Supporting information for

The latest AI breakthroughs in structural biology: protein binder design and conformational state prediction

Luciano A. Abriata – luciano.abriata@epfl.ch

School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne

CH-1015 Lausanne, Switzerland

SI material 1: Full chat behind Figure 1D

I opened PDB 1XBB in the PDB, it is supposed to have imatinib bound to it but where is it? The image is what I see right now

Analysis

I completely understand the frustration—you load a "bound complex" and it looks like just a lone protein floating in the void.

In the case of **PDB 1XBB** (Syk kinase bound to Imatinib/Gleevec), the ligand is definitely there, but it often hides behind default visualization settings. Here is how to find it and why it might be missing from your current view:

1. The "Hidden" Ligand (Ligand ID: STI)

In PDB 1XBB, Imatinib is listed under the chemical ID **STI**. If you are looking at a ribbon or "cartoon" diagram of the protein, the software often hides small molecules to keep the view clean.

*How to reveal it in the Mol viewer:**

- **Look at the right-hand panel:** Under the "Components" or "Sequence" tab, look for **STI**.
- **Toggle Visibility:** Click the small "eye" icon next to **STI** or **Ligand**.
- **Focus:** Click on the word "STI" to center the camera on the binding pocket. The ligand should appear as a "stick" model nestled between the N-terminal and C-terminal domains of the kinase.

2. Check the "Asymmetric Unit" vs. "Biological Assembly"

Sometimes the default view only shows one chain (Chain A). In 1XBB, the asymmetric unit contains one molecule of the Syk kinase domain and one molecule of Imatinib.

- If you see multiple proteins but no drug, you might be looking at a symmetry-related view.
- Ensure the "Model" or "Assembly" setting is set to **Deposited** or **Biological Assembly 1**.

3. Binding Site Location

If you are zoomed out, look for the "hinge region" of the kinase.

In 1XBB, Imatinib binds in the **ATP-binding pocket**, which is the deep cleft between the small N-terminal lobe (mostly β-sheets) and the larger C-terminal lobe (mostly α-helices).

> **Note:** Interestingly, 1XBB is famous because Imatinib binds to Syk in a "cis" (folded) conformation, which is different from the "trans" (extended) conformation it takes when binding to its primary target, Abl kinase.

Quick Troubleshooting Table

If you see...	Try this...
Only ribbons/tubes	Change the style to "Ball and Stick" for "Ligands" in the display menu.
A blank screen	Click the **"Center View"** (house icon) to reset the camera.
"Heterogens" are off	Check your settings to ensure "HETATM" or "Heteroatoms" are being loaded.

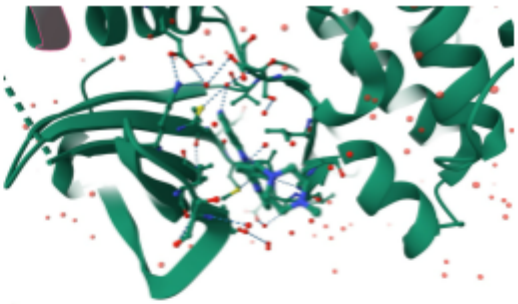
Export to Sheets

Would you like me to generate a specific list of the amino acids that form the binding pocket so you can select and highlight them manually?

Visualizing Imatinib (STI) in Mol*

This video provides a step-by-step guide on how to navigate the Mol* viewer, specifically showing how to identify, focus on, and change the representation of ligands like Imatinib.

Exploring PDB Structures in 3D with MolStar (Mol*): Introductory Guide - YouTube
RCSBProteinDataBank · 32K views



oh thanks, I got it! now can you tell me what interactions you see between imatinib and the protein?

Analysis

Great find! Now that you've zoomed into the pocket, you can see how snugly Imatinib (STI) fits into the ATP-binding site. In **PDB 1XBB**, Imatinib is actually in a rare "cis" conformation, which allows it to make a very specific set of contacts with the **Syk kinase** domain.

Based on the structural data for this specific complex, here are the key interactions you are seeing:

1. The Hydrogen Bond Network

The most critical interactions for binding affinity are the hydrogen bonds, often shown as dashed lines in your viewer.

- **The Hinge Region:** Imatinib mimics ATP by forming hydrogen bonds with the "hinge" of the kinase. Specifically, the N-atom of the pyridine ring and the adjacent NH group interact with the backbone of residues like **Ala453**.
- **The Gatekeeper:** There is often a crucial interaction near the "gatekeeper" residue, which in Syk is **Met448**. This residue helps define the shape of the pocket Imatinib occupies.
- **The Catalytic Lysine:** Look for a bond between the drug and **Lys395**. This is a highly conserved residue in kinases that Imatinib often "anchors" to.

2. Hydrophobic "Sandwiching"

While H-bonds provide specificity, hydrophobic interactions provide the "glue."

- The aromatic rings of Imatinib (the benzene and pyrimidine rings) are sandwiched between hydrophobic side chains from the top (N-lobe) and bottom (C-lobe) of the pocket.
- Residues like **Leu377**, **Val385**, and **Leu501** create a greasy environment that stabilizes the non-polar parts of the drug.

3. The Specific "Cis" Conformation

What makes 1XBB unique is that Imatinib is "folded."

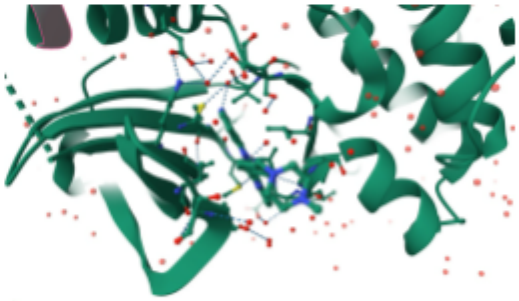
- In many other kinases (like Abl), Imatinib binds in an extended "trans" shape.
- In Syk, the drug folds back on itself so that its own hydrophobic groups can interact with each other while fitting into a slightly deeper, more compact pocket within the protein.

How to see these labels in Mol*

To confirm exactly which amino acids are involved in the image you uploaded:

1. **Hover your mouse** over the amino acid "sticks" that the dashed lines are pointing to. The bottom left or right of the screen will usually display the residue name and number (e.g., `ASP 512` or `ALA 453`).
2. **Click a residue** to select it, then use the "Component" menu to change its color so you can distinguish the "Hinge" from the "Catalytic Loop."

Would you like me to explain why Imatinib is a "selective" inhibitor and why it binds to Syk differently than it does to BCR-ABL?



please focus only on what you see in the image, not what you already know about the system... what interactions do you see? list them please

Analysis

Based on the close-up view of the binding pocket in your images, here are the specific visual interactions currently rendered:

1. Hydrogen Bonds (Blue Dashed Lines)

The software is explicitly highlighting several hydrogen bonds between the Imatinib molecule and the surrounding protein residues:

- **The Hinge Region:** There are two distinct blue dashed lines connecting the nitrogen atoms of the drug's heterocyclic rings to the backbone of a specific protein loop (the hinge).
- **Side Chain Anchors:** You can see multiple dashed lines extending from the polar groups (nitrogens and oxygens) of the drug to various nearby amino acid side chains, particularly in the upper and left regions of the pocket.
- **Water-Mediated Bonds:** Some blue dashed lines connect the drug to the small red spheres (water molecules), which then connect to the protein, indicating a water-bridged interaction network.

2. Steric and Geometric Complementarity

- **Cleft Fitting:** The drug molecule is physically nestled in a deep "V-shaped" cleft between a series of β-sheets (on the left) and α-helices (on the right).
- **Conformational Shape:** The Imatinib molecule is not straight; it is bent into a "U" or "cis" shape, allowing it to wrap around a central protein protrusion.

3. Hydrophobic Proximity

- **Van der Waals Packing:** Several non-polar side chains (rendered as green sticks) are positioned very close to the carbon rings of the drug without blue dashed lines. Their physical proximity suggests hydrophobic packing, particularly around the central aromatic rings of the molecule.

4. Coordination with Water

- **Solvent Interaction:** There are several red spheres (water molecules) clustered directly around the Imatinib molecule, especially near its polar tail (the right side of the molecule in your zoom-in).

How to identify the specific atoms

If you want to know the exact names of these residues:

- **Hover your cursor** over any of the green "sticks" that are connected to the drug by a blue dashed line.
- A tooltip will pop up in the viewer (usually at the bottom) telling you the residue name and number (e.g., **ASP 512** or **ALA 453**).

Would you like me to help you identify the specific amino acid names for those "hinge" residues shown in the hinge loop?

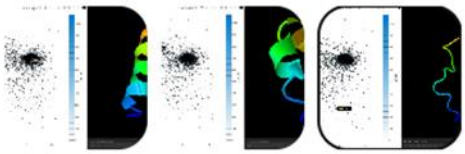
Nice, but summarize this further into one line per type of interaction:
1. Hydrogen Bonds (Blue Dashed Lines)
The software is explicitly highlighting several hydrogen bonds between the Imatinib molecule and the...

Here is the condensed summary of the interactions visible in your images of the **1XBB** complex:

- **Hydrogen Bonds:** Blue dashed lines connect the drug's nitrogen/oxygen atoms directly to the protein hinge, side chains, and bridging water molecules.
- **Steric Complementarity:** The drug adopts a "U-shaped" cis-conformation to fit precisely within the V-shaped cleft between β-sheets and α-helices.
- **Hydrophobic Proximity:** Non-polar green sticks (side chains) pack closely against the drug's carbon rings without forming dashed lines, indicating Van der Waals contact.
- **Coordination with Water:** Multiple red spheres (solvent molecules) cluster around the drug's polar tail and mediate connections to the protein structure.

Would you like me to identify the specific residue numbers for those hydrogen-bonded "side chain anchors" seen in your zoom-in?

SI material 2: Another chat with Gemini 3, here asking help to interpret a free energy landscape computed by BioEmu and inspected with the PDBMS' free energy inspector



I'm showing you here snapshots from some MD analysis I'm doing. What can we conclude about the studied peptide?

Analysis

Based on the snapshots you've shared from your Molecular Dynamics (MD) analysis, it looks like you are exploring the **folding landscape** of a small peptide. The plots show a 2D projection of the Free Energy Landscape (FEL) alongside representative structures at specific time frames.

Here is what we can conclude about the peptide's behavior:

---

1. Structural Transitions

The snapshots illustrate three distinct conformational states:

- **Frame 4715 (Image 1):** The peptide is in a highly ordered **alpha-helical state**. This corresponds to the deep, high-density "well" on the right side of the plot.
- **Frame 2126 (Image 2):** This represents a **transition or partially folded state**. You can see the helix beginning to break or "fray," moving away from the central cluster.
- **Frame 3032 (Image 3):** This is a **random coil or unfolded state**. The structure is extended and disordered, corresponding to a point far outside the high-density region in the "basin of attraction."

2. Stability and the Free Energy Landscape

The FEL (the scatter plot) provides a "map" of where the peptide likes to spend its time:

- **The Main Basin:** The dense black cluster (around Coordinate 1 = 3, Coordinate 2 = 1) is clearly the **global free energy minimum**. This indicates that the alpha-helix is the most thermodynamically stable state under your simulation conditions.
- **Conformational Entropy:** The wide scattering of points in the negative Coordinate 2 region (where Frame 3032 sits) suggests a broad, flat landscape for the unfolded state. While reachable, these states are much less populated (lower density) than the helix.
- **Folding Pathway:** The peptide seems to follow a specific trajectory to reach the helical state. Frame 2126 suggests the "path" to folding involves specific intermediates rather than a completely random collapse.

---

3. Preliminary Summary

The peptide you are studying is a **helix-former**. It spends the majority of its time in a compact helical fold but possesses enough thermal energy to occasionally unfold into a disordered string before eventually finding its way back to the stable basin.

Technical Considerations

To refine this analysis, you might want to consider:

- **Defining Coordinates:** What are "Coordinate 1" and "Coordinate 2"? Usually, these are the first two Principal Components (PCA) or variables like Root Mean Square Deviation (RMSD) and Radius of Gyration (R_g).
- **Convergence:** Does the peptide transition back and forth between these states multiple times? If it only unfolded once and stayed there, your simulation might not be fully converged.

Would you like me to help you interpret the specific PCA eigenvectors or calculate the free energy barriers (ΔG) between these states?

here the defining coordinates are PCA. Might it make a difference if I plot RMSD vs helix and Gyration radius?

Switching from **PCA (Principal Components Analysis)** to specific physical descriptors like **RMSD** and **Radius of Gyration (R_g)** will definitely change how you view the landscape. While PCA is great for capturing the largest motions, it's an abstract mathematical projection. Using RMSD and R_g gives you a more "physical" intuition.

Here is how that shift would likely impact your analysis:

---

1. Interpretability: Abstract vs. Physical

- **PCA (Current):** You know that Frame 4715 and Frame 3032 are "different," but you don't inherently know *how* based on the axes. You have to look at the structures to realize Coordinate 2 represents "unfolding."
- **RMSD vs. R_g:** The axes become self-explanatory.
  - **Low RMSD + Low R_g:** A compact, native-like fold (your helix).
  - **High RMSD + High R_g:** An extended, unfolded string.
  - **High RMSD + Low R_g:** A "molten globule" or misfolded compact state.

2. Capturing the "Folding Funnel"

Using RMSD and R_g is the classic way to visualize a **Folding Funnel**.

- **RMSD** acts as a progress coordinate (how close am I to the goal?).
- R_g acts as a compactness coordinate (how collapsed am I?).

This often reveals if the peptide collapses first and then forms the helix, or if the helix forms while the peptide is still relatively extended.

3. Potential Downsides

There is a reason you used PCA first:

- **Sensitivity:** PCA is specifically designed to find the coordinates where the peptide moves the *most*. RMSD and R_g are "blind" to specific internal rotations. For example, two different helical orientations might have the same R_g and RMSD, causing them to overlap on your plot, whereas PCA would likely separate them.
- **Resolution:** PCA might show you two distinct sub-basins (e.g., a left-handed helix vs. a right-handed helix) that would just look like one big blob on an RMSD plot.

Comparison Summary

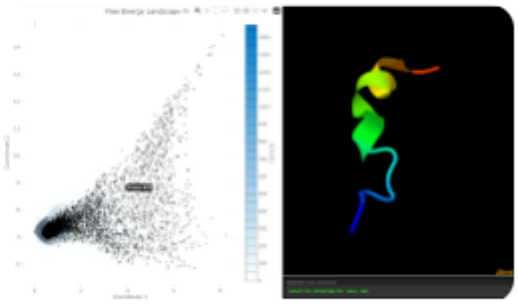
Feature	PCA	RMSD vs. R_g
Primary Goal	Maximize variance/motion	Monitor structural similarity & size
Clarity	Requires "decoding" of axes	Immediate physical meaning
Basin Separation	Excellent for distinct modes	Can "smear" different structures together
Standard Use	Exploratory data analysis	Traditional folding mechanism study

Export to Sheets

Recommendation:

Keep both! Use PCA to see how many distinct "states" exist, then use **RMSD vs.** R_g to explain those states to others in a way that is biologically intuitive.

Would you like me to help you draft a Python script (using MDTraj or MDAnalysis) to calculate the RMSD and R_g from your trajectory files?



good idea. here's what i got (coordinate 1 is RMSD to an ideal helix, and coordinate 2 is Rg). What can you infer now? And I have one specific question, would you say the helix is folding from N to C terminus or from C to N terminus?

Show thinking

This new plot provides a much clearer physical picture of your peptide's folding mechanism. By using **RMSD to an ideal helix** (Coordinate 1) and **Radius of Gyration (R_g)** (Coordinate 2), we can now map the structural transition directly to physical dimensions.

Structural Inference from the Landscape

- **The Native Basin:** The extremely dense, dark blue cluster at the bottom left (RMSD ≈ 0.5–1.0 Å, R_g ≈ 7.0–7.5 Å) represents the fully folded, compact **alpha-helix**. This is your global energy minimum.
- **The Unfolded Region:** As Coordinate 1 (RMSD) increases, we see the landscape broaden significantly. The points with RMSD > 6 Å and R_g > 10 Å represent **extended, disordered states**.
- **The "V-Shape" Distribution:** The data follows a diagonal trend. This tells us that as the peptide loses its helical character (higher RMSD), it simultaneously becomes less compact (higher R_g). There is no significant population of "compact but non-helical" states, which suggests a relatively cooperative folding process.

Direction of Folding: N to C or C to N?

Looking at **Frame 874** in your structure viewer, we can determine the folding direction. In standard protein coloring (like the "group" or "chain" scheme in JSmol/PyMOL), the **Blue end is the N-terminus** and the **Red/Orange end is the C-terminus**.

Based on the structure shown:

- The **N-terminal (blue/cyan)** region is a disordered, flexible tail.
- The **C-terminal (yellow/orange/red)** region is clearly wound into a well-defined alpha-helix.

Conclusion

The peptide is **folding from the C-terminus to the N-terminus**.

In Frame 874, the "nucleation" of the helix has occurred at the C-terminal end, while the N-terminal end remains a random coil. If the simulation continues toward the native basin, you would expect to see that helical winding "propagate" toward the blue N-terminal tail until the entire chain is folded.