**Algorithm-Based Insights**

1. **LCS or DP-based similarity**
   * Use LCS to find common subsequences **within each class** (e.g., all class-0 GPCRs).
   * Compare LCS length between different classes (e.g., class-0 vs. class-1) to infer sequence-level differences.
2. **Divide and Conquer**
   * Split sequences by class and process each group recursively.
   * Or apply divide-and-conquer to segment sequences into functional motifs.
3. **Greedy or Approximate Methods**
   * Try greedy heuristics to match patterns in transcription factors (class-6), which often have repeating binding motifs.
4. **Backtracking**
   * Find all possible subsequences of a given length and classify which ones appear more frequently in certain classes.
5. **Branch and Bound**
   * Prune impossible or irrelevant subsequence combinations based on class-specific features.

**📊 Potential Project Outputs**

* A **heatmap or matrix** of LCS similarity scores across classes.
* A table like:

| **Compared Classes** | **Avg LCS Length** | **Interpretation** |
| --- | --- | --- |
| 0 vs 0 | 125 | High similarity (same family) |
| 0 vs 6 | 45 | Very different structure |
| 3 vs 4 | 80 | Some overlap (both synth enzymes) |

* Charts showing class-wise subsequence patterns or frequency.

**🔄 Bonus Tip: Biological Context**

Knowing what each class **does** biologically can even help shape your hypotheses:

* GPCRs (0) and Ion channels (5) both involve **signal transduction** → might show some similarity.
* Kinases (1) and phosphatases (2) perform **opposing functions**, but structurally may share catalytic domain motifs.
* Transcription factors (6) are usually very **sequence-specific** → may show less similarity to others.