Identification of an Archaeal α-L-Fucosidase Encoded by an Interrupted Gene

PRODUCTION OF A FUNCTIONAL ENZYME BY MUTATIONS MIMICKING PROGRAMMED −1 FRAMESHIFTING*S

Received for publication, November 20, 2002, and in revised form, February 4, 2003
Published, JBC Papers in Press, February 4, 2003, DOI 10.1074/jbc.M211834200

Beatrice Cobucci-Ponzano‡, Antonio Trincone§, Assunta Giordano§, Mosè Rossi‡¶, and Marco Moracci‡¶

From the ‡Institute of Protein Biochemistry, Consiglio Nazionale delle Ricerche, Via P. Castellino 111, 80131 Naples, Italy, the §Istituto di Chimica Biomolecolare, Consiglio Nazionale delle Ricerche, Via Campi Flegrei 34, 80078 Pozzuoli (NA), Italy, and the ¶Dipartimento di Chimica Biologica Università di Napoli “Federico II,” Via Mezzocannone 16, 80131 Naples, Italy

The analysis of the complete genome of the thermoacidophilic Archaeon Sulfolobus solfataricus revealed two open reading frames (ORF), named SSO11867 and SSO3060, interrupted by a −1 frameshift and encoding for the N- and the C-terminal fragments, respectively, of an α-L-fucosidase. We report here that these ORFs are actively transcribed in vivo, and we confirm the presence of the −1 frameshift between them at the cDNA level, explaining why we could not find α-fucosidase activity in S. solfataricus extracts. Detailed analysis of the region of overlap between the two ORFs revealed the presence of the consensus sequence for a programmed −1 frameshifting. Two specific mutations, mimicking this regulative frameshifting event, allow the expression, in Escherichia coli, of a fully active thermophilic and thermostable α-L-fucosidase (EC 3.2.1.51) with micromolar substrate specificity and showing transfucosylating activity. The analysis of the fucosylated products of this enzyme allows, for the first time, assigning a retaining reaction mechanism to family 29 of glycosyl hydrolases. The presence of an α-fucosidase putatively regulated by programmed −1 frameshifting is intriguing both with respect to the regulation of gene expression and, in post-genomic era, for the definition of gene function in Archaea.

It has been known that carbohydrates can serve as structural components of natural products, as energy sources, or more interestingly, as key elements in various molecular recognition processes. In this regard, α-L-fucose is an important constituent of the carbohydrate chains of glycoconjugates involved in a variety of biological events as growth regulators and receptors in signal transduction, cell-cell interactions, and antigenic response (1).

In plants, α-L-fucosylated oligosaccharides derived from xyloglucan, a plant cell wall component that controls cell expansion, have been shown to regulate auxin- and acid pH-induced growth (2). In mammals, oligosaccharides containing fucose have been found, for instance, in human milk and in blood group substances (3), and they are reported to play important roles in fertilization (4) and in adhesion processes of viruses, bacteria, and other parasites (5). Changes in fucosylation patterns have been observed in several physiological events including pregnancy (6), programmed cell death of different cell types (7), and in a variety of pathological events including diabetes (8) and colon and liver carcinomas (9, 10). In addition, the determination of α-fucosidase activity can be used to predict the development of colorectal, ovarian, and hepatocellular carcinomas (11–13), whereas the deficiency in this enzyme causes fucosidosis, a well known lysosomal storage disorder (14). The central role of fucose derivatives in biological events explains the interest in α-L-fucosidase and fucosyltransferase activities.

Family 29 of glycosyl hydrolases classification (GH29)1 groups α-fucosidases (EC 3.2.1.51) from a variety of sources, including human and several pathogenic bacteria (15). No structural data are available about this class of enzymes, and the residues involved in catalysis are still unknown. Recently, it has been shown (16) that the α-fucosidase from Thermus sp. Y5 performs the hydrolytic reaction with the retention of the anomeric configuration. The analysis of the genome of the hyperthermophilic Archaeon Sulfolobus solfataricus (17) revealed the presence of two ORFs, annotated as SSO11867 and SSO3060, separated by a −1 frameshift, that are homologous to the N- and the C-terminal fragments, respectively, of full-length bacterial and eukaryal GH29 fucosidases.

In the genome of this Archaeon several interrupted genes, conserved in distantly related Archaea, have been identified, arguing that these interruptions were not sequencing mistakes. Rather, since no multiple interruptions could be found, a selection toward a single conserved shift occurs, suggesting a

*S This work was supported by Agenzia Spaziale Italiana Project "Extremophilic Archaea as Model Systems to Study Origin and Evolution of Early Organisms: Molecular Mechanisms of Adaptation to Extreme Physical-Chemical Conditions" Contract I/R/365/02 and by MIUR Project RBAU015B47 “Folding di Proteine: l’Altra Meta ‘ del Codice Genetico.” The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

This paper is dedicated to the memory of Eraldo Antonini, eminent biochemist, prematurely deceased 20 years ago, on March 19th, 1983. The on-line version of this article (available at http://www.jbc.org) contains the NMR spectra of the products obtained.

1 To whom correspondence should be addressed: Institute of Protein Biochemistry-CNR, Via P. Castellino 111, 80131, Naples, Italy. Tel.: 39-081-6132271; Fax: 39-081-6132277; E-mail: moracci@dafne.ibpe.na.cnr.it.

1 The abbreviations used are: GH29, family 29 of glycosyl hydrolases; ORF, open reading frames; Ssa-fuc, mutant S. solfataricus α-fucosidase; pNP-Fuc, p-nitrophenyl-α-L-fucopyranoside; pNP-Glc, p-nitrophenoxy-α-L-glucopyranoside; Sβ-gly, S. solfataricus β-glycosidase; PBS, phosphate-buffered saline; RT, reverse transcriptase.
conserved translational regulation mechanism. The maintenance of a correct reading frame is fundamental to the integrity of the translation process; nevertheless, an increasing number of cases have been described in which localized deviations from the standard translational rules are used to regulate the correct expression of a minority of genes (18). These events, named recoding, are used to increase the diversity in gene expression or for its regulation, and they include programmed ribosome frameshifting from different reading frames, ribosome hopping over nucleotides, and reading of stop codons as sense codons (readthrough). Among recoding events, programmed –1 frameshifting are by far the most prevalent (19); they have been well characterized in RNA viruses, but the basic molecular mechanisms governing these events are almost identical from yeast to humans (20). In Archaea, documented recoding events are limited to readthrough (21); no proofs of programmed –1 frameshifting have been reported in this domain so far. To test the hypothesis that SSO11867 and SSO3060 ORFs could express a functional α-fucosidase, we have cloned and expressed these genes in Escherichia coli. Two specific mutations, designed on the basis of the programmed –1 frameshifting mechanism, allow the expression of a full-length thermophilic and thermostable α-fucosidase, with micromolar substrate specificity, which promotes transfucosylation reactions by following a retainer reaction mechanism. This is the first evidence of an α-fucosidase from Archaea, and our data unequivocally demonstrate that GH29 enzymes follow a retaining reaction mechanism. Furthermore, the data presented give support to the hypothesis that translational recoding could be used to regulate gene expression in Archaea.

EXPERIMENTAL PROCEDURES

Archaeal Strain and Cultivation—S. solfataricus cells, strain MT4, were grown at 87 °C, pH 3.0, as reported previously (22); instead, strain P2 was grown at 80 °C, pH 3.5, in a minimal salts culture medium supplemented with yeast extract (0.1%), casamino acids (0.1%), plus glucose (0.1%). Growth was monitored spectrophotometrically at 600 nm and stopped when it reached OD 0.6–0.8.

Substrates—All commercially available substrates were purchased from Sigma.

α-Fucosidase Expression Analysis in S. solfataricus—Genomic DNA from S. solfataricus MT4 and P2 strains was prepared by the NucleoSpin Tissue kit (Macherey-Nagel, Germany). The DNA fragment of 833 nucleotides (positions 1–833), containing the region overlapping the ORFs SSO11867 and SSO3060, was prepared by PCR, by using the following synthetic oligonucleotides (Genenco, Florence, Italy): α-fuc-3, 5′-GAGGAAGATCTATCTATCTTAT-CTTCTAGGTATACCC-3′; α-fuc-2, 5′-CGCGGATCCCTATCTATCTTTAATCTAGGTATACCC-3′; 5′-fuc-1, 5′-CGCGGATCCCTATCTATCTTTAATCTAGGTATACCC-3′; 5′-fuc-2, 5′-CGCGGATCCCTATCTATCTTTAATCTAGGTATACCC-3′; 5′-fuc-3, 5′-CGCGGATCCCTATCTATCTTTAATCTAGGTATACCC-3′; 5′-fuc-4, 5′-CGCGGATCCCTATCTATCTTTAATCTAGGTATACCC-3′. The DNA fragment containing the mutations was subcloned in the vector pGEX-11867/3060, GST was fused to the N-terminal of the SSO11867 gene product; the fusion and the entire DNA fragment obtained after amplification were controlled by DNA sequencing.

Site-directed Mutagenesis—The vector pGEX-11867/3060, described above, was used as template for site-directed mutagenesis experiments (24). The synthetic oligonucleotides used were α-fuc-3 and α-fuc-4, and were used to amplify mutagenic oligonucleotide FrameFuc-1: 5′-ACTGGGCCGAAATTCTTTAGGTGATATTGG-3′ and FrameFuc-2: 5′-GGAGGAAGATCTATCTATCTTTAATCTAGGTATACCC-3′. Both oligonucleotides were synthesized by Bioneer (Daedeok, Korea) and by DNA synthesis (Geneco, Florence, Italy) using a standard program, and the synthetic oligonucleotide α-fuc-3 was subcloned in the pGEX-2TK plasmid (Novagen), obtaining the vector pET-3060, in which SSO3060 ORF is under the control of the isopropyl-1-thio-β-D-galactopyranoside-inducible T7 RNA polymerase promoter that drives high expression levels in bacterial hosts. The ORF obtained after amplification was controlled by DNA sequencing.

Protein Purification—E. coli BL21(DE3)pET-3060 was grown in 2 liters of Super Broth at 37 °C. Gene expression was induced by the addition of 1 mM isopropyl-1-thio-β-D-galactopyranoside when the culture reached an A600 of 1.0. Growth was allowed to proceed for 16 h, and cells were harvested by centrifugation at 5,000 × g and frozen at −20 °C. The resulting cell pellet was thawed, resuspended in 2 ml g-1 of 100 mM sodium phosphate buffer, pH 7.4, 150 mM NaCl, 0.1 mM β-galactoside (pNp-Fuc) at 4 °C, and homogenized by treatment with a cell disruption equipment (Constant Systems Ltd., Warwick, UK). After disruption, the homogenate was diluted 1:1 with the same buffer and centrifuged for 30 min at 30,000 × g; cell debris was discarded, and the crude extract was diluted 10-fold in the same buffer; no activity on pNP-Fuc at 65 °C was observed (see below for the standard enzymatic assay).

Growth of E. coli BL21(RB791/pGEX-11867/3060 and total proteins extraction was performed as described for pET-3060. The crude extract was applied to a glutathione-Sepharose 4B column (Amersham Biosciences) that had been equilibrated with the same buffer. After 10 column volumes washing with PBS buffer (without detergent), the fusion protein was eluted from the column by the addition of 500 mM Tris-HCl, pH 8.0, supplemented with 10 mM reduced glutathione, at room temperature (22–25 °C); the eluate is collected in 1.5-mL volume fractions and assayed for GST activity at 25 °C. Active fractions were collected and stored at 4 °C. The active pool was then subjected to thrombin treatment. For this aim, pooled fractions (about 30 mL) were incubated at 4 °C overnight with 30 units of thrombin solution (Amersham Biosciences). This sample, analyzed by SDS-PAGE, did not reveal any band compatible with the molecular weight of the full-length α-fucosidase but showed low activity on pNP-Fuc at 65 °C.

Mutant α-fucosidase was expressed from E. coli BL21(RB791/pGEX-11867/3060 and extracted as for pET-3060. The GST binding was performed by adding 3 ml of the glutathione-Sepharose 4B matrix (Amersham Biosciences), equilibrated with the same buffer, to the crude extract and incubated overnight at 4 °C. After the binding, the matrix was packed, and after 30 column volumes washing with PBS, (each at 0.1%), and different combinations of these). Enzymatic assays in S. solfataricus extracts were performed by using up to about 0.6 mg of crude extracts in standard conditions (see below for the standard enzymatic assay) on p-nitrophenyl-α-D-fucopyranoside (pNP-Fuc) at 75 °C for up to 5 h; at these conditions the hydrolysis of the substrate was identical to that of the blank mixture without protein.

Primer extension experiments were performed as reported previously (23): the synthetic oligonucleotides (Genenco, Florence, Italy) used are as follows: Fext1, 5′-CAGCTTTTTCATGAAACCTTAG-3′; Fext2, 5′-GACTATATCAGTTTTATAGAAC-3′.
buffer (without Triton), the matrix was resuspended in 1 volume of PBS buffer and incubated overnight at 4 °C with 60 units of thrombin solution. The efficiency of thrombin cleavage was tested by loading onto SDS-PAGE an amount of the matrix slurry before and after the thrombin treatment. Thereafter, the soluble and GST-free /H9251-fucosidase protein was recovered by 5 column volumes washes with PBS buffer. Washes containing the /H9251-fucosidase protein were pooled and concentrated by ultrafiltration on an Amicon YM30 membrane (cut-off 30,000 Da). After this treatment, the /H9251-fucosidase was 95% pure by SDS-PAGE and was used for all the subsequent characterizations. The purification procedure yielded about 2 mg of pure protein from 13.7 g of wet cell pellet. The sample stored at 4 °C is stable for several months. Direct sequencing of the N-terminal of the purified enzyme produced the sequence: Ser-Val-Gly-Ser-Met-Ser-Gln-Asn-Ser-Tyr-Lys-Ile-Leu-Lys-, in which the underlined amino acids correspond to the N terminus of SSO11867.

Enzyme Characterization—The standard assay of /H9251-fucosidase activity was performed at 65 °C in 50 mM sodium phosphate buffer at pH 6.3, with pNp-Fuc substrate at the final concentration of 1 mM. The molar extinction coefficient of p-nitrophenol is 9340 M^-1cm^-1 measured at 405 nm, at 65 °C, in 50 mM sodium phosphate buffer, pH 6.5. One unit of enzyme activity was defined as the amount of enzyme catalyzing the hydrolysis of 1 µmol of substrate in 1 min at the conditions described. Spontaneous hydrolysis of the substrate (about 0.1%) was subtracted by using appropriate blank mixtures without enzyme. The kinetic constants of the /H9251-fucosidase mutant (Ss/H9251-fuc) on pNp-Fuc were measured at 65 °C in 50 mM sodium phosphate buffer, pH 6.3, by using substrate concentrations in the range 0.005–3 mM. The protein concentration in the reaction mixture was 0.4 g ml^-1. All kinetic data were calculated as the average of at least two experiments and were plotted and refined with the program GraFit (25).

Thermal activity of Ss/H9251-fuc was analyzed by assaying the enzyme (0.8 g) on pNp-Fuc substrate concentrations of 1 and 3 mM in the temperature ranges of 40–65 and 70–95 °C, respectively. Thermal stability was tested by incubating pure enzyme (0.08 mg ml^-1) in PBS buffer at the indicated temperatures, as reported previously (26).

Molecular mass of denatured Ss/H9251-fuc was determined by SDS-PAGE in both reducing and non-reducing conditions. Molecular mass of native Ss/H9251-fuc was determined by gel filtration on a Superose 6 column 26/60 HiLoad (Amersham Biosciences) runs in PBS buffer at a 0.3 ml min^-1 flow rate; molecular weight markers were run under the same conditions.

Enzymatic Synthesis of Fucosides—Thin layer chromatography and NMR spectra were performed as reported previously (27). The pNp-Fuc
The reaction was started at 75°C; enzyme preparation (0.08 mg of total protein) were added, and the consumption of donor as monitored by TLC (CHCl₃/MeOH/H₂O, 65:25:4, v/v). The total reaction mixture was cooled, and the product was purified by RP-18 column chromatography; the fractions containing fucosylated products were purified by preparative TLC. After acetylation, the following compounds were obtained in 14% total yield with respect to the donor.

For α-L-Fuc(1′-3)-α-L-Gloce-O-pNp, COSY and 1H-13C NMR correlations permit the assignment of 1H and 13C signals of carbohydrate moiety as follows. Gloc moiety: 5.74 (94.2) H1, 5.09 (72.2) H2, 4.37 (74.1) H3, 5.19 (69.1) H4, 3.89 (69.16) H5, 4.13–4.03 (61.6) H6a, H6b; α-L-Fuc moiety: 5.46 (96.5) H1, 5.06 (68.6) H2, 5.24 (67.2) H3, 5.27 (70.8), H4, 4.12 (65.3) H5, 1.13 (15.9) H6. [α]D₂₀ = -29.8 CHCl₃, c = 2.2.

For α-L-Fuc(1′-2)-α-L-Gloce-O-pNp, COSY and 1H-13C NMR correlations permit the assignment of 1H and 13C signals of carbohydrate moiety as follows. Gloc moiety: 5.77 (95.6) H1, 3.95 (76.7) H2, 5.67 (70.9) H3, 5.11 (70.6), H4, 3.99 (68.5) H5, 4.03–4.24 (64.3) H6a, H6b. α-L-Fuc moiety: 5.24 (98.3) H1, 5.08 H2, 5.17 H3, 5.19 H4 (67.9,67.4, 70.4 overlapping correlation), 4.95 (65.5) H5, 0.71 (15.6) H6. [α]D₂₀ = -18.7 CHCl₃, c = 2.2.

During pNp-Fuc hydrolysis reactions, the formation of a disaccharidic component was noticed along with fucose. Purification, acetylation, and NMR characterization of the disaccharidic compound as reported above revealed it to be α-L-Fuc(1′-3)-α-L-Fuc-O-pNp.

For α-L-Fuc(1′-2)-α-L-Fuc-O-pNp, CH₃ (CO)S is of disaccharidic moiety; 5.90 (H-1 aryl-linked fucose), 5.32 (H-2 aryl-linked fucose), 5.30 (H-4 aryl-linked fucose), 5.27 (2H, H3-H4 external fucose unit), 5.22 (H2 external fucose) 5.20 (H-1 external fucose), 4.45 (H-3 aryl-linked fucose), 4.41 (H-5 aryl-linked fucose) 4.06 (H-5 external fucose) 1.21 (H-6 aryl-linked fucose) 1.11 (H-6 external fucose). 13C NMR signals: 94.5, 93.9, 92.8, 71.0, 69.8, 69.1, 69.0, 67.7, 67.6, 66.4, 64.8, 15.9, 16.0.

RESULTS

Identification and Sequence Analysis of the α-Fucosidase Locus—S. solfataricus, strain P2, is a hyperthermophilic archaeon able to grow at acidic pH (pH 3–5) and at high temperatures (80–87°C). In an effort to determine the full set of glycosyl hydrodases produced by this archaeon, we analyzed the ORFs putatively encoding for these enzymes in the sequenced genome (17). Two of these ORFs, SSO11867 and SSO3060, encode for 81 and 426 amino acids polypeptides, respectively, and are homologous to the N- and C-terminal parts, respectively, of GH29 α-fucosidases (Fig. 1) (15). SSO11867 and SSO3060 are separated by a 1-frame shift in a 40-base overlap (Fig. 2); this is consistent with the observation that no α-fucosidase activity could be found in S. solfataricus extracts obtained from cells grown on different media (glucose, fucose, or sucrose as the only energy sources or in combination with yeast extract and casamino acids). To test if the frameshift was due merely to sequencing errors, DNA fragments containing the overlapping region were amplified from the genome of S. solfataricus P2 and MT4 strains, which are strictly taxonomically related (29) and directly sequenced. The sequences were identical to the one published (Fig. 3A), confirming the presence of a 1-frame shift between these two ORFs. Moreover, a cDNA fragment was amplified by RT-PCR from a total RNA preparation of S. solfataricus, strain P2, and directly sequenced (Fig. 3B). Again, the obtained sequence turned out to be identical to that available from the data bank with no ambiguities (Fig. 3C), indicating that the population of RNA amplified by RT-PCR was identical to the genomic DNA; this result excludes the possibility of RNA-editing events and demonstrates that SSO11867 and SSO3060 are co-transcribed in vivo.

Experiments of primer extension on the same total RNA preparations were performed with oligonucleotides that anneal on the SSO11867 and SSO3060 ORFs (Fig. 4). The primer Fext2 showed a specific transcriptional initiation site nine nucleotides upstream from the first ATG of the SSO11867 ORF (Fig. 4); this was confirmed by primer Fext1, which, however, showed additional initiation sites (Fig. 4). The strongest signal showed by Fext1 maps on the third base of the first putative ATG of the C-terminal ORF SSO3060 (Fig. 4). However, the expression of the ORF SSO3060 in E. coli as pET3060 did not produce any protein band of the expected molecular weight (46.5 kDa), and no detectable α-fucosidase activity on the pNp-Fuc substrate at 65°C was found. This suggests that this ORF could not express a functional enzyme independently; this is not surprising because in the N-terminal ORF SSO11867 an amino acidic sequence conserved among GH29 α-fucosidases can be found (Fig. 1).

Potential promoter sequences were found at −26 nucleotides from the transcription initiation site upstream from SSO11867, whereas no such consensus could be identified in the region upstream from SSO3060 (Fig. 4).

Preparation of a Full-length α-Fucosidase—The complete S. solfataricus α-fucosidase locus, with SSO11867 and SSO3060 in different reading frames (pGEX-11867/3060), drives the expression in E. coli of trace amounts of α-fucosidase activity; after removal of GST a specific activity of 2.3 × 10⁻⁴ units mg⁻¹ on pNp-Fuc at 65°C was found. A detailed analysis of the DNA sequence of the region of overlap between the two ORFs revealed the presence of a stretch of six adenosines followed by a thimine (Fig. 4) that resembles one of the heptamers that are involved in programmed −1 frameshifting (19). Typically, the sites cis-regulating these events consist of a "slippery" heptameric sequence of the general form X-XY-YZ (codons are in the zero frame, and X, Y, and Z can be identical or different nucleotides) and often include an upstream Shine-Dalgarno sequence and a downstream mRNA secondary structure. In fact, it has been reported that a Shine-Dalgarno sequence along the mRNA and pseudoknots or stem-loops promote the pausing of the ribosome (19). The sequence of overlap between the ORFs SSO11867 and SSO3060 presents a similar organization. The slippery sequence AAA-AAT was immediately followed by a stem loop, and the rare codon CAC, which is used at low frequency by S. solfataricus (4.7 per...
thousand in the *S. solfataricus* genome), was found upstream from the slippery sequence (Fig. 2). This rare codon presumably plays the function of the Shine-Dalgarno sequence, which is rarely observed in isolated genes in this Archaeon (30), by inducing the pausing of the ribosome and increasing the frequency of the frameshifting event. These observations raised the hypothesis that a programmed frameshift could promote the expression of an active enzyme in *S. solfataricus*; however, testing this hypothesis in vivo is impaired by the lack of molecular genetic tools for hyperthermophilic Archaea.

To test whether a functional α-fucosidase could be produced, a single frame between the ORFs was restored by site-directed mutagenesis. In the mechanism of programmed −1 frameshifting proposed for Eukarya and bacteria, two tRNAs, hybridized to the XXY and YYZ codons of the X-XXY-YYZ sequence, are proposed to slip simultaneously backwards on the mRNA to the −1 frame, hybridizing to XXX and YYY codons. The AAT triplet, coding for Asn-78 in SSO11867, corresponds to the YYZ codon, and is the last one decoded in the zero frame (19); after this triplet, the ribosome would shift onto the TTC codon of the Phe-10 (SSO3060 numbering) continuing the translation in the −1 frame (Fig. 2). To obtain the fused gene, we performed site-directed mutagenesis in the pGEX-11867/3060 vector in which glutathione S-transferase (GST) enzyme was fused to

---

**Fig. 3.** Sequence analysis of the α-fucosidase locus in *S. solfataricus*. A, electropherogram of the DNA sequence of the PCR products obtained from *S. solfataricus* genome, strain P2; the region of overlap between the two ORFs SSO11867 and SSO3060 is underlined. B, agarose gel showing the products of RT-PCR from total cellular RNA; lane 1, control (amplification of total RNA without reverse transcriptase enzyme); lane 2, α-fucosidase; lane 3, XylS (positive control); lane 4, molecular weight markers, 100-bp DNA ladder. C, electropherogram of the DNA sequence of the RT-PCR product shown in B, lane 2; the region of overlap between the two ORFs SSO11867 and SSO3060 is underlined.
the N-terminal of SSO11867. On the basis of the mechanism proposed, a T in the region following the slippery heptamer was introduced; moreover, we introduced the conservative mutation AAAAAG (encoding for Lys-77 in SSO11867) to increase the translational fidelity by disrupting the heptameric sequence. These mutations changed the putative frameshifting site from CTA-AAA-AAT-TCG-GCC (zero frame, slippery heptamer underlined) to CTA-AAG-AAT-TTC-GGC (the nucleotides in boldface were originally in the frame; the mutations are underlined). As a result, the ORFs are in the same translation frame producing a single polypeptide.

The obtained pGEX-frameFuc plasmid was used to express the mutant enzyme in *E. coli*. An intense protein band of molecular weight compatible with the GST-\(\alpha\)-fucosidase fusion (84.4 kDa) was observed after affinity chromatography of the crude extracts on glutathione-Sepharose 4B (Fig. 5); the removal of GST by thrombin cleavage was performed directly onto the column producing, by a single purification step, a protein more than 95% pure (Fig. 5). The recombinant enzyme showed the expected 57-kDa molecular mass (corresponding to the predicted full-length polypeptide of 495 amino acids) and revealed a specific activity of 32.1 units mg\(^{-1}\) on pNP-Fuc at 65 °C; it was termed Ssa-fuc.

Characterization of the \(\alpha\)-Fucosidase—The native molecular weight of the enzyme, expressed and purified as described above, was analyzed by gel filtration under native conditions: a single peak containing thermophilic \(\alpha\)-fucosidase activity eluted between thyroglobulin (660 kDa) and ferritin (490 kDa) for a molecular mass of about 32.1 units mg\(^{-1}\) on pNP-Fuc at 65 °C; it was termed Ssa-fuc.

**Fig. 4.** Analysis of transcriptional initiation sites in SSO11867 and SSO3060 ORFs by primer extension. Top, primer extension: lanes 1 and 2, primer extension controls (without total RNA) of Fext1 and Fext2 oligonucleotides, respectively; lanes 3 and 4, primer extension of Fext1 and Fext2 oligonucleotides, respectively. The signals corresponding to the initiation sites upstream from SSO11867 (site A) and SSO3060 (site B) are indicated by closed arrows. The asterisks indicate an aspecific signal of Fext1 observed also in the control. Bottom, \(\alpha\)-fucosidase locus in S. solfataricus: the TATA box and the TFB-responsive element (BRE) are boxed; the first ATGs of SSO11867 and SSO3060 are underlined with a dark line; the slippery heptameric sequence is underlined; the amino acids involved in the programmed –1 frameshift are shaded. The initiation sites A and B are shown with closed arrows. The sequences corresponding to Fext1 and Fext2 oligonucleotides, mapping in the C-terminal SSO3060 and N-terminal SSO11867 ORFs, respectively, are underlined with open arrows.
buffer systems (data not shown). The thermal activity of Ss-fuc is reported in Fig. 7A; the activity on pNP-Fuc increased sharply up to the optimal temperature of 95 °C, the highest temperature tested. The calculated value of activation energy ($E_a$) for the hydrolytic reaction of this substrate was $91 \pm 2 \text{ kJ mol}^{-1}$.

The residual activity of Ss-fuc after preincubation at temperatures ranging from 75 to 95 °C was followed for up to 2 h on pNP-Fuc at 65 °C (Fig. 7B); the enzyme displayed high stability at 75 °C, showing even a 40% activation after 30 min of incubation and maintaining 60% residual activity after 2 h at 80 °C. The calculated activation energy for the inactivation reaction, calculated from the Arrhenius plot shown in the inset of Fig. 7B, turned out to be $1120 \pm 167 \text{ kJ mol}^{-1}$, more than 10-fold the $E_a$ measured for the catalyzed reaction, confirming the extreme thermal stability of Ss-fuc.

**Fig. 5.** SDS-PAGE analysis of expressed α-fucosidase. Lane 1, slurry of glutathione-Sepharose 4B matrix with bound GST-α-fucosidase fusion expressed from *E. coli* BL21(RB791)/pGEX-frameFuc, before the thrombin cleavage; lanes 2 and 3, same as lane 1, first and second elution, respectively, after thrombin treatment; lanes 4 and 5, purified Ss-fuc 2 and 4 µg, respectively. The molecular weight markers used are as follows: phosphorylase b (97,000), bovine serum albumin (66,000), ovalbumin (45,000), and carbonic anhydrase (30,000).

**Fig. 6.** Gel filtration analysis of Ss-fuc. Elution profile at 280 nm from a Superose 6 column. The elution volumes are as follows: 7.39 ml of blue dextran; 12.09 ml of thyroglobulin (669,000); 13.53 ml of Ss-fuc; 14.03 ml of ferritin (490,000); 15.48 ml of aldolase (158,000); 16.01 ml of bovine serum albumin (67,000); 18.46 ml of ribonuclease A (13,700).

**Fig. 7.** Activity and stability of Ss-fuc. A, thermal activity and derived Arrhenius plot (inset) of Ss-fuc on pNP-Fuc. B, thermal stability of Ss-fuc at 75 (open circles), 80 (closed circles), 85 (open squares), 90 (closed squares), and 95 °C (open triangles). The Arrhenius plot derived from the data in the temperature range 80–95 °C is reported in the inset.
The enzyme followed the classical Michaelis-Menten behavior on pNp-Fuc at 65 °C, with $K_m$ and $k_{cat}$ values of 0.028 ± 0.004 mM and 287 ± 11 s⁻¹, respectively, which led to an extremely high $k_{cat}/K_m$ (10,250 s⁻¹ mM⁻¹). Ssa-fuc shows high selectivity for the pNp-Fuc substrate; in fact, even after extensive incubation (more than 2 h) at the optimal conditions, no activity was observed on aryl substrates such as pNp-α-L-arabinoside, pNp-α-D-galactoside, pNp-α-D-glucoside, pNp-α-D-mannoside, pNp-α-D-xylloside, and pNp-α-D-xyloside. No activity was observed on the sulfated polysaccharide fucoidan from Fucus vesiculosus, formed by a repetition of α-(1,3)- and α-(1,4)-fucose residues, and on its products obtained by acetolysis (31) and deprotected by alkaline methanolysis.

Ssa-fuc Promotes Transfucosylation Reactions—In order to test the transfucosylating activity of Ssa-fuc, we investigated the pyranosidic acceptor pNp-Glc. In the reaction in which pNp-Fuc and pNp-Glc were the donor and the acceptor substrates, respectively, in 5 h Ssa-fuc cleaved the donor, forming fucosylated products with 14% total yield with respect to pNp-Fuc. Interglycosidic linkages were determined by NMR spectroscopy (COSY and $^1$H-$^1$C NMR correlation), and the signals were assigned as reported under “Experimental Procedures.” In the COSY spectrum of one of the products, following the correlations through pyranosidic protons of α-D-Glc unit and starting from the anomeric signal at 5.74 ppm ($J = 3.9$ Hz), it is easy to detect H3 proton, which correlates with carbon signal at 74.1 ppm (glycosylation shifts), identifying this compound as α-L-Fucp-(1-3)-α-D-Glc-O-pNp. Using the same reasoning, the interglycosidic linkages of the other disaccharide (α-1–2, 3.95/76.7) were easily assigned to α-L-Fucp-(1-2)-α-D-Glc-O-pNp.

The analysis by TLC of the hydrolysis reaction mixture containing only pNp-Fuc revealed the presence of fucose, of pNp, and of a compound that, after acetylation and NMR spectroscopy, was identified as α-L-Fuc-(1-3)-α-L-Fuc-O-pNp. This disaccharide was used as substrate for enzymatic hydrolysis; interestingly, this reaction mixture revealed, by TLC analysis, the presence of fucose and pNp-Fuc products after incubation at 65 °C for 3 min and the complete hydrolysis of the substrate after 10 min (Fig. 8). This indicated that the enzyme cleaved the aryl disaccharide starting from its non-reducing end.

**DISCUSSION**

Among the disrupted genes in *S. solfataricus*, we have identified two ORFs, SSO11867 and SSO3060, homologous to eukaryal and bacterial α-fucosidases from family 29 of glycosyl hydrolases; the presence of a −1 shift between these ORFs was confirmed, and we found that they are actively co-transcribed in *S. solfataricus*. A complete programmed −1 frameshifting regulation site has been identified, suggesting that SSO11867 and SSO3060 could be expressed in vivo by following this mechanism. Remarkably, the two ORFs express in *E. coli* trace amounts of α-fucosidase activity, confirming that the frameshifting cassette promotes non-regulated frameshifting during overexpression of foreign genes in *E. coli* (19). Primer extension analysis showed two major transcriptional initiation sites: the first mapped upstream from the two ORFs, and the second started from the G of the first ATG of the SSO3060 ORF. This ORF could not express a functional protein with α-fucosidase activity in *E. coli*, suggesting that post-transcriptional cleavage may occur on the longer transcript. However, the possibility that SSO3060 expresses in *S. solfataricus* a polypeptide with a different function, thus explaining the programmed −1 frameshifting regulation, cannot be ruled out; experiments are in progress to study the expression in vivo of these genes.

Two mutations, in key positions to mimic programmed −1 frameshifting, made it possible to drive the expression of the full-length α-fucosidase Ssa-fuc that is optimally active at 95 °C. Remarkably, the enzyme is stable at 80 °C, the temperature at which *S. solfataricus* optimally grows; these observations are strong evidences that the enzyme produced by following the programmed −1 frameshifting could be stable and active in vivo. These results indicate that SSO11867 and SSO3060 ORFs encode for the first archaeal α-fucosidase identified so far, suggesting that they may not be pseudogenes and that programmed −1 frameshifting could be present in this living domain.

The substrate characterization revealed that Ssa-fuc cata-

![Fig. 8. TLC analysis of the hydrolysis of α-L-Fuc-(1-3)-α-L-Fuc-O-pNp.](image)

![Fig. 9. Proposed retaining reaction mechanism of Ssa-fuc.](image)
alyzes the hydrolysis of pNp-Fuc with high efficiency and specificity; in the framework of our mechanistic studies on therophilic glycosyl hydrolases, the reaction mechanism of Ss-α-fucosidase was analyzed in detail. Glycosyl hydrolases follow two distinct mechanisms that are termed retaining or inverting if the enzymatic cleavage of the glycosidic bond liberates a product with the same or the opposite anomeric configuration of the substrate, respectively (32). Both classes of enzymes employ a pair of carboxylic acid residues in catalysis and operate via transition states with substantial oxocarbenium ion character. The alignment of the amino acid sequences of the 24 GH29 α-fucosidases shown in Fig. 1 revealed two residues of aspartic/glutamic acid that are completely conserved in this family (Asp-124 and Asp-146 in Ssa-α-fuc numbering) and are the best candidates to be involved in catalysis as the nucleophile or the acid/base of the reaction (not shown); experiments are currently in progress to test the functional role of these residues.

In retaining enzymes, when acceptors different from water intercept the reactive glycosyl-enzyme intermediate, they promote transglycosylation reactions (Fig. 9). The ability of Ssa-α-fucosidase to function in the transglycosylation mode allowed us to demonstrate experimentally that the enzyme catalyzes the formation of the α-(1,2) and α-(1,3) bonds between fucose and the glucose of pNp-Glc. The α-anomeric configuration of the interglycosidic linkages in the products unequivocally indicates that GH29 α-fucosidases follow a retaining reaction mechanism; this is the first time that the mechanism followed by glycosyl hydrolases from this family has been experimentally demonstrated by following transglycosylation reaction. Moreover, the activity observed on α-L-Fuc-(1–3)α-L-Fuc-O-pNp revealed that Ss-α-fucosidase is an exo-glycosyl hydrolase that attacks the substrates from their non-reducing end. The disaccharide α-L-Fuc-(1–3)α-D-Glc-O-pNp was synthesized previously (33) by using a mesophilic α-fucosidase in a reaction using the fluorodepyrophosphate of the donor and glucose as acceptor in 34% yield. In our case no efforts were made for the optimization of the reaction conditions (keeping as low as possible the acceptor/donor equivalent ratio useful for synthetic application with rare acceptors). However, the thermophilic nature of Ssa-α-fucosidase is of interest for biotechnological exploitation of its transferring capability (34).

Family 29 of glycosyl hydrolases includes enzymes from plants, vertebrates, and pathogenic microbes of plants and humans. The data presented here demonstrate that this enzymatic activity is present in all the three living domains. The ORF TM0306 from the bacterium Thermotoga maritima, putatively encoding for an α-fucosidase, is the only other known example from hyperthermophiles (35); recently, an α-L-fucosidase secreted by the moderate thermophile Thermus sp. Y5 has been isolated and characterized (16). Interestingly, Ssa-α-fucosidase showed the highest amino acid sequence identity (40%) with a putative α-fucosidase from the phytopathogen bacterium Xanthomonas (36), whereas only 25% identity with the T. maritima enzyme was observed; this is surprising because higher similarity between enzymes from hyperthermophiles would be expected.

Structural comparisons of Ssa-α-fucosidase with Thermus sp. enzyme, which is a secreted tetrameric protein, are hampered by the only partial sequence available for the eubacterial enzyme; however, Ssa-α-fucosidase did not show signal peptides for secretion, and no α-fucosidase activity was found in S. solfataricus media.

α-Fucosidases in higher plants and in mammals are associated with different mechanisms of cell growth and regulation, because they are involved in the modification of fucosylated glucans (1). By contrast, at present, there are only limited data on α-fucosidases from eukaryal and bacterial microorganisms. The presence, in S. solfataricus, of glycosylated proteins has been reported recently (37); although the composition of their carbohydrate moiety is unknown, different enzymatic activities for the synthesis, modification, and hydrolysis of these glycosidic bonds should be present in this Archaeon and regulated in some manner. The activity of Ssa-α-fucosidase (efficient hydrolysis of aryl fucosides, very high substrate specificity; transglycosylation capability) may suggest its involvement in vivo in these biological processes.

Relatively little is known about the function of the 22 glycosyl hydrolases of S. solfataricus; several genes encoding for this class of enzymes, including SSO11867 and SSO3060, an α-glucosidase (SSO3051), a β-glucuronidase (SSO3036), a β-xylidosidase (SSO3032), and the clustered α-xylidosidase (XylS), and β-glucosidase (SSp-gly) (SSO3022 and SSO3019, respectively) map in the same region of about 50 kb and are likely to be involved in the degradation of sugars for energy metabolism. In particular, we reported previously that XylS and Ssp-gly hydrolyze xyloglucan oligosaccharides in vitro in a cooperative way (22). In this regard, a similar situation is present in T. maritima, in which the α-fucosidase gene is part of a cluster of six ORFs potentially involved in xyloglucan utilization (22); remarkably, S. solfataricus, strain P2, hydrolyzes xyloglucan from tamarind seed, which is not fucosylated.3 The ability of Ssa-α-fucosidase to hydrolyze short fucosylated oligosaccharides is revealed by the rapid hydrolysis observed using α-L-Fuc-(1–3)α-L-Fuc-O-pNp substrate. The hypothesis that Ssa-α-fucosidase, XylS, and Ssp-gly cooperate in the sequential hydrolytic steps on fucosylated xyloglucans from plant cell walls (38) is tempting and is currently under investigation, in the hope of shedding some light on a novel metabolic pathway in S. solfataricus for the utilization of this hemicellulose as a carbon source.

Acknowledgments—We are grateful to V. Salerno for the total RNA preparations. We thank V. Carratore and L. Camardella for protein sequencing.

REFERENCES

1. Staedtcher, E., Altmann, F., Wilson, I. B., and Marx, L. (1999) Biochim. Biophys. Acta 1473, 216–236
2. de la Torre, F., Sampedro, J., Zarra, I., and Revilla, G. (2002) Plant Physiol. 128, 427–435
3. Stahl, B., Thurl, S., Henker, J., Siegel, M., Finke, B., and Sawatzki, G. (2001) Adv. Exp. Med. Biol. 501, 299–306
4. Mori, E., Hedrick, J. L., Wardrip, N. J., Mori, T., and Takasaki, S. (1998) Glycoconjug. J. 15, 447–456
5. Litsinskij, J. J., Siegel, G. P., and Litsinskij, C. M. (1998) Am. J. Clin. Pathol. 110, 425–440
6. Wiese, T. J., Dunlap, J. A., and Yorek, M. A. (1997) Intern. Med. 345, 827–931
7. Russell, L., Waring, P., and Beaver, J. P. (1998) Biochem. Biophys. Res. Commun. 250, 449–453
8. Wiese, T. J., Dunlap, J. A., and Yorek, M. A. (1997) Biochim. Biophys. Acta 1335, 61–72
9. Rapoport, E. and Pendra, J. L. (1999) Glycobiology 9, 1337–1345
10. Noda, K., Miyoshi, E., Uozumi, N., Gao, C. X., Suzuki, K., Hayashi, N., Hori, M., and Taniguchi, N. (1999) Int. J. Cancer. 75, 444–450
11. Fernandez-Rodriguez, J., Ayude, D., de la Cadena, M. P., Martinez-Zorrano, V. S., de Carlos, A., Caride-Castro, A., de Castro, G., and Rodriguez-Berrollo, F. J. (2000) Cancer Detect. Prev. 24, 143–149
12. Abdel-Alem, H., Ahmed, A., Sabra, A. M., Zakhari, M., Soliman, M., and Hamed, H. (1996) Int. J. Gynecol. Obstet. 55, 273–279
13. Ishizuka, H., Nakayama, T., Matsuoaka, S., Gotoh, I., Ogawa, M., Suzuki, K., Tanaka, N., Tsuchiya, K., Okubo, H., Arakawa, Y., and Okano, T. (1999) Intern. Med. 38, 927–931
14. Michalski, J. C., and Klein, A. (1999) Biochem. Biophys. Acta 1455, 69–84
15. Henrisatt, B. (1991) Biochem. J. 280, 309–316
16. Enevnyaeva E. V., Kulminskaya A. A., Kalikinens N., Nifantiev N. E., Arkhatskii N. P., Saenko A. I., Cheparnaya O. V., and Arutyunyan, A. V. (2001) Glycoconjug. J. 10, 827–834
17. She, Q., Singh, R. K., Confalonieri, F., Zivanovic, Y., Allard, G., Aways, M. J., Chepurnaya, O. V., and Arutyunyan, A. V. (2001) Glycoconjug. J. 10, 827–834
18. B. Cobucci-Ponzano, unpublished results.
18. Baranov, P. V., Gurvich, O. L., Fayet, O., Prere, M. F., Miller, W. A., Gesteland, R. F., Atkins, J. F., and Giddings, M. C. (2001) *Nucleic Acids Res.* 29, 264–267
19. Hammell, A. B., Taylor, R. C., Peltz, S. W., and Dinman, J. D. (1999) *Genome Res.* 9, 417–427
20. Higuchi, R., Krummel, B., and Saiki, R. K. (1988) *Nucleic Acids Res.* 16, 7351–7367
21. Leatherbarrow, R. J. (1992) *GraFit*, Version 3.0, Erithacus Software Ltd., Staines, UK
22. Prisco, A., Moracci, M., Rossi, M., and Ciaramella, M. (1995) *J. Bacteriol.* 177, 1614–1619
23. Higuchi, R., Krummel, B., and Saiki, R. K. (1988) *Nucleic Acids Res.* 16, 7351–7367
24. Leatherbarrow, R. J. (1992) *GraFit*, Version 3.0, Erithacus Software Ltd., Staines, UK
25. Cobucci-Ponzano, B., Moracci, M., Di Lauro, B., Ciaramella, M., D’Avino, R., and Rossi, M. (2000) *Proteins* 48, 98–106
26. Trincone, A., Perugini, G., Rossi, M., and Moracci, M. (2000) *Bioorg. Med. Chem. Lett.* 10, 365–368
27. Carpent, F. (1988) *Nucleic Acids Res.* 16, 10881–10890
28. de Rosa, M., Gambacorta, A., and Bu’Lock, J. D. (1975) *J. Gen. Microbiol.* 86, 156–164
29. Tolstrup, N., Sensen, C. W., Garrett, R. A., and Clausen, I. G. (2000) *Extremophiles* 3, 175–179
30. Wang, L. X., Sakairi, N., and Kuzuhara, H. (1991) *Carbohydr. Res.* 219, 133–148
31. Koshland, D. E., Jr. (1953) *Biol. Rev. Camb. Philos. Soc.* 28, 416–436
32. Parkes, E., Thiem, J., and Ajsaoka, K. (2000) *Carbohydr. Res.* 328, 293–299
33. Fischer, L., Bromann, R., Kengen, S. W. M., de Vos, W. M., and Wagner, F. (1996) *Biotechnology* 14, 88–91
34. Nelson, K. E., Clayton, R. A., Gill, S. R., Gwinn, M. L., Dodson, R. J., Haft, D. H., Hickey, E. K., Peterson, J. D., Nelson, W. C., Ketchum, K. A., McDonald, L., Utterback, T. R., Malek, J. A., Linher, K. D., Garrett, M. M., Stewart, A. M., Cotten, M. D., Pratt, M. S., Phillips, C. A., Richardson, D., Heidelberg, J., Sutton, G. G., Fleischmann, R. D., White, O., Salzberg, S. L., Smith, H. O., Venter, J. C., and Fraser, C. M. (1999) *Nature* 399, 323–329
35. da Silva, A. C., Ferro, J. A., Reinaich, F. C., Farah, C. S., Furlan, L. R., Quaggio, R. B., Monteiro-Vitorello, C. B., Van Sluys, M. A., Almeida, N. F., Alves, L. M., de Amaral, A. M., Bertozzi, M. C., Camargo, L. E., Camarote, G., Cannavan, F., Cardozo, J., Chamberto, F., Ciapina, L. P., Cicarelli, R. M., Coutinho, L. L., Cusiano-Santos, J. R., El-Dorry, H., Faria, J. B., Ferreira, A. J., Ferreira, R. C., Ferro, M. I., Formiglia, E. F., Franco, M. C., Greggor, C. C., Gruber, A., Katsuyama, A. M., Kishi, L. T., Leite, R. P., Lemos, E. G., Lemos, M. V., Locali, E. C., Machado, M. A., Madeira, A. M., Martinez-Rossi, N. M., Martino, E. C., MEDHANIS, J., Meneck, C. F., Miyaki, C. Y., Moon, D. H., Moreira, L. M., Novo, M. T., Okura, V. K., Oliveira, M. C., Oliveira, V. R., Pereira, H. A., Ross, A., Sena, J. A., Silva, C., da Souza, R. F., Spinola, L. A., Takita, M. A., Tamura, R. E., Teixeira, E. C., Teza, R. I., Trindade dos Santos, M., Truffi, D., Tsai, S. M., White, P. F., Setubal, J. C., and Kitajima, J. P. (2002) *Nature* 417, 459–463
36. Lower, B. H., and Kennelly, P. J. (2002) *J. Bacteriol.* 184, 2614–2619
37. Crombie, H. J., Chengappa, S., Hellayer, A., and Reid, J. S. G. (1998) *Plant J.* 15, 27–38
Identification of an Archaeal \(\alpha\)-l-Fucosidase Encoded by an Interrupted Gene:
PRODUCTION OF A FUNCTIONAL ENZYME BY MUTATIONS MIMICKING
PROGRAMMED –1 FRAMESHIFTING
Beatrice Cobucci-Ponzano, Antonio Trincone, Assunta Giordano, Mosè Rossi and Marco Moracci

J. Biol. Chem. 2003, 278:14622-14631.
doi: 10.1074/jbc.M211834200 originally published online February 4, 2003

Access the most updated version of this article at doi: 10.1074/jbc.M211834200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2003/03/24/M211834200.DC1

This article cites 37 references, 7 of which can be accessed free at
http://www.jbc.org/content/278/17/14622.full.html#ref-list-1
Addition and Correction

Vol. 278 (2003) 14622–14631

Identification of an archaeal α-l-fucosidase encoded by an interrupted gene. Production of a functional enzyme by mutations mimicking programmed −1 frameshifting.

Beatrice Cobucci-Ponzano, Antonio Trincone, Assunta Giordano, Mosè Rossi, and Marco Moracci

Page 14629, Fig. 9: The anomic configuration of the α-l-fucoside and of the α-l-fucose is wrong. A corrected version of the figure is shown below.

![Corrected Figure]

R = 4-nitrophenyl
R' = H ➔ hydrolysis
R' = Acceptor ➔ transfucosylation

We suggest that subscribers photocopy these corrections and insert the photocopies at the appropriate places where the article to be corrected originally appeared. Authors are urged to introduce these corrections into any reprints they distribute. Secondary (abstract) services are urged to carry notice of these corrections as prominently as they carried the original abstracts.