Project Acronym:	**IDPfun**
Project Full Title:	**Driving functional characterization of intrinsically disordered proteins**
Grant Agreement:	**778247**
Project Duration:	**48 months (01/03/2018 - 28/02/2022)**

Deliverable D1.2

Software for automatic detection of IDRs and linear motifs from protein structures

Work Package:	**WP1**
Lead Beneficiary:	**UCD**
Due Date:	**30 November 2019 (M6)**
Submission Date:	**30 November 2019 (M6)**
Deliverable Type:	**D**
Dissemination Level:	**P**

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 778247

Table of Content

Executive summary 3

List of Acronyms 4

Project overview 5

Availability 6

Executive summary

This document describes the *"D1.2 Software for automatic detection of IDRs and linear motifs from protein structures"* for the IDPfun project. The deliverable was a software pipeline for the automatic extraction of X-ray derived IDR information (missing electron densities), linear motifs and consensus generation. The pipeline was completed on time and as described.

List of Acronyms

Acronym	Definition
IDD	Intrinisically Disordered Domains
SLiM	Short Linear Motif

Project overview

Intrinsically disordered regions contain numerous functional modules that mediated key cellular functions. These function modules are known as
Intrinisically Disordered Domains (IDDs) and Short Linear Motifs (SLiMs). Many interactions mediated by the IDDs and SLiMs within the disordered regions of proteomes have been captured in complex with their binding partners. These interactions generally involve linear interactions of the disordered regions with a pocket or pockets on a globular binding partner.

These disorder-globule complexes have two major properties that can be used to identify them from other classes of interaction interface: (i) The disordered binding partner will have a high ratio of intramolecular interactions to intermolecular interactions; (ii) The globular binding partner will have high levels of intramolecular interactions. The tool uses these properties to parse, classify and annotate pdb structures based on the structural properties of the interactors. The tool takes as an input a PDB identifier and outputs the disordered interface-containing protein, disordered interface start and end, disordered interface-binding protein, disordered interface-binding domain. The tool also provides detailed motif interface information including atomic resolution details of the contacts within the interface to generate a binding consensus.

instances-The tool was benchmarked on the ELM database using a collection of 448 manually curated ELMcontaining PDB structures. The tools collections ran without issue on the dataset. The results were as follows:
- 59 structures did not return a motif:

- 39 structures returned an did not return any interface information. In all cases the output of the tools was correct. 3 Entry is marked as obsolete and 36 contained only a single entity in the structure and were therefore not a complex.
- 20 structures found the interface but incorrectly classified the in interfaces class as containing no motif. In these cases, the motif-containing chain have a high number of intramolecular contact and fall below the threshold for recognition of a disordered binding partner in a structure.

- 389 structures returned a motifs

- In total, there are 582 (Motif, Elm instance) pairs to compare (Most of them are 1 to 1
- 510 returned the correct ELM instance.
- 61 returned the correct ELM instance but the protein mapping was different to ELM.
- 11 Motif region do not overlaps with ELM instance.

The tool is built in Python using BioPython for PDB file parsing. The tools has been containerised as a Docker image and can be run as a local or web-based webservice using the python FLASK web framework.

Availability

The project source code is available at: *https://gitlab.com/idpfun/rise-find-interfaces/*