Crystallization of GreA, a Transcript Cleavage Factor from *Escherichia coli*

Seth A. Darst\(^1\), Charles E. Stebbins\(^1\), Sergei Borukhov\(^2\), Marianna Orlova\(^2\), Guoha Feng\(^3\), Robert Landick\(^2\) and Alex Goldfarb\(^2\)

\(^1\)Rockefeller University
1230 York Avenue, New York, NY 10021, U.S.A.

\(^2\)Public Health Research Institute
New York, NY 10016, U.S.A.

\(^3\)Department of Biology
Washington University, St. Louis, MO 63130, U.S.A.

GreA is a 17.6-kDa protein from *Escherichia coli* that induces cleavage of the nascent transcript in the preinitiation complex of RNA polymerase, followed by release of the 3’-terminal fragment. Crystals of GreA have been obtained from polyethylene glycol 4000, 2-propanol and sodium citrate, pH 5-6 and have been propagated by a novel seeding procedure. The crystals diffract beyond 2 Å resolution and belong to the orthorhombic space group P2_12_1, with cell dimensions a = 101.7 Å, b = 42.2 Å, c = 40.0 Å and with one molecule in the asymmetric unit.

Keywords: protein crystallization; X-ray diffraction; transcription elongation factor; transcript cleavage factor

Analysis of *in vitro* transcription reactions reveals that the DNA-dependent RNA polymerases do not elongate RNA at a uniform rate. The polymerase pauses at certain sites on the DNA template before resuming elongation, and a fraction of the polymerase molecules can become trapped at these sites, resulting in arrested ternary complexes that can neither propagate nor dissociate (Levin & Chamberlin, 1987). Transcription elongation factors, such as the eukaryotic SII, increase the overall rate of transcript elongation by greatly mitigating pausing and rescuing ternary complexes from the arrested state (Reines et al., 1989; Sluder et al., 1990; SivaRaman et al., 1990; Borukhov et al., 1993).

In studies of ternary complexes using purified *Escherichia coli* RNA polymerase stalled at defined template positions by nucleotide deprivation, Surratt et al. (1991) observed the hydrolytic cleavage of 2 to 10 nucleotide fragments from the 3’ end of the growing transcript. The 5’-terminal fragment of the transcript, bearing a free 3’-OH, remains bound in the complex and can be extended by the polymerase upon the addition of nucleotides, indicating that its 3’ end is positioned in the active site of the polymerase. The 3’-terminal fragment of the transcript is released. Similar reactions, dependent on SII, were observed in eukaryotic systems (Ihab & Luse, 1992; Reines, 1992). Subsequent studies identified two homologous factors, GreA and GreB, normally present as trace contaminants in preparations of *E. coli* RNA polymerase, that induced the transcript cleavage reaction (Borukhov et al., 1992, 1993). Reverse translocation of the RNA polymerase along the DNA template accompanies the cleavage reaction. None of the transcript cleavage factors interact with, or modify, free RNA on their own; the transcript cleavage reaction occurs only in the context of the ternary complex with RNA polymerase. This and other observations (S. Borukhov & A. Goldfarb, unpublished results) suggests that the RNA cleavage activity is an intrinsic property of the RNA polymerase that is induced by these factors. While GreA and GreB have no apparent sequence homology with SII, stimulation of nascent RNA cleavage in ternary transcription complexes appears to be an evolutionarily conserved function of transcription elongation factors. The ubiquity of transcript cleavage, which has also been observed in transcription complexes of vaccinia virus RNA polymerase (Hagler & Shuman, 1983) and eukaryotic RNA polymerase II (Whitehall et al., 1984), suggests that it must play an important role in transcription or its regulation.

The *greA* gene, encoding a 158-residue (17,630 Da) protein, was initially isolated by virtue of its ability to suppress, at high copy number, a temperature-sensitive mutation in the RNA polymerase β subunit (Sparkowski & Das, 1990, 1991), suggesting that GreA is involved in some vital processes.
aspect of transcription. GreA crosslinks efficiently to the 3’ end of the nascent transcript, indicating an intimate interaction with the region around the catalytic site itself (S. Borukhov et al., unpublished results). The biological role of GreA and transcript cleavage factors in general remain obscure. Two possibilities include: (1) increasing transcriptional fidelity by removal of misincorporated nucleotides (Erie et al., 1993; Kassavetis & Geiduschek, 1993; Wang & Hawley, 1993); (2) rescuing arrested transcription complexes (Reines, 1992; Borukhov et al., 1993).

We are interested in the structure of GreA to aid in understanding its role in the transcription process by relating the structure to the biochemical and biophysical data that are accumulating (S. Borukhov & A. Goldfarb, unpublished results). Other than GreB, only one GreA homolog has been identified, a 158 residue protein from *Pseudomonas pseudozeaxixii* with 50% amino acid identity to GreA (Marks & Wood, 1992). The function or biochemical activity of this protein has not been identified. Thus, the small family of sequence-related, prokaryotic transcript cleavage factors do not appear to be related to any proteins with known structures.

To prepare GreA for crystallization, *E. coli* XLBlue cells, transformed with pDLN278 plasmid carrying the *greA* gene (Feng et al., 1994), were grown in 10 L of LB medium with ampicillin (100 μg/ml) to an absorbance of *A* 660 = 0.2, and the expression of GreA was induced by the addition of IPTG† to 1 mM. The growth was continued for two hours and the cells were harvested by centrifugation. For purification of GreA, all procedures were carried out at 4°C. Frozen cell paste (20 g) containing overexpressed GreA was suspended in 155 ml of lysis buffer (40 mM Tris-HCl (pH 8-0), 0-2 M NaCl, 5% (v/v) glycerol, 10 mM EDTA, 1 mM DTT, 0-1 M PMSF), blended at low speed for 5 to 10 minutes until the mixture was homogeneous and sonicated for five one-minute timings at maximum power settings using a preparative 1 inch tip of a Sonics & Materials ultrasonicator. The mixture was centrifuged at 13,000 rpm for 30 minutes in a Sorvall RC-4B centrifuge. NaCl (4 M) was added to the supernatant to a final concentration of 0-8 M, followed by addition, with vigorous stirring but without foaming, of 10% Polymem P; pH 7-9, to a final concentration of 0-25% (v/v). After five minutes of stirring, the mixture was centrifuged as above and 4 M ammonium sulfate was added to the supernatant to a final concentration of 1-5 M. The mixture was incubated for 15 minutes and centrifuged as above. The resulting supernatant (about 200 ml) was loaded at 80 ml/hour onto a 100 ml phenyl *Toyopera* TSK 650 M column (2-6 × 20 cm) pre-equilibrated with buffer A (40 mM Tris-HCl (pH 8-0), 1-5 M ammonium sulfate, 10 mM EDTA, 1 mM DTT). The column was washed with 500 ml of the same buffer and eluted with buffer B (40 mM Tris-HCl (pH 8-0), 5% glycerol, 10 mM EDTA, 1 mM DTT) using an FPLC system and a 600 ml linear gradient (100% buffer A–100% buffer B) at 120 ml/hour. Fractions of 20 ml were collected and a sample of each fraction was analyzed by SDS/14% PAGE. The fractions containing GreA (eluting at about 50 to 70% buffer B) were pooled, diluted twofold with 2 M ammonium sulfate, and the mixture was loaded at 80 ml/hour onto a 30 ml butyl *Toyopera* TSK 650 M column (1-6 × 15 cm), pre-equilibrated with buffer A. The column was washed with 100 ml of the same buffer and eluted as above, with 200 ml of a linear gradient at 80 ml/hour. The fractions enriched in GreA (eluting at 30 to 50% buffer B) were pooled and concentrated by ultrafiltration in a Centriprep 10 device (Amicon) to a final volume of 3 ml. The material (650 μl samples) was fractionated by gel filtration on an FPLC Superdex 75 10/30 HR column in buffer B containing 0-3 M NaCl. The fractions containing about 90% pure GreA (as judged by SDS/14% PAGE) were pooled, concentrated as above, diluted fivefold with buffer B and applied at 2 ml/minute on a Waters Protein-Pak Q 8 HR column (1 cm × 10 cm) pre-equilibrated with buffer B. After the column was washed with 20 ml of the same buffer, the material was eluted with 80 ml of a 0 M to 0-8 M NaCl gradient in buffer B. Electrophoretically homogeneous GreA eluted from the column as the major peak at about 0-15 M NaCl. The fractions containing pure GreA (about 55 mg) were pooled and concentrated to 30 to 100 mg/ml using Centricron-10 devices (Amicon) and stored at 4°C in 10 mM Tris-HCl (pH 7-9), 1 mM EDTA, 200 mM NaCl, 10 mM DTT and 5% glycerol. For crystallization, the protein was diluted to 15 to 50 mg/m1 with 10 mM DTT.

Three crystal forms of GreA were grown using the method of hanging-drop vapour diffusion. Rod-shaped crystals, up to 0.8 mm × 0.2 mm × 0.2 mm in size, were grown by mixing equal volumes of protein with 18 to 20% (v/v) polyethylene glycol 8000, 0-2 M CaCl2, 10 mM DTT and 0-1 M Mes (pH 6-5) and equilibrating at room temperature against a reservoir containing the same solution. Precession photographs showed that the crystals were tetragonal, with a unit cell of *a* = 162 Å, *c* = 194 Å. Diffraction from the crystal, measured on an RAXIS IIC image plate area detector connected to a Rigaku RU200 copper target rotating-anode X-ray source, extended to about 3.5 Å resolution.

Another form of crystals grew into thick rods about 0.3 mm × 0.2 mm × 0.2 mm from 24 to 26% polyethylene glycol 4000, 0-2 M CaCl2, 10 mM DTT and 0-1 M Hepes (pH 7-5). Diffraction from the crystals, measured on the area detector, extended to about 2.8 Å resolution. These data, and additional measurements by precession photography, showed that these crystals were also tetragonal but with a unit cell of *a* = 162 Å, *c* = 173 Å.

Finally, the highest-quality diffraction was obtained with crystals grown from solutions of

† Abbreviations used: IPTG, isopropyl-β-D-thiogalactopyranoside; PMSF, phenylmethylsulfoxyl fluoride; FPLC, fast protein liquid chromatography.
polyethylene glycol 4000, 2-propanol and sodium citrate, (pH 5-6). Initially, a few silver-like crystals, 0.8 mm × 0.1 mm × 0.05 mm, grew over a period of several weeks at 4°C from a precipitate formed in 20% polyethylene glycol 4000, 20% (v/v) 2-propanol, 10 mM DTT, 0.1 M sodium citrate, (pH 5-6). Although rather thin, these crystals diffracted beyond 2.5 Å resolution. Despite repeated efforts, we have been unable to reproduce the nucleation of this crystal form and have only been able to produce crystals by seeding from previously grown crystals. Various methods of macroseeding (Thaller et al., 1985) and microseeding (Stura & Wilson, 1990) resulted in showers of microrystals, clusters of long, thin plate-like crystals or twinned crystals, none of which was suitable for data collection. The best crystals were grown using a crystallization solution of 10 mM DTT, 0.05 M sodium citrate (pH 5-6) and varying amounts of polyethylene glycol 4000 and 2-propanol, hereafter referred to simply as X/Y, with X being the weight percent of polyethylene glycol 4000, and Y being the volume percent of 2-propanol. The following procedure was used. One or a few small crystals, grown by one of the above-mentioned seeding techniques, were added to a 10 µl drop of protein (15 mg/ml) in 10/10. The drop was incubated at 4°C over a reservoir of 7.5/7.5 and monitored several times a day by light microscopy. Over a period of several days, the crystals slowly dissolved. Within 24 hours of the crystals first becoming invisible at the highest power of the microscope (about 70 ×), the drop was transferred over a reservoir of 10/10 and incubated for a period of at least 48 hours (or up to about one week). The drops were then transferred successively over reservoirs of 12.5/12.5, 15/15 and 17.5/17.5, with similar incubation periods between each step. A few small crystals would begin to appear after the incubation over 10/10, or sometimes during the 12.5/12.5 incubation. After the final equilibration over 17.5/17.5, the drops contained 5 to 10 well-formed crystals, each about 0.3 mm × 0.2 mm × 0.15 mm. Larger crystals could sometimes be grown, but these were usually cracked or twinned. Despite their small size, the crystals diffract beyond 2 Å resolution.

Native data sets (Table 1) were measured on the area detector from single crystals held at 4°C or −180°C using a Molecular Structure Corporation nitrogen gas delivery system. For the low-temperature data collection, the crystals were cryoprotected by slowly transferring to a solution of 20/35 before freezing in a loop of ophthalmological suture material (10-0 Ethilon, Ethicon). At 4°C, diffraction from the crystals degrades noticeably during the course of the data collection, contributing to the higher Ram and lower resolution. The crystals belong to the orthorhombic space group P212121. A heavy-atom derivative has been prepared by soaking crystals in CH3HgCl. Inspection of Patterson maps has located the heavy-atom binding position and also has confirmed the spacegroup assignment. A search for other derivatives is underway.

Using a protein density of 1.3 g/cm³ and the unit-cell volume of 172,000 Å³, it is clear that there is one molecule per asymmetric unit. The solvent content then is 49%. The calculated cell volume per unit mass, Vd, is 2.5 Å³, which is within the range found for other protein crystals (Matthews, 1968).

We are extremely grateful to John Kurtiany and Stephen Burley and the members of their laboratories for their help and encouragement in all aspects of this work, and in particular A. F. D’Amore for help with precision photography. T. S. R. Krishna for help with cryocrystallography methods, and Kirk Clark and Joseph Kim for help with data analysis and general advice. S.A.D. is a Lucille P. Markey Scholar. This work was supported in part by a grant from the Lucille P. Markey Charitable Trust and from the Irma H. Hirschl Trust.

References


Edited by A. Klug

(Received 24 June 1994; accepted 28 June 1994)