Organic & Biomolecular Chemistry

www.rsc.org/obc

Volume 10 | Number 24 | 28 June 2012 | Pages 4629–4808

ISSN 1477-0520

RSC Publishing

PERSPECTIVE

S. Müller *et al.*

Mixed oligonucleotides for random mutagenesis: best way of making them

1477-0520(2012)10:24;1-D

Organic & Biomolecular Chemistry

Dynamic Article Links
View Online

Cite this: *Org. Biomol. Chem.*, 2012, **10**, 4641

www.rsc.org/obc

PERSPECTIVE

Mixed oligonucleotides for random mutagenesis: best way of making them

Tamil Selvi Arunachalam, Claudia Wichert, Bettina Appel and Sabine Müller*

Received 14th February 2012, Accepted 23rd March 2012
DOI: 10.1039/c2ob25328c

The generation of proteins, especially enzymes, with pre-deliberated, novel properties is a big challenge in the field of protein engineering. This aim, over the years was critically facilitated by newly emerging methods of combinatorial and evolutionary techniques, such as combinatorial gene synthesis followed by functional screening of many structural variants generated in parallel (library). Libraries can be generated by a large number of available methods. Therein the use of mixtures of pre-formed trinucleotide blocks representing codons for the 20 canonical amino acids for oligonucleotide synthesis stands out as allowing fully controlled partial (or total) randomization individually at any number of arbitrarily chosen codon positions of a given gene. This has created substantial demand of fully protected trinucleotide synthons of good reactivity in standard oligonucleotide synthesis. We here review methods for the preparation of oligonucleotide mixtures with a strong focus on codon-specific trinucleotide blocks.

Introduction

Over the past decade, the interest in efficient methods for site-directed mutagenesis has strongly grown. Research in the area of protein functional studies and design as well as in the growing field of white biotechnology require procedures that allow preparation of protein and peptide libraries to be screened for function[1] in order to characterize and/or optimize biomolecules, or even design new enzymes with potential for a number of applications in molecular biology, medicine or industry.[2,3] In many cases, the sequence of a protein is known but sufficient information about the three-dimensional structure and structure–function relation is lacking. Therefore, protein engineering by rational design is a less suitable option. On the contrary, directed evolution based on the preparation of libraries by random mutagenesis techniques has become the method of choice.

Conventional chemical or physical mutagenesis[4,5] and classical error-prone PCR[6,7] belong to the non-recombining methods, which due to the dominant randomness of events are impossible to control, and thus have a rather limited application spectrum. *In vitro* recombination techniques are much better suited to the purpose of controlled randomization, and over the past years many recombining methods like DNA shuffling,[8] StEP (Staggered Extension Process),[9] ITCHY (Incremental Truncation for the Creation of Hybrid Enzymes)[10] and SCRATCHY (ITCHY combined with DNA shuffling)[11] and related strategies were developed for the generation of gene libraries.[12,13] These methods allow for randomization of a pre-defined domain of a gene with the degree and location of randomization being adjustable, although full control over mutagenesis is still rather limited. Oligonucleotide-based methods provide a better possibility to pre-determine and to control randomization and to significantly reduce wild-type background.[14] Here, sophisticated techniques like cassette mutagenesis,[15,16] iterative CASTing (Combinatorial Active site Saturation Test)[17] or SeSaM (Sequence Saturation Mutagenesis),[18] which have been developed with the aim to reduce library size, as well as *in vivo* mutation methods like delitto perfetto[19] or DuARCheM (Dual Approach to Random Chemical Mutagenesis)[20] belong to the ever growing repertoire of techniques in the field of directed evolution for protein engineering.

Generally, in oligonucleotide-directed mutagenesis, subsets of the 20 amino acids are introduced at a defined position of a protein by using chemically synthesized oligonucleotides of mixed composition for initial preparation of the respective gene library. Routinely, the required randomized oligonucleotides are synthesized using a mixture of nucleotides at each step of the synthesis.[21] However, with this strategy it is impossible to avoid incorporation of undesired amino acids and stop codons as well as to construct a desired subset of codons in a defined position. Furthermore, the frequency of mutants produced by this method is biased towards those amino acids that are encoded by redundant codons. To overcome this hurdle, the use of spiked oligonucleotides has been suggested. Alternatively, chemical strategies like resin splitting[22] or the use of trinucleotide synthons[23] for the synthesis of oligonucleotide mixtures of stronger controlled composition were introduced. The host of methods of combinatorial and evolutionary protein engineering that combine random mutagenesis or combinatorial gene synthesis with functional screening or genetic selection has been extensively reviewed previously.[24] Here, we focus on oligonucleotide based-methods

Institut für Biochemie, Ernst Moritz Arndt Universität, Felix Hausdorff Strasse 4, Greifswald, D-17487, Germany

for the preparation of ensembles (libraries) of many structural variants generated in parallel. In particular, we will review the work done in the field of synthesis of trinucleotide synthons representing codons for the 20 canonical amino acids and their use for fully controlled partial or total randomization individually at any number of arbitrarily chosen codon positions of a given gene.

Spiked oligonucleotides

As already mentioned above, the chemical synthesis of fully randomized oligonucleotides includes coupling of a mixture of nucleotides at each step of the synthesis. Therein, the length of the randomized region is variable; synthesis of oligonucleotides up to the lengths of about 100 nucleotides is certainly possible nowadays.[25,26] Moreover, further enlargement of the chemically synthesized fragments containing randomized regions, can be achieved by enzymatic ligation *in vitro*.[27] A sincere drawback of this method however, is the resulting enormous bias in the library, due to the different incorporation frequency of the monomer building blocks caused by their inherent reactivity, making statistical random mutation impossible. Furthermore, the degenerated genetic code causes silent mutations and over- or under-representation of several amino acids (*e.g.* more serine but less methionine). Other problems are the generation of stop codons and possible new open reading frames. Furthermore, a completely randomized library is inconveniently large; because of its size the handling is difficult and the percentage of active mutants is low.

One possibility to reduce the occurrence of stop codons and the library size is to operate with NNS codons instead of NNN codons (N = A, C, G or T; S = G or C).[28] This procedure takes advantage of the third position redundancy in most codons. Thus, by using all four nucleotides at the first two positions but only C or G at the last, only 32 codons instead of 64, but still one stop codon (instead of four) are generated. Variations of this method were applied for generation of combinatorial libraries of oligonucleotides mutated at the codon level. A more ingenious approach is the application of spiked oligonucleotides, allowing for reducing the bias, considering the reactivity of building blocks and the degenerated genetic code. Using spiked oligonucleotides, the mutation frequency can be adjusted and the resulting sequences are closer to the wild-type sequence.

The term 'spiked oligonucleotides' was introduced in 1989 by Hermes *et al.*,[29] but the concept was applied for several years prior.[21,30–32] Spiked oligonucleotides (equivalent terms are doped mixtures or doped libraries) are the product of a synthesis with solutions of building blocks deliberately contaminated with each of the other three building blocks (Fig. 1).

Applying a simple formula, the volume of spiking mixture required for a desired number of average replacements per oligonucleotide can be easily calculated:[33]

$$V_{\text{aliquots of spiking mix}} = \text{error rate} \times 1.33 \times V_{\text{pure building blocks}}$$

The spiking mixture is prepared by combining equal aliquots of the pure solution of monomer building blocks. Then, a defined volume of the spiking mixture, corresponding to the aliquot taken before from the pure solution, is transferred back to each bottle containing the pure solution of monomer building blocks. In the formula above, the character 1.33 is a correction factor required for the calculation, because the spiking mixture is made of all four building blocks instead of just the three needed (error rate = mistakes per oligonucleotide/length of oligonucleotide). Following this formula, a binominal distribution of mistakes per oligonucleotide is observed.[34]

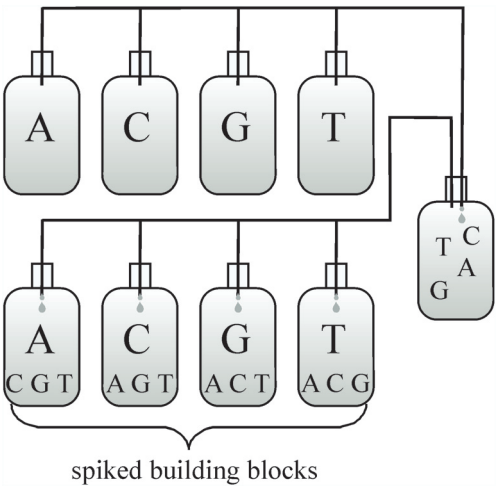

Fig. 1 Scheme of monomer doping for preparation of spiked oligonucleotides.

With the simple scheme described above, the library size can be limited and randomization can be restricted. However, the bias resulting from the different reactivity of the monomer building blocks is not reduced. More sophisticated algorithms enable the inclusion of the nucleotide reactivity in the calculation, and thus to restrict the bias with respect to amino acid distribution. Moreover, it is possible to preferably incorporate or to discriminate specific amino acids.[35,36] Thus, not only the library size is reduced, but also the mutation rate and quality are more stringently directed, which in turn leads to an increase of the number of potentially successful mutants. Tomandl *et al.*[37] have developed a mathematical algorithm that even allows for reverse translation of the amino acid sequence in the composition of spiked building block mixtures. This algorithm was also applied in combination with the method of resin splitting (see below), leading to further improvement of the predictability of the resulting oligonucleotide composition.

Resin splitting

The strategy of resin-splitting for preparation of peptide libraries was first established by Furka *et al.*[38] and Lam *et al.*[39] in 1991. This method was used for the preparation of libraries containing about 10^6 peptides attached to the resin beads. It is also called split and mix or mix and split method and follows the one bead–one compound principle. Later, the methodology was transferred to the synthesis of oligonucleotides.[40] First, the 3′-fixed region of the library is synthesized and placed in various ports on the synthesizer. Each port is treated with the individual nucleotide solution for coupling and chain assembly to $n + 1$. After the

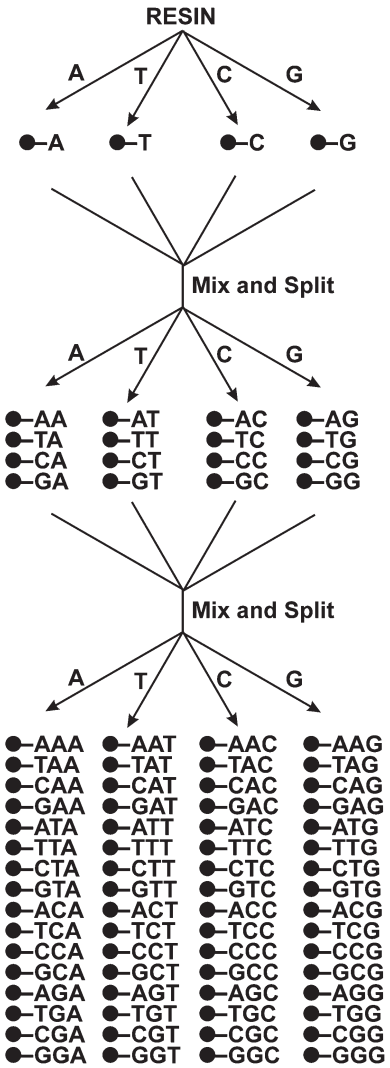

Fig. 2 Synthesis of randomized oligonucleotide sequences by the resin-splitting approach.

coupling cycle has finished, the programme is interrupted and the resin in the columns is dried by passing air through the column. Next, the resin of all columns is combined, followed by splitting to equal portions for the next round of chain assembly. This iterative cycle of mixing and splitting can be continued as long as needed to incorporate the randomized sequence (Fig. 2). Needles *et al.*[40] have mentioned that libraries of greater diversity can be created by reducing bead dimensions, and according to Glaser *et al.*,[41] variation of the ratios of the beads in the columns allows controlling the level of mutation with respect to wild-type sequence. Cormack *et al.*[42] applied the resin-splitting method to a codon-based mutagenesis approach. They used two portions of silica matrix for mutant and wild-type columns. By split and mix they were able to obtain oligonucleotides with defined mutation frequency, which is mainly determined by the ratio of silica beads in the mutant *versus* that in the wild-type column. As a variation of codon-mutagenesis, the NNK or NNS method (where N represents A, G, C or T, K represents T or G and S represents G or C) was developed and used together with wild-type codons as described above.[42] Hence, applying this kind of codon-based mutagenesis, the number of clones required to mutate a pre-defined site with a specific amino acid is lower than in nucleotide-based mutagenesis. Based on this approach, Neuner *et al.*[43] have used dinucleotide phosphoramidite building blocks in resin-splitting. Therein it was shown that the minimum number of seven dinucleotide building blocks is sufficient to encode all 20 natural amino acids. The technology can also be used for the construction of oligonucleotides limited to degenerate codons, and to the construction of complex gene libraries.[43,44] Recently, Yin *et al.*[45] showed the construction of a randomized human scFv antibody library by using randomized oligonucleotides that were prepared with the help of the split and mix approach.

This method has its own merit due to the construction of libraries that are arbitrarily complex with degenerate oligonucleotides. Although the randomization of the codon position can be modulated by varying the ratios of starting materials and/or the reaction products, the drawback of this method lies in the use of several columns and several reactions. Moreover, it is time consuming, tedious, difficult to automate, deleterious to the overall yield and moreover, requires large amounts of support and large amounts of deoxynucleoside phosphoramidites. Due to these drawbacks, the application of the resin-splitting approach for the synthesis of randomized libraries is rather limited.

Trinucleotide synthons

As an alternative to the use of mixtures of monomeric building blocks or resin splitting strategies, it was suggested to use trinucleotide synthons that would allow a codon-based synthesis of randomized oligonucleotides.[23] The beauty of the procedure lies in the fact that the usage of codon-specific trinucleotides permits the synthesis of oligonucleotide mixtures capable of inducing all possible amino acid substitutions across a defined region of a gene in a pre-defined and strictly controlled fashion. Hence, codon redundancy and stop codons are completely eliminated. Furthermore, the method allows for randomization with fewer than 20 amino acids, making possible the choice of specific amino acids out of 20 for randomization. In 1992, Sondek and Shortle introduced a general strategy for random insertion and substitution mutagenesis based on the use of trinucleotide phosphoramidites.[23] Whereas this initial study was still limited to the use of dGCT and dGGT of low quality obtained from a commercial supplier, the first report on the synthesis of a full set of trinucleotides representing codons of all 20 amino acids appeared two years later.[46] Ever since, several methodologies for the synthesis of trinucleotide synthons were developed, varying in the protecting group scheme (Fig. 3) as well as in the preparation strategy.

The key element of trinucleotide synthesis is the definition of a suitable orthogonal protecting scheme for starting materials and intermediate products, allowing selective deblocking of the functionality required for the next reaction step. In addition, all remaining protecting groups in the trinucleotide need to be stable during the final introduction of the phosphoramidite and throughout the following oligonucleotide synthesis. In that, the clever choice of the protecting groups, in particular for the phosphorous and for the 3′-OH group is of significant importance.

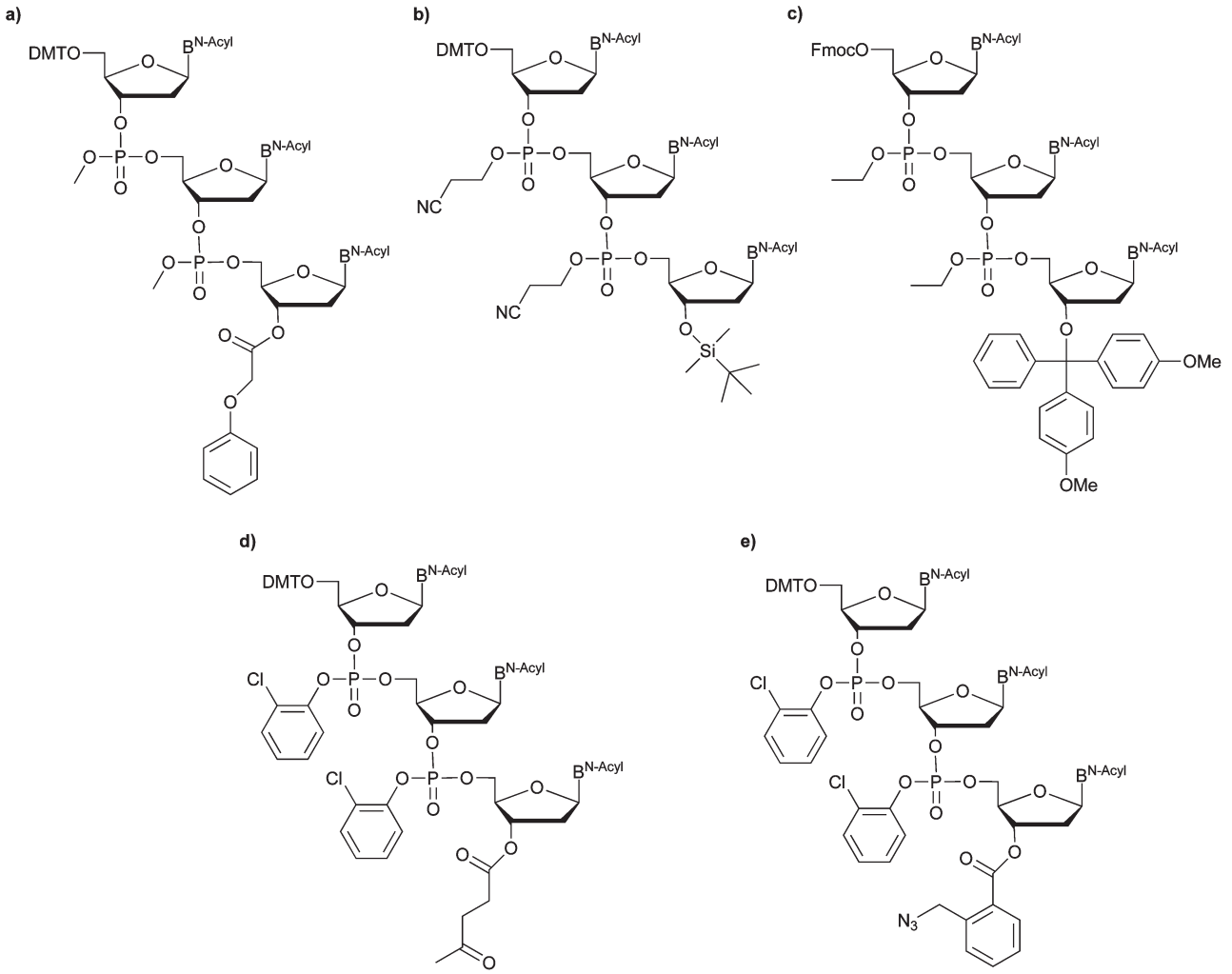

Fig. 3 Structures of trinucleotides with varying protecting groups at 5′- and 3′-OH groups and at the phosphate according to (a) Virnekäs *et al.*,[46] (b) Lyttle *et al.*[48] and Müller *et al.*,[61] (c) Yanez *et al.*,[52] (d) Zehl *et al.*[58] and (e) Yagodkin *et al.*[59]

Trinucleotide synthesis by phosphite triester chemistry

Chronologically, trinucleotide synthesis efforts started with the phosphoramidite procedure, but later switched to phosphotriester chemistry, very likely because of their higher stability. The first described trinucleotide synthesis by Virnekäs and co-workers[46] relies on phosphite triester chemistry, using a methyl group for protection of the phosphite/phosphate moiety and phenoxyacetyl for blocking of the 3′-hydroxyl function (Fig. 3a and 4). An *N*-acyl-3′-*O*-phenoxyacetyl (Pac) protected monomer was coupled to the methyl protected phosphoramidite of an *N*-acyl-5′-*O*-dimethoxytrityl (DMT) protected nucleoside followed by oxidation of the phosphite triester to the phosphate triester. The obtained dinucleotide then was extended in 5′-direction, first removing the 5′-*O*-DMT group followed by coupling to another methyl protected phosphoramidite of an *N*-acyl-5′-*O*-DMT protected nucleoside and oxidation. Finally, the 3′-*O*-Pac group was removed by treatment with NH_3/MeOH, and the phosphoramidite was prepared (Fig. 4). As mentioned above, in particular the orthogonality between phosphate and 3′-OH-protection is very important for successful preparation of trinucleotide synthons being decorated with all protecting groups and functionalities required for the following use in standard oligonucleotide synthesis. In the procedure reported by Virnekäs *et al.*,[46] the 3′-*O*-Pac group had to be removed with a strong nucleophile, conditions that are known to cause also cleavage of phosphotriesters.[47] In order to reduce this side reaction to a minimum, Virnekäs *et al.*[46] replaced the routinely used β-cyanoethyl group at the phosphate by the more stable methyl group. Nevertheless, the final trinucleotides contained a significant amount of by-products, certainly also due to side reactions during treatment with NH_3/MeOH. Moreover, removal of the Pac group from the 3′-OH was troublesome in general, delivering only 75% yield in the best case (TTT), and 47% in the worst case (dGAG). Upon final introduction of the phosphoramidite function, overall yields of trinucleotide synthons ranged from 25 to 40%. It was further reported that the synthesized trinucleotide phosphoramidites can be stored at −20 °C for at least one year without significant decomposition. The phosphoramidites were used as 0.1 M solutions in acetonitrile for coupling reactions; with tetrazole activation, coupling yields were 80–85% at 15 s coupling time. However, extending the coupling time to 1 min and using

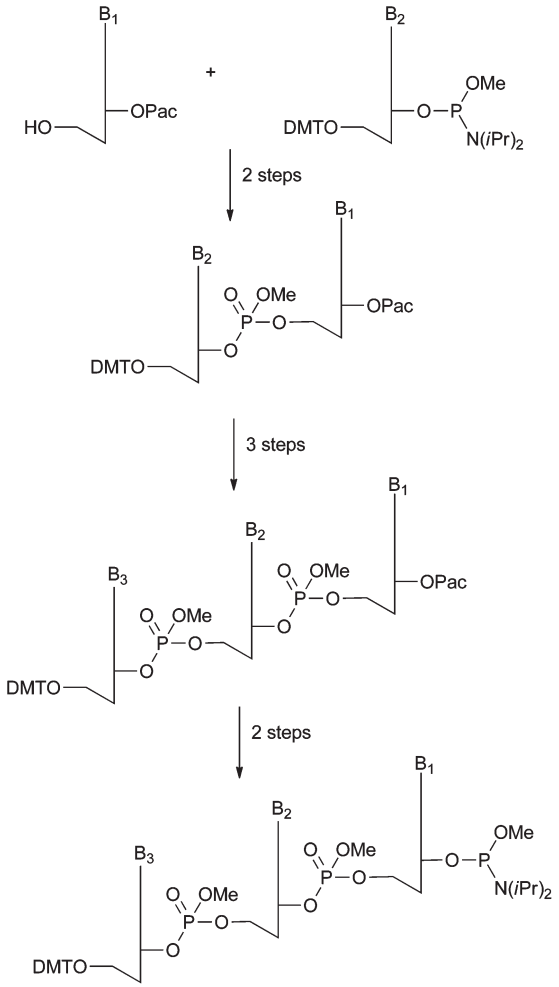

Fig. 4 Synthesis of trinucleotide phosphoramidites following the procedure described by Virnekäs *et al.*[46] with phenoxyacetyl (Pac) for protection of the 3′-OH group and methyl for phosphate protection.

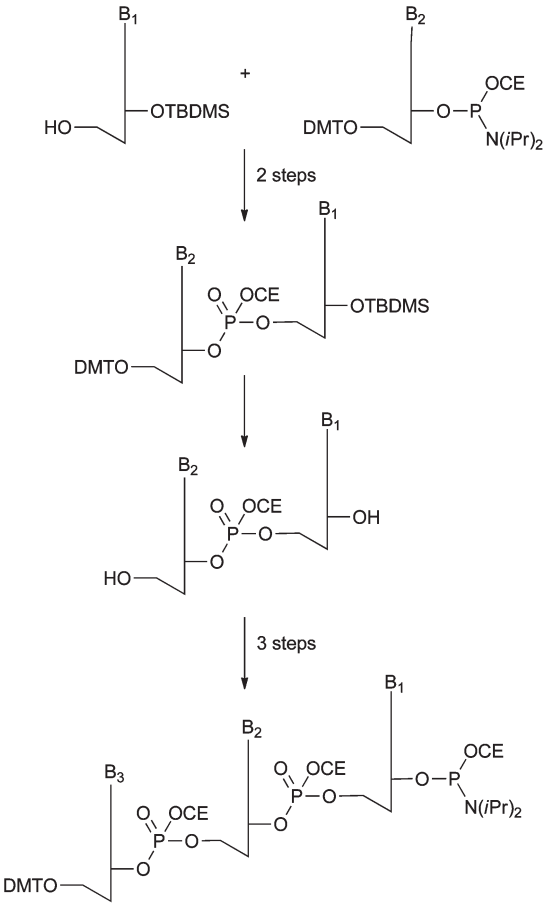

Fig. 5 Synthesis of trinucleotide phosphoramidites following the procedure described by Lyttle *et al.*[48] with *tert*-butyldimethylsilyl (TBDMS) for protection of the 3′-OH group and β-cyanoethyl for phosphate protection.

double-coupling cycles, the coupling yield was increased to 96–98.5%.[46]

The problems associated with the 3′-OH-protecting group, was tried to overcome by the use of *tert*-butyldimethylsilyl (TBDMS) instead of Pac (Fig. 3b).[48] Routinely, the TBDMS group is removed by treatment with fluoride ions. For a long time, tetrabutyl ammonium fluoride (TBAF) has been the reagent of choice for this purpose. However, TBAF was found to also attack phosphotriesters, and therefore it is not suitable to be applied here. Assumingly, for this reason, Lyttle *et al.*[48] removed the 3′-*O*-TBDMS group by treatment with 6 N HCl. In detail, the procedure involved coupling of an *N*-acyl-3′-*O*-TBDMS protected nucleoside to the 3′-*O*-phosphoramidite of an *N*-acyl-5′-*O*-DMT protected nucleoside followed by oxidation of the phosphite triester to a phosphotriester (Fig. 5). Here, β-cyanoethyl was used for protection of the phosphite/phosphate, because it was shown earlier that methyl phosphoramidite coupling chemistry in conjunction with methyl protected internucleotide phosphate linkages leads to methylation of thymidine and guanosine during DNA synthesis.[49,50] However, care must be taken, since the β-cyanoethyl group at the oxidized phosphotriester linkage is rather labile compared with the more stable alkyl groups. The resulting dinucleotide was treated with 6 N HCl leading to simultaneous removal of the 3′-*O*-TBDMS group and the 5′-*O*-DMT group (Fig. 5). This however, is little beneficial to the following step, where the free 5′- and 3′-OH functions compete in the coupling reaction with an activated nucleotide to make the trinucleotide complete. Furthermore, strong acidic conditions are known to cause depurination of nucleosides. For coupling of the dinucleotide to the remaining third nucleoside, Lyttle *et al.*[48] decided to extend the dinucleotide in 5′-direction, trusting on the selectivity of the reaction for the primary rather than for the secondary alcohol. The final trinucleotide phosphoramidite was prepared in three steps: coupling, oxidation and phosphitylation (Fig. 5). Not surprisingly however, the authors mentioned the observation of by-products such as isomeric trimers resulting from 3′–3′-coupling. Nevertheless, following this protocol, phosphoramidites of the five trinucleotides dATA, dCTT, dATC, dATG and dAGC were prepared. Using these synthons for the synthesis of oligonucleotides, the average coupling efficiency was about 71%. The final yield of oligonucleotides was rather low, certainly due to side reactions mentioned above. Moreover, a significant amount of single base insertions was observed.[48]

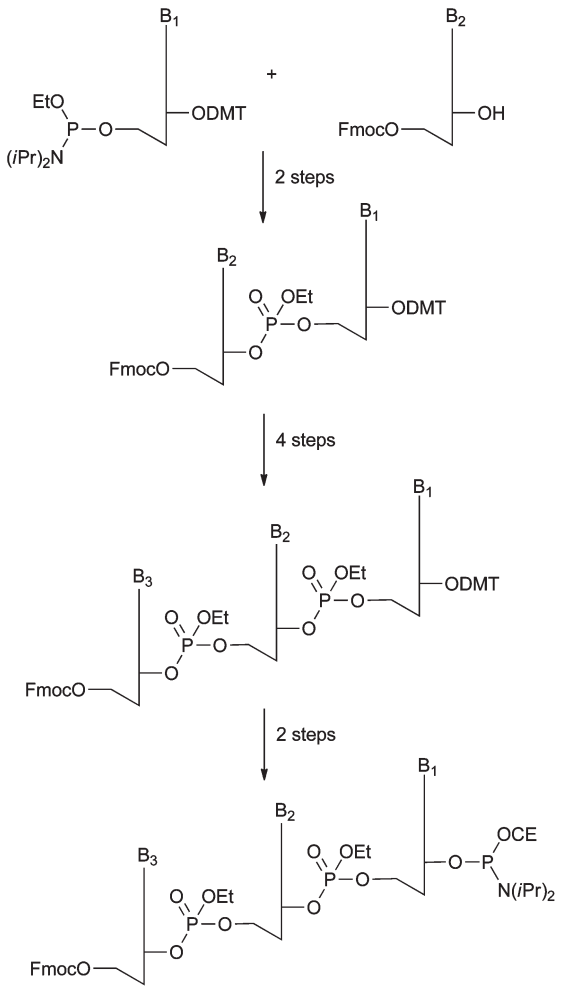

Fig. 6 Synthesis of trinucleotide phosphoramidites following the procedure described by Yanez *et al.*[52] with dimethoxytrityl (DMT) for protection of the 3′-OH group, fluorenyloxymethyl (Fmoc) for protection of the 5′-OH-group and ethyl for phosphate protection.

Gaytán and co-workers tried to solve the protecting group issue by the use of fluorenyloxymethyl (Fmoc) for blocking of the 5′-hydroxyl function, and DMT for 3′-OH protection (Fig. 3c).[51,52] After first attempts with phosphotriester chemistry,[51] they synthesized 5′-*O*-Fmoc protected trinucleotides in a two-step procedure based on phosphoramidite chemistry, although here 5′-*O*-phosphoramidites of nucleosides/nucleotides were coupled to the 3′-OH of 5′-*O*-Fmoc protected nucleosides (Fig. 6). In brief, a 3′-*O*-DMT protected monomer was converted to the 5′-*O*-(*N*,*N*-diisopropylamino-ethyl)-phosphoramidite, followed by coupling to a 5′-*O*-Fmoc protected nucleoside. After oxidation of the internucleotide phosphite, the resulting dinucleotide was treated with triethyl amine to remove the Fmoc group, followed by activation of the set free 5′-OH as phosphoramidite. This was coupled to another 5′-*O*-Fmoc protected nucleoside, after oxidation of the internucleotide phosphite resulting in the 5′-*O*-Fmoc-3′-*O*-DMT protected trinucleotide carrying *N*-acyl groups for protection of the nucleobases and ethyl groups at the internucleotide phosphates. The ethyl group was used for protection of the phosphite because of its higher stability compared to methyl or β-cyanoethyl groups.[53,54] Finally, the 3′-*O*-DMT group was cleaved off by treatment with acid, and the 3′-*O*-phosphoramidite of the protected trinucleotide was prepared. Following this protocol, Gaytán and co-workers synthesized 20 Fmoc protected trinucleotide phosphoramidites.[52] A pool of these synthons was incorporated into oligonucleotides at substoichiometric levels by combination with ordinary 5′-*O*-DMT protected monomers. The distinct nature of the 5′-*O*-protecting group at the trinucleotides required modification of the program for automated assembly of mutagenic oligos, implementing the extra step for removal of the Fmoc group. The pool of randomized oligonucleotides was used for generation of libraries of TEM-1-β-lactamase variants. Sequence analysis of several randomly picked clones showed an uneven representation of codons correlating with the reactivity of the individual trinucleotide synthons, demonstrating the necessity of empirically adjusting the concentration of trimers in a mixture for equal, or more general, for controlled representation of codons.[52]

Trinucleotide synthesis by phosphotriester chemistry

As mentioned above, trinucleotides have been also prepared by phosphotriester couplings. Ono *et al.*[55] synthesized seven dimer precursors following a protocol described by Broka *et al.*[56] These dimers were extended in 3′-direction by converting the 3′-OH group of the dinucleotide to an 2-chlorophenylphosphoric acid ester and coupling to an *N*-acylated nucleoside affording the trimer block (Fig. 7). For the coupling reaction, the authors trusted on the selectivity for primary alcohols, and therefore leaving the 3′-OH group unprotected. Finally, the trinucleotide phosphoramidite was prepared. Four trimer blocks were synthesized following this protocol and then used for preparation of a mixed triplet-amidite solution, which then was applied in oligonucleotide synthesis. A coupling yield of about 90% was reported for the trimers. The authors mentioned weak solubility of the trimer phosphoramidites in CH_3CN, and therefore dissolved the mixed triplet amidites in CH_3CN–DMF (5 : 1, v/v) at a concentration of 0.02–0.03 M. This is in agreement with our own experience, even though previous reports mentioned good solubility of trinucleotides in CH_3CN and their use as 0.1 M solution in the solvent.[46] Possibly, the distinct nature of the phosphate protecting groups applies for the observed solubility differences. Furthermore, Ono *et al.*[55] address the question of synthesis economy regarding the trinucleotide design. In their strategy, they minimized the number of dimer precursors needed for the synthesis of a complete set of triplet-amidite blocks encoding all 20 amino acids. Their approach is based on the fact that the third base of a codon is the most degenerate. Thus, a set of trinucleotides representing anticodons of all 20 amino acids can be designed bearing a fewer number of different bases at the first position and hence require a smaller number of dimers compared to the corresponding codons. The antisense sequences are then converted to codons in template mediated replication. According to this principle, only seven dimer blocks (instead of 16 for codon sequences) were needed to be prepared for the synthesis of 20 triplet-amidite blocks for antisense sequences.

Another route to trinucleotide synthons *via* phosphotriester chemistry has been described by Kayushin *et al.*[57] Here too, trinucleotide synthesis proceeds in 5′–3′-direction, starting from

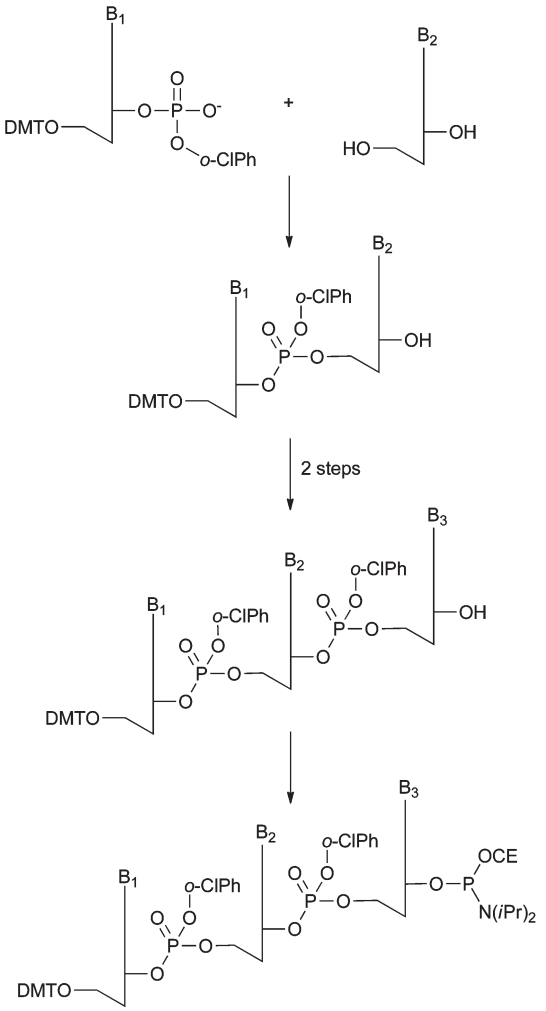

Fig. 7 Synthesis of trinucleotide phosphoramidites following the procedure described by Ono *et al.*[55] and by Kayushin *et al.*[57] with free 3′-OH group and *o*-chlorophenyl (*o*-ClPh) for phosphate protection.

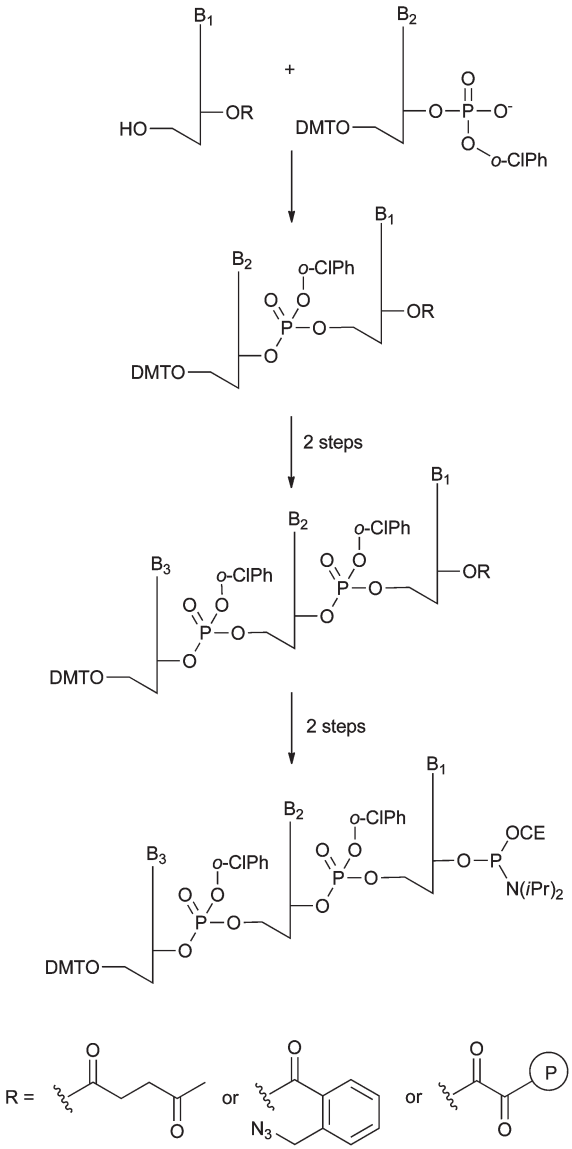

Fig. 8 Synthesis of trinucleotide phosphoramidites following the procedure described by Zehl *et al.*,[58] by Yagodkin *et al.*[59] with levulinoyl or 2-azidomethylbenzoyl for protection of the 3′-OH group and *o*-chlorophenyl (*o*-ClPh) for phosphate protection, and by Kayushin *et al.*[66] for solid phase synthesis of trinucleotide blocks.

5′-*O*-DMT-*N*-acyl-3′-*O*-(*o*-chlorophenylphosphate)nucleosides that were coupled to *N*-acylated nucleosides with unprotected 3′-OH function, again trusting on the selectivity of the reaction for and 3′-*O*-phosphitylation (Fig. 7). Not surprisingly, the authors observed a number of side products; in particular 3′–3′ coupled dimers and trimers, resulting from the non-protected 3′-hydroxyl group. However, Kayushin *et al.*[57] have developed an elaborate analytical protocol in order to reliably confirm the identity and to separate their products. The trinucleotide phosphoramidites were coupled to all of the monomers, for assigning of reaction factors and determination of coupling yields. With double and triple couplings and an extended coupling time of 120 s, yields between 94 and 98% were obtained. As already mentioned above,[52] the individual trinucleotides showed distinct coupling efficiency, and it was once more pointed out that the different reactivity has to be considered for the preparation of trinucleotide mixes in order to reach an even distribution of codons in the library or to adjust the mutation rate.

In order to circumvent the problem of side reactions by the non-protected 3′-hydroxyl, further work in this area has addressed the question of an appropriate blocking group for the 3′-functionality in combination with phosphotriester chemistry and 2-chlorophenyl as phosphate protecting group.[58,59] Zehl *et al.*[58] used the levulinoyl (Lev) group for 3′-*O* protection (Fig. 3d and 8), and synthesized a set of trinucleotides by extension of synthesized dimers in either 5′- or 3′-direction. The 3′-*O*-Lev group was removed by treatment with a 0.5 M solution of hydrazine in pyridine–acetic acid (3 : 5, v/v), the trimers were converted to 3′-*O*-phosphoramidites and tested for their ability to be incorporated in oligonucleotide chains. With coupling times of 15 min and all further conditions as in standard oligonucleotide synthesis, coupling yields of trimers were found being in the range of 66 to 99%, with considerable variation depending on the particular batch. The Lev group was also used in combination with phosphoramidite chemistry to synthesize trimeric

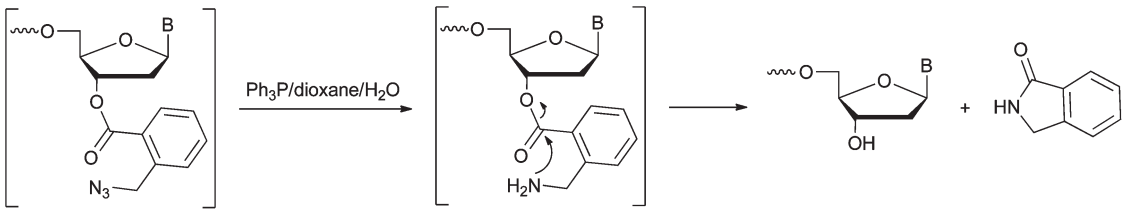

Fig. 9 Cleavage mechanism of the 3′-*O*-(2-azidomethylbenzoyl) group according to ref. 59.

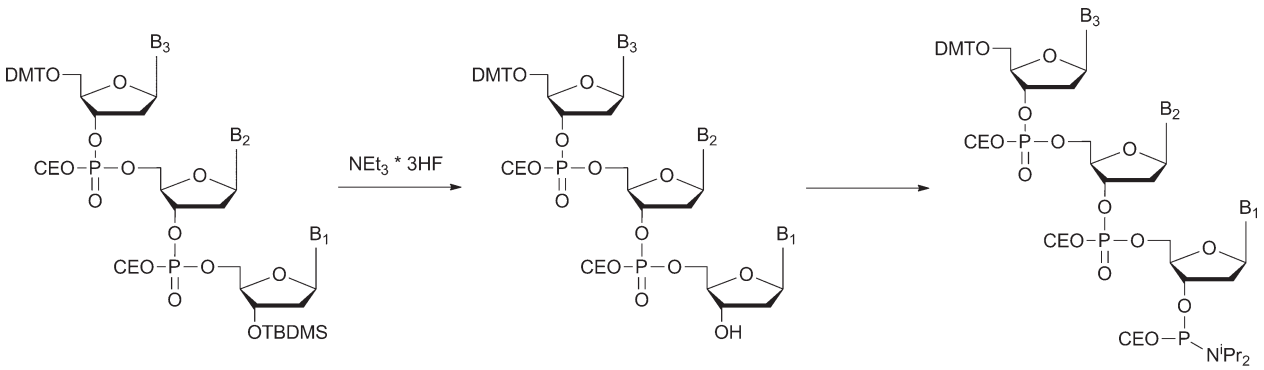

Fig. 10 Synthesis of trinucleotide phosphoramidites following the procedure described by Janczyk *et al.*[61] with *tert*-butyldimethylsilyl (TBDMS) for protection of the 3′-OH group and β-cyanoethyl for phosphate protection.

phosphoramidite synthons that were used for the synthesis of oligodeoxyribonucleotide phosphorothioates with substantial reduction of (*N*−1)-mer content.[60] In 2007, Yagodkin *et al.*[59] published a paper extending their previous work[57] by investigating a number of suitable 3′-*O*-protecting groups for trinucleotide synthesis. In particular, 2,4-dichlorophenoxyacetyl, 4-azidobutanoyl and 2-azidomethylbenzoyl groups were tested. Initial experiments showed that among the three groups, the 2-azidomethylbenzoyl function is most suitable (Fig. 3e). It was removed by reduction with triphenylphosphine in dioxane in the presence of water, thereby first producing 2-aminomethylbenzoic acid ester, which undergoes spontaneous intramolecular cyclization to generate the free 3′-hydroxylgroup (Fig. 9). The synthesis of trimers proceeded *via* phosphotriester chemistry in solution starting with condensation of *N*-acyl-3′-*O*-(*o*-chlorophenylphosphate)nucleosides to 3′-*O*-(*o*-azidomethylbenzoyl) protected nucleoside derivatives (Fig. 8). Upon cleavage of the 5′-*O*-DMT group, dinucleotides were extended to trimers by coupling of another *N*-acyl-3′-*O*-(2-chlorophenylphosphate)nucleosides. The 3′-*O*-(2-azidomethylbenzoyl) group was selectively removed under neutral conditions as explained above (Fig. 9), followed by phosphitylation of the set free 3′-OH group (Fig. 8). A full set of trimers was synthesized and used as 0.1–0.15 M solutions in CH_3CN–CH_2Cl_2 (1 : 3, v/v) in oligonucleotide synthesis. With double and triple couplings and a coupling time of 120 s, coupling yields of 71–98% were obtained.

Re-investigating trinucleotide synthesis by phosphite triester chemistry

Recently, our lab has reinvestigated trinucleotide synthesis by the phosphoramidite approach in solution with DMT and TBDMS as pair of orthogonal protecting groups and β-cyanoethyl for protection of the phosphate (Fig. 3b, 5 and 10).[61] Even though, the β-cyanoethyl group is rather labile, the advantage of trinucleotide synthons with β-cyanoethyl protection at the phosphate is their easy use in oligonucleotide synthesis without additional steps required for removal as it is the case with methyl-, ethyl-, or 2-chlorophenyl groups used in the strategies described above. The use of the TBDMS group previously was hampered by the problems associated with its selective removal.[48] However, we decided to have another look at this problem and to try using the TBDMS group for 3′-*O*-protection. Our synthesis proceeded in 3′–5′-direction, starting with an *N*-acyl-5′-*O*-DMT protected nucleoside-3′-*O*-phosphoramidite that is coupled to an *N*-acyl-3′-*O*-TBDMS protected nucleoside under standard conditions of phosphoramidite chemistry in solution. After oxidation of the internucleotide phosphorous and purification, the DMT group was removed from the 5′-hydroxyl function followed by coupling of the 5′-deprotected dinucleotide to an *N*-acyl-5′-*O*-DMT protected nucleoside phosphoramidite, affording the fully protected trinucleotide. The 3′-*O*-TBDMS group was cleaved off with fluoride ions, and finally the free 3′-OH group was phosphitylated. In contrast to problems reported in the past, we were able to selectively remove the TBDMS group without considerable side-reactions by treatment with triethylamine/3HF, a reagent used in modern RNA chemistry.[62–64] Due to its aggregated structure, triethylamine/3HF presents the fluoride ion as a softer nucleophile compared with the traditionally used reagent TBAF. Thus, TBDMS was cleaved with virtually quantitative yield leaving the protected trinucleotide undamaged. In order to keep the β-cyanoethyl group at the phosphates intact, strict control of the pH in all reaction steps was necessary. Following this protocol, we have synthesized a set of 20 trimer phosphoramidites that were successfully coupled to short test sequences.[61]

Trinucleotide synthesis on solid phase

All methods described above involve trinucleotide synthesis in solution. Alternatively, trinucleotide blocks can be also assembled on solid support using standard phosphoramidite chemistry. However, a key element in trinucleotide synthesis on solid phase is to link the 3′-start nucleoside *via* a suitable functionality that can be cleaved after synthesis without damaging all other protecting groups of the trinucleotide. Hence, the routinely used succinate anchor[65] cannot be used. A possible scenario was suggested by Kayushin *et al.*,[66] who adapted their previous method for trinucleotide preparation to large scale synthesis on solid phase. Using CPG with 70 Å pore size, 120–130 μmol g^{-1} nucleotide was loaded on the support *via* an oxalyl anchor (Fig. 8). After synthesis, the oxalyl anchor was cleaved with a 5% solution of 25% aqueous ammonia in methanol for 20 to 40 min, leaving the phosphotriester linkages and the acyl groups at the nucleobases unharmed. Alternatively, the oxalyl linker can be cleaved with 20% pyridine in methanol over 12 h. The procedure allowed for synthesis of 3′-unprotected trinucleotides at 5 g scale with a total yield of 75–90%. We also have started an effort to synthesize protected trinucleotides on solid phase. Our procedure is based on standard phosphoramidite chemistry in conjunction with a disulphide linkage to connect the 3′-start nucleoside to the polymer. The disulphide bridge can be cleaved under neutral conditions preserving all other protecting groups of the trinucleotide. Our initial results show that trinucleotides of high quality can be synthesized in short time with excellent yields (M. Janczyk, B. Appel, S. Müller, publication in preparation).

Final preparation of gene libraries

For further generation of gene libraries, the prepared randomized oligos usually are used as primers in PCR amplification followed by assembly to the full lengths gene segment and cloning into an appropriate vector.[29,30,33] Over the past years, a host of methods has been developed allowing for efficient cloning, as for example the recently reported one-pot methodology for cassette randomisation and recombination including megaprimer PCR.[3]

For an overview of cloning strategies the reader is referred to references.[67–69] Finally, the generated gene libraries are expressed to protein libraries to be screened for functional mutants.

Conclusion

Among a host of related methods for combinatorial gene synthesis, the use of trinucleotide synthons representing codons for the 20 canonical amino acids stands out as allowing fully controlled partial (or total) randomization individually at any number of arbitrarily chosen codon positions of a given gene. Pools of oligonucleotides with random sequence are synthesized by combination of natural mononucleoside phosphoramidites and trinucleotide phosphoramidites that are coupled at the DNA synthesizer under standard conditions of the DNA phosphoramidite approach with small variations in the coupling protocol for the trimers, such as longer coupling times and multiple couplings. In contrast to first reports that trinucleotide blocks due to their steric bulk are coupled in solid-phase DNA synthesis with yields of less than 5%,[23] the examples reported in this review demonstrate that steric bulk is not a major issue, and that when using slightly longer reaction times and double or triple coupling cycles, trinucleotides perform well in standard DNA synthesis with excellent coupling yields of >98%. Basically, the length of oligonucleotides to be synthesized by trinucleotide couplings is limited by the same criteria as for normal DNA or RNA synthesis. With an average coupling yield of 98%, about 40 to 50 trinucleotide couplings at 1 μmol scale would be theoretically possible. Alternatively, there exists always the possibility to divide the randomized oligonucleotide into two or more fragments that can be efficiently synthesized and then joined together enzymatically.

The development of improved strategies for trinucleotide synthesis goes hand in hand with the ever growing interest in protein studies at the molecular and submolecular level. Therein, trinucleotide synthons and their use for the preparation of controlled randomized libraries will greatly enhance the prospects of analysis of protein structure–function relationship, and moreover, of the generation of proteins, especially enzymes, with pre-deliberated, novel properties.

References

1 C. Jäckel and D. Hilvert, Biocatalysts by evolution, *Curr. Opin. Biotechnol.*, 2010, **21**, 753.
2 K. L. Morley and R. J. Kazlauskas, *Trends Biotechnol.*, 2005, **23**, 231.
3 A. Hidalgo, A. Schließmann, R. Molina, J. Hermoso and U. T. Bornscheuer, *Protein Eng., Des. Sel.*, 2008, **21**, 567.
4 J. M. Chalker and B. G. Davis, *Curr. Opin. Chem. Biol.*, 2010, **14**, 781.
5 A. Kodym and R. Afza, *Methods Mol. Biol.*, 2003, **236**, 189.
6 Y. Zhou, X. Zhang and R. H. Ebright, *Nucleic Acids Res.*, 1991, **19**, 6052.
7 E. O. McCullum, B. A. R. Williams, J. Zhang and J. C. Chaput, *Methods Mol. Biol.*, 2010, **634**, 103.
8 W. P. Stemmer, *Proc. Natl. Acad. Sci. U. S. A.*, 1994, **91**, 10747.
9 H. Zhao, L. Giver, Z. Shao, J. A. Affholter and F. H. Arnold, *Nat. Biotechnol.*, 1998, **16**, 258.
10 S. Lutz, M. Ostermeier, G. L. Moore, C. D. Maranas and S. J. Benkovic, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 11248.
11 H. Zhao, K. Chockalingam and Z. Chen, *Curr. Opin. Biotechnol.*, 2002, **13**, 104.
12 T.-W. Wang, H. Zhu, X.-Y. Ma and D.-Z. Wie, *Mol. Biotechnol.*, 2006, **34**, 55.
13 S. Sen, V. V. Dasu and B. Mandal, *Appl. Biochem. Biotechnol.*, 2007, **143**, 212.
14 T. Lanio and A. Jeltsch, *BioTechniques*, 1998, **25**, 958.
15 A. R. Oliphant, A. L. Nussbaum and K. Struhl, *Gene*, 1986, **44**, 177.
16 A. Martinez, T. Opperman and J. P. Richardson, *J. Mol. Biol.*, 1996, **257**, 895.
17 M. T. Reetz, L. Wang and M. Bocola, *Angew. Chem.*, 2006, **118**, 1258.
18 T. S. Wong, K. L. Tee, B. Hauer and U. Schwaneberg, *Nucleic Acids Res.*, 2004, **32**, e26.
19 F. Storici and M. A. Resnick, *Methods Enzymol.*, 2006, **409**, 329.
20 U. Mohan and U. C. Banerjee, *ChemBioChem*, 2008, **9**, 2238.
21 K. M. Derbyshire, J. J. Salvo and N. D. F. Grindley, *Gene*, 1986, **46**, 145.
22 S. M. Glaser, D. E. Yelton and W. D. Huse, *J. Immunol.*, 1992, **149**, 3903.
23 J. Sondek and D. Shortle, *Proc. Natl. Acad. Sci. U. S. A.*, 1992, **89**, 3581.
24 T. S. Wong, D. Roccatano and U. Schwaneberg, *Environ. Microbiol.*, 2007, **9**, 2645.
25 C. B. Reese, *Org. Biomol. Chem.*, 2005, **3**, 3851.
26 S. L. Beaucage and R. P. Iyer, *Tetrahedron*, 1992, **48**, 2223.
27 A. Xiong, R. Peng, J. Zhunag, J. Liu, F. Gao, J. Chen, Z. Cheng and Q. Yao, *Biotechnol. Adv.*, 2008, **26**, 121.
28 M. Isalan, *Nat. Protoc.*, 2006, **1**, 468.

29 J. D. Hermes, S. M. Parekh, S. C. Blacklow, H. Koster and J. R. Knowles, *Gene*, 1989, **84**, 143.
30 C. A. Hutchison III, S. K. Nordeen, K. Vogt and M. H. Edgell, *Proc. Natl. Acad. Sci. U. S. A.*, 1986, **83**, 710.
31 M. D. Matteucci and H. L. Heyneker, *Nucleic Acids Res.*, 1983, **11**, 3113.
32 I. S. Dunn, R. Cowan and P. A. Jennings, *Protein Eng., Des. Sel.*, 1988, **2**, 283.
33 S. J. Dale, M. Belfield and T. C. Richardson, *Methods*, 1991, **3**, 145.
34 P. Gaytán, J. Yanez, F. Sanchez and X. Soberon, *Nucleic Acids Res.*, 2001, **29**, e9.
35 A. P. Arkin and D. C. Youvan, *Nat. Biotechnol.*, 1992, **10**, 297.
36 L. J. Jensen, K. V. Andersen, A. Svendsen and T. Kretzschman, *Nucleic Acids Res.*, 1998, **26**, 697.
37 D. Tomandl, A. Schober and A. Schwienhorst, *J. Comput.-Aided Mol. Des.*, 1997, **11**, 29.
38 A. Furka, F. Sebestyen, E. M. Asgedom and G. Dibo, *Int. J. Pept. Protein Res.*, 1991, **37**, 487.
39 K. S. Lam, S. E. Salmon, E. M. Hersh, V. J. Hruby, W. M. Kazmierski and R. J. Knapp, *Nature*, 1991, **354**, 82.
40 M. C. Needles, D. G. Jones, E. H. Tate, G. L. Heinkel, L. M. Kochersperger, W. J. Dower, R. W. Barrett and M. A. Gallop, *Proc. Natl. Acad. Sci. U. S. A.*, 1993, **90**, 10700.
41 S. M. Glaser, D. E. Yelton and W. D. Huse, *J. Immunol.*, 1992, **149**, 3903.
42 B. P. Cormack and K. Struhl, *Science*, 1993, **262**, 244.
43 P. Neuner, R. Cortese and P. Monaci, *Nucleic Acids Res.*, 1998, **26**, 1223.
44 S. J. Lahr, A. Broadwater, C. W. Carter Jr., M. L. Collier, J. C. Waldner, G. J. Pielak and M. H. Edgell, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 14860.
45 C.-C. Yin, L.-L. Ren, L. L. Zhu, X.-B. Wang, Z. Zhang, H.-L. Huang and X.-Y. Yan, *J. Biochem.*, 2008, **144**, 591.
46 B. Virnekäs, L. Ge, A. Plückthun, K. C. Schneider, G. Wellnhofer and S. E. Moroney, *Nucleic Acids Res.*, 1994, **22**, 5600.
47 *Oligonucleotide Synthesis: A Practical Approach*, ed. M. J. Gait, IRL Press, Oxford, 1984, p. 13.
48 M. H. Lyttle, E. W. Napolitano, B. L. Calio and L. M. Kauvar, *BioTechniques*, 1995, **19**, 274.
49 X. Gao, B. L. Gaffney, M. Senior, R. R. Riddel and R. A. Jones, *Nucleic Acids Res.*, 1985, **13**, 573.
50 A. T. Yerung, W. J. Dinehart and B. K. Jones, *Nucleic Acids Res.*, 1988, **16**, 4539.
51 P. Gaytán, J. Yanez, F. Sanchez, H. Mackie and X. Soberón, *Chem. Biol.*, 1998, **5**, 519.
52 J. Yanez, M. Argüello, J. Osuna, X. Soberón and P. Gaytán, *Nucleic Acids Res.*, 2004, **32**, e158.
53 K. A. Gallo, K. L. Shao, L. R. Phillips, J. B. Regan, M. Koziolkiewicz, B. Uznanski, W. J. Stec and G. Zon, *Nucleic Acids Res.*, 1986, **14**, 7405.
54 N. D. Sinha, J. Biernat, J. McManus and H. Koster, *Nucleic Acids Res.*, 1984, **12**, 4539.
55 A. Ono, A. Matsuda, J. Zhao and D. Santi, *Nucleic Acids Res.*, 1995, **23**, 4677.
56 C. Broka, T. Hozumi, R. Arentzen and K. Itakura, *Nucleic Acids Res.*, 1980, **8**, 5461.
57 A. L. Kayushin, M. D. Korosteleva, A. I. Miroshnikov, W. Kosch, D. Zubov and N. Piel, *Nucleic Acids Res.*, 1996, **24**, 3748.
58 A. Zehl, A. Starke, T. Hartsch, R. Merkl and H.-J. Fritz, *Chem. Commun.*, 1996, 2677.
59 A. Yagodkin, A. Azhayev, J. Roivainen, M. Antopolsky, A. Kayushin, M. Korosteleva, A. Miroshnikov, J. Randolph and H. Mackie, *Nucleosides, Nucleotides Nucleic Acids*, 2007, **26**, 473.
60 A. Eleuteri, D. C. Capaldi, D. L. Cole and V. T. Ravikumar, *Nucleosides Nucleotides*, 1999, **18**, 475.
61 M. Janczyk, B. Appel, D. Springstubbe, H.-J. Fritz and S. Müller, *Org. Biomol. Chem.*, 2012, **10**, 1510.
62 D. Gasparutto, T. Livache, H. Bazin, A. M. Duplaa, A. Guy, A. Khorlin, D. Molko, A. Roget and R. Téoule, *Nucleic Acids Res.*, 1992, **20**, 5159.
63 E. Westman and R. Strömberg, *Nucleic Acids Res.*, 1994, **22**, 2430.
64 S. Müller, J. Wolf and S. A. Ivanov, *Curr. Org. Synth.*, 2004, **1**, 293.
65 S. P. Adams, K. S. Kavka, E. J. Wykes, S. B. Holder and G. R. Galuppi, *J. Am. Chem. Soc.*, 1983, **105**, 661.
66 A. Kayushin, M. Korosteleva and A. Miroshnikov, *Nucleosides, Nucleotides Nucleic Acids*, 2000, **19**, 1967.
67 J. W. Zyskind, in *Encyclopedia of Microbiology*, ed. J. Lederberg, Academic Press, San Diego, 2000, vol. 4, p. 55.
68 R. Levis, G. L. Costa and M. P. Weiner, in *PCR Primer – A Laboratory Manual*, ed. C. W. Dieffenbach and G. S. Dveksler, Cold Spring Harbor Laboratory Press, Boca Raton, 1995, p. 537.
69 M. Dani, in *Combinatorial Chemistry and Technologies*, ed. S. Miertus and G. Fassina, CRC Press, New York, 2005, p. 41.

