Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry Letters 14 (2004) 2407–2410
Bioorganic & Medicinal Chemistry Letters

Site-specific conjugation of oligonucleotides to the C-terminus of recombinant protein by expressed protein ligation

Shuji Takeda, Shinya Tsukiji and Teruyuki Nagamune*

Department of Chemistry and Biotechnology, School of Engineering, The University of Tokyo, Tokyo 113-8656, Japan

Received 24 February 2004; revised 8 March 2004; accepted 9 March 2004

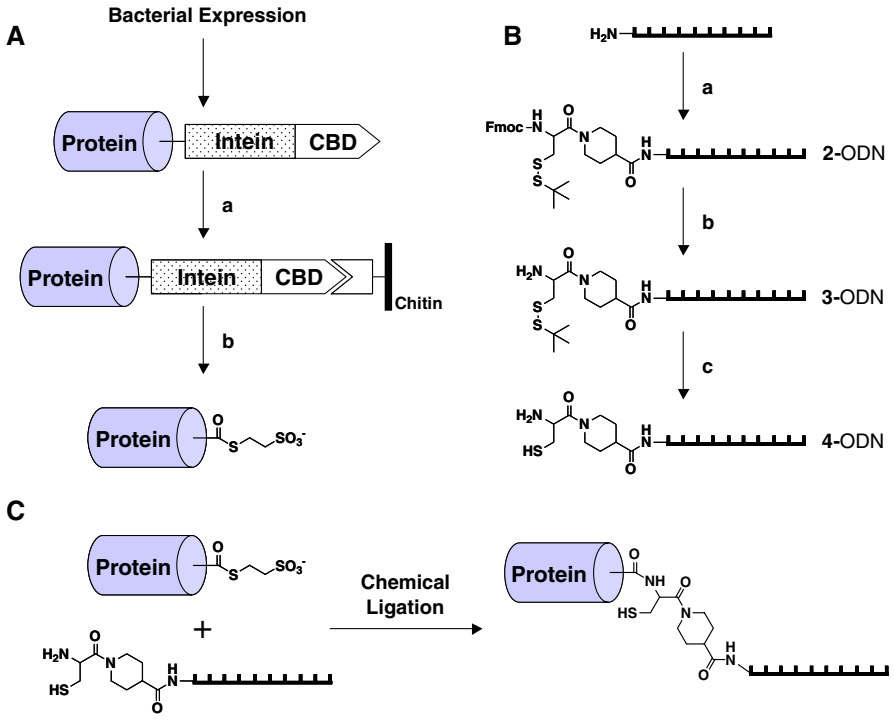
Abstract—We developed a convenient method for synthesizing homogeneous DNA–protein conjugates. The method is based on expressed protein ligation of intein-fusion proteins and oligonucleotides derivatized with a cysteine. A range of cysteinyl oligonucleotides were synthesized by using a new reagent **1** and were successfully applied to expressed protein ligation to attach the oligonucleotides specifically at the C-terminus of a recombinant protein.

Semisynthetic DNA-protein/enzyme conjugates hold great promise as versatile molecular tools for biotechnology and materials science.[1] By coupling the unique target-specific hybridization properties of DNA molecules with the almost unlimited functionality of proteins, a number of hybrid molecules for a variety of applications have been generated so far. In a pioneering work by Corey and Schultz, an oligonucleotide–*Staphylococcus* nuclease chimera was prepared as a synthetic restriction nuclease.[2] The specific triple-helix formation of the single stranded DNA moiety of the chimera enabled the site-selective cleavage of the plasmid DNA. Recently, Bayley and co-workers constructed a DNA-α-hemolysin nanopore biosensor capable of identifying individual DNA strands with single-base resolution.[3] These and other works[4] clearly demonstrate that the rational introduction of an oligonucleotide binding domain into a protein represents a powerful strategy for the construction of artificial proteins/enzymes, which selectively target (cleave or detect) specific nucleic acids. Such molecules are extremely difficult to design using protein components alone. On the other hand, the accurate and robust hybridization ability of DNA molecules has proven to be useful for the DNA-programmed assembly of proteins into a spatially designed arrangement. This approach has recently attracted considerable attention for the fabrication of protein microarrays,[5] nanometer-sized supramolecular devices,[6] and immunological bioassays.[7]

However, compared to recent significant advances in the total chemical synthesis of oligonucleotide–peptide conjugates by solution-phase[8] or solid-phase synthesis,[9] conjugation strategies for incorporating an oligonucleotide into a larger recombinant protein are still poorly developed. DNA–protein conjugates studied so far have been prepared by three conventional strategies. The first is the direct attachment of a nucleotide oligomer containing a thiopyridyl disulfide to the reactive cysteine residue of a protein via a disulfide exchange reaction.[2,3] Although this strategy can produce a homogeneous protein carrying an oligonucleotide at a specific position, the inherent chemical instability of the disulfide bond against reducing reagents may restrict its broad use. The second strategy uses bifunctional crosslinkers, such as *N*-[γ-maleimidobutyryloxy]succinimide ester (GMBS)[7a] and sulfosuccinimidyl 4-[*p*-maleimidophenyl] butyrate (sulfo-BMPB),[5] to covalently bridge an oligonucleotide and a protein. This approach is applicable to any proteins with no reactive cysteine residue. However, chemical modification of protein surfaces in a nonspecific manner generates conjugates with high heterogeneity. In the third strategy, biotinylated proteins are prepared to produce noncovalent assemblies with biotinylated DNA through the strong interaction with streptavidin.[5,7b] The biotinylation of proteins again relies on random chemical modification. Therefore, development of a novel methodology for efficient and site-selective conjugation of oligonucleotides to proteins

Keywords: Expressed protein ligation; Protein–DNA conjugate; Cysteine modification.
* Corresponding author. Tel.: +81-3-5841-7328; fax: +81-3-5841-8657; e-mail: nagamune@bio.t.u-tokyo.ac.jp

0960-894X/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2004.03.023



Scheme 1. A. Expression and purification of recombinant proteins containing an α-thioester. (a) Chitin beads equilibrated in buffer A (50 mM Tris–HCl, 500 mM NaCl, 1 mM EDTA, pH 8.5); (b) Thiolysis in buffer B (50 mM Tris–HCl, 500 mM NaCl, 1 mM EDTA, 50 mM MESNA, pH 8.5), 4 °C, 24 h; **B**. Synthesis of cysteinyl-oligonucleotides. (a) **1** (70–80 equiv) in a borate buffer (10 mM, pH 8.5) containing 50% DMF; (b) aqueous ammonia, 37 °C, 12 h; (c) 2 M DTT aq, rt, 1 h; **C**. Conjugation of **4**-ODNs to proteins at their C-terminus. Buffer B, 4 °C, 48 h.

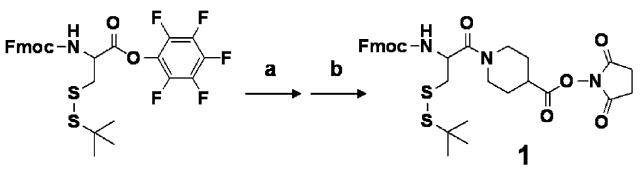
is a challenge that would allow further investigation into the scope of DNA–protein hybrids.

Here we report a facile and efficient method for coupling synthetic oligonucleotides with recombinant proteins specifically at their C-terminus. Our strategy is based on expressed protein ligation (EPL), in which a recombinantly expressed protein containing a C-terminal α-thioester reacts with a synthetic peptide containing an N-terminal cysteine to form a native peptide bond (Scheme 1).[10] This technique has previously been used for the chemical semisynthesis of a variety of proteins.[11]

For chemoselective ligation with C-terminal α-thioesters, oligonucleotides require a cysteine moiety to be incorporated. However, no convenient route for introduction of a cysteine or its analogue at the end of oligonucleotides has been reported until recently.[12] For this purpose, we designed a new reagent **1**, a Fmoc-protected dipeptide carrying active *N*-hydroxysuccinimide (NHS) ester, for solution-phase coupling with amine-functionalized oligonucleotides. Isonipecotic acid was used as a connecter to prevent facile epimerization of the cysteine moiety and rapid hydrolysis of the NHS ester. One advantage of this approach is the simplicity of preparation of the cysteine-appended oligonucleotides, as described below. A second advantage is that, since synthetic 5′- and 3′-amino-oligonucleotides are now available at a low cost, **1** allows recombinant proteins to be conjugated at either end of DNA strands.

Compound **1** was synthesized as shown in Scheme 2 in a straightforward manner from commercially available materials. Coupling of the pentafluorophenyl ester of *N*-α-Fmoc-*S*-*tert*-butylsulfenylcysteine with 1.0 equiv of isonipecotic acid was accomplished under standard conditions. The intermediate was then reacted with NHS using HBTU to yield a white solid **1** in 38% yield for the two steps.[13]

Synthesis of cysteine-modified oligodeoxynucleotides, **4**-ODN, is outlined in Scheme 1B. The reaction mixture of 5′-amino-ODN**a** (5′-NH_2–$(CH_2)_3$–O–$(CH_2)_2$–pODN, Table 1) and excess (ca. 80 equiv) of **1** was incubated in a borate buffer (10 mM, pH 8.5) containing 50% DMF at room temperature for 10 h, and purified by reversed-phase HPLC (RP-HPLC).[14] RP-HPLC indicates that the coupling reaction takes place cleanly with no significant side reactions except moderate hydrolysis of the NHS ester group of compound **1**. Fmoc protection of the resulting conjugate **2**-ODN**a** was removed in aqueous ammonia at 37 °C for 12 h, followed by lyophilization, to obtain the conjugate **3**-ODN. Subsequently, **3**-ODN**a** was treated with 2 M DTT solution at room temperature for 1 h to deprotect the *S*-*tert*-butylsulfenyl



Scheme 2. (a) Isonipecotic acid (1 equiv), DIEA (2 equiv), DMF, rt, 6 h; (b) NHS (1 equiv), HBTU (1.2 equiv), DIEA (1 equiv), DMF, rt, 5 h.

Table 1. Sequences of ODNs used in this study

ODN	Sequence	Modification
a	AAAAAAAAAAAA	5′
b	AAAAAAAAAAAA	3′
c	TTTTTTTTTTT	5′
d	TTTTTTTTTTT	3′
e	CCCCCCCCCCC	5′
f	CCCCCCCCCCC	3′
g	TTTCCCAGTCACGACGTTGTA	5′
h	GTGGATAACCGTATTACCGCC	5′
i	TTTTTGATGAATTCGGA	5′
j	TTTTTTCCGAATTCATC	5′
k	TCCTGTGTGAATTTGTTATCCGCT	5′
l	TTTTTTTTTTTTTCACAGCTGAG-GATAGGACAT	5′

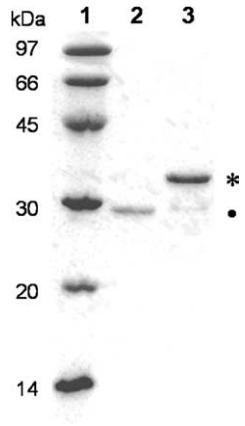


Figure 1. SDS-PAGE analysis of ECFP-ODN**c** conjugate. Lane 1, molecular weight marker; 2, ECFP-thioester alone; 3, Reaction mixture. The ECFP-thioester and the ligation product are denoted by ● and ∗, respectively.

group, yielding the desired conjugate **4**-ODN**a**. These two deprotection steps also produce no by-products as evaluated by RP-HPLC. All products were satisfactorily characterized by MALDI-TOF mass spectrometry.[15] It should be noted that, owing to the clean reaction, highly pure **2**-ODNs and subsequent **3**-ODNs can be easily prepared by purification with size-exclusion chromatography using a NAP-5 column (Amersham Bioscience), in total yields over 60%. With this procedure, several **3**-ODNs with various lengths and sequences were efficiently synthesized (Table 1), indicating the general applicability of this conjugation chemistry.

Recombinant protein α-thioesters can be generated by thiolysis of the corresponding protein–intein fusion where the intein has a mutation to block the final step of protein splicing (Scheme 1A).[16] As a demonstration of our semi-synthetic strategy, we attempted to ligate the **4**-ODNs to the C-terminus of a green fluorescent protein variant, ECFP. The full-length ECFP sequence was cloned into a commercially available intein expression system, pTWIN1 vector (New England Biolabs). Subsequent overexpression in *Escherichia coli* resulted in the generation of a protein chimera in which the desired ECFP fragment was fused at its C-terminus to a *Mycobacterium xenopi* GyrA intein linked to a chitin binding domain (CBD). The recombinant fusion protein was purified by affinity chromatography on chitin beads and was then cleaved using 2-mercaptoethanesulfonic acid (MESNA) to afford the ECFP-thioester with >95% homogeneity as indicated by SDS-PAGE analysis.

Next we investigated EPL between the ECFP-thioester and the **4**-ODN**c** to give the covalent DNA–protein conjugate (Scheme 1C).[17] Because *S-tert*-butylsulfenyl groups are smoothly released in situ in the ligation buffer containing MESNA, **3**-ODN**c** was directly used in the following EPL experiments. The ligation reaction was initiated by addition of **3**-ODN**c** to the ECFP-thioester solution obtained above and was monitored by SDS-PAGE.[18] It was clearly demonstrated that the ligation proceeds to approximately 80% completion after incubation at 4 °C for two days (Fig. 1). MALDI-TOF-MS analysis of the ligation product after agarose gel purification indicated a mass consistent with the expected product.[19] Furthermore, we have successfully applied this strategy to couple ECFP and 3′-cysteinyl-ODN (**b**,**d** and **f**), and the C1 domain of Protein G and 5′-cysteinyl-ODN (**a**, **b** and **l**) (data not shown).

In summary, we demonstrated for the first time the site-specific conjugation of oligonucleotides to recombinant proteins at their C-terminus by expressed protein ligation. We are now applying this method to the synthesis of other DNA–protein hybrids for future applications.

Acknowledgements

We would like to thank Profs. Kazuhiko Saigo and Takeshi Wada for generously allowing us to use their facilities and fruitful discussion in the initial stage of this study. We are grateful to all members of the Saigo Laboratory, particularly Dr. Yasuhiro Ishida for helpful assistance. We thank Tsutomu Tanaka for helpful discussions.

References and notes

1. (a) Niemeyer, C. M. *Chem. Eur. J.* **2001**, *7*, 3189; (b) Niemeyer, C. M. *Trends Biotechnol.* **2002**, *20*, 395.
2. Corey, D. R.; Pei, D.; Schultz, P. G. *J. Am. Chem. Soc.* **1989**, *111*, 8523.
3. Howorka, S.; Cheley, S.; Bayley, H. *Nat. Biotechol.* **2001**, *19*, 636.
4. Saghatelian, A.; Guckian, K. M.; Thayaer, D. A.; Ghadiri, M. R. *J. Am. Chem. Soc.* **2003**, *125*, 344.
5. (a) Niemeyer, C. M.; Sano, T.; Smith, C. L.; Cantor, C. R. *Nucleic Acids Res.* **1994**, *22*, 5530; (b) Niemeyer, C. M.; Boldt, L.; Ceyhan, B.; Blohm, D. *Anal. Biochem.* **1999**, *268*, 54.
6. (a) Keren, K.; Krueger, M.; Gilad, R.; Ben-Yoseph, G.; Sivan, U.; Braun, E. *Science* **2002**, *297*, 72; (b) Yan, H.; Park, S. H.; Finkelstein, H.; Reif, J. H.; LaBean, T. H. *Science* **2003**, *301*, 1882; (c) Keren, K.; Berman, R. S.;

Buchstab, E.; Sivan, U.; Braun, E. *Science* **2003**, *301*, 1380; (d) Li, H.; Park, S. H.; Reif, J. H.; LaBean, T. H.; Yan, H. *J. Am. Chem. Soc.* **2004**, *126*, 418.

7. (a) Schweizer, B.; Roberts, S.; Grimwande, B.; Shao, W.; Wang, M.; Fu, Q.; Shu, Q.; Laroche, I.; Zhou, Z.; Tchernev, V. T.; Christiansen, J.; Velleca, M.; Kingsmore, S. F. *Nat. Biotechnol.* **2002**, *20*, 359; (b) Niemeyer, C. M.; Wacker, R.; Adler, M. *Nucleic Acids Res.* **2003**, *31*, e90.
8. (a) Zanta, M. A.; Belguise-Valladier, P.; Behr, J. P. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 91; (b) Harrison, J. P.; Balasubramanian, S. *Nucleic Acids Res.* **1998**, *26*, 3136; (c) Bongartz, J.-P.; Aubertin, A.-M.; Milhaud, P. G.; Lebleu, B. *Nucleic Acids Res.* **1994**, *22*, 4681.
9. (a) Kubo, T.; Morikawa, M.; Ohba, H.; Fujii, M. *Org. Lett.* **2003**, *5*, 2623; (b) Stetsenko, D. A.; Gait, M. J. *Bioconjugate Chem.* **2001**, *12*, 576; (c) Stetsenko, D. A.; Malakhov, A. D.; Gait, M. J. *Org. Lett.* **2002**, *4*, 3259; (d) Peyrottes, S.; Mestre, B.; Burlina, F.; Gait, M. J. *Tetrahedron* **1998**, *54*, 12513; (e) Napoli, L. D.; Messere, A.; Montesarchio, D.; Piccialli, G.; Benedetti, E.; Bucci, E.; Rossi, F. *Bioorg. Med. Chem.* **1999**, *7*, 395; (f) Schwope, I.; Bleczinski, C. F.; Richert, C. *J. Org. Chem.* **1999**, *64*, 4749.
10. (a) Muir, T. W. *Annu. Rev. Biochem.* **2003**, *72*, 249; (b) Hofmann, R. M.; Muir, T. W. *Curr Opin. Biotech.* **2002**, *13*, 297; (c) Cotton, G. J.; Muir, T. W. *Chem. Biol.* **1999**, *6*, R247.
11. Very recently, Niemeyer and co-workers reported the synthesis of PNA–protein conjugates using EPL. However, the synthesis of DNA–protein conjugates by EPL has not yet been reported. Lovrinovic, M.; Seidel, R.; Wacker, R.; Schroeder, H.; Seitz, O.; Engelhard, M.; Goody, R. S.; Niemeyer, C. M. *Chem. Commun.* **2003**, 822.
12. Stetsenko, D. A.; Gait, M. J. *J. Org. Chem.* **2000**, *65*, 4900.
13. The compound **1** reported herein was satisfactorily characterized by ^{1}H NMR and MALDI-TOF-MS spectrometry. ^{1}H NMR ($CDCl_3$): δ 1.34 (s, 9H), 1.94–2.10 (m, 4H), 2.82 (s, 4H), 2.88–2.97 (m, 4H), 3.20–3.42 (m, 2H), 3.99 (m, 1H), 4.22 (d, 2H), 4.37 (t, 1H), 5.03 (q, 1H), 5.68 (t, 1H), 7.26 (t, 2H), 7.34 (t, 2H), 7.59 (d, 2H), 7.76 (d, 2H). MALDI-TOF-MS (dithranol): calcd for $[M+Na]^+ =$ 662.20, obsd 662.00.
14. HPLC was carried out on a μBondasphere C18 column (5 μ, 150 × 3.8 mm, Waters) eluted with 0.1 M ammonium acetate buffer (pH 7.0) containing 0–90% acetonitrile in a linear gradient over 60 min at a flow rate of 1 mL/min, detected at 260 nm.
15. Each product of the modification steps of ODN **a** and ODN**c** was identified by MALDI-TOF-MS (Matrix: 3-hydroxypicolinic acid). ODN **a**: **2**-ODN**a**, calcd for $[M+H]^+ = 4399.0$, obsd 4401.0; **3**-ODN**a**, calcd for $[M+H]^+ = 4179.0$, obsd 4179.8; **4**-ODN**a**, calcd for $[M+H]^+ = 4190.9$, obsd 4090.2. ODN**c**: **2**-ODN**c**, calcd for $[M+H]^+ = 4292.4$, obsd 4289.9; **3**-ODN**c**, calcd for $[M+H]^+ = 4070.8$, obsd 4070.0; **4**-ODN**c**, calcd for $[M+H]^+ = 3982.7$, obsd. 3982.8.
16. Southworth, M. W.; Amaya, K.; Evans, T. C.; Xu, M. Q.; Perler, F. B. *BioTechniques* **1999**, *27*, 110.
17. All reactions were carried out in a total volume of 10 μL of buffer B (50 mM Tris–HCl, 500 mM NaCl, 1 mM EDTA, 50 mM MESNA, pH 8.5). Final concentrations of ECFP and **3**-ODNs were 13 and 100 μM, respectively.
18. All samples were boiled in a loading buffer containing 2-mercaptoethanol to completely cleave disulfide bonds prior to SDS-PAGE analysis.
19. The reaction solution was subjected to 3% agarose gel electrophoresis. The gel was run in TEA buffer for 1.5 h. The product, ECFP-ODN**c**, was extracted from the gel using Freezen'N Squeeze (Bio-Rad). MALDI-TOF-MS (Matrix: sinapic acid): ECFP-thioester: calcd for $[M+Na]^+ = 27014.3$, obsd 27016.0; ECFP-ODN**c**: calcd for [M+2-mercaptoethanesulfonic acid]$^+ = 30975.04$, obsd 30955.0. DTT can be used after the ligation step to reduce any unwanted disulfide bonds. Nyanguile, O.; Dancik, C.; Blackmore, J.; Mulgrew, K.; Kaleko, M.; Stevenson, S. C. *Gene Therapy* **2003**, *10*, 1362.