한국농화학회지 제 40권 제 1호(1997)
Agricultural Chemistry and Biotechnology
Vol. 40, No. 1, pp. 1~11(1997)

REVIEW

Recent Development of Analytical Methods for the Biopolymers Built on a Solid Support

Hunwoo Shin

P-project team, Yu-kong, Ltd. Wonchon dong 140-1, Yusung gu, Daejon

Recently, combinatorial chemistry is blooming in the area to search for biologically active molecules.[1-2] Contrary to the traditional "rational" approach based on structure and mechanism, this method generate hundreds to millions of "randomly" combinated oligomers (library) in a very short time using solid phase (peptide) chemistry[3-5] and split synthesis tactic.[6,7] (Fig. 1).

The libraries generated are screened[8-11] using fluoroscent group attached antibody and few resin beads bound with the antibody are separated from the library pools and sequenced using several method. Using the obtained sequence, several lead compounds are generated with the aid of molecular modeling which are expected to show similar binding properties but with improved efficacy, stability and bioavailabilty.

The pertinence of this method as a fast and efficient substitute for the traditional method is not yet fully proven as application of combinatorial chemistry for the development of biologically active molecules began only few years ago. Also, as solid phase chemistry has not been fully established as in the case of solution chemistry, the appropriate solid phase reaction should be established prior to the development of the new lead. This may limits the diversity of the library and obviate the benefits of the combinatorial chemistry (fast development). However, the first phase I clinical test for a CNS (central nervous system) compound developed by this method is already known in progress (Eli-Lilly) and reportedly, it took less than 2 years from target identification to the beginning of clinical trials.[12]

Application of combinatorial chemistry for the development of agricultural medicines is not much known so far as in the case of human drug, probably due to the lack of the biochemical knowledge. But with appropriate biological screening method, it will be a good area to apply this new method.

One of the important issue for the use of library for the drug discovery is the sequencing of the single bead selected via screening.

A prerequisite for the single bead analysis is the high sensitivity of the detection methods due to the limited amount of the materials attached on a resin bead. The size of a resin bead is as small as 10 μm (5×10^9 beads/g) and loading capacity of the resin is about 20 fmol. The loading capacity of the widely used 80 μm resin bead (3×10^6 beads/g) is 80~100 pmol. Except Edman degradation analysis of the peptides, only small portions of the reaction site of the resin bead is used for the sequencing, and the amount of the materials ranges between few femtomol to picomol. Thus, for the single bead analysis, below picomol detection limit is very important. Another requirement for the sequencing is the wide applicability of the sequencing method to the diverse oligomers. Traditionally, most of the libraries were built using solid phase peptide synthesis (SPPS) and the resulting oligopeptides were analyzed using Edman degradation. However, as Edman degradation is applicable almost only to the oligomers composed of natural amino acids, recently developed diverse oligomers[13] such as retro-inverso peptides,[14-17] peptoids,[18-20] peptidyl phosphonates,[21-22] oligo-ureas,[23-24] and oligosaccharides[25] can not be analyzed using this method. To solve this problem, various sequencing methods were developed recently and in the following section, those methods are discussed.

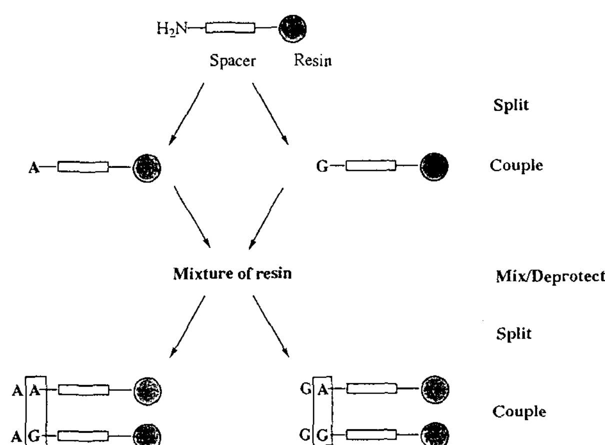


Fig. 1. Generation of a peptide library by split synthesis. Resins are divided into n portions, coupled with n-building blocks, mixed and deprotected. By 1 more cycle, n^2 number of combinations are generated.

Key words : combinatorial chemistry, solid phase peptide synthesis, enconding method, Edman degradation

Direct Analysis Method

Edman degradation-HPLC method

Edman degradation[26,27] is a method that cleaves the N-terminal amino acid of a peptide step by step to its phenylthiohydantoin analogs by the reaction of the phenylisothiocyanate with N-terminal amino acid unit of a peptide. (Fig. 2)

Phenylisothiocyanate react with the αNH_2 functional group to form the phenylthiocarbamoyl peptide and this intermediate decompose in acidic condition to a 2-anilinothiazolin-5-one and the peptide which has one less amino acid. The 2-anilinothiazolin-5-one is further transformed to a more stable 3-phenyl-2-thiohydantoin (PTH amino acid) through phenylthiocarbamoyl amino acid (PTC amino acid) and it is identified by HPLC by comparison with the standard materials derived from the known amino acids. For the analysis of a polypeptide on a solid phase (resins), same principles can be applied.[28] However, in this case, the sequenceable chain length is limited by small quantity of the oligopeptide and the deletion of the chain during the Edman degradation. In the case of TantaGel S resin which could load 80~100 pmol per 80 μm size bead, ~10 residue is the maximum sequenceable length despite the ~1 pmol quantity detection limit of the HPLC method.

A major limitation of the Edman degradation in the usual peptides analysis is the low sample throughput and with automated facility equipped with on-line PC, only about 60 residue can be analyzed per day. For the oligopeptide library generated for the screening purpose, residues usually do not exceed 5~6 mers as the peptide epitopes are usually composed of 5~6 amino acid residues. Thus, the low sample throughput is not a big problem. Rather, limitation lies in that it is applicable almost only to the peptides composed of natural L-amino acids.[28] Analysis of some phosphorylated peptides by this method were reported but it was possible only after especially elaborated analytical procedures. Thus, direct analysis of diverse oligomers that are designed to contain unnatural amino acid or non-peptide backbone such as carbamate and urea to improve binding properties, solubility and proteolysis of peptide is impossible using this method.(Fig. 3)

This is an important problem in the analysis of the solid phase bound oligomers because of the increasing diversity of the oligomers.

Matrix-Assisted Laser Desorption Ionization Time-of-Flight (MALDI-TOF) Mass Spectroscopic method.

Recently S. Kent used Edman degradation-MALDI-TOF method[30-33] to analyze a ploypeptide. Different from the Edman degradation-HPLC method, the mixture

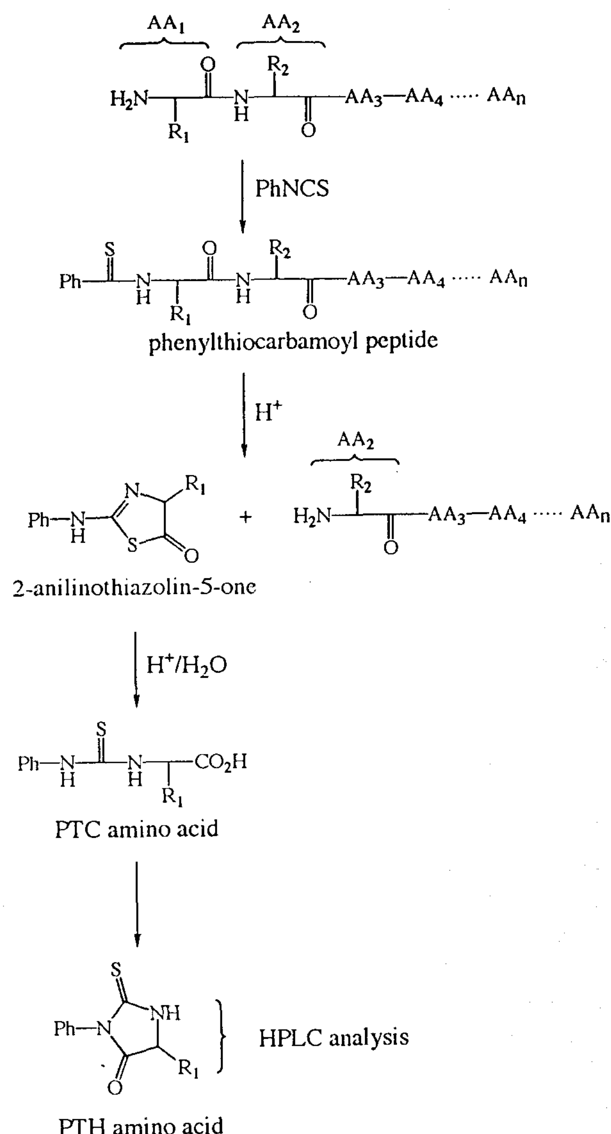


Fig. 2. Edman degradation of a peptide. The N-terminal amino acid building block is cleaved stepwise by the reaction with PhNCS. Resulting unstable anilinothiazolinone is transformed into more stable PTH amino acid and analyzed by HPLC.

of phenylisocyanate (5%) and phenylisothiocyanate (95%) was used instead of pure phenylisothiocyanate. Phenylisocyanate react with the $^{\alpha}NH_2$ of the N-terminal amino acid unit to form N^{α}-phenylcarbamoylpeptide but does not further decompose to the shortened peptide as in the case of the phenyliso-thiocyanate. Thus after a round, there exist 5% of the N^{α}-phenylcarbamoylpeptide along with 95% of the shortened peptides. This mixture is in turn subjected to the next cascading degradation round without separation, eventually resulting in the mixtures of the N^{α}-phenyl carbamoyl peptides of the different chain length. The mixture is analyzed by MALDI-TOF mass spectroscopy without fragmentation of the parent ion peaks and thus, the chromatogram of the mixture is appeared as a "ladder" as modeled in the following figure. (Fig. 4)

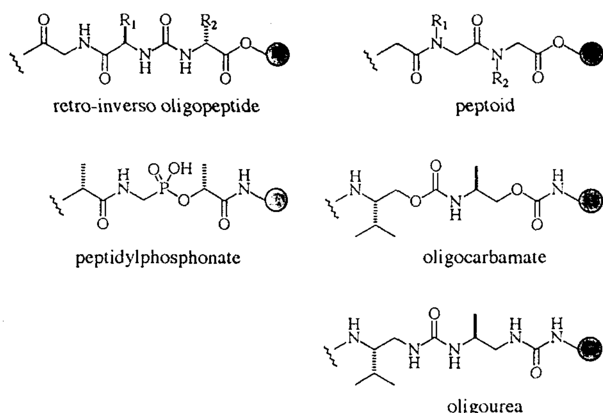


Fig. 3. Some examples of recently developed solid bound oligomers. These structures can not be analyzed by conventional Edman degradation method.

The mass difference of each peak stands for the amino acid building unit and read out of the mass difference ("ladder sequencing") assign the sequence. This method is faster and convenient as it does not analyze each fragment one by one as in the case of conventional Edman degradation-HPLC method. Also this method is more sensitive than Edman degradation-HPLC method. MALDI-TOF mass spectrometry is possible to identify 2~5 fmol quantity of the substrate and using this technique, portions (5%) of the fragments derived from 10 pmol peptides immobilized on a solid support could be analyzed after degradation.[30]

Above method was also applied to the analysis of a phosphorylated peptides. The degradation method is milder as it does not need to use strong acidic condition to form PTH amino acid (25% TFA, 60°C) from PTC amino acid and acid-labile phosphorylated amino acids like phosphoserine remain intact and can be identified directly. Using this technique, site of post translational modification present in a 16 residue peptide LRRAS(Pi) GLI-YNNPLMAR-NH_2 was identified by comparison with the unphosphorylated analog. However, for this method, there are some limitations. Firstly, isobaric materials like Leu (Δm=113.2) and Ile (Δm=113.2) can not be differentiated as sequencing is based on the mass difference of each building block. Secondly, it is difficult to sequence high mass oligopeptides because of the limit of the resolution. The present resolution accuracy is ~1/10,000 for MALDI-TOF method. For peptides analysis, because of the presence of the amino acid residues Gln (Δm=128.1), Glu (Δm=129.1) and Asn (Δm=114.1), Asp (Δm=115.1) which are different by 1 Da, the deviation ought be in the range of ±0.3 Da to assign the sequence of the peptide without misinterpretation. This limits the allowable mass to ~3,000 Da (3,000×1/10,000=0.3) which corresponds roughly to the peptides of 30 residues. For Lys (Δm=128.2), ε-amino group is coupled to phenylisothiocyanate during the analysis and it is distinguished from Gln and Glu.

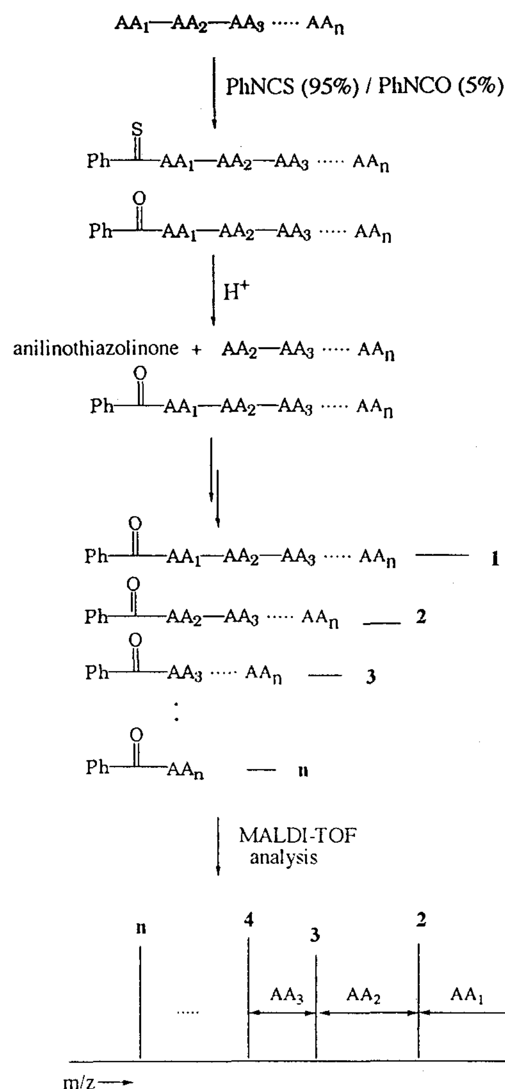


Fig. 4. Peptide analysis using MALDI-TOF mass spectroscopy. Phenylisocyanate (PIC) adduct of a peptide fragments generated in the presence of PITC (95%) and PIC(5%) are used for the analysis. Sequence is revealed by the mass difference of phenylisocyanate adducts.

T. Keough[35] group also used MALDI-TOF technique to sequence resin bound peptides. The fragments were generated during the sequence construction by capping certain portions (10%) of the sequence at each building step using acetyl protected Ala. The sequences were built on the H_2N-XXXRM (X=arbitrary amino acid as a linker, R=Arg, and M=Met) coupled to the TantaGel S resin bead and C-terminal methionine was selectively cleaved from the resin by treating it with CNBr and 0.1N HCl for 16 h in the dark at room temperature. Arg (R) was introduced to make all the fragment contain at least one charge to enhance the MALDI sensitivity and linker (X) amino acids were used to make the fragments heavier (>500 Da) to avoid the 'chemical noise' produced by the desorption of the UV-absorbing matrix. When the

14 single beads of synthetic penta peptides built on 88 μm (~100 pmol/bead) TantaGel S were analyzed as unknowns, 95% success rate was claimed only with the use of 5% of the fragments, and no significant impurities originated from resin were found. By this method, isobaric materials like D and L-amino acid should be differentiated by employing different capping materials such as acetyl, pivaloyl, and trifluoroacetyl group, thus, making their mass different from each other.

This method, however, suffers a problem originated from the relatively small amount of the capped fragments compared with the parent sequence. After cleavage of the materials from the resin, there would exist uncapped parent sequence [as a major component], capped fragments, and side products generated from the side reaction and/or incompletion of the coupling reaction. As the each capped partial sequences is only ~10% of the parents sequences in the given example, if the reaction is 90% complete or if 10% side reaction occurs, it will generate equal amount of the side products to the capped fragments, and it would be enough to scramble the spectrum. Thus, if the sequence building reaction is not very clean, the scrambling will be very significant problem. Also, scrambling by the fragmentation of the parents ion during the mass analysis is another problem. Significant fragmentation is not reported in peptides analysis but it is significant in the case of oligonucleotides depending on the matrix and nucleobase.[36-38] Capping % ratio may be increased for these problems but it will limit the chain length and may intensify the erroneous binding to the capping sequences for both cases.

As an additional merit of MALDI-TOF approach, time and effort which had to be poured in for the cleavage of the fragments may be saved by direct mass analysis of peptides (oligomers) covalently attached to a resin bead. G. Siuzdak group recently reported direct analysis of the undecapeptide KPAFLKPQFLG built on a photolabile group attached to a resin bead using the UV laser source [355 nm] in the mass spectrometer without any prior treatment.[39-42] (Fig. 5).

Moreover, coupling reaction of the undecapeptide KPAFLKPQFLG with Boc-Arg(Tos) was monitored according to the time by this method and single bead analysis at 0, 1, 6 min. clearly showed the appearance of the product and the disappearance of the starting material. This method seems somewhat similar to the above mentioned method that use CNBr to cleave Arg site of the peptide, but greatly different in that it does not need any prior treatment, thus, making monitoring of the solid phase reaction possible. At this stage, use of this technique for ladder sequencing method are not reported but there is no reason why this is not applicable to the sequence analysis.

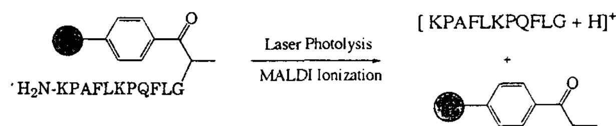


Fig. 5. Schematic representation of the product generated upon laser photolysis and MALDI ionization of a peptide covalently attached to a solid phase resin through a photolabile a-methylphenacyl-ester linker.

Encoding method

Aside from the mass spectroscopic analysis to overcome the limit of the present analytical method (Edman degradation), another widely accepted methods are encoding. Encoding is a method that inserts some kind of information corresponding to the each oligomer building block during the sequence building process and the information are read when necessary to provide the clue of the sequence. The informations (code) are inserted in the form of combinated oligopeptides, oligonucleotide and haloaromatics (that differ in retention time) but in some cases, a predetermined oligomers are built on a specified sector of glass plate to use the spatial location of each oligomer as an information source. Recently, radio frequency recording chips were used for information storage. With these methods, diverse oligomers could be analyzed but on the other hand, it evoked problems to be solved such as excess labor to build the coding sequence, undesired binding interaction between antibody and coding sequence, and chemical compatibility between the construction condition of coding and binding strand.

Peptide and oligonucleotide code.

A strategy that use natural peptide as a code for the sequencing of unnatural oligomers were developed by Zuckerman and K. S. Lam respectively based on the same principle.[43,44] That is, to the appropriate linker attached to the resin bead, Lys group that is protected with two orthogonal protecting group (acid labile Boc for N^α and base labile Fmoc for N^ε or Fmoc for N^α and weak acid labile Moz for N^ε, vice versa) is coupled and coding and screening strands are selectively constructed on one of the amine functional group by alternative parallel synthesis (Fig. 6).

For example, to the resin coupled N^α-Fmoc-N^ε-Ddz-Lys, piperidine is treated to deprotect N^α-Fmoc protecting group and Fmoc protected (un)natural building block (usually amino acid) is coupled using standard peptide coupling protocol to build screening strand. Then, N^ε-Ddz protecting group is deprotected using 2% trifluoroacetic acid (TFA), and the corresponding combinated and Ddz protected peptide sequence, for example, GlyAlaPhe (GAF) for ornithine (in screening strand) and LeuGlyAla (LGA) for N-sec-butylglycine

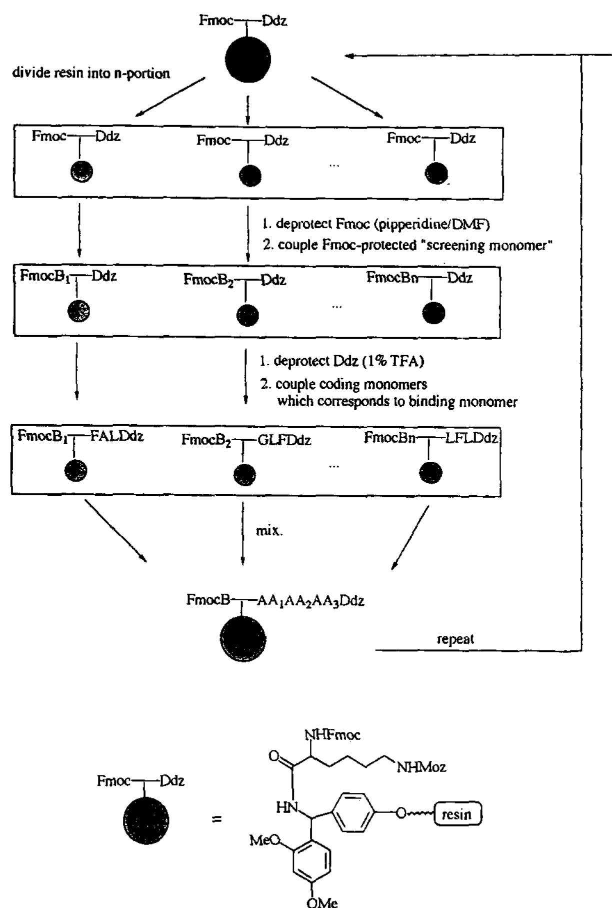


Fig. 6. The synthesis of a peptide encoded library by alternative parallel synthesis strategy. Coding and binding strands were built respectively using orthogonal protecting groups during the library generation. The unknown building block (denoted as B) can be identified by reading the 3 amino acid decoding sequence $AA_1AA_2AA_3$.

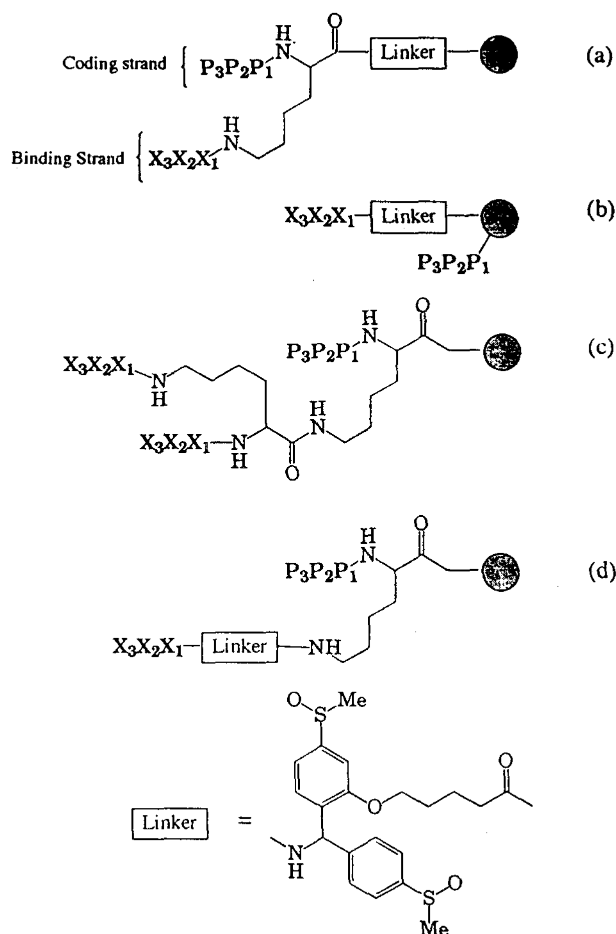


Fig. 7. Alternative ways of attachment of coding sequence to the solid-phase particle. Typically both coding and binding strands are built on the Lys as shown in (a) and problems by the interaction between coding strand and antibody would be minimized by the statistical distribution of both strand as shown in (b). The ratio of both strand also could be controlled by multiple use of Lys as shown in (c). By insertion of the linker molecule, the binding strand can be released into solution in the later stage for the solution phase testing as shown in (d).

(in screening strand) are coupled as a code for the unnatural building block in the screening strand. When necessary, sequence of the binding strand is decoded by reading the sequence of the coding strand. This alternative parallel synthesis is repeated and at the end, the binding strands are blocked by acetyl group to prevent the Edman degradation during the analysis of coding sequence. The number of the unnatural building block is not actually limited because tripeptides composed of 10 amino acids are enough for the assignment of 1,000 building blocks. (10^3=1,000)

One of the problem of the Lys based alternative parallel synthesis is the binding interaction of the coding strand with the screening antibody (coding : binding strand=1 : 1) and/or the cooperative effect of both sequence which mislead the sequence information. This problem may be diminished by statistical distribution of both strand on the resin bead. (Fig. 7) and in this case, any possible ratio can be achieved and misleading can be minimized. Moreover, by employing the different types of linker, multiple screening strand can be released at different condition while the coding strand is remaining on the bead, which could be advantageous for the screening and sequencing purpose.[45)]

Another problem of this method is the excess labor to build the coding strand which are not necessary in the case of other method. When 3 codes are used, for example, for the pentamer library composed of 10 unnatural amino acid, 30 coupling and deprotection step will be required for the first building blocks and 150 coupling and deprotection step in total. This problem would be exaggerated for bigger library.

Similar to the above method, oligonucleotides were used for the coding of the unnatural peptide sequence analysis.[46-50)] As an example,[46] on a 10 μm polystyrene resin bead (20 fmol loading capacity), C-terminal heptapeptide (RQFKVVT) of dynorphin B was successfully synthesized using automatic synthesizer in parallel with

69-mer oligonucleotide containing the code for the heptapeptide. Two nucleotides were used for the assignment of each amino acid building block and they were TA, TC, CT, AT, TT, TT, and AC for Arg (R), Gln (Q), Phe (F), Lys (K), Val (V), and Thr (T) respectively. The sequence of the oligonucleotides was **5'-ATC CAA TCT CTC CAC** *ATC TCT ATA CTA TCA TCA CC* [TA TC CT AT TT TT AC] CTC AC**T CAC TTC CAT TCC AC-3'** and bold portion of this sequence corresponds to PCR-priming site while the region in italics is homologous to the primer used for sequencing this template. The portion in brackets corresponds to the peptide sequence. The ratio between oligonucleotide and peptide codes was ~1 : 20 but after polymerase chain reaction (PCR), oligonucleotides were successfully analyzed to reveal the corresponding peptide sequences. This method was successfully applied for the sequencing of 10 μm beads obtained from heptapeptide library composed of 7 amino acids Arg, Gln, Phe, Lys, Val, (D)-Val and Thr (7^7=823,543)

For the parallel synthesis of the peptides and nucleotides, polystyrene beads (10 μm diameter, 20 fmol loading capacity) was first treated with a mixture of Fmoc-Thr(tert-butyl)-OH (for peptide coding) and 4-O-DMT-oxybutyrate (for oligonucleotide coding) and each codes were built on each side using N^{α}-Fmoc protected amino acid and 5-O-DMT protected-2-deoxynucleoside phosphoramidites respectively. The ratio between peptide strands and oligonucleotides code was ~20 : 1. The chemical compatibility of the oligonucleotide and peptide was achieved by slight modification of the standard oligonucleotide synthesis process. Though 1% trichloroacetic acid (TCA)/ CH_2Cl_2 for deblocking of dimethoxytrityl (DMT) protecting group, I_2/collidine/ H_2O/ CH_3CN for oxidation of phosphite triester, thiophenolate for phosphate O-demethylation and ethanloic ethylenediamine for debenzoylation of protected cytidine and 7-deazaadenine residue are known inert for the peptide chain, deprotection condition for the Boc protecting group of the side chain of amino acid leads to rapid depurination of oligonucleotides containing either 2'-deoxyadenosine (dA) or 2'-deoxyguanosine (dG). This problem was avoided by using 7-deaza-2'-deoxyadenosine (c^7dA) in the template oligonucleotide tag. Also, the use of piperidine [10% in dimethoxyformamide (DMF), 5-10 min] for the deprotection of Fmoc protecting group is known to result in partial demethylation of phosphotriester and thus inducing aberrant phosphitylation during the chain elongation but it was reversed by the final oligonucleotide deprotection steps.[51]

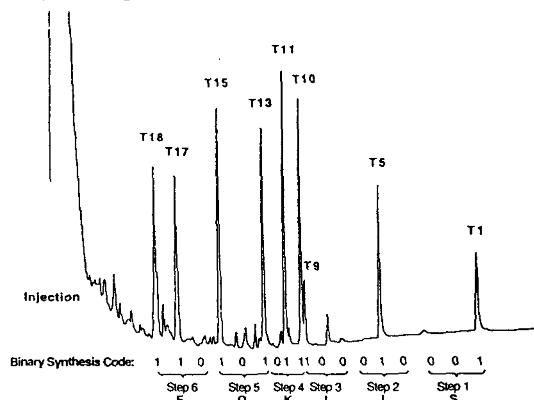
A major advantage of this method is the sensitivity via PCR priming. Thus use of small resin (10 μm, 20 fmol capacity. Less than 5% of this loading capacity was used for oligonucleotides) was possible to build a huge library and this minimized the problems caused by antibody-oligonucleotide binding. Also, large number of the building blocks could be accommodated by using more oligonucleotides. More than 1000 building blocks could be assigned using pentanucleotide (4^5=1024), which may be useful for the diverse library. However, chemical compatibility of the oligonucleotide for the diverse reaction condition is still in question.

Binary encoding method.

Another encoding method that use haloaromatic compounds as binary code was recently developed by Still and Ohlmeyer.[52-54] As codes, inert haloaromatic compounds which are different in their retention time during the GC analysis were introduced as a combinated mixture to represent building blocks. For example, for the peptide library composed of 7 amino acid, combinations of 3 tags T1, T2, T3, T1+T2, T1+T3, T2+T3, T1+T2+T3 were coupled for the first building blocks after activation as their acyl carbonate, followed by the coupling of corresponding amino acid. Combination of T4, T5, ⋯⋯, T4+T5+T6 were used for the 2nd building block. The sequence of a selected bead could be interpreted by the characteristic presence and absence of the tags as shown in the chromatogram (Fig. 8). The chromatogram was obtained from the 50~80 μm single bead selected from hexapeptides library composed of 7 amino acid Glu (E), Gln (Q), Lys (K), Leu (L), Ile (I), and Ser (S) screened with mAb 9E10. With this method, small number of codes were used for large library and in general, with X.N number of tag molecules, $(2^X-1)^N$ member library can be assigned. In the above mentioned library, 18 codes (3·6) tag molecules were successfully used for the above 117,649 [i.e., $(2^3-1)^6$] hexapeptide library. The presence and absence of each code can be denoted as binary code [0 and 1] and identity of a sequence can be represented as sequences of binary code.

The tags were prepared from halophenols and bromoalkanol with various chain length as shown in Fig. 8 (b). For differentiation of tagging molecule's retention time, chain length and aromatic units were adjusted. Employment of haloaromatic unit is also to increase the sensitivity to electron-capture (EC) detection.[55] The library built on 50~80 μm bead coupled with 1% binary codes were detected without significant problem. For the detection, the codes were cleaved from the resin as a haloaromatic alcohols by 3 h UV (366 nm) irradiation followed by heating (2 h, 90℃) and they were silylated with bis(trimethylsilyl)acetamide and analyzed by gas chromatography equipped with EC.

Recently a method that attach the codes to the resin's aromatic region using the acyl carbene/benzene coupling reaction was developed.[55]



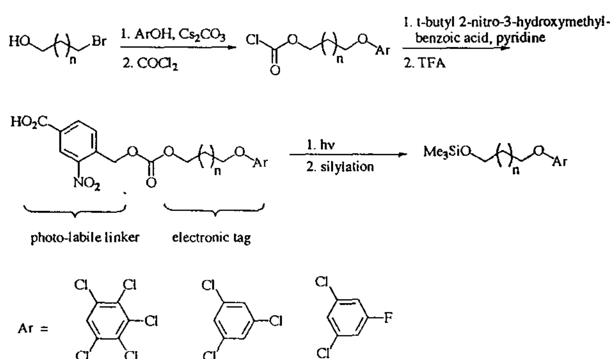


Fig. 8. (a) An example of single bead analysis using molecular tag. Each amino acid was assigned with the combination of molecular tags and reading of this tags revealed the corresponding sequence. Reprinted with permission. (b) Molecular tags were prepared from halophenols and bromoalcohols with various chain length. Tag molecules were decomposed to its alcohol by irradiation of UV and silylated for the EC-GC analysis.

This method does not couple the tags to the screening chain but to the benzene ring of the polystyrene resin bead by acyl carbene/ arene coupling reaction in the presence of catalytic amount $Rh_2(OAc)_4$ or $Rh_2(O_2CCF_3)_4$. This method avoid the employment of the photolabile o-nitrobenzyl fuctional group and weak carbonate functional group of the aforementioned tags, thus, increase the insensitivity of the tagging molecule to the general organic reaction condition. The carbene insertion reaction and carbene/arene coupling reaction of the tag to the screening chain is not significant due to the large excess of the polymer bound benzene moiety to the screening chain and minute amount of the tagging molecules employed. In a typical combinatorial synthesis using 25 tags, overall tag induced destruction of the screening chain was expected <5%. The tags were released by sonication in the presence of ceric ammonium nitrate and silylated with bis(trimethylsilyl)acetamide for subjection for EC-GC. The yield of oxidative cleavage of the tagging molecule was reported >90% yield.

The binary encoding method was also successfully used for the receptor binding study of the dye-coupled synthetic tyrosyl receptor to the library composed of D and L-amino acids which are capped with various group.[57,58]

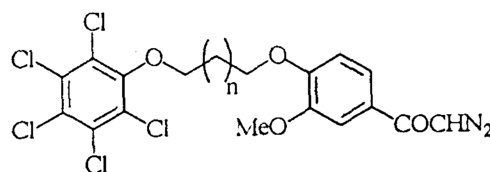


Fig. 9. Molecular tags that can be coupled to polystyrene solid supports.

Spatially Addressable Biooligomer Sequencing Method.

Another revolutionary method that use the spatial location of a biooligomer, built on a glass plate was also developed.[59-61] That is, for the assignment of oligomer sequences active to a fluorescein conjugated monoclonal antibody (mAb), each different oligomers were constructed in a different sector of a small glass plate using photolabile nitroveratryloxycarbonyl (NVOC) group protected amino acid and photolithography to make a library that looks like a checker board (Fig. 10, 11).

The glass surface was first treated with the NVOC protected 3-(aminopropyl)-triethoxy silane and NVOC group of this spacer in a selected area was deprotected using 365 nm UV light while the rest of the area was masked from the light source. The deprotected spacer was coupled with NVOC protected amino acid by dipping the plate in a solution containing peptide and coupling reagent. This process was repeated with the masked area until all the area was completely coupled with the first building block. Then 2nd building blocks were coupled in the orthogonal direction to generate the combinated peptide sequence. The size of a sector can be as small as 20 $\mu m \times 20$ μm and library of >250,000/cm^2 peptide was constructed on a glass plate. Using electron-beam lithography, 10^{10} peptide library was generated on 1 cm^2 glass plate. For the binding study and sequencing, the library plate is treated with a fluorescein bound antibody and the sectors bound with it are brightly colored under UV. The sequences of the sectors are revealed by the location in a glass plate and thus, sequencing is almost instantaneous and no special treatment or analytical tools are necessary. This is very important advantage of this method and this method avoid the problem associated with the encoding method.

This method was successfully applied also for the construction of a carbamate library whose building blocks are amino alcohols derived from the natural amino acid.[62] For the carbamate bond construction, solid phase bound amine functional group was reacted with p-nitrophenyl chloroformate monomer prepared from the reaction of NVOC protected amino alcohol and the p-nitrophenyl

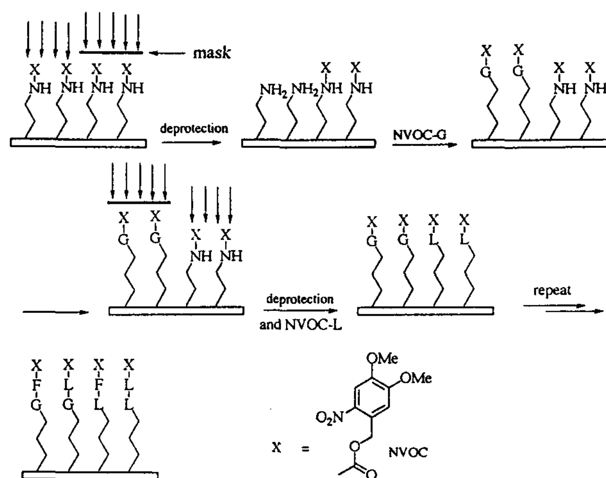


Fig. 10. The generation of a library using lithography. Half of the glass bound spacer was masked during the deprotection of photolabile protecting group (NVOC, denoted as X) and an amino acid building block was coupled to the deprotected amine group using standard peptide chemistry. Then the masked area was deprotected and coupled with different building block. These steps were repeated orthogonally to generate a desired library. Carbamate and oligonucleotide libraries were also generated using this methodology. (see text)

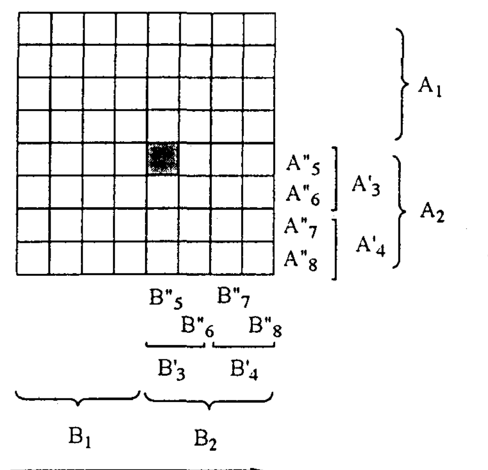


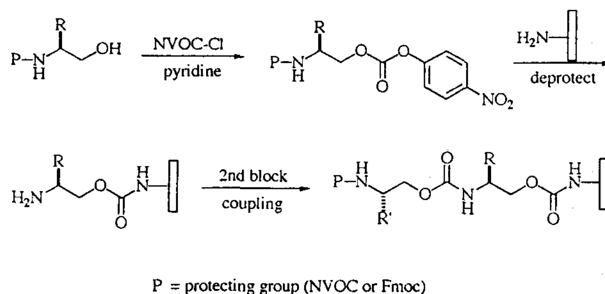
Fig. 11. Top view model of hexapeptide library generated on a glass plate. It is composed of 2 amino acids building blocks at each step and the combinations are 64. (2^6) Each checkerboard sector has different combinations. For the library construction, half of the plate (A_2 row in the figure) is masked, and photolabile group is deprotected by UV. Photolabile group only in A_1 row is deprotected and this area is coupled with a first building unit (named as A_1). With the masking of A_2 area, same process is applied to dope all the area (first building block is completed) . Then, 2nd building blocks are introduced orthogonally using the same process. By this cycle, library of 4 dipeptides, B_1A_1, B_2A_1, B_1A_2, A_2B_2, is generated. For the smaller sector, thin strips A'_1 (not written), ..., A'_4 are made by co-deprotection and co-coupling of A'_2 and A'_4 strips with the masking of A'_1 and A'_3 strips (vice versa) and orthogonal strips are followed. At this cycle, formation of strip B_1 generate 4 combinations. This cycle is repeated until the desired size library is obtained. For the shaded area, the sequence is $B''_5A_5''B'_3A'_3B_2A_2$ (from surface to bottom) and in the presence of fluorescein coupled antibody, the matching sector will be brightly colored under UV.

chloroformate in the presence of pyridine. (Fig. 12)

This method was further applied for the synthesis of short oligonucleotide library[62] to sequence some short oligonucleotide using the hybridization between the complementary oligonucleotides. Oligonucleotide library was built under same principle using photolabile group protected 5'-O'-(a-methyl-6-nitro-piperonyloxycarbonyl)-N-acyl-2'-deoxynucleoside phosphoramidites (MeNPoc-N-acyl-2'-deoxynucleoside phosphoramidites). (Fig. 13)

MeNPoc- protecting group was used to insure the similar deprotection time between the nucleotides and using 365 nm light (14.5 mW), the deprotection was 99% complete in 4.5 min. However, photolithographic synthesis provided oligonucleotide with diminished yield and quality[64] and in the case of oligonucleotide CTTT, average yield of MeNPoc based yield was 92.7%, while the traditional DMTr based yield was 97.9%. This was not derived from thymine dimerization,[65] and only partially attributable to the benzoylcytidine degradation, which could be fixed by changing protecting groups.

For the sequencing, short fluorescein attached oligonucleotide to be sequenced is treated to the surface of the oligonucleotide library and as it binds to its complementary sequence most tightly, the location of its complementary sequence is revealed by checking the intensity of the fluorescence with epifluorescence microscopy. The sequences of the complementary nucleotide are obtained from its location and sequence of the target oligonucleotide is obtained from this. However, as other mismatching complementary nucleotide also bind to the fluorescein attached oligonucleotide, care



P = protecting group (NVOC or Fmoc)

Fig. 12. Generation of an oligocarbamates on solid supports.

should be taken to conclude the sequence and perhaps, this would limit the sequencing of the long chain. In the case of octanucleotide 5'-GCGGCGGC-fluorescein, 5-35 times stronger fluorescence signal was obtained than those with single or double base-pair hybridization mismatches. On the other hand, this behavior may act as a useful technique for the binding study of the oligonucleotide.

Radio Frequency Encoding Method

Very recently, methods that use small radio frequency (RF) recording chip were developed.[66,67] Similar to the

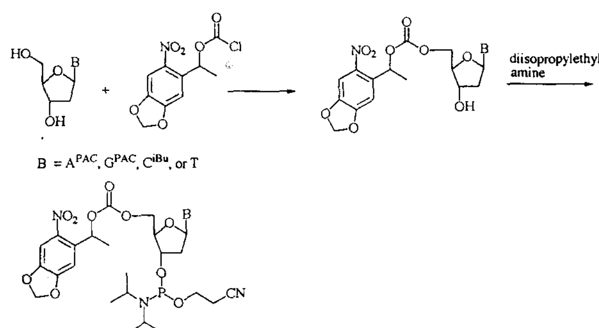


Fig. 13. Synthesis of 5'-O'-MeNPoc-N-acyl-2'-deoxynucleotide for the application of the photolithographic method to the synthesis of oligonucleotide library. This protecting group is cleaved at 365 nm and thus was chosen to avoid side reactions originated from the π-π^* transition of the nucleotide bases by UV (280 nm) during the photolysis.

"tea bag approach", which seal resins inside the porous polypropylene bags for the individual peptide coupling reaction and common washing and deprotection reaction, few mmol of resins were included in a porous polypropylene capsule but with a small radio frequency recordable chip coated with glass for the information storage. The information is stored or retrieved using radio frequency and libraries are built using split synthesis with the corresponding information storage of the reaction.

This method has several advantages over the method previously discussed. Firstly, retrieval of sequencing information by this method is unambiguous and instantaneous. 100% retrieval ratio were reported for both 125 and 24 member peptide library. Secondly, it can record very big and diverse information such as reaction temperature, pH, reaction time, etc. Thirdly, the non invasive transmission and retrieval of information from any capsule is compatible virtually with all synthetic methods and thus, almost any chemical reaction is compatible with this method. Also it does not require excess labor or special tools.

However, it is not clear how to determine the completion of the chemical reaction with this method and presumably this is why these methods are limited to the synthesis of peptide library so far.

Acknowledgments

We greatly thanks to professor Jung Jin (Department of Agricultural Chemistry, Seoul National University) for the helpful discussion for this review and professor W. Clark Still (Department of Chemistry, Columbia University) for the permission and supply of EC-GC chromatogram for the binary encoding method.

References

1. Gallop, M. A., R. W. Barrett, W. J. Dower, Fodor, S. P. A. Fodor, and E. M. Gordon (1994) Application of Combinatorial Technologies to Drug Discovery. 1. Background and Peptide Combinatorial Libraries. *J. Med. Chem.* **37**, 1233-1251.
2. Gordon, E. M., R. W. Barrett, W. J. Dower, S. P. A. Fodor, and M. A. Gallop (1994) Applications of Combinatorial Technologies to Drug Discovery. 2. Combinatorial Organic Synthesis, Library Screening Strategies and Future Directions. (1994) *J. Med. Chem.* **37**, 1385-1401.
3. Jung, G ther and S. G. Beck-Sickinger (1992) Multiple Peptide Synthesis and Their Application. *Angew. Chemie. Int. Ed. Engl.* **31**, 367-486.
4. Merrifield, R. B. (1963) Solid Phase Peptide Synthesis *J. Am. Chem. Soc.* **85**, 2149-2152.
5. Frühtel, J. S. and G. Jung (1996) Organic Chemistry on Solid Supports. *Angew. Chem. Int. Ed. Engl.* **35**, 17-42.
6. Lam, K. S., V. J. Hruby, M. Lebl, R. J. Knapp, W. M. Kazmierski, E. M. Hersh, and S. E. Salmon (1993) The Chemical Synthesis of Large Random Peptide Libraries and Their Use for the Discovery of Ligands for Macromolecular Acceptors. *Bioorg. Med. Chem. Lett.* **3**, 419-424.
7. Lam, K. S., S. E. Salmon, E. M. Hersh, V. J. Hruby, M. Lebl, W. M. Kazmierski, and R. J. Knapp (1991) A New Type of Synthetic Peptide Library for Identifying Ligand-Binding Activity. *Nature* **354**, 82-84.
8. Chen, J. K., W. S. Lane, A. W. Brauer, A. Tanaka, and S. L. Schreiber (1993) Biased Combi- natorial Libraries: Novel Ligands for the SH3 Domain of Phosphatidylinositol 3-Kinase. *J. Am. Chem. Soc.* **115**, 12591-12592.
9. Houghten, R. A., C. Pinilla, S. E. Blondelle, J. R. Appel, C. T. Dooley, and J. H. Cuervo (1991) Generation and Use of Synthetic Peptide Combinatorial Libraries for Basic Research and Drug Discovery. *Nature* **354**, 84-86.
10. Geysen, H. M. and T. J. Mason (1993) Screening Chemically Synthesized Peptide Libraries for Biologically Relevant Molecules. *Bioorg. Med. Chem. Lett.* **3**, 397-404.
11. Geysen, H. M., R. H. Meloen, and S. J. Barteling (1984) Use of Peptide Synthesis to Prove Viral Antigens for Epitopes to a Resolution of a single Amino Acid. *Proc. Natl. Acad Sci. USA* **81**, 3998-4002.
12. Borman, S. (1996) Combinatorial Chemistry. *Chemical & Engineering News* **February 12**, 28-54.
13. Liskamp, R. M. J. (1994) Opportunities for New Chemical Libraries: Unnatural Biopolymers and Diversomers. *Angew. Chem. Int. Ed. Engl.* **33**, 633-6.
14. Dür, H., M. Goodman, and G. Jung (1992) Retro-Inverso Amide Bonds between Trifunctional Amino Acids. *Angew. Chem. Int. Ed. Engl.* **31**, 785-787.
15. Goodman, M. and M. Chorev (1979) On the Concept of Linear Modified Retro-Peptide Structures. *Acc. Chem. Res.* **12**, 1-7. (c)
16. Cushman, M., J. Jurayj, and J. D. Moyer (1990) Synthesis,

Biological Testing, and Stereochemical Assignment of an End Group Modified Retro-Inverso Bombesin C-Terminal Nonapeptide. *J. Org. Chem.* **55**, 3186-3194.

17. Gante, J. (1994) Peptidomimetics-Tailored Enzyme Inhibitors. *Angew. Chem. Int. Ed. Engl.* **33**, 1699-1720.
18. Zuckermann, R. N., J. M. Kerr, S. B. H Kent, and W. H. Moos (1992) Efficient Method for the Preparation of Peptoids [Oligo(N-Substituted Glycines)] by Submonomer Solid-Phase Synthesis. *J. Am. Chem. Soc.* **114**, 10646-7.
19. Simon, R. J., R. S. Kania, R. N. Zuckermann, V. D. Huebner, D. A. Jewell, S. Banville, S. Ng, L. Wang, S. Rosenberg, C. K. Marlowe, D. C. Spellmeyer, R. Tan, A. D. Frankel, D. V. Santi, F. E. Cohen, and P. A. Bartlett (1992) Peptoids: A modular approach to drug discovery. *Proc. Natl. Acad. Sci. USA* **89**, 9367-9371.
20. Zuckermann, R. N., E. J. Martin, D. C. Spellmeyer, G. B. Stauber, K. R. Shoemaker, J. M. Kerr, G. M. Figliozzi, D. A. Goff, M. A. Siani, R. J. Simon, S. C. Banville, E. G. Brown, L. Wang, L. S. Richter, and W. H. Moos (1994) Discovery of Nanomolar Ligands for 7-Transmembrane G-Protein-Coupled Receptors from a Diverse N-(Substituted) Glycine Peptoid Library. *J. Med. Chem.* **37**, 2678-2685.
21. David A. Campbell and J. C. Bermak (1994) Solid-Phase Synthesis of Peptidylphosphonate. *J. Am. Chem. Soc.* **116**, 6039-6040.
22. Cho, C. Y., E. J. Moran, S. R. Cherry, J. C. Stephans, S. P. A. Fodor, C. L. Adams, A. Sundaram, J. W. Jacobs, and P. G. Schultz (1993) An unnatural Biopolymer. *Science* **261**, 1303-1305.
23. Burgess, k., D. S. Linthicum, and H. Shin (1995) Solid Phase Synthesis of Unnatural Biopolymers Containing Repeating Urea Units. *Angew. Chem. Int. Ed. Engl.* **34**, 907-909.
24. Hutchins, S. M. and K. T. Chapman (1994) A General Method for the Solid Phase Synthesis of Urea. *Tetrahedron. Lett.* **35**, 4055-4058.
25. Danishefsky, S. J., K. F. McClure, J. T. Randolph, and R. B. Ruggeri (1993) A Strategy for the Solid Phase Synthesis of Oligosaccharide. *Science* **260**, 1307-1309.
26. Lehninger, A. L., D. L. Nelson, and M. M. Cox (1993) In "Principles of Biochemistry" 2nd Ed. p 149-50. Worth Publishers, NY, U. S. A.
27. Adam S. Inglis (1991) Chemical Procedures for C-Terminal Sequencing of Peptides and Proteins. *Anal. Biochem.* **195**, 183-196.
28. Aebersold, R. H., Nika, H., Pipes, G. D., Wettenhall, R. E. H., Clark, S. M., Hood, L. E., and S. B. H. Kent (1988) In 'Methods in Protein Sequence Analysis' p 79 Springer-Verlag, Berlin.
29. Meyer, H. E., E. Hoffmann-Posorske, A. Donella-Deana, and H. Korte (1991) Sequence Analysis of Phosphotyrosine-Containing Peptides. *Methods Enzymol.* **201**, 206-224.
30. Chait, B. T., R. Wang, R. C. Beavis, and S. B. Kent (1993) Protein Ladder Sequencing. *Science* **262**, 89-92.
31. Chait, B. T. and S. B. H. Kent (1992) Weighing Naked Proteins: Practical, High-Accuracy Mass Measurement of Peptides and Proteins. *Science* **257**, 1885-1894.
32. Beavis, R. C., B. T. Chait, H. S. Creel, M. J. Fournier, T. L. Mason, and D. A. Tirrell (1992) Analysis of Artificial Proteins by Matrix-Assisted Laser Desorption Mass Spectrometry, *J. Am. Chem. Soc.* **114**, 7584-7585.
33. Metzger, J. W. (1994) Ladder Sequencing of Peptides and Proteins-A Combination of Edman Degradation and Mass Spectrometry. *Angew. Chem. Int. Ed. Engl.* **33**, 723-725.
34. Metzger, J. W., K. H. Wiesm ler, V. Gnau, J. Br jes, and G. Jung (1993) Ion-Spray Mass Spectrometry and High-Performance Liquid Chromatography-Mass Spectrometry of Synthetic Peptide Libraries. *Angew. Chem. Int. Ed. Engl.* **32**, 894-896.
35. Youngquist, R. S., G. R. Fuentes, M. P. Lacey, and T. Keough (1994) Matrix-assisted Laser Desorption ionization for Rapid Determination of the Sequences of Biologically Active Peptides Isolated from Support-bound Combinatorial Peptide libraries. *Rapid. Comm. Mass Spec.* **8**, 77-81.
36. Zhu, Lin., G. R. Parr, M. C. Fitzgerald, C. M. Nelson and L. M. Smith (1995) Oligodeoxynucleo-tide Fragmentation in MALDI/TOF Mass Spectrometry Using 355-nm Radiation. *J. Am. Chem. Soc.* **117**, 6048-6056.
37. Little, D. P., R. A. Chorush, J. P. Speir, M. W. Senko, N. L. Kelleher and F. W. McLafferty (1994) Rapid Sequencing of Nucleotides by High Resolution Mass Spectrometry. *J. Am. Chem. Soc.* **116**, 4893-4897.
38. Little, D. P. and F. W. McLafferty (1995) Sequencing 50-mer DNAs Using Electronspray Tandem Mass Spectrometry and Complementary Fragmentation Methods. *J. Am. Chem. Soc.* **117**, 6783-6784.
39. Fitzgerald, M. C., K. Harris, C. G. Shevlin and G. Siuzdak (1996) Direct Characterization of Solid Phase Resin-Bound Molecules By Mass Spectrometry. *Biorg. Med. Chem. Lett.* **6**, 979-982.
40. Egner, B. J., G. J. Langley, and M. Bradly (1995) Solid Phase Chemistry: Direct monitoring by Matrix-Assisted Laser Desorption/ Ionization Time of Flight Mass Spectrometry. A Tool for Combinatorial Chemistry. *J. Org. Chem.* **60**, 2652-2653.
41. Zambias, R. A., D. A. Boulton, and P. R. Griffin (1994) Microchemical Structural Determination of a Peptoid Covalently Bound to a Polymeric Bead by Matrix-Assisted Laser desorption Ionization Time-of-Flight mass Spectrometry. *Tetrahedron Lett.* **25**, 4283-6.
42. Brummel, C. L., I. N. W. Lee, Y. Zhou, S. J. Benkovic, and N. Winograd (1994) A Mass-Spectro-metric Solution to the Address Problem of Combinatorial Libraries. *Science* **264**, 399-402.
43. Kerr, J. M., S. C. Banville and, R. N. Zuckermann (1993) Encoded Combinatorial Peptides Libraries Containing Non-Natural Amino Acids. *J. Am. Chem. Soc.* **115**, 2529-2531.
44. Nikolaiev, V., A. Stierandova, V. Krchnak, B. Seligmann, K. S. Lam, S. E. Salmon, and M. Lebl (1993) Peptide En-

coding for Structure Determination of Nonsequenceable Polymers Within Libraries Synthesized and tested on Solid Phase Supports. *Pept. Res.* **6**, 161-170.

45. Salmon, S. E., K. S. Lam, M. Lebl, A. Kandora, P. S. Khattri, S, Wade, M. Patek, P. Kocis, V. Krchnak, D. Thorpe, and S. Felder (1993) Discovery of Biologically Active Peptides in Random Libraries: Solution Phase Testing after Staged Orthogonal Release from Resin Beads. *Proc. Natl. Acad. Sci. USA* **90**, 10708-10712.
46. Needels, M. C., D. G. Jones, E. H. Tate, G. L. Heinkel, L. M. Kochersperger, W. J. Dower, R. W. Barrett, and M. A. Gallop (1993) Generation and Screening of an Oligonucleotide Encoded Synthetic Peptide Library. *Proc. Natl. Acad. Sci. USA* **90**, 10700-10704.
47. Brenner, S., and R. A. Lerner (1993) Encoded Combinatorial Library *Natl. Acad. Sci. USA.* **89**, 5181-5183. Nielson, J., S. Brenner, and Janda, K. D. (1993) Synthetic Method for the implementation of Encoded Combinatorial Chemistry. *J. Am. Chem. Soc.* **115**, 9812-9813.
48. Caruthers, M. H. (1991) Chemical Synthesis of DNA and DNA Analogues. *Acc. Chem. Res.* **24**, 278-284.
49. Hayakawa, Y., S. Wakabayashi, H. Kato, and R. Noyori. (1990) The Allylic protection Method in Solid Phase Oligonucleotide Synthesis. An Efficient Preparation of Solid-Anchored DNA Oligomers. *J. Am. Chem. Soc.* **112**, 1691-1696.
50. Reddy, M. P., M. A. Michael, F. Faroqui, and N. S. Girgis (1994) New and Efficient Solid Support for the Synthesis of Nucleic Acids. *Tetrahedron. Lett.* **35**, 5771-5774.
51. Lehmann, C., Y.-Z. Xu, C. Christodoulou, Z.-K. Tan, and M. J. Gait (1989) Solid-Phase Synthesis of Oligonucleotides using 9-fluorenylmethoxy carbonyl (Fmoc) for 5'-hydroxy Protection. *Nucleic Acids Res.* **17**, 2379-2390.
52. Ohlmeyer, M. H. J., R. N. Swanson, L. W. Dillard, J. C. Reader, G. Asouline, R. Kobayashi, M. Wigler and W. C. Still (1993) Complex Synthetic Chemical Libraries Indexed with Molecular Tags. *Proc. Natl. Acad. Sci. USA* **90**, 10922-6.
53. Eckes, P. (1994) Binary Encoding of Compound Libraries *Angew. Chem. Int Ed. Engl.* **33**, 1573-75.
54. Borchardt, A. and W. C. Still (1994) Synthetic Receptor Binding Elucidated with an Encoded Combinatorial Library. *J. Am. Chem. Soc.* **116**, 373-4.
55. Grimsrud, E. P (1992) in. *Detectors for capillary chromatography,* eds. Hill, H. H. & McMinn, D. G. (Wiley, New York), pp 83-107.
56. Nestler, H. P., P. A. Bartlett, and W. S. Still. (1994) A General Method for Molecular Tagging of Encoded Combinatorial Chemistry Libraries. *J. Org. Chem.* **59**, 4723-4724.
57. Borchardt, A. and W. Clark Still (1994) Synthetic Receptor Binding Elucidated with an Encoded Combinatorial Library. *J. Am. Chem. Soc.* **116**, 373-374.
58. Erickson, S. D., J. A. Simon, and W. C. Still (1992) Practical Synthesis of a Highly Enantioselective Receptor for Peptides. *J. Org. Chem.* **58**, 1305-1308.
59. Fodor, S. P. A., J. L. Read, M. C. Pirrung, L. Stryer, A. T. Lu., and D. Solas (1991). Lightdirected, Spatially Addressable Parallel Chemical Synthesis. *Science* **251**, 767-773.
60. Rozsnyai, L. F., D. R. Benson, S. P. A. Fodor, and P. G. Schultz (1992) Photolithographic Immobilization of Biopolymers on Solid Supports. *Ang. Chem. Int. Ed. Engl.* **31**, 759-761.
61. Kiederowski, G. V (1991) Light Directed Parallel Synthesis of Up to 250,000 Different Oligopeptides and Oligonucleotides. *Angew. Chem. Int. Ed. Engl.* **30**, 822-823.
62. Cho, C. Y., E. J. Moran, S. R. Cherry, J. C. Stephans, S. P. A. Fodor, C. L. Adams, A. Sundaram, J. W. Jacobs, and P. G. Schultz (1993) An unnatural Biopolymer. *Science* **261**, 1303.
63. Pease, A. C., D. Solas, E. J. Sullivan, M. T. Cronin, C. P. Holmes, and S. P. A. Fodor (1993) *Proc. Natl. Acad. Sci. USA.* **91**, 5022-5026.
64. Pirrung, M. C. and J.-C. Bradly (1995) Comparison of Methods for Photochemical Phosphoramidite-Based DNA Synthesis. *J. Org. Chem.* **60**, 6270-6276.
65. M. M. Greenberg and J. L. Gilmore (1994) Cleavage of Oligonucleotides from Solid-Phase Supports Using O-Nitrobenzyl Photochemistry. *J. Org. Chem.* **59**, 746-53.
66. Moran, E. J., S. Sarsha, J. F. Cargill, M. M. Shahbaz, A. Lio, A. M. M. Mjalli, and R. W. Armstrong (1995) Radio Frequency Tag Encoded Combinatorial Library Method for the Discovery of Tripeptide-Substituted Cinnamic Acid inhibitors of the Protein Tyrosine Phosphatase TPB 1B. *J. Am. Chem. Soc.* **117**, 10787-10788.
67. Nicolaou, K. C., X. Y. Xiao, A. S. Parandoosh, and M. P. Nova (1995) Radio Frequency Encoded Combinatorial Chemistry. *Angew. Chem. Int. Ed. Engl.* **34**, 2289-2291.