Prediction of antigenic determinants of trichosanthin by molecular modeling

HE YONG NING* ZONG XIANG XIA*¹, YIN WANG**, YONG YONG JI**, MING YEH** * Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, Shanghai 200032, China ** Shanghai Institute of Cell Biology, Chinese Academy of Sciences, Shanghai 200031, China

ABSTRACT

The antigenic determinants of trichosanthin were predicted by molecular modeling. First, the threedimensional structure model of the antigen-binding fragment of anti-trichosanthin immunoglobulin E was built on the basis of its amino-acid sequence and the known three-dimensional structure of an antibody with similar sequence. Secondly, the preferable antigen-antibody interactions were obtained based on the known three-dimensional structure of trichosanthin and of the hypervariable regions of anti-trichosanthin immunoglobulin E. Two regions in the molecular surface of trichosanthin were found to form extensive interactions with the hypervariable regions of the antibody and have been predicted to be the possible antigenic determinants: one is composed of two polypeptide segments, Ile201-Glu210 and Ile225-Asp229, which are close to each other in the three-dimensional structure; and the other is the segment Lys173-Thr178. The former region seems to be the more reasonable antigenic determinant than the latter one.

Key words: Antigenic determinants, trichosanthin, anti-trichosanthin immunoglobulin E, molecular modeling.

^{1.} Corresponding author.

INTRODUCTION

Trichosanthin (TCS) is a single-chained ribosome-inactivating protein. It has been used as a traditional Chinese medicine for inducing abortion and the treatment of gynaecological diseases[1, 2]. In recent years, TCS was found to possess anti-HIV activities[3] and has been brought into clinical practice in the U.S.A.. However, TCS is a toxic protein, and the therapeutic effect accompanied with allergic reactions, such as headache and fever, for some patients. Therefore, the prediction of the antigenic determinants of TCS is of importance for the protein engineering in order to reduce the side effects.

The crystallization and three-dimensional structure determination of the complex of TCS with the anti-TCS antibody would be the straightforward method to determine the antigenic determinants of TCS. Immunoglobulin E (IgE) is a class of antibody directly involved in allergic reactions in human body, but it exists in serum at very low concentration. At present stage, there is difficulty in growing the crystals of TCS-IgE complex. In this paper we report the prediction of the antigenic determinants of TCS by molecular modeling.

One of the authors' groups has determined the monoclinic crystal structure of TCS at 2.7 Å resolution[4, 5], which consists of a large N-terminal domain and a small C-terminal domain containing 181 and 66 amino-acid residues, respectively.

Other group of the authors has succeeded in the cDNA cloning and the gene sequence determination of both the light chains and the heavy chains of five anti-TCS mouse monoclonal IgEs from different fussions[6]. It has been found that all the five IgEs and one monoclonal IgG1 recognize the same site of TCS[7], suggesting the existence of an allergenic determinant in the surface region of TCS molecule. The amino-acid sequences of both chains have been derived from the gene sequences.

The known structures of all antibodies are very similar to each other. The immunoglobulin is a Y-shaped molecule with two Fab fragment arms and one Fc fragment stem. Each Fab fragment is composed of one constant domain (C_L) and one variable domain (V_L) from the light chain as well as one constant domain (C_{H1}) and one variable domain (V_H) from the heavy chain. C_L tightly associates with C_{H1} and V_L with VH, and a flexible "elbow" connects the two domains of each chain in the Fab fragment[8, 9]. Each variable domain consists of a conserved β sheet framework and some short loops. Six of the loops, three from the light chain (L1, L2, L3) and three from the heavy chain (H1, H2, H3), vary in length and in amino-acid sequence among different immunoglobulins, which are the hypervariable regions, referred to as complementarity-determining regions (CDRs) and located at the tip of each Fab arm in the Y-shaped antibody molecule. These hypervariable regions are the antigen-binding sites, which recognize different antigens specifically and result in the specific immune responses[10].

Based on the comparison of known antibody structures and sequences, the "canonical structure model" has been proposed about the conformation of the hypervariable regions of antibodies[11]. In this model, antibodies have only a few main-chain conformations for each hypervariable region, and most sequence variations would only modify the surface provided by the side chains in a canonical main-chain structure. Some experiments[12] even showed that some of the hypervariable regions in the antibodies have the same main-chain conformation in spite of several differences in sequences. However, H3 usually do not comply with the canonical structure model[13].

METHODS

The first step is to build the three-dimensional structure model of anti-TCS IgE based on the known three-dimensional structures of antibodies and the primary amino-acid sequence of anti-TCS IgE, assuming that the backbone structure of the homologous domains were the same except in the regions where there are length differences between the amino-acid sequences[14].

One of the five anti-TCS IgEs, TE1, was chosen as the representative of the anti-TCS IgEs since it has been studied more detailed than the others and shows good binding activity with TCS.

There are about thirty antibody fragment structures in Brookhaven Protein Data Bank (PDB, 1993), but none of them contains the variable regions of IgE. The structure of a peptide-complexed Fv fragment of IgG1, B13I2[15] (PDB entry 2igf) was used as the starting model, which is very similar to TEl in both the length and the sequence (80% identical for the heavy chain and 64% for the light chain). The lengths of CDRs are also quite similar to each other in the two primary structures. In the light chain, CDR1 of TEl is one residue shorter and CDR3 is one residue longer than those in B13I2 while CDR2 has the same length as in the latter. In the heavy chain, the lengths of both CDR1 and CDR2 are the same as in B13I2, but CDR3 are four residues different in length from that in B13I2. This difference probably does not bias our modeling, because unlike the other CDRs the lengths of CDR3 loops of the heavy chains are quite different among the six anti-TCS antibodies while they recognize the same antigen site, suggesting that the length of CDR3 of the heavy chain may not be critical for recognizing the antigenic determinant of TCS.

The primary structure of the starting model was replaced by the amino acid sequence of TE1. Then the model was energy-minimized (500 cycles) to give the three-dimensional structure of the Fv fragment of anti-TCS IgE.

The second step is the prediction of the antigen determinants of TCS by looking for the preferable interactions between TCS and the CDR loops of anti-TCS IgE.

Antigenic determinants are usually located in the surface regions of proteins. We carefully examined all the different surface regions of TCS to look for the regions with the surface roughly complementary to that of the antigen -binding pocket formed by the CDRs of TEl. Then, the sidechain conformation was modified for the residues in these regions of TCS and the corresponding residues in CDRs of TEl to see whether reasonable interactions could be formed (hydrogen bonds, salt bridges and van der Waals contacts were considered). Finally, the surface regions of TCS that can well fit the antigen-binding pocket of TEl were taken as the possible antigenic determinants, and the further adjustment of the side chain conformation and the energy minimization (100 cycles) of the whole structure were performed.

The model building and the manual adjustment were carried out using the graphics software TOM/FRODO[16] and the energy minimization using the program package XPLOR[17], both on an INDIGO XS24 workstation.

RESULTS

Fig 1 shows the comparison of the amino-acid sequences of the hypervariable regions in TE1 compared with those in B13I2.

L1 L2 L3 TE1: ...KASQSVDY_DGDSYMN..... AASNLES....QQSNEDPTWT... B1312: ...RSNQTILLSDGDTYLE......KVSNRFS.....FQGS_HVPPT... H1 H2 H3 TE1: ...SFGMH...YISSGSSTIYYADTVKG...SYYSYFSYEGAMDY.. B1312: ...RCAMS....GISSGGSYTFYPDTVKG...TS_SD_PF_YF_DY..

Fig 1. Sequence of the hypervariable regions of TEl compared with those of B13I2. The positions of insertion or deletion are marked by underlines.

The energy-minimized Fv fragment of anti-TCS IgE model does not show obvious conformational change for main-chain except H3, compared with the B13I2 starting model (the root mean square shift of atomic positions is 0.46 Å, the total energy change is 114,032 kcal/mol). Two molecular surface regions in TCS were found to form extensive interactions with the hypervariable regions of TEl and have been predicted to be the possible antigenic determinants, as shown in Fig 2 (the root mean square shifts of atomic positions and the total energy change of the minimization of region 1 and 2 are 0.27 Å and 38346 kcal/mol and 0.29 Å and 25410 kcal/mol, respectively. After the minimization, there are no unreasonable contacts in the structures.).

The first region is a discontinuous antigenic determinant[18], composed of two polypeptide segments, Ile201 ~ Glu210 and Ile225 ~ Asp229, which are from different parts of the amino-acid sequence and are brought together by the folding of the polypeptide chain (Fig 2). The segment Ile201 ~ Glu210 contains two continuous β -turns while the segment Ile225 ~ Asp229 is a β -strand of a hairpin[4]. This region

- \triangleright Fig 2. C a drawing of TCS (in green) with the two predicted antigenic determinants indicated (in red).
 - Fig 3. a) The space filling model showing the complementarity of TCS (red for region 1 and green for the rest of TCS) to TEl (yellow for the light chain and blue for the heavy chain).
 b) C α drawing of TCS and TEl as well as the side-chains of region 1 of TCS. Region 1 of TCS is shown in red and the rest of TCS structure in green. The light chain and the heavy chain of TEl are shown in yellow and blue respectively.
 - Fig 4. a) The space filling model showing the complementarity of TCS (red for region 2 and green for the rest of TCS) to TEl (yellow for the light chain and blue for the heavy chain).

b) C α drawing of TCS and TEl as well as the side-chains of region 2 of TCS. Region 2 of TCS is shown in red and the rest of TCS structure in green. The light chain and the heavy chain of TEl are shown in yellow and blue respectively.



is located in the small domain of TCS, with a quite protrusive main-chain conformation in the three-dimensional structure, and the surface of this region is well complementary to that of the hypervariable region of TEl, as shown in Fig 3. Nine of the fifteen residues in this region are polar or charged (Ser, Thr, Asn, Gln and Glu), forming a highly hydrophilic surface. These hydrophilic residues form extensive hydrogen bonds with TEl, which are listed in Tab la. While this region protrudes into the antigen-binding pocket of the antibody, the hydrophobic residues in this region make van der Waals contacts with TEl within 4 Å distances (Tab 2a). The segment Ile201 ~ Glu210 interacts with the CDR loops L3, H2 and H3, and the segment Ile225 ~ Asp229 with L1.

The second region is a continuous antigenic determinant[18], the polypeptide segment Lysl70 Thr178 (Fig 2), which agrees with the antigenic determinant predicted from the hydrophilicity calculation by CHARGPRO program in PC/GENE[6]. This segment is in the large domain of TCS, and basically consists of a β -turn in the structure[4]. Its main-chain conformation is not so protrusive and the surface is not so complementary to the CDRs of TEl as in region 1 (Fig 4), but it contains some residues with long and positively charged side-chains, such as Lys173, Arg174 and Lys177, in addition to the polar residues such as Asp176 and Thr178, forming a hydrophilic surface with higher polarity than that in region 1. The extensive hydrogen bonding interactions and van der Waals interactions of this region with TE1 are shown in Tab lb and Tab 2b. Among these residues, Asp176 makes important contribution to the antibody-binding, which form three hydrogen bonds with TE1. The residues from L1, H2 and H3 of TEl are involved in the hydrogen bonding interactions with TCS.

(a) Region 1				<u>_</u>
	Atom 1 in TCS	Atom 2 in TE1	(CDR)	distance (Å)
	Ile 201 O	Tyr 59 OH	(H2)	3.00
	Thr 204 OG1	Tyr 59 OH	(H2)	3.14
	Asn 205 ND2	Tyr 100 OH	(H3)	2.88
	Gln 208 OE1	Ser 105 O	(H3)	2.60
	Glu 210 OE2	Asp 98 OD1	(L3)	2.56
	Thr 226 O	Tyr 31 OH	(L1)	2.99
	Asn 227 OD1	Asp 32 OD1	(L1)	2.61
	Asn 227 OD1	Asp 32 OD2	(L1)	3.00
(b) Region 2)				
	Atom 1 in TCS	Atom 2 in TE1	(CDR)	distance (Å)
	Lys 173 NZ	Ser 52 OG	(H2)	2.75
	Lys 173 NZ	Ser 53 OG	(H2)	2.95
	Arg 174 NH1	Tyr 59 OH	(H2)	2.94
	Arg 174 NH2	Tyr 59 OH	(H2)	2.70
	Arg 176 OD1	Tyr 103 OH	(H3)	2.94
	Arg 176 OD2	Tyr 36 OH	(L1)	3.10
	Asp 176 N	Ser 105 OG	(H3)	2.51
	Lys 177 NZ	Tyr 31 OH	(L1)	2.75
	Thr 178 OG1	Asp 32 OD1	(L1)	2.74

Tab 1. The hydrogen bonding interactions between TCS and TE1.

(a) Region 1			
	Residues in TCS	Residues in TE1	
	Ile 201	Tyr 59H	
	Thr 204	Tyr50H, Ser52H, Thr57H, Tyr59H	
	Asn 205	Tyr100H, Ser105H, Tyr106H	
	Asn 206	Tyr106H, Ser105H	
	Gln 208	Tyr31L, Asn96L, Trp101L	
	Phe 209	Tyr31L	
	Glu 210	Asp98L	
	Thr 226	Tyr31L	
	Asn 227	Asp32L	
	Asp 229	Asp 32L	
(b) Region 2			
	Residues in TCS	Residues in TE1	
	Lys 173	Ser52H, Ser53H, Thr57H	
	Arg 174	Tyr59H	
	Asp 175	Tyr103H, Ser105H	
	Asp 176	Tyr36L, Tyr103H, Phe104H	
	Lys 177	Tyr31L	
	Thr 178	Tyr31L, Asp32L	

Tab 2. The Van der Waals interactions between TCS and TEl. (the distance cutoff is 4.0 Å H and L represent heavy chain and light chain, respectively.)

From Tab 1 and Tab 2, we can see that CDRH3 of TEl is involved in the interaction with TCS. In our model, Fv model is built upon the sequence of TEl, which has the longest CDRH3 (14 residues) among the six anti-TCS antibodies, and the other CDRH3s have no more than 10 residues. In TE2, the CDRH3 is so short (only 6 residues) that it will not be close enough to TCS in space to form any effective interactions as in the case of TE1. So, we think that the CDRH3s are probably much more related to the affinity of these antibodies than to the recognizing speciality of them.

DISCUSSION

Molecular modeling, as a method to predict the antigenic determinants, has some substantial uncertainty and limitation. However, it provides a structurally reasonable and energetically favorable model for the design of further experiments.

Here we predict two molecular surface regions of TCS to be the possible antigenic determinants. From the results, we find that the region 1 has a better complementary surface and more extensive interactions with TEl than region 2. Therefore, region 1 seems to be a more reasonable region which can be recognized by these antibodies. However, the experimental evidence is needed to make the conclusion.

The following experimental results support our prediction:

1) The investigation on the sequence conservation and the ELISA determination with the light chain replacement showed that the light chain plays a more important role than the heavy chain in the antigen-binding[6], which agrees with the results from our modeling study that light chain has a more important role to interact with

TCS in region 1 than in region 2.

2) The mutant N227K of TCS showed immune response change (Chen HB, personal communication) and this residue is located in region 1.

REFERENCES

- [1] Wang Y. Trichosanthin (in Chinese). Science Press: Beijing. 1990.
- [2] Zhang JS, Liu WY. The mechanism of action of trichosanthin on eukaryotic ribosomes-RNA N-glycosidase activity. Nucleic Acids Res 1992; 20:1271-5.
- [3] MeGrath MS, Hwang KM, Caldwell SE, et al. GLQ233: An inhibitor of human immunodeficiency virus replication in acutely and chronically infected cells of lymphocyte and mononuclear phagocyte lineage. Proc Natl Acad Sci USA 1989; 86:2844-8.
- [4] Xia ZX, Zhang L, Zhang ZM, Wu S, Dong YC. The three-dimensional structure of trichosanthin relined at 2.7 Å resolution. Chinese J Chem 1993; 11:280-8.
- [5] Zhang L, Xia ZX, Wu S, Dong YC. The further crystallographic refinement of trichosanthin at 2.7 Å resolution and the comparison among the three molecular structures in two crystal forms. Chinese J Chem 1994; 12:223-30.
- [6] Wang Y, Yeh M. Molecular characterization of the V regions of four antibodies specific for trichosanthin. Immunology 1996; submitted for publication.
- [7] Gu H, Yeh M, Yao Z. Investigation of antigenic determinants on trichosanthin by antibody competitive binding assay. Acta Biol Exp Sinica 1986; **19**:121.
- [8] Poljak R J, Amzel LM, Avey HP, Chen BL, Phizackerly RP, Saul F. Three-dimensional structure of the Fab fragment of a human immunoglobulin at 2.8 Å resolution. Proc Natl Acad Sci USA 1973; 70:3305-10.
- [9] Amzel LM, Poljak RJ. Three-dimensional structure of immunoglobulin. Annu Rev Biochem 1979; 48:961-97.
- [10] Jones PT, Dear PH, Foote J, Neuberger MS, Winter G. Replacing the complementarity-determining regions in an human antibody with those from a mouse. Nature 1986; **321**:522-5.
- [11] Chothia C, Lesk AM, Tramontano A, et al. Conformations of immunoglobulin hypervariable regions. Nature 1989; 342:877-83.
- [12] de la Paz P, Sutton BJ, Darsley MJ, Rees AR. Modeling of the combining sites of three antilysozyme monoclonal antibodies and of the complex between one of the antibodies and its epitope. EMBO J 1986; 5:415-25.
- [13] Wu TT, Johnson G, Kabat EA. Length distribution of CDRH3 in antibodies. Proteins 1993; 16:1-7.
- [14] Padlan EA, Kabat EA. Model-building study of the combining sites of two antibodies to (1→6)dextran. Proc Natl Acad Sci USA 1988; 85:6885-9.
- [15] Stanfield RL, Fieser TM, Lerner RA, Wilson IA. Crystal structure of an antibody to a peptide and its complex with peptide antigen at 2.8 Å. Science 1990; 248:712-9.
- [16] Jones TA. Interactive computer graphics:FRODO. In: Wyckoff HW, Hirs CHW. eds. Methods in enzymology. Academic Press: Orlando. 1985; 115:157-71.
- [17] Brunger AT, Clore GM, Gronenborn AM, Karplus M. Three-dimensional Structure of proteins determined by molecular dynamics with interproton distance restraints:application to crambin. Proc Natl Acad Sci USA 1986; 83:3801-5.
- [18] Barlow DL, Edwards MS, Thornton JM. Continuous and discontinuous protein antigenic determinants. Nature 1986; 322:747-8.

Received 17-5-1996. Revised 10-10-1996. Accepted 24-10-1996.