Computer Aided Vaccine Design

Dr G P S Raghava

Concept of Drug and Vaccine

- Concept of Drug
  - Kill invaders of foreign pathogens
  - Inhibit the growth of pathogens
- Concept of Vaccine
  - Generate memory cells
  - Trained immune system to face various existing disease agents

VACCINES

A. SUCCESS STORY:

- COMPLETE ERADICATION OF SMALLPOX
- WHO PREDICTION : ERADICATION OF PARALYTIC POLIO THROUGHOUT THE WORLD BY YEAR 2003
- SIGNIFICANT REDUCTION OF INCIDENCE OF DISEASES: DIPTHERIA, MEASLES, MUMPS, PERTUSSIS, RUBELLA, POLIOMYELITIS, TETANUS

B.NEED OF AN HOUR

1) SEARCH FOR NONAVAILABILE EFFECTIVE VACCINES FOR DISEASES LIKE:
MALARIA, TUBERCULOSIS AND AIDS

2) IMPROVEMENT IN SAFETY AND EFFICACY OF PRESENT VACCINES

3) LOW COST

4) EFFICIENT DELIVERY TO NEEDY

5) REDUCTION OF ADVERSE SIDE EFFECTS

DEVELOPMENT OF NEW VACCINES: REQUIREMENT

A.

1. BASIC RESEARCH: Sound Knowledge of Fundamentals

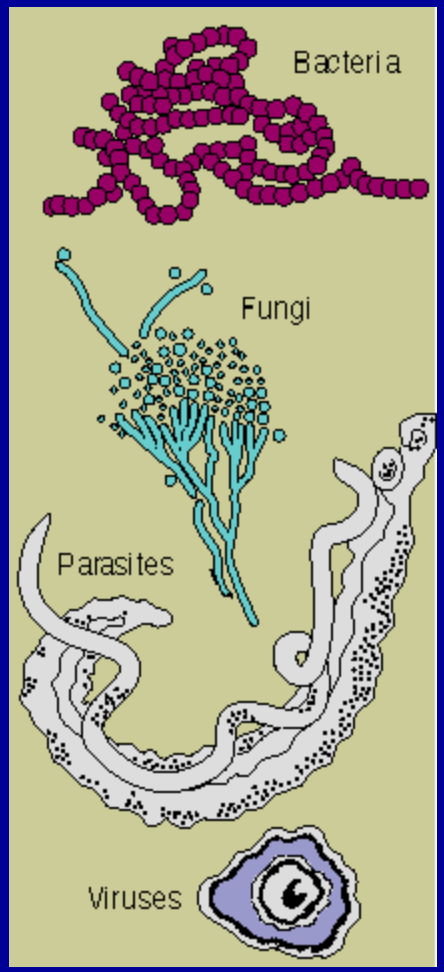
2. Combination of computer and Immunology

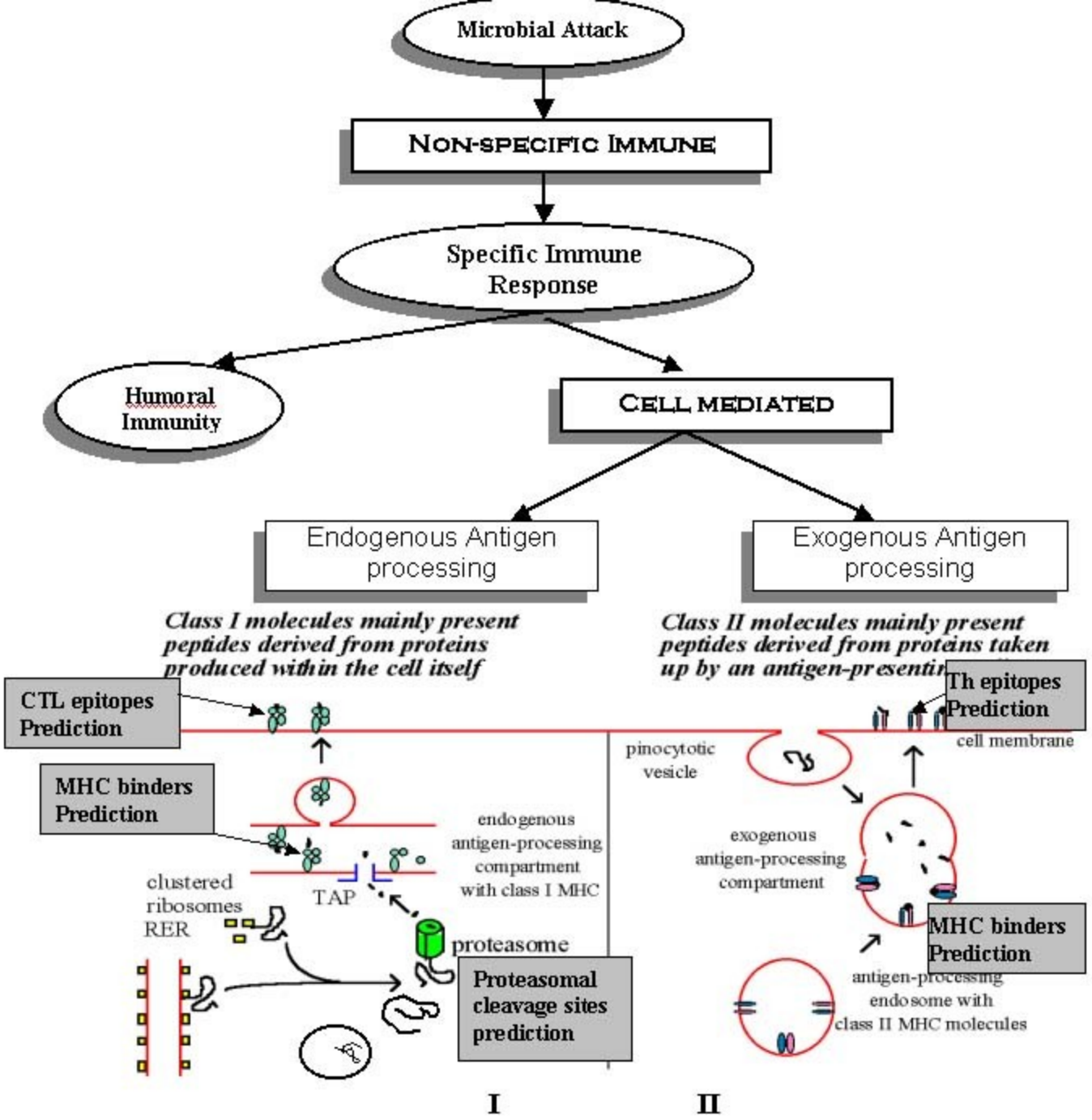
B.

1.Prediction of T and B cell epitopes

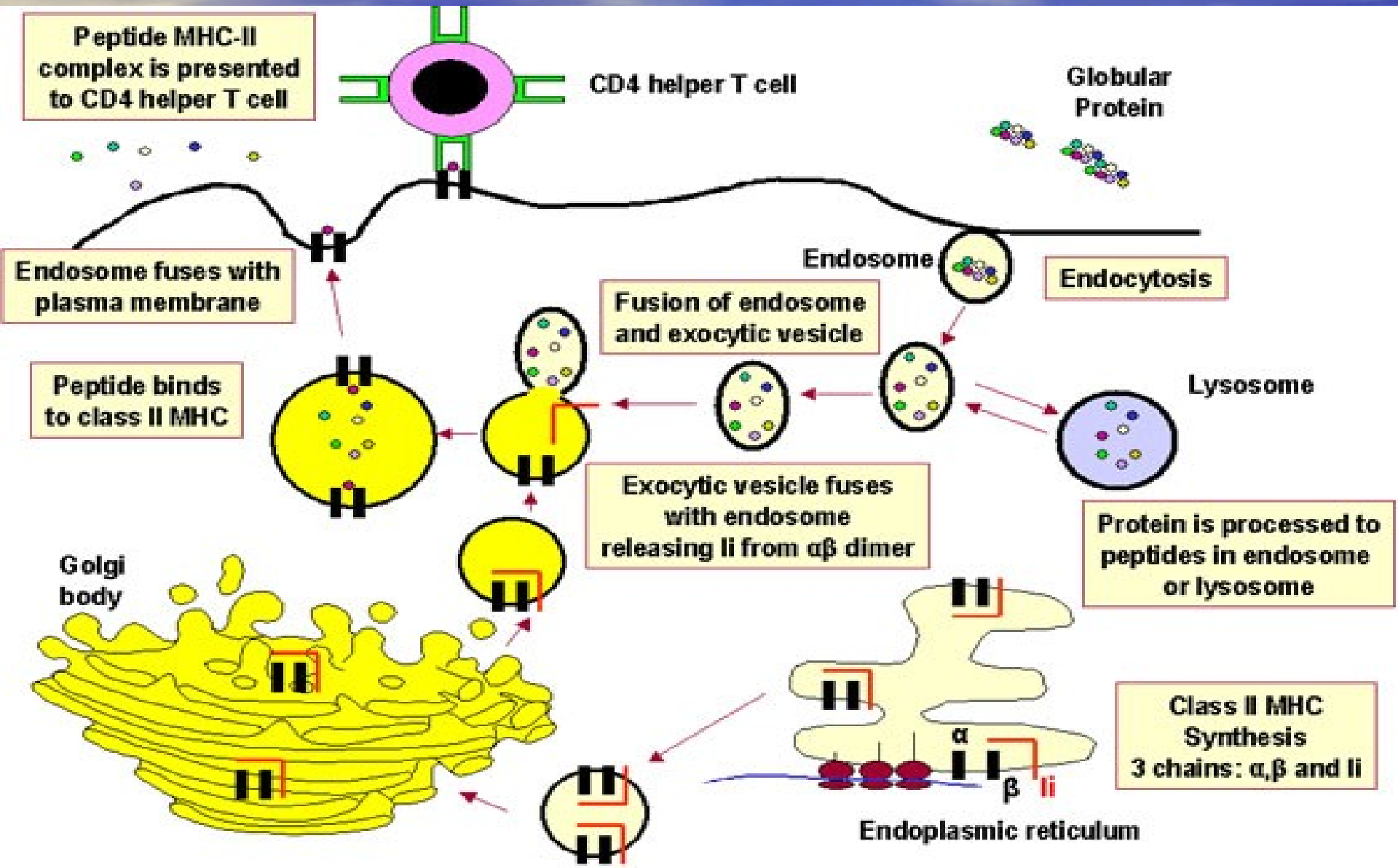
2. Prediction of Promiscuous MHC binders

Foreign Invaders or Disease Agents

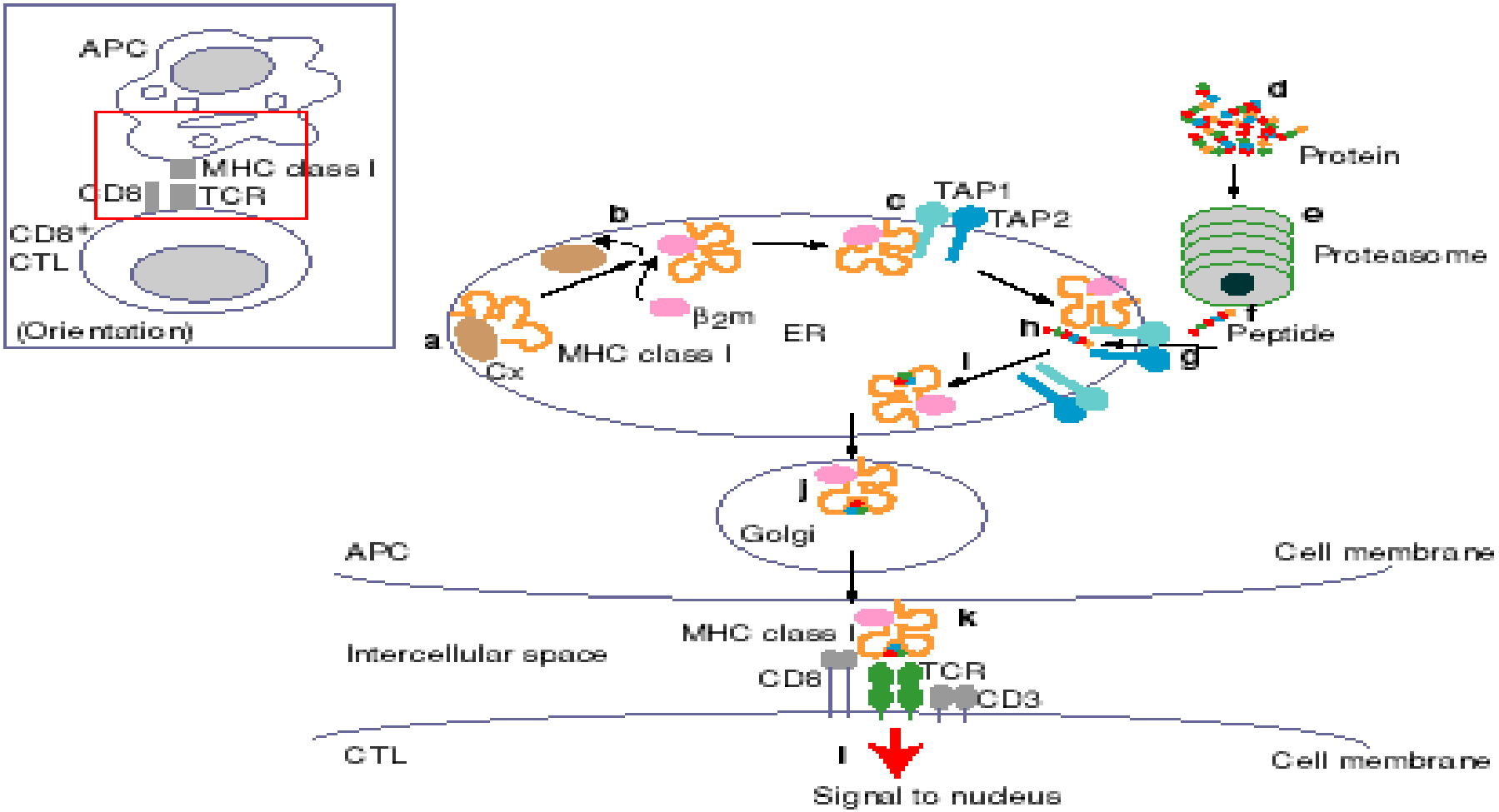



Microbial Attack
NON-SPECIFIC IMMUNE
Specific Immune Response
Humoral Immunity
CELL MEDIATED
Endogenous Antigen processing
Exogenous Antigen processing
Class I molecules mainly present peptides derived from proteins produced within the cell itself
Class II molecules mainly present peptides derived from proteins taken up by an antigen-presentir
CTL epitopes Prediction
MHC binders Prediction
Th epitopes Prediction
cell membrane
pinocytotic vesicle
endogenous antigen-processing compartment with class I MHC
clustered ribosomes RER
TAP
proteasome
Proteasomal cleavage sites prediction
exogenous antigen-processing compartment
MHC binders Prediction
antigen-processing endosome with class II MHC molecules
I
II

Exogenous Antigen processing



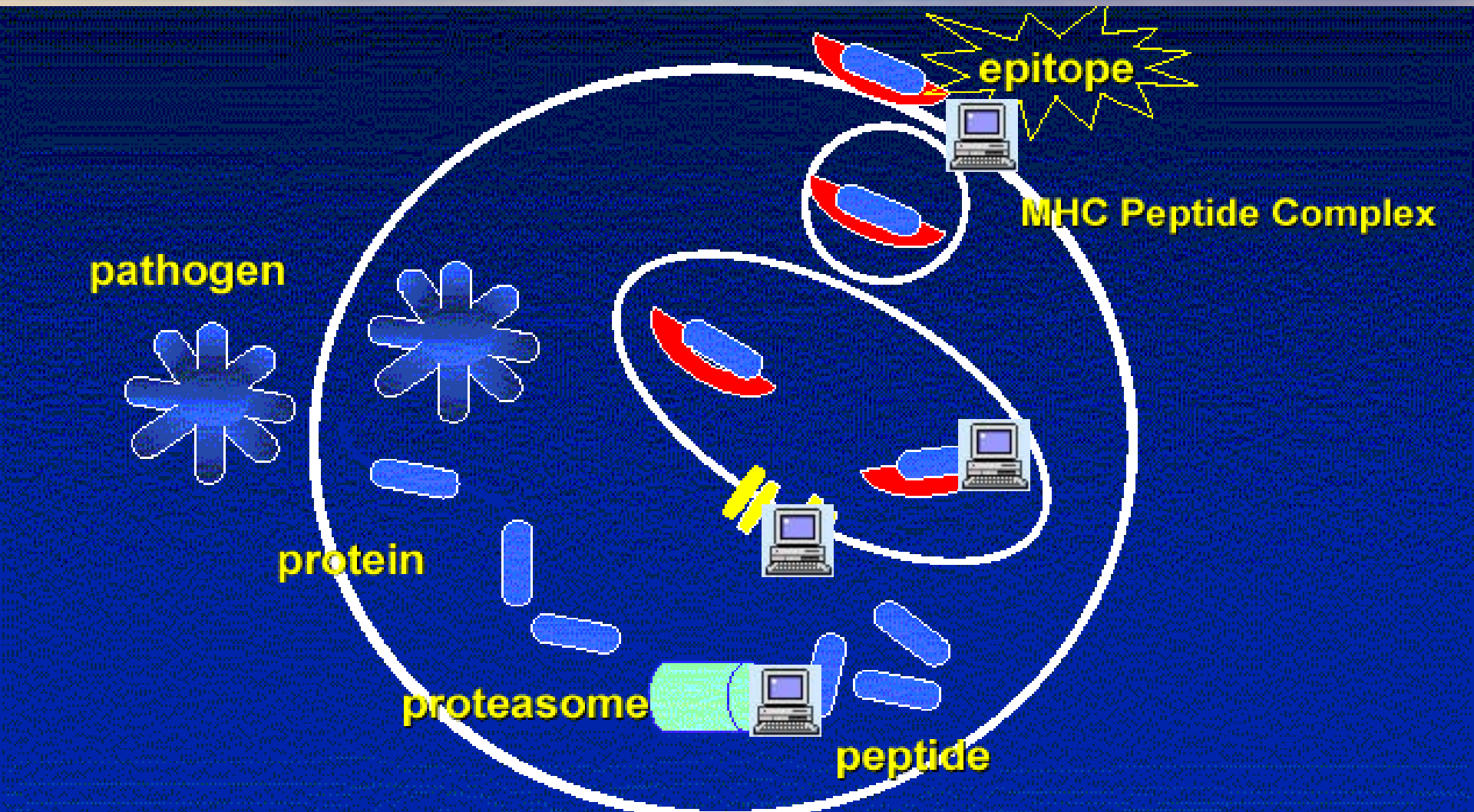
Animated Endogenous antigen processing



Degradation & transport of antigens that bind major histocompatibility complex (MHC) class I molecules

Expert Reviews in Molecular Medicine

Major steps of endogenous antigen processing



Why computational tools are required for prediction.

200 aa proteins

Chopped to overlapping peptides of 9 amino acids

Bioinformatics Tools

192 peptides

10-20 predicted peptides

invitro or *invivo* experiments for detecting which snippets of protein will spark an immune response.

Computer Aided Vaccine Design

- Whole Organism of Pathogen
  - Consists more than 4000 genes and proteins
  - Genomes have millions base pair
- Target antigen to recognise pathogen
  - Search vaccine target (essential and non-self)
  - Consists of amino acid sequence (e.g. A-V-L-G-Y-R-G-C-T ……)
- Search antigenic region (peptide of length 9 amino acids)

Computer Aided Vaccine Design

- Problem of Pattern Recognition
  - ATGGTRDAR Epitope
  - LMRGTCAAY Non-epitope
  - RTTGTRAWR Epitope
  - EMGGTCAAY Non-epitope
  - ATGGTRKAR Epitope
  - GTCVGYATT Epitope
- Commonly used techniques
  - Statistical (Motif and Matrix)
  - AI Techniques

Prediction Methods for MHC-I binding peptides

- Motifs based methods
- Quantitative matrices based methods
- Machine learning techniques based methods
  - ANN
  - SVM
- Structural based methods

Introduction of MHC molecules

- Composed of two anti-parallel alpha helices arranged on beta sheets
- Peptide binds in between the two alpha helices
- Difficulties associated with developing prediction methods
- Available methods

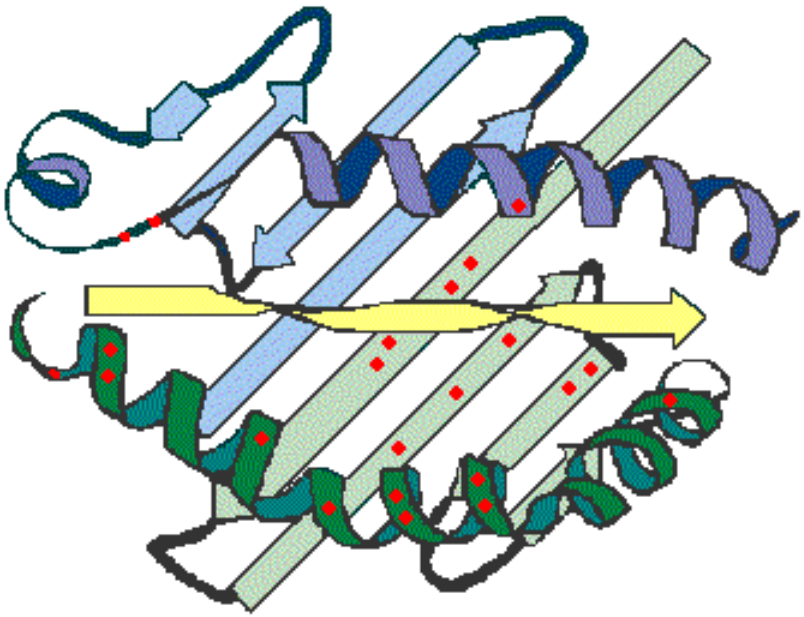
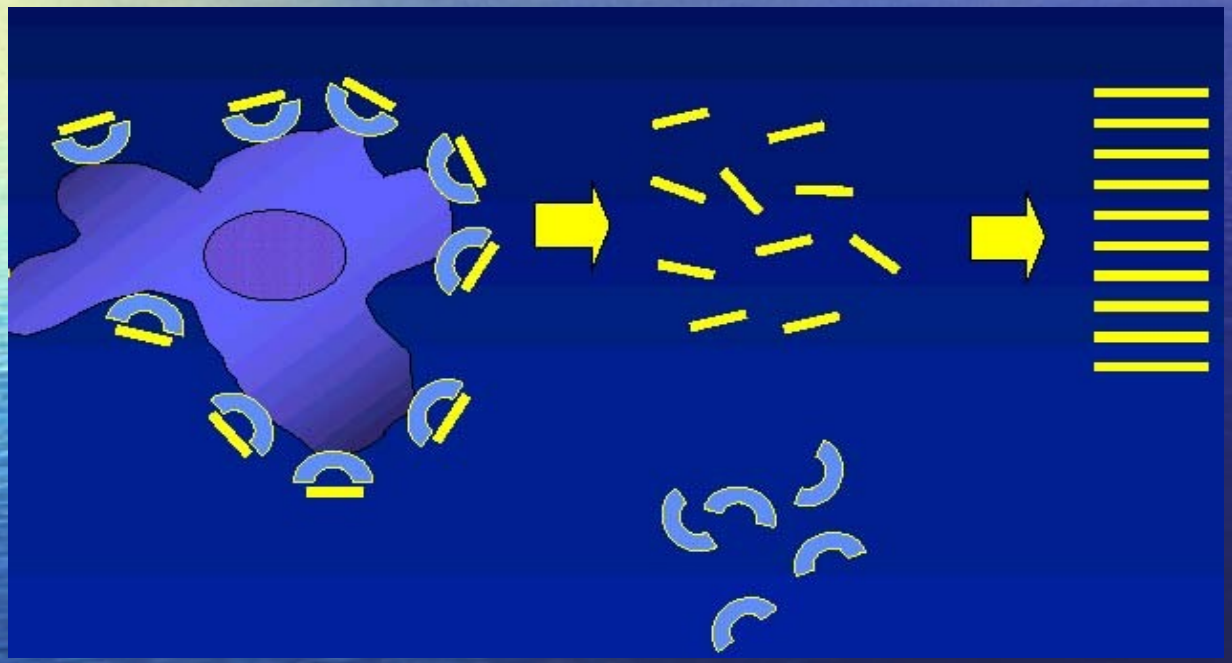


Figure 2

Ribbon diagram showing a top view of the class II antigen binding site with the bound peptide in yellow. Most of the polymorphism in class II proteins are in this domain (marked in red) and many contact the bound peptide. In this way the polymorphisms of the MHC class II molecule can influence the type of peptide it can bind.

1: Motif based Methods :

The occurrence of certain residues at specific positions in the peptide sequence is used to predict the MHC ligands. These residues are known as anchor residues and their positions as anchor positions.



? L ? ? ? ? ? V ?

Prediction accuracy - 60–65%

Limitations

- ALL binders don't have exact motifs.
- Ignorance to secondary anchor residues.
- Ignorance to residues having adverse effect on binding.

These limitations are overcome by the use of quantitative matrices. These are essentially refined motifs, covering the all amino acid of the peptide.

2 : Quantitative matrices:

In QM, the contribution of each amino acid at specific position within binding peptide is quantified.The QM are generated from experimental binding data of large ensemble of sequence variants.

Amino acid/Position	P1	P2	P3	P4	P5	P6	P7	P8	P9
A	0.69	-0.99	0.52	-0.27	0.29	0.31	0.62	0.27	-0.09
C	-0.62	-1.47	0.08	0.71	-0.75	0.32	0.22	0.61	-0.40
D	-1.05	-1.75	0.18	0.19	-0.14	-0.53	-1.05	-1.06	-1.54
E	-1.52	-1.73	-1.14	0.56	-0.72	-0.86	-0.84	0.00	-1.79
F	1.08	-1.80	-0.15	-0.47	0.70	0.32	0.92	0.36	-1.67
G	-0.51	-1.74	-0.26	-0.22	-0.26	-0.91	-1.19	-0.58	-1.86
H	0.38	-1.80	-0.24	-0.40	0.00	-0.34	0.40	0.27	-2.00
I	0.00	0.12	-0.47	-0.47	-0.21	0.14	0.20	-0.18	0.38
K	0.14	-1.75	-1.14	-0.01	-1.17	-1.12	-1.69	-0.65	-1.94
L	0.09	6.31	0.54	-0.06	0.26	0.45	0.50	0.62	6.03
M	0.31	6.22	0.63	-0.86	0.08	0.33	-0.29	-0.43	-0.37
N	-0.67	-2.00	0.75	-0.04	-0.04	-0.20	0.10	-0.35	-2.00
P	-0.93	-1.88	-0.29	0.76	0.57	0.87	0.42	0.28	-2.00
Q	-0.69	-1.33	-0.61	-0.02	-0.38	-0.58	-0.76	-0.47	-1.50
R	0.31	-2.00	-0.67	0.20	0.00	-0.70	-0.31	-0.14	-1.82
S	0.72	-1.87	0.60	0.33	0.04	0.45	0.24	0.92	-1.62
T	-0.48	-0.16	-0.73	-0.35	-0.03	0.00	-0.11	0.27	-0.29
V	-0.22	-0.07	0.09	-0.26	0.49	0.58	0.43	-0.39	6.28
W	-0.67	-2.00	0.80	-0.74	0.50	-0.80	-0.40	-0.40	-2.00
X	0.00	2.00	2.00	0.00	2.00	2.00	2.00	2.00	2.00
Y	1.25	-1.58	0.88	-0.67	0.67	-0.34	0.22	-0.29	-2.00

Available quantitative matrices for MHC class I :-

- Sette et al ., 1989
- Ruppert et al., 1993
- Parker et al., 1994
- Gulukota et al., 1997
- Bhasin and Raghava 2003 (submitted).

The score of the peptide is calculated by summing up the scores of each amino acid of the peptide at specific position.

Score of peptide ILKE PVHGV will be calculated as follows:

Amino acid/Position	P1	P2	P3	P4	P5	P6	P7	P8	P9
A									
C									
D									
E				0.56					
F									
G								-0.58	
H							0.40		
I	0.00								
K			-1.14						
L		6.31							
M									
N									
P									
Q									
R									
S									
T									
V						0.58			6.28
W									
Y					0.68				

Peptide score=I+L+K+E+P+V+G+V

Peptide score < threshold score = predicted binder

Peptide score > threshold score = predicted non-binder

In few cases the peptide score is calculated by multiplying the score of each amino acid of peptide.

The matrices based methods can predict peptides having canonical motifs with fair accuracy.

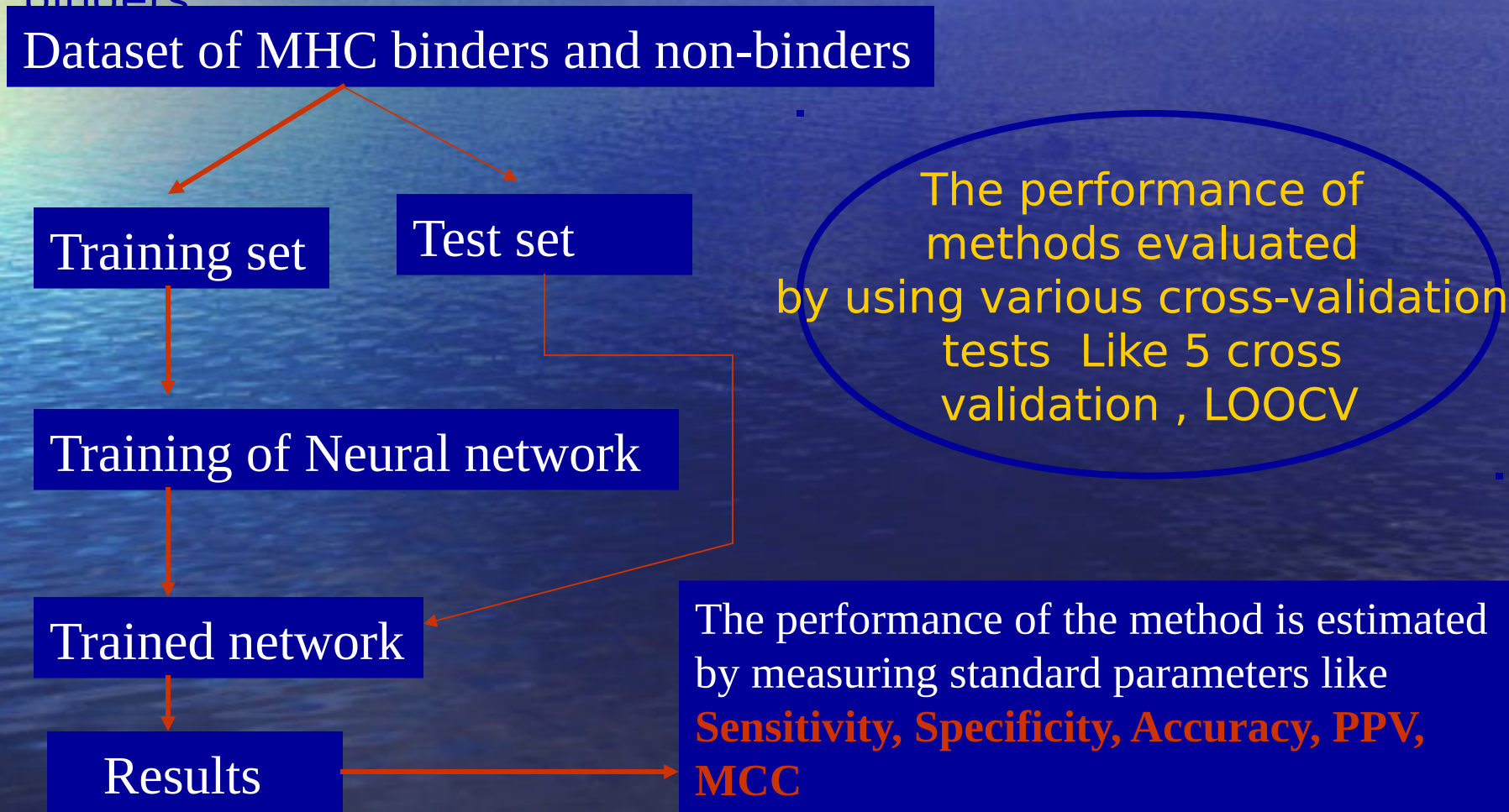
Online methods based on quantitative matrices

Program	URL	Service available
ProPred		47 MHC alleles
nHLAPred		67 MHC alleles
SYFPEITHI		> 200 MHC alleles
LpPEP		1 MHC allele
RANKPEP		>40 MHC alleles
BIMAS		>46 MHC alleles
MAPPP		>50 MHC alleles

Limitations: These methods are not able to handle the non-linearity in data of MHC binders and non-binders.

3: Machine learning Approach

ARTIFICAL NEURAL NETWORKS :In order to handle the non-linearity of data artificial neural network based approach has been applied to classify the data of MHC binders and non-binders



4: Structure Based MHC binders prediction

Based on the known structure of MHC molecules and peptide, these methods evaluates the compatibility of different peptides to fit into the binding groove of distinct MHC molecule. The MHC ligands are chosen by threading the peptide in the binding groove of MHC and getting an estimate of energy. The peptide with lowest binding energy is considered as best binder.

Advantages:

Large set of experimentally proven peptides for each MHC allele is not required.

Limitations:

- Very less amount data about 3D structure of MHC and Peptide.
- Computation is very slow
- Large number of false positive results because each pocket of MHC allele can bind with side chain of many amino acids.

Thankyou