

GENE 07340

Sequence of the proteinase-inhibitor cystatin homologue from the pollen of *Ambrosia artemisiifolia* (short ragweed)*

(Amino acid alignment; allergy; polymerase chain reaction; sequence comparison; recombinant DNA)

Bruce L. Rogers^a, Joanne Pollock^a, David G. Klapper^b and Irwin J. Griffith^a

^aImmuLogic Pharmaceutical Corporation, Waltham, MA 02154, USA; and ^bDepartment of Microbiology and Immunology, University of North Carolina School of Medicine, Chapel Hill, NC 27599, USA

Received by R. Wu: 26 February 1993; Revised/Accepted: 23 April/23 April 1993; Received at publishers: 17 June 1993

SUMMARY

The nucleotide sequence of a cDNA (designated *IPC1/5*) encoding a cystatin (Cyt) proteinase-inhibitor homologue from short ragweed (*Ambrosia artemisiifolia*) pollen was determined and compared to other plant and animal Cyt. The absence of disulfide bonds in the predicted translation product of the *IPC1/5* sequence suggests that it most resembles family-I members of the Cyt superfamily. Significant amino acid (aa) sequence identity was found when comparing the translated sequence of *IPC1/5* to rice seed Cyt, human keratocyte Cyt A and human liver Cyt B.

INTRODUCTION

The Cyt superfamily of proteins is comprised of several related family groups that share the characteristic of inhibiting cysteine proteinases. Type-I Cyt (sometimes referred to as stefins) consist of approximately 100 aa with no disulfide bonds or carbohydrate (reviewed in Barrett, 1987). Two well-characterized mammalian Cyt within this group have been designated Cyt A and B (Takio et al., 1983; Takio et al., 1984). Type-II Cyt differ from type-I in that they have two disulfide loops and type-III (sometimes referred to as kininogens) are even

more complex proteins comprised of multiple domains totaling over 350 aa (Barrett, 1987).

EXPERIMENTAL AND DISCUSSION

(a) Cloning of a cystatin-like proteinase inhibitor

A major allergen of *A. artemisiifolia*, designated *Amb a 1*, was biochemically purified from pollen and examined by tryptic digestion (Rafnar et al., 1991). Analysis of one tryptic fragment revealed a sequence that did not match that of *Amb a 1*. The gene encoding this sequence, designated *IPC1/5*, was cloned and found to have an open reading frame of 276 nt that encoded a protein of 92 aa (Fig. 1). A search of the SwissProt database showed homology of *IPC1/5* with the type-I Cyt (Fig. 2). The type-I Cyt have been isolated from several species and tissue types including human liver (Ritonja et al., 1985), human epidermal keratinocytes (Kartasova et al., 1987), and rice seeds (Kondo et al., 1990). So far, the rice seed Cyt are the only other plant Cyt that have been described.

Correspondence to: Dr. B.L. Rogers, ImmuLogic Pharmaceutical Corp., 610 Lincoln Street, Waltham, MA 02154, USA. Tel. (1-617) 466-4766; Fax (1-617) 466-6010.

*On request the authors will supply detailed experimental evidence for the conclusions reached in this Short Communication.

Abbreviations: aa, amino acid(s); *A.*, *Ambrosia*; *Amb a 1*, allergen from *Ambrosia artemisiifolia* named according to the IUIS approved nomenclature (Marsh et al., 1988); bp, base pair(s); Cyt, cystatin(s); *IPC 1/5*, gene encoding Cyt homologue; cDNA, DNA complementary to mRNA; kb, kilobase; nt, nucleotide(s); oligo, oligodeoxynucleotide; PCR, polymerase chain reaction.

AAAAAATCGATCTCTTGGAGGAATACCGAAGTAAAGACAATGATAACAGGTCGATTT 60
 M S I L G G I T E V K D N D N S V D F 19
 CGACGAGCTTGGTAAATTCGGCATCTGCTGAACACACAAAGGAGAAATGCTGCTGGA 120
 D E L A K F A I A E H D R K E R A A L E 39
 GTTTCGAAAATGATAGAAAAGAGCAGGCGGTACAGGCGACCATGTATTATATAAA 180
 F G K V I E K K Q Q A V Q G T M Y Y I K 59
 AGTGAAGCAATGATGCTGCTGAGAGAAAATCTATGAAGCAAGGTGCGGTTAAGCT 240
 V E A N D G G E K K T Y E A K V W V K L 79
 ATGGGAATTTCAAGGAATTCAGCAACTCAAACTGTTTGTATGTTGCCACCTCACTT 300
 W E N F K E L Q E L K L V * 92
 AACTCCATATGACGG 316

Fig. 1. The nt sequence and deduced aa sequence of the *A. artemisiifolia* Cyt homologue. The putative start codon is underlined and the stop codon is indicated by an asterisk. The nt sequence has GenBank accession No. L16624.

(b) Sequence homology with other cystatins

Recently, crystallographic data from a 1:1 stoichiometric complex of Cyt with papain confirmed the previously postulated contact residues of these two molecules (Stubbs et al., 1990). Significantly, the regions that contain these contact residues are conserved between the translated cDNA clone *IPC1/5* and the other Cyt (Fig. 2). For example, there is conservation of a G residue near the N terminus of the Cyt superfamily that probably corresponds to Gly⁵ in the translated *IPC1/5* sequence (arrowed in Fig. 2). Another region of significant homo-

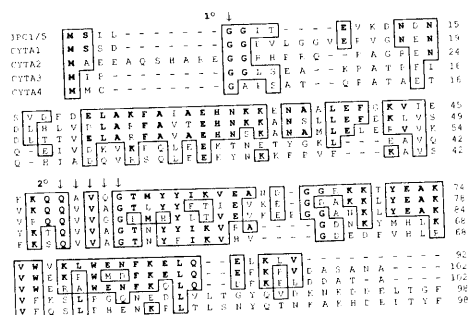


Fig. 2. Comparison of the aa sequences of the *A. artemisiifolia* Cyt homologue and other members of the Cyt proteinase inhibitor family of proteins. Identical residues are in bold and regions of sequence homology are boxed. Dashes represent gaps introduced to maximize sequence alignment. The aa numbering is derived from the complete sequence starting with the N-terminal Met assigned the number +1. The known Cyt sequences are listed as follows: CYTA1, Cyt proteinase inhibitor I (oryzaeystatin-I) from rice seeds (Kondo et al., 1990); CYTA2, a second Cyt proteinase inhibitor (oryzaeystatin-II) from rice seeds (Kondo et al., 1990); CYTA3, human Cyt from epidermal keratocytes (Kartasova et al., 1987); CYTA4, human liver Cyt (Ritonja et al., 1985). The % aa sequence identity with the *A. artemisiifolia* homologue is 53.3% (CYTA1), 47.8% (CYTA2), 25.0% (CYTA3), and 20.7% (CYTA4). The total % of aa sequence similarity with the *A. artemisiifolia* homologue is 64.1% (CYTA1), 61.9% (CYTA2), 37.0% (CYTA3) and 31.6% (CYTA4). The putative Cyt primary (1°) and secondary (2°) contact residues are arrowed. Computer programs contained in PCGENE (IntelliGenetics, Inc., Mountain View, CA) were used for both aa and nt sequence analysis.

logy referred to as the N-terminal 'trunk' comprised of the aa residues QVVAG (Stubbs et al., 1990), is also conserved in *IPC1/5* (Fig. 1, panel B). Although this region is highly conserved overall, the translated *IPC1/5* sequence has an Ala⁵⁰ rather than a V, and a Q⁵² rather than an A/G at these positions (Fig. 2). These differences may suggest some structural latitude at these positions.

(c) *A. artemisiifolia* cystatin homologue is an allergen

A. artemisiifolia pollen is a complex protein source that is responsible for one of the most important aeroallergens in North America (King, 1976). The cDNAs encoding the major allergens from this pollen, designated *Amb a* I and *Amb a* II, have been cloned and extensively characterized (Rafnar et al., 1991; Griffith et al., 1991; Rogers et al., 1991). To determine whether the *IPC1/5* cDNA encodes an allergenic protein, it was expressed in *E. coli* as a β -galactosidase fusion protein using the pSEM-2 vector (Knapp et al., 1990). The recombinant protein was subjected to Western blot analysis using ragweed allergic human IgE as a probe. One in three patients tested had IgE that bound specifically to the recombinant *IPC1/5* fusion protein (data not shown). Therefore, the cysteine proteinase inhibitor analog in ragweed pollen encoded by the *IPC1/5* cDNA is an allergen and has been designated *Amb a* VIII according to IUIS nomenclature (Marsh et al., 1988).

(d) Conclusions

(1) A cDNA encoding a Cyt homologue of 92 aa in length has been isolated from *A. artemisiifolia* pollen.

(2) Sequence comparison of translated *IPC 1/5* with other known Cyt proteins detected considerable sequence homology with known Cyt. This level of homology was highest (64.1%) for the rice seed Cyt proteinase inhibitor I. Those aa known from crystallographic data to be contact residues between Cyt and papain are conserved in the aa sequence predicted from the nt sequence of *IPC 1/5*.

(3) Ragweed allergic human IgE binds to a β -galactosidase fusion protein containing the translated sequence of *IPC 1/5*.

REFERENCES

- Barrett, A.J.: The cystatins: a new class of peptidase inhibitors. *Trends Biochem. Sci.* 12 (1987) 193–196.
- Griffith, I.J., Pollock, J., Klapper, D.G., Rogers, B.L. and Nault, A.: Polymorphism of *Amb a* I and *Amb a* II, the major allergens of *Ambrosia artemisiifolia* (short ragweed). *Int. Arch. Allergy Appl. Immunol.* 96 (1991) 296–304.
- Kartasova, T., Cornelissen, B.J.C., Belt, P. and Van de Putte, P.: Effects

- of UV, 4-NQO and TPA on gene expression in cultured human epidermal keratinocytes. *Nucleic Acids Res.* 15 (1987) 5945–5962.
- King, T.P.: Chemical and biological properties of some atopic allergens. *Adv. Immunol.* 23 (1976) 77–105.
- Knapp, S., Broker, M. and Amann, E.: pSEM vectors: high level expression of antigenic determinants and protein domains. *BioTechnology* 8 (1990) 280–281.
- Kondo, H., Abe, K., Nishimura, I., Watanabe, H., Emori, Y. and Arai, S.: Two distinct cystatin species in rice seeds with different specificities against cysteine proteinases. *J. Biol. Chem.* 265 (1990) 15832–15837.
- Marsh, D.G., Goodfriend, L., King, T.P., Lowenstein, H. and Platt, Mills, T.A.E. Allergen nomenclature. *Ann. Allergy* 60 (1988) 499–503.
- Rafnar, T., Griffith, I.J., Kuo, M.-C., Bond, J.F., Rogers, B.L. and Klapper, D.G.: Cloning of *Amb a* I (antigen E), the major allergen family of short ragweed pollen. *J. Biol. Chem.* 266 (1991) 1229–1236.
- Ritonja, A., Machleidt, W. and Barrett, A.J.: Amino acid sequence of the intracellular cysteine proteinase inhibitor cystatin B from human liver. *Biochem. Biophys. Res. Commun.* 131 (1985) 1187–1192.

- Rogers, B.L., Morgenstern, J.P., Griffith, I.J., Yu, X.-B., Counsell, C.M., Brauer, A.W., King, T.P., Garman, R.D. and Kuo, M.C.: Complete sequence of the allergen *Amb a* II: recombinant expression and reactivity with T cells from ragweed allergic patients. *J. Immunol.* 147 (1991) 2547–2552.
- Stubbs, M.T., Laber, B., Bode, W., Huber, R., Jerala, R., Lenarcic, B. and Turk, V.: The refined 2.4 Å X-ray crystal structure of recombinant human stefin B in complex with the cysteine proteinase papain: a novel type of proteinase inhibitor interaction. *EMBO J.* 9 (1990) 1939–1947.
- Takio, K., Kominami, E., Wakamatsu, N., Katunuma, N. and Titani, K.: Amino acid sequence of rat liver thiol proteinase inhibitor. *Biochem. Biophys. Res. Commun.* 115 (1983) 902–908.
- Takio, K., Kominami, E., Bando, Y., Katunuma, N. and Titani, K.: Amino acid sequence of rat epidermal thiol proteinase inhibitor. *Biochem. Biophys. Res. Commun.* 121 (1984) 149–154.