RESEARCH FRONT

Highlight

CSIRO PUBLISHING

Aust. J. Chem. **2009**, *62*, 1328–1332 www.publish.csiro.au/journals/ajc

Synthesis and Applications of Covalent Protein-DNA Conjugates

Patrick M. Schaeffer[A,C] *and Nicholas E. Dixon*[B]

[A]Comparative Genomics Center, School of Pharmacy and Molecular Sciences, James Cook University, Townsville, Qld 4811, Australia.
[B]School of Chemistry, University of Wollongong, NSW 2522, Australia.
[C]Corresponding author. Email: patrick.schaeffer@jcu.edu.au

Synthetic protein-DNA conjugates are valuable tools with applications in fields including nanobiotechnology, bioanalytical chemistry, and molecular diagnostics, and various synthetic methods for their production have been developed during the past three decades. The present article reviews current methodologies for the synthesis of covalent protein-DNA conjugates with particular focus on the regiospecificity and stoichiometry of these reactions.

Manuscript received: 18 March 2009.
Manuscript accepted: 4 May 2009.

Introduction

Synthetic DNA-protein conjugates generated either by covalent or non-covalent coupling chemistry have recently attracted attention as tools for nanobiotechnology, bioanalytical chemistry, and molecular diagnostics. Their potential applications as biosensors, artificial nucleases, and in the design of protein microarrays through DNA-directed immobilization have recently been reviewed.[1] Moreover, double-stranded DNA has numerous attractive features for the bottom-up design of diverse supramolecular building blocks for the production of nanostructures, i.e. its rigidity, predictable structure, and assembly through complementary hybridization. Examples of nanostructures include cubes,[2] tetrahedra,[3] octahedra,[4–6] and 2D arrays,[6–9] with the most impressive being the scaffolded DNA origamis.[10]

The introduction of proteins into DNA nanostructures can afford new functions and functional groups at their surfaces. The inherent properties of proteins, e.g. recognition elements and catalytic activities, can be exploited by their introduction at defined positions. A major current area of application of protein-DNA conjugates recently reviewed by Niemeyer[1] is in the production of DNA-labelled antibodies for immunodiagnostics. Such devices offer ways to increase the sensitivity of immunoassays by several orders of magnitude using the power of polymerase chain reaction (PCR) as a means for detection.[11] Protein-DNA conjugates have also been used in proximity ligation assays,[12–14] in immuno-PCR detection methods,[15–17] to organize biocatalysts in series for sequential reactions to study proximity effects,[18,19] and for the reassembly of a split luciferase directed through DNA hybridization.[20] While these conjugates are useful at present, they will be amenable to many additional applications once their synthesis is made more routine.

Covalently linked protein-DNA conjugates are of particular interest as they do not dissociate over time when diluted or faced with inappropriate buffer or temperature conditions. Covalent coupling is a *condicio sine qua non* for multiplex applications (e.g. simultaneous detection of multiple analytes using different protein-DNA conjugates). Methods for covalent attachment of proteins to nucleic acids include chemical cross-linking of oligonucleotides to protein lysine or cysteine residues,[21–23] expressed protein-ligation,[16,24] chemoenzymatic methods,[25–28] and the use of photoaptamers.[29] Although outside the scope of this review, it is worth mentioning that the great majority of non-covalent protein-DNA conjugates still exploit the streptavidin-biotin interaction ($K_D = 10^{-15}$ M),[1] with streptavidin as a protein connector between a biotinylated DNA and a biotinylated protein; the non-covalent protein-DNA conjugates typically have variable stoichiometry because streptavidin is a tetramer.

Our objective here is to review the current state of methods for the synthesis of covalently linked protein-DNA conjugates with particular focus given to the regiospecificity and stoichiometry of the reactions. Limitations of methods and applications of the conjugates are also discussed. We will not discuss the synthesis of peptide oligonucleotide conjugates, as they have recently been thoroughly reviewed,[30] nor will we focus much on classic methods based on commercially available bifunctional cross-linkers.

Using the Protein's Functional Groups

A little more than 20 years ago, Corey and Schultz reported the synthesis of a conjugate consisting of a nuclease covalently attached to a single-stranded DNA moiety, which enabled it to hybridize to a complementary sequence located in a plasmid to form a triple helix, then to specifically cleave this site.[21] Staphylococcal nuclease was covalently attached to a 3′-thiol containing oligonucleotide through a disulfide link to a unique cysteine that had been engineered in the enzyme.[21,31] Cysteine is a relatively rare amino acid and thiol chemistry offers simple ways to site-specifically attach oligonucleotides. Unfortunately, mutagenesis and knowledge about the structure of the protein is prerequisite. Gianneschi and co-workers recently used the

 10.1071/CH09164 0004-9425/09/101328

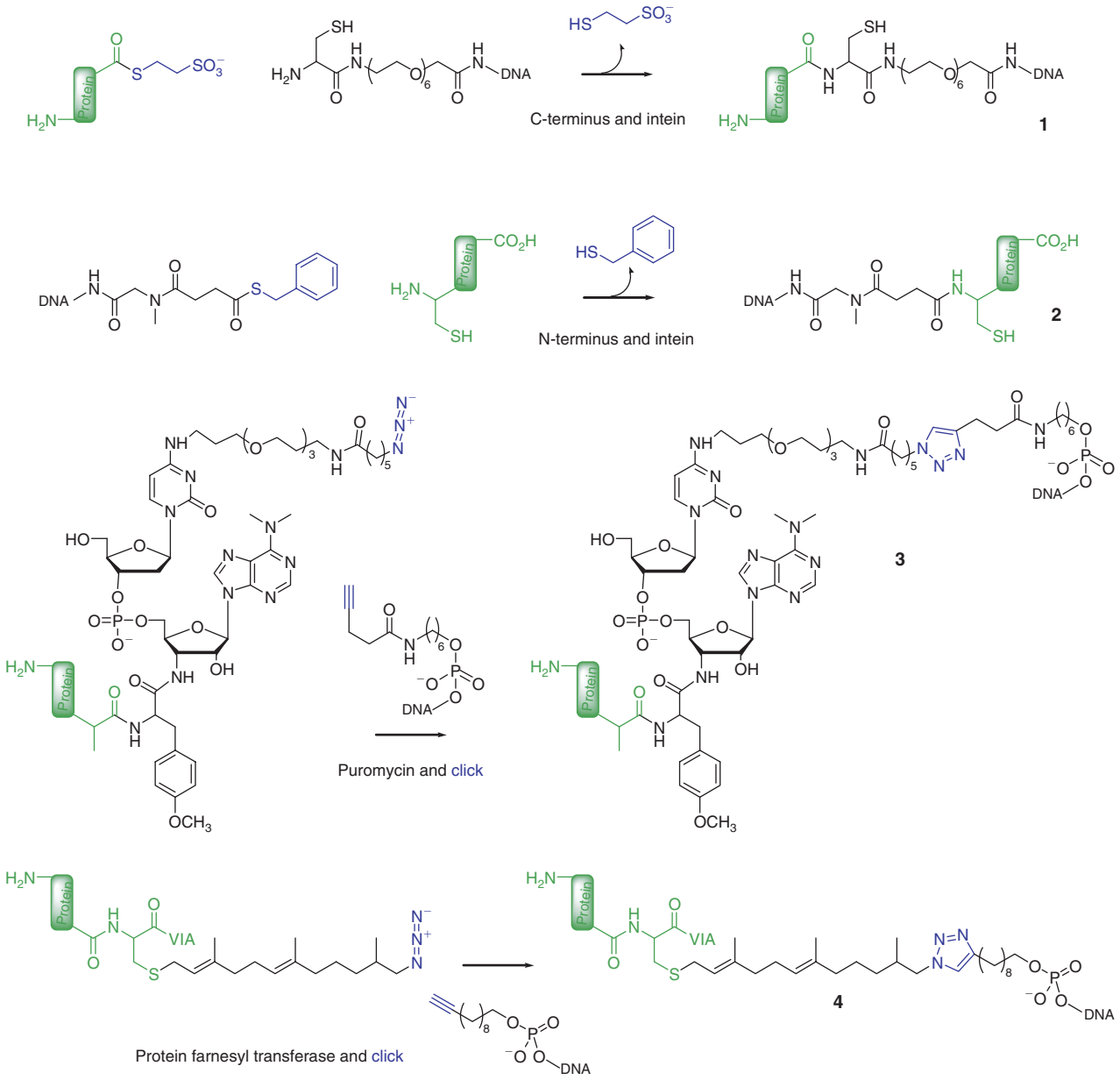

Fig. 1. Homogeneous covalent protein-DNA conjugates. Green lozenges represent the protein moieties.

same strategy for the design of a programmable DNA-regulated semisynthetic enzyme.[32]

Proteins also offer exposed lysine residues containing reactive amines that can be used for cross-linking with commercial bifunctional reagents to generate protein-DNA conjugates. These methods are acceptable so long as they do not functionally inactivate the protein. As many lysine residues are present at protein surfaces, DNA labelling is often a random process that leads to heterogeneous conjugates with variable stoichiometry and regiospecificity.[14] These methods have nevertheless been successfully applied to the production of antibody-DNA conjugates for use in ultrasensitive diagnostics including immuno-PCR.[17]

Intein-Based Strategies for the Production of Homogeneous Conjugates

In contrast, protein-DNA conjugates with well characterized stoichiometry and regiospecificity are required for the bottom-up design of supramolecular architectures, and recent advances have focussed particularly on synthetic methods for their generation. One example of such regiospecific conjugation took advantage of a technique called expressed protein ligation originally developed for the semi-synthesis of proteins.[33] The target protein was genetically fused to an intein with protein splicing activity, followed by a chitin-binding domain.[34] The protein was bound on a column to a chitin matrix, then liberated from it by reaction with mercaptoethansulfonic acid, generating a C-terminal thioester through which the target protein was ligated to a cysteine-PNA conjugate. These conjugates were successfully used for the production of a protein chip through DNA-directed immobilization. Later the same group reported the production of the protein-DNA conjugate (**1**) in Fig. 1,[35] and Burbulis and coworkers have produced protein-DNA 'tadpoles' using a similar approach.[16] Until recently, expressed protein ligation was restricted to labelling of the C-terminus of a protein, but Takeda and coworkers have developed a novel intein-based method for

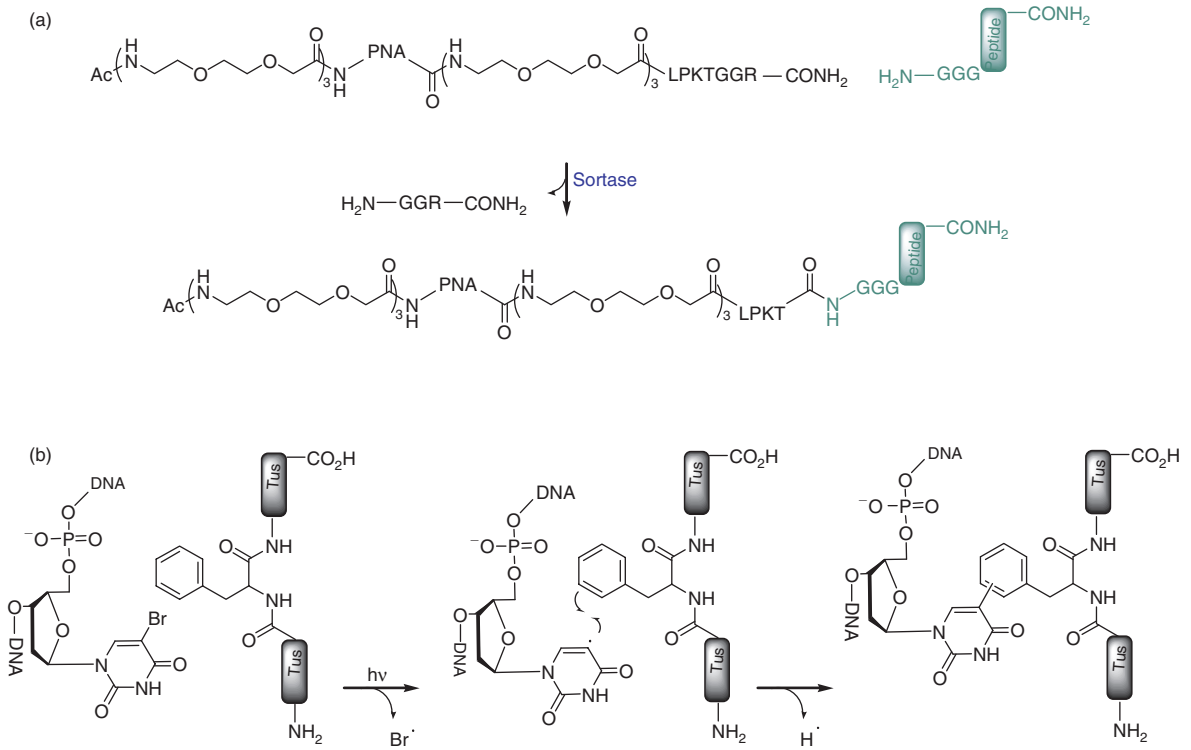

Fig. 2. Mechanistic aspects of protein-DNA conjugation. (a) Sortase mediated ligation. (b) Putative photocross-linking mechanism of the Tus-*Ter* variant complex.

the labelling of the N-terminus via native chemical ligation to afford the protein-DNA conjugate (**2**).[36]

Chemoenzymatic Methodologies for the Production of Homogeneous Conjugates

Puromycin is a fungal secondary metabolite that inhibits protein translation on ribosomes by mimicking an aminoacylated tRNA. This results in translation abortion through covalent bond formation between puromycin and the C-terminal amino acid of a growing peptide chain. About 10 years ago, Roberts and Szostak[37] used this property to link proteins to mRNAs for in vitro evolution experiments. Inspired by this pioneering work, Humenik and coworkers recently described the ribosome-mediated C-terminal labelling of an esterase by a 2′-deoxycytidylyl-(3′-5′)-puromycin analogue bearing an azide group. The azide modified esterase was then reacted with a 5′-alkyne modified oligonucleotide by Cu^{I}-catalyzed [3+2] cycloaddition (click) chemistry to yield the protein-DNA conjugate (**3**).[27] Also, the recent development of a method for the production of internally alkyne modified oligonucleotides[38] could be very useful as it would offer a way to free both 5′ and 3′ ends of an oligonucleotide in a conjugate.

A chemoenzymatic approach recently developed by Duckworth and coworkers[26] produced various protein-oligonucleotide conjugates (**4**) that can spontaneously assemble into defined nanoarchitectures by cDNA hybridization. The enzyme, protein farnesyltransferase, was used to label an engineered substrate protein containing a C-terminal tetrapeptide (SVIA) tag with an azide-modified isoprenoid diphosphate, and site specific DNA conjugation was achieved by a click reaction of the protein azide with a 5′-alkyne-modified DNA. This strategy yielded a short and compact protein-DNA linkage that allowed precise control over protein spacing and orientation in the assembled nanostructure. Also of relevance to the use of click chemistry is a recent report of the in vivo site specific incorporation of an alkyne moiety into calmodulin using a pyrrolysine analogue.[39] This technique could potentially be useful for the synthesis of protein-DNA conjugates as azide-labelled oligonucleotides are now commercially available.

Another enzymatic method described in 2007 uses the commercially available SNAP-Tag system for the covalent linkage of proteins with their encoding DNA sequences for in vitro protein evolution;[28] an O^6-benzylguanine moiety incorporated into DNA in a PCR primer serves as a substrate for a fusion protein containing a mutant form of O^6-alkylguanine-DNA alkyltransferase, thus establishing a covalent link to the DNA.

In *Staphyloccocus aureus*, a transpeptidase called sortase, site-specifically attaches proteins to the surface of the bacterium. It did not take long after the discovery of sortase for development of sortase-based methods for the synthesis of protein fusions and labelling of proteins with chemical probes to transpire. Recently, Pritz and coworkers[25] used sortase to ligate a synthetic peptide-PNA conjugate containing the sortase LPXTG recognition motif with the N-terminal GGG substrate motif of a synthetic peptide (Fig. 2a). Although the production of a genuine protein-DNA conjugate by this approach has apparently not been described, sortase is obviously a great tool to use to make one.

Photochemical Methods

Photocross-linking reactions have been used for decades to locate protein-binding sites in nucleic acids for transcription factors, ribosomal and replicative proteins, and to map amino acids in proteins at nucleic acid interaction sites. Photo-activation of

the aromatic nucleobases can lead to covalent bonds with nearby electron rich amino acid side chains in a bound protein. Cross-linking yields are usually low, but can be improved by use of the more photo-reactive 5-bromodeoxyuridine (BrdU) in place of thymidine in the DNA molecules.

In one approach to exploitation of photocross-linking, single-stranded nucleic acid photo-aptamers that bind proteins with high affinity have been selected by SELEX for the production of protein microarrays.[29] Unlike the earlier examples of methods specifically designed to produce homogenous protein-DNA conjugates as soluble reagents, photo-aptamers were selected to bind and covalently capture proteins at a surface to increase the sensitivity of aptamer-based microarrays. Some of these photo-aptamers can be cross-linked to their target proteins with yields that approach 80%.[40] However, there has been no report yet of their use for the large-scale synthesis of protein-DNA conjugates.

We have recently investigated photo-conjugation methods that take advantage of the strong site-specific interaction of the Tus replication terminator protein with its 23-bp recognition sequence, *Ter*.[41] A *Ter* DNA oligonucleotide bearing a single BrdU substitution at a site predicted from X-ray crystal structures[42,43] to bring it close to the aromatic ring of a phenylalanine residue (F140) in the DNA binding site of Tus, was used for the large scale synthesis of protein-DNA conjugates. Tus and *Ter* associate easily and rapidly, and at low salt concentrations the complex dissociates with a half-life of many hours. When the complexes were UV-irradiated at 312 nm, up to 65% could be covalently cross-linked within minutes under mild conditions.[44] A Tus-GFP fusion protein was also efficiently DNA-labelled using the same method; such a Tus fusion-based method thus offers promise for the easy preparation of protein-DNA conjugates that are well defined with respect to 1:1 stoichiometry and regiospecificity. A putative cross-linking mechanism involving F140 of Tus and the BrdU-substituted *Ter* variant is shown in Fig. 2b.

Conclusion

As our knowledge of the structures and properties of biomolecules expands, so does our desire to manipulate them. Recent years have seen strong interest in the development of highly specific protein-DNA conjugation methods for various applications. This interest is stimulated, for example, by the desire of nanoscientists to manipulate or design new molecular materials and tools from bottom-up or by self-assembly processes. Without doubt, we will see further increase in novel applications especially in the multidisciplinary field of nanotechnology. Many of these conjugates are already employed in frontier technologies and it is anticipated that many more unanticipated uses of them will come to light in the near future. This will be facilitated by the improved spatial control offered by recently described methodologies.

Acknowledgement

Our work in this area is supported by a project grant from the NHMRC, Australia.

References

[1] C. M. Niemeyer, *Trends Biotechnol.* **2002**, *20*, 395. doi:10.1016/S0167-7799(02)02022-X

[2] J. H. Chen, N. C. Seeman, *Nature* **1991**, *350*, 631. doi:10.1038/350631A0

[3] R. P. Goodman, I. A. T. Schaap, C. F. Tardin, C. M. Erben, R. M. Berry, C. F. Schmidt, A. J. Turberfield, *Science* **2005**, *310*, 1661. doi:10.1126/SCIENCE.1120367

[4] Y. W. Zhang, N. C. Seeman, *J. Am. Chem. Soc.* **1994**, *116*, 1661. doi:10.1021/JA00084A006

[5] W. M. Shih, J. D. Quispe, G. F. Joyce, *Nature* **2004**, *427*, 618. doi:10.1038/NATURE02307

[6] E. Winfree, F. R. Liu, L. A. Wenzler, N. C. Seeman, *Nature* **1998**, *394*, 539. doi:10.1038/28998

[7] T. H. LaBean, H. Yan, J. Kopatsch, F. R. Liu, E. Winfree, J. H. Reif, N. C. Seeman, *J. Am. Chem. Soc.* **2000**, *122*, 1848. doi:10.1021/JA993393E

[8] H. Yan, S. H. Park, G. Finkelstein, J. H. Reif, T. H. LaBean, *Science* **2003**, *301*, 1882. doi:10.1126/SCIENCE.1089389

[9] Y. He, Y. Chen, H. P. Liu, A. E. Ribbe, C. D. Mao, *J. Am. Chem. Soc.* **2005**, *127*, 12202. doi:10.1021/JA0541938

[10] P. W. K. Rothemund, *Nature* **2006**, *440*, 297. doi:10.1038/NATURE04586

[11] C. M. Niemeyer, M. Adler, R. Wacker, *Trends Biotechnol.* **2005**, *23*, 208. doi:10.1016/J.TIBTECH.2005.02.006

[12] S. M. Gustafsdottir, E. Schallmeiner, S. Fredriksson, M. Gullberg, O. Soderberg, M. Jarvius, J. Jarvius, M. Howell, U. Landegren, *Anal. Biochem.* **2005**, *345*, 2. doi:10.1016/J.AB.2005.01.018

[13] M. Gullberg, S. Fredriksson, M. Taussig, J. Jarvius, S. Gustafsdottir, U. Landegren, *Curr. Opin. Biotechnol.* **2003**, *14*, 82. doi:10.1016/S0958-1669(02)00011-3

[14] E. R. Hendrickson, T. M. H. Truby, R. D. Joerger, W. R. Majarian, R. C. Ebersole, *Nucleic Acids Res.* **1995**, *23*, 522. doi:10.1093/NAR/23.3.522

[15] I. Burbulis, K. Yamaguchi, R. H. Yu, O. Resnekov, R. Brent, *Nat. Methods* **2007**, *4*, 1011. doi:10.1038/NMETH1127

[16] I. Burbulis, K. Yamaguchi, A. Gordon, R. Carlson, R. Brent, *Nat. Methods* **2005**, *2*, 31. doi:10.1038/NMETH729

[17] T. Sano, C. L. Smith, C. R. Cantor, *Science* **1992**, *258*, 120. doi:10.1126/SCIENCE.1439758

[18] S. H. Park, P. Yin, Y. Liu, J. H. Reif, T. H. LaBean, H. Yan, *Nano Lett.* **2005**, *5*, 729. doi:10.1021/NL050175C

[19] C. M. Niemeyer, J. Koehler, C. Wuerdemann, *ChemBioChem* **2002**, *3*, 242. doi:10.1002/1439-7633(20020301)3:2/3<242::AID-CBIC242>3.0.CO;2-F

[20] K. A. Cissell, Y. Rahimi, S. Shrestha, S. K. Deo, *Bioconjug. Chem.* **2009**, *20*, 15. doi:10.1021/BC8003099

[21] D. R. Corey, P. G. Schultz, *Science* **1987**, *238*, 1401. doi:10.1126/SCIENCE.3685986

[22] S. Howorka, S. Cheley, H. Bayley, *Nat. Biotechnol.* **2001**, *19*, 636. doi:10.1038/90236

[23] F. Kukolka, C. M. Niemeyer, *Org. Biomol. Chem.* **2004**, *2*, 2203. doi:10.1039/B406492E

[24] M. Lovrinovic, C. M. Niemeyer, *ChemBioChem* **2007**, *8*, 61. doi:10.1002/CBIC.200600303

[25] S. Pritz, Y. Wolf, O. Kraetke, J. Klose, M. Bienert, M. Beyermann, *J. Org. Chem.* **2007**, *72*, 3909. doi:10.1021/JO062331L

[26] B. P. Duckworth, Y. Chen, J. W. Wollack, Y. Sham, J. D. Mueller, T. A. Taton, M. D. Distefano, *Angew. Chem. Int. Ed.* **2007**, *46*, 8819. doi:10.1002/ANIE.200701942

[27] M. Humenik, Y. W. Huang, Y. R. Wang, M. Sprinzl, *ChemBioChem* **2007**, *8*, 1103. doi:10.1002/CBIC.200700070

[28] V. Stein, I. Sielaff, K. Johnsson, F. Hollfelder, *ChemBioChem* **2007**, *8*, 2191. doi:10.1002/CBIC.200700459

[29] C. Bock, M. Coleman, B. Collins, J. Davis, G. Foulds, L. Gold, C. Greef, J. Heil, J. S. Heilig, B. Hicke, M. N. Hurst, G. M. Husar, D. Miller, R. Ostroff, H. Petach, D. Schneider, B. Vant-Hull, S. Waugh, A. Weiss, S. K. Wilcox, D. Zichi, *Proteomics* **2004**, *4*, 609. doi:10.1002/PMIC.200300631

[30] N. Venkatesan, B. H. Kim, *Chem. Rev.* **2006**, *106*, 3712. doi:10.1021/CR0502448

[31] R. Zuckermann, D. Corey, P. Schultz, *Nucleic Acids Res.* **1987**, *15*, 5305. doi:10.1093/NAR/15.13.5305
[32] N. C. Gianneschi, M. R. Ghadiri, *Angew. Chem. Int. Ed.* **2007**, *46*, 3955. doi:10.1002/ANIE.200700047
[33] T. W. Muir, D. Sondhi, P. A. Cole, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 6705. doi:10.1073/PNAS.95.12.6705
[34] M. Lovrinovic, R. Seidel, R. Wacker, H. Schroeder, O. Seitz, M. Engelhard, R. S. Goody, C. M. Niemeyer, *Chem. Commun.* **2003**, 822. doi:10.1039/B212294D
[35] M. Lovrinovic, M. Spengler, C. Deutsch, C. M. Niemeyer, *Mol. Biosyst.* **2005**, *1*, 64. doi:10.1039/B503839A
[36] S. Takeda, S. Tsukiji, H. Ueda, T. Nagamune, *Org. Biomol. Chem.* **2008**, *6*, 2187. doi:10.1039/B720013G
[37] R. W. Roberts, J. W. Szostak, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 12297. doi:10.1073/PNAS.94.23.12297
[38] J. Gierlich, G. A. Burley, P. M. E. Gramlich, D. M. Hammond, T. Carell, *Org. Lett.* **2006**, *8*, 3639. doi:10.1021/OL0610946
[39] T. Fekner, X. Li, M. M. Lee, M. K. Chan, *Angew. Chem. Int. Ed.* **2009**, *48*, 1633. doi:10.1002/ANIE.200805420
[40] T. H. Koch, D. Smith, E. Tabacman, D. A. Zichi, *J. Mol. Biol.* **2004**, *336*, 1159. doi:10.1016/J.JMB.2004.01.001
[41] C. Neylon, A. V. Kralicek, T. M. Hill, N. E. Dixon, *Microbiol. Mol. Biol. Rev.* **2005**, *69*, 501. doi:10.1128/MMBR.69.3.501-526.2005
[42] M. D. Mulcair, P. M. Schaeffer, A. J. Oakley, H. F. Cross, C. Neylon, T. M. Hill, N. E. Dixon, *Cell* **2006**, *125*, 1309. doi:10.1016/J.CELL.2006.04.040
[43] K. Kamada, T. Horiuchi, K. Ohsumi, N. Shimamoto, K. Morikawa, *Nature* **1996**, *383*, 598. doi:10.1038/383598A0
[44] D. B. Dahdah, I. Morin, M. J. J. Moreau, N. E. Dixon, P. M. Schaeffer, *Chem. Commun.* **2009**, 3050. doi:10.1039/B900905A