J|A|C|S
COMMUNICATIONS

Published on Web 02/10/2004

Sortase-Mediated Protein Ligation: A New Method for Protein Engineering

Hongyuan Mao,* Scott A. Hart, Amy Schink, and Brian A. Pollok

Ansata Therapeutics, Inc. 505 Coast Boulevard South, La Jolla, California 92037

Received December 3, 2003; E-mail: Helen@ansatainc.com

Specific incorporation of novel functionality in proteins through protein engineering is becoming a widely used tool in protein biochemistry. One of the most powerful methods is through a controlled protein ligation, where peptide analogues, unnatural amino acids, stable isotopes, fluorophores, and other biochemical and biophysical probes can be specifically incorporated into recombinantly expressed proteins. Through native chemical ligation, two fully deprotected synthetic peptide fragments can be selectively linked via a native peptide bond.[1,2] A 166-amino acid polymer-modified erythropoiesis protein has been synthesized using this ligation method.[3] However, native chemical ligation relies on efficient synthesis of peptide thioesters, which can be technically difficult for large polypeptides such as proteins. An intein-based protein ligation system can generate a protein thioester by thiolysis of a corresponding protein−intein fusion,[4] and has been successfully applied to unnatural modification of proteins.[5−8] However, difficulties remain because the target protein must be expressed as a fusion with an intein.[9] Currently there are very few enzyme-based approaches for peptide ligation besides the intein fusion system.[10−12] One is subtiligase, an engineered subtilisin that is capable of catalyzing the ligation of peptide fragments.[10,11] Since the engineering of subtiligase, few reports of further development have appeared, likely because the approach is cumbersome, requiring stepwise esterification and subsequent ligation.

We present here a novel approach for polypeptide ligation using a sortase-catalyzed reaction. Sortase, a transpeptidase found in the cell envelope of many Gram-positive bacteria, anchors surface proteins to the peptidoglycan cross bridge of the cell wall.[13,14] Schneewind, et al., has identified a *Staphylococcus aureus* sortase (SrtA) that catalyzes the transpeptidation by cleaving between threonine and glycine at an LPXTG recognition motif and subsequently joining the carboxyl group of threonine to an amino group of pentaglycine on the cell wall peptidoglycan.[13,15] Variants of the LPXTG recognition motif were found only at the X and T positions (e.g. T → A, and X = D, E, A, N, Q, or K).[16] In vitro, sortase slowly hydrolyzes an LPXTG peptide in the absence of a suitable nucleophile, but catalyzes exclusively a transpeptidation in the presence of a triglycine.[17] Recently, a kinetic study suggested the nucleophile binding site of sortase prefers a diglycine.[18]

Previous reports of sortase activity and mechanism have established the reactivity of glycine, diglycine, and triglycine as substrates of ligation in vitro.[17,18] However, further examples of the ability of sortase to couple longer polypeptides in vitro is very limited.[19] Toward this aim, we first synthesized four peptides G_nPep (G_n-RRNRRTSKLMLR, n = 1, 2, 3, or 5) as potential sortase substrates to determine the number of glycines required at the N terminus for efficient conjugation in vitro. Each was incubated with an equimolar amount of an LPXTG-containing fluorescent peptide RE(*Edans*)-LPKTGK(*Dabcyl*)R in the presence of sortase. The rate of sortase-mediated LPXTG ligation was monitored through the increase of *Edans* fluorescence upon the spatial separation of the two fluorophores. It appeared that peptide with only a single aminoglycine

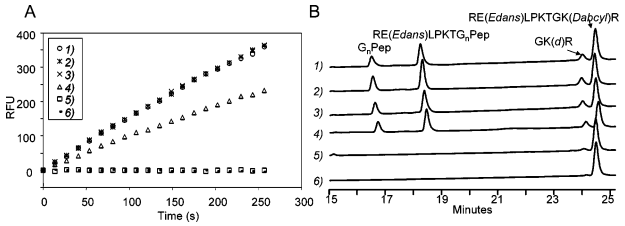

Figure 1. Sortase-mediated ligation of RE (*Edans*)LPKTGK(*Dabcyl*)R with G_nPep (n = 1, 2, 3, or 5): (A) Fluorescence measurement of sortase cleavage rate of 0.1 mM RE(*Edans*)LPKTGK(*Dabcyl*)R in the presence of 0.1 mM G_nPep peptide (***1***−***4***; n = 5, 3, 2, and 1 respectively). ***5*** and ***6*** are RE(*Edans*)LPKTGK(*Dabcyl*)R in the presence and absence of sortase, respectively. (B) Reverse phase HPLC chromatography of the ligation reactions after 30 min at 37 °C shows similar amounts of ligation product RE(*Edans*)LPKTG_nPep formation (***1*** to ***4***; n = 5, 3, 2, and 1 respectively). ***5*** and ***6*** are RE(*Edans*)LPKTGK(*Dabcyl*)R in the presence and absence of sortase, respectively.

Table 1. MALDI-TOF Mass Analysis of the Sortase-Mediated Conjugation of RE(*Edans*)LPKTGK(*Dabcyl*)R with Various Substrates

nucleophile	ligation product[a]	observed	calculated
-	RE(***e***)LPKTGK(***d***)R	1,624.59 ± 0.04	1,624.85
H_2O	RE(***e***)LPKT	1,033.25 ± 0.10	1,033.51
G_1Pep	RE(***e***)LPKT-G_1Pep	2,658.10 ± 0.20	2,658.46
G_2Pep	RE(***e***)LPKT-G_2Pep	2,715.50 ± 0.10	2,715.49
G_3Pep	RE(***e***)LPKT-G_3Pep	2,772.56 ± 0.15	2,772.51
G_5Pep	RE(***e***)LPKT-G_5Pep	2,886.58 ± 0.06	2,886.55
Tat	RE(***e***)LPKT-Tat	2,631.56 ± 0.06	2,631.48
(D)-Tat	RE(***e***)LPKT-(D)-Tat	2,631.54 ± 0.08	2,631.48

[a] ***e*** = *Edans* and ***d*** = *Dabcyl*.

was capable of nucleophilic attack at the LPXTG substrate at a rate more than 50 times faster than that of H_2O (Figure 1A). The rate increased slightly in the presence of a peptide with two or more N-terminal glycines. This observation is consistent with the proposed diglycine nucleophile-binding site in sortase.[18] Nonetheless, the final product yield (approximately 30% after 30 min) was not affected by the number of glycines presented at the N terminus (Figure 1B and Table 1).

We next evaluated the suitability of sortase in protein−peptide and protein−protein ligation. When GFP-LPETG-6His was incubated with sortase in the absence of an appropriate aminoglycine nucleophile, it was hydrolyzed to GFP-LPET and formed a low level of nonspecific conjugates (Table 2 and Figure 2A). However, in the presence of an aminoglycine-containing peptide (with a 1 to 5 molar ratio of GFP-LPETG-6His to peptide) GFP-LPETG-6His formed only specific conjugates with the peptide (Table 2 and Figure 2A). The conjugation yield was approximately 50% after 6 h and increased to 90% within 24 h (Figure 2A). Furthermore, sortase successfully mediated conjugation of GFP-LPETG-6His to a 29 kDa protein (Gly-emGFP) with an N-terminal glycine (Table 2 and Figure 2A). The rate of protein−peptide or protein−protein ligation appeared to be slower than that of peptide−peptide ligation,

10.1021/ja039915e CCC: $27.50

Table 2. MALDI-TOF Mass Analysis of the Sortase-Mediated Conjugation of GFP-LPETG-6His with Various Substrates

nucleophile	ligation product	observed	calculated
–	GFP-LPETG-6His	28,496 ± 8	28,503
H_2O	GFP-LPET	27,370 ± 8	27,381
G_1Pep	GFP-LPET-G_1Pep	29,013 ± 14	29,008
G_2Pep	GFP-LPET-G_2Pep	29,060 ± 5	29,065
G_3Pep	GFP-LPET-G_3Pep	29,124 ± 12	29,122
Gly-emGFP	GFP-LPET-G-emGFP	56,216 ± 41	56,290
(D)-Tat	GFP-LPET-(D)-Tat	28,952 ± 14	28,962
G_3K(folate)	GFP-LPET-G_3K(folate)	28,106 ± 2	28,103
G_2Y−PTD5	GFP-LPET-[G_2Y−PTD5]	29,246 ± 6	29,255
AT-P-022	GFP-LPET-[AT-P-022]	31,341 ± 18	31,335

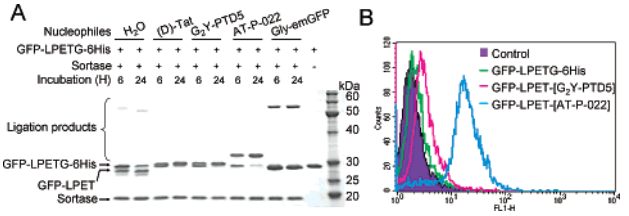

Figure 2. Application of sortase-based ligation to protein. (A) Sortase-mediated ligation of GFP-LPETG-6His with peptide (D)-Tat (gygrkkrrqrrr), G_2Y-PTD5 (GGYRRQRRTSKLMKR), a branched peptide AT-P-022 {(PTD5-Ahx)$_2$KYK(G_2)}, or protein Gly-emGFP. (B) Flow cytometry profile measuring the efficiency of transduction of GFP conjugates in NIH3T3 cells.

presumably because of the increased steric hindrance provided by the protein.

Having demonstrated its utility in the ligation of native peptide sequences, we asked whether sortase was able to ligate nonnative peptide fragments, including (D)-peptides and nonpeptidic molecules to an LPXTG motif. First, we repeated the above peptide−peptide and protein−peptide ligation with a (D)-Tat peptide (gygrkkrrqrrr) as the nucleophile. Remarkably, in both cases sortase was able to conjugate (D)-Tat to the C terminus of the LPXTG motif (Tables 1 and 2, Figure 2A, and Supporting Information) although the ligation rate was cut in half (see Supporting Information). We then hypothesized that by adding a di- or triglycine to the N terminus of a small molecule, it should be conjugated by sortase to the C terminus of an LPXTG motif. Indeed, by derivatizing folate with an N-terminal triglycine via a lysine {G_3K(folate)}, we have successfully obtained a GFP−folate conjugate through the sortase ligation (Table 2).

To demonstrate the functional utility of the sortase ligation, we have applied the process to the synthesis of protein−peptide conjugates that would be difficult to obtain by other means. The synthetic cationic peptide RRQRRTSKLMKR (PTD5) has been shown to possess protein transduction activity.[20] An improved version, a branched synthetic peptide (PTD5-Ahx)$_2$KYK, was found to confer 14-fold better efficiency for cell loading than a single PTD5 moiety (manuscript in preparation). While a fusion protein containing a single PTD5 sequence may be easily generated by recombinant expression, a protein containing more complex and efficient PTD moieties cannot be easily made using either the recombinant expression or chemical synthesis. We sought to use our sortase-based ligation method to generate a conjugate between the synthetic branched PTD peptide AT-P-022 {(PTD5-Ahx)$_2$KYK-(G_2)} and the recombinant GFP-LPETG-6His. As a comparator, we also conjugated a linear peptide G_2Y-PTD5 with GFP-LPETG-6His. We found that the conjugation efficiency with the branched peptide was similar to that of a linear peptide (G_2Y-PTD5) (Figure 2A). The single linear PTD5- and branched PTD5-conjugated proteins were subsequently purified and incubated with NIH3T3 cells. Cells incubated with the GFP-branched PTD conjugate contain significantly more GFP fluorescence (13-fold) than those incubated with an equal concentration of GFP-linear PTD conjugate (Figure 2B). Together these results demonstrate that sortase can be used to generate biologically useful protein conjugates that are difficult or impossible to make otherwise.

In conclusion, we have demonstrated a site-specific, robust, and reliable method for protein ligation using bacterial sortase. This method can be applied to specific protein conjugation with a wide range of polypeptides bearing different biological or biophysical properties and appears to be applicable to the efficient production of novel natural−nonnatural hybrid macromolecules.

Acknowledgment. We appreciate Dr. Thomas Machleidt for his assistance in the areas of cell culture and flow cytometry analysis. We thank Dr. David Stolow and Dr. Karin Zeh for their critical comments of the manuscript. We also thank the entire Ansata Therapeutics team for their valuable suggestions on applications to the sortase protein ligation technology.

Supporting Information Available: Experimental materials and methods including the protein expression, conjugation, fluorescence measurement, MALDI-TOF data, HPLC profiles as well as cell assays (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Tam, J. P.; Xu, J.; Eom, K. D. *Biopolymers* **2001**, *60*, 194−205.
(2) Dawson, P. E.; Kent, S. B. *Annu. Rev Biochem.* **2000**, *69*, 923−960.
(3) Kochendoerfer, G. G.; Chen, S. Y.; Mao, F.; Cressman, S.; Traviglia, S.; Shao, H.; Hunter, C. L.; Low, D. W.; Cagle, E. N.; Carnevali, M.; Gueriguian, V.; Keogh, P. J.; Porter, H.; Stratton, S. M.; Wiedeke, M. C.; Wilken, J.; Tang, J.; Levy, J. J.; Miranda, L. P.; Crnogorac, M. M.; Kalbag, S.; Botti, P.; Schindler-Horvat, J.; Savatski, L.; Adamson, J. W.; Kung, A.; Kent, S. B.; Bradburne, J. A. *Science* **2003**, *299*, 884−887.
(4) Blaschke, U. K.; Silberstein, J.; Muir, T. W. *Methods Enzymol.* **2000**, *328*, 478−496.
(5) Ayers, B.; Blaschke, U. K.; Camarero, J. A.; Cotton, G. J.; Holford, M.; Muir, T. W. *Biopolymers* **1999**, *51*, 343−354.
(6) Cotton, G. J.; Muir, T. W. *Chem. Biol.* **2000**, *7*, 253−261.
(7) Iakovenko, A.; Rostkova, E.; Merzlyak, E.; Hillebrand, A. M.; Thoma, N. H.; Goody, R. S.; Alexandrov, K. *FEBS Lett.* **2000**, *468*, 155−158.
(8) Sydor, J. R.; Mariano, M.; Sideris, S.; Nock, S. *Bioconjug. Chem.* **2002**, *13*, 707−712.
(9) Nyanguile, O.; Dancik, C.; Blakemore, J.; Mulgrew, K.; Kaleko, M.; Stevenson, S. C. *Gene Ther.* **2003**, *10*, 1362−1369.
(10) Jackson, D. Y.; Burnier, J.; Quan, C.; Stanley, M.; Tom, J.; Wells, J. A. *Science* **1994**, *266*, 243−247.
(11) Chang, T. K.; Jackson, D. Y.; Burnier, J. P.; Wells, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 12544−12548.
(12) Rall, K.; Bordusa, F. *J. Org. Chem.* **2002**, *67*, 9103−9106.
(13) Mazmanian, S. K.; Liu, G.; Ton-That, H.; Schneewind, O. *Science* **1999**, *285*, 760−763.
(14) Novick, R. P. *Trends Microbiol.* **2000**, *8*, 148−151.
(15) Perry, A. M.; Ton-That, H.; Mazmanian, S. K.; Schneewind, O. *J. Biol. Chem.* **2002**, *277*, 16241−16248.
(16) Mazmanian, S. K.; Ton-That, H.; Schneewind, O. *Mol. Microbiol.* **2001**, *40*, 1049−1057.
(17) Ton-That, H.; Mazmanian, S. K.; Faull, K. F.; Schneewind, O. *J. Biol. Chem.* **2000**, *275*, 9876−9881.
(18) Huang, X.; Aulabaugh, A.; Ding, W.; Kapoor, B.; Alksne, L.; Tabei, K.; Ellestad, G. *Biochemistry* **2003**, *42*, 11307−11315.
(19) Kruger, R. G.; Dostal, P.; McCafferty, D. G. *Chem. Commun. (Cambridge)* **2002**, 2092−2093.
(20) Mi, Z.; Mai, J.; Lu, X.; Robbins, P. D. *Mol. Ther.* **2000**, *2*, 339−347.

JA039915E