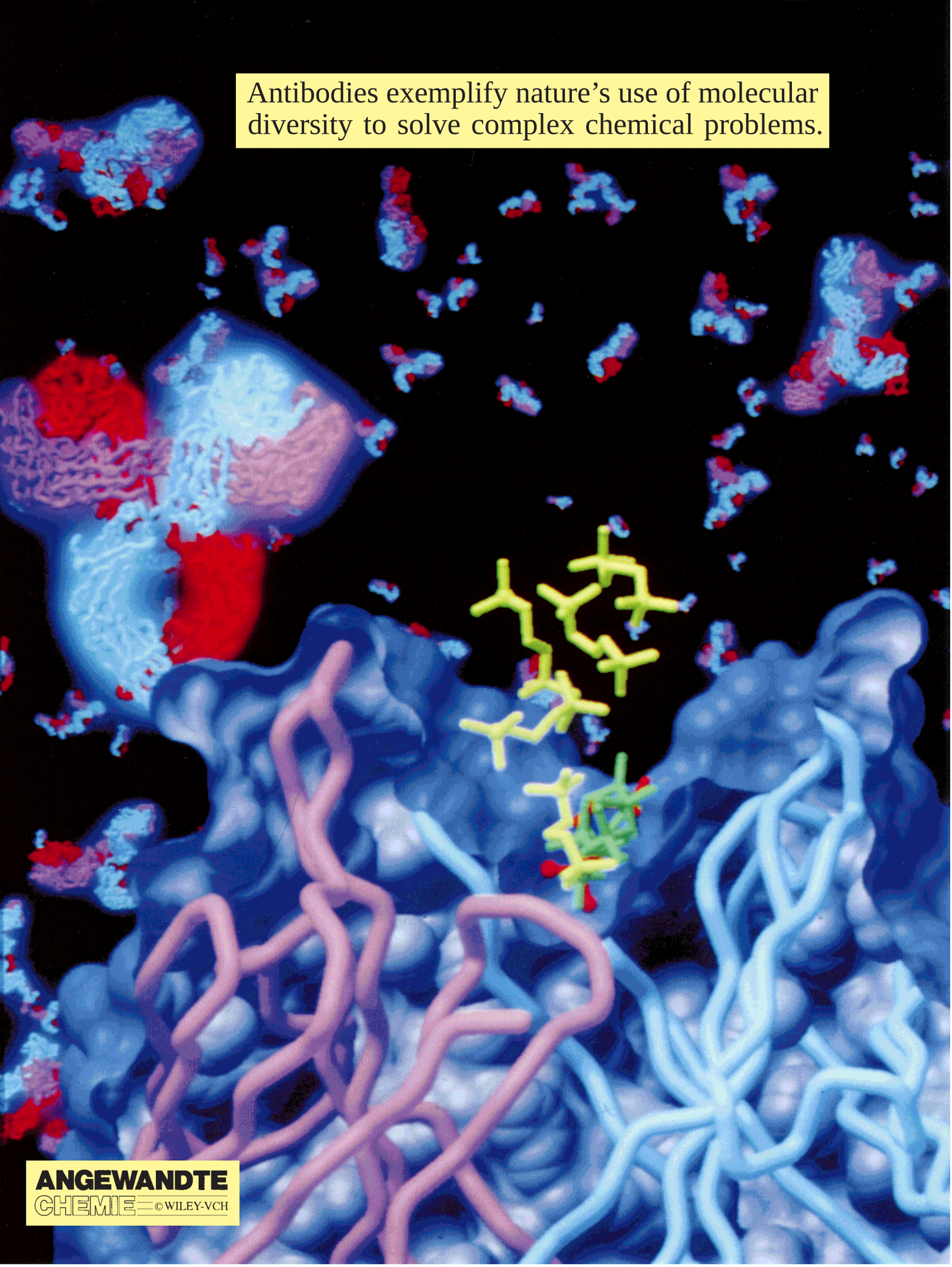


ANGEWANDTE
CHEMIE

Generating New Molecular Function: A Lesson from Nature

David R. Liu and Peter G. Schultz*

As the focus in molecular synthesis shifts increasingly from structure to function, more efficient strategies must be developed to generate molecules with defined biological, chemical, or physical properties. Nature has produced molecules with a wide number of complex biological functions by synthesizing large, diverse collections or libraries of chemical structures and screening or selecting for a desired function. This same approach is now being applied in conjunction with the tools and principles of chemistry, biology, and physics to generate synthetic molecules with novel or enhanced properties, ranging from biological catalysts and bioactive small molecules to luminescent metal oxides. Analysis of the structures and properties of these molecules, together with the wealth of related information produced in these experiments, is leading to an increased understanding of the relationship between molecular structure and function.

Keywords: combinatorial chemistry · materials science · nucleic acids · proteins · synthetic methods

1. Introduction

Our ability to create new molecular structures with novel physical, chemical, or biological properties has had a major impact on many fields of science including the biomedical sciences, chemistry, and solid-state physics. Examples range from the synthesis of antibiotics and genes to photoresists and high-temperature superconductors. As the structural and functional complexity of target molecules increases, our ability to define the precise structural requirements that result in a desired set of properties becomes increasingly limited. This in turn leads to a large increase in the number of molecules that must be iteratively synthesized and tested to identify those structures with a desired function—a time- and resource-intensive process. Among the strategies being developed to meet the challenges created by the ever growing need for molecules with new or enhanced properties are a number of biologically inspired approaches. The latter derive from our recognition that nature has produced an impressive array of complex molecular structures and assemblies with functions ranging from gene regulation and signal transduction to photosynthesis and protein biosynthesis. As we begin to understand the structures and molecular mechanisms involved in these processes, we can combine these biological insights, as well as the processes themselves, with the tools of the physical sciences to create new molecules with functions found neither in nature nor the laboratory.

One such example of this synergy between chemistry and biology is the development and application of combinatorial strategies. This approach, in which large, diverse collections, or "libraries", of molecules are generated and subsequently screened or selected for novel functions, stems from the combinatorial processes in nature. For example, both the humoral and cellular arms of the immune system have developed highly sophisticated combinatorial genetic mechanisms for generating molecular diversity and selecting receptors that can recognize foreign antigens with high affinity and selectivity. The mechanisms of these processes are relatively well understood.[1–3] The three-dimensional framework of the antibody molecule consists of two polypeptide chains (heavy (H) and light (L)) containing six loops on an eight-stranded β-sheet framework.[4] This structure is assembled from four different gene segments, variable (V), diversity (D), joining (J), and constant (C), each of which can be chosen from a number of distinct genes. The combinatorial association of V, D, and J gene segments with additional junctional diversity occurring at the V_L–J_L, V_H–D, and D–J_H joining regions leads to a structurally diverse population of germ-line antibodies (Figure 1). After a "lead" antibody structure is selected from this pool based on its ability to bind a foreign substance, its affinity is increased as the immune response proceeds by somatic mutation. This process, which alters bases throughout the DNA sequences encoding the variable region, results in additional diversity. Thus, in its most basic form, the combinatorial process involves the synthesis of large numbers of distinct structures around a central frame-

[*] P. G. Schultz, D. R. Liu
Howard Hughes Medical Institute
and
Department of Chemistry
University of California, Berkeley
Berkeley, CA 94720–1460 (USA)
Fax: (+1) 510-643-6890
E-mail: pgschultz@lbl.gov

  1433-7851/99/3801-0037 $ 17.50+.50/0

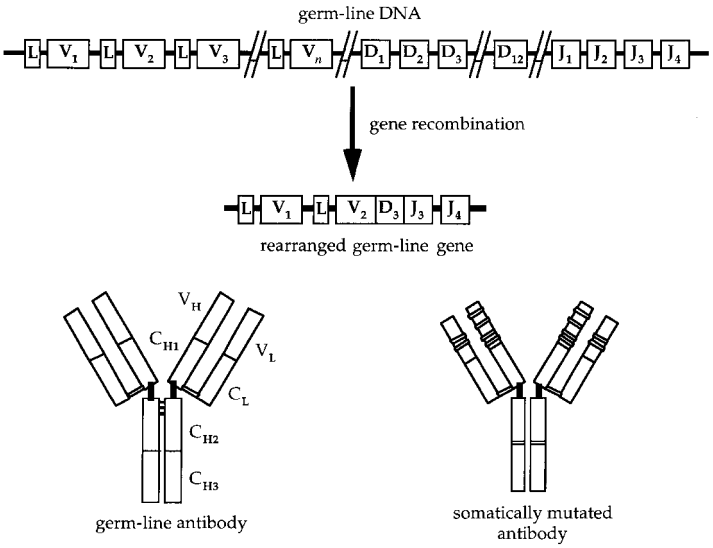


Figure 1. Combinatorial association of V, D, and J genes to form an antibody molecule.

work from sets of building blocks. The choice of framework structure and building blocks is best made based on empirical and theoretical models at hand. Those structures with a desired function then are identified and further optimized by repeated rounds of screening or selection and mutation.

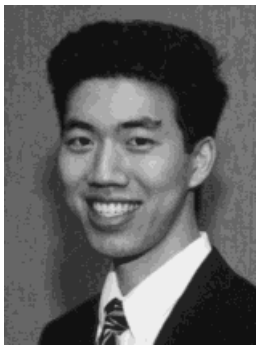
This combinatorial approach is applicable to any molecular structure that can be assembled either in a stepwise or concerted fashion from a set of molecular precursors, and where a screen or selection for a desired function exists. This includes oligomeric molecules such as nucleic acids and peptides, nonpolymeric molecules such as natural products, or even solid-state materials such as superconductors or polymers. To illustrate the utility and scope of this new synthetic strategy, as well as the challenges that remain, a number of examples, both from our laboratory and many others, will be described. These applications cover a broad spectrum, ranging from biological catalysis and drug discovery to materials science.

2. Protein and Polypeptide Protein Diversity

2.1. Immunological Diversity and Catalytic Antibodies

The notion that natural immunological diversity can be used to generate novel chemical function was first illustrated with the generation of catalytic antibodies.[5–7] Immunological diversity is clonally selected on a time scale of weeks (in contrast to evolutionary diversity), and is thus "programmable" in the laboratory setting. However, binding affinity rather than catalytic activity serves as the basis for antibody selection, and therefore mechanistic principles such as selective binding and stabilization of the transition state must be used to generate catalytic antibodies. For example, to generate an antibody that catalyzes the insertion of a metal ion into porphyrin (the last chemical step in heme biosynthesis), antibodies were generated against *N*-methylprotoporphyrin (**1**). This molecule mimics the putative transition state, a conformationally distorted porphyrin ring, of the metalation reaction. An antibody specific for **1** catalyzes the insertion of metal ion into mesoporphyrin with rates similar to that of the

P. G. Schultz



D. R. Liu

Peter G. Schultz did his undergraduate and graduate work at the California Institute of Technology. His thesis work resulted in the first synthetic molecules that sequence-selectively cleave DNA. In 1985, after postdoctoral studies at the Massachusettes Institute of Technology, he joined the faculty of the University of California at Berkeley, where he is currently Professor of Chemistry, a Principal Investigator at Lawrence Berkeley National Laboratory and an Investigator of the Howard Hughes Medical Institute. His research—which spans the interface of biology, chemistry, and materials science—includes a) the development of catalytic antibodies; b) the application of molecular diversity to problems in biomolecular recognition and catalysis, drug discovery, and materials science; c) the development of methods for incorporating unnatural amino acids selectively into proteins; and d) most recently single-molecule biological imaging. He was a founding scientist of Affymax Research Institute and is a cofounder of Symyx Technologies. His awards include the Waterman Award of the National Science Foundation, membership in the National Academy of Sciences, and the 1994 Wolf Prize in Chemistry. Recently, he became a Professor of Chemistry at the Scripps Research Institute, and has been named as the director of the new Novartis Institute for Functional Genomics in San Diego. This review is based on his opening lecture at the 36th IUPAC Congress in Geneva in the summer of 1997.

David R. Liu was born in Riverside, California in 1973. He received his B. A. in chemistry from Harvard University in 1994, where he conducted research on steroid biosynthesis under the guidance of Professor E. J. Corey. Presently he is a graduate student in Professor Schultz's group at the University of California, Berkeley, and investigates in vitro and in vivo methods for the site-specific incorporation of unnatural amino acids into proteins.

natural biosynthetic enzyme, ferrochelatase.[8] Resonance Raman studies revealed that both the enzyme and antibody facilitate the reaction by distorting the planar porphyrin substrate, which is consistent with the proposed reaction mechanism. This distortion involves doming of the porphyrin ring in the case of the enzyme, and for the antibody an up–down distortion mimicking that in *N*-methylmesoporphyrin[9] (Figure 2). Crystallographic analysis of the antibody·*N*-methylprotoporphyrin complex has revealed the structural basis for this distortion, and shown that somatic mutations in response to **1** play an important mechanistic role.[10]

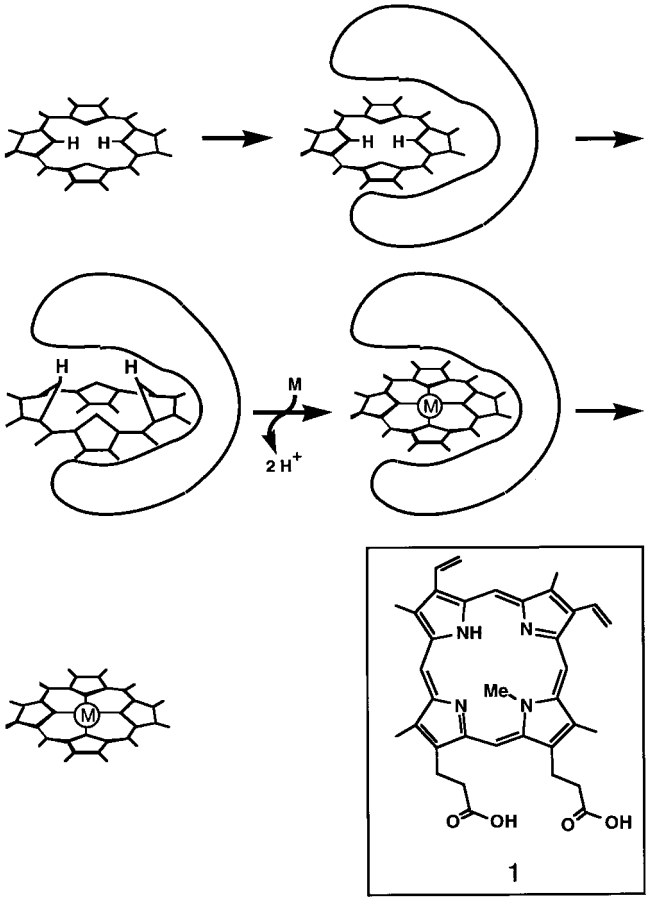


Figure 2. Representation of *N*-methyl protoporphyrin (**1**) distortion induced by an antibody with ferrochelatase activity.[10]

A second example involves a transesterification reaction catalyzed by an antibody generated against a phosphonate diester transition state analogue.[11] The antibody catalyzes the corresponding acyl transfer reaction of thymidine and an alanyl ester with an effective molarity of 3×10^4 M ($(k_{cat}/K_m)/k_{uncat} \approx 10^8$). Moreover, the antibody does not catalyze acyl transfer to water. NMR studies of the Michaelis complex suggest that the high catalytic efficiency and selectivity of this antibody result largely from an optimal orientation of the acyl donor and nucleophile in the antibody active site, reflecting the orientation of the phosphonate moiety in the hapten.[12] Thus, by selecting the natural diversity of the immune system on the basis of mechanistic criteria (for example, distortion of a planar porphyrin ring or proximity of reactive groups), it is possible both to generate antibodies with enzyme-like properties as well as to test basic principles of enzymic catalysis.

Other approaches have been developed to generate catalytic antibodies including covalent catalysis, proximity effects, and general acid–base catalysis.[7] In addition, strategies have been developed to directly select immunological diversity for catalytic function both in vivo and in vitro. Such approaches have been used to generate antibodies that catalyze a wide array of enzymic and nonenzymic reactions. These include stereoselective redox reactions, pericyclic reactions, rearrangement and acyl transfer reactions, as well as a number of "difficult" chemical transformations including kinetically disfavored exo-Diels–Alder and anti-Baldwin cyclization reactions.[13] A recent example involved the generation of an aldolase antibody by means of a mechanism-based selection with a hapten capable of forming a covalent bond with active-site lysine residues. The catalytic efficiency and stereoselectivity of this antibody are remarkably similar to those of Class I aldolase enzymes.[14] Catalysis proceeds through imine formation between the ketone substrate and the ε-amino group of an active site-lysine residue, followed by enamine formation, condensation with an aldehyde, and subsequent hydrolysis to give the aldol product—a mechanism similar to that of the enzyme. At the same time the antibody tolerates a remarkably broad range of substrate structures (Figure 3) and thus represents one of the most general stereoselective catalysts to date for carbon–carbon bond formation.[15] This example beautifully illustrates the catalytic potential of immunological diversity that can be harnessed with proper chemical instruction.

The characterization of catalytic antibodies also provides important new insights into the evolution of binding and catalytic function, as well as the combinatorial processes of the immune system itself. For example, structural and mechanistic analyses of antibody-catalyzed oxy-Cope rearrangement and Diels–Alder reactions have illustrated a number of mechanisms whereby binding energy can be used affect the stereoelectronics of pericyclic reactions.[16–18] Detailed biophysical and structural studies have also been carried out to examine the affinity maturation of an esterolytic antibody generated against a nitrophenyl phosphonate transition state analogue. Nine somatic mutations resulted in a catalytic antibody that binds hapten with 35 000-fold greater affinity than its germ-line precursor. In contrast to the affinity-matured antibody, which binds hapten in a "lock-and-key" fashion, the germ-line antibody can adopt more than one combining site conformation Both antigen binding and somatic mutation stabilize the conformation with optimal hapten complementarity (Figure 4). This result suggests that conformational diversity (a key element of Pauling's chemical instruction model of antibody specificity) may play an important role in expanding the binding potential of the primary immune response.[19–21] This study further showed that the binding and catalytic properties of the antibody are greatly affected by mutations distant from the bound ligand, an important lesson for those involved in efforts to modify protein function. The study of catalytic antibodies has also provided insights into the structural basis for the polyspecificity of families of germ-line antibodies, another factor that likely contributes to the remarkable binding potential of the primary antibody repertoire.[17] Thus, by generating new

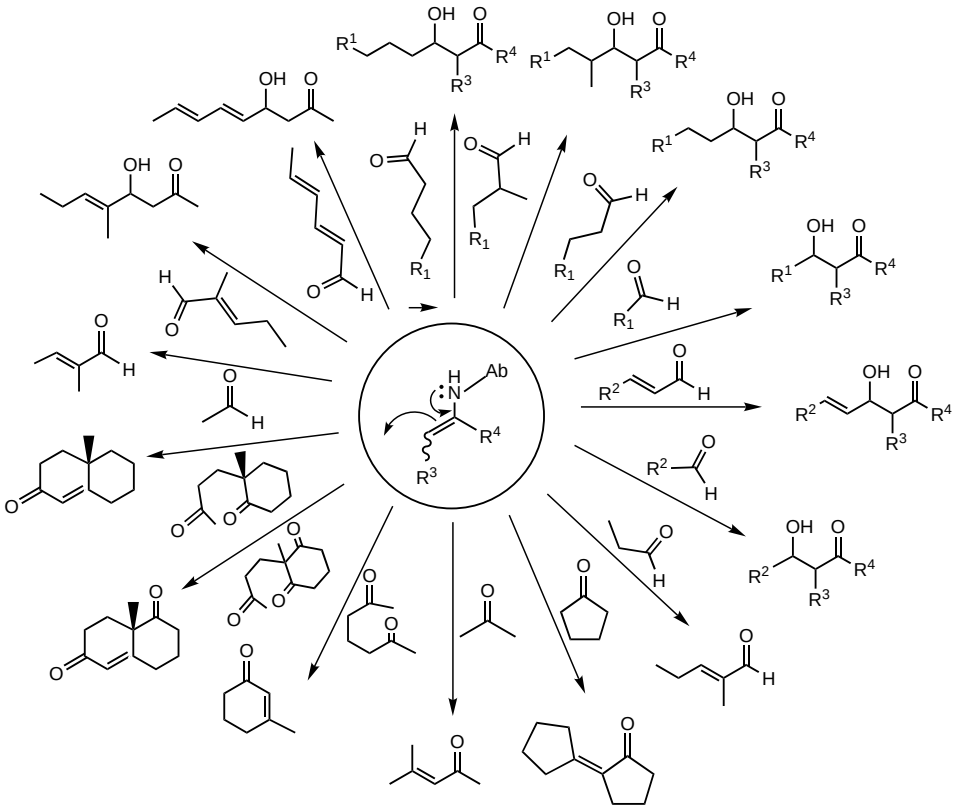


Figure 3. Schematic representation of an antibody (Ab) aldolase with broad substrate specificity.[14] Shown in the sphere is the covalent enamine intermediate that is formed from the ketone substrate and the ε-amino group of an active-site lysine residue.

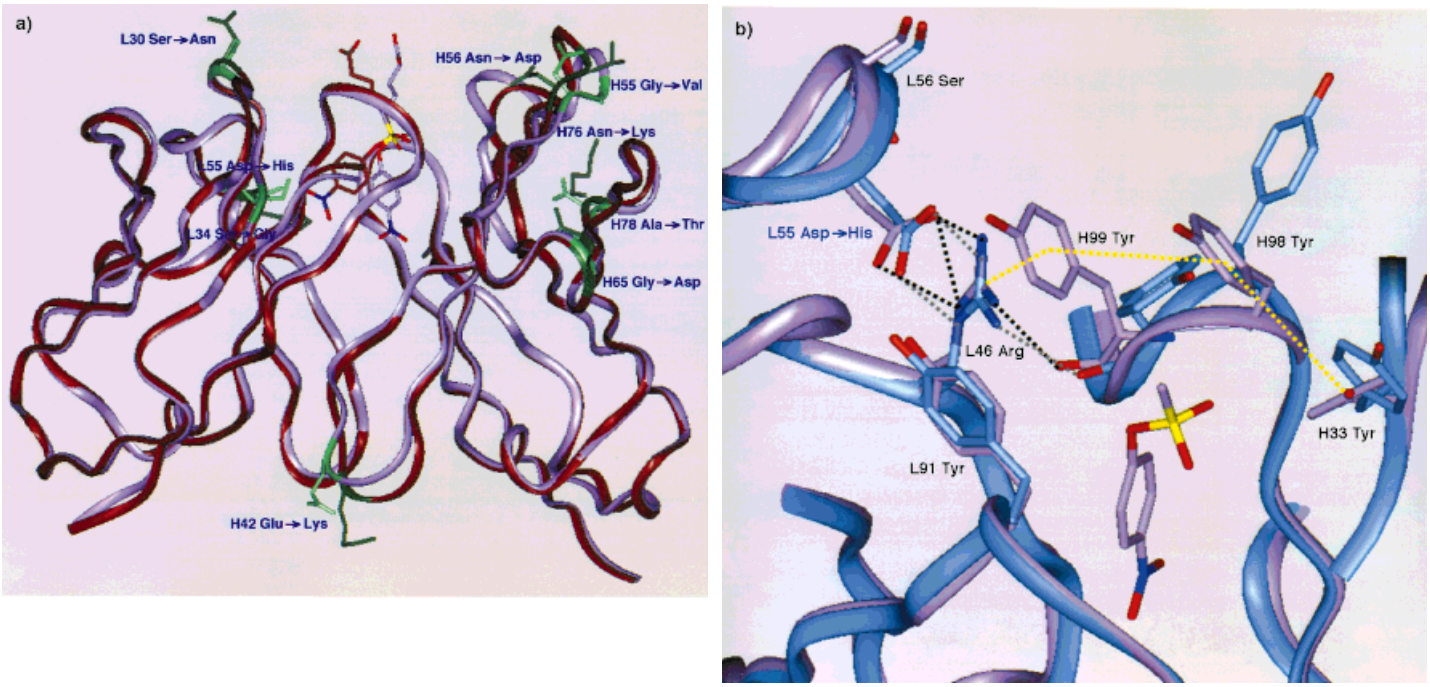


Figure 4. Conformational diversity in an antibody esterase. Whether the amino acid belongs to a heavy or light chain is indicated with the letter H or L before the sequence number. a) Superposition of the germ-line Fab–hapten complex (Fab = antigen-binding fragment of an antibody; purple) and the affinity-matured Fab–hapten complex (red). Somatically mutated residues (X → Y) are shown in green. b) Superposition of the structures of the germ-line Fab domain without hapten (blue) and the germ-line Fab–hapten complex (purple). Gray dotted lines denote hydrogen bonds in the structure of the germ-line Fab without hapten, while black dotted lines denote hydrogen bonds in the germ-line Fab–hapten complex.[19] Yellow dotted lines show the formation of a "double T-stack" arrangement between the side chains of three tyrosine residues.

functions from antibody diversity, new insights are being gained into the relationship between structure and function in the immune system itself.

Researchers have expanded upon the natural diversity generated by the immune system. In 1989 a novel vector system based on the bacteriophage lambda was used to

express a synthetic combinatorial library of Fab fragments of the mouse antibody repertoire in *Escherichia coli*.[22] This system allowed rapid and easy identification of monoclonal Fab fragments that bind a given antigen, in a form suitable for genetic manipulation. For example, monoclonal Fab fragments against a transition state analogue hapten were generated using this technique.

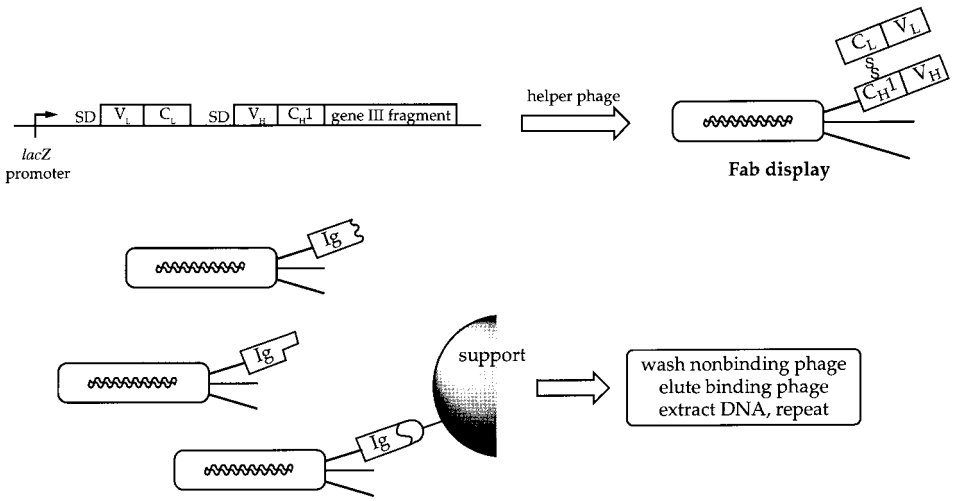


Figure 5. Phage display and screening of an antibody library. SD = Shine–Dalgarno sequence (ribosome binding site on prokaryotic mRNA); Ig = immunoglobulin domain.

Phage display,[23–26] in which a peptide or protein is expressed on the surface of filamentous phage, as well as other schemes linking the polypeptide chain to the encoding DNA,[27–29] have also proven to be very powerful techniques for generating diversity in vitro. By mutagenizing codons randomly or at specific sites in a gene with synthetic oligonucleotides, libraries of greater than 10^8 different polypeptide sequences can be generated and expressed on the phage surface. Subsequent screening by affinity-based techniques using immobilized ligands, followed by elution of bound phage, amplification, and additional rounds of screening can lead to the isolation of high-affinity receptors or ligands. The identity of a particular member of the library is determined by sequencing the phage DNA.

With phage display techniques, synthetic combinatorial libraries of antibody genes from multiple species (including human), or from large synthetic repertoires of antibody encoding genes, have been expressed on phage as single-chain Fv or Fab fragments (Figure 5).[30, 31] High-affinity antibodies have been isolated that bind a wide variety of small ligands and macromolecules including viral and tumor antigens, and that catalyze a number of chemical reactions.[24, 32] More recently, methods have been developed that allow the direct chemical selection of catalysts from phage displayed Fab libraries.[33] For example, an antibody glycosidase was isolated using a mechanism-based inhibitor that produced a reactive quinone methide, resulting in the covalent trapping of catalytic clones onto a solid support for further rounds of amplification and selection. In vitro methods for generating antibody diversity may become as important as hybridoma technology in the development of monoclonal antibodies for diagnostic, therapeutic, and chemical applications. Phage display methods are also being used to identify other protein frameworks, such as the loops of four-helix bundle proteins, which when randomized might function as miniantibodies.[34]

Most recently, we have developed a general scheme for the in vitro evolution of protein catalysts that directly links substrate turnover to a selective advantage in a biologically amplifiable system.[35] Substrate is covalently and site-specifically attached by a flexible tether to the pIII coat protein of a filamentous phage that also displays the protein catalyst. Intramolecular conversion of substrate into product provides a basis for isolating active catalysts from a library of proteins, either by release or attachement of the phage to a solid support, or by the capture of the product by a product-specific antibody (Figure 6). This methodology was developed using the enzyme staphylococcal nuclease (SNase) as a model. Phage

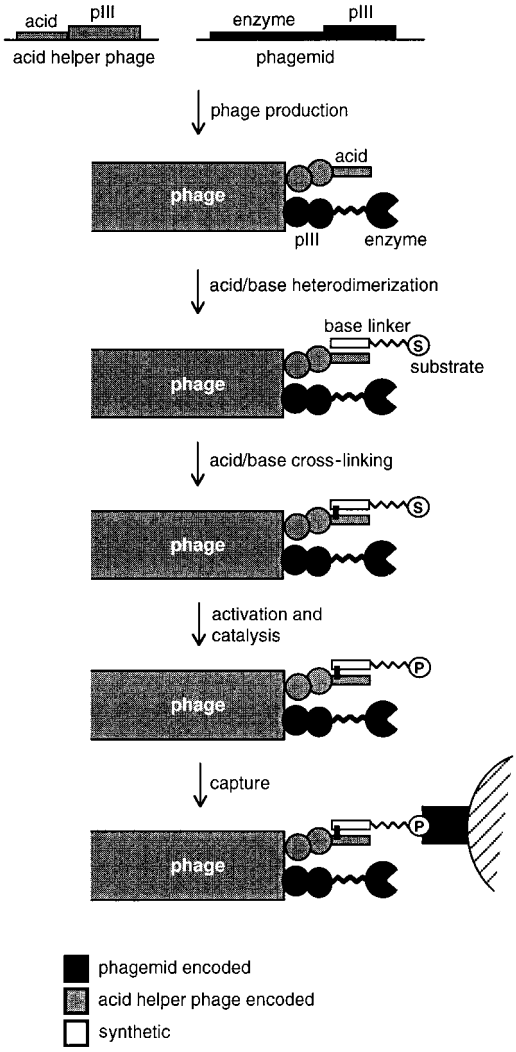


Figure 6. A general scheme for the in vitro evolution of protein catalysts.

displaying SNase can be enriched up to 1000-fold in a single step from a librarylike ensemble of phage displaying noncatalytic proteins. Additionally, this approach should allow one to functionally clone natural enzymes based on their ability to catalyze specific reactions (for example, glycosyl transfer, sequence-specific proteolysis or phosphorylation, polymerization), rather than their sequence or structural homology to known enzymes.

2.2. Peptide Diversity

Libraries of linear and cyclic peptides displayed on phage have been used to identify high-affinity, selective ligands which bind a range of biological receptors including antibodies, enzymes, lectins, cell surface receptors, signal transduction proteins, and even nucleic acids.[24] One of the most impressive examples was the isolation of a disulfide-bonded cyclic peptide containing 14 amino acids with the consensus sequence YXCXXGPXTWXCXP (where X can be several amino acids), that binds and activates the receptor for the cytokine erythropoietin (EPO).[36] The successful isolation of this peptide required both variations in the stringency of the screening (by varying binding avidity and elution conditions) and the generation and screening of libraries derived from low-affinity peptides. The isolated peptide binds the extracellular domain of the EPO receptor (EPOR) with an IC_{50} of roughly 0.2 μM, and stimulates erythropoiesis in mice through a signaling pathway that appears to be identical to that induced by the natural ligand. The molecular basis for the agonist activity of this peptide was determined by analysis of the three-dimensional crystal structure of the complex of EPOR and a related 20 amino acid peptide (Figure 7).[37] The

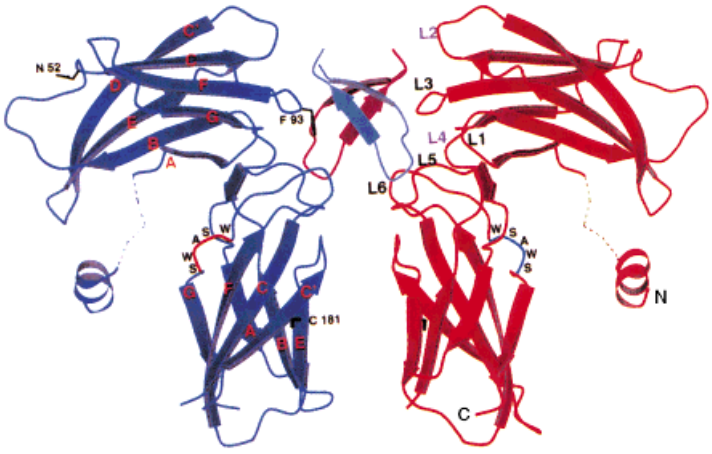


Figure 7. Complex of a 20-residue peptide isolated from a phage peptide library with the extracellular domain of the EPO receptor.[37]

structure reveals that the peptide forms a dimer consisting of a four-stranded, anti-parallel β-pleated sheet and two type I β turns. The peptide induces an almost perfect twofold dimerization of the receptor through a combination of hydrophobic and hydrogen-bonding interactions with segments of four loop regions of EPOR. The successful isolation of this novel peptide, which may define a minimal epitope for activation of EPOR, suggests that peptides may be able to recapitulate many of the other functions of larger proteins.

Another interesting application of phage display techniques is the determination of substrate specificities of proteases by screening peptide libraries with an assay based on peptide hydrolysis.[38] In this case the amino terminal domain of the pIII protein of filamentous phage is fused to a tag by a randomized peptide linker. Phage are then bound to an affinity support specific for the tag. Upon protease treatment, phage displaying peptides that are good substrates are cleaved from the support. After several rounds of binding, proteolysis, and amplification, sensitive and resistant substrate sequences were identified for the proteases subtilisin BPN′ and factor X_a. Peptide libraries presented on phage have also been panned directly against cells to identify peptide sequences that bind and/or enter specific cell types. These peptides may ultimately provide the basis for intravenously delivered gene therapy vectors[39] or tissue-specific therapeutic agents.[40]

2.3. Evolving Protein Function

The natural rate of genetic diversification by point mutation can be increased with UV, chemical, or enzymatic mutagenesis, or genetic methods such as the use of mutator strains. Proteins with random mutations generated by these methods can be subjected to an appropriate screen or selection to isolate mutants with altered properties. For example, these methods have been used to change the substrate specificity of β-galactosidase,[41] ribitol dehydrogenase,[42] and alkylamidases.[43]

More recently, a number of highly efficient methods have been developed for generating larger and more diverse libraries of mutants including cassette mutagenesis,[44, 45] error-prone polymerase chain reaction (PCR),[46] and the technique of DNA shuffling[47, 48] These methods have accelerated our ability to carry out artificial evolution by introducing diversity at a higher rate, and have been used to modify the stability, catalytic activity, and specificity of proteins.[49–54] In particular, DNA shuffling mimics the combinatorial processes of the immune system and natural evolution by combining both random point mutagenesis of a gene with in vitro homologous recombination to generate libraries of structurally diverse mutants (Figure 8). This approach allows one to more efficiently search large sequence spaces for mutations that lead to additive or even cooperative enhancements in binding or catalytic function. When directly compared with processes such as error-prone PCR that iteratively build up beneficial mutations without recombination, DNA shuffling has yielded evolved proteins with higher desired activities in fewer rounds of selection.[47] DNA shuffling has been used to enhance the properties of a number of proteins including glycosidases, β-lactamases, antibodies, green fluorescent protein (GFP), and growth factors.[48] For example, three cycles of DNA shuffling and two cycles of backcrossing (recombination with the wild-type gene) with selection on increasing concentrations of the antibiotic cefotaxime resulted in a β-lactamase with six mutations which conferred to the cell a 32000-fold higher resistance to the antibiotic.[47] The mutations, like the somatic mutations in affinity-matured

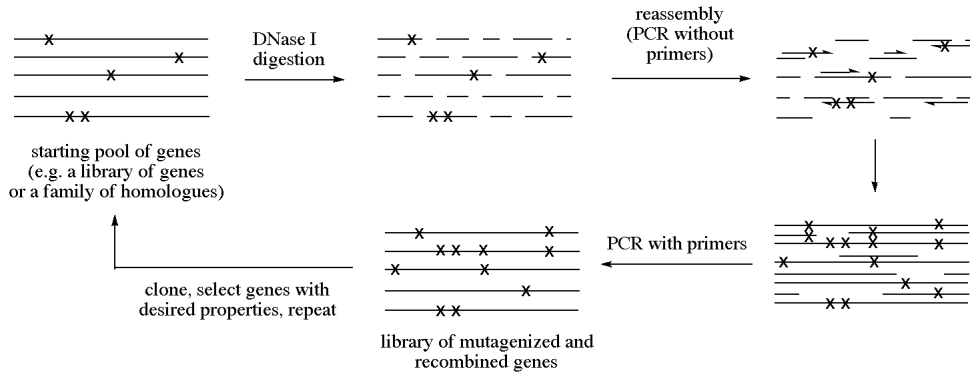


Figure 8. DNA shuffling involves random fragmentation of a gene into discrete fragments, reassembly of these fragments by PCR without primers, final PCR amplification with primers, and cloning of the mutated and recombined genes.[47]

antibodies, are distributed throughout the protein, making the rational design of the mutant proteins difficult, if not impossible.

Recently, the technique of DNA shuffling has been applied to the evolution of an "orthogonal" suppressor tRNA and aminoacyl-tRNA synthetase pair that may allow one to site-specifically incorporate into proteins in vivo amino acids with novel structural or electronic properties not specified in the genetic code.[55] The orthogonal suppressor tRNA was constructed by introducing eight mutations into $tRNA_2^{Gln}$ based on analyses of the X-ray crystal structure of the GlnRS–$tRNA_2^{Gln}$ complex (GlnRS = glutaminyl-tRNA aminoacyl synthetase) and previous biochemical data. The resulting tRNA satisfies the minimal requirements for the delivery of an unnatural amino acid: It is not acylated by any endogenous *E. coli* aminoacyl-tRNA synthetase, including GlnRS, and it functions efficiently in protein translation. Repeated rounds of DNA shuffling and oligonucleotide-directed mutagenesis were then used to generate a large library of directed and random mutants of GlnRS. This library was selected for mutant enzymes that efficiently acylate the engineered tRNA with glutamine in vivo, based on suppression of an amber codon (termination codon UAG) in the *lacZ* gene (Figure 9). With this approach an "orthogonal" mutant GlnRS/tRNA pair has been generated that functions in vivo. Currently, DNA shuffling and cassette mutagenesis are being used to select for synthetase mutants that acylate orthogonal tRNAs with novel amino acids.[56] Clearly, both structural information and diversity play a critical role in the design of these experiments.

A variety of other proteins with modified binding affinities and specificities, including growth hormones[57] and DNA-binding proteins,[58–60] have been isolated from large libraries of mutants. For example, the zinc finger DNA-binding motif of Zif268 protein has been expressed on filamentous phage, and libraries of mutants have been selected against duplex DNA sequences containing wild-type and mutant binding sites.[61] This approach has resulted in the generation of a DNA-binding peptide comprising three zinc fingers that binds to a unique nine base pair region of a BCR-ABL fusion oncogene with a dissociation constant K_d of 6.2×10^{-7} M. Its binding affinities for genomic BCR and C-ABL sequences (which differ by as little as one base pair) differ by an order of magnitude. Binding of the peptide to the target oncogene in transformed cells in vitro resulted in blockage of transcription. Techniques other than phage display can also be used to generate DNA-binding proteins with altered specificities. For example, helix-turn-helix and leucine zipper proteins with novel specificities have been generated from large libraries of mutants using selections based on interference with transcription of antibiotic resistance genes.[62, 63] Studies of this sort are also helping to define the recognition codes between the amino acid sequences of DNA-binding proteins and their cognate recognition sites.

Large libraries of mutant proteins have also been used to investigate the factors that influence the structure and stability of specific protein folds. For example, cassette mutagenesis has been used to randomly alter seven residues in the hydrophobic core of the N-terminal domain of phage λ repressor.[64] By selecting for functional repressors (phage-resistant transformants), it was shown that many different sequences can form a stable core. The main determinant of whether a particular sequence is compatible with the wild-type fold was hydrophobicity, although the van der Waals volume of the core and steric interactions between residues also limited the number of functional sequences. More recently, the sequence determinants that dictate the four helix bundle structure have been investigated.[65] A library of cytochrome b_{562} mutants was generated in which amino acids at defined positions were replaced with any of the hydrophobic amino acids Phe, Leu, Ile, Met, or Val, or any of the

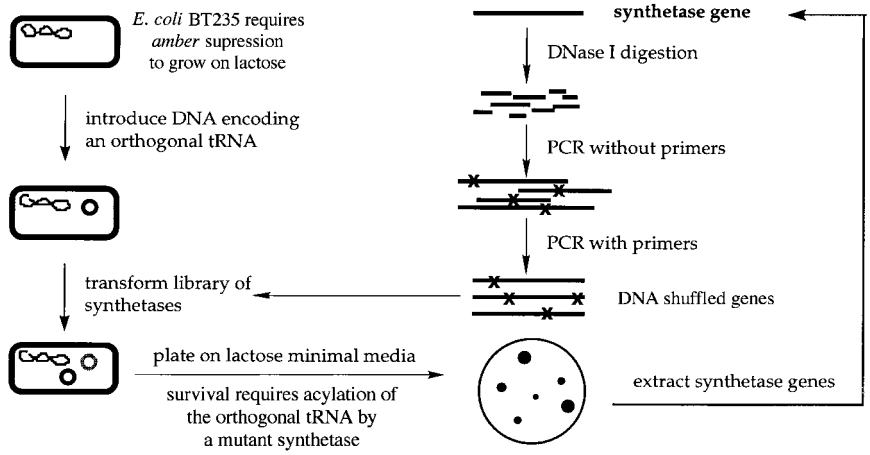


Figure 9. Selection of an orthogonal GlnRS/tRNA pair from a library of GlnRS mutants.[55]

hydrophilic amino acids Glu, Asp, Lys, Asn, Gln, or His to generate a binary pattern of nonpolar and polar amino acids. Bacterial expression of soluble, protease-resistant protein was assayed spectrophotometrically by monitoring heme absorbance. It was found that 29 of 48 sequences examined expressed protein that was both soluble and resistant to intracellular degradation. Two proteins that were characterized by urea denaturation experiments were stabilized by 3.7 and 4.4 kcal mol^{-1} relative to the unfolded form. A similar strategy has been applied to a random sequence of 80 to 100 amino acids consisting of Gln, Leu, and Arg. Bacterial expression of these "QLR proteins" revealed that 5% of the sequences were expressed at readily detectable levels. Several of the characterized proteins possessed features of native proteins such as α-helical content and highly cooperative folding, suggesting that folded proteins occur frequently in libraries of random amino acid sequences.[66]

The relationship between protein sequence and structure has also been examined using libraries created neither in vivo nor in vitro, but rather by computational methods. Recently, a design algorithm based on chemical potential functions and stereochemical constraints was used to search 1.9×10^{27} amino acid sequences for those that adopt a β-α-β zinc finger motif. The resulting designed sequence (FSF-1) possesses little or no identity to any known protein sequence. The synthesis and subsequent NMR spectroscopic investigation of FSF-1 in solution revealed a compact structure in excellent agreement with the design target. This experiment demonstrates the potential power of computational methods to screen combinatorial libraries ten or more orders of magnitude larger than any created in the laboratory.[67]

3. Nucleic Acid Diversity

Combinatorial methods have played an important role in investigating the scope of nucleic acid function and its possible role in the prebiotic world. In these experiments libraries containing up to 10^{15} random single-stranded RNA or DNA sequences are generated using a combination of chemical synthesis, PCR, and runoff transcription (in the case of RNA).[68–71] These libraries are then selected for individual molecules that either 1) bind specific ligands using affinity-based methods or 2) carry out specific chemical transformations using in vitro selections or screenings. Amplification of an enriched population of binders or catalysts is followed by subsequent rounds of selection. With these techniques, nucleic acids have been generated that selectively bind small molecules, proteins, and nucleic acids; and catalyze a number of reactions ranging from phosphoryl and acyl transfer reactions to metalation and pericyclic reactions.

3.1. Selecting for Binding Function

Single-stranded RNA and DNA oligomers (or aptamers) have been isolated from large, random oligonucleotide libraries that selectively bind a wide array of molecules including T4 DNA polymerase, human immunodeficiency virus (HIV), HIV Rev protein, thrombin, amino acids, organic dyes, and cofactors.[68, 69, 72–78] For example, by screening a library of 10^{13} transcripts with ATP immobilized on agarose, an RNA was isolated that binds adenosine with a dissociation constant K_d of 0.7 μM.[74] The solution structure of a 40 nucleotide RNA containing the consensus ATP-binding motif complexed with AMP (which binds with affinity similar to that of ATP) has been determined by NMR spectroscopy.[79] The AMP molecule is bound in a "GNRA-like" hairpin fold[80] with the intercalated adenine involved in an A·G mismatch pair and the 2′ and 3′ hydroxyl groups of the ribose sugar forming hydrogen bonds to the RNA. The binding of AMP was shown to induce a conformational transition in the ATP binding site on the RNA. RNA aptamers have also been isolated from oligonucleotide libraries that bind the purine theophylline with high affinity ($K_d = 0.1$ μM). These aptamers do not bind appreciably to caffeine ($K_d = 3500$ μM), which differs from theophylline by only a methyl group at nitrogen atom N7.[81] Experiments of this sort demonstrate that RNA, like proteins, can bind a wide array of small molecules with extremely high specificity, by a combination of van der Waals packing, hydrogen-bonding, and electrostatic interactions.

Aptamers have also been isolated that bind selectively to macromolecules, including proteins and nucleic acids. For example, single-stranded RNAs have been isolated that bind sequence-selectively to duplex DNA through the formation of triple-helical structures involving Hoogsteen base pairs.[82] Single-stranded DNAs have been isolated that bind to thrombin through the formation of a highly conserved guanine-rich structure of 14–17 nucleotides. In this case the binding affinity was 25–200 nM, and several aptamers inhibited thrombin-catalyzed fibrin clot formation in vitro at nanomolar concentrations.[77]

3.2. RNA Catalysis

The earliest use of in vitro selection schemes to isolate RNAs with novel phenotypes from pools of mutants were those of Spiegelman et al.[83] In these experiments Qβ replicase, a template-specific RNA-directed polymerase, was used in a Darwinian selection system to isolate a structured RNA that had reduced affinity for the intercalating dye ethidium bromide. The mutant arose by sequential mutation during the course of the experiments, and was not a preexisting variant of the original RNA pool.

The use of in vitro selections to isolate RNAs with novel catalytic activities has involved either libraries generated by mutagenesis of naturally occurring ribozymes, or synthetic libraries of randomized RNA sequences. For example, the *Tetrahymena* group I ribozyme has been used as a starting point to isolate a new ribozyme that cleaves DNA.[70, 71] The selection required the ribozyme to complete a phosphorylation reaction with an alternative DNA substrate that results in the addition of nine bases to the 3′-terminal hydroxyl group of the ribozyme (Figure 10). This 3′-terminal extension permits successful annealing of an amplification primer necessary to complete the cycle of selection. With a much larger pool of mutants (10^{13}), variants were isolated that catalyze the

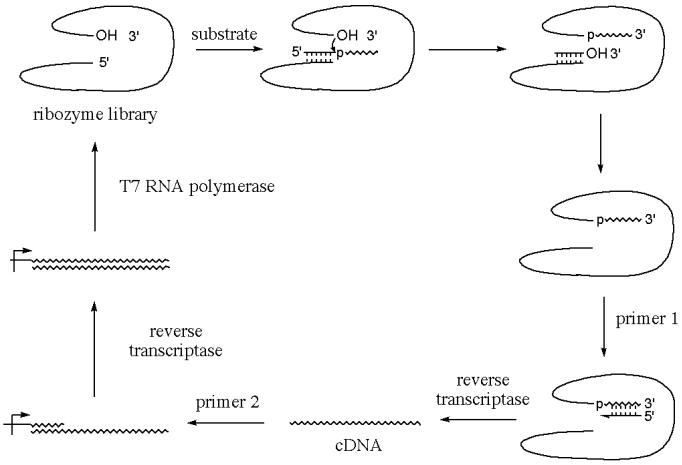


Figure 10. Selection of a *Tetrahymena* ribozyme mutant that uses an alternative DNA substrate.[70, 71] Wavy lines represent DNA strands.

reaction under physiological conditions at a rate up to 65 times faster than the wild-type ribozyme. Other selection schemes using naturally occurring ribozymes have yielded RNAs with self-copying functions[84] and altered metal requirements.[85]

One of the most impressive examples of the selection of RNA catalysts from complex libraries of random sequences was the isolation of a RNA ligase that can act as a RNA polymerase.[86, 87] Beginning with a pool of approximately 10^{15} distinct transcripts, iterative rounds of amplification and selection were carried out based upon the ability of the ribozyme to ligate its own 5′-terminal triphosphate group to a substrate oligonucleotide containing a 5′ "tag" (Figure 11). A ribozyme that catalyzes 3′,5′-phosphodiester bond formation with k_{cat} greater than 1 s^{-1} was isolated;[86, 87] when fused to a template strand, this RNA ligase was shown to catalyze RNA primer extension by up to six nucleotides in a template-directed fashion.[88] Nucleotides complementary to the template are added up to 1000 times more efficiently than mismatched oligonucleotides. Although a number of obstacles remain to generate an RNA replicase (for example, strand separation after polymerization, nonspecific binding of template and primer, and sequence fidelity) these experiments illustrate the potential to evolve RNAs that catalyze reactions which may have been required in the prebiotic world.

Selections and affinity-based methods have also been used to expand the scope of RNA-catalyzed reactions to other classes of reactions including isomerization, carbon–carbon bond forming, metalation, and alkylation reactions. For example, by exploiting the same notions of transition state theory that were used in the initial generation of catalytic antibodies, an RNA was isolated that catalyzes the isomerization of a bridged biphenyl substrate to its diastereomer (Figure 12).[89] In this case a library of 10^{13} random RNA molecules was screened based on binding affinity to the planar biphenyl transition state analogue **2** immobilized on a solid support. A similar approach has been used to generate single-stranded RNAs and DNAs that catalyze porphyrin metalation.[90, 91] More recently, selections based on covalent linkage of a ribozyme to a solid support have been used to generate RNAs that catalyze a Diels–Alder reaction between a diene and dienophile.[92]

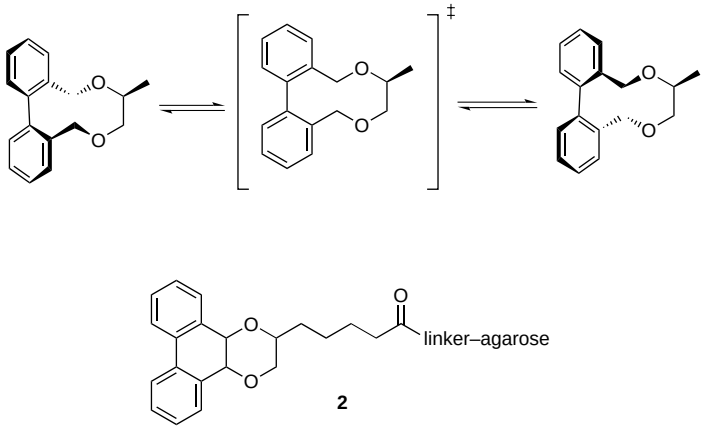


Figure 12. Atropisomerization between diastereoisomeric 1,1′-biphenyls. Structure **2** is an analogue of the transition state for this reaction **2**.[89]

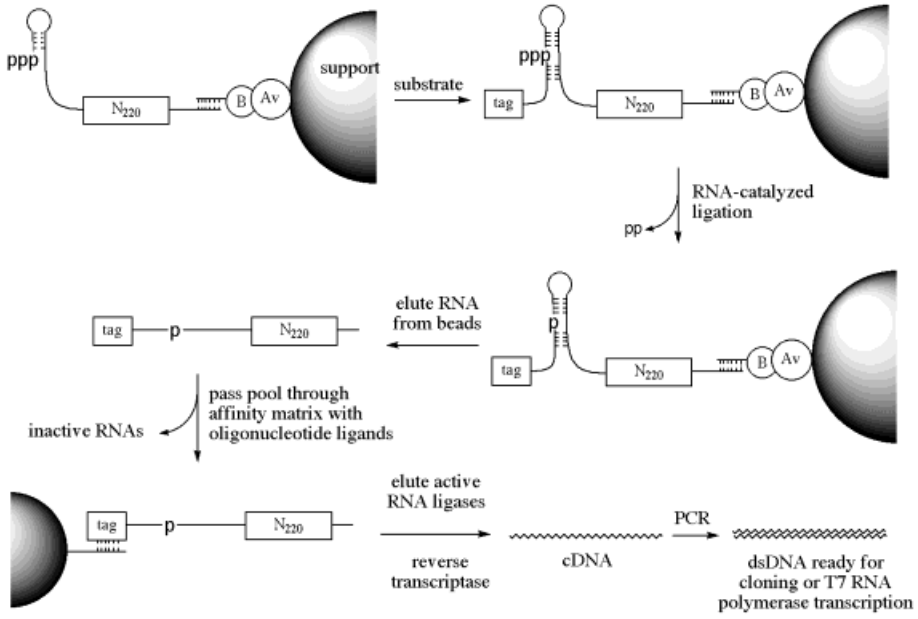


Figure 11. Selection scheme for generation of an RNA ligase.[86, 87] Biotin and avidin are represented by B and Av, respectively.

3.3. Other Applications

Many other novel uses of nucleic acid libraries continue to appear including in vivo applications. One such example involves the use of genomic expression libraries for immunization, a technique that makes use of genetic immunization (immunization with the gene encoding the protein antigen rather than the protein itself) and the fact that

all antigens of a pathogen are encoded in its DNA.[93] Indeed, it has been shown that even partial expression libraries made from the DNA of *Mycoplasma pulmonis*, a natural pathogen in rodents, provide protection against challenges from the pathogen. Libraries of nucleic acids have also been used in vivo to identify cis-acting sequences that lead to localization of RNAs in the nucleus of *Xenopus laevis* oocytes.[94] Clearly, libraries of nucleic acids, like those of peptides and proteins, will find many applications in the search for new biological and chemical functions, and in our efforts to better understand complex biological processes.

4. Diversity in Chemical Synthesis

The biosynthetic machinery used by living cells to generate diverse populations of molecules is limited to specific classes of chemical structures including oligomeric molecules such as polypeptides, oligonucleotides, and polysaccharides, and a variety of natural products such as the polyketides. Moreover, these structures are generated from restricted sets of building blocks. The development of methods for generating large, diverse populations of synthetic molecules has extended the scope of this approach to new classes of molecules with a broader range of chemical, biological, and physical properties.

4.1. Approaches to Synthetic Diversity

Virtually all of the efforts aimed at generating and screening large populations of synthetic diversity are based on the method of solid-phase synthesis developed by Merrifield.[95] The initial work focused on polypeptides, which can be rapidly assembled by iterative chemical synthesis from a diverse set of commonly occurring and unnatural amino acids. Methods were first developed that simply involved the simultaneous synthesis of multiple discrete peptide sequences using either polyacrylic acid grafted polyethylene pins arranged in microtiter format,[96] or later resins sealed in porous polypropylene bags (the "tea bag" technique).[97]

These experiments were followed by efforts aimed at generating more diverse populations of molecules by connecting sets of building blocks in all possible combinations, again mimicking the combinatorial strategies used by nature to efficiently generate molecular diversity. These methods involve either the generation of spatially separated libraries of molecules, or mixtures of resins each containing a discrete structure or small number of related structures. One example of the former approach involves the "mimotope" strategy in which peptides are again synthesized on pins, but in this case by randomly incorporating residues from mixtures of activated amino acids to give all peptide sequences.[98, 99] Two positions in a peptide containing n residues are then iteratively defined by multipin synthesis of 400 discrete mixtures, each containing all possible $(n-2)$ peptide sequences with one of the 400 possible pairs of the common L-amino acids defined. This technique has been used in conjunction with ELISA (enzyme-linked immunosorbent assay) screening methods to identify peptides that selectively bind antibodies and other receptors.

A second method for the spatially addressable synthesis of combinatorial libraries makes use of photolithography together with solid-phase synthesis to generate high-density, combinatorial arrays of peptides on a glass substrate.[100, 101] By illuminating specific regions of the glass substrate containing growing peptide chains, in which the terminal amino group is protected with a photolabile group, the spatial distribution of individual amino acid coupling steps can be controlled (Figure 13). With n binary masks in which half of the

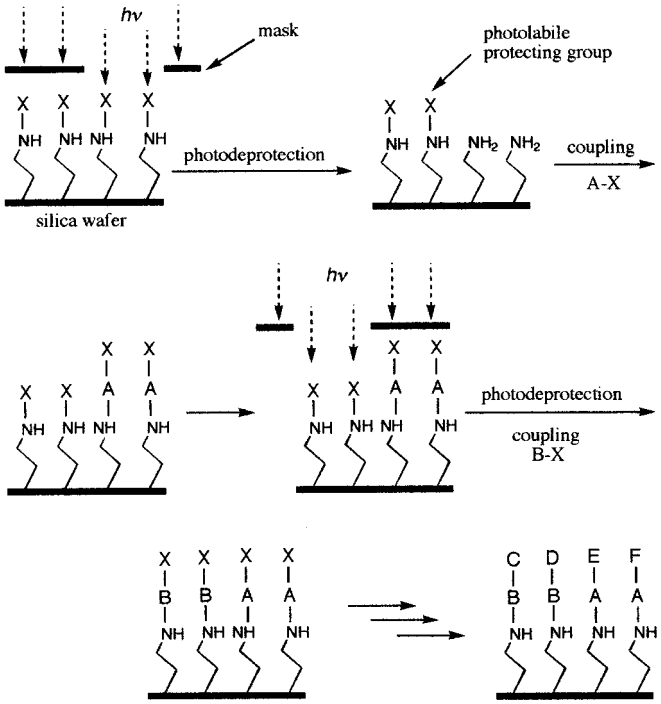


Figure 13. Light-directed synthesis of combinatorial arrays.

surface is photolyzed during each coupling step, 2^n compounds can be synthesized. The sequence of masks and coupling steps defines the identity of the peptide at each site. Binding of fluorescently labeled receptors to the surface is detected by fluorescence emission. Although this technique was developed for peptide synthesis, it has proven to be most useful for generating high-density arrays of oligonucleotides (see below).

Another approach that has been used to generate combinatorial libraries of peptides is the split synthesis method.[102–104] This technique involves dividing the resin support into n equal fractions, coupling each fraction with a single activated monomer (or in some cases, a small number of monomers), and then recombining the fractions (Figure 14). Iteration for x cycles leads to a stochastic population of n^x peptides. This approach has been used in conjunction with the "tea bag" technique and beaded resins equipped with non-cleavable linkers to generate large combinatorial libraries of up to 50 million different sequences. However, since mixtures of resins are generated, this technique also requires methods to identify the sequence of a high-affinity ligand isolated from an affinity-based assay.

A number of strategies have been developed to determine the identity of a peptide ligand isolated from pools of random

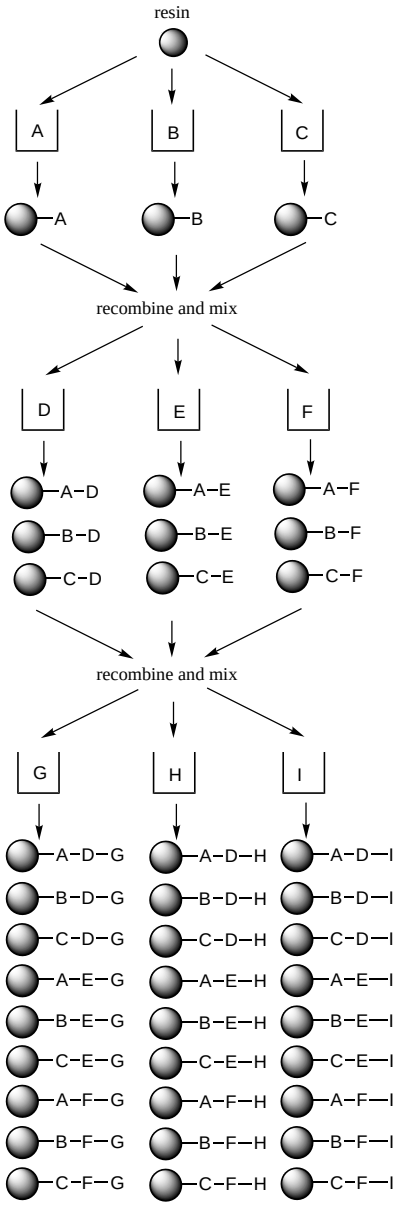


Figure 14. Preparation of a combinatorial library by the split synthesis method.

sequences. These include methods based on the iterative "mimotope" deconvolution strategy described above,[105] as well as direct Edman peptide microsequencing, which can be used with beads containing 80 – 100 pmol of peptide.[106] Mass spectrometry can also be used to identify the sequence of a peptide, if in each step of the synthesis a small fraction of the growing polypeptide chain is capped, resulting in a small amount of truncated chain at each step in the synthesis of the full length peptide.[107]

More recently, a number of methods have been developed for generating libraries of peptides on resin beads in which each bead is attached to a molecular "tag" that allows rapid and sensitive identification of the associated polypeptide.[108–111] The tags are attached to the bead coincident with each monomer coupling step. Consequently, they must be stable under the conditions of polymer synthesis and deprotection, their attachment must not interfere with the synthesis of the polypeptide, and they must have high information content that can be detected with high sensitivity. A number of tagging techniques have been developed (Figure 15), including the use of oligonucleotide tags that can be detected by PCR amplification and DNA sequencing,[109] and a set of related halocarbon tags that can be separated and detected be electron capture capillary gas chromatography.[111] The high detection sensitivity of both these tags should facilitate the synthesis of increasingly diverse chemical libraries.

The techniques described above have been used to identify natural and unnatural polypeptides that bind a host of molecules including antibodies, enzymes, and receptors.[24] For example, antimicrobial peptides with activities against both gram-positive and gram-negative bacteria have been generated.[112] The binding specificities of synthetic steroid- and macrocycle-based hosts for peptide ligands have also been determined.[113–115] In addition, these techniques have been used to analyze protein – protein interactions involved in cellular processes.[116] In one such study, two different SH3 domains conjugated to fluorescein were used to screen a peptide library for ligands that possessed affinity for SH3 domains. Analysis of the resulting sequences suggested a set

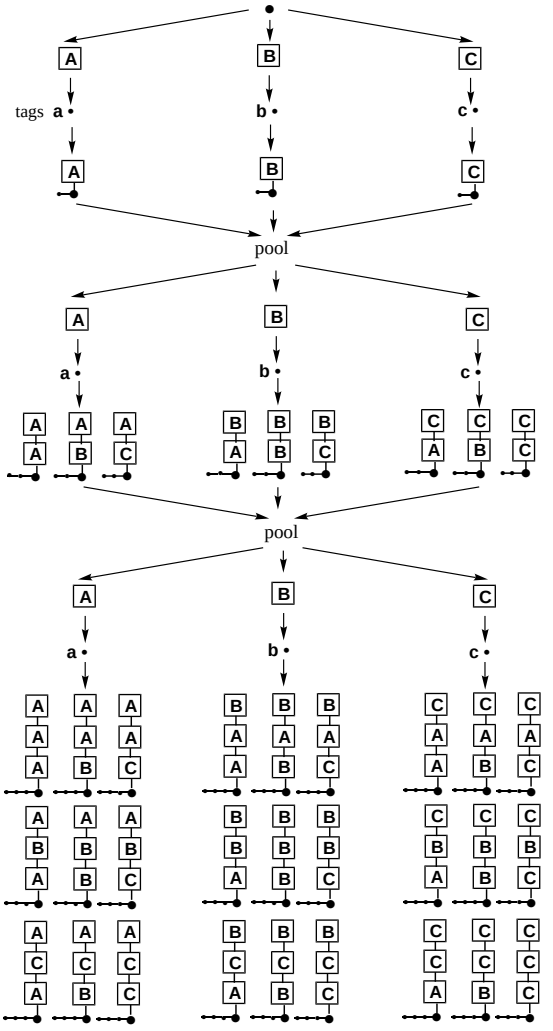


Figure 15. The use of molecular tags to identify members of a library.[108–111]

of rules governing SH3 – ligand interactions, and subsequent structural studies revealed a detailed mode of ligand binding by the SH3 domain.[116]

4.2. Extension to Other Polymeric and Nonpolymeric Frameworks

Molecular diversity has also been used to explore the chemical and biological properties of heteropolymers composed of building blocks other than amino acids. These unnatural polymeric frameworks may have novel properties such as improved pharmacokinetic properties (including membrane permeability and biological stability) that increase bioavailability, or altered conformational or hydrogen-bonding properties that provide insights into biomolecular structure and folding. A number of novel "unnatural biopolymers" have been generated including oligocarbamates,[117] peptoids,[118, 119] oligoureas,[120–122] oligopyrrolinones,[123] oligoazatides,[124] oligosulfonamides,[125] and β-peptides.[126–128] Because these oligomeric compounds can be assembled from a set of diverse building blocks, large molecular libraries can be rapidly generated and screened for desired properties. For example, oligocarbamates consisting of a chiral ethylene backbone with a variety of side chains linked through relatively rigid carbamate groups have been efficiently synthesized by the stepwise coupling of N-protected amino-*p*-nitrophenyl carbonate monomers. The monomers in turn are readily derived from the corresponding amino acids (Figure 16).[117] Libraries of these molecules have been generated, and high-affinity selective ligands have been identified that bind antibodies and integrin receptors.[111, 129] Libraries of polypeptoids,[118, 119] a polypeptide backbone composed of N-substituted glycine units, have also been generated. Peptoids have been identified that selectively bind the α-adrenergic receptor and a number of enzymes.[118] Efforts are ongoing to determine the conformational and pharmacological properties of these molecules.

The application of combinatorial approaches to creating nonpolymeric small molecules came with the realization that many such molecules can also be synthesized by the iterative stepwise addition of building blocks on a solid support. The most notable early example was that of the combinatorial synthesis of a library of 1,4-benzodiazepines,[130, 131] a pharmacophore common to a wide variety of clinically important therapeutic agents. A library of approximately 2000 structurally diverse 1,4-benzodiazepine derivatives containing a variety of chemical functionalities including amides, carboxylic acids, amines, phenols, and indoles was constructed by multipin synthesis from three components: 2-aminobenzophenones, amino acids, and alkylating agents (Figure 17). Ligands that selectively bind p60src, a cholecytokinin A receptor, and DNA-binding antibodies have been isolated from this library.

Combinatorial libraries for a large number of small organic molecules have since been generated including β-turn mimetics,[132] hydantoins,[133] protease inhibitors,[134] pyrrolidines,[135] β-lactams,[136] and tyrosine kinase inhibitors.[137] In addition, a library of 1300 di- and trisaccharides was recently synthesized on beads and screened against *Bauhinia purpurea* lectin conjugated to a reporter enzyme.[138] Two ligands were isolated that bound to the lectin more tightly than the natural ligand. Similar approaches provide promising routes to the identification of carbohydrate-based ligands for other receptors.

Finally, considerable effort is also being focused on the development of new solid-phase synthetic methodologies for generating small-molecule libraries,[139, 140] as well as theoretical and empirical methods for library design. For example, the structure-based design of nonpeptide libraries containing the hydroxyethylamine pharmacophore of pepstatin yielded potent inhibitors ($K_i = 9$–15 nM) of cathepsin D.[141] Clearly, the use of libraries of synthetic oligomeric and nonpolymeric molecules will continue to have a significant impact on the pharmaceutical industry.

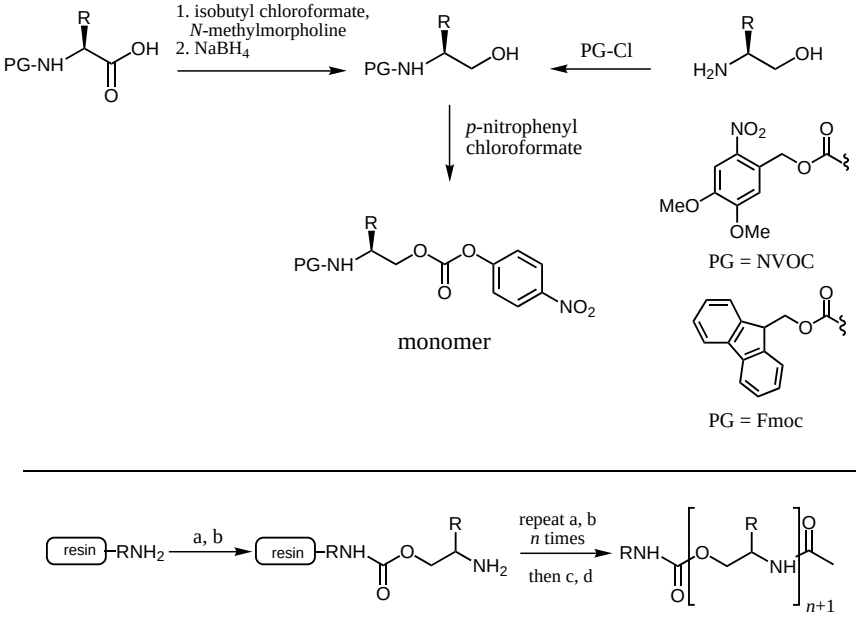


Figure 16. Synthesis of oligocarbamates from amino acid starting materials.[117] a) Monomer, 1-hydroxy-1*H*-benzotriazole (HOBt), diisopropylethylamine, *N*-methylpyrrolidine (NMP); b) piperidine, NMP; c) acetic anhydride, pyridine; d) trifluoroacetic acid (TFA), CH_2Cl_2. PG = protecting group, NVOC = nitroveratrylmethoxycarbonyl, Fmoc = 9-fluorenylmethoxycarbonyl.

4.3. Natural Products

Not only can chemical synthesis be used to generate libraries of small organic molecules, recently it has been shown that the biosynthetic machinery of cells can be exploited to produce libraries of natural products such as the polyketides, which include the well-known antibiotics erythromycin, avermectin, and spiramycin.[142] These compounds are synthesized by modular polyketide synthases (PKSs), large multifunctional enzyme assemblies that

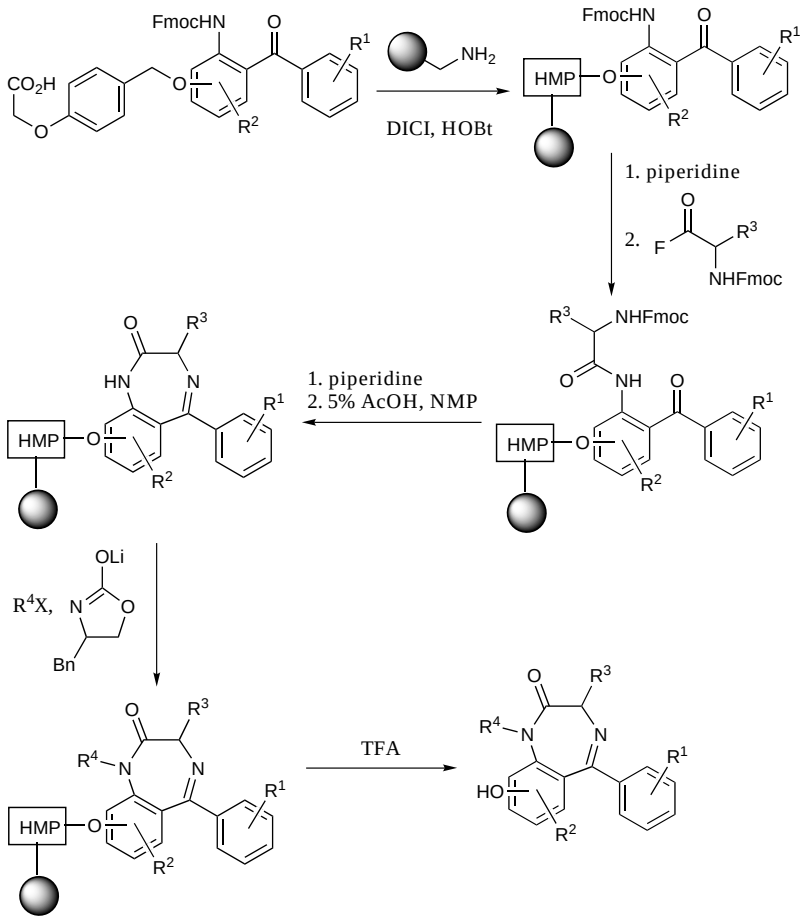


Figure 17. Synthesis of a library of 1,4-benzodiazepines on solid support.[130] Bn = benzyl, DICI = diisopropylcarbodiimide, HMP = 4-(hydroxymethyl)phenoxyacetic acid.

catalyze the stepwise biosynthesis of these natural products.[143, 144] These modules encode sets of enzymes (ketosynthases, acyl transferases, ketoreductases, dehydratases, and enoyl reductases) that processively carry out rounds of polyketide chain elongation with concomitant alterations in oxidation level and stereochemistry after each elongation step. Deleting individual modules, altering the specificity of individual enzymes in a module, or adding new enzymatic activities results in the synthesis of a large number of different macrolides.[142] Most recently, it has been shown that exogenous addition of synthetic diketide mimics to deoxyerythronolide B synthase containing a genetic block in the KS-1 gene results in the synthesis of a novel polyketide containing the synthetic building block (Figure 18).[145] As more nontemplated biosynthetic pathways are isolated and characterized (e.g., cyclic peptides, ene-diynes, and bleomycins) and our ability to manipulate their function increases, this approach is likely to provide an increasingly diverse collection of chemical structures to screen for novel biological properties.

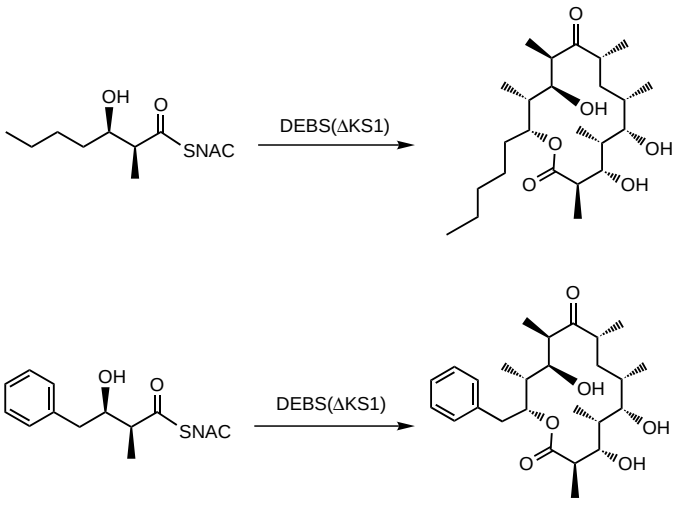


Figure 18. Precursor-directed biosynthesis of novel polyketides containing synthetic building blocks.[145] NAC = *N*-acetylcysteamine, DEBS = deoxyerythronolide B synthase.

4.4. Genomics

Combinatorial synthetic methods have also found important applications in genomics. For example, the photolithographic chemical synthesis methods developed for peptides have made it possible to synthesize spatially addressable, high-density arrays of oligonucleotides of known sequence.[146, 147] This technique has been used to generate more than 260 000 specifically chosen oligonucleotide probes (in an area of about 6.5 cm^2) that hybridize to all open reading frames (ORFs) in the yeast genome (twenty 25-mers complementary to segments within each ORF; Figure 19).[148] With this array one can monitor the expression levels of nearly all yeast genes in a quantitative, sensitive, and highly reproducible fashion due to the redundancy in the detection and analysis of the data. Arrays of this sort are likely to have a major impact on our ability to analyze the information content of genomic DNA, including mismatch scanning, homology comparisons, and the detection of genetic differences between strains. Other parallel, high-throughput methods have also been developed for generating oligonucleotide arrays,[149–156] including spotting methods with presynthesized oligonucleotides.[157]

A recent example demonstrates how combinatorial libraries of small molecules, structural information, and oligonucleotide arrays can be used synergistically to identify and characterize selective kinase inhibitors.[158] A combinatorial library of 2,6,9-trisubstituted purines was generated based on structural analysis of the novel binding mode of the purine olomoucine to the ATP binding pocket of the cell cycle kinase CDK2.[159, 160] By iterating chemical library synthesis and biological screening, high-affinity, selective inhibitors of human CDK2/cyclin A and yeast cdc28 kinases were identified. The CDK2/cyclin A kinase inhibitors include compounds that selectively inhibit the growth of colon

Figure 19. Fluorescence image of an array of oligonucleotide probes following hybridization of a fluorescently labeled RNA sample prepared from mRNA extracted from yeast cells.

cancer cell lines. The structural basis for the affinity and selectivity of these purine derivatives was determined by analysis of the three-dimensional crystal structure of their complex with CDK2, providing a basis for further enhancing the affinity and selectivity of the compounds. The cellular effects of this class of compounds were further examined and compared to those of the CDK2 inhibitor flavopiridol, as well as to those arising from mutations in cdc28 itself, by monitoring changes in mRNA expressions levels for all genes in treated cells of *Saccharomyces cerevisiae* using high-density oligonucleotide probe arrays.[148, 161] In particular, of the 105 transcripts that changed greater than threefold in response to both compounds (335 for the purine and 267 for flavopiridol) only seven were down regulated, all of which were associated with progression of the cell cycle. These experiments also began to delineate other common effects of these compounds on cellular metabolism, stress response, signaling, and growth, as well as reveal significant differences in the effects of flavopiridol and the purine, despite their apparent in vitro selectivity for CDKs. Information of this sort may, in general, be useful in comparing compounds prior to clinical studies or in identifying targets whose inhibition might potentiate the effects of a primary drug. Moreover, the use of gene expression profiles together with multiple copy overexpression libraries[162] may allow one to rapidly identify the targets of compounds with interesting biological activities isolated from phenotypic screening of whole cells. The combined use of chemical libraries, novel in vitro and cellular screens,[163–165] and genomics is likely to provide important new insights into complex cellular processes.

5. Materials Science

The most recent application of combinatorial approaches has been to solid-state and materials science. The properties of many functional materials, such as high-temperature superconductors, heterogeneous catalysts, ferroelectric materials, magnets, and even structural materials such as alloys arise from complex interactions involving the host structure, dopants, defects, and interfaces, all of which are highly dependent on composition and processing. Unfortunately, the current level of theoretical and empirical understanding does not generally allow one to predict the structures and resulting properties of these materials.[166] The situation is further complicated by the complex compositions of many modern materials (four or more elements) and the fact that materials synthesis, unlike organic synthesis, is generally not a kinetically controlled process. Given approximately 60 elements in the periodic table that can be used to make compositions consisting of three, four, five, or even six elements, the universe of possible new compounds with interesting physical and chemical properties remains largely uncharted. Combinatorial methods should make it possible to rapidly synthesize, process, and analyze large libraries (hundreds to hundreds of thousands of members) of inorganic and organic materials and devices. This may lead to significant increases both in the efficiency of materials discovery and optimization, and in our information base relating materials structure and properties.

5.1. Synthesis and Screening

The first application of combinatorial methods to materials science involved the synthesis of libraries of thin-film copper oxides containing high-temperature superconductors.[167] Libraries designed to explore many different materials compositions were synthesized by sequentially sputtering precursors at different sites on a substrate using a series of precisely positioned shadow masks (Figure 20). Low-temperature annealing followed by high-temperature processing resulted in formation of superconducting thin films. More recently, quaternary and shutter masking systems, together with photolithographic masking techniques and pulsed laser deposition, have made possible the synthesis of high-quality, diverse thin-film libraries of some 1000 to 10000 samples on an area of about 6.5 cm^2.[168–170] Solution-based methods have also been applied to materials library synthesis. For example, scanning fluid delivery stems (inkjets) have been used to rapidly and accurately deliver nanoliter volumes of precursor solutions to generate libraries of metal oxides and organic polymers.[171, 171] Processing conditions play an important role in determining materials properties, and it is important to include these variables in the design of a materials library. Identical libraries synthesized simultaneously can be processed under different conditions to enhance diversity. Gradient temperature ovens have been used to identify optimal processing conditions for a library of metal oxides; similarly, it may be possible to vary oxygen partial pressure by introducing oxygen on one side of the chip.[173]

A number of detection systems have been developed to date for screening materials libraries for optical, electronic, magnetic, or chemical properties of interest.[173] Optical imaging systems have been used to evaluate libraries of photoluminescent materials. For example, a scanning spectrophotometer and CCD array detectors (CCD = charge-coupled device) were used to evaluate the photon output and chromaticity of each member in phosphor libraries upon excitation with monochromatic UV light.[168, 169] A novel

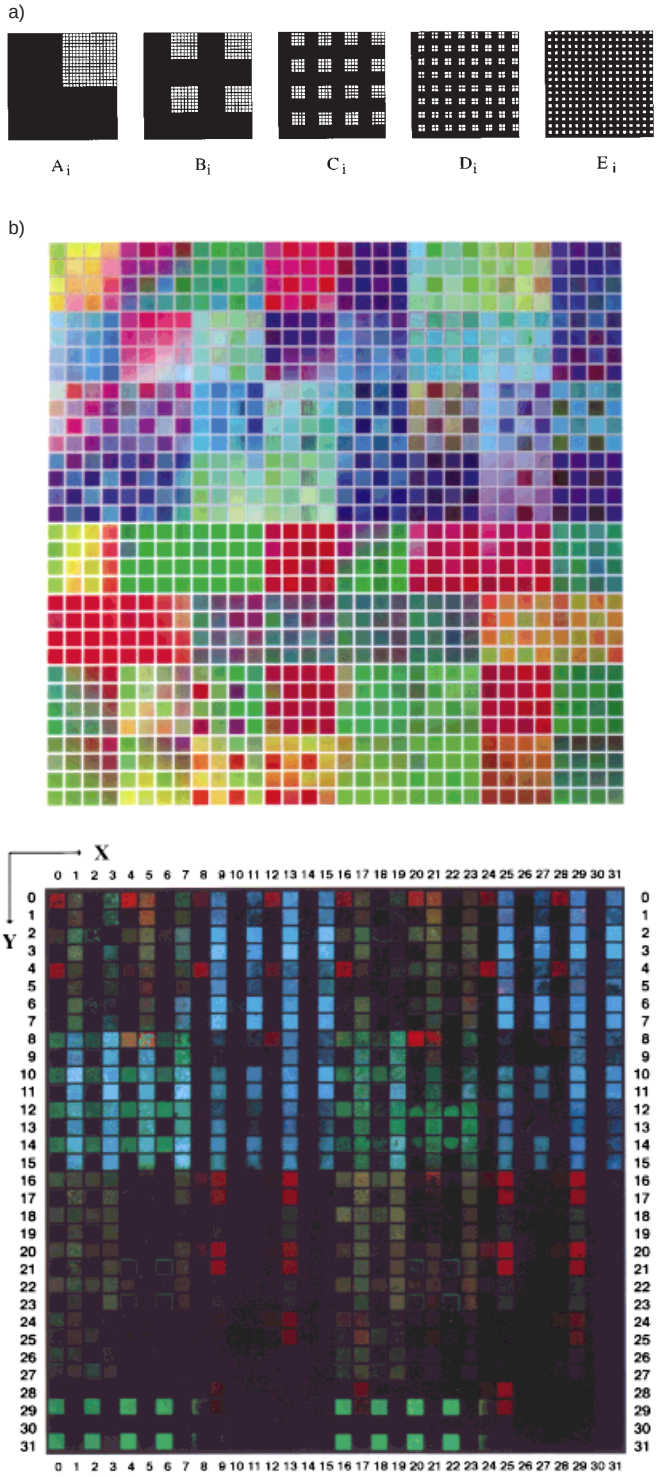


Figure 20. a) In the quaternary masking scheme, deposition is carried out using a series of n different masks which subdivide the substrate into a series of nested quadrants. Each mask is used in four sequential deposition steps with a 90° rotation of the mask in each step, generating up to 4^n compositions in $4n$ deposition steps. b) Photograph of the 1024-membered library (deposited on a Si substrate with an area of $2.54 \times 2.54\,cm^2$) under ambient light (top) and UV irradiation (bottom).

scanning-tip microwave near-field microscope (STMNM) has been developed to nondestructively measure dielectric constant and tangent loss of a library of ferroelectric materials with submicron spatial resolution and high sensitivity (Figure 21).[170, 174] More recently, X-ray microbeam techniques with a spot size of $3 \times 20\ \mu m^2$ have been used to characterize the composition and structure of samples in thin-film libraries.[175] Infrared imaging thermography has been used to qualitatively evaluate catalyst performance in a small member library of hydrogen oxidation catalysts,[176] and mass spec-

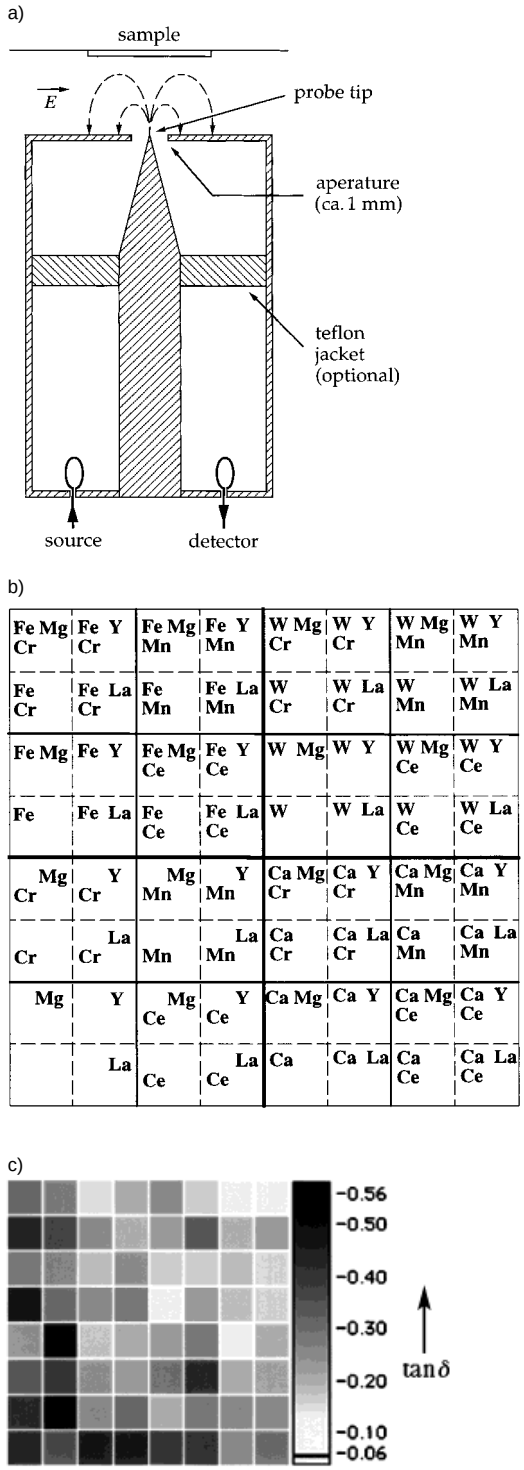


Figure 21. a) Schematic representation of a scanning tip microwave near-field microscope for analyzing ferroelectric and dielectric libraries. b) Composition of a region of a $BaTiO_3$ ferroelectric library with dopands indicated. c) Image of tangent loss ($\tan\delta$) from this library obtained with the near-field scanning microwave microscope.[174]

trometry has been used to rapidly screen a library of oxidation catalysts with high sensitivity.[177] Magneto-optical detectors are being developed to image libraries of magnetic materials, and many other detector systems can be envisaged to measure a range of material properties (IR and Raman spectroscopy, surface plasma resonance, nanoindentation, light scattering, and polarized light microscopy).

5.2. Applications

Although there remains a great deal of work to be done in developing methods for the combinatorial synthesis, processing, and detection of materials libraries, a number of applications of this approach has already been reported. Combinatorial approaches have been used to identify a class of cobalt oxide magnetoresistive materials of the form $(La_{0.88}S_{0.12})CoO_3$.[178] Magnetoresistance was found to increase as the size of the dopant ion increased, in contrast to Mn-containing compounds, in which the magnetoresistive effect increases as the size of the alkaline earth ion decreases. Combinatorial methods have also been applied to the optimization and identification of luminescent materials. Recently, two novel blue phosphorescent compounds, $SrCeO_4$[179] and $Gd_3Ga_5O_{12}/SiO_x$,[168] were discovered in combinatorial libraries. Similar approaches have also been applied to ferroelectric materials where the effects of transition metal dopants on the dielectric constant and tangent loss of a library of thin films of $(Ba_xSr_{1-x})TiO_3$ were determined.[170] Combinatorial methods are likely to have a significant impact on catalysis and polymer chemistry, and the first attempts to apply combinatorial methods in this area have already surfaced.[173–176, 180–182] It may also be possible to generate combinatorial libraries of entire devices such as capacitors or electroluminescent displays.[183]

6. Conclusion

The development of strategies for generating large diverse libraries of biomolecules, small organic molecules, and solid-state materials, together with novel screens and selections for specific biological, chemical, or physical properties, is having an enormous impact on science. These approaches are significantly increasing the rate and efficiency with which the scientific method can be used both to identify molecules with novel properties and to understand the structural basis for their function. Clearly, many challenges lie ahead including 1) improved theoretical and computational methods for designing libraries and analyzing the output of library experiments; 2) the development of methods for increasing library size (especially for polypeptides, proteins, and synthetic molecules); 3) high-throughput screens for a broader range of biological (cellular), chemical, and physical properties; 4) extension of biological selections and amplification to synthetic diversity; and 5) further engineering improvements in miniaturization and robotics. Nonetheless, molecular diversity has already been exploited to produce a large number of molecules with functions that would have been difficult and, in many cases, impossible to realize using more traditional synthetic approaches.

A portion of this work was supported by the Department of Energy, the National Institutes of Health, and the Office of Naval Research. D.R.L. would like to acknowledge a predoctoral fellowship from the Howard Hughes Medical Institute. P.G.S. is a Howard Hughes Medical Institute Investigator and W. M. Keck Foundation Investigator.

Received: February 18, 1998
Revised version: July 15, 1998 [A274IE]
German version: *Angew. Chem.* **1999**, *111*, 36 – 56

[1] F. M. Burnet, *The Clonal Selection Theory of Acquired Immunity*, Vanderbilt University Press, Nashville, TN, **1959**, p. 53.
[2] D. W. Talmage, *Science* **1959**, *129*, 1649.
[3] S. Tonegawa, *Nature* **1983**, *302*, 57.
[4] D. R. Davies, S. Chacto, *Acc. Chem. Res.* **1993**, *26*, 421.
[5] S. J. Pollack, J. W. Jacobs, P. G. Schultz, *Science* **1986**, *234*, 1570.
[6] A. Tramontano, K. D. Janda, R. A. Lerner, *Science* **1986**, *234*, 1566.
[7] P. G. Schultz, R. A. Lerner, *Science* **1995**, *269*, 1835.
[8] A. G. Cochran, P. G. Schultz, *Science* **1990**, *249*, 781.
[9] M. E. Blackwood, Jr., T. S. Rush III, F. Romesberg, P. G. Schultz, T. G. Spiro, *Biochemistry* **1998**, *37*, 779.
[10] F. E. Romesberg, B. D. Santarsiero, B. Spiller, J. Yin, D. Barnes, P. G. Schultz, R. C. Stevens, *Biochemistry* **1998**, *37*, 14404.
[11] J. R. Jacobsen, J. R. Prudent, L. Kockersperger, S. Yonkovich, P. G. Schultz, *Science* **1992**, *256*, 365.
[12] E. Driggers, P. G. Schultz, unpublished results.
[13] P. G. Schultz, R. A. Lerner, *Acc. Chem. Res.* **1993**, *26*, 391 – 395.
[14] C. F. Barbas III, A. Heine, G. Zhong, T. Hoffmann, S. Gramatikova, R. Bjornestedt, B. List, J. Anderson, E. A. Stura, I. A. Wilson, *Science* **1997**, *278*, 2085 – 2092.
[15] G. F. Zhong, T. Hoffmann, R. A. Lerner, S. Danishefsky, C. F. Barbas III, *J. Am. Chem. Soc.* **1997**, *119*, 8131 – 8132.
[16] H. D. Ulrich, E. Mundorff, B. D. Santarsiero, E. M. Driggers, R. C. Stevens, P. G. Schultz, *Nature* **1997**, *389*, 271 – 275.
[17] F. E. Romesberg, B. Spiller, P. G. Schultz, R. C. Stevens, *Science* **1998**, *279*, 1929 – 1933.
[18] A. Heine, E. A. Stura, J. T. Yli-Kauhaluoma, C. Gao, Q. Deng, B. R. Beno, K. N. Houk, K. D. Janda, I. A. Wilson, *Science* **1998**, *279*, 1934.
[19] G. J. Wedemayer, P. A. Patten, L. H. Wang, P. G. Schultz, R. C. Stevens, *Science* **1997**, *276*, 1613 – 1756.
[20] P. A. Patten, N. S. Gray, P. L. Yang, C. B. Marks, G. J. Wedemayer, J. J. Boniface, R. C. Stevens, P. G. Schultz, *Science* **1996**, *271*, 1086.
[21] G. J. Wedemayer, L. H. Wang, P. A. Patten, P. G. Schultz, R. C. Stevens, *J. Mol. Biol.* **1997**, *268*, 390.
[22] W. D. Huse, L. Sastry, S. A. Iverson, A. S. Kang, M. Alting-Mees, D. R. Burton, S. J. Benkovic, R. A. Lerner, *Science* **1989**, *246*, 1275 – 1281.
[23] S. Cwirla, E. A. Peters, R. W. Barrett, W. J. Dower, *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 6378 – 6382.
[24] M. A. Gallop, R. W. Barrett, W. J. Dower, S. P. A. Fodor, E. M. Gordon, *J. Med. Chem.* **1994**, *37*, 1233 – 1251.
[25] J. K. Scott, C. P. Smith, *Science* **1990**, *249*, 386 – 390.
[26] J. J. Delvin, L. C. Panganiban, P. E. Delvin, *Science* **1990**, *249*, 404 – 406.
[27] M. G. Cull, J. F. Miller, P. J. Schatz, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 1865 – 1869.
[28] L. C. Mattheakis, R. Bhatt, W. J. Dower, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 9022.
[29] R. W. Roberts, J. W. Szostak, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 12297 – 12302.
[30] A. S. Kang, C. F. Barbas, K. D. Janda, S. J. Benkovic, R. A. Lerner, *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 4363 – 4366.
[31] J. McCafferty, A. D. Griffiths, G. Winter, D. J. Chiswell, *Nature* **1990**, *348*, 552 – 554.

[32] K. D. Janda, C. H. Lo, T. Li, C. F. Barbas III, P. Wirsching, R. A. Lerner, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 2532.
[33] K. D. Janda, L.-C. Lo, C.-H. Lo, M.-M. Sim, R. Wang, C.-H. Wong, R. A. Lerner, *Science* **1997**, *275*, 945 – 948.
[34] J. Ku, P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 6552.
[35] H. Pedersen, S. Holder, D. S. Sutherlin, U. Schwitter, D. S. King, P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 10523.
[36] N. C. Wrighton, F. X. Farrell, R. Chang, A. K. Kashyap, F. P. Barbone, L. S. Mulcahy, D. L. Johnson, R. W. Barrett, L. K. Jolliffe, W. J. Dower, *Science* **1996**, *273*, 458 – 463.
[37] O. Livnah, E. A. Stura, D. L. Johnson, S. A. Middleton, L. S. Mulcahy, N. C. Wrighton, W. J. Dower, L. K. Jolliffe, I. A. Wilson, *Science* **1996**, *273*, 464.
[38] D. J. Matthews, J. A. Wells, *Science* **1993**, *260*, 1113 – 1117.
[39] M. A. Barry, W. J. Dower, S. A. Johnston, *Nat. Med.* **1996**, *2*, 299 – 305.
[40] M. Barinaga, *Science* **1998**, *279*, 323 – 324.
[41] B. G. Hall, *Biochemistry* **1981**, *20*, 4042 – 4049.
[42] M. S. Neuberger, B. S. Hartley, *J. Gen. Microbiol.* **1981**, *122*, 181 – 191.
[43] A. Paterson, P. H. Clarke, *J. Gen. Microbiol.* **1979**, *114*, 75 – 85.
[44] J. Reidhaar-Olson, J. Bowie, R. M. Breyer, J. C. Hu, K. L. Knight, W. A. Lim, M. C. Mossing, D. A. Parsell, K. R. Shoemaker, R. T. Sauer, *Methods Enzymol.* **1991**, *208*, 564 – 586.
[45] J. A. Wells, M. Vasser, D. B. Powers, *Gene* **1985**, *34*, 315.
[46] R. C. Caldwell, G. F. Joyce, *PCR Methods Appl.* **1992**, *2*, 28 – 33.
[47] W. P. C. Stemmer, *Nature* **1994**, *370*, 389 – 391.
[48] P. A. Patten, R. J. Howard, W. P. C. Stemmer, *Curr. Opin. Biol.* **1997**, *8*, 724 – 733.
[49] C. Ho, M. Jasin, P. Schimmel, *Science* **1985**, *229*, 389.
[50] A. R. Oliphant, K. Struhl, *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 9094.
[51] H. Liao, T. McKensie, R. Hageman, *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 576.
[52] J. F. Reidhaar-Olson, R. T. Sauer, *Science* **1988**, *241*, 53.
[53] J. C. Moore, F. H. Arnold, *Nat. Biotechnol.* **1996**, *14*, 458.
[54] K. W. Munir, D. C. French, L. A. Loeb, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 4012.
[55] D. R. Liu, T. J. Magliery, M. Pastrnak, P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 10092 – 10097.
[56] D. R. Liu, M. Pastrnak, P. G. Schultz, unpublished results.
[57] S. Bass, R. Green, J. A. Wells, *Proteins: Struct. Funct. Genet.* **1990**, *8*, 309 – 314.
[58] Y. Choo, A. Klug, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 1163 – 1167.
[59] E. J. Rebar, C. O. Pabo, *Science* **1994**, *263*, 671 – 673.
[60] A. C. Jamieson, S.-H. Kim, J. A. Wells, *Biochemistry* **1994**, *33*, 5689 – 5695.
[61] Y. Choo, I. Sánchez-García, A. Klug, *Nature* **1994**, *372*, 642 – 645.
[62] L. Huang, T. Sera, P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 3969 – 3973.
[63] T. Sera, P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 2920 – 2925.
[64] W. A. Lim, R. T. Sauer, *Nature* **1989**, *339*, 31 – 36.
[65] S. Kamtekar, J. M. Schiffer, H. Xiong, J. M. Babik, M. H. Hecht, *Science* **1993**, *262*, 1680 – 1685.
[66] A. R. Davidson, R. T. Sauer, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 2146 – 2150.
[67] B. I. Dahiyat, S. L. Mayo, *Science* **1997**, *278*, 82 – 87.
[68] A. D. Ellington, J. W. Szostak, *Nature* **1990**, *346*, 818 – 822.
[69] C. Tuerk, L. Gold, *Science* **1990**, *249*, 505 – 510.
[70] D. L. Robertson, G. F. Joyce, *Nature* **1990**, *344*, 467.
[71] A. A. Beaudry, G. F. Joyce, *Science* **1992**, *257*, 635.
[72] C. Tuerk, S. MacDougal, L. Gold, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 6988.
[73] D. P. Bartel, M. L. Zapp, M. R. Green, J. Szostak, *Cell* **1991**, *67*, 529.
[74] M. Sassanfar, J. W. Szostak, *Nature* **1993**, *364*, 550 – 553.
[75] M. Famulok, J. W. Szostak, *J. Am. Chem. Soc.* **1992**, *114*, 3990.
[76] G. J. Connell, M. Illangesekare, M. Yarus, *Biochemistry* **1993**, *32*, 5497.
[77] L. Bock, L. Griffin, J. Latham, E. Vermaasm, J. Toole, *Nature* **1992**, *355*, 564 – 566.
[78] A. D. Ellington, J. W. Szostak, *Nature* **1992**, *355*, 850.
[79] F. Jiang, R. A. Kumar, R. A. Jones, D. J. Patel, *Nature* **1996**, *382*, 183 – 186.
[80] C. R. Woese, S. Winker, R. R. Gutell, *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 8467 – 8471.
[81] R. D. Jenison, S. C. Gill, A. Pardi, B. Polisky, *Science* **1994**, *263*, 1425 – 1429.
[82] D. Pei, H. D. Ulrich, P. G. Schultz, *Science* **1991**, *253*, 1408 – 1411.
[83] I. Haruna, S. Spiegelman, *Science* **1965**, *150*, 886.
[84] R. Green, *Science* **1992**, *258*, 1910 – 1915.
[85] N. Lehman, G. F. Joyce, *Nature* **1993**, *361*, 182 – 185.
[86] D. P. Bartel, J. W. Szostak, *Science* **1993**, *261*, 1411 – 1418.
[87] E. H. Ekland, J. W. Szostak, D. P. Bartel, *Science* **1995**, *269*, 364 – 370.
[88] E. H. Ekland, D. P. Bartel, *Nature* **1996**, *382*, 373 – 376.
[89] J. R. Prudent, T. Uno, P. G. Schultz, *Science* **1994**, *264*, 1924 – 1927.
[90] M. M. Conn, J. R. Prudent, P. G. Schultz, *J. Am. Chem. Soc.* **1996**, *118*, 7012 – 7013.
[91] Y. Li, D. Sen, *Nat. Struct. Biol.* **1996**, *3*, 743 – 747.
[92] T. M. Tarasow, S. L. Tarasow, B. E. Eaton, *Nature* **1997**, *389*, 54 – 57.
[93] M. A. Barry, W. C. Lai, S. A. Johnston, *Nature* **1995**, *377*, 632 – 635.
[94] C. Grimm, E. Lund, J. E. Dahlberg, *EMBO J.* **1997**, *16*, 793 – 806.
[95] R. B. Merrifield, *J. Am. Chem. Soc.* **1963**, *85*, 2149 – 2154.
[96] H. M. Geysen, R. H. Meloen, S. J. Barteling, *Proc. Natl. Acad. Sci. USA* **1984**, *81*, 3998 – 4002.
[97] R. A. Houghten, *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 5131 – 5135.
[98] H. M. Geysen, S. J. Rodda, T. J. Mason, G. Tribbick, P. G. Schoofs, *J. Immunol. Methods* **1987**, *102*, 259 – 274.
[99] H. M. Geysen, S. J. Rodda, T. J. Mason, *Mol. Immunol.* **1986**, *23*, 709 – 715.
[100] S. P. A. Fodor, J. L. Read, M. C. Pirrung, L. Stryer, A. T. Lu, D. Solas, *Science* **1991**, *251*, 767.
[101] J. W. Jacobs, S. P. A. Fodor, *Trends Biotechnol.* **1994**, *12*, 19 – 26.
[102] A. Furka, F. Sebestyen, M. Asgedom, G. Dibo, *Int. J. Pept. Protein Res.* **1991**, *37*, 487 – 493.
[103] K. S. Lam, S. E. Salmon, E. M. Hersh, V. J. Hruby, W. M. Kazmierski, R. J. Knapp, *Nature* **1991**, *354*, 82 – 84.
[104] R. A. Houghten, C. Pinilla, S. E. Blondelle, J. R. Appel, C. T. Dooley, J. H. Cuervo, *Nature* **1991**, *354*, 84 – 86.
[105] C. Pinilla, J. R. Appel, P. Blanc, R. A. Houghten, *Biotechniques* **1992**, *13*, 901 – 905.
[106] K. S. Lam, V. J. Hruby, M. Lebl, R. J. Knapp, W. M. Kazmierski, E. M. Hersh, S. E. Salmon, *Bioorg. Med. Chem. Lett.* **1993**, *3*, 419 – 424.
[107] G. P. Smith, D. A. Schultz, J. E. Ladbury, *Gene* **1993**, *128*, 37 – 42.
[108] S. Brenner, R. A. Lerner, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 5181 – 5183.
[109] M. N. Needels, D. G. Jones, E. H. Tate, G. L. Heinkel, L. M. Kochersperger, W. J. Dower, R. W. Barrett, M. A. Gallop, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 10700 – 10704.
[110] V. Nikolaiev, A. Stierandova, V. Krchnak, B. Sekigmann, K. S. Lam, S. E. Salmon, M. Lebl, *Pept. Res.* **1993**, *6*, 161 – 170.
[111] M. H. J. Ohlmeyer, R. N. Swanson, L. W. Dillard, J. C. Reader, G. Asouline, R. Kobayashi, M. Wigler, W. C. Still, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 10922 – 10926.
[112] R. A. Houghten, J. R. Appel, S. E. Blondelle, J. H. Cuervo, C. T. Dooley, C. Pinilla, *Biotechniques* **1992**, *13*, 412 – 421.
[113] M. Torneiro, W. C. Still, *J. Am. Chem. Soc.* **1995**, *117*, 5887 – 5888.
[114] S. S. Yoon, W. C. Still, *Tetrahedron* **1995**, *51*, 567 – 578.
[115] A. Borchardt, W. C. Still, *J. Am. Chem. Soc.* **1994**, *116*, 373 – 374.
[116] J. K. Chen, S. L. Schreiber, *Angew. Chem.* **1995**, *107*, 1041; *Angew. Chem. Int. Ed. Engl.* **1995**, *34*, 953 – 969.
[117] C. Y. Cho, E. J. Moran, S. R. Cherry, J. C. Stephans, S. P. A. Fodor, C. L. Adams, A. Sundaram, J. W. Jacobs, P. G. Schultz, *Science* **1993**, *261*, 1303 – 1305.
[118] R. J. Simon, R. S. Kania, R. N. Zuckermann, V. D. Huebner, D. A. Jewell, S. Banville, S. Ng, L. Wang, S. Rosenberg, C. K. Marlowe, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 9367 – 9371.
[119] R. N. Zuckermann, E. J. Martin, D. C. Spellmeyer, G. B. Stauber, K. R. Shoemaker, J. M. Kerr, G. M. Figliozzi, D. A. Goff, M. A. Siana, R. J. Simon, *J. Med. Chem.* **1994**, *37*, 2678 – 2685.
[120] J. S. Nowick, N. A. Powell, E. J. Martinez, E. M. Smith, G. Noronha, *J. Org. Chem.* **1992**, *57*, 3763 – 3765.
[121] K. Burgess, J. Ibarzo, D. S. Linthicum, D. H. Russell, H. Shin, A. Shitangkoon, R. Totani, A. J. Zhang, *J. Am. Chem. Soc.* **1997**, *119*, 1556 – 1564.
[122] J.-M. Kim, T. E. Wilson, T. C. Norman, P. G. Schultz, *Tetrahedron Lett.* **1996**, *37*, 5305 – 5308.

[123] A. B. Smith, T. P. Keenan, R. C. Holcomb, P. A. Sprengeler, M. C. Guzman, J. L. Wood, P. J. Carroll, R. Hirschmann, *J. Am. Chem. Soc.* **1992**, *114*, 10672 – 10674.
[124] H. Han, K. D. Janda, *J. Am. Chem. Soc.* **1996**, *118*, 2539 – 2544.
[125] C. Gennari, B. Salom, D. Potenza, A. Williams, *Angew. Chem.* **1994**, *106*, 2181; *Angew. Chem. Int. Ed. Engl.* **1994**, *33*, 2067 – 2069.
[126] M. Hagihara, N. J. Anthony, T. J. Stout, J. Clardy, S. L. Schreiber, *J. Am. Chem. Soc.* **1992**, *114*, 6568 – 6570.
[127] D. H. Appella, L. A. Christianson, D. A. Klein, D. R. Powell, X. Huang, J. J. Barchi, S. H. Gellman, *Nature* **1997**, *387*, 381 – 384.
[128] J. L. Matthews, M. Overhand, F. N. M. Kühnle, P. E. Ciceri, D. Seebach, *Liebigs Ann.* **1997**, 1371 – 1379.
[129] C. Y. Cho, R. S. Youngquist, S. J. Paikoff, M. H. Beresini, A. R. Hebert, C. W. Liu, D. E. Wemmer, T. Keough, P. G. Schultz, *J. Am. Chem. Soc.* **1998**, in press.
[130] B. A. Bunin, J. A. Ellman, *J. Am. Chem. Soc.* **1992**, *114*, 10997 – 10998.
[131] B. A. Bunin, M. J. Plunkett, J. A. Ellman, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 4708 – 4712.
[132] A. A. Virgilio, A. A. Bray, W. Zhang, L. Trinh, M. Snyder, M. M. Morrissey, J. A. Ellman, *Tetrahedron* **1997**, *53*, 6635 – 6644.
[133] S. H. De Witt, J. S. Kiely, C. J. Stankovic, M. C. Schroeder, D. M. Cody, M. R. Pavia, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 6909 – 6913.
[134] D. A. Campbell, J. C. Bermak, T. S. Burkoth, D. V. Patel, *J. Am. Chem. Soc.* **1995**, *117*, 5381 – 5382.
[135] M. M. Murphy, J. R. Schullek, E. M. Gordon, M. A. Gallop, *J. Am. Chem. Soc.* **1995**, *117*, 7029 – 7030.
[136] B. Ruhland, A. Bhandari, E. M. Gordon, M. A. Gallop, *J. Am. Chem. Soc.* **1996**, *118*, 253 – 254.
[137] J. Green, *J. Org. Chem.* **1995**, *60*, 4287 – 4290.
[138] R. Liang, L. Yan, J. Loebach, M. Ge, Y. Uozumi, K. Sekanina, N. Horan, J. Gildersleeve, C. Thompson, A. Smith, *Science* **1996**, *274*, 1520 – 1522.
[139] P. H. H. Hermkens, H. C. J. Ottenheijm, D. Rees, *Tetrahedron* **1996**, *52*, 4527 – 4554.
[140] P. H. H. Hermkens, H. C. J. Ottenheijm, D. Rees, *Tetrahedron* **1997**, *53*, 5643 – 5678.
[141] E. K. Kick, D. C. Roe, A. G. Skillman, G. Liu, T. J. A. Ewing, Y. Sun, I. D. Kuntz, J. A. Ellman, *Chem. Biol.* **1997**, *4*, 297 – 307.
[142] C. Khosla, *Chem. Rev.* **1997**, *97*, 2577 – 2590.
[143] J. Cortes, S. F. Haydock, G. A. Roberts, D. J. Bevitt, P. F. Leadlay, *Nature* **1990**, *348*, 176 – 178.
[144] S. Donadio, M. J. Staver, J. B. McAlpine, S. J. Swanson, L. Katz, *Science* **1991**, *252*, 675 – 679.
[145] J. R. Jacobsen, C. R. Hutchinson, D. E. Cane, C. Khosla, *Science* **1997**, *277*, 367 – 369.
[146] A. C. Pease, D. Solas, E. J. Sullivan, M. T. Cronin, C. P. Holmes, S. P. A. Fodor, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 5022 – 5026.
[147] M. S. Chee, X. Huang, R. Yang, E. Hubbell, A. Berno, D. Stern, J. Winkler, D. J. Lockhart, M. S. Morris, S. P. A. Fodor, *Science* **1996**, *274*, 610 – 614.
[148] L. Wodicka, H. Dong, M. Mittman, M.-H. Ho, D. J. Lockhart, *Nat. Biotechnol.* **1997**, *15*, 1359 – 1367.
[149] M. D. Adams, J. M. Kelley, J. D. Gocayne, M. Dubnick, M. H. Polymeropoulos, H. Xiao, C. R. Merril, A. Wu, B. Olde, R. F. Moreno, A. R. Kerlavage, W. R. McCombie, J. C. Venter, *Science* **1991**, *252*, 1651 – 1656.
[150] G. G. Lennon, H. Lehrach, *Trends Genet.* **1991**, *7*, 314 – 317.
[151] P. Liang, A. B. Pardee, *Science* **1992**, *257*, 967 – 971.
[152] K. Okubo, N. Hori, R. Matoba, T. Niiyama, A. Fukushima, Y. Kojima, K. Matsubara, *Nat. Genet.* **1992**, *2*, 173 – 179.
[153] S. Meier-Ewert, E. Maier, A. Ahmadi, J. Curtis, H. Lehrach, *Nature* **1993**, *361*, 375 – 376.
[154] V. E. Velculescu, L. Zhang, B. Vogelstein, K. W. Kinzler, *Science* **1995**, *270*, 484 – 487.
[155] N. Zhao, H. Hashida, N. Takahashi, Y. Misumi, Y. Sakaki, *Gene* **1995**, *156*, 207 – 213.
[156] C. Nguyen, D. Rocha, S. Granjeaud, M. Baldit, K. Bernard, P. Naquet, B. R. Jordan, *Genomics* **1995**, *29*, 207 – 216.
[157] M. Schena, D. Shalon, R. W. Davis, P. O. Brown, *Science* **1995**, *270*, 467 – 470.
[158] N. S. Gray, L. Wodicka, A.-M. Thunnissen, T. C. Norman, S. Kwon, F. H. Espinoza, D. O. Morgan, G. Barnes, S. LeClerc, L. Meijer, S.-H. Kim, D. J. Lockhart, P. G. Schultz, *Science* **1998**, *281*, 533 – 538.
[159] N. S. Gray, S. Kwon, P. G. Schultz, *Tetrahedron Lett.* **1997**, *38*, 1161 – 1164.
[160] T. C. Norman, N. S. Gray, J. T. Koh, *J. Am. Chem. Soc.* **1996**, *118*, 7430 – 7431.
[161] J. L. DeRisi, V. R. Iyer, P. O. Brown, *Science* **1997**, *278*, 680 – 686.
[162] D. Hirata, K. Yano, T. Miyakawa, *Mol. Gen. Genet.* **1994**, *242*, 250.
[163] G. Zlokarnik, P. A. Negulescu, T. E. Knapp, L. Mere, N. Burres, L. Feng, M. Whitney, K. Roemer, R. Y. Tsien, *Science* **1998**, *279*, 84 – 88.
[164] S. D. Liberles, S. T. Diver, D. J. Austin, S. L. Schreiber, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 7825 – 7830.
[165] E. J. Licitra, J. O. Liu, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 12817 – 12821.
[166] F. J. DiSalvo, *Science* **1990**, *247*, 649.
[167] X.-D. Xiang, X. Sun, G. Briceño, Y. Lou, K.-A. Wang, H. Chang, W. G. Wallace-Freedman, S.-W. Chen, P. G. Schultz, *Science* **1995**, *268*, 1738 – 1740.
[168] J. X. Wang, *Science* **1998**, in press.
[169] E. Danielson, J. H. Golden, E. W. McFarland, C. M. Reaves, W. H. Weinberg, X. D. Wu, *Nature* **1997**, *389*, 944.
[170] H. Chang, C. Gao, I. Takeuchi, Y. Yoo, J. Wang, P. G. Schultz, X.-D. Xiang, *Appl. Phys. Lett.*, submitted.
[171] X. Sun, K.-A. Wang, Y. Yoo, W. G. Wallace-Freedman, C. Gao, X.-D. Xiang, P. G. Schultz, *Adv. Mater.* **1998**, *9*, 1046.
[172] R. Nielsen, personal communication, **1997**.
[173] P. G. Schultz, X.-D. Xiang, *Curr. Opin. Solid State Mater. Sci.* **1998**, in press.
[174] Y. Lu, T. Wei, F. Duewer, Y. Lu, N. B. Min, P. G. Schultz, X.-D. Xiang, *Science* **1997**, *276*, 2004.
[175] E. D. Isaacs, M. Kao, G. Aeppli, X.-D. Xiang, X. Sun, P. G. Schultz, M. A. Marcus, G. S. I. Cargill, R. Haushalter, *Appl. Phys. Lett.*, in press.
[176] F. C. Moates, M. Somani, J. Annamalai, J. T. Richardson. *Ind. Eng. Chem. Res.* **1996**, *35*, 4801.
[177] H. Weinberg, personal communication, **1998**.
[178] G. Briceño, H. Chang, P. G. Schultz, X.-D. Xiang, *Science* **1995**, *270*, 273 – 275.
[179] H. Weinberg, *Science* **1998**, in press.
[180] C. L. Hill, R. D. Gall, *J. Mol. Catal. A* **1996**, *114*, 103.
[181] S. Brocchini, K. James, V. Tangpasuthadol, J. Kohn, *J. Am. Chem. Soc.* **1997**, *119*, 4553.
[182] T. A. Dickinson, D. R. Walt, J. While, J. S. Kauer, *Anal. Chem.* **1997**, *69*, 3413.
[183] I. Takeuchi, H. Weinberg, unpublished results.