Biotecnología Aplicada, Vol. 8, No. 2, 174-181 (1991)

Artículos originales cortos

Characterization of the primary structure of recombinant human epidermal growth factor*

L. CASTELLANOS and **R. ESTRADA**

Department of Protein Chemistry, Vaccine Division, Center for Genetic Engineering and Biotechnology, P.O.Box. 6162, Havana, Cuba

Received in June 1990
Approved in January 1991

SUMMARY

Human Epidermal Growth Factor (EGF) expressed in *S. cerevisiae* was analyzed by reversed phase liquid chromatography. Two main components were isolated and studied by aminoacid analysis, NH_2- terminal sequence and peptide mapping. The lower retention time component corresponded to human EGF lacking two residues (Leu- Arg) at the C-terminus while the higher retention time component corresponded to human EGF lacking the C-terminal residue (Arg).

RESUMEN

El factor de crecimiento epidermico humano (EGF) recombinante, expresado en S. cerevisiae, fue analizado por cromatografía líquida de alta eficiencia en fase inversa. Se aislaron dos especies principales, que fueron estudiadas por análisis de aminoácidos, secuenciación NH_2- terminal y mapeo peptídico. La especie de menor tiempo de retención correspondió al EGF humano carente de los dos últimos residuos en su extremo C-terminal (Leu-Arg). La especie de mayor tiempo de retención corresponde al EGF humano carente del extremo C-terminal (Arg).

INTRODUCTION

An important number of polypeptides sharing as a common property their effect on cell growth have been described, the first one to be isolated and characterized was the Epidermal Growth Factor, EGF (Cohen, 1962). Human Epidermal Growth Factor is a polypeptidic hormone comprising 53 aminoacids and presenting three disulfide bridges (fig. 1). This paper reports the characterization of the primary structure of recombinant Human Epidermal Growth Factor cloned and expressed in our Center in *Saccharomyces cerevisiae*.

* Presented at the 7th International Conference on Methods in Protein Sequence Analysis, Berlin, 1989.
** To whom correspondence should be addressed.

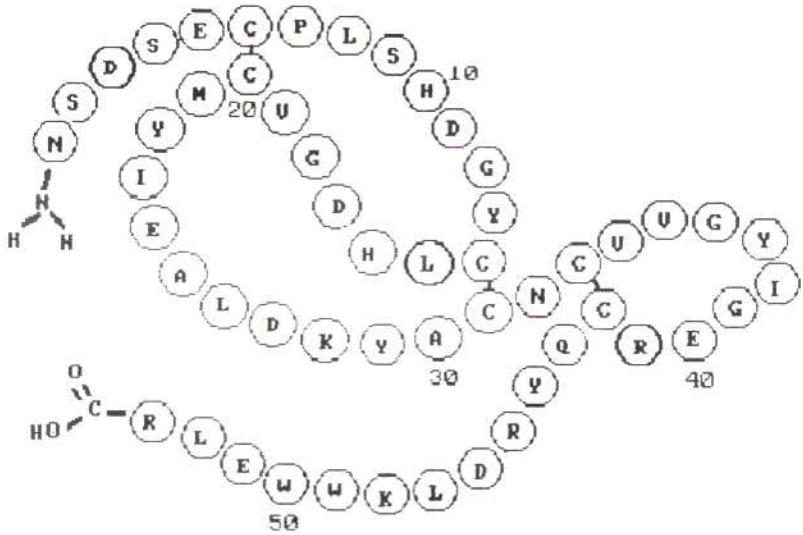

FIG. 1. Primary structure of Human Epidermal Growth Factor.

MATERIALS AND METHODS

Epidermal Growth Factor was cloned in *S. cerevisiae*, and produced by the Pilot Plant Division at the Center for Genetic Engineering and Biotechnology in Havana. A detailed report concerning its purification is being published (Cinza *et al.*, 1991). All reagents used were analytical grade from commercial suppliers.

Chromatography

Recombinant EGF was analyzed by reversed phase high performance liquid chromatography (RP-HPLC) on an LKB (Sweden) dual pump chromatograph model 2151, coupled with two UV detectors at 226 nm and 280 nm respectively (LKB model 2158 and Knauer Variable Wavelength Monitor,F.R.G.). A C-8 wide pore column, 250 mm x 4.6 mm (Baker, U.S.A.) was developed with a linear gradient (0% to 80% B in 40 minutes) of acetonitrile (mobile phase B) in water (mobile phase A) containing 0.1% trifluoroacetic acid at a flow of 1 ml per minute. Data were processed by a Shimadzu CR3A integrator.

Aminoacid analysis

Fractions isolated by RP-HPLC from three consecutive production batches were gas phase hydrolyzed for 24, 48 and 72 hours in a Ciba-Corning reactor, with 6N HCl containing 0.1% phenol and 0.1% beta-mercaptoethanol. After evaporation, free aminoacids were derivatized in 0.1 M sodium borate pH 10.0 with ortho-phthalaldehyde and 3-mercaptopropionic acid, and were analyzed on a reversed phase aminoacid analyzer (AminoSys, Pharmacia, Sweden). Data were processed by the Nelson PE 2600 Chromatography Data Software (Perkin Elmer, U.S.A.).

Reduction and carboxymethylation: Recombinant EGF was reduced for 4 hours (1% beta-mercaptoethanol, 60°C) and carboxymethylated (10 equivalents of sodium iodoacetate per equivalent of -SH group) at room temperature in the dark. The reaction was monitored upon completion by RP-HPLC. The solution was dialyzed against buffer (0.1 M Tris, 0.1 M NaCl, pH 7.0; membrane cut-off 1000 Da, Spectrapor, U.S.A) and then analyzed by RP-HPLC as described.

Peptide mapping

Fractions isolated by RP-HPLC were reduced as previously described, hydrolyzed for 30 hours with 2% aqueous acetic acid in vials sealed under vacuum and the peptide maps were analyzed on reversed phase chromatography. Alternatively, reduced fractions were cleaved at the methionine residue by treatment with a large excess of cyanogen bromide (300 equivalents per methionine residue) in 0.1 N hydrochloric acid during 24 hours, and then chromatographed as described.

Aminoacid sequencing

Automatic NH_2-terminal sequences were obtained on a Knauer 830 sequencer. Selected peptides were manually sequenced in the presence of polybrene by the Edman phenylisothiocyanate method as described by Tarr (Tarr, 1982) and phenylthio-hydantoine (PTH-) derivatives were identified by RP- HPLC (Castellanos, 1989).

RESULTS AND DISCUSSION

Preliminary analysis of recombinant EGF on reversed phase chromatography showed two main components (fig. 2). A similar profile was observed when comparing three different production batches. When chromatograms were obtained from previously reduced and carboxymethylated EGF samples, the two major components were observed again, with the same relative intensities, but at higher retention times (about two minutes later). Retardation in elution on HPLC columns for reduced and carboxymethylated proteins has also been observed for other proteins under study in our laboratory (data not shown). From this result it was concluded that the existence of two major components is independent from the presence of S-S bonds in the protein and then, their presence cannot be related to differences in the location of S-S bridges. The main components from three different production batches were separated by RP-HPLC (EGF-1: lower retention time, EGF-2: higher retention time) and their aminoacid composition was determined after 24 hours of acid hydrolysis (table 1). Aminoacid analysis indicated that all analyzed EGF-1 peaks were lacking one Leucine and one Arginine, while all EGF-2 peaks were lacking one Arginine. Acid hydrolysis during 48 hours and 72 hours confirmed the preceding

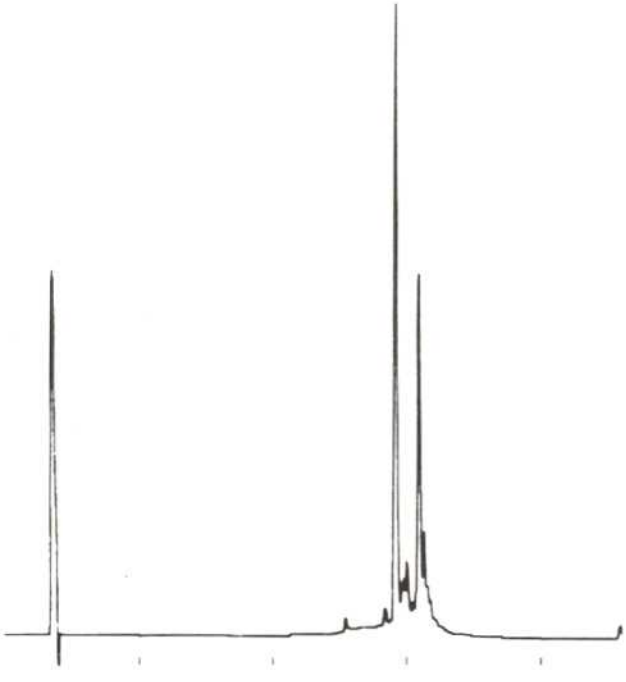

FIG. 2. Reversed phase chromatogram of recombinant EGF. (For experimental conditions, see text.)

results (table 2), recovery of Leucine and Arginine being time independent, so that their low recovery could not be explained by an hydrolytically resistant peptide bond.

results were identical and corresponded to the expected sequence. Cysteine was identified as its S-carboxymethyl-phenylthiohydantoine derivative on these

Table 1
AMINOACID COMPOSITION OF MAJOR COMPONENTS ISOLATED FROM THREE DIFFERENT PRODUCTION BATCHES OF RECOMBINANT EGF

Aminoacid		Batch-1	Batch-1	Batch-2	Batch-2	Batch-3	Batch-3
	E.V.	EGF-1	EGF-2	EGF-1	EGF-2	EGF-1	EGF-2
B	7	7.37	7.27	7.21	7.47	7.37	7.43
Z	5	5.22	5.54	5.14	5.32	5.13	4.96
G	4	4.26	4.19	4.16	4.22	4.01	4.24
A	2	2.47	2.08	2.29	2.07	2.23	2.18
H	2	1.91	1.68	2.33	2.34	2.45	2.43
K	2	2.01	2.21	1.91	1.77	1.76	2.12
R	3	2.01	2.19	2.11	2.01	2.05	2.12
T	0	0.09	0.08	0.11	0.06	0.08	0.13
S	3	2.71	2.72	2.57	2.82	2.63	2.74
Y	5	5.21	5.47	4.68	5.26	4.91	5.21
F	0	0.12	0.09	0.06	0.14	0.08	0.05
V	3	3.11	3.45	3.41	3.33	3.56	3.18
I	2	2.18	2.15	2.15	1.85	2.14	1.93
L	5	4.48	5.24	3.97	5.26	4.17	5.32

EGF-1: Lower retention time, *EGF-2*: Higher retention time, *E.V.*: Expected values, *B*: Aspartic acid plus asparagine, *Z*: Glutamic acid plus glutamine. Results were obtained after 24 hours hydrolysis. Proline, cysteine and tryptophane are not determined by this method. Proline and tryptophane were confirmed from sequence data, cysteine content was verified in an independent experiment (Besada *et al.*, 1991)

NH_2-terminal sequence was accomplished over the cyanogen bromide peptides, arising from the cleavage at residue 21 (methionine). Values obtained for the first 10 aminoacids of the NH_2-terminal (1-20) cyanogen bromide peptides are recorded on table 3. Further confirmation for homogenicity at the NH_2- terminus was evidenced by sequence analysis of the carboxymethyl derivatives of EGF-1 and EGF-2 for the first 23 residues, analyses (data not shown). Results obtained from automatic sequencing of the intact protein were coincident for EGF-1 and EGF-2. Identification of the sequence was possible up to residue 50. Table 4 presents sequence data for EGF-1; for EGF-2, data were essentially equivalent.

From the above mentioned results we suspected a deletion at the C-terminus originating two EGF species. C-terminus is

Table 2

RECOVERY OF LEUCINE AND ARGININE. COMPARISON OF AMINOACID ANALYSIS DATA AFTER 24, 48 AND 72 HOURS HYDROLYSIS. RECOVERY FOR THESE AMINOACIDS WAS TIME-INDEPENDENT

LEUCINE				
		Batch-1	Batch-2	Batch-3
EGF-1	24 hrs	4.48	3.97	4.17
EGF-1	48 hrs	4.28	4.11	4,19
EGF-1	72 hrs	4.08	4.13	4.13
EGF-2	24 hrs	5.24	5.26	5.32
EGF-2	48 hrs	5.19	5.56	5.46
EGF-2	72 hrs	5.03	4.93	5.55
ARGININE				
		Batch-1	Batch-2	Batch-3
EGF-1	24 hrs	2.01	2.11	2.05
EGF-1	48 hrs	2.08	2.13	2.11
EGF-1	72 hrs	2.01	2.21	2.05
EGF-2	24 hrs	2.19	2.01	2.12
EGF-2	48 hrs	2.11	2.24	2.43
EGF-2	72 hrs	2.04	2.06	2.14

Table 3

SEQUENCE DATA FOR THE FIRST 10 RESIDUES FROM THE NH2- TERMINAL CYANOGEN BROMIDE PEPTIDES OBTAINED FROM EGF-1 AND EGF-2. NH2-TERMINUS WERE HOMOGENEOUS AND CORRESPONDED TO THE EXPECTED SEQUENCE*

Cycle	A.A.	EGF-1	A.A.	EGF-2
1	N	68.5	N	57.2
2	S	36.8	S	26.5
3	D	50.6	D	45.7
4	S	28.6	S	18.8
5	E	55.6	E	37.3
6	(C)	N.D.	(C)	N.D.
7	P	19.4	P	28.2
8	L	23	L	18.6
9	S	9	S	6.7
10	H	N.Q.	H	N.Q.

* PTH-histidine appears as a wide peak and was not quantified (N.Q.: not quantified), PTH-cysteine is not observed under these conditions (N.D.: not determined.) recovery of pth-aminoacid for each edman cycle is reported in picomol.

Table 4
NH2-TERMINAL SEQUENCE DATA FOR EGF-1 (RESIDUES 1 TO 50)*

Cycle	A.A.	pmol	Cycle	A.A.	pmol
1	N	169	26	L	25.4
2	S	152	27	D	30.7
3	D	160	28	K	16.8
4	S	110	29	Y	26.2
5	E	139	30	A	33.7
6	(C)	N.D.	31	(C)	N.D.
7	P	46	32	N	14.4
8	L	101	33	(C)	N.D.
9	S	45	34	V	21.1
10	H	N.Q.	35	V	32.2
11	D	53	36	G	16.1
12	G	56	37	Y	24.5
13	Y	13	38	I	12.9
14	(C)	N.D.	39	G	23.6
15	L	45	40	E	17.3
16	H	N.Q.	41	R	4.7
17	D	41.2	42	(C)	N.D.
18	G	38.5	43	Q	5.9
19	V	27.6	44	Y	23.1
20	(C)	N.D.	45	R	N.Q.
21	M	24.7	46	D	6.8
22	Y	24.3	47	L	8.5
23	I	16.7	48	K	7.3
24	E	24.9	49	W	N.Q.
25	A	40.8	50	W	N.Q.

* For abbreviations, see footnote in table 3. PTH-Tryptophane at positions 49 and 50 was detected at the same retention time of diphenyl thiourea and is observed as an increase in its peak area; the quantitation being only approximative, its value is not included.

an Arginine preceded by a Leucine. As the DNA sequence of the cloned plasmid corresponded to the expected one (data not shown), such a modification could be originated by a carboxypeptidase-like activity acting on the expressed protein at some stage of the process, generating the C-terminal deletion of one or two residues.

In order to confirm the C-terminal sequence, EGF-1 and EGF-2 were digested with diluted acid (2% acetic acid) for 30 hours at 110°C. Under these conditions, selective cleavage at Aspartic acid is obtained, which was accompanied by some cleavage at Asparagine residues, probably preceded by their conversion to Aspartic acid. Each peptide was

identified by aminoacid analysis and by NH_2 -terminal analysis. A careful evaluation of the RP-HPLC maps evidenced a differentiating peptide in both species with a slight shift in their retention times. Figure 3 shows a co- chromatography of both peptide maps. These peptides were manually sequenced in the presence of polybrene and corresponded to the C-terminus as expected, giving the following sequences:

C-terminal peptide sequence

CYCLE	1	2	3	4	5	6	7
EGF-1	L	K	W	W	E	N.D.	N.D.
EGF-2	L	K	W	W	E	L	N.D.

Note: N.D., Not determined

Those results agree with aminoacid composition data, confirming that EGF-1 corresponds to (1-51) EGF and EGF-2 corresponds to (1-52) EGF. In no case the (1-53) EGF was identified. Both components EGF-1 and EGF-2 displayed an equivalent affinity for the EGF receptor (results not shown). An evaluation of the biological activity of recombinant EGF, containing the characterized components, has been reported (Cinza *et al.*, 1991) and proved to be equivalent to the natural EGF. During the execution of this study, Nascimento reported similar results for human EGF cloned also in *S. cerevisiae* (Nascimento *et al.*, 1988).

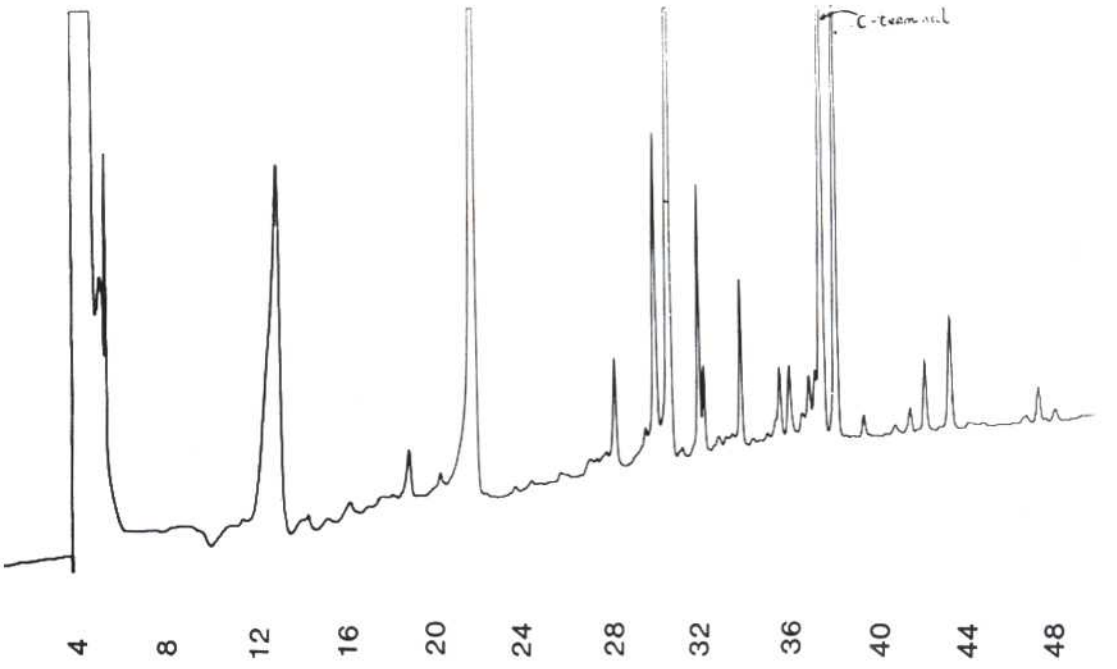

FIG. 3. Co-chromatography of the peptide maps obtained from acetic acid hydrolysis of EGF-1 and EGF-2. Two c-terminal peptides were obtained, showing a slight shift in their retention times. All other peptides were coincident for both EGF species and overlap on the chromatogram.

REFERENCES

BESADA, V.; W. ANTUCH; A. M. CINZA; I. ROJAS; M. QUINTANA; G. PADRON; T. TAKAO and Y. SHIMONISHI (1991). Chemical characterization of recombinant human Epidermal Growth Factor. *Analyt. Chem. Acta* **239**: 301-305.

CASTELLANOS, L. (1989). Separación de PTH-aminoácidos mediante cromatografía en fase inversa en columnas cortas. *Interferón y Biotecnología* **6**: 67-71.

CINZA, A. M.; M. QUINTANA, J. LOMBARDERO; R. POUTOU; E. PEREZ; L. C. PEREZ; C. M. MELLA; V. BESADA; G. PADRON; L. CASTELLANOS; R. ESTRADA and J. MORALES-GRILLO (1991). Establecimiento de un cultivo discontinuo para la producción del factor de crecimiento epidérmico humano en levadura. Caracterización del producto. *Biotecnología Aplicada* **8**: (this issue).

COHEN, S. (1982). Isolation of a mouse submaxillary gland protein accelerating incisor eruption and eyelid opening in the new-born animal. *J. Biol. Chem.* **237**: 1555-1562.

NASCIMENTO, C. G.; A. GYENE; S. M. HALLORAN; J. MERRYWEATHER; P. VALENZUELA; K. S. STEINER; F. R. MAISIARZ and A. RANDOLPH (1988). Characterization of Recombinant Human Epidermal Growth Factor produced in yeast. *Biochemistry* **27**: 797-802.

TARR, G. (1982). "Manual batchwise sequencing methods", In: *Methods in Protein Sequence Analysis*, Ed. by Marshal Elzinga, Humana Press, Clifton, N.J., U.S.A.