



**HAL**  
open science

## Correlation between the location of antigenic sites and the prediction of turns in proteins

Jean-Luc Pellequer, Eric Westhof, Marc H.V. van Regenmortel

► **To cite this version:**

Jean-Luc Pellequer, Eric Westhof, Marc H.V. van Regenmortel. Correlation between the location of antigenic sites and the prediction of turns in proteins. *Immunology Letters*, 1993, 36 (1), pp.83-99. 10.1016/0165-2478(93)90072-A . hal-03232176

**HAL Id: hal-03232176**

**<https://hal.science/hal-03232176>**

Submitted on 26 May 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Correlation between the location of antigenic sites and the prediction of turns in proteins

Jean-Luc Pellequer<sup>a,b</sup>, Eric Westhof<sup>b</sup> and Marc H.V. Van Regenmortel<sup>a</sup>

*UPR Structure des Macromolécules Biologiques et Mécanismes de Reconnaissance, <sup>a</sup>Laboratoire d'Immunochimie and <sup>b</sup>Equipe de Modélisation et de Simulation des Acides Nucléiques, Institut de Biologie Moléculaire et Cellulaire du CNRS, 67084 Strasbourg Cedex, France*

(Received 2 March 1993; accepted 5 March 1993)

## 1. Summary

In the present study, we developed new turn scales based on the occurrence of amino acids at each of the four positions of a turn using a structural database comprised of 87 proteins. We found that the scales correctly predicted a fraction of the turn regions in proteins with approximately 80% confidence. We used the turn scales for predicting the location of antigenic sites in proteins. The method was developed with the specific aim of predicting only a few peaks for each protein (two or three). We found that it leads to a high level of accurate prediction (70% of correct prediction of known epitopes). Our method should be useful for selecting protein regions to be synthesized in order to produce anti-peptide antibodies cross-reacting with the parent protein.

## 2. Introduction

The regions of a protein antigen that are recognized by the combining sites of an antibody are known as antigenic determinants or epitopes. Two types of epitopes are generally distinguished:

*Key words:* Epitope; Antigenicity prediction; Peptide; Turn; Secondary structure

*Correspondence to:* M.H.V. Van Regenmortel, UPR Structure des Macromolécules Biologiques et Mécanismes de Reconnaissance, Laboratoire d'Immunochimie, Institut de Biologie Moléculaire et Cellulaire du CNRS, 15 rue René Descartes, 67084 Strasbourg Cedex, France.

continuous epitopes that are constituted of consecutive residues in the protein sequence and discontinuous epitopes made up of residues distant in the sequence but brought together in the three-dimensional space by the folding of the polypeptide chain. It is generally assumed that most antigenic regions in proteins are constituted of discontinuous epitopes [1–5], a view reinforced by recent X-ray crystallographic studies of antibody-antigen complexes. Until now, six protein-antibody complexes have been solved: three Fab-lysozyme complexes [6–8], two Fab-influenza neuraminidase complexes [9,10] and a Fab-Fab idiotypic complex [11]. In each case, the epitope was clearly discontinuous.

Another commonly used method for delineating epitopes consists in preparing a large number of synthetic peptide fragments of the protein antigen and testing their ability to be recognized by antibodies raised against the protein. This method only identifies so-called continuous epitopes in proteins [4]. A synthetic peptide that mimics a protein epitope will react with the antiprotein antibodies and in addition it is also expected to be able to induce anti-peptide antibodies that recognize the parent protein.

Although some attempts have been made to predict the location of epitopes by analyzing the known three-dimensional structure of a few proteins [12–16] most efforts have concentrated on predicting the location of continuous epitopes only on the basis of the amino acid sequence of the protein [17–19].

A recent comparative study of prediction algorithms [20] has shown that the classical antigenicity prediction methods, which use certain propensity scales of the amino acids such as hydrophilicity [21,22], accessibility [23,24] or flexibility scales [25,26] have a fairly low success rate. The highest level of correct prediction observed in that study [20] was obtained with the turn scale of Levitt [27] which led to 61% of residues being correctly predicted. We decided therefore to investigate further the use of secondary structure prediction, concentrating mainly on turns, for the purpose of antigenicity prediction. Since available turn scales have been constructed using only a few proteins [27–30] we developed a new turn database using a set of well-refined proteins from which the four positions of the turns could be extracted.

### 2.1. Use of turns in antigenicity prediction

A turn is a sequence of the polypeptide backbone involved in a change of direction of the main chain. A turn can be assigned when the distance between the  $C\alpha$  of residue  $i$  and the  $C\alpha$  of residue  $i+3$  is less than 7 Å, except in the case of the central residues of a helix [31]. Only  $\beta$ -turns which are made of four residues [32] were examined in our study. In such turns, a hydrogen bond between the CO of residue  $i$  and the NH of residue  $i+3$  is generally present. The different classes of  $\beta$ -turns (type I, II, IV, VI) were not distinguished except type III, which is a helix  $3_{10}$  fragment and for which a separate scale was made [32].

The location of turns in proteins has been used previously for antigenicity prediction [20,33,34] since turns present properties of surface accessibility, hydrophilicity and mobility known to be correlated with antigenicity. The importance of turns as recognition sites in proteins has been stressed by Rose [35] and it seems that many continuous epitopes are located within turns. It should be noted that turn sequences may also contain loop sequences as discussed by Leszczynski and Rose [36].

NMR and X-ray diffraction studies of peptides have shown that many peptides corresponding to turns in proteins have a high probability of adopting a turn-like structure in water [37–40]. The an-

tigenic properties of such turn peptides have been shown to resemble those of the corresponding regions in the protein [41–44] and correlations between turns and cross-reactive antigenicity have been established in the case of many proteins [45–48]. Krchnak et al. [34] also showed that peptide sequences that are predicted to have a turn conformation, tend to induce antibodies able to cross-react with the parent protein.

The important role played by peptide conformation in determining immunogenicity and antigenicity has been clearly demonstrated in the case of antigenic site A influenza hemagglutinin, since only cyclized synthetic peptides could elicit antibodies capable of reacting with influenza virus [49] and of inducing protective immunity [50]. The superiority of cyclic peptides compared to linear peptides for raising antibodies able to react with the cognate protein has also been demonstrated in other systems [51–53].

Various synthetic strategies have been developed to induce a turn conformation in peptides [54,55] including the incorporation of D amino acids [56,57].

In the present study, we developed new turn scales for improving the accuracy of antigenicity prediction. We shall first describe how our turn database was devised and will then describe how antigenicity prediction scores can be obtained with these scales.

We did not attempt to improve antigenicity prediction by identifying as many epitopes as possible in a particular protein but concentrated instead on the prediction of a few epitopes with the lowest possible rate of incorrect predictions. Indeed, in many cases, an investigator is mainly interested in identifying a cross-reactive continuous epitope that will allow the production of a useful immunological reagent for detecting a gene product, and his main concern is to avoid incorrect epitope assignments that would lead to the synthesis of ineffective peptides.

## 3. Establishment of the databases

### 3.1. Turn database

The turn database contains 87 high resolution ( $<2\text{\AA}$ ) proteins, extracted from the Brookhaven

Protein Data Bank [58], each with a crystallographic R-factor below 0.20 (Table 1). From this database composed of 15938 residues, a total of 793 turns (four positions) could be identified. The turn identification was made with the program of Kabsch and Sander [59] based on the hydrogen bonding pattern. We identified three classes of turns: first the classical turn named '33' which has no preferred localization, second the hairpin turn named 'EE' which is localized between two strands of a  $\beta$ -sheet and, third the  $3_{10}$  helical turn named '10'. The one letter code of Kabsch and Sander was used, i.e., T for turn, B for isolated  $\beta$ -bridge, S for bend, E for sheet and G for  $3_{10}$  helix. The turn class '33' has the hydrogen bond pattern: >33< with the Kabsch and Sander letter code corresponding to TT or TTS or BTT or TTTT. The hairpin class 'EE' has the same hydrogen bond pattern (>33<) and a letter code corresponding to ETTE or ETT or TTE. The  $3_{10}$  turns have a particular hydrogen bond pattern like >>33<< and a letter code like GGG or GGGG.

A bias is present in our database since only hydrogen-bonded turns were taken into account. However, this was done deliberately since it seemed likely that synthetic peptides corresponding to turns in proteins would have a higher probability of adopting a  $\beta$ -turn structure in solution if turn stability could be enhanced by a hydrogen-bond [35,37]

The conformational parameters were determined in the usual manner [27,28,31] using the conformational parameter  $P_{ti} = \frac{f_{ti}}{\langle f_{tj} \rangle} = \frac{(n_i/N_i)}{(N_j/N)}$  where  $f_{ti}$  is an AA frequency;  $n_i$  is the number of residues of a particular kind in position  $i$  (from 1 to 4) and  $N_i$  is the total number of these residues in the database;  $\langle f_{tj} \rangle$  is a secondary structure frequency;  $N_j$  is the total number of residues that occurs in turns and  $N$  is the total number of residues in the whole database.

The method used for predicting the location of turns was based on an individual assignment. This means that, for a four-position turn, we calculated four profiles corresponding, respectively, to the first, second, third and fourth position of a turn. The final value assigned to each residue in the sequence was obtained after adding (ADDI-

TIO routine) or multiplying (MULTIPLI routine) these four files in the following way. The first AA in the first file (which corresponds to the first position of the turn) was added to (or multiplied by) the second AA in the second file and so on until the fourth residue. All the positions were then shifted by one residue and the calculation was continued until the end of the sequence. In order to eliminate the uncertain assignments, a null value was assigned when at least two of the four values were of opposite sign. At the end, all the values were normalized between +3 and -3 in order to facilitate comparisons between different scales [60].

### 3.2. Epitope database

The antigenic database is comprised of 14 proteins in which 82 continuous epitopes have been identified (Table 2). Only continuous epitopes shorter than 18 residues were taken into account. This does not mean that all 18 residues in the peptide were considered to be contact residues of the epitopes [61] but simply that these peptides contain at least one continuous epitope responsible for their antigenic reactivity. In addition, it seemed that the likelihood of finding antigenic reactivity in peptides was bound to increase the longer the peptide [62] and thus that any prediction would increasingly lose its significance with very long peptides. The existence of many continuous epitopes in proteins does present a problem in the evaluation of antigenicity prediction since random allocation of epitopes should lead to a not insignificant level of correct predictions [3].

The tertiary structure of 11 of the proteins has been established either by X-ray study or by homology prediction with known structures. We found that 50% of antigenic residues occurred in turn and loop regions while nearly 35% occurred in helices (Table 3). As epitopes are usually located at the surface of proteins, it could be expected that helical epitopes would belong to amphiphilic helices. Such helices can also be predicted with the PREDITOP program [60].

### 3.3. Evaluation of the accuracy of predictions

The use of a window of five residues in length

TABLE 1

The 87 three-dimensional structures of proteins used in this study<sup>a</sup>.

PDB code	Proteins	Resolution	R-Factor	Length
1 AAP	PROTEASE INHIBITOR DOMAIN OF ALZHEIMER AMYLOID	1.5	0.177	58
1 ACX	ACTINOXANTHIN	2	—	108
1 ALC	α-LACTALBUMIN	1.7	0.22	123
1 BBP	BILIN BINDING PROTEIN	2	0.2	173
1 BP2	PHOSPHOLIPASE A2	1.7	0.171	123
1 CRN	CRAMBIN	1.5	0.114	46
1 CSE	SUBTILISIN CARLSBERG chain E	1.2	0.178	274
1 ECA	HEMOGLOBIN	1.4	0.19	136
1 ER8	ENDOTHAPEPSIN chain E	2	0.17	330
1 GCR	g-II CRYSTALLIN	1.6	0.23	174
1 GD1	GLYCERALDEHYDE 3 P DEHYDROGENASE	1.8	0.177	334
1 GOX	GLUCOLATE OXIDASE	2	0.189	369
1 GPI	GLUTATHIONE PEROXIDASE	2	0.171	198
1 HOE	α-AMYLASE INHIBITOR	2	0.199	74
1 MBD	MYOGLOBIN	1.4	—	153
1 NXB	NEUROTOXIN B	1.38	0.24	62
1 PAZ	PSEUDOAZURIN	1.55	0.18	123
1 PCY	PLASTOCYANIN	1.6	0.17	99
1 PPT	AVIAN PANCREATIC POLYPEPTIDE	1.37	—	36
1 PSG	PEPSINOGEN	1.65	0.173	370
1 R69	434 REPRESSOR	2	0.193	69
1 RNH	SELENOMETHIONYL RIBONUCLEASE H	2	0.198	155
1 SGT	TRYPSIN	1.7	0.161	223
1 SN3	SCORPION NEUROTOXIN 3	1.8	0.16	65
1 UBQ	UBIQUITIN	1.8	0.176	76
1 XY1	β-MERCAPTOPROPIONATE-OXYTOCIN	1.04	0.088	8
2 ACT	ACTINIDIN	1.7	0.171	219
2 ALP	α-LYTIC PROTEASE	1.7	0.131	198
2 APP	ACID PROTEINASE	1.8	0.136	323
2 APR	ACID PROTEINASE	1.8	0.143	325
2 AZA	AZURIN	1.8	0.157	129
2 B5C	CYTOCHROME C5	2	—	93
2 CAB	CARBONIC ANHYDRASE B	2	0.193	260
2 CCY	CYTOCHROME C	1.67	0.188	128
2 CDV	CYTOCHROME C3	1.8	0.176	107
2 CGA	CHYMOTRYPSINOGEN A	1.8	0.173	245
2 CI2	CHYMOTRYPSIN INHIBITOR 2	2	0.198	82
2 CNA	CONCAVALIN A	2	—	237
2 CPP	CYTOCHROME P450	1.63	0.19	414
2 CTS	CITRATE CYNTHASE COMPLEX	2	0.161	437
2 CYP	CYTOCHROME C PEROXIDASE	1.7	0.202	294
2 FB4	FAB	1.9	0.189	445
2 LH1	LEGHEMOGLOBIN	2	—	153
2 LHB	HEMOGLOBIN V	2	0.142	149
2 LZM	LYSOZYME	1.7	0.193	164
2 MHR	MYOHEMERYTHRIN	1.7	0.158	118
2 OVO	OYOMUCOID THIRD DOMAIN	1.5	0.199	56
2 PRK	PROTEINASE K	1.5	0.167	279
2 PTN	TRYPSIN	1.55	0.193	223
2 RHE	BENCE JONES PROTEIN	1.6	0.149	114

PDB code	Proteins	Resolution	R-Factor	Length
2 SGA	SUBTILISIN CARLSBERG COMPLEX WITH EGLIN	1.5	0.126	181
2 SNS	STAPHYLOCOCCAL NUCLEASE COMPLEX	1.5	0.19	149
2 TRX	THIOREDOXIN	1.68	0.165	108
2 TSC	THYMIDYLATE SYNTHASE	1.97	0.18	264
2 UTG	UTEROGLOBULIN	1.64	0.19	70
3 BLM	$\beta$ -LACTAMASE	2	0.163	257
3 C2C	CYTOCHROME C2	1.68	0.175	112
3 CLA	CHLORAMPHENICOL ACETYLTRANSFERASE	1.75	0.183	213
3 DFR	DIHYDROFOLATE REDUCTASE	1.7	0.152	162
3 EBX	ERABRUTOXIN B	1.4	0.176	62
3 EST	NATIVE ELASTASE	1.65	0.169	240
3 GRS	GLUTATHIONE REDUCTASE	1.54	0.186	478
3 INS	INSULIN	1.5	0.182	51
3 LYM	LYSOZYME	2	0.149	129
3 RNT	RIBONUCLEASE T1	1.8	0.137	104
3 RP2	RAT MAST CELL PROTEASE II	1.9	0.191	224
3 SGB	PROTEINASE B chain E	1.8	0.125	241
3 WGA	WHEAT GERM AGGLUTININ	1.8	0.179	170
4 51C	CYTOCHROME C551	1.6	0.187	82
4 DFR	DIHYDROFOLATE REDUCTASE	1.7	0.155	159
4 FD1	FERREDOXIN	1.9	0.192	106
4 FXN	FLAVODOXIN	1.8	0.2	138
4 HHB	HEMOGLOBIN	1.74	0.135	287
4 I1B	INTERLEUKIN-1 $\beta$	2	0.19	153
4 PTI	TRYPSIN INHIBITOR	1.5	0.162	58
5 CPA	CARBOXYPEPTIDASE A	1.54	0.19	307
5 CYT	CYTOCHROME C	1.5	0.159	103
5 HVP	HIV-1 PROTEASE	2	0.176	99
5 RXN	RUBREDOXIN	1.2	0.096	54
5 TNC	TROPONIN C	2	0.155	162
6 LDH	APO LACTATE DEHYDROGENASE	2	0.202	329
6 TMN	THERMOLYSIN COMPLEX	1.6	0.171	316
7 RSA	RIBONUCLEASE A	1.26	0.15	124
8 ABP	L-ARABINOSE BINDING PROTEIN	1.49	0.175	306
8 DFR	DIHYDROFOLATE REDUCTASE	1.7	0.188	189
9 PAP	PAPAIN	1.65	0.161	212
9 XIA	D-XYLOSE ISOMERASE	1.9	0.141	388

\*The mean resolution of this database is 1.72 Å with a mean R-factor of 0.173. The hyphen sign means that the R-factor is not present in the PDB file. Care has been taken not to choose proteins which have more than 30% residue identity. The minimum acceptable resolution was 2 Å.

centered on the third residue was found to be better than a seven residue length window for predicting turns (result not shown). The smoothing procedure has been described elsewhere [18]. The test for evaluating the accuracy of turn prediction was performed as follows. Only residues above the threshold of 0.7 times the standard deviation around the mean of all points were counted. Two categories were formed, the A class (predicted residues localized in known turn regions) and the C

class (predicted residues outside any known turn structure). The prediction results were expressed by the ratio  $\frac{A}{A+C}$  which is equal to 1 if there is no incorrect prediction and to 0 if not a single predicted residue belongs to any known turn region. It should be pointed out that residues and not regions are counted, which allows a rigorous evaluation of the prediction. Only peaks were taken into account since the major application of

TABLE 2

List of the 85 continuous selected epitopes<sup>a</sup>.

Proteins	Epitopes	References	Proteins	Epitopes	References		
CHO	12-21	ref. 73	MYO	1-6	ref. 91		
	32-40	ref. 73		15-22	ref. 92		
	48-63	ref. 73		48-55	ref. 93		
	83-97	ref. 74		56-62	ref. 92		
CYT	1-4	ref. 75		94-99	ref. 92		
	13-25	ref. 76		113-119	ref. 92		
	42-50	ref. 77		121-127	ref. 91		
	56-73	ref. 78		145-151	ref. 92		
	HBV	2-16		ref. 79	PIL	5-12	ref. 94
22-35		ref. 80		38-50		ref. 94	
48-65		ref. 81	65-75	ref. 94			
69-79		ref. 81	93-104	ref. 94			
95-109		ref. 79	103-116	ref. 94			
125-139		ref. 79	119-131	ref. 94			
HCG	139-147	ref. 82	RAS	1-18	ref. 95		
	1-7	ref. 83		29-44	ref. 95		
	40-52	ref. 84		64-76	ref. 96		
	92-105	ref. 83		91-108	ref. 95		
	IFB	110-122	ref. 85	REN	50-60	ref. 97	
134-145		ref. 85	63-71		ref. 97		
7-22		ref. 86	81-90		ref. 97		
40-47		ref. 86	118-126		ref. 97		
73-81		ref. 86	133-144		ref. 98		
LEG	105-112	ref. 86	162-169		ref. 97		
	117-120	ref. 86	180-188		ref. 98		
	15-23	ref. 87	211-224		ref. 98		
	52-59	ref. 87	247-255		ref. 97		
	92-98	ref. 87	287-295		ref. 97		
LYS	107-116	ref. 87	SCO	1-14	ref. 99		
	132-142	ref. 87		27-35	ref. 99		
	38-54	ref. 88		36-46	ref. 99		
	64-80	ref. 89		55-63	ref. 99		
MHR			TMV	1-10	ref. 100		
				4-9	ref. 90	19-32	ref. 101
				16-21	ref. 90	34-39	ref. 102
				37-46	ref. 13	55-61	ref. 100
				54-58	ref. 90	62-68	ref. 103
				63-72	ref. 90	76-88	ref. 101
				80-85	ref. 90	103-112	ref. 104
				90-95	ref. 90	134-146	ref. 101
				110-115	ref. 90	149-158	ref. 102

<sup>a</sup>Abbreviations are: CHO, cholera toxin; CYT, cytochrome c; HBV, surface antigen of the hepatitis B virus; HCG, human chorionic gonadotropin hormone; IFB,  $\beta$ -interferon; LEG, leghemoglobin; LYS, hen-egg lysozyme; MHR, myohemerythrin; MYO, myoglobin; PIL, Gal-Gal pyelonephritis *E. coli* pili; RAS, h-RAS p21 oncogene; REN, human renin; SCO, scorpion neurotoxin; TMV, tobacco mosaic virus protein. All these fragments have been selected for the cross-reactivity of antibodies raised against peptides with the native proteins. All the three dimensional structures have been extracted from the PDB except IFB which is a model.

TABLE 3

Secondary structure composition of epitopes in 11 known three-dimensional structures<sup>a</sup>.

Protein codes <sup>b</sup>	Number of antigenic amino acids in			Total	Protein length	Structure references
	Helices	Sheets	Turns			
CHO	5	24	21	50	103	ref. 105
CYT	21	0	23	44	104	ref. 106
IFB	22	0	23	45	166	ref. 86
LEG	40	0	5	45	153	ref. 107
LYS	4	13	38	55	129	ref. 108
MHR	29	0	26	55	118	ref. 109
MYO	36	0	20	56	153	ref. 110
RAS	41	16	28	85	189	ref. 111
REN	4	29	78	111	340	ref. 112
SCO	6	10	27	43	65	ref. 113
TMV	35	6	49	90	158	ref. 114
TOTAL	243	98	338	679	1678	
RATIO	35.79	14.43	49.79	100	40.46	

<sup>a</sup>The protein codes are listed in Table 2. <sup>b</sup>The ratio cells show that 50% of antigenic determinants are located in turns while 35% are found in helices and 15% in  $\beta$ -sheets. The total cells indicate the sum of the column. The last ratio of the protein length column indicates that nearly 40% of amino acids of those proteins are antigenic. This reveals that our antigenic database is not biased toward too small proteins.

antigenicity prediction is to select peptides for synthesis and this is based only on positive correlation.

The test for the accuracy of antigenicity prediction was performed as for the turn prediction, except that the comparisons are now made with the epitopes of Table 2. It should be noted that the precision of data on epitope localization in peptides, that may be as long as 18 residues, is not as good as that of the localization of turns in crystallographic structures of proteins. This means that the sensitivity of the A/A+C criterion for accurate prediction is not the same in both cases.

## 4. Results

### 4.1. Description of the turn databases

The comparison of our turn frequency values with those of Chou and Fasman [30] presents some discrepancies for some amino acids. Unfortunately, it is not possible to distinguish errors attributable to the limitation of the size of the Chou and Fasman database from those caused by the presence in that database of poorly refined structures [63].

The most striking difference concerns the underestimation of Gly and the overestimation of Met and Cys at the 2nd and the 3rd position in the Chou and Fasman frequencies (Table 4). Interestingly the value of Trp at the 4th position is the highest in both the Chou and Fasman database and in our own database (Table 4).

In general, the frequencies of the different residues at the 2nd and 3rd position of a turn are very similar in both databases. It is well-known that the two middle positions of a turn ( $i+1$ ,  $i+2$ ) are the most important ones for a correct chain reversal [32]. The important role of these two positions is supported by a comparison of the four positions in the turns in each database (i.e., '33', '10', 'EE'). The frequency distribution of each AA in the three databases gave a lower standard deviation for the 2nd and the 3rd position (Table 4). This is in line with the observation that only a few amino acids (Pro, Gly, Ser, Asn, Asp for instance) occurred in these positions.

The database corresponding to TURN10 is closer to a helix database than to a typical turn class since component AAs like Ala, Glu, Asp, Ser are mainly found in helices [27]. However, some typical turn AAs such as Pro and Asn were



TABLE 4

List of the conformational parameters obtained for each of the four turn databases<sup>a</sup>.

Amino acids	Position 1 in the turn database			MEAN	SDV	Amino acids	Position 2 in the turn database			MEAN	SDV
	TURN33	TURNE E	TURN10				TURN3	TURNE E	TURN10		
A	0.67	0.62	1.00	0.76	0.21	A	0.86	0.62	1.24	0.91	0.31
C	1.37	0.85	1.29	1.17	0.28	C	0.29	1.13	0.64	0.69	0.42
D	2.30	1.46	2.39	2.05	0.51	D	0.93	1.68	0.77	1.13	0.49
E	0.62	0.77	0.39	0.59	0.19	E	1.20	1.03	1.66	1.30	0.33
F	1.09	0.70	1.59	1.13	0.45	F	0.54	0.52	0.80	0.62	0.16
G	0.78	0.87	1.16	0.94	0.20	G	0.63	2.24	0.94	1.27	0.85
H	2.11	0.91	3.24	2.09	1.17	H	0.21	0.91	0.46	0.53	0.35
I	0.36	1.56	0.20	0.71	0.74	I	0.32	0.52	0.49	0.44	0.11
K	0.56	0.76	0.33	0.55	0.22	K	1.31	0.86	1.32	1.16	0.26
L	0.58	0.70	0.87	0.72	0.15	L	0.33	0.44	0.87	0.55	0.29
M	0.38	0	0	0.13	0.22	M	0.38	0	0.83	0.40	0.42
N	1.72	1.92	1.36	1.67	0.28	N	0.62	2.06	0.31	1.00	0.93
P	1.29	0.60	2.50	1.46	0.96	P	4.19	1.64	3.64	3.16	1.34
Q	1.06	0.54	0.14	0.58	0.46	Q	0.75	0.54	0.41	0.57	0.17
R	0.52	1.31	1.14	0.99	0.42	R	1.37	0.19	1.00	0.85	0.60
S	1.59	1.08	0.70	1.12	0.45	S	2.00	1.50	1.27	1.59	0.37
T	1.13	0.51	0.54	0.73	0.35	T	1.31	0.92	0.47	0.90	0.42
V	0.66	1.98	0.34	0.99	0.87	V	0.34	0.36	0.82	0.51	0.27
W	0.60	1.30	2.31	1.40	0.86	W	0.45	0	0.99	0.48	0.50
Y	0.89	0.68	1.04	0.87	0.18	Y	0.83	0.17	0.65	0.55	0.34
SUM	20.28	19.12	22.53	20.64	1.73	SUM	18.86	17.33	19.58	18.59	1.15

Amino acids	Position 3 in the turn database			MEAN	SDV	Amino acids	Position 4 in the turn database			MEAN	SDV
	TURN33	TURNE E	TURN10				TURN3	TURNE E	TURN10		
A	0.57	0.54	0.95	0.69	0.23	A	0.83	0.93	1.30	1.02	0.25
C	0.49	0.28	1.29	0.69	0.53	C	1.47	0.56	1.72	1.25	0.61
D	1.52	1.35	2.22	1.70	0.46	D	1.44	0.79	1.88	1.37	0.55
E	1.07	0.9	2.25	1.41	0.74	E	1.07	0.77	0.98	0.94	0.15
F	0.73	0.87	0.13	0.58	0.39	F	0.66	0.87	1.46	1.00	0.41
G	3.42	4.05	0.33	2.60	1.99	G	1.93	1.30	0.50	1.24	0.72
H	0.95	0.3	1.85	1.03	0.78	H	0.84	0.30	1.85	1.00	0.79
I	0.09	0.13	0	0.07	0.07	I	0.45	1.04	0.69	0.73	0.30
K	0.94	0.43	1.32	0.90	0.45	K	0.67	1.19	1.23	1.03	0.31
L	0.27	0.26	0.33	0.29	0.04	L	0.67	0.18	1.07	0.64	0.45
M	0.25	0.37	0.28	0.30	0.06	M	1.01	1.83	0.56	1.13	0.64
N	2.24	1.79	1.99	2.01	0.23	N	0.86	0.96	1.67	1.16	0.44
P	0.16	0.3	0.68	0.38	0.27	P	0	0	0	0	0
Q	0.69	0.54	1.23	0.82	0.36	Q	1.13	2.52	1.10	1.58	0.81
R	0.85	1.13	0.43	0.80	0.35	R	1.04	1.13	1.71	1.29	0.36
S	1.1	1.42	2.6	1.71	0.79	S	1.13	1.25	0.70	1.03	0.29
T	0.67	0.61	0.78	0.69	0.09	T	1.06	0.92	0.39	0.79	0.35
V	0.09	0.09	0.27	0.15	0.10	V	0.87	1.17	0.14	0.73	0.53
W	0.75	1.3	0	0.68	0.65	W	1.20	2.17	1.65	1.67	0.49
Y	0.95	0.34	0.36	0.55	0.35	Y	1.24	0.85	1.82	1.30	0.49
SUM	17.8	17	19.29	18.03	1.16	SUM	19.57	20.73	22.42	20.91	1.43

<sup>a</sup>The values are given by the formula:  $(n_i/N_i)/(N_j/N)$ ; where N is the total number of amino acids in the data base (N = 15938), N<sub>j</sub> is the total number of residues occurring in turns (N<sub>j</sub> = 793 for the turn column), n<sub>i</sub> is the number of residues in position i (n<sub>1</sub> = 50 for Ala in the first position of the turn scale) and N<sub>i</sub> is the total number of such a residue in the whole database (N<sub>1</sub> = 1342 for Ala). The mean column corresponds to the mean value for each AA of the three turn database, one position at a time. It should be noted that the standard deviations of the means are lower for the two middle positions (namely the positions 2 and 3).

found but no significant quantity of Gly.

There were strong differences between the TURN33 and TURNEE databases. Most differences appeared in the first and the fourth position where hydrophobic AAs like Val and Ile, which have a typically  $\beta$ -sheet AA preference, are more prevalent in the TURNEE database. Furthermore, the two middle positions in the TURNEE database have few large AA like Trp and Tyr in the 2nd position or His and Tyr in the third position (Table 4) and charged AAs are nearly always excluded from those two positions compared to the TURN33 database.

The diversity in amino acid composition for each turn position implies that a global database (i.e., the mean of the frequencies in the four databases) would lead to a smoothing of those conformational subtleties and could lead to possible misinterpretation in the location of turns in proteins.

#### 4.2. Description of the turn predictions

The three turn databases (TURN10, TURNEE, TURN33) were normalized between the values +3 and -3 by the PREDITOP package [60] and were used as propensity scales for secondary structure predictions. The PREDITOP package

was run for all the prediction curves and the ADDITIO and MULTIPLI routines were used as described [60]. Cross validation of our prediction method was not considered to be necessary since the removal of one protein from a database of the size we used would have led to only a very small change in the data base.

We used the calculated files to localize turns or loops in all the proteins of the database. The results for the three scales are given in Table 5. We obtained a mean, with all the scales, of 80% of correctly predicted peaks using the MULTIPLI routine and nearly 70% when using the ADDITIO routine. The result files obtained by the multiplication routine have only a few predicted peaks but with only few incorrect predictions, while the addition result files predict five times more peaks than the multiplication routine but with a ten percent point increase in incorrect predictions. It is interesting to note that the best prediction results obtained by Chou and Fasman [30] was 70% which corresponds to our addition procedure score. This means that the ADDITIO routine could be useful for predicting turns in a secondary structure prediction. If we compare the addition results, our results are closer to those obtained with the neural network method [64,65]

TABLE 5

Turn predictions using our own turn databases<sup>a</sup>.

Turn database		A	C	A/A+C
TURN10	multiplication	127	22	0.85
	addition	1555	990	0.61
TURNEE	multiplication	450	148	0.75
	addition	1818	912	0.67
TURN33	multiplication	320	72	0.82
	addition	1994	811	0.71
Mean ratio A/A+C for	multiplication	<b>0.81</b>		
	addition	<b>0.66</b>		

<sup>a</sup>The predictions are obtained by multiplying (first lines) or by adding (second lines) the profiles corresponding to positions 1 to 4 in a turn (detailed in the text). The three turn databases have been tested on the 87 proteins from the BRK database. The A column corresponds to the correctly predicted amino acids above the threshold of 0.7 times the standard deviation; this means that the predicted amino acids are located in known loop regions (non-helix and non-sheet). The C column corresponds to the incorrect predictions above the threshold of 0.7 times the standard deviation; this means that the predicted peaks are located outside of any known loops. The accuracy ratio  $\frac{A}{A+C}$  gives the percentage of correct predictions.

which gave 70% of correctly predicted residues in turns. However, the results obtained with the multiplication are closer to those obtained with the pattern approach developed with well known protein families [66,67] which gave up to 90% of correct prediction using homologous proteins. Nevertheless, our method is less accurate regarding specific turns than the turn assignment prediction done by Wilmot and Thornton [31] who analyzed the different types of  $\beta$ -turns in proteins by searching locations of typical AAs in well known types of turns namely, type I, type II as well as in non-specific turns where specific side-chain interactions could be the basis for turn conformation. It should be emphasized that our method does not attempt to predict the location of all turns in a protein but only tries to identify regions where turns have a high probability of occurrence. Our results indicate that our turn prediction method is indeed sufficiently accurate for the purpose of antigenicity prediction.

#### 4.3. Correlation between predicted turns and antigenic sites

There is considerable evidence that antigenic sites are often located in the turns of proteins. In the present study we did not attempt to examine the correlation between antigenicity and the location of actual turns in proteins of known structure, but to use the likely correlation between predicted turns and the location of antigenic sites in order to predict antigenicity. Furthermore, our major aim was not to use the turn scales for identifying all antigenic regions of a protein but to use them for identifying with a high level of accuracy at least a couple of the antigenic regions of a protein.

For this purpose we computed four files, each one representing a position in a four residue turn. The calculation was based on a window assignment of five residue length centered on the third one [60]. This window size was preferred over the four residue window size because of the proximity effect obtained by an odd centered window. Each position of the window was smoothed by a Gaussian function as already described [60]. Then we added or multiplied those four result files in order to eliminate incorrect predictions. One value

in the result file corresponds to one turn in the protein (four residues), leading to sharp predicted peaks. The MULTIPLI routine was principally used to evaluate antigenicity predictions. All the results are presented in Tables 5-9. The left columns of the tables represent the predictions obtained with the MULTIPLI routine while the right columns correspond to the ADDITIO routine.

The TURN10 scale was the least successful in antigenicity prediction. This may be due to the fact that such turns are fragments of  $3_{10}$  helices. This is revealed by the fact that peaks fall in completely different regions from those predicted with the other turn scales. However, we can notice that

TABLE 6

Ratio between correctly and incorrectly predicted amino acids in epitopes using the TURN10 scale<sup>a</sup>.

Protein codes	Multiplication			Addition		
	A	C	A/(A+C)	A	C	A/(A+C)
CHO	0	0	—	11	4	0.73
CYT	0	1	0	9	5	0.64
HBV	0	2	0	15	14	0.52
HCG	4	1	0.80	12	6	0.67
IFB	1	3	0.25	4	17	0.19
LEG	0	1	0	12	17	0.41
LYS	0	1	0	9	5	0.64
MHR	0	0	—	4	11	0.27
MYO	0	0	—	6	10	0.38
PIL	0	1	0	13	19	0.41
RAS	1	1	0.50	18	9	0.67
REN	5	1	0.83	26	18	0.59
SCO	2	3	0.40	3	4	0.43
TMV	3	0	1	18	7	0.72
Mean	<b>1.14</b>	<b>1.07</b>	<b>0.34<sup>b</sup></b>	<b>11.43</b>	<b>10.43</b>	<b>0.52<sup>b</sup></b>

<sup>a</sup>The A and C columns correspond to the same values as in the legend of Table 5, except that the prediction is compared to the epitope location. It is noteworthy that the A/(A+C) ratios do not correspond to a global antigenicity prediction, but rather to a ratio of no incorrect prediction. The three scales are represented as well as the total turn scale. The first columns correspond to the multiplication of the four predicted files of each turn position and the second columns correspond to the addition. <sup>b</sup>The removal of the five proteins (CYT, LEG, LYS, MHR and MYO) that are common to the turn and antigenicity data bases led to a change of the value 0.34 to 0.47 and of the value 0.52 to 0.55, indicating an even better prediction score.

TABLE 7

Ratio between correctly and incorrectly predicted amino acids using the TURNEE scale<sup>a</sup>.

Protein codes	Multiplication			Addition		
	A	C	A/(A+C)	A	C	A/(A+C)
CHO	2	1	0.67	10	3	0.77
CYT	2	0	1	6	3	0.67
HBV	7	1	0.88	20	19	0.51
HCG	4	2	0.67	12	7	0.63
IFB	9	1	0.90	14	8	0.64
LEG	3	8	0.27	9	20	0.31
LYS	2	0	1	5	5	0.50
MHR	3	0	1	7	8	0.47
MYO	2	0	1	10	14	0.42
PIL	8	2	0.80	23	4	0.85
RAS	4	4	0.50	15	20	0.43
REN	4	10	0.29	21	40	0.34
SCO	1	1	0.50	9	2	0.82
TMV	2	3	0.40	18	7	0.72
Mean	<b>3.79</b>	<b>2.36</b>	<b>0.70<sup>b</sup></b>	<b>12.79</b>	<b>11.43</b>	<b>0.58<sup>b</sup></b>

<sup>a</sup>Same comments as in Table 6.

<sup>b</sup>Without the five proteins present in both data bases, the value 0.70 became 0.62 and the value 0.58 became 0.63 (see Table 6).

the TURN10 scale is the only one that gave a prediction ratio of 1 for TMV protein (Table 6). It is also surprising that this scale gave better results with the ADDITIO routine than with the MULTIPLI routine. This indicates that this turn scale is not very reliable for antigenicity predictions.

The TURNEE scale (hairpin) gave the best result when the multiplication routine was used, i.e., a mean ratio (A/A+C) of 70% (Table 7). One can notice that the predicted antigenic peaks are often found in hairpin turns in proteins of known three-dimensional structures. If a peptide adopts a hairpin structure in solution, such a scale could be of great advantage in antigenicity prediction. In the many instances where the TURNEE scale leads to incorrect antigenicity predictions for proteins of known three-dimensional structure, the incorrectly predicted peaks fall within hairpin turns of the three-dimensional structure (Fig. 1) indicating that the incorrect antigenicity predictions are not a consequence of wrong turn assignments. As expected we observed better results with the multiplication than with the addition

TABLE 8

Ratio between correctly and incorrectly predicted amino acids using the TURN33 scale<sup>a</sup>.

Protein codes	Multiplication			Addition		
	A	C	A/(A+C)	A	C	A/(A+C)
CHO	2	0	1	6	2	0.75
CYT	0	1	0	8	5	0.62
HBV	2	4	0.33	25	14	0.64
HCG	6	0	1	17	10	0.63
IFB	7	0	1	14	16	0.47
LEG	0	2	0	6	13	0.32
LYS	4	0	1	18	0	1
MHR	1	1	0.50	5	7	0.42
MYO	1	0	1	12	8	0.60
PIL	1	2	0.33	23	3	0.88
RAS	1	0	1	10	9	0.53
REN	5	1	0.83	20	22	0.48
SCO	1	0	1	7	0	1
TMV	2	1	0.67	12	6	0.67
Mean	<b>2.36</b>	<b>0.86</b>	<b>0.69<sup>b</sup></b>	<b>13.07</b>	<b>8.21</b>	<b>0.64<sup>b</sup></b>

<sup>a</sup>Same comments as in Table 6.

<sup>b</sup>Without the five proteins present in both data bases, the value 0.69 became 0.79 and the value 0.64 became 0.67 (see Table 6).

TABLE 9

Ratio between correctly and incorrectly predicted amino acids using the Lewitt scale<sup>a</sup>.

Protein codes	A	C	A/(A+C)
CHO	16	8	0.67
CYT	12	12	0.50
HBV	41	21	0.66
HCG	17	15	0.53
IFB	18	19	0.49
LEG	17	18	0.49
LYS	29	2	0.94
MHR	17	7	0.71
MYO	16	17	0.48
PIL	27	12	0.69
RAS	23	18	0.56
REN	35	48	0.42
SCO	13	1	0.93
TMV	28	8	0.78
Mean	<b>22.07</b>	<b>14.71</b>	<b>0.63</b>

<sup>a</sup>A window length of five residues centered on the third was used. The window is smoothed by a Gaussian function as in the addition or multiplication routines. The threshold is 0.7 times the standard deviation.

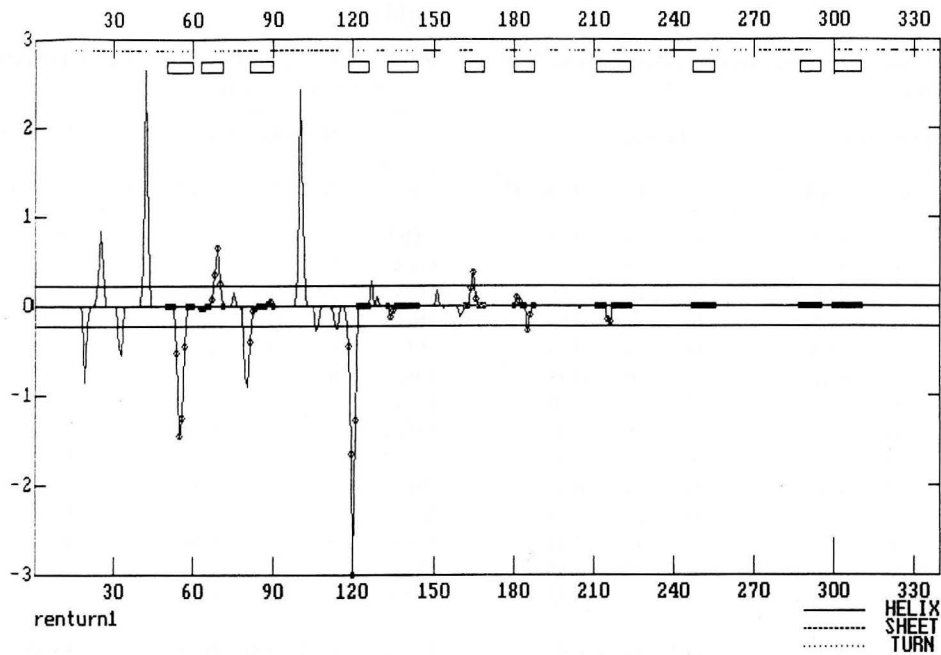


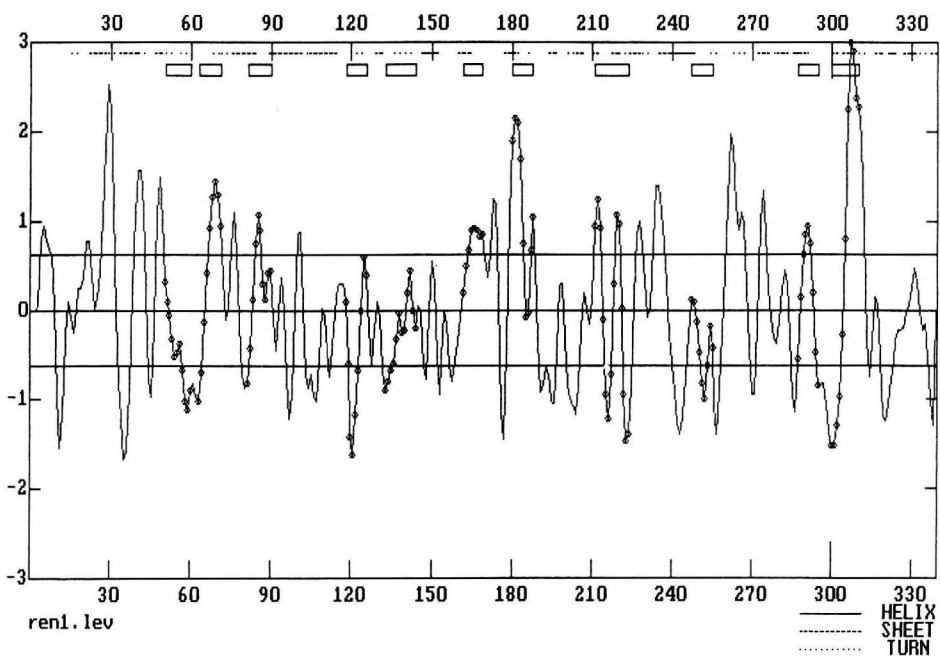
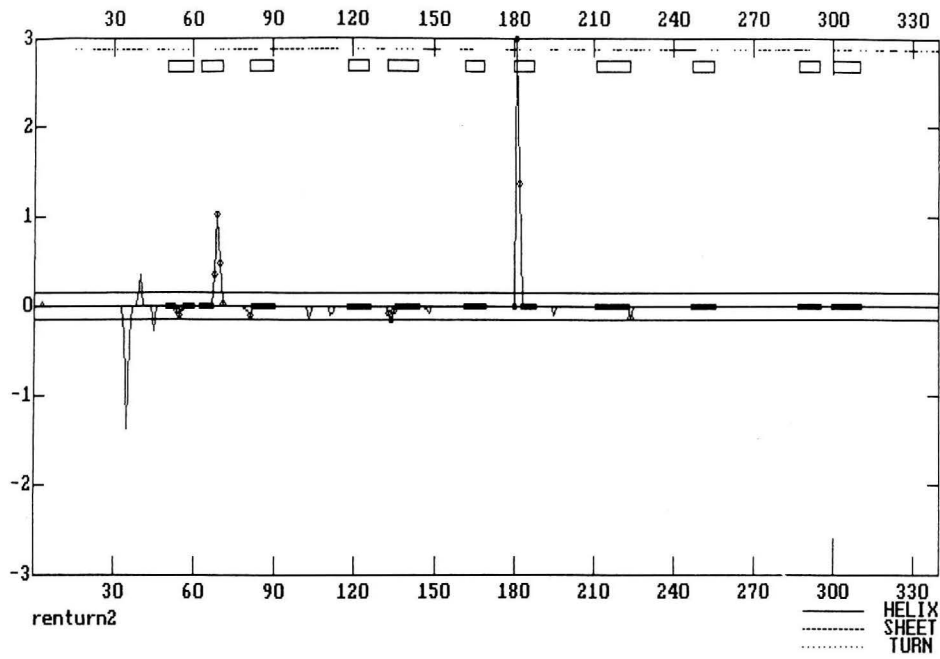
Fig. 1. Antigenicity prediction profile of renin constructed by the scale TURNEE as explained in the text. This graph was normalized between +3 and -3. The two plain lines around the mean correspond to  $\pm 0.7$  time the standard deviation. Rectangles at the top of the curves correspond to the known protein epitopes with the circles drawn on the curves corresponding to these residues. The secondary structure pattern, if known, is shown above the rectangle. The plain line corresponds to a helix, a dashed line corresponds to a sheet and a dotted line to a turn. The percentage of correct prediction is 29%. However, we can note that the two highest peaks fall in turn regions near a  $\beta$ -sheet structure. In this case, an incorrect antigenic prediction is not caused by an incorrect turn prediction.

routine. In fact, the ADDITIO routine gave results as reliable as those obtained with the turn scale of Levitt [27] but always with fewer peaks (Fig. 2). The capacity of our method to identify only a small number of turns fulfills our major aim which was to predict few peaks but with a high degree of reliability.

The TURN33 scale (classical turns, Table 8) led to similar results as the TURNEE scale (Table 7). This scale produced, with seven proteins, a fully correct prediction (i.e., a ratio  $A/A+C$  of 1, for the multiplication routine) (Fig. 3). The totally incorrect prediction obtained for the leghemoglobin protein may be due to the fact that this pro-

tein belongs to the helical class of proteins (A class) in which few hydrogen-bonded turns are present. The excellent results obtained with myoglobin may be due to the fact that, because of extensive immunochemical studies, all epitopes of this molecule have been identified. After the TURN10 scale, the TURN33 scale gave the lowest number of peaks, i.e., a mean of 2.36 per protein in the case of the multiplication routine (Table 8). One should remember that turn predictions lead to one residue peak for the four positions of a turn. Surprisingly, the results obtained with the addition procedure are similar (mean ratio  $A/A+C$  of the 14 proteins equal to 0.64) to those

Fig. 2. Antigenicity prediction profile of renin constructed by the scale TURN33 (top) and by the turn scale of Levitt (bottom) as explained in the text. The plot description is identical to that of Fig. 1. In the second plot, 21 peaks were predicted (standard method leading to a correct prediction of 42%) while in the first plot only 3 peaks were predicted (multiplication method leading to a correct prediction of 83%). Our calculation method is thus well suited for predicting a small number of peaks with a high level of confidence.



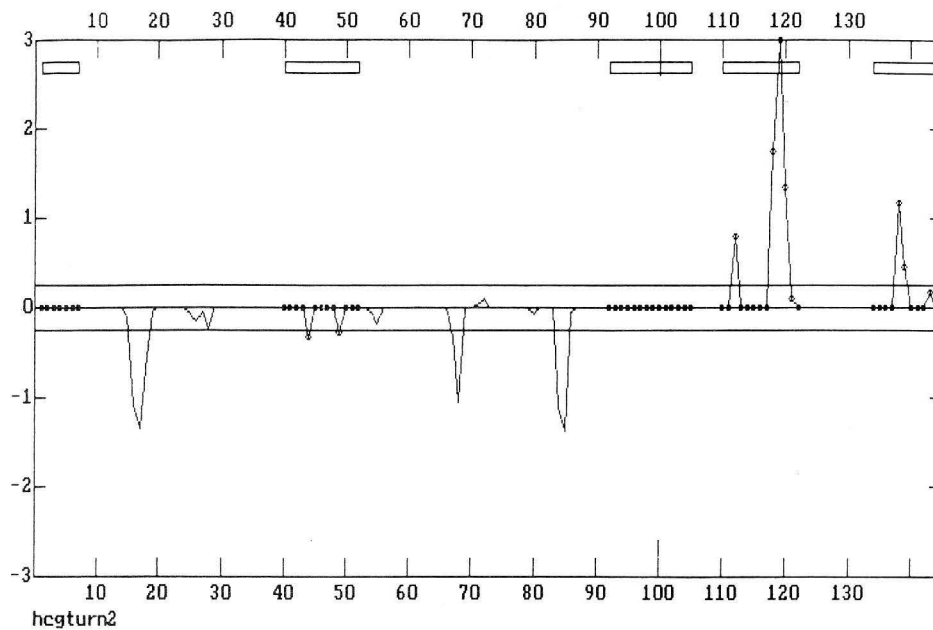


Fig. 3. Antigenicity prediction profile of human chorionic gonadotropin hormone constructed by the scale TURN33 as explained in the text. The plot description is identical to that of Fig. 1. In this case, the percentage of correct prediction is 100%.

obtained with the multiplication procedure (mean ratio of 0.69). However, the drawback of the addition procedure is that five times more peaks were predicted (13.07 for the addition and 2.36 for the multiplication procedure).

In order to ascertain the value of a new predictive method it is necessary to compare its results with those obtained with standard methods. We therefore compared our results with those obtained with the Levitt scale which was found earlier to give the best antigenicity prediction [20]. The mean ratio (0.63) obtained in the present study (Table 9), using 14 proteins, is nearly identical to the ratio of 0.61 found with 11 proteins in our previous study [20]. However, the Levitt scale gave 7% more incorrect predictions and also predicted 10 times more peaks than the multiplication procedure on the four-position turn scales (a mean of 22.07 amino acids predicted per profile).

## 5. Discussion

Many algorithms have been developed to predict the position of continuous epitopes in proteins from certain features of their primary struc-

ture. The relative effectiveness of the various methods has been difficult to assess mainly because different authors have used different criteria for evaluating the level of correct prediction. In the present study, as well as in an earlier comparison of various algorithms [20] we used an evaluation method (A/A + C) similar to that used in the first study of Hopp and Woods [21] except that we counted amino acids instead of simply counting the highest peaks.

Previous studies have shown that some of the best results were obtained by combining several parameters such as hydrophilicity, surface accessibility, backbone flexibility or secondary structure [22,33]. However, different authors have combined various scales in different ways. Jameson and Wolf [33] for instance, gave the different parameters an arbitrary weight before adding the separate curves to compute the so-called antigenic index. Unfortunately, the integrated version of the antigenic index into the UWGCC package [68] does not provide any accuracy testing routine [69].

Parker et al. [22] used the superimposition of three profiles namely their HPLC hydrophilicity

profile, the surface profile according to Janin [70] and the flexibility profile according to Karplus and Schulz [25]. They reported a high level of correct prediction although they did not consider incorrect predictions in the evaluation of their method. Other methods of predicting antigenicity by combining different scales have also been proposed [71,72], but no information was provided regarding the obtained level of incorrect predictions.

In the present study, we developed new turn scales based on the occurrence of amino acids at each of the four positions of a turn using a database comprised of 87 proteins. We found that these scales correctly predicted turn regions in proteins with approximately 80% confidence. When the turn scales were used to predict antigenicity in proteins, it was found that the level of correct prediction was 70% (with 14 proteins). The major difference compared to previous results is the smaller number of predicted peaks of antigenicity, approximately two peaks per prediction for each protein. This high level of accurate prediction shows a posteriori that turn prediction does correlate with antigenic sites. Such a score has never been reached by any technique in which the accuracy was estimated by counting correctly and incorrectly predicted amino acids.

The principal assumption made in this study is that short peptides corresponding to turns in proteins will tend to have a turn conformation in solution. The advantage of using turns for antigenicity prediction is that they incorporate structural information which cannot be represented by a single physico-chemical parameter such as hydrophilicity. Indeed, it seems at least as important for antigenicity to take into account the form of the peptide in solution as to consider the chemical properties of its constituents.

It may be surprising that our antigenicity prediction based on turns is so successful since, in our antigenic database, only 50% of the epitopes are constituted of turns. However, our method detects the most probable turns and it is conceivable that these turns are also the major epitopes of a protein. Moreover, turns possess all the well-known parameters considered to be important in antigen-antibody recognition, namely hydrophilicity, accessibility and flexibility.

## References

- [1] Benjamin, D.C., Berzofsky, J.A., East, I.J., Gurd, F.R.N., Hannum, C., Leach, S.J., Margoliash, E., Michael, J.G., Miller, A., Prager, E.M., Reichlin, M., Sercarz, E.E., Smith-Gill, S.J., Todd, P.E. and Wilson, A.C. (1984) *Ann. Rev. Immunol.* 2, 67.
- [2] Van Regenmortel, M.H.V. (1986) *Trends Biochem. Sci.* 11, 36.
- [3] Getzoff, E.D., Tainer, J.A., Lerner, R.A. and Geysen, H.M. (1988) *Adv. Immunol.* 43, 1.
- [4] Van Regenmortel, M.H.V. (1989) *Phil. Trans. R. Soc. Lond. B* 323, 451.
- [5] Appel, J.R., Pinilla, C., Niman, H. and Houghten, R. (1990) *J. Immunol.* 144, 976.
- [6] Amit, A.G., Mariuzza, R.A., Phillips, S.E.V. and Poljak, R.J. (1986) *Science* 233, 747.
- [7] Sheriff, S., Silverton, E.W., Padlan, E.A., Cohen, G.H., Smith-Gill, S., Finzel, B.C. and Davies, D.R. (1987) *Proc. Natl. Acad. Sci. USA* 84, 8075.
- [8] Padlan, E.A., Silverton, E.W., Sheriff, S. and Cohen, G.H. (1989) *Proc. Natl. Acad. Sci. USA* 86, 5938.
- [9] Colman, P.M., Laver, W.G., Varghese, J.N., Baker, A.T., Tulloch, P.A., Air, G.M. and Webster, R.G. (1987) *Nature* 326, 358.
- [10] Tulip, W.R., Varghese, J.N., Webster, R.G., Air, G.M., Laver, W.G. and Colman, P.M. (1989) *Cold Spring Harbor Symp. Quant. Biol.* 54, 257.
- [11] Bentley, G.A., Boulot, G., Riottot, M.M. and Poljak, R.J. (1990) *Nature* 348, 254.
- [12] Westhof, E., Altschuh, D., Moras, D., Bloomer, A.C., Mondragon, A., Klug, A. and Van Regenmortel, M.H.V. (1984) *Nature* 311, 123.
- [13] Tainer, J.A., Getzoff, E.D., Alexander, H., Houghten, R.A., Olson, A.J., Lerner, R.A. and Hendrickson, W.A. (1984) *Nature* 312, 127.
- [14] Barlow, D.J., Edwards, M.S. and Thornton, J.M. (1986) *Nature* 322, 747.
- [15] Novotny, J., Handschumacher, M., Haber, E., Bruccoleri, R.E., Carlson, W.B., Fanning, D.W., Smith, J.A. and Rose, G.D. (1986) *Proc. Natl. Acad. Sci. USA* 83, 226.
- [16] Hofmann, H.-J., Hädige, D., Höltje, M. and Höltje, H.-D. (1991) *Quant. Struct. Act. Relat.* 10, 300.
- [17] Hopp, T.P. (1986) *J. Immunol. Methods* 88, 1.
- [18] Van Regenmortel, M.H.V. and de Marcillac, G.D. (1988) *Immunol. Lett.* 17, 95.
- [19] Stern, P.S. (1991) *Trends Biotechnol.* 9, 163.
- [20] Pellequer, J.L., Westhof, E. and Van Regenmortel, M.H.V. (1991) *Methods Enzymol.* 203, 176.
- [21] Hopp, T.P. and Woods, K.R. (1981) *Proc. Natl. Acad. Sci. USA* 78, 3824.
- [22] Parker, J.M.R., Guo, D. and Hodges, R.S. (1986) *Biochemistry* 25, 5425.
- [23] Chothia, C. (1976) *J. Mol. Biol.* 105, 1.
- [24] Emini, E.A., Hughes, J.V., Perlow, D.S. and Boger, J. (1985) *J. Virol.* 55, 836.



- [25] Karplus, P.A. and Schulz, G.E. (1985) *Naturwissenschaften* 72, S. 212.
- [26] Ragone, R., Facchiano, F., Facchiano, A., Facchiano, A.M. and Colonna, G. (1989) *Prot. Eng.* 2, 497.
- [27] Levitt, M. (1978) *Biochemistry* 17, 4277.
- [28] Chou, P.Y. and Fasman, G.D. (1974) *Biochemistry* 13, 211.
- [29] Garnier, J., Osguthorpe, D.J. and Robson, B. (1978) *J. Mol. Biol.* 120, 97.
- [30] Chou, P.Y. and Fasman, G.D. (1979) *Biophys. J.* 26, 367.
- [31] Wilmot, C.M. and Thornton, J.M. (1988) *J. Mol. Biol.* 203, 221.
- [32] Richardson, J.S. (1981) *Adv. Prot. Chem.* 34, 167.
- [33] Jameson, B.A. and Wolf, H. (1988) *Comput. Appl. Biosci.* 4, 181.
- [34] Krchnak, V., Mach, O. and Maly, A. (1989) *Methods Enzymol.* 178, 586.
- [35] Rose, G.D., Gierasch, L.M. and Smith, J.A. (1985) *Adv. Prot. Chem.* 37, 1.
- [36] Leszczynski, J.F. and Rose, G.D. (1986) *Science* 234, 849.
- [37] Dyson, H.J., Rance, M., Houghten, R.A., Lerner, R.A. and Wright, P.E. (1988) *J. Mol. Biol.* 201, 161.
- [38] Wood, S.P., Tickle, I.J., Treharne, A.M., Pitts, J.E., Mascarenhas, Y., Li, J.Y., Husain, J., Cooper, S., Blundell, T.L., Hruby, V.J., Buku, A., Fischman, A.J. and Wyssbrod, H.R. (1986) *Science* 232, 633.
- [39] Sorimachi, K., Craik, D.J., Lloyd, E.J., Beyreuther, K. and Masters, C.L. (1990) *Biochem. Int.* 22, 447.
- [40] Urge, L., Gorbics, L. and Otvos, L.J. (1992) *Biochem. Biophys. Res. Commun.* 184, 1125.
- [41] Schulze-Gahmen, U., Prinz, H., Glatter, U. and Beyreuther, K. (1985) *EMBO J.* 4, 1731.
- [42] Rothbard, J.B., Fernandez, R., Wang, L., Teng, N.N.H. and Schoolnik, G.K. (1985) *Proc. Natl. Acad. Sci. USA* 82, 915.
- [43] Dyson, H.J., Cross, K.J., Houghten, R.A., Wilson, I.A., Wright, P.E. and Lerner, R.A. (1985) *Nature* 318, 480.
- [44] Larvor, M.P., Djavadi-Ohanian, L., Friguet, B., Baleux, F. and Goldberg, M.E. (1991) *Mol. Immunol.* 28, 532.
- [45] Mendz, G.L. and Moore, W.J. (1985) *Biochem. J.* 229, 305.
- [46] Williamson, M.P., Hall, M.J. and Handa, B.K. (1986) *Eur. J. Biochem.* 158, 527.
- [47] Laczko, I., Hollosi, M., Üрге, L., Ugen, K., Weiner, D.B., Mantsch, H.H., Thurin, J. and Otvös, L.J. (1992) *Biochemistry* 31, 4282.
- [48] LaRosa, G.J., Davide, J.P., Weinhold, K., Waterbury, J.A., Profy, A.T., Lewis, J.A., Langlois, A.J., Dreesman, G.R., Boswell, R.N., Shaddock, P., Holley, L.H., Karplus, M., Bolognesi, D.P., Matthews, T.J., Emini, E.A. and Putney, S.D. (1990) *Science* 249, 932.
- [49] Schulze-Gahmen, U., Klenk, H.-D. and Beyreuther, K. (1986) *Eur. J. Biochem.* 159, 283.
- [50] Muller, S., Plaué, S., Samama, J.P., Valette, M., Briand, J.P. and Van Regenmortel, M.H.V. (1990) *Vaccine* 8, 308.
- [51] Dorow, D.S., Shi, P.T., Carbone, F.R., Minasian, E., Todd, P.E.E. and Leach, S.J. (1985) *Mol. Immunol.* 22, 1255.
- [52] Fehrentz, J.-A., Heitz, A., Seyer, R., Fulcrand, P., Devilliers, R. and Castro, B. (1988) *Biochemistry* 27, 4071.
- [53] Joisson, C., Kuster, F., Plaué, S. and Van Regenmortel, M.H.V. (1993) *Arch. Virol.* 128, 229.
- [54] Plaué, S. (1990) *Int. J. Peptide Protein Res.* 35, 510.
- [55] Vuilleumier, S. and Mutter, M. (1992) in: *Structure of Antigens*. Vol. 1 (M.H.V. Van Regenmortel, Ed.) pp. 43–54.
- [56] Imperiali, B., Fisher, S.L., Moats, R.A. and Prins, T.J. (1992) *J. Am. Chem. Soc.* 114, 3182.
- [57] Bean, J.W., Kopple, K.D. and Peishoff, C.E. (1992) *J. Amer. Chem. Soc.* 114, 5328.
- [58] Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer Jr., E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) *J. Mol. Biol.* 112, 535.
- [59] Kabsch, W. and Sander, C. (1983) *Biopolymers* 22, 2577.
- [60] Pellequer, J.L. and Westhof, E. (1993) *J. Mol. Graph. in press.*
- [61] Van Regenmortel, M.H.V. (1992) in: *Structure of Antigens* (M.H.V. Van Regenmortel, Ed.) Vol. 1, pp. 1–27.
- [62] Van Regenmortel, M.H.V., Altschuh, D. and Klug, A. (1986) in: *Synthetic Peptides as Antigens*. Ciba Foundation Symposium, Vol. 119, pp. 76–92.
- [63] Rooman, M.J. and Wodak, S.J. (1988) *Nature* 335, 45.
- [64] Qian, N. and Sejnowski, T.J. (1988) *J. Mol. Biol.* 202, 865.
- [65] McGregor, M.J., Flores, T.P. and Sternberg, M.J.E. (1989) *Prot. Eng.* 2, 521.
- [66] Cohen, F.E., Abarbanel, R.M., Kuntz, I.D. and Fletterick, R.J. (1983) *Biochemistry* 22, 4894.
- [67] Cohen, F.E., Abarbanel, R.M., Kuntz, I.D. and Fletterick, R.J. (1986) *Biochemistry* 25, 266.
- [68] Devereux, J., Haeblerli, P. and Smithies, O. (1984) *Nucleic Acids Res.* 12, 387.
- [69] Wolf, H., Modrow, S., Motz, M., Jameson, B.A., Herrmann, G. and Förtsch, B. (1988) *Comput. Appl. Biosci.* 4, 187.
- [70] Janin, J. (1979) *Nature* 277, 491.
- [71] Menéndez-Arias, L. and Rodriguez, R. (1990) *Comput. Appl. Biosci.* 6, 101.
- [72] Kolaskar, A.S. and Tongaonkar, P.C. (1990) *FEBS Lett.* 276, 172.
- [73] Kazemi, M. and Finkelstein, R.A. (1991) *Mol. Immunol.* 28, 865.
- [74] Jacob, C.O., Sela, M. and Arnon, R. (1983) *Proc. Natl. Acad. Sci. USA* 80, 7611.
- [75] Jemmerson, R., Morrow, P.R., Klinman, N.R. and Paterson, Y. (1985) *Proc. Natl. Acad. Sci. USA* 82, 1508.
- [76] Corradin, G., Juillerat, M.A., Vita, C. and Engers, H.D. (1983) *Mol. Immunol.* 20, 763.
- [77] Atassi, M.Z. (1981) *Mol. Immun.* 18, 1021.
- [78] Jemmerson, R. and Paterson, Y. (1986) *BioTechniques* 4,

- [79] Gerin, J.L., Alexander, H., Wai-Kuo, S.J., Purcell, R.H., Dapolito, G., Engle, R., Green, N., Sutcliffe, J.G., Shinnick, T.M. and Lerner, R.A. (1983) *Proc. Natl. Acad. Sci. USA* 80, 2365.
- [80] Lerner, R.A., Green, N., Alexander, H., Liu, F., Sutcliffe, J.G. and Shinnick, T.M. (1981) *Proc. Natl. Acad. Sci. USA* 78, 3403.
- [81] Neurath, A.R., Kent, S.B.H. and Strick, N. (1984) *Virus Res.* 1, 321.
- [82] Bhatnagar, P.K., Papas, E., Blum, H.E., Milich, D.R., Nitecki, D., Karels, M.J. and Vyas, G.N. (1982) *Proc. Natl. Acad. Sci. USA* 79, 4400.
- [83] Bidart, J.-M., Troalen, F., Salesse, R., Bousfield, G.R., Bohuon, C.J. and Bellet, D.H. (1987) *J. Biol. Chem.* 262, 8551.
- [84] Stevens, V.C., Chou, W.-S., Powell, J.E., Lee, A.C. and Smoot, J. (1986) *Immunol. Lett.* 12, 11.
- [85] Bidart, J.-M., Bellet, D.H., Alberici, G.F. and van Besien, F. (1987) *Mol. Immunol.* 24, 339.
- [86] Redlich, P.N., Hoepflich, P.D., Colby, C.B. and Grossberg, S.E. (1991) *Proc. Natl. Acad. Sci. USA* 88, 4040.
- [87] Hurrell, J.G.R., Smith, J.A. and Leach, S.J. (1978) *Immunochemistry* 15, 297.
- [88] Takagaki, Y., Hirayama, A., Fujio, H. and Amano, T. (1982) *Arch. Biochem. Biophys.* 214, 750.
- [89] Teicher, E., Maron, E. and Arnon, R. (1973) *Immunochemistry* 10, 265.
- [90] Getzoff, E.D., Geysen, H.M., Rodda, S.J., Alexander, H., Tainer, J.A. and Lerner, R.A. (1987) *Science* 235, 1191.
- [91] Schmitz, H.E., Atassi, H. and Atassi, Z. (1983) *Immunol. Commun.* 12, 161.
- [92] Atassi, M.Z. (1975) *Immunochemistry* 12, 423.
- [93] Rodda, S.J., Geysen, H.M., Mason, T.J. and Schoofs, P.G. (1986) *Mol. Immunol.* 23, 603.
- [94] Schmidt, M.A., O'Hanley, P. and Schoolnik, G.K. (1984) *J. Exp. Med.* 161, 705.
- [95] Tanaka, T., Slamon, D.J. and Cline, M.J. (1985) *Proc. Natl. Acad. Sci. USA* 82, 3400.
- [96] Willumsen, B.M., Papageorge, A.G., Kung, H.-F., Bekesi, E., Robins, T., Johnsen, M., Vass, W.C. and Lowy, D.R. (1986) *Mol. Cell. Biol.* 6, 2646.
- [97] Bouhnik, J., Galen, F.-X., Menard, J., Corvol, P., Seyer, R., Fehrentz, J.-A., Le Nguyen, D., Fulcrand, P. and Castro, B. (1987) *J. Biol. Chem.* 262, 2913.
- [98] Evin, G., Galen, F.-X., Carlson, W.D., Handschumacher, M., Novotny, J., Bouhnik, J., Menard, J., Corvol, P. and Haber, E. (1988) *Biochemistry* 27, 156.
- [99] Granier C. (1992) Personal Communication.
- [100] Altschuh, D., Hartman, D., Reinbolt, J. and Van Regenmortel, M.H.V. (1983) *Mol. Immunol.* 20, 271.
- [101] Al Moudallal, Z., Briand, J.P. and Van Regenmortel, M.H.V. (1985) *EMBO J.* 4, 1231.
- [102] Altschuh, D., Al Moudallal, Z., Briand, J.P. and Van Regenmortel, M.H.V. (1985) *Mol. Immunol.* 22, 329.
- [103] Milton De, L.R.C. and Van Regenmortel, M.H.V. (1979) *Mol. Immunol.* 16, 179.
- [104] Morrow, P.R., Rennick, D.M., Leung, C.Y. and Benjamini, E. (1984) *Mol. Immunol.* 21, 301.
- [105] Sixma, T.K., Pronk, S.E., Kalk, K.H., Wartna, E.S., Van Zanten, B.A.M., Witholt, B. and Hol, W.G.J. (1991) *Nature* 351, 371.
- [106] Takano, T. and Dickerson, R.E. (1981) *J. Mol. Biol.* 153, 79.
- [107] Arutyunyan, E.G., Kuranova, I.P., Vainshtein, B.K. and Steigemann, W. (1980) *Sov. Phys. Crystallog.* 25, 43.
- [108] Kundrot, C.E. and Richards, F.M. (1987) *J. Mol. Biol.* 193, 157.
- [109] Sheriff, S., Hendrickson, W.A. and Smith, J.L. (1987) *J. Mol. Biol.* 197, 273.
- [110] Phillips, S.E.V. and Schoenborn, B.P. (1981) *Nature* 292, 81.
- [111] Pai, E.F., Krengel, U., Petsko, G.A., Goody, R.S., Kabsch, W. and Wittinghofer, A. (1990) *EMBO J.* 9, 2351.
- [112] Sielecki, A.R., Hayakawa, K., Fujinaga, M., Murphy, M.E.P., Fraser, M., Muir, A.K., Carilli, C.T., Lewicki, J.A., Baxter, J.D. and James, M.N.G. (1989) *Science* 243, 1346.
- [113] Almasy, R.J., Fontecilla-Camps, J.C., Suddath, F.L. and Buggs, C.E. (1983) *J. Mol. Biol.* 170, 497.
- [114] Namba, K., Pattanayek, R. and Stubbs, G. (1989) *J. Mol. Biol.* 208, 307.