Experiment 4: Protein Structure Prediction using SWISS-MODEL

AIM

To predict the 3D structure of a target protein using SWISS-MODEL automated homology modeling server, analyze template selection parameters, and evaluate model quality using integrated assessment tools including GMQE and QMEAN scoring functions.

THEORY

Principle of Homology Modeling

Homology modeling, also known as comparative modeling, is a computational approach for predicting the three-dimensional structure of proteins based on a fundamental principle observed in molecular evolution: protein structure is more conserved than sequence. This conservation arises because protein function depends critically on structure, and natural selection maintains structural integrity even as sequences diverge over evolutionary time. When two proteins share significant sequence similarity (typically above 30% identity), they almost invariably adopt similar three-dimensional folds. This remarkable relationship between sequence and structure forms the foundation of homology modeling, enabling researchers to predict the structure of a target protein by using experimentally determined structures of homologous proteins as templates.

SWISS-MODEL Pipeline Overview

SWISS-MODEL is a fully automated protein structure homology-modeling server that has pioneered accessible structural prediction since 1993. The server implements a sophisticated four-step workflow that transforms a protein sequence into a reliable three-dimensional model. The process begins with template identification, where SWISS-MODEL employs two complementary search strategies against its curated Template Library (SMTL). BLAST provides rapid identification of closely related templates with high sequence similarity, while HHblits extends the search capability to detect remote homologs through sensitive profile-based methods. This dual approach ensures comprehensive coverage across the spectrum of evolutionary relationships.

Once suitable templates are identified, the pipeline proceeds to the critical target-template alignment phase. This step goes beyond simple sequence comparison by incorporating evolutionary information from multiple sequence alignments, ensuring that structurally and functionally important residues are correctly matched between the target and template sequences. The alignment quality directly impacts the final model accuracy, making this step crucial for successful modeling.

The actual model construction is performed by the ProMod3 modeling engine, which employs a sophisticated multi-step approach. Quality assessment represents the final critical component of the SWISS-MODEL pipeline. The server employs multiple complementary metrics to evaluate model reliability at both global and local levels. The Global Model Quality Estimate (GMQE) provides an overall confidence score ranging from 0 to 1 by combining properties from the target-template alignment, including sequence identity, coverage, and alignment quality. QMEAN (Qualitative Model Energy ANalysis) offers detailed per-residue quality estimation using statistical potentials derived from high-resolution crystal structures, helping users identify which regions of the model are most reliable. Additionally, Ramachandran analysis validates the stereochemical quality by examining backbone dihedral angles, ensuring that the model conforms to allowed conformations observed in experimental structures. Together, these assessment tools enable users to make informed decisions about model reliability and suitability for different applications, from hypothesis generation to structure-based drug design.

EXERCISE

Part A: Target Sequence Selection

1. **Access UniProt Database**
    - Navigate to https://www.uniprot.org
    - For this tutorial, we will use **Superoxide Dismutase [Cu-Zn] from Drosophila melanogaster**
    - Search using UniProt accession code: **P61851**
    - Alternative proteins for independent practice:
        - Mouse MDM2 (P23804)
        - Hemoglobin subunit beta (P09905)
        - Any protein from your research interest
2. **Retrieve Protein Information**
    - Note that function of SOD
    - Check the "Structure" section for existing experimental structures
    - Note any prior structural details available for SOD
    - Download sequence in FASTA format
    - Record protein details: length and localization

Part B: Template Search

1. **Access SWISS-MODEL**
    - Go to https://swissmodel.expasy.org
    - Click "Start Modelling"
2. **Input Target Sequence**
    - Enter UniProt accession code: **P61851**

- OR paste the SOD FASTA sequence
- Project title will auto-populate: "Superoxide dismutase [Cu-Zn]"
- Enter email address for notification
- Click "Search for Templates"

3. **Monitor Search Progress**
   - Observe the search log showing:
     - What do you see?
   - Wait for completion

Part C: Template Analysis and Selection

1. **Review Template List** For SOD, you will find multiple high-quality templates with >67% identity Create a table documenting top templates (example expected results):

Template ID	Organism	Resolution	Identity	Coverage	GMQE	QSQE

2. Template Selection Decision For SOD with high sequence identity (>67%):

- Select top-ranking template
- Note: Multiple templates can be selected for comparison
- Consider biological relevance (presence of metal ions)

3. **Build Model**

- Select template(s) by checking boxes
- Click "Build Model"
- Wait for model generation (5-15 minutes)

Part D: Model Analysis (40 minutes)

1. **Global Quality Assessment**
   - Record GMQE score (0-1 scale)
   - Note confidence level:
2. **Local Quality Analysis (QMEAN)**
   - Examine per-residue plot showing:
   - Identify poorly modeled regions (loops, termini)

Part E: Mutant Modeling Task

1. Edit the FASTA sequence to delete "HGAPVDENRHLGDLGNIEATGDCPT" amino acids. The sequence after

```
>sp|P61851|SODC_DROME Superoxide dismutase [Cu-Zn] OS=Drosophila
melanogaster OX=7227 GN=Sod1 PE=1 SV=2
MVVKAVCVINGDAKGTVFFEQESSGTPVKVSGEVCGLAKGLHGFHVHEFGDNTNGCMSSGPHFNPYGKEKVNIT
DSKITLFGADSIIGRTVVVHADADDLGQGGHELSKSTGNAGARIGCGVIGIAKV
```

2. Paste the mutant sequence into SWISS-MODEL and repeat the modeling workflow.

3. Record new quality scores and annotate any structural changes in regions surrounding the deletion.
4. Compare results to the original model and discuss reliability.