Selection approaches to catalytic systems

Paul A. Brady and Jeremy K. M. Sanders*

Cambridge Centre for Molecular Recognition, University Chemical Laboratory, Lensfield Road, Cambridge, UK CB2 1EW E-mail: jkms@cam.ac.uk

CHEMICAL SOCIETY REVIEWS

The key feature of enzymic catalysis is recognition of the transition state. Synthesis of designed systems rarely leads to successful catalysts as the rules for conformation and intermolecular interactions are too imperfectly understood. This review describes several current 'selection' approaches to the generation of systems that can recognise transition state analogues. Examples covered include catalytic antibodies, ribozymes, imprinted polymers. combinatorial chemistry, and thermodynamic templating. All have the potential to yield effective catalysts without prior design of every detail.

1 Introduction

How can chemists produce new catalysts? In particular, can we mimic the astonishing efficiency and selectivity of enzymic catalysis and apply the same approach to important synthetic reactions for which there are no naturally occurring enzymes? In this review, we start from the principle that the key feature of enzymic catalysis is recognition of the transition state, and then describe several current approaches to the application of that principle. Our examples are taken from molecular biology, organic and inorganic synthesis, and polymer and solid-state chemistry. Many of the literature references are to reviews, rather than to original research papers.

At one level, we already 'understand' enzyme action; as early as 1894, Emil Fischer proposed that catalysis arose as a result of a binding process.[1] He termed this the 'Lock and Key' principle: the key (substrate) fits into the lock (enzyme) and is then modified in some way. More subtly, Pauling suggested in 1948 that an enzyme catalyses a reaction by selectively binding and stabilising the transition state more than the starting materials or products.[2] This idea, which is essentially the view held today, is summarised in the energy profiles shown in Fig. 1. The enzyme **E** binds the substrate **S** to form a complex **ES** with binding free energy ΔG_{ES}. The two transition states for the reaction are $\mathbf{TS}_{cat}$ for the catalysed reaction and $\mathbf{TS}_{uncat}$ for the reaction in the absence of catalyst. The corresponding activation energies for the reaction are ΔG_{cat} and ΔG_{uncat}. It is clear from this picture that ΔG_{ETS}, the free energy for binding the transition state to the enzyme, is larger than ΔG_{ES}. This is the origin of the rate enhancement.

Despite understanding enzyme catalysis at this level of principle, we fall very short of the more severe, practical test—

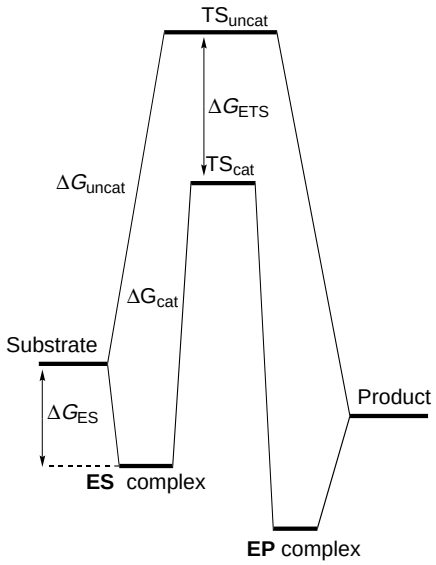


Fig. 1 Free energy profiles for a catalysed and an uncatalysed reaction

Paul Brady was born in Manchester, UK in 1972 but spent much of his youth living in Switzerland. He obtained his BA (Chemistry) from the University of Cambridge in 1993, carrying out some research with Ian Paterson and working at Ciba-Geigy and FMC Corporation. He remained in Cambridge to work with Jeremy Sanders, making oligosteroid receptors as the first step towards the thermodynamic templating approach that is described in this article. He was awarded his PhD for this work in 1996 and will embark on a career in Patent Law in autumn 1997.

Paul A. Brady

Jeremy Sanders was born in London in 1948, and obtained his BSc from Imperial College, London, in 1969. He moved to Cambridge to work for his PhD with Dudley Williams on lanthanide shift reagents. After a postdoctoral year (1972–1973) in the USA, working on protein NMR, he joined the staff of the University of Cambridge; he was appointed to his present Chair in 1996. He has been awarded the Meldola Medal, Hickinbottom Award, two Pfizer Awards, the Loschmidt Prize and the Pedler Lectureship, and in 1995 was elected a Fellow of the Royal Society. His research interests are mainly centred on various aspects of molecular recognition, but also have included chemical and biological applications of NMR spectroscopy. In 1996, he joined the Editorial Board of Chemical Society Reviews; *his previous reviews in this journal have covered lanthanide shift reagents, chlorophyll coordination chemistry and spectroscopy, biodegradable plastics and spiders' webs.*

Jeremy K. M. Sanders

that of designing and making our own artificial catalytic systems to rival natural enzymes. Our ability to predict complex molecular shapes has improved with the development of computer modelling, but it is not yet at a level that enables us to predict accurately the three-dimensional structure of a protein or even of relatively small supramolecular complexes. This is because such structures are the finely balanced outcome of many weak non-covalent binding interactions that are themselves inadequately understood. Transition states have only a transient existence, which means that good predictions about their molecular recognition properties are even more elusive. As a result, most attempts to synthesise working catalysts by rational design of a particular structure have failed. It is perhaps inevitable that we will make the wrong molecules in an effort to uncover the design rules for making the right ones.

Many scientists have therefore embarked on a different line of attack. This focuses directly on finding systems that recognise transition states or their stable analogues, and is inspired by nature's evolutionary methods: instead of a single structure being designed, a large array of different molecules is created simultaneously. From this pool of different binders or catalysts, the best one is selected. Until the design rules are known, this selection approach should have an increased chance of success since vastly more molecules are generated than in a design approach. While this idea is conceptually straightforward, putting the principles into practice has proved difficult. In the following sections we describe the many selection approaches that are being explored. Despite the great diversity in the systems and their methods of generation, all face the same challenge, which is to achieve efficient transition state stabilisation.

2 Catalytic antibodies[3–5]

As a large part of the catalytic activity of enzymes results from stronger binding of the transition state than of the starting material(s), any molecule that can bind the transition state species selectively should be a catalyst for that reaction. With this in mind, Jencks suggested in 1969 that antibodies generated in the mammalian immune response should function as enzymes.[6] Antibodies are proteins that are produced in the body in response to an alien species, called an antigen, and they bind to such a molecule or particle strongly and selectively.

2.1 Antibody structure and generation

Antibodies are proteins consisting of a number of domains (Fig. 2). The variable domains consist of peptide chains of diverse amino acid compositions, giving rise to a binding site area of *ca.* 20 × 20 Å. This site may bind small antigens by encapsulation or may present a more open recognition cleft for large antigens. The different compositions of the variable regions are the source of antibody diversity. Blood serum contains around 10^8 different antibodies, which makes the likelihood of being able to bind a given antigen quite high.

Small molecules are often poor initiators of the immune response, or are simply broken down by metabolism, so antibodies to a desired antigen (termed hapten) are usually raised by covalently conjugating the hapten to a large protein molecule. A sample of this antigen conjugate is then injected into the rat or mouse, usually repeatedly. After around a week, antibodies to the hapten can be detected. Repeated exposure to the antigen gives an enhanced response as improved, second generation antibodies are produced. The spleen is then removed from the mouse and the antibodies are extracted. A given immunisation leads to recognition of many different parts of the hapten and so leads to a diverse, 'polyclonal', set of antibodies. In order to study antibody–substrate interactions, a sample of pure single (monoclonal) antibody is required. This is now possible using biotechnological manipulation and relatively large amounts of a single monoclonal antibody can be prepared, albeit at considerable cost. Antibodies have good bioavailabilty

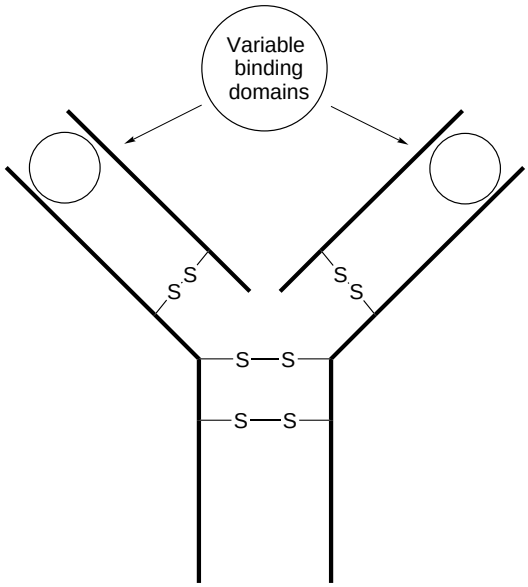
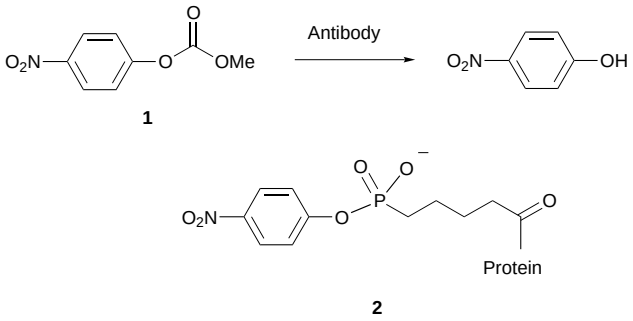


Fig. 2 Schematic view of an antibody

which makes them potentially suitable for use as drugs; they are however readily degraded within the body, which could be a problem.

The first catalytic antibodies were reported simultaneously in 1986 by Schultz and by Lerner. In each case, a tetrahedral phosphonate ester group was used as a transition state analogue (TSA) to mimic the tetrahedral shape and charge distribution of the oxy-anion intermediate in ester hydrolysis. So, antibodies for the hydrolysis of substrate **1** were raised using TSA **2**. These

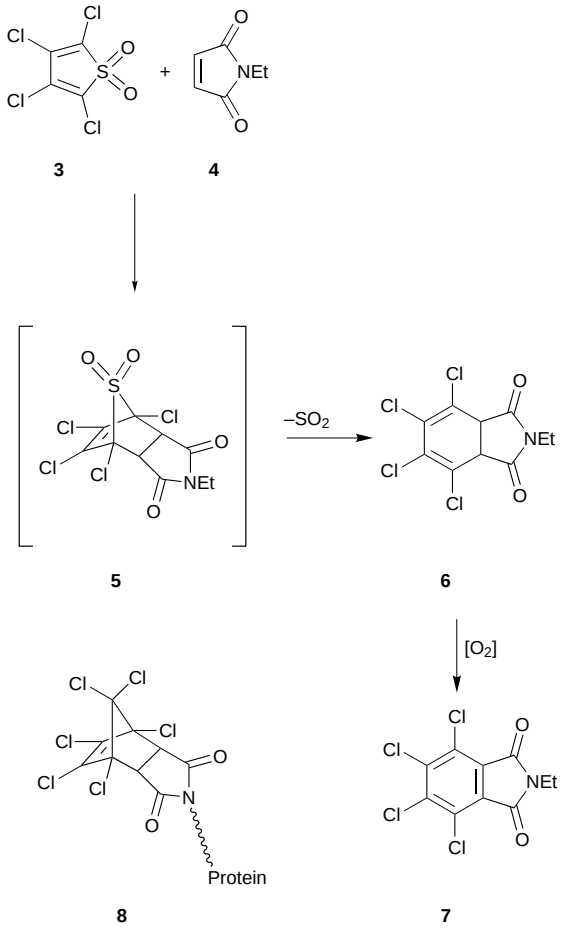


1

2

antibodies bind anionic tetrahedral groups well and thus stabilise the transition state for the hydrolysis reaction; they showed enzyme-like kinetic behaviour, were inhibited by the hapten and accelerated the reaction by factors of up to 10^4. As with enzymes, good substrate selectivity was observed: antibodies raised to **2** show discrimination for **1** over the *o*-nitrophenyl substrate isomer.

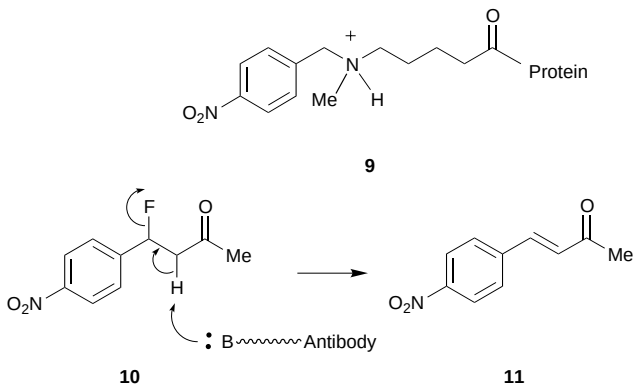
Catalysis with turnover is relatively easy to achieve in hydrolysis reactions as the small product molecules bind to the enzyme less strongly than the larger starting material. They are therefore released into solution, allowing the enzyme to bind further substrate molecules. Bond formation reactions are more difficult to catalyse as the products tend to be more strongly bound than the starting materials, resulting in inhibition of the catalyst. A good example of a case where this problem has been overcome is Hilvert's 'Diels–Alderase' antibody which accelerates the reaction between the thiophene dioxide **3** and maleimide **4**. The cycloaddition is followed by spontaneous elimination of sulfur dioxide from the initial product **5** to give **6**, and atmospheric oxidation to the aromatic species **7**. This change in geometry allows product release. The transition state in both of these steps resembles the high energy intermediate **5**,



and so the norbornene derivative **8** was a suitable hapten. Antibodies raised to this hapten show multiple turnovers (> 50) and a reasonable rate enhancement (an effective molarity of 100 M).

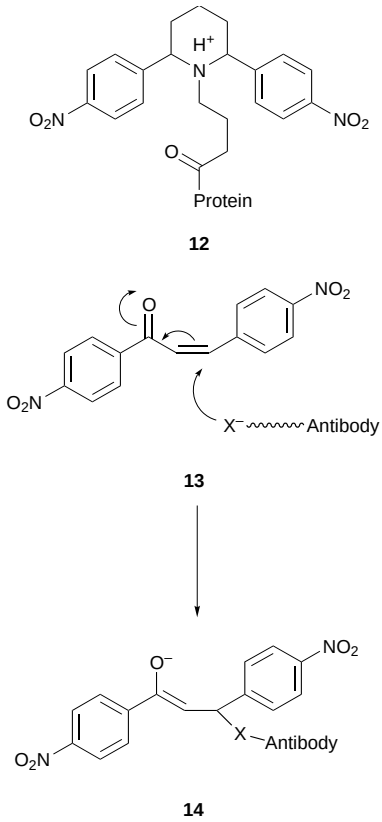
The rate enhancements of 10^3–10^4 found in early systems are quite modest when compared with natural enzymes. This results largely from the fact that most of the catalytic antibody (cAB) enhancement is due to transition state stabilisation by geometrical complementarity. However, shape complementarity is only part of the key to enzyme catalysis. It is also useful to have appropriately positioned catalytic groups present at the active site, and a more sophisticated approach is required to introduce these.

The first example of a designed hapten placing a desired catalytic group in an antibody was **9**, which was used to



generate cABs for the catalysis of HF β-elimination from **10** to give **11**. It contains charged functionality to introduce oppositely charged groups in the antibody. The tertiary ammonium group in the hapten ensures placement of a basic group in the antibody, which is able to abstract the proton when substrate **10** is bound. The strategy is effective and results in a 8.8×10^4-fold rate increase for the elimination.

A charged functional group present at the active site can do more than simply stabilise a charge electrostatically. Hapten **12**,

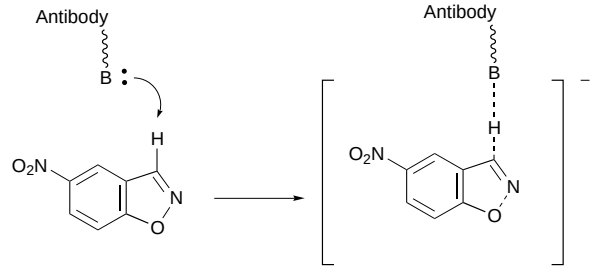


which is protonated at physiological pH, was used for the generation of antibodies for the *cis–trans* isomerisation of the α,β-unsaturated ketone **13**; it is believed that an anionic nucleophile in the active site enhances the catalysis as shown (k_{cat}/k_{uncat} = 15 000). A similar ammonium cation was used to induce a carboxyl group in the active site of an antibody that catalyses an aldol addition;[7] the carboxylate is thought to act as a general base in the reaction, contributing to the 2.0×10^5-fold rate enhancement.

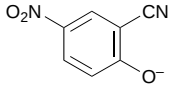
A similar general base mechanism was inferred for the antibody-catalysed decomposition of **15** to **16**.[8] The hapten **17** was used to generate antibodies containing a negatively charged group. Substantial antibody catalysis was observed for the reaction ($k_{cat}/k_{uncat} \approx 10^8$) and this was attributed mainly to the 'precise positioning' in the antibody of a carboxylate group that is able to act as a general base. Catalysis was inhibited by added hapten and covalent modification confirmed that the active site did indeed contain the carboxyl group important for catalysis.

However, the extraordinary efficiency of this antibody does not prove that it works in the way intended. Hilvert's group had shown that the reaction is catalysed by acetate ion in a dipolar, aprotic solvent (acetonitrile) to almost the same level as by the antibody in water. This is due to the activation of acetate ion once it is removed from a hydrogen bonding environment. As the carboxylate function in the antibody is in a cavity, and hence in an aprotic environment, it is possible that the main part of the rate acceleration observed is due to this medium effect.

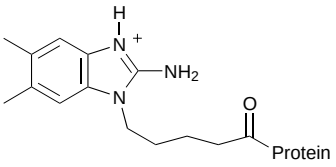
This logic led Hollfelder *et al.*[9] to investigate other proteins with hydrophobic binding sites containing general base residues. They found that commercially available serum albumins catalysed the reaction with turnover numbers and rate accelerations that are very similar to the antibodies raised against the transition state analogue. This result suggests that the precise positioning of the base in the antibody is much less important



15



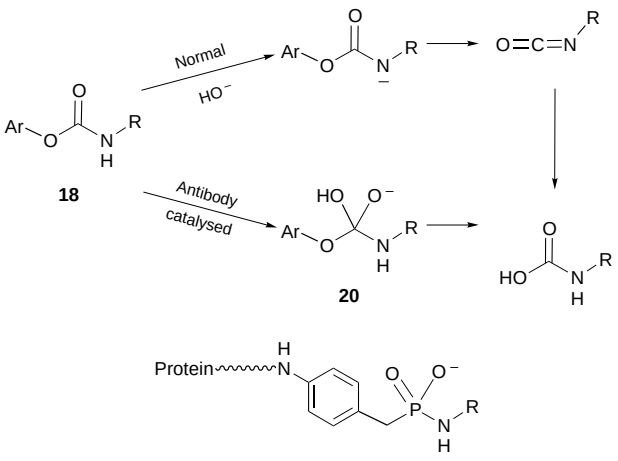
16



17

than medium effects within the cavity. It is conceivable that similar effects have been overlooked in the past during attempts to rationalise results in an over-optimistic fashion.

Nevertheless, there are now several examples of antibodies that catalyse inaccessible or otherwise unfavourable processes; a number of these involve redirecting the course of cationic cyclisation reactions.[5] A recent example involving a change of mechanism, rather than product is the base-catalysed hydrolysis of carbamate **18** which normally occurs by the top route



illustrated (initial deprotonation of the NH group followed by expulsion of ArO^-) but use of the TSA **19** led to antibodies that catalyse a pathway involving initial nucleophilic attack by hydroxide to give the intermediate **20**.[10]

Note that, even when racemic haptens are used, the resulting antibodies are generally highly enantioselective because one individual molecule of an arbitrary chirality has always initiated the response leading to a particular monoclonal antibody.

The success rate for generating and selecting good hapten-binding antibodies is high, but catalytic activity is quite rare and even the best antibodies do not rival enzymic efficiency. Expansion of the repertoire of available functional groups is necessary, and there is clearly a need for methods of isolating an antibody directly based on catalytic ability rather than binding strength; such approaches are beginning to appear.[11] All the selection approaches we describe here share other common problems, and these are discussed at the end of the article.

3 Ribozymes[12]

Effective enzymic catalysis has evolved over millions of years through mutation and selection. In order to use evolution in the laboratory to create a better catalyst or binder from a good one, a class of molecules is required that can be replicated under 'sloppy' conditions—that is, which can be amplified to give multiple copies of itself with some mutations. Every so often a mutation will lead to an increase in activity for a particular trait and so evolution occurs if the improved activity can be selected for.

There is a widely held view that the present biological era in which information flows from DNA to RNA to protein was preceded by a simpler system with fewer components: a chain of RNA (ribonucleic acid) is simultaneously capable of storing information in its sequence and folding into a three-dimensional structure that carries out some vital catalytic roles in the cell, so it seems feasible that there may have been an 'RNA world' from which ours evolved. Ribozymes are RNA molecules which have some natural catalytic activity: for example the *Tetrahymena* ribozyme can cleave certain RNA sequences with a specificity similar to that of RNA processing enzymes.

The approach to carrying out an evolution process in the laboratory is generally as shown in Fig. 3. A random library of RNA molecules of a suitable size (generally >50 bases) is taken. A subset of these is selected on the basis of the desired function. This subset is then amplified by reverse transcription to cDNA and then transcribed using RNA polymerase under sloppy conditions (a fault every *ca.* 100–1000) bases to obtain a new set of RNAs based on the active molecules.

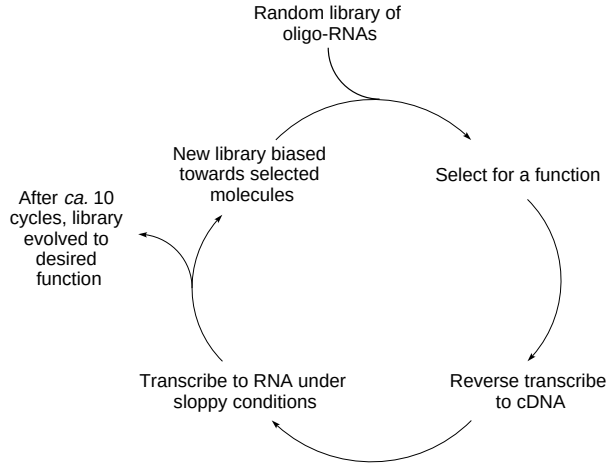


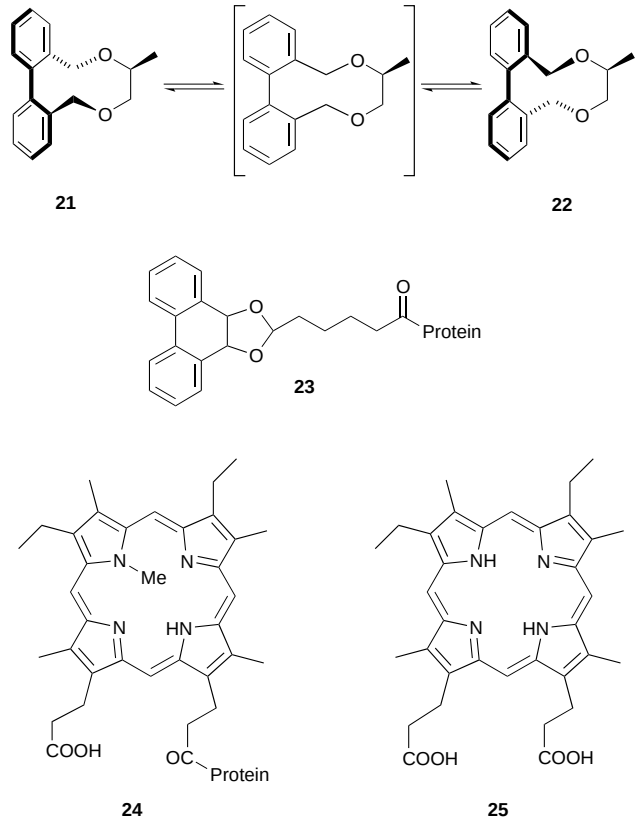
Fig. 3 Schematic sequence for the evolution of an RNA library

Conceptually, this sequence is straightforward and it has been used to good effect to isolate RNAs with desired binding properties: the selection step takes place by affinity chromatography so that molecules which bind the given substrate well are selectively retained on a column. Given the successful use of transition state analogues in the generation of catalytic antibodies, it was originally hoped that evolution of a binder to a transition state analogue would result in a ribozyme with catalytic activity. This has generally proved not to be the case, with a few exceptions. This lack of success is probably due to the lack of functional diversity resulting from the availability of only four building blocks in RNA molecules. Thus, a more dynamic selection process, screening for catalytic activity rather than binding has been employed.

The first example of an *in vitro* evolved ribozyme was that of Beaudry and Joyce.[13] Starting from the *Tetrahymena* ribozyme, which has RNA cleavage activity, a ribozyme with DNA cleavage activity was evolved. The selection procedure employed was a DNA cleavage reaction, in which part of the cleaved DNA remained bound to the 3′ terminal of the ribozyme. The DNA involved was designed to be a primer sequence for cDNA (complementary DNA) synthesis. After incubation of the ribozymes with the cleavage DNA, they were amplified by reverse transcription into cDNA followed by transcription to RNA with an RNA polymerase. Only ribozymes containing the primer were reverse transcribed and amplified, so the amplified ribozyme pool was enhanced in DNA-cleaving molecules. Repetition of this procedure ten times resulted in the generation of ribozymes which accelerated the DNA cleavage reaction by a factor of 100-fold over the wild type. More recently, Tsang and Joyce have improved this acceleration to 10^5 using a more stringent selection procedure.[14]

A pool of RNA polymerase ribozymes was more recently evolved by Bartel and Szostak. They screened for the ability of a ribozyme to join a substrate to itself. Ribozymes that performed this reaction successfully were isolated from the pool by affinity column chromatography for the substrate and then amplified. After ten rounds of evolution, a fairly homologous ribozyme family which showed a rate acceleration of 7×10^6 for RNA polymerisation was obtained from a starting pool of over 10^{15} different random RNA sequences. This is not actually catalysis, but rather templated ligation, and still falls a long way short of the rate acceleration of 3×10^{11} achieved by the enzyme. Similar procedures have now been used to evolve ribozymes for RNA hydrolysis, polynucleotide kinase, self-alkylation and self-acylation.[15]

The first example of a ribozyme obtained as a strong binder to a transition state analogue was for the isomerisation of biphenyl **21** to its diastereomer **22**.[16] Ribozymes were selected



for their ability to bind the transition state analogue **23** attached to a support. An initial library of 10^{15} 195-mers was passed though the affinity matrix and good binders were retained. These were subsequently washed off and amplified. This selection was carried out seven times to afford ribozymes that showed a rate enhancement of 90-fold and catalytic turnover. Hapten **23** has also been used to generate catalytic antibodies for the same reaction.[17] These show greater activity than the ribozyme (k_{cat}/k_{uncat} = 2000), perhaps illustrating the better resolution, and hence better binding, available to an oligopeptide with 20 potential building blocks than to an oligonucleotide with four. Using a similar procedure, TSA **24** has been used to select ribozymes that insert copper into the porphyrin **25**.[18] Analogue **25** is an *N*-alkylated porphyrin, which had previously been shown to be a good structural analogue of the transition state for metal insertion.

Ribozymes clearly show potential in the field of catalysis, but it is likely that their structural make-up does not give enough diversity and specificity for them to rival peptide-based catalysts. They can rival peptides in schemes involving nucleotides as substrates, but fare less well in other areas. Their evolutionary capacity should go some way to offset this deficiency if appropriate selection procedures for catalytic activity can be developed.

4 Imprinted polymers[19]

Antibodies lack features that are important for many practical applications, such as thermal stability and chemical robustness. They have a short lifetime and are expensive to produce, so a synthetic analogue would be of much interest. Copying their mode of production synthetically is clearly not practical, but molecular-sized cavities can be generated in the solid state by polymerisation in the presence of a guest template. This is known as molecular imprinting or 'footprinting'.

The process leading to an imprinted polymer is illustrated schematically in Fig. 4: a bulk polymerisable monomer is mixed with a binding monomer and the guest molecule. Then polymerisation is initiated and the polymer forms around the imprint molecule. After removal of the guest, the polymer should contain cavities of the correct size and shape for the imprint concerned.

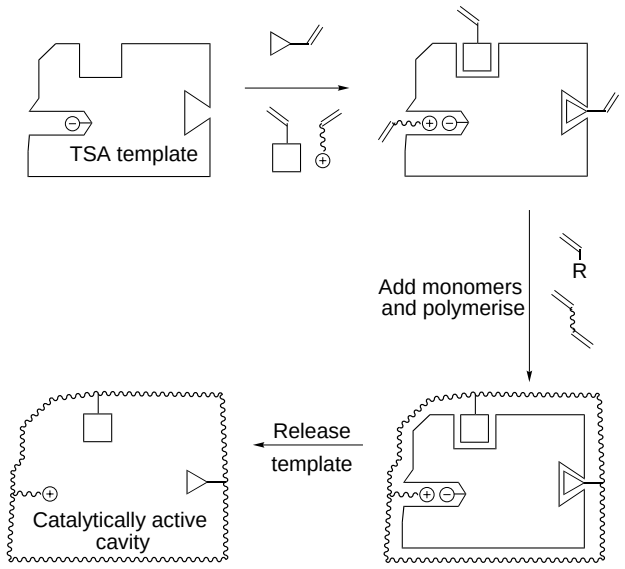
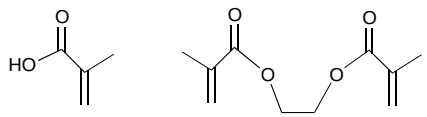


Fig. 4 Schematic sequence for the generation of an imprinted polymer

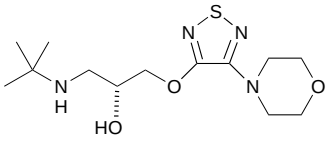
Whilst conceptually very simple, this approach has been difficult to realise in practice. Dickey pioneered the method in the 1940s by preparing silica gels in the presence of methyl orange and homologues, and demonstrating selective adsorption of the appropriate dye.[20] However, the silicas were not very stable, and lacked a diversity of functional groups to participate in binding or catalysis. In a carbon-based polymerisation

scheme, any functional groups can be attached to monomer units, but they need to be appropriately positioned. There are two general strategies to accomplish this, the self-assembly method and the controlled distance method.

An example of a self-assembly directed polymerisation is the generation of a chiral stationary phase for separation of β-adrenergic blockers by polymerisation of methacrylic acid **26** and a crosslinker **27** in the presence of a (*S*)-(−)-timolol



26 **27**

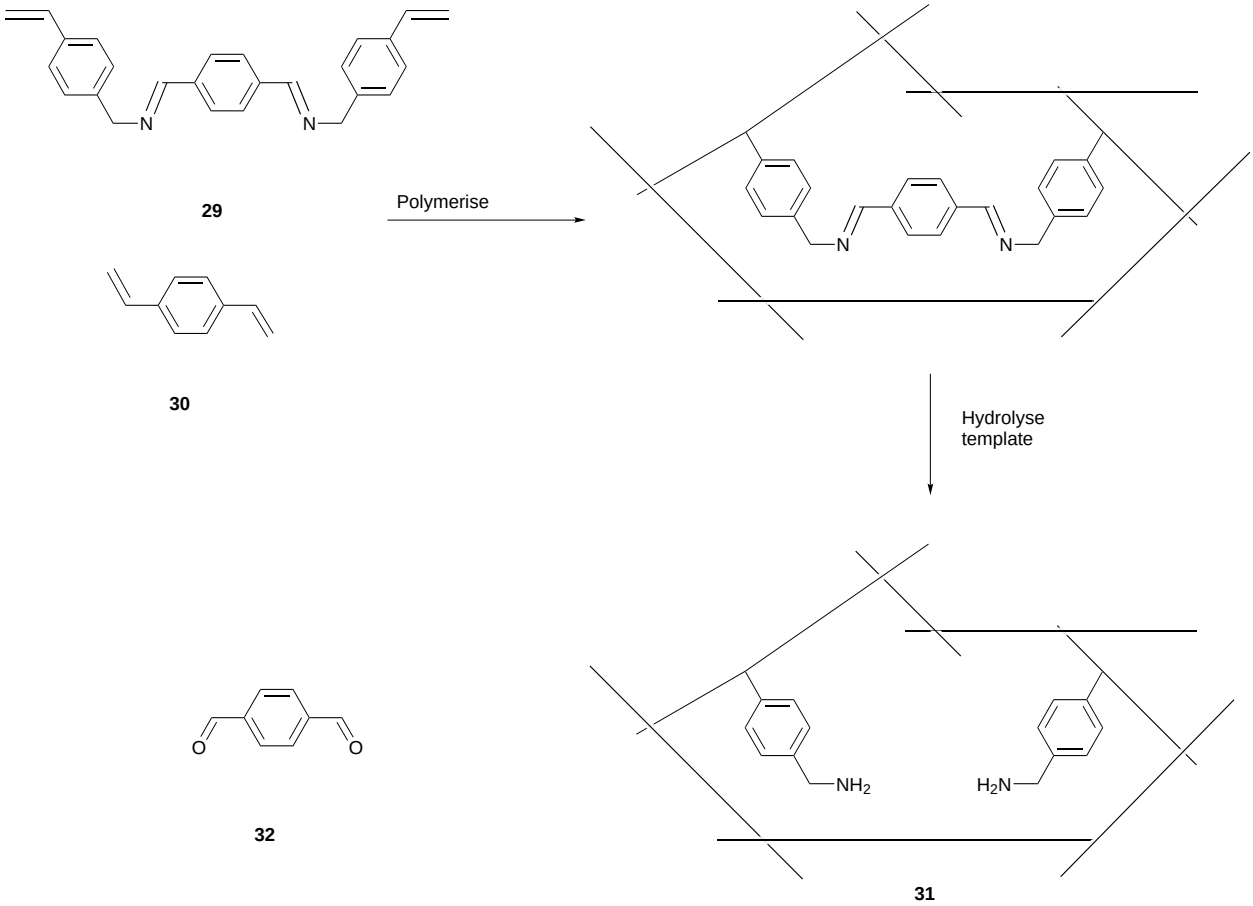


28

template **28**.[21] Monomer **26** forms noncovalent linkages (mainly hydrogen bonds) with the template in organic solution and this self-assembly holds the monomer units in appropriate positions during the polymerisation process. The template is removed from the finished polymer by washing with acetic acid. The polymer generated in this way has a three-dimensional network, with template-complementary binding sites. When used in powdered form as a stationary phase for chromatography, the polymer enables easy enantioselective separation of (*S*)-(−)-timolol from a racemic mixture.

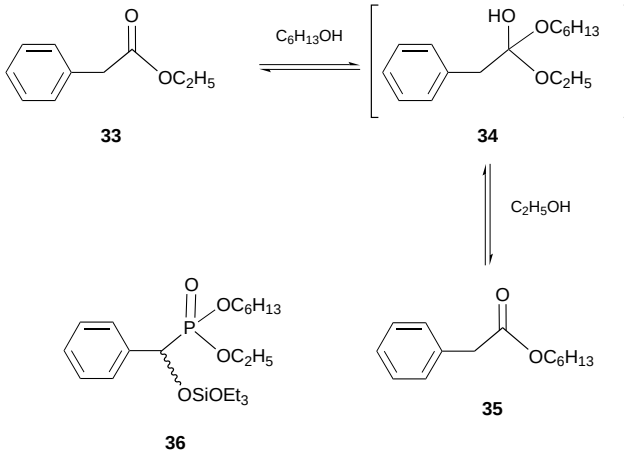
The controlled distance approach uses a monomer that is a conjugate of the template and recognition site. An example is illustrated in Fig. 5. The conjugate monomer **29** contains two benzylic groups for polymerisation spanned by an aromatic diimine. The monomer is co-polymerised with the inert scaffolding monomer **30** and when reaction is complete, the central aromatic spacer is removed by washing with acid to give polymer **31**. This polymer has a cavity with amine binding groups situated in the correct positions and orientations to recognise the dialdehyde guest **32**. The controlled distance method is more likely to be successful than the self-assembly directed method described above if the intermolecular forces are quite weak. On the other hand, as hydrogen bonds and van der Waals forces are strongly dependent on bond length and the controlled method may not optimise these lengths, the self-assembly method could potentially create a better fit if the forces are strong enough. At this stage, neither method has shown itself to be generally superior.

Thus, it is now possible to create polymers incorporating custom-designed binding sites. Such polymers have been used for selective extractions and as chiral stationary phases. From this position, footprinted polymers can progress to the more ambitious challenge of catalysis. Maier has recently generated a silica-based polymer using the controlled distance approach to form cavities capable of catalysing the transesterification of **33** to **35** *via* intermediate **34**:[22] the silyl ether monomer unit **36** contains a phosphonate side chain analogous to those used in catalytic antibodies for mimicking ester hydrolysis. The side chain is thermally cleavable and after integration into the polymer is removed by heating. The polymer is produced by co-polymerisation of $Si(OEt)_4$ and 1 mol% of **36**. At pH 7 no transesterification is observed in the absence of catalyst, but on addition of the imprinted polymer rapid reaction ensues. No detailed kinetics have been reported, but selectivity was



29 **30** **32** **31**

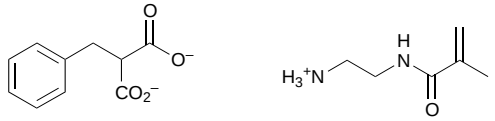
Fig. 5 The controlled distance approach to polymer generation



33 **34** **35** **36**

observed: in experiments where hexanol and phenylethanol compete for the ester, there is tenfold selectivity for the alkyl alcohol, despite the fact that both alcohols react at virtually the same rate in the presence of sulfuric acid as catalyst.

Catalysis by the imprinted polymer was achieved by using cavities that were the correct shape to bind the transition state well. As with antibodies, better rate acceleration would be expected if catalytic functional groups were positioned in the correct positions in the cavity. This has recently been achieved by Shea:[23] for the dehydrofluorination of **10**, a polymer was prepared in the presence of transition state analogue **37**. The template was expected to position the amino-substituted acrylamide **38**, present at 3 mol% in the polymerisation mixture,

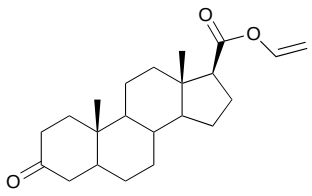


37 **38**

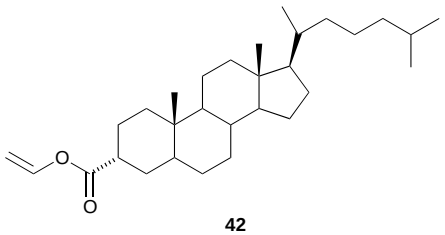
in such a way that there would be a basic amino group in position in the cavity to remove a proton from the substrate **10**. A rate acceleration of 13 times relative to reaction in free solution was observed. This is significantly less than the catalytic antibody's 10^5-fold increase and suffered from product inhibition but the approach does show some promise.

There have also been attempts to impart reaction stereo- and regio-specificity using imprinted polymers. A good example is in the reduction of steroid 3- and 17-ketones with $LiAlH_4$ reported by Byström *et al.*[24] Divinyl benzene was co-polymerised with steroid **39**, which contains a vinyl group linked to the 17β position. After polymerisation, the steroid–polymer ester linkage was cleaved with $LiAlH_4$ and the steroid template was washed away to leave steroid-shaped cavities containing specifically positioned hydroxy groups. Hydride was then attached to the hydroxy groups by treating with further $LiAlH_4$. When the steroid diketone **40** was added to the polymer, it was reduced specifically at the 17 position to give product **41** with good stereochemical control (β : α = 80 : 5). This contrasts with the reduction in solution which has a 99% preference for the 3 position. Using a polymerisation template **42** with the vinyl group attached at the 3α position, a polymer is generated which effects the reduction specifically at the 3 position. The cholestanol product obtained in this way was predominantly the less readily available 3α-isomer (α : β = 72 : 28) in contrast to the solution reaction (α : β = 10 : 90).

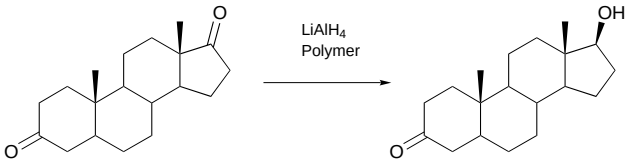
Specifically imprinted polymers can now be prepared with some reliability. By comparison with their biological rivals, catalytic antibodies, they offer many advantages. They are relatively simple and cheap to prepare and can be formed rapidly (2–3 days). They show good mechanical, chemical and thermal stability, can be re-used almost indefinitely, and can be



39



42



40 **41**

stored for years without loss of activity. On the other hand, the observed binding has not approached that of catalytic antibodies in either specificity or strength. In chiral separations, selectivities have usually been less than tenfold. This represents a binding energy difference of *ca.* 5 kJ mol^{-1}, which is less than a single hydrogen bond, implying that there is a non-integral number of binding interactions at each site. How is this possible? One answer lies in the heterogeneity of binding sites: a small fraction of the binding sites may possess specificity similar to antibodies, but the majority are non-specific.[25] Polymer systems currently in use tend to possess floppy backbones that lack any structural or bonding features leading to useful secondary structure, so there is little to prevent formation of an ill-defined continuum of cavity shapes and sizes. The presence of many and diverse non-selective binding sites is the main factor limiting the applications of imprinted polymers. Heterogeneity has the further consequence that the precise interactions responsible for binding cannot be elucidated. This is unsatisfactory from an intellectual point of view and means that systematic changes to improve the activity are not easy to envisage.

5 Combinatorial chemistry[26]

One promising approach to the synthesis of large collections of diverse molecules is combinatorial chemistry. This allows vast 'libraries' of different molecules to be synthesised simultaneously. Strictly speaking, these are not self-designed systems in the same sense as the others described here, but the idea of selection from diversity is similar. In molecular biology, it has long been common practice to use molecular libraries, and techniques for generating similar libraries of organic compounds have been developed over the past few years. The key problem is how to separate and identify a compound with the desired properties: although sub-milligram quantities of pure compounds can be characterised by NMR, a synthetic library of around 100 000 compounds will contain only tiny amounts of each.

The first approach to this problem was spatial segregation. Geysen synthesised different oligopeptides on polyethylene pins arranged in a matrix, whereby each rod was dipped into different reagents. The same conceptual operation can also be employed on a smaller scale by use of photolithography.

However, there is a practical limit to the size of library which can be generated by these methods: it is not feasible to have, for example, 10^6 pins on a support.

A larger number of compounds can be generated by the split synthesis method, where the target molecules are bound to small beads. These can be as small as 50 μm in diameter, so in theory 600 000 beads, each carrying a unique compound, could be present in a 1 cm^3 sample. The beads are separated into equally sized portions and each portion is treated with a different building block reagent in the first step. The beads are then recombined and mixed before being separated again after which a second building block is added. This process can be repeated as many times as required. Each bead in the final library has a product from a specific reaction sequence bound to it as shown in Fig. 6. All the molecules on a given bead are the same and thus, by an appropriate screening method, an active bead can be selected. To identify the molecule on the bead, a method of tagging or a deconvolution process has to be used.

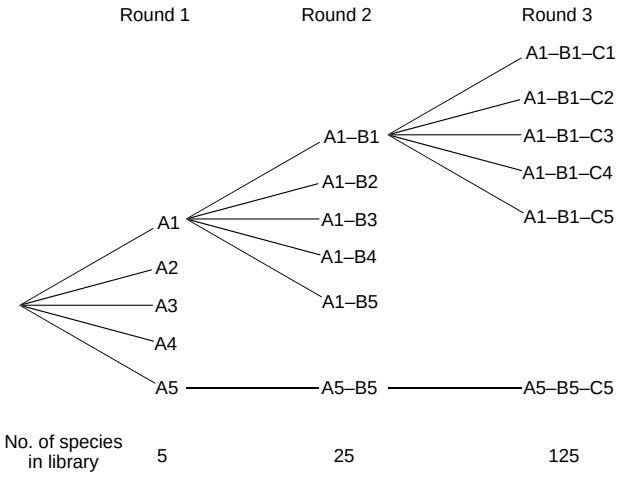


Fig. 6 Schematic of a split synthesis protocol involving five different reagents at each step

These methods allow efficient generation and screening of relatively complex libraries of compounds. Most of the effort to date has concentrated on the generation of binders for drug discovery and this is becoming more successful with the extension of the methods to more diverse sets of molecules. However, the application to the generation of catalysts is still underdeveloped. The methods used for the generation of catalysts can be the same as those in combinatorial synthesis of drug leads, but a more demanding selection procedure is necessary to identify an active catalyst.

Although many libraries have been prepared, they have been of limited structural diversity and topology as they have generally been prepared in a linear fashion and using a limited set of reactions: many reactions that work well in solution are much less effective when applied to beads, and there is a great need for more efficient solid-state organic chemistry. The combinatorial approach will have still greater potential when more methods have been developed for its application to broader and more diverse sets of molecules. Many in the pharmaceutical industry are now attacking this task.

Combinatorial chemistry covers a wider range of approaches than is often realised. Menger has recently investigated polymer-bound polyamides as catalysts for the hydrolysis of a phosphodiester.[27] Various amounts of a range of different acids were condensed with polyallylamine as shown in Fig. 7. Solutions of the combinatorially generated polymers were then screened for catalytic activity in the hydrolysis reaction. In the presence of Zn^{2+} ions, one particular polymer was found to increase the rate of the reaction by a factor of 3×10^4. This is five times more effective than the rate acceleration achieved by a catalytic antibody generated for the same reaction. No rate

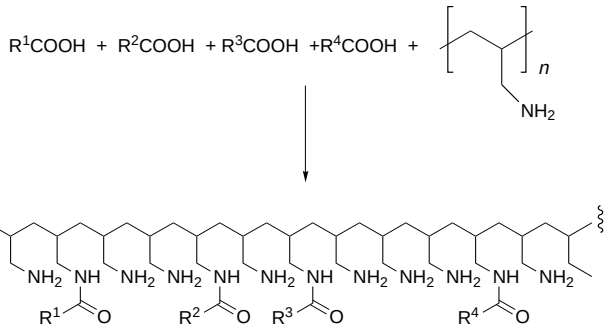


Fig. 7 Menger's combinatorial synthesis of polymeric catalysts for the hydrolysis of a phosphodiester

enhancement was found for polymers bearing only a single substituent. A similar approach yielded polymeric reducing agents.[28] These catalysts were easy to prepare and show good activity, but as with imprinted polymers, their structure is heterogeneous and unknown, and physical selection of the active sequences appears impractical so the key structural features leading to success are also unknown.

A combinatorial approach can also be taken to the optimisation of catalysts previously found to be active. For example, polyoxometalate oxidation catalysts[29] and metal-promoted insertions of carbenes into C–H bonds[30] have each been optimised combinatorially.

6 Thermodynamic templating[31]

In the imprinted polymer approach, the binding of the hapten stabilises a particular spatial arrangement of the building blocks during the assembly process: the best final cavities arise because they were thermodynamically stabilised at some point in their creation. However, heterogeneity is inevitable when irreversible, kinetically controlled reactions are used for bond creation: 'incorrect' bond formation cannot be proof-read and corrected. The problem is exacerbated in the polymer approach which prevents separation of successful from unsuccessful cavities. To avoid these problems, several groups have been working on a new approach that will utilise reversible, thermodynamically controlled bond making to generate hosts. Aliseev and Nelen have developed a protocol for the selection of the best of three isomeric hosts. An affinity chromatography column with a guest attached is connected to an equilibration chamber and the host molecules are circulated around the system. The reaction employed was a photochemical *cis*/*trans* isomerisation, and after several cycles, the amount of the strongly binding isomer increased from 3 to 85%, due to its retention on the high affinity column.[32] The more difficult process of carrying out the equilibration in the same reaction vessel as the binding process has recently been explored by Huc and Lehn.[33] Several components which were able to bond to each other by the reversible formation of imines were mixed together. In the presence of carbonic anhydrase the equilibrium distribution of imines was shifted towards the guest structure that fitted best into the enzyme binding site. In this way a potential inhibitor for the enzyme was generated combinatorially from a 'virtual' combinatorial library. The same conceptual process can be envisaged the other way around, *i.e.* the combinatorial formation of a host in the presence of a template: binding of this guest should stabilise a particular product thermodynamically in a mixture and hence increase the amount of that product formed.

Thermodynamic templating has been used for a variety of purposes,[31,34,35] but not, as far as we know, for generating catalysts in the way proposed here. This concept is shown pictorially in Fig. 8. If several building blocks are assembled in the presence of a guest, then at least one combination would be expected to bind the guest. This product should be reduced in energy ($\Delta H_{binding}$) relative to non-binding products and should

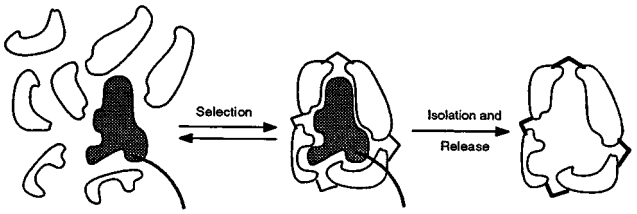
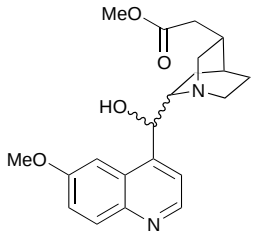


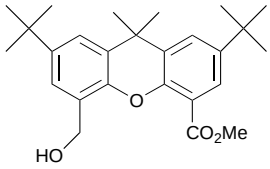
Fig. 8 Thermodynamic templating for lead generation: the host synthesis reaction takes place in the presence of the template, favouring a host which binds the template well

therefore be preferred if its assembly proceeds in a reversible and thermodynamically controlled fashion. Non-binding hosts produced should on average be proof-read and recycled into other products assuming that all possible hosts can be accessed without kinetic barriers. Any strongly binding product will become more concentrated in the reaction mixture through a process of thermodynamic templating and after isolation it could be identified as the best of all possible hosts. Furthermore, use of a different template should allow isolation of a different host from the same reaction mixture. These principles have been utilised in the generation of inorganic coordination complexes such as Lehn's self sorting helices,[36] but as these compounds are held together only by relatively weak interactions in solution, they lack the robust character of covalent molecules.

The approach outlined in Fig. 8 depends on finding a reaction that is fast even in the absence of an excess of one reactant, and on devising suitable building blocks. Our work in this area began with transesterification of steroids (Fig. 9).[37] The starting materials are furnished with an alcohol and a methyl ester group, so removal of methanol from the reaction mixture leads to a dynamic mixture of the cyclic compounds shown. When the reaction is carried out in the presence of metal ions as templates, the distribution of rings is shifted towards the macrocyclic steroid that binds best to that particular metal best.[38] Binding preferences were confirmed by electrospray mass spectrometry[39] In order to obtain good resolution in the binding of potential guests, a wide repertoire of building blocks will be required. The variety of building blocks has recently been extended to alkaloid derivatives **43** and the xanthene derivative **44**, which also undergo efficient reversible cyclisation when



43



44

placed under the transesterification conditions.[40] When rather rigid building blocks are utilised, the distribution of products obtained is small, but for more flexible starting materials a combinatorial mixture containing many different products can be obtained.[41] A major attraction of this approach is that the products are discrete compounds that can be characterised and modified rationally.

Issues that will become important in the future development of these concepts include the nature of the guest (a transition state analogue or a drug for which a receptor is required) and whether the guest should be in free solution or attached to a solid support as shown in Fig. 8. In solution, it will mimic the real binding process required better than when on a solid support, but the latter will allow for easier isolation of a binding product. The current building blocks are quite large, so the resolution they are able to achieve at the binding site is not very high. Either a wide diversity of such building blocks or smaller units will be required to improve this.

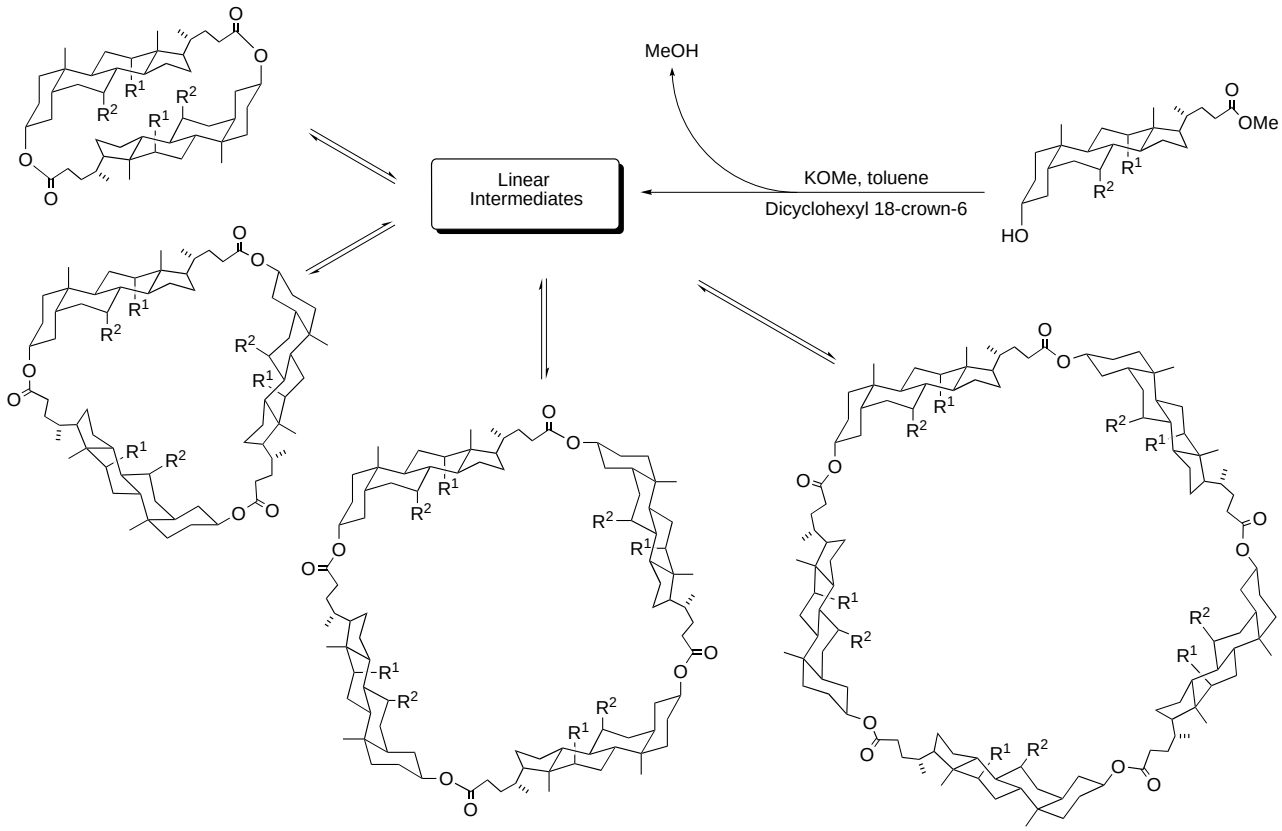


Fig. 9 The thermodynamically controlled cyclisation of steroid derivatives by transesterification

A similar approach has been common for some time in the synthesis of zeolites. It is now possible to design structure-directing templates for the synthesis of microporous solids, and it is surely only a matter of time before a transition state analogue is used to generate a catalytic zeolite.[42] Hill's 'self-repairing' polyoxyanion oxidation catalysts[43] are superficially similar in concept in the sense that catalyst synthesis and catalysis occur simultaneously, but it does not appear that templating by the substrate plays a major rôle.

7 Conclusions

These selection methods have made considerable progress and are now in regular use for lead generation. Each method has inherent problems: there is a numerical and analytical limit to the diversity accessible by library methods; imprinted polymers have heterogeneous and unknown binding sites; antibodies are expensive, inefficient, and their peptidic nature brings problems of thermal and enzymic instability; and ribozymes, in addition to expense and instability, may lack the structural diversity necessary for effective host generation. The more recent idea of thermodynamic templating combines several of the best features of the other techniques but it too will require a wide range of building blocks and careful hapten design in order to fulfil its potential.

In addition to the specific problems peculiar to each of these techniques, all share a general problem: our understanding of transition state analogues and hapten/template design needs to improve in order to take full advantage of all the potentially powerful techniques described in this article, and ways need to be found for inducing catalysts for multi-step reactions. So, despite the fears expressed by some chemists, this type of approach does not give the intellectual responsibility for design away to the molecules themselves: it shifts the challenge to devising better ways of selecting the right molecule.

However, given the relatively short history of the 'selection approach', major progress has been made. In just ten years, catalysts for a variety of reactions have been created and the number of active systems developed in this way may exceed those generated by more conventional designed approaches. We believe that selection approaches to self-designed systems will be strategically important in decades to come.

8 Acknowledgements

We thank the EPSRC and Rhône-Poulenc Rorer for financial support, and numerous colleagues for stimulating discussions.

9 References

1 E. Fischer, *Ber. Dtsch. Chem. Ges.*, 1894, **27**, 2985.
2 L. Pauling, *Nature*, 1948, **161**, 707.
3 N. R. Thomas, *Nat. Prod. Rep.*, 1996, **13**, 479.
4 P. G. Schultz and R. A. Lerner, *Science*, 1995, **269**, 1835.
5 T. Li, R. A. Lerner and K. Janda, *Acc. Chem. Res.*, 1997, **30**, 115.
6 W. P. Jencks, *Catalysis in Chemistry and Enzymology*, McGraw Hill NY, 1969, page 16.
7 T. Koch, J.-L. Reymond and R. A. Lerner, *J. Am. Chem. Soc.*, 1995, **117**, 9383.
8 S. N. Thorn, R. G. Daniels, M.-T. M. Auditor and D. Hilvert, *Nature*, 1995, **373**, 228.
9 F. Hollfelder, A. J. Kirby and D. S. Tawfik, *Nature*, 1996, **383**, 60.
10 P. Wentworth Jr., A. Datta, S. Smith, A. Marshall, L. J. Partridge and G. M. Blackburn, *J. Am. Chem. Soc.*, 1997, **119**, 2315.
11 K. D. Janda, L.-C. Lo, C.-H. L. Lo, M.-M. Sim, R. Wang, C.-H. Wong and R. A. Lerner, *Science*, 1997, **275**, 945.
12 J. R. Lorsch and J. W. Szostak, *Acc. Chem. Res.*, 1996, **29**, 103.
13 A. A. Beaudry and G. F. Joyce, *Science*, 1992, **257**, 635.
14 J. Tsang and G. F. Joyce, *Biochemistry*, 1994, **33**, 5966.
15 E. H. Ekland, J. W. Szostak and D. P. Bartel, *Science*, 1995, **269**, 364.
16 J. R. Prudent, T. Uno and P. G. Schultz, *Science*, 1994, **264**, 1924.
17 T. Uno, J. Ku, J. R. Prudent, A. Huang and P. G. Schultz, *J. Am. Chem. Soc.*, 1996, **118**, 3811.
18 M. M. Conn, J. R. Prudent and P. G. Schultz, *J. Am. Chem. Soc.*, 1996, **118**, 7012.
19 G. Wulff, *Angew. Chem. Int. Ed. Engl.*, 1995, **34**, 1812.
20 F. H. Dickey, *Proc. Natl. Acad. Sci. USA*, 1949, **35**, 227.
21 M. Kempe and K. J. Mosbach, *Chromatogr.*, 1995, **694**, 3.
22 J. Heilmann and W. F. Maier, *Angew. Chem. Int. Ed. Engl.*, 1994, **33**, 471.
23 K. J. Shea and J. V. Beach, *J. Am. Chem. Soc.*, 1994, **116**, 379.
24 S. Byström, A. Börje and B. Akermark, *J. Am. Chem. Soc.*, 1993, **115**, 2081.
25 L. I. Anderssen, R. Müller, G. Vlatakis and K. Mosbach, *Proc. Natl. Acad. Sci. USA*, 1995, **92**, 4788.
26 G. Lowe, *Chem. Soc. Rev.*, 1995, **24**, 309.
27 F. M. Menger, A. V. Eliseev and V. A. Migulin, *J. Org. Chem.*, 1995, **60**, 6666.
28 F. M. Menger, C. A. West and J. Ding, *Chem. Commun.*, 1997, 633.
29 C. L. Hill and R. D. Gall, *J. Mol. Catalysis*, 1996, **114**, 103.
30 B. M. Cole, K. D. Shimizu, C. A. Krueger, J. P. A. Harrity, M. L. Snapper and A. H. Hoveyda, *Angew. Chem. Int. Ed. Engl.*, 1996, **35**, 1668.
31 S. Anderson, H. L. Anderson and J. K. M. Sanders, *Acc. Chem Res.*, 1993, **26**, 469.
32 A. V. Eliseev and M. I. Nelen, *J. Am. Chem. Soc.*, 1997, **119**, 1147.
33 I. Huc and J-M. Lehn, *Proc. Natl. Acad. Sci. USA*, 1997, **94**, 2106.
34 W. C. Still, P. Hauck and D. Kempf, *Tetrahedron Lett.*, 1987, **28**, 2817.
35 J. T. Goodwin and D. G. Lynn, *J. Am. Chem. Soc.*, 1992, **114**, 9197.
36 V. C. M. Smith and J.-M. Lehn, *Chem. Commun.*, 1996, 2733.
37 P. A. Brady, R. P. Bonar-Law, S. J. Rowan, C. Suckling and J. K. M. Sanders, *Chem. Commun.*, 1996, 221.
38 P. A. Brady and J. K. M. Sanders, *J. Chem. Soc., Perkin Trans. 1*, 1997, 3237.
39 P. A. Brady and J. K. M. Sanders, to be submitted.
40 S. J. Rowan, D. G. Hamilton, P. A. Brady and J. K. M. Sanders, *J. Am. Chem. Soc.*, 1997, **119**, 2578.
41 S. J. Rowan and J. K. M. Sanders, *Chem. Commun.*, 1997, 1407.
42 D. W. Lewis, D. J. Willock, C. R. A. Catlow, J. M. Thomas and G. J. Hutchings, *Nature*, 1996, **382**, 604.
43 C. L. Hill and X. Zhang, *Nature*, 1995, **373**, 324.

Received, 4th April 1997
Accepted, 23rd May 1997