Vol. 211, No. 1, 1995 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
June 6, 1995 Pages 14-20

SEQUENCE AND STRUCTURAL HOMOLOGIES BETWEEN *M.tuberculosis* CHAPERONIN 10 AND THE MHC CLASS I\II PEPTIDE BINDING CLEFT

Edith Chan, Gianluca Fossati, Paola Giuliani, Pierluigi Lucietto, Andrea Zaliani, Antony R.M. Coates[1] and Paolo Mascagni*

Italfarmaco Research Centre, Via Lavoratori 54,Cinisello B., 20092 Milan, Italy

[1] Division of Molecular Microbiology, St George's Hospital, Medical School, London, UK

Received April 20, 1995

SUMMARY: The peptide corresponding to the C-terminal half of *M.tuberculosis* hsp10 was synthesised based on the prediction that it might represent an independent structural region of the protein. This hypothesis was confirmed by aggregation and CD studies using this peptide and longer sequences of the protein. The peptide shares about 40-50% sequence homology with $\alpha 2$ and $\beta 1$ chains of MHC class I and II antigens. This and the CD results which indicated that the peptide at acidic pHs folds into an anti-parallel β-sheet were used to generate a 3D model which has the same "W" fold contained in the MHC peptide binding groove. These data suggest that the hypothesis of molecular mimicry proposed to be one of the mechanisms which triggers autoimmune diseases may be extended to hsp10 proteins. Furthermore the suggested evolutionary relationship between hsp's and MHC antigens may find support from these data.

The sequence related chaperonin 10 (cpn10, hsp 10 or 10kDa antigens) class of proteins mediate the folding of either newly synthesised proteins or proteins that following cellular stress have undergone denaturation (1). In addition to their chaperonin activity several cpn10s (e.g. *Mycobacterium tuberculosis* and *M. leprae* cpn10s) are amongst the most potent stimulators of the immune system known (see for example ref. 2). Since infectious microorganisms are associated with many forms of chronic inflammatory disease, it has been proposed that stress proteins might trigger autoimmunity in genetically susceptible individuals (3). The concept of molecular mimicry has been suggested as one possible mechanism by which this can occur (4). This hypothesis proposes that the host's immune system recognises self-epitopes which are shared by, for example, microbial heat shock proteins (hsp) and so initiate an autoimmune response.

Recently (5,6) a correlation at the structural level, between MHC and hsp70 has been suggested. This was based on the deduced consensus secondary structure from a set of aligned hsp70 sequences. The hsp70 substrate-binding domain was proposed to be similar to the peptide binding groove of MHC class I\II antigens. Although this proposal needs confirmation it remains attractive due to the several common characteristics between these two classes of

*Corresponding author. Telefax: ++39-2-66011579.

0006-291X/95 $5.00

molecules, namely, (i) they both bind short peptides with diverse sequences (7), (ii) hsp's are involved in antigen processing (8) and (iii) one of the inducible forms of the hsp70 family is encoded in the MHC class III region (9). Collectively these observations were used to propose an evolutionary relationship between stress proteins and MHC antigens (5,6).

In this study structural investigations on a synthetic peptide corresponding to the C-terminal half (sequence 59-99) of *M.tuberculosis* (Mt) cpn10 reveal that both structure and sequence homologies exist between the Mt fragment and the peptide binding site of MHC class I and II antigens. These findings add support to the hypothesis that MHC antigens and heat shock proteins are evolutionarily linked. Furthermore they suggest that the hypothesis of molecular mimicry between hsp molecules and self molecules which has been suggested to play a role in the pathogenesis of autoimmune diseases (3), may also apply to the cpn10 family.

MATERIALS AND METHODS

Chemical Synthesis. The 59-99 peptide was synthesised on solid phase using t-BOC chemistry. After HF cleavage, 28 mg of crude peptide was dissolved in 50 ml of 8M urea , the solution was transferred to the Rotofor cell (BioRad), ampholites (2%) were added (pH range 3-10) and focusing was allowed to take place at 4°C for 4 hr.s. At the end of the run the fractions corresponding to the calculated pI of the peptide [4.26] were pooled together the volume brought to 50 ml with 8M urea and focused for additional 4 hr. At the end of this second run those fractions having a similar HPLC profile were combined and further fractionated by semi-preparative RP-HPLC. This gave material with the expected molecular weight (Mw expected 4,477, found 4,477), amino acid composition and N-terminal sequence. The purity of the peptide was larger than 95% as judged by analytical HPLC.

Circular Dichroism. Spectra were recorded on a Jasco J600 instrument using a rectangular quartz cuvette (0.1 cm). The concentration of the polypeptides in 0.1M phosphate buffer was 0.1 mg/ml. Spectra were the average of 8 scans at 50 nm/min each. The observed ellipticity was converted to mean residue weight ellipticity $[\Theta]$\(deg·cm^2·d·mol^{-1}). The variable selection method (10) and the CCA method (11) were used for the analysis of secondary structure.

Computer Modeling. Homology building was done using the protein package in Quanta 3.3 (Molecular Simulation). The 59-99 peptide was aligned to the α2 chain of class I MHC by sequence homology. Only the backbone coordinates (12) of the homologous part of class I MHC HLA-A2 were copied onto the unknown structure of 59-99. Side chains were built from the coordinates of the corresponding residues in the template structure. Refinement of the structure was carried out saturating each of its residue in a 6 Å shell of water and by (i) 2000 steps of ABNR minimization, (ii) molecular dynamics calculations (5ps of heating to 300°K, 5ps of equilibration at 300°K, and 100ps of production simulation) and (iii) a further step of minimisation. All energy calculations were performed with the CHARMm (13) force field using polar hydrogens. The cut-off distance for the non-bonded interactions was 15.0 Å.

RESULTS AND DISCUSSION

The sequence of Mt cpn10 protein is conveniently divided in two halves approximately identical in size by the 50-56 region which has been identified as the protein monoclonal antibody binding site (14). Furthermore secondary structure predictions assigned sequence 46-59 to a long loop region (data not shown). Thus the 59-99 peptide was synthesised based upon the assumption that it might represent a region of the protein which may independently fold into stable structures.

Gel filtration studies. Aggregation studies on the Mt cpn10 protein and fragments were carried out by size exclusion chromatography and analytical ultracentrifugation and are described elsewhere (15). The salient points from these studies can be summarised as follows: (i) Unlike all cpn10s known to date, the Mt protein is tetrameric in dilute solutions, (ii) In the presence of a large molar excess of divalent ions the protein forms heptameric structures which are functional in refolding assays, (iii) Mt cpn10 fragments corresponding to sequences 26-99, 51-99 and 59-99 are monomeric at neutral pH and dimeric (26-99 and 51-99) and monomeric (59-99) at acidic pH. Thus, the assumption that sequence 59-99 might represent a defined structural region of the protein seemed confirmed by its aggregation behavior at acidic pHs which contrasted with that of the other larger fragments.

CD spectroscopy. The CD spectrum of 59-99 at neutral pH had minima at about 215 and 202 nm (Figure 1A). By contrast whole protein and the other fragments had minima in the 198-203 nm range and shoulders between 217 and 220 nm (Figure 1A). Interpretation of the protein spectrum (15) suggested that the structure of Mt cpn10 contains β-pleated sheets. This conclusion is consistent with the indication that the nearly complete X-ray structure of *E.coli* cpn10 is made of identical subunits with a β-barrel fold (16).

Decreasing the pH of 59-99 solutions led to a decrease of the contribution at 202 nm and an increase in the band at 215 nm.. The latter, which became maximal at pH 3.0, was interpreted as deriving from anti parallel β-sheets (17). This conclusion was consistent with both secondary structure predictions (Table 2) and spectral deconvolution carried out using two

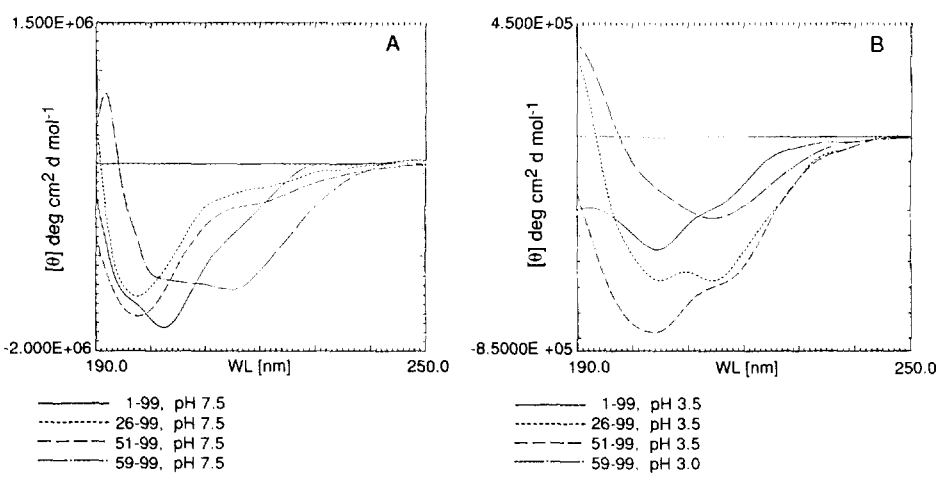

Figure 1.
CD spectra of Mt cpn10 and fragments corresponding to sequences 59-99, 51-99 and 26-99. Spectral data below 195 nm are not reliable due to interferences of the solvent (0.1 M phosphate) in this wavelength range. The spectra of Mt cpn10, peptide 26-99 and peptide 51-99 at neutral pH are multiplied by a factor of 300, 10 and 5, respectively. Those of the protein and fragment 26-99 at acidic pH by a factor of 100 and 10, respectively.
A) Spectra at neutral pH. B) Spectra at acidic pH.

independent methods (19,11): the first gave 10% helix, 41% β-sheet, 14% turn and 27% random coil whilst the second 0% helix, 21% β-sheet, 29% and 50% for turn and random coil respectively. The structure of the peptide at acidic pH was extremely stable as shown by the invariance of the spectrum in the temperature interval 0°-60°C (data not shown).

Extending the sequence of the peptide by further 8 residues had an important effect on the CD spectrum of the resulting 51-99 peptide. Thus the latter showed, in addition to the β-contribution at 215-217 nm, a new intense minimum at 204 nm (Figure 1B). The nature of this new secondary structure element, present also in the 26-99 fragment and whole protein, was not identified unequivocally although it was found that it correlates with the ability of these polypeptides to aggregate (15).

In conclusion, the CD data in addition to showing that at acidic pH peptide 59-99 has a stable structure containing a β-sheet, reinforced the initial hypothesis that the region C-carboxyl to the protein antibody binding site might represent a structurally independent region either when part of whole protein or when in an isolated form as is the case of the peptide described here.

Molecular modeling. Sequence homology searches revealed that a stretch of seven residues (aa 80-86) shares 70% identity and 86% similarity with β1 chains of class II MHC molecules (sequence 32-38) (Table 1). Although the consensus sequence Y\h-t-t-q-q-Y\h-h was derived for the 80-86 region of all cpn10s, proteins other than the mycobacterial ones did not show such a high homology, their similarities with the β1 chain ranging between 30% and 50% (Table 1). Within the homologous region of MHC II residues 35 and 36 (E-E) are conserved

Table 1. Sequence Homology between Cpn10 and β1 MHC II Antigens

	80						86
M10[a]	Y	N	G	E	E	Y	L
Cpn10[b]	L(7)	D(14)	G(19)	E(15)	E(20)	Y(16)	L(12)
	V(7)	N(4)	N(5)	K(6)	D(7)	L(7)	V(9)
	Y(6)	G/E(3)	D(2)	D(3)		V/I(1)	I(2)
	I(6)	A(2)	S(1)	Q(2)		S(1)	M/F(1)
	F(1)	S(1)		S(1)		F(1)	S(1)
Consensus[c]	**Y/h**	**t**	**t**	**q**	**q**	**Y/h**	**h**
	32						38
MHCII β1[d]	Y(38)	N(41)	R(22)	E(44)	E(44)	Y(24)	V(21)
	H(3)	H(2)	Q(15)			F(8)	L(12)
	F(2)	P(1)	L(6)			N(6)	A(8)
	R(1)		D(1)			S(3)	W/T(1)
						W/H/I(1)	G(1)
Consensus	**Y**	**N**	**R/Q**	**E**	**E**	**Y**	**h**

Data base retrieval and sequence homology were carried out using MacDNASIS Pro (Hitachi) with multiple alignments with Quanta 3.3 (Molecular Simulation). [a]M10 refers to the three mycobacterial cpn10s (i.e. *M.tuberculosis*, *leprae* and *bovis*). [b]27 sequences of Cpn10 other than those of the mycobacterial family were extracted from SWISS-PROT. The frequency of occurrence of each amino acid in the sequences is shown in brackets. [c]The consensus sequence was obtained from a combination of sequence and secondary structure composition. For the prediction of secondary structure see Table 2. [d]A total of 44 β1 class II MHC sequences were extracted from SWISS-PROT. The consensus sequence was derived from the most frequently occurring residues at each position.

whilst residues 37 and 38 are polymorphic although mainly aromatic and hydrophobic respectively. Residues 32 and 33 are also conserved and mainly Y and N respectively (Table 1). MHC I antigens have an inferior degree of identity, although the consensus sequence from above applies also to this class of molecules (see Table 2). The region homologous with cpn10 comprises the β-hairpin between the second and third strands of the β-sheet forming half of the pavement of the MHC I/II peptide binding groove (Figure 2) (12,18). These data and the CD results which show that 59-99 contains an anti parallel β-sheet prompted an alignment between the 59-99 peptide and the corresponding regions of class I and II MHC molecules. This revealed that both types of molecules share a similar degree of identity with the cpn10 protein although the overall average homology (i.e. identity plus similarity) was larger in the case of MHC I (a maximum of 53% vs. 39%) (Table 2). These homologies, although within plausible limits (20), are too low to be significant on their own. However our CD data strengthens the case for sequence and structural homology between Mt cpn10 and MHC. Accordingly, the sequence of 59-99 was copied onto the coordinates of MHC I and a 3D model obtained. The final structure retained the fold of the MHC molecule and was a β-sheet composed of 4 up-and-down anti parallel strands (Figure 2). Using the Kabsch and Sander's definition (21), the model contained 30% of the residues in a β-sheet conformation, 29% in β-turn, 7% in α and 34% with undefined structure. These figures were in general agreement with those obtained from deconvolution of the CD spectra.

From the results described here we propose two hypotheses. Firstly, the molecular mimicry between other hsp molecules (e.g. 60kDa, Reference 1) and self-molecules which has been suggested to play a role in autoimmune diseases, may also apply to the cpn10 family. Recent data which indicate that the Mt cpn10 molecule is capable of modulating the severeness of adjuvant arthritis in rats following administration of killed *M.tuberculosis* in oil (22) is consistent with this conclusion. Secondly, although the structural data presented here need

Table 2. Sequence and structural homologies between the 59-99 peptide from *M.tuberculosis* cpn10 and α2 and β1 chains of MHC antigens class I and II, respectively

Secondary structure(Mt cpn10)[a]		tbbbbbt-tbbbbbbb--ttbbbb--bttbbbbbbtt-llllllll
MHC Class II (β1 chain)[b]	9	EYSTSE-CHFFNGTER--VRFLDRYFYNQEEYVRFDS-DVGEFRAV 50
Mt Cpn10	59	PLDVAEG-DTVIYSKY--GGTEIK--YNGEEYLILSARDVLAVVSK 99
MHC Class I (α2 chain)[b]	122	TIQIMYGCDVGSDGRFLRGYRQDA--YDGKDYIALNE-DLRSWTAA 160
Secondary structure (MHC I)[c]		bbbbbbbbbltttllbbbbbbbbbb--bttlllbbbbt-tllbbbbl

[a]The secondary structure of Cpn10 was predicted by first aligning 27 sequences of Cpn10, then considering the secondary structure of each section (separated by gaps) using two different algorithms: Chou-Fasman and GORII. Secondary structure representations are as follows: b = β-sheet, t = turn and l = loop. Residue match was done using the Dayhoff matrix and a cut-off of 9%. Identical residues are underlined whilst similarity is denoted by |. [b] The 41 (out of 107) most representative human MHC I/II sequences present in the SWISS-PROT data bank were used for the alignment with the Mt cpn10 peptide. HLA-DR5 β1 chain and HLA-A3 α2 chain are shown here as examples of MHC class II and I, respectively. [c]The secondary structure was taken from 3HLA PDB file.

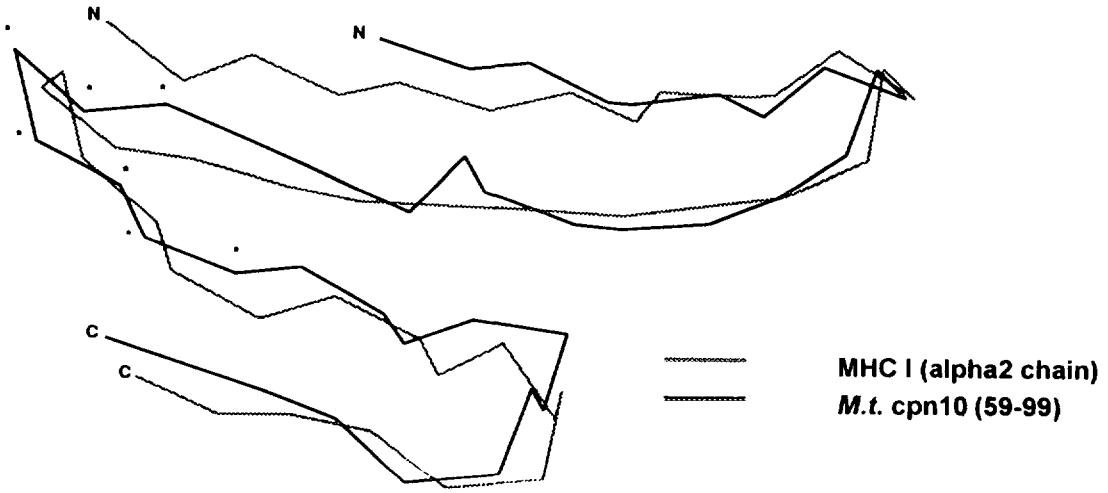

Figure 2.
Cα trace of the 3D model of peptide 59-99 superimposed with the α2 chain of MHC class I antigen. Residues marked with * correspond to the region of MHC class I and II molecules highly homologous with sequence 80-86 of Mt cpn10.

confirmation at the atomic level, they support an evolutionary relationship between hsp molecules and MHC antigens. This agrees with the work from other laboratories which have suggested that, for example, the structural similarities between the hsp70 substrate-binding domain and MHC may be of evolutionary significance (5,6) although these data also need atomic level verification.

REFERENCES

1. Ellis RJ and Van der Vies (1991) Annu. Rev. Biochem. **60**, 321-347.
2. Coates, A.R.M. (1995) In: The Chaperonins (Ellis R.J Ed.), Academic Press, London, in press.
3. Young D.B., Melhert A. and Smith D.F. (1990) In Stress proteins in biology and medicine (Morimoto R.I., Tissieres A. and Gerorgopoulos C., Ed.s) pp131-166 Cold Spring Harbor.
4. Cohen, I.R. (1991) Annu. Rev. Immunol. 9, 567-589.
5. Flajnik M.F., Canel C., Kramer J. and Kasahara M. (1991) Immunogenetics 33, 295.
6. Rippman F., Taylor W.R., Rothbard J.B. and Green N.M. (1991) EMBO J. 10, 1053.
7. Flynn G.C., Chappell T.G. and Rothman J.E. (1989) Science 245, 385-390.
8 De Nagal D.C. and Pierce S.K. (1991) Semin. Immunol. 2, 65-71.
9. Sargent, C.A., Durham, I., Trowsdale, I.; and Campbell, R.D. (1989) Proc. Natl. Acad. Sci., 86, 1972-1986.
10. Manavalan P. and Johnson W.C. Jr. (1983) Nature 305, 831-832.
11. Perczel, A., Hollosi, M., Tusnady, G., and Fasman G.D.(1991) Prot. Eng. 4, 669-679.
12. Bjorkman P.J., Saper M.A., Samraoui B., Bennett W.S., Strominger J.L. and Wiley D.C. (1987) Nature 329, 106-512.
13. Brooks B.R., Bruccloeri R.E., Olafson B.D., States D.J., Swaminathan S. and Karplus M. (1983) J. Comput. Chem 4, 187.
14. Verbon, A., Hartskeerl, R. A., and Kolk, A.H.J. (1991) Clin. exp. Immunol. 86, 6-12.

15. Fossati, G., Lucietto, P., Giuliani, P., Coates, A.R.M., Harding, S., Colfen, H., Legname, G., Chan, E., Zaliani, A., and Mascagni, P. (1995), J. Biol. Chem. Submitted.
16. Saibil, H. R. (1994) Nature: Structural biology 1, 838-842.
17. Perczel, A., Park, K., and Fasman, G. D.(1992). Proteins: Structure, Function, and Genetics. **13**, 57-69.
18. Brown J.H., Jardetzky T.S., Gorga J.C., Stern L.J., Urban R.G., Strominger J.L. and Wiley D.C. (1993) Nature 364, 33-39.
19. Doolittle R.F. (1990) Methods in Enzymology 183, 99-110.
20. Sanders, C., and Schneider R. (1991) Proteins 9, 56-58.
21. Kabsch W. and Sander C. (1983) Biopolymers 22, 2577-2637.
22. Ragno, S., Winrow, V.R., Morris, C.J., Blake, D.R., Del Pierro, F., Gromo, G., and Mascagni P. (1995), J. Immunol. Submitted.