Prediction of RNA and DNA binding sites: preliminary presentation

Quirin Heiss[1,2] Pandu Raharja [1,2] Julian Schmidt [1,2]

[1]Technische Universität München

[2]Ludwig-Maximilians-Universität München

November 10, 2016

Outline

- CAFA.
- Background: RNA and RNA binding proteins.
- Background: datasets.
- Project plan.

Definition

- **RNA/DNA Binding Protein prediction**: given a protein, determine whether a protein is RNA/DNA binding.
- **RNA/DNA binding site prediction**: given a protein sequence, determine side chains that bind with a DNA/RNA.

Methods: abstraction level

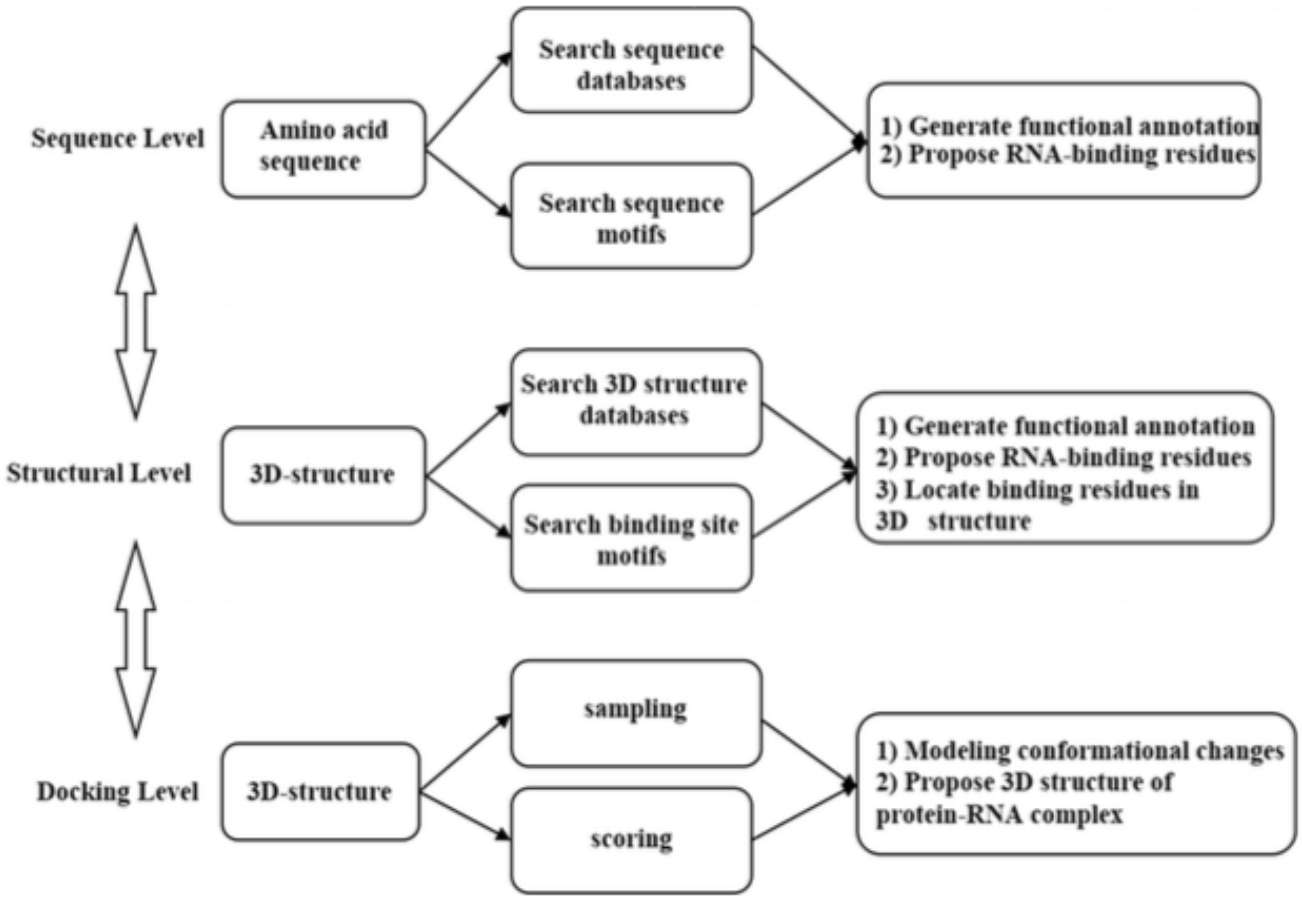

Figure 1. Strategies for RNA-binding site and RBP prediction.

1

[1] Si, Wu *et al.*, 2015, *Int. J. Mol. Sci.*

Possible Features

Sequence-based features:

- **Amino acid composition**.
- **Sequence similarity,** such as MSA.
- **Evolutionary invormation,** such as PSSM.
- **Evolutionary invormation,** such as PSSM.

Structure-based features:

- **Secondary structure:** experimental (assigned using DSSPcont) or predicted.
- **Accessible surface area,** in percnent (%).

Chemical and physical features:

- **Hydrophobicity.**
- **Electrostatic patches.**
- **Cleft Size.**

Methods: abstraction level

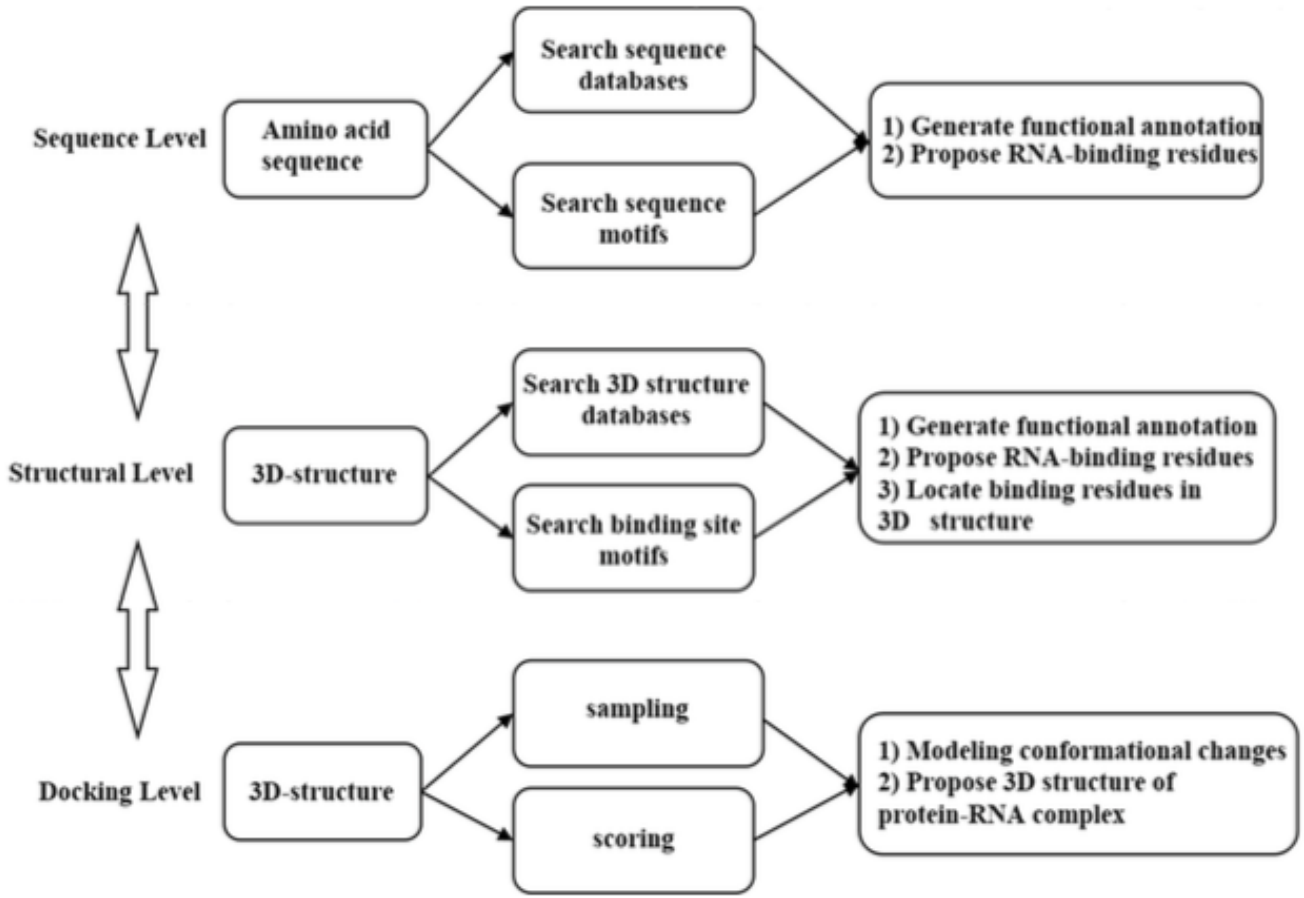

Figure 1. Strategies for RNA-binding site and RBP prediction.

[2]

[2] Si, Wu *et al.*, 2015, *Int. J. Mol. Sci.*

Methods: previously used algorithm

- **Naive Bayes (NB)** classifier.
- **Support Vector Machine (SVM)**.
- **Random Forest**.
- **Neural Network (NN)**.