

REFRACT

Update presentation

Carlos Vigil Vásquez

μ Platypus Lab, CABD. Sevilla, Spain.

15 of November, 2022

Introduction

ACE1



- ▶ Structural proteins, e.g., nuclear pores

Figure: ACE-1 yeast protein
Sec13/31

Introduction

ACE1



- ▶ Structural proteins, e.g., nuclear pores
- ▶ β propeller + α -solenoid

Figure: ACE-1 yeast protein
Sec13/31

Introduction

ACE1



- ▶ Structural proteins, e.g., nuclear pores
- ▶ β propeller + α -solenoid
- ▶ Head, trunk & tail

Figure: ACE-1 yeast protein
Sec13/31

Introduction

ACE1



- ▶ Structural proteins, e.g., nuclear pores
- ▶ β propeller + α -solenoid
- ▶ Head, trunk & tail
- ▶ ↓ **Conservation signal from amino acid sequence**
- ▶ ↓ **Conservation signal from structural alignment**

Figure: ACE-1 yeast protein
Sec13/31

Results

Foldseek optimization

Can we improve the understanding of this ACE1-like proteins?

Results

Foldseek optimization

Use global structural alignment as a proxy,

Results

Foldseek optimization

Use global structural alignment as a proxy, then refine interesting cases with local structural alignment.

Results

Foldseek optimization

Use global structural alignment as a proxy, then refine interesting cases with local structural alignment.

Methodology.

Fast-search with Foldseek, then refine with MOMA2



Results

Foldseek optimization

Use **global structural alignment as a proxy**, then refine interesting cases with local structural alignment.

Methodology.

Fast-search with Foldseek, then refine with MOMA2



Results

Foldseek optimization

Can Foldseek find ACE1-like proteins?

What is Foldseek?

Foldseek optimization

Foldseek is a novel structural alignment tool that "enables **fast and sensitive** comparisons of large structure sets.

¹*Foldseek: fast and accurate protein structure search*, 10.1101/2022.02.07.479398

What is Foldseek?

Foldseek optimization

Foldseek is a novel structural alignment tool that "enables **fast and sensitive** comparisons of large structure sets. It reaches **sensitivities similar to state-of-the-art**

¹*Foldseek: fast and accurate protein structure search*, 10.1101/2022.02.07.479398

What is Foldseek?

Foldseek optimization

Foldseek is a novel structural alignment tool that "enables **fast and sensitive** comparisons of large structure sets. It reaches **sensitivities similar to state-of-the-art** structural aligners while being **four to five orders of magnitude faster**"¹.

¹*Foldseek: fast and accurate protein structure search*, 10.1101/2022.02.07.479398

Results

Foldseek optimization

Optimized Foldseek's parameters for ACE1 protein search:

Results

Foldseek optimization

Optimized Foldseek's parameters for ACE1 protein search:

- ▶ Grid-search optimization

Results

Foldseek optimization

Optimized Foldseek's parameters for ACE1 protein search:

- ▶ Grid-search optimization
- ▶ Foldseek parameters *sensitivity* and *k-mer size*

Results

Foldseek optimization

Optimized Foldseek's parameters for ACE1 protein search:

- ▶ Grid-search optimization
- ▶ Foldseek parameters *sensitivity* and *k-mer size*
- ▶ Complete PDB database

Results

Foldseek optimization

Optimized Foldseek's parameters for ACE1 protein search:

- ▶ Grid-search optimization
- ▶ Foldseek parameters *sensitivity* and *k-mer size*
- ▶ Complete PDB database

Objective: Maximize performance based in finding target proteins with same UniProt accession code

Results

Foldseek optimization

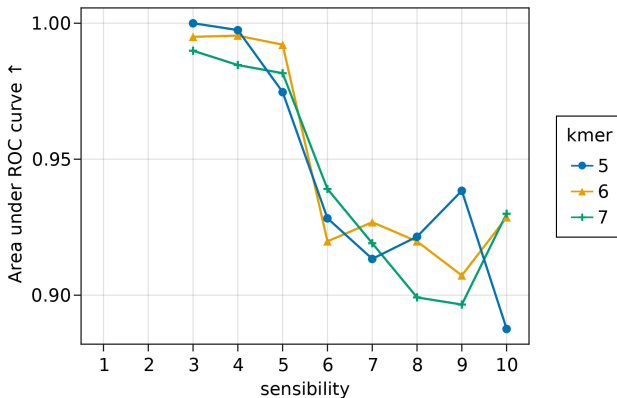


Figure: Parameter optimization for ACE1-like protein search based in AuROC

Results

Foldseek optimization

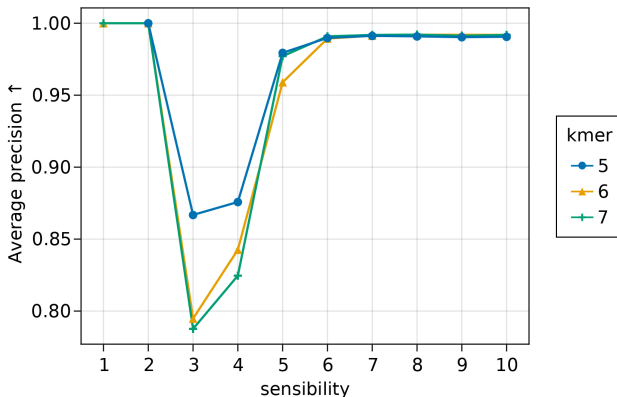


Figure: Parameter optimization for ACE1-like protein search based in AuPRC

Results

Foldseek optimization

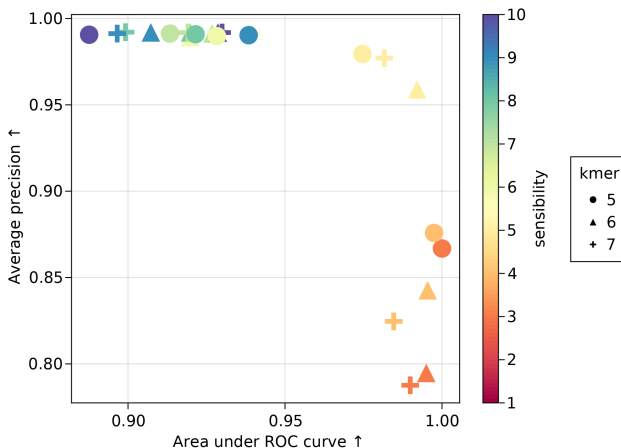


Figure: Parameter optimization for ACE1-like protein search

Results

Foldseek optimization

The optimal parameters for Foldseek for searching for ACE1-like proteins are:

- ▶ Sensibility of 5
- ▶ k-mer size of 5, 6 or 7

Results

Foldseek exploration

Can Foldseek find novel ACE1-like proteins?

Results

Foldseek exploration

Optimized Foldseek's parameters for ACE1 protein search:

Results

Foldseek exploration

Optimized Foldseek's parameters for ACE1 protein search:

- ▶ Grid-search optimization of *sensitivity* & *k-mer size*

Results

Foldseek exploration

Optimized Foldseek's parameters for ACE1 protein search:

- ▶ Grid-search optimization of *sensitivity* & *k-mer size*
- ▶ Complete GO terms used as protein description proxy

Results

Foldseek exploration

Optimized Foldseek's parameters for ACE1 protein search:

- ▶ Grid-search optimization of *sensitivity* & *k-mer size*
- ▶ Complete GO terms used as protein description proxy
- ▶ Jaccard index of GO terms to assess similarity between proteins

Results

Foldseek exploration

Optimized Foldseek's parameters for ACE1 protein search:

- ▶ Grid-search optimization of *sensitivity* & *k-mer size*
- ▶ Complete GO terms used as protein description proxy
- ▶ Jaccard index of GO terms to assess similarity between proteins

Objective: Minimize GO term similarity maintaining performance

Results

Foldseek exploration

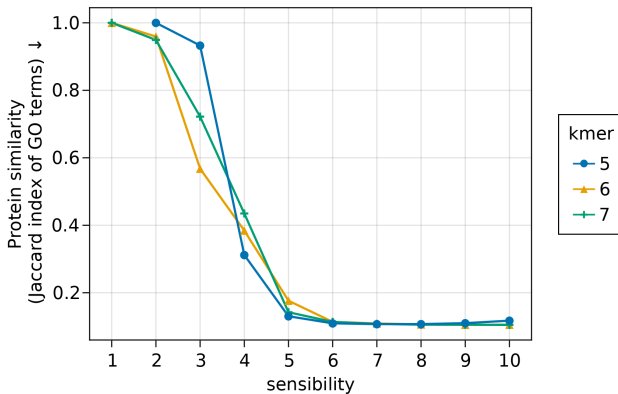


Figure: Foldseek novelty optimization

Results

Foldseek exploration

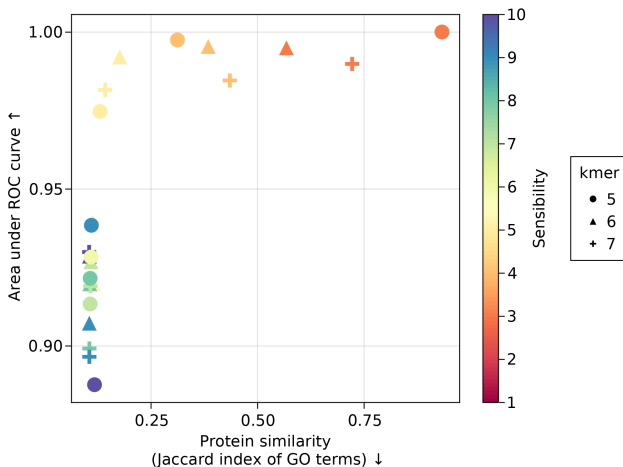


Figure: Foldseek novelty optimization as a function of AuROC

Results

Foldseek exploration

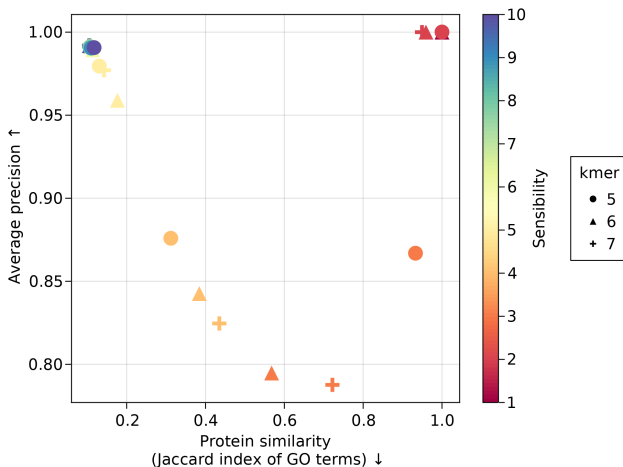


Figure: Foldseek novelty optimization as a function of AuPRC

Results

Foldseek exploration

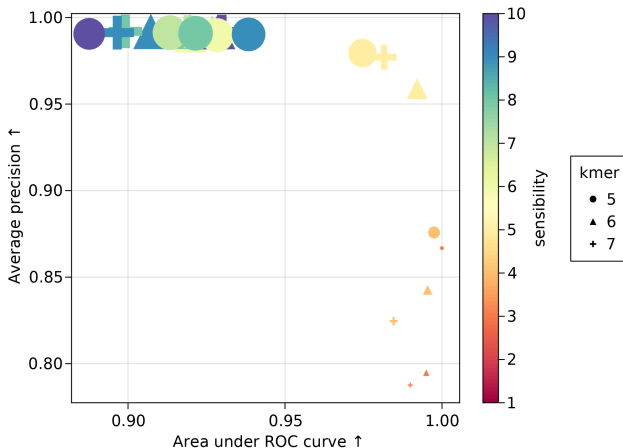


Figure: Foldseek novelty optimization as a function of AuROC and AuPRC