Vol. 215, No. 1, 1995 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
October 4, 1995 Pages 48-56

# MOLECULAR CLONING OF A NEW ERYTHROPOIESIS-STIMULATING ACTIVITY (ESA36) FROM *STREPTOMYCES THERMOVIOLACEUS* AND EXPRESSION IN *ESCHERICHIA COLI*[1]

Makiko Sugie, Syuichi Oka†* and Hideoki Tanaka

Biomolecular Engineering Department and *Biosignalling Department,
National Institute of Bioscience and Human-Technology,
Higashi 1-1, Tsukuba, Ibaraki 305, Japan

Received August 22, 1995

**SUMMARY**: Previous studies have shown that an Actinomycetes strain produces a novel protein with a molecular weight of 87,000 which stimulated the growth of murine erythroid progenitors *in vitro*, and we purified the protein. In this study, we have isolated and sequenced a clone encoding a polypeptide which had the equivalent activity. The clone contained an open reading frame of 334 amino acids. The activity was confirmed by the expression in *E. coli* as a fusion protein with glutathion S-transferase (GST). Based on the differences of molecular weight and N-terminal amino acid sequences, this protein was considered to be different from that previously reported and was named "ESA36". 

Erythropoietin (EPO) is a glycoprotein, which sustains survival, proliferation and differentiation of immature erythroid cells, thus regarded to be a dominant regulator of erythropoiesis. In our course of search for a novel molecule which modulates the activity of EPO, we found (1) that a thermophilic Actinomycetes strain F-423 (*Streptomyces thermoviolaceus* IFO 13905) produced a protein which stimulated the growth of murine erythroid progenitors *in vitro*. The protein was purified to homogeneity and was named Erythropoiesis-Stimulating Activity (ESA). ESA was an acidic glyco-

[1]Nucleotide sequence reported in this article has been submitted to DNA Data Bank of Japan (DDBJ) with accession number D63904.

†To whom correspondence should be addressed. FAX: 0298-54-6095.

**ABBREVIATIONS:** SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis; IPTG, isopropylthio-ß-D-galactoside; APMSF, (4-amidinophenyl)methane-sulfonyl fluoride; BFU-E, burst forming unit-erythroid; CFU-C, colony forming unit in culture.

0006-291X/95 $12.00

protein with a molecular weight of 87,000 on SDS-PAGE. Recently we found that partially purified ESA inhibited the proliferation of endothelial cells from human umbilical cord vein and HT1080 human fibrosarcoma cells and facilitated migration of endothelial cells (unpublished data). Thus ESA was expected to be a multifunctional protein.

To investigate the mechanism of activities of ESA an abundant pure ESA was necessary. Molecular cloning and expression are effective approaches to ensure the supply of pure proteins. In this paper, we report the isolation of a ESA gene which had stimulatory activity on erythroids but was small in molecular size. This protein was considered to be different from that previously reported, thus was named "ESA36".

## MATERIALS AND METHODS

Bacterial strains, plasmids: Plasmid pUC18 was used for cloning in *E. coli* JM109, and pUC118 and pUC119 were used as sequencing vectors to produce single strand DNA in *E. coli* MV1184. pUCE27 is a derivative of pUC18 and contains a 10.1 kb insert that includes ESA gene.

Cloning and expression: For Southern hybridization, three types of oligonucleotide guessmers were designed as probes. N-terminal sequence analysis of purified ESA (1) revealed 10 amino acids, Ser-Thr-Lys-Pro-Thr-Thr-Thr-Phe/Arg-Leu-Glu. As the codon usage in *Streptomyces* is known to be highly biased for G and C, two types of 30-mers, TCIACIAAGCCIACIACIACI-TTCCTIGAG and TCIACIAAGCCIACIACIACIGACCTIGAG, and one 17-mer, AC(G/C)AAGCC(G/C)AC(G/C)AC(G/C)AC, were synthesized. Chromosomal DNA of *S. thermoviolaceus* IFO 13905 was digested to completion with *BamH* I, DNA fragments of around 9.2 kb were size-fractionated, ligated to pUC18, then this genomic library was used to transform *E. coli* JM109 and Southern hybridization was performed. For the expression of ESA, *E. coli* JM109 harboring recombinant plasmids were grown on an LB medium.

DNA sequencing: DNA fragments were subcloned into pUC118 and pUC119, nested deletion sets were made, single-stranded phagemid DNA were prepared and nucleotide sequencing was performed by the dideoxy method with -21M13 dye primer and *Taq* DNA polymerase.

Production of ESA in *E. coli* as a fusion protein with GST: In order to avoid two stop codons upstream from the putative ATG start codon, pUCE27 was digested with *Nsp* I which recognized six nucleotides containing ATG. The 3.5 kb *Nsp* I-*Kpn* I fragment carrying ESA gene was subcloned into pUC18 that had been cut with *Sph* I and *Kpn* I, thus restored the ATG in *Sph* I/*Nsp* I site. This subclone was cut with *Hin*d III, blunted, and *Eco*R I linker was ligated. Finally the *Eco*R I-*Sal* I fragment of this reconstruction was ligated to pGEX-4T-1, 2 or 3. These recombinant plasmid constructs were called pGESA-1, 2 and 3. *E. coli* JM109 transformed with pGESA-1, 2 or 3 were

grown in 2xYT-G medium until the $A_{600}$ reached about 1.0, then IPTG (0.1 mM) was added and incubation was continued for an additional 1 hr. Cells were collected by centrifugation, suspended in phosphate buffered saline (PBS) containing 0.02% of Tween 20 and protease inhibitors ( 20 μg/ml APMSF, 0.5 μg/ml leupeptin, 0.01 mM EDTA, 1 μg/ml pepstatin), and disrupted with a sonicator. The solubilized fusion proteins were recovered by centrifugation and applied to GSH Sepharose 4B columns. Bound fusion proteins were eluted from the columns with 10 mM GSH. The proteins were cleaved with thrombin in the case of the biological assay for ESA; fusion protein solutions were digested with thrombin, and passed through another GSH Sepharose column to adsorb GST. In order to adsorb thrombin, the solution was passed through an Antithrombin III-Agarose column which had been equilibrated with heparin.

Biological assay for ESA: The method used to determine erythropoiesis-stimulating activity (ESA) was a modification of the colony forming unit-erythroid (CFU-E) assay (1), but activity units were not determined.

Identification of C-terminal amino acids of expressed ESA: Purified GST-ESA fusion protein (2.8 mg) was denatured in 1% SDS at 95 °C for 5 min, then digested with 25 μg of carboxypeptidase A and 20 μg of carboxypeptidase B in 20 mM Hepes (pH 8.2) at 37 °C. Aliquots were drawn out at intervals and the enzyme reaction was stopped by acidifying below pH 2.0. Released amino acids were determined using an amino acid analyzer.

## RESULTS AND DISCUSSION

For cloning of ESA, we used a hybridization approach first. Three types of guessmers were used to hybridize to the *Bam*H I digests of genomic DNA. As Southern analysis revealed a single signal of approximately 9.2 kb, DNA fragments around this size were recovered and a genomic library was made. 6600 plasmids were isolated from this library and 427 clones were selected by dot hybridization. 24 of these selected clones were mapped and partially sequenced. All clones examined contained sequences which are highly similar to the guessmers, but none of them encoded the expected N-terminal 10 amino acids.

Then we used an expressional approach. It is well known that the expression of a *Streptomyces* gene in *E. coli* is difficult, because very few DNA sequences serving as promoters in *Streptomyces* are recognized by the RNA polymerase of *E. coli* (2). But a few examples of expressions have been reported, though their expression levels were low when compared with expression in *Streptomyces* (3, 4). *E. coli* JM109 harboring "selected clones" were divided into pools, each with 20 different recombinants, and cultured for ESA activity assay on their cell extracts. Positive pools were partitioned

into smaller pools and cultured. Thus we obtained *E. coli* JM109 (pUCE27) from 427 recombinant clones.

The recombinant plasmid pUCE27 contained an insert of 10.1 kb (Fig. 1). Fig. 1 shows that the transformant of pUCE278, where the plasmid pUCE27 insert was cloned to pUC118 in the same orientation to the *lacZ* promoter, had equal activity to the original, while the transformant of pUCE279, which was in the opposite orientation, had no activity. Furthermore, two deletion mutants of pUCE27, pUCE20 and pUCE21, also had equivalent activity. These results imply that the ESA gene was encoded in a 3.9 kb fragment from *Bam*H I to *Spl* I (shown by grayish bar in Fig. 1) and this gene was not transcribed from its own promoter, but from *lacZ* promoter in the pUC18.

The *Bam*H I-*Kpn*I fragments (3.7 kb) were subcloned into pUC118 and pUC119, and the nucleotide sequences were determined on both strands. The 1500 base nucleotide sequence from the 5' terminus of pUCE27 and the predicted amino acid sequence of ESA are shown in Fig. 2. The open reading frame (ORF) was assumed to start with the initiator codon (ATG) at 156-158 bases from the 5' end (*Bam*H I site), because a putative ribosome binding site was situated -10 nucleotide in front of the ATG, where the sequence AGGAAG exhibits strong complementarity to the 3' end of *Streptomyces lividans* 16S rRNA (5). Putative -35 and -10 regions of the promoter

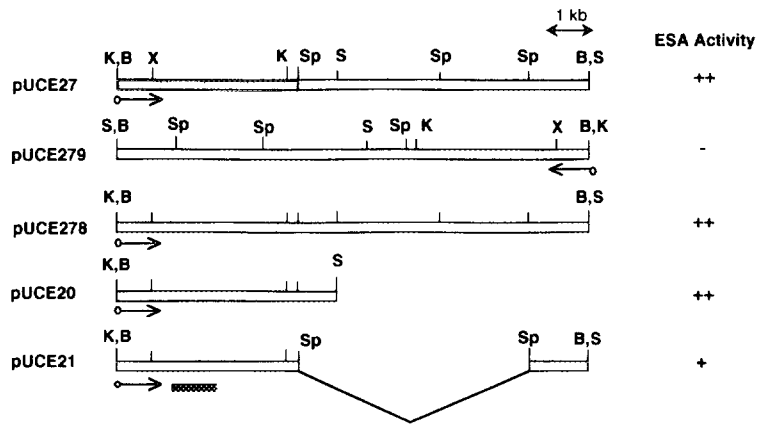


Figure 1. Restriction enzyme map of pUCE27 and its derivatives. The arrows represent the orientation of the ESA gene in relation to the *lacZ* promoter *Ptacz* on pUC18, pUC19 and pUC118. ▬ indicates a fragment which hybridized with the guessmers. Relative ESA activities, determined by the CFU-E colony formation assay, are also shown. B, *Bam*H I; S, *Sse* 8387 I; X, *Xho* I; Sp, *Spl* I; K, *Kpn* I.

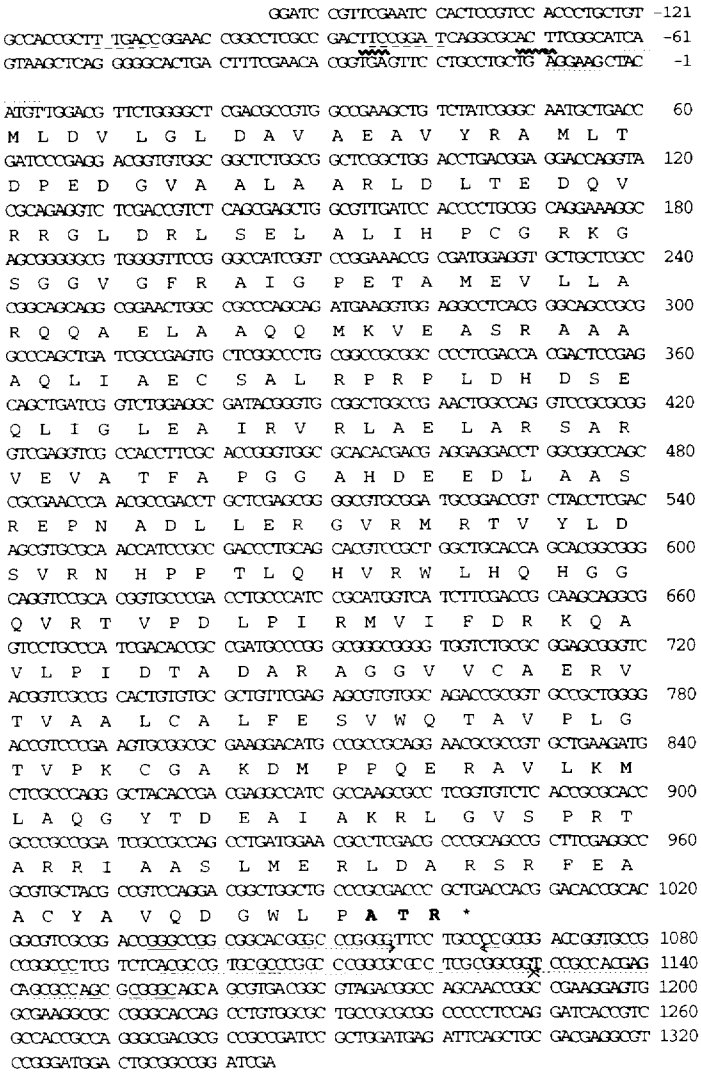


Figure 2. Nucleotide sequence of the ESA36 gene. An ORF of 1002 bp (334 aa) is shown with the deduced amino acid sequence. The putative ribosome binding site is underlined, and dotted underlines are the putative -35 and -10 regions. Wavy uplines are two stop codons which exist up-stream from the start codon and are in frame, and the dotted upline is the *Nsp* I site used to construct the fusion protein. An asterisk indicates the stop codon, and amino acids shown in bold letters are confirmed ones in C-terminal determination. Two palindromes in the 3' noncoding region are shown by broken lines with arrows.

were assumed to be the nucleotide sequences TTGACC (-111 to -106) and TCCGGAT (-86 to -80), which were in good agreement with those reported (6, 7). The ORF (334 amino acids) had high G+C content (72% overall) especially in the third position (90%), which is the typical characteristics of

*Streptomyces* genes. The G+C content of the putative promoter region (-111 to -11) was 61%. Two palindromes were found downstream from the TGA stop codon, which could serve to terminate transcription.

The deduced ESA contained 334 amino acids and had a calculated Mr of 36,231, while the Mr of ESA, which was conventionally purified from cell extract of *S. thermoviolacus*, was 87,000. Downstream from the stop codon and the two palindromes, a nucleotide sequences was found which was highly homologous to the guessmers used as a probe, and 16 nucleotides of 17-mer were identical. Hybridization analysis revealed that the restriction fragment of 0.8 kb in which this homologous sequence exists (shown in Fig. 1) was the sole hybridizing fragment in pUCE27. These results suggest that pUCE27 was selected artificially by hybridization but the clone included a nucleotide sequence encoding a polypeptide which had ESA activity.

To confirm the ESA activity, we constructed three expression vectors in which the ESA gene was ligated to the GST gene in three translational reading frames (Fig. 3). These vectors can be expressed as fusion proteins of GST and ESA, and these two polypeptide moieties can be separated after

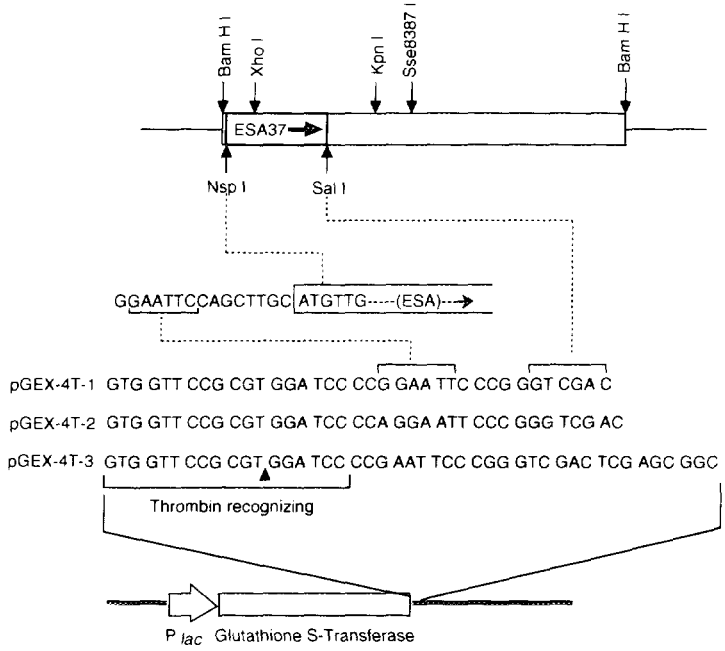


Figure 3. Constructs of pGESA-1, 2 and 3, and the nucleotide sequence which extends from *Eco*R I site of pGEX vectors to *Nsp* I site of the ESA gene. Details are described in Materials and Methods.

cleavage with thrombin that recognizes the connecting region of the fusion proteins.

Among three recombinant plasmids of different reading frames, only pGESA-2, which was presumed to be in frame, expressed a 66 kDa fusion protein and showed ESA activity when assayed after separation from GST and thrombin. Activity was shown in the concentration from 4 pM to 4 nM (Fig. 4), which was in good agreement with that of ESA reported previously (1). However, the molecular weight of the expressed ESA was 36,500 on SDS-PAGE after cleavage with thrombin, and C-terminal amino acids were identified to be -Ala-Thr-Arg-OH, thus it was considered to be another protein different from the previous one. We named this new protein "ESA36", while the other one was re-named "ESA87" according to their Mr. When we purified ESA in the previous report (1), the activities for stimulating the growth of murine erythroid progenitors were found in some fractions which eluted from ion-exchange or hydrophobic columns in different conditions. It implied that the microbe used produced several types of proteins having the correspondent activities. Present results are consistent with the previous observations.

ESA36 showed no activity when used as the fusion protein with GST. This seemed to be caused by GST moiety, because GST itself was inhibitory on the CFU-E colony formation in the presence of EPO. On the other hand, thrombin stimulated the CFU-E colony formation in the case of the presence of low concentration of EPO (Fig. 5A). Thrombin showed no effect on the

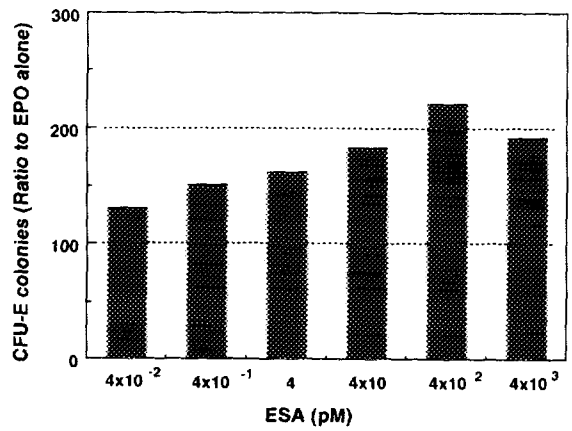


Figure 4. Dose response for ESA36. Recombinant plasmid pGESA-2 was expressed in *E. coli*, and ESA36 was recovered after cleavage of the fusion protein. Colonies derived from CFU-E were counted after cultivation of murine bone marrow cells for 2 days in the presence of EPO (0.025 units/ml) and various concentrations of ESA36.

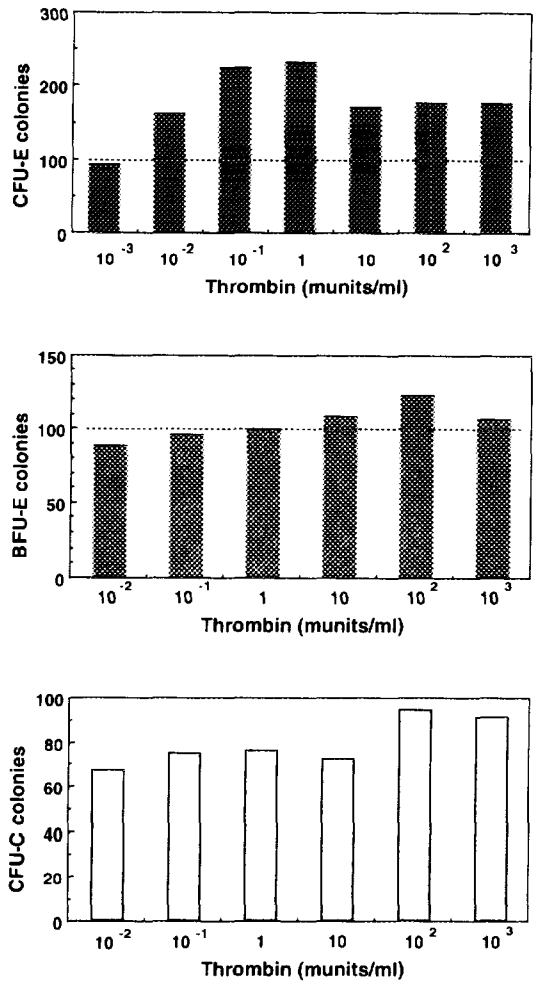


Figure 5. Dose responses for thrombin. CFU-E colony assay was done in the presence of rhEPO (0.025 units/ml), while BFU-E was done in the presence of rmIL-3 (20 ng/ml) with EPO (0.25 units/ml) and counted on day 7. CFU-C colony assay was done in the presence of rhM-CSF (2 units /ml) and counted on day 7. Ratios to the number of colonies, which were formed in the indicated conditions without thrombin, are shown.

BFU-E colony formation, while the CFU-C colony was inhibited by lower concentration of thrombin (Fig. 5B, 5C). Thrombin is known to affect various cells such as platelets, fibroblasts, endotherial cells or leukemic cells (8) through its unique receptors on the cell surface; i.e. thrombin cleaves the specific site in N-terminal peptide chain of its receptor, and the emerging N-terminus acts as an anchored ligand (9). Thus "thrombin receptor peptides" (TRP) (5, 6, 7 or 14 amino acids from the new N-terminus) are known to show correspondent activities on these cells *in*

*vitro* (8). When we examined the synthetic hexapeptide Ser-Phe-Leu-Leu-Arg-Asn, which is one of the TRPs, in the colony formation experiments, the peptide induced the same effects as thrombin. The hexapeptide was effective in the concentration range from 2 pM to 2 nM.

## ACKNOWLEDGMENTS

We are grateful to Drs. Hiroyuki Hosoya and Kazuya Nakagomi for helpful discussions and experimental advice.

## REFERENCES

1. Sugie, M., Oka, S. and Tanaka, H. (1989) *Biochem. Biophys. Res. Commun. 163*, 888-894.
2. Buttner, M.J. and Brown, N.L. (1985) *J. Mol. Biol. 185*, 177-188.
3. Bahri, S.M. and Ward, J.M. (1990) *J. Gen. Microbiol. 136*, 811-818.
4. Ogawara, H., Kasama, H., Nashimoto, K., Ohtsubo, M., Higashi, K. and Urabe, H. (1993) *Gene 125*, 91-96.
5. Bibb, M.J. and Cohen, S.N. (1982) *Mol. Gen. Genet. 187*, 265-277.
6. Delic, I., Robbins, P. and Westpheling, J. (1992) *Proc. Natl. Acad. Sci. 89*, 1885-1889.
7. Arisawa, A., Kawamura, N., Tsunekawa, H., Okamura, K., Tone, H. and Okamoto, R. (1993) *Biosci. Biotech. Biochem., 57*, 2020-2025.
8. Mari, B., Imbert, V., Belhacene, N., Far, D.F., Peyron, J.F., Pouysségur, J., Obberghen-Schilling, E.V., Rossi, B. and Auberger, P. (1994) *J. Biol. Chem. 269*, 8517-8523.
9. Vu, T.H., Hung, D.T., Wheaton, V.I. and Coughlin, S.R. (1991) *Cell 64*, 1057-1068.