Volume 292, number 1,2, 9–12 FEBS 10302 November 1991
© 1991 Federation of European Biochemical Societies 00145793/91/$3.50
*ADONIS* 0014579391010108

# The aromatic amino acid content of the bacterial chaperone protein groEL (cpn60)

## Evidence for the presence of a single tryptophan

Nicholas C. Price[1], Sharon M. Kelly[1], Stephen Wood[2] and Arlene auf der Mauer[2]

[1]*Department of Biological and Molecular Sciences, University of Stirling, FK9 4LA, Scotland, UK* and [2]*Department of Crystallography, Birkbeck College, Malet Street, London, WC1E 7HX, UK*

Received 12 July 1991

Studies of the absorption and fluorescence properties of the chaperone protein groEL (cpn60) from *Escherichia coli* show that tryptophan is present, in contrast to the proposed amino acid sequence of the protein (Hemmingsen, S.M. et al. (1988) Nature 333, 330–334). By determining a suitable value for the specific absorption coefficient of the protein at 280 nm, it has been shown that the content of the aromatic amino acids corresponds to a single tryptophan and (most probably) seven tyrosines per subunit ($M_r$ 57 200).

Chaperone protein; Fluorescence; Spectrophotometry

## 1. INTRODUCTION

The proteins encoded by the *groE* gene of *E. coli* were discovered as critical components for the assembly of large bacteriophages such as lambda and T4 [1,2]. These proteins, groEL, a tetradecamer of subunit $M_r$ 57 200 (also known as cpn60), and groES, a heptamer of subunit $M_r$ 10 300 (also known as cpn10) have been purified [3,4]. They are the subject of much current research designed (i) to establish the way in which they assist in the correct folding and assembly of other proteins, and (ii) to explore the relationships between the various groups of chaperone proteins (for recent reviews, see for example [5–7]). The amino acid sequences (derived from the DNA sequence) of the groEL and groES proteins have been published [8]. This work indicates that the groEL protein contains 7 tyrosines and no tryptophan per subunit. However, earlier amino acid composition data [3] had indicated that tryptophan was present in small amounts (0.19 mol% or approximately 1 per subunit) in the groEL protein. In addition values quoted for the specific absorption coefficient of the groEL protein at 280 nm (using either nitrogen analysis [9] or amino acid analysis [10]) are significantly different from each other, and are both incompatable with the proposed amino acid sequence data (see section 4).

*Abbreviations:* GdnHCl, guanidinium chloride; c.d., circular dichroism.

*Correspondence address:* N.C. Price, Department of Biological and Molecular Sciences, University of Stirling, Stirling, FK9 4LA, Scotland, UK. Fax: (44) (786) 64994.

In view of the well known difficulties in amino acid analysis in the quantitative estimation of tryptophan in proteins, because of the lability of the amino acid on hydrolysis, we have used the more sensitive spectrophotometric (absorption and fluorescence) techniques to show conclusively that tryptophan is present in groEL preparations. By establishing a suitable value for the specific absorption coefficient of the protein at 280 nm, we are able to show that the content corresponds to a single tryptophan per subunit. This information is of interest, not only because of the need to establish protein concentrations in order to calculate, for example, stoichiometries of association with other proteins, but also because the intrinsic fluorescence of groEL should be taken into account when fluorescence is used to monitor the folding of a target protein.

## 2. EXPERIMENTAL

The groEL protein was purified from *E. coli* strain MC1061 carrying the plasmid pND5 described by Jenkins et al. [11] using methods described by Hendrix [3,4]. The product eluted as a single symmetrical peak (of high $M_r$) on gel filtration on Superose 6, on anion-exchange chromatography at pH 7.5 on MonoQ and was at least 98% homogeneous as judged by Coomassie blue staining following SDS-PAGE on 12% acrylamide gels [12]; the subunit $M_r$ was 56 500 ± 2000, consistent with the published value [8].

GdnHCl (ultrapure grade) was purchased from Gibso BRL, Paisley, Scotland, UK. The concentrations of solutions of GdnHCl were checked by refractive index measurements [13].

Fluorescence measurements were made at 20°C on a Perkin Elmer LS50 spectrofluorimeter. C.d. spectra were recorded at 20°C on a JASCO J600 spectropolarimeter; determination of secondary structure was undertaken using the CONTIN procedure [14]. For these measurements, the groEL protein was in the buffer system used by

Lissin et al. [9] in their studies of the self assembly of groEL, namely 50 mM potassium phosphate, 20 mM potassium acetate, 5 mM magnesium acetate, 0.1 mM dithiothreitol, pH 7.5.

Determinations of protein concentrations were undertaken using a variety of methods: (i) far u.v. measurements [15], (ii) Coomassie blue binding [16], (iii) bicinchoninic acid [17], (iv) amino acid analysis. Methods (ii) and (iii) were performed with reference to bovine serum albumin as standard. Method (iii) was performed using the buffer noted above but without dithiothreitol. Method (i) could not be performed using this buffer, because of the high absorption due to acetate in the far u.v.; measurements were therefore made in 50 mM sodium phosphate, pH 7.5. There was no detectable difference in the absorbance at 280 nm of a solution of groEL when these different buffers were used. Amino acid analysis was performed at the WELMET Protein Characterisation Facility at the University of Edinburgh. All spectrophotometric measurements were performed at least 5 times, and the amino acid analysis was performed in duplicate. The error in the various determinations is estimated to be less than ±5%.

## 3. RESULTS

### 3.1. *Evidence for the presence of tryptophan in the groEL protein*

#### 3.1.1. Measurements of absorbance

The ratio of absorbances of a protein at 280 and 288 nm depends critically on the content of tyrosine and tryptophan in the protein [18]. Using the published values for these amino acids at these wavelengths [18], it is possible to calculate the ratio for a protein in the presence of 6M GdnHCl. If tryptophan were absent and only tyrosine present, as in the published amino acid sequence [8], the ratio should be 3.325. If tryptophan is present, the ratio is lowered (for tryptophan itself the ratio is 1.182). From seven separate determinations for the groEL protein in the presence of 6 M GdnHCl, we find a ratio of 1.96±0.05. (It should be noted that the value is slightly lower (1.83±0.05) in the absence of GdnHCl, reflecting environmental effects on the chromophores in the intact protein.) The ratios calculated for various combinations of ratios of tryptophan (W) and tyrosine (Y) are, for example, 1.951 (1W + 7Y); 2.018 (1W + 8Y); 1.650 (2W + 7Y) (these data refer to a protein in the presence of 6 M GdnHCl). Thus, the observed results clearly indicate the presence of tryptophan in the groEL protein.

#### 3.1.2. Fluorescence measurements

The emission wavelengths for tryptophan and tyrosine derivatives are about 355 and 305 nm, respectively [19]. In proteins, there can be shifts from these values because of environmental effects; these effects can be particularly marked in the case of tryptophan side chains where emission maxima lower than 325 nm have been reported [19,20]. The fluorescence spectrum of the groEL protein in buffer (excited at 280 nm) is shown in Fig. 1 (curve A). There is a peak at about 310 nm due predominantly to tyrosine side chains, but significant fluorescence is still observed at wavelengths up to about 400 nm. The spectrum is difficult to interpret unequivocally, because of the environmental effects noted above. The situation is made much clearer when (i) the groEL protein is denatured in 6 M GdnHCl, so that the environmental effects are essentially removed, and (ii) excitation is performed at 290 nm, where tryptophan absorbs much more strongly than tyrosine. Under these circumstances, the fluorescence spectra show distinct maxima at 355 nm (Fig. 1, curves B and C) characteristic of tryptophan emission. Curve D in Fig. 1 shows the spectrum obtained by exciting at 290 nm in the absence of GdnHCl; the emission maximum is 334 nm, corresponding to a moderate degree of exposure of a tryptophan side chain [19,20].

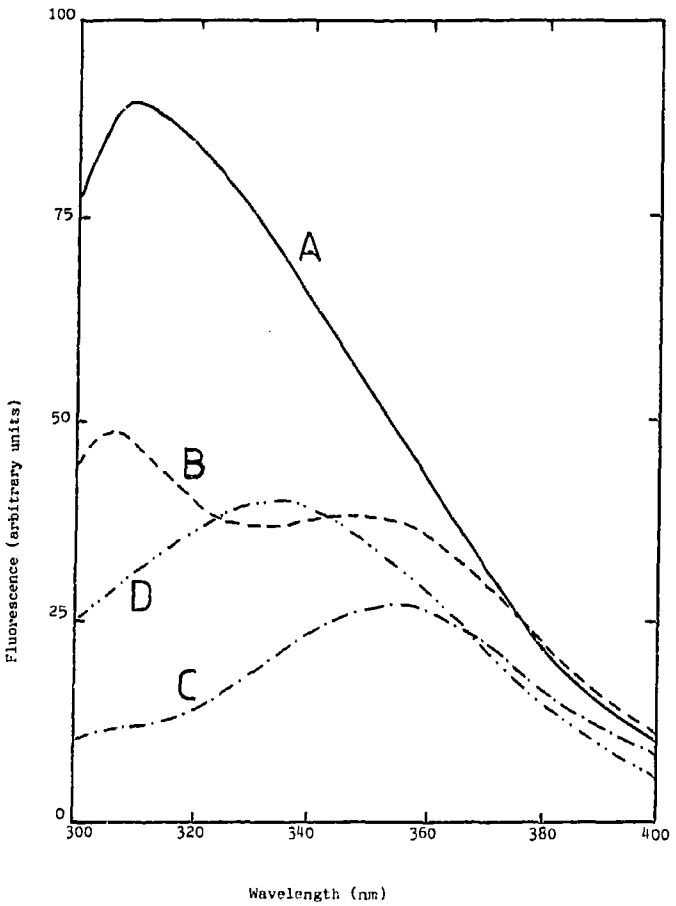


Fig. 1. Fluorescence emission spectra of groEL protein. Spectra were recorded at 20°C in 50 mM potassium phosphate, 20 mM potassium acetate, 5 mM magnesium acetate, 0.1 mM dithiothreitol, pH 7.5. The protein concentration was 61 μg/ml. Curve A, excitation at 280 nm; curve B, as A, but in the presence of 6 M GdnHCl; curve C, as B, but with excitation at 290 nm; curve D, as curve A, but with excitation at 290 nm. Spectra were corrected for the effects of Raman scattering.

### 3.2. *The specific absorption coefficient of the groEL protein at 280 nm*

Using the various methods outlined in the Experimental section, values for the specific absorption coefficient of the groEL protein at 280 nm were determined.

(i) Far u.v. measurements [15]. In 50 mM sodium phosphate buffer, pH 7.5, we find that the absorbance

of the groEL protein at 205 nm is 88.1 times that of the absorbance at 280 nm. In order to evaluate the molar absorption coefficient of the groEL protein at 205 nm, it is necessary to consider the chromophores concerned [15]. The major contribution is made by the peptide bond; from the molar coefficient (2400 [15]), this is calculated to represent 22.95 for a 1 mg/ml solution of the groEL protein assuming the published values [8] for the number of amino acids and the subunit $M_r$. Smaller contributions are made by the side chains of tyrosine, tryptophan, phenylalanine, histidine, arginine and methionine. Using the absorption coefficients for these amino acids [15] and their percent occurrence either from the sequence data [8] or the composition data [3], values of 26.43 or 26.21, respectively, can be calculated for the absorbance of a 1 mg/ml solution of the groEL protein at 205 nm. Using an average value of 26.32, the value of the absorbance at 280 nm for a 1 mg/ml solution is 0.299.

(ii), (iii) The methods involving Coomassie blue [16] or bicinchoninic acid [17] are comparative, in the sense that a standard protein (bovine serum albumin) is used. However, as discussed elsewhere [17,21], they are reasonably reliable measures of protein concentration, since they reflect the occurrence of positively charged side chains and peptide bonds, respectively, and in both respects the properties of the groEL protein are reasonably typical. The values obtained by the two methods for the absorbance at 280 nm of a 1 mg/ml solution of the groEL protein are 0.305 and 0.282, respectively.

(iv) Amino acid analysis. In order to relate the results of amino acid analysis to the content of protein, it is necessary to have accurate composition data. As mentioned in section 4, there is very close overall agreement between the results of direct amino acid analysis [3] and DNA (and hence amino acid) sequencing [8]. In particular, the data for the contents of lysine and arginine agree within 5%. These amino acids were chosen for analysis in view of their stability upon hydrolysis and their relative abundance in the groEL protein (7.3 and 4.0 mol%, respectively). On the basis of the amino acid analysis, the absorbance at 280 nm of a 1 mg/ml solution of groEL protein in buffer is 0.280 (based on the lysine content) or 0.275 (based on the arginine content). The tyrosine content was found to be 7.0 per subunit, a value consistent with the amino acid sequence (7) [8] and amino acid analysis (7.55) [3] information.

Taking all the values above into account, an average value of 0.288 for the absorbance at 280 nm of a 1 mg/ml solution of groEL is obtained. This value corresponds reasonably closely with that reported by Lissin et al. [9], based on nitrogne analysis. These authors reported a value of 0.25 at 276.5 nm; this would correspond to a value of 0.242 at 280 nm. However, it is likely that Lissin et al. [9] based their calculations on nitrogen content of 16%, the typical value for proteins. In fact, because of the preponderance of lighter amino acids in the groEL protein, the actual nitrogen content is 16.9%, which would have the effect of raising the absorbance value to 0.255. If this value is included along with our data, an average value of 0.283 is derived. In conclusion, we consider that a value of 0.285 for the specific absorption coefficient (litre $\cdot g^{-1} \cdot cm^{-1}$) at 280 nm is appropriate for the groEL protein.

### 3.3. *Content of tryptophan and tyrosine in the groEL protein*

Assuming the above value for the absorbance of the groEL protein at 280 nm, the content of the aromatic amino acids can be calculated. The absorbance at 280 nm in the presence of 6 M GdnHCl is found to be 0.95±0.01 times that in the absence of GdnHCl, i.e. 0.271 for a 1 mg/ml solution; at 288 nm, using the ratio determined earlier, the value is 0.138. Taking into account the reference values for tryptophan and tyrosine at these wavelengths [18], the numbers of these amino acids per subunit ($M_r$ 57 200) of the groEL protein are 1.04 and 7.47, respectively. Additional estimates of the numbers of these amino acids can be derived from fluorescence spectra in the presence of 6 M GdnHCl (see Fig. 1), using the *N*-acetyl amides of tryptophan and tyrosine as reference compounds. From five separate determinations the contents of these amino acids, based on the fluorescence at 355 and 305 nm, are 0.72 and 6.84 per subunit of the groEL protein, respectively. It is probably that the somewhat lower values than those determined by absorbance measurements reflect a degree of quenching by neighbouring portions of the polypeptide chain in the unfolded protein.

## 4. DISCUSSION

The data in this paper show that there is a single tryptophan and (most probably) seven tyrosines per subunit of the groEL protein. Whereas the latter value is in accord with the published amino acid sequence [8], the presence of tryptophan indicates that there is an error in the sequence data, arising presumably from a DNA sequencing error. The presence of tryptophan in the protein had been indicated by Hendrix [3], but the discrepancy with the sequence data had not been commented on. We believe that any DNA sequencing error(s) is (are) likely to be minor, since a comparison of the amino acid composition data determined directly [3] and by the DNA (and hence amino acid sequence) [8] using the method of Cornish-Bowden [22] shows that the compositions are very similar indeed, with a difference index ($S\Delta n$) of only $0.11N$, where $N$ is the number of amino acids in the polypeptide chain. In fact, the only serious deviations (outside 10%) are for histidine, cysteine and tryptophan, all of which are present in very small amounts.

It should be noted that amino acid sequences (derived from DNA sequences) have been determined for a num-

ber of groEL homologues in other organisms. A number, including *Mycobacterium leprae*, *M. tuberculosis*, *M. bovis* and human mitochondrial cpn60 show a conserved tryptophan at position 44 (groEL numbering), while the plastic derived products show a conserved tryptophan at position 475. Both of these regions of the *E. coli groE* gene have been sequenced more than once [8,23,24] and it seems unlikely that any error would be duplicated. An alternative explanation to a sequencing error is that a mutation has occurred leading to an amino acid change in the groEL protein. It would require careful sequence analysis of both the original gene and derived plasmids to exclude this possibility, but it should be noted that evidence from specific absorption values (see below) suggests that tryptophan is present in the groEL protein products of a number of different constructed plasmids [9,10].

The value for the specific absorption coefficient at 280 nm (0.285) for the groEL protein has been derived by a number of independent techniques, which, taken together, give a coherent picture. As further (albeit indirect) evidence for this value, it should be noted that when this value of the coefficient is used to calculate the mean residue ellipticities from the far u.v. c.d. spectrum, of the groEL protein, an $\alpha$-helical content of 56% is derived; this is close to the value obtained (57%) by applying the PREDICT set of programs [25] to the amino acid sequence of the protein, and to the value (57%) previously obtained by c.d. measurements [9].

It is of interest that the discrepancies between the observed absorption properties of the protein [9] and those predicted on the basis of the amino acid sequence [8] had not been commented on previously. If groEL contained seven tyrosines and no tryptophan per subunit, the absorbance of a 1 mg/ml solution at 280 nm can be calculated [18] as 0.157 (in 6 M GdnHCl), i.e. 0.165 in buffer. When a single tryptophan is included, the values become 0.256 and 0.270, respectively (very close to the values determined in this paper). The value reported by Viitanen et al. [10] for the absorption coefficient of the groEL protein at 280 nm (23800 $M^{-1} \cdot cm^{-1}$) would correspond to 0.416 for a 1 mg/ml solution; however, these authors gave relatively few details about the determination of this value, and made no comment on the discrepancy from the value predicted on the basis of the published amino acid sequence [8].

A final possible explanation for our results is that tryptophan is not present in the groEL protein, but is present in a contaminating protein in the preparation. We consider this unlikely in view of our analytical data on the preparation, since such a contaminant would either possess the same $M_r$ as groEL under denaturing (SDS-PAGE) and non-denaturing (gel filtration) conditions or would be rich in tryptophan. (Assuming a 98% homogeneity value and a content of 0.19 mol% tryptophan overall, the contaminant would possess a tryptophan content of 50 × 0.19 mol%, i.e. 9.5 mol%.) Such a contaminant would bemost unusual.

The data in this paper will help in the further characterisation of the groEL protein and its role in assisting protein folding and assembly, not only in calculating the concentration of the protein, but also in indicating the need to consider its intrinsic fluorescence when studying the folding of other proteins in reconstituted systems [25]. The location of the tryptophan residue in the protein will be undertaken by a combination of specific cleavage and sequencing methods, both at the DNA and protein level.

*Acknowledgements:* We wish to thank the SERC and AFRC for financial support, Prof. John Ellis for encouragement and Mr. A. Cronshaw for help with amino acid analysis.

## REFERENCES

[1] Georgopoulos, C.P., Hendrix, R.W., Casjens, S.R. and Kaiser, A.D. (1973) J. Mol. Biol. 76, 45–60.
[2] Georgopoulos, C.P., Hendrix, R.W., Kaiser, A.D. and Wood, W.B. (1972) Nature New Biol. 239, 38–41.
[3] Hendrix, R.W. (1979) J. Mol. Biol. 129, 375–392.
[4] Chandrasekhar, G.N., Tilly, K., Woolford, C., Hendrix, R. and Georgopoulos, C. (1986) J. Biol. Chem. 261, 12414–12419.
[5] Ellis, R.J. and Hemmingsen, S.M. (1989) Trends Biochem. Sci. 14, 339–342.
[6] Ellis, R.J. (1990) Science 250, 954–959.
[7] Ellis, R.J. and van der Vies, S.M. (1991) Annu. Rev. Biochem. 60, 321–347.
[8] Hemmingsen, S.M., Woolford, C., van der Vies, S.M., Tilly, K., Dennis, D.T., Georgopoulos, C.P., Hendrix, R.W. and Ellis, R.J. (1988) Nature 333, 330–334.
[9] Lissin, N.M., Venyaminov, S. Yu. and Grishovich, A.S. (1990) Nature 348, 339–342.
[10] Viitanen, P.V., Lubben, T.H., Reed, J., Goloubinoff, P., O'Keefe, D.P. and Lorimer, G.H. (1990) Biochemistry 29, 5665–5671.
[11] Jenkins, A.J., March, J.B., Oliver, I.R. and Masters, M. (1986) Mol. Gen. Genet. 202, 446–454.
[12] Laemmli, U.K. (1970) Nature 227, 680–685.
[13] Nozaki, Y. (1972) Methods Enzymol. 26, 43–50.
[14] Provencher, S.W. and Glöckner, J. (1981) Biochemistry 20, 33–37.
[15] Scopes, R.K. (1974) Anal. Biochem. 79, 277–282.
[16] Sedmak, J.J. and Grossberg, S.E. (1977) Anal. Biochem. 79, 544–552.
[17] Smith, P.K., Krohn, R.I., Hermanson, G.T., Mallia, A.K., Gartner, F.H., Provenzano, M.D., Fujimoto, E.K., Goeke, N.M., Olson, B.J. and Klenk, D.C. (1985) Anal. Biochem. 150, 76–85.
[18] Edelhoch, H. (1967) Biochemistry 6, 1948–1954.
[19] Brand, L. and Witholt, B. (1967) Methods Enzymol. 11, 776–856.
[20] Eftink, M.R. and Ghiron, C.A. (1976) Biochemistry 15, 672–680.
[21] Peterson, G.L. (1983) Methods Enzymol. 91, 95–119.
[22] Cornish-Bowden, A. (1983) Methods Enzymol. 91, 60–75.
[23] Chanda, P.K., Ono, M., Kuwano, M. and Kung, H.-F. (1985) J. Bacteriol. 161, 446–449.
[24] Miki, T., Orita, T., Furuno, M. and Horiuchi, T. (1988) J. Mol. Biol. 201, 327–328.
[25] Eliopoulos, E.E., Geddes, A.J., Brett, M., Pappin, D.J.C. and Findlay, J.B.C. (1982) Int. J. Biol. Macromol. 4, 263–268.
[26] Martin, J., Langer, T., Boteva, R., Schramel, A., Horwich, A.L. and Hartl, F.-U. (1991) Nature 352, 36–42.