

Pattern Recognition Among T-Cell Epitopes

J. B. Rothbard¹, A. Townsend², M. Edwards³, and W. Taylor³

A. Introduction

T cells recognize protein antigens by mechanisms qualitatively different from those used by B cells. B cells are capable of binding antigens via their surface immunoglobulins, as in other well-understood ligand-receptor interactions. In contrast, T cells are unable to bind antigen in the absence of the major histocompatibility class I or class II gene products [1]. The details of this possible tertiary interaction are still poorly understood. One puzzling feature of the contrasting recognition processes is that the antigen receptor on T cells exhibits great structural similarity with immunoglobulins [2, 3].

One possible resolution of this apparent contradiction is the proposal that the form of the antigen which the two populations of lymphocytes see is different. Recent work from a number of laboratories has revealed that B cells recognize protein antigens with their native conformation intact, while the majority of T cells recognize protein antigens with their native conformation disrupted [4, 5]. The most lucid demonstration supporting this generalization was the experiment by Watts and McConnell, who successfully stimulated an ovalbumin specific T-helper clone with a peptide corresponding to a region of the protein and the correct class II antigen anchored in a lipid monolayer. The intact protein in its native

conformation was completely ineffective [6]. Work in other laboratories examining other T-cell clones has revealed that both enzymatically derived peptides and synthetic peptides are excellent stimulators of helper and cytolytic T cells.

The interaction between antigen, T-cell receptor and class II proteins was first extensively studied by Heber-Katz and Schwartz, who used cytochrome-specific helper clones [7, 8]. The region of the molecule which stimulated their clones was the linear region corresponding to residues 88–103. This C-terminal helical region of the protein stimulated the clones as well as the intact protein. By examining the cross-reaction between their cytochrome-specific clones and cytochromes from other species they concluded that residues between 98 and 103 were necessary for the specificity. However, these residues alone were insufficient for stimulation. Longer peptides were necessary for maximum stimulation of their clones. The authors concluded that the longer peptides could stabilize a preferred conformation, in this case a helix, and could therefore bind the receptor with higher affinity. Various modifications to the sequence provided additional support for this concept.

Concurrently with this work, Berzofsky [9] and Livingstone et al. [10] were analysing the T-cell epitopes within myoglobin. In this molecule as well, the epitopes seen by T cells were located within linear regions of the protein sequence composing helical regions. The ovalbumin epitope previously mentioned could not be unequivocally designated as helical because its X-ray structure was not known; however, it was composed

¹ Imperial Cancer Research Fund, Lincoln's Inn Fields, London WC2A 3PX, England

² John Radcliffe Hospital, Oxford, England

³ Birkbeck College, London, England

of residues which could easily be modeled to be an amphipathic helix.

In compiling these and the other epitopes known at the time, Berzofsky postulated that T cells interact with amphipathic regions of protein antigens [11]. A large number of the epitopes he examined can adopt a conformation with separate polar and non-polar faces. In his model, the T-cell receptor would interact with one facade, while the class II or class I molecule would interact with the other. The analysis he used was that of Eisenberg, who used vector analysis to generate moments of inertia (in this case, moments of hydrophobicity) to quantify the amphipathic character of a region of protein sequence [12].

In this paper, we have analysed 30 T-cell epitopes and have discovered a pattern that is present in 29. The pattern is present in the linear sequence and does not require the region to adopt a particular conformation. Using the known epitopes as a data base, we have generated a template for predicting T-cell epitopes in other protein sequences and have predicted a likely region within the sequence of the nucleoprotein of influenza. This region was synthesized and shown to be the principal region seen by cytotoxic T cells isolated from CBA mice. We believe that this simple motif can be of general use for prediction and is an interesting model on which experiments analysing the interaction of T cells and proteins can be based.

B. Methods

Analyses of the hydrophobic moment for the peptides and proteins were made as described by Eisenberg [12]. The protrusion indices were calculated as described by Thornton et al. [13]. Sequence alignments and generation of templates were carried out by methods created by Thornton and Taylor [14]. Synthesis of peptides was done by standard Merrifield techniques [15] on an Applied Biosystems 430A peptide synthesizer [16]. T-cell lines were isolated from CBA mice and maintained in vitro as described [17]. A standard procedure was used for ^{51}Cr release assay [16] and the transfected NP target cells were prepared as previously described [18].

C. Results, Analyses and Discussion

The known helper and cytotoxic T-cell epitopes composing the database of this analysis are shown in Table 1. They are a combination of human and murine epitopes which have been either published or communicated to us.

We were dissatisfied, for several reasons, with Delisi and Berzofsky's generalization that T-cell epitopes are localized to regions of proteins that are amphipathic, particularly amphipathic helices. The first was that any generalization is only as accurate as its database. In this case, the number of epitopes examined was small, and the particular type of protein from which many of the epitopes were derived was not representative of all protein structures. The proteins exclusively composed of α -helices were over-represented. Secondly, as published, the correlation with amphipathic character was purely qualitative. Using the vector analysis of Eisenberg, discrete values are generated for linear regions of sequence. If their correlation is correct, the highest values should be recognized most often. Such a correlation is not seen (see below). When the angle used in generating the vectors of hydrophobicity is restrained at 100° (that consistent with standard α -helix of 3.6 residues per turn), only approximately 60%–75% of epitopes are consistent with the correlation. By varying the angle, and consequently increasing the areas of the protein that are possible, the correlation does improve. When analysed critically, the correlation simply implies that all areas of globular proteins are possible T-cell epitopes, with the exception of loops and turns – the prominent sites for B-cell recognition [13, 34, 35].

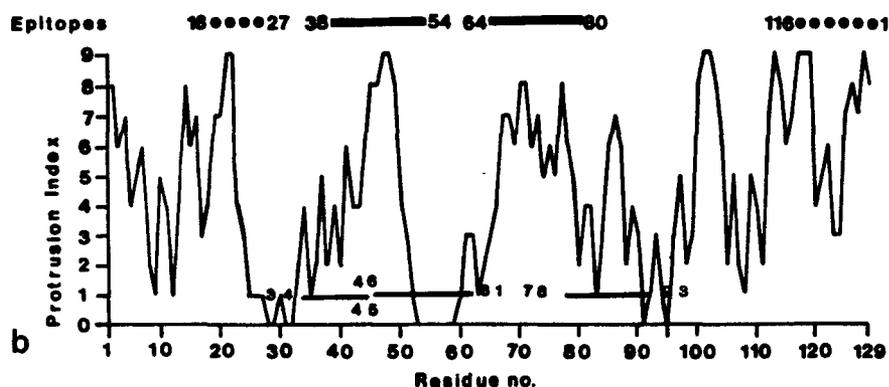
To examine this possibility, we plotted both the known linear B-cell epitopes and T-cell epitopes on to the protrusion index profile for myoglobin and lysozyme. As can be vividly seen, the B-cell epitopes map to highly exposed areas, whereas the areas preferentially recognized by T cells are poorly exposed (Fig. 1 a, b).

If proteolytic events are involved in antigen processing in order for T cells to recognize protein antigens, the most sterically available sites would be preferentially cleaved by enzymes for the same reasons

Table 1. Compilation of reported T-cell epitopes

Position	1	2	3	4	5	6	7	8	9	Reference											
SPERM WHALE MYOGLOBIN 69-78	L	T	A	L	G	A	I	L	K	K	[10]										
SPERM WHALE MYOGLOBIN 106-118	E	F	I	S	E	A	I	I	H	V	L	H	S	R	[9]						
SPERM WHALE MYOGLOBIN 110-121		A	I	I	H	V	L	H	S	R	H	P	G	[10]							
SPERM WHALE MYOGLOBIN 132-145		N	K	A	L	E	L	F	R	K	D	I	A	A	K	[9]					
INSULIN B-CHAIN 5-16	H	L	C	G	S	H	L	V	E	A	L	[19]									
CYTOCHROME PIGEON 93-104	K	S	E	R	V	D	L	I	A	Y	L	K	D	A	T	S	K	[7]			
CYTOCHROME BOVINE 13-25		K	C	A	Q	H	T	V	E	K	G	G	K	H	K	[20]					
OVALBUMIN 323-329	S	I	S	Q	A	V	H	A	A	H	A	E	I	N	E	A	G	R	[21]		
FLU NUCLEOPROTEIN (34/68) 335-349	I	A	A	F	E	D	L	R	V	L	S	F	I	R	G	[16]					
FLU NUCLEOPROTEIN (1968) 365-379	I	A	S	M	E	N	M	D	A	M	E	S	S	T	L	[16]					
FLU NUCLEOPROTEIN (1934) 365-379		A	S	M	E	N	M	E	T	M	E	S	S	T	L	[16]					
FLU NUCLEOPROTEIN (34/68) 50-63		S	D	Y	E	G	R	L	I	Q	N	S	L	T	I	[22]					
FLU HAEMAGGLUTININ PR/8 111-120		F	E	R	F	E	I	F	P	K	E	[23]									
FLU HAEMAGGLUTININ A/TEXAS/1/77 115-128	S	S	G	T	L	E	F	I	N	E	G	F	N	W	[24]						
FLU HAEMAGGLUTININ PR8/34 302-313				C	P	K	Y	V	R	S	A	K	L	R	M	[24]					
FLU HAEMAGGLUTININ A/NT/60/68 302-313				C	P	K	Y	V	K	Q	N	T	L	K	L	A	T	G	M	[24]	
FLU HAEMAGGLUTININ A/TEXAS/1/77 311-324		K	Q	N	T	L	K	L	A	T	G	M	R	M	V	[24]					
RAT MYELIN BASIC PROTEIN 5-20	P	S	Q	R	H	G	S	K	Y	L	A	T	A	[25]							
RAGWEED ALLERGEN Ra3 51-65	E	V	W	R	E	E	A	Y	H	A	C	D	I	K	D	[26]					
RAGWEED ALLERGEN Ra 51-65		E	V	W	R	E	E	A	Y	H	A	C	D	I	K	D	[26]				
HUMAN AChR GAMMA 125-147			K	S	Y	C	E	I	I	V	T	H	F	P	F	D	Q	Q	N	C	[27]
HEN EGG LYSOZYME 34-45		F	E	S	N	F	N	T	E	A	T	N	R	[28]							
HEN EGG LYSOZYME 46-61		N	T	D	G	S	T	D	Y	G	I	L	Q	I	N	S	R	[29]			
HEN EGG LYSOZYME 78-93	I	P	C	S	A	L	L	S	S	D	I	T	A	S	V	N	[30]				
HERPES GLYCOPROTEIN D 8-23					S	L	K	M	A	D	P	N	R	F	R	G	K	D	L	P	[31]
STAPH. NUCLEASE 61-80				F	T	K	K	M	V	E	N	A	K	K	I	E	V	E	F	D	[32]
STAPH. NUCLEASE 86-100	G	L	A	Y	I	Y	A	D	G	K	M	V	N	[32]							
STAPH. NUCLEASE 91-105			Y	I	Y	A	D	G	K	M	V	N	E	A	L	V	R	[32]			
VP1 FOOT-MOUTH VIRUS 141-160	R	G	D	L	Q	V	L	A	Q	K	V	A	R	T	L	P	[33]				

LYSOZYME - HEN EGG



MYOGLOBIN - SPERM WHALE

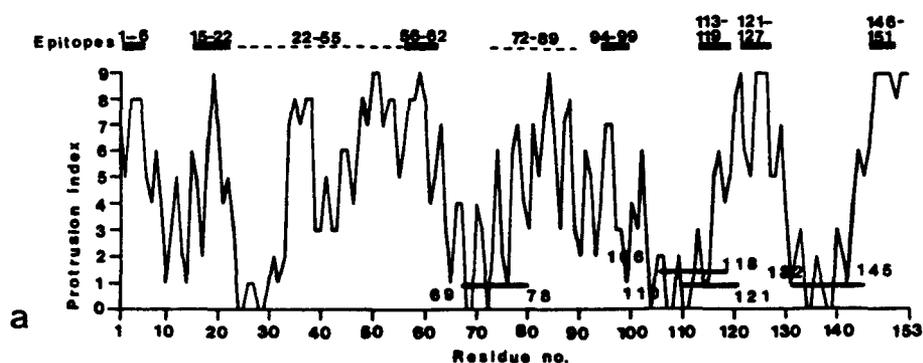


Fig. 1 a, b. Known B-cell epitopes (*upper bars, a, b*) and T-cell epitopes (*lower bars, a, b*) plotted on protrusion index profile [13] of myoglobin and lysozyme

Table 2. Amino acid compositions of different positions within defined T-cell epitopes

Position	
1	3A 2L 3F Y W 2T 6E K H 2N Q S 2C G 12 phobic 8 charged 6 polar 1 gly
2	6A 2I Y 3T 4D E 3R 3N 2S 2C 2G P 12 phobic 8 charged 7 polar 3 gly and pro
3	2V 6L I T F 2Y 2M E K H 4S 2Q C 3G 2P 15 phobic 3 charged 6 polar 5 gly and pro
4	3D 8E 9K 5H 3R 2G 28 charged 2 gly
5	7A 3V 5L 5I 2T F 3Y 4M 30 hydrophobic
6	4A 7V 5L 6I 2T 2F Y C 2M 29 hydrophobic 1 polar
7	3A V L I T 3D 5E K 3H 3R 4N 2Q S P 6 phobic 15 charged 7 polar 1 pro
8	3A 2I F Y 3T 2E 4K H R S 2N Q P 2G 10 phobic 8 charged 4 polar 3 pro and gly
9	4A 2V 3L 2I M D 2E 2K H R 3S C 2G 12 phobic 7 charged 4 polar 2 gly

that they are seen by B cells. Consequently, these regions would no longer be intact and could not compose a T-cell epitope. Only those areas not preferentially cleaved would be possible epitopes. We feel that this will be an important pattern which dramatically differs from the sites preferentially seen by B cells. Obviously, the T-cell sites in this model are negatively, not positively, selected for, as are the B-cell sites. We must be cautious in carrying this concept too far, because there are T-cell clones that apparently see similar areas of the influenza haemagglutinin as antibodies (D. B. Thomas, personal communication). A useful model is to view the two groups of epitopes as composing two sets that have overlapping areas, but most of each set is unique.

A major concern with analyses based on either amphipathic character or the secondary and tertiary structure of protein antigens is that if T cells do indeed see protein antigens with their native conformation disrupted, then if there is a pattern, it must manifest itself in the *primary* structure. The epitopes in Table 1 are so listed as to exhibit

the pattern we have discerned. This pattern was determined by examining the residues composing the epitopes as members of characteristic sets based on their physical properties [14]. Table 2 lists the amino acid compositions of the nine positions within the known epitopes listed in Table 1.

As can be seen in Table 2, each of the epitopes has within it a line at sequence composed of (a) a charged residue or gly, followed by (b) two hydrophobic residues. In 22 out of 29 cases, the residue in the next position (7 in the table) was either charged or polar. In the six cases where it was hydrophobic, all had a polar residue in the next position (tyrosine and threonine can act as either polar or hydrophobic residues) [14]. An examination of the literature has revealed several interesting features about this pattern. As can be seen from the way in which the epitopes are listed in Table 1, the pattern exists within the centre of the epitope; rarely is it seen on an extreme edge. In those cases where peptides were synthesized as nested sets to map the minimum residues necessary to stimulate T-cell clones or lines,

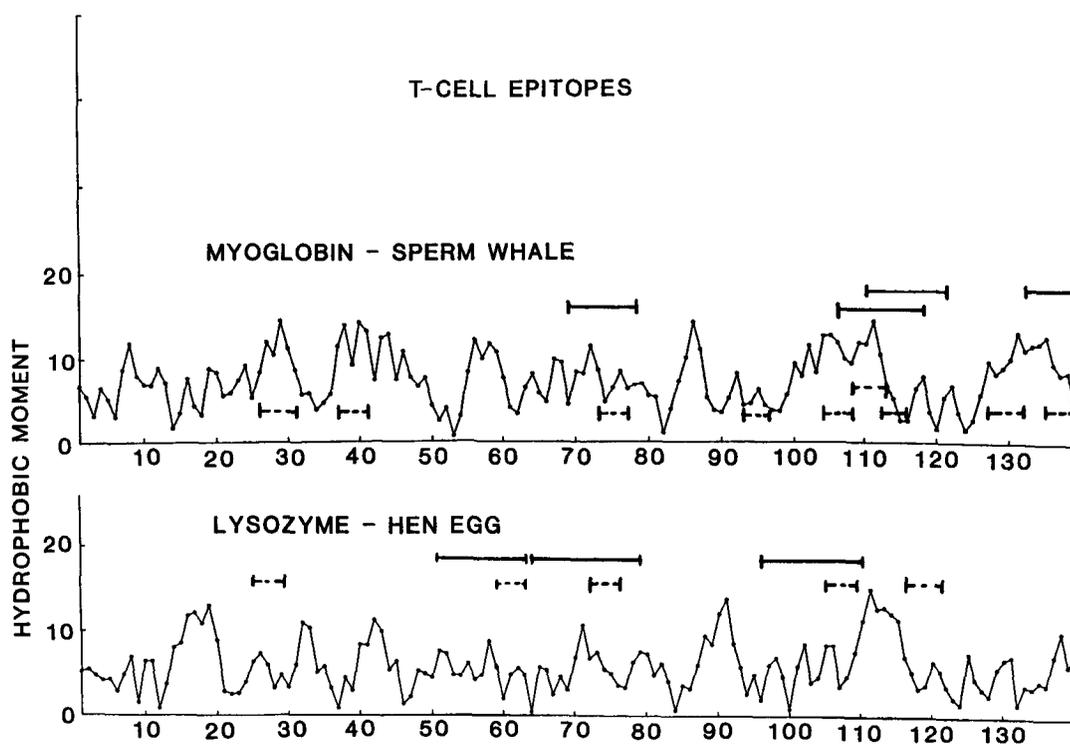


Fig. 2. Hydrophobic moments calculated for sperm whale myoglobin and hen egg lysozyme, plotted against their residue number [12]. The window size was 6 residues and the angle was 100° (hen egg lysozyme sequence contained the leader

of 19 amino acids). Known T-cell epitopes are marked with *solid bars*, while predicted areas containing the pattern of charged or gly, hydrophobic, hydrophobic are marked with *broken bars*

never has a peptide lacking residues composing this pattern been able to stimulate.

In three cases, the pattern exists in two separate areas of a known epitope (ragweed, influenza, haemagglutinin and myoglobin), and we believe that they constitute two overlapping but distinct epitopes which could explain differences in the fine specificities of the clones stimulated by the large peptide. In order to examine both how often this pattern occurs in proteins and how well it correlates with known epitopes, we generated a template for the pattern, using the known compositions of amino acids at the nine sites listed in Tables 1 and 2. We imposed the further restriction that all areas demarcated had to have a charged residue or glycine at position four and hydrophobic residues at positions five and six. The sites predicted for myoglobin and lysozyme are shown in Fig. 2.

As can be seen, many of the T-cell epitopes map to regions with high hydrophobic moments; however, they do not simply correlate with the areas with highest values. In fact, much of both molecules has high (8)

amphipathicity. The pattern described in this report is limited to well-defined regions (9 in myoglobin and 5 in lysozyme) that correlate well with the known epitopes. As previously mentioned, there are separate patterns for the two overlapping epitopes in region 110 in myoglobin. This is illuminating, but it is not a stringent test of the model, because the template used was created from known epitopes. A more useful test is its ability to predict previously unknown epitopes.

The system chosen for analysis was the recognition of influenza nucleoprotein by murine cytotoxic T cells resulting from an infection with intact virus. Previous work in our laboratory has demonstrated that the principal region recognized by the strain of mice lies within the first 77 residues. Figure 3 shows the known (16) and the predicted sites of T-cell recognition in two areas of the protein. On the basis of the pattern described, we synthesized a peptide corresponding to residues 50-63. When used in the chromium release assay, it acted as a substitute for the intact virus, the intact protein and the de-

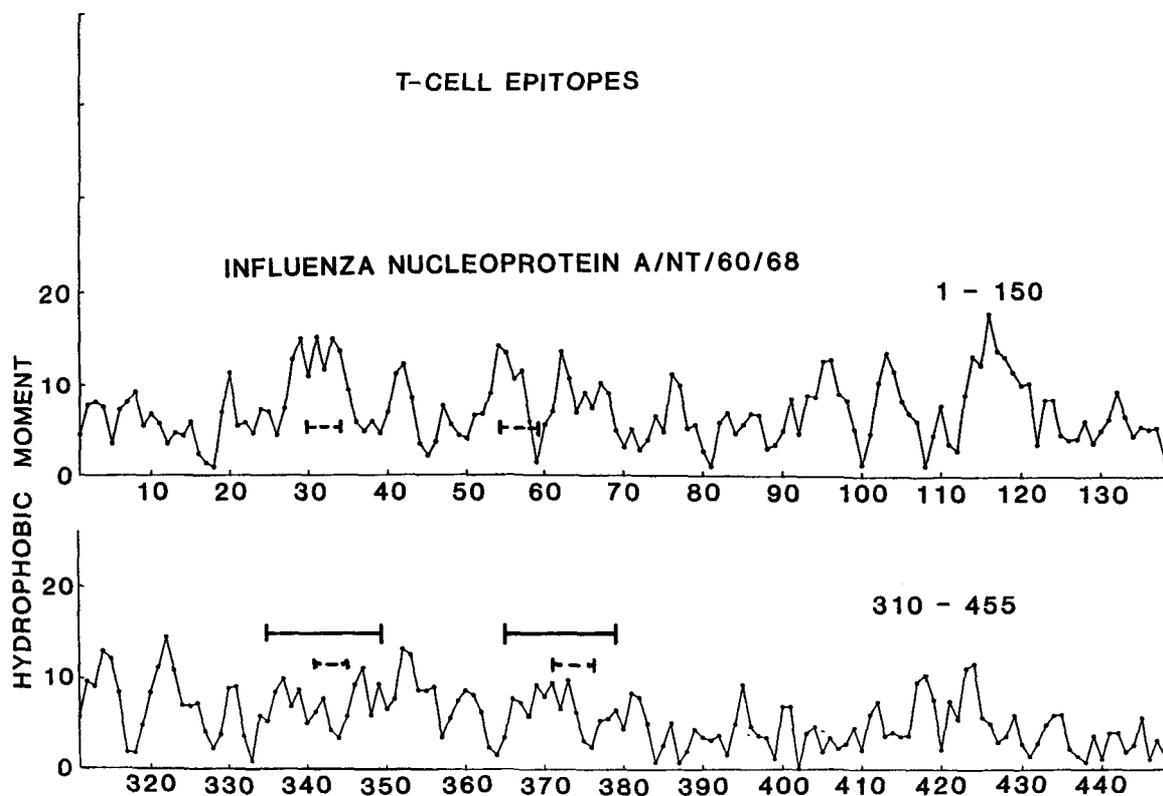


Fig. 3. The plot of hydrophobic moments against residue number of two regions of influenza nucleoprotein. The known epitopes are delineated

with *solid bars*; the areas containing the predicted pattern of charged or glycine, hydrophobic, hydrophobic are delineated with *broken bars*

SPECIFICITY OF CTL LINES FROM CBA MICE

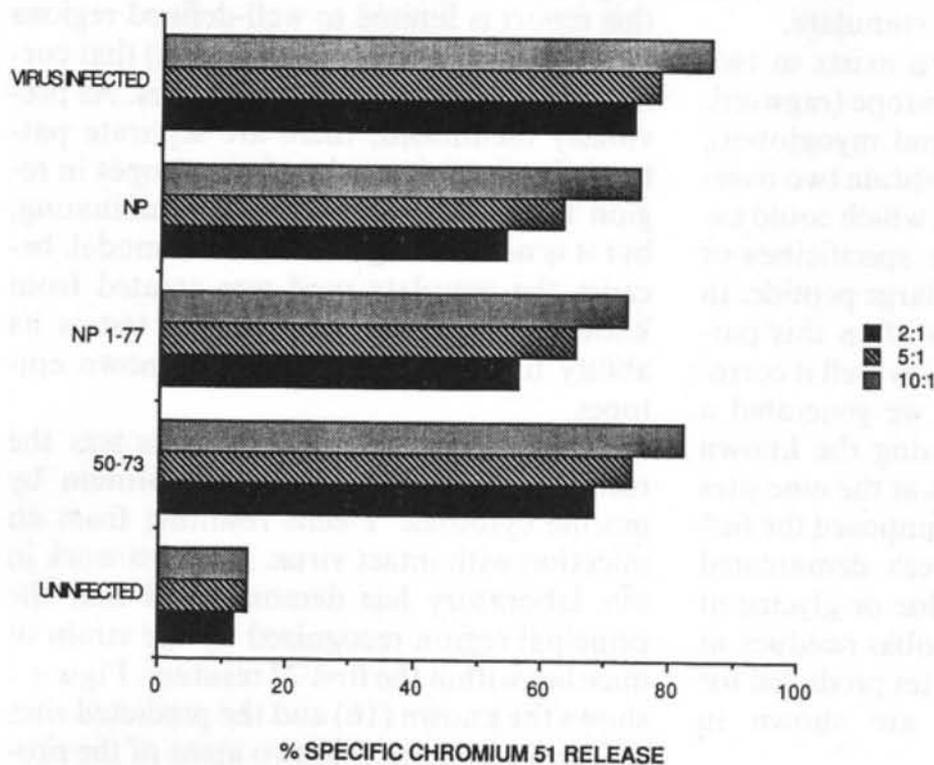


Fig. 4. Results of lysis of chromium-labelled target 2 cells ($H-2^k$) infected with virus, transfected with intact nucleoprotein (*NP*) or a fragment of *NP*, 1–

77, or simply pulsed with peptide corresponding to residues 50–73 as described [16]

leted protein corresponding to residues 1–77 (Fig. 4). From these data, we are able to conclude that this peptide contains all the necessary residues to stimulate the cytotoxic T lymphocytes from this strain of mouse directed at the nucleoprotein of influenza.

In conclusion, we have identified a pattern within the known epitopes of protein antigens recognized by T cells. It is present in 29 out of 30 areas defined. By using the compositions of the residues within these regions from the known epitopes, we have constructed a template with predictive value. Using this method, we have identified a previously undefined epitope within the nucleoprotein of influenza. At this time, we are examining other proteins and other T-cell lines and clones in order to confirm or to contradict the theories described in this report.

References

1. Schwartz RH (1985) *Annu Rev Immunol* 3:237
2. Hendrick S, Nielson E, Kavalier J, Cohen D, Davis MM (1984) *Nature* 308:153

3. Yanagi Y, Yoskikai Y, Leggett K, Clark S, Aleksander I, Mak T (1984) *Nature* 308:145
4. Benjamin D, Berzofsky J, East I, Gurd F, Hannum C, Leach S, Morgoliash E, Michael J, Miller A, Prager E, Reichlin M, Sercarz E, Smith-Gill SJ, Todd P, Wilson A (1984) *Annu Rev Immunol* 2:67
5. Unanue E (1984) *Annu Rev Immunol* 2:395
6. Watts T, Brian A, Kappler J, Marrack P, McConnell H (1985) *Proc Natl Acad USA* 81:7564
7. Hedrick S, Matis L, Hecht T, Samelson L, Longo D, Herber-Katz E, Schwartz R (1982) *Cell* 30:141
8. Herber-Katz E, Hansburg D, Schwartz R (1983) *J Mol Cell Immunol* 1:3
9. Berkower I, Buckenmeyer G, Berzofsky J (1986) *J Immun*
10. Livingstone A, Fathman CG (1986) *Ann Rev Immunol* 5:477
11. DeLisi C, Berzofsky J (1985) *Proc Natl Acad Sci USA* 82:7048
12. Eisenberg D, Weiss R, Terwilliger T (1982) *Nature* 299:371
13. Thornton JM, Edwards MS, Taylor WR, Barlow DJ (1986) *EMBO J* 5:409
14. Taylor WR (1986) *J Mol Biol* 188:233
15. Erickson B, Merrifield B (1976) *The proteins*, vol 2, 3rd edn. Academic, New York, pp 257–493

16. Townsend A, Rothbard J, Gotch F, Bahadut G, Wraith D, McMichael AJ (1986) *Cell* 44:959
17. Townsend A, Skehel J (1984) *J Exp Med* 160:552
18. Townsend A, McMichael A, Carter N, Huddleston J, Brownlee G (1984) *Cell* 39:13
19. Thomas JW, Danho W, Bullesbach E, Fohles J, Rosenthal A (1981) *J Immunol* 126:1095
20. Corradin GP, Juillerat M, Vita C, Ehgers H (1983) *Mol Immunol* 20:763
21. Watts T, Gariepy J, Schoolnik G, McConnell H (1986) *Proc Natl Acad Sci USA* 82:5480
22. Townsend A, Rothbard J (unpublished results)
23. Hackett C, Dietzschold B, Gerhard W, Ghrist B, Knorr R, Gillessen D, Melchers F (1983) *J Exp Med* 158:294
24. Lamb J, Green N (1983) *Immunology* 50:659
25. Zamvil S, Mitchell D, Moore A, Kitamura K, Steinman L, Rothbard J (1986) *Nature* 324:258
26. Ku isaki J, Atassi H, Atassi M (1986) *Eur J Immunol* 16:236
27. Lennon V, McCormick D, Lambert E, Griesmann G, Atassi M (1985) *Proc Natl Acad Sci USA* 82:8805
28. Allen P, Strydom D, Unanue E (1984) *Proc Natl Acad Sci USA* 81:2489
29. Allen P, Matsueda G, Haber E, Unanue E (1986) *Proc Natl Acad Sci USA* 83:4509
30. Manca F, Clarke J, Miller A, Sercarz E, Shastri N (1984) *J Immunol* 133:2075
31. Heber-Katz E, Hollosi M, Dietzschold B, Hudecz F, Fasman G (1986) *J Immunol* 135:1385
32. John Smith (private communication)
33. Francis MJ, Fry CM, Rowlands D, Brown F, Bittle J, Houghton R, Lerner R (1985) *J Gen Virol* 66:2347
34. Westof E, Altschuh D, Moras D, Bloomer A, Mondragon A, Klug A, van Regenmortel M (1984) *Nature* 311:123
35. Tainer J, Getzoff E, Alexander H, Houghton R, Olsen A, Lerner R, Hendrickson W (1984) *Nature* 312:127