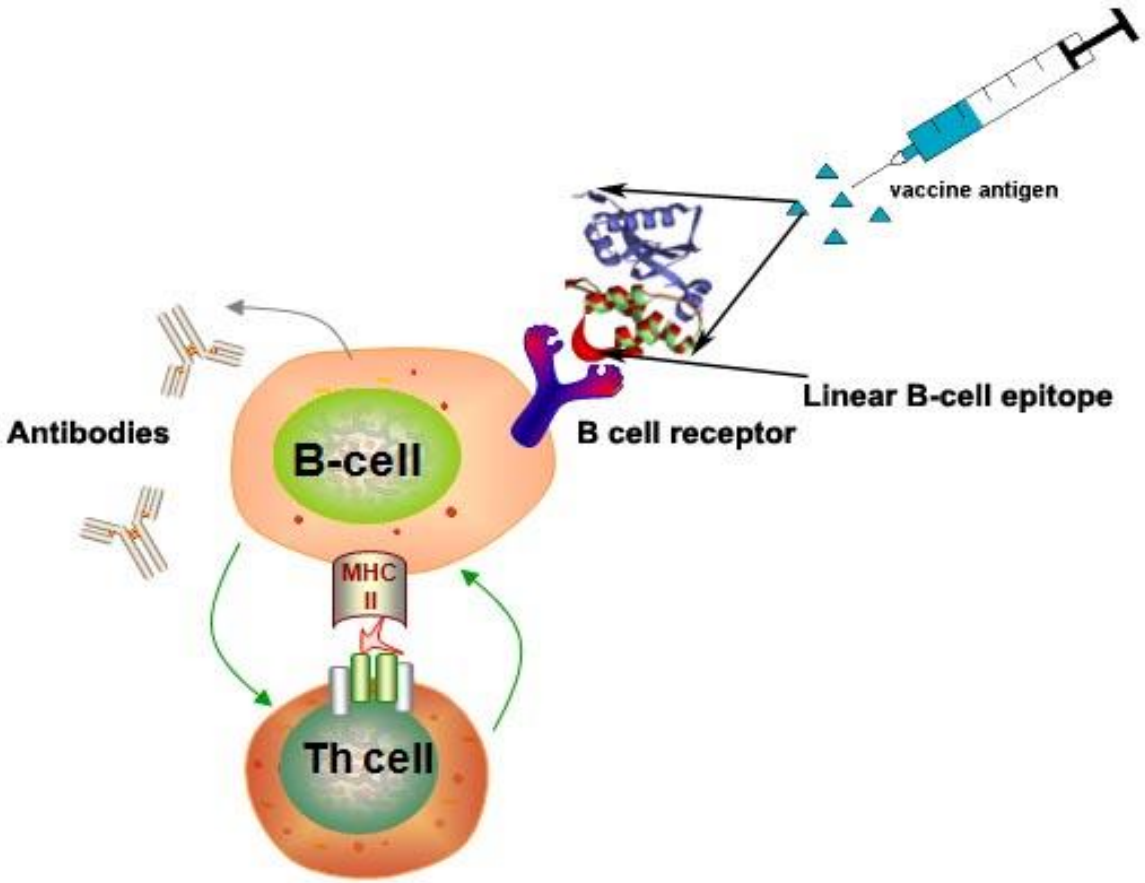
Fundamentals and Methods for B-Cell Epitope Prediction



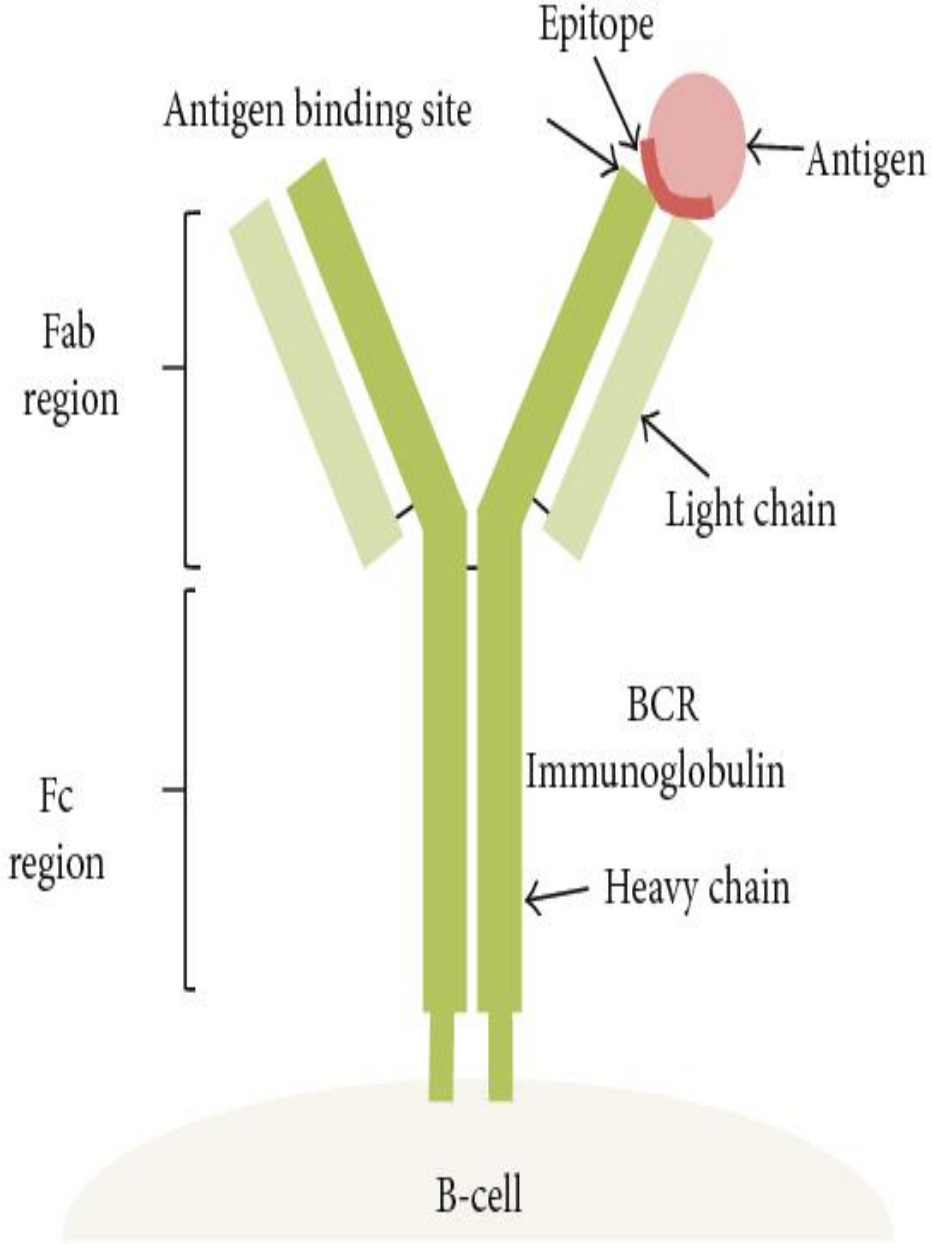
Purnima

Department of Bioinformatics

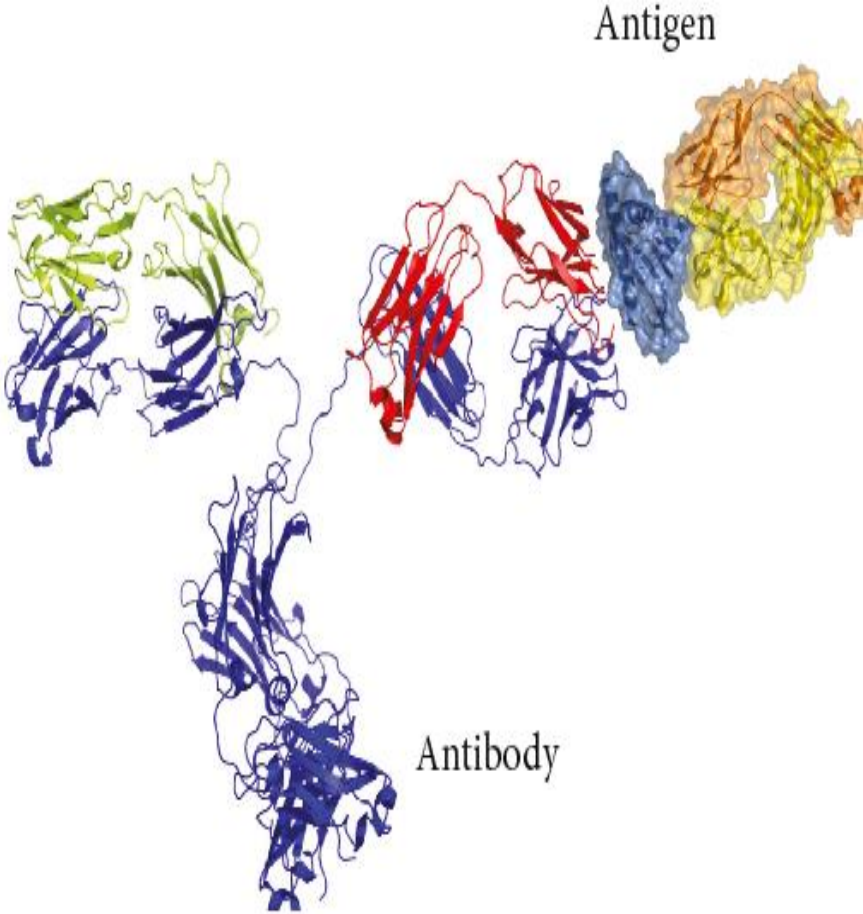
Introduction

- Antigen-antibody interaction is a key event in humoral immune response to invading pathogen. A specific antibody (Ab) recognizes antigen (Ag) at discrete regions known as **antigenic determinants or B-cell epitopes**.
- B-cell epitopes can be defined as a surface accessible clusters of amino acids, which are recognized by secreted antibodies or B-cell receptors and are able to elicit cellular or humoral immune response.
- B-cells form an essential part of the adaptive immune system, as they are capable of providing long-term protection against pathogens and harmful molecules.
- Their extremely specific B-cell receptors, named immunoglobulins or antibodies, are key components in this process. Antibodies recognize their molecular targets, termed antigens, via interactions between their binding site (paratope) and a specific region of the antigen (epitope).
- Most B-cell epitopes are discontinuous in sequence, meaning that they are composed of residues that might be far apart in sequence and are brought together in spatial proximity by the protein folding .

- E. D. Getzoff, J. A. Tainer, R. A. Lerner, and H. M. Geysen, “The chemistry and mechanism of antibody binding to protein antigens,” Advances in Immunology, vol. 43, pp. 1–98, 1988
- (Kringelum JV, Nielsen M, Padkjaer SB, Lund O. Structural analysis of B-cell epitopes in antibody:protein complexes. *Mol Immunol.* (2013) 53:24–34. doi: 10.1016/j.molimm.2012.06.001)


Epitope
Antigen binding site
Antigen
Fab region
Light chain
BCR
Immunoglobulin
Fc region
Heavy chain
B-cell

(a)


Antigen
Antibody

(b)

Jose L. Sanchez-Trincado, Marta Gomez-Perosanz, and Pedro A. Reche
Laboratory of Immunomedicine, Faculty of Medicine, Complutense University of Madrid, Ave Complutense S/N, 28040 Madrid, Spain(2017)

- Adaptive immunity is mediated by T- and B-cells, which are immune cells capable of developing pathogen-specific memory that confers immunological protection.
- Memory and effector functions of B- and T-cells are predicated on the recognition through specialized receptors of specific targets (antigens) in pathogens.
- More specifically, B- and T-cells recognize portions within their cognate antigens known as epitopes. There is great interest in identifying epitopes in antigens for a number of practical reasons, including understanding disease etiology, immune monitoring, developing diagnosis assays, and designing epitope-based vaccines.
- Epitope identification is costly and time-consuming as it requires experimental screening of large arrays of potential epitope candidates.
- Fortunately, researchers have developed in silico prediction methods that dramatically reduce the burden associated with epitope mapping by decreasing the list of potential epitope candidates for experimental testing.

W. E. Paul, Fundamental Immunology, Lippincott Williams & Wilkins, 2012.

- The immune system is typically divided into two categories, **innate** and **adaptive**.
- **Innate immunity** involves nonspecific defense mechanisms that act immediately or within hours after a microbe appearance in the body. All multicellular beings exhibit some kind of innate immunity.
- In contrast, **adaptive immunity** is only present in vertebrates and it is highly specific. In fact, the adaptive immune system is able to recognize and destroy invading pathogens individually. Moreover, the adaptive immune system remembers the pathogens that fights, acquiring a pathogen-specific long-lasting protective memory that enables stronger attacks each time the pathogen is reencountered .
- Nonetheless, innate and adaptive immune mechanisms work together and adaptive immunity elicitation is contingent on prior activation of innate immune responses.

W. E. Paul, Fundamental Immunology, Lippincott Williams & Wilkins, 2012.

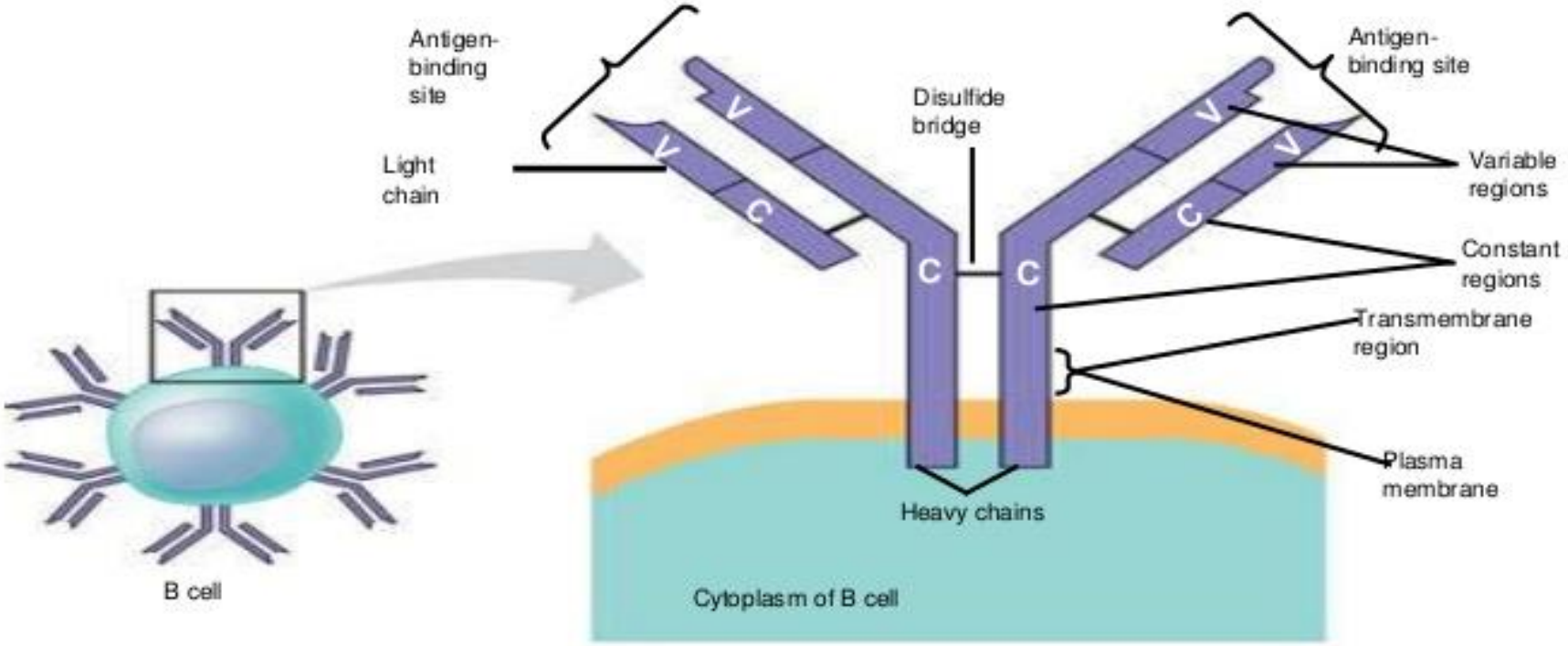
- Adaptive immunity is articulated by lymphocytes, more specifically by B- and T-cells, which are responsible for the humoral and cell-mediated immunity.
- B- and T-cells do not recognize pathogens as a whole, but molecular components known as **antigens**. These antigens are recognized by **specific receptors** present in the cell surface of B- and T-cells.
- Antigen recognition by these receptors is required to activate B- and T-cells but not enough, as second activation signals stemming from the activation of the innate immune system are also needed.
- The specificity of the recognition is determined by genetic recombination events that occur during lymphocyte development, which lead to generating millions of different variants of lymphocytes in terms of the antigen-recognizing receptors .
- Antigen recognition by B- and T-cells differ greatly.

B-cell epitope

- B-cells recognize solvent-exposed antigens through antigen receptors, named as B-cell receptors **(BCR),** consisting of membrane-bound immunoglobulins.
- Upon activation, B-cells differentiate and secrete soluble forms of the immunoglobulins, also known as antibodies, which mediate humoral adaptive immunity.
- A B-cell epitope is the antigen portion binding to the immunoglobulin or antibody. These epitopes recognized by B-cells may constitute any exposed solvent region in the antigen and can be of different chemical nature. However, most antigens are proteins and those are the subjects for epitope prediction methods.

B Cell Receptors for Antigens

- B cell receptors
  - Bind to specific, intact antigens
  - Are often called membrane antibodies or membrane immunoglobulin's



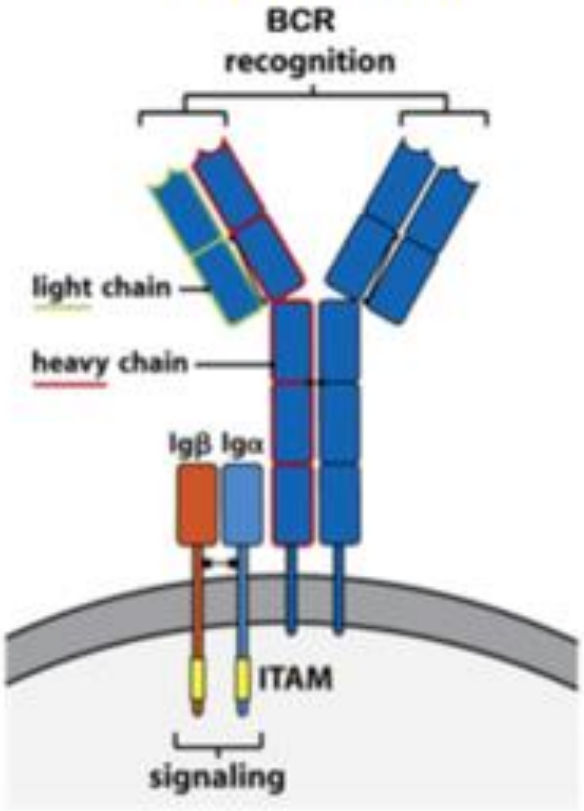
(a) A B cell receptor consists of two identical heavy chains and two identical light chains linked by several disulfide bridges.

T-cell epitope

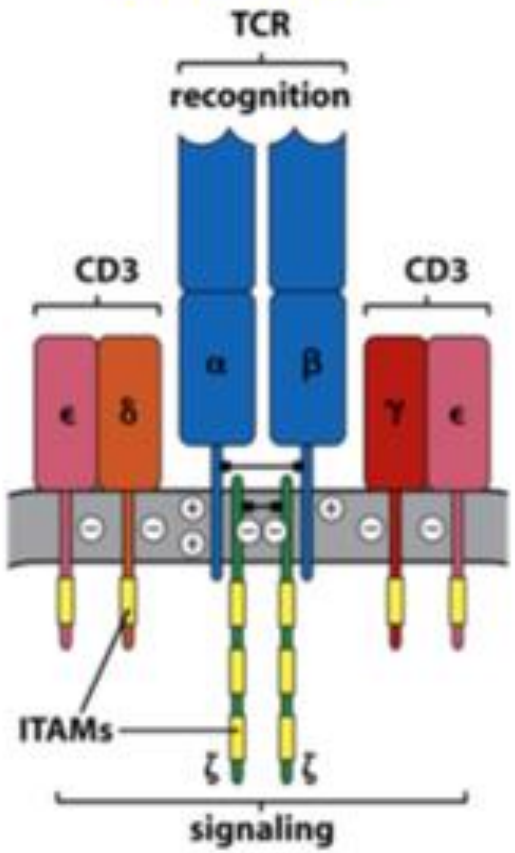
- **T-cells** present on their surface a specific receptor known as T-cell receptor **(TCR**) that enables the recognition of antigens when they are displayed on the surface of antigen-presenting cells (APCs) bound to major histocompatibility complex (MHC) molecules.
- **T-cell epitopes** are presented by class I (MHC I) and II (MHC II) MHC molecules that are recognized by two distinct subsets of T-cells, CD8 and CD4 T-cells.
- Subsequently, there are CD8 and CD4 T-cell epitopes. CD8 T-cells become cytotoxic T lymphocytes (CTL) following T CD8 epitope recognition.
- Meanwhile, primed CD4 T-cells become helper (Th) or regulatory (Treg) T-cells . Th cells amplify the immune response, and there are three main subclasses: **Th1** (cell-mediated immunity against intracellular pathogens), **Th2** (antibody-mediated immunity), and **Th17** (inflammatory response and defense against extracellular bacteria)

T Cell and B Cell Antigen Receptors (TCR and BCR)

B cell receptor



T cell receptor



M. H. Van Regenmortel, "What is a B-cell epitope?," Methods in Molecular Biology, vol. 524, pp. 3–20, 2009

- Identifying epitopes in antigens is of great interest for a number of practical reasons, including:
- understanding disease etiology,
- immune monitoring,
- developing diagnosis assays,
- designing epitope-based vaccines.

J. Ponomarenko and M. Van Regenmortel, "B-cell epitope prediction," in Structural Bioinformatics, pp. 849–879, John Wiley & Sons, Inc, 2009.

B-cell epitopes can be identified by different methods including :

- solving the 3D structure of antigen-antibody complexes,
- peptide library screening of antibody binding
- performing functional assays in which the antigen is mutated and the interaction antibody-antigen is evaluated .

Properties of B cell epitopes

B cell epitopes tend to be located in the flexible regions of an immunogen and display site mobility.

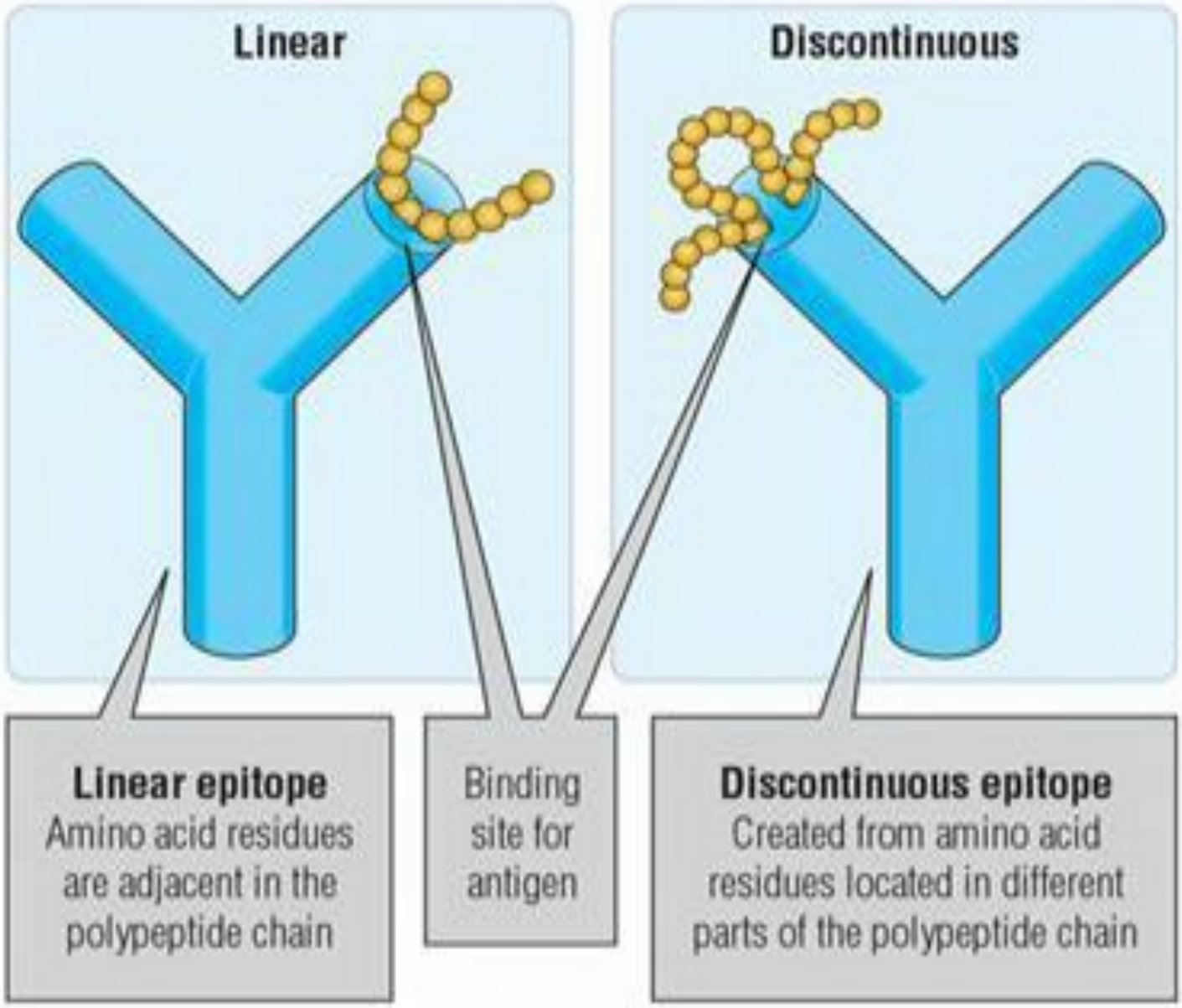
- **Major Antigenic determinants** in proteins generally located in the most **mobile regions**.
- **Site mobility** of epitopes **maximizes complementarity** with the Ag binding site**.**
- **But is of lower affinity due to loss of entropy.**

Complex proteins contain multiple overlapping B cell epitopes, some of which are immunodominant

- **Most of the surface** of a protein is potentially **antigenic**.
- **Subset of antigenic sites** on a given protein recognized by the immune system is much **smaller** than the **potential antigenic repertoire.**
- **Immunodominant epitopes** induce a more **pronounced immune response** than other epitopes of the same protein.

Prediction of B-Cell Epitopes

- B-cell epitope prediction aims to facilitate B-cell epitope identification with the practical purpose of replacing the antigen for antibody production or for carrying structure function studies.
- Any solvent-exposed region in the antigen can be subject of recognition by antibodies. Nonetheless, B-cell epitopes can be divided in two main groups: **linear and conformational .**
- **Linear B-cell epitopes** consist of sequential residues, peptides, whereas **conformational B-cell epitopes** consist of patches of solvent-exposed atoms from residues that are not necessarily sequential .
- Therefore, linear and conformational B-cell epitopes are also known as **continuous and discontinuous B-cell epitopes**, respectively. Antibodies recognizing linear B-cell epitopes can recognize denatured antigens, while denaturing the antigen results in loss of recognition for conformational B-cell epitopes.
- Most B-cell epitopes (**approximately a 90%) are conformational** and, in fact, only a minority of native antigens contains linear B-cell epitopes


Linear
Discontinuous
Linear epitope
Amino acid residues are adjacent in the polypeptide chain
Binding site for antigen
Discontinuous epitope
Created from amino acid residues located in different parts of the polypeptide chain

Prediction Methods

Several **experimental methods** for epitope identification are available including:

- Protein crystallography
- ELISA
- Peptide-chip

❖ But in general they are expensive, time consuming, low-throughput, or have low accuracy.

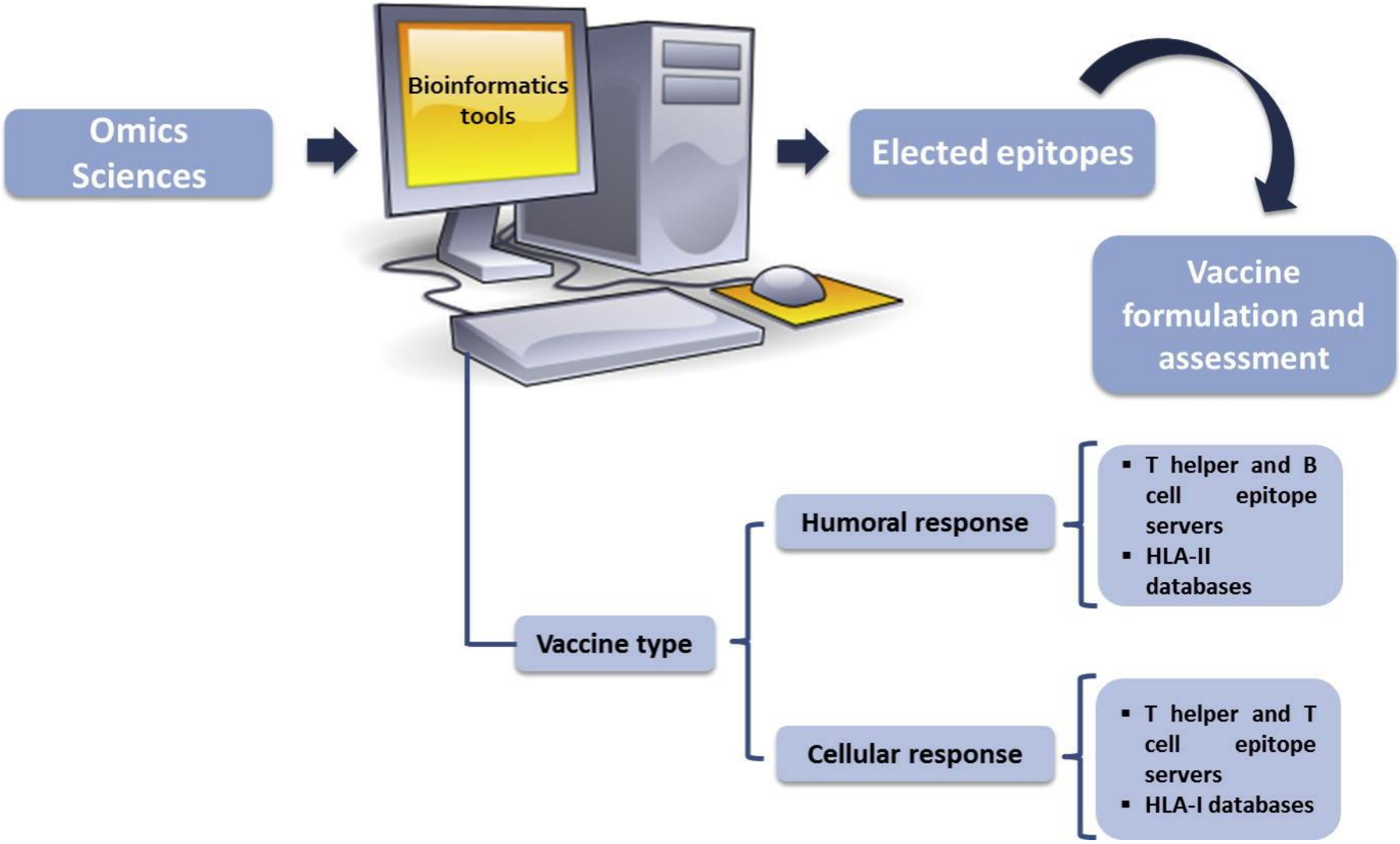
Immunoinformatics

- **The immunoinformatics** is a discipline whose main objective is to convert large-scale immunological data, using computational and mathematical approaches, to understand and organize these large scale data to obtain immunologically meaningful interpretations .
- The tools in this field are based on statistical and machine learning system and are used for studies in modeling molecular interactions (such as antigen processing and presentation) and also plays a role in defining new hypotheses related to understand the immune system mechanisms.

Computational methods

- Several computational methods have been developed to assist or substitute the experimental approaches, including BepiPred, DiscoTope, CBtope, and ABCpred .
- In absence of information on the cognate antibody, B-cell epitope prediction tools can broadly be categorized in two groups: **sequence- and structure-based** methods.
- As the names suggest, **sequence-based methods** predict the B-cell epitopes from the protein sequence of the antigen alone.
- Whereas **structure-based methods** take into account also their 3D structure.

Schematic representation of the workflow to identify epitopes for vaccine development



In Silico B-Cell Epitope Prediction

- Earlier prediction methods were **monoparametric** (based on single residue property or propensity scale) calculating average propensity value along a sliding window .
- It was demonstrated that methods based on **propensity profiling** yield poor results in the practice.
- To improve the performance of prediction of both continuous and discontinuous epitopes, machine learning methods were evolved.
- Currently used methods for continuous epitope prediction combine two or more residue properties with machine learning approaches .
- In general, prediction methods can be divided based on the level of input information to methods based on **antigen sequence** and methods based on **3D structure of antigen**.

Prediction of Continuous B-Cell Epitopes

- The first prediction method using recurrent neural network, **ABCPred,** has been trained on B-cell epitopes obtained from BciPep database and nonepitopes obtained randomly from Swiss-Prot database.
- The ABCPred is a neural network based method for prediction of continuous B-cell epitopes using fixed length pattern .
- The ABCPred dataset contains data of epitopes from viruses, bacteria, parasites, and fungi that are stored in BciPep database with the prediction accuracy of 65.9%. ABCPred, AAP method and BCPred, and BayesB predict only short peptide fragments.
- The B-cell epitopes of the Emy162 protein of *Echinococcus multilocularis* (the causative agent of zoonotic helminthosis) were predicted using BCPred and ABCPred

Prediction of Continuous B-Cell Epitopes

- **BCPred method** employs subsequence kernel-based SVM classifier and was trained on homology-reduced dataset of linear B-cell epitopes (with <80% sequence identity) derived from dataset previously used to test ABCPred. The performance of BCPred (AUC 0.758) outperforms implementation of AAP (AUC 0.7).
- **FBCPred** is a novel method developed for prediction of B-cell epitopes with flexible length. Homology-reduced dataset is publicly available for comparing existing linear B-cell epitope prediction methods and testing of new prediction software.
- **BepiPred** predicts continuous epitopes by combining two residues properties with Hidden Markov Model. BepiPred was evaluated on dataset of epitopes extracted from the literature, AntiJen, and HIV databases. This method has a quite low sensitivity [19].
- The **server BcePred** is used for prediction of continuous B-cell epitopes based on physicochemical properties and allows user to select any residue property or combination of two or more properties employed in prediction. The performance of BcePred was evaluated on dataset containing epitopes obtained from BciPep database and dataset of randomly chosen nonepitopes from Swiss-Prot. The accuracy of BcePred combining four amino acid properties (hydrophilicity, flexibility, polarity, and exposed surface) is 58.70%
- **B-cell epitope prediction using support vector machine tool (BEST)** is sequence-based tool designed for prediction of both linear and conformational epitopes from full antigen sequence. Prediction is based on averaging of selected scores (sequence conservation, similarity to experimentally validated B-cell epitopes, predicted secondary structure, and relative solvent accessibility) generated from 20-mers. BEST achieves AUC at 0.81 and 0.85 for the fragment-based prediction and 0.57 and 0.6 for full antigen. BEST outperforms several modern sequence-based B-cell epitope predictors including ABCPred, BCPred, COBEpro, and CBTOPE

Prediction of Discontinuous B-Cell Epitopes

- **BEpro server** (formerly known as PEPITO) uses a combination of amino acid propensity scores along with side chain orientation and solvent accessibility information using half sphere exposure values at multiple distances to predict discontinuous B-cell epitopes.
- **PEPOP** is structure-based method, which identifies clusters of accessible surface residues and segments that might form putative discontinuous epitopes and can be used to design immunogenic peptides.
- Computational prediction tool **EPITOPIA** employs Naïve Bayes classifier to predict epitopes in linear sequence or 3D structure. It distinguishes the nonepitope and epitope regions by computing an immunogenicity score (reflecting the immunogenic potential of a certain residue relative to all residues in the antigen) for each solvent accessible residue or a score for every amino acid. EPITOPIA yields higher success rate of 89.4% (mean AUC value of 0.60) when compared to ElliPro and DiscoTope.
- **CBTOPE** was proposed for the prediction of discontinuous epitopes from antigen primary structure. This SVM-based predictor combines traditional features of physicochemical profiles and sequence-derived inputs including composition and collocation of amino acids. It outperformed other structure methods using binary profile of pattern and physicochemical profile of patterns with better sensitivity and AUC on the same benchmark dataset.
- Improved Spatial Epitope Prediction of Protein Antigens server (**SEPPA)** focusing on single residue propensity scales and continual segment clustering was developed in 2009 by Sun and colleagues . SEPPA employs a novel concept of unit patch of residue triangle and spatial clustering coefficient to describe local spatial context in protein antigen surface and 3D characteristic of epitopes. A parameter of 4 Å was chosen in the definition of unit patch of residue triangle. Curated data of nonredundant spatial epitopes from PDB database was used for method testing. SEPPA outperforms popular prediction tools, CEP, DiscoTope, and BEpro, and achieves an average AUC over 0.742.

B-Cell Epitope Databases

- The B-cell epitope databases can be classified as multifaceted database such as IEDB and AntiJen.
- B-cell oriented database such as BciPep, Epitome, and SDAP.
- Single pathogenic organism oriented database such as the HIV Molecular Immunology Database, FLAVIdB, and Influenza Sequence and Epitope Database.

❖ An Introduction to B-Cell Epitope Mapping and In Silico Epitope Prediction, Lenka Potocnakova,1 Mangesh Bhide,1,2 and Lucia Borszekova Pulzova1