

## Nucleotide sequence of cDNA coding for dianthin 30, a ribosome inactivating protein from *Dianthus caryophyllus*

Giuseppe Legname<sup>1,2</sup>, Paola Bellosta<sup>2</sup>, Gianni Gromo<sup>2</sup>, Daniela Modena<sup>2</sup>,  
Jeff N. Keen<sup>3</sup>, Lynne M. Roberts<sup>1</sup> and J. Michael Lord<sup>1</sup>

<sup>1</sup> Department of Biological Sciences, University of Warwick, Coventry (U.K.),

<sup>2</sup> Centro Ricerche Italfarmaco SpA, Milan (Italy) and

<sup>3</sup> Department of Biochemistry and Molecular Biology, University of Leeds, Leeds (U.K.)

(Received 4 June 1991)

Key words: Dianthin 30; Ribosome inactivating protein; Saporin 6

**Rabbit antibodies raised against dianthin 30, a ribosome inactivating protein from carnation (*Dianthus caryophyllus*) leaves, were used to identify a full length dianthin precursor cDNA clone from a  $\lambda$ gt11 expression library. N-terminal amino acid sequencing of purified dianthin 30 and dianthin 32 confirmed that the clone encoded dianthin 30. The cDNA was 1153 basepairs in length and encoded a precursor protein of 293 amino acid residues. The first 23 N-terminal amino acids of the precursor represented the signal sequence. The protein contained a carboxy-terminal region which, by analogy with barley lectin, may contain a vacuolar targeting signal.**

Many plant tissues contain proteins that specifically inhibit protein synthesis by susceptible ribosomes [1]. These ribosome inactivating proteins (RIPs) are a highly specific RNA *N*-glycosidase that catalytically removes a single adenine residue from a highly conserved loop structure present in 23S, 26S or 28S ribosomal RNA [2]. Generally, RIPs occur as monomeric proteins of molecular mass around 30 kDa which may be *N*-glycosylated [1]. In some cases the RIP (in this case called the A chain) is part of a heterodimeric protein in which it is joined to a galactose-binding lectin (the B chain) by a single disulphide bond. Ricin and abrin are examples of such heterodimers and are amongst the most cytotoxic compounds known [3].

Recently, there has been considerable interest in single chain RIPs or the A chains of cytotoxic lectins. This interest springs from their antiviral properties [4,5] and because of their widespread use as the toxic moiety of cell-type specific cytotoxic conjugates formed by chemically or genetically linking the RNA *N*-glycosidase to a range of antibodies, lymphokines or growth factors [6].

Carnation (*Dianthus caryophyllus*) leaves contain two single chain RIPs, dianthin 30 and 32. These proteins

were purified as described previously [7] except that a final step involving gel filtration HPLC was included. Polyclonal antibodies were raised in rabbits against dianthin 30. The antibodies were used to screen a cDNA library produced from carnation leaf mRNA in the expression vector  $\lambda$ gt11. Several positive clones were identified and Fig. 1 shows a restriction map and sequencing strategy for one of the clones. This clone is 1153 basepairs in length and contains a single open reading frame encoding a protein of 293 amino acids (Fig. 2). The first 23 amino acids represent the N-terminal signal sequence. The protein encoded by the clone was positively identified as dianthin 30. Purified dianthin 30 and dianthin 32, electro eluted from an SDS polyacrylamide gel, were subjected to N-terminal amino acid sequencing. The first 31 residues were

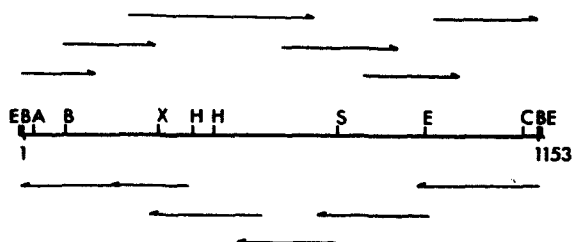


Fig. 1. Restriction endonuclease map of a dianthin precursor cDNA clone and strategy of sequencing. The abbreviations used are, E, *EcoRI*; B, *BamHI*; A, *AseI*; X, *XhoI*; H, *HpaI*; S, *SspI*; C, *ClaI*.

```

          30                                60
CAA AAA TCA ATA TAA TTT TTC ACG CAT ATT AAT TCC AAC GTT CAA CCA GCT TCA CTC GAA

          90                                120
ATG AAG ATT TAT TTA GTG GCC GCG ATA GCA TGG ATC CTG TTT CAG TCT TCA TCT TGG ACA
Met Lys Ile Tyr Leu Val Ala Ala Ile Ala Trp Ile Leu Phe Gln Ser Ser Ser Trp Thr
-23

          150                                180
ACT GAT GCG GCC ACA GCA TAC ACA TTA AAT CTC GCA AAT CCA TCC GCG AGT CAA TAC TCA
Thr Asp Ala Ala Thr Ala Tyr Thr Leu Asn Leu Ala Asn Pro Ser Ala Ser Gln Tyr Ser
-1      1

          210                                240
TCT TTT CTG GAT CAA ATC CGA AAC AAT GTG AGG GAT ACC AGC CTC ATA TAC GGT GGG ACA
Ser Phe Leu Asp Gln Ile Arg Asn Asn Val Arg Asp Thr Ser Leu Ile Tyr Gly Gly Thr
20

          270                                300
GAC GTA GCC GTG ATT GGT GCG CCT TCT ACT ACT GAT AAA TTC CTT AGA CTT AAT TTC CAA
Asp Val Ala Val Ile Gly Ala Pro Ser Thr Thr Asp Lys Phe Leu Arg Leu Asn Phe Gln
40

          330                                360
GGT CCT CGA GGA ACG GTC TCT CTT GGC CTT AGG CGC GAG AAC TTA TAC GTG GTC GCG TAT
Gly Pro Arg Gly Thr Val Ser Leu Gly Leu Arg Arg Glu Asn Leu Tyr Val Val Ala Tyr
60

          390                                420
CTT GCA ATG GAT AAC GCA AAT GTT AAC CGT GCA TAT TAC TTC AAA AAC CAA ATC ACT TCT
Leu Ala Met Asp Asn Ala Asn Val Asn Arg Ala Tyr Tyr Phe Lys Asn Gln Ile Thr Ser
80

          450                                480
GCT GAG TTA ACC GCC CTT TTC CCC GAG GTT GTG GTT GCA AAT CAA AAA CAA TTA GAG TAC
Ala Glu Leu Thr Ala Leu Phe Pro Glu Val Val Val Ala Asn Gln Lys Gln Leu Glu Tyr
100

          510                                540
GGG GAA GAT TAC CAG GCG ATA GAA AAG AAC GCC AAG ATA ACA ACA GGC GAT CAA AGT AGA
Gly Glu Asp Tyr Gln Ala Ile Glu Lys Asn Ala Lys Ile Thr Thr Gly Asp Gln Ser Arg
120

          570                                600
AAG GAA CTC GGT TTG GGG ATC AAT CTA CTT ATA ACG ATG ATT GAT GGA GTG AAT AAG AAG
Lys Glu Leu Gly Leu Gly Ile Asn Leu Leu Ile Thr Met Ile Asp Gly Val Asn Lys Lys
140

          630                                660
GTA CGT GTA GTC AAA GAC GAG GCA AGG TTT TTG TTA ATC GCA ATT CAA ATG ACG GCT GAG
Val Arg Val Val Lys Asp Glu Ala Arg Phe Leu Leu Ile Ala Ile Gln Met Thr Ala Glu
160

          690                                720
GCC GCG CGA TTT AGG TAC ATA CAG AAC TTG GTT ACC AAG AAC TTC CCA AAC AAG TTC GAC
Ala Ala Arg Phe Arg Tyr Ile Gln Asn Leu Val Thr Lys Asn Phe Pro Asn Lys Phe Asp
180

          750                                780
TCA GAA AAT AAG GTT ATT CAA TTT CAA GTT AGT TGG AGT AAG ATT TCT ACG GCA ATA TTT
Ser Glu Asn Lys Val Ile Gln Phe Gln Val Ser Trp Ser Lys Ile Ser Thr Ala Ile Phe
200

          810                                840
GGG GAT TGC AAA AAC GGC GTG TTT AAT AAA GAT TAT GAT TTC GGG TTT GGG AAA GTG AGG
Gly Asp Cys Lys Asn Gly Val Phe Asn Lys Asp Tyr Asp Phe Gly Phe Gly Lys Val Arg
220

          870                                900
CAG GCA AAA GAC CTT CAA ATG GGG CTC CTT AAG TAT TTA GGT AGA CCG AAG TCG TCG TCA
Gln Ala Lys Asp Leu Gln Met Gly Leu Lys Tyr Leu Gly Arg Pro Lys Ser Ser Ser
240

          930                                960
ATC GAG GCG AAT TCC ACT GAC GAC ACA GCT GAT GTG CTT TGA AGC TCA CTT CAT ACT CAA
Ile Glu Ala Asn Ser Thr Asp Asp Thr Ala Asp Val Leu End
260

          990                                1020
ATG ATC ATA TGA GTT GAT TAG CCC AAC ATA TAT ATG CAA CTT CGG ACG ATA TAT ATT TAT

          1050                                1080
ATT GTA TCC ACG TTT GTT CAT GTT TCC GAT AAC GGT GTC GTT TGT GTT TGG ACG AGT AAT

          1110                                1140
GAA TAA AGA TTG TAC TCC GTA TAT ATA ATA CGG AGT ATA TGT TAT TAT CGA TCG TAA AAA
-----
AAA AAA AAA AAA A

```

Fig. 2. Nucleotide sequence and deduced amino acid sequence of dianthin precursor cDNA. The first 23 amino acid residues represent the N-terminal signal sequence. The mature dianthin 30 sequence starts at *Ala1*. A potential polyadenylation signal is underlined.

	1					60
DIAN.	.....MKIY	LVAIAWILF	QSSW.TTDA	ATAYTLNLAN	PSASQYSSFL	
SAPORIN.	.....MKIY	VVATIAWILL	QFSAWTTDA	VTSITLIDLVN	PTACQYSSFV	
TRICH.	.....M	IRFLVLSLLI	LTLFLTPAV	EGDVSFRLSG	ATSSSYGVFI	
RICIN.	MKPGGNTIVI	WMYAVATWLC	FGSTSGWSFT	LEDNNIFPKQ	YPIINFTAG	ATVQSYTNFI
Cons	-----	-----	-----	-----	-----	-----Y--F-
	61					120
DIAN.	DQIRNNVRDT	SLIYGGTDV.	.AVIGAPSTT	DKFLRLNFGQ	PRGTVSLGLR	RENLYVVAYL
SAPORIN.	DKIRNNVKDP	NLKYGGTDI.	.AVIGPPS.K	EKFLRINFQS	SRGTVSLGLK	RDNLVYVAYL
TRICH.	SNLRKALPNE	RKLYDIPLLR	SSLPGSQRYA	..LIHLT.NY	ADETISVAID	VTNVYIMGYR
RICIN.	RAVRGRLTTG	ADVREHPIVL	PNRVGLPINQ	RFILVELSNH	AELSVTLALD	VTNAYVVGYR
Cons	---R-----	-----	---G-----	-----	-----	--N-Y---Y-
	121					180
DIAN.	AMDNAVNRRA	YYFKNQITSA	ELTALFPEVV	VANQKQLEYG	EDYQAIKNA	KITTDGQSRK
SAPORIN.	AMDNTNVNRA	YYFRSEITSA	ESTALFPEAT	TANQKALEYT	EDYQSIKNA	QITQGDQSRK
TRICH.	AGDT.....S	YFFNEASATE	AAKYVFKDAM	..RKVTLPYS	GNYERLQ...	..TAAGKIRE
RICIN.	A.....GNSA	YFFHPDNQED	AEAITHLFTD	VQNRYTFAPG	GNYDRLEQ..	...LAGNLR
Cons	A-----	Y-F-----	-----	-----	--Y-----	-----R-
	181					240
DIAN.	ELGLGINLLI	TMIDGV...N	KKVRVVKDEA	RFLLIAIQMT	AEAARFRYIQ	NLVTK..NFP
SAPORIN.	ELGLGIDLLS	TSMEAV...N	KKARVVKDEA	RFLLIAIQMT	AEAARFRYIQ	NLVIK..NFP
TRICH.	NIPLGLPALD	SAITTLF..Y	YN...ANSA	SALMVLIQST	SEAARYKFIE	QQIGKRVD..
RICIN.	NIELGNPGL	EASALYYS	TGGTQLPTLA	RSFIICIQMI	SEAARFQYIE	GEMRTRIRYN
Cons	---LG---L-	-----	-----A	-----IQ--	-EAAR---I-	-----
	241					300
DIAN.	NKFDSENKVI	QFQVSWSKIS	TAI.FGDCKN	GVFNKDYDFG	FGKVRQAKDL	QMGLLKYLGR
SAPORIN.	NKFNSENKVI	QFEVNWKKIS	TAI.YGDAKN	GVFNKDYDFG	FGKVRQVKDL	QMGLLMYLGK
TRICH.	KTFPLSLAII	SLENSWSALS	KQIQIASTNN	GQFESPVVLI	NAQNQRVTIT	NVDAGVVTSN
RICIN.	RRSAPDPSVI	TLENSWGRLS	TAIQ..ESNQ	GAFASPIQLQ	RRNGSKFSVY	DVSILIPIIA
Cons	-----I	-----W---S	---I-----	G-F-----	-----	-----
	301					360
DIAN.	PKSSSIEANS	TDDTADVL..	.....	.....	.....	.....
SAPORIN.	PKSSNEAN..	.....	.....	.....	.....	.....
TRICH.	IALLNRNRM	AAMDDVPMT	QSPGCGSYAI	.....	.....	.....
RICIN.	LMVYRCAPP	SSQF.....	.....	.....	.....	.....
Cons	-----	-----	-----	-----	-----	-----

Fig. 3. Optional alignment of the dianthin 30 sequence with those of saporin 6 [8], trichosanthin [9] and ricin A chain [10].

obtained for dianthin 30 and these matched exactly the deduced residues from the cDNA, while the N-terminus of dianthin 32 differed considerably (data not shown).

Comparison of the amino acid sequence deduced for dianthin 30 with those of other RIPs revealed a very high degree of homology (75%) with saporin 6. While the degree of homology was much lower in the case of trichosanthin (17%) or ricin A chain (20%) several key residues, including the active site tetrapeptide -EAAR- were absolutely conserved (Fig. 3).

Further analysis of the dianthin 30 sequence revealed a carboxy-terminal region containing an *N*-

glycosylation site and showing homology to a carboxy-terminal propeptide present in several plant vacuolar proteins such as wheat germ agglutinin and barley lectin (Fig. 4). In the case of barley lectin this carboxy-terminal propeptide has been shown to function in

Barley lectin	VFAEAIANSTLVAE
Wheatgerm agglutinin	VFAEAIANSTLLQE
Dianthin-30	PKSSSIEANSTDDTADVL

Fig. 4. Amino acid sequence comparison of the carboxy-terminal propeptides of barley lectin [11] and wheat germ agglutinin [12] with a carboxy-terminal region of dianthin 30. *N*-glycosylation signals are boxed.

sorting the lectin to the plant cell vacuole [11]. Although the intracellular location of dianthin 30 is unknown at present, it may well be a vacuolar protein and the carboxy-terminal region may play a necessary role in targeting the RIP to this organelle.

## References

- 1 Stirpe, F. and Barbieri, L. (1986) *FEBS Lett.* 195, 1–8.
- 2 Endo, Y. and Tsurugi, K. (1987) *J. Biol. Chem.* 262, 8128–8130.
- 3 Olsnes, S. and Pihl, A. (1982) in *Molecular Action of Toxins and Viruses* (Cohen, P. and Van Heyningen, S., eds.), pp. 51–105, Elsevier, Amsterdam.
- 4 Irvin, J.D. (1975) *Arch. Biochem. Biophys.* 169, 522–528.
- 5 McGrath, M.S., Hwang, K.M., Caldwell, S.E., Gaston, I., Luk, K.C., Wu, P., Ng, V.L., Crowe, S., Daniels, J., Marsh, J., Deinhart, T., Lekas, P.V., Vennari, J.C., Yeung, H.W. and Lifson, J.D. (1989) *Proc. Natl. Acad. Sci. USA* 86, 2844–2848.
- 6 Spooner, R.A. and Lord, J.M. (1990) *Trends Biotech.* 8, 189–193.
- 7 Stirpe, F., Williams, D.G., Onyon, L.J., Legg, R.F. and Stevens, W.A. (1981) *Biochem. J.* 195, 399–405.
- 8 Benatti, L., Saccardo, M.B., Dani, M., Nitti, G., Sassano, M., Lorenzetti, R., Lappi, D.A. and Soria, M. (1989) *Eur. J. Biochem.* 183, 465–470.
- 9 Chow, T.A., Feldman, R.A., Lovett, M. and Piatak, M. (1990) *J. Biol. Chem.* 265, 8670–8674.
- 10 Lamb, F.I., Roberts, L.M. and Lord, J.M. (1985) *Eur. J. Biochem.* 148, 265–270.
- 11 Bednarek, S.Y., Wilkins, T.A., Dombrowski, J.E. and Raikhel, N.V. (1990) *Plant Cell* 2, 1145–1155.
- 12 Raikhel, N.V. and Wilkins, T.A. (1987) *Proc. Natl. Acad. Sci. USA* 84, 6745–6749.