Isolation and DNA Sequence of a Gene Encoding α-Trichosanthin, a Type I Ribosome-inactivating Protein*

Theresa P. Chow, Robert A. Feldman, Michael Lovett, and Michael Piatak‡
From Genelabs Incorporated, Redwood City, California 94063

α-Trichosanthin (α-TCS) is a ribosome-inactivating protein that has recently been shown to inhibit the replication of human immunodeficiency virus. We have isolated a gene encoding α-TCS and have determined its DNA sequence. The data indicate that α-TCS is synthesized as a preprotein consisting of 289 amino acids, the first 23 residues of which comprise a putative secretory signal peptide. The last 19 residues comprise a carboxyl extension that has not been reported to be associated with the mature protein and that may be processed in the endoplasmic reticulum or Golgi apparatus of cells producing α-TCS. The mature protein consists of 247 amino acids. The sequence predicted by translation of the DNA sequence agrees with and confirms the primary sequence determined recently on the protein. The molecular clone for α-TCS will facilitate directed mutational analyses that may provide information on how this peptide, and other ribosome-inactivating proteins, function. These studies may also lead to the development of therapeutic agents with altered activities and/or improved properties for in vivo use.

EXPERIMENTAL PROCEDURES

Materials-Leaves of T. kirilowii Maxim. were collected under contract in South Korea. The leaves were harvested, shipped on dry ice, and stored at -70 °C. Chemicals, enzymes, and radioisotopes were obtained from various commercial sources and were used in accordance with the suppliers' recommendations. Oligonucleotides were synthesized using standard chemistry on instruments from Biosearch, San Rafael, CA, or Applied Biosystems Inc., Foster City, CA, and were supplied by Nucleotide Collective Services, Genelabs Inc. Random priming kits for generating radioactively labeled probes were obtained from Boehringer Mannheim. GeneAmp® DNA amplification reagent kit was from Cetus/Perkin-Elmer Inc., Norwalk, CT; Sequenase® DNA sequencing kit was from U.S. Biochemicals Corp.; T7 DNA sequencing kit was from Pharmacia LKB Biotechnology Inc.; Lambda ZAP® II/EcoRI CIAP vector kit and Gigapack® Plus phage packaging kit were from Stratagene, La Jolla, CA. The Escherichia coli strain, XL1-Blue (endA1, hsdR17 (rk-, mk+), supE44, thi, recA1, gyrA96, relA1, lacI, lacZΔM15, Tn10(tev5)) was also obtained from Stratagene.

Methods and protocols for recombinant DNA manipulations are generally referenced by Ausubel et al. (10) and Maniatis et al. (11).

Isolation of Genomic DNA—Genomic DNA was isolated from frozen leaves by a modification of published methods (12). Briefly, 36 g of frozen leaves were ground to a fine powder using a mortar and pestle kept on dry ice. 22-Mercaptoethanol was added to 2% (w/v) followed by an equal volume of chloroform:isoamyl alcohol (24:1) by gently mixing for 30 min followed by low speed centrifugation sufficient to separate the aqueous and chro-
organic phases. The organic phase was removed, and the extraction was repeated. One-tenth volume of 10% (v/v) CTAB was then added to the isolated aqueous phase, and the extraction was repeated once again. The upper aqueous phase was removed, and the DNA was precipitated by adding an equal volume of precipitation buffer (1% CTAB, 50 mM Tris-Cl, pH 8.0, 10 mM EDTA). The DNA was collected by spooling it on a glass rod and then converted to the sodium salt by resuspending it in 0.3 M sodium acetate. The precipitate was rinsed in 95% ethanol, dried, and dissolved in 10 mM Tris-Cl, 1 mM EDTA, pH 8.0. To further eliminate contaminants, the DNA was reprecipitated by adding an equal volume of 2× extraction buffer followed by an equal volume of 10 mM Tris-Cl, pH 8.0, 1 mM EDTA. Greater than 5 mg of nucleic acid per mg of ground leaf tissue was obtained.

Amplification of Specific Probe Sequence—The degenerate oligonucleotides shown in Fig. 1 were used to prime the amplification of an α-TCS-specific DNA fragment using DNA isolated from T. kirilowii as template. A typical reaction consisted of 100 pmol of primer MPQP-1, 50 pmol each of primers MPQP-2 and MPQP-3, 200 μM concentration each of dATP, dCTP, dGTP, and dTTP, 1 μg of genomic DNA as template, 2.5 units of Taq polymerase, 50 mM KCl, 10 mM Tris-Cl, pH 8.3, 1.5 mM MgCl₂, and 0.01% (w/v) gelatin in a total reaction volume of 100 μl. The reaction mixture was overlaid with 76 μl of mineral oil to prevent evaporation. The cycle profile used with the DNA Thermal Cycler (Cetus/Perkin-Elmer) was: denaturation at 94 °C for 1 min, annealing at 45 °C for 2 min, and primer extension at 72 °C for 3 min assuming an extension rate of 1000 bp/min. Primer extension was carried out for 25 to 30 cycles.

To generate sufficient amounts of DNA for use as a probe, DNA amplification reactions were carried out at 4 times the amounts and volumes noted above. The resultant products were fractionated on a 6% (w/v) polyacrylamide gel (acrylamide:bisacrylamide, 19:1) in 0.04 M Tris acetate; pH 7.8, 0.002 M EDTA and heated to 100 °C for 5 min to denature the DNA. The mixture was then quick frozen on dry ice and centrifuged briefly to collect condensate. Two μl of 5× Sequenase™ sequencing buffer was added, and the mixture annealed at 37 °C for 5 min. The standard sequencing protocol was then followed.

T. kirilowii Genomic Library and Clone Selection—A library of complete digest EcoRI genomic DNA fragments in lambda ZA104s II (Stratagene) was generated following the manufacturers' suggestions and protocols. Recombinant phage were packaged in vitro using a Gigapack® Plus packaging mix (Stratagene). Approximately 1×10⁶ recombinant phage were lifted onto nitrocellulose filters and probed with 6.5×10⁶ cpm of 32P-labeled probe fragment per ml of 3× SSC, 0.1% sodium dodecyl sulfate at 65 °C. Recombinant plasmids were derived from selected phage by in vivo excision and circularization mediated by the addition of helper phage as per instructions of the manufacturer.

RESULTS

Generation of an α-TCS-specific Probe—The complete primary sequence for α-TCS had been determined (9), and oligonucleotide probe/primers could be designed from regions showing minimum degeneracy in coding potential. The probe/primers designed and the protein sequences from which they were derived are shown in Fig. 1. To allow for the generation of sequences significantly longer than 20 nucleotides of manageable complexity, deoxynosine residues were incorporated at positions where all four nucleotides were possible (13, 14). One oligonucleotide pool (MPQP-1) was derived from amino acid residues 80–101 and consisted of 128 isomers of a 35-mer containing 4 deoxynosine residues. Two other pools (MPQP-2 and -3) were derived from amino acid residues 164–174 and consisted of 128 isomers each of a 32-mer containing 3 deoxynosine residues. The latter two pools of synthetic sequences differed in G or T at the second position from the 5′ end. The first set of oligonucleotides and the other two sets taken together were also designed to face one another as primers on a genomic DNA template and to be used for amplification of a specific DNA fragment in a polymerase chain reaction. This was the approach taken to isolate a sequence-specific probe for α-TCS.

It was assumed that the gene sequence encoding α-TCS would not contain any introns as was found for ricein, a Type II RIP (15). On this assumption and from the positions of the oligonucleotide primers relative to the determined protein sequence, it was predicted that a DNA fragment of approximately 255 bp would be amplified in a polymerase chain reaction using T. kirilowii genomic DNA as template. The results of such a reaction are shown in Fig. 2. One amplified DNA fragment of the expected size was detected after agarose gel electrophoresis by ethidium bromide staining and by autoradiography when 32P-labeled primer was used. No significant “background” products were noted even after 40 cycles of polymerase chain reaction.

To confirm the identity of the amplified DNA fragment, approximately 100 ng of gel-purified fragment was subjected to a DNA sequencing reaction using the MPQP-1 oligonucleotide pool as primer. A sequence for 116 bases was determined and translated in all six potential reading frames (Fig. 3). One open reading frame showed a translation product which matched exactly the sequence of α-TCS from amino acid residue 128 through 163. The translated residues at the beginning and the end of the DNA sequence did not match the determined protein sequence, but these regions were at the limits of the interpretable DNA sequence data, and errors may have been made. Nevertheless, it was clear that the amplified fragment did correspond to an α-TCS-specific or -like sequence.

![Fig. 1. Design of oligonucleotide primers/probes specific for α-TCS](https://example.com/fig1.png)
Genomic Clone for α-Trichosanthin

Fig. 2. Polymerase chain reaction amplification of a DNA fragment corresponding to an α-TCS coding sequence. As detailed under "Experimental Procedures," a polymerase chain reaction was run on genomic DNA from T. kirilowii using 32P-labeled MFQP-1 and unlabeled MFQP-2 and -3 as primers. The reaction products were separated on a 6% polyacrylamide gel in TAE and detected by exposure to x-ray film. The developed autoradiogram is shown. Lane a, HaeIII-digested αX174 DNA end-labeled with 32P. Lane b, the total polymerase chain reaction products.

Fig. 3. Determined DNA sequence and corresponding translation for the α-TCS-amplified fragment. The DNA sequence determined from the approximate 255-bp fragment amplified underlined. (a), Genomic Clone (b), a Genomic Clone (c), Genomic Clone (d), a Genomic Clone

Isolation and Sequence Determination of a Genomic Clone for α-TCS—Without introns, a continuous DNA sequence of about 750 bp would be sufficient to code for mature α-TCS. However, for the coding of a secretory signal peptide, a minimum total coding sequence of about 850 bp might be expected. On Southern blot analysis using 32P-labeled amplified fragment as probe positive-hybridizing bands of about 7, 4, and about 750 bp would be sufficient to code for mature α-TCS.

A library of EcoRI fragments in lambda ZAP II was generated as described under "Experimental Procedures." Approximately 106 recombinants were screened with 500 ng of the specifically amplified DNA fragment labeled with 32P to a specific activity of about 1.3 x 106 cpmp/μg. One clone, pQ21D, was strongly positive; a second clone, pQ30E, yielded a fainter signal. Both clones were plaque-purified and rescued as plasmids as described under "Experimental Procedures." Restriction analyses (data not shown) indicated that the cloned inserts were approximately 4 and 0.6 kbp, respectively, in size. The clone pQ21D, but not pQ30E, contained a SalI site predicted from the DNA sequence of the specifically amplified fragment. This information, together with the observation that pQ21D hybridized much more strongly than pQ30E, suggested a higher probability that pQ21D contained a DNA sequence specific to α-TCS.

Preliminary sequence analysis of pQ21D confirmed that it contained a sequence potentially coding for α-TCS beginning 409 bp from one end. The sequence information was then extended to cover the entire coding region for α-TCS as shown in Fig. 4. No attempt was made to sequence the entire non-coding flanking regions in the gene, although 339 bp preceding the first in-frame ATG codon were confirmed.

The DNA sequence determined for pQ21D, a translation showing the coding of α-TCS, and a comparison of the determined protein sequence to that translation are shown in Fig. 5. First of all, it is clear that pQ21D contains an authentic gene sequence for α-TCS. Aligned with the determined protein sequence, which also indicates the limits of coding for the mature sequence, there are only two amino acid differences, a Thr for a Ser at position 211 and a Met for a Thr at position 224, both relatively conservative changes. The translation information indicates that the gene sequence encodes a precursor protein. The open reading frame begins with an ATG at nucleotide position 340 and continues through nucleotide 1206. Nucleotides 340 through 405 likely encode a putative secretory signal peptide; nucleotides 1150 through 1206 encode a carboxyl-terminal extension of the mature protein. There are no potential N-linked glycosylation sites (Asn-X-Ser/Thr) and, as expected, there are no indications of introns contained within the coding sequence. The sequence upstream of the open reading frame is 72% A + T, like that in the gene for ricin (15). There are several sequences that resemble a TATA box found in eukaryotic genes 30–35 nucleotides upstream of transcription start sites (16, 17), but there were no consensus sequences upstream from the transcription start sites for α-TCS.
Genomic Clone for α-Trichosanthin

**FIG. 5.** Gene sequence for α-trichosanthin. (a), the coding strand sequence determined for pQ21D. The sequence is numbered above for convenience, and relevant restriction sites are shown. Potential control sequences are underlined. (b), the translation of the encoded precursor protein for α-TCS. (c), the reported primary sequence of α-TCS taken from Collins et al. (9). The protein sequences are numbered below. The mature protein sequence is numbered from 1 through 247; the putative secretory signal peptide is numbered in a negative fashion as it precedes the mature sequence; and the additional carboxyl-terminal sequence noted in the precursor is numbered in parentheses to represent a continuation of the mature sequence. It will not be possible to identify any of these as actual control sequences until a transcript for α-TCS can be isolated and mapped on the gene.

**DISCUSSION**

Although gene sequences for several RIPs have been cloned or synthesized (18–21), this report describes the first isolation and DNA sequence of a complete gene for a Type I RIP. RIPs constitute a rather large group of proteins. In cloning α-TCS, we chose an approach which should be generally applicable to the cloning of other RIPs. The basis of this approach is to rapidly evaluate the prevalence of a RIP sequence in a DNA (or RNA) preparation and to generate a highly specific probe by using degenerate primers in a polymerase chain reaction. For α-TCS, we had the benefit of having a complete protein sequence, allowing us to select the optimum regions for primer design. However, even with limited protein sequence information, specific probe sequences might be amplified. As noted previously and shown in Fig. 6, RIPs share extensive sequence homology. If, for example, only amino terminal sequence information is available, degenerate primers for PCR might be designed from the known sequence and coupled with primers designed from a region showing extensive sequence homology among RIPs. In Fig. 6, one such region would correspond to α-TCS residues 160 through 167. A consensus sequence, EAARF(K/Q)YI, might be taken, and a degenerate primer might be designed. After polymerase chain reaction, amplified products of a predicted size, based on alignment of the new sequence with the other known sequences, would be isolated and characterized by DNA sequencing. As additional protein sequences are determined, it may even be possible to generate “universal” RIP primers to be used to screen for RIPs and facilitate their cloning when no protein sequence is available.

With limited or no sequence information, however, the degeneracy of the primers to be used could be extreme. In the example above, the consensus primer could contain as many as 4096 isomers of a 23-mer. Incorporation of deoxyinosine at selected positions might reduce the complexity considerably. For example, the potential number of isomers for the sequence from which we derived MIPQP-2 plus -3 was 9216. The final primer pools used contained only 256 isomers. Codon usage tables for plants or preferably for RIPs might also be employed to further restrict primer degeneracy. At this time, we cannot make estimates or recommendations as to the limit of the complexity of primer pools that may be used successfully in polymerase chain reaction. Primer pools containing greater than 4000 isomers have been used successfully to amplify mammalian sequences (22), but colleagues were unsuccessful using similarly complex primers applied to another plant-derived DNA. A complication in the latter instance, however, was that the genome complexity was estimated to be at least 10-fold greater than that of T. kirilowii or mammalian DNA which is about 10⁶. This would increase the chances of spurious

**FIG. 6.** Alignment of complete RIP sequences. Available RIP primary sequences are aligned to show maximum homologies. All positions matched and identical residues are shaded. The sequence of α-TCS is numbered for reference. TCS, α-TCS sequence taken from pQ21D; ricA, ricin A-chain sequence (19); abrA, abrin A-chain sequence (32); BPSI, barley protein synthesis inhibitor (33); MIRA, mirabilis sequence (18).

K. Wood and M. Pietak, personal communication.
arious or nonspecific priming requiring the testing of several different primer combinations. Problems not withstanding, this approach and method quickly led to our isolation of a RIP gene and may be useful when applied to the cloning of others.

It is typical to find multiple RIPS in a plant. In Ricinum communis, several related genes exist, coding for at least two toxins and one agglutinin (15, 19, 20, 23). Preliminary Southern blot analysis of EcoRI-digested T. kirilowii DNA showed three hybridizing bands of about 7, 4, and 1.4 kbp in size, respectively. Follow-up analyses on gels run for shorter periods showed an additional band of about 0.6 kbp. Of these, the 4- and 0.6-kbp bands correspond to the clones, pQ21D and pQ30E, respectively. The clone, pQ30E, has been sequenced in part, and the data show that it contains an RIP-like sequence. The insert is too small to encode a complete protein, however, and it is possible that it does not derive from a functional gene. We plan to isolate a full length clone from an alternatively restricted library and determine the identity of this sequence. A second RIP, trichokin, has been isolated from seeds of T. kirilowii indicating that at least one additional, active gene exists (24). Thus, it is probable that RIPS in T. kirilowii, like those in R. communis and likely in other plants, comprise a multigene family.

The translation of pQ21D indicates that α-TCS is produced as a preproprotein. The mature protein is preceded by 23 codons which resemble a consensus secretory signal peptide (25). This is consistent with the premise that all RIPS are secreted proteins as they would otherwise diminish or stop protein synthesis in the cells in which they are produced. Ribosomes taken from a plant producing a RIP are particularly resistant to that RIP, but the resistance is not absolute (26-28). It is reasonable to expect that these proteins, which are not only potent inhibitors of protein synthesis but are produced in relatively large amounts, would be secreted and compartmentalized. Another feature of α-TCS revealed by the gene translation may also relate to this. There are 19 codons following the mature protein sequence which, by analogy to preproprerin, may be processed after translocation to the endoplasmic reticulum (29). Prorcin contains 12 amino acids which link the carboxyl end of the A-chain sequence to the amino end of the B-chain sequence. This sequence may function in one aspect to facilitate the formation of the disulfide bond between the A- and the B-chain, but it may also function to maintain the protein in an inactive state until it is safely placed across the membrane of the endoplasmic reticulum. Nonreduced ricin is much reduced in its ability to block protein synthesis in an in vitro translation assay (30), and prorcin, produced by injection of RNA transcripts into oocytes, is essentially inactive (31). It will be of interest to test our hypothesis by expressing both the mature and the proprotein forms of α-TCS in E. coli or in vitro and compare their activities on eukaryotic ribosomes.

Acknowledgments—We wish to thank Laura Christensen and Cam Hoover for synthetic oligonucleotides and help with DNA sequencing, Dr. Seung Mann Park for help in obtaining T. kirilowii Maxim. leaves, and Dr. Jeffery Lifson for helpful discussions and personal support.

REFERENCES

1. Jimenez, A., and Vasquez, D. (1985) Annu. Rev. Microbiol. 39, 649-672

2. Stirpe, F., and Barbieri, L. (1986) FEBS Lett. 195, 1-8

3. Yeung, H. W., Li, W. W., Feng, Z. Barbieri, L., and Stirpe, F. (1988) Int. J. Peptide Protein Res. 31, 265-268

4. Wang, Y., Qian, R. Q., Gu, Z.-W., Jin, S.-W., Zhang, L.-Q., Xia, Z.-X., Tian, G.-Y., and Ni, C.-Z. (1986) Pure Appl. Chem. 58, 279-288

5. Jin, Y.-C. (1985) in Advances in Chinese Medicinal Materials Research (Chang, H. M., Yeung, H. W., Tsao, W.-W., and Koo, A., eds), pp. 319-326, World Scientific Publishing Co., Singapore

6. Tsao, S. W., Yan, K. T., and Yeung, H. W. (1986) Toxicon 24, 801-840

7. Zhang, X., and Wang, J. (1986) Nature 321, 477-478

8. McGrath, M. S., Hwang, K. M., Caldwell, E. E., Gaston, I., Lok, K.-C., Wu, P., Ng, V. L., Crowe, S., Daniels, J., Marsh, J., Dehart, T., Lekas, P. V., Vennari, J. C., Yeung, H.-W., and Lifson, J. D. (1988) Proc. Natl. Acad. Sci. U. S. A. 85, 2944-2948

9. Collins, K. J., Robertson, J. D., LoPresti, M., Stone, K. L., Williams, K. K., Wu, P., Hwang, K., and Piatnak, M. (1990) J. Biol. Chem. 265, 8665-8669

10. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, A. A., and Struhl, K. (eds) (1988) Current Protocols in Molecular Biology, J. Wiley and Sons, New York

11. Maniatis, T., Fritsch, E. F., and Sambrbrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

12. Taylor, B., and Powell, A. (1983) Focus (Bethesda Res. Lab.) 4, 6

13. Ohbika, E., Matsuaki, S., Ikehara, M., Takahashi, Y., and Matsubara, K. (1985) J. Biol. Chem. 260, 2605-2608

14. Takahashi, Y., Kato, K., Hayashizaki, Y., Wakabayashi, T., Ohtaka, E., Matsuaki, S., Ikehara, M., and Matsubara, K. (1986) Proc. Natl. Acad. Sci. U. S. A. 83, 1921-1925

15. Halling, K. C., Halling, A. C., Murray, E. E., Ladin, B. F., Houston, L. L., and Weaver, R. F. (1985) Nucleic Acids Res. 13, 8019-8033

16. Grosschedl, R., and Birnstiel, M. L. (1980) Proc. Natl. Acad. Sci. U. S. A. 77, 1432-1436

17. Benoist, C., and Chambon, P. (1981) Nature 286, 304-310

18. Habuka, N., Murakami, Y., Noma, M., Kudo, T., and Horikoshi, K. (1989) J. Biol. Chem. 264, 6625-6637

19. Lanabi, F. L., Roberts, L. M., and Lord, J. M. (1965) Eur. J. Biochem. 148, 265-270

20. Roberts, L. M., Lamb, F. I., Pappin, D. J. C., and Lord, J. M. (1980) J. Biol. Chem. 255, 19652-19656

21. Benatti, L., Saccardo, M. B., Dani, M., Nitti, G., Sassano, M., Lorenzetti, R., Lappi, D. A., and Soria, M. (1989) Eur. J. Biochem. 183, 460-470

22. Larrick, J. W., Danielsson, L., Brenner, C. A., Wallace, E. F., Abrahamson, M., Fry, K. E., and Borrebekke, C. A. K. (1989) BioTechnol. 7, 934-938

23. Araki, T., and Funatsu, G. (1987) Biochim. Biophys. Acta 911, 191-200

24. Casellas, P., Dussossoy, D., Falasco, A. I., Barbieri, L., Guillenot, J. C., Ferrara, P., Bolognesi, A., Cenini, P., and Stirpe, F. (1988) Eur. J. Biochem. 176, 581-588

25. von Heine, G. (1981) Eur. J. Biochem. 116, 419-422

26. Hui, R. A., Beuming, G., and Shepherd, R. J. (1973) Virology 56, 390-393

27. Coleman, W. H., and Roberts, W. K. (1981) Biochim. Biophys. Acta 654, 57-66

28. Harley, S. M., and Beevers, H. (1982) Proc. Natl. Acad. Sci. U. S. A. 79, 5395-5398

29. Butterworth, A. G., and Lord, J. M. (1983) Eur. J. Biochem. 137, 57-65

30. Olades, S., Sandvig, K., Rines, K., and Pihl, A. (1976) J. Biol. Chem. 251, 3985-3992

31. Richardson, P. T., Westby, M., Roberts, L. M., Gould, J. H., Colman, A., and Lord, J. M. (1989) FEBS Lett. 255, 15-20

32. Funatsu, G., Taguchi, Y., Kamoshino, M., and Yanaka, M. (1988) Agric. Biol. Chem. 52, 1095-1097

33. Asano, K., Svensson, B., Svensson, I., Poulsen, F. M., and Roesporff, P. (1986) Carlsberg Res. Commun. 51, 129-141
Isolation and DNA sequence of a gene encoding alpha-trichosanthin, a type I ribosome-inactivating protein.
T P Chow, R A Feldman, M Lovett and M Piatak

J. Biol. Chem. 1990, 265:8670-8674.

Access the most updated version of this article at http://www.jbc.org/content/265/15/8670

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/265/15/8670.full.html#ref-list-1