THE PRIMARY STRUCTURE OF THE INITIATION FACTOR IF-3 FROM *ESCHERICHIA COLI*

Dieter BRAUER and Brigitte WITTMANN-LIEBOLD
Max-Planck-Institut für Molekulare Genetik, Abt. Wittmann, Berlin-Dahlem, Germany

Received 18 May 1977

1. Introduction

The initiation factor IF-3 plays an important role during the initiation step of protein biosynthesis in *E. coli* and other organisms (see ref. [1] for a review), although its exact mode of action is still under investigation.

In an accompanying paper [2] the existence of at least two forms is described which differ in their electrophoretic mobilities and their molecular weights. It is not yet clear how these forms (IF-3ℓ and IF-3s) are related to two other forms (IF-3α and IF-3β) reported previously [3,4].

As shown by the protein-chemical methods illustrated in this paper the various forms of IF-3 differ in their N-termini. The long form, IF-3ℓ, consists of 181 amino acids whereas the short form, IF-3s, has only 175 amino acids: the first six amino acids of IF-3ℓ are absent in IF-3s. In addition to these two main forms, the existence of at least one minor form has been shown which differs from IF-3ℓ by the absence of the first amino acid only, namely *N*-monomethylmethionine.

2. Materials and methods

2.1. *Isolation of the different forms of IF-3*

E. coli K cells were ruptured by grinding with alumina and the ribosomes pelleted by centrifugation. After dialysis against 1 mM Mg^{2+} the subunits were separated by zonal centrifugation in Beckman Ti-15 rotors. Proteins were extracted from 30 S subunits and separated by CM-cellulose chromatography as described previously [5]. The IF-3s protein eluted somewhat earlier from the CM-cellulose column than IF-3ℓ and was separated from the coeluting ribosomal protein S16 by gel filtration on Sephadex G-100 in 15% acetic acid. The isolated proteins were checked for their purity by two-dimensional gel electrophoresis [6] and stored in 1% acetic acid at −20°C until lyophilisation before use. All protein samples needed for the determination of the primary structure of IF-3 were kindly provided by Dr H. G. Wittmann.

2.2. *Enzymic digestions*

The proteins were digested with trypsin pretreated with TPCK (1-chloro-4-phenyl-3-tosyl-amidobutan-2-one) at pH 7.8, at 37°C for 4 h. Digestion with *Staphylococcus aureus* protease [7] from Miles Biochemicals (Frankfurt/Main) was made in 0.1% ammonium acetate buffer, pH 4.0, at 37°C for 40 h and with thermolysin (Serva, Heidelberg) at pH 7.5, at 50°C for 90 min. Details of the enzymic digestion procedures have been described previously [8–10]. Digestion with carboxypeptidase C (Boehringer, Mannheim) was performed by a modified procedure [11] at pH 5.3, at 50°C for various lengths of times up to 45 min.

2.3. *Modified enzymic digestions*

The proteins were reversibly modified with ETPA (Exo-cis-3, 6-endoxo-Δ^4-tetrahydrophthalic anhydride) by protecting the ϵ-amino groups of the lysine residues [12] and then digested with TPCK-trypsin. Modification of the arginine residues with 1,2 cyclohexandione and tryptic digestion was made by a slight variation of the procedure described in ref. [13].

2.4. *Chemical digestion with CNBr*

Cleavage of IF-3 with CNBr was done for 48 h at room temperature in the dark as described [14].

2.5. *Isolation of peptides*

The fingerprint technique on cellulose thin-layer plates [8,10] gave information about the number and the amino acid composition of the peptides. The separation of peptides on a preparative scale was made either by gel filtration on Sephadex G-50 (superfine, 140 × 1 cm) in 10% acetic acid [10] or on a micro-column of Dowex M71 (0.3 × 10 cm) at 55°C using pyridine formate gradients [10,15]. The fractions were examined either photometrically or by spotting aliquots of every second fraction onto cellulose thin-layer plates. Fractions containing more than one peptide were further subjected to chromatography, electrophoresis or the fingerprint technique on thin-layer plates.

2.6. *Amino acid analyses*

Analyses of protein samples and of peptides were performed on a Durrum D-500 and on a LKB-Biocal 3201 analyser, as described previously [10,16]. For determination of cysteine the proteins or peptides were oxidized with performic acid before acid hydrolysis. The presence of tryptophan was tested by amino acid analysis after hydrolysis with methane sulphonic acid [17] and by spraying of fingerprint plates with *p*-aminobenzaldehyde [18].

2.7. *Methods for sequencing*

The amino acid sequences were determined by the following methods:

(a) Automated Edman degradation [19] of intact proteins in an improved Beckman sequenator [20].

(b) Solid-phase sequencing [21] with attachment of the C-terminal carboxyl groups to amino–polystyrene resins [22,23] and with attachment of the N-terminal and side-chain amino groups to aminopropyl–glass [24,25].

(c) Manual Edman degradation combined with dansylation of the free N-terminal amino acid residue after each degradation step [26].

The determination of aspartic acid, glutamic acid and their amides was achieved by converting the released 2-anilino-5-thiazolinone derivatives to the phenylthiohydantoin derivatives followed by identification with chromatographic [20] or electrophoretic methods [10,27]. In cases where the identification after dansylation was uncertain and also for quantitative determinations, the free amino acids were liberated from their 2-anilino-5-thiazolinones or phenylthiohydantoin derivatives by hydrolysis and identified on an amino acid analyser.

3. Results and discussion

3.1. *Sequence analysis of IF-3ℓ*

The protein sample containing IF-3ℓ was subjected to digestion with various proteases and to chemical cleavage with CNBr. The resulting peptides were isolated and sequenced as illustrated in fig.1. Furthermore, automatic Edman degradation was performed with the intact proteins in an improved Beckman sequenator. Combination of all results which will be described in detail elsewhere led to the primary structure of IF-3ℓ as shown in fig.1. The longest form IF-3ℓ consists of 181 amino acid residues, with the amino acid composition of NMM_1, Asp_4, Asn_7, Thr_5, Ser_7, Glu_{21}, Gln_{13}, Pro_7, Gly_{14}, Ala_{10}, Cys_1, Val_{15}, Met_6, Ile_{12}, Leu_{14}, Tyr_3, Phe_5, His_1, Lys_{20} and Arg_{15}. Its N-terminal amino acid was determined as *N*-monomethylmethionine (NMM) from amino acid analyser data [28] and by mass spectrometry of the PTH-derivative.

During these studies it became clear that the protein sample contained, besides IF-3ℓ as the major form, also a minor form, called IF-3ℓ′, which differed from IF-3ℓ only by the absence of the first amino acid, namely *N*-monomethylmethionine. This was shown by the isolation and sequence analysis of several peptides lacking the N-terminal amino acid (fig.2). Automatic Edman degradation gave two amino acids in each step: one main spot for IF-3ℓ and a minor spot for IF-3ℓ′. Since the two forms differ in only one neutral amino

Fig.1. The primary structure of IF-3ℓ from *Escherichia coli*. (SQ) Liquid-phase Edman degradation in a modified sequenator. (⁻₊) Means unambiguously identified residue by thin-layer chromatography. (T) Tryptic peptides. (ET) Tryptic peptides after blocking the lysines by ETPA. (A) Tryptic peptides after blocking the arginines by 1,2 cyclohexandione. (SP) Peptides from digestion with Staphylococcus protease. (TH) Peptides from thermolysin digestion. (CNBr) Peptides resulting from CNBr cleavage. (Case C) Carboxypeptidase C digestion.

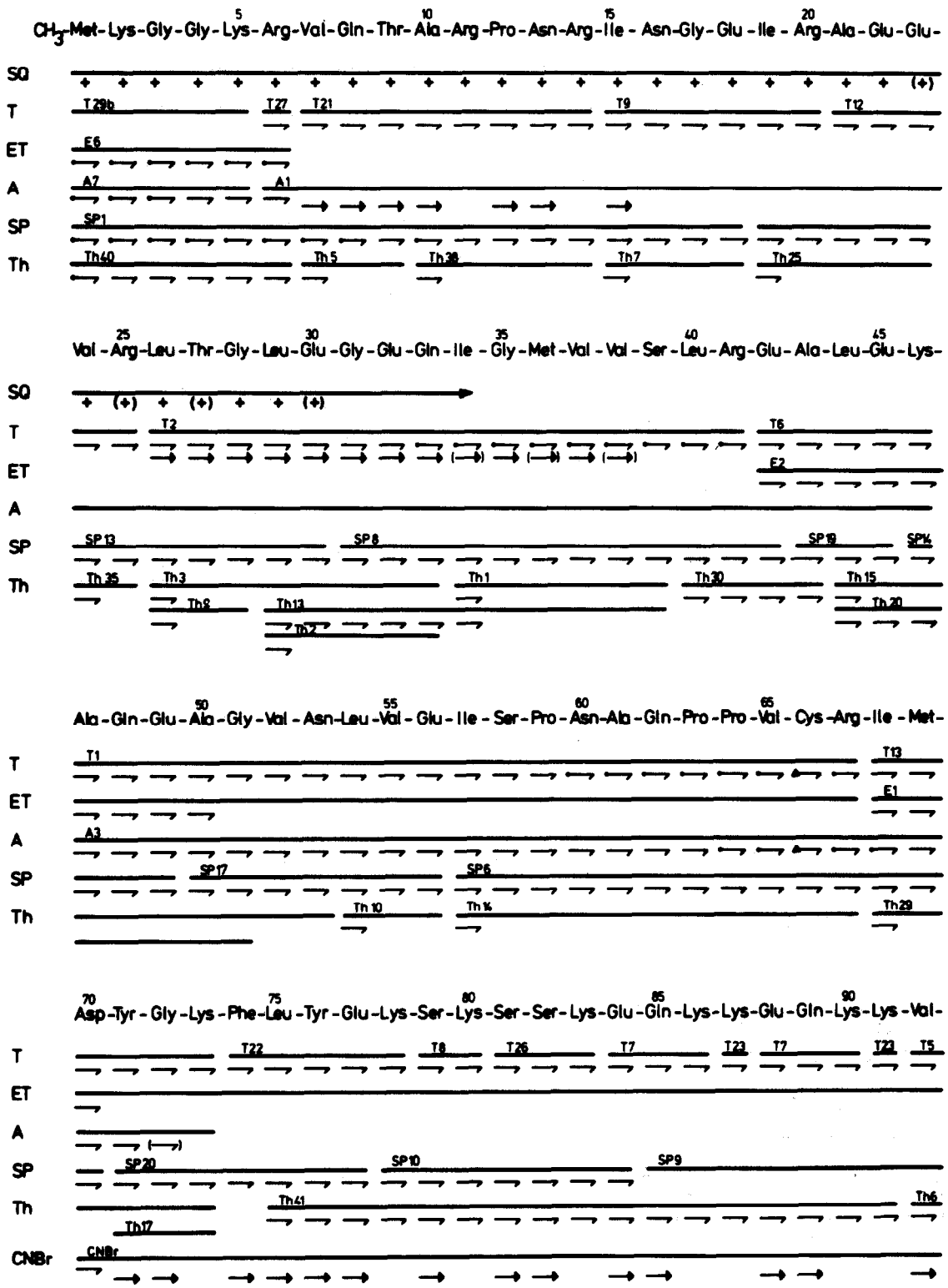
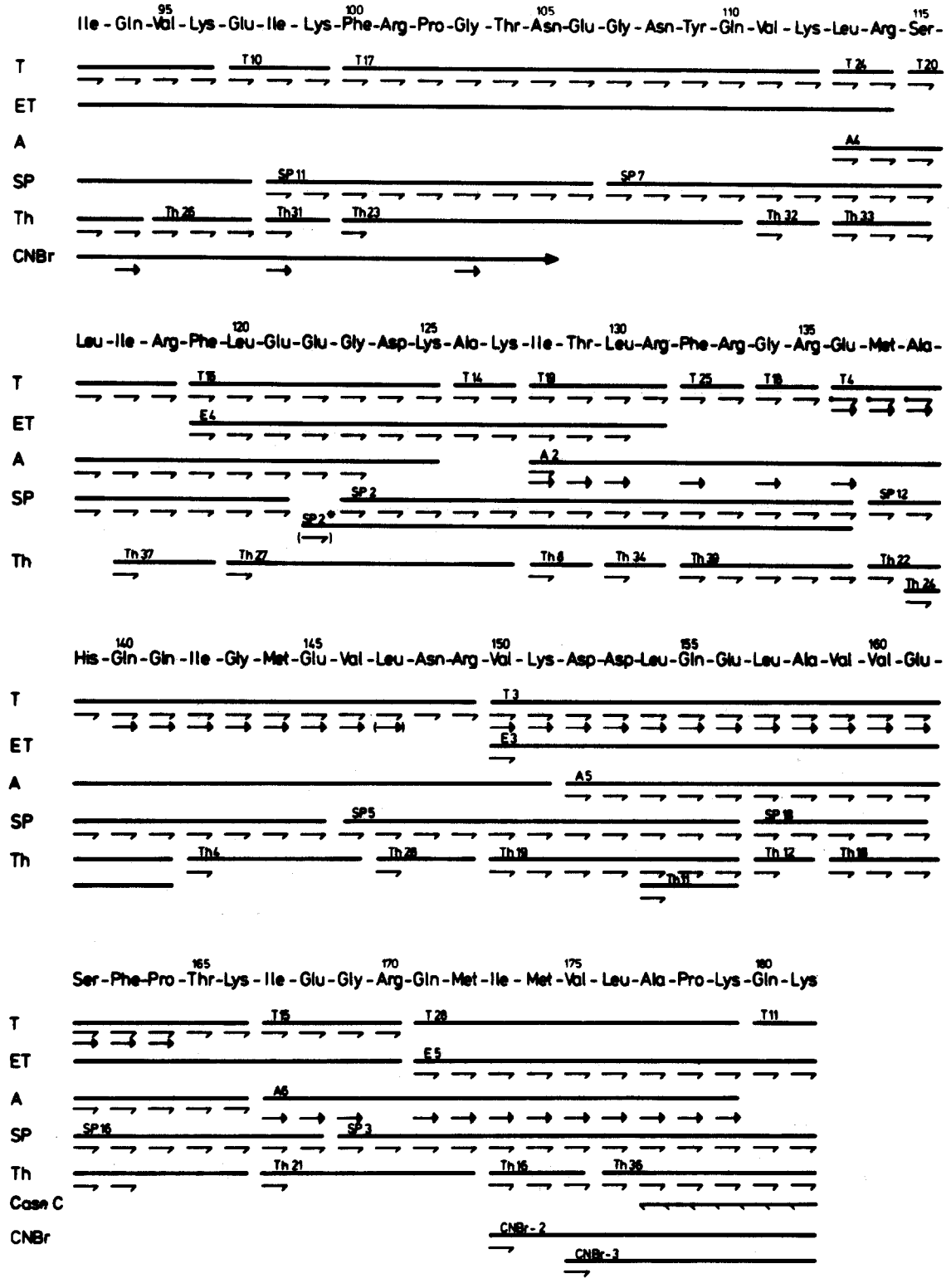


Fig.1 continued overleaf


Ile - Gln - Val - Lys - Glu - Ile - Lys - Phe - Arg - Pro - Gly - Thr - Asn - Glu - Gly - Asn - Tyr - Gln - Val - Lys - Leu - Arg - Ser -
95
100
105
110
115
T
ET
A
SP
Th
CNBr
T10
T17
T24
T20
A4
SP11
SP7
Th26
Th31
Th23
Th32
Th33
Leu - Ile - Arg - Phe - Leu - Glu - Glu - Gly - Asp - Lys - Ala - Lys - Ile - Thr - Leu - Arg - Phe - Arg - Gly - Arg - Glu - Met - Ala -
120
125
130
135
T15
T14
T18
T25
T19
T4
E4
A2
SP2
SP2*
SP12
Th37
Th27
Th8
Th34
Th39
Th22
Th24
His - Gln - Gln - Ile - Gly - Met - Glu - Val - Leu - Asn - Arg - Val - Lys - Asp - Asp - Leu - Gln - Glu - Leu - Ala - Val - Val - Glu -
140
145
150
155
160
T3
E3
A5
SP5
SP18
Th6
Th28
Th19
Th31
Th12
Th38
Ser - Phe - Pro - Thr - Lys - Ile - Glu - Gly - Arg - Gln - Met - Ile - Met - Val - Leu - Ala - Pro - Lys - Gln - Lys
165
170
175
180
T15
T28
T11
E5
A6
SP16
SP3
Th21
Th16
Th36
Casn C
CNBr
CNBr-2
CNBr-3

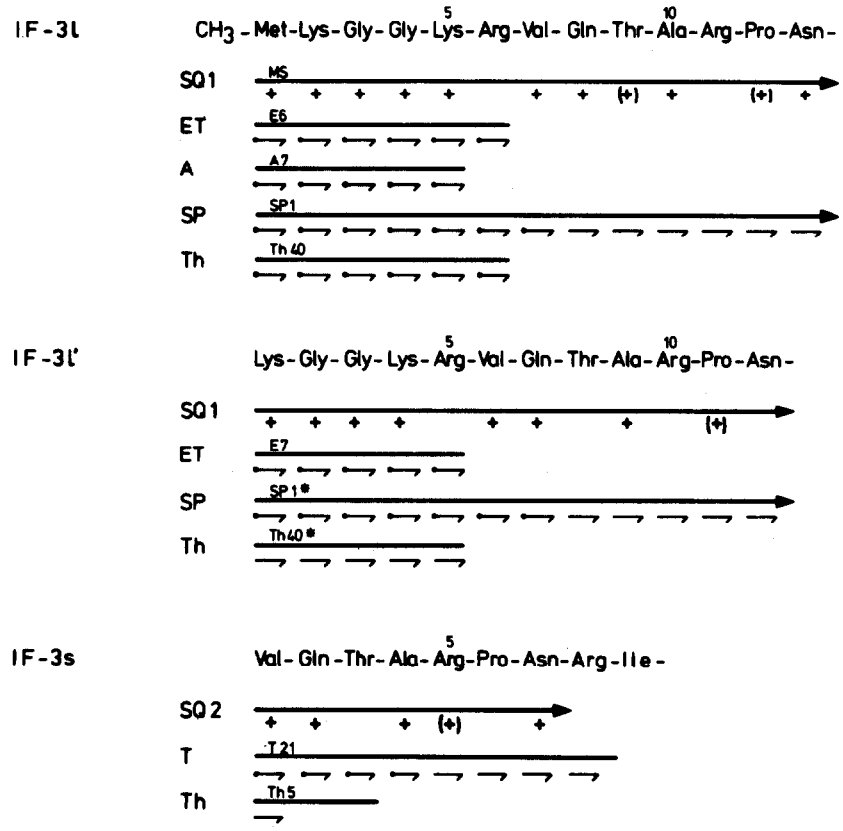


Fig.2. N-Terminal sequences of the different forms IF-3ℓ, IF-3ℓ', IF-3s of the initiation factor from *Escherichia coli*. For abbreviations see fig.1.

acid, i.e., they contain 181 or 180 amino acids, a separation between them could not be achieved by chromatographic or electrophoretic methods; even not by two-dimensional gel electrophoresis [2].

3.2. *Sequence analysis of IF-3s*

Analogous sequence studies as described for IF-3ℓ were also done with the protein sample containing IF-3s. The sequenator runs with this protein resulted in the following order of amino acids: Val–Gln–Thr–Ala–Arg–Pro–Asn ... This sequence is identical with that of IF-3ℓ, beginning at position 7, i.e., the first six amino acids of the IF-3ℓ protein chain are absent in IF-3s (fig.2). Comparison of the fingerprint maps obtained by tryptic digestion of each of the two forms also showed no other difference between them. The conclusion that both forms have identical sequences (except the N-terminal region) is strengthened by the finding that no peptides were found which differed between the two forms and that treatment of IF-3ℓ and IF-3s with carboxypeptidase C released the same amino acids in both cases. All these results support the finding that IF-3s differs from IF-3ℓ only by the lack of the first six amino acids (fig.2).

One protein sample contained (besides IF-3s as the major form) a minor form whose sequence starts with Arg–Val–Gln–Thr–Ala–Arg–Pro–Asn. . . , as shown by a sequenator run; i.e., the first five amino acids of IF-3ℓ are absent in this minor form. Further studies are necessary to study the frequency of this minor form.

3.3. *Molecular weights*

The molecular weights of the two main forms, IF-3ℓ and IF-3s, are 20 668 and 19 997, respectively. These chemical molecular weights are approximately

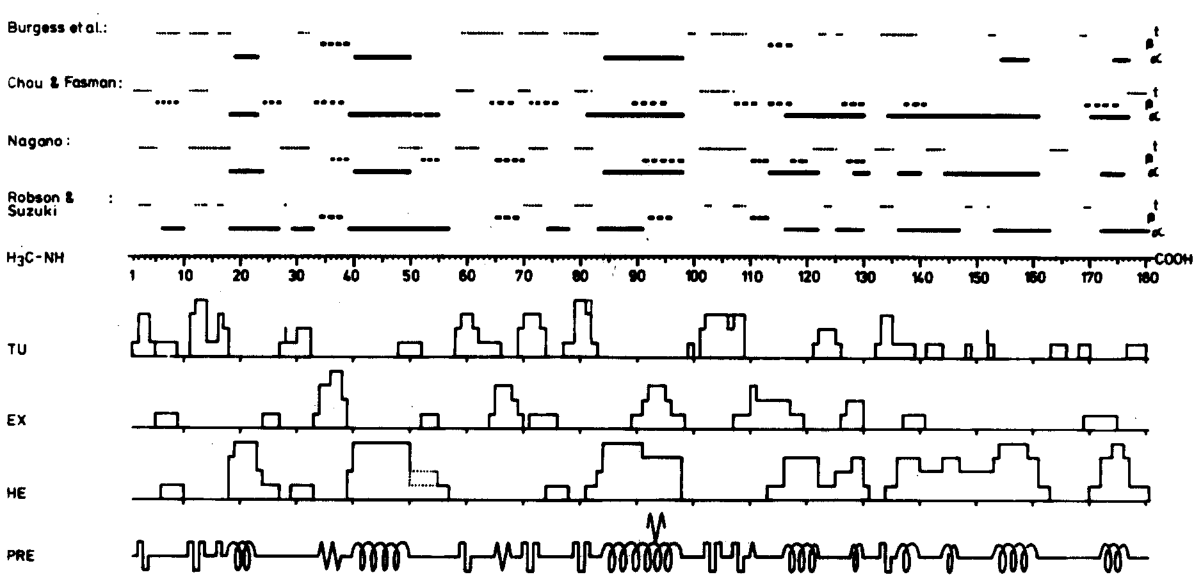


Fig.3. Prediction of the secondary structure of protein IF-3ℓ. In the upper part the predictions are given according to four different methods. Turns or loops are indicated by (.....) (t). β-Structure is indicated by (– – –) (β). α-Helix is indicated by (——) (α). In the lower part the histograms of turns or loops (TU), extended structure (EX) and helix (HE) configurations are presented. In the last line, denoted by 'PRE', the average prediction when at least 3 methods are in full agreement, is given: ( ) turn or loops, ( ) extended structure, ( ) helix regions.

10% lower than those determined by SDS gel electrophoresis [1]. However, it is a general experience that the latter technique often results in molecular weight values for basic proteins, e.g., ribosomal proteins, which are somewhat too high.

3.4. *Structure comparison with ribosomal proteins*

To test the possibility that there exist structure homologies between IF-3 and ribosomal proteins, the primary structure of IF-3ℓ was compared by a computer programme with those of 25 *E. coli* ribosomal proteins whose amino acid sequences have been completely determined. Besides a number of tripeptides, only eight tetrapeptides and no peptides of longer lengths were found to be identical in IF-3ℓ and any of the ribosomal proteins. This represents a relatively low degree of sequence homology and can easily be explained on a random basis [29].

3.5. *Secondary structure*

The knowledge of the amino acid sequence of a protein allows the prediction of its secondary structure. Four computer programmes [30,31] were used to predict the secondary structure of IF-3ℓ according to four different methods [32–37], and the results are shown in fig.3. There is a reasonable agreement among the four methods; the resulting secondary structure was predicted ('PRE') according to at least three programmes resulting in the same conformation. The values for the α-helix content of IF-3ℓ range from 34–37%, for β-structure from 6–8%, for turns (or loops) from 15–17%, and for random coil from 40–43%. Experimentally determined values as obtained from circular dichroism studies on IF-3 are 27% for α-helix and 18% for β-structure [38]. This is in relatively good agreement with the predicted values mentioned above.

Acknowledgements

We thank Mr Alfred Franz and Mr Arnold Lehmann for excellent technical assistance and Dr R. Brimacombe for kindly reading the manuscript.

References

[1] Ochoa, S. and Mazumder, R. (1974) in: The Enyzmes (Boyer, P. D. ed) Vol. 10, pp. 1–51, Academic Press, New York.
[2] Suryanarayana, T. and Subramanian, A. R. FEBS Lett. (1977) 79, 264–268.
[3] Lee-Huang, S. and Ochoa, S. (1971) Nature New Biol. 234, 236–239.
[4] Lee-Huang, S. and Ochoa, S. (1973) Arch. Biochem. Biophys. 156, 84–96.
[5] Hindennach, I., Stöffler, G. and Wittmann, H. G. (1971) Eur. J. Biochem. 23, 7–12.
[6] Kaltschmidt, E. and Wittmann, H. G. (1970) Anal. Biochem. 36, 401–412.
[7] Houmard, J. and Drapeau, G. R. (1972) Proc. Natl. Acad. Sci. USA 69, 3506–3509.
[8] Bitar, K. G. and Wittmann-Liebold, B. (1975) Hoppe-Seyler's Z. Physiol. Chem. 356, 1343–1352.
[9] Stadler, H. and Wittmann-Liebold, B. (1976) Eur. J. Biochem. 66, 49–56.
[10] Heiland, I., Brauer, D. and Wittmann-Liebold, B. (1976) Hoppe-Seyler's Z. Physiol. Chem. 357, 1751–1770.
[11] Tscheche, H. and Kupfer, S. (1972) Eur. J. Biochem. 26, 33–36.
[12] Riley, M. and Perham, S. (1970) Biochem. J. 118, 733–739.
[13] Toi, K., Bynum, E., Norris, E. and Itano, H. (1967) J. Biol. Chem. 242, 1036–1043.
[14] Gross, E. and Witkop, B. (1961) J. Am. Chem. Soc. 83, 1510.
[15] Chen, R. and Wittmann-Liebold, B. (1975) FEBS Lett. 52, 139–140.
[16] Hitz, H. Schäfer, D. and Wittmann-Liebold, B. (1977) Eur. J. Biochem. in press.
[17] Liu, T. Y. and Chang, Y. H. (1971) J. Biol. Chem. 246, 2842–2848.
[18] Needleman, S. B. (1975) in: Protein Sequence Determination; Molecular Biology, Biochemistry and Biophysics 8, p. 201, Springer Verlag, Berlin, Heidelberg, New York.
[19] Edman, P. and Begg, G. (1967) Eur. J. Biochem. 1, 80–91.
[20] Wittmann-Liebold, B. (1973) Hoppe-Seyler's Z. Physiol. Chem. 354, 1415–1431.
[21] Laursen, R. A. (1971) Eur. J. Biochem. 20, 89–102.
[22] Previero, A., Derancourt, J., Coletti-Previero, M. A. and Laursen, R. A. (1975) FEBS Lett. 33, 135–138.
[23] Wittmann-Liebold, B. and Lehmann, A. (1975) in: Solid Phase Methods in Protein Sequence Analysis (Laursen, R. A. ed) pp. 81–90, Pierce Chemical Company, Rockford, USA.
[24] Robinson, P. J., Dunhill, P. and Lilly, M. D. (1971) Biochim. Biophys. Acta 242, 659–662.
[25] Machleidt, W., Wachter, E., Scheulen, M. and Otto, J. (1973) FEBS Lett. 37, 217–220.
[26] Gray, W. R. and Hartley, B. S. (1963) Biochem. J. 89, 379.
[27] Brauer, D., details to be published.
[28] Chen, R., Brosius, J., Wittmann-Liebold, B. and Schäfer, W. (1977) J. Mol. Biol. 111, 173–181.
[29] Wittmann-Liebold, B. and Dzionara, M. (1976) FEBS Lett. 65, 281–283.
[30] Wittmann-Liebold, B., Robinson, S. M. L. and Dzionara, M. (1977) FEBS Lett. 77, 301–307.
[31] Dzionara, M., Robinson, S. M. L. and Wittmann-Liebold, B., paper submitted.
[32] Burgess, A. W., Ponnuswamy, P. K. and Sheraga, H. A. (1974) Israel J. Chem. 12, 239–286.
[33] Chou, P. Y. and Fasman, G. D. (1974) Biochemistry 13, 211–222.
[34] Chou, P. Y. and Fasman, G. D. (1974) Biochemistry 13, 222–244.
[35] Chou, P. Y., Adler, A. J. and Fasman, G. D. (1975) J. Mol. Biol. 96, 29–45.
[36] Nagano, K. (1977) J. Mol. Biol. 109, 251–274.
[37] Robson, B. and Suzuki, E. (1976) J. Mol. Biol. 107, 327–356.
[38] Heiland, I. (1976) unpublished results.