2 Viewing and Analyzing Proteins, Ligands and their Complexes

Overview

Analyzing the properties of proteins containing thousands of atoms is best accomplished by representing molecular properties in simplified graphical forms. In addition, drug design focuses on only a small portion of the protein structure. To design drugs, you need to analyze the interactions of small molecules with proteins using simple tools to find and visualize the key interactions -- fast. For example it is important to be able to:

- Define and manipulate groups of atoms,

- Display hydrogen bonds and atoms that bump together,

- Manipulate sequences,

- Identify active sites,

- Color by properties, like hydrophobicity and hydrophilicity,

- Locate crevices,

- Extracting information from sequence alignments

In this exercise, you will be introduced to BioMedCAChe's unique Sequence View, a powerful analysis tool for navigating stuctures containing tens of thousands of atoms and visualizing.

In this exercise, you will use CAChe to

- view the protein's accessible surface

- view and analyze the sequence of a protein

- locate the active site

- show and measure hydrogen bonds between the ligand and the protein

- display surface of the binding pocket and look at docking interactions

Viewing the accessible surface

An accessible surface is the surface of a protein that could be touched by a 1.4 Angstrom sphere.

✥ **To create and view the accessible surface**

1. Choose **File | Open**, move into the folder `3-Viewing and Analyzing Proteins and Ligands` and select `TB-PNP-All-Refined.csf`.

   `TB-PNP-All-Refined.csf` opens and displays the cleaned structure after refinement of the hydrogen positions and then subsequent refinement of all atom positions.

2. Using the Select Molecule Tool, click an atom in the protein.

   The protein highlights.

3. Choose **View | Hide Unselected**.

   All except the selected protein disappear. Only visible atoms - both selected and unselected - contribute to the surface. The selected atoms define the limits of the box in which the surface is drawn. Areas of the surface outside a box around the selected atoms are not drawn.

4. Choose **Analyze | Accessible Surface**.

   After a few seconds, the accessible surface file is created, read into memory and displayed. The small black text - **asf 0.01-** is the surface label. You select a surface by clicking on the center of the label.

   Mauve (purple-red) areas and blue areas are hydrophilic while the cream area is hydrophobic. Mauve indicates hydrogen acceptors (e.g. C=O) and blue indicates areas of hydrogen donors (e.g. -XH).

   As expected hydrophilic regions dominate the outside of the protein. Proteins tend to fold so that hydrophobic residues are on the interior and hydrophilic groups are on the exterior. Large patches of hydrophobic areas (cream colored) on the surface of a protein, suggest that a protein is involved in interactions with other proteins.

   You should also note that there is no tunnel leading to the active site. This is a crystal of TB-PNP with a bound inibitor: ImmH. TB-PNP has folded around the ligand, closing the tunnel and making it difficult to identify an active site from the solvent accessible surface.

   The labels (+1 and -1) showing through the surface are the atom charges.

✥ **Hide the accessible surface**

1. Choose **Analyze | Show Surfaces**.

   The Show Surfaces dialog opens.

2. Uncheck `accessible1.asf` and choose **OK**.

   The surface disappears.

✥ **Show the hidden atoms and bonds**

1. Choose **View | Show Hidden**.

All hidden atoms and bonds are displayed in preparation for the next step in the exercise.

Viewing the protein sequence

The sequence data for a protein 3D structure that is open in CAChe is displayed in the Sequence View window. When you open multiple proteins, their sequence data are added to the Sequence View window. Only one Sequence View window can be open at one time. You use the sequence window to:
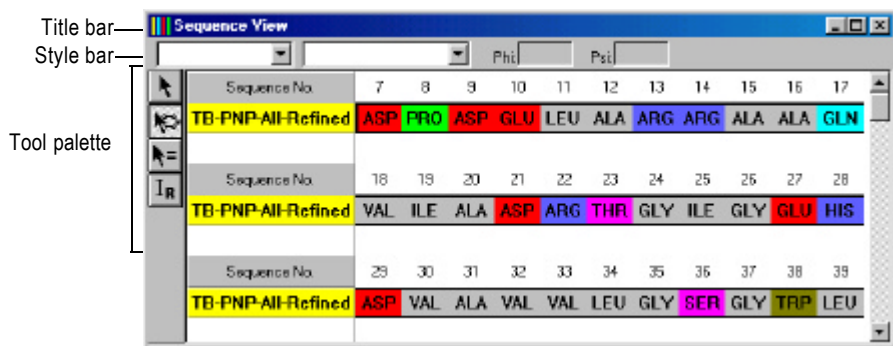
- view the sequence
- color the sequence by property
- mutate residues
- edit the sequence
- build peptides and proteins from residues
- analyze secondary structure
- adjust the conformation or secondary structure of a protein
- align sequences
- superimpose sequences.

	letter code	
Residue	**1**	**3**
Alanine	A	ALA
Arginine	R	ARG
Asparagine	N	ASN
Aspartic acid	D	ASP
Cysteine	C	CYS
Glutamine	Q	GLN
Glutamic acid	E	GLU
Glycine	G	GLY
Histidine	H	HIS
Isoleucine	I	ILE
Leucine	L	LEU
Lysine	K	LYS
Methionine	M	MET
Phenylalanine	F	PHE
Proline	P	PRO
Serine	S	SER
Threonine	T	THR
Tryptophan	W	TRP
Tyrosine	Y	TYR
Valine	V	VAL

✥ **To view the sequence of TB-PNP**

1. From a workspace window, choose **Analyze | Sequence** to display the sequence window.

   The Sequence View window displays sequences using 1-letter codes in rows that begin with the chemical sample name. The menu bar changes.

2. Choose **View | 3-Letter-Code**.

   3-letter codes are displayed for residues colored by chemical type.

The sequence has a direction, starting from the N terminus of the first residue to the C terminus of the last one. Hetero groups are always displayed, but water molecules (HOH or DOD) are never displayed.
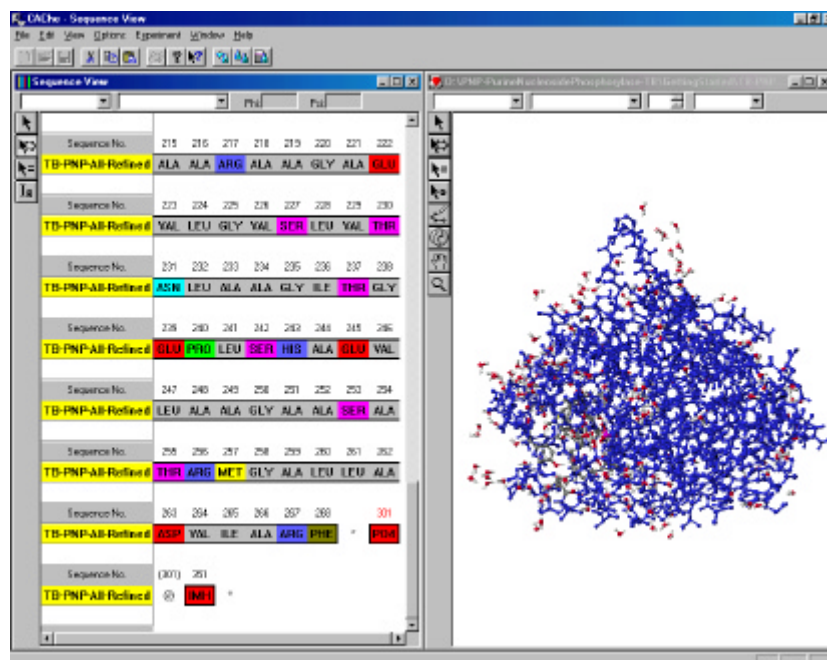
When you view a sequence, the Sequence View window comes to the front and the menus and toolbar change.

3.   Choose **Window | Tile**.

The Sequence and 3D Structure Windows are displayed side-by-side.

4.   Scroll the Sequence View Window to the bottom.

Your view should look like this

An asterisk '*' called the "chain terminal" marks the end of each protein chain. Note that the chain terminal is not part of the structure and you cannot select the chain terminal.

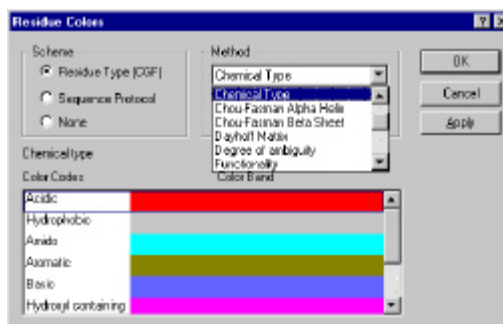## Analyzing the protein sequence

Initially residues in the Sequence View are colored by the chemical type method. You can change the coloring method to display other properties.

☝ **To color a sequence by property or residue type**

1.   Click on the title bar of the Sequence View to activate it and choose **View | Residue Colors...**

The Residue Colors dialog appears.

The Chemical Type method is highlighted.

2. Choose the color **Scheme | Sequence Protocol**.

    The list of choices in the **Method** pulldown menu changes.

3. Choose Method **Karplus and Schultz Flexibility** and select **OK**.

    The dialog closes and the residue colors change to indicate flexibility
    based on the local 7 residue sequence. A red residue is predicted to be in
    a highly flexible portion of the protein. A blue residue is in a relatively
    inflexible portion of the protein.

Method details are explained in the User Guide Chapter 21,
"Understanding Sequence Property Predictions"

    The **Method Window** specifies the number of adjacent residues that are
    grouped together for analysis. For example, a window of 7 will result in
    the sequence being analyzed in blocks of 7 residues to determine
    properties such as the antigenicity or hydropathy. Note that residues at the
    beginning and end of the sequence are not colored because they are
    outside Method Window.

Locating the active site
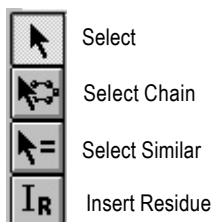
⮕ **To locate residues in the active site**

Select

Select Chain

Select Similar

Insert Residue

1. In the Sequence View Window, choose the Select Tool from the tool palette and click **IMH**.

   The bound ImmH ligand highlights and the rest of the sample becomes grey. The sequence view provides fast access to structures by residue and group.

2. Click on the title bar of the 3D Structure Window to activate it.

   The title bar becomes blue indicating the 3D Structure Window is active.

3. In the 3D Structure Window, choose **Edit | Select Neighbors**.

   The **Select Neighbors** dialog appears.
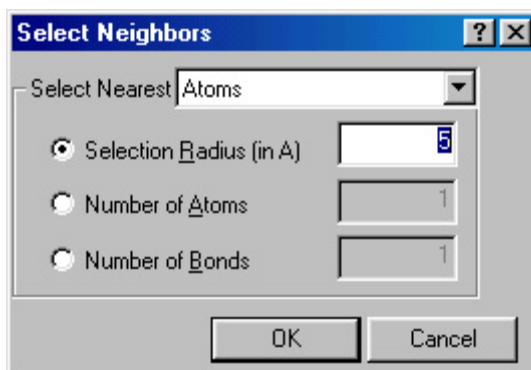
4. Chose **Residues, waters, HETs** from the **Select Nearest** pull down menu in the Select Neighbors dialog, enter a **Selection Radius** of 3 Angstroms and choose **OK**.

   The dialog closes and atoms within 3 Angstroms of any atom in ImmH are selected. In addition, residues in Sequence View that contain a selection atom are highlighted and all others are dimmed.

5. Scroll through the sequence and observe which residues have been highlighted.

   Note that residues near the ligand are often far from each other in the sequence. The three dimensional structure is necessary to understand active sites. Also note that no residues have very high flexibility. The residues associated with the active site have low to moderate flexibility.

At this point you should have all the residues in the active site selected. The 16 residues are:
Ser36, His90, Tyr92, Ala120, Ala121, Gly122, Tyr188, Glu189, Val205, Gly206, Met207, Ser208, Thr230, Asn231, His243, Val246,

PO_4^{3-} is also selected.

Differences in protein-ligand active sites may be important in determining binding selectivity. In Exercise 6, you examine the structure of human PNP and determine whether the residues in TB-PNP are conserved in human PNP.

Ꝃ **To name the active site**

1. In the 3D Structure Window, choose **Edit | Group Atoms**.

   The Group Atoms dialog appears.

2. Type the name "ActivesiteWithImmH" into the **Group Name** text box.
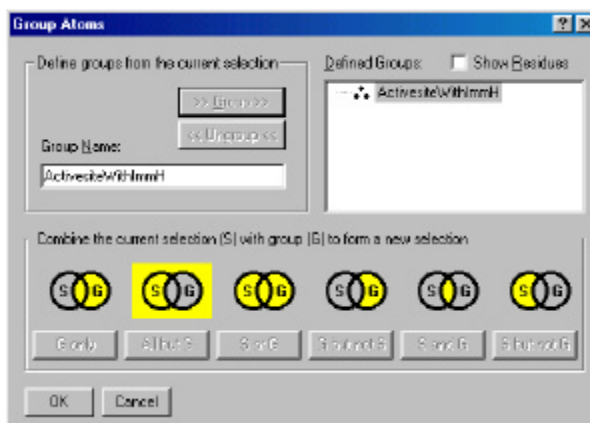
   The **>>Group>>** button highlights as you type.

3. Choose **>>Group>>**.

   The ActivesiteWithImmH group is created and added to the **Defined Groups** list on the right.

4. Choose **OK**.

   The Group Atoms dialog closes.

In the future, you will be able to pick the active site quickly from **Edit | Atom Groups** or by using the Select Group Tool.

Select Tool

Select Molecule

Select Similar

Select Group

Atom Tool

Rotate

Translate

Zoom
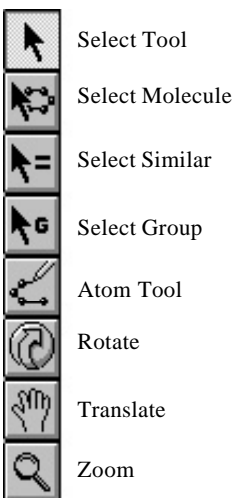
Locating the hydrogen bonds

ImmH has specific hydrogen bond interactions with TB-PNP. Understanding these interactions is important when designing new inhibitors. If other interactions are equal, inhibitors with more specific hydrogen bond interactions will bind better than those with fewer interactions.

✎ **To display hydrogen bonds between ImmH and TB-PNP**

1. In the Sequence View Window, choose the Select Tool click **IMH**.

    ImmH highlights in the 3D Structure Window.

📖 TIP

To turn off the numbers on the H-Bond labels, choose the Select Similar Tool, click on an H-Bond label to select all H-Bond labels, then choose **View | Geometry Label Attributes**. Uncheck **Value** on the **Label** tab and choose **OK**.

2. Click the title bar in the 3D Structure Window to activate it and choose **Analyze | Label H-Bonds**.

    Green distance labels appear indicating the distance between hydrogen bond donors and acceptors. Each green distance label measures the length of a hydrogen bond interaction.

3. Click **View | Fit in Window**.

    ImmH fills the window.

✎ **To identify the hydrogen bonding residues in TB-PNP**

1. In the 3D Structure Window click a protein atom at the end of an H-bond distance label. Then shift-click each atoms at the end of each of the remaining protein H-bonds to ImmH. Ignore the H-bonds to PO_4^{3-}.

    The selected atoms highlight in the 3D Structure Window and their residues highlight in the Sequence View Window. Other residues dim.

2. Scroll through the Sequence View and note the residue type and name.

    You should identify bonding to 5 different residues: Tyr188, Glu189, Met207, Asn231 (2 bonds), and His243. The hydrogen bond to Met207 is relatively long and expected to be weaker than the other hydrogen bonds.

Changes in ligand-protein hydrogen bonding may be important in determining binding selectivity. In Exercise 6, you examine the structure of human PNP and determine whether these residues important to hydrogen bonding are conserved.

Displaying the active site pocket

The protein active site is a pocket in the protein that contains the bound ligand. The surface of the protein near a bound ligand maps out the pocket. To generate this surface in BioMedCAChe, you select the surface first and then the accessible surface of the protein in a box enclosing the ligand is drawn. This is the surface of the protein adjacent to the ligand and the surface that forms a pocket around the ligand.

✍ **To view the surface of the binding site adjacent to the ligand**

1. In the Sequence View Window, choose the Select Tool and then choose residue **IMH**.

    ImmH highlights in the Sequence View Window and in the 3D Structure Window.

2. Activate the 3D Structure Window by clicking on its title bar, then choose **View | Show Hidden.**

    You do this to ensure that all atoms are included in the surface determination. Hidden atoms do not contribute to accessible surfaces.

3. Next choose **Analyze | Adjacent Surface - Pocket**.

    After several seconds, a wireframe surface is drawn around ImmH.

4. Choose **View | Hide Unselected**.

    All atoms and bonds except ImmH disappear. The adjacent surface and ImmH remain visible.

5. Click the background with the Select Tool.

    The adjacent accessible surface for the ImmH pocket is highlighted while ImmH remains selected and highlighted. You see a view similar to this.

📖 NOTE

You may have to rotate the structure to orient it like this view.
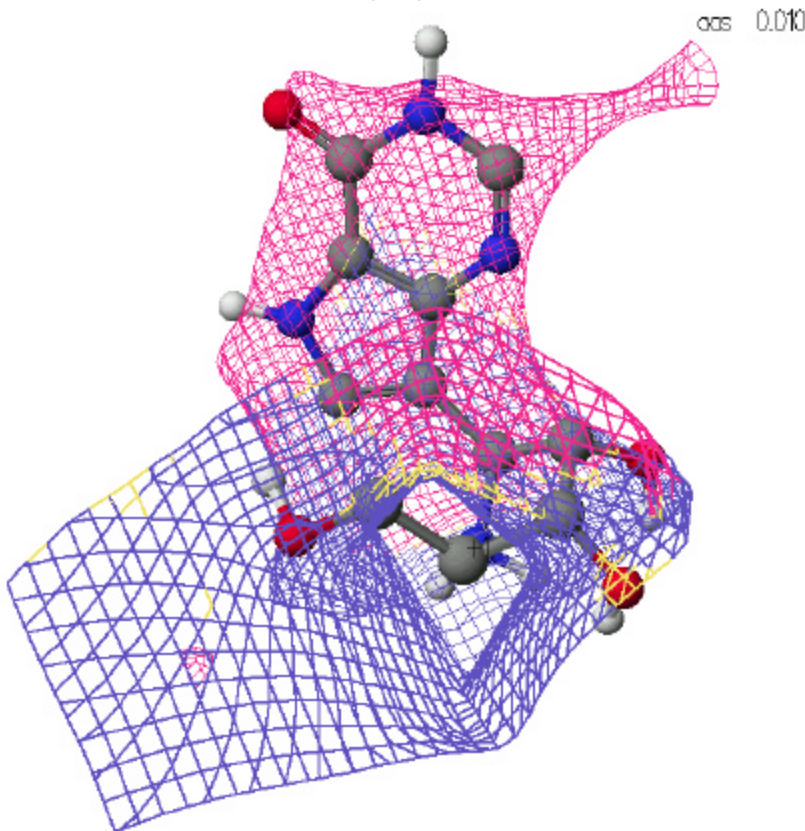
Red marks a surface where the protein has H-acceptors (e.g. -C=O), blue

marks a surface where the protein has H-donors and cream marks the hydrophobic surface. Note ImmH's hydrogen bond donors hook through the surface and ImmH's hydrogen bond acceptors touch the surface. ImmH's hydrophobic atoms are far from the surface.

Notice the red N1-C2 region in the upper righthand corner. This suggests that ImmH might be modified to yield compounds of higher affinity by the addition of functional groups to C2 that can donate hydrogens to acceptor groups on the protein which do not currently form hydrogen bonds. You will check for this region in human PNP. If it is not present, then you may be able to exploit this region for selectivity.

This completes the viewing and analyzing proteins and ligands introductory exercise. As you work with BioMedCAChe you will discover many additional views and analysis methods.