The primary structure of UK114 tumor antigen

Fabrizio Ceciliani[a], Ludovica Faotto[a], Armando Negri[a], Irma Colombo[b], Bruno Berra[b], Alberto Bartorelli[c], Severino Ronchi[a,*]

[a]*Istituto di Fisiologia Veterinaria e Biochimica, Università di Milano, Via Celoria 10, 20133 Milan, Italy*
[b]*Istituto di Fisiologia Generale e Chimica Biologica, Università degli Studi di Milano, via Trentacoste 2, 20134 Milan, Italy*
[c]*Istituto Ricerche 'G. Sisini', Università degli Studi di Milano, via Parea 4, 20138 Milan, Italy*

Received 31 May 1996

Abstract UK114 is a tumor antigen expressed by various malignant neoplasms. The complete amino acid sequence of UK114 purified from goat liver has been determined by automated Edman degradation of CNBr and endoproteinase Lys-C peptides. The protein contains 137 amino acid residues, which corresponds to a molecular mass of 14229 Da. MALDI-TOF analysis resulted in a molecular weight of 14290, suggesting that the N-terminal Met residue is acetylated. Sequence comparison shows that UK114 from goat liver (1) has 77% identity with a previously described 23 kDa protein from rat liver (Levy-Favatier et al. (1993) *Eur. J. Biochem.* 212, 665–673), (2) shares a very high degree of similarity with a family of prokaryotic and eukaryotic hypothetic proteins whose function have not yet been characterized, and (3) exhibits a significant similarity to a group of tumor-associated antigens which belongs to a superfamily of heat shock proteins, acting as possible targets for the host's antitumor immunity.

Key words: UK114; Tumor antigen; Amino acid sequence

1. Introduction

The presence and activity in neoplastic cells of different tumor-associated antigens has been widely demonstrated (for a review see [2]): nevertheless their molecular nature remains in many cases largely unknown. Bartorelli et al. [3,4] have recently described a new group of tumor-associated antigens, extracted from goat liver with perchloric acid and indicated as UK101. Policlonal antisera and monoclonal antibodies anti-UK101 react against 90% of human tumors tested, confirming the hypothesis that antigens very similar to the proteins contained in UK101 are expressed in neoplastic tissues. Moreover, UK101 induces an in vivo antineoplastic effect, reducing in several cases the metastatic masses [4,5].

UK 101 contains three major proteins [6]: (1) a 10 kDa protein (UK110), (2) a 14 kDa protein (UK114), and (3) a 50 kDa protein (UK150). The larger component, UK150, was shown to be a mannose-rich glycoprotein and did not exhibit any of the potential immunological and antineoplastic properties of UK101 fractions [6,7]. UK110 consists of a mixture of integer ubiquitin and its degradation products, derived from partial removal of C-terminal residues. However, ubiquitin alone does not produce any therapeutic effect. Thus we focused our attention on the 14 kDa protein. Bartorelli et al., using immunocytochemical studies [8], have demonstrated the presence of UK114 in various, adenocarcinoma type, human malignant neoplasms. The same authors located UK114 on the cell membrane of several human cancer cell lines. Moreover, they showed that anti-UK114 antibodies induce cytolysis in vitro and cause a marked delay of tumor growth in vivo. These experiments suggest that human tumors expressing UK114 on the surface of the cell membrane may undergo antibody-mediated cytolysis. Thus, UK114 is thought to be the main significant component of UK101 with potential pharmacological activity. Preliminary structural results suggested that UK114 might be a protein not yet characterized [6]. In this paper we present the complete primary structure of UK114 as a preliminary step to investigation of its biological activity.

2. Material and methods

2.1. Materials

Endoproteinase Lys-C was from Boehringer Mannheim. TFA, 6 N HCl and CNBr were from Pierce Chemical Company. Immobilon-P membrane and AQC were from Millipore. All other reagents used were of the highest purity commercially available.

2.2. Purification and structural characterization of UK114

Purification of UK114 from goat liver was performed as previously described [7]. The primary structure of the protein was determined after (1) in situ CNBr fragmentation of the native protein blotted on PVDF membrane as described in [9] and (2) CNBr fragmentation of UK114 in 70% formic acid [10] followed by reduction and alkylation of CNBr peptides with iodoacetamide and endoproteinase Lys-C digestion [11]. Peptides were purified by RP-HPLC as described by [10]. Amino acid sequence was determined using an Applied Biosystems Mod. 477/A pulse liquid-phase sequencer (Table 1, Fig. 3).

2.3. Miscellaneous methods

Amino acid analysis was performed after gas-phase hydrolysis by pre-column derivatization with AQC [12], as previously described [13].

Molecular weight was determined either by SDS-PAGE or by MALDI-TOF mass spectrometry using a focus/LD-TOF spectrometer (VESTEC Instruments) and a saturated solution of sinapinic acid in 0.1% TFA/acetonitrile (2 : 1, v/v) as matrix.

3. Results and discussion

Approximately 1–5 nmol of native and carbamidomethylated protein was subjected to automated Edman degradation but no phenylthiohydantoin derivative could be identified, indicating that the amino-terminal residue of the protein was blocked. Thus the amino acid sequence of UK114 had to be determined after characterization and sequence alignment of the peptides obtained by CNBr fragmentation alone and CNBr fragmentation followed by endoproteinase Lys-C

*Corresponding author. Fax: (39) 2-2666301.

Abbreviations: CNBr, cyanogen bromide; MALDI-TOF, matrix-assisted laser desorption/ionization time of flight; TFA, trifluoroacetic acid; PVDF, polyvinylidene fluoride; AQC, 6-aminoquinolyl-*N*-hydroxysuccinimidyl carbamate

PII S0014-5793(96)00850-2

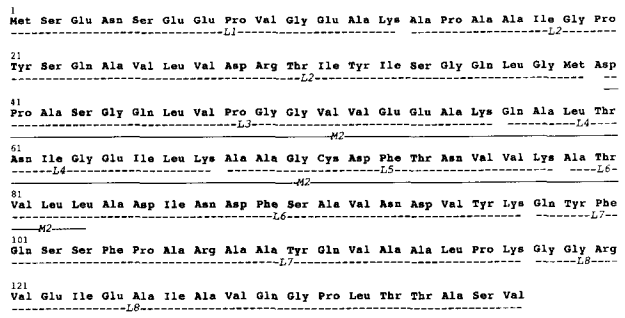

Fig. 1. The primary structure of UK114. The continuous line indicates the sequence derived from CNBr fragmentation of the protein. The dashed line indicates the sequence derived from the endoproteinase Lys-C digestion.

digestion. Fig. 1 summarizes the data used to determine the complete primary structure of UK114. The protein consists of 137 amino acid residues, and the calculated molecular mass of 14229 Da agrees with the value of 14290 determined by MALDI-TOF analysis. The 60 Da difference between theoretically and experimentally determined masses could be explained by the amino-acetylation of N-terminal residue of methionine. The amino acid composition deduced from the sequence agrees well with that determined experimentally (data not shown). It should be pointed out that in situ CNBr fragmentation failed to cleave the peptide bond between Met^{1} and Ser^{2}, allowing the unambiguous determination of the sequence from Asp^{40} to Leu^{83}.

The alignments between Gly^{38} and Met^{39}, Lys^{97} and Gln^{98}, Lys^{117} and Gly^{118} were deduced from the similarity between UK114 and the closely related proteins shown in Fig. 2.

A BLAST [14] search of UK114 with other proteins in EMBL/Genebank and SwissProt databases revealed a very high degree of identity (77%) between UK114 and a rat perchloric acid-soluble 23 kDa protein [1]. The latter was originally described as a dimer of two identical 99 amino acid polypeptides (starting from Met^{39} to Val^{137}). However, Oika (D49363, GeneBank) showed that the published sequence of the 23 kDa protein is partially incorrect, since insertion of a guanosine base at position 28 of the cDNA would results in a new open reading frame which codes for a 137 amino acid polypeptide.

Furthermore, UK114 exhibits a very high degree of identity with a group of proteins belonging to the so-called YER057C/YIL051C/Y5GF family. Fig. 2A shows the alignment between UK114 and the sequence of the two most similar proteins belonging to this family, the YABJ from *Bacillus subtilis* and the Y5GF from *E. coli* (58% and 52% identity, respectively).

The functions of the proteins belonging to this family have not yet been characterized, but it is interesting to note that the sequence of UK114 is very well conserved during evolution from prokaryotic to mammalian.

Comparison of the amino acid sequence of UK114 with protein sequences in the NCBI Entrez database (Beauty algorithm [15]), revealed also a high degree of similarity between portions of UK114 and of two closely related MHC-binding proteins, MBP-1 and MBP-2, which belong to a family of transcription factors involved in the regulation of MHC class I gene expression [16] (Fig. 2B). However, such portions in MBP-1 and MBP-2 do not belong to those regions which are suggested to be related to their functions.

Finally, the region from residue 46 to residue 100 exhibits a significant (35%) similarity with a highly conserved region of human hsp90, as already reported by Levy-Favatier et al. [1] for the 23 kDa protein (Fig. 2C). Heat shock proteins (hsp) play several essential roles within the cell. In particular hsp90 and hsp70 can elicit a tumor-specific immunity [17,18]. Hsp peptides complexed with MHC class I molecules may be presented to T cells [19]; hsp themselves may act as a presenting

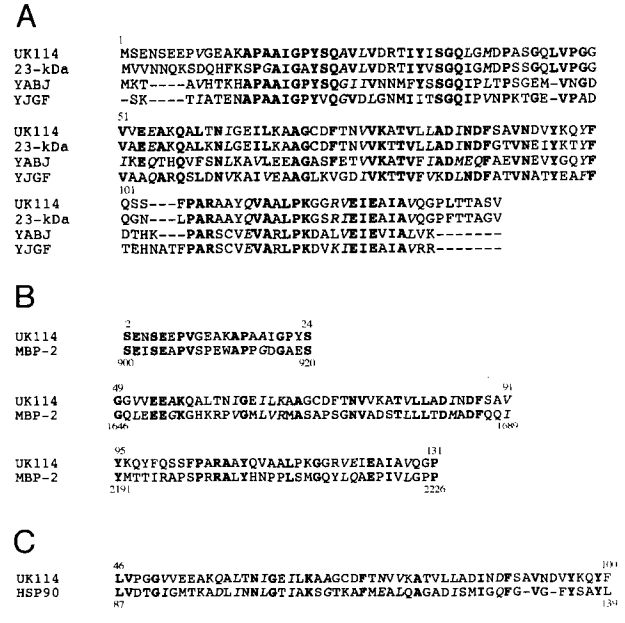

Fig. 2. Sequence alignment between the primary structure of UK114 and the sequence of other closely related proteins, the YER057C/YIL051C/Y5GF family and 23 kDa (A), the MHC-binding protein MBP-2 (B) and hsp90 (C). The identities are indicated in bold. Conservative substitutions are indicated in italics.

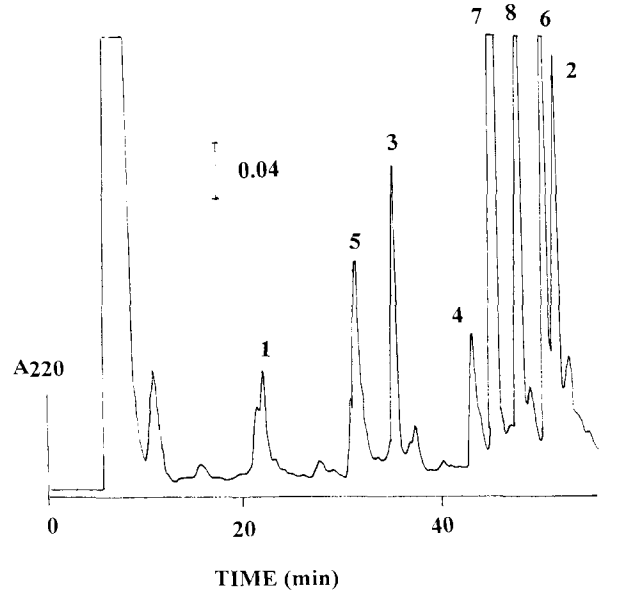

Fig. 3. HPLC chromatography of UK114 following CNBr fragmentation and Lys-C digestion. Peak numbers refer to the alignment shown in the test.

Table 1
Amino acid sequence of peptides obtained from in situ CNBr fragmentation of UK114 (C2) or CNBr fragmentation followed by Lys-C digestion (L1, etc.) and HPLC chromatography (see Fig. 3)

C2			C2 (cont.)			L1			L2			L3		
		(pmol)			(pmol)			(pmol)			(pmol)			(pmol)
1	Asp	879	27	Leu	349	1	Ser	52	1	Ala	258	1	Asp	89
2	Pro	746	28	Lys	108	2	Glu	24	2	Pro	226	2	Pro	142
3	Ala	1112	29	Ala	439	3	Asn	15	3	Ala	177	3	Ala	207
4	Ser	152	30	Ala	380	4	Ser	16	4	Ala	182	4	Ser	52
5	Gly	692	31	Gly	159	5	Glu	17	5	Ile	160	5	Gly	292
6	Gln	753	32	(Cys)	n.d.	6	Glu	22	6	Gly	166	6	Gln	106
7	Leu	252	33	Asp	63	7	Pro	39	7	Pro	153	7	Leu	161
8	Val	486	34	Phe	219	8	Val	34	8	Tyr	98	8	Val	200
9	Pro	380	35	Thr	58	9	Gly	46	9	Ser	45	9	Pro	105
10	Gly	527	36	Asn	78	10	Glu	22	10	Gln	68	10	Gly	238
11	Gly	524	37	Val	284	11	Ala	19	11	Ala	90	11	Gly	192
12	Val	556	38	Val	319	12	Lys	6	12	Val	114	12	Val	162
13	Val	721	39	Lys	59				13	Leu	65	13	Val	216
14	Glu	214	40	Ala	398				14	Val	130	14	Glu	30
15	Glu	256	41	Thr	29				15	Asp	35	15	Glu	42
16	Ala	590	42	Val	210				16	Arg	15	16	Ala	90
17	Lys	112	43	Leu	194				17	Thr	32	17	Lys	11
18	Gln	361	44	Leu	228				18	Ile	92			
19	Ala	520							19	Tyr	40			
20	Leu	510							20	Ile	70			
21	Thr	91							21	Ser	30			
22	Asn	120							22	Gly	56			
23	Ile	298							23	Gln	28			
24	Gly	326							24	Leu	26			
25	Glu	72							25	Gly	26			
26	Ile	284							26	(Met)	n.d.			

L4			L5			L6			L7			L8		
		(pmol)			(pmol)			(pmol)			(pmol)			(pmol)
1	Gln	1161	1	Ala	302	1	Ala	297	1	Gln	178	1	Gly	424
2	Ala	1050	2	Ala	361	2	Thr	79	2	Tyr	218	2	Gly	372
3	Leu	1022	3	Gly	427	3	Val	260	3	Phe	251	3	Arg	100
4	Thr	463	4	CAM-Cys	2	4	Leu	165	4	Gln	154	4	Val	296
5	Asn	520	5	Asp	144	5	Leu	185	5	Ser	43	5	Glu	174
6	Ile	672	6	Phe	225	6	Ala	155	6	Ser	43	6	Ile	231
7	Gly	613	7	Thr	108	7	Asp	91	7	Phe	156	7	Glu	159
8	Glu	427	8	Asn	170	8	Ile	184	8	Pro	69	8	Ala	176
9	Ile	509	9	Val	166	9	Asn	37	9	Ala	172	9	Ile	143
10	Leu	348	10	Val	234	10	Asp	51	10	Arg	34	10	Ala	109
11	Lys	36	11	Lys	5	11	Phe	114	11	Ala	90	11	Val	150
						12	Ser	25	12	Ala	100	12	Gln	171
						13	Ala	72	13	Tyr	58	13	Gly	132
						14	Val	163	14	Gln	26	14	Pro	36
						15	Asn	27	15	Val	94	15	Leu	63
						16	Asp	19	16	Ala	76	16	Thr	21
						17	Val	114	17	Ala	100	17	Thr	58
						18	Tyr	32	18	Leu	68	18	Ala	76
						19	Lys	4	19	Pro	17	19	Ser	27
									20	Lys	5	20	Val	11

Numbers following residues report the amount in pmol detected in each cycle. n.d., not detected; CAM-Cys, carbamidomethyl cysteine.

molecule, complexed with endogenously derived cellular peptides [20]. Furthermore, it has been demonstrated [21,22] that hsp are maintained at low levels in normal cells and are confined inside the cytoplasm; the stress induces an unusual expression of these proteins and a targeting both in cell compartments and on the cell surface.

On the basis of sequence similarities and immunocytochemical studies we can speculate about the putative role of UK114; as previously described for hsp, UK114 is expressed on the surface of the cell as an effect of stress induced by the neoplastic transformation and recognition of UK114 by anti-UK114 antibodies induces an antibody-mediated cytolysis of the tumoral cell. This proposed mechanism could be consistent with the immunological properties and localization of UK114 described in [8]. Although the regulation of UK114 expression and correlation with the neoplastic transformation of the cell have not been investigated in detail, both the primary structure of UK114 and expression of this protein as a cancer-dependent phenomenon confirms the hypothesis that UK114 belongs to a group of tumor-associated antigens acting as a possible target for the host's antitumor immunity. Nevertheless, it is not clear if the antigenity of UK114 derives from the protein per se or, alternatively, from associated peptides, this hypothesis suggesting a role as molecular chaperon of antigenic peptides, as described for hsp90 [17,18].

Acknowledgements: We thank Prof. Alessandra Albertini from the University of Pavia for the BLAST search of UK114 in EMBL/

Gene Bank and SwissProt databases.The protein sequence reported here has been submitted to the EMBL Data Bank with Accession no. P80601.

References

[1] Levy-Favatier, L., Cuisset, L., Nedelec, B., Tichonicky, L., Kruh, J. and Delpech, M. (1993) Eur. J. Biochem. 212, 665–673.
[2] Hakomori, S. (1989) Adv. Cancer Res. 52, 257–331.
[3] Bartorelli, A., Berra, B., Ronchi, S., Biancardi, C., Cavalca, V., Bailo, M., Mor, C., Ferrara, R., Botta, M., Arzani, C. and Clemente, C. (1994) J. Tumor Marker Oncol. 9/3, 37–46.
[4] Bartorelli, A., Zanolo, G., Fassio, F., Botta, M., Ferrara, R., Bailo, M., Biancardi, C., Cavalca, V., Arzani, C. and Maraschin, R. (1994) J. Tumor Marker Oncol. 9/3, 49–56.
[5] Bartorelli, A., Berra, B., Clemente, C., Liverani, A., Santi, C., Ronchi, S., Botturi, M., Corchia, P.L., Orefice, S. and Mor, C. (1994) J. Tumor Marker Oncol. 9/3, 57–64.
[6] Ceciliani, F., Biancardi, C., Cavalca, V., Ferrara, R., Botta, M., Arzani, C., Bailo, M., Berra, B., Ronchi, S. and Bartorelli, A. (1996) J. Tumor Marker Oncol., in press.
[7] Bartorelli, A., Biancardi, C., Cavalca, V., Ferrara, R., Botta, M., Arzani, C., Colombo, I., Berra, B., Ceciliani, F., Ronchi, S. and Bailo, M. (1996) J. Tumor Marker Oncol., in press.
[8] Bartorelli, A., Bussolati, B., Millesimo, M., Gugliotta, P. and Bussolati, G. (1966) Int. J. Oncol. 8, 543–548.
[9] Yuen, W.S., Chui, A.H., Wilson, K.J. and Yuan, P.M. (1988) Appl. Biosyst. User Bull. 36.
[10] Negri, A., Ceciliani, F., Tedeschi, G., Simonic, T. and Ronchi, S. (1992) J. Biol. Chem. 267/17, 11865–11871.
[11] Stone, K.L. and Williams, K.R. in (P. Matsudaira ed.), A Practical Guide to Protein and Peptide Purification for Microsequencing, pp. 54–56, 1993, Academic Press.
[12] Cohen, S. and Michaud, D.P. (1993) Anal. Biochem. 211, 279–287.
[13] Ceciliani, F., Bortolotti, F., Menegatti, E., Ronchi, S., Ascenzi, P. and Palmieri, S. (1994) FEBS Lett. 342, 221–224.
[14] Altschul, S.F., Gish, W., Myers, E.W. and Lipman, D.J. (1990) J. Mol. Biol. 215, 403–410.
[15] Worley, K.C., Wiese, B.A. and Smith, R.F. (1995) Genome Res. 5, 173–184.
[16] Van 't Veer, L.J., Lutz, P.M., Isselbacher, K.J and Bernards, R. (1992) Proc. Natl. Acad. Sci. USA 89, 8971–8975.
[17] Udono, H. and Srivastava, P.K. (1993) J. Exp. Med. 178, 1391–1396.
[18] Udono, H. and Srivastava, P.K. (1993) J. Cell. Biochem. 17D, 113.
[19] Jarditzky, T.S., Lane, R.A., Robinson, R.A., Madden, D.R. and Wiley, D.C. (1991) Nature 353, 326.
[20] Tamura, Y., Tsuboi, N., Sato, N. and Kikuchi, K. (1993) J. Immunol. 151, 5516–5524.
[21] Ferrarini, M., Heltai, S., Zocchi, M.R. and Rugarli, C. (1992) Int. J. Cancer 51, 613–619.
[22] Chouchane, L., Bowers, F.S., Sawasdikosol, S., Simpson, R.M. and Kindt, M.J. (1994) J. Infect. Dis. 169, 253–259.