3DPatch: fast sequence and structure conservation annotation in a web browser

David Jakubec

Institute of Organic Chemistry and Biochemistry of the CAS

National Bioinformatic Conference ENBIK 2018
June 11, 2018

Introduction: 3DPatch motivation

- Evolutionary conservation → powerful indicator of functionally important regions of protein structures

Introduction: conservation example

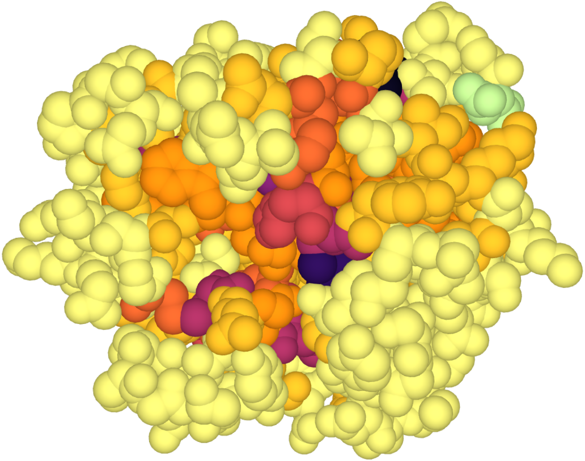


Figure 1: Structure of trypsin (PDB ID 5fxl) marked-up according to relative residue conservation level.

Introduction: 3DPatch motivation

- Evolutionary conservation → powerful indicator of functionally important regions of protein structures
    - Problem: calculation and projection of residue conservation scores onto 3D structures is difficult.

Introduction: 3DPatch

- 3DPatch
  - Web application facilitating conservation level-based mark-up of protein 3D structures
    - Built upon established infrastructure.
    - Modular, transferable, & extensible.

Introduction: 3DPatch

- 3DPatch
  - Web application facilitating conservation level-based mark-up of protein 3D structures
    - Built upon established infrastructure.
    - Modular, transferable, & extensible.
  - Interactive speed — $\mathcal{O}(10^1)$ s.

Introduction: 3DPatch

- 3DPatch
  - Web application facilitating conservation level-based mark-up of protein 3D structures
    - Built upon established infrastructure.
    - Modular, transferable, & extensible.
  - Interactive speed — $\mathcal{O}(10^1)$ s.
  - Minimal user input requirements & configuration
    - Only a sequence required.
    - No manual multiple sequence alignment (MSA) construction and transfer necessary.
    - No prior knowledge about homologous 3D structures required.

Introduction: back-end services and databases

- HMMER web server
  - Web front-end and interface (API) to HMMER suite programs
    - HMMER: package for sequence analysis using profile hidden Markov models (HMMs).
    - Profile HMM: statistical model of a MSA.
  - Single sequence (*phmmer*) and profile HMM (*hmmsearch*) searches against sequence databases.

Introduction: back-end services and databases

- HMMER web server
  - Web front-end and interface (API) to HMMER suite programs
    - HMMER: package for sequence analysis using profile hidden Markov models (HMMs).
    - Profile HMM: statistical model of a MSA.
  - Single sequence (*phmmer*) and profile HMM (*hmmsearch*) searches against sequence databases.
- Skylign
  - Calculates sequence logo and per-position information content (IC) for a profile HMM query

Introduction: profile HMM logo example

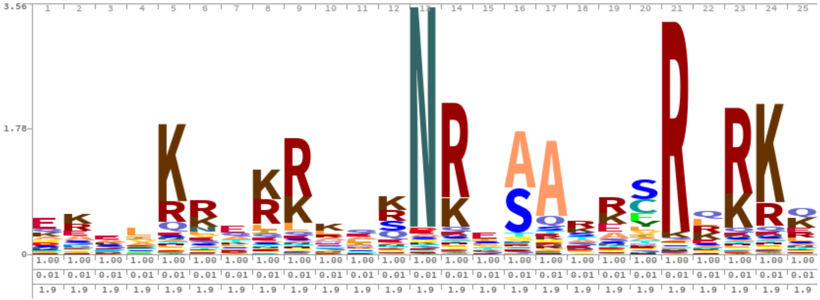
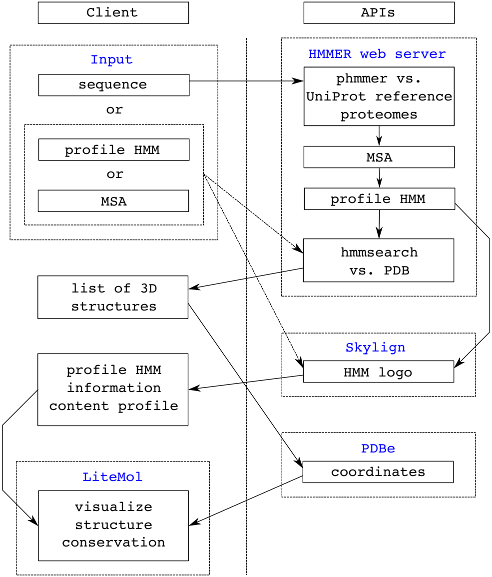


Figure 2: Profile HMM logo for a part of the `bZIP_1` (PF00170) Pfam profile HMM.

Introduction: back-end services and databases

- HMMER web server
    - Web front-end and interface (API) to HMMER suite programs
        - HMMER: package for sequence analysis using profile hidden Markov models (HMMs).
        - Profile HMM: statistical model of a MSA.
    - Single sequence (*phmmer*) and profile HMM (*hmmsearch*) searches against sequence databases.
- Skylign
    - Calculates sequence logo and per-position information content (IC) for a profile HMM query
        - $\mathrm{IC}(X) \approx$ relative entropy(X).
        - IC related to position conservation level.

3DPatch: how does it work?



Client

APIs

Input

sequence

or

profile HMM

or

MSA

HMMER web server

phmmer vs. UniProt reference proteomes

MSA

profile HMM

hmmsearch vs. PDB

list of 3D structures

Skylign

HMM logo

profile HMM information content profile

PDBe

coordinates

LiteMol

visualize structure conservation

3DPatch: video demonstration

3DPatch.avi

3DPatch: availability and implementation

- Client-side web application, pure JavaScript.
- 3DPatch URL: `http://skylign.org/3DPatch/`
- Mozilla Firefox, Google Chrome, Safari.
- Source code: `https://github.com/davidjakubec/3DPatch`
- MIT License.
- *Bioinformatics*.

3DPatch-tools

- Python scripts for performing residue-level IC calculation and structure mark-up locally.
- Goal: cache up-to-date annotations for large sequence/3D structure/profile HMM collections
    - PDBe, Pfam/InterPro profile HMMs, ...
    - ... but also your own.
- Source code: `https://github.com/davidjakubec/3DPatch-tools`
- MIT License.

Acknowledgements

- Jiří Vondrášek (IOCB AS CR)
- Rob Finn & Sequence Families team members (EMBL-EBI)
- Funding/support:

elixir

ÚOCHB AV ČR
IOCB PRAGUE

EMBL-EBI

Thank you for your attention.