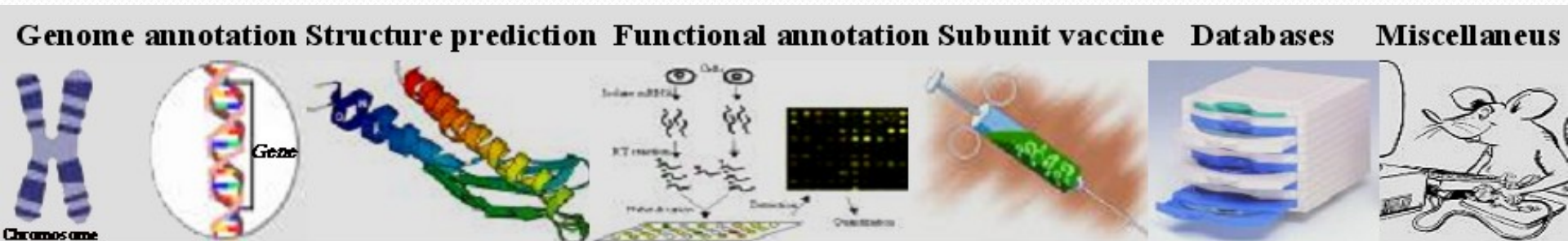


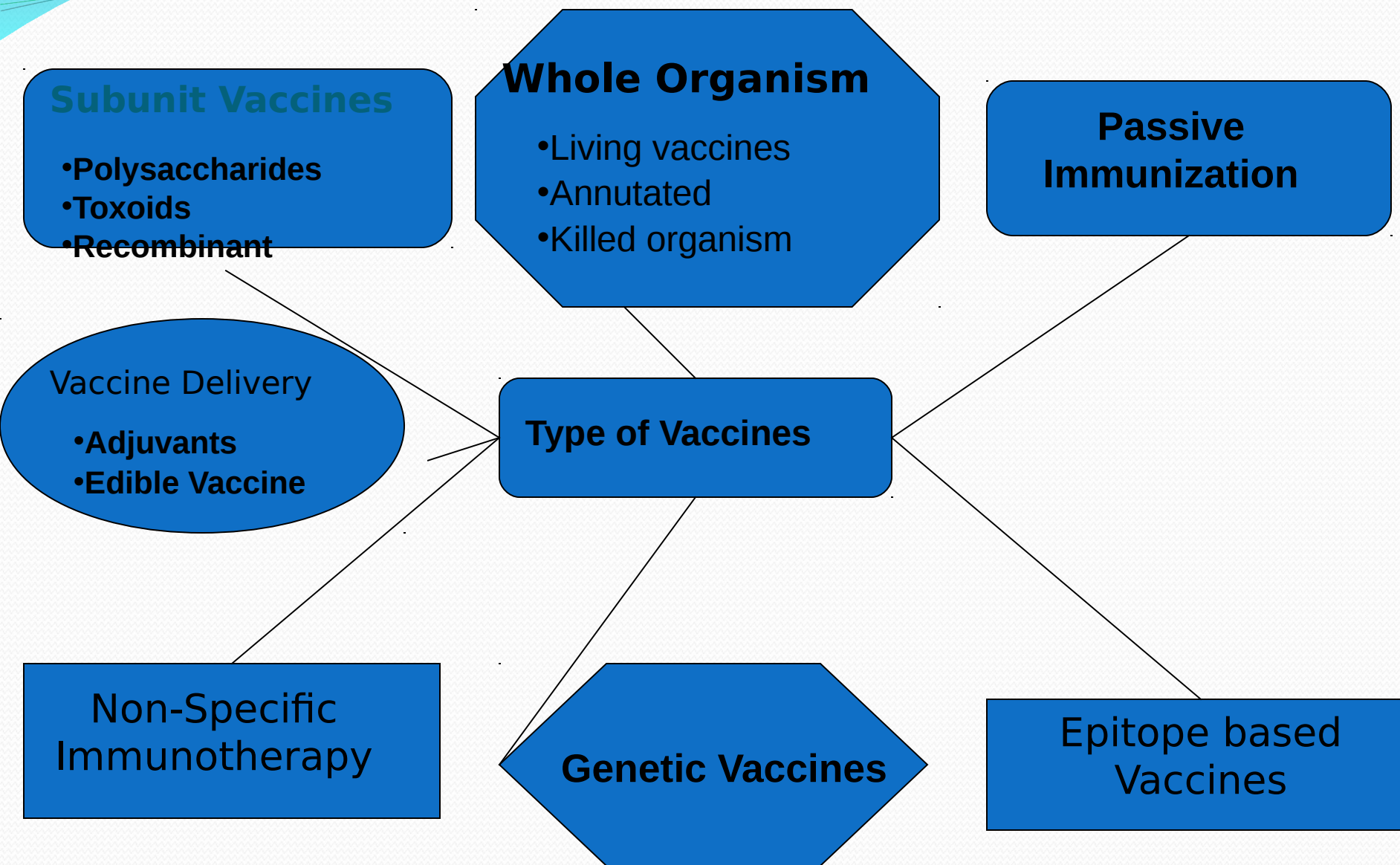
Prediction of Linear and Conformational B-cell Epitopes



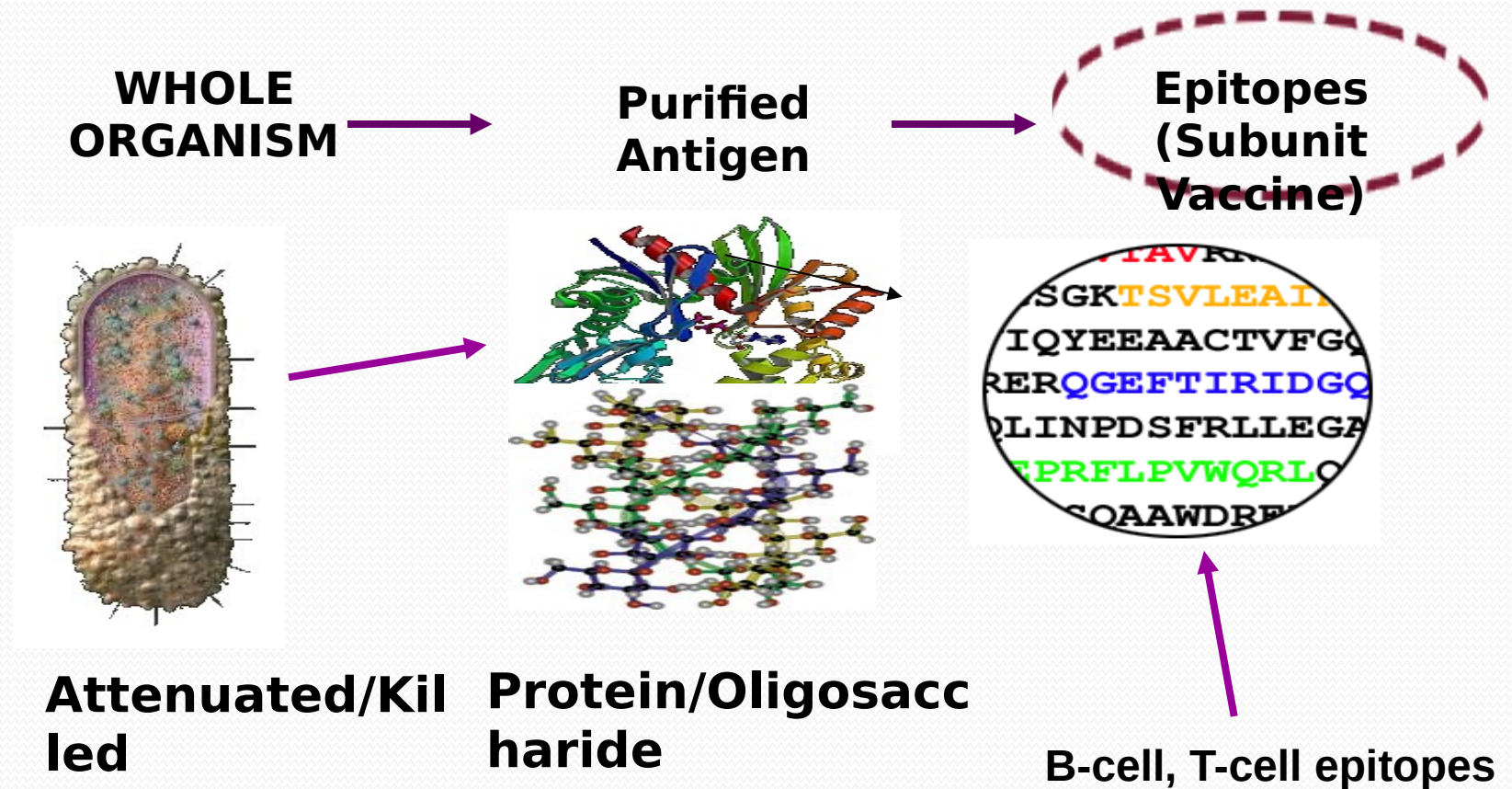
Dr. G P S Raghava, Head Bioinformatics Centre
Institute of Microbial Technology, Chandigarh, India

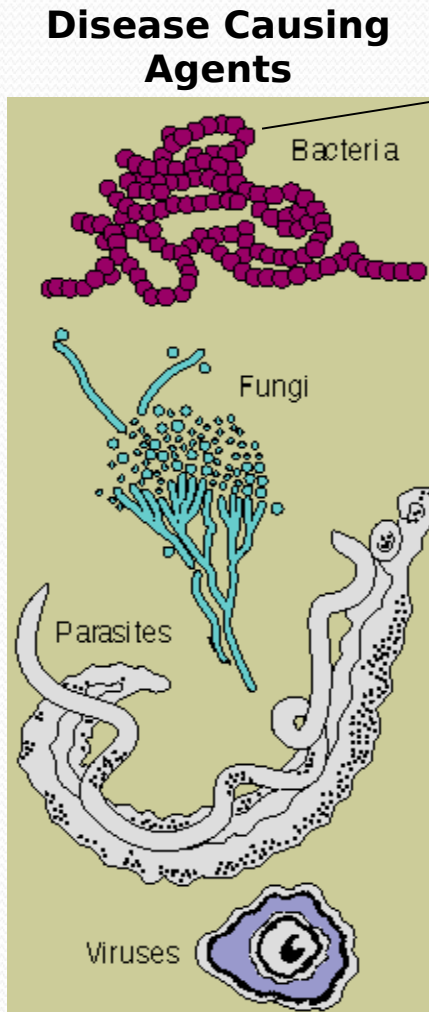
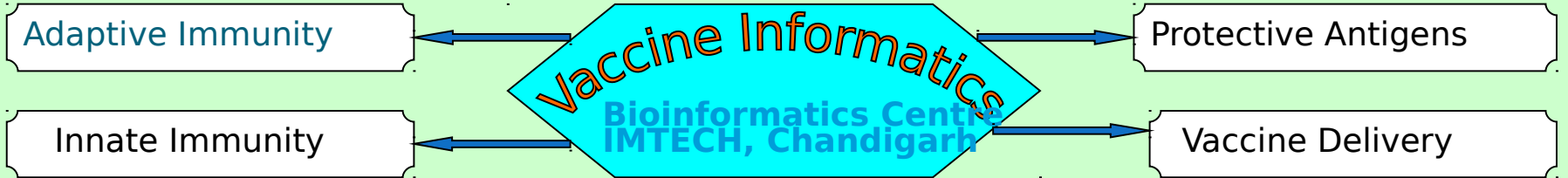


Web Site: <http://www.imtech.res.in/raghava/>

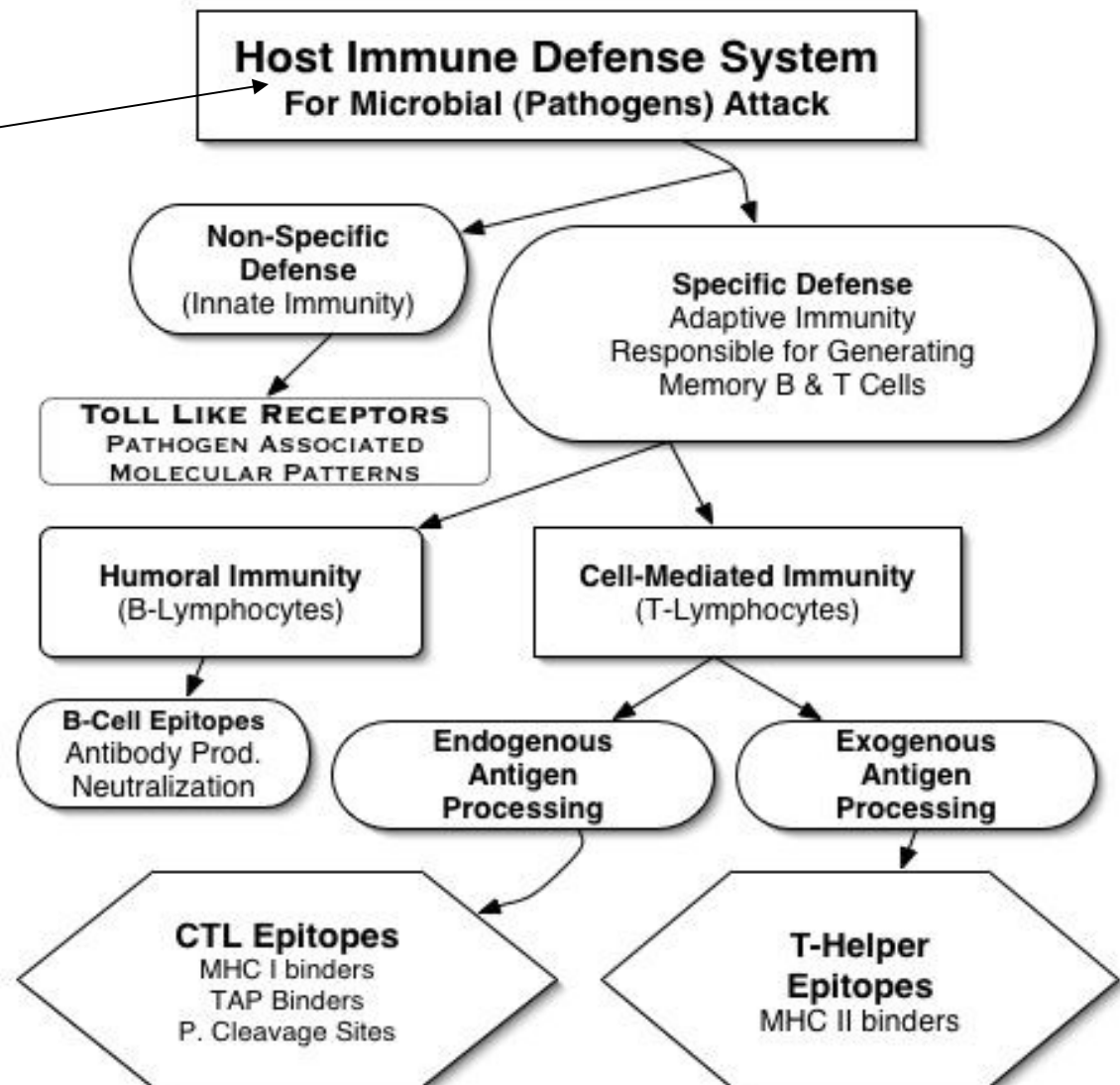


Strategies of vaccine development



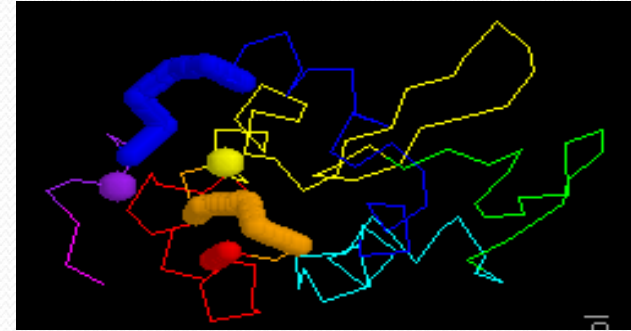


Pathogens/Invaders



Types of epitopes

B cell Epitope

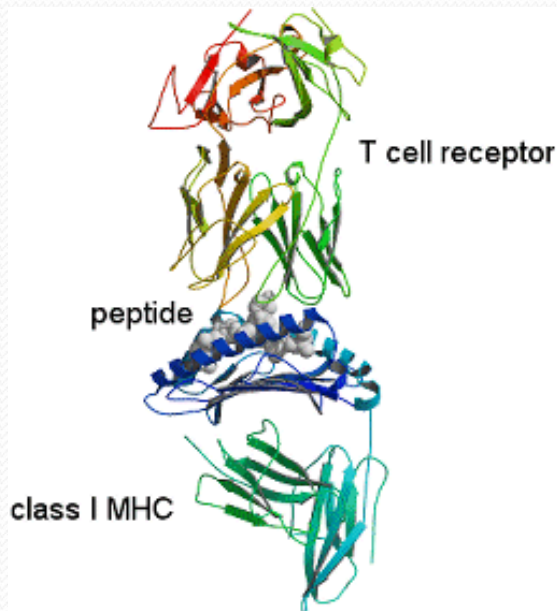


Conformational

Non-
Conformational

Linear B cell Epitope

T cell Epitope



CTL epitope

Class I MHCs

Endogenous Antigen
Processing

Helper T cell
epitope

Class II MHCs

Exogenous Antigen
Processing

Linear B-cell epitope prediction

- Linear B-cell epitopes are as important or even more applicative than conformational epitopes.
- Most of the existing methods for linear B-cell epitopes are based on old and very small dataset (~1000 epitopes).
- Performance of all the methods are not better than 0.75 AUC.
- None of the methods used experimentally verified negative datasets.

Adaptive Immunity

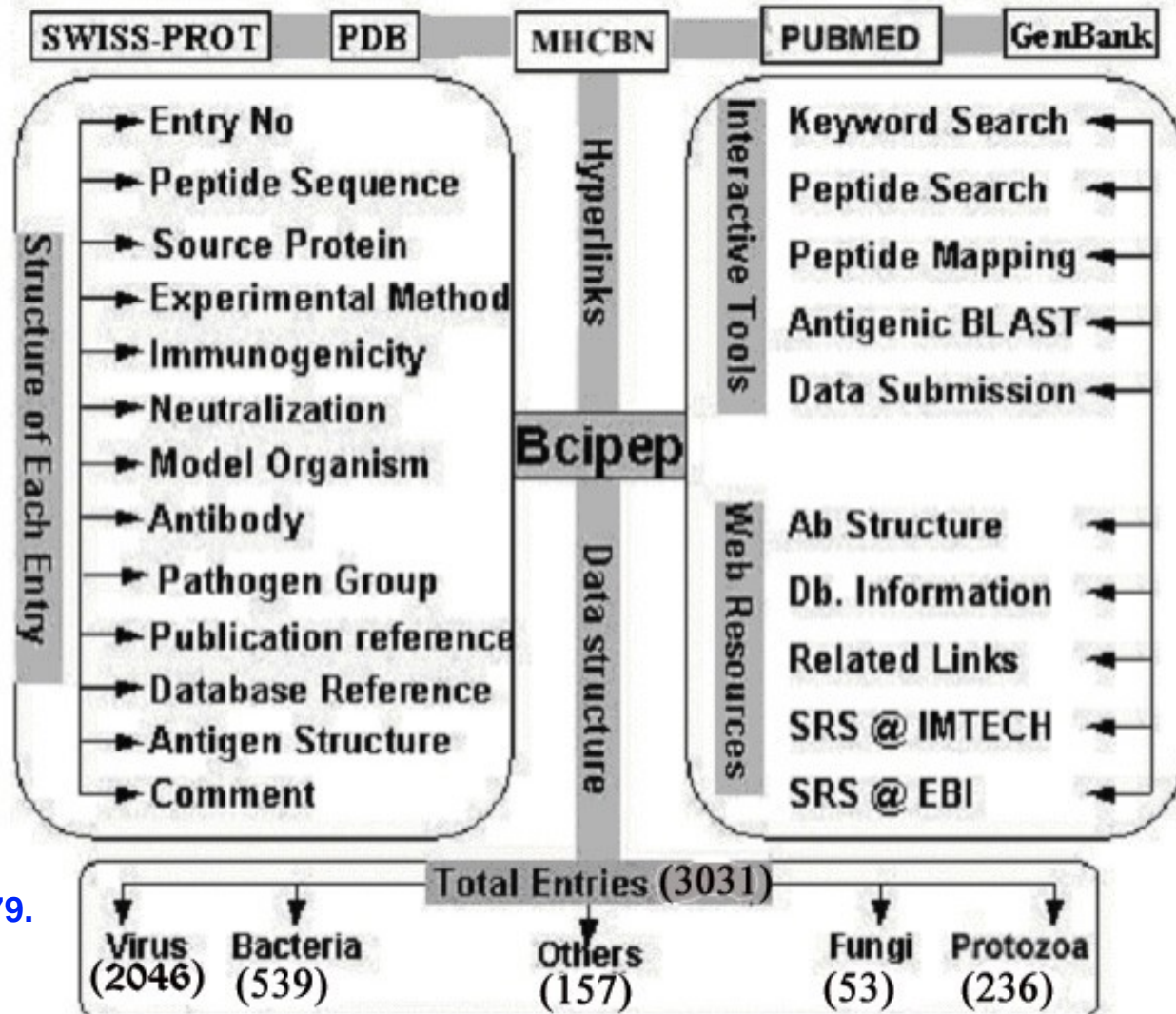
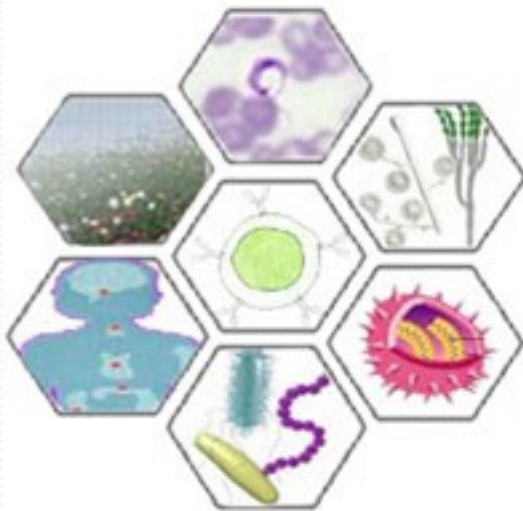
Innate Immunity

Vaccine Informatics
Bioinformatics Centre
IMTECH, Chandigarh

Protective Antigens

Vaccine Delivery

**BCIPEP: A database of
B-cell epitopes.**



Saha et al.(2005) BMC Genomics 6:79.

Saha et al. (2006) NAR (Online)

BCEpred: Benchmarking of physico-chemical properties used in existing B-cell epitope prediction methods

In 2003, we evaluate parameters on 1029 non-redundant B-cell epitopes obtained from BCIpep and 1029 random peptide

Saha and Raghava (2004) ICARIS 197-204.

Physico-chemical Properties	Threshold	Sensitivity	Specificity	Accuracy% (Max)
Hydrophilicity [1]## (Parker et al., 1986)**	2.00	33	76	54.47
Accessibility[2] (Emini et al., 1985)	2.00	65	46	55.49
Flexibility [3] (Karplus and Schulz, 1985)	1.90	47	68	57.53
Surface [4] (Janin and Wodak, 1978)	2.40	37	74	55.73
Polarity [5] (Ponnuswamy et al., 1980)	2.30	2.8	81	54.08
Turns [6] (Pellequer et al., 199)	1.90	17	89	52.92
Antigenic Scale [7] (Kolaskar and Tongaonkar, 1990)	1.80	59	52	55.59
[3]+[1]+[5]+[4]	2.38	56	61	58.70

ABCpred: ANN based method for B-cell epitope prediction

Challenge in Developing Method

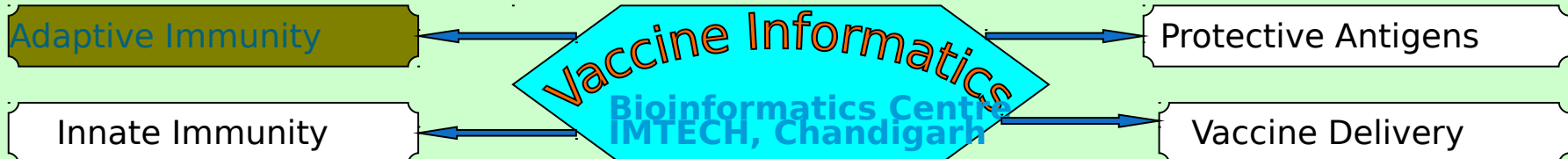
1. Machine learning technique needs fixed length pattern where epitope have variable length
2. Classification methods need positive and negative dataset
3. There are experimentally proved B-cell epitopes (positive) dataset but not Non-epitopes (negative)

Assumptions to fix the Problem

1. More than 90% epitope have less than 20 residue so we fix maximum length 20
2. We added residues at both end of small epitopes from source protein to make length of epitope 20
3. We generate random peptides from proteins and used them as non-epitopes

Creation of fixed pattern of 20 from epitopes

Window Length/ Peptide	<u>AEFPLDIT</u>	<u>ACVPTDPNPQEVVLNVNVTEN</u> (20 amino acid length)
20	PKG YVG <u>AEFPLDIT</u> AGTEAA	<u>ACVPTDPNPQEVVLNVNVTEN</u>
18	KGY VG <u>AEFPLDIT</u> AGTEA	<u>CVPTDPNPQEVVLNVNTE</u>
16	GY VG <u>AEFPLDIT</u> AGTE	<u>VPTDPNPQEVVLNVNVT</u>
14	Y VG <u>AEFPLDIT</u> AGT	<u>PTDPNPQEVVLNVN</u>
12	V G <u>AEFPLDIT</u> AG	<u>TDPNPQEVVLVN</u>
10	G <u>AEFPLDIT</u> A	<u>DPNPQEVVLV</u>



Prediction of B-Cell Epitopes

- **BCEpred: Prediction of Continuous B-cell epitopes**
 - Benchmarking of existing methods
 - Evaluation of Physico-chemical properties
 - Poor performance slightly better than random
 - Combine all properties and achieve accuracy around 58%
 - Saha and Raghava (2004) ICARIS 197-204.
- **ABCpred: ANN based method for B-cell epitope prediction**
 - Extract all epitopes from BCIPEP (around 2400)
 - 700 non-redundant epitopes used for testing and training
 - Recurrent neural network
 - Accuracy 66% achieved
 - Saha and Raghava (2006) Proteins,65:40-48
- **ALGpred: Mapping and Prediction of Allergenic Epitopes**
 - Allergenic proteins
 - IgE epitope and mapping
 - Saha and Raghava (2006) Nucleic Acids Research 34:W202-W209

B-cell epitope classification

B-cell epitope – structural feature of a molecule or pathogen, accessible and recognizable by B-cells

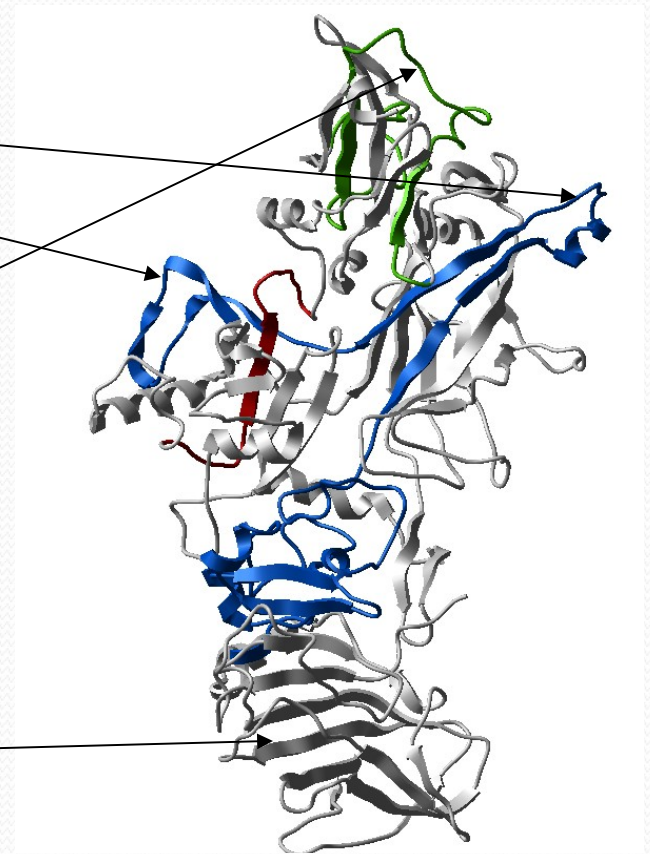
Linear epitopes

One segment of the amino acid chain

Discontinuous epitope (with linear determinant)

Discontinuous epitope

Several small segments brought into proximity by the protein fold



3D Structure



Sequence??

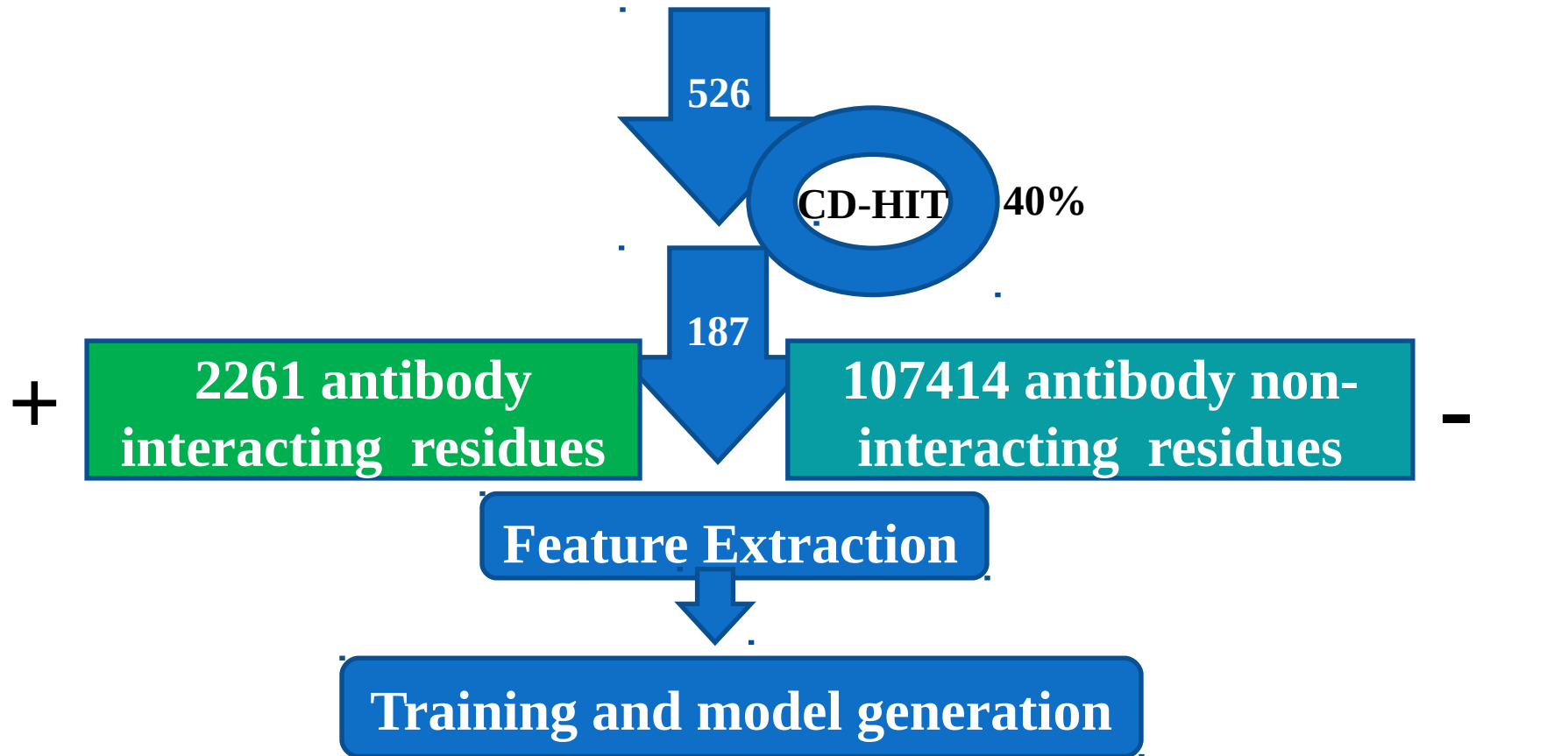
MTANKFIPNKF**SIK**TFSVLLFAISSSQAIENVAMNEHYTESDIKRNHKTEKNKTEKEKF

Data collection and processing

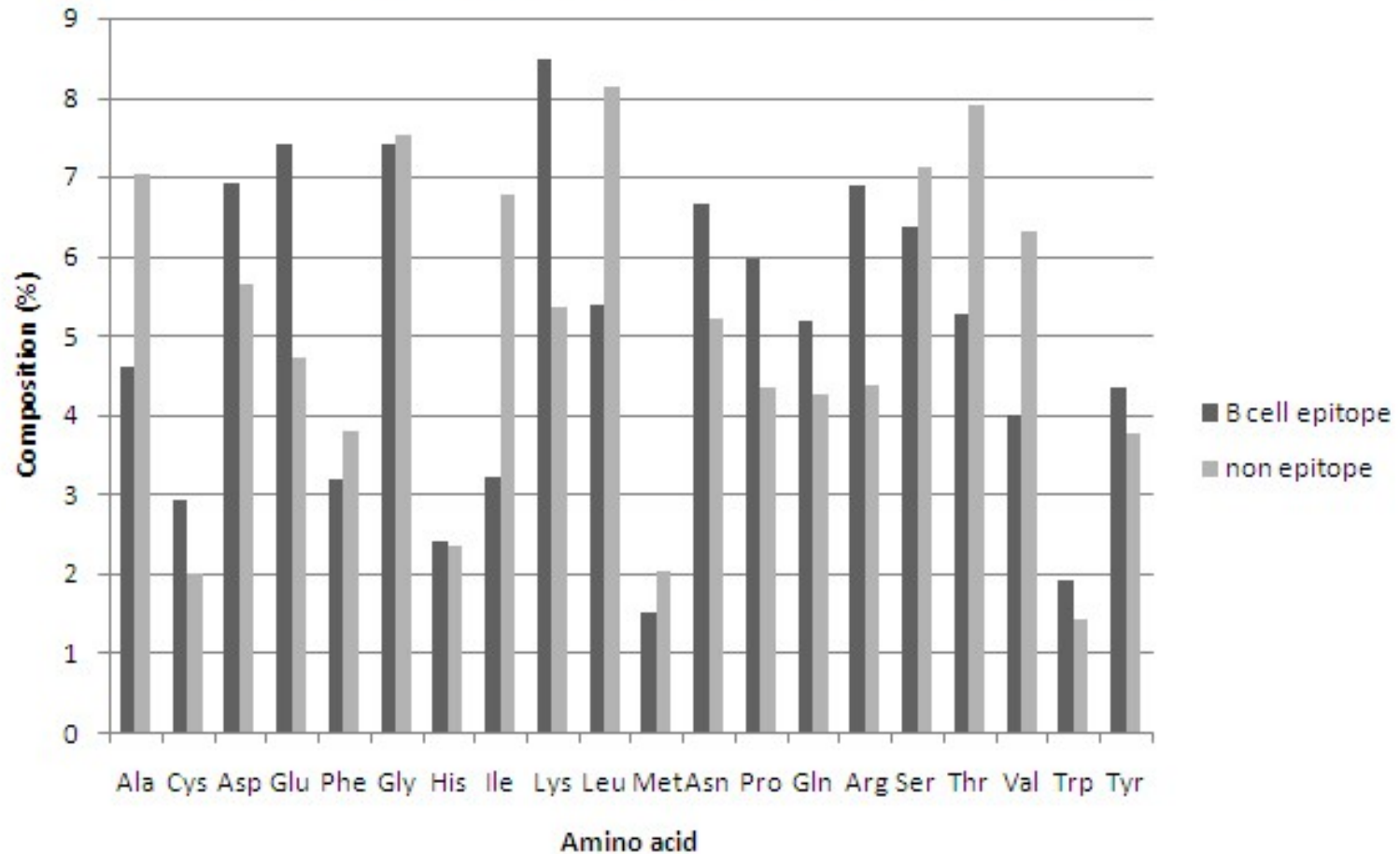
**Immune Epitope Database
(365 sequences)**

**Ponomarenko *et al* 2007
(161)**

Antigenic sequences with conformational B-cell epitope information

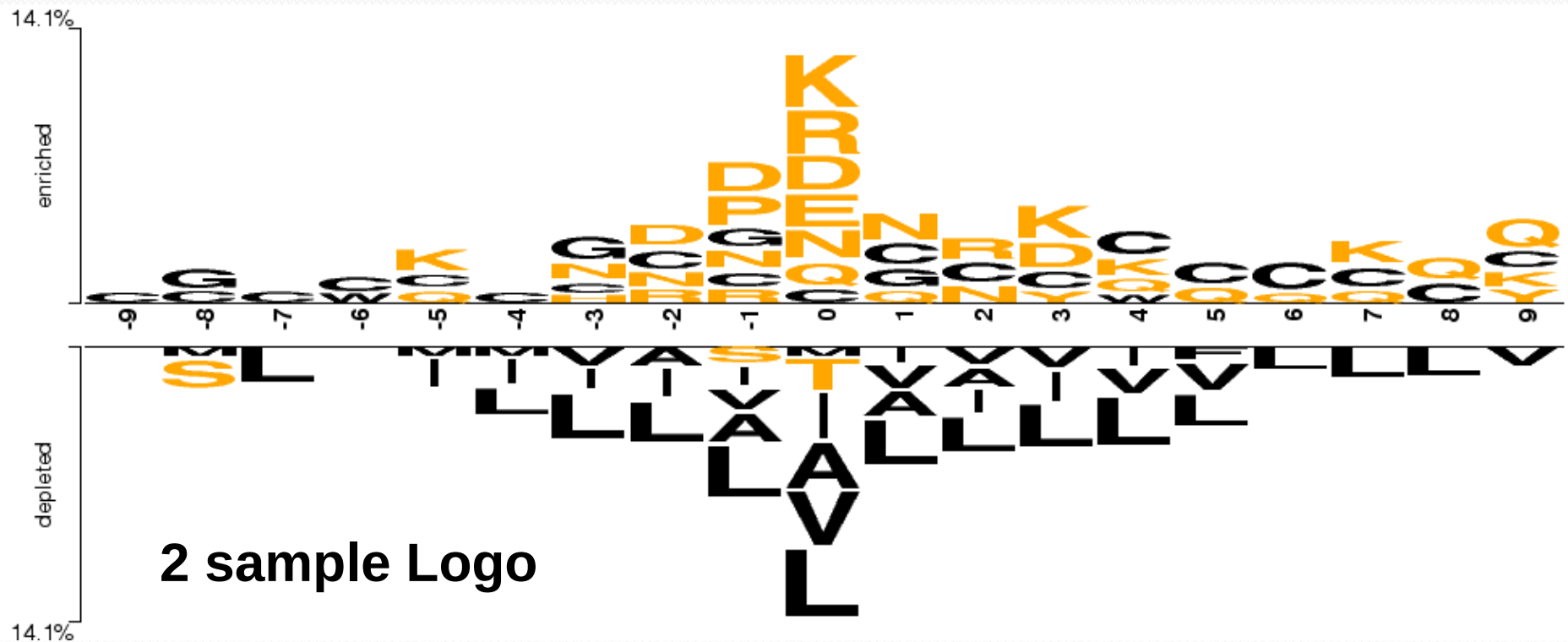


Amino acid distribution at central residue

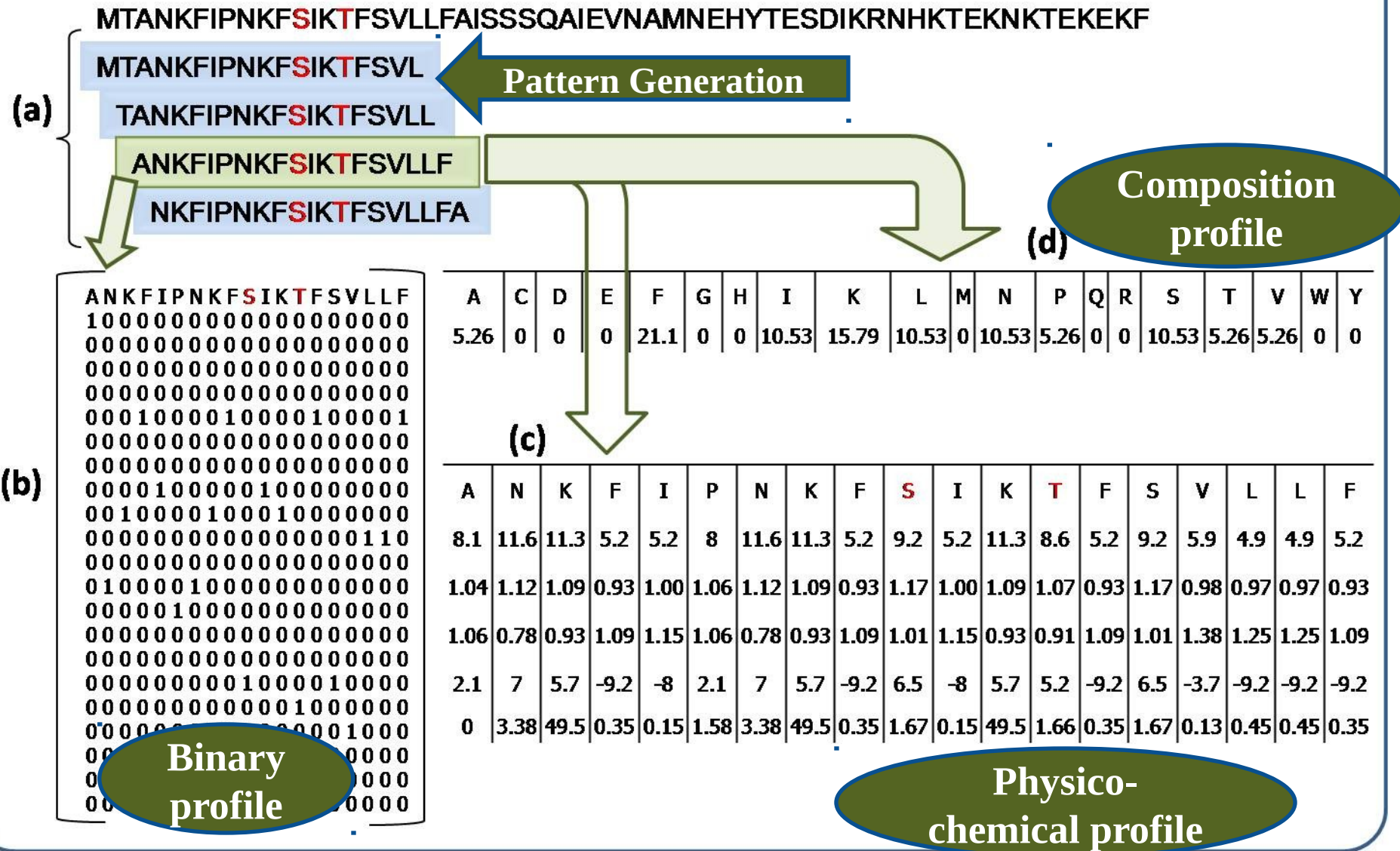


Polar and charged residues are preferred in antibody interaction

Propensities of surface-exposed and buried residues



Feature extraction from an antigen primary sequence



Comparison of structure based and CBTOPE algorithm on Benchmark dataset*

Value	ProMate	PSI-PRED best patch	Patch Dock best model	ClusPro (DOT) best model	CEP	DiscoTope	CBTOPE (This Study)
Sen	0.09	0.33	0.43	0.45	0.31	0.42	0.80
1-Spe	0.08	0.14	0.11	0.07	0.22	0.21	0.15
PPV	0.10	0.19	0.26	0.39	0.11	0.16	0.31
Acc	0.84	0.82	0.85	0.89	0.74	0.75	0.84
AUC	0.51	0.60	0.66	0.69	0.54	0.60	0.89

*Ponomarenko *et al* 2007, BMC Structural Biology

PPV - positive predictive value



CBTOPE-Menu

- Home
- Submit Sequence
- Algorithm
- Developers
- Help
- Supplementary Data
- Contact Us

Query Submission Page

Submit protein sequence(s) for prediction ::

Sequence Name (Optional):

E-mail Address (Optional):

(Please enter your email address if you want to receive your result via email)

Sequence: (Type/paste FASTA format amino acid sequences, multiple sequences can be submitted at a time)

```
>P40136|CYAA_BACAN Calmodulin-sensitive adenylate cyclase -  
Bacillus anthracis.  
MTRNKFIPNKFISIISFSVLLFAISSSQAIENVAMNEHYTESDIKRNHKTENKTEKEKFK  
DSINNLVKTEFTNETLDKIQQTQDLLKKIPKDVLEIYSELGGEIYFTDIDLVEHKELQDL  
SEEEKNSMNSRGEKVPFASRFVFEKKRETPKLIINIKDYAINSEQSKEVYIEIGKGISLD
```

OR Upload Sequence file No file chosen

SVM threshold:

[CBTOPE Home](#)

CBTOPE:: Prediction of Conformational B-cell epitopes No of sequences = 1

Threshold Selected: -.3

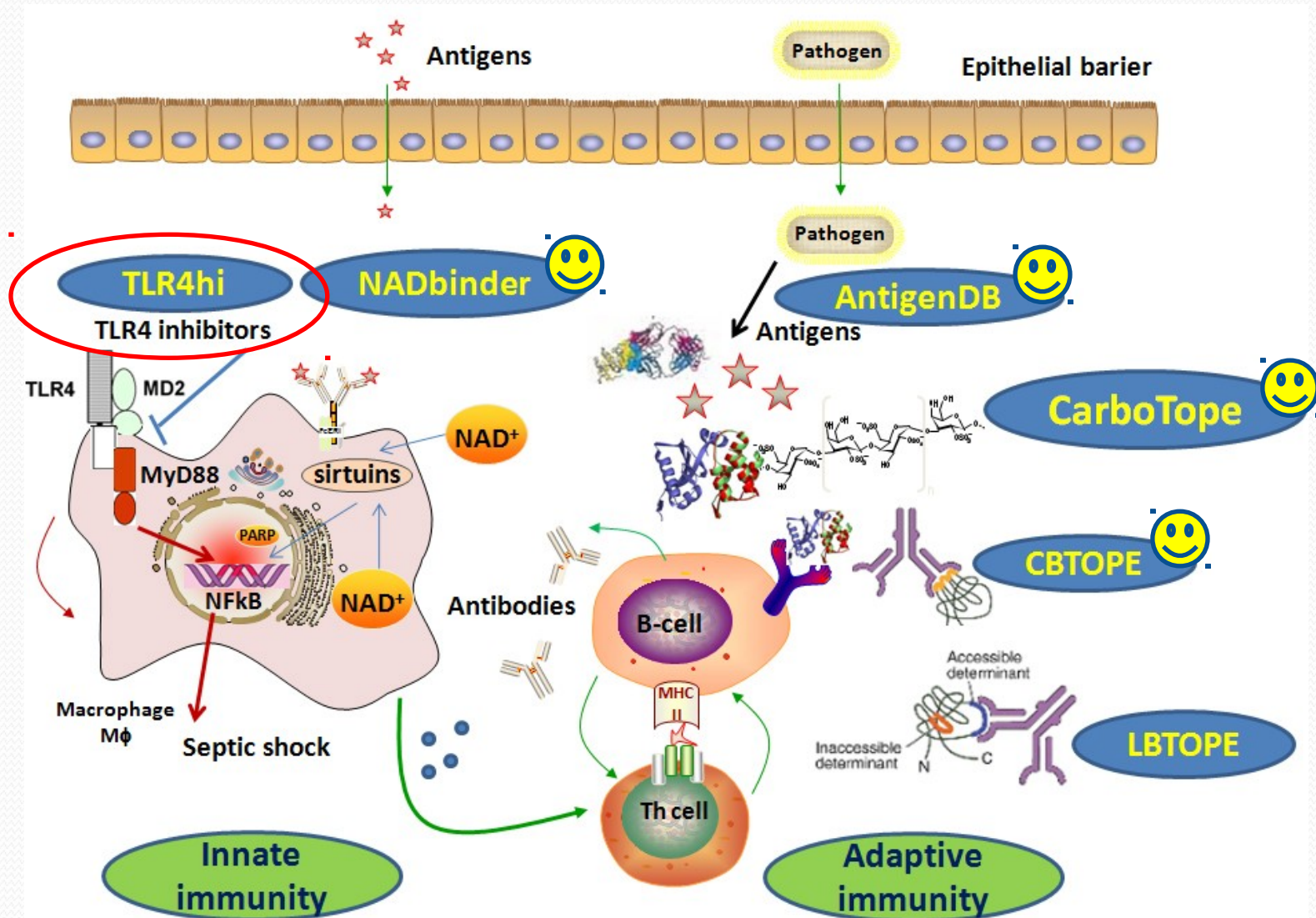
Legends:

- 1=amino acid position
- 2=Amino acid Sequence
- 3= probability scale (0-9) for each amino acid [Above 4 scale can be considered as epitope residue]
- 4= Secondary structure state of each residue as predicted by 'psipred' standalone program.
- H= Helical , E= Beta Sheet , C= Coil/turns.

>P40136_CYAA_BACANCalmodul			Length = 800			
1	010.....020.....030.....040.....050.....06					
2	MTRNKFIPNK	FSIISFSVLL	FAISSSQAI	ENVAMNEHYTE	SDIKRNHKT	ENKTEKEKFK
3	4443333333	3333333333	333332334	3333343333	3333333333	3333333333
4	CCCCCCCCC	ECEEEHHHHH	EEECCEEE	EEECCEECH	HHHHCCCCC	CCCCCHHHH

B-cell epitope prediction related resources

Database/server	URL	Description
ABcheck	http://www.bioinf.org.uk/abs/seqtest.html	Aligned antibody sequences to identify sequencing errors
ABCpred	http://www.imtech.res.in/raghava/abcpred/	Artificial neural network linear B cell epitope predictor
ABG	http://www.ibt.unam.mx/vir/structure/structures.html	Directory of antibody structures and sequence alignments
ANTIGENIC	http://bioinfo.bgu.ac.il/cgi-bin/emboss.pl?action	Identification of B cell epitopes
AntiJen	http://www.jenner.ac.uk/antijen/	Quantitative binding data for proteins, including B cell epitopes
BCEPred	http://www.imtech.res.in/raghava/bcepred/	Linear B cell epitopes prediction by physico-chemical properties
BCIPEP	http://www.imtech.res.in/raghava/bcipep	Database of B cell epitopes of varying immunogenicity
BepiPred	http://www.cbs.dtu.dk/services/BepiPred	Linear B cell epitope prediction
BEPITOPE	jlpelequer@cea.fr	Sequence based tool for continuous B cell epitope prediction
BEPro(PEPITO)	http://pepito.proteomics.ics.uci.edu/	Discontinuous B cell epitope prediction
CED	http://202.41.70.74:8080/cgi-bin/cep.pl	Predicts conformational epitopes for proteins
CEP	http://bioinfo.ernet.in/cep.htm	Linear and conformational epitope prediction using 3D structure
CID	http://ludwig-sun5.unil.ch/CancerImmunomeDB/	Documents antibody eliciting antigens in cancer patients
ClusPro	http://nrc.bu.edu/cluster/	Rigid protein-protein docking (Fast-Fourier transform correlation)
ConSurf	http://consurf.tau.ac.il	Identification of functional regions in proteins of known structure
DiscoTope	http://www.cbs.dtu.dk/services/DiscoTope	Sequence/structure based discontinuous epitope prediction
DNAPLOT	http://vbase.mrc-cpe.cam.ac.uk	Align a rearranged V gene to closest V, D and J germ line genes
EMT	elro@novozymes.com	Phage-display based linear, conformational epitopes prediction
EPIMAP	mumey@cs.montana.edu	Phage-display based discontinuous epitopes prediction
Epitome	http://www.rostlab.org/services/epitome	Database of antigenic residues and interacting antibodies
HaptenDB	http://www.imtech.res.in/raghava/haptendb/	A listing of haptens, structural similarity searches
HIV Immunology	http://www.hiv.lanl.gov/immunology	B cell HIV epitopes, proteome linear epitope maps, citations
IEDB	http://epitope2.immuneepitope.org/home.do	T- and B cell epitopes and non-epitopes database
IEDB Ab Epitope	http://www.immuneepitope.org/tools/bcell/iedb_input	Sequence based tool for the identification of continuous epitopes
IMGT/IG	http://imgt.cines.fr/cgi-bin/IMGTlect.jv	Immunoglobulin structures and annotated sequences
JenPep	http://www.jenner.ac.uk/JenPep/	Binding data for immunological protein-peptide interactions
MMDB	http://www.ncbi.nlm.nih.gov/Structure/MMDB/mmdb.shtml	Listing of crystal structures, including antibodies, HLA, and TCRs
PatchDock	http://bioinfo3d.cs.tau.ac.il	Rigid-body protein-protein docking based on local shape feature
PIER	n/a	Binding interface prediction using structure and sequence properties
PPI-PRED	http://bmbpcu36.leeds.ac.uk/ppi_pred/index.html	Protein-protein interface prediction
ProMate	n/a	Binding interface prediction using structure and sequence properties
SACS	http://www.bioinf.org.uk/abs/sacs	Summary of antibody structures
SDAP	http://fermi.utmb.edu/SDAP/	Structural database of allergenic proteins



. Epitope Mapping Using Phage Display Peptides

Tool	Description	Web site
LocaPep	Selection of seeds and clusters searching	http://atenea.montes.upm.es
MimoPro	Dynamic programming, branch & bound and compactness factor	http://informatics.nenu.edu.cn/MimoPro
Pep3DSearch	Implementation of ant colony optimization algorithm	http://kyc.nenu.edu.cn/Pep3DSearch/
PEPITOPE	Implementation of PepSurf and Mapitope	http://pepitope.tau.ac.il/
PepSurf	Stochastic-based color-coding method	http://pepitope.tau.ac.il/
MEPS	Surface ensemble and C β distances	http://www.caspur.it/meps
Mapitope	Physicochemical properties of mimotopes	http://pepitope.tau.ac.il/
MIMOP	MimAlign and MimCons	Software available upon request
MIMOX	Mimotope alignments and residue clustering	http://web.kuicr.kyoto-u.ac.jp/~hjian/mimox
3DEX	Physicochemical neighborhood of C α or C β atoms	Windows-based software

Linear B-Cell Epitope Tools

Tool	Web site	Reference
LEPS	http://leps.cs.ntou.edu.tw/	Wang et al. (28)
BayesB	http://www.immunopred.org/bayesb/	Wee et al. (31)
COBEpro	http://scratch.proteomics.ics.uci.edu	Sweredoski & Baldi (30)
BCPREDS/ FBCPRED	http://ailab.cs.iastate.edu/bcpreds	El-Manzalawy et al. (27, 29)
ABCpred	http://www.imtech.res.in/raghava/abcpred	Saha & Raghava (24)
BepiPred	http://www.cbs.dtu.dk/services/BepiPred	Larsen et al. (25)
Bcepred	http://www.imtech.res.in/raghava/bcepred	Saha & Raghava (22)
BEPITOPE/ PREDITOP	Standalone for Windows systems	Odorico & Pellequer (20)

Conformational B-Cell Epitope Tools

Tool	Description	Web site
EPSVR	Support Vector Regression and Meta server consensus	http://sysbio.unl.edu/EPSVR/
SEPPA	Concept of “unit patch of residue triangle” with spatial clustering coefficient	http://lifecenter.sgst.cn/seppa/
Epitopia	Physicochemical and structural geometrical features with Naïve Bayes	http://epitopia.tau.ac.il/
EPCES	Use of six different scoring functions	http://sysbio.unl.edu/EPCES/
BEpro	Combination of amino acid propensity scores and half-sphere exposure	http://pepito.proteomics.ics.uci.edu/
ElliPro	Residue protrusion index and neighbor clustering	http://tools.immuneepitope.org/tools/ElliPro/
PEPOP	Clustering of surface accessible segments	http://pepop.sysdiag.cnrs.fr/PEPOP/
DiscoTope	Combination of amino acid statistics, spatial information, and surface exposure	http://www.cbs.dtu.dk/services/DiscoTope/
CEP	Surface accessibility of residue	http://115.111.37.205/cgi-bin/cep.pl



Thanks for your attention