Biology Knowledge

Contents

1 Basic Knowledge

1.1 Evolutionary information

Evolutionary information refers to data and insights gained from studying how biological molecules, like proteins and DNA, have changed over time. This includes:

1. Sequence Conservation: Identifying regions that remain unchanged across different species, indicating important functional or structural roles.
2. Phylogenetic Relationships: Understanding the evolutionary history and connections between organisms through genetic similarities and differences.
3. Mutational Patterns: Analyzing how mutations affect function and structure, which can reveal adaptive changes or disease mechanisms.

1.2 Central Dogma

The **Central Dogma of Molecular Biology** describes the flow of genetic information:

$$\text{DNA} \rightleftharpoons \text{RNA} \rightarrow \text{Protein}.$$

This process can be divided into four major steps: replication, transcription, RNA processing, and translation. Each step has distinct molecular mechanisms, but all are constrained by the directionality of nucleic acids: synthesis occurs strictly in the $5' \rightarrow 3'$ **direction**.

1.2.1 Replication

Replication ensures faithful inheritance of genetic material when cells divide.

- DNA helicase (解旋酶) unwinds the double helix.
- DNA polymerase (聚合酶) synthesizes complementary strands using existing strands as templates.
- Synthesis proceeds continuously on the **leading strand** ($3' \rightarrow 5'$) and discontinuously on the **lagging strand** ($5' \rightarrow 3'$) (Okazaki fragments).

- Result: two daughter DNA molecules, each consisting of one parental and one newly synthesized strand (semi-conservative replication). Typically, two daughter DNA molecules are the same as their original DNA.

1.2.2 Transcription (DNA → RNA)

Transcription converts DNA-encoded information into RNA transcripts.

- RNA polymerase binds to promoter regions upstream of genes.
- It synthesizes an RNA copy of the coding strand, complementary to the template strand.
- In prokaryotes, transcription directly produces functional mRNA.
- In eukaryotes, the initial product is **pre-mRNA**, which requires further processing.

1.2.3 RNA Processing (pre-mRNA → mature mRNA)

In eukaryotic cells, primary transcripts undergo several modifications:

- **5 capping**: a modified guanine nucleotide is added to protect RNA from degradation and assist ribosome binding.
- **Splicing**: introns (non-coding regions) are removed, exons are joined together.
- **Polyadenylation**: a poly(A) tail is added at the 3 end, enhancing stability and export from the nucleus.

The final mature mRNA is exported to the cytoplasm for translation.

1.2.4 Translation (RNA → Protein)

Translation converts mRNA into a polypeptide sequence.

- Ribosomes initiate translation at the start codon (AUG).
- Transfer RNAs (tRNAs) bring amino acids to the ribosome, matching codons with anticodons.

- The ribosome catalyzes peptide bond formation as it moves codon by codon along the mRNA.
- Translation terminates when a stop codon (UAA, UAG, or UGA) is encountered.
- The resulting polypeptide folds into its functional three-dimensional structure.

Proteins carry out structural, enzymatic, and regulatory roles in the cell, completing the flow of genetic information.

1.2.5 Reverse Transcription (RNA → DNA)

Although the Central Dogma describes the usual flow of information from DNA to RNA to Protein, there are important exceptions. One of the most notable is **reverse transcription**, where genetic information flows from RNA back to DNA.

Mechanism.

- The enzyme **reverse transcriptase** synthesizes complementary DNA (cDNA) using RNA as a template.
- The newly synthesized cDNA can integrate into the host genome.
- RNase H activity often removes the RNA strand after cDNA synthesis.

Biological Context.

- Retroviruses (e.g., HIV) use reverse transcription to replicate within host cells.
- Retrotransposons within eukaryotic genomes also employ this process.
- In biotechnology, reverse transcription is applied in **RT-PCR** to convert RNA into cDNA for expression studies.

Summary. Reverse transcription demonstrates that information flow is not strictly one-way. In certain biological and experimental contexts, RNA can be converted back into DNA, broadening the flexibility of genetic information transfer.

1.3 propensity score of residues

In bioinformatics and structural biology, the **propensity score of residues** refers to the *likelihood or tendency of specific amino acid residues* to adopt certain structural features, participate in specific interactions, or be located in particular regions (like on the surface or in the core) of a protein. Propensity scores help predict and understand protein structure, stability, and function.

Types of Propensity Scores

1. **Secondary Structure Propensity**:
    - Describes the likelihood of an amino acid residue to form a specific type of secondary structure, like an *alpha helix*, *beta sheet*, or *coil*.
    - For example, *alanine* has a high propensity to form alpha helices, while *valine* and *isoleucine* favor beta sheets.
2. **Surface Accessibility Propensity**:
    - Indicates the likelihood of a residue to be found on the *protein surface* (exposed to solvent) or *buried* within the protein core.
    - For example, *hydrophobic residues* like leucine and valine have a lower surface propensity, while *hydrophilic residues* like lysine and arginine are more likely to be exposed on the surface.
3. **Hydrophobicity/Hydrophilicity Propensity**:
    - Measures the tendency of a residue to interact with water (hydrophilicity) or avoid water (hydrophobicity).
    - This property is often used to predict surface residues and binding sites, as hydrophilic residues are more likely to be solvent-accessible.
4. **Binding Propensity**:
    - Indicates the likelihood of a residue to be involved in *binding interactions*, such as with ligands, DNA, RNA, or other proteins.
    - Specific residues like *tyrosine* and *tryptophan* often have a higher binding propensity in protein interfaces.

5. **Epitope Propensity**:

    - Used in immunology to predict *B-cell or T-cell epitopes* by identifying residues that are likely to be antigenic or recognized by antibodies.
    - Epitopes often have specific residues with a higher propensity for being located in accessible and flexible regions of the protein.

Applications of Propensity Scores in Structural Biology

- **Protein Structure Prediction**: Helps predict secondary structures and surface accessibility, contributing to computational modeling of protein structures.
- **Epitope Mapping**: Identifies regions of antigens with high epitope propensity to predict potential B-cell or T-cell binding sites, useful in vaccine design.
- **Binding Site Prediction**: Propensity scores help determine likely binding sites for small molecules, DNA, RNA, and other proteins by identifying residues with high binding propensity.
- **Stability and Folding Studies**: Analyzes residue-specific propensities to infer which regions stabilize the protein structure or contribute to folding pathways.

Example

Suppose we' re analyzing a protein structure to identify likely binding sites. Propensity scores for binding can highlight residues with a high likelihood of forming stable interactions. If *tyrosine* has a high binding propensity score in a particular region, that site may be a promising candidate for binding interactions.

Summary

The **propensity score of residues** is a predictive measure indicating how likely a residue is to adopt specific structural roles or participate in interactions. It is an essential tool in structural biology, allowing researchers to make educated predictions about protein structure, function, and interactions based on residue behavior.

1.4 Cα (alpha carbon) atoms

The **number of Cα (alpha carbon) atoms within a certain distance** is a metric used in structural biology to describe *local residue density, packing, or spatial proximity* of residues within a protein.

- **Cα Atom**: The central carbon atom in each amino acid, connected to the amino, carboxyl, and side chain groups, often used as a reference point in protein structures.
- **Interpretation of Cα Distance Metric**:
  - **Residue Density**: A higher number of Cα atoms within a specified distance (e.g., 5 Å) indicates dense packing, while a lower number suggests a more exposed or flexible area.
  - **Structural Packing and Interaction Sites**: Dense Cα regions often imply stable structural cores or potential binding sites in protein-protein interactions.

For example, a Cα atom with *10 other Cα atoms within a 5 Å radius* suggests a densely packed core, while one with *3 Cα atoms within the same radius* may indicate an exposed region.

1.5 Post-translational modifications (PTMs)

Post-translational modifications (PTMs) are chemical alterations that proteins undergo after their synthesis (translation) in the cell. These modifications can significantly influence a protein's function, activity, stability, localization, and interactions with other molecules. Common types of PTMs include:

- **Phosphorylation:** Addition of a phosphate group, often regulating enzyme activity and signaling pathways.
- **Glycosylation:** Attachment of carbohydrate moieties, which can affect protein folding, stability, and cell-cell interactions.
- **Ubiquitination:** Addition of ubiquitin proteins, typically marking a protein for degradation by the proteasome.

- **Methylation:** Addition of methyl groups, influencing gene expression and protein interactions.
- **Acetylation:** Addition of acetyl groups, often impacting gene expression and protein stability.

These modifications enhance the functional diversity of the proteome, allowing a single gene to give rise to multiple protein forms with distinct functions. PTMs are crucial for regulating cellular processes and responding to environmental changes."

2 Immunity

2.1 Pathogen (病原体), antigen, antibodies:

Pathogens are harmful organisms like **viruses, bacteria, fungi, or parasites** that can cause diseases in the body. These pathogens are made up of many components, including **proteins, lipids, and carbohydrates.**

Antigens are small fragments, often **pieces of proteins** from the pathogen. These antigens are what the immune system recognizes as **foreign and harmful**. They act like ”flags” that alert the immune system to the presence of an invader.

Antibodies are proteins produced by **B cells** (a type of immune cell). Their job is to specifically recognize and bind to antigens. Once bound, antibodies help neutralize the pathogen or mark it for destruction by other immune cells. Each antibody is highly specific to a particular antigen, like a lock and key, and this specificity helps the immune system target and eliminate pathogens more efficiently.

2.2 Self and Antigenic Nonamer Peptides:

Nonamer peptides are peptides made up of 9 amino acids. These peptides can be derived from self-proteins (normal proteins from the body) or antigenic proteins (proteins from pathogens like viruses or bacteria).

Self peptides help the immune system distinguish normal body cells from infected or abnormal cells, while **antigenic peptides** signal the presence of a pathogen.

2.3 Cell Types

B Cells:

- **Role:** Part of humoral immunity; responsible for producing antibodies.
- **Function:** Recognize specific antigens via B-cell receptors (BCRs) and differentiate into plasma cells to secrete antibodies.

T Cell: T cell cannot bind free-floating pathogens, which can only bind with major histocompatibility complex (MHC) on the surface of antigen-presenting cells (APC)

Helper T Cells (CD4+):

- **Role:** Coordinate the immune response by activating other immune cells.
- **Function:** Recognize antigens presented by MHC Class II molecules and provide signals to activate B cells and cytotoxic T cells.

Cytotoxic T Cells (CD8+):

- **Role:** Part of cell-mediated immunity; directly kill infected or cancerous cells.
- **Function:** Recognize antigens presented by MHC Class I molecules and induce apoptosis in target cells.

2.4 V(D)J Recombination

V(D)J recombination is the process by which B cell receptors (BCRs) and T cell receptors (TCRs) achieve their enormous diversity. Instead of each receptor being encoded by a single fixed gene, the immune system generates diversity by *randomly combining* gene segments.

2.4.1 Genetic Segments

- **V (Variable)** segments: provide the main variable part of the receptor.
- **D (Diversity)** segments: only exist in heavy chains (BCR) and in the TCR β/δ chain.
- **J (Joining)** segments: connect to the constant region.
- Light chains (Igκ, Igλ) of BCR/Antibody and TCR α chains use only V and J, while heavy chains and TCR β use V + D + J.

Number of segments. In the human genome, each of the V, D, and J categories is represented by multiple gene segments. For example, the immunoglobulin heavy chain locus contains ∼40 functional V segments, ∼23 D segments, and ∼6 J segments. Light chains (Igκ, Igλ) have ∼30–40 V and a few J segments but no D. TCR loci are similarly organized: the β chain has ∼48 V, 2 D, and 13 J segments, while the α chain has ∼45 V and 50 J segments. This multiplicity of choices, combined with random recombination, is the basis for the enormous receptor diversity. Meanwhile, the pairing of heavy (β) chain and light (α) chain is also random.

2.4.2 Mechanism

- **Step 1: Recognition.** Enzymes RAG1 and RAG2 recognize special recombination signal sequences (RSS) flanking each segment.
- **Step 2: Cutting.** DNA is cut at these positions, and intervening DNA is removed.
- **Step 3: Joining.** The selected V, (D), and J segments are joined together by a DNA repair process called non-homologous end joining (NHEJ).
- **Step 4: Junctional diversity.** Extra randomness is introduced:
  - Deletion of nucleotides at the ends of segments.
  - Addition of random nucleotides by the enzyme TdT (N nucleotides).
  - Formation of short palindromic (P) nucleotides during repair.

Non-Homologous End Joining (NHEJ). NHEJ is the major pathway by which cells repair DNA double-strand breaks (DSBs). During V(D)J recombination, the RAG1/2 enzymes cleave DNA flanking the V, D, and J segments, leaving free DNA ends. The Ku70/Ku80 heterodimer quickly binds to these ends, protecting them and recruiting DNA-PKcs. Artemis and other nucleases then process the broken ends, while TdT (terminal deoxynucleotidyl transferase) can randomly add nucleotides, introducing additional diversity. Finally, the XRCC4–Ligase IV complex completes the ligation step. Because NHEJ does not rely on long regions of homology, it often results in small insertions or deletions at the junctions, which significantly contribute to the diversity of the CDR3 region.

Recombination Signal Sequences (RSS). RSS are short conserved DNA motifs located adjacent to each V, D, and J gene segment, serving as recognition sites for the RAG enzymes. A canonical RSS consists of a conserved heptamer sequence (e.g., `CACAGTG`), a spacer of either 12 or 23 base pairs, and a conserved nonamer (e.g., `ACAAAAACC`). V(D)J recombination follows the so-called "12/23 rule": only an RSS with a 12 bp spacer can pair with one carrying a 23 bp spacer, ensuring proper joining of segments. This rule prevents inappropriate recombination and enforces correct assembly of V, D, and J elements. Conceptually, RSS can be thought of as "cutting marks" in the DNA, with RAG1/2 acting as the molecular scissors that recognize and cleave at these sites.

2.4.3 Functional Outcome

- The final recombined gene encodes the receptor's **variable region**.
- CDR1 and CDR2 loops are encoded within the V segment, which is known if V is given.
- The highly diverse **CDR3 loop** spans the V–(D)–J junction and is the most important for antigen specificity.
- This mechanism allows the immune system to create up to 10^{13} possible receptor sequences, enough to recognize a vast range of antigens.

2.4.4 Summary (Intuition)

V(D)J recombination is like *shuffling and recombining Lego blocks*:

- Choose one V block, (optionally one D block), and one J block.
- Cut and paste them together.
- Add some random glue pieces (extra nucleotides).

The result is a unique receptor gene that no other cell has.

2.5 Procedure of Two Kinds of Adaptive Immunity

Cellular Immunity (Cell-Mediated): During the process of maturation, billions of variations of T-cells are formed, each carrying a unique surface protein, called T-cell receptor, TCR. Both kinds of T-cells are included.

Basically, specific immunity relies on the invading pathogen finding a match among these billions of T-cell variations. Only the ones that can bind to the pathogen, are selectively activated. T-cells, however, cannot bind free-floating pathogens. They can only bind to pieces of the pathogen bound to a certain host molecule called major histocompatibility complex, or MHC, on the surface of so-called "antigen-presenting cells" . There are two classes of MHC:

MHC Class I: MHC class I molecules are expressed by all nucleated cells of the body. These molecules are constantly produced in the cytoplasm and, on their way to the cell membrane, pick up pieces of peptides and display them on the cell surface. If a cell is infected by a

virus or is cancerous, a foreign or an abnormal antigen is displayed; and the cell can bind and activate a matching T-cell. (Recognized by CD8+ T Cells (Cytotoxic T Cells))

1. **Length:** Peptides that bind to MHC Class I molecules are typically 8 to 10 amino acids long, with 9-mers (nonamers) being the most common length.
2. **Binding Groove:** The binding groove of MHC Class I is closed at both ends, which restricts the peptide length and positions the ends of the peptide in defined positions within the groove.
3. **Anchor Residues:**
    - The key anchor residues for MHC Class I binding are typically found at position 2 (P2) and position 9 (P9).
    - These residues are essential for fitting into specific pockets within the MHC Class I binding groove, ensuring stable binding.
    - Common anchor residues at P2 include Leucine (L) or Methionine (M), and at P9, you often find Valine (V) or Leucine (L).
4. **Function:** These peptides are typically derived from intracellular proteins (e.g., viral proteins) and presented to CD8+ cytotoxic T cells, which then target and destroy the infected or abnormal cells.

MHC Class II: Only presented on professional antigen-presenting cells, including dendritic cells, B-cells, macrophages and reticular cells. These cells ”eat” invaders like bacteria, break them down, and show pieces of the invader on their MHC Class II. They are recognized by CD4+ T Cells (Helper T Cells). This helps alert other immune cells to come and destroy the infection.

1. **Length:** Peptides that bind to MHC Class II molecules are longer, typically ranging from 12 to 25 amino acids. This is due to the open-ended nature of the MHC Class II binding groove, which accommodates longer peptides.
2. **Binding Groove:** The binding groove of MHC Class II is more flexible and can accommodate peptides of various lengths by letting the peptide hang out from either end.
3. **Anchor Residues:**

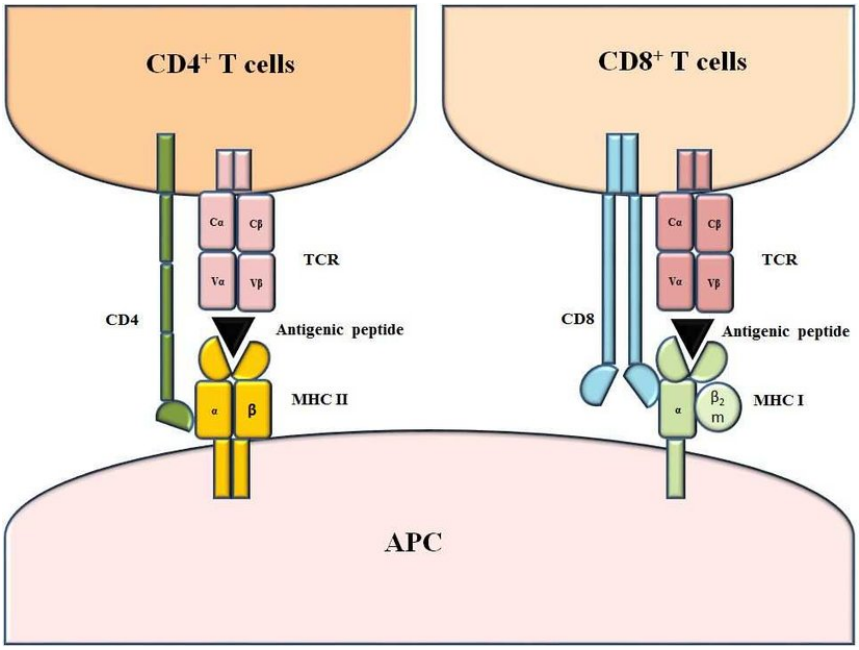

图 1: TCR Binding

- Anchor residues in MHC Class II peptides are more distributed along the length of the peptide. The positions can vary depending on the MHC Class II allele.
- Unlike MHC Class I, there are not just two key anchor positions, but rather a series of motifs that interact with the binding pockets along the groove.

4. **Function:** MHC Class II peptides are presented to **CD4+ helper T cells**, which help activate other immune cells, including B cells and macrophages. These peptides are typically derived from extracellular pathogens (e.g., bacteria) that have been engulfed and processed by antigen-presenting cells.

Activation of T-cells requires a second binding between the two cells. This is the verification step, a safeguard mechanism serves to prevent the immune system from overreacting.

Once activated, T-cells undergo repeated cycles of mitosis in a process called clonal expansion. This process produces clones of identical cytotoxic and helper T-cells, both of which are specific to the pathogen. Some of these cells differentiate into effector cells, while other become memory cells.

Helper T-cells produce interleukins which help with the activation of cytotoxic T-cells, B cells, and other immune cells. Cytotoxic T-cells, on the other hand, are the main actors of cellular immunity. They release toxins and directly kill infected or cancerous host cells. While effector

cells die during or shortly after the infection, memory cells live for much longer periods of time. Memory T-cells are also more numerous than the original naïve T-cells. Upon reexposure to the same pathogen, they can mount a much faster immune response, destroying the pathogen so quickly that no signs of illness are noticeable.

- **Antigen Presentation:** Infected cells present antigens on MHC Class I molecules.
- **Recognition:** Cytotoxic T cells recognize these antigens and receive additional activation signals from helper T cells.
- **Target Cell Destruction:** Activated cytotoxic T cells induce apoptosis in infected or cancerous cells.

Humoral Immunity (Anitbody-Mediated): B-cells develop in the bone marrow and complete their maturation in the spleen. B-cells are formed in billions of variations, each carrying a unique surface protein, called B-cell receptor, BCR. T-cells and B-cells are usually separated into defined T-cell and B-cell zones within these organs.

Specific immunity relies on the invading pathogen finding a match among these many variations of B-cells. Only cells that can bind to the pathogen, can be activated to produce antibodies. B-cell surface receptors, BCRs, are actually membrane-bound antibodies.

Each B-cell has thousands of identical copies of BCR on its surface. When a pathogen binds, it usually binds to several of these receptors, linking them together, and triggering endocytosis of the pathogen. B-cells then cut the pathogen into pieces and display them on MHC-II molecules on their surface. Thus, B-cells now become antigen-presenting cells, but are **not yet activated.** In most cases, activation of antigen-primed B-cells does not happen until they are stimulated by antigen-specific T-helper cells.

Nearby, in the T-cell zone, T-helper cells are activated by dendritic cells carrying antigens of the same pathogen and become effector T-helper cells. Some of these effector cells leave lymph nodes for the site of infection, while other, namely the follicular helper cells, migrate to T-cell B-cell borders, and bind to the antigens presented by B-cells. This interaction triggers T-cells to produce helper factors, which activate B-cells.

Activated B-cells undergo the first rounds of proliferation and differentiation, giving rise to the first batch of plasma cells producing antibodies, mainly of IgM class; and a group of cells that are committed to become memory B-cells. The latter undergo antibody class switching; and form a so-called germinal center, where they go through cycles of multiplication and

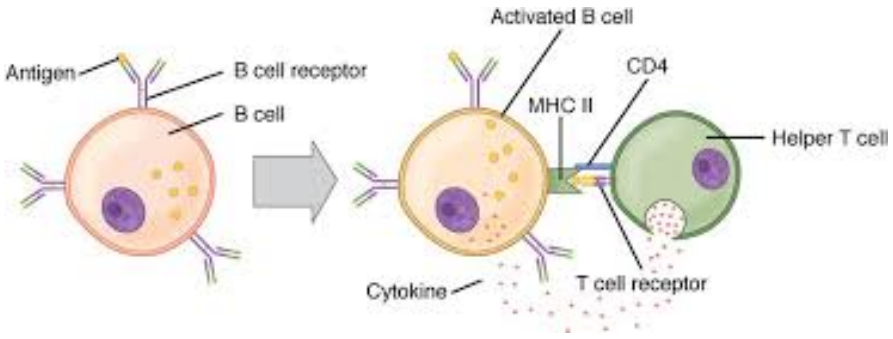

图 2: Activation of B cell and T cell.

hypermutation in the immunoglobulin gene. This process produces slightly different variations of the same antibody, which are then subject to a binding test to the same antigen. Those that no longer bind are discarded, while the remaining compete for binding to antigen-specific T-helper cells. B-cells with the highest affinity to the antigen win the interaction with T-helpers and exit the germinal center. They can either become long-lived memory B-cells, or differentiate into antibody-producing plasma cells. This second batch of plasma cells produces better antibodies and lives longer than the first batch. They also make antibodies of different classes (predominantly IgG), which neutralize the pathogen in many different ways. Upon re-exposure to the same pathogen, memory B-cells mount a much faster immune response.

Video

- **Antigen Recognition:** B cells bind directly to antigens via BCRs.
- **Activation:** Helper T cells recognize antigen-MHC Class II complexes on B cells and provide activation signals.
- **Antibody Production:** Activated B cells differentiate into plasma cells, producing antibodies that neutralize pathogens.

2.6 Major Histocompatibility Complex (MHC):

这是抗原内部蛋白质上的某一段序列，他们都是被 T cell 所识别而激发免疫系统。

MHC molecules are like display boards on the surface of your body’ s cells. They grab small pieces (called antigens) of proteins from inside the cell and show them on the cell surface.

These antigens can be from normal proteins or from invaders like viruses or bacteria.

T cells then "check" what the MHC is displaying to see if it' s normal or dangerous.

An **MHC Class I molecule**, such as HLA-A2, presents peptides to CD8+ T cells. It consists of three main parts:

- A heavy chain (HLA-A2 itself).
- A peptide (typically 8-10 amino acids long), which is loaded onto the HLA-A2 molecule.
- A small protein called β2-microglobulin (p2m), which helps stabilize the HLA-peptide complex.

The stability of the peptide-HLA-A2 complex is often assessed by measuring the rate of dissociation of β2-microglobulin (p2m). In a stable complex, p2m remains associated with the HLA-A2 molecule, ensuring the peptide stays bound and presented to T cells.

2.7 Complementarity Determining Region (CDR)

这是 T 细胞受体，或者 B 细胞受体的一段序列

Each numbering scheme segments antibodies into framework regions (FRs) and complementarity-determining regions (CDRs). There are three complementarity-determining regions (CDR1, CDR2, and CDR3) in each receptor, but CDR3 is the most important for antigen recognition.

How CDR3 Contributes to Binding Specificity:

- The unique amino acid sequence of CDR3 allows for the precise fit between the receptor and its specific antigen.
- In T cells, CDR3 interacts with the peptide presented by the MHC molecule. In B cells, it interacts with the surface of the antigen.

Why CDR3 is crucial:

- CDR3 is generated by VDJ recombination, a process that shuffles different gene segments together to create a highly diverse set of TCRs (T cell receptors) and BCRs (B cell receptors).
- As a result, CDR3 has the greatest variability among the three CDR regions. This variability is critical because the antigenic peptides presented by the MHC can be very

diverse. The CDR3 region directly contacts and recognizes these diverse peptides, making it essential for the specific recognition of a wide range of antigens.

- CDR1 and CDR2 (VJ) interact more with the MHC itself, which is more conserved across individuals, while CDR3 recognizes the highly variable antigenic peptide presented by the MHC.

Method to localize the CDR3 ——Sequence Alignment

1. Smith-Waterman algorithm for local alignment or the Needleman-Wunsch algorithm for global alignment.
2. Multiple Sequence Alignment: MSA compares multiple TCR sequences at once to find common patterns and differences, which helps identify the boundaries of the CDR3 region by aligning other conserved regions like framework regions (FR1, FR2, and FR3).

Tools: IMGT and IgBlast

2.8 T Cell and B Cell Epitopes

T Cell Epitopes:

- Presentation: T cell epitopes are short fragments of proteins (peptides) that are presented to T cells by MHC (Major Histocompatibility Complex) molecules on the surface of cells.
- Recognition: T cells recognize these peptide epitopes through their T cell receptors (TCRs), but only when the peptides are bound to an MHC molecule. T cells cannot recognize free-floating antigens.
- Source of Peptides: The peptides presented as T cell epitopes come from processed proteins, either from inside the cell (for MHC Class I and CD8+ T cells) or from extracellular pathogens that have been internalized and processed (for MHC Class II and CD4+ T cells).
- Epitope Length: T cell epitopes are usually short peptides:
  - Class I MHC epitopes: 8-10 amino acids long.

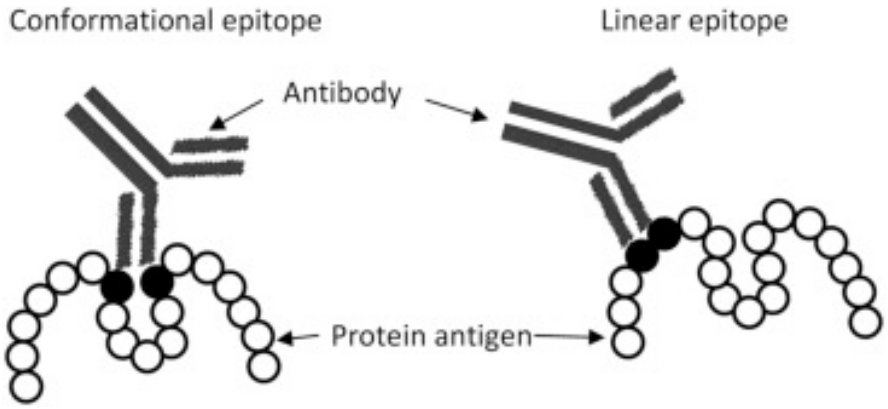

图 3: B Cell Binding

- Class II MHC epitopes: 12-25 (13-18) amino acids long.

B Cell Epitopes:

- Recognition: B cell epitopes are recognized directly by B cell receptors (BCRs) or by antibodies. Unlike T cells, B cells do not require antigen processing and presentation by MHC molecules. B cells can bind directly to antigens in their native, unprocessed form.
- Nature of the Epitope: B cell epitopes can be proteins, carbohydrates, or lipids and are often found on the surface of pathogens. They are recognized by the three-dimensional structure or shape of the antigen, not just by a specific sequence of amino acids.
- Types of B Cell Epitopes:
  - Linear epitopes: These consist of a sequence of amino acids in a continuous stretch (like a T cell epitope).
  - Conformational epitopes: These are made up of amino acids that are not continuous in sequence but come together in the protein's three-dimensional structure.
- Length and Structure: B cell epitopes can be larger and more complex than T cell epitopes, often involving 3D shapes on the surface of proteins.

2.9 Affinity/Stability:

- The affinity of the MHC for a peptide refers to how tightly the peptide binds to the MHC molecule. Strong affinity peptides are typically better at stimulating T cells,

because they stay bound to the MHC molecule for longer, allowing more opportunities for interaction with T cell receptors (TCRs).

- The affinity of the CDR for the peptide describes how strongly the TCR or BCR binds to the peptide-MHC complex、

2.10 TCR and BCR

TCR has α and β chain which have distinct roles and characteristics:

Alpha (α) Chain

- **Gene Segments**: Comprised of V (variable) and J (joining) segments.
- **CDR Regions**: Contains CDR1, CDR2, and CDR3 regions for antigen recognition.
- **Recombination**: Undergoes VJ recombination to generate diversity.

Beta (β) Chain

- **Gene Segments**: Composed of V, D (diversity), and J segments.
- **CDR Regions**: Also contains CDR1, CDR2, and CDR3 regions.
- **Recombination**: Utilizes VDJ recombination, contributing to greater diversity compared to the alpha chain.

BCR (B Cell Receptor) consists of heavy and light chains:

Heavy Chain

- **Gene Segments**: Comprised of V, D, and J segments.
- **CDR Regions**: Contains CDR1, CDR2, and CDR3 regions for antigen binding.
- **Recombination**: Undergoes VDJ recombination to generate diversity.

Light Chain

- **Gene Segments**: Contains V and J segments.

- **CDR Regions**: Includes CDR1, CDR2, and CDR3 regions.
- **Recombination**: Utilizes VJ recombination, similar to the TCR alpha chain.

2.11 Role of BCR and TCR binding

BCR Binding: Involves direct recognition of antigens, often free in bodily fluids. It leads to antibody production.

TCR Binding: Recognizes processed antigens presented by MHC molecules on the surface of other cells.

2.12 Cross-Reactivity

Cross-reactivity refers to the ability of an immune receptor, such as a T-cell receptor (TCR) or antibody, to recognize and bind to multiple different antigens or epitopes that share similar structural features. While this can enhance immune flexibility, it may also lead to unintended immune responses if the receptor binds to non-target antigens.

2.13 Allergic Reactions

Allergic reactions are immune responses to harmless substances, known as allergens. These reactions occur when the immune system mistakenly identifies these substances as threats. Symptoms can include inflammation, itching, and swelling. Management strategies include avoidance, medications, and immunotherapy.

2.14 Specificity of BCR

BCRs are highly specific due to their unique variable regions, which are shaped precisely to recognize specific antigens. This specificity arises from the exact fit between the antigen-binding site and the antigen's structure, minimizing interactions with unrelated molecules.

2.15 Unique Molecular Identifier (UMI)

It is a short, random sequence of nucleotides added to each DNA or RNA molecule during library preparation in sequencing experiments. Following are the reasons why low UMI counts

might indicate weak interactions:

1. Fewer UMI counts suggest that the TCR-peptide pair is present in low quantities, which can imply that there are fewer binding events or interactions.
2. In experiments, strong interactions typically result in higher UMI counts because they occur more frequently.
3. Low counts are more susceptible to noise, making it harder to distinguish from background levels.

2.16 Eluted Ligands (EL), 洗脱配体:

从 MHC 复合体上取下来的相结合的多肽

The term eluted ligands (EL) refers to peptides that are naturally bound to MHC molecules and are released or "eluted" from these molecules for further study.

2.17 Mass spectrometry (MS)，质谱技术：

用于识别所洗脱的多肽是什么

X-ray Crystallography for Epitope Detection

X-ray crystallography is a powerful technique to determine the 3D structure of antibody-antigen (Ab-Ag) complexes, allowing researchers to identify **B-cell epitopes**—the regions on an antigen that are recognized and bound by antibodies.

Process of Using X-ray Crystallography to Detect Epitopes

The process involves several key steps:

1. **Crystallization of the Antibody-Antigen Complex**: The antibody and antigen are co-crystallized to form an ordered crystal lattice. Crystallization is often the most challenging step due to the specific conditions required for stable complex formation.

2. **X-ray Diffraction**: The resulting crystal is exposed to a focused X-ray beam. When X-rays interact with the atoms in the crystal, they are scattered, creating a diffraction pattern that is recorded on a detector.

3. **Data Analysis and Structure Solution**: The diffraction pattern is used to construct an **electron density map**, a 3D representation of the electron distribution in the complex. This map enables researchers to identify the positions of individual atoms within the antigen-antibody complex with atomic-level accuracy.

4. **Identifying the Epitope on the Antigen**: By analyzing the 3D structure of the antibody-antigen complex, researchers can determine the specific amino acid residues on the antigen that are in close contact with the antibody. These contact residues define the **B-cell epitope**, as they are the exact parts of the antigen recognized by the antibody.

Mass spectrometry (MS) is used to identify these peptides by measuring their mass and charge, which helps in understanding immune recognition.

3 High-Throughput Sequencing

3.1 Unique Molecular Identifier (UMI)

3.2 PCR