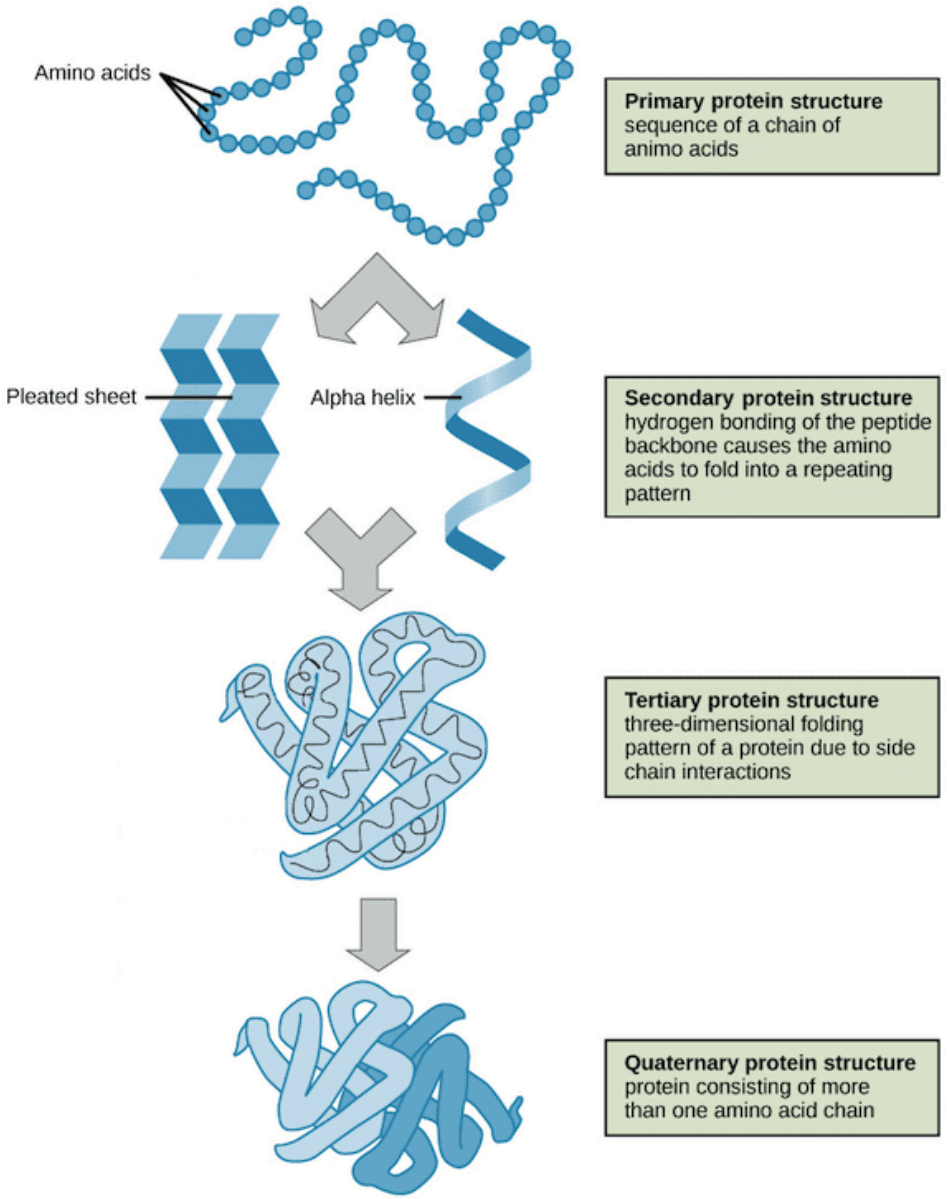
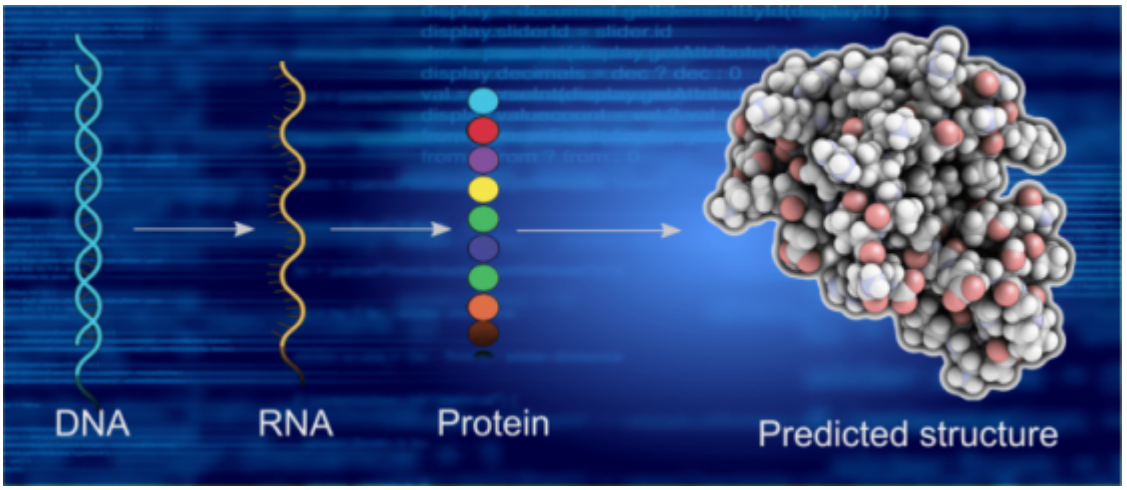

Amino acids
Pleated sheet
Alpha helix
Primary protein structure
sequence of a chain of animo acids
Secondary protein structure
hydrogen bonding of the peptide backbone causes the amino acids to fold into a repeating pattern
Tertiary protein structure
three-dimensional folding pattern of a protein due to side chain interactions
Quaternary protein structure
protein consisting of more than one amino acid chain


DNA
RNA
Protein
Predicted structure

Protein Extraction Methods:

Proteins are essential macromolecules that perform a variety of functions in the body, like DNA replication, catalyzing reactions, and providing structural support. They are studied in three main ways:

1. **In Vivo**: Studying proteins within the organism to understand how they interact.

2. **In Vitro**: Studying purified proteins in controlled lab settings to avoid interference from other factors.

3. **In Silico**: Using computer simulations to study proteins, saving time and resources.

Proteins are classified into types like **extracellular matrix proteins** (e.g., elastin, collagen) and **globular proteins** (e.g., enzymes, antibodies). Purifying proteins from other cellular components is crucial for research, whether for large-scale production (e.g., insulin) or analysis of small protein amounts.

Uses of Isolated Proteins:

Protein extraction is widely applied in both research and industry. The purification process requires multiple steps and detection methods, including absorbance, spectrometry, and antibody-based techniques.

In **clinical applications**, isolated proteins can help diagnose diseases like diabetes or be used in treatments (e.g., collagen in skincare). In **research**, purified proteins enable various studies:

- **Immunoprecipitation (IP)**: Isolates proteins using antibodies.

- **Proteomics**: Studies the entire set of proteins in an organism.

- **Enzyme Assays**: Measure enzyme activity, including different experiment types like relaxation or transient kinetics.

- **Western Blot**: Detects specific proteins in a sample.

- **Gel Electrophoresis**: Separates proteins by size and charge.

- **Biomarkers**: Used to track biological processes like disease or treatment effects.

These techniques are crucial for advancing both clinical diagnostics and scientific understanding.

What is Protein Sequencing?
Protein sequencing is a technique used to determine the exact order of amino acids in a protein. This sequence is known as the protein's primary structure, and understanding it helps us know how a protein functions in the body.

Significance of Protein Sequencing:

1. Decoding Genetic Information: It translates DNA's genetic code into proteins, revealing the exact sequence of amino acids in a protein.

2. Unveiling Protein Function: The sequence affects how a protein folds and interacts with other molecules, helping scientists understand its role.

3. Biotechnology: It aids in designing drugs, enzymes, and other proteins for medical and industrial uses.

4. Personalized Medicine: By sequencing proteins, doctors can identify genetic mutations linked to diseases and create personalized treatments.

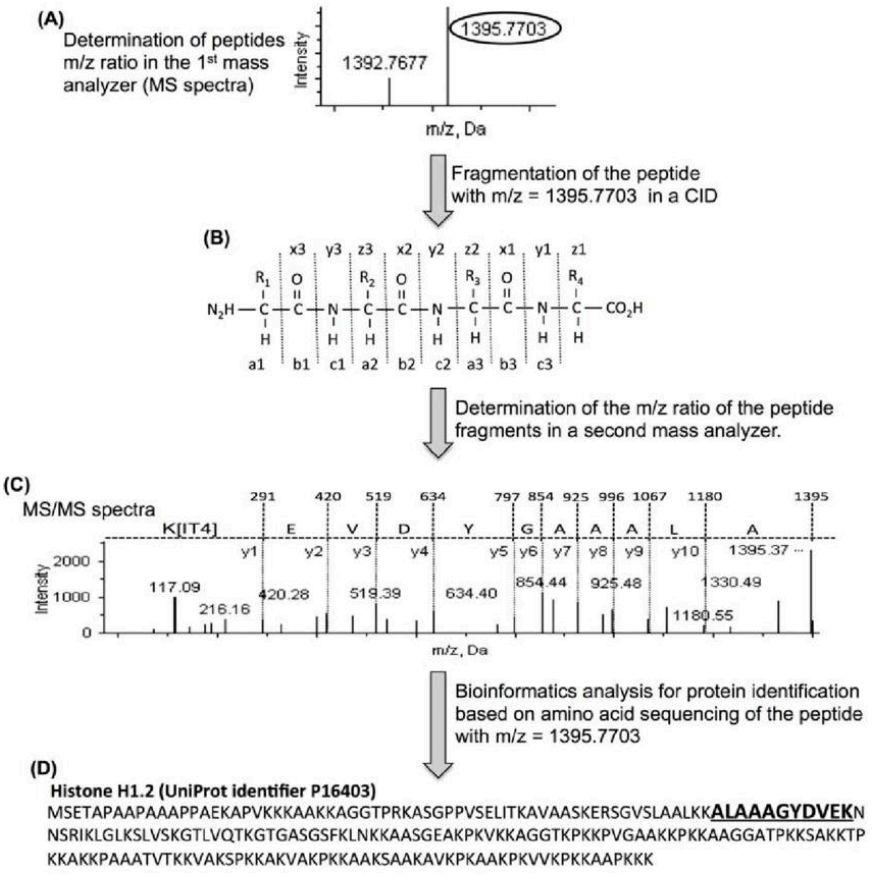
5. Structural Biology: Protein sequencing is essential for understanding the three-dimensional shape of proteins, which is key to drug design.

6. Proteomics Advancements: It helps in studying all proteins in an organism, uncovering complex biological processes.

Methods of Protein Sequencing:

- Edman Degradation: This older method removes and identifies amino acids from the protein's starting point, but requires large samples.

- Mass Spectrometry: It analyzes protein fragments to determine their sequence and is useful for complex mixtures. Mass Spectrometry (MS) measures the mass-to-charge ratio of ions to identify and quantify molecules.



- Next-Generation Sequencing (NGS): This modern technique uses mRNA to indirectly infer protein sequences, offering high throughput.

Applications of Protein Sequencing:

- Drug Development: Helps create targeted therapies by identifying protein structures linked to diseases.
- Structural Biology: Reveals a protein's 3D shape, crucial for designing drugs that target specific protein structures.
- Proteomics Research: Identifies proteins in biological samples, providing insights into diseases like cancer and neurological disorders.
- Biotechnology: Protein sequencing is key in designing and producing therapeutic proteins and enzymes.
- Personalized Medicine: It helps tailor treatments by identifying genetic variations in proteins.

Challenges and Solutions:

- Sample Preparation: Extracting and purifying proteins can be difficult, but new techniques are making it easier.
- Data Analysis: The large amount of data requires advanced software to interpret protein sequences accurately.
- Cost: Protein sequencing can be expensive, but efforts are ongoing to reduce costs and improve efficiency.

Technological Advancements:

- High-Throughput Sequencing: Allows simultaneous analysis of many proteins, speeding up research.
- Mass Spectrometry Improvements: Enhanced sensitivity helps detect low-abundance proteins and modifications.
- Hybrid Approaches: Combining different sequencing methods improves accuracy and efficiency.

Bioinformatics:
Advanced software tools help analyze sequencing data, identify proteins, and even predict how proteins fold and function, contributing to more precise results in proteomics research.

In summary, protein sequencing is essential for understanding the role of proteins in biology, advancing medicine, and developing biotechnologies.

RCSB Protein Data Bank (PDB):

The RCSB Protein Data Bank (PDB) is a key resource for scientific research and education, providing detailed 3D structures of macromolecules (like proteins and nucleic acids) and small molecules (such as drugs and cofactors). Researchers, educators, and students use it to explore how biological molecules interact, which is crucial in fields like molecular biology, medicine, and biotechnology. RCSB in RCSB PDB stands for Research Collaboratory for Structural Bioinformatics. It is the US data center for the global Protein Data Bank (PDB) archive, providing access to 3D structures of biological macromolecules.

The PDB was established in 1971 and has grown to include over 113,000 entries. It is used by a wide range of people, from scientists to students. The PDB archive includes data from experiments like X-ray crystallography, nuclear magnetic resonance (NMR), and 3D electron microscopy (3DEM). The information is carefully annotated and validated to ensure consistent and reliable data for research and education.

- The PDB supports the Worldwide Protein Data Bank (wwPDB), a network of data centers that collaborate on data deposition, annotation, and validation. RCSB PDB is based at Rutgers University and the University of California, San Diego.
- Tools for searching, visualizing, and analyzing data are available on the RCSB PDB website, which helps users explore molecular interactions and the functions of biological molecules.
- It provides a wealth of free educational resources for teachers and students to explore molecular chemistry and biology.

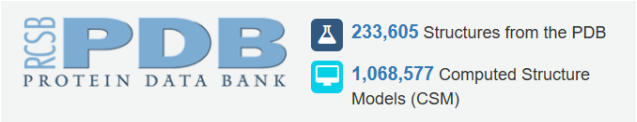
Data Annotation and Validation:

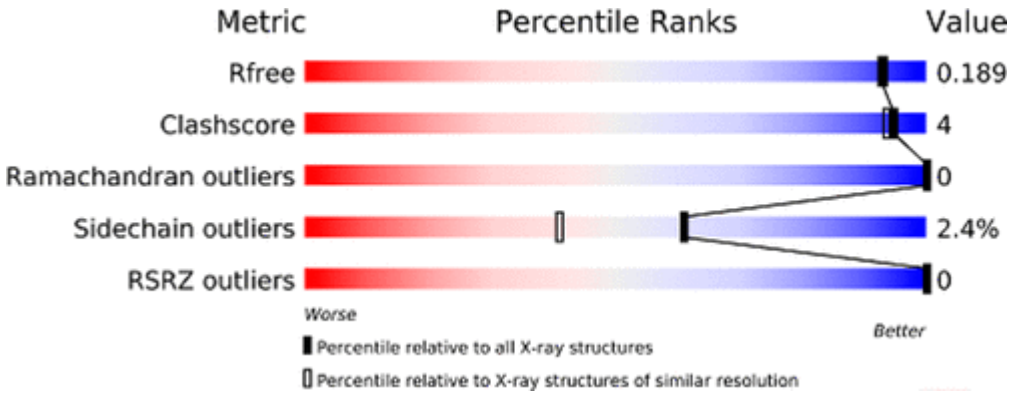
When researchers submit a new structure, it is reviewed by biocurators, who check the data for accuracy and consistency. The data is annotated with detailed experimental information and cross-referenced with other resources. This includes ensuring uniform representation of all molecules, such as proteins and small molecules. Validation reports assess the quality of structures and highlight areas that need attention.

The PDB uses standardized data dictionaries for organizing and accessing information, such as:

- PDB Exchange (PDBx)/mmCIF: Describes crystallographic, NMR, and 3DEM data.
- Chemical Component Dictionary (CCD): Describes small molecules, including their properties and structure.
- Biologically Interesting molecule Reference Dictionary (BIRD): Includes molecules with unique biological or pharmaceutical functions.

This detailed and consistent data curation makes it easier for users to search, analyze, and understand the structural data for various biological molecules.





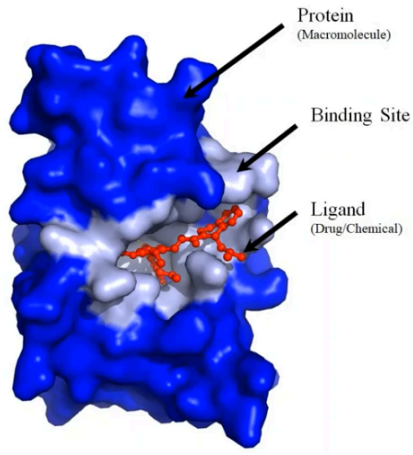
The validation report slider graphic shows how the quality of a PDB entry compares to others. In this example, the slider for entry 1cbs (a small protein with a ligand) indicates better overall quality compared to other X-ray structures, with a resolution of 1.8 Å.

https://www.rcsb.org/

Ligands in Coordination Chemistry

A metal ion in solution usually combines with **ligands—molecules or ions like solvent molecules or simple ions—which form complex ions or coordination compounds.** These complexes have a central metal atom (often a transition metal) surrounded by ions or neutral molecules. Ligands are ions or molecules that bond to the metal atom or ion. They act as Lewis bases (electron pair donors), while the central metal acts as a Lewis acid (electron pair acceptor). Ligands have at least one atom that donates an electron pair to bond with the metal.

The word "ligand" comes from the Latin "ligare," meaning "to bind," and was first used by Alfred Stock in 1916 in relation to silicon chemistry. **Ligands can be anions, cations, or neutral molecules, and are classified by the number of donor atoms**: **monodentate** (one donor atom), **bidentate** (two donor atoms), and **polydentate** (multiple donor atoms).



https://www.youtube.com/watch?v=v-RmqwGRCa4

Molecular recognition is how biological molecules, like proteins, specifically and strongly bind to other molecules (including small ones) to form a stable complex. This process is essential for all life activities—from DNA replication to metabolism and cell signaling.

Proteins are one of the most important molecules in living systems. They do their jobs—such as acting as enzymes, building structures, or sending signals—by binding to other molecules. These other molecules, called *ligands*, can be anything from another protein to a small molecule like oxygen or a drug.

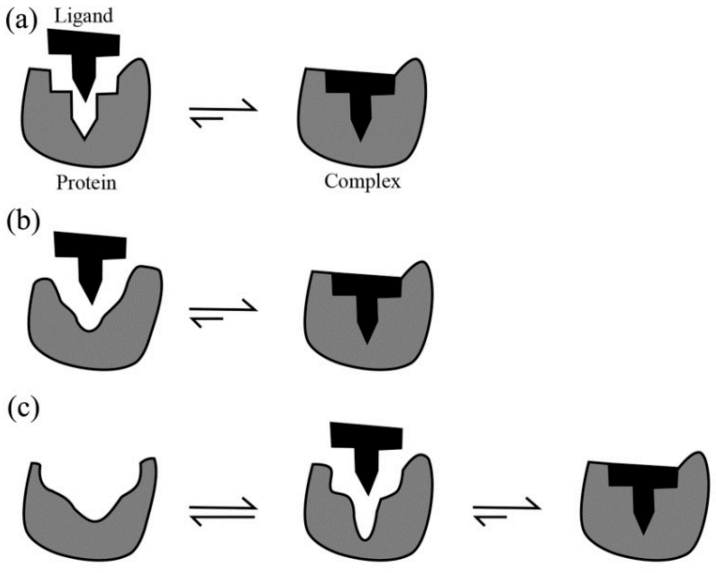
To truly understand how proteins work, we need to understand how they interact with ligands. This includes knowing what forces drive binding, how fast it happens (kinetics), and how favorable it is (thermodynamics). This knowledge is also crucial for designing new drugs that can bind specifically and strongly to a target protein.

This review is divided into three parts:

1. **Mechanisms Behind Protein–Ligand Binding:**
   It explains the basics of how proteins bind to ligands, including the concepts of binding speed (kinetics), energy changes (thermodynamics), and the forces involved. It also covers the idea of *enthalpy–entropy compensation*, where gains in one type of energy are balanced by losses in another.

2. **Binding Models:**
   Three models explain how proteins bind to ligands:

   - **Lock-and-Key**: The protein and ligand fit perfectly from the start.
   - **Induced Fit**: The protein changes shape to fit the ligand.
   - **Conformational Selection**: The protein exists in many shapes, and the ligand binds to the one that already fits.



3. **Techniques to Study Binding:**
   This includes both experimental and computer-based methods:

   - **Experimental Methods**:

        - *Isothermal Titration Calorimetry (ITC)*
        - *Surface Plasmon Resonance (SPR)*
        - *Fluorescence Polarization (FP)*
          These help measure how strong and fast the binding is.
    - **Computational Methods**:
        - *Docking*: Predicts how a ligand fits into a protein. It's fast and used in drug discovery.
        - *Free Energy Calculations*: More accurate but slower, based on physics and extensive simulations.

🔥 Enthalpy (ΔH) – Think: Energy of Interactions

Enthalpy is the total heat energy in a system. In binding, it mostly comes from how strongly molecules attract each other.

- **When molecules bind**, they often form **noncovalent interactions** like:
    - Hydrogen bonds
    - Van der Waals forces (weak interactions)
    - Ionic or dipole interactions

If forming these interactions **releases energy**, the process has **negative enthalpy** ($\Delta H < 0$), which **favors binding**.

Simple analogy:
Imagine two magnets snapping together — that releases energy. That's a **favorable enthalpy** change.

🌪️ Entropy (ΔS) – Think: Disorder or Freedom

Entropy measures how **disordered** or **spread out** energy is in a system. More freedom = more entropy.

In binding:

- Molecules (like proteins and ligands) are **free to move** when unbound (high entropy).

- When they bind, they **lose freedom** — rotation and movement are restricted (lower entropy).

- BUT, **when water molecules are released** during binding, **they gain freedom**, which **increases entropy**.

So, entropy is a trade-off:

- ❌ Binding reduces movement of the protein & ligand → **bad for entropy**

- ✅ But releases trapped water → **good for entropy**

Simple analogy:
You tie two people together. They can't move as freely (entropy goes down), but if you let a crowd of people go free at the same time (like water molecules), overall disorder increases (entropy goes up).

🧠 In Binding: It's a Balance

Binding is favorable when the overall **free energy (ΔG)** is negative:

$\Delta G = \Delta H - T\Delta S$

- **ΔH** = Enthalpy (energy from bonds)

- **TΔS** = Temperature × Entropy (freedom/disorder)

There are **two ways binding can be favorable**:

1. **Enthalpy-driven**: Strong new interactions give off energy (big negative ΔH)

2. **Entropy-driven**: Lots of water is released → system becomes more disordered (big positive ΔS)

Example: When protein and ligand meet, they displace water molecules around them. This increases the freedom of water (solvent entropy) and favors binding. Although breaking water's hydrogen bonds requires energy (positive enthalpy), this is compensated by the entropy gain. For the lock-and-key model, the entropy gain must be large enough to also offset the loss of motion when the ligand binds. So while enthalpy (from van der Waals or hydrogen bonds) contributes, entropy gain—mainly from water release—dominates the binding.

Monodentate Ligands

Monodentate ligands bind to the metal ion through one donor atom. Examples include chloride ions (chloro), water (aqua), hydroxide ions (hydroxo), and ammonia (ammine).

Bidentate Ligands

Bidentate ligands have two donor atoms, allowing them to bind to the metal ion at two points. An example is **ethylenediamine (en)**, which has nitrogen atoms that can bond to the metal.

Polydentate Ligands

Polydentate ligands have multiple donor atoms, such as **EDTA**, which has six donor atoms that can bind to a metal ion. These ligands form more stable complexes.

Ambidentate Ligands

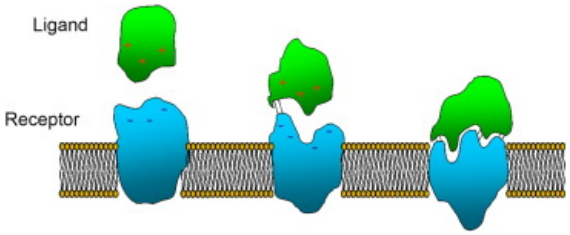
Ambidentate ligands can bind to the central atom in two places. An example is **thiocyanate (SCN−)**, which can attach at either the sulfur or nitrogen atom.

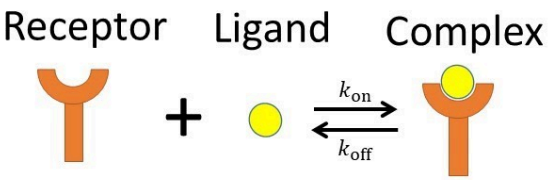
Chelation

Chelation occurs when a polydentate ligand bonds to a metal ion, forming a ring. This complex is called a **chelate**, and the ligand is called a **chelating agent**. Chelating agents have a high affinity for metal ions compared to monodentate ligands. Examples of chelating agents are **ethylenediamine** and **EDTA**. The term "chelate" comes from the Greek word "chela," meaning claw, describing the way the ligand "grips" the metal like a lobster's claw.

Diseases and Tools in Molecular Docking for Disease Research

Molecular docking has been applied to various diseases to identify the molecular targets of nutraceuticals.Molecular docking is a computer-based method used to predict how **small molecules (ligands) interact with proteins (receptors).** It helps researchers identify potential drug candidates by simulating their binding interactions/ binding kinetics.





$$\frac{dR}{dt} = -k_{on}RL + k_{off}C$$

$$\frac{dL}{dt} = -k_{on}RL + k_{off}C$$

$$\frac{dC}{dt} = k_{on}RL - k_{off}C$$

Key Diseases & Natural Compounds (Nutraceuticals)

1️⃣ Sickle Cell Disease

- **Compounds:** Polyphenols (quercetin, curcumin)
- **Target Proteins:** Hemoglobin variants, adhesion molecules
- **Goal:** Reduce oxidative stress & prevent sickling of red blood cells

2️⃣ Cancer

- **Compounds:** Resveratrol, curcumin, EGCG, genistein
- **Target Proteins:** p53, EGFR, Bcl-2, COX-2, VEGF
- **Goal:** Trigger cancer cell death, inhibit tumor growth & spread

3️⃣ Cardiovascular Diseases

- **Compounds:** Omega-3s, flavonoids, polyphenols
- **Target Proteins:** ACE, eNOS, LDL receptors
- **Goal:** Lower blood pressure, cholesterol & inflammation

4️⃣ Gut Disorders (IBD, Ulcers, Microbiome Imbalance)

- **Compounds:** Probiotics, prebiotics, polyphenols
- **Target Proteins:** NF-κB, COX-2, gut enzymes
- **Goal:** Reduce gut inflammation & improve microbiome balance

5️⃣ Neurodegenerative Diseases (Alzheimer's, Parkinson's)

- **Compounds:** Curcumin, resveratrol, omega-3s
- **Target Proteins:** Beta-amyloid, tau, AChE, alpha-synuclein
- **Goal:** Slow neurodegeneration by reducing oxidative stress & inflammation

6️⃣ Reproductive Health Disorders

- **Compounds:** Phytoestrogens, antioxidants (vitamin E, flavonoids)
- **Target Proteins:** Estrogen & androgen receptors, FSH receptors
- **Goal:** Balance hormones & improve fertility

Virtual Screening in Drug Discovery 🧪🔬

A **computational method** to quickly find drug candidates, complementing lab testing.

Types of Virtual Screening:

1. **Ligand-Based Screening** (Uses known active compounds to find similar ones)
    - **Methods:**
        - Similarity & substructure searching
        - Pharmacophore matching (identifies key binding features)
        - 3D shape matching
2. **Structure-Based Screening** (Uses known protein structures to predict binding)
    - **Uses molecular docking** to rank potential drugs based on their predicted binding strength.

Databases Used:

✅ Biomolecular (genes, proteins)
✅ Organic molecules
✅ Biological/disease targets
✅ Natural products

Molecular Docking Process

1️⃣ **Ligand Sampling:** Tests different shapes & positions in the protein's binding site.
2️⃣ **Scoring Function:** Ranks ligand poses based on binding strength & stability.

Search Algorithms for Docking

🔹 **Systematic (Direct) Methods** (step-by-step adjustments)

- **Conformational Search:** Gradual shape/rotation changes
- **Fragmentation:** Docks small parts separately & rebuilds (e.g., FlexX, DOCK, LUDI)
- **Database Search:** Screens large libraries for matches (e.g., FLOG)

🔹 **Stochastic (Random) Methods** (randomized optimization)

- **Monte Carlo:** Random ligand placements & refinements (e.g., MCDOCK, ICM)

- **Genetic Algorithm:** Evolutionary selection for best binding poses (e.g., AutoDock, GOLD)
- **Tabu Search:** Avoids repeating past solutions to explore new ones (e.g., Molegro Virtual Docker)

Models to choose! QSAR (Quantitative Structure-Activity Relationship)

📌 **Introduced in 1964** as an early drug design method.

✅ **Goal:** Find patterns between a molecule's structure & biological activity.
✅ **Useful when:** Receptor data is limited, but activity data is available for many compounds.
✅ **Application:** Helps identify potential drug compounds by analyzing chemical properties.

🚀 Key Takeaway:

Virtual screening & QSAR accelerate drug discovery, reducing costly trial-and-error lab testing! 🔬💊

QSAR, Monte Carlo, and Machine Learning in Drug Discovery 🚀🔬

QSAR (Quantitative Structure-Activity Relationship) models predict a drug's biological activity based on its chemical structure.

🔹 **Monte Carlo Optimization/prediction** improves QSAR models by randomly adjusting molecular **descriptors** (numerical values representing binding interactions like size, charge, hydrophobicity).
🔹 **Combining QSAR & Docking** helps identify promising drug candidates faster.

QSAR & Machine Learning in Drug Discovery

✅ **QSAR models** help predict drug effectiveness and reduce lab testing.
✅ **Challenges:** Large datasets, complex descriptors, and data errors.

🔹 **Monte Carlo in Docking:**

- **Random Sampling:** Tests different ligand positions in a protein.
- **Metropolis Algorithm:** Accepts better positions based on energy and sometimes worse ones to avoid bad fits.
- **Simulated Annealing:** Gradually refines ligand positions for stability.

🔹 **Scoring Function:** Ranks ligand positions—lower energy = better binding.

Applications of Monte Carlo in Drug Discovery:
✅ **Lead Optimization:** Improves drug molecules.

✅ **Database Screening:** Finds potential drugs in large libraries.
✅ **QSAR Modeling:** Enhances prediction accuracy.

Why Use Machine Learning in QSAR? 🤖

🔹 **Random Forest (RF):** A popular algorithm for QSAR due to its accuracy, simplicity, and ability to handle noisy data.
🔹 **Neural Networks:** Used in drug discovery (e.g., Merck Kaggle competition).
🔹 **Ensemble Learning:** Combining multiple models improves accuracy.

QSAR Model Development Steps:

1️⃣ Select training data (known compounds).
2️⃣ Input biological activity data.
3️⃣ Generate molecular conformations.
4️⃣ Calculate descriptors (size, charge, hydrophobicity).
5️⃣ Choose a statistical method (regression, machine learning).
6️⃣ Develop QSAR equation.
7️⃣ Validate model.
8️⃣ Use a model to predict unknown compounds.

✅ **Outcome:** Faster, cheaper, and more efficient drug discovery!

How Random Forest Works in QSAR

✅ **Multiple Decision Trees:** RF builds many decision trees and averages their predictions.
✅ **Random Sampling & Bootstrapping:** Improves generalization.

🔹 **Applications:**

- Predicting drug activity (binding strength).
- Assessing toxicity (environmental safety).

🔹 **Key Takeaway:** RF is **powerful, reliable, and widely used** in drug discovery! 🌱💊

Improving QSAR with Advanced Machine Learning

🔹 **Limitations of Existing QSAR Models:**

- Most focus on only one aspect (data sampling, method variety, input representation).
- This **reduces prediction accuracy**.

🔹 **New Approach: Multi-Subject Ensemble Learning**

- Uses **1D-CNNs & RNNs** to extract features from **SMILES notation** (chemical structure representation).
- Implements **meta-learning** to combine the best models.
- The **combined model outperforms individual models**, making QSAR predictions more accurate. Comprehensive Ensemble in QSAR Prediction for Drug Discovery.

🔹 **Try it here:** http://data.snu.ac.kr/QSAR/

✅ **Benefits:** More reliable drug predictions, better accuracy, and faster discovery. 🚀

Scoring Functions: Ranking Ligands in Docking

Scoring functions measure how well a ligand binds to a protein.

🔹 **Types:**
1️⃣ **Force Field-Based:** Uses physics-based calculations (e.g., AutoDock, DOCK, GoldScore).
2️⃣ **Empirical:** Predicts binding using real protein-ligand data (e.g., ChemScore).
3️⃣ **Knowledge-Based:** Uses statistical data (e.g., PMF, DrugScore).
4️⃣ **Consensus:** Combines multiple scoring methods for better accuracy.

Popular Molecular Docking Software

✅ **AutoDock & AutoDock Vina** – Free, widely used for flexible docking.
✅ **DOCK** – Grid-based scoring, developed by UCSF.
✅ **GOLD** – Uses genetic algorithms for ligand docking.
✅ **MolDock** – Fast docking via **Fourier Transform**.
✅ **Discovery Studio** – Commercial software with various docking tools.
✅ **Chimera** – Visualization & docking analysis.

🔬 Bottom Line: Accelerating Drug Discovery

✅ **QSAR + Machine Learning + Monte Carlo** = Faster, more accurate drug discovery! 🚀💊

Simplified Guide: Molecular Docking🚀🔬

Key Steps in Molecular Docking

1️⃣Retrieving Protein & Ligand Structures

- Proteins come from databases like **PDB** or are modeled using **Swiss-Model**.
- Ligands are sourced from **PubChem, ZINC, or ChemDraw**.

2️⃣ Protein Preparation

- Clean the protein by removing unwanted molecules and optimizing its structure.

3️⃣ Lead Identification

- Select ligands with potential biological activity.

4️⃣ Active Site Prediction

- **Site-Directed Docking:** Targets a known binding site.
- **Blind Docking:** Searches for possible binding sites.
- **Docking with a Standard:** Compares with a reference molecule.

5️⃣ Protein-Ligand Docking

- The ligand is positioned into the protein, and interactions are analyzed.

6️⃣ Post-Docking Analysis

- The best molecules are ranked based on **binding affinity and interactions** (e.g., hydrogen bonds, hydrophobic forces).

Why is Molecular Docking Important?

✔️ Screens potential drug molecules efficiently.
✔️ Predicts how well a drug binds to a target protein.
✔️ Supports drug discovery and development.

Applications of Molecular Docking

✅ Hit Identification & Virtual Screening

- Finds drug candidates by screening large compound libraries.

✅ Lead Optimization

- Modifies drug molecules for better binding and selectivity.

✅ Bioremediation

- Helps design molecules that assist in environmental cleanup.

✅ ADMET Prediction

- Evaluates **Absorption, Distribution, Metabolism, Excretion, and Toxicity (ADMET)**.
- Tools: **AutoDock Vina, GOLD, Glide, Schrödinger Suite**.

✅ **Molecular Dynamics Simulation**

- Studies ligand behavior over time to ensure **binding stability**.

✅ **Structure Elucidation**

- Predicts protein structures and binding sites for unknown proteins.

Nutraceuticals: Food-Based Health Solutions

🔹 **What are Nutraceuticals?**

- A combination of **"nutrition" and "pharmaceutical"**.
- Food-derived compounds with **health benefits**.
- Used in **disease prevention and treatment**.

🔹 **Types of Nutraceuticals**
1️⃣ **Conventional:** Vitamins, minerals, herbal extracts, probiotics, enzymes.
2️⃣ **Non-Conventional:** Algae, fungi, animal by-products (e.g., spirulina, mushroom extracts).
3️⃣ **Fortified Foods:** Vitamin D-fortified milk, iron-enriched cereals.
4️⃣ **Recombinant Nutraceuticals:** Genetically engineered molecules like **recombinant insulin and vitamin B12**.

🔬 Protein-Ligand Docking Overview

A computational technique used to predict how a small molecule (ligand) binds to a protein's active site, crucial in **drug discovery** and **structural bioinformatics**.

🎯 Key Objectives

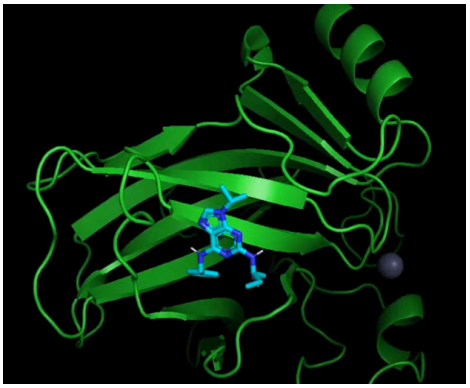
- 🔗 **Predict Binding Mode**
  Determines the most likely orientation and position of a ligand within a protein's binding site.
- 💊 **Drug Discovery**
  Simulates interactions to identify promising drug candidates before synthesis and testing.

- 🧬 **Understand Protein Function**
  Reveals how proteins interact with ligands and perform biological roles.

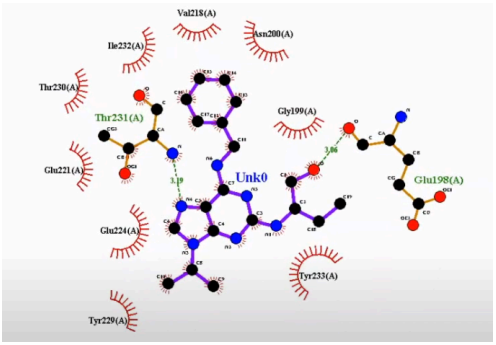
- 📚 **Virtual Screening**
  Screens large compound libraries computationally to find high-affinity binders to target proteins.

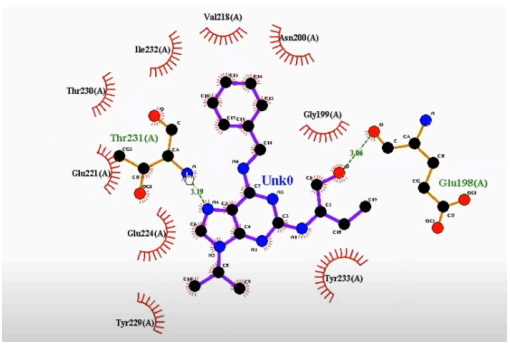
⚙️ How Docking Works

1. 📥 **Input**
    - 3D structure of the protein (from X-ray, NMR, or cryo-EM)
    - Structure of the ligand (drug-like molecule)



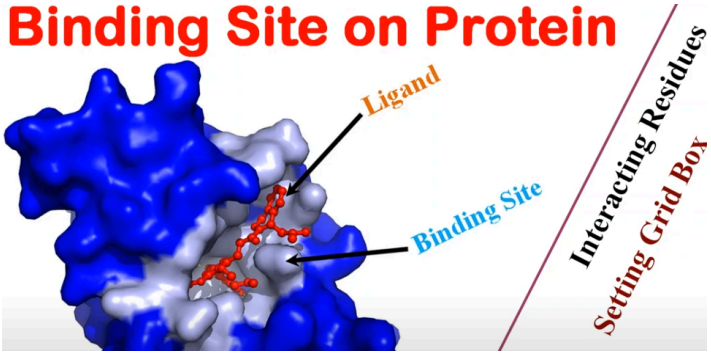
2. 🔍 **Search Algorithm**
    - Explores different ligand poses within the protein's binding site
    - Considers translations, rotations, and sometimes internal flexibility





3. 📊 **Scoring Function**
    - Evaluates each pose based on interaction energy (e.g., hydrogen bonds, hydrophobic effects, van der Waals forces)

4. 🏆 **Pose Selection**
    - Identifies top-scoring binding conformations as the most probable



5. 🧠 **Analysis**
    - Examines key molecular interactions for biological insights or drug optimization

---

🧱 Types of Docking

- 🧊 **Rigid Docking**
  Assumes fixed structures for both protein and ligand — faster but less flexible.
- 🌀 **Flexible Docking**
  Allows conformational changes in the ligand for more realistic predictions.
- ⚙️ **Protein Flexibility**
  Advanced methods account for movements in the protein's binding site, increasing accuracy.

📌 Key Considerations

- ✅ **Accuracy**
  Depends on input quality, algorithm sophistication, and scoring function reliability.
- 💻 **Computational Cost**
  Increases with system size and flexibility; high-throughput docking requires optimization.
- 📊 **Virtual Screening Use**
  Enables the evaluation of thousands to millions of compounds for potential activity.

Trends in the Nutraceutical Industry

🌿 **Cannabis-Based Products** – Increasing acceptance.
💄 **Nutricosmetics** – Supplements for **skin, hair, and nails**.
♻️ **Sustainable Packaging** – Eco-friendly materials.
💪 **Sports Nutrition** – Growth in fitness supplements.
🐾 **Pet Nutrition** – High-quality food for animals.
🛍️ **Online Sales Growth** – Direct-to-consumer sales booming.
🌱 **Seed Oils** – Flaxseed & chia oil in supplements.
🚫 **Endocrine Protection** – Counteracts harmful chemicals like **BPA**.

Molecular Docking in Nutraceutical Research

🔬 **How It Helps:**
✔️ Identifies how nutraceuticals interact with disease-related proteins.
✔️ Reduces animal testing by using computational models.

🔹 **Common Targets in Nutraceutical Research**

Target	Example	Effect
Enzymes	Curcumin → **COX-2**	Reduces inflammation
Receptors	Resveratrol → **Sirtuin**	Linked to aging & longevity
RNA/DNA	Quercetin → **Topoisomerase**	Inhibits DNA replication
Epigenetic Markers	HDACs, DNMTs	Regulates gene expression
Other Proteins	Hesperidin → **Alpha-glucosidase**	Helps carb digestion

✅ **Why It Matters:** Molecular docking speeds up nutraceutical research, helping discover and validate health benefits efficiently! 🚀

Using AutoDock Vina with Chimera

🔹 **AutoDock Vina** is a docking tool that predicts how ligands (small molecules) bind to proteins. AutoDock Vina (in **UCSF [**University of California, San Francisco] Chimera) is one of the computationally fastest and most accurate software employed in docking.

🚨 **Important:**

- The **AutoDock Vina web service** was discontinued on **April 30, 2020**.
- To use it, you **must install AutoDock Vina locally** and set its executable path in Chimera.
- Chimera **only supports docking one ligand at a time** with limited sampling.
- For large-scale docking, **use AutoDock Vina directly** instead of Chimera.
- Chimera can still **visualize** docking results.

How to Use AutoDock Vina in Chimera

1️⃣ Prepare Structures

- **Receptor (Protein)** → Download from **PDB** or prepare using Chimera.
- **Ligand** → Obtain from **PubChem, ZINC, or ChemDraw**.
- **Run Dock Prep** in Chimera to:
  - Fix missing atoms.
  - Add hydrogens (for H-bonding calculations).
  - Remove unnecessary molecules.

2️⃣ Define Docking Parameters

- **Select Receptor & Ligand** in Chimera's AutoDock Vina tool.
- **Set Search Box** (where docking occurs):
  - Adjust manually or enter **X, Y, Z coordinates**.
- **Advanced Options:**
  - **Binding Modes**: Up to **10** (default: **9**).
  - **Search Exhaustiveness**: 1 (fast) to 8 (slow, accurate).
  - **Energy Difference**: Filters out weak binding poses (default: **3 kcal/mol**).

3️⃣ Run Docking

- Click **OK** (runs & closes the window) or **Apply** (runs but keeps the window open).

- Docking runs as a **background task**—you can cancel it if needed.

4️⃣ View Results

- Docked molecules open in **ViewDock** for analysis.
- Output files:
    - **name.pdbqt** → Docking results.
    - **name.receptor.pdbqt** → Processed receptor.
    - **name.ligand.pdbqt** → Processed ligand.
    - **name.conf** → Configuration file.

Limitations & Workarounds

🚧 Peptides & Multi-Residue Ligands Appear Fragmented

- AutoDock Vina **changes atom order**, which **scrambles peptides**.
- Fix: **Delete & re-add bonds** in Chimera.

🚧 Ring Structures Can Break

- Some ring bonds may be mistakenly marked as rotatable, leading to **incorrect structures**.
- **Fix:** Use AutoDock Vina directly instead of Chimera.

Key Takeaways

✅ AutoDock Vina helps predict ligand-protein interactions.
✅ Chimera supports docking **one ligand at a time** with **basic settings**.
✅ For **large-scale docking**, use **AutoDock Vina directly**.
✅ **Check structures carefully** after docking to avoid errors.

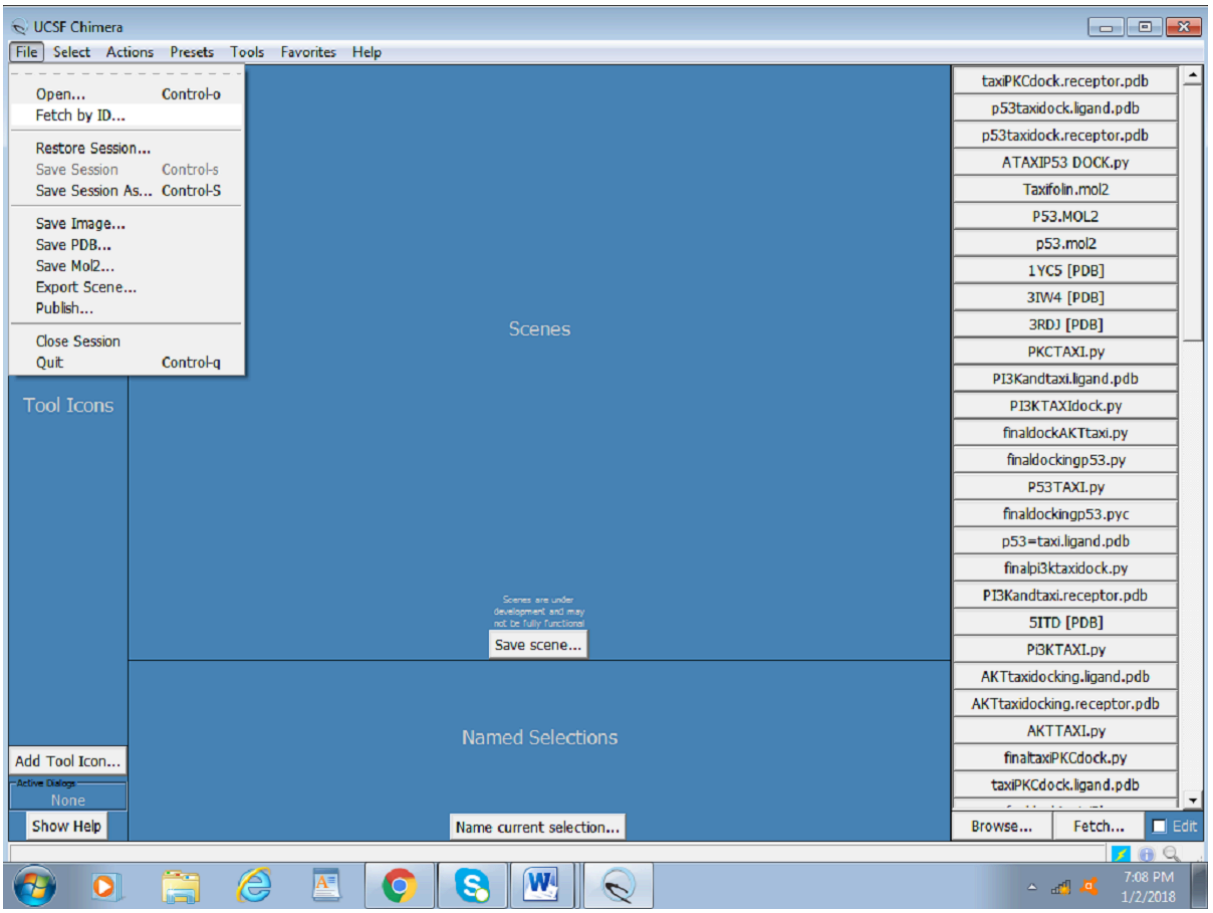
🚀 **Ready to dock?** Install AutoDock Vina locally and start predicting drug interactions! 🎯

Simplified Guide: Molecular Docking Using UCSF Chimera & AutoDock Vina

---

Requirements

✔ **Operating System:** Windows 7, 8, 10, Mac, or Linux
✔ **Software:** UCSF Chimera 1.12 https://bioinform.jmir.org/2020/1/e14232



Step-by-Step Docking Procedure

1️⃣ Retrieve the Target Protein

- Get the protein **PDB file** from the **Protein Data Bank (PDB)**.
- In **Chimera**, go to **File > Fetch by ID**, enter the **PDB ID** (e.g., **3QKK** for Akt).
- Ensure an **internet connection** or **manually download** the PDB file.
- Save the file in a **working directory** (e.g., `Users/Desktop/Docking/`) as **Akt.pdb**.

2️⃣ Prepare the Target Protein

1️⃣ Identify the Active Site:

- Select the inhibitor: **Select > Residue > [Inhibitor Name]** (e.g., SMH).
- Highlight it in a different **color** for visibility (**Actions > Color > Red**).

2️⃣ Optimize the Protein:

- Open **Dock Prep**: **Tools > Structure Editing > Dock Prep**.
- In the **Dock Prep box**, select all options **except** "Delete non-complexed ions" and click **OK**.
- **Add Hydrogens** (Tools > Structure Editing > AddH).
- **Assign Charges** (Use **Gasteiger charges**). Gasteiger charges are a method of assigning partial charges to atoms in a molecule, used in computational chemistry and molecular modeling. They are based on a model of the molecule's electrostatic potential and are calculated by summing over atomic electron densities.
- Ensure **net charge is zero**.
- Save as **preped_Akt.PDB**.

3️⃣ Prepare the Ligand

- Fetch ligand from **PubChem** using **PubChem CID**:
  - **Tools > Structure Editing > Build Structure > PubChem CID**
  - Enter **CID** (or use **SMILES format** for novel compounds).
- Optimize the ligand just like the protein (remove solvents, add hydrogens, assign charges).
- Save the ligand as **prep_fisetin.mol2**.

4️⃣ Docking Using AutoDock Vina

1️⃣ Open AutoDock Vina:

- **Tools > Surface or Binding Analysis > AutoDock Vina**.

2️⃣ Set Docking Parameters:

- Define the **grid box** around the **active site** (where the inhibitor was).
- If the active site is unknown, set box size based on literature.

- Save the output file as **Akt_Fisetin.pdbqt**.

3️⃣**Prepare Receptor & Ligand**:

- **Delete the original inhibitor** (Actions > Atoms and Bonds > Delete).
- Choose:
  - **Receptor:** `preped_Akt.PDB`
  - **Ligand:** `prep_fisetin.mol2`
- Set all receptor and ligand options to **TRUE**.

4️⃣**Run Docking**:

- Use **Opal Web Service** or set the local path for AutoDock Vina.
- Click **OK** to start docking.

5️⃣Viewing Results

- After docking, results appear in a **dialog box** showing:
  - **Score** (binding affinity).
  - **RMSD (Root-Mean-Square Deviation)** values.
- To visualize **hydrogen bonds**:
  - **H Bonds > Add Count to Entire Receptor**.
  - Adjust **bonding parameters** (Intermodel for receptor-ligand bonds).

💾 **Save the session**: **File > Save Session as > "Akt Fisetin Docking"**

✅ **Chimera** is used to **prepare, visualize, and analyze docking** results.
✅ **AutoDock Vina** predicts how the **ligand binds to the protein**.
✅ **Docking results** include **binding affinity, hydrogen bonds, and RMSD values**.
✅ **Save your work** to revisit docking results later.

🚀 **Start docking and analyze your molecular interactions today!** 🎯

https://static.igem.org/mediawiki/2021/9/99/T--UChicago--ADVina.pdf