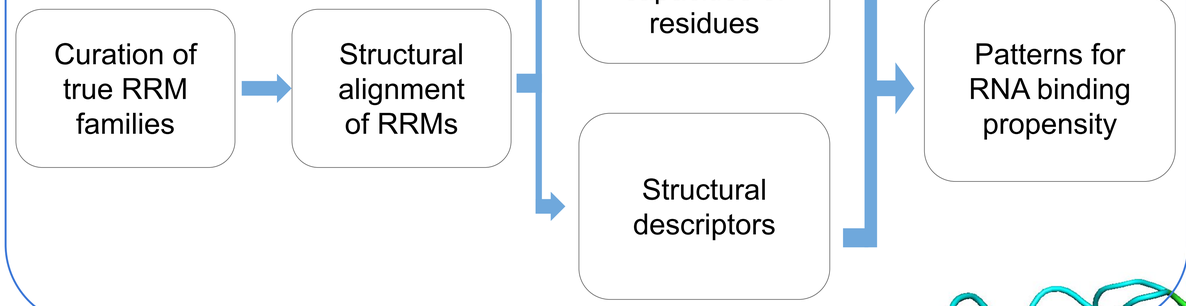
Data-driven identification of structural patterns associated with RNA binding in RNA Recognition Motif domains

Hrishikesh Dhondge[1], Joel Roca Martínez[2], Wim Vranken[2], Marie-Dominique Devignes[1], Isaure Chauvot de Beauchene[1]
1 Université de Lorraine, CNRS, Inria, LORIA, F-54000 Nancy, France
2 Interuniversity Institute of Bioinformatics in Brussels, ULB-VUB, Brussels, Belgium

Abstract

Proteins containing one or more RNA Recognition Motif (RRM) domains perform important regulation functions through binding to specific RNA sequences. The European RNAct project consists iof designing new RRM proteins for therapeutic purposes. Our task is to provide data-driven rules and knowledge, based on the 3D structures of RRM domains. Protein domain classifications were screened to identify the most comprehensive set of RRM families for which one or more 3D structures are available. We created a master alignment for all available RRM 3D structures using Pfam (PF00076) seed alignment and the Kpax tool. Amino-acid properties and structural descriptors like torsion angles or accessibility were linked to the nucleotide binding capacity at each position, and compared in RNA bound versus unbound structures. The extracted patterns will help to predict RNA binding propensity at the domain and residue level, and to model the 3D structure of the complex. Such patterns will be tested in vitro/vivo by the RNAct partners.

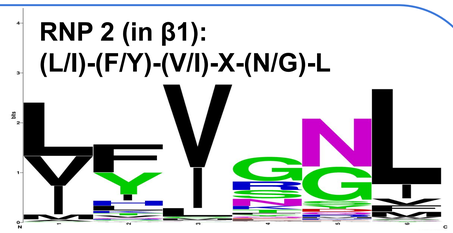
Workflow



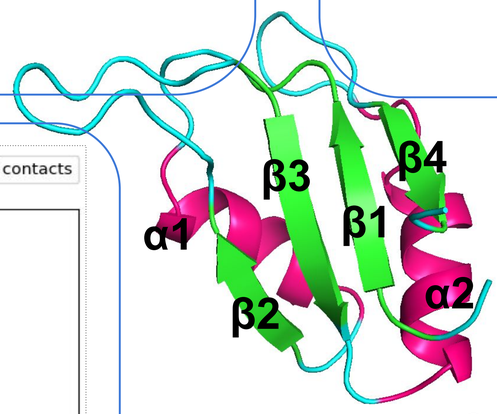
Curation and alignment

- 19 True RRM families from Pfam (Structural alignment and RNP positions)
- 1274 RRM structures from 218 protein sequences
- 289 unique RRM structures

RNP 2 (in β1):
(L/I)-(F/Y)-(V/I)-X-(N/G)-L

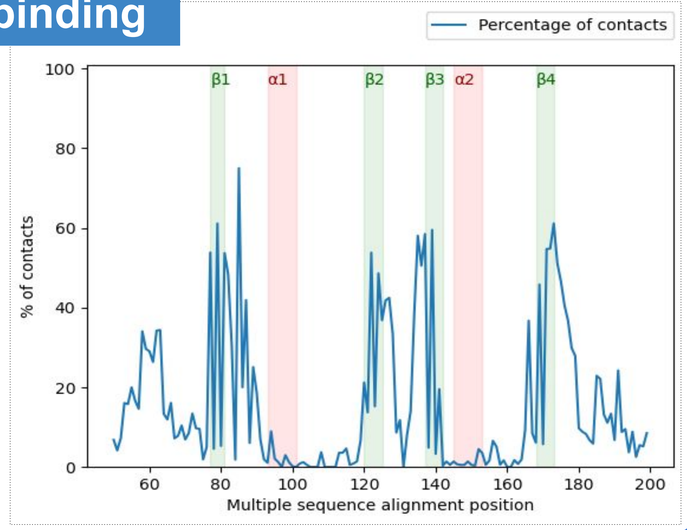


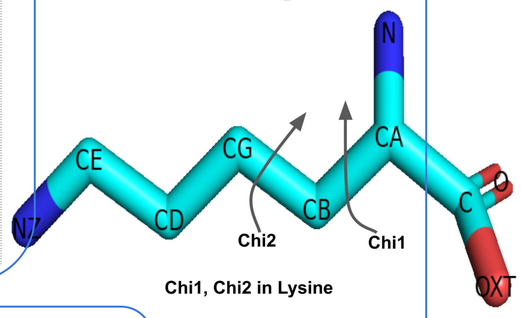
RNP 1 (in β3):
(R/K)-G-(F/Y)-(G/A)-(F/Y)-V-X-(F/Y)



Nucleotide binding

- The contact percentage from 578 structures
- Contact threshold: < 5 Å distance
- RNA binds mostly on the β-strands
- Some loops are involved in binding

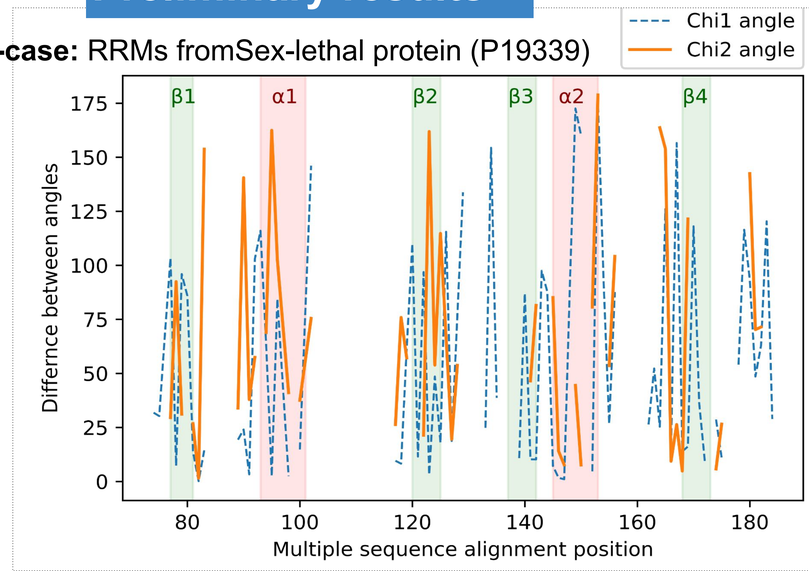




Chi1, Chi2 in Lysine

Preliminary results

Test-case: RRMs fromSex-lethal protein (P19339)



- Higher the difference higher the variability in side-chain angles
- No value represents gaps in the alignment or no side-chain at that position

Ongoing work

- Finding patterns between different descriptors and amino acid properties that will help to predict RNA binding propensity at the residue and domain level.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 813239.