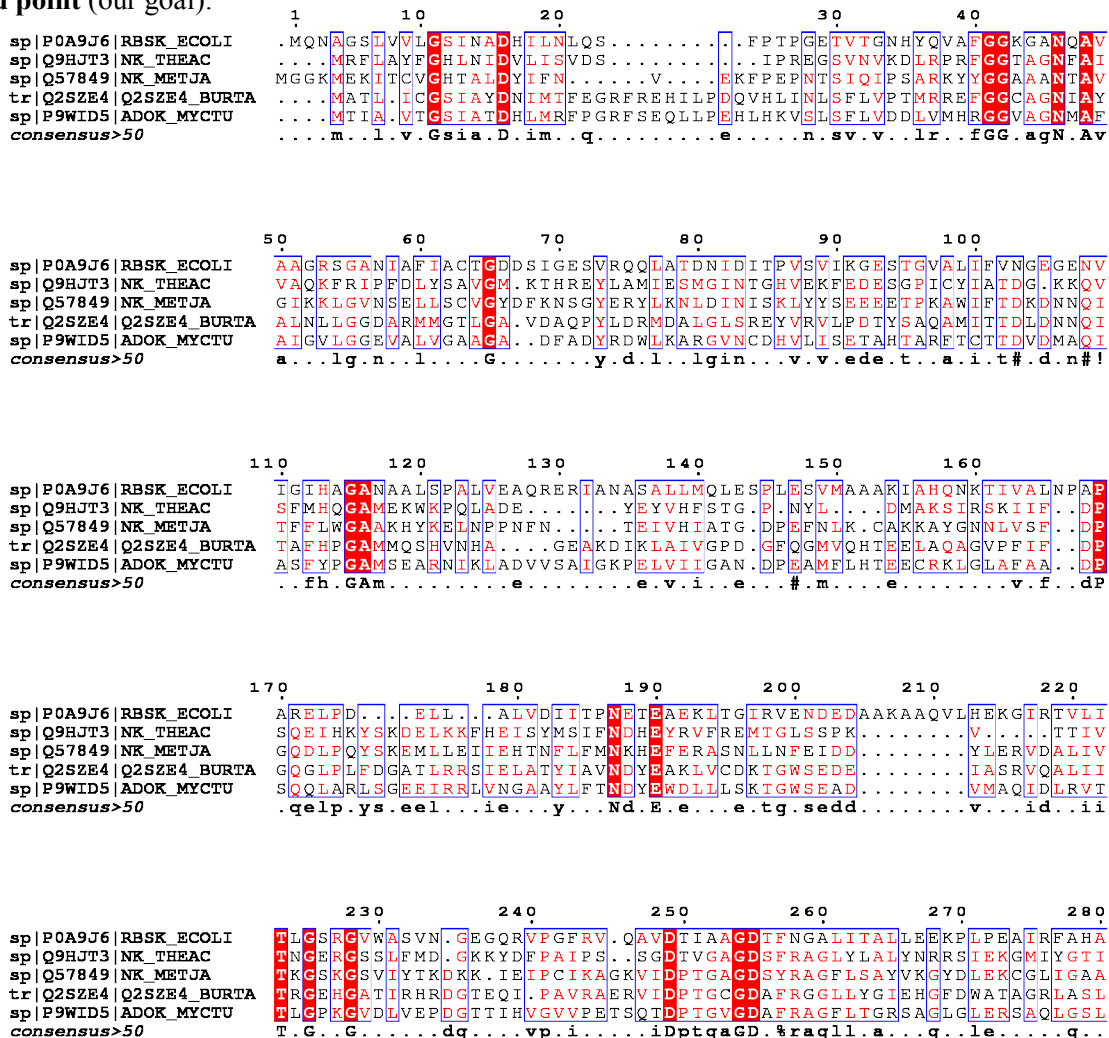# In-Class Activity, II:
## Further Sequence and Structural Analysis of Nucleoside Kinases

In Part I (last week), we began learning PyMOL basics, including how to create and customize molecular 'representations' (cartoon diagrams, coloring, etc.). Now, in Part II, we'll cover three AREAS of utility in analyzing TaNK in order to propose a mutant: (i) protein (3D) structure comparison, (ii) sequence comparison, and (iii) detailed analysis/visualization of the active site of an NK homolog from *Methanocaldococcus jannaschii* (the same tools and approaches apply for TaNK). AREA I—protein structure comparison—was our topic last time, and AREAS II and III are described below.

---

**Area II**: **Sequence analysis**: Comparative analysis of protein sequences by creating, analyzing, and visualizing multiple sequence alignments (MSA)

**Starting point**: (i) We want to perform a sequence analysis of TaNK (PDB ID `3bf5`); and, (ii) motivated by Elkin et al., we wish to compare the TaNK sequence to four specific NK homologs of known 3D structure. These four are from *M. jannaschii* (`2c4e`; MjNK), *Mycobacterium tuberculosis* adenosine kinase (`2pkm`; MtAK), *Burkholderia thailandensis* (`3b1o`; BtNK), and an *E. coli* ribokinase (EcRK). PDB codes are indicated after each of the above, except for EcRK. So, let's take this as our start point: How can we find EcRK's PDB ID (so that we can get its sequence)? (In bioinformatics there's usually >1 'right' way!)

**End point** (our goal):
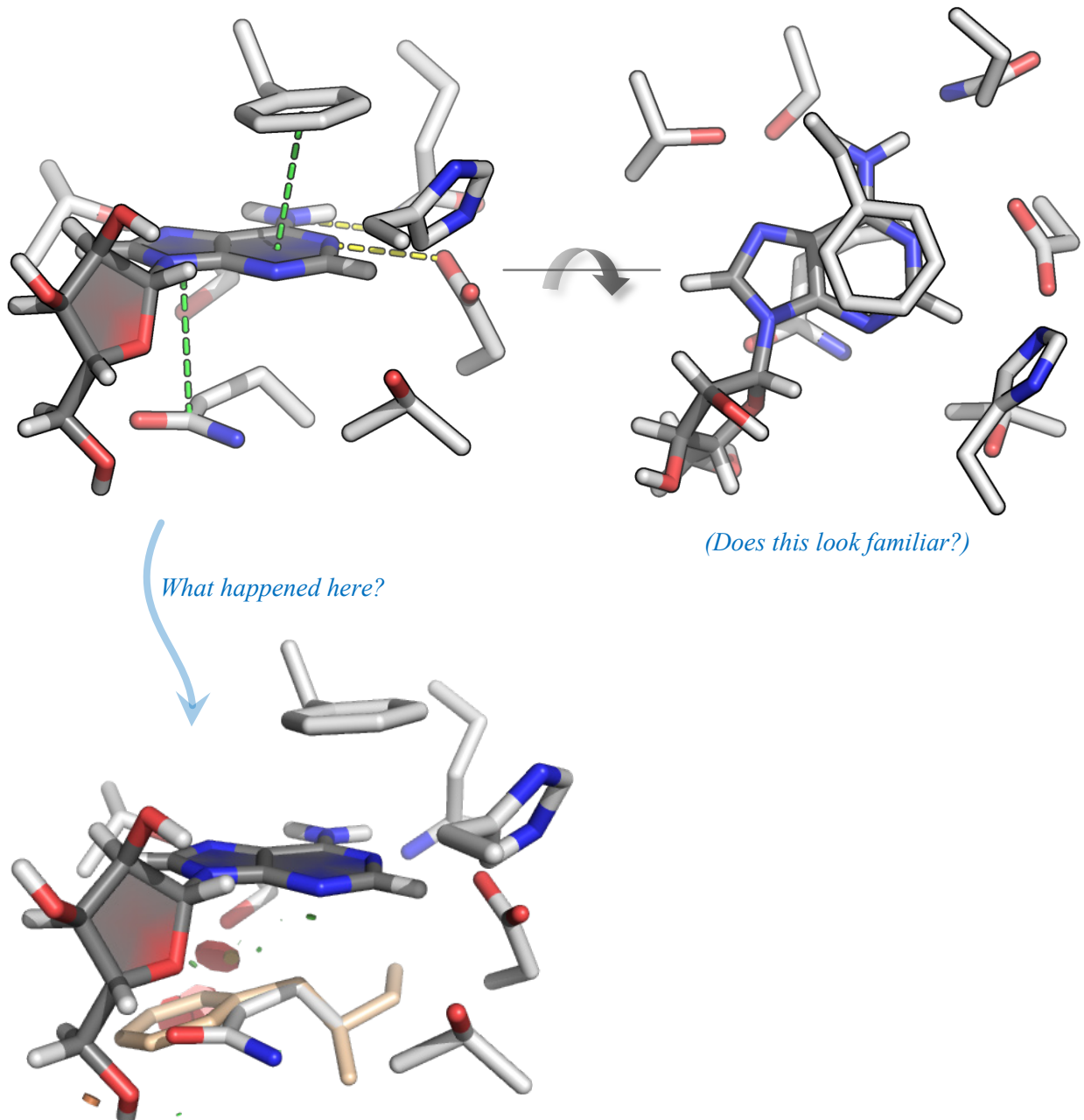


©2017 Cameron Mura; originated oct2017; revised fa2017

How we got there…

➢ Step 1: (i) Have a copy of Elkin et al. handy, and (ii) Google "pdb 3bf5"

➢ Step 2: Let's find our missing PDB code (for EcRK; done mostly in the PDB website)
- Search for a homolog of `3bf5` (ideally EcRK will crop-up in the results) using the PDB's built-in search tools. In particular, 'Sequence Similarity' → BLAST link. (Have a quick look; see all the *pairwise alignments*?). Now prune the search results to just *E. coli* via the menu options on the left ('Refinements' -> 'Organism'). Which of these five hits should we use? Let's use `1rka`. (Why?)
- So, now our final list of PDB IDs for sequence analysis is: `3bf5`, `2c4e`, `3b1o`, `2pkm`, `1rka`.

➢ Step 3: Gather all the desired protein sequences
- Visit the UniProt website, click "Retrieve/ID mapping" menu (near top)
- Enter the 5 PDB codes. Check the 'Options' section below (enter PDB as 'from'). Now, you can either download them in bulk (in FASTA format), by clicking the radio-box to select all 5 seqs and then 'Download' button, or you can utilize the built-in 'Align' button. Now, let's do the former—download (save) the file locally (i.e., to your computer) as '`my_NK_seqs.fasta`'. (For fun, you can also try UniProt's built-in align approach [it uses Clustal-Omega]; highlight the resultant MSA by 'similarity', explore around. It is pretty feature-rich, let's you add more sequences and do another round of alignment; also can 'edit and resubmit' for alignment, and so on.)

➢ Step 4: Compute the MSA, using a web-server implementation of the Clustal-Omega MSA method.
- Google search "clustal omega"; should return the EMBL-EBI web-server as top (or near-top) hit.
- Upload the sequences from your local file, `my_NK_seqs.fasta`; click 'Submit' the launch the calculation; should take less than a couple minutes.
- Play with the results ('Results for job `clustalo-####`' window will appear), download them via the "Download Alignment Results" button. Save to a local file '`my_NK_seqs_ClustalO.aln`'; note the extension here (which indicates Clustal's file-format [not FASTA], which is fine for the next step… for fun, can try opening that file in WordPad).

➢ Step 5: Display/visualize a 'publication-quality' MSA, using either ESPript or BoxShade (or another approach of your own choosing).
- Perform online searches, 'esprit' and 'boxshade alignment'; these should return the appropriate sites as the top hit. Both of these are webserver-based tools that enable users to download the results in various formats (PostScript, PDF, PNG or TIFF; even RTF [BoxShade] for editing in Word). As new users, we advise using the default settings for most of these webservers and informatics tools. That said, in the case of ESPript (above) the 'Global score threshold' was changed to 0.5, 'display consensus line' was activated, and PNG alignment output file was requested.

➢ Step 6: Polish/post-processing of the results (optional, depending on purpose of the MSA—for manuscript, poster, etc., versus actively exploring the results)
- This can be done in PowerPoint, MS Word, or using various desktop publishing tools (Adobe Illustrator, etc.). If you want some ideas about how to schematize the secondary structural elements (above the MSA, like in Elkin's Fig 1), make sure to check the PDB webpage for entry 3BF5—click on the 'Sequence' tab at the top, scroll down, see how you can use that information.

**Area III**: **Active site analysis**: Detailed visualization & analysis of the adenosine–binding site in *M. jannaschii* homolog (MjNK w/ bound Ado; PDB `2c49` [`2c4e` is *apo* form])

**Starting point**: To enable us to design and propose a mutation that we predict will alter the substrate specificity of TaNK, a key step will entail detailed study of the 3D structure and biochemistry of the active-site; ideally, this will involve any ligands, substrates, cofactors, etc. bound at such sites. To begin, let's visually examine and analyze the detailed ('zoomed-in') structure of the MjNK enzyme bound to adenosine (Ado). Note that one excellent place to start is by combing through all the deep information available in the PDB entry; e.g., click on "Poseview Image of ADN in 2C49" under "2D diagrams and interactions" in order to quickly assess the ligand-binding site.

**End point** (our goal): A detailed image of atomic interactions mediating Ado binding



*(Does this look familiar?)*

*What happened here?*

## How did we generate these images (and, in the process, can we analyze the active site)?

- ➢ `fetch 2c49`

- ➢ Rotate the view a bit, then Display → Sequence. Scan the sequence for '`ADN`', '`MG`', any other ligands (non–amino-acid residues; known as 'heteroatoms' in PDB terminology)

- ➢ Create atom selections for the Ado and for a region of interest (near the Ado; see also the Elkin paper):
  ```
  select myAdo, (chain A and resn ADN)
  select myROI, (resi 13,138,164,33,31,111,113,163)
  ```

- ➢ Notice something that may cause trouble? Let's try defining the atom selection this way instead:
  ```
  select myROI, ( (chain A AND resi 13,138,164,111,113,163) OR
                  (chain B AND resi 31,33) )
  ```

- ➢ Let's also create a myROI atom selection which is pruned to just the $C^{\alpha}$ atoms + side-chains atoms (no main-chain):
  ```
  select myROI_sc, (myROI AND NOT name N,C,O)
  ```

- ➢ Middle-click '`myROI`' named atom selection (right-hand menu) and then zoom-in a bit with mouse

- ➢ `hide everything, all`

- ➢ Add hydrogens to `myAdo`: '`h_add myAdo`'; adjust default setting for display of hydrogen sticks:
  ```
  set stick_h_scale, 1.0
  ```

- ➢ Now, show the sticks (and adjust appearance of double-bonds):
  ```
  show sticks, (myAdo OR myROI_sc); set valence, 0
  ```

- ➢ From the atom-selection menus for `myROI_sc` and `myAdo`: Color by element, pick a scheme you like. If, after doing that, you want to darken the carbon atoms in the myAdo rings (to, say, a dark grey), try the following: '`color grey50, myAdo and elem C`')

- ➢ To slightly thin the sticks, versus default setting: '`set stick_radius, 0.2`'

- ➢ Fancy adenine and ribose rings:
  ```
  set cartoon_ring_finder, 3
  set cartoon_ring_mode, 1
  show cartoon, myAdo
  ```

- ➢ Make the background white (as usual, can also use menu sequence): '`bg_color white`'

- ➢ This is a good juncture at which to play around with measuring some distances, e.g. between the two fused purine rings in the adenine base and the neighboring side-chains. To do this, use the Wizard → Measurement → 'Distances to Rings' menu sequence

- ➢ Let's render (ray-trace) the scene, after adjusting a setting to make nice outlines around bonds…
  ```
  set ray_trace_mode, 1   (optionally, also 'set ray_trace_gain, 0')
  ray
  ```
  …and write the image to a file:
  ```
  png myFileName.png
  ```

- ➢ Finally, explore the mutagenesis wizard to examine effects of possible side-chain substitutions/ mutations (Wizard → 'Mutagenesis' menu sequence)