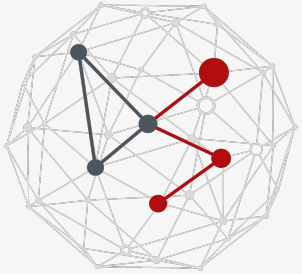






DATA SCIENCE
UNIVERSITY OF PADOVA

# Residue Interaction Network Generator
# - RING -

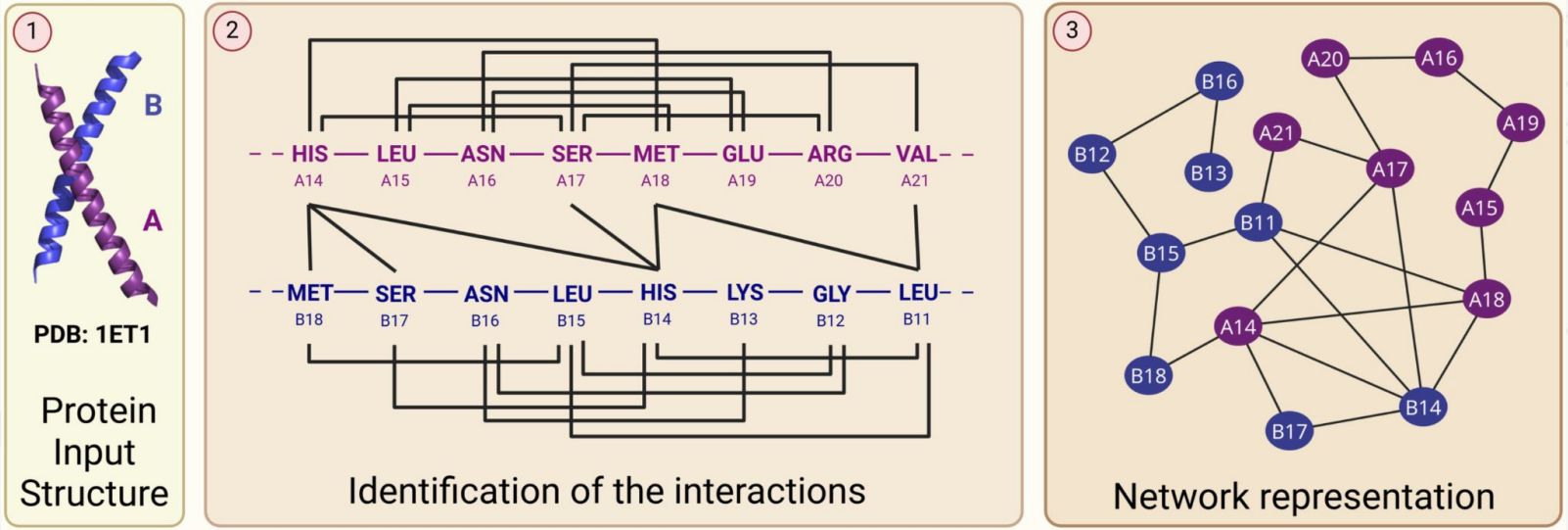
Master of Science in Data Science

Damiano Piovesan

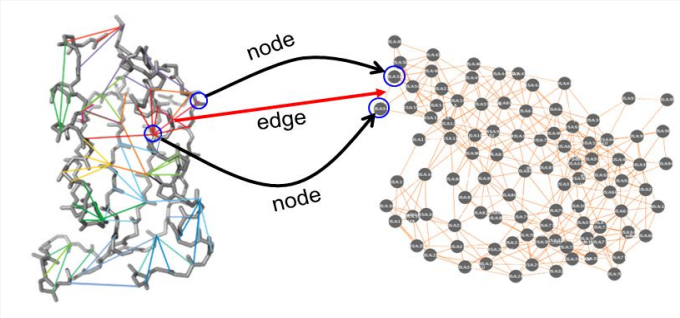
# Concept

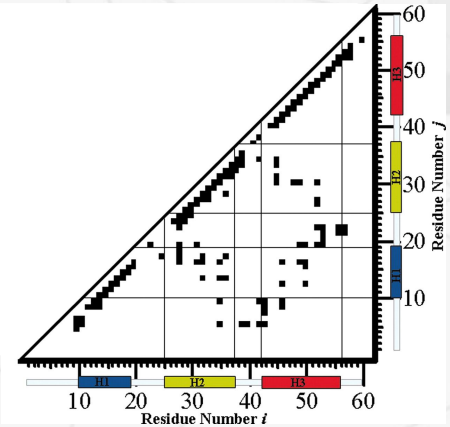
A graph representation of a protein structure, where:

- Residues → Nodes
- Interactions → Edges

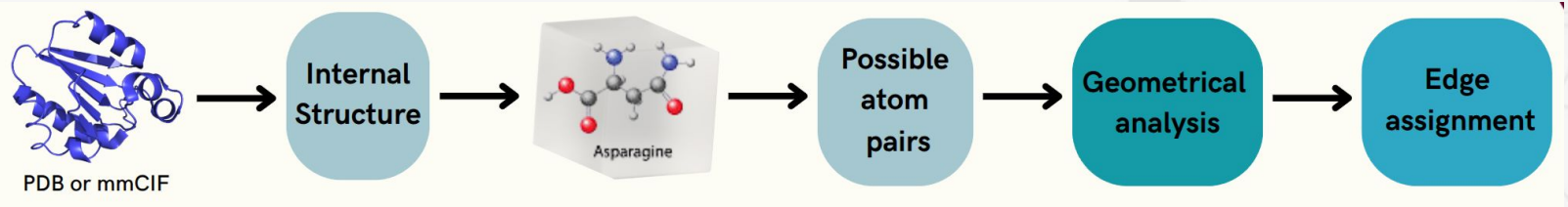


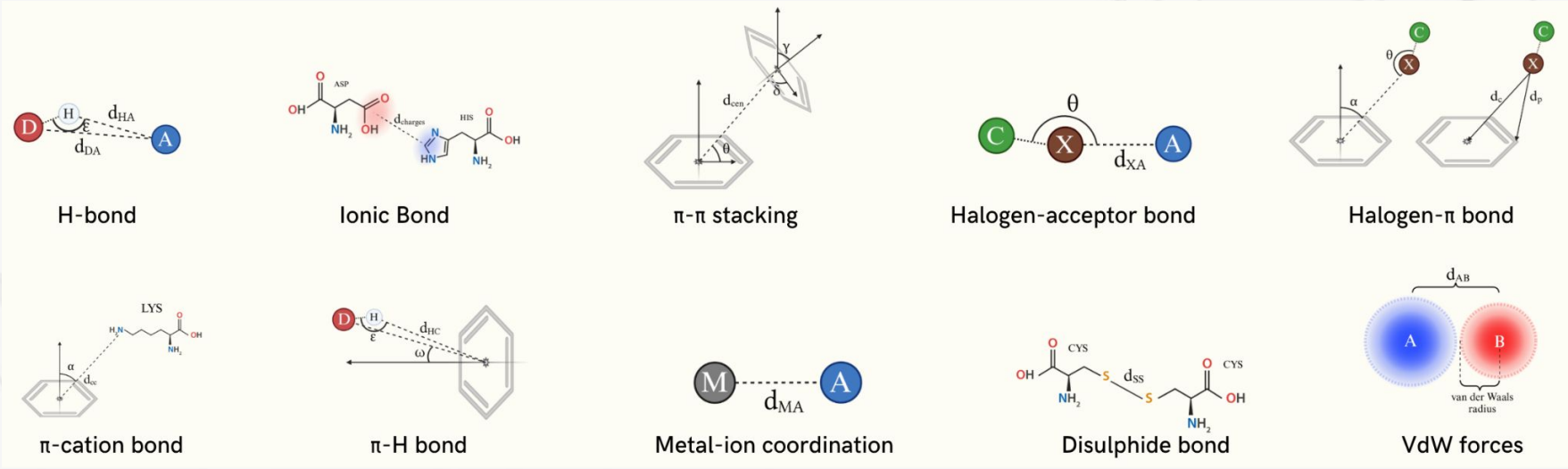
# RIN and contact maps



# Possible interactions



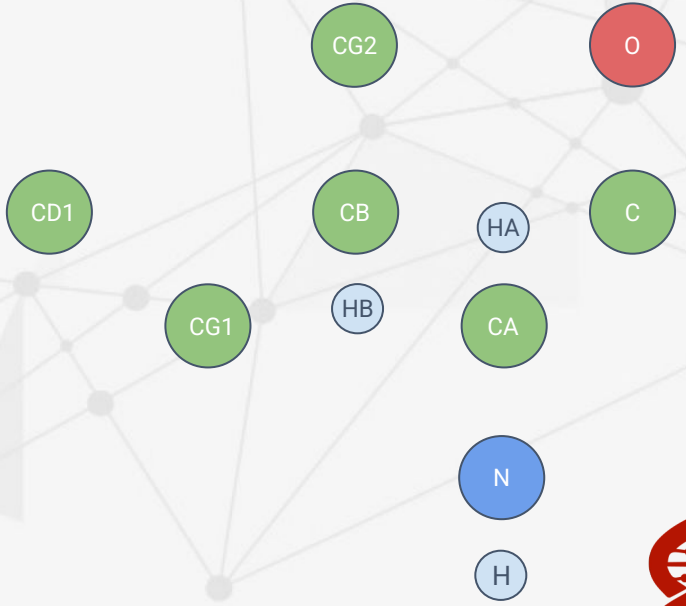



DIPARTIMENTO MATEMATICA

# How to build the protein structure

| id | type_symbol | atom_id | comp_id | asym_id | seq_id | | x | y | z |
|---|---|---|---|---|---|---|---|---|---|
| ATOM | 61 | N | N | . | ILE | A | 1 | 5 | ? | -28.199 | -33.564 | -15.366 |
| ATOM | 62 | C | CA | . | ILE | A | 1 | 5 | ? | -27.521 | -32.693 | -16.369 |
| ATOM | 63 | C | C | . | ILE | A | 1 | 5 | ? | -26.502 | -31.786 | -15.674 |
| ATOM | 64 | O | O | . | ILE | A | 1 | 5 | ? | -26.618 | -31.493 | -14.501 |
| ATOM | 65 | C | CB | . | ILE | A | 1 | 5 | ? | -28.641 | -31.864 | -16.998 |
| ATOM | 66 | C | CG1 | . | ILE | A | 1 | 5 | ? | -28.041 | -30.881 | -18.008 |
| ATOM | 67 | C | CG2 | . | ILE | A | 1 | 5 | ? | -29.379 | -31.084 | -15.908 |
| ATOM | 68 | C | CD1 | . | ILE | A | 1 | 5 | ? | -27.509 | -31.651 | -19.217 |
| ATOM | 69 | H | H | . | ILE | A | 1 | 5 | ? | -28.909 | -33.198 | -14.799 |
| ATOM | 70 | H | HA | . | ILE | A | 1 | 5 | ? | -27.038 | -33.293 | -17.123 |
| ATOM | 71 | H | HB | . | ILE | A | 1 | 5 | ? | -29.335 | -32.521 | -17.502 |
| ATOM | 72 | N | N | . | GLN | A | 1 | 6 | ? | -25.505 | -31.340 | -16.389 |

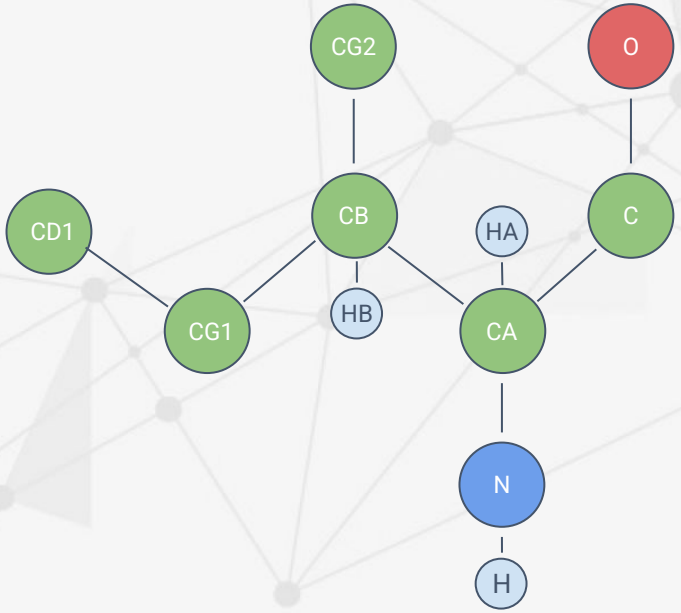
Isoleucine ?



# How to build the protein structure

Connect the atoms using the
**chemical component dictionary - CCD**

```
RESIDUE    ILE
CONECT       N      CA    H
CONECT       CA     N     C     CB    HA
CONECT       C      CA    O
CONECT       O      C
CONECT       CB     CA    CG1   CG2   HB
CONECT       CG1    CB    CD1
CONECT       CG2    CB
CONECT       CD1    CG1
CONECT       H      N
CONECT       HA     CA
CONECT       HB     CB
END
FORMUL       ILE     C6 H13 N O2
```



# How to build the protein structure

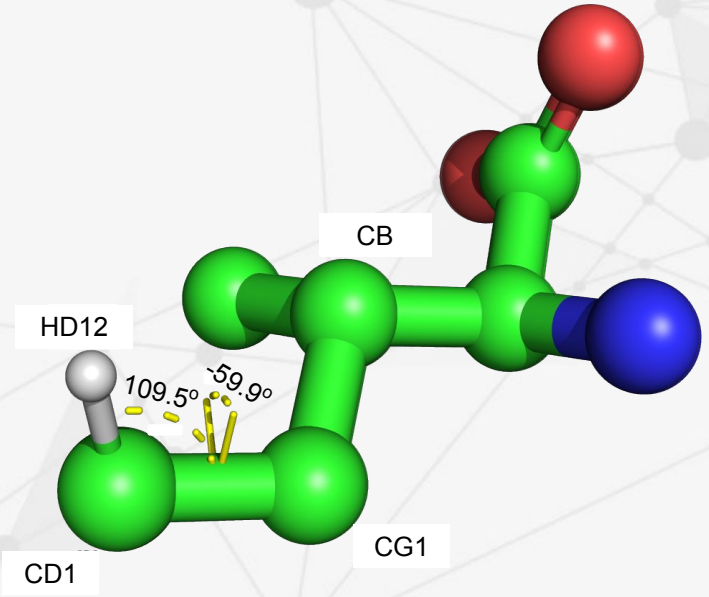
DIPARTIMENTO MATEMATICA

Add the hydrogens

```
ILE
...
HD12 CD1 CG1 109.5 CB -59.9
...
```

The algorithm converts the internal representation (z-matrix) to cartesian with some trigonometry formulas based on

- Bond length
- Bond angle
- Dihedral angle



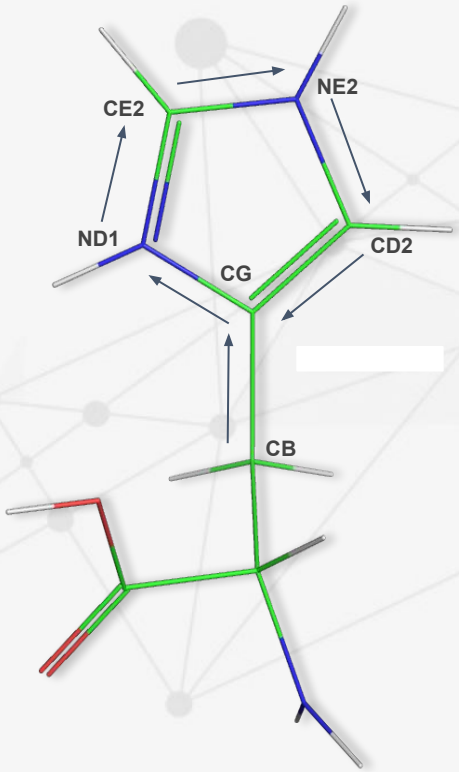
# How to build the protein structure

DIPARTIMENTO MATEMATICA

Find aromatic rings (Depth First Search)

| Visited |
| --- |
| [CB] |
| . . . |
| [CB, CG, ND1, CE2, NE2, CD2, CG] |

Histidine




Additional filters

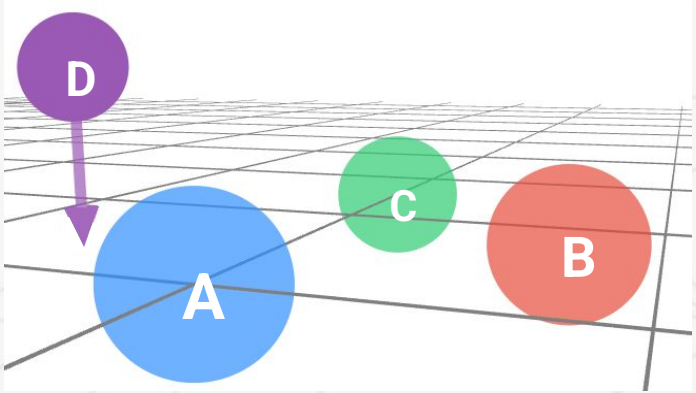
- **Coplanarity** of ring atoms using the dot product
- Atoms names in extracted dictionary from the **CCD**

# How to build the protein structure

DIPARTIMENTO MATEMATICA

Find aromatic rings (Depth First Search)

| Visited |
| --- |
| [CB] |
| . . . |
| [CB, CG, ND1, CE2, NE2, CD2, CG] |




Additional filters

- **Coplanarity** of ring atoms using the dot product
- Atoms names in extracted dictionary from the **CCD**

# Calculate the interactions

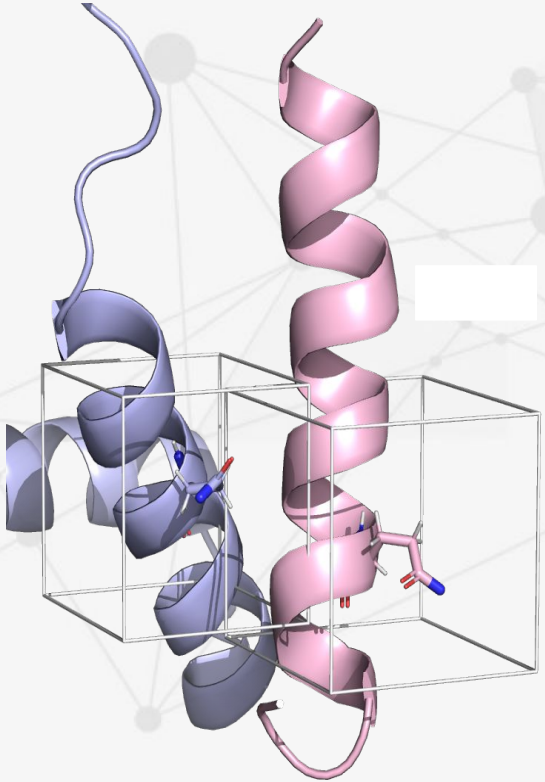
DIPARTIMENTO MATEMATICA

**An all vs all calculation?**

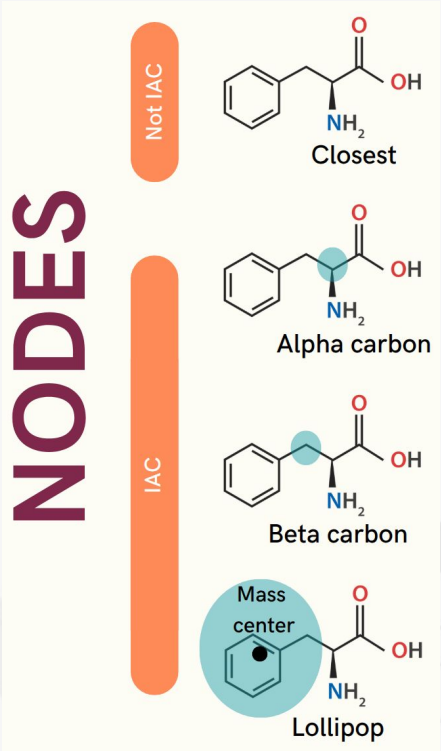
Interactions have a maximum range (to be relevant). We can use this knowledge and the structural information to skip a lot of calculations and **reduce runtime**

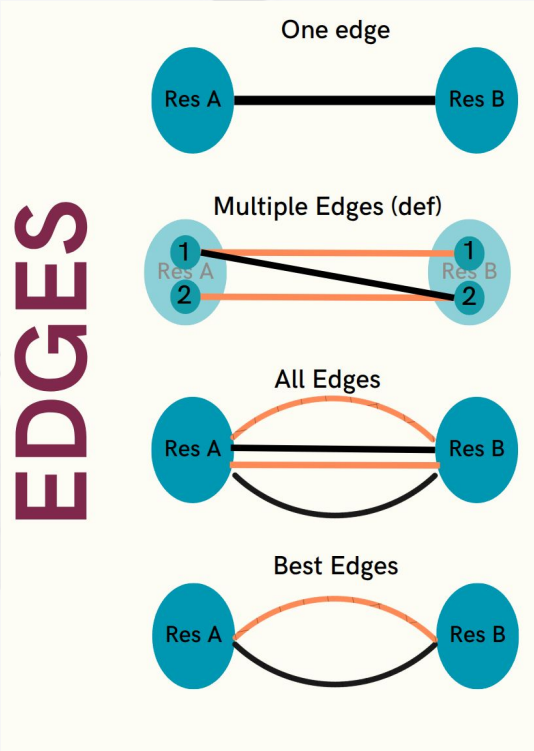
Use bounding boxes

- At chain level
- Residue level



# Nodes definition and edge cardinality



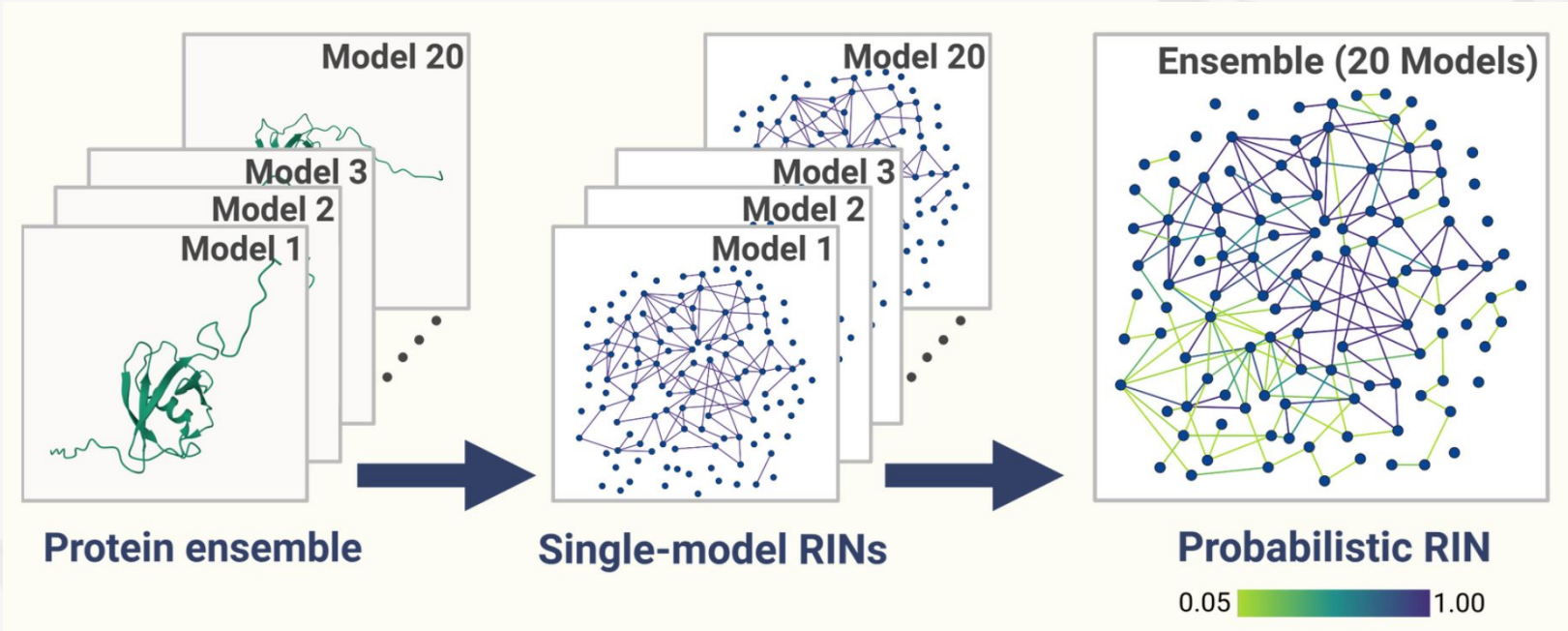
DIPARTIMENTO MATEMATICA

# Contacts in high-resolution PDB structures

| | Intra-chain | | | | Inter-chain | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Interaction partners | AA-AA | | AA-Ligand^ | | AA-AA | | AA-Ligand | | AA-Nucleic acid | |
| Number of PDBs | 30 244 | | 3166 | | 12 605 | | 25 062 | | 462 | |
| Number of contacts | 9 431 609 | | 33 074 | | 300 687 | | 324 296 | | 9311 | |
| Hydrogen bond | 8 442 301 | 89.5% | 31 410 | 95.0% | 229 295 | 76.3% | 229 855 | 70.9% | 8517 | 91.5% |
| π-π stacking | 620 405 | 6.6% | 1475 | 4.5% | 36 301 | 12.1% | 33 175 | 10.2% | 608 | 6.5% |
| π-cation | 87 262 | 0.9% | 102 | 0.3% | 7278 | 2.4% | 2836 | 0.9% | 180 | 1.9% |
| Ionic bond | 234 741 | 2.5% | – | – | 25 025 | 8.3% | – | – | – | – |
| Disulfide | 23 526 | 0.2% | – | – | 890 | 0.3% | – | – | – | – |
| π-hydrogen bond* | 23 374 | 0.2% | 35 | 0.1% | 1898 | 0.6% | 962 | 0.3% | 6 | 0.1% |
| Halogen bond* | – | – | 43 | 0.1% | – | – | 831 | 0.3% | – | – |
| Metal ion coordination* | – | – | – | – | – | – | 56 637 | 17.5% | – | – |

Del Conte et al.
**RING 4.0: faster residue interaction networks with novel interaction types across over 35,000 different chemical structures**
NAR

# Probabilistic RIN



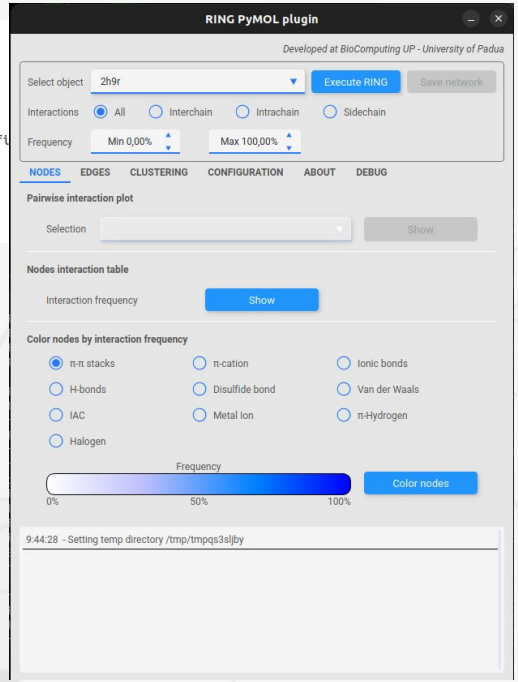
DIPARTIMENTO MATEMATICA

# How to use RING

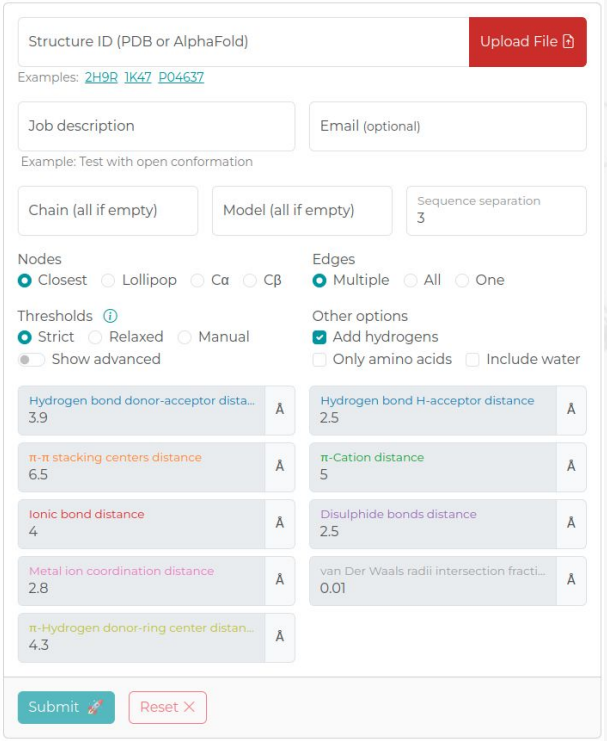
## CLI

```
❯ ring -h

RING $ -- Residue Interaction Network Generator - Version v4.0-4-gfca5770


Input options:
  -i  <filename>   Input PDB/mmCIF file
  -I  <filename>   Input text file with a list of PDB/mmCIF files, absolute path needs to be specifi
 --all_models      Executes RING on all models (default false)
  -m <number>      Model number to use (default first model)
  -c  <id>         Chain identifier to read (default all chains)

Output options:
 --out dir <dirname> Output node and edges to the selected dir. (default input dir)
 --export_structure  Export the internal structure representation of RING to mmCIF/PDB

Network options:
  -n <string>       Network policy to define a contact between 2 groups/residues
                    Options: closest, lollipop, ca, cb (default closest)
  -t <number>       Network distance threshold for the specified network type (default 5.0)
  -g <number>       Sequence separation (gap) between residues (default 3)
 --get_iac          Include generic Inter-Atomic Contacts (IAC) (default false)
 --no_het           Skip calculation of hetero atoms connections (default false)
 --water            Include water molecules (default false)
 --no_specific      Skip calculation of bond specification (default false)
 --energy           Calculates TAP and FRST potentials for each residue, slower. (default false)
 --no_add_H         Skip the addition of H atoms to amino acids and nucleotides (default false)

Edges options:
 default:           Return multiple edges for a pair of nodes, those that should be the most int
 --all_edges        Return all edges found for a pair of nodes. (default false)
 --best_edge        Return only the most valuable connection between two nodes (default false)
```

## PyMOL plugin




## Web server



DIPARTIMENTO MATEMATICA

# Command Line Interface

## Geometric parameters

- Strict
- Relaxed
- Manual
  - Distance thresholds
  - Angles

## Additional settings

- Gap between residues (default 3)
- IAC?
- Water molecules?
- Hidrogen adding (default True)

## OUTPUT FILES

### RING representation

- Plain text file. Similar to the original PDB input file, but containing the structure as used by RING.
- Can be opened with Pymol for visualization.
- Useful for checking hydrogen placement.

### Nodes file

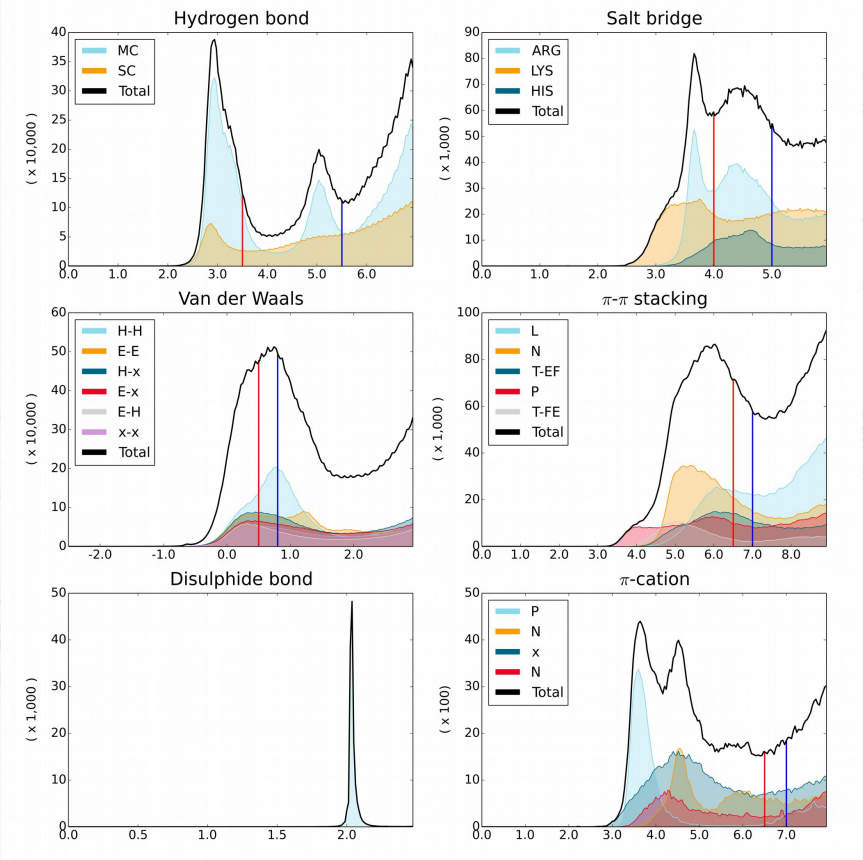
- Plain text file. Contains information about the nodes.
- Organized in columns: NodeId, Chain, Position, Residue, Type, DSSP, **Degree**, Bfactor_CA, Coordinates, pdbFileName, Model.

### Edges File

- Plain text file. Contains information about the edges.
- Indicates: interacting nodes and atoms, interaction type, geometrical parameters, model number.

# Distance cutoff

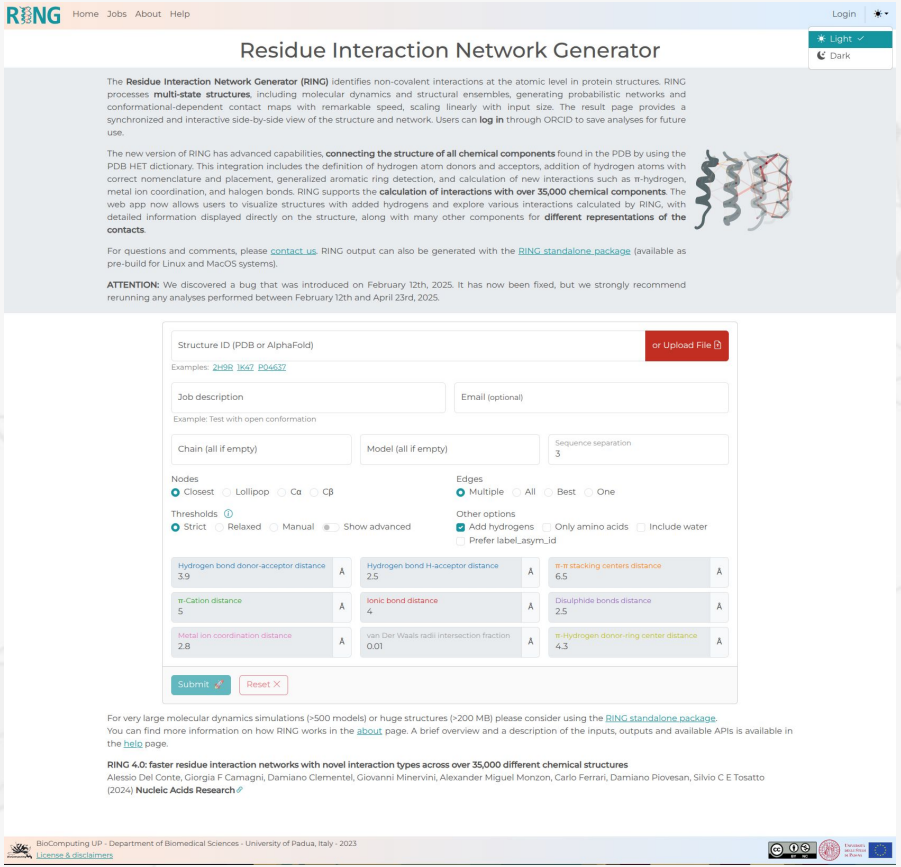
- Relaxed
- Strict (default)



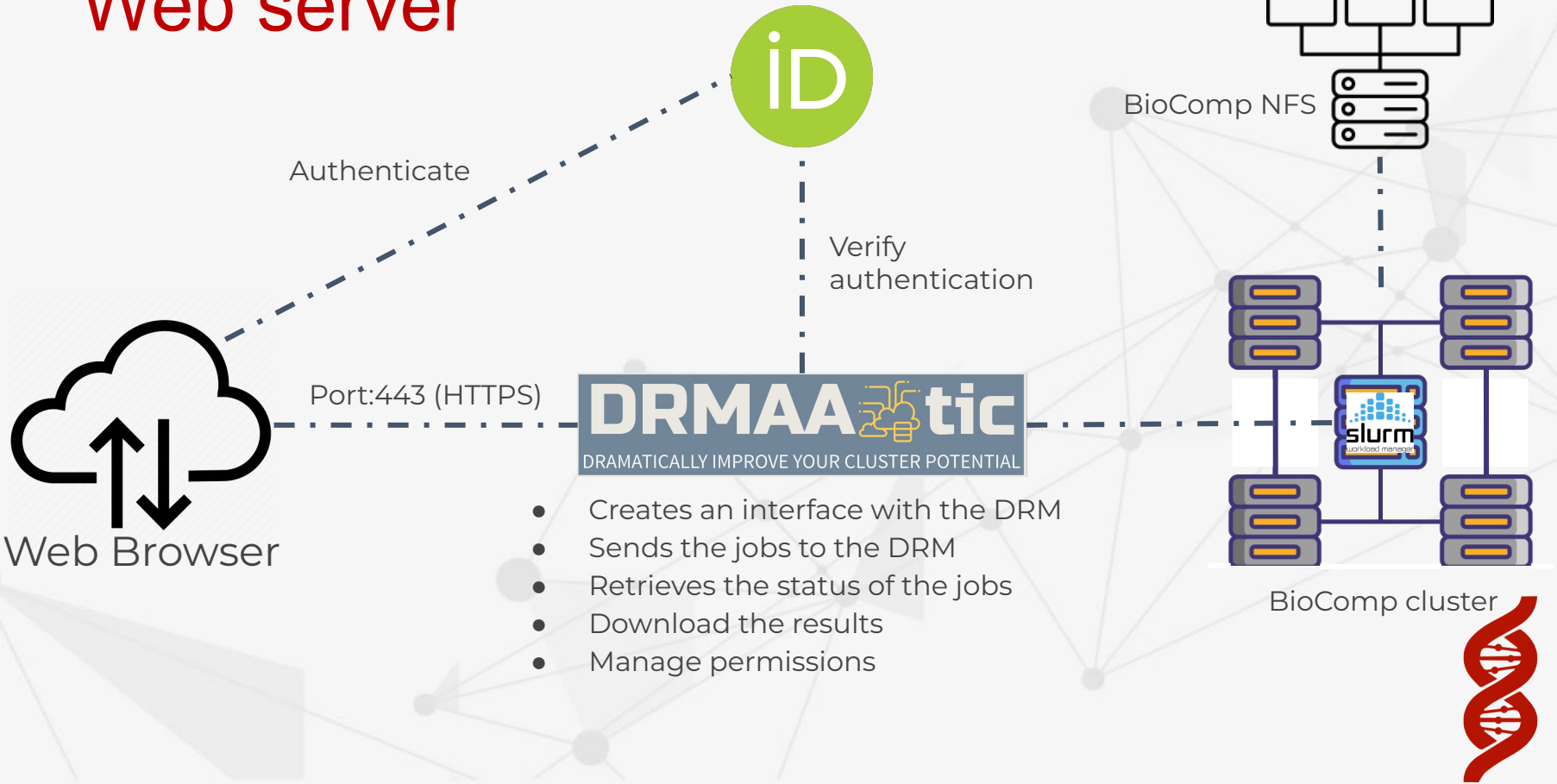

*The RING-2.0 web server for high quality residue interaction networks.* Piovesan D, Minervini G, Tosatto, SCE. **NAR**, 2016, 44(W1), W367-74

# Web server

https://ring.biocomputingup.it/



# Web server




- Creates an interface with the DRM
- Sends the jobs to the DRM
- Retrieves the status of the jobs
- Download the results
- Manage permissions

DIPARTIMENTO MATEMATICA

# RING-PyMol (not working with latest PyMol)

Repo and Wiki

https://github.com/BioComputingUP/ring-pymol

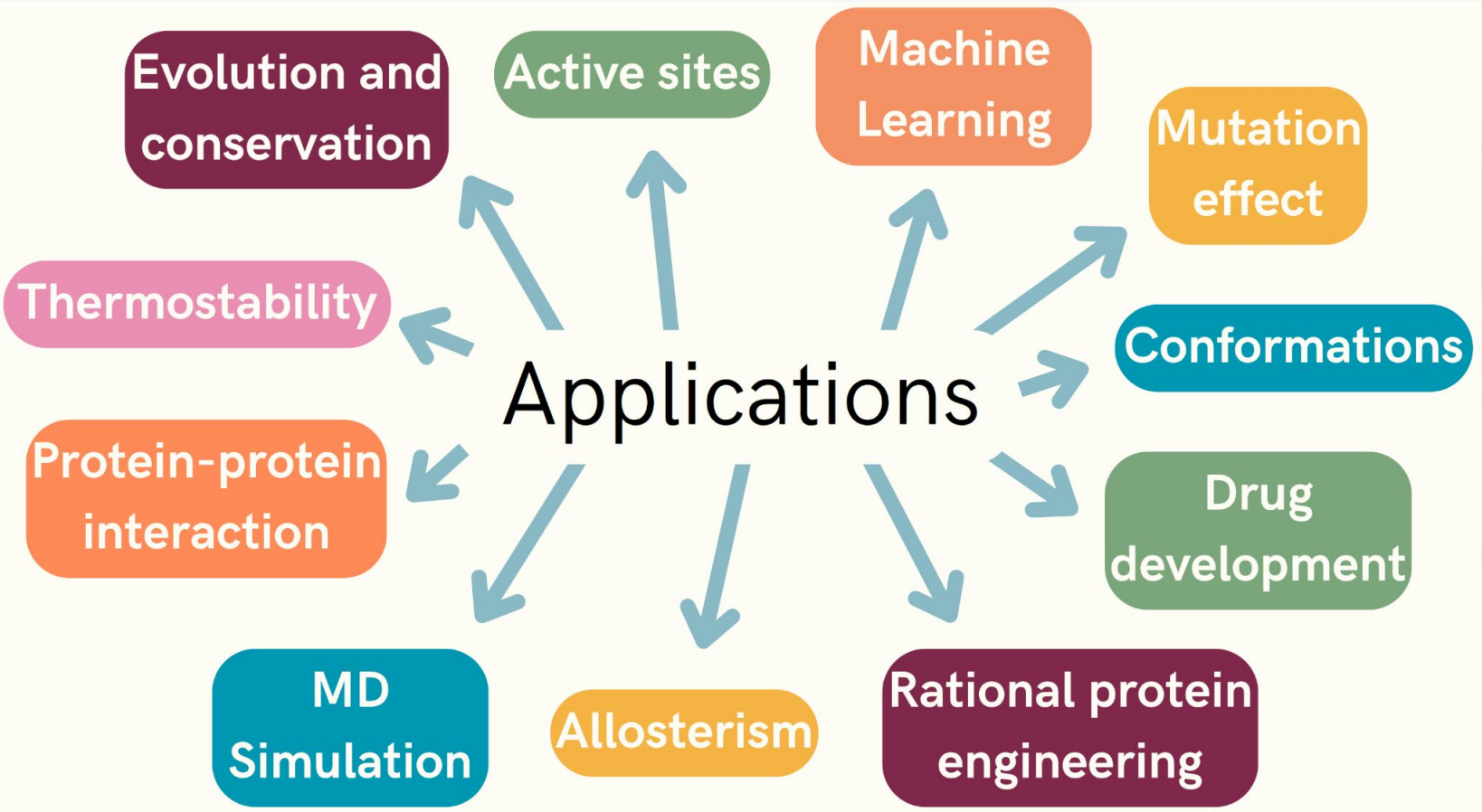
Plugin code archive

https://biocomputingup.it/shared/ring-plugin/

DIPARTIMENTO MATEMATICA

# Which is best?

| Local RING | Web Server |
| --- | --- |
| ✓ FAST | ✗ Slow(er) |
| ✓ Great for very large structures and huge ensembles | ✗ Not suitable for very large structures and huge ensembles |
| ✓ Great for large scale analysis | ✗ Not suitable for large scale analysis |
| ✗ Limited output for single-model analysis | ✓ Comprehensive output |
| ✗ Not very user friendly | ✓ USER FRIENDLY |
| ✗ Network can't be directly imported to Cytoscape | ✓ Network can be directly imported to Cytoscape as a JSON file |


Applications
Evolution and conservation
Active sites
Machine Learning
Mutation effect
Thermostability
Conformations
Protein-protein interaction
Drug development
MD Simulation
Allosterism
Rational protein engineering