A characterization by sequencing of the termini of the polypeptide chain of cyclic AMP receptor protein from *Escherichia coli* and the subtilisin produced N-terminal fragment

A. Tsugita, B. Blazy*, M. Takahashi* and A. Baudras*

European Molecular Biology Laboratory, 6900 Heidelberg, FRG and * *Centre de Recherche de Biochimie et de Génétique Cellulaires du CNRS, 118, Route de Narbonne, 31062 Toulouse Cedex, France*

Received 7 June 1982

Amino acid composition *CDP (CAP)* *Terminal sequence* *cAMP-binding fragment*

1. INTRODUCTION

The cyclic AMP (cAMP) receptor protein (CRP), also called catabolite activator protein (CAP), is a regulatory protein which controls expression of several catabolite-sensitive operons in *Escherichia coli* [1–6]. Typically, the cAMP • CRP complex stimulates transcription of certain catabolite genes but it is also known to shut off expression of several other genes [7–9] and to interfere with natural polarity of polycistronic units [10]. The protein has M_r 45 000 and it is composed of 2 apparently identical subunits [11]. In the presence of cAMP limited proteolytic digestion of CRP with subtilisin results in a resistant N-terminal core, α-CRP, which retains the dimeric structure and cAMP binding activity but has lost cAMP-dependent DNA-binding capacity at pH 8 as judged by the nitrocellulose filter assay [12,13]. However, from a thermal denaturation study on double-stranded poly [d(AT)] we have shown that α-CRP stabilizes the double-stranded structure of the polynucleotide and that this effect is modulated by cAMP [14]. This observation demonstrates that part of the cAMP-dependent DNA binding capacity is still present in the N-terminal fragment. The unique nucleic acid sequence of the *crp* gene coding for CRP has been established by 2 independent groups [15,16]. Although the agreement is good between the amino acid compositions predicted from this sequence and the published chemical analysis, some doubt exists as to the precise nature of the amino acid sequence at the N- and C-termini [15,16]. Therefore, we have determined the amino acid composition and the N- and C-terminal partial sequence of CRP purified from a wild-type strain and a super-producing strain [15] and have located the unique cleavage point which generates α-CRP by subtilisin limited digestion.

2. EXPERIMENTAL

CRP was purified to homogeneity as outlined in [17] from two *E. coli* K-12 strains, namely KLF 41/JC 1553 employed previously and a strain containing recombinant plasmid pBS*crp2* in [15]. In cell extracts of the latter strain, the protein concentration, estimated from cAMP binding capacity, increased ~40-fold, and this value was consistent (N. Guiso, personal communication) with the value obtained by radioimmuno-assay [18]. Purity of both protein preparations of CRP was of the order of 99% as estimated from SDS–polyacrylamide gel electrophoresis.

α-CRP was prepared essentially as in [13] in the following way. For each particular batch of subtilisin (Sigma type VII n. P-5255), small-scale digestion assays were carried out in order to optimize the ratio of the protease to CRP. Under optimal conditions non-digested CRP was practically undetectable and a single polypeptide band (app. M_r 12 500 ± 500) without any visible contaminant was observed on a 15% SDS–polyacrylamide gel. Typically, a solution 0.4 mg CRP/ml in 0.1 M NaCl, 10 mM Tris–HCl, 1 mM EDTA (pH 8) containing 1 mM cAMP was digested with 8.5 μg subtilisin/ml for 15 min at

Published by Elsevier Biomedical Press

00145793/82/0000–0000/$2.75

37°C. The reaction was stopped with 1 mM phenylmethanesulfonyl fluoride (PMSF), and the mixture dialysed against 50 mM NaCl, 15 mM phosphate, 0.5 mM EDTA, 1 mM PMSF (pH 6.8). From the dialysed solution α-CRP was adsorbed on a column of Bio-Rex 70, washed with the same buffer (except for PMSF) and eluted by implementing the buffer with 1 M NaCl. A second chromatography step may be necessary to remove any trace of cAMP which can be ascertained from the value of the absorbance ratio A_{280}/A_{260}.

Proteins and peptides (~5 μg) were hydrolysed with a mixture of CF_3CO_2H/HCl (1:2) for 25 and 50 min at 166°C [19], which proved highly efficient for the hydrolysis of hydrophobic proteins. Amino acid analysis was conducted with Durrum D500 analyser having a 2.5 nmol amino acid full scale detection. The value of cystine were determined with performic acid oxidized proteins after hydrolysis [20] and those of tryptophan were analysed after micro-scale hydrolysis (unpublished) with 3 N methane sulfonic acid [21]. N-Terminal sequences were analysed with a Beckman 890C spinning cup sequanator [22] 3 mg polybrene (Pierce) was used as a carrier [23] in the spinning cup and washed 3 times by the normal protein program (0.1 M quadrol program from Beckman). The proteins, 30 nmol [^{14}C]carboxymethylated (CM-) [24] α-CRP, 3 nmol ^{14}CM-CRP from the super producing strain and 0.3 nmol CRP of the wild strain, were boiled in 1% SDS at 100°C for 5 min and subjected to double coupling in the first step. Dithioerythritol (Sigma) was added to the solvents [25]. After automatic conversion (P6 sequamat) to phenylthiohydantoin amino acids the analysis was made with high-pressure chromatography (Pye Unicam LC3) [22] and ^{14}CM-cysteine was identified by radioactivity.

C-Terminal sequences were analysed by carboxypeptidase (Cp) digestion followed by direct amino acid analysis [26]. The digestion buffers were 0.1 M pyridine–collidine acetate (pH 8.5) for CpA (Worthington Biochemicals, diisopropylfluorophosphate (DFP)-treated)) and for CpB (Sigma, PMFS-treated and re-treated with DFP), 0.1 M pyridine acetate (pH 5.5) for CpY (Pierce) and 0.1 M pyridine formate (pH 2.5) for CpP (Takara-Shuzo, Kyoto). CP (5 μg) was added to 1–2 nmol proteins and incubated for 6 h at 37°C.

3. RESULTS AND DISCUSSION

Table 1 shows the amino acid composition of CRP and of the subtilisin digestion product α-CRP. Both compositions of CRPs from the super-producing and the wild strains are essentially identical and comparable with [11,27]. Major differences of the compositions from those derived from DNA sequence [15,16] are lower values of ~2 mol isoleucine and 1 mol methionine. The N-terminal methionine has been processed (see below) and isoleucine may be due to insufficient hydrolysis. This was found also for α-CRP indicating the resistant isoleucyl peptide bonds to be in the α-CRP. Values of histidine and lysine agreed with those deduced from the DNA sequence, while these amino acids were found to have lower values in [11,28], possibly due to the poor hydrolysis. Also our compositions of CRPs indicate the absence of processing at the C-terminal 3 amino acids, glycine, threonine and arginine (see below) and the loss of the N-terminal methionine. The partial N- and C-terminal sequences of both CRP proteins and α-CRP are presented in fig.1. CRP purified from the super-producing strain and α-CRP prepared from it show redundancy at the N-termini, where 90% valine and ~7% methionine were found. This N-terminal redundancy as well as an acid-labile peptide bond Asp–Pro (position 9–10) limited the extent of analysis up to residue 37–41, as the average yield per step was ~97.5%. The N-terminal sequence we obtained here is completely coincident with the sequence deduced from the DNA sequence including the presence of a glutamic acid residue at position 13. Except for this position (reported as a glutamine) the sequence of the first 14 residues was known [11,27] for CRP from a conventional strain, KLF 41/JC 1553. With CRP purified from this strain the first step of Edman degradation revealed only traces of methionine (< 1.2%). The exact origin of the N-terminal redundancy observed with CRP from an over-producing strain cannot be ascertained from this sequence analysis. It may be caused by irregular posttranslational processing [29].

The C-terminal amino acid of CRP purified from the 2 strains, arginine, of heat denatured protein (100°C for 1 min) was liberated almost quantitatively by digestion with CpP. The other CpA and CpB digested the denatured protein resulting in the

Table 1

Amino acid composition of CRP and α-CRP

Hydrolysis (min):	CRP of the super-producing strain			CRP of the wild-type strain			CRP			αCRP of the super-producing strain			αCRP
	25	50	Integer	24	50	Integer	From DNA sequence	Previously reported		25	50	Integer	From DNA sequence
Asp	13.0	13.0	13	13.0	13.0	13	13	14	15	6.0	6.0	6	6
Thr	12.2	11.7	13	12.4	12.0	13	13	13	12	4.8	4.6	5	5
Ser	10.1	8.8	11	9.6	8.6	11	11	9	10	6.9	6.2	7–8	7
Glu	27.2	29.6	30	27.2	29.6	30	30	31	31	17.3	17.8	18	18
Pro	6.0	6.0	6	5.8	6.0	6	6	7	6	4.2	4.4	4	4
Gly	16.2	16.3	16	16.4	16.1	16	16	16	16	8.4	8.4	8	8
Ala	13.5	13.3	13	13.3	13.4	13	13	12	14	6.5	6.4	6–7	6
Val	11.5	13.7	14	11.1	14.1	14	14	14	13	6.2	7.2	7	7
Met	5.9	3.7	6	5.5	5.2	6	7	5	5	1.9	1.4	2	2
Ile	12.5	15.2	15	12.3	15.0	15	17	16	15	5.8	7.3	7	9
Leu	18.3	22.0	22	20.0	21.6	22	22	21	22	11.1	11.5	12	12
Tyr	5.8	5.7	6	5.7	5.4	6	6	5	5	4.7	4.2	5	5
Phe	4.7	4.9	5	4.6	4.5	5	5	5	5	3.6	3.5	4	4
His	5.4	5.7	6	5.0	5.8	6	6	5	5	2.8	3.5	4	4
Lys	14.2	14.6	15	14.3	14.7	15	15	13	14	8.7	9.5	9–10	10
Arg	10.8	10.6	11	11.3	11.0	11	11	10	9	4.4	4.5	4–5	4
Cys/2	3.0		3[a]	2.9		3	3	2	2	1.6		2[a]	2
Trp	1.7		2	1.9		2	2	2	2	1.8		2	2

[a] These values are also observed by titration with DTNB of the denatured proteins (unpublished)

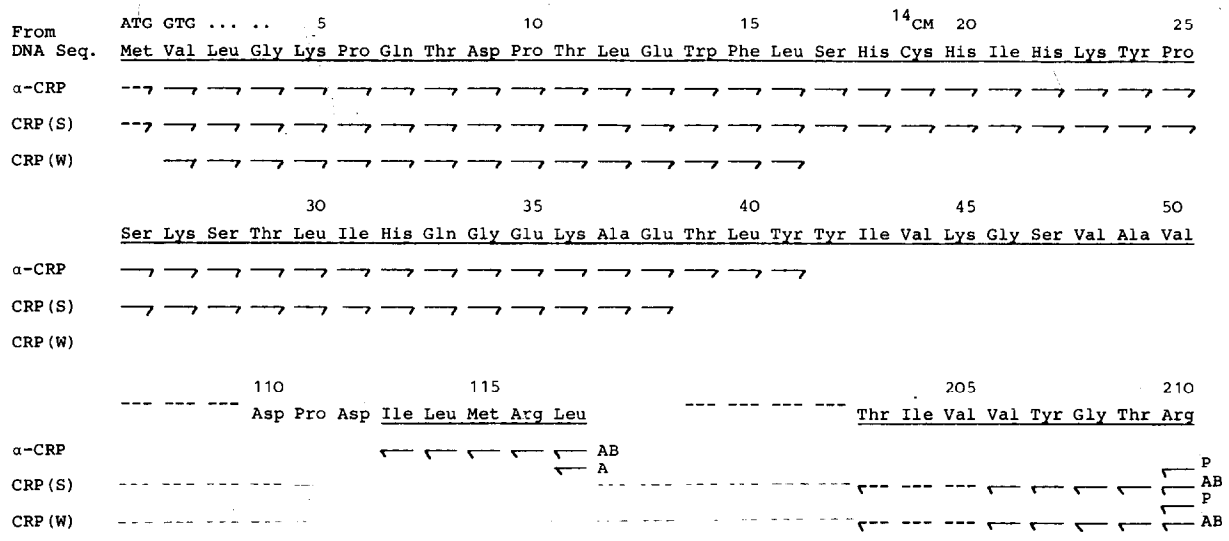

Fig.1. Partial sequence of CRP from the super producing strain (S) and the wild strain (W) and α-CRP.

C-terminal sequence as (Ile,Val)–Tyr–Glu–Thr–Arg with a recovery of 40% arginine (see fig.1). The results clearly indicate that the C-terminus of the CRP polypeptide chain is an arginine residue, and this is consistent with the DNA sequence (Arg 210). In contrast, CpY digestion resulted in the amino acids of the C-terminal region to be Ile, Leu, Val and Tyr [28]. Together with the DNA sequence data their results suggest that some kind of processing or cleavage of a C-terminal tripeptide may have occurred either in vivo or in the course of the extraction and purification procedures [15,16]. We have also digested CRP with CpY and observed liberation of similar amino acids but only in very small amounts ($< 5\%$). In the presence of pepstatine A (Sigma), which is known to inhibit contaminating proteases present in preparation of carboxypeptidase Y [30], the liberation of these amino acids was markedly reduced. Altogether, the results suggest that the hypothesis of a processing at the C-terminal of the CRP chain is rather unlikely.

We have found the N-terminal sequence of α-CRP identical with that of the native protein including N-terminus redundancy in the case of the preparation from the super-producing strain. This is at variance with [27] where as much as 65% of the α-CRP molecules had lost a N-terminal dipeptide. We show here that carefully adjusted conditions for subtilisin digestion, derived from the original conditions in [13], permits preparation of α-CRP with the N-terminus retaining intact. Kinetics of digestion of α-CRP with CpA and a mixture of A and B revealed that the C-terminal sequence of the core fragment is (Ile–Leu)–Met–Arg–Leu (fig.1). This sequence overlaps with positions 113–117 in the sequence of the native protein predicted from the DNA sequence. From the amino acid composition (table 1) and the N- and C-terminal sequences we thus conclude that α-CRP is the intact N-terminal fragment of the CRP polypeptide chain spanning the sequence from Val (2) to Leu (117). The C-terminal leucine of α-CRP is consistent with the known specificity of subtilisin. From the yield of C-terminal leucine, we can estimate that homogeneity of the α-CRP is $\geq 95\%$.

From a 2.9 Å resolution X-ray crystallographic analysis, a model of the structure of the cAMP · CRP complex has been published [31]. With this model and the available sequence [15,16] we observe that α-CRP lacks a large part (⅔–¾) of the long C α-helix of the polypeptide backbone along which contacts are established between the 2 chains in the dimer. However, we note that the stability of the dimeric state of α-CRP in solution is at least as good as that of native CRP (unpublished). Though not active on transcription in vitro (unpublished) this well-defined derivative of CRP is worthy of interest for structure analysis and it may help unravel important aspects of structure–function relationships in this regulatory protein. Particularly puzzling is the way by which structural information originating in the cAMP binding domain may be transferred to the C-terminal region of the molecule with a modulation of its DNA-binding properties.

ACKNOWLEDGEMENTS

The authors express their gratitude to Drs A. Ullmann and B. Gicquell-Sanzey who made available the over-producing strain used in this study. Excellent technical assistance of J.J. Scheffler and F. Viboirs (EMBL) in the protein sequence and amino acid analysis and M.M. Bardet (CNRS) in the preparation of CRP and α-CRP are gratefully acknowledged.

REFERENCES

[1] Zubay, G., Schwartz, D. and Beckwith, J. (1970) Proc. Natl. Acad. Sci. USA 66, 104–110.
[2] De Crombrugghe, B., Chen, B., Anderson, W., Nissley, P., Gottesman, M. and Pastan, I. (1971) Nature New Biol. 231, 139–142.
[3] Epstein, W., Rothman-Denes, L.B. and Hesse, J.(1975) Proc. Natl. Acad. Sci. USA 72, 2300–2304.
[4] De Crombrugghe, B. and Pastan, I. (1978) in: The Operon (Miller, J.H. and Reznikoff, W.S. eds) pp. 303–324, Cold Spring Harbour Laboratory, New York.
[5] Taniguchi, T., O'Neill, M. and De Crombrugghe, B. (1979) Proc. Natl. Acad. Sci. USA 76, 5090–5094.
[6] Ogden, S., Haggerty, D., Stoner, C.M., Kolodrubetz, D. and Schleif, R. (1980) Proc. Natl. Acad. Sci. USA 77, 3346–3350.
[7] Musso, R.E., Di Lauro, R., Adhya, S. and De Crombrugghe, B. (1977) Cell 12, 847–854.
[8] Mallick, U. and Herrlich, P. (1979) Proc. Natl. Acad. Sci. USA 76, 5520–5523.
[9] Movva, R.N., Green, P., Nakamura, K. and Inouye, M. (1981) FEBS Lett. 128, 186–190.

[10] Ullmann, A., Joseph, E. and Danchin, A. (1979) Proc. Natl. Acad. Sci. USA 76, 3194–3197.
[11] Anderson, W.B., Schneider, A.B., Emmer, M., Perlman, R.L. and Pastan, I. (1971) J. Biol. Chem. 246, 5929–5937.
[12] Krakow, J.S. and Pastan, I. (1979) Proc. Natl. Acad. Sci. USA 70, 2529–2533.
[13] Eilen, E., Pampeno, C. and Krakow, J.S. (1978) Biochemistry 17, 2469–2473.
[14] Takahashi, M., Gronenborn, A.M., Clore, G.M., Blazy, B. and Baudras, A. (1982) FEBS Lett. 139, 37–40.
[15] Cossart, P. and Gicquel-Sanzey, B. (1982) Nucleic Acids Res. 10, 1363–1378.
[16] Aiba, H., Fujimoto, S. and Ozaki, N. (1982) Nucleic Acids Res. 10, 1345–1362.
[17] Takahashi, M., Blazy, B. and Baudras, A. (1980) Biochemistry 19, 5124–5130.
[18] Guiso, N. and Blazy, B. (1980) Biochem. Biophys. Res. Commun. 94, 278–283.
[19] Tsugita, A., Scheffler, J.J. (1982) Proc. Japan Acad. 58B 1; Eur. J. Biochem. (1982) in press.
[20] Hirs, C.H.W. (1956) J. Biol. Chem. 219, 611–621.
[21] Penke, B., Ferenezi, R. and Kovacs, K. (1974) Anal. Biochem. 60, 45–50.
[22] Scheffler, J.J., Tsugita, A., Linden, G., Schweitz, H. and Lazdunski, M. (1982) Biochem. Biophys. Res. Commun. in press.
[23] Tarr, E.G., Beecher, J.F., Bell, M. and McKaac, D.J. (1978) Anal. Biochem. 84, 622–627.
[24] Wittman-Liebold, B. (1973) Hoppe Seyler's Z. Physiol. Chem. 354, 1415–1431.
[25] Creastfield, A.M., Moore, S. and Stein, W.H. (1963) J. Biol. Chem. 238, 622–627.
[26] Tsugita, A., Gregor, I., Kobota, I. and v.d. Broek, R. (1979) Cytochrome Oxidase (King, T.E. et al. eds) pp. 67–77, Elsevier Biomedical, Amsterdam, New York.
[27] Schlessinger, D.H. (1978) Fed. Proc. Fed. Am. Soc. Exp. Biol. 37, 1619.
[28] Aiba, H. and Krakow, J.S. (1981) Biochemistry 20, 4774–4780.
[29] Schürmann, P., Maeda, K. and Tsugita, A. (1981) Eur. J. Biochem. 116, 37–45.
[30] Lee, H. and Risrdam, J.F. (1978) Biochem. Biophys. Res. Commun. 85, 1135–1142.
[31] McKay, D.B. and Steitz, T.A. (1981) Nature 290, 744–749.