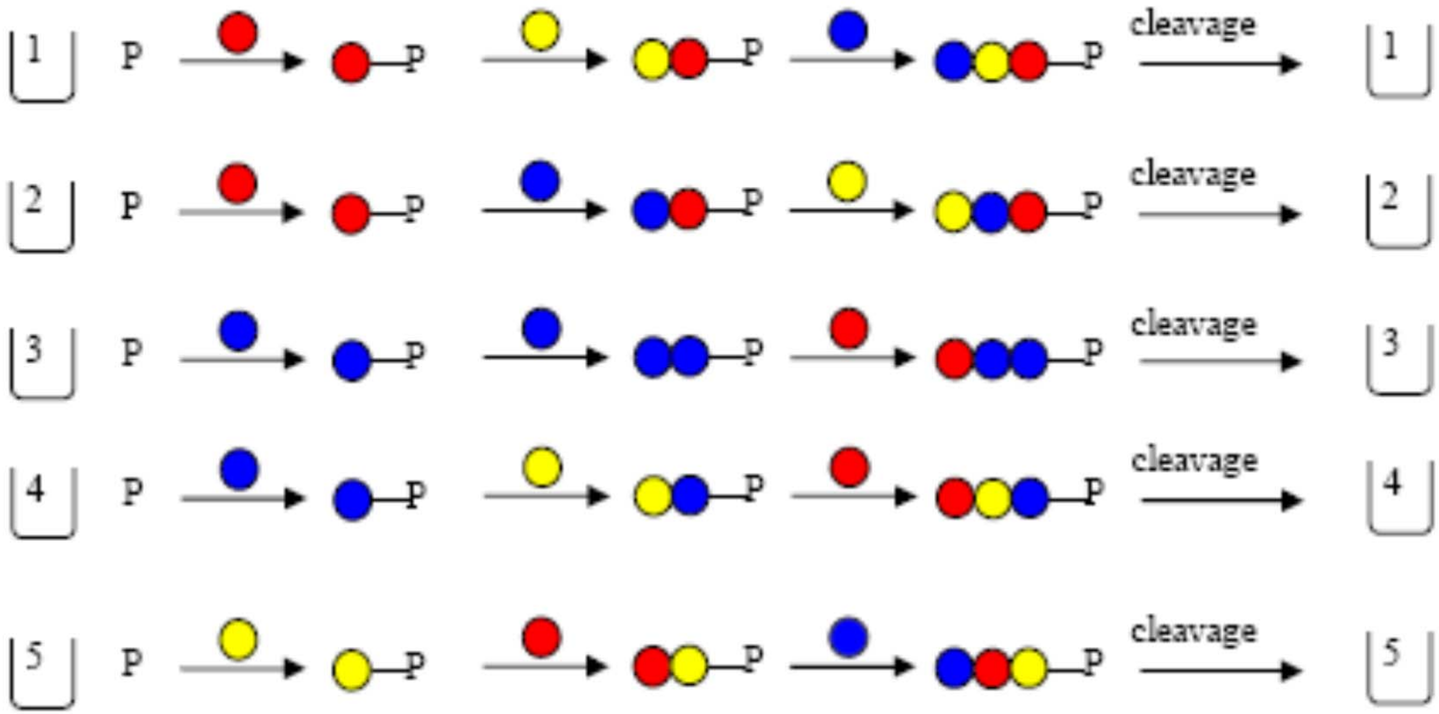
# Advantage

building block set of $N$ members

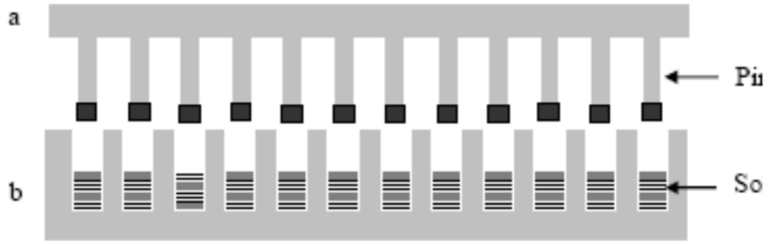
number of possible sequences containing $X$ members = $N^X$

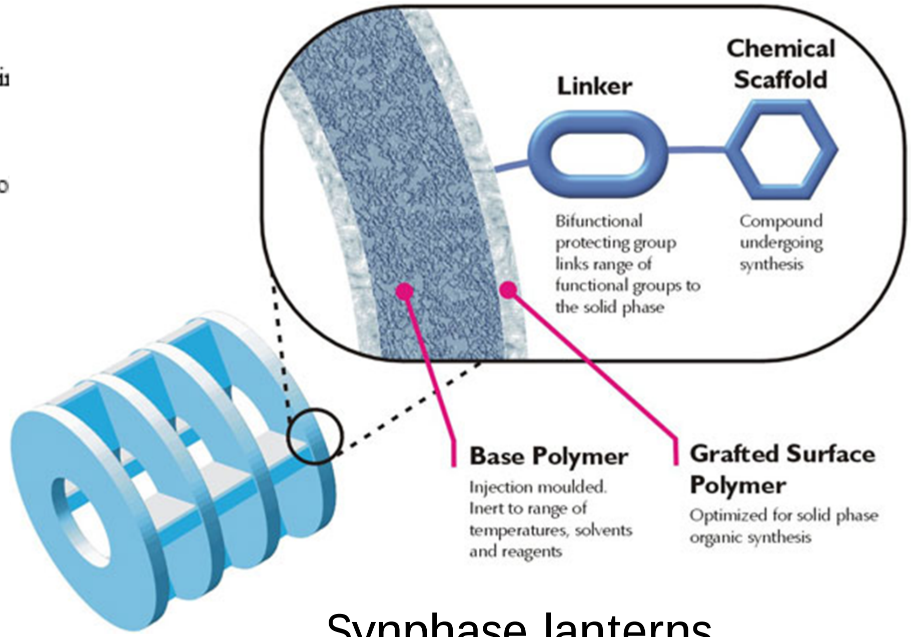
number of coupling steps required using split/pool/mix = $N \times X$

# Parallel synthesis

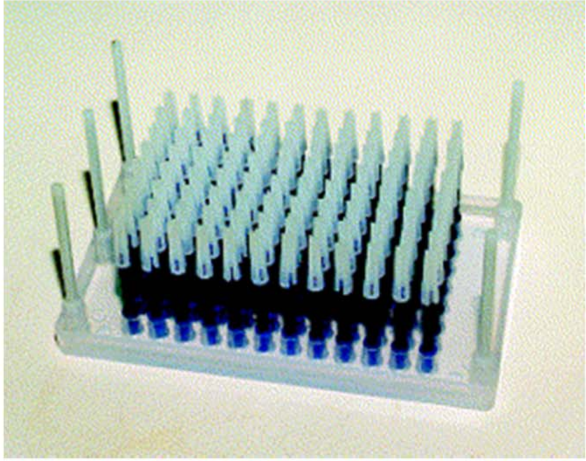


# Peptides on pins





Synphase lanterns



# Other techniques for parallel synthesis: Houghtens' 'tea-bag' method

- one compound per bag
- after each synthesis, regroup the bags according to the peptide sequence

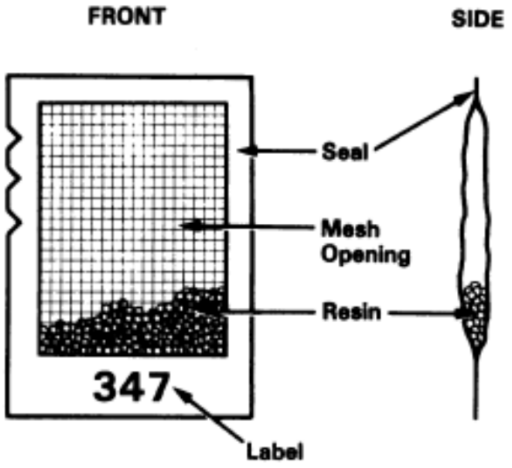
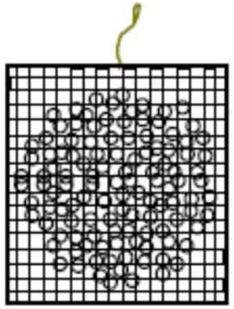
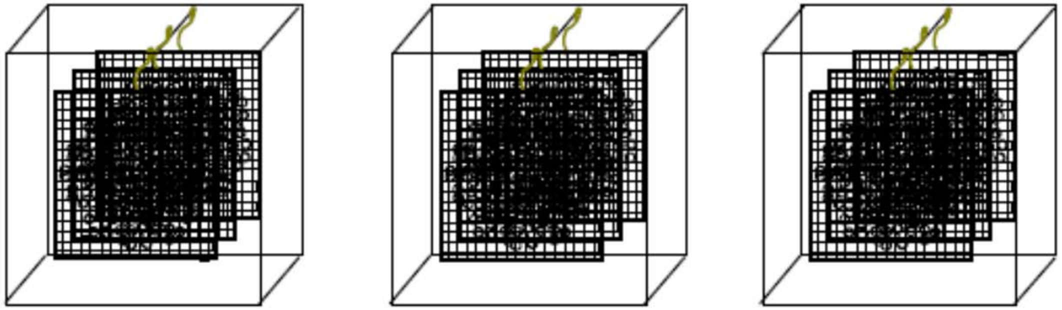


FIG. 1. Illustration of a mesh packet containing resin.

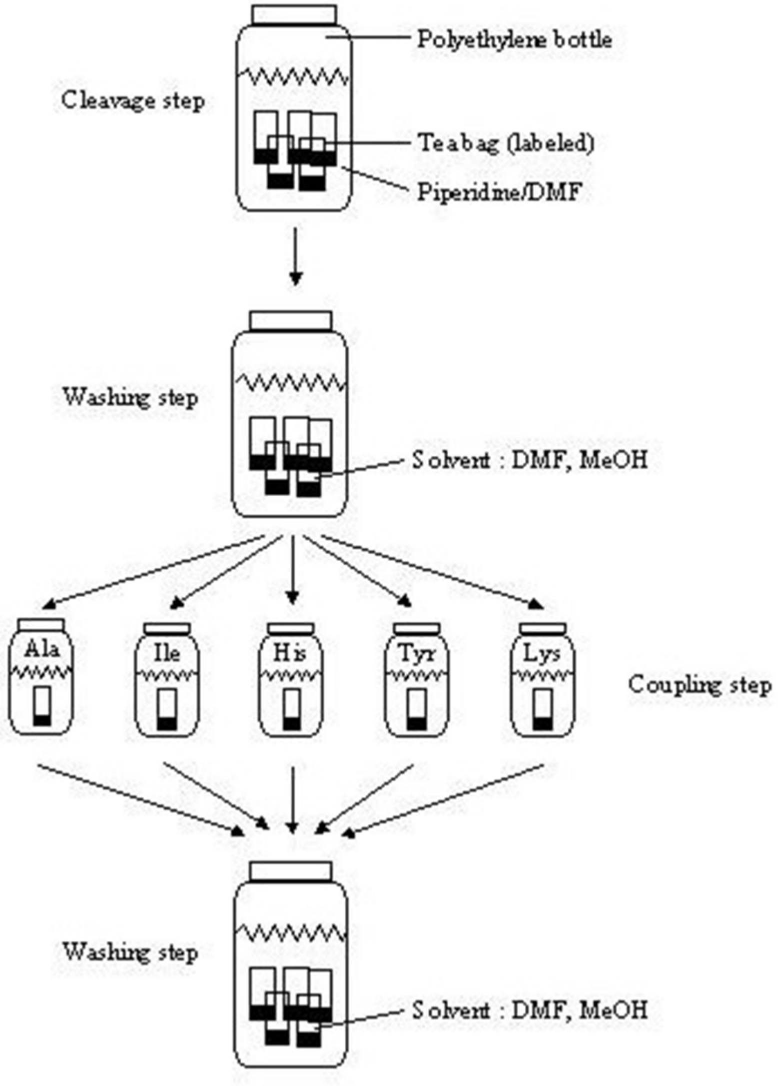


tea-bag



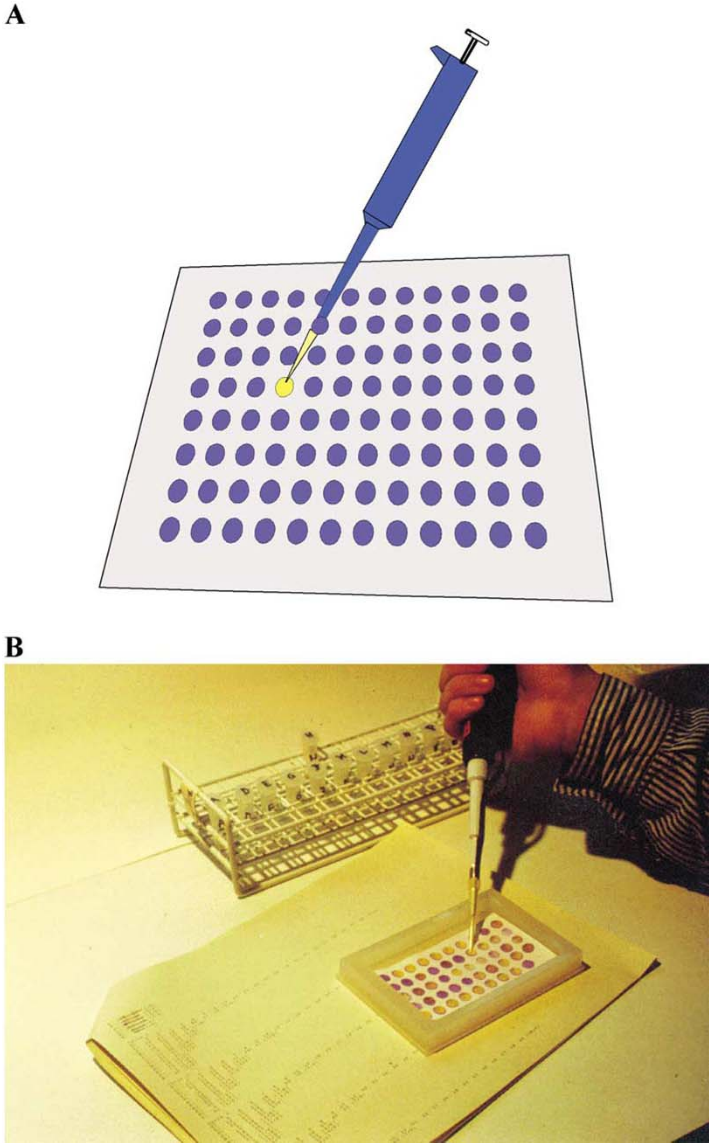
reaction vessels with bags

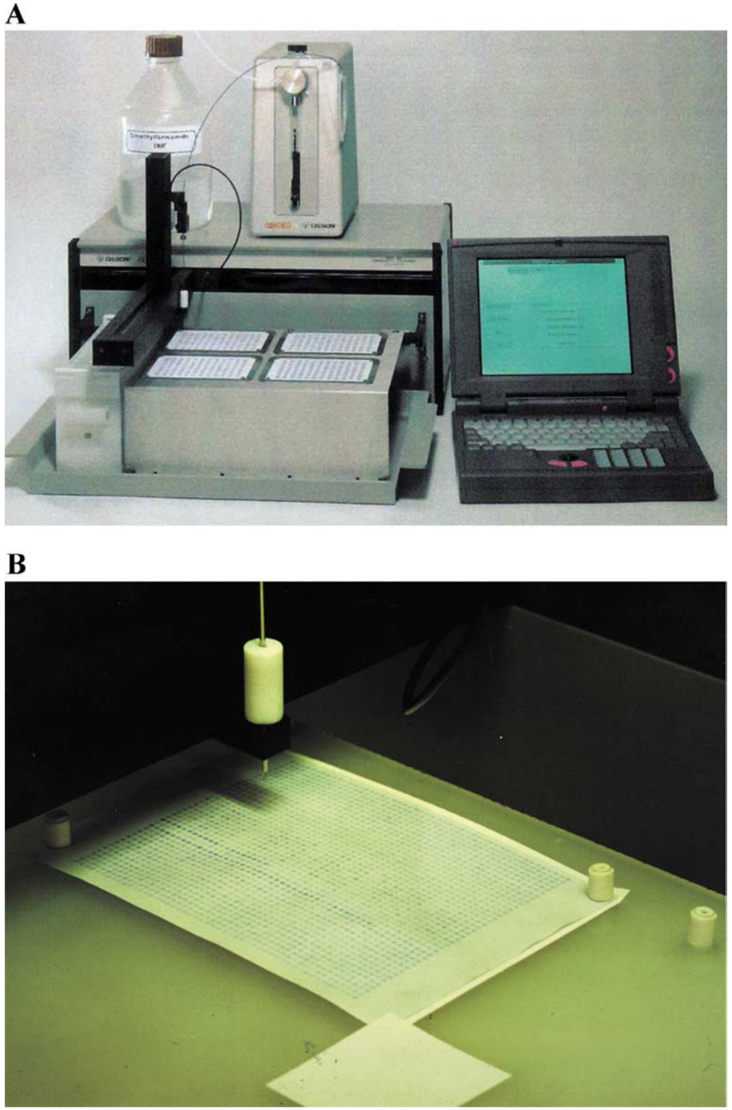
# ‘Tea bag’ parallel synthesis



**Houghten, 1984**

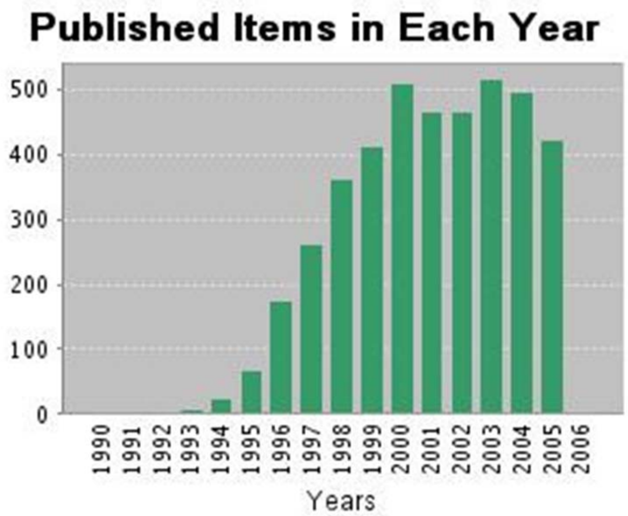
# Other techniques for parallel synthesis: SPOT synthesis





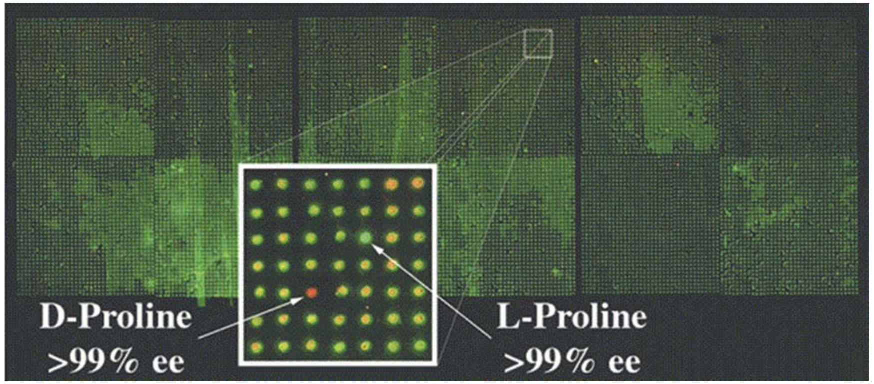
Synthesis on membranes

# Accomplishments



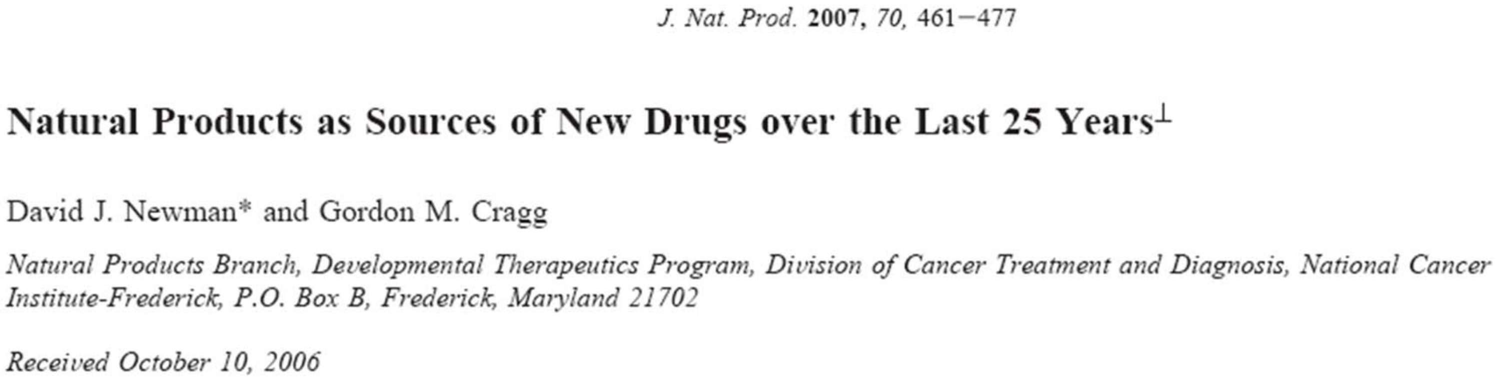
scientific production

robot synthesizers



miniaturization

# Does CombiChem work ?



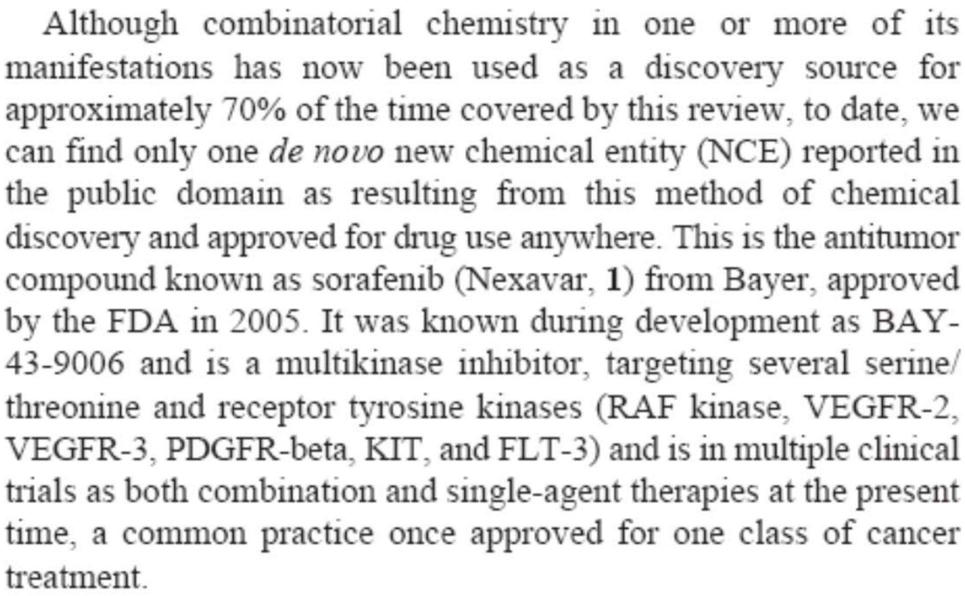
*J. Nat. Prod.* **2007,** *70,* 461−477

**Natural Products as Sources of New Drugs over the Last 25 Years**[⊥]

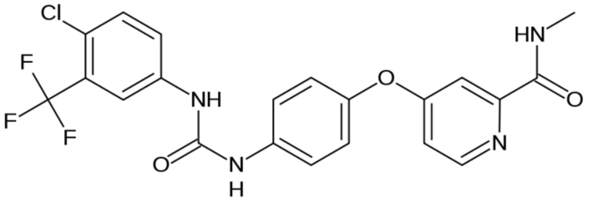
David J. Newman* and Gordon M. Cragg

*Natural Products Branch, Developmental Therapeutics Program, Division of Cancer Treatment and Diagnosis, National Cancer Institute-Frederick, P.O. Box B, Frederick, Maryland 21702*

*Received October 10, 2006*



Although combinatorial chemistry in one or more of its manifestations has now been used as a discovery source for approximately 70% of the time covered by this review, to date, we can find only one *de novo* new chemical entity (NCE) reported in the public domain as resulting from this method of chemical discovery and approved for drug use anywhere. This is the antitumor compound known as sorafenib (Nexavar, **1**) from Bayer, approved by the FDA in 2005. It was known during development as BAY-43-9006 and is a multikinase inhibitor, targeting several serine/threonine and receptor tyrosine kinases (RAF kinase, VEGFR-2, VEGFR-3, PDGFR-beta, KIT, and FLT-3) and is in multiple clinical trials as both combination and single-agent therapies at the present time, a common practice once approved for one class of cancer treatment.



sorafenib

Volume 10, Number 3 May/June 2008

## Perspective

# Application of Combinatorial Chemistry Science on Modern Drug Discovery

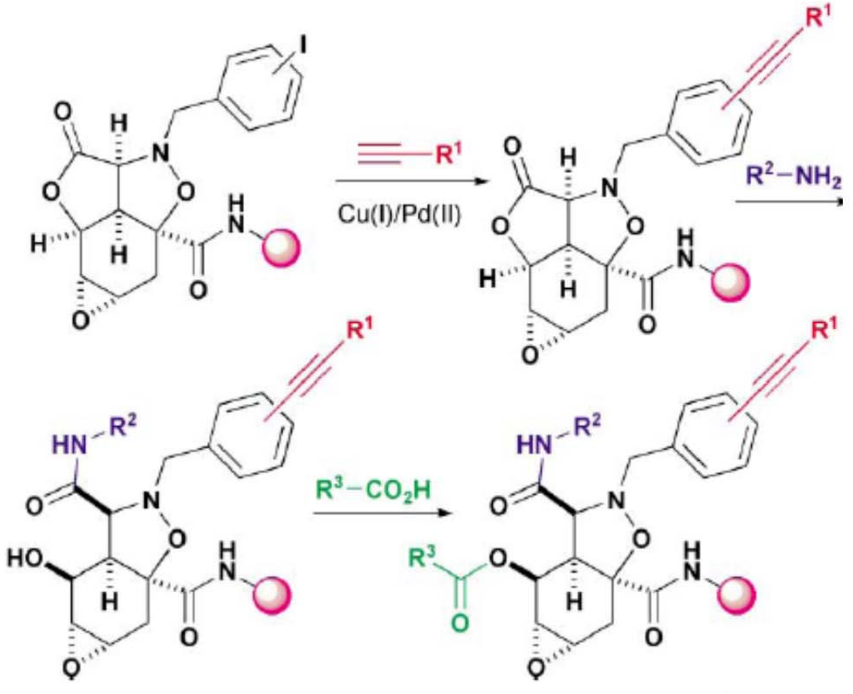
J. Phillip Kennedy, Lyndsey Williams, Thomas M. Bridges, R. Nathan Daniels,
David Weaver, and Craig W. Lindsley*

*Departments of Pharmacology and Chemistry, Vanderbilt Program in Drug Discovery, Vanderbilt Institute of Chemical Biology, Vanderbilt University Medical Center, Vanderbilt University, Nashville, Tennessee, USA 37232*

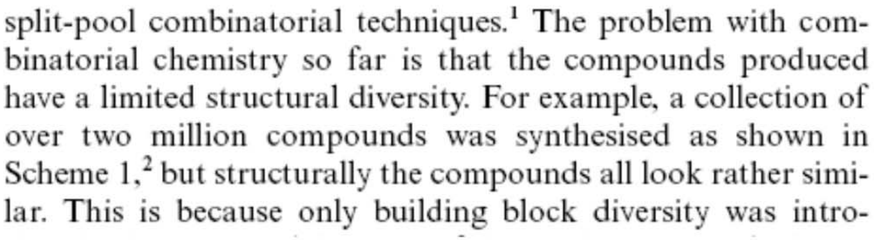
*Received November 19, 2007*

The 1990s witnessed a surge in combinatorial chemistry that infiltrated both academic and industrial laboratories. In the pharmaceutical industry, combinatorial chemistry, in the form of classical solid-phase organic synthesis and large compound libraries, promised to rapidly deliver new clinical candidates and drugs for company's struggling pipelines. By the early 2000s, it was clear to many companies that, after huge investments, combinatorial chemistry failed to deliver on its promises and most industrial combinatorial chemistry laboratories were disbanded.[1] Why? Drug discovery is not

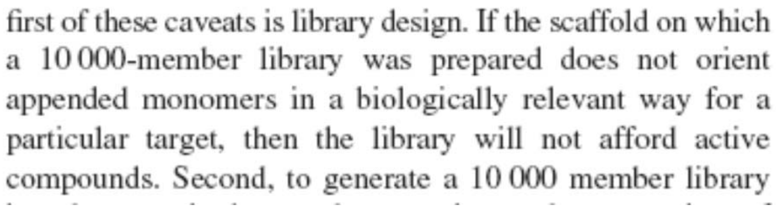
# Reason 1: Library design



**Scheme 1** Synthesis of over two million compounds.[2] The first structure shown represents three different spacers to the solid support (represented as a shaded sphere), both enantiomers and *ortho*-, *meta*- and *para*-iodo derivatives. The collection was constructed using 30 alkynes, 62 amines and 62 carboxylic acids (and skip codons).

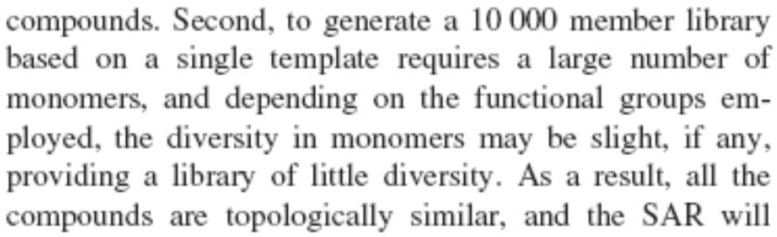


split-pool combinatorial techniques.[1] The problem with combinatorial chemistry so far is that the compounds produced have a limited structural diversity. For example, a collection of over two million compounds was synthesised as shown in Scheme 1,[2] but structurally the compounds all look rather similar. This is because only building block diversity was intro-

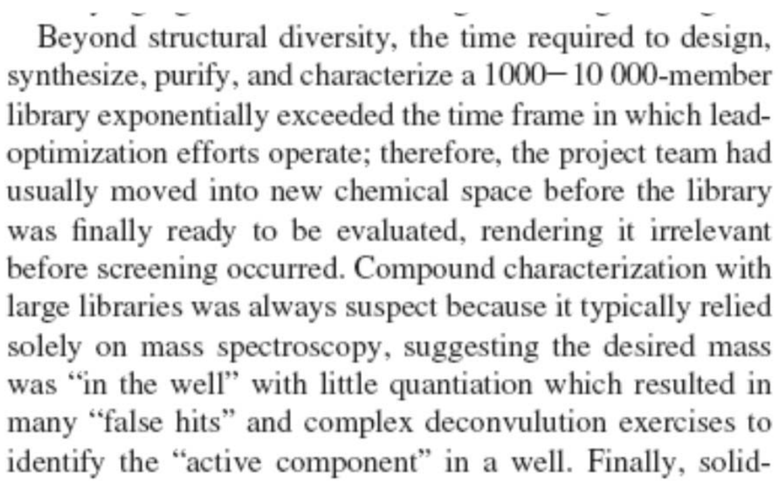


first of these caveats is library design. If the scaffold on which a 10 000-member library was prepared does not orient appended monomers in a biologically relevant way for a particular target, then the library will not afford active compounds. Second, to generate a 10 000 member library

# Other reasons

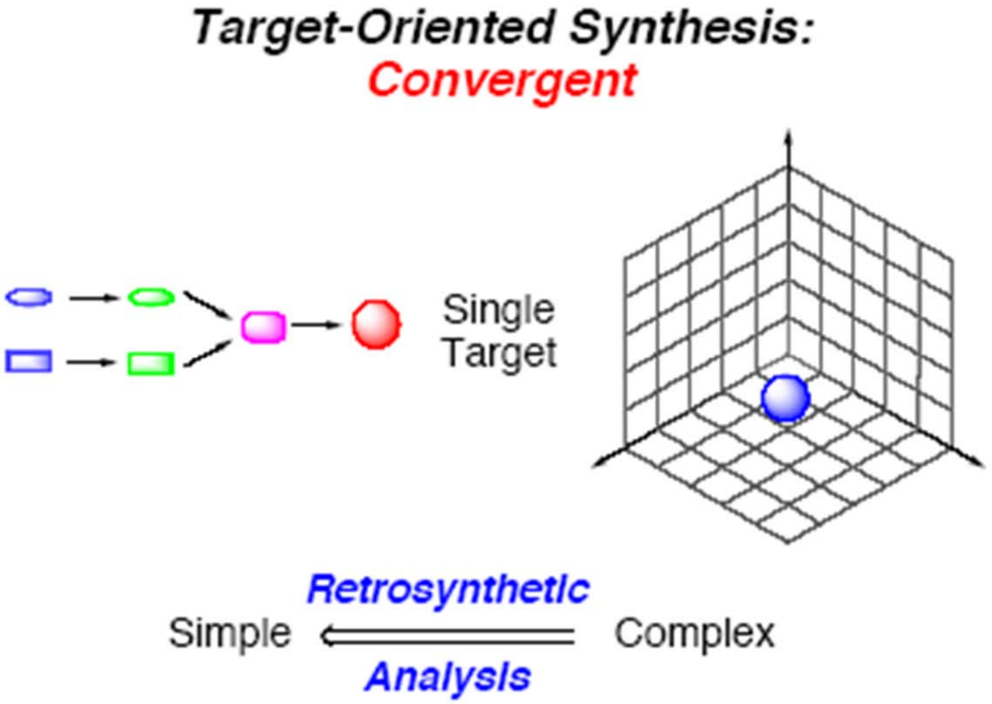


compounds. Second, to generate a 10 000 member library based on a single template requires a large number of monomers, and depending on the functional groups employed, the diversity in monomers may be slight, if any, providing a library of little diversity. As a result, all the compounds are topologically similar, and the SAR will

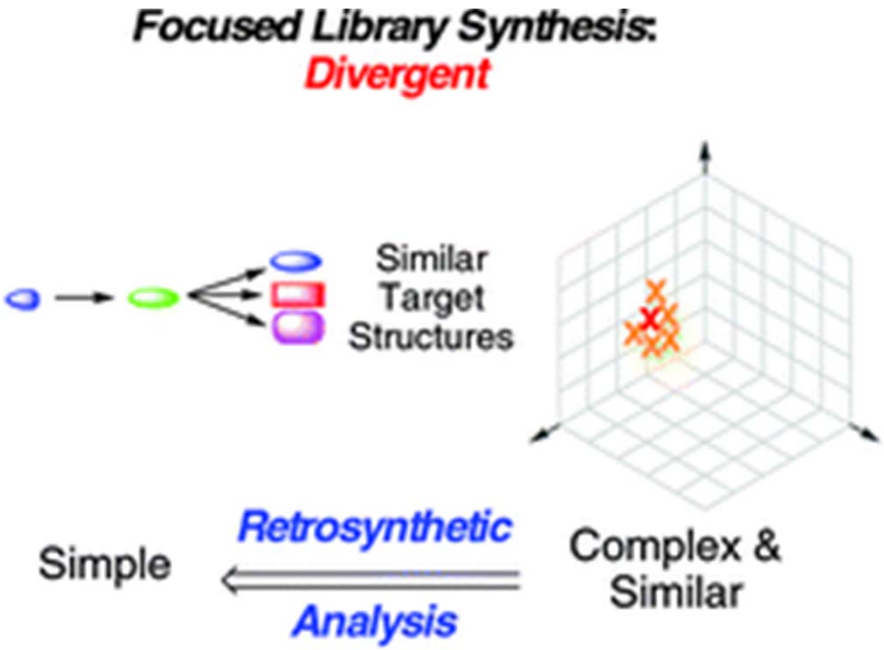


Beyond structural diversity, the time required to design, synthesize, purify, and characterize a 1000–10 000-member library exponentially exceeded the time frame in which lead-optimization efforts operate; therefore, the project team had usually moved into new chemical space before the library was finally ready to be evaluated, rendering it irrelevant before screening occurred. Compound characterization with large libraries was always suspect because it typically relied solely on mass spectroscopy, suggesting the desired mass was "in the well" with little quantiation which resulted in many "false hits" and complex deconvulution exercises to identify the "active component" in a well. Finally, solid-

# The dogma of organic chemistry: Target Oriented Synthesis



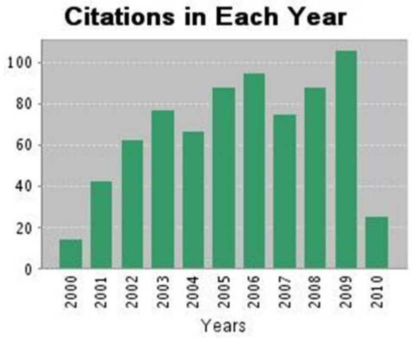
# Focused libraries



# Difficulties

- The target is not always known
  (secondary metabolites; systems biology)
- The target may have limited disponibility
- The target may be synthetically too complex
- Chemical space around the target may be too limited

# A revolutionary idea !



743 citations

20. R. P. Ahlquist, *Am. J. Physiol.* 1, 100 (1948).      *Hypertens.* 2, 162(1996).      63. I am indebted to J. C

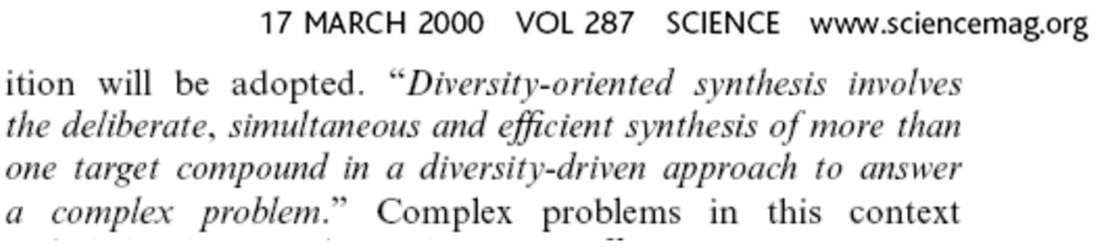
REVIEW

## Target-Oriented and Diversity-Oriented Organic Synthesis in Drug Discovery

**Stuart L. Schreiber**

Modern drug discovery often involves screening small molecules for their ability to bind to a preselected protein target. Target-oriented syntheses of these small molecules, individually or as collections (focused libraries), can be planned effectively with retrosynthetic analysis. Drug discovery can also involve screening small molecules for their ability to modulate a biological pathway in cells or organisms, without regard for any particular protein target. This process is likely to benefit in the future from an evolving forward analysis of synthetic pathways, used in diversity-oriented synthesis, that leads to structurally complex and diverse small molecules. One goal of diversity-oriented syntheses is to synthesize efficiently a collection of small molecules capable of perturbing any disease-related biological pathway, leading eventually to the identification of therapeutic protein targets capable of being modulated by small molecules. Several synthetic planning principles for diversity-oriented synthesis and their role

provided the means to synthesize not only single target compounds or collections of related targets but also collections of structurally diverse compounds. Target-oriented syntheses are used in drug discovery efforts involving preselected protein targets, whereas diversity-oriented syntheses are used in efforts to identify simultaneously therapeutic protein targets and th
ulators. Target-orien
fited from a power
named retrosynthetic
rable algorithm for di
sis is only now begi

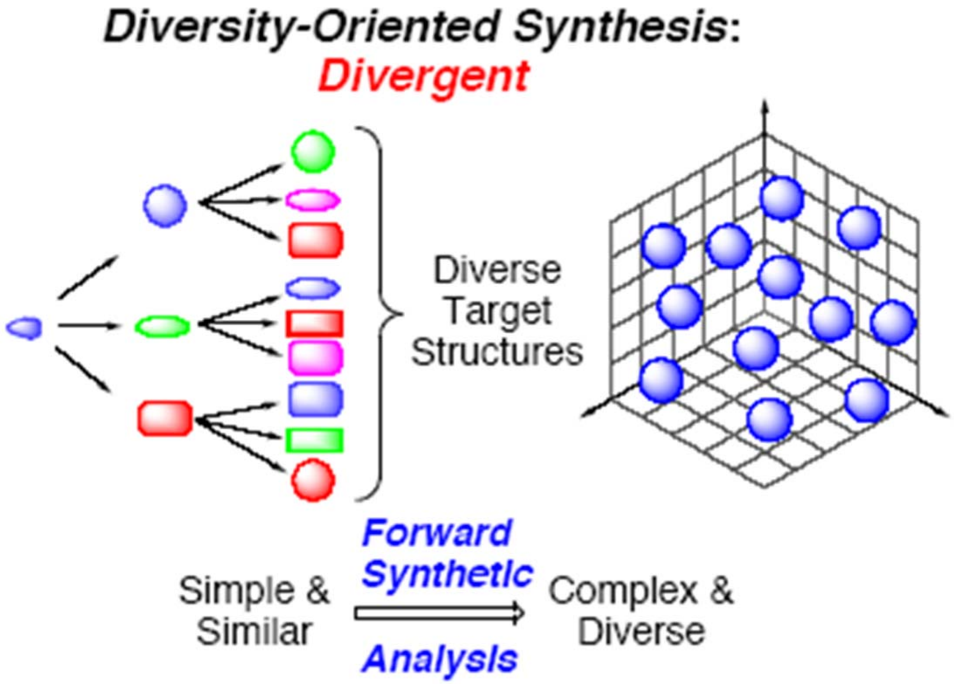


17 MARCH 2000 VOL 287 SCIENCE www.sciencemag.org

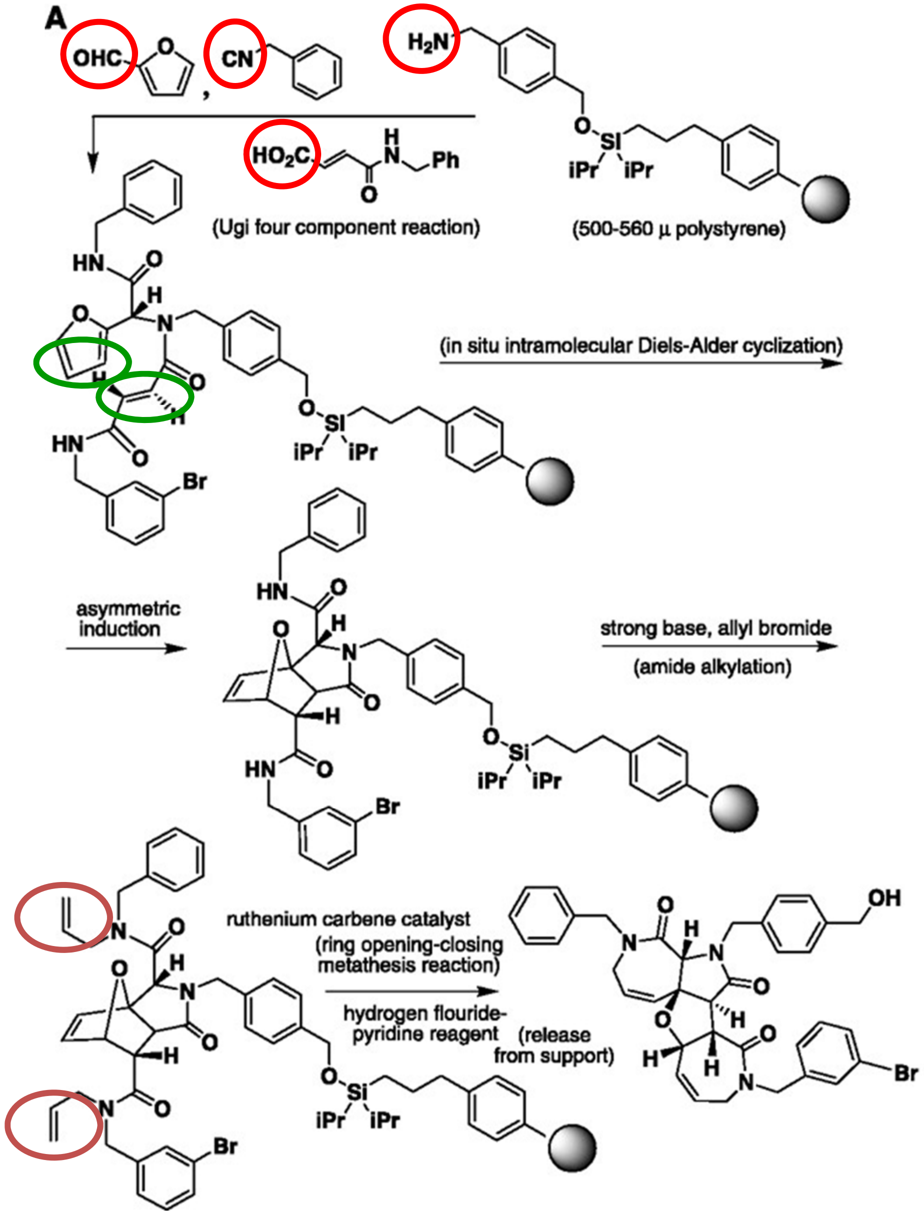
ition will be adopted. “*Diversity-oriented synthesis involves the deliberate, simultaneous and efficient synthesis of more than one target compound in a diversity-driven approach to answer a complex problem.*” Complex problems in this context

Stuart Schreiber
Harvard, USA

# Explore Chemical Space

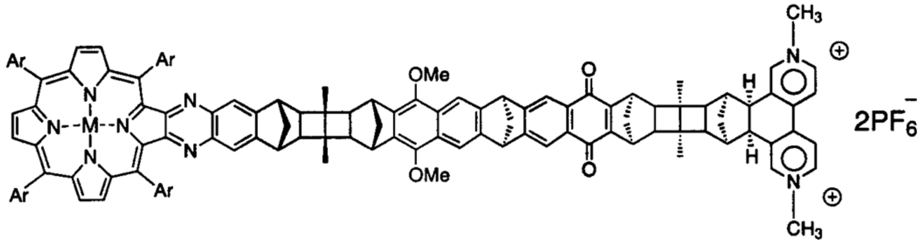


# Synthetic strategies

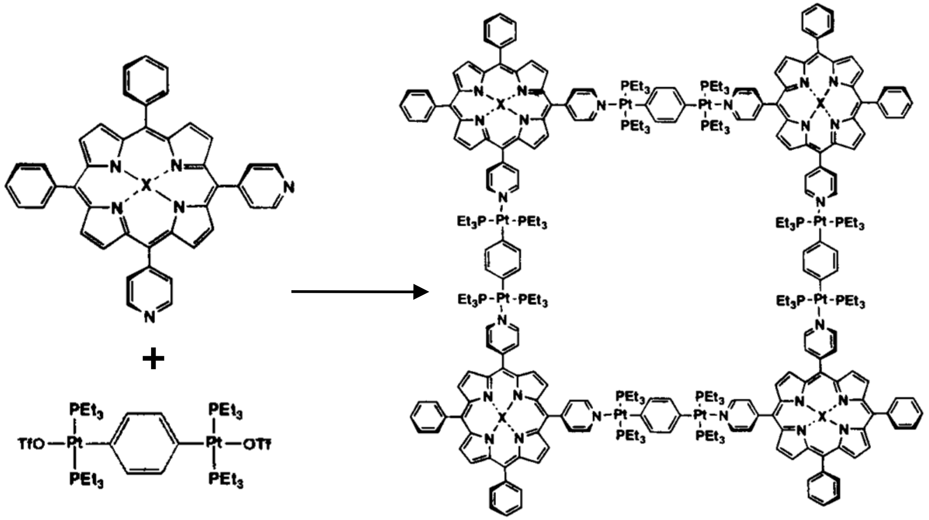


tandem
complexity-generating
reactions

# Covalent versus Noncovalent Synthesis



Paddon-Row *et al.* *Chem.Eur.J.* **1999**, 2518



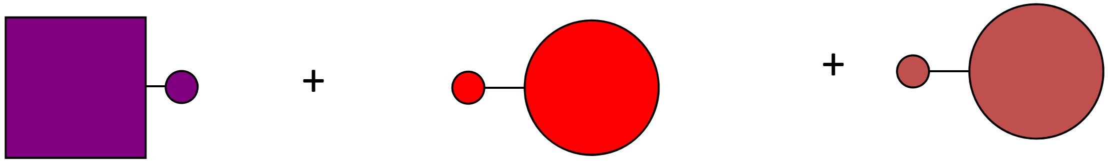
Stang *et al.* *JACS* **1999**, 2741

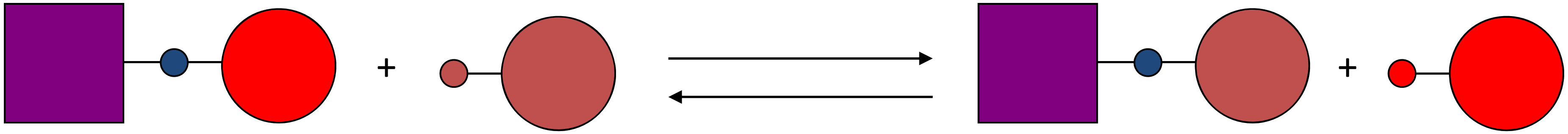
covalent ——— Characteristics ——— noncovalent

| | | |
|---|---|---|
| building block | atom | molecule |
| target | molecules | assemblies |
| bond type | covalent | ionic, hydrophobic, metal-coordination, H-bond |
| bond energy | 35-135 kcal/mol | 2 - 20 kcal/mol |
| kinetic stability | high | low |

Angewandte Chemie, **2001,** 2383

# Equilibrium reactions



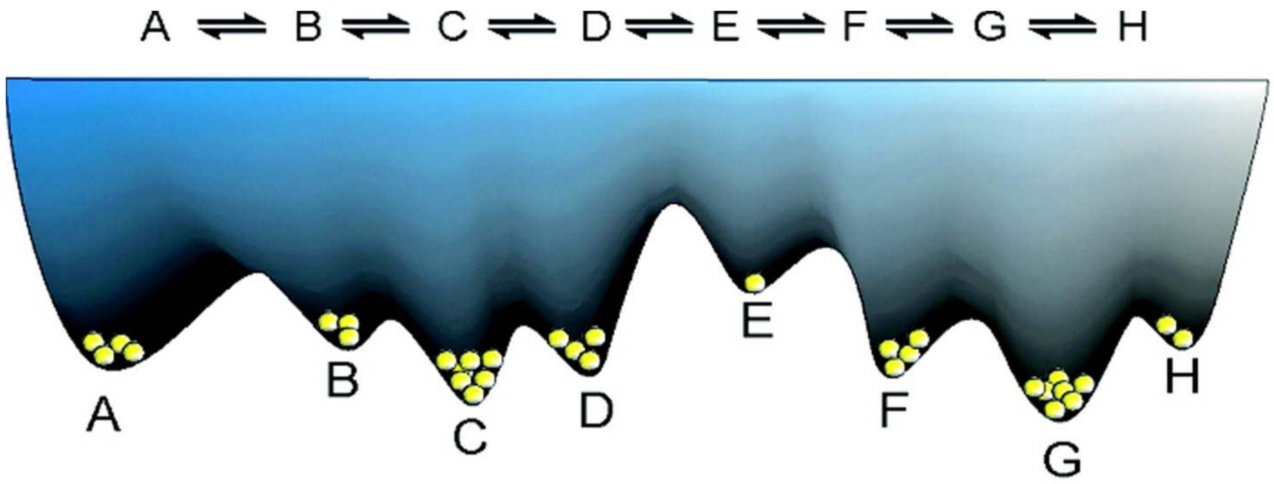


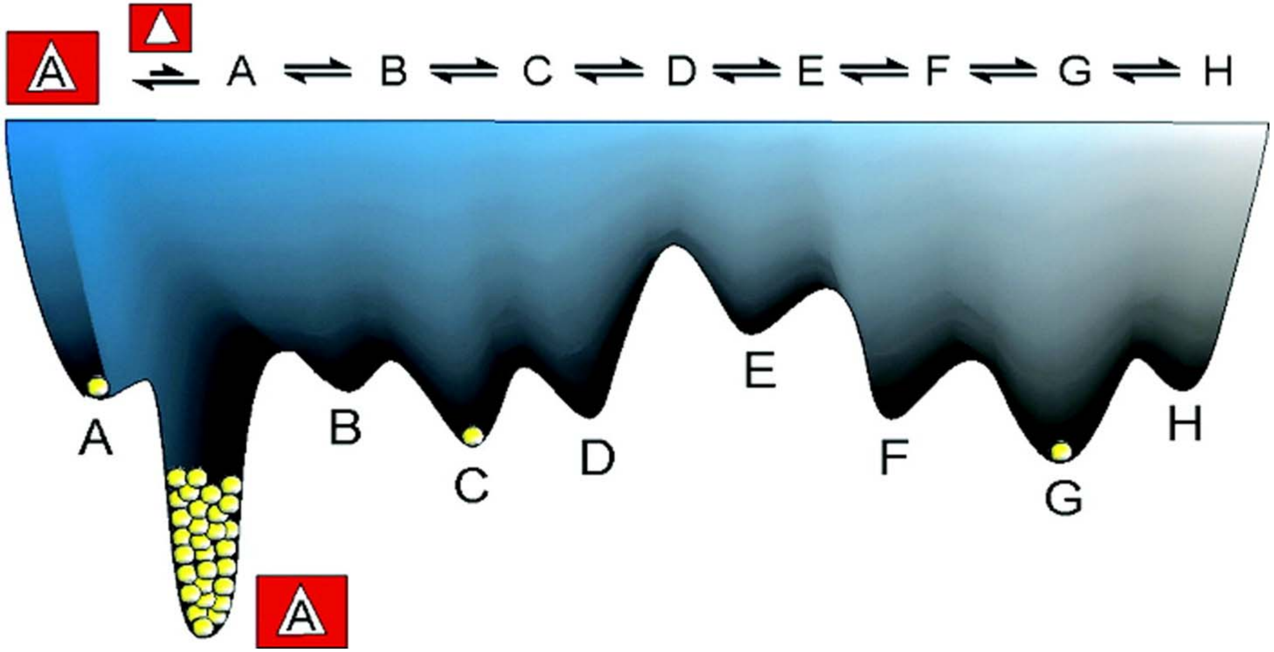
Henry Louis Le Chatelier
(1850-1936)

## Le Chatelier’s principle :

If a chemical system at equilibrium experiences a change in concentration, temperature, volume, or partial pressure, then the equilibrium shifts to counteract the imposed change and a new equilibrium is established.

# An idea in the mid-90ies

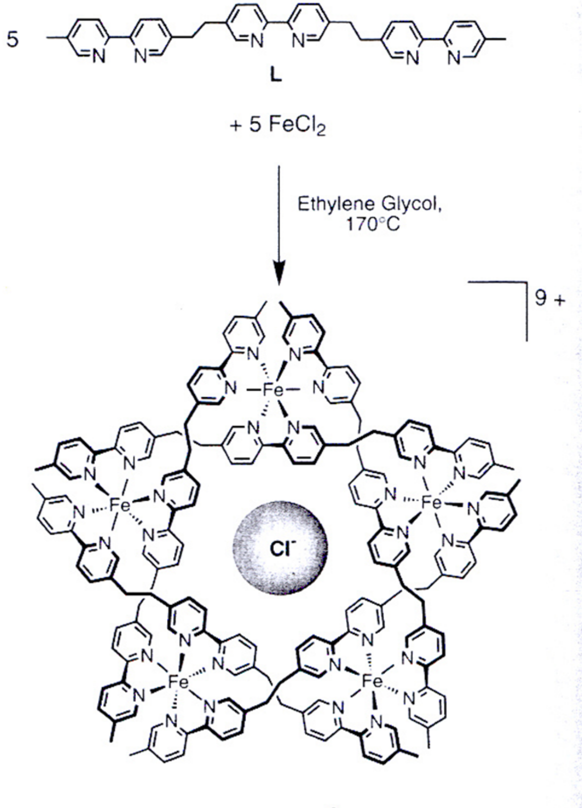




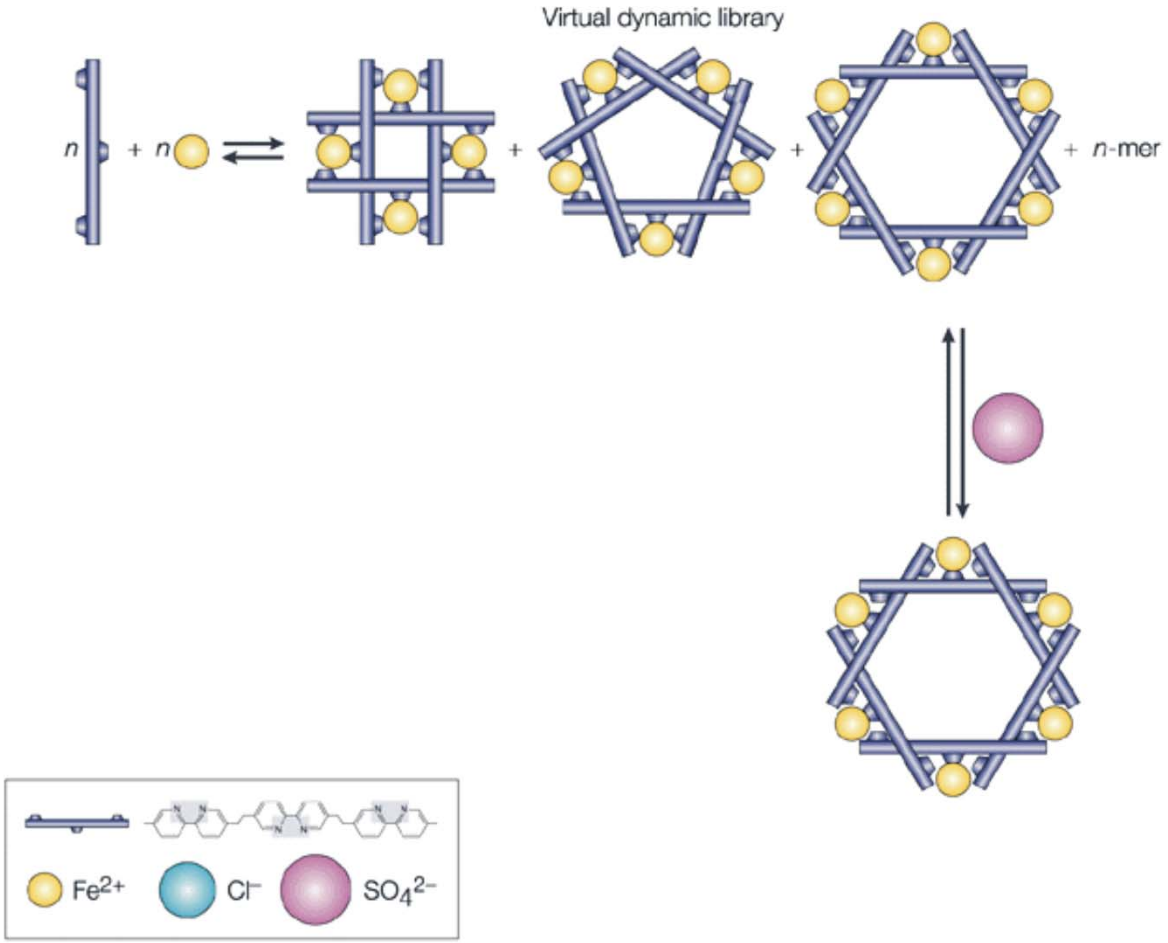
Jean-Marie Lehn
(Nobel Prize 1987)

Jeremy Sanders

# A first example

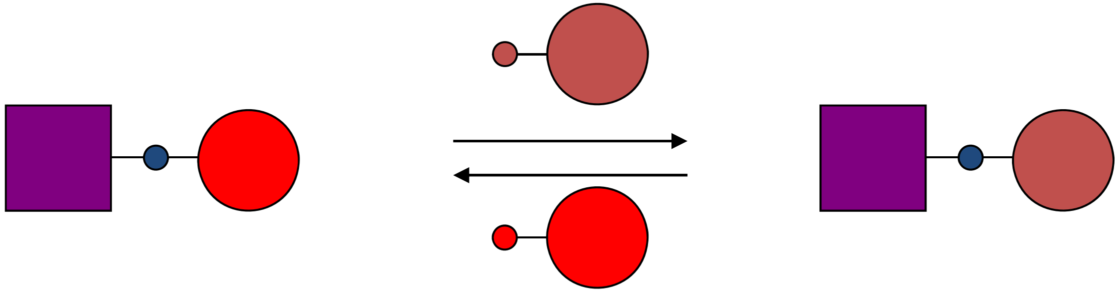


$[(Fe_5L_5)Cl]^{9+}$



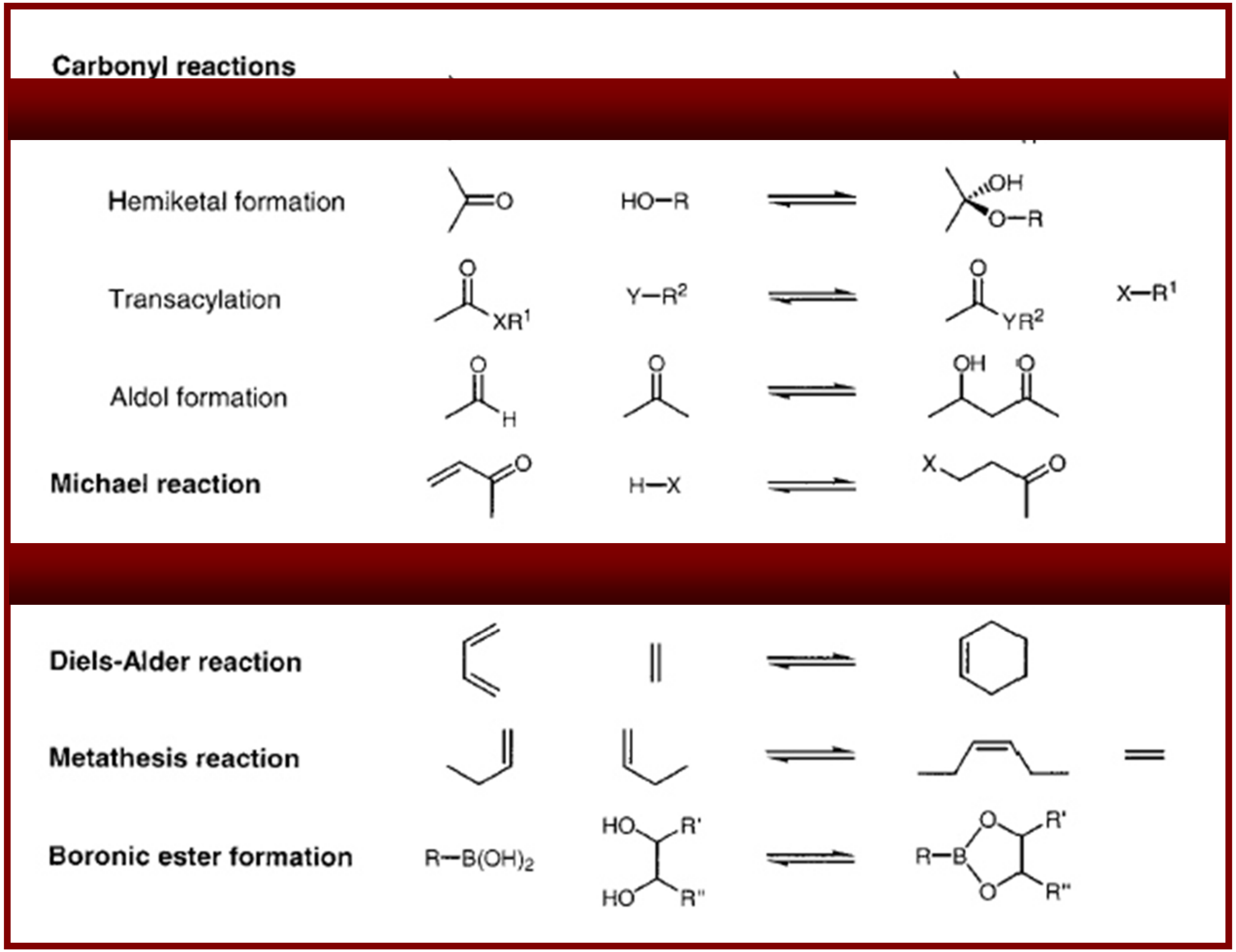
Lehn, Angewandte Chemie, **1996**

# Criteria for the exchange reaction



- reversible on a reasonable time scale
- compatability with the screening conditions
  (functional groups, solvent, pH, etc)
- mild reaction conditions
  (temperature, pressure, concentration)
- solubility of all library members under all conditions
- possibility to turn off the reaction – stop exchange
  for analysis and isolation
- isoenergetic library members (ideally)

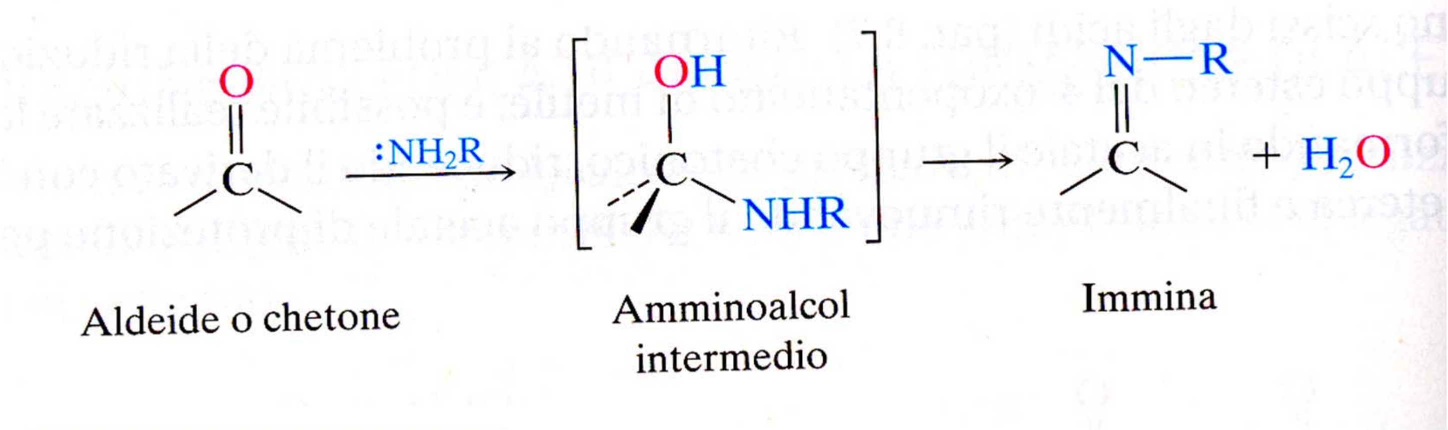
# Reversible covalent bond formation

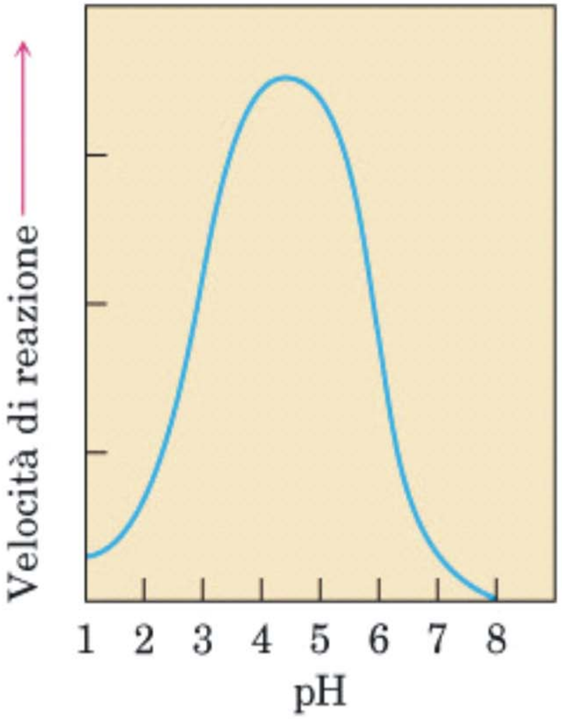


[redacted]

[redacted]

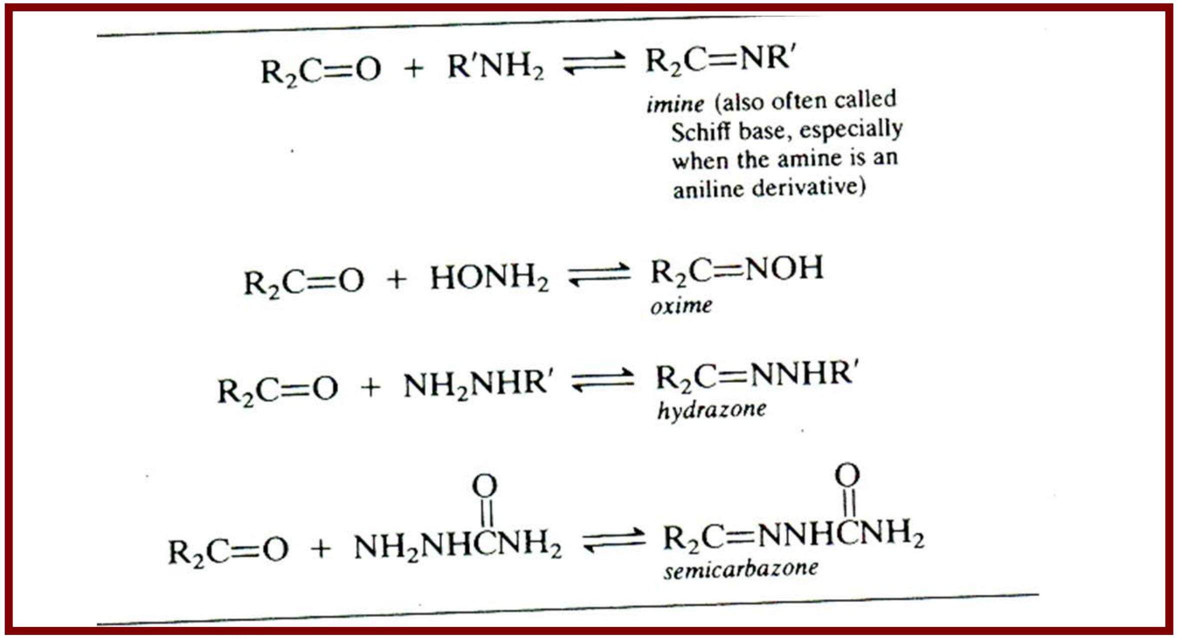
# Imine formation





imine formation is acid catalysed

# More stable analogs of imines



$$R_2C{=}O + R'NH_2 \rightleftharpoons R_2C{=}NR'$$

*imine* (also often called Schiff base, especially when the amine is an aniline derivative)

$$R_2C{=}O + HONH_2 \rightleftharpoons R_2C{=}NOH$$

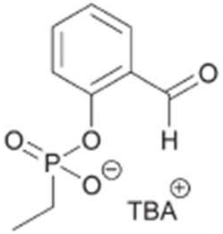
*oxime*

$$R_2C{=}O + NH_2NHR' \rightleftharpoons R_2C{=}NNHR'$$

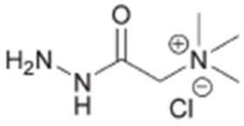
*hydrazone*

$$R_2C{=}O + NH_2NH\overset{\overset{\displaystyle O}{\|}}{C}NH_2 \rightleftharpoons R_2C{=}NNH\overset{\overset{\displaystyle O}{\|}}{C}NH_2$$

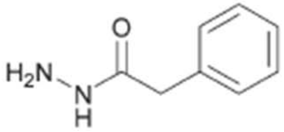
*semicarbazone*



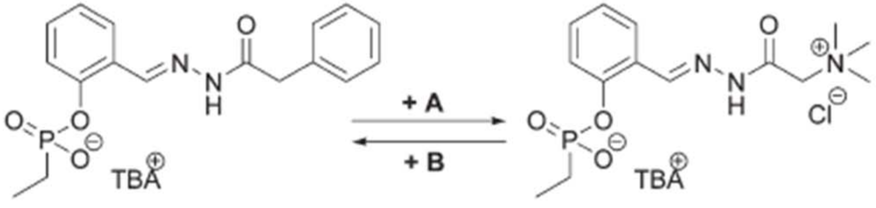
**1**



**A**



**B**

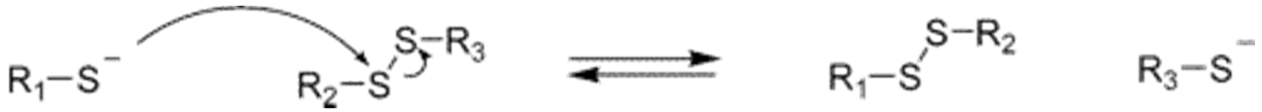


**1B** **1A**

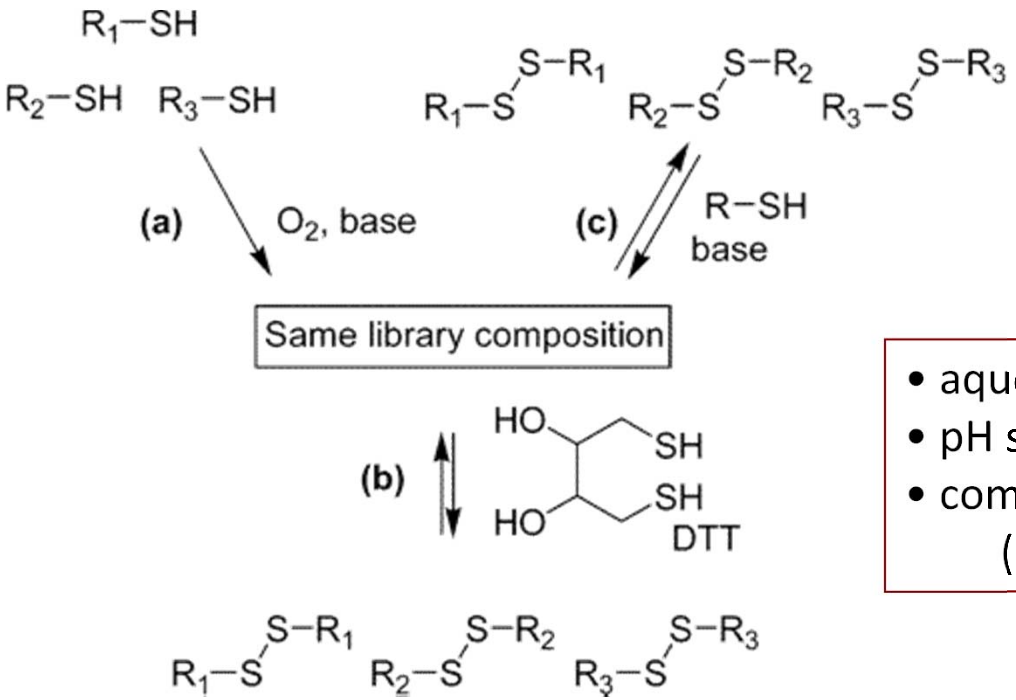
Imine exchange is generally ‘frozen’ by reduction to amine.

# Disulfide exchange

exchange



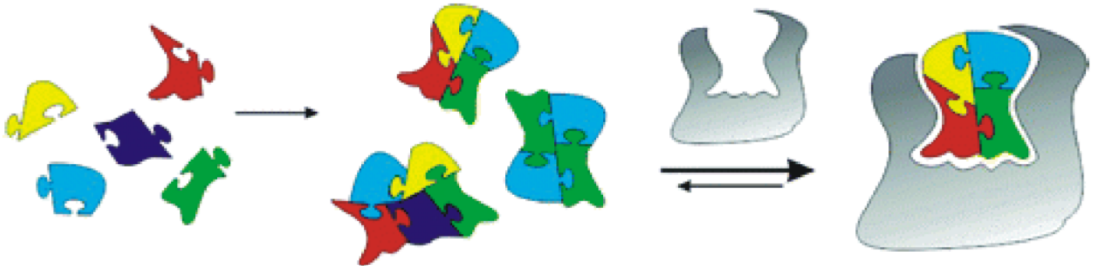
dynamic libraries



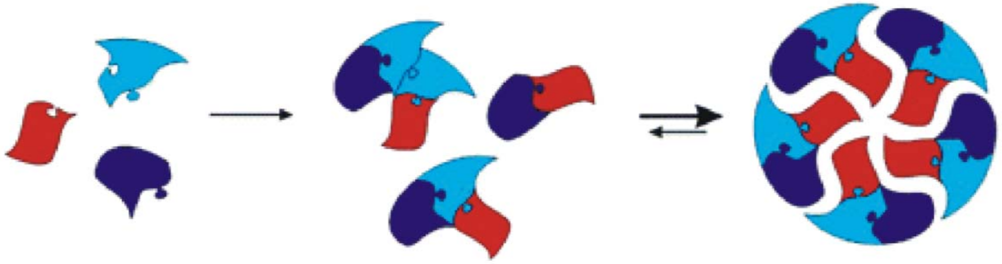
- aqueous solutions
- pH sensitive
- compatible with functional groups (biomolecules)

# Applications of DCC

Receptors



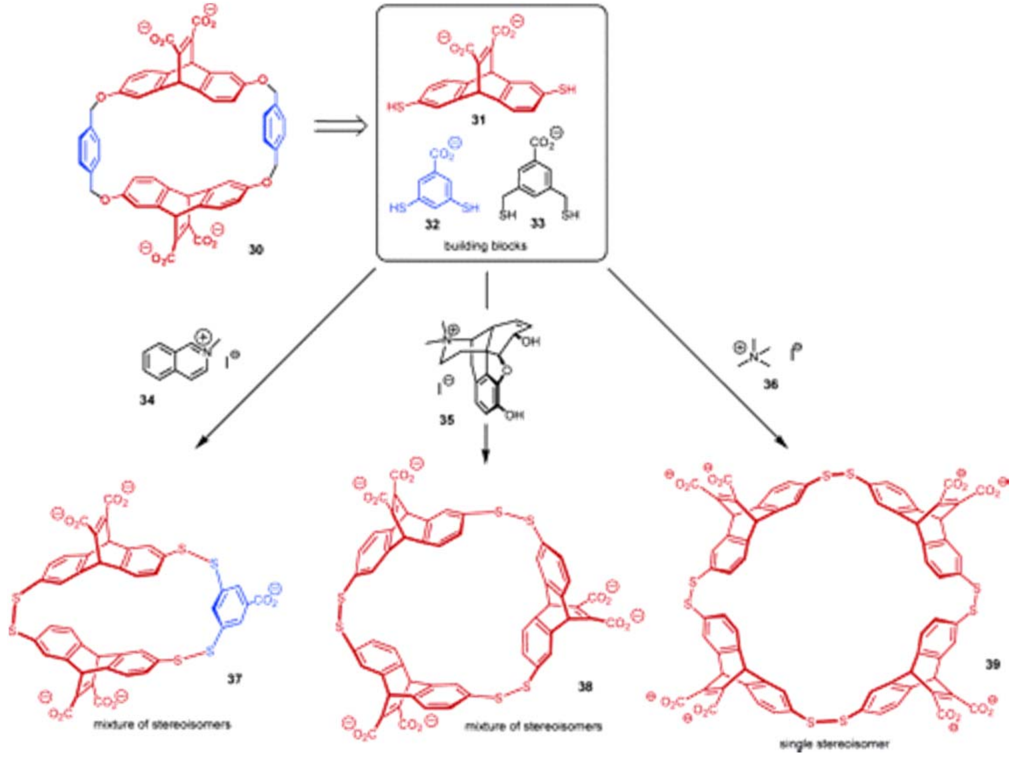
Guests

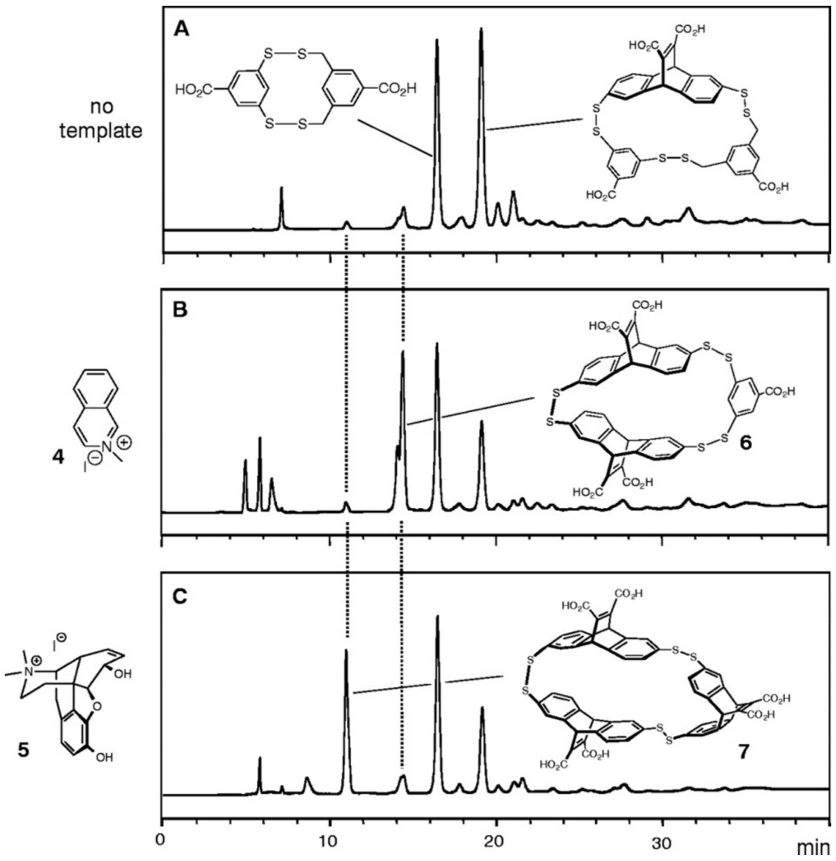


Materials

Sanders, Otto, ChemRev, **2006,** 3652

# Receptors

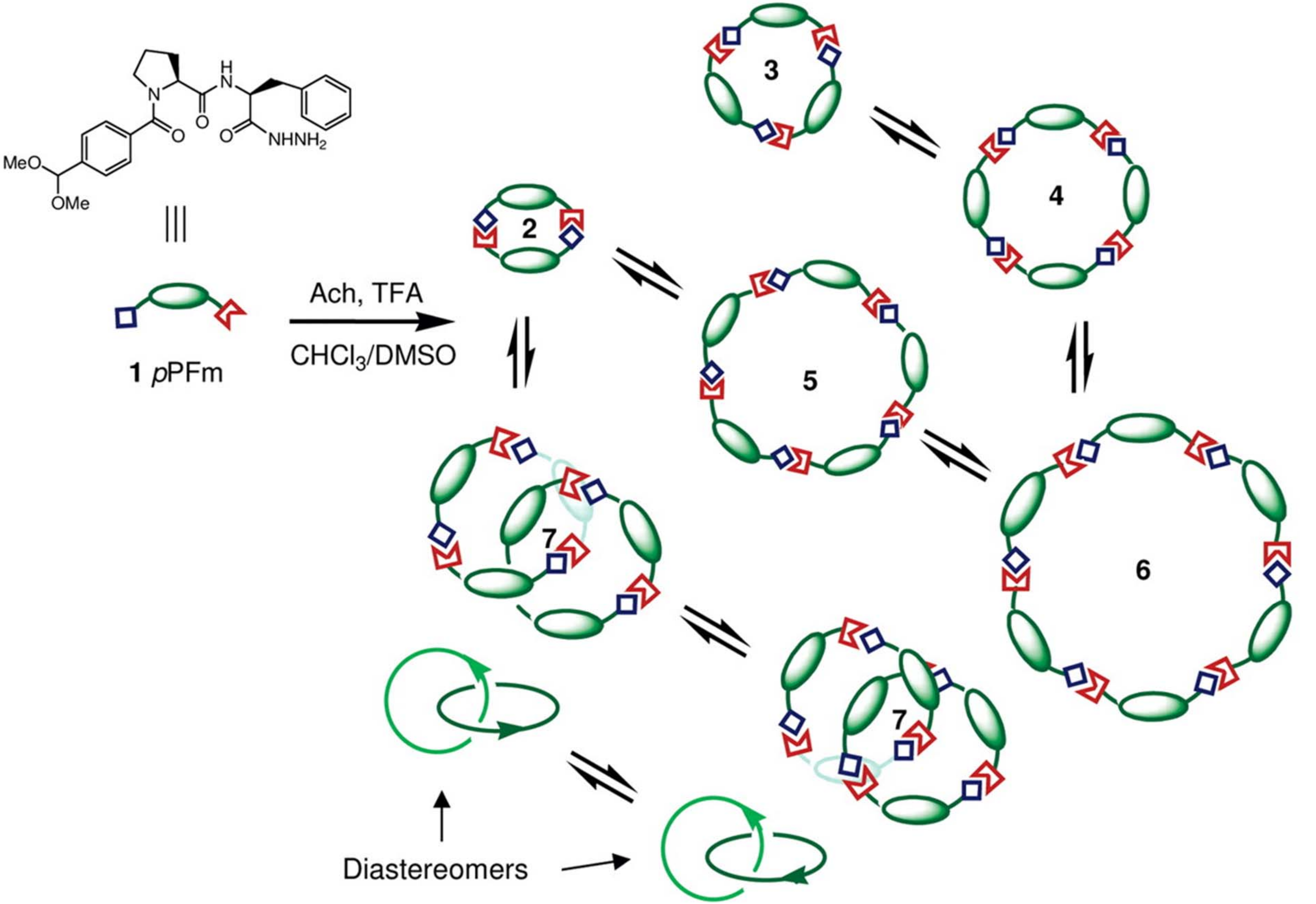




$K_{ass} = 2.5 \times 10^5\ M^{-1}$ $K_{ass} = 7.1 \times 10^5\ M^{-1}$

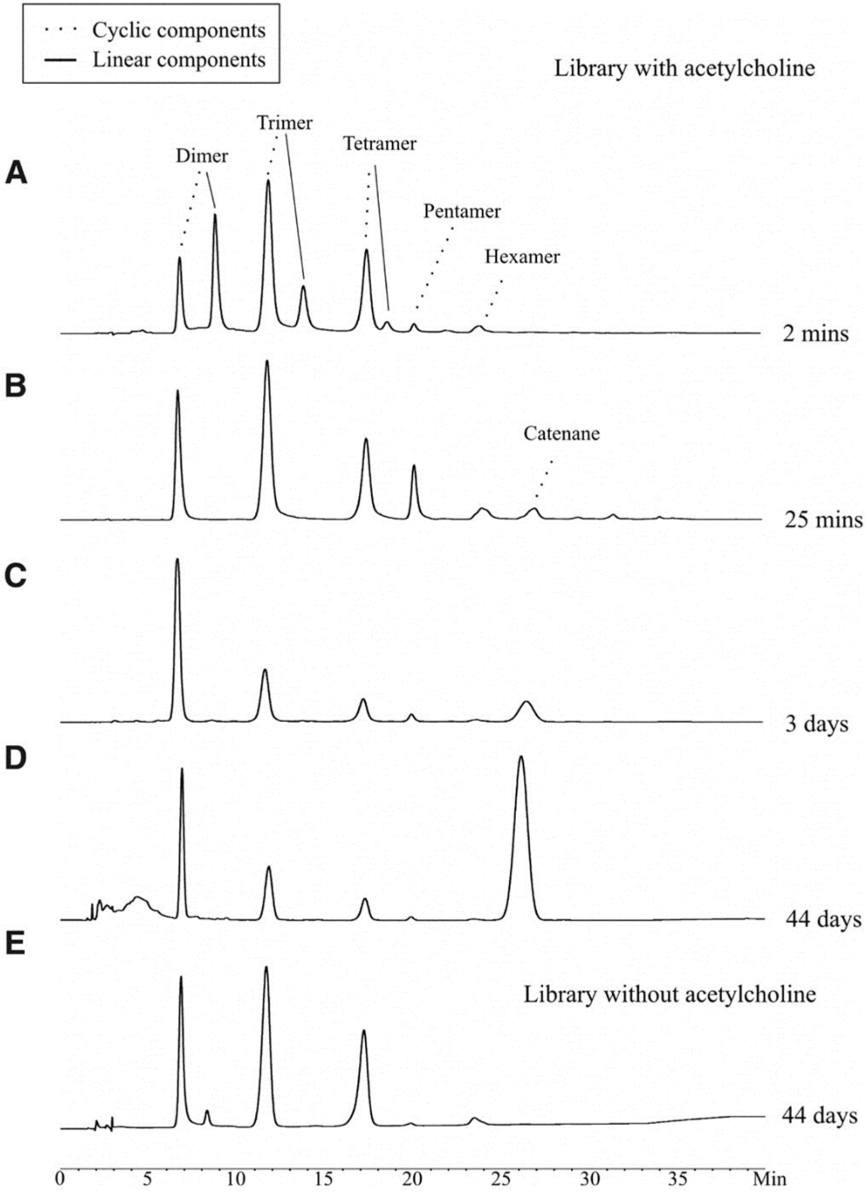
Sanders, Science, **2002**

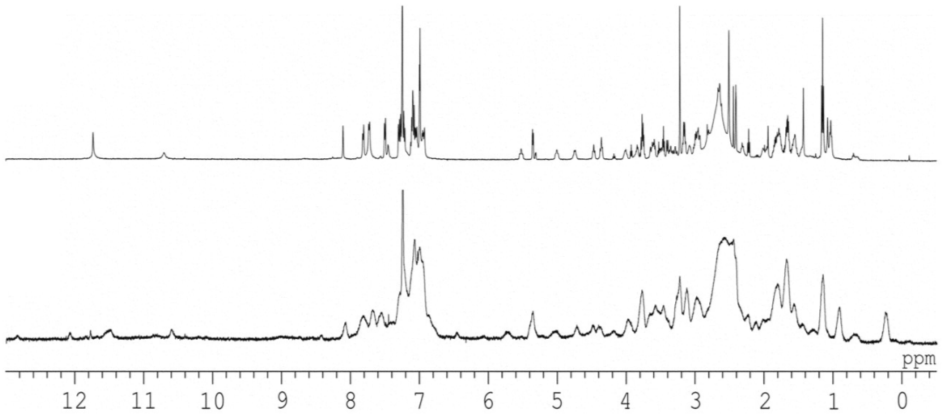
# A self-selected catenane

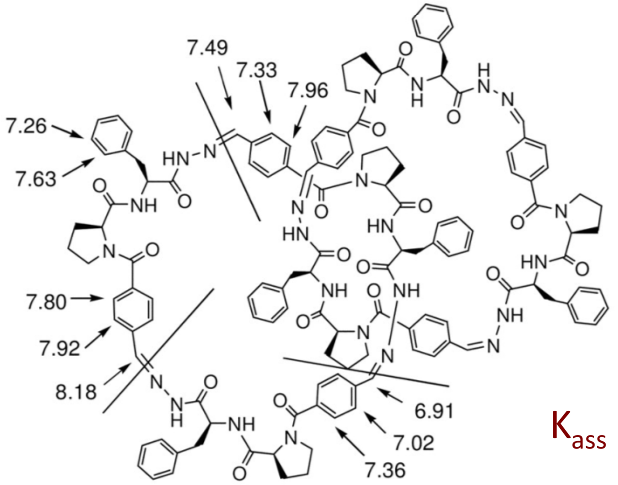


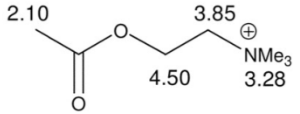
Sanders, Otto, Science, **2005**

# Library shift



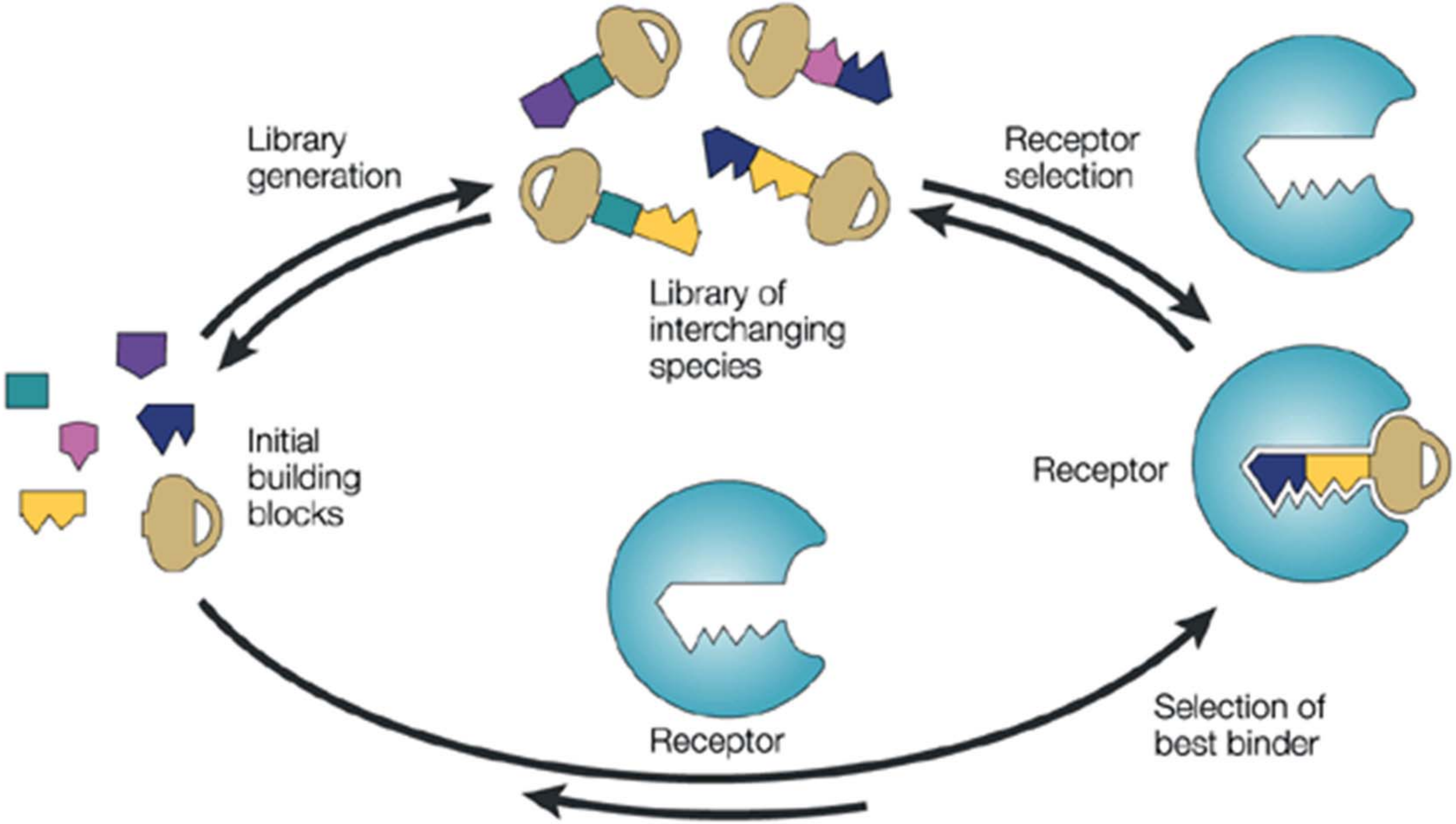






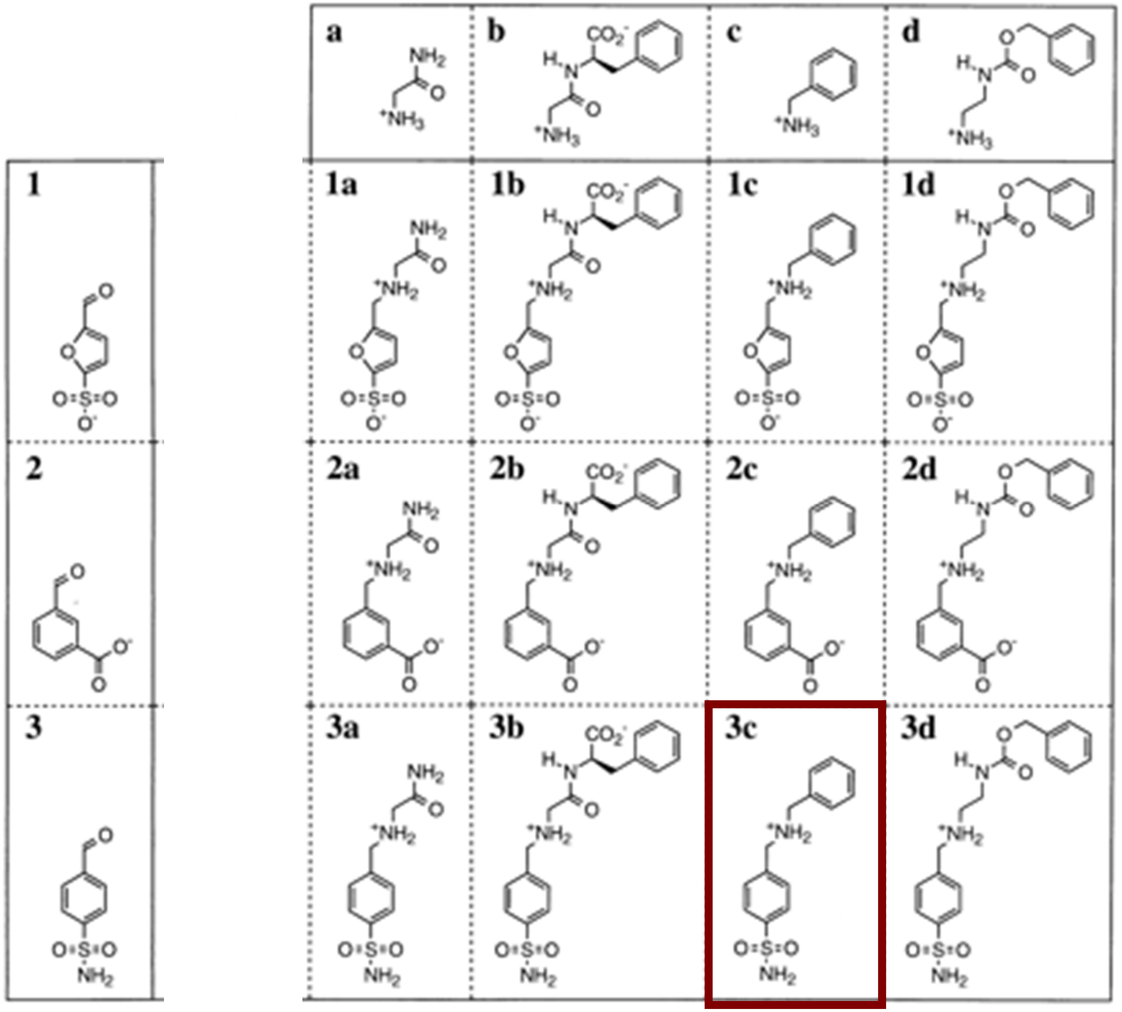
$K_{ass} = 1.4 \times 10^{7}\ M^{-1}$

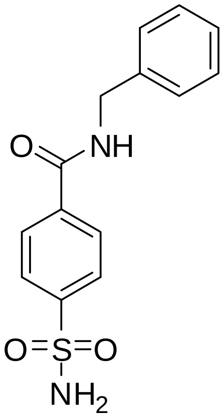
# Guest Development



Lehn, NatRevDrugDiscov, **2002**

# A DCL of potential inhibitors for Carbonic Anhydrase

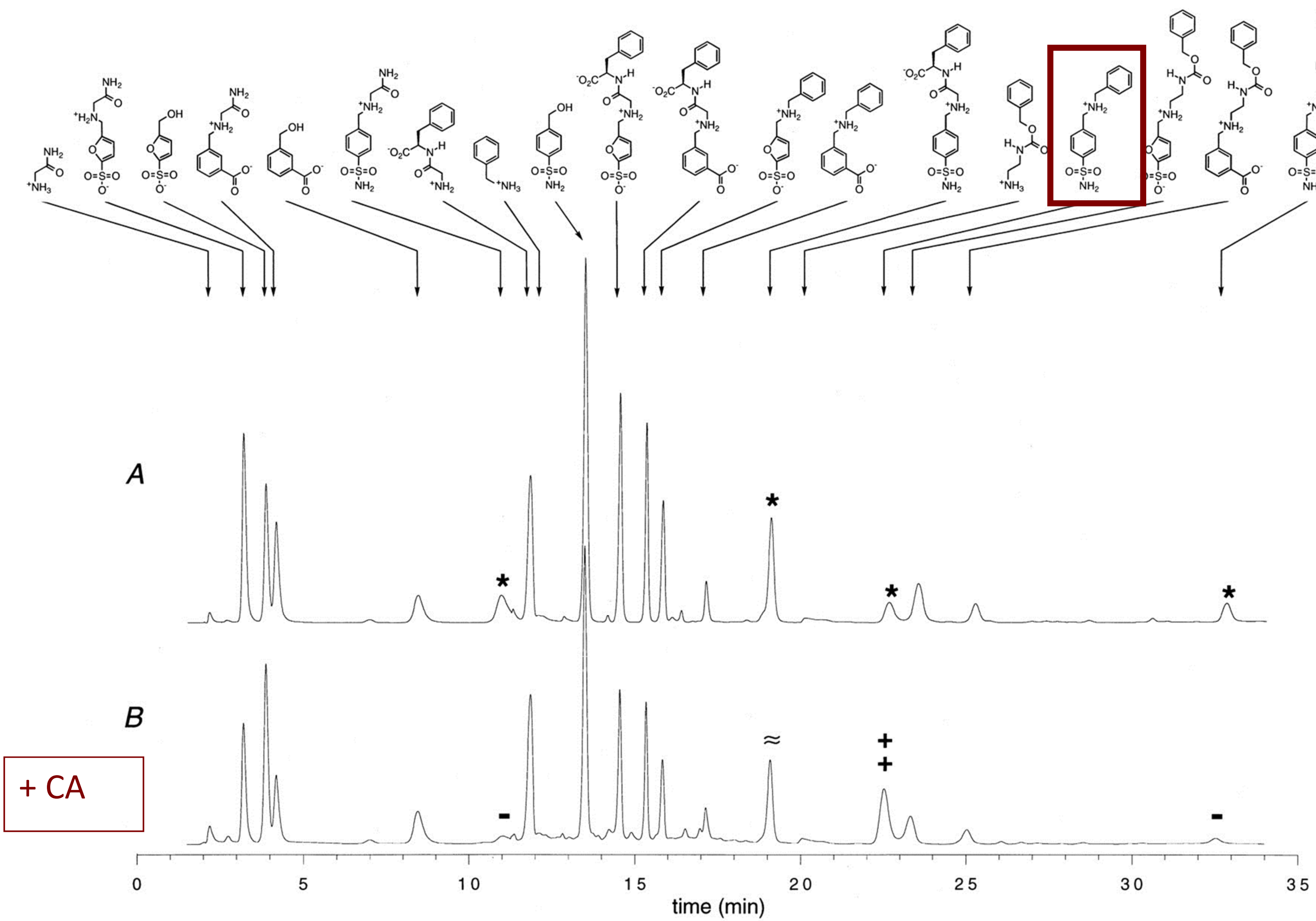




known inhibitor for CA with $K_d$ = 1.1 nM

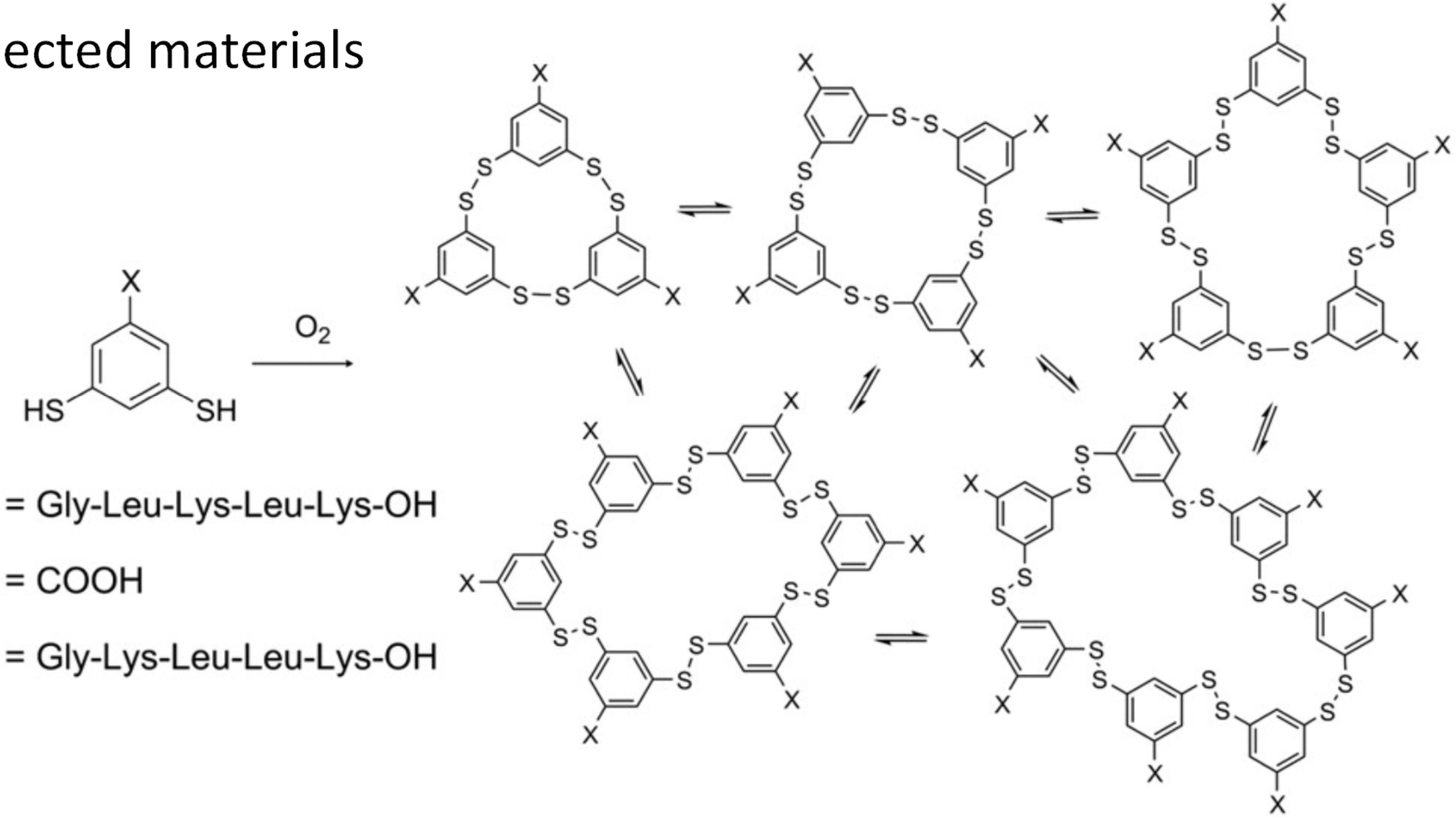
Lehn, PNAS, **1997**

# HPLC Analysis



+ CA

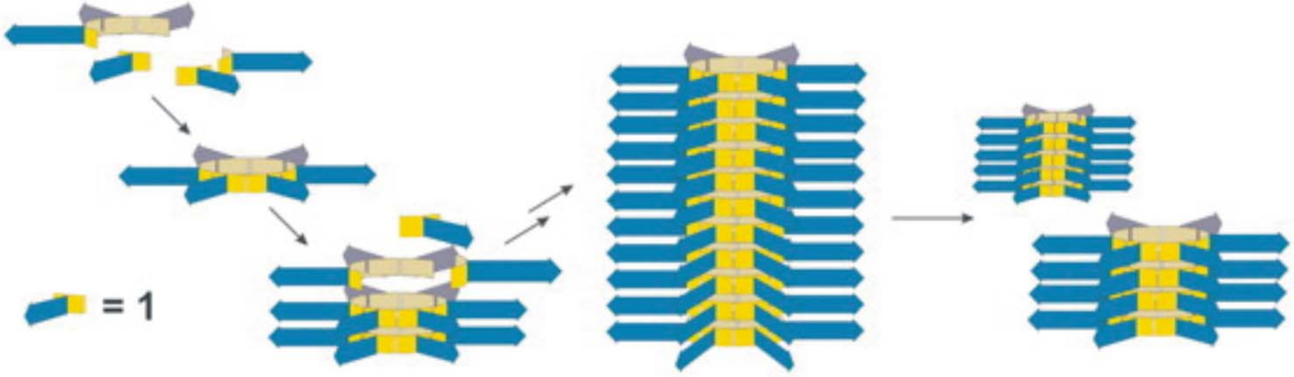
# Self-selected materials



**1** X = Gly-Leu-Lys-Leu-Lys-OH

**2** X = COOH

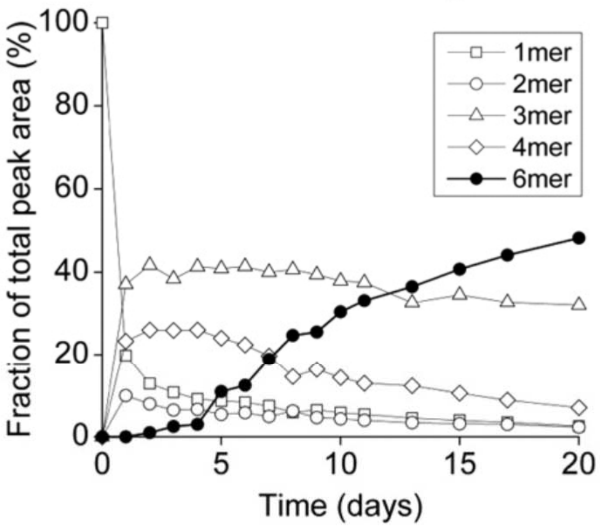
**3** X = Gly-Lys-Leu-Leu-Lys-OH



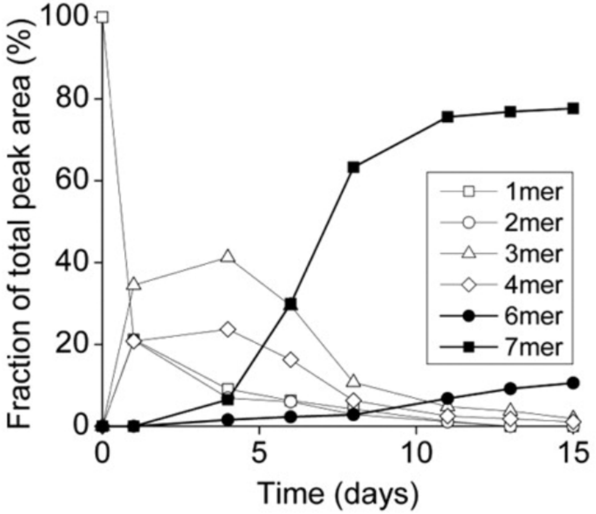
Otto, Science, **2010**

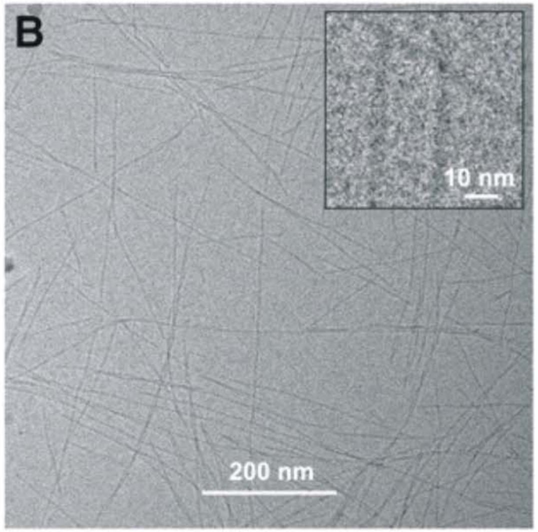
# Shaken or stirred

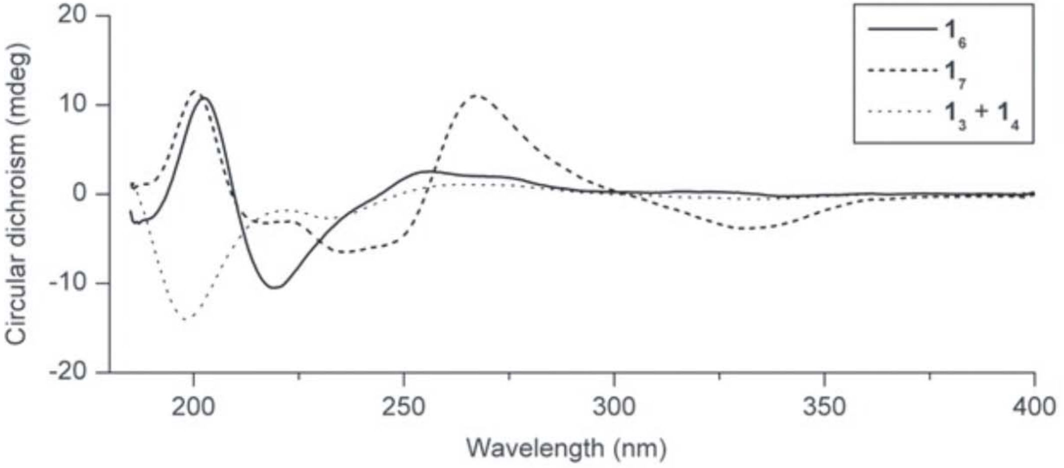
shaken



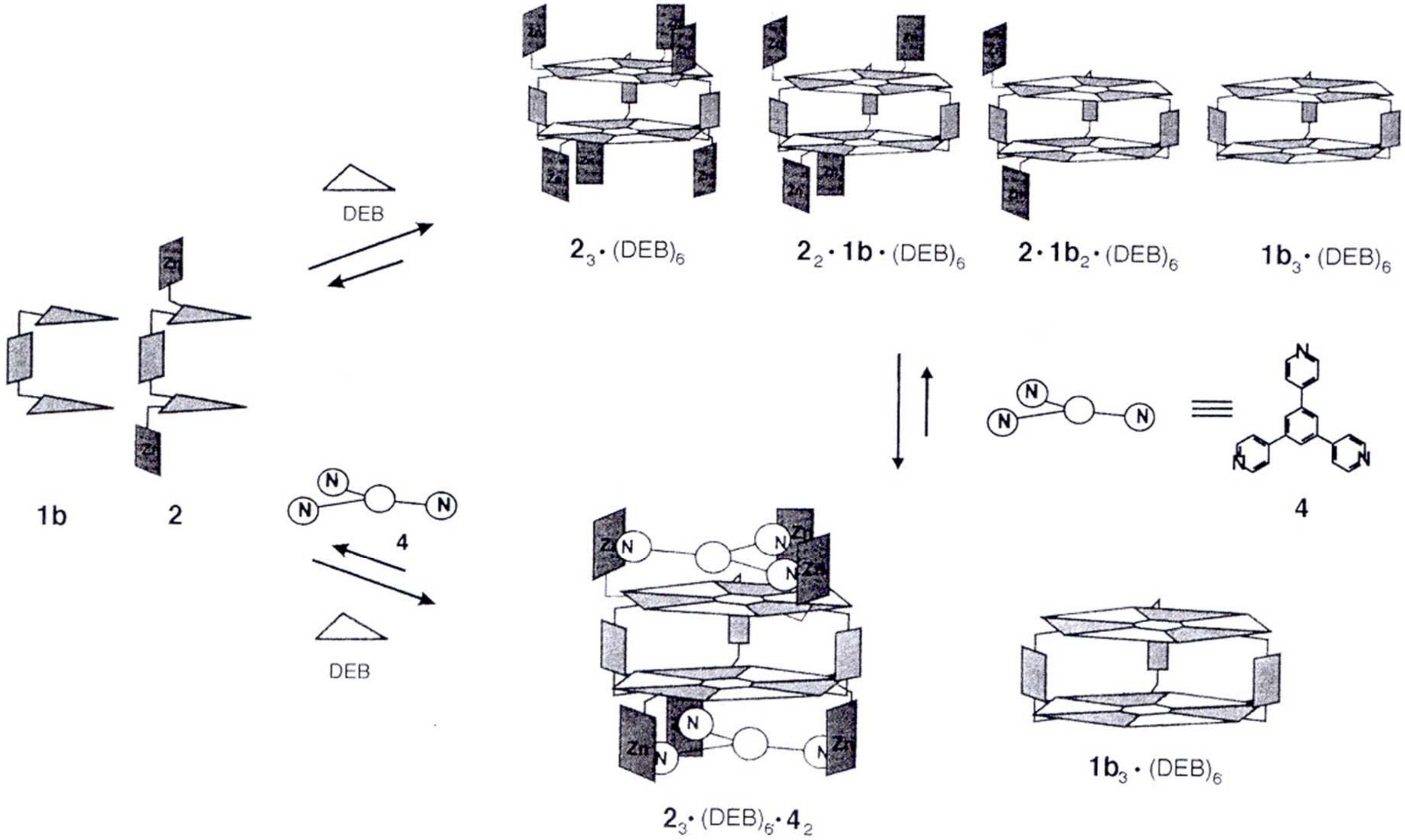
stirred





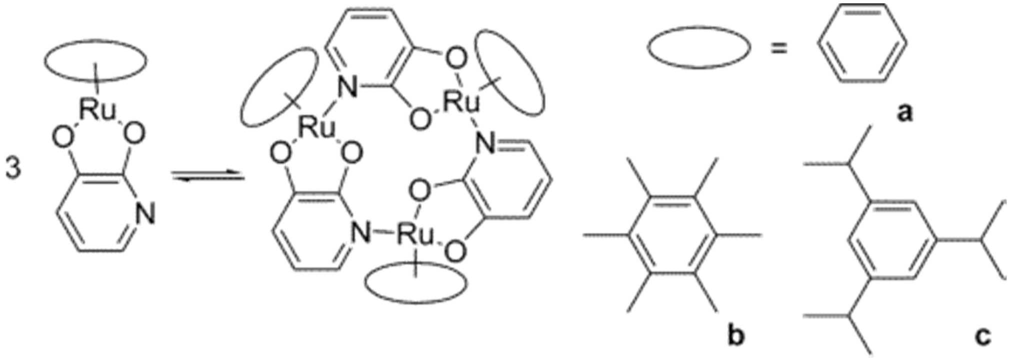


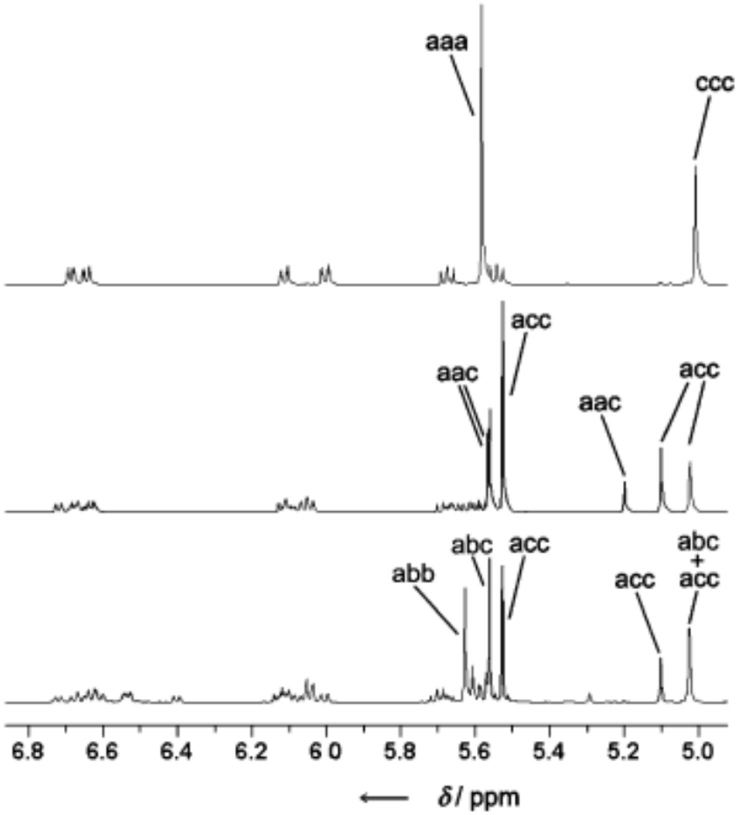
# Amplification of the inert species

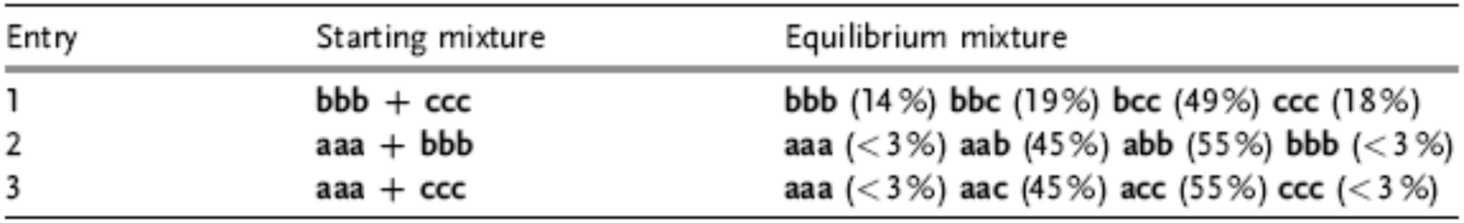


Reinhoudt, Angewandte Chemie, **2000**

# Is life (research) really that easy ?



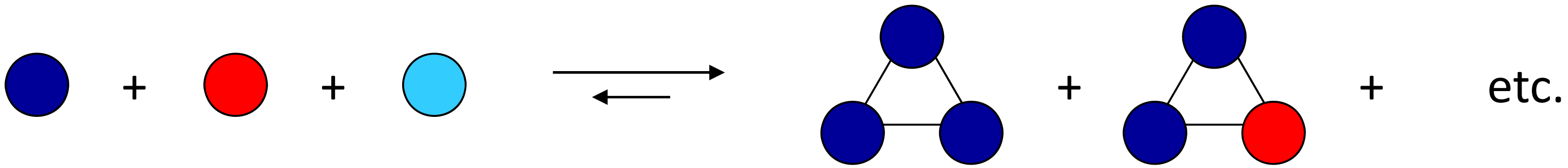


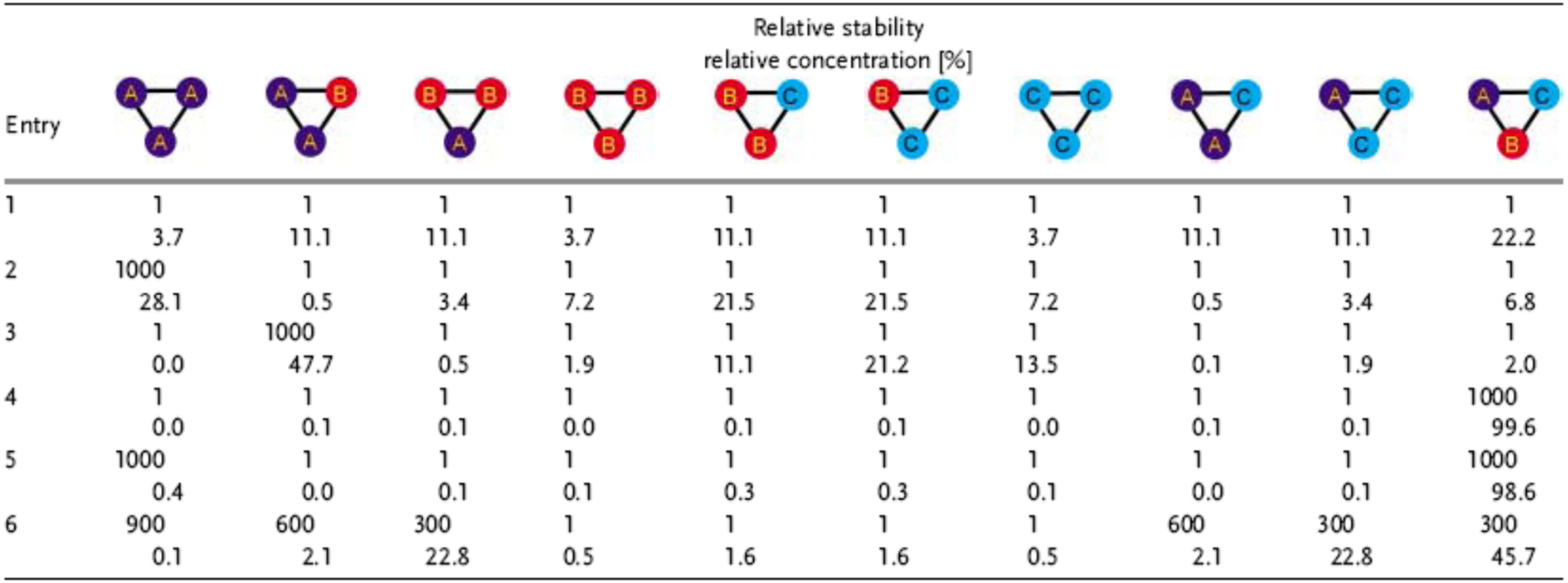


| Entry | Starting mixture | Equilibrium mixture |
|---|---|---|
| 1 | **bbb** + **ccc** | **bbb** (14%) **bbc** (19%) **bcc** (49%) **ccc** (18%) |
| 2 | **aaa** + **bbb** | **aaa** (<3%) **aab** (45%) **abb** (55%) **bbb** (<3%) |
| 3 | **aaa** + **ccc** | **aaa** (<3%) **aac** (45%) **acc** (55%) **ccc** (<3%) |

Severin, Angewandte, **2003**

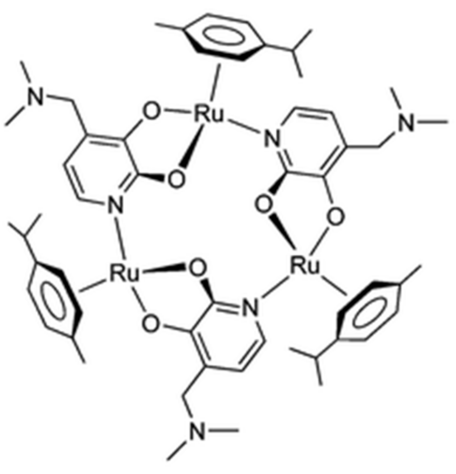
# Simulations



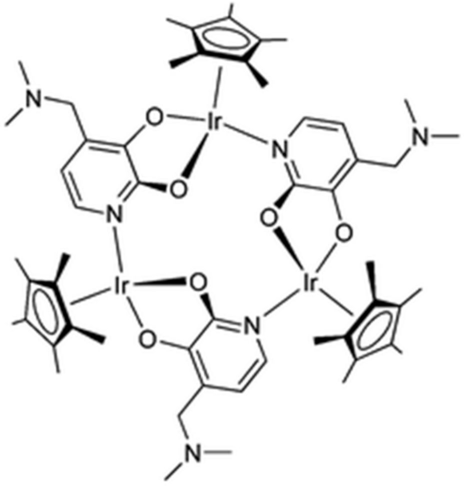


| Entry | A A A | A B A | B B A | B B B | B C B | B C C | C C C | A C A | A C C | A C B |
|---|---|---|---|---|---|---|---|---|---|---|
| | Relative stability<br>relative concentration [%] | | | | | | | | | |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| | 3.7 | 11.1 | 11.1 | 3.7 | 11.1 | 11.1 | 3.7 | 11.1 | 11.1 | 22.2 |
| 2 | 1000 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| | 28.1 | 0.5 | 3.4 | 7.2 | 21.5 | 21.5 | 7.2 | 0.5 | 3.4 | 6.8 |
| 3 | 1 | 1000 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| | 0.0 | 47.7 | 0.5 | 1.9 | 11.1 | 21.2 | 13.5 | 0.1 | 1.9 | 2.0 |
| 4 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1000 |
| | 0.0 | 0.1 | 0.1 | 0.0 | 0.1 | 0.1 | 0.0 | 0.1 | 0.1 | 99.6 |
| 5 | 1000 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1000 |
| | 0.4 | 0.0 | 0.1 | 0.1 | 0.3 | 0.3 | 0.1 | 0.0 | 0.1 | 98.6 |
| 6 | 900 | 600 | 300 | 1 | 1 | 1 | 1 | 600 | 300 | 300 |
| | 0.1 | 2.1 | 22.8 | 0.5 | 1.6 | 1.6 | 0.5 | 2.1 | 22.8 | 45.7 |

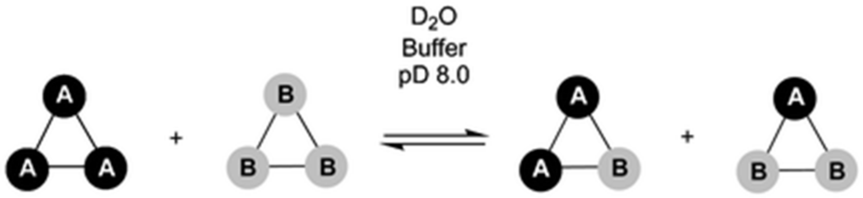
# The importance of template concentration

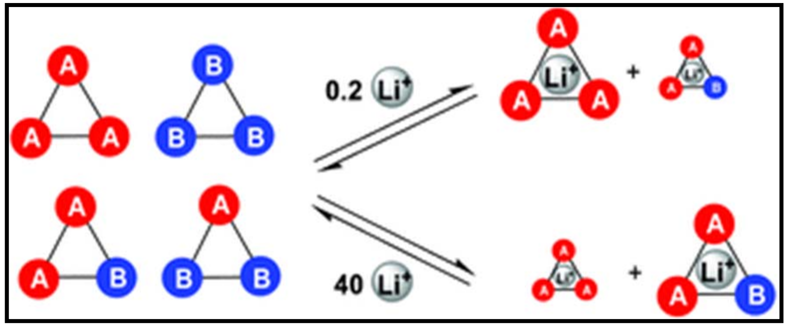


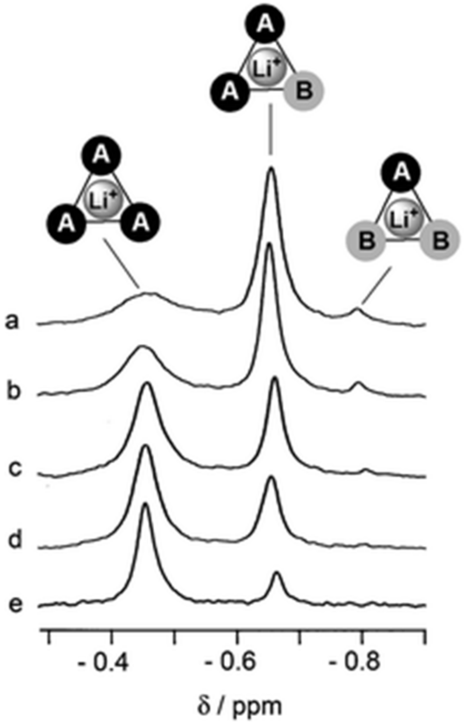
**AAA**



**BBB**

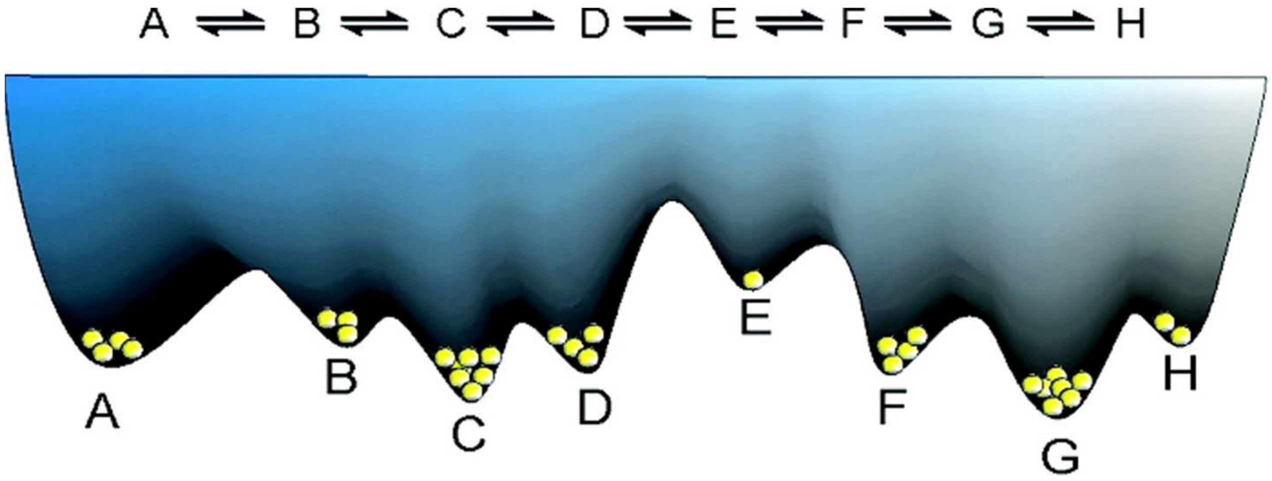


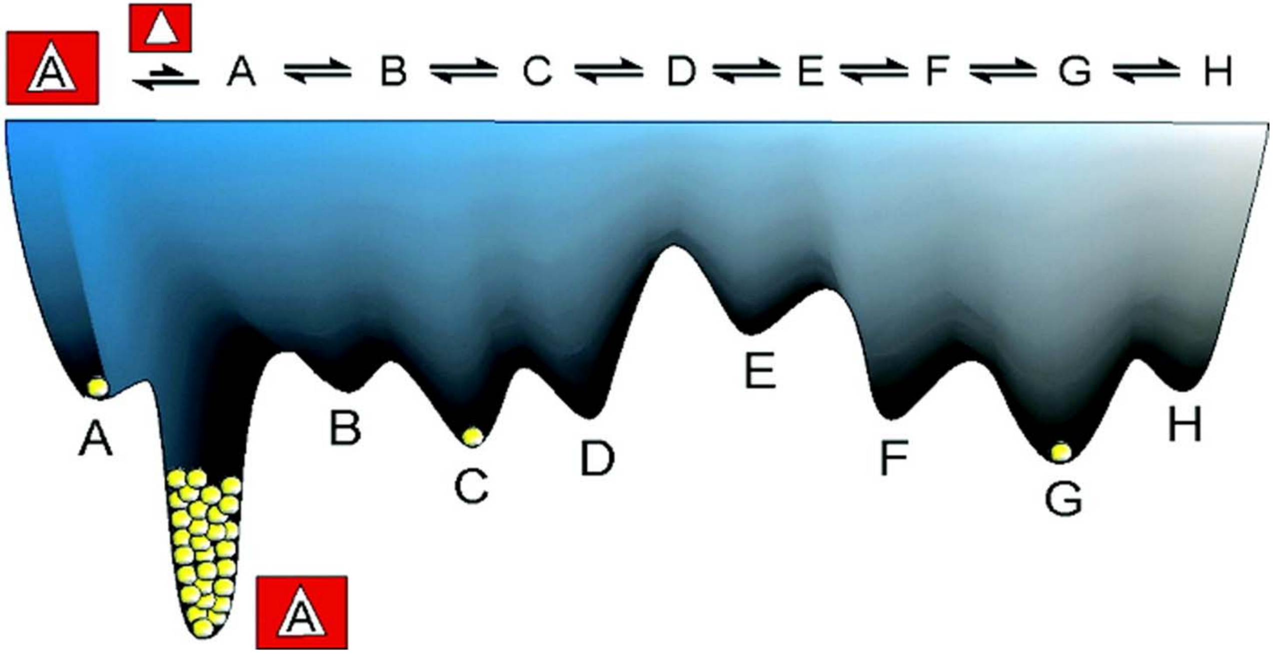




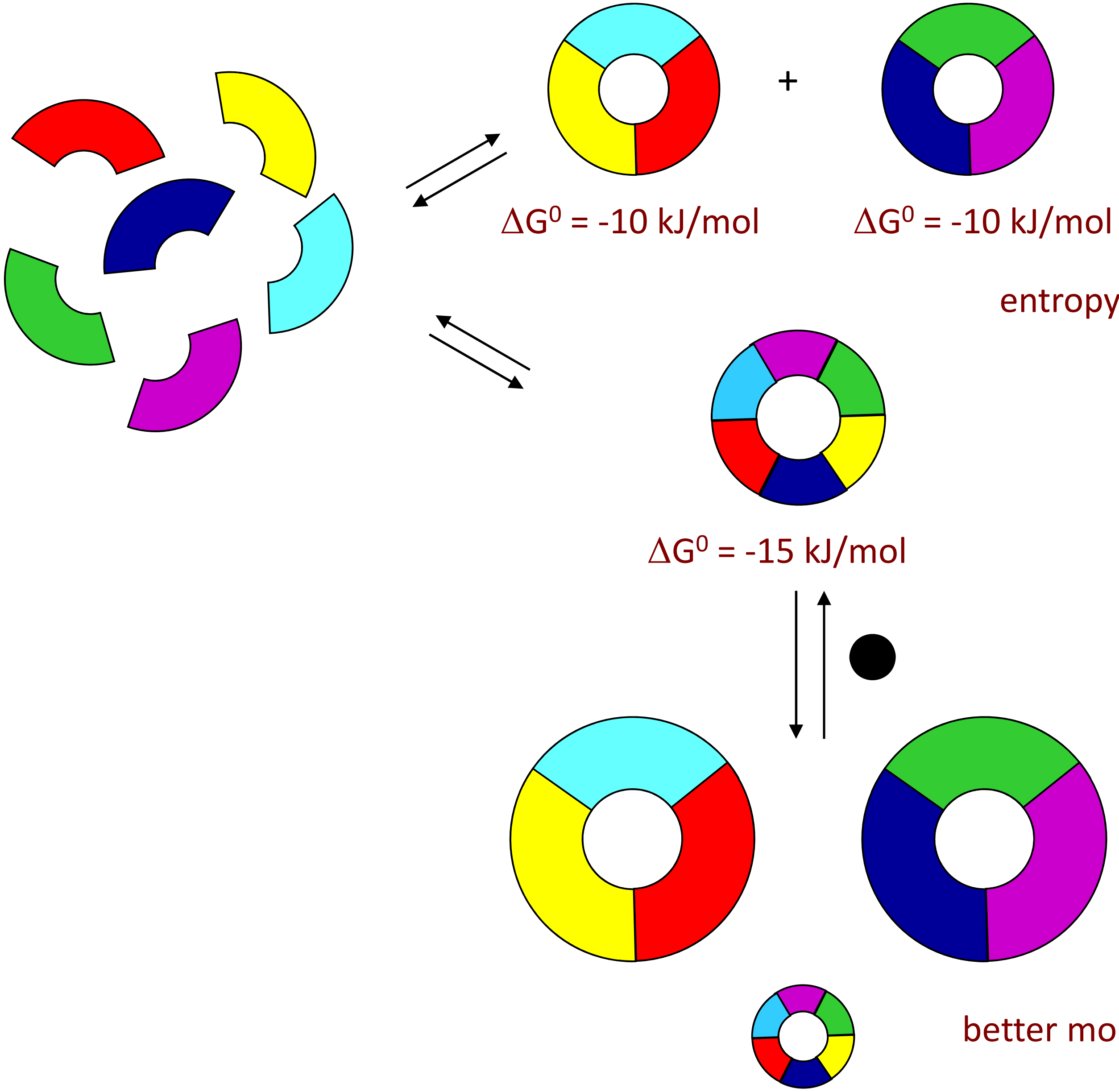
Severin, ChemComm, **2005**

# This picture is too simple





# Critical issues



$\Delta G^0$ = -10 kJ/mol

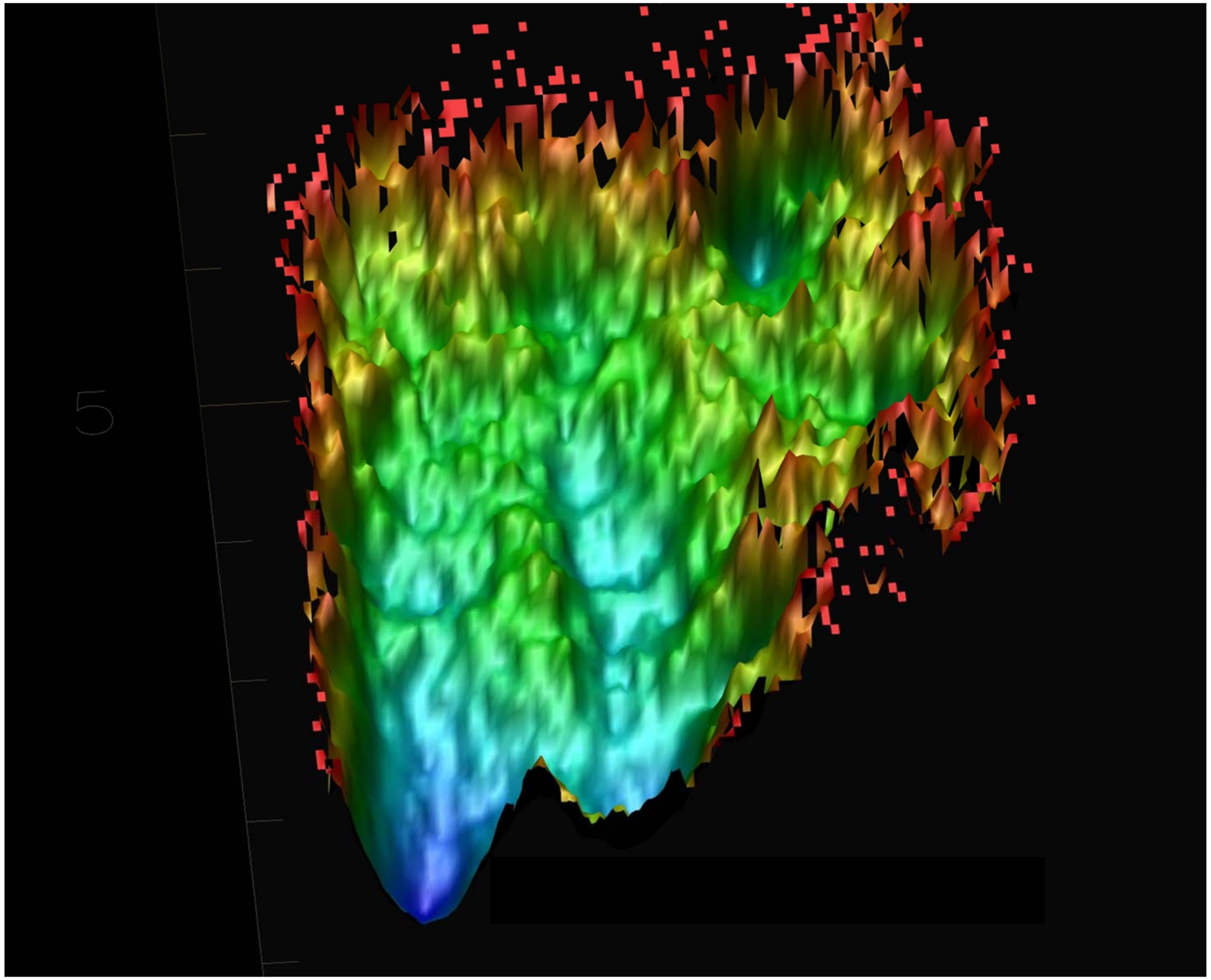
$\Delta G^0$ = -10 kJ/mol

$\Delta G^0$ = -15 kJ/mol

entropy favors small macrocycles

better more intermediate binders than
a few strong ones

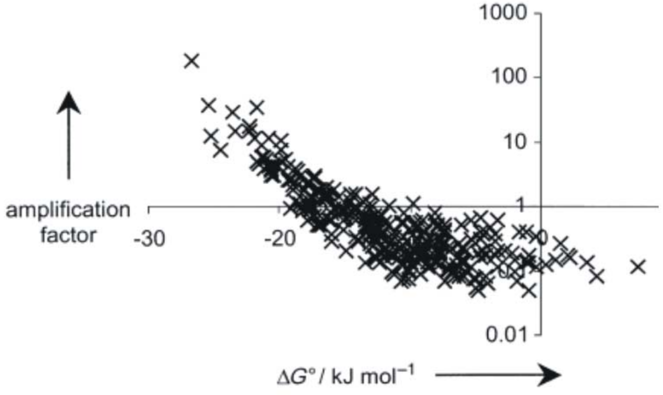
# Important

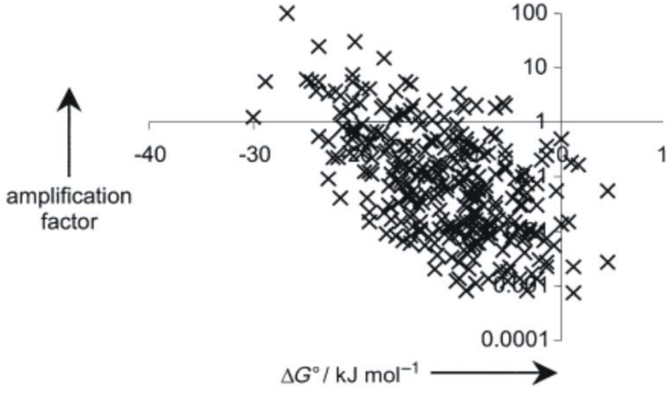


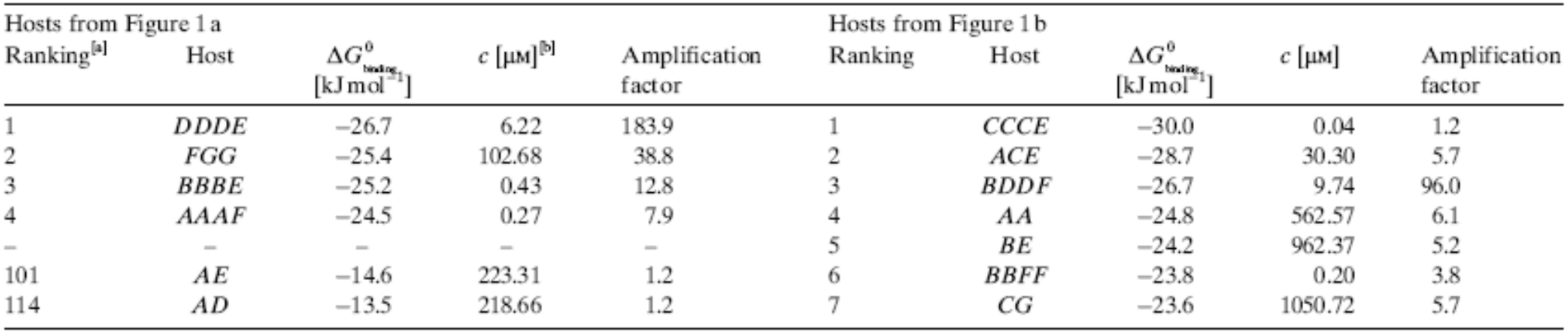
[redacted]

The overall change in free energy in the library will determine the new distribution

# Extended simulations





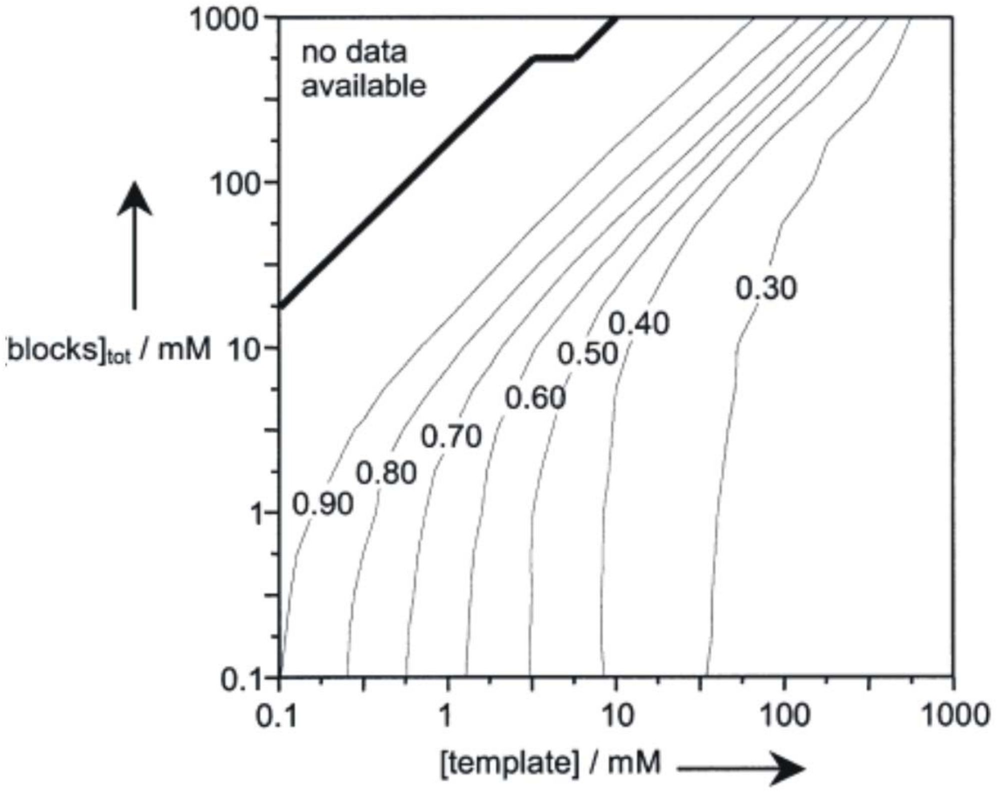


| Hosts from Figure 1 a Ranking[a] | Host | $\Delta G^0_{\text{binding}}$ [kJ mol$^{-1}$] | $c$ [µM][b] | Amplification factor | Hosts from Figure 1 b Ranking | Host | $\Delta G^0_{\text{binding}}$ [kJ mol$^{-1}$] | $c$ [µM] | Amplification factor |
|---|---|---|---|---|---|---|---|---|---|
| 1 | *DDDE* | –26.7 | 6.22 | 183.9 | 1 | *CCCE* | –30.0 | 0.04 | 1.2 |
| 2 | *FGG* | –25.4 | 102.68 | 38.8 | 2 | *ACE* | –28.7 | 30.30 | 5.7 |
| 3 | *BBBE* | –25.2 | 0.43 | 12.8 | 3 | *BDDF* | –26.7 | 9.74 | 96.0 |
| 4 | *AAAF* | –24.5 | 0.27 | 7.9 | 4 | *AA* | –24.8 | 562.57 | 6.1 |
| – | – | – | – | – | 5 | *BE* | –24.2 | 962.37 | 5.2 |
| 101 | *AE* | –14.6 | 223.31 | 1.2 | 6 | *BBFF* | –23.8 | 0.20 | 3.8 |
| 114 | *AD* | –13.5 | 218.66 | 1.2 | 7 | *CG* | –23.6 | 1050.72 | 5.7 |

28 dimers, 84 trimers, 210 tetramers; affinities randomly assigned

Sanders, Otto, ChemEurJ, **2004**

# Correlation between amplification and concentrations



# Guidelines for a successful application of DCC

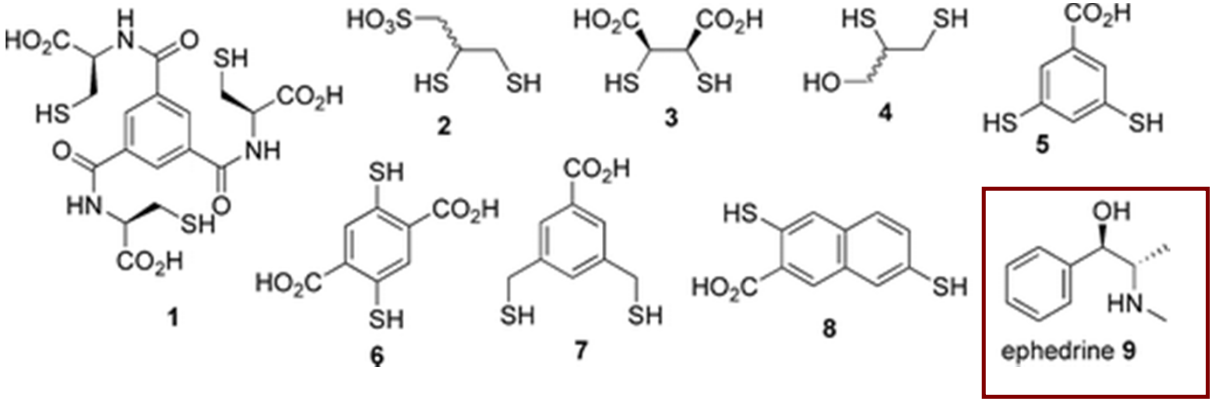
- work at low target concentration
- detect only target-bound species
- use *virtual* libraries (no receptor in the absence of target)
- use a common building block in all receptors

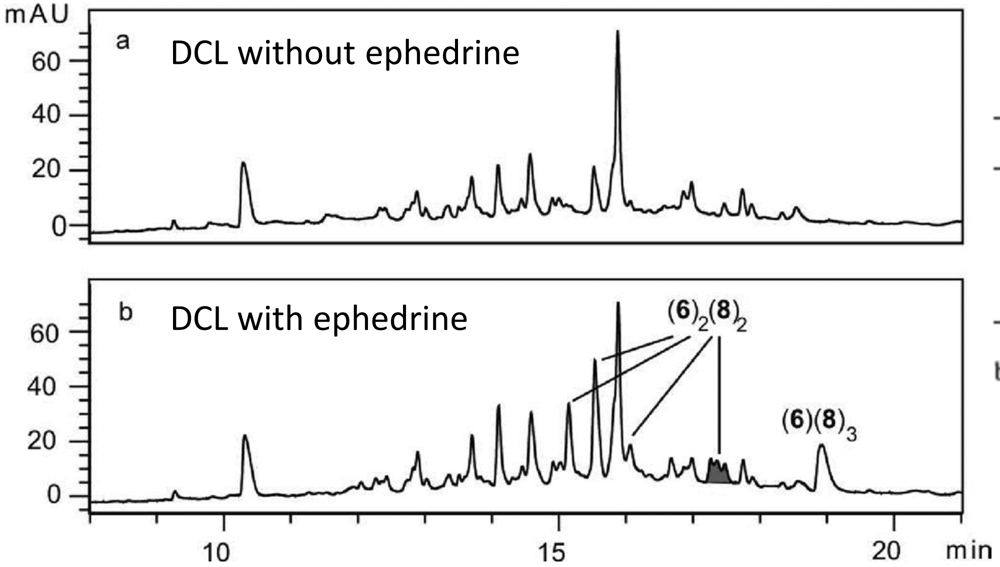
Then….

- good binders have a high probability of being amplified
- any significantly amplified compound is essentially guaranteed to be a good binder

# Screening of large libraries

## DCL composed of > 9000 members




a DCL without ephedrine

b DCL with ephedrine

| | $(6)_2(8)_2$ | $(6)(8)_3$ |
|---|---|---|
| $K$ ($M^{-1}$) | $1.3 \times 10^4$ | $1.5 \times 10^4$ |
| $\Delta G^\circ$ ($kJ \cdot mol^{-1}$) | −23.5 | −23.8 |
| $\Delta H^\circ$ ($kJ \cdot mol^{-1}$) | −24.2 | −23.6 |
| $T\Delta S^\circ$ ($kJ \cdot mol^{-1}$) | −0.7 | 0.2 |

[a] Determined using isothermal titration calorimetry in 50 mM borate buffer pH 8.0 at 298 K.

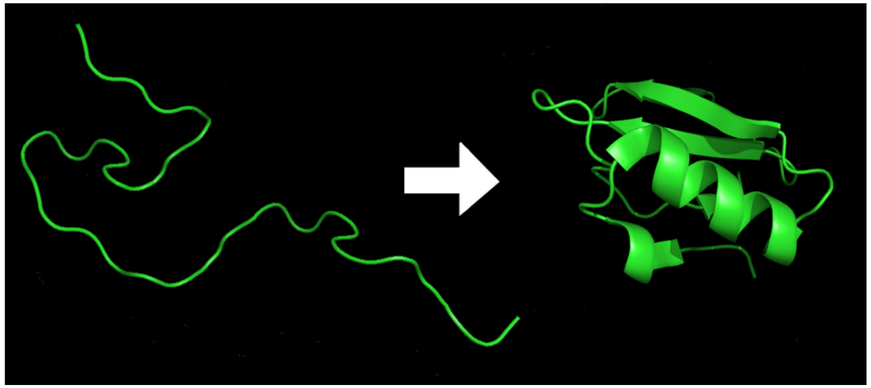
Otto, JACS, **2008**

# Current developments

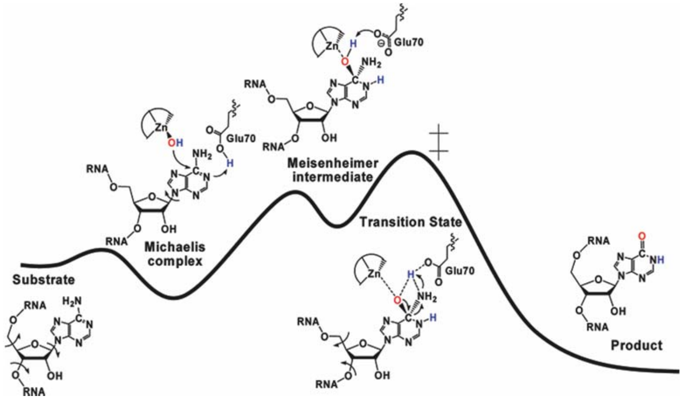
- Detecting weak rather than strong noncovalent interactions
  - Fragment based drug discovery (Tethering)
  - Protein folding
  - Catalyst Discovery
- Systems chemistry

# DCC is survival of the fittest

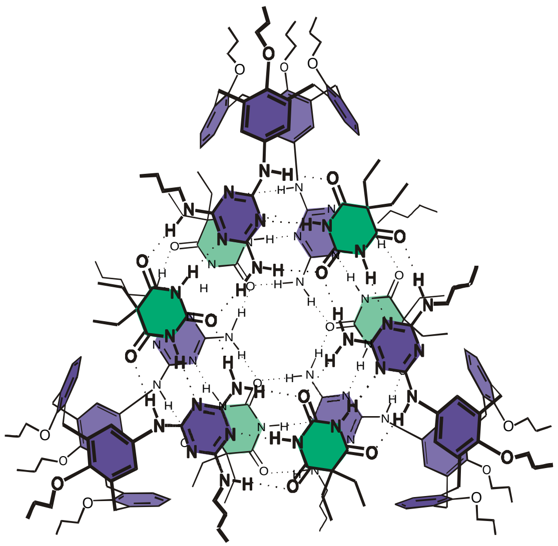
# The importance of weak interactions



protein folding

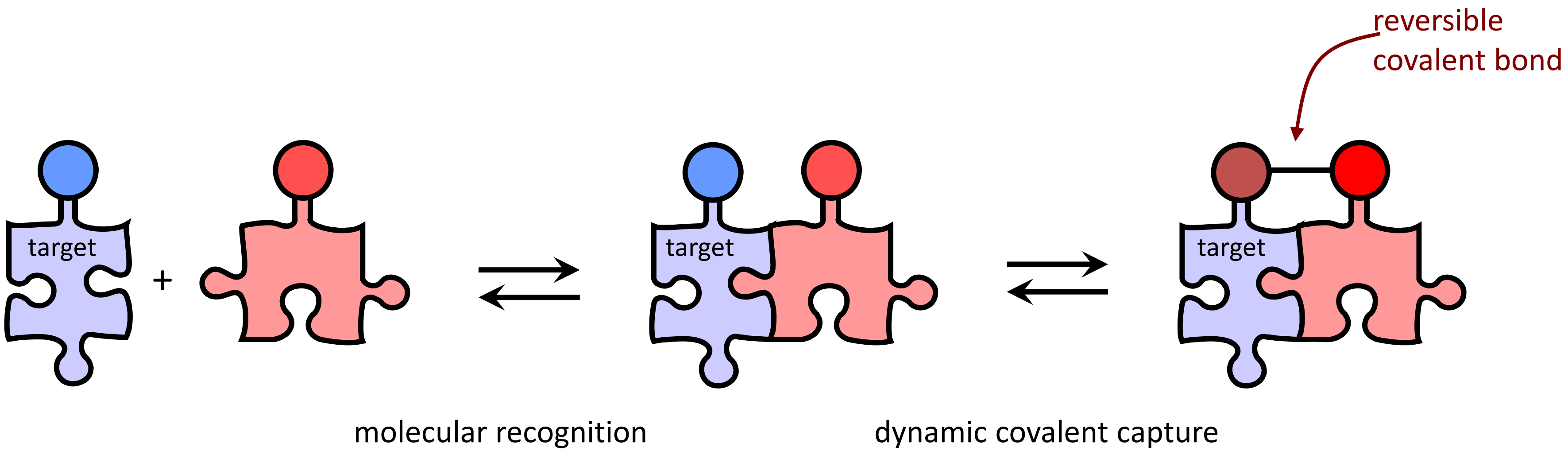


enantioselective catalysis



self-assembly

# Dynamic Covalent Capture



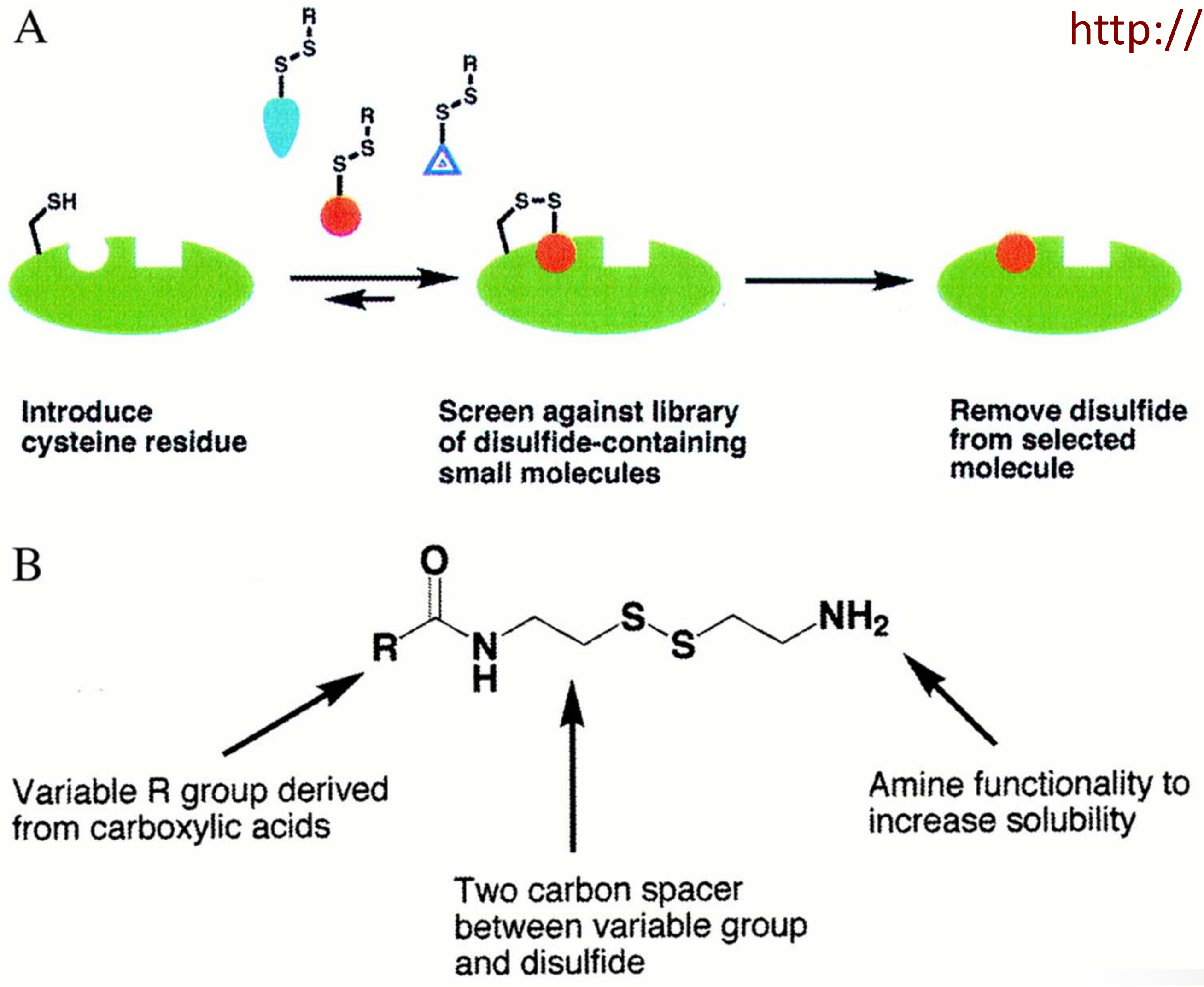
## key advantages

- intramolecular recognition
- thermodynamic control (self-selection !!)
- general applicability

Angewandte Chemie, **2009**, 2288 (review)

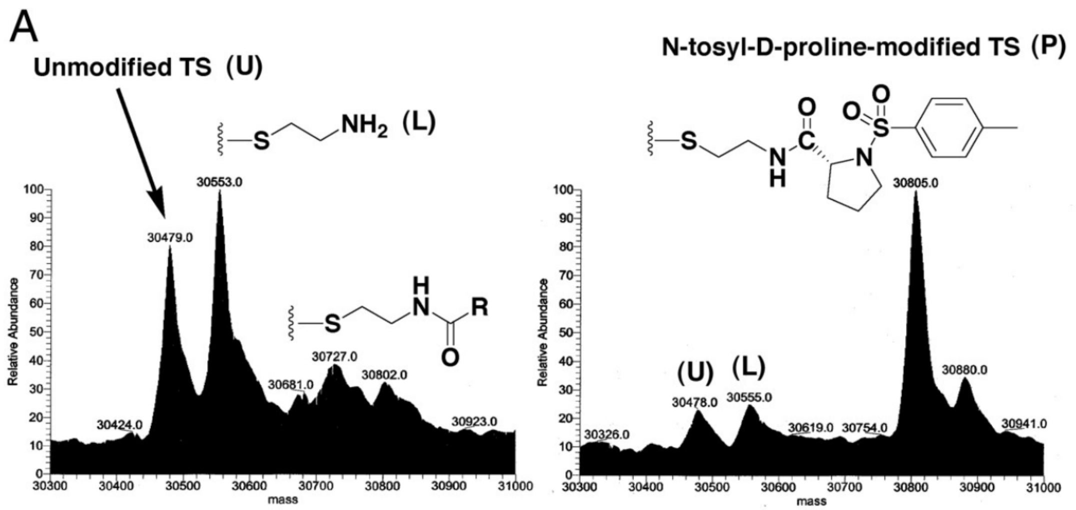
# Tethering

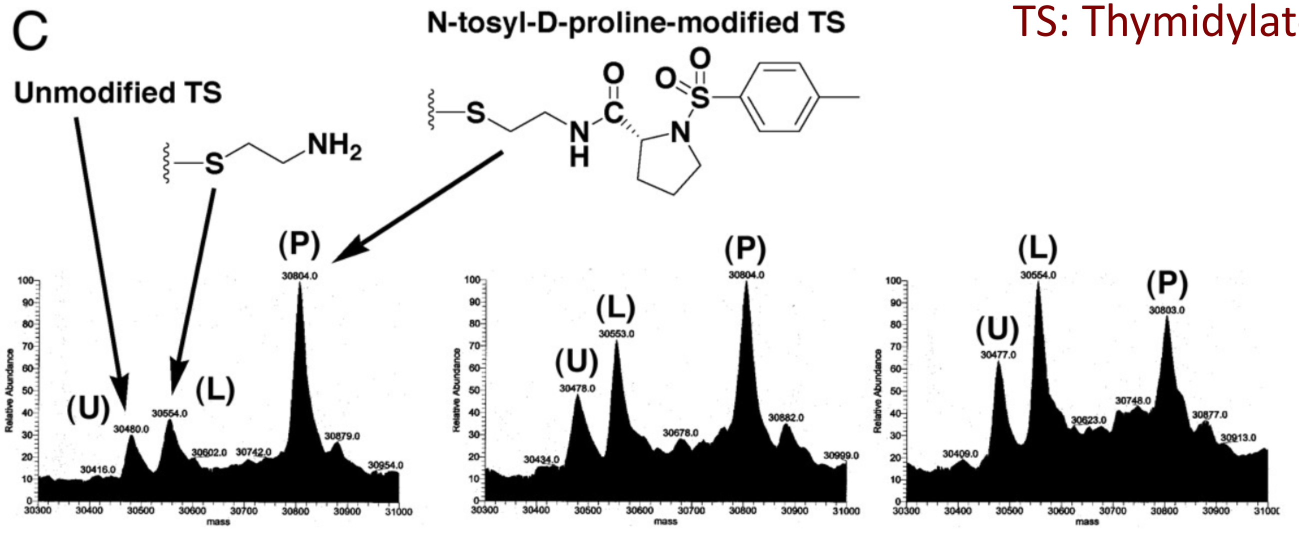
http://www.sunesis.com/



Erlanson, Wells, PNAS, **2000**

# Screening





20 Compounds,
100 μM each

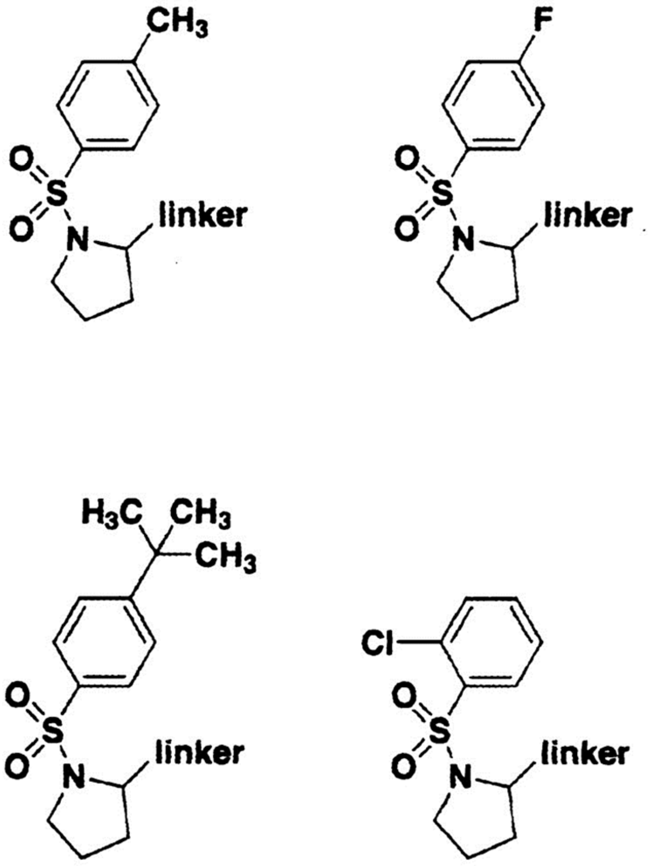
50 Compounds,
40 μM each

100 Compounds,
20 μM each

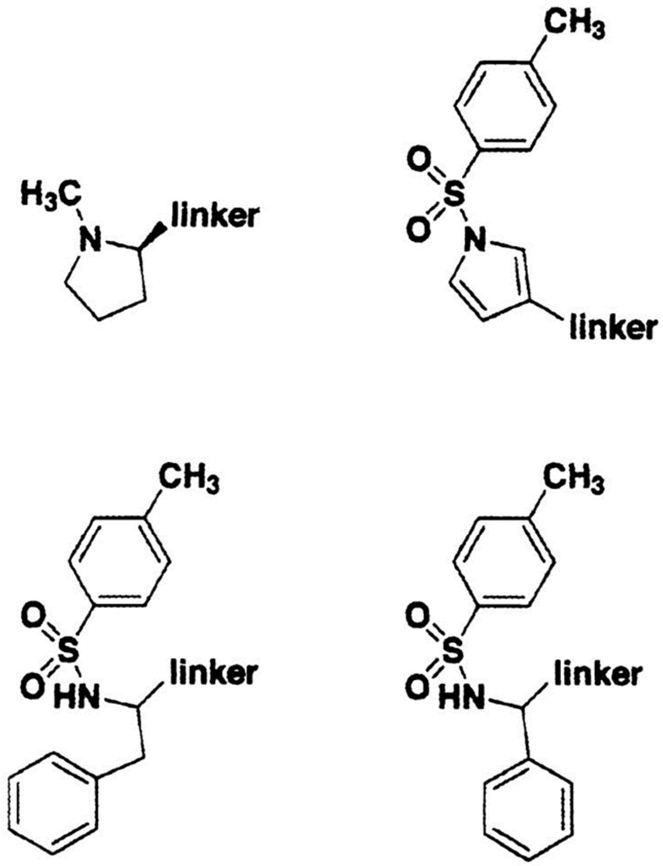
TS: Thymidylate synthase

# Results

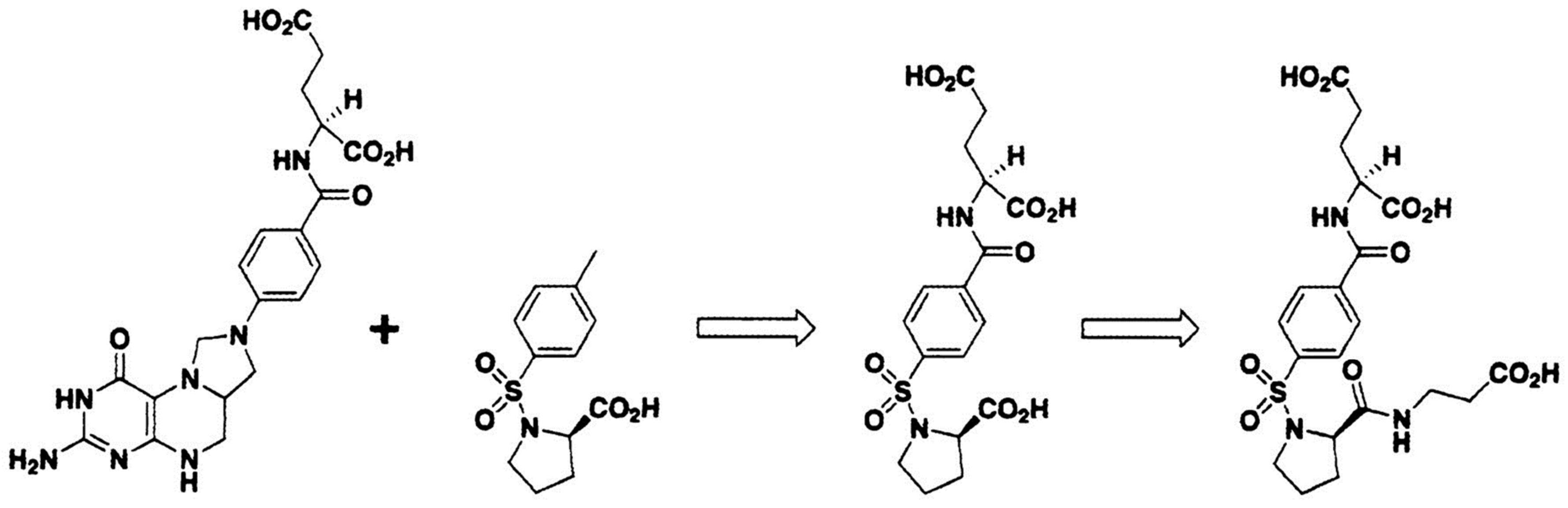
## Selected



## Not Selected



# Fragment based drug discovery



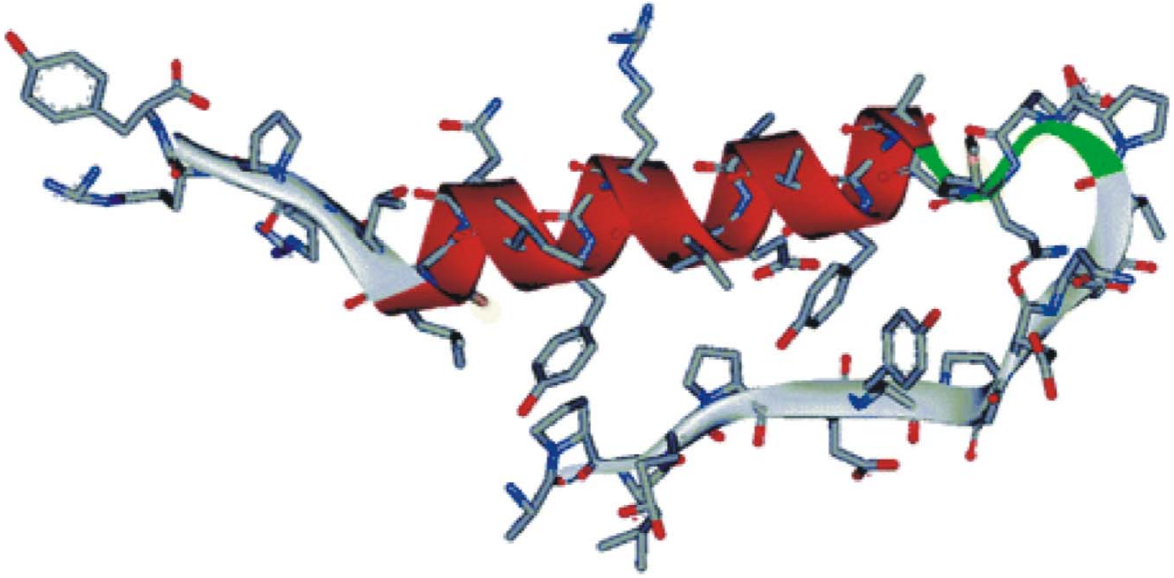
**Methylenetetrahydrofolate**
**$K_m = 14\ \mu M$**

**N-tosyl-D-proline**
**$K_I = 1.1 \pm 0.25\ mM$**

**$K_I = 24 \pm 7\ \mu M$**

**$K_I = 330 \pm 40\ nM$**

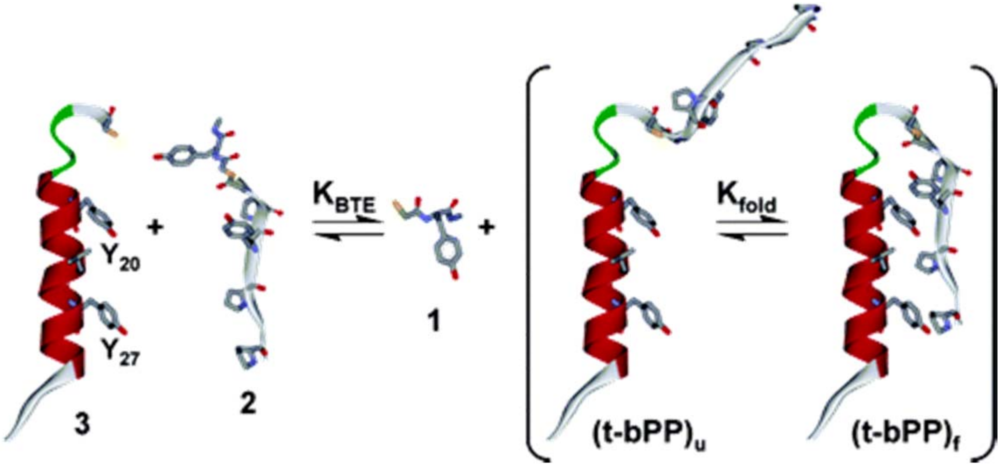
# Protein folding

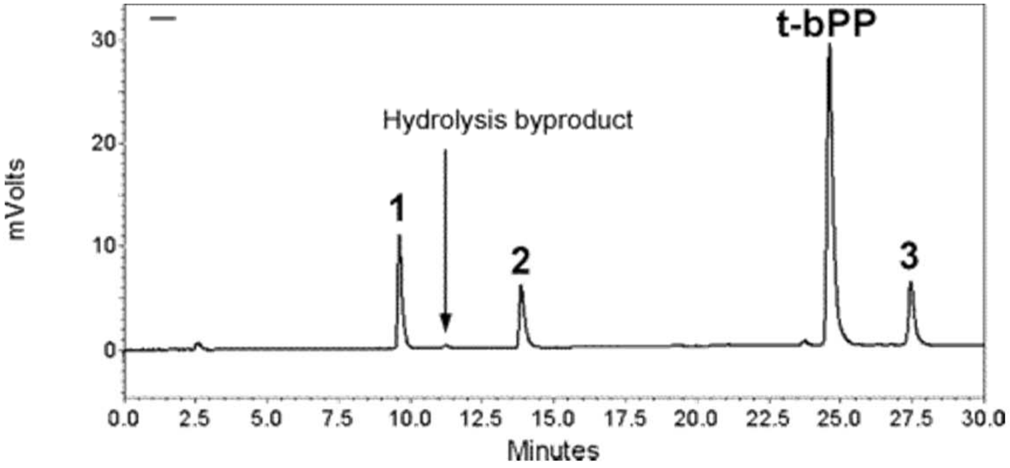


bPP: bovine pancreatic polypeptide

Gellman, JACS, **2004**

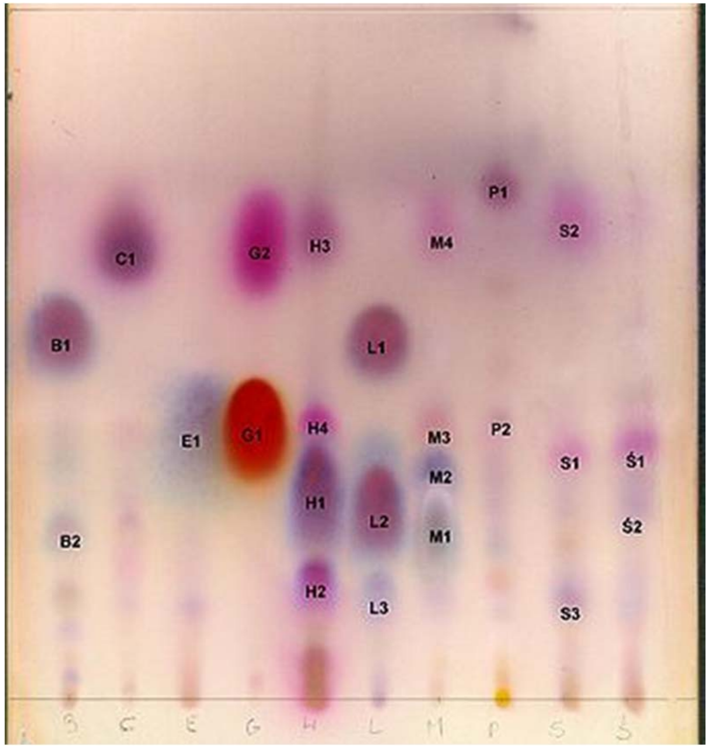
# Backbone thioester exchange





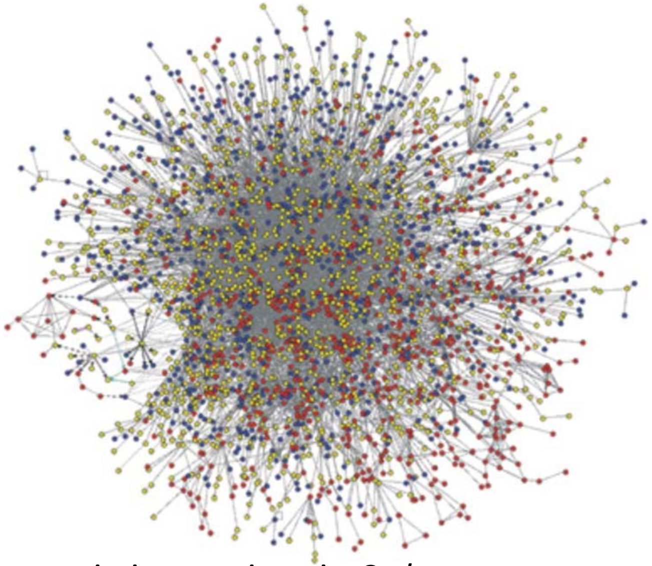
| peptide | $\Delta G_{fold/BTE}$ | $\Delta G_{fold/Gdm}$ |
|---|---|---|
| bPP | | −3.0 |
| bPP (L3R, D10G) | | −2.7 |
| t-bPP | −1.4 | −2.5 |
| t-bPP (Y20A) | +0.8 | |
| t-bPP (Y27A) | +0.1 | |
| t-bPP (Y20,27A) | +1.1 | |
| t-bPP (A22G) | −0.7 | −0.8 |

# Systems Chemistry



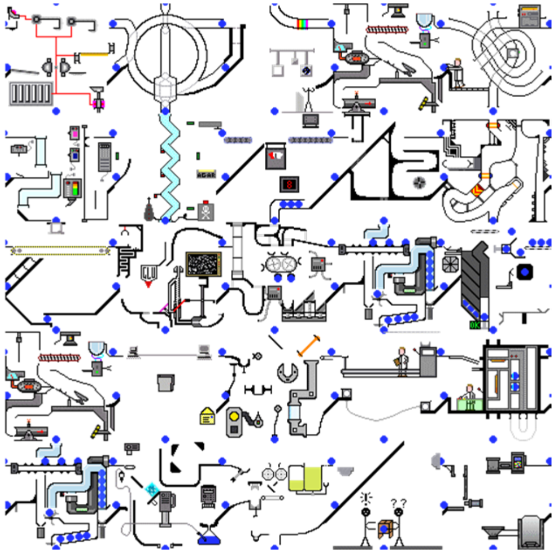
Chemists don't like mixtures

# Networks are everywhere



protein interactions in *C.elegans*

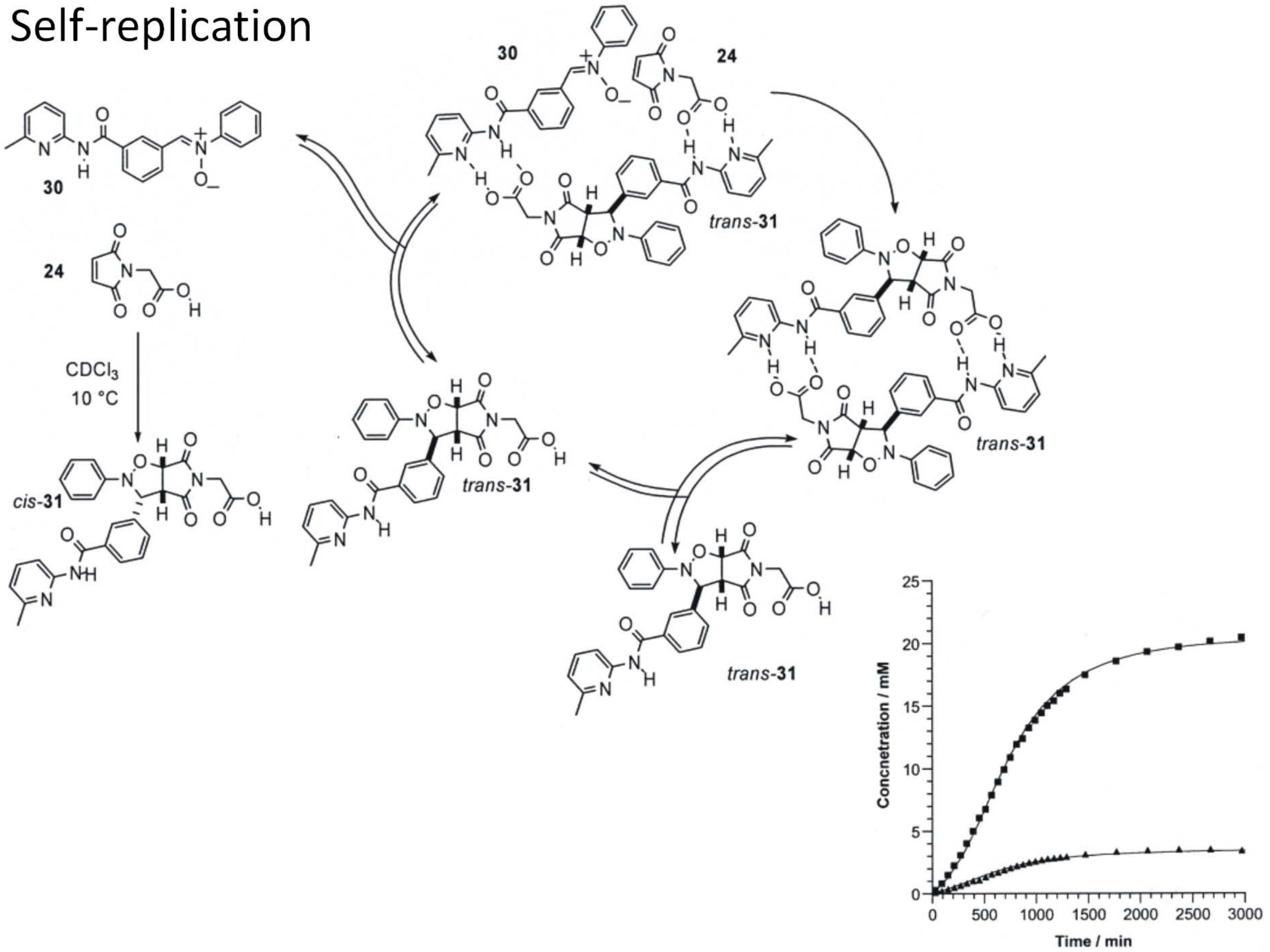
Life ?



stock exchange

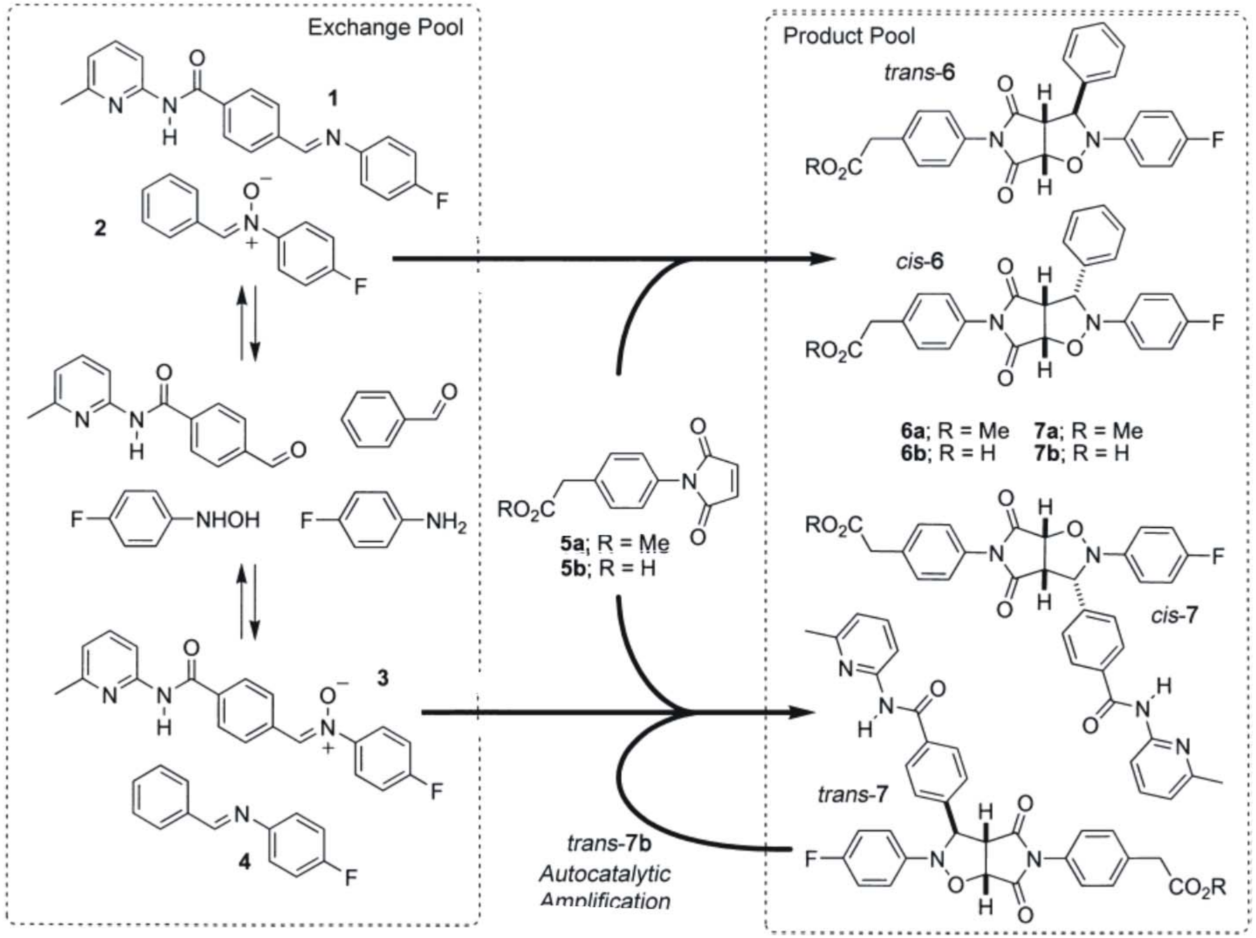
social networks

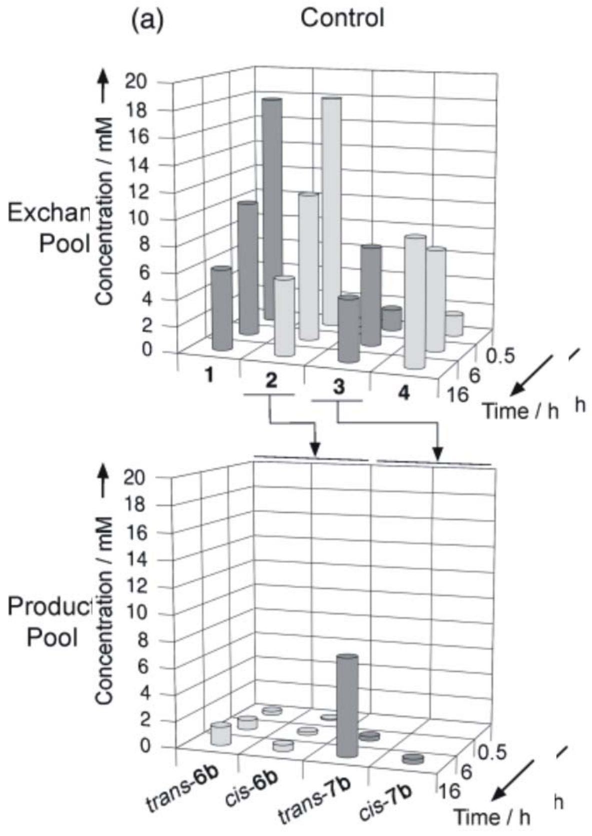
# Self-replication



Philp, Org.Lett, **2001**

# Self-amplification of a catalyst





Philp, Angewandte Chemie, **2008**

# Perspective Dynamic Combinatorial Chemistry

- Proof-of-Principle studies
- DCC (like Combinatorial Chemistry) is not a panacea

Challenges…

- Large libraries
- Analytical techniques
- Functional systems in water
- Commercial applications
- Molecular Systems