Prokaryotic Translation Initiation Factor IF3 Is an Elongated Protein Consisting of Two Crystallizable Domains†

Jadwiga H. Kycia, Valerie Biou, Fong Shu, Sue Ellen Gerchman, Vito Graziano, and V. Ramakrishnan*

Biology Department, Brookhaven National Laboratory, Upton, New York 11973

Received January 9, 1995[⊗]

ABSTRACT: We show that translation initiation factor IF3 can be split into two fragments of nearly equal size by the *Escherichia coli* outer membrane protease omptin. Circular dichroism and small-angle neutron scattering show that the two fragments are structured as domains. Each domain is relatively compact, and they are separated by about 45 Å in intact IF3. Thus IF3 is an elongated protein that consists of two well-separated domains. We suggest that these two domains are involved in ribosome binding across the cleft of the 30S ribosome. We also report the crystallization of each domain of IF3.

Initiation of protein synthesis in prokaryotes involves the recognition of a specific site on messenger RNA by the ribosome and the binding of the initiator fMet-tRNA to the P site of the ribosome. In addition to mRNA, tRNA, and the ribosome, initiation requires the involvement of three initiation factors, IF1, IF2, and IF3, and is accompanied by the hydrolysis of GTP [reviewed by Gualerzi and Pon (1990)].

The initiation factor IF3 binds to the 30S ribosomal subunit with an association constant in excess of 10^7 M^{-1} (Weiel & Hershey, 1982). Initially, IF3 was shown to prevent the association of ribosomal subunits by binding to free 30S subunits (Godefroy-Colburn *et al.*, 1975). Previous reports that IF3 promoted the binding of mRNA to the 30S ribosome have been contradicted in a recent study using defined oligonucleotides containing the initiation site (Laughrea & Tam, 1989). Recent "toeprinting" experiments show that IF3 is involved in the selection and inspection of the initiator tRNA and recognizes the anticodon stem−loop of tRNA (Hartz *et al.*, 1990; Hartz *et al.*, 1989). Thus, IF3 may have a dual function: binding to free 30S ribosomes, thus preventing 70S ribosome formation in the absence of message, and proofreading of the initiation complex.

The approximate site of interaction of IF3 in the 30S ribosome has been determined in several studies. Cross-linking experiments show that IF3 cross-links primarily to nucleotides 819−859 in the central domain of 16S rRNA, with a minor cross-link to the region 1506−1529 in the 3′ end (Ehresmann *et al.*, 1986). It has been UV cross-linked to ribosomal proteins S7, S11, S12, S18, and S21 (MacKeen *et al.*, 1980). Chemical protection experiments are also consistent with IF3 interacting with the central domain (Muralikrishna & Wickstrom, 1989). This information is consistent with IF3 interacting with a region of the 30S ribosome that includes part of the platform, the cleft, and the head, facing the 50S subunit. These results also agree with the localization observed by immune electron microscopy of 30S subunits cross-linked to IF3 (Stöffler *et al.*, 1980).

In this study, we show that IF3 can be cleaved into two fragments by omptin, the *Escherichia coli* outer membrane protease coded by the *ompT* gene. We describe circular dichroism and small-angle neutron scattering experiments on both intact IF3 and individual fragments. Our studies show that the individual fragments are structured as domains, and that IF3 is an elongated, two-domained protein. We also report the results of our efforts to crystallize the individual domains.

MATERIALS AND METHODS

Cloning and Expression of the Bacillus stearothermophilus IF3 Gene. The gene for *Bacillus stearothermophilus* IF3 was amplified by PCR using the known protein sequence (Kimura *et al.*, 1983) and following previously described methods applied to various ribosomal genes (Ramakrishnan & Gerchman, 1991). The amplified gene was cloned into the T7 expression vector pet-13a (Gerchman *et al.*, 1994), which is a kanamycin-resistant derivative of pet-11a (Studier *et al.*, 1990). The gene was sequenced with Sequenase (U.S. Biochemicals) using the protocols provided by the manufacturer. The gene sequence predicted an amino acid sequence that is identical to that determined directly (Kimura *et al.*, 1983) and was also the same as the gene sequence subsequently reported (Pon *et al.*, 1989). When the IF3 gene in pet-13a was expressed in the *E. coli* strain BL21(DE3), very high levels of expression were obtained.

Purification of IF3. IF3 was purified using a procedure similar to that for other ribosomal proteins overexpressed in *E. coli* (Ramakrishnan & Gerchman, 1991). All buffers contained 0.05 mM phenylmethanesulfonyl fluoride and 0.05 mM benzamidine as protease inhibitors, and all procedures were done at 0−4 °C unless otherwise noted. Cells were grown to an OD_{600} of 0.8 and induced with 0.5 mM IPTG. The cells were harvested and stored at −70 °C. For extraction, the cells were thawed, resuspended in 5 vol of lysis buffer (50 mM Tris-HCl, pH 8.0, and 1 mM EDTA), and treated with lysozyme at a final concentration of 1 mg/mL for 1 h. Cells were lysed by the addition of sodium deoxycholate to 0.08%. The resulting viscosity was reduced by treatment with DNAse I (2 μg/mL) for 15 min in the presence of 0.01 M $MnCl_2$ and 0.01 M $MgCl_2$.

The salt concentration was raised to 0.8 M NaCl, and the extract was centrifuged at 12 000 rpm in a Sorvall SS34 rotor

† Supported by the Office of Health and Environmental Research, U.S. Department of Energy, and Grant GM 44973 (to Stephen W. White, Duke University, and V.R.) from the NIH.

⊗ Abstract published in *Advance ACS Abstracts,* April 15, 1995.

0006-2960/95/0434-6183$09.00/0

for 20 min. The supernatant was diluted with lysis buffer to bring the salt concentration down to 0.2 M NaCl and loaded onto an S-Sepharose Fast-Flow (Pharmacia) column equilibrated in lysis buffer containing 0.2 M NaCl. A gradient of 0.25−0.8 M NaCl in lysis buffer was applied. The total gradient was 10 column volumes. Fractions containing IF3 were pooled, concentrated, and applied to a Sephacryl S-100 HR (Pharmacia) gel filtration column equilibrated in lysis buffer containing 0.5 M NaCl. The fractions containing IF3 from this column were essentially pure IF3 as judged by Coomassie gel staining of 8−25% polyacrylamide-SDS gels.

Digestion of IF3 with Omptin. Pure IF3 at a concentration of 10 mg/mL in 50 mM phosphate and 0.2 M NaCl, pH 7.2, was treated with omptin (gift of Dr. Walter F. Mangel) at a final concentration of 40 μg/mL. The mixture was incubated at 0 °C for 30 min and then moved to 37 °C for 15 min. Aliquots of the reaction mixture were stopped at various time points by adding of an equal volume of 2× SDS sample buffer (40% glycerol, 0.25 M Tris-HCl, pH 6.8, 12 mM 2-mercaptoethanol, and 0.1% bromophenol blue) and boiling for 2 min.

The products of digestion of IF3 by omptin were run on an 8−25% gradient SDS−polyacrylamide gel on a Pharmacia Phast system. Electroblotting on the Phast System was used to transfer the proteins to a nitrocellulose membrane. The transfer buffer used was 25 mM Tris, pH 8.3, 192 mM glycine, and 20% methanol. After the transfer, the nitrocellulose membrane was stained with 0.25% Coomassie Blue in 20% methanol. The two bands corresponding to the cleavage products were cut out of the membrane and submitted for N-terminal sequence determination.

Once the omptin cleavage site in IF3 had been determined, oligonucleotide primers were synthesized to amplify by PCR the regions of the IF3 gene coding for the two fragments. These regions were separately amplified and cloned into pet-13a and were expressed and purified using methods similar to those for the intact IF3 gene.

Circular Dichroism Studies. Circular dichroism spectra of intact IF3 and its fragments were measured to determine the secondary structure content.

The concentration of the protein solutions was about 1 mg/mL in 30 mM phosphate buffer at pH 7.0. Phosphate rather than chloride ions were used to minimize absorption. The concentration was determined from the UV absorption spectrum, and the following extinction coefficients were used: intact IF3, ϵ_{258} = 1840 M^{-1} cm^{-1} ; N-terminal fragment, ϵ_{280} = 1490 M^{-1} cm^{-1}; C-terminal fragment, ϵ_{258} = 770 M^{-1} cm^{-1}.

Circular dichroism spectra between 178 and 300 nm were measured for the various IF3 samples in Hellma quartz cells with a path length of 0.1 mm. The measurements were done on beamline U9B at the NSLS at Brookhaven National Laboratory (Sutherland *et al.*, 1980). Secondary structure content was estimated using the program Varselec (Manavalan & Curtis Johnson, 1987).

Neutron Scattering. Samples containing intact IF3 or each of the two fragments were dialyzed into a buffer containing 10 mM Tris-HCl and 0.5 M NaCl, apparent pH 8.0, in 99.6% D_2O. The samples were then loaded into a quartz cell of 5 mm path length. Neutron scattering from these samples was measured on the small-angle diffractometer on beamline H9B at the High Flux Beam Reactor at Brookhaven National Laboratory (Schneider & Schoenborn, 1984). Scattering was also measured from a cell containing the D_2O buffer, one containing pure H_2O for calibration, and an empty cell, and the background was measured when the beam was blocked. Data from the two-dimensional position-sensitive detector were radially integrated to obtain the scattered intensity as a function of q, $q = 4\pi \sin \theta/\lambda$, where 2θ is the scattering angle and λ is the wavelength. The radially integrated intensity from each of the protein samples was corrected for the contributions from the buffer and the empty cell and normalized for beam intensity, transmission, and protein concentration to obtain the corrected intensity, $I(q)$. Protein concentrations were estimated by amino acid analysis directly on the various neutron-scattering samples. This was important in estimation of the molecular weight, because the extinction coefficient of intact IF3 appears to depend on the buffer it is in. Moreover, the low extinction coefficients of IF3 and its domains mean that even a small amount of impurity in the samples could change the apparent concentration considerably. The intact IF3 and N-terminal fragment were at concentrations of about 2 mg/mL, and the C-terminal fragment was at a concentration of about 7 mg/mL.

The radius of gyration of each protein sample was determined by Guinier plots of the data ($\ln[I(q)]$ vs q^2). The linear regions at low angles were fit to a straight line, and the radius of gyration and the molecular weight were determined using the slope and intercept, respectively (Jacrot & Zaccai, 1981; Porod, 1982).

The length distribution function $P(r)$ measures the distribution of distances between pairs of points in the molecule. The maximum chord d_{max} in the molecule was estimated from the data using the indirect Fourier transform algorithm of Moore (1980); this value was then used in the program GNOM (Svergun *et al.*, 1988) to calculate $P(r)$.

RESULTS

IF3 Cleavage by Omptin. The gene for IF3 from *B. stearothermophilus* codes for a protein with a molecular mass of 19.8 kDa. The result of digesting intact IF3 by omptin is shown in Figure 1. As can be seen from the gel, omptin cuts IF3 into two roughly equal fragments, which are relatively resistant to further attack. This suggests that IF3 consists of two protease-resistant domains connected by a linker easily accessible to protease.

The omptin cleavage site was determined by sequencing the N-terminal region of each of the fragments produced by digestion of IF3. The smaller fragment had the same N-terminal sequence as intact IF3, while the larger fragment had a sequence that began with Gln-80 (assuming the first residue to be methionine). Thus omptin cleaves IF3 between Lys-79 and Gln-80 to produce an N-terminal fragment with a molecular mass of 9.1 kDa and a C-terminal fragment of 10.8 kDa.

Circular Dichroism. The CD spectra of intact IF3 and the N- and C-terminal fragments are shown in Figure 2. The estimated secondary structures from these spectra are shown in Table 1. The percentage of total residues in a particular secondary structure, as well as the actual number of residues N in that structure are shown in each case. The sum row is obtained from the data on the N- and C-terminal fragments by a weighted average for the percentages and a simple sum for the number of residues. The results show that the sum

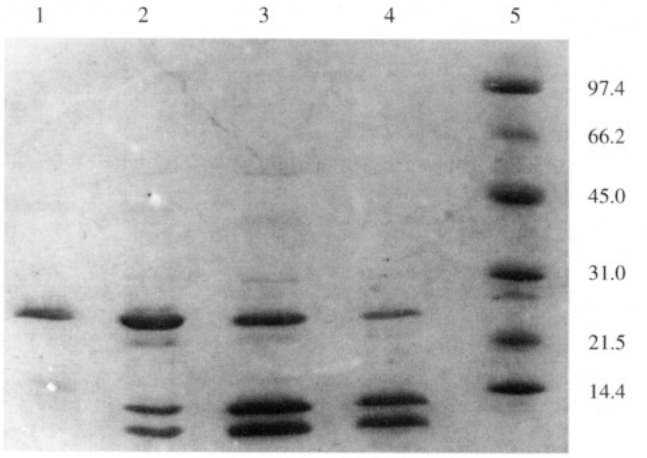

FIGURE 1: SDS−polyacrylamide gel showing the time course of digestion of IF3 from *B. stearothermophilus* by the *E. coli* outer membrane protease omptin. The gel shows that IF3 can be cleaved into two fragments that are relatively resistant to further cleavage. The lanes show various digestion times as follows: lane 1, undigested IF3; lane 2, 30 min at 0 °C; lane 3, 30 min at 0 °C followed by 10 min at 37 °C; lane 4, 30 min at 0 °C followed by 15 min at 37 °C; lane 5, molecular mass standards. The figures on the right are the molecular masses of the standards in kilodaltons.

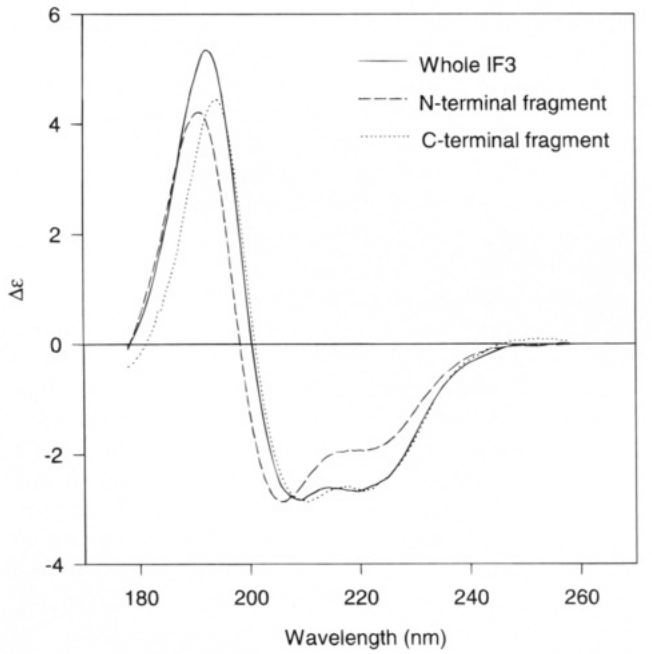

FIGURE 2: Circular dichroism data showing the molar ellipticity as a function of wavelength, on intact IF3 and the N- and C-terminal fragments of IF3.

Table 1: Secondary Structure Content of IF3 and Its Fragments

	α-helix		anti-parallel β		parallel β		turn		other	
	%	N	%	N	%	N	%	N	%	N
N-terminal fragment	25	19	18	14	8	6	19	15	31	24
C-terminal fragment	30	28	18	17	6	6	17	16	28	26
sum	27	47	18	31	7	12	18	31	29	50
intact IF3	28	48	19	32	7	12	17	29	28	48

of the secondary structure content of the two fragments is roughly equal to that of the intact protein.

Neutron Scattering. Figure 3a shows Guinier plots of the small-angle region of the neutron-scattering data. The straight lines represent fits to the linear region at low angles. Molecular weights calculated from the intercepts were close

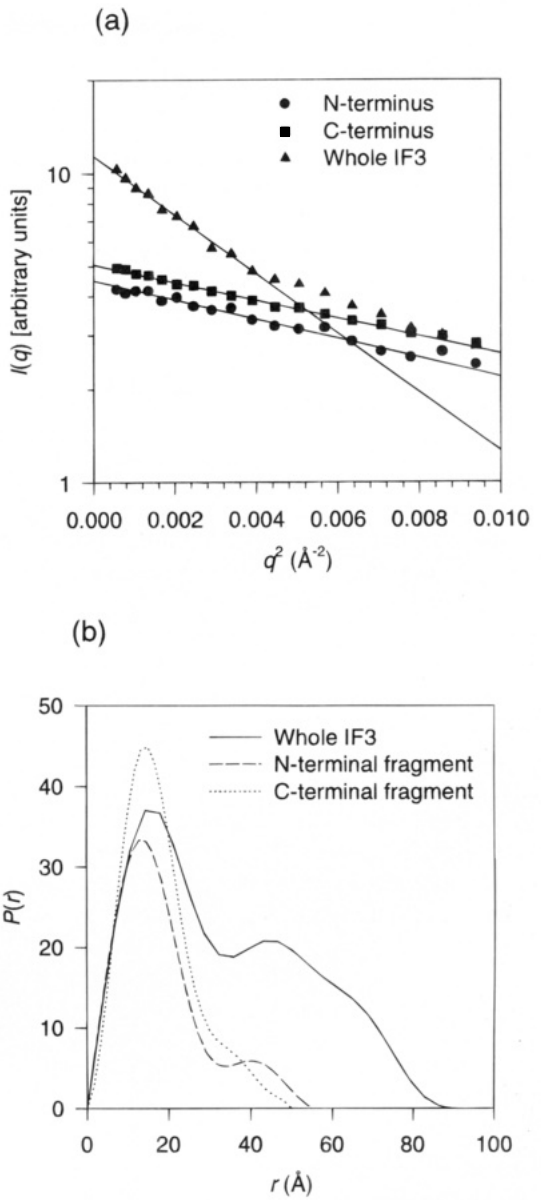

FIGURE 3: (a) Guinier plots of small-angle neutron-scattering data on intact IF3 and the N- and C-terminal fragments. The slopes are related to the radius of gyration, and the intercepts to the molecular weight. (b) Length distribution functions derived from neutron-scattering data on IF3 and its N- and C-terminal fragments. The curves describe the probability distribution of distances between pairs of points inside the molecule.

to the values expected if IF3 and its fragments existed as monomers in solution. As can be seen from the slopes, the curves for the two fragments of IF3 have very similar slopes. However, the curve for the intact protein is characteristic of a two-domained protein, with a steeper inner slope corresponding to the shape of the whole protein, and an outer one that is similar to the slope for the individual fragments. This dichotomy becomes clearer in Figure 3b, which shows the length-distribution function $P(r)$ for each of the fragments and the whole protein. This function measures the distribution distances between all pairs of points inside the particle, weighted by the contrast at those points. The length distribution of each of the fragments is mostly unimodal, with a possible tail at larger distances, especially for the N-terminal fragment. However, the distribution of the intact protein is significantly bimodal, with the first peak corresponding to that of the distribution within each domain, and

the second peak corresponding to distances between domains.

The radius of gyration of each fragment and of the intact protein can be obtained by the Guinier analysis shown in Figure 3a; it can also be obtained from the second moment of $P(r)$. These two methods gave very similar values, with those from the second moment of $P(r)$ being slightly larger. The radii of gyration obtained were N-terminal domain, 15.0 ± 0.5 Å; C-terminal domain, 14.7 ± 0.5 Å; and intact IF3, 27.5 ± 1.0 Å. From these data, and knowing the fractional contribution of each domain to the scattering, the distance between the centers of mass of the domains can be calculated (Jacrot, 1976) using the formula

$$R_g^2 = f_1 R_1^2 + f_2 R_2^2 + f_1 f_2 \Delta^2$$

where R_g, R_1, and R_2 are the radii of gyration, respectively, of the intact protein and the N- and C- terminal fragments; f_1 and f_2 are the scattering fractions of the N- and C-terminal fragments; and Δ is the distance between the centers of mass.

Using values of 0.46 and 0.54 for f_1 and f_2, respectively, we get a value of 46 ± 3 Å for the separation between the centers of mass of the N- and C-terminal domains. This corresponds rather well to the second mode in the $P(r)$ distribution for intact IF3 shown in Figure 3b.

Crystallization Studies. Despite extensive trials, we have not yet been able to get high-quality crystals of intact IF3 from *B. stearothermophilus.* This has been the experience of other workers as well (Gualerzi & Pon, 1990). However, once we determined that IF3 could be cleaved easily into two domains and that each domain could be expressed well, we tried to crystallize each domain separately. For crystallization, we screened various conditions using the hanging-drop vapor diffusion method. We were able to obtain crystals of each domain. Crystals of both domains diffracted to beyond 2-Å resolution and were suitable for structure determination.

Crystals of the N-terminal fragment were obtained in two conditions: Form 1 was obtained in 2.8 M phosphate, pH 7.8. The phosphate solution was made by mixing equimolar solutions of sodium monobasic and potassium dibasic salts in the appropriate ratio to get the desired pH. These crystals were orthorhombic, with cell dimensions $a = 26.7$, $b = 39.8$, and $c = 65.2$ Å. The crystals were radiation sensitive and difficult to stabilize or freeze. Form 2 was obtained in 2.7 M ammonium sulfate and 0.1 M Tris-HCl, pH 7.4. These crystals were also orthorhombic, with $a = 25.3$, $b = 39.2$, and $c = 63.3$ Å. They were less radiation sensitive and could be stabilized easily in ammonium sulfate mother liquor. From the fact that both crystal forms are orthorhombic and have similar cell dimensions, we feel that the packing in the two forms is probably very similar.

Crystals of the C-terminal fragment were obtained in 2.4 M phosphate, pH 8. The crystals are in space group C_2, with cell dimensions of $a = 72.0$, $b = 30.0$, $c = 43.3$ Å, and $\beta = 100.15°$.

DISCUSSION

Although initiation factors were discovered two decades ago, progress on understanding their mode of action has been severely hampered by a lack of structural information. We present here some of the first such information for IF3.

As we have pointed out in the introduction, most biochemical information is consistent with IF3 interacting with a wide region surrounding the cleft of the 30S ribosomal subunit. It is difficult to imagine how a compact protein could make such widespread contacts.

The CD spectra of the individual fragments and the fact that the fragments can be crystallized are evidence that the fragments exist as well-structured domains. The $P(r)$ curves and radii of gyration of the individual fragments suggest that these domains are relatively compact and globular. The plateau at the high end of the $P(r)$ curve for the N-terminal domain suggests that it may have a protuberance. On the other hand, if IF3 were a compact globular protein, it would have a radius of gyration of about 15 Å, rather than the much larger 27 Å observed. Moreover, the bimodal nature of the $P(r)$ curve for intact IF3, with a primary mode at lower distances and a shoulder at longer distances, is even more pronounced than in the case of calmodulin (Kataoka *et al.*, 1991; Trewhella *et al.*, 1990). The scattering curves and the $P(r)$ functions are indicative of a structure of IF3 consisting of two domains separated by a linker accessible to protease, with a separation of 46 Å between the centers of mass of the two domains.

This separation between the two domains is such that it would permit the binding of one domain on each side of the cleft of the 30S ribosomal subunit. For example, S7 and S11 are on either side of the cleft, and both cross-link to IF3 (MacKeen *et al.*, 1980); these two ribosomal proteins have a center-of-mass separation of about 70 Å (Ramakrishnan *et al.*, 1981). From the $P(r)$ function shown in Figure 3b, the maximum chord in whole IF3 is at least 80−90 Å. Thus whole IF3 is long enough to span the cleft and make contact with both S7 and S11.

Our circular dichroism studies confirm that the secondary structure of the separately expressed domains has not changed from the structure that the domains have in the intact molecule. This is not surprising, as the boundary between the two domains was determined by protease cleavage rather than in some arbitrary manner. In any case, the results provide assurance that the structural work that we are carrying out on the individual domains will be of relevance to the structure of whole IF3. The failure of intact IF3 to crystallize despite extensive trials with the protein from a moderate thermophile suggests that the two domains may not be held in a rigid orientation relative to one another, but may have some flexibility.

Our picture for IF3 of two ribosome binding domains connected by a linker suggests that the structures of these domains will be of interest even separately. Moreover, in combination with the small-angle scattering data, a high-resolution picture of each domain could give us a very good idea of the relative disposition of the two domains and hence of the structure of the whole molecule.

ACKNOWLEDGMENT

We thank Dieter Schneider for help with neutron scattering and John Trunk and Denise Monteleone for help with circular dichroism measurements.

REFERENCES

Ehresmann, C., Moine, H., Mougel, M., Dondon, J., Grunberg-Manago, M., Ebel, J. P., & Ehresmann, B. (1986) *Nucleic Acids Res. 14*, 4803−4821.

Gerchman, S. E., Graziano, V., & Ramakrishnan, V. (1994) *Protein Expression Purif. 5*, 242−251.
Godefroy-Colburn, T., Wolfe, A. D., Dondon, J., Grunberg-Manago, M., Dessen, P., & Pantaloni, D. (1975) *J. Mol. Biol. 94*, 461−487.
Gualerzi, C. O., & Pon, C. L. (1990) *Biochemistry 29*, 5881−5889.
Hartz, D., McPheeters, D. S., & Gold, L. (1989) *Genes Dev. 3*, 1899−1912.
Hartz, D., Binkley, J., Hollingsworth, T., & Gold, L. (1990) *Genes Dev. 4*, 1790−1800.
Jacrot, B. (1976) *Rep. Prog. Phys. 39*, 911−953.
Jacrot, B., & Zaccai, G. (1981) *Biopolymers 20*, 2413−2426.
Kataoka, M., Head, J. F., Persechini, A., Kretsinger, R. H., & Engelman, D. M. (1991) *Biochemistry 30*, 1188−1192.
Kimura, M., Ernst, H., & Appelt, K. (1983) *FEBS Lett. 160*, 78−81.
Laughrea, M., & Tam, J. (1989) *Biochem. Cell Biol. 67*, 812−7.
MacKeen, L. A., Kahan, L., Wahba, A. J., & Schwartz, I. (1980) *J. Biol. Chem. 255*, 10526−10531.
Manavalan, P., & Curtis Johnson, W., Jr. (1987) *Anal. Biochem. 167*, 76−85.
Moore, P. B. (1980) *J. Appl. Crystallogr. 13*, 168−175.
Muralikrishna, P., & Wickstrom, E. (1989) *Biochemistry 28*, 7505−7510.
Pon, C. L., Brombach, M., Thamm, S., & Gualerzi, C. O. (1989) *Mol. Gen. Genet. 218*, 355−357.
Porod, G. (1982) in *Small angle x-ray scattering* (Glatter, O., & Kratky, O., Eds.) pp 17−33, Academic Press, London.
Ramakrishnan, V., & Gerchman, S. E. (1991) *J. Biol. Chem. 266*, 880−885.
Ramakrishnan, V., Yabuki, S., Sillers, I.-Y., Schindler, D. G., Engelman, D. M., & Moore, P. B. (1981) *J. Mol. Biol. 153*, 739−760.
Schneider, D. K., & Schoenborn, B. P. (1984) in *Neutrons in Biology* (Schoenborn, B. P., Ed.) Vol. 27, pp 119−142, Plenum Press, New York.
Stöffler, G., Bald, R., Kastner, B., Lührmann, R., Stöffler-Meilicke, M., & Tishendorf, G. (1980) in *Ribosomes: Structure, Function and Genetics* (Chambliss, G., Craven, G. R., Davies, J., Davis, K., Kahan, L., & Nomura, M., Eds.) pp 171−205, University Park Press, Baltimore.
Studier, F. W., Rosenberg, A. H., Dunn, J. J., & Dubendorff, J. W. (1990) *Methods Enzymol. 185*, 61−89.
Sutherland, J. C., Desmond, E. J., & Takacs, P. Z. (1980) *Nucl. Instrum. Methods 172*, 195−199.
Svergun, D. I., Semenyuk, A. V., & Feigin, L. A. (1988) *Acta Crystallogr. A44*, 244−250.
Trewhella, J., Blumenthal, D. K., Rokop, S. E., & Seeger, P. A. (1990) *Biochemistry 29*, 9316−9324.
Weiel, J., & Hershey, J. W. B. (1982) *Biochemistry 20*, 5859−5865.

BI950042D