Identification of possible functionality of an unknown protein sequence

Abhishek Dasgupta (06MS07)
Sambit Bikas Pal (06MS03)

Indian Institute of Science Education and Research, Kolkata

April 18, 2008

What we've done: in a nutshell.

- Picked a protein from SWISSPROT (ideally this would be an unknown protein sequence; however without access to an unknown sequence, we took a known one.)

What we've done: in a nutshell.

- Picked a protein from SWISSPROT (ideally this would be an unknown protein sequence; however without access to an unknown sequence, we took a known one.)
- Did a sequence search on PFAM.

What we've done: in a nutshell.

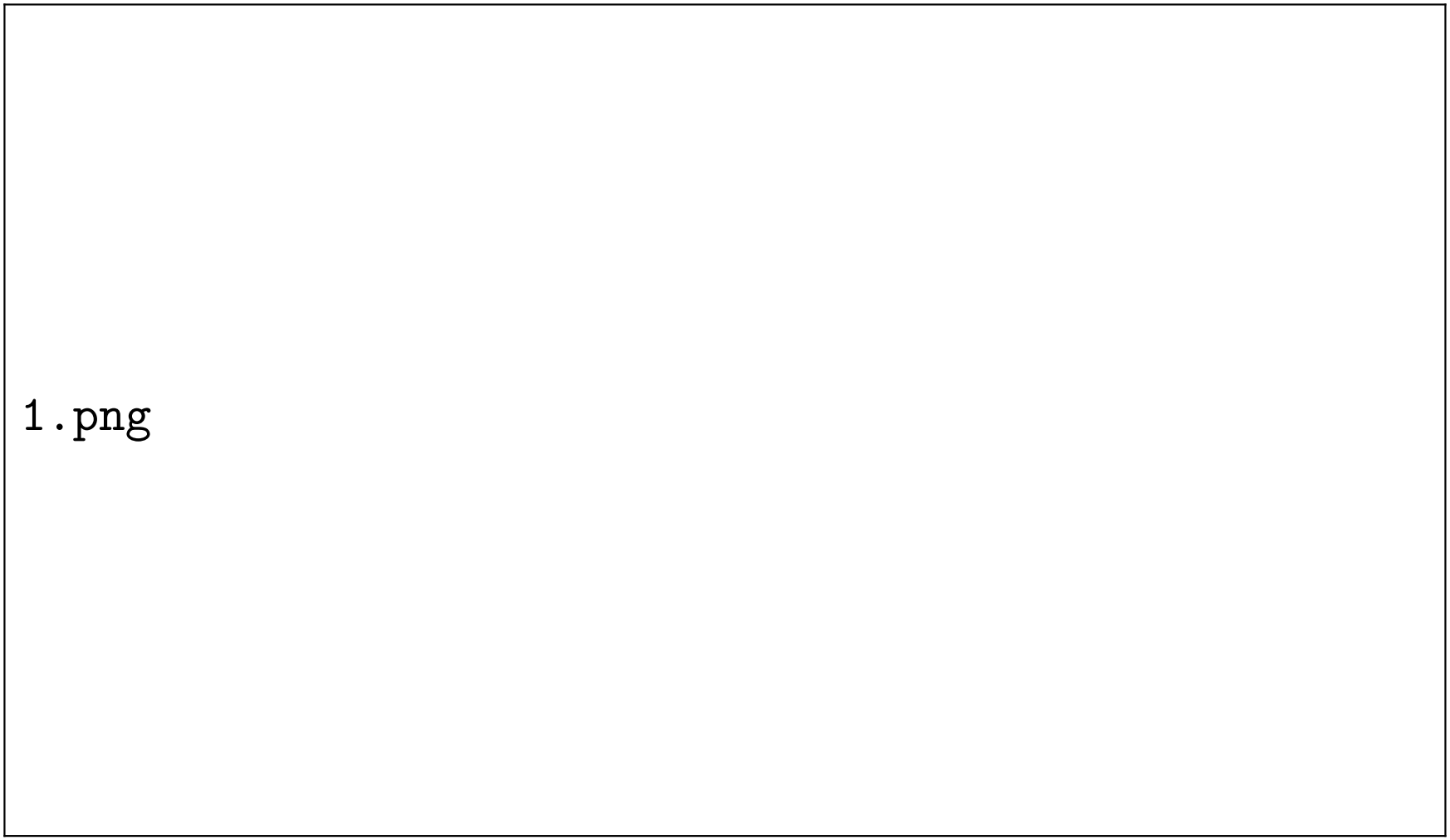
- Picked a protein from SWISSPROT (ideally this would be an unknown protein sequence; however without access to an unknown sequence, we took a known one.)
- Did a sequence search on PFAM.
- Found a match! Topoisomerase IV (PF00521)
  (what are topoisomerases?)

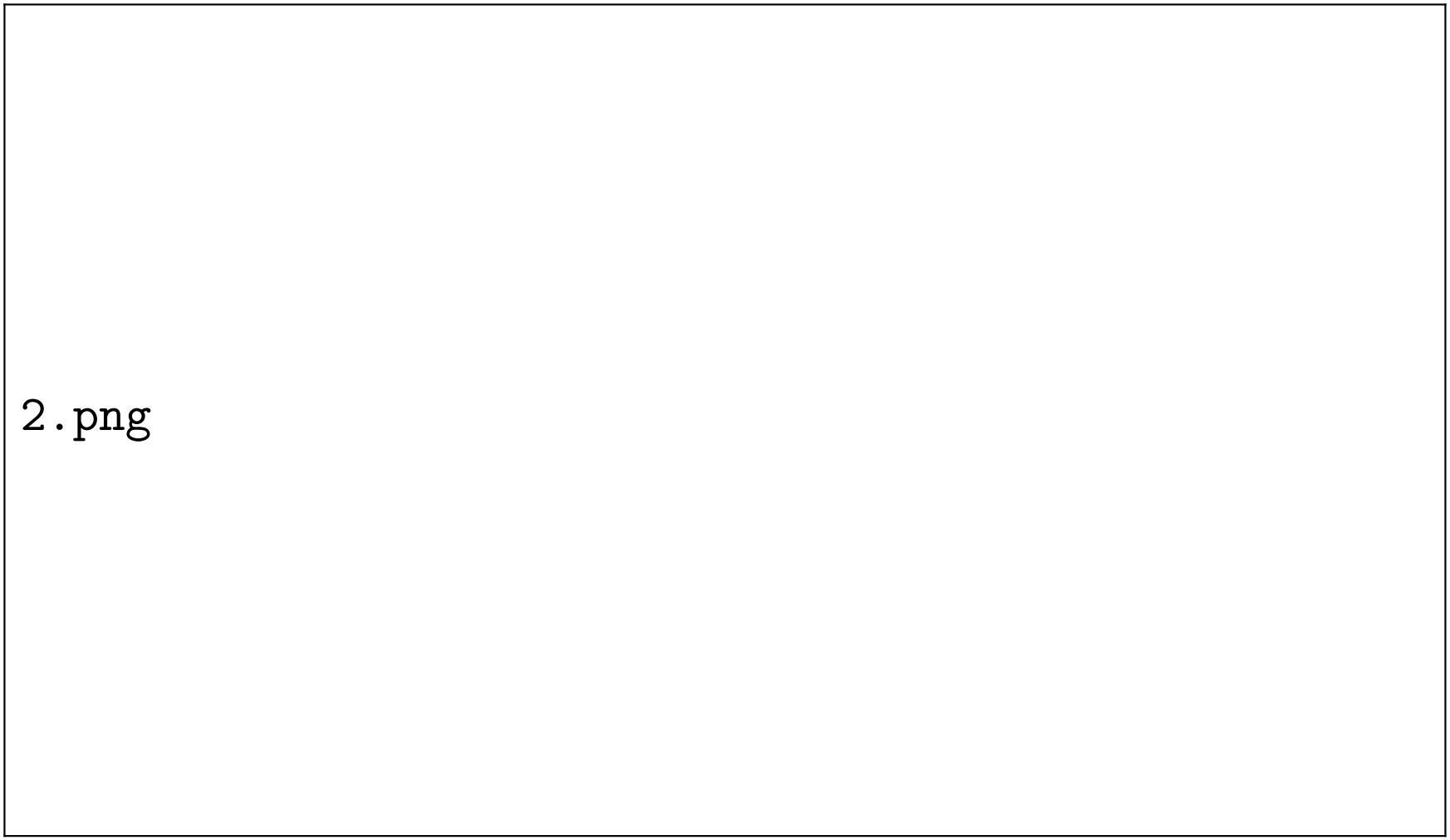
What we've done: in a nutshell.

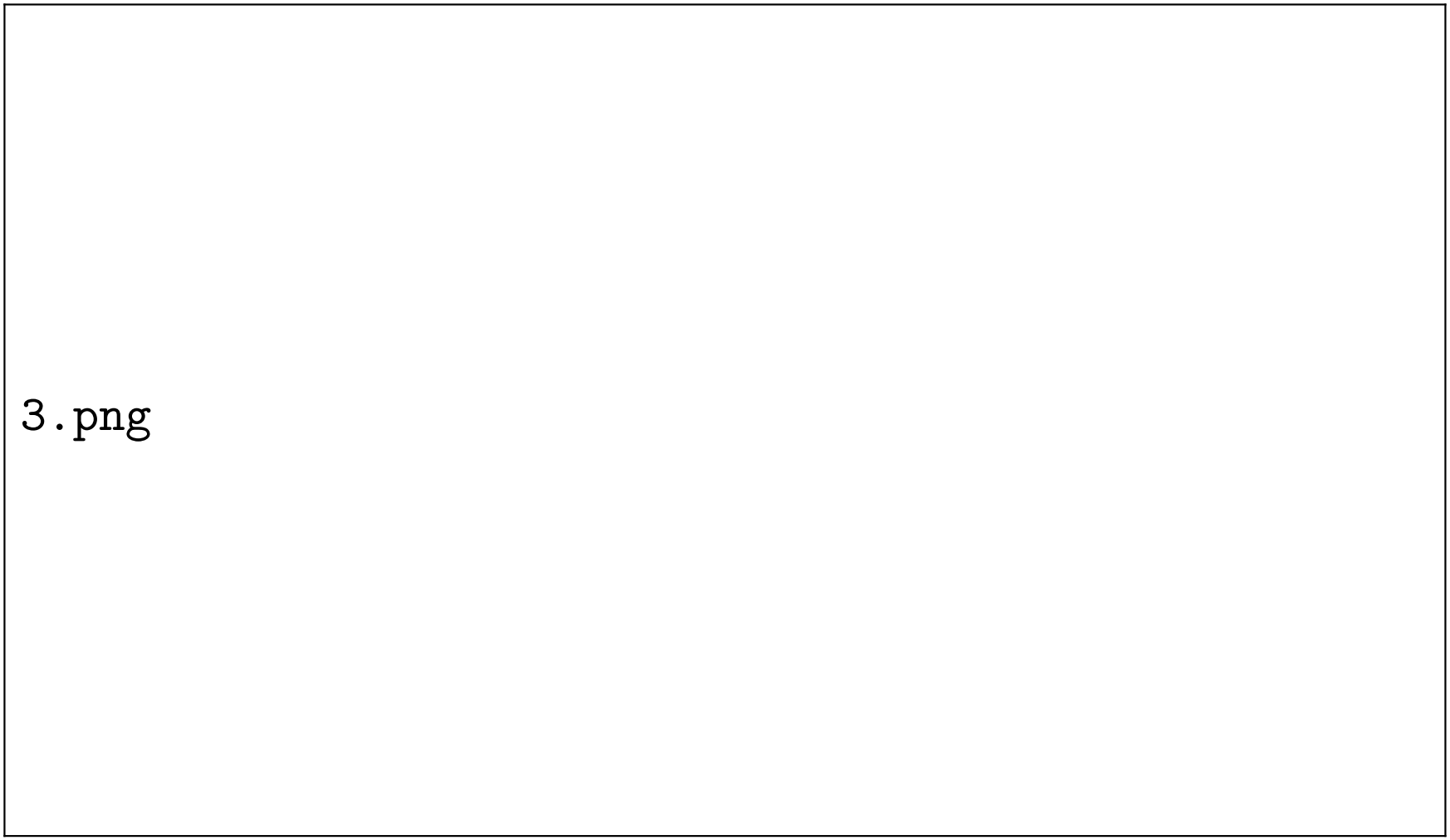
- Picked a protein from SWISSPROT (ideally this would be an unknown protein sequence; however without access to an unknown sequence, we took a known one.)
- Did a sequence search on PFAM.
- Found a match! Topoisomerase IV (PF00521)
  (what are topoisomerases?)
- Did CLUSTALW alignment with the PFAM **seed**. Identified conserved residues.

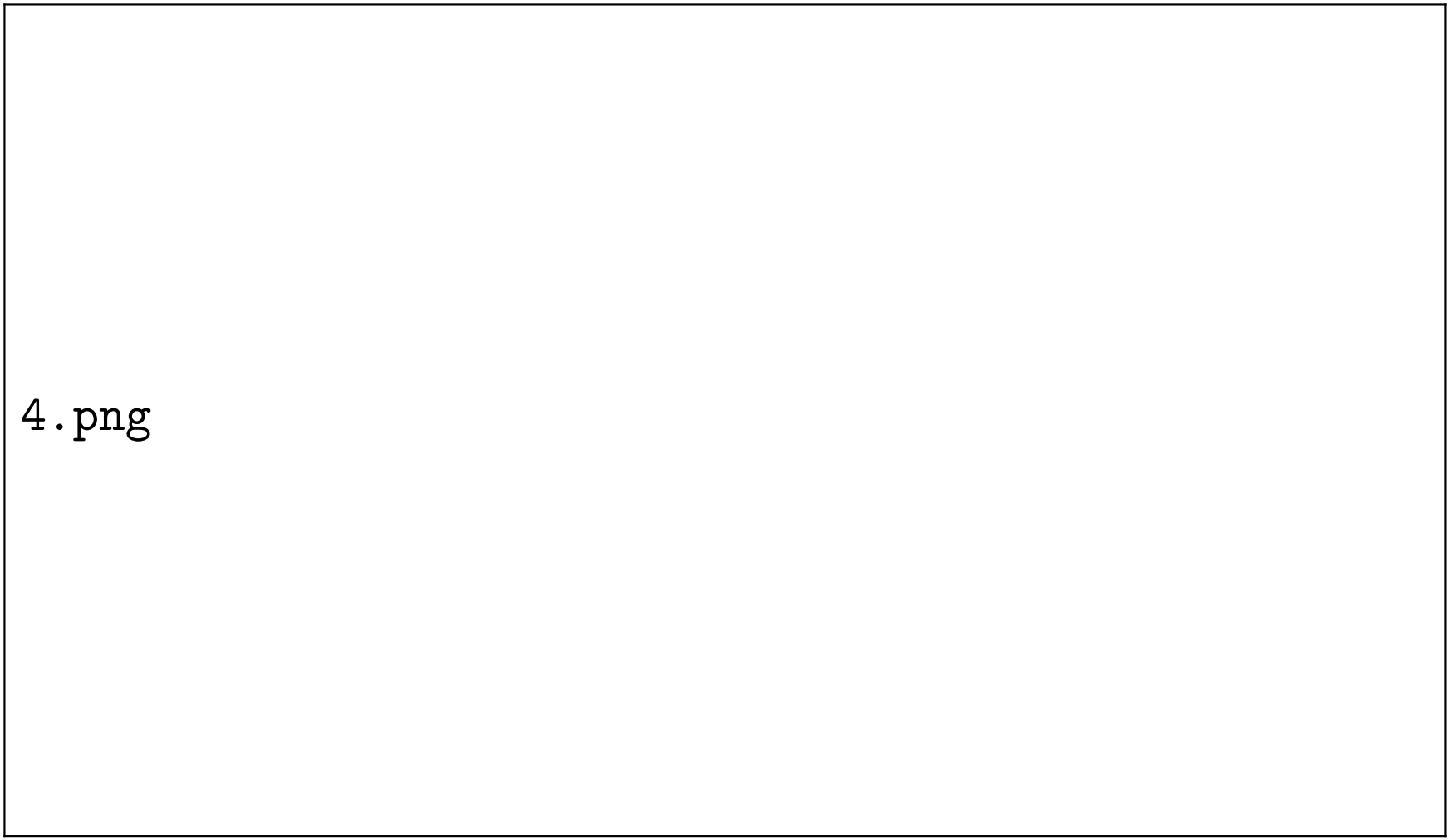
What we've done: in a nutshell.

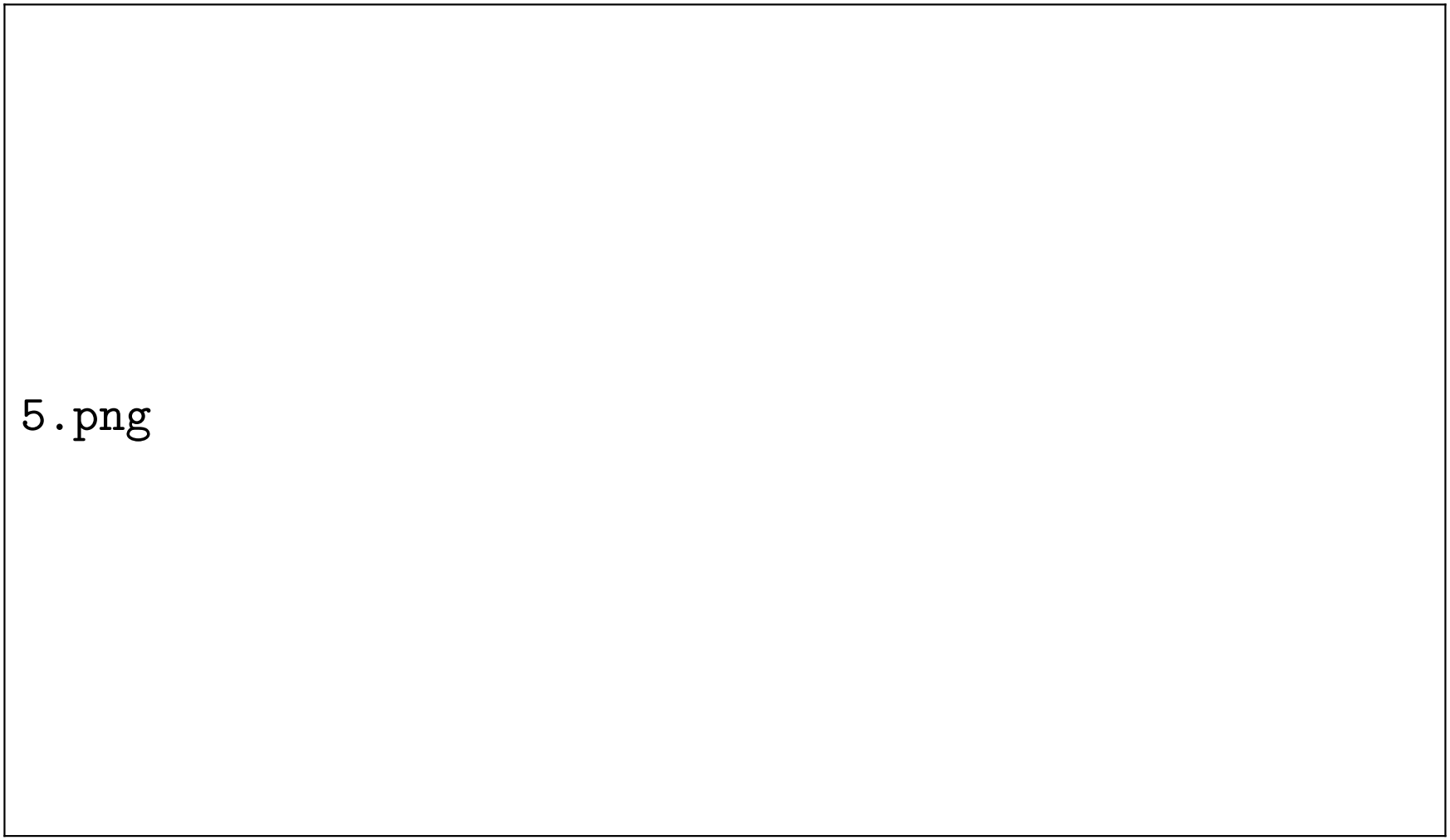
- Picked a protein from SWISSPROT (ideally this would be an unknown protein sequence; however without access to an unknown sequence, we took a known one.)
- Did a sequence search on PFAM.
- Found a match! Topoisomerase IV (PF00521)
  (what are topoisomerases?)
- Did CLUSTALW alignment with the PFAM **seed**. Identified conserved residues.
- Read the structure paper $\longrightarrow$ important residues $\longrightarrow$ are they there?

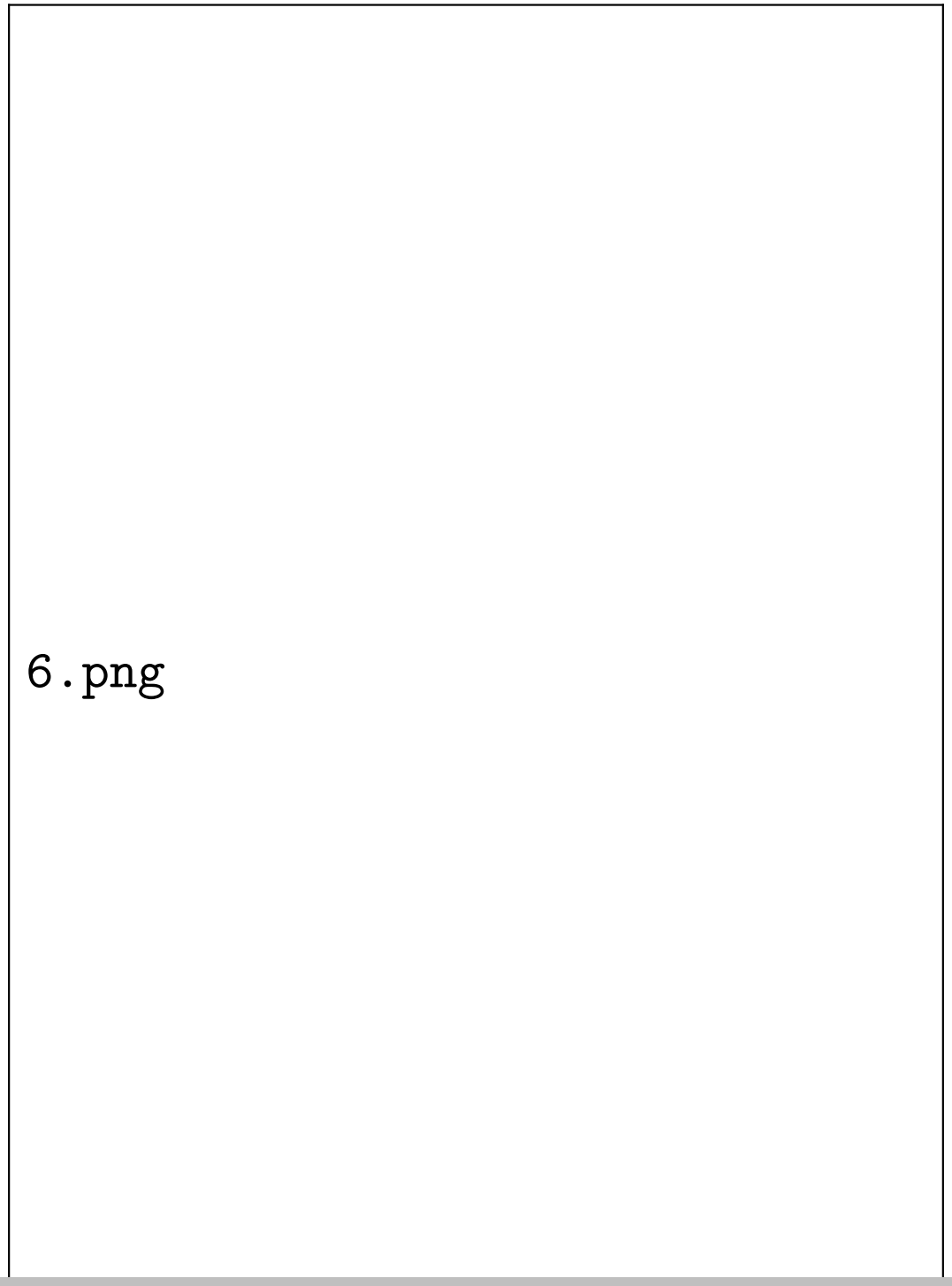

1.png


2.png


3.png

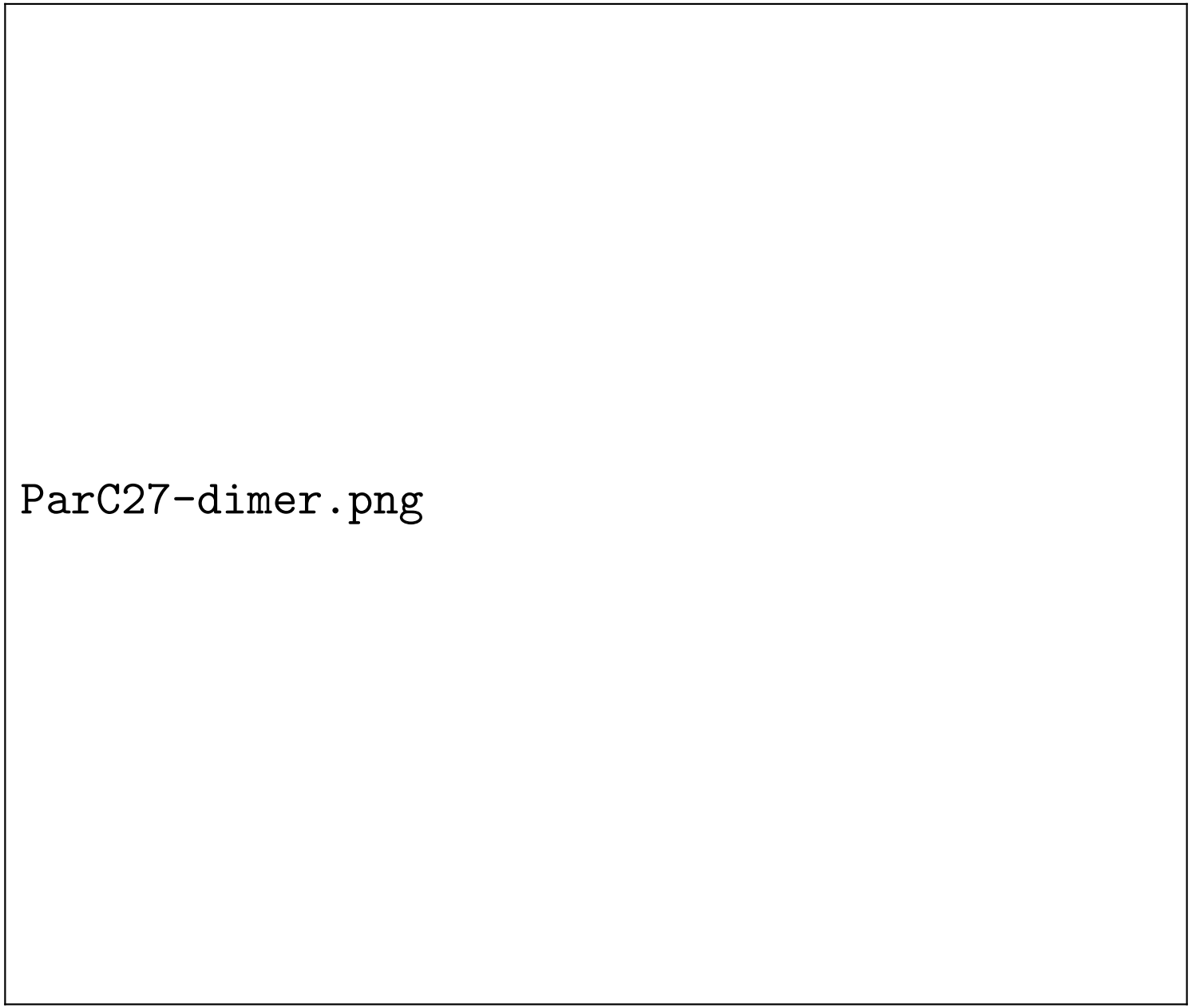

4.png


5.png


6.png

From the structure paper for 1ZVT (doi:10.1016/j.jmb.2005.06.029) we see that topo IV is mainly encoded by **parC**, **parE**
parC comprises two domains:

- (28-158) helix turn helix motif similar to CAP; contains active sites needed for DNA cleavage: Arg119 and Tyr120. Arg119 and Tyr120 was found exactly conserved across the test sequences (taken from Pfam seed PF00521) and our unknown protein.
- (159-340) **tower** packs against CAP domain; structural support.
- also there's a compact α-helical bundle connected by long α-helices to the tower domain and the C-terminal domain.


ParC27-dimer.png

Knowing the unknown

So what is the unknown protein?

- It is most probably a topoisomerase IV. It contains the Arg119 and Trp120 residues as mentioned earlier.
- The topoisomerase portion starts only from sequence 42, as is evidenced by the lack of alignments before that residue.

Thanks!

We are especially grateful to Dr. Rana Bhadra without whose help this project would not have been realised.