Purification, characterization and molecular cloning of trichoanguin, a novel type I ribosome-inactivating protein from the seeds of *Trichosanthes anguina*

Lu-Ping CHOW*, Ming-Huei CHOU*, Cheng-Ying HO*, Chyh-Chong CHUANG†, Fu-Ming PAN†, Shih-Hsiung WU† and Jung-Yaw LIN*[1]

*Institute of Biochemistry, College of Medicine, National Taiwan University, No. 1, Section 1, Jen-Ai Road, Taipei, Taiwan, Republic of China, and †Institute of Biological Chemistry, Academia Sinica, Nankang, Taipei, Taiwan, Republic of China

The seeds of the plant *Trichosanthes anguina* contain a type I ribosome-inactivating protein (RIP), designated trichoanguin, which was purified to apparent homogeneity by the combined use of ion-exchange chromatographies, i.e. first with DE-52 cellulose and then with CM-52 cellulose. The protein was found to be a glycoprotein with a molecular mass of 35 kDa and a pI of 9.1. It strongly inhibits the protein synthesis of rabbit reticulocyte lysate, with an IC_{50} of 0.08 nM, but only weakly that of HeLa cells, with an IC_{50} of 6 μM. Trichoanguin cleaves at the A4324 site of rat 28 S rRNA by its N-glycosidase activity. The cDNA of trichoanguin consists of 1039 nt and encodes an open reading frame coding for a polypeptide of 294 amino acid residues. The first 19 residues of this polypeptide encode a signal peptide sequence and the last 30 residues comprise an extension at its C-terminus. There are four potential glycosylation sites, located at Asn-51, Asn-65, Asn-201 and Asn-226. A comparison of the amino acid sequence of trichoanguin with those of RIPs such as trichosanthin, α-momorcharin, ricin A-chain and abrin A-chain reveals 55%, 48%, 36% and 34% identity respectively. Molecular homology modelling of trichoanguin indicates that its tertiary structure closely resembles those of trichosanthin and α-momorcharin. The large structural similarities might account for their common biological effects such as an abortifacient, an anti-tumour agent and anti-HIV-1 activities. Trichoanguin contains two cysteine residues, Cys-32 and Cys-155, with the former being likely to be located on the protein surface, which is directly amenable for conjugation with antibodies to form immunoconjugates. It is therefore conceivable that trichoanguin might be a better type I RIP than any other so far examined for the preparation of immunotoxins, with a great potential for application as an effective chemotherapeutic agent for the treatment of cancer.

Key words: inhibition of protein synthesis, N-glycosidases, ribosome-inactivating proteins, Cucurbitaceae.

INTRODUCTION

Ribosome-inactivating proteins (RIPs) are ubiquitous in the plant kingdom, with great abundance found in some plant families, such as the Cucurbitaceae [1]. RIPs inhibit protein synthesis by cleaving the N-glycosidic bond of adenine at position 4324 of rat liver 28 S rRNA and preventing the binding of elongation factor 2 [1]. The single adenine residue is removed from a highly conserved loop structure in the ribosomal RNA that render its 5′- and 3′-phosphodiester bonds very susceptible to acid-aniline cleavage and release the diagnostic RNA fragment of 420 nt [1]. RIPs are classified into two subgroups on the basis of their structures and functions: type I proteins consist of a single polypeptide chain of molecular masses ranging between 28 and 35 kDa and alkaline isoelectric points (pI) of pH 8–10 with or without carbohydrates [2]; type II RIPs consist of a catalytically active A chain linked to a cell-binding B chain. The B chain, possessing lectin properties, binds to the D-galactose moieties of the cell surface, leading to endocytosis and delivery of the A chain into the cell, where the latter can attack ribosomes enzymically [3,4]. Among type II RIPs, the cDNA species of ricin from *Ricinus communis* and abrin from *Abrus precatorius* have been cloned [5,6] and expressed in an *Escherichia coli* system [7,8].

Several type I RIPs have been purified and characterized from plants, e.g. pokeweed antiviral protein (PAP; *Phytolacca americana*) [9], momordin (*Momordica charantia*) [10], luffin (*Luffa cylindrica*) [11], bryodin (*Bryonia dioica*) [12] and dianthin (*Dianthus caryophyllus*) [13]. Trichosanthin (*Trichosanthes kirilowii*), α-momorcharin (*Momordia charantia*), saporin (*Saponaria officinalis*), PAP and bryodin have also been shown to have abortifacient activities [14].

Trichosanthin and α-momorcharin have been shown to be effective against T cells and macrophages infected with HIV-1 [15,16]. Clinical trials on trichosanthin also showed a decrease in the p24 antigen and an increase in CD4-positive cells in some patients [16,17]. There is great interest in their potential application as immunotoxins that can be selectively targeted to a particular cell type, such as cancer cells [18]. Here we describe the purification, characterization and molecular cloning of a new RIP from seeds of *T. anguina* in the Cucurbitaceae family, and study three-dimensional molecular models of it, based on the tertiary structure of trichosanthin and α-momorcharin.

EXPERIMENTAL

Materials

The seeds of snake gourds (*T. anguina*) were purchased from a local store. Restriction enzymes and T4 DNA ligase were obtained from New England Biolabs (Beverly, MA, U.S.A.). The

Abbreviations used: PAP, pokeweed antiviral protein; RACE, rapid amplification of cDNA ends; RIP, ribosome-inactivating protein.

[1] To whom correspondence should be addressed (e-mail lupin@ha.mc.ntu.edu.tw).

The nucleotide sequence data reported will appear in DDBJ, EMBL and GenBank Nucleotide Sequence Databases under the accession number AF055086.

Trizol kit for RNA extraction was obtained from Life Sciences (Petersburg, FL, U.S.A.). Oligo(dT)–cellulose was purchased from Pharmacia (Uppsala, Sweden). The Marathon™ cDNA amplification kit was from Clontech (Palo Alto, CA, U.S.A.). Deoxyribonucleotide primers were synthesized by the phosphoramide method with an Applied Biosystems (Foster City, CA, U.S.A.) automated DNA synthesizer. *Taq* DNA polymerase, *p*GEM-T vector, rabbit reticulocyte lysate and L-[^{3}H]leucine were obtained from Promega (Madison, WI, U.S.A.). The AmpliTag FS Prism Ready Reaction Cycle sequencing kit was from Applied Biosystems. Abrin A-chain was isolated and purified as described previously [6]. Other chemicals were of analytical grade.

Purification of trichoanguin

All purification procedures were performed at 4 °C. *T. anguina* seeds were homogenized with a Waring blender by using 10 mM sodium phosphate buffer, pH 7.2. A floating layer of solidified fat was removed with cheesecloth. After centrifugation of the suspension at 15000 ***g*** for 30 min, solid $(NH_4)_2SO_4$ was added to the supernatant to 95 % saturation. After being left for 1 h, the precipitates were collected by centrifugation, dissolved in 10 mM sodium phosphate buffer, pH 7.8, and dialysed against the same buffer. After dialysis the clear supernatant was applied to a DE-52 cellulose column (2.2 cm × 10 cm) pre-equilibrated with the buffer. The flow-through fractions were collected and applied to a CM-52 cellulose column (2.2 cm × 10 cm) pre-equilibrated with 10 mM sodium acetate buffer, pH 5.0. The column was eluted with a linear gradient of 0–0.4 M NaCl in the same buffer. Fractions with inhibitory activity towards protein synthesis were pooled, dialysed extensively against distilled water and freeze-dried. The purified fraction was analysed by SDS/PAGE.

Gel filtration

Gel filtration of trichoanguin was performed with a Superose 12 column HR 10/30 (Pharmacia), which was eluted with 50 mM sodium phosphate buffer, pH 6.8, containing 150 mM NaCl; the flow rate was 0.4 ml/min. The column was calibrated with the following molecular mass markers: BSA (67 kDa), ovalbumin (45 kDa) and chymotrypsinogen (25 kDa).

Electrophoresis

Active fractions isolated from each purification step were analysed by SDS/PAGE [12.5 % (w/v) gel], as described by Laemmli [19]. The protein bands were revealed by being stained with Coomassie Brilliant Blue R-250. Carbohydrate-containing bands were detected with periodic acid/Schiff reagent [20]. The pI of trichoanguin was estimated from isoelectric focusing electrophoresis performed with a pH 3.5–10 gel with a Pharmacia Multiphor II system.

Protein sequence analysis

The N-terminal amino acid sequence of trichoanguin was determined by using the automated Edman degradation method with an Applied Biosystems model 477 A protein sequencer and an on-line phenylthiohydantoin analyser. The C-terminal residues were determined by using the method of Kamo and Akira [21]. The reaction was performed by carboxypeptidase A digestion in 0.1 M pyridine/acetate/collidine buffer, pH 8.2, at 37 °C for 6 h, and the reaction products were dried and analysed directly with an A-5500 amino acid analyser (Irica, Kyoto, Japan).

Cell-free inhibition of protein synthesis

Assay of inhibition of protein synthesis *in vitro* was performed as described [22], with a cell-free rabbit reticulocyte lysate (Promega). Various amounts of toxin were added to the reaction mixture, and the reaction was performed at 30 °C for 1 h. The trichloroacetic acid-insoluble products were collected on a glass fibre disc by filtration with Whatman GF/C, then processed for liquid-scintillation counting. Each inhibition point is calculated as the mean for three individual tests.

Cytotoxicity assays

HeLa cells were grown in RPMI 1640 medium supplemented with 4 mM non-essential amino acids, streptomycin (100 i.u./ml), penicillin (100 μg/ml) and 10 % (v/v) fetal calf serum. Cells were plated in 24-well plates at a concentration of 10^5 cells per well and incubated at 37 °C under CO_2 for 24 h. The medium was then replaced by serum-free RPMI 1640 medium containing various amounts of toxin. Cells were further incubated at 37 °C for 18 h; protein synthesis was measured by incubating the cells for 1 h in serum-free, leucine-free RPMI 1640 containing 0.5 μCi/ml L-[^{3}H]leucine. The radioactivity incorporated into protein was determined as described previously [23]. Each point is the mean for triplicate assays.

RNA N-glycosidase activity

Rat liver ribosomes were prepared by the method of Wettstein et al. [24]; the ribosomes were incubated for 15 min with abrin A-chain or trichoanguin (10 nM) at 37 °C in a final volume of 100 μl of reaction buffer [113 mM KCl/10 mM $MgCl_2$/0.05 % (v/v) β-mercaptoethanol/2 units of RNasin]. The reaction was terminated by the addition of 0.5 % SDS; the reaction products were extracted with phenol, precipitated with alcohol and then treated with 0.8 M aniline to cleave 28 S rRNA selectively at the depurinated site by β-elimination. The reaction products were analysed by using 7 M urea/3.5 % (w/v) PAGE; the gels were stained with ethidium bromide [25].

Sequencing of trichoanguin cDNA

Total RNA was extracted from the maturing seeds of *T. anguina* in late summer with the Trizol reagent kit [26]. Poly(A)-rich RNA species were purified with a oligo(dT) column (Pharmacia). mRNA (1 μg) was reverse-transcribed with the Marathon™ cDNA amplification kit (Clontech), and the double-stranded cDNA species were ligated to Marathon™ cDNA adaptors. Two degenerate primers were synthesized based on the N-terminal and internal conserved sequences of trichoanguin: 5′ primer A, encoding the first eight residues (DVSFDLST), and 3′ primer B, encoding the highly conserved amino acids [EAARY(F)KYI] (Table 1), were used, and the reactions were subjected to 30 cycles of heat denaturation at 94 °C for 1 min, annealing the primers to the DNAs at 50 °C for 1 min, and DNA chain extension with *Taq* polymerase at 72 °C for 2 min, followed by a final extension at 72 °C for 10 min.

Rapid amplification of cDNA ends (RACE) on the 3′ end was performed with the Marathon™ cDNA amplification kit by using the flanking primer AP-1 and gene-specific primer C (Table 1). The 5′ end was amplified by 5′ RACE in essentially the same

Table 1 Oligonucleotide primers used for the isolation and cloning of trichoanguin cDNA

Technique	Primer	Sequence
Cloning	5′ Primer A	5′-GATGTTAGCTTCGATTTGTCGAC-3′ T C A A
	3′ Primer B	5′-ATATATTTATACCTTGCAGCTTC-3′ C GA T A T C C
RACE	Adaptor AP-1	5′-CCATCCTAATACGACTCACTATAGGGC-3′
	Primer C	5′-GCGCTTCTTGTACTCATTCAGTGT-3′
	Primer D	5′-ATCATAATTACCCGAATAAGGAAG-3′

manner as that for the 3′ end, by using the flanking primer AP-1 and the gene-specific primer D [27] (Table 1). All PCR products were subcloned into the *p*GEM-T vector (Promega), then transformed into *E. coli* strain JM109. DNA sequencing was performed with the Tag dye primer cycle sequencing kit (Perkin Elmer) and subjected to electrophoresis on a 373 A Stretch ABI DNA sequencer.

Molecular modelling of trichoanguin protein

A sequence search against the SCOP [28] database revealed the high degree of similarity of the trichoanguin protein sequence to those of RIPs. In the RIP superfamily, several three-dimensional structures have been solved by X-ray crystallography. Four sequences from RIPs, α-trichosanthin (PDB code 1TCS [29]), α-momorcharin (PDB code 1MGR [30]), abrin A chain (PDB code 1ABR [31]), and ricin A chain (PDB code 1RTC [32]), were obtained from the PDB for pairwise sequence alignment, and were compared with the trichoanguin sequence by using the GAP and BESTFIT programs of the GCG package.

The three-dimensional model of trichoanguin was built by using the X-ray structures of α-trichosanthin and α-momorcharin as templates. The multiple sequence alignments generated from the PILEUP program were checked manually and then used in the comparative homology modelling process because of their great similarity. Starting with the alignment, a method of automatic comparative modelling by means of satisfying spatial constraints as implemented in the MODELLER (version 4.0) program [33] was used to produce a trichoanguin model containing all main chains and side chain atoms without further manual intervention. First, MODELLER was used to derive many distance and dihedral angle restraints on the trichoanguin sequence from its alignment with the template RIP structures. Then the spatial restraints and CHARMM energy [34] terms enforcing the proper stereochemistry were combined into an objective function. The variable target function procedures, employing the methods of conjugate gradients and molecular dynamics with simulated annealing, were used to obtain three-dimensional models by optimizing the objective function. Twenty slightly different three-dimensional models of trichoanguin were calculated by varying the initial structure. The structure with the lowest value of the objective function was selected as the representative model. Assessment of the reliability of the model was performed residue by residue. The deviation from the standard geometry and atomic overlap was determined and evaluated more rigorously, residue by residue, with the PROCHECK program [35]. PROSAII [36] and Profile-3D [37,38] were used to test the suitability of the derived three-dimensional conformation for the amino acid sequence of trichoanguin and to develop an energetic profile of the modelled structure. Secondary structures of the trichoanguin model were calculated with the DSSP program [39]. Graphics were displayed by the InsightII package from MSI/Biosystem Technologies (San Diego, CA, U.S.A.) and the MOLMOL program [40].

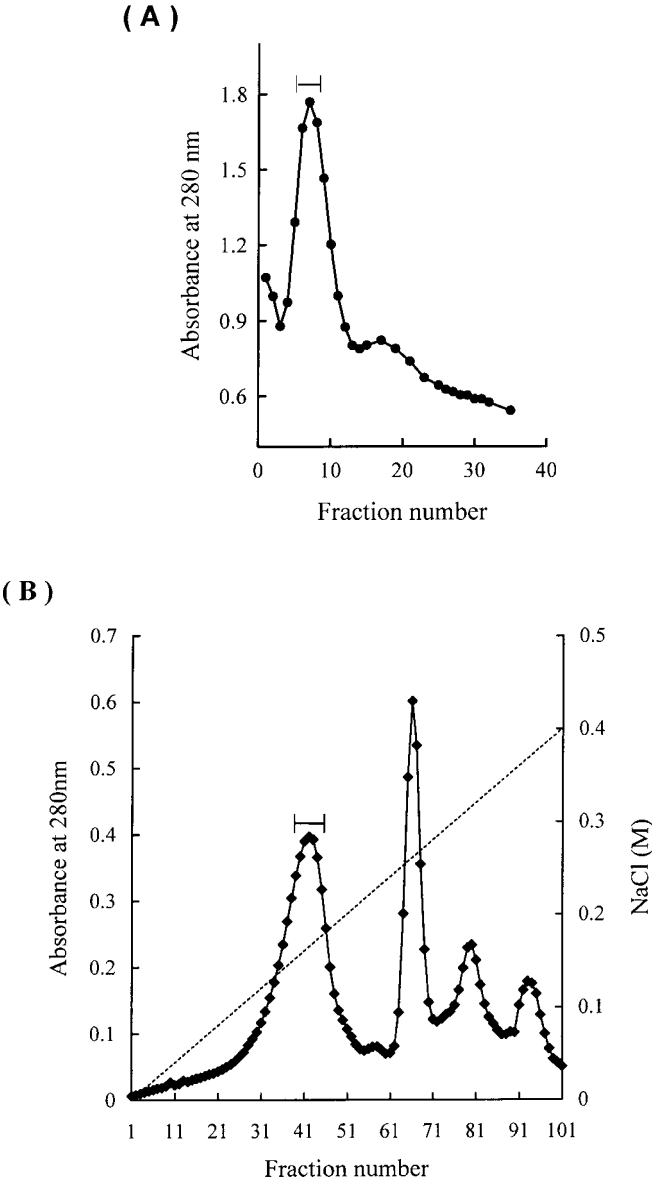

Figure 1 Purification of trichoanguin by ion-exchange chromatography

(**A**) Elution profile of trichoanguin from a DE-52 cellulose column. After precipitation by $(NH_4)_2SO_4$ and dialysis against equilibrium buffer (10 mM sodium phosphate buffer, pH 7.8), the supernatant of the dialysate was applied to a DE-52 cellulose column and eluted with equilibrium buffer. The A_{280} of each fraction was measured. Fractions showing activity in the inhibition of protein synthesis were pooled (as indicated by the horizontal bar). (**B**) Elution profile of trichoanguin from a CM-52 cellulose column. Fractions containing protein synthesis inhibitory activity from (**A**) were applied to a CM-cellulose column previously equilibrated with 10 mM sodium acetate buffer, pH 5.0. The protein was eluted with a linear gradient of 0–0.4 M NaCl with the same eluting buffer. The A_{280} of each fraction was measured. Fractions with inhibitory activity in a cell-free rabbit-reticulocyte system were collected (as indicated by the horizontal bar).

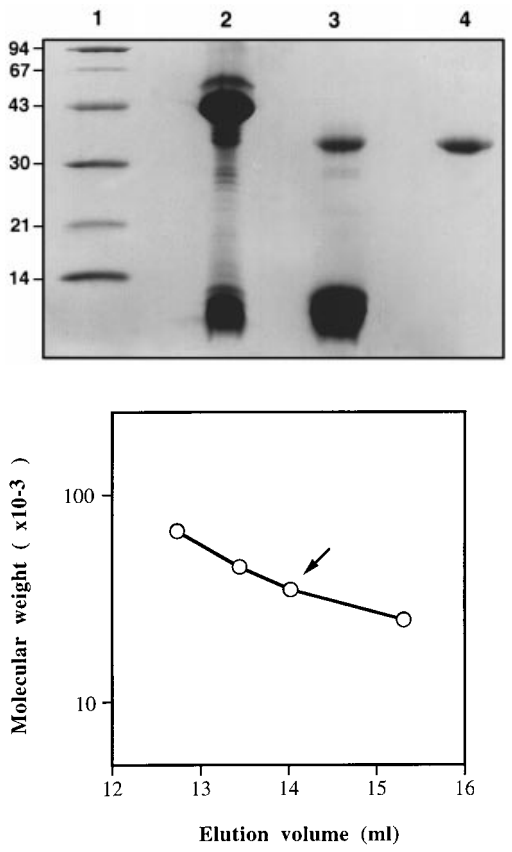

Figure 2 Molecular masses of the purified fractions

Upper panel: SDS/PAGE of various purified fractions. Lane 1, molecular mass markers; lane 2, crude extract; lane 3, active fractions from DE-52 cellulose column chromatography; lane 4, active fractions from CM-52 cellulose column chromatography. The gel was stained with Coomassie Brilliant Blue R-250. The positions of molecular mass standards (in kDa) are shown at the left. Lower panel: estimation of apparent molecular mass of native protein by gel filtration. Approx. 0.2 mg of trichoanguin was subjected to gel filtration on a Superose 12 column (Pharmacia) equilibrated and eluted at a flow rate of 0.4 ml/min with 50 mM sodium phosphate buffer, pH 6.8, containing 150 mM NaCl. Each collected fraction was 0.2 ml. The column was calibrated with the following molecular mass standards: BSA (67 kDa), ovalbumin (45 kDa), and chymotrypsinogen (25 kDa). The elution of the proteins was monitored by measuring A_{280}. The logarithm of protein molecular mass was plotted against elution volume. The arrow indicates the elution volume of trichoanguin.

Table 2 Purification of trichoanguin from *T. anguina* seeds

The preparation was from 250 g of *T. anguina* seeds as described in the Experimental section. One unit is defined as the the amount of enzyme necessary to inhibit protein synthesis by 50% in 1 ml of rabbit reticulocyte lysate reaction mixture.

Preparation	Total protein (mg)	Total activity (10^6 units)	Specific activity (10^3 units/mg)	Yield (%)
Crude extract	2851	81	29	100
DE-52 cellulose	315	45	143	55
CM-52 cellulose	31	14.8	477	18

RESULTS

Isolation of trichoanguin from *T. anguina*

Trichoanguin was purified to homogeneity from seeds of *T. anguina* by two simple steps. One major peak obtained in the first step of DE-52 cellulose chromatography exhibited the protein synthesis inhibitory activity (Figure 1A). In this fraction, a major protein of 35 kDa and a minor protein of 10 kDa were present, as revealed by SDS/PAGE analysis. The 10 kDa protein was removed in the second step by CM-52 cellulose chromatography (Figure 1B), resulting in a protein fraction (the first peak) with a subunit molecular mass of 35 kDa and apparent homogeneity (Figure 2, upper panel). The protein thus obtained was characterized and designated 'trichoanguin'. The fraction containing purified trichoanguin was analysed further by size-exclusion chromatography. Trichoanguin was eluted as a single chain of 35 kDa under non-denaturing conditions (Figure 2, lower panel). A pI of 9.1 was estimated with the Pharmacia Multiphor II system (results not shown), in agreement with the alkaline property of most type I RIPs. The inhibitory activity and relative yields in the various steps of purification are summarized in Table 2. Trichoanguin showed a positive reaction to staining with periodic acid/Schiff reagent (results not shown), indicating its glycoprotein nature. Trichoanguin strongly inhibited the incorporation of labelled amino acids into the rabbit reticulocyte lysate cell-free system, and the degree of inhibition was com-

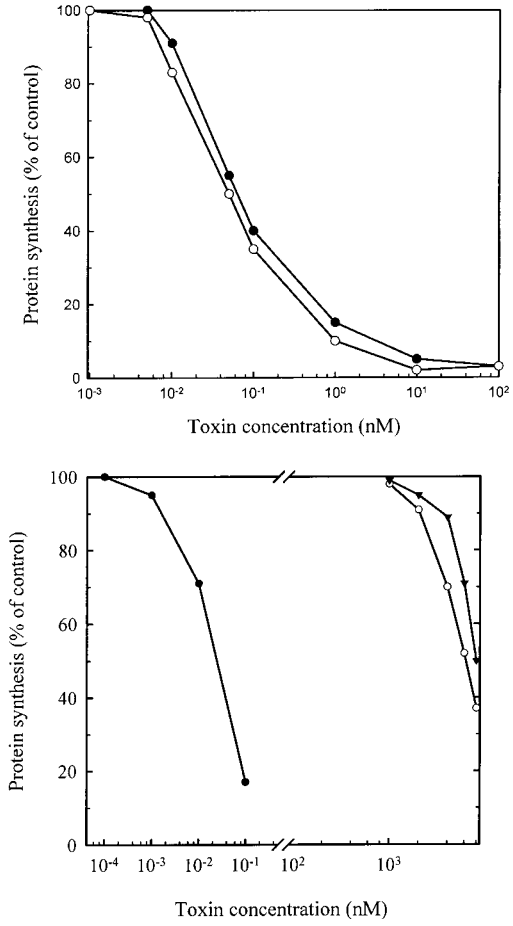

Figure 3 Effects of trichoanguin on protein synthesis

Upper panel: inhibitory effect of varying the concentration of trichoanguin on protein synthesis in a cell-free rabbit reticulocyte lysate system. The incorporation of radioactivity in L-[^{3}H]leucine into protein was measured; each point denotes the mean for triplicate assays. Symbols: ●, trichoanguin; ○, abrin A chain. Lower panel: inhibition of protein synthesis in HeLa cells. Various concentrations of toxin were incubated with HeLa cells and the incorporation of L-[^{3}H]leucine into cellular proteins was subsequently determined. Each value is the mean for triplicate samples. Symbols: ○, trichoanguin; ▼, abrin A chain; ●, abrin.

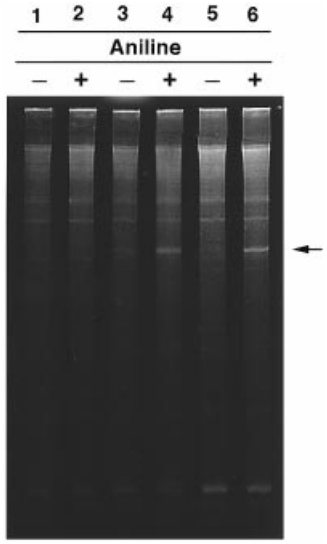

Figure 4 Analysis by gel electrophoresis of N-glycosidase activity

Rat liver ribosomes were used in each assay. Extracted RNA from those ribosomes treated with aniline (+) or untreated (−) were separated by electrophoresis on a denaturing polyacrylamide gel. Lanes 1 and 2, samples in the absence of toxin control; lanes 3 and 4, samples treated with 10 nM trichoanguin; lanes 5 and 6, samples treated with 10 nM abrin A chain. The arrow indicates the position of the RNA fragment generated by aniline treatment of the ribosomal RNA.

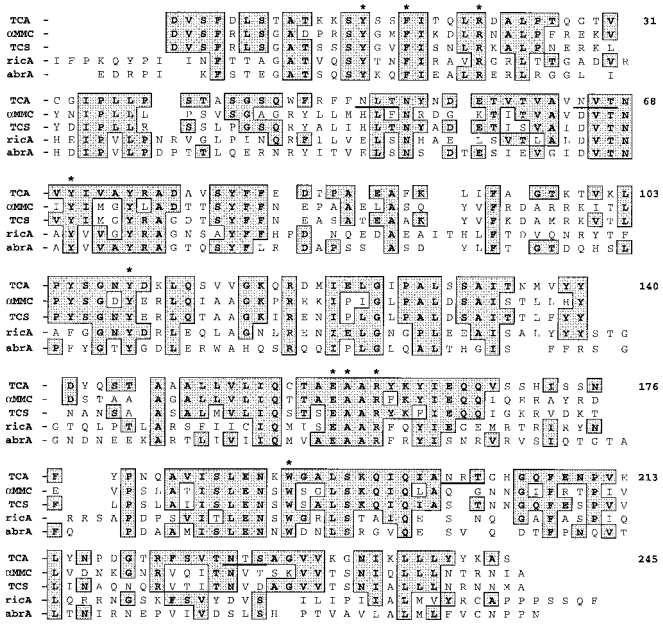

Figure 6 Sequence comparison of trichoanguin (TCA) and other RIPs

Sequence alignment was performed with the PILEUP program of GCG by using published sequences of α-momorcharin (αMMC) [30], trichosanthin (TCS) [29], ricin A chain (ricA) [32] and abrin A chain (abrA) [31]. The numbering system is based on the sequence of trichosanthin as reference. Gaps have been introduced for optimal alignment and maximum similarity between all compared sequences. Identical amino acids are shown in shaded boxes. The highly conserved and consensus amino acid residues involved in the active site are indicated by asterisks. Potential N-glycosylation sites are underlined.

parable with that of abrin. As shown in Figure 3 (upper panel), the IC_{50} values of trichoanguin and abrin A-chain were determined as 0.8 and 0.06 nM respectively.

Cytotoxicity

The addition of trichoanguin to cultures of HeLa cells resulted in the weak inhibition of protein synthesis, with an IC_{50} of 6 μM (Figure 3, lower panel). At relatively high concentrations, trichoanguin causes a moderate decrease in protein synthesis. Abrin, a type II RIP, had a very strong toxicity towards HeLa cells, with an IC_{50} of 0.05 nM, but abrin A-chain exhibited a weak toxicity towards the intact HeLa cells with an IC_{50} of 7 μM, similar to that of trichoanguin.

```
1              M  A  L  S  F  F  F  L  A  I  S  L  G  S  P  T  A  I  G  D  V  S   22
1    c taa ttc ATG GCA CTC TCC TTT TTC TTT CTC GCC ATC TCT CTT GGC TCT CCT ACT GCC ATT GGT GAT GTT AGC   73

23  F  D  L  S  T  A  T  K  K  S  Y  S  S  F  I  T  Q  L  R  D  A  L  P  T  Q   47
74  TTC GAT TTG TCG ACA GCT ACT AAA AAA TCC TAT TCA TCT TTC ATC ACA CAA CTC AGG GAT GCT CTT CCA ACT CAA  148

48  G  T  V  C  G  I  P  L  L  P  S  T  A  S  G  S  Q  W  F  R  F  F  N  L  T   72
149 GGC ACA GTG TGC GGC ATT CCA TTG CTG CCT TCC ACC GCA TCC GGC TCA CAA TGG TTC AGA TTC TTC AAT CTT ACC  223

73  N  Y  N  D  E  T  V  T  V  A  V  N  V  T  N  V  Y  I  V  A  Y  R  A  D  A   97
224 AAT TAT AAC GAT GAA ACC GTC ACC GTG GCT GTA AAT GTA ACC AAT GTC TAC ATC GTG GCG TAT CGT GCC GAT GCT  298

98  V  S  Y  F  F  E  D  T  P  A  E  A  F  K  L  I  F  A  G  T  K  T  V  K  L  122
299 GTA TCC TAC TTT TTT GAA GAC ACT CCA GCT GAA GCT TTC AAG CTC ATA TTT GCA GGT ACT AAG ACG GTA AAA CTT  373

123 P  Y  S  G  N  Y  D  K  L  Q  S  V  V  G  K  Q  R  D  M  I  E  L  G  I  P  147
374 CCT TAT TCG GGT AAT TAT GAT AAG CTT CAA AGT GTA GTA GGC AAA CAA AGA GAT ATG ATT GAG CTT GGA ATC CCG  448

148 A  L  S  S  A  I  T  N  M  V  Y  Y  D  Y  Q  S  T  A  A  A  L  L  V  L  I  172
449 GCT TTA AGC AGT GCC ATT ACG AAC ATG GTT TAT TAC GAC TAC CAA AGT ACT GCA GCC GCG CTT CTT GTA CTC ATT  523

173 Q  C  T  A  E  A  A  R  Y  K  Y  I  E  Q  Q  V  S  S  H  I  S  S  N  F  Y  197
524 CAG TGT ACT GCA GAA GCT GCA AGG TAT AAA TAT ATT GAG CAA CAA GTT TCT TCA CAT ATT AGC TCT AAT TTT TAT  598

198 P  N  Q  A  V  I  S  L  E  N  K  W  G  A  L  S  K  Q  I  Q  I  A  N  R  T  222
599 CCA AAT CAA GCA GTC ATA AGC TTA GAA AAC AAG TGG GGT GCT CTT TCC AAA CAA ATC CAG ATA GCA AAT AGG ACG  673

223 G  H  G  Q  F  E  N  P  V  E  L  Y  N  P  D  G  T  R  F  S  V  T  N  T  S  247
674 GGA CAT GGA CAA TTC GAA AAC CCT GTT GAG CTA TAT AAC CCT GAT GGC ACA CGA TTT AGT GTA ACC AAT ACT TCG  748

248 A  G  V  V  K  G  N  I  K  L  L  L  Y  Y  K  A  S  V  G  S  E  Y  D  I  P  272
749 GCT GGA GTT GTA AAA GGC AAT ATC AAA CTC CTA CTA TAC TAC AAA GCC AGT GTT GGT AGT GAA TAT GAT ATC CCT  823

273 T  T  I  L  H  P  G  A  M  G  M  L  H  N  Q  N  G  N  Y  V  T  M  STOP       294
824 ACT ACA ATC TTG CAC CCT GGA GCT ATG GGA ATG CTT CAC AAT CAG AAT GGA AAT TAT GTT ACA ATG TGA ctt tta  898

899 att ctc aca gct ctg cct tcg gtt caa cac cca cgc ata gat tta agg cat tgt tca tag tca ccc atg ttg tga  973

974 atg tgc ttt att tca tta aat aaa taa gtg tgg aat ctt ttc cgc caa aaa aaa aaa aaa aaa aaa          1039
```

Figure 5 Nucleotide and amino acid sequences encoding trichoanguin

The nucleotides and the predicted amino acid residues are numbered at the right and at the left. The mature protein sequence is numbered from 20 to 264, the signal peptide runs from 1 to 19, and the additional extension region at the C-terminus extends from 265 to 294. The regions encoding the polyadenylation signal (AATAAA) and the poly(A) tail are underlined.

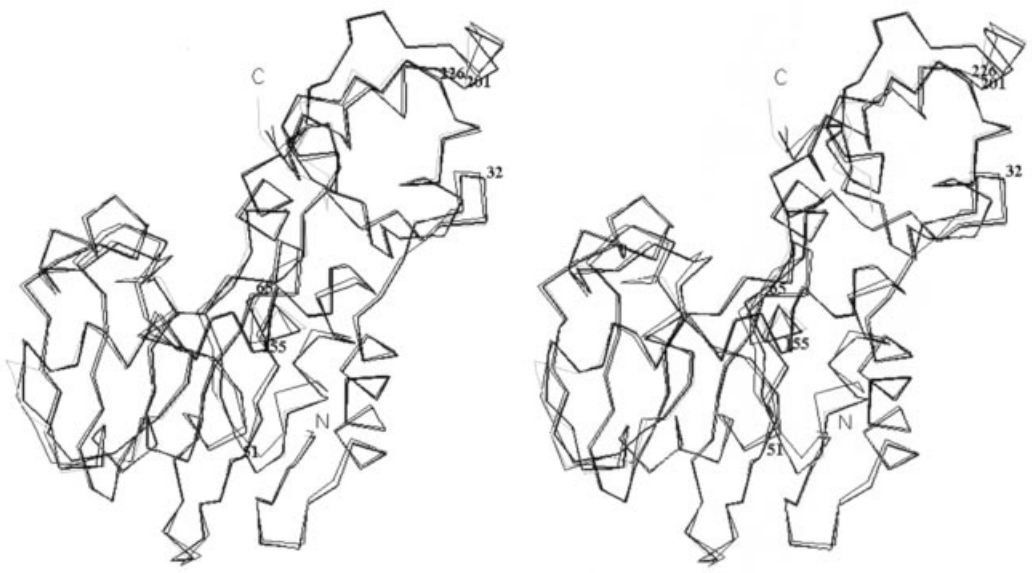

Figure 7 Superposition of three-dimensional models for homologous proteins

Comparison of trichoanguin (thick line) and its closest structurally known homologous RIPs, α-momorcharin (medium line) and trichosanthin (thin line).

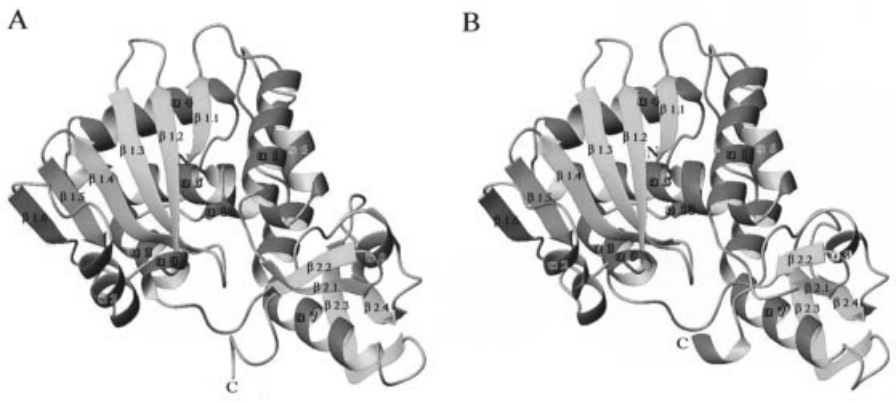

Figure 8 Ribbon representations of trichosanthin (A) and trichoanguin (B) molecules

The N-terminus and C-terminus of each secondary structure region are labelled. The helices are labelled α1 to α8 along the chain; the β-sheets are labelled β1.1 to β2.2. The computer program MOLMOL was used to generate these diagram.

N-glycosidase activity

The N-glycosidase activity of trichoanguin was examined by incubating ribosomes with various amounts of trichoanguin or abrin A-chain, and the extracted rRNA was analysed by gel electrophoresis. As shown in Figure 4, when the rRNA from trichoanguin-treated ribosomes was treated with aniline at acidic pH, a cleaved fragment of approx. 420 nt was obtained, similar to that found in abrin A chain/aniline-treated ribosomes.

Molecular cloning and sequence analysis of trichoanguin

PCR amplification of total cDNA mixtures prepared from seeds of *T. anguina* with the Marathon™ cDNA amplification protocol coupled with designed primers for 5′ RACE and 3′ RACE achieved the amplification of a full-length cDNA fragment of approx. 1000–1100 nt encoding trichoanguin. Sequence analysis of the cDNA clone revealed that it is 1039 nt in length, a cDNA sequence containing an open reading frame of 882 nt, corresponding to a polypeptide of 294 residues with a calculated molecular mass of 27066 Da (Figure 5).

In most cases, type I RIPs are cleaved post-translationally to yield the mature form. The deduced polypeptide chain of trichoanguin contains a segment of 19 residues at the N-terminus coding for a signal peptide. The C-terminal residues were identified as Ala-Ser by carboxypeptidase A digestion. The cDNA sequence also contained a 30-residue extension at the C-terminus, followed by a translation termination codon (TGA) and a 105 nt 3′-untranslated sequence with an AATAAA polyadenylation site.

There are four potential glycosylation sites at residues 51, 65, 201 and 226 (Figure 6); the presence of carbohydrate might account for the discrepancy observed between the actual coding

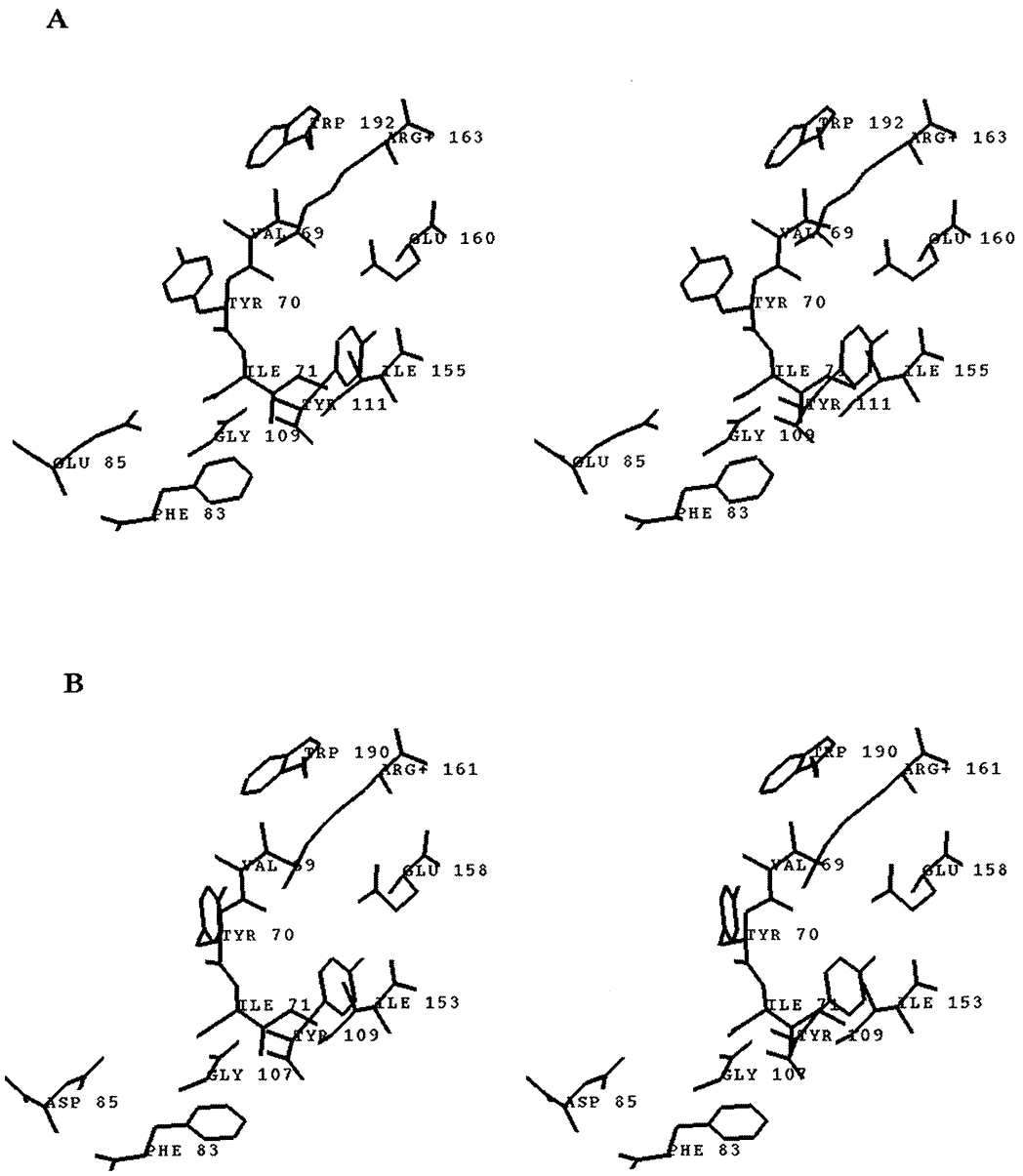

Figure 9 Stereo diagrams of the active site in trichosanthin (A) and trichoanguin (B)

The diagrams show the amino acid residues involved in the binding of adenine and the residues associated with the surroundings of the active site.

molecular mass and that estimated by SDS/PAGE. Unlike trichosanthin and α-momorcharin, trichoanguin contained two free cysteine residues, one located at residue 32 and the other at residue 155.

Sequence similarity and comparison between various RIPs

A comparison of the amino acid sequence of trichoanguin with sequences of trichosanthin and other RIPs (Figure 6) revealed that there is a high degree of similarity between them. A number of gaps and insertions have been made in the sequences to optimize the alignment, and the percentages of identity between trichoanguin and trichoanthin, α-momorcharin, ricin A-chain and abrin A-chain were found to be 55%, 48%, 36% and 34% respectively. There are several major regions of high sequence similarity in these RIPs, i.e. regions of matching amino acids: residues 4–25, 66–85, 104–134, 146–166 and 182–190 of trichoanguin, with the last four regions clustering around the putative active-site cleft. In these regions, absolutely conserved amino acids are: Tyr-14, Phe-17, Arg-22, Tyr-70, Gly-107, Tyr-109, Ala-146, Glu-158, Ala-159, Arg-161, Glu-187, Asn-188 and Trp-190 in trichosanthin. It is interesting to note that the conserved residues Tyr-70, Tyr-109, Glu-158, Ala-159, Arg-161, Glu-187 and Trp-190 are clustered together around the proposed active-site cleft in the three-dimensional crystal structure of trichosanthin. The aromatic amino acids (Tyr-14, Phe-17, Tyr-70, Tyr-109 and Trp-190) are highly conserved in RIPs, and these aromatic residues are important in stabilizing the interactions between the bases of RNA and trichoanguin, which are involved in the N-glycosidase activity.

Molecular modelling of trichoanguin

The trichoanguin sequence was modelled with the coordinates of trichosanthin and α-momorcharin because these two proteins exhibited the highest sequence identity (55% and 48%) in the sequence comparison with trichoanguin [30]. Figure 7 shows that the α-carbon backbones of trichoanguin and α-momorcharin can be superimposed on the backbone of trichosanthin. The root-mean-square difference for aligned α-carbon positions between trichosanthin and trichoanguin was 0.520 Å, whereas that between α-momorcharin and trichosanthin was 0.502 Å. However, positional differences between the individual side chain positions might be substantially larger, particularly for the less constrained residues on the molecular surface. A schematic ribbon drawing of the known structure of trichosanthin is shown in Figure 8. To facilitate the discussion of the modelling of trichoanguin in relation to this structure, the secondary structural elements are numbered as described previously [30]. The model contains eight α-helices and a six-stranded β-sheet with a left-handed twist similar to that found in trichosanthin.

In Figure 8, trichoanguin was divided into two domains in accordance with the structural description of trichosanthin [30]. The main differences between trichoanguin and trichosanthin in domain 1 are located in the middle portion and in the loops connecting the secondary structural elements. Internally, there are differences at two residues between the two: deletions of residues 89 and 98 of trichosanthin. The first deletion removes one residue from helix 2 in trichosanthin. Because this helix is located on the molecular surface, the deletion is easily accommodated in trichoanguin. The second deletion shortens a surface loop connecting helix 2 and β-sheet 1.6 by the four-residue loop. It implies that this region might not be specifically crucial to the structure or function of trichoanguin. Trichoanguin has an insertion of one residue, Arg-202, in a loop connecting helix 7 and β-sheet 2.1. The antiparallel β-sheets (β2.1 and β2.2) of trichoanguin in C-terminal regions differ slightly from those in trichosanthin. The C-terminal part of trichoanguin is one residue shorter than that of trichosanthin and is predicted to be a 3_{10} helix.

Trichoanguin contains two free thiol groups: one, Cys-32, is located at the surface loop region; the other, Cys-155, located adjacent to the active site, seems to interfere with disulphide linkage formation. The four putative N-glycosylation sites Asn-51, Asn-65, Asn-201 and Asn-226 are located at the solvent-exposed surface or flexible loop in the modelled structure.

It has been suggested that amino acid residues lining the active site cleft are generally conserved within the RIP family, which might be important for substrate binding and catalysis (Figure 9). Eleven residues were found to be highly conserved in trichosanthin and trichoanguin; five of them, Tyr-70, Tyr-109, Glu-158, Arg-161 and Trp-190, directly form the major cleft in the crystal structure of trichosanthin, whereas the others, including Val-69, Ile-71, Phe-83, Asp-85, Gly-107 and Ile-153, are also located at the active-site cleft and are highly conserved between various RIPs.

DISCUSSION

In the present study it was found that the 19 residues at the N-terminal extension and the extra 30 residues at the C-terminal end of trichoanguin are removed post-translationally to yield the mature form. The 19-residue leader segment is a secretory signal sequence containing a higher content of hydrophobic amino acids, which is expected to direct transport of the nascent polypeptide chain across the endoplasmic reticulum membrane into the endoplasmic reticulum lumen [41]. Similar post-translational processing mechanisms of a C-terminal extension for these RIPs were recently observed for the precursors of trichosanthin [42] and saporin-6 [43]. Four putative N-glycosylation sites, Asn-51, Asn-65, Asn-201 and Asn-226 (Figure 6), occur along the amino acid sequence of trichoanguin. In the modelled structure, all of these are located on the exposed surface of trichoanguin and are thus expected to be glycosylated (Figure 7).

It is noteworthy that modelling studies of these proteins have allowed us to visualize the prominent cleft, which has been suggested to comprise the active site of various RIPs [30]. The presence of conserved residues of similar amino acids around the proposed active-site cleft among trichoanguin, trichosanthin and α-momorcharin was clearly identified and confirmed. This similarity strengthens the notion that there could be a strong preservation of three-dimensional structure in these proteins with similar catalytic functions, with critical amino acid residues being conserved especially in the region of the active site.

Figure 9 shows a close-up view of the active centres of trichoanguin and trichosanthin. The residues constituting the active site of trichosanthin (Tyr-70, Tyr-111, Glu-160, Arg-163 and Trp-192) are fully conserved in trichoanguin. In trichosanthin, the key active-site residues, including Glu-160 and Arg-163, are directly involved in catalysis, whereas Tyr-70 and Tyr-111 have a crucial role in binding the rRNA loop. Trichoanguin possesses the same residues at its catalytic site and the rRNA loop-binding site. Most RIPs contain an acidic amino acid residue at position 85, which might provide a proton for protonating adenine. In trichosanthin, the N-7 atom of the adenine is protonated by Glu-85, which is replaced by Asp in trichoanguin and the abrin A-chain.

Many immunoconjugates of RIPs and specific antibodies have been evaluated *in vitro* and *in vivo* as potential therapeutic agents for the treatment of cancer and autoimmune diseases. When trichosanthin was conjugated to a hepatoma-associated antibody, the resultant immunotoxin was 500-fold more cytotoxic than free trichosanthin and only one order of magnitude less cytotoxic than free ricin [44]. By linking monoclonal anti-Thy1.1 antibodies to PAP or ricin A-chain through a disulphide bond, both conjugates were shown to specifically inhibit protein synthesis of Thy1.1-positive target leukaemic cells [45]. Cross-linking of saporin to an anti-CD4 antibody leads to effective killing of $CD4^+$ cells [46]. Bryodin conjugated with anti-CD40 antibody was shown to be potently cytotoxic against CD40-expressing B-lineage non-Hodgkin's lymphoma and multiple myeloma cells [47]. Furthermore a conjugate of PAP with the anti-CD4 antibody was found to be very effective in inhibiting HIV-1 production [48].

From the above examples of immunotoxins applied to therapeutic uses, an important consideration for immunoconjugate assembly is the nature of the linkage between antibody and RIP. A disulphide linkage is usually thought to be essential for maximal cytotoxicity. Most type 1 RIPs do not have any free cysteine residues, which necessitates the modification of both antibody and RIP with chemical agents to produce the disulphide bond. Fortunately, trichoanguin contains two cysteine residues, one of which is located at the surface loop and can directly form a disulphide bond with an activated antibody thiol group via a disulphide-exchange reaction. Therefore trichoanguin is a novel free-cysteine-containing RIP, which might be ideal for the preparation of immunoconjugates with great potential as a chemotherapeutic agent for the treatment of various cancers or AIDS.

We thank Professor S.-H. Chiou (Institute of Biochemical Sciences, National Taiwan University, Taipei, Taiwan) for critical reading and comments on the manuscript. This

work was supported by grant NSC-87-2314-B-002-026 (to L.-P.C.) from the National Science Council, Taipei, Taiwan.

REFERENCES

1 Stirpe, F. and Barbieri, L. (1986) FEBS Lett. **195**, 1–8
2 Stirpe, F., Gasperi-Campani, A., Barnieri, L., Falasca, A., Abbondanza, A. and Stevens, W. A. (1983) Biochem. J. **216**, 617–625
3 Barbieri, L. and Stirpe, F. (1982) Cancer Surv. **1**, 489–520
4 Stirpe, F., Barbieri, L., Battelli, M. G., Soria, M. and Lappi, D. A. (1992) Bio/Technology **10**, 405–412
5 Lamb, F. I., Roberts, L. M. and Lord, J. M. (1985) Eur. J. Biochem. **148**, 265–270
6 Hung, C. H., Lee, M. C., Lee, T. C. and Lin, J. Y. (1993) J. Mol. Biol. **229**, 263–267
7 Piatak, M., Lane, J. A., Laird, W., Bjorn, M. J., Wang, A. and Williams, M. (1988) J. Biol. Chem. **263**, 4837–4843
8 Hung, C. H., Lee, M. C., Chen, J. K. and Lin, J. Y. (1994) Eur. J. Biochem. **219**, 83–87
9 Irvin, J. D. (1975) Arch. Biochem. Biophys. **169**, 522–528
10 Barbieri, L., Zamboni, M., Lorenzoni, E., Montanaro, L., Sperti, S. and Stirpe, F. (1980) Biochem. J. **186**, 443–452
11 Kishida, K., Masuho, Y. and Hara, T. (1983) FEBS Lett. **153**, 209–212
12 Stirpe, F., Barbieri, L., Battelli, M. G., Falasca, A. I., Abbondanza, A., Lorenzoni, E. and Stevens, W. A. (1986) Biochem. J. **240**, 659–665
13 Stirpe, F., Williams, D. G., Onyon, L. J., Legg, R. F. and Stevens, W. A. (1981) Biochem. J. **195**, 399–405
14 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. et al. (1989) Proc. Natl. Acad. Sci. U.S.A. **86**, 2844–2848
15 Huang, S. L., Huang, P. L., Nara, P. L., Chen, H. C., Kung, H. F., Huang, P., Huang, H. I. and Huang, P. L. (1990) FEBS Lett. **272**, 12–18
16 Byers, V. S., Levin, A. S., Malvino, A., Waites, L., Robins, R. A. and Baldwin, R. W. (1994) AIDS Res. Hum. Retroviruses **10**, 413–420
17 Kahn, J. O., Gorelick, K. J., Gatti, G., Arri, C. J., Lifson, J. D., Gambertoglio, J. G., Bostrom, A. and Williams, R. (1994) Antimicrob. Agents Chemother. **38**, 260–267
18 Lambert, J. M., Blattler, W. A., McIntyre, G. D., Goldmacher, V. S. and Scott, Jr., C. F. (1988) Cancer Treat. Res. **37**, 175–209
19 Laemmli, U. K. (1970) Nature (London) **227**, 680–685
20 Zacharius, R. M., Zell, T. E., Morrison, J. H. and Woodlock, J. J. (1969) Anal. Biochem. **30**, 148–152
21 Kamo, M. and Akira, T. (1987) J. Biochem. (Tokyo) **102**, 243–246

Received 7 September 1998/20 October 1998; accepted 30 November 1998

22 Thorpe, P. E., Brown, A. N. F., Ross, W. C. J., Cumber, A. J., Detre, S. I., Edwards, D. C., Davies, A. J. S. and Stirpe, F. (1981) Eur. J. Biochem. **116**, 447–454
23 Oda, T., Aizono, Y. and Funatsu, G. (1984) J. Biochem. (Tokyo) **96**, 377–384
24 Wettstein, F. O., Staehelin, T. and Noll, H. (1963) Nature (London) **197**, 430–435
25 May, M. J., Hartley, M. R., Roberts, L. M., Krieg, P. A., Osborn, R. W. and Lord, J. M. (1989) EMBO J. **8**, 301–308
26 Chomczynski, P. and Sacchi, N. (1987) Anal. Biochem. **162**, 156–159
27 Frohman, M. A., Dush, M. K. and Martin, G. R. (1988) Proc. Natl. Acad. Sci. U.S.A. **85**, 8998–9002
28 Hubbard, T. J. P., Murin, A. G., Brenner, S. E. and Chothia, C. (1997) Nucleic Acids Res. **25**, 236–239
29 Xiong, J. P., Xia, Z. X. and Wang, Y. (1994) Nat. Struct. Biol. **1**, 695–700
30 Huang, Q. C., Liu, S. P., Jin, Y. Q. and Wang, Y. (1995) Biochem. J. **309**, 285–298
31 Tahirov, T. H., Lu, T. H., Liaw, Y. C., Chen, Y. L. and Lin, J. Y. (1995) J. Mol. Biol. **250**, 354–367
32 MIsna, D., Monzingo, A. F., Katzin, B. J., Ernst, S. and Robertus, J. D. (1993) Protein Sci. **2**, 429–435
33 Sali, A. and Blundell, T. L. (1993) J. Mol. Biol. **234**, 779–815
34 Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. and Karplus, M. (1983) J. Comp. Chem. **4**, 187–217
33 Laskowski, R. A., MacArther, M. W., Moss, D. J. and Thornton, J. M. (1993) J. Appl. Crystallogr. **26**, 283–291
36 Sippl, M. J. (1993) Proteins **17**, 355–362
37 Bowie, J. U., Luthy, R. and Eisenberg, D. (1991) Science **253**, 164–170
38 Luthy, R., Bowie, J. U. and Eisenberg, D. (1992) Nature (London) **356**, 83–85
39 Kabsch, W. and Sander, C. (1983) Biopolymers **22**, 2577–2637
40 Koradi, R., Billeter, M. and Wuthrich, K. (1996) J. Mol. Graphics **14**, 51–55
41 Butterworth, A. G. and Lord, J. M. (1983) Eur. J. Biochem. **137**, 57–65
42 Chow, T. P., Feldman, R. A., Lovett, M. and Piatak, M. (1990) J. Biol. Chem. **265**, 8670–8674
43 Benatti, L., Nitti, G., Solinas, M., Valsasina, B., Vitale, A., Ceriotti, A. and Soria, M. R. (1991) FEBS Lett. **291**, 285–288
44 Wang, Q. C., Ying, W. B., Xie, H., Zhang, Z. C., Yang, Z. H. and Ling, L. Q. (1991) Cancer Res. **51**, 3353–3355
45 Ramakrishnan, S. and Houston, L. L. (1984) Cancer Res. **44**, 1389–1404
46 Ramakrishnan, S., Fryxell, D., Mohanraj, D., Olson, M. and Li, B. Y. (1992) Annu. Rev. Pharmacol. Toxicol. **32**, 579–621
47 Francisco, J. A., Gawlak, S. L. and Siegall, C. B. (1997) J. Biol. Chem. **272**, 24165–24169
48 Zarling, J. M., Moran, P. A., Haffar, O., Sias, J., Richman, D. D., Spina, C. A., Meyers, D. E., Kuebelbeck, V., Ledbetter, J. A. and Uckun, F. M. (1990) Nature (London) **347**, 92–95