

Image Processing in cryoEM:

Introduction to Scipion

José-Maria CARAZO lab

Centro Nacional de Biotecnología-CSIC

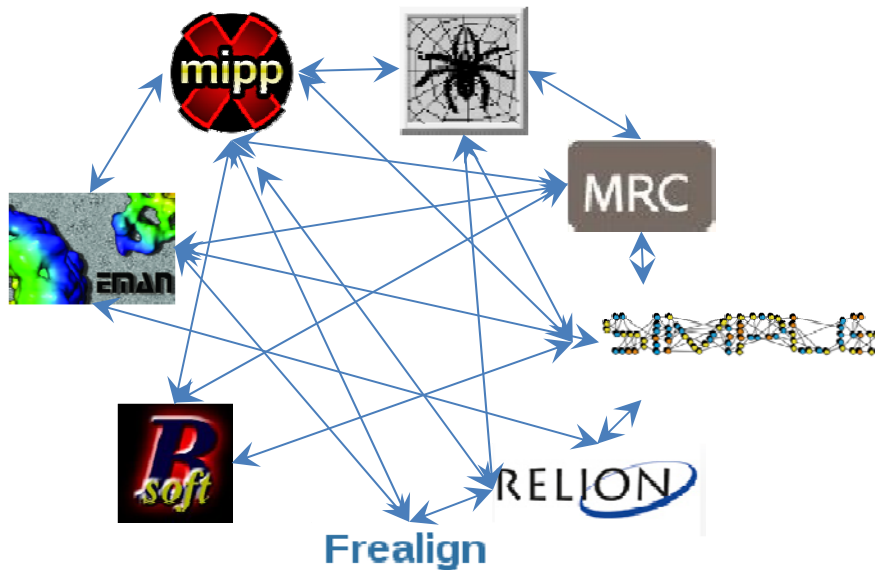
Instruct Image Processing Center

carazo@cnb.csic.es



The EM field needs software integration

Using different EM software packages may be difficult, specially to have a clear “reproducibility track record”





SCIPION

A Software Framework
for Integration, Reproducibility
and Validation in Cryo-EM



Scipion Principles

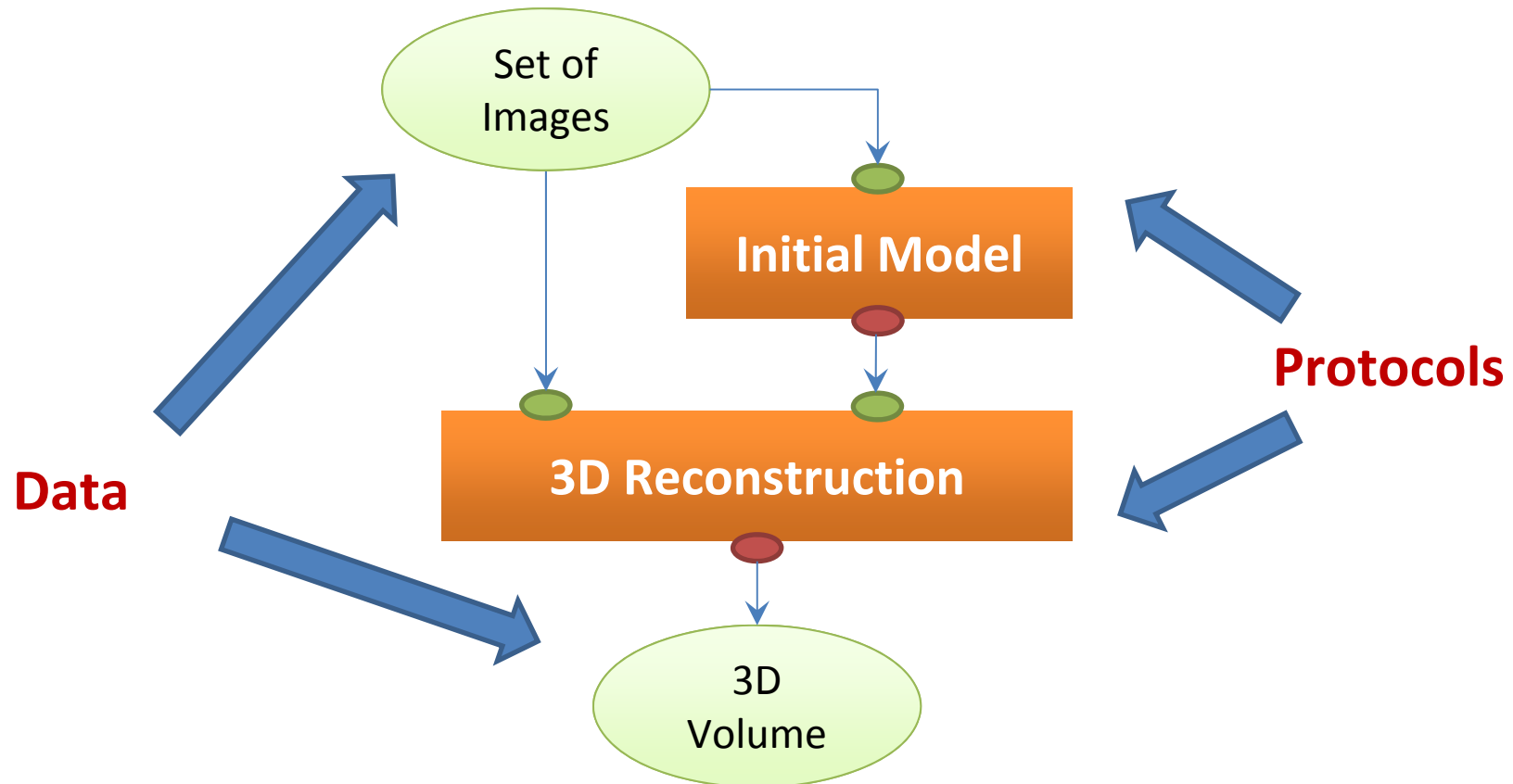
1. **Integrate** EM software packages to be used in the same project.
2. Full project traceability, improving **reproducibility**.
3. Execute complete workflows in an **automated** manner.
4. Local and **Cloud** deployment
5. Easy to install and use. Easy to extend with new protocols.



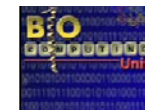
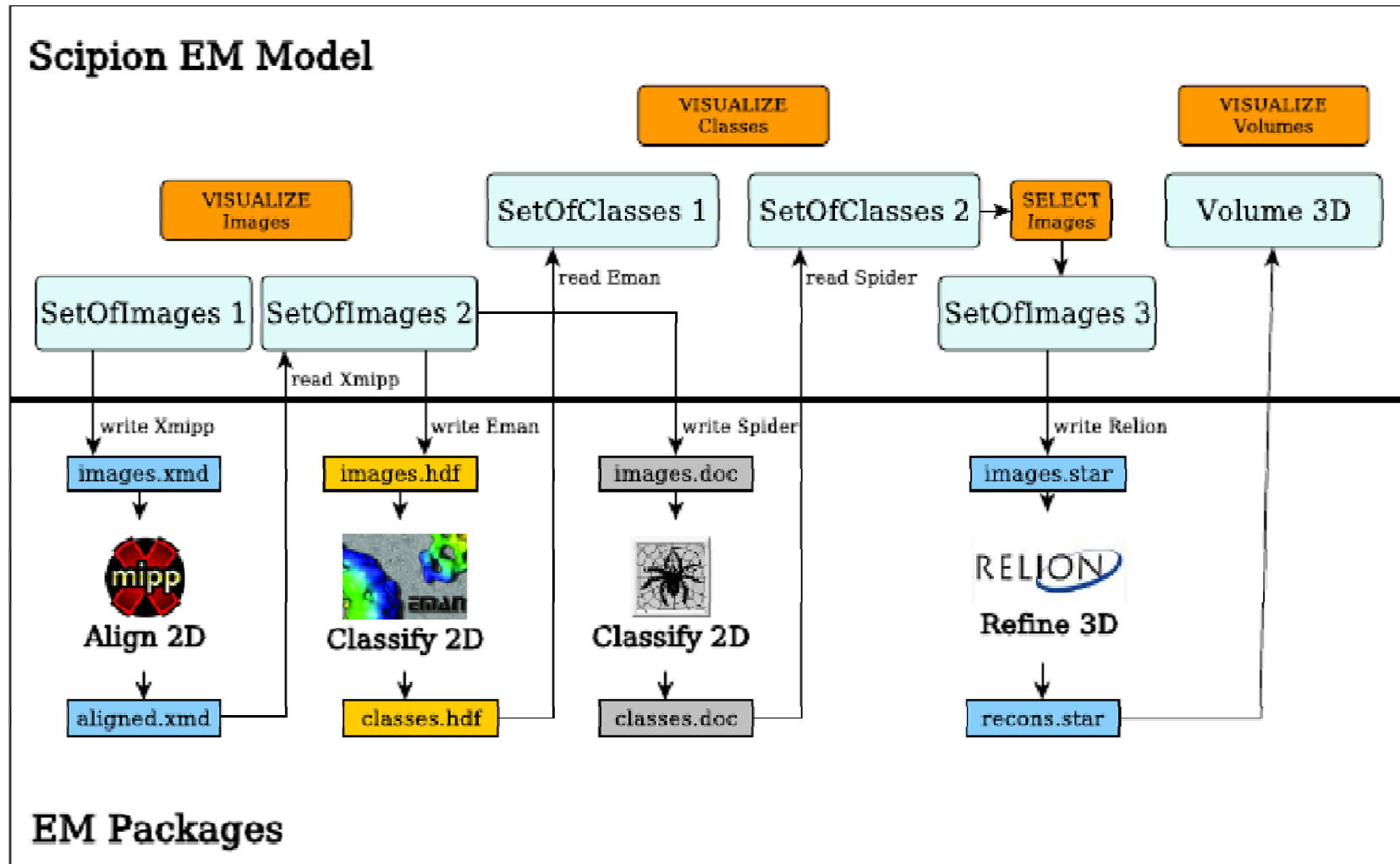
Goal 1: Integrate EM software packages to be used in the same project.



We created a model of the EM domain



Modelling but need of conversion functions for each package



Goal 2: Full project traceability,
improving reproducibility.



We track all the steps performed in a project

The screenshot displays the SCIPION software interface. On the left is a sidebar with a tree view of protocols, including 'Imports', 'Micrographs', 'Particles', '2D', and '3D'. The main window shows a hierarchical workflow tree for a project, starting with 'PROJECT' and proceeding through steps like 'scipion - import particles', 'scipion - filter particles', 'scipion - align apsr', 'scipion - align pairwise', 'scipion - custom mask', 'scipion - capca', and finally three classification steps: 'scipion - classify ward', 'scipion - classify kmeans', and 'scipion - classify dlad'. Below the tree is a summary panel with tabs for 'Summary', 'Methods', and 'Output Log'. The 'Summary' tab is active, showing input and output details for the 'scipion - align pairwise' step, including a 'SUMMARY' section with 'Radius range: 5 - 44' and 'Search range (px): 8'. A red 'Analyze Results' button is visible in the top right of the summary panel.



Scipion Cloud in production

Most courses now use Scipion Cloud

NeCEN

[ABOUT NECEN](#)

[SERVICES](#)

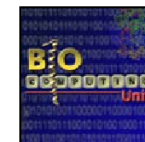
[USER ACCESS](#)

[MICROSCOPES](#)

[PEOPLE](#)

[RESEARCH PROJECTS](#)

CRYO-EM SCHOOL



Production connection with Amazon

Project Help

SCIPION FEI_course (2017-01-11) fb24dbd Project EM_Course_Dec_04 Protocols | Data

View: All

Edit Copy Delete Steps Browse Db Collapse Labels View: Tree Refresh

Workflow steps (from top to bottom):

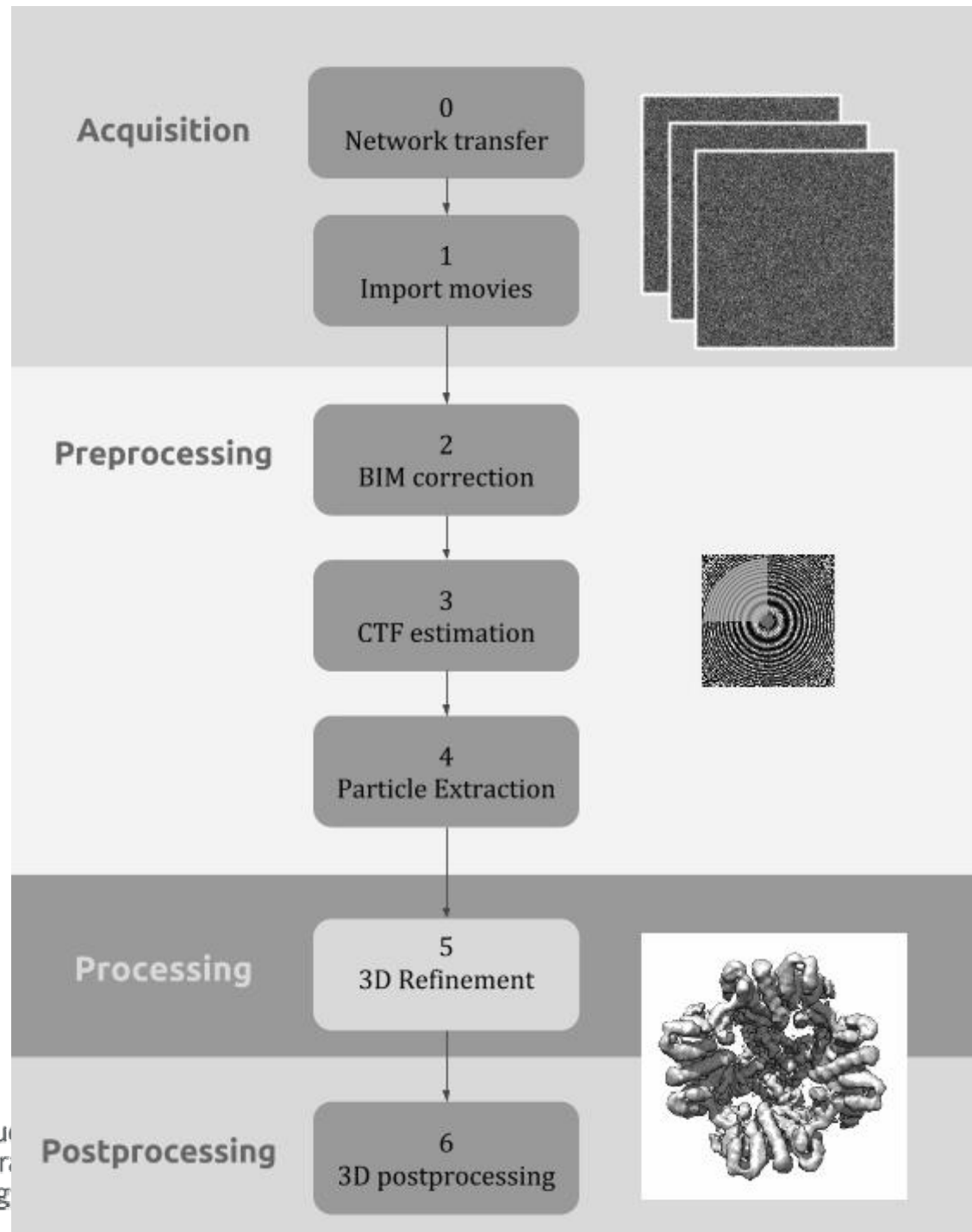
- PROJECT
- scipion - import movies (finished)
- motioncorr - motioncorr alignment (finished)
- xmipp3 - movie gain (finished)
- grigoriefflab - ctffind (finished)
- xmipp3 - manual-picking (step 1) interactive
- eman2 - sparx gaussian picker (finished)
- create subset by focus (finished)
- xmipp3 - auto-picking (step 2) (finished)
- xmipp3 - consensus picking (copy) (finished)
- xmipp3 - extract particles (finished)
- reion - 3D classification
- xmipp3 - extract particles 2

Summary Methods Output Log

Input		
InputObject (from reion - 3D classification)		reion - 3D classification
Output		
Volume 1 16463 -> outputRepresentatives		SetOfVolumes (1 items, 360 x 360 x 360, 1.34 A/px)

Analyze Results





Production connection with Amazon

The screenshot displays the SCIPION software interface for Project Taipei_2017. The top navigation bar includes the SCIPION logo, the project name, and a 'Protocols | Data' link. A left sidebar lists protocol categories: Imports, Micrographs, Particles, 2D, 3D, and Tools. The main area shows a workflow tree starting with 'PROJECT', followed by 'scipion - import movies finished', 'xmipp3 - optical alignment finished', and a parallel split into 'grigoriefflab - ctffind finished', 'xmipp3 - manual-picking (step 1) interactive', and 'xmipp3 - preprocess micrographs finished'. Further steps include 'xmipp3 - auto-picking (step 2) finished', 'eman2 - boxer interactive', 'xmipp3 - consensus picking finished', and 'xmipp3 - extract particles finished'. Below the tree, a summary table shows input and output data.

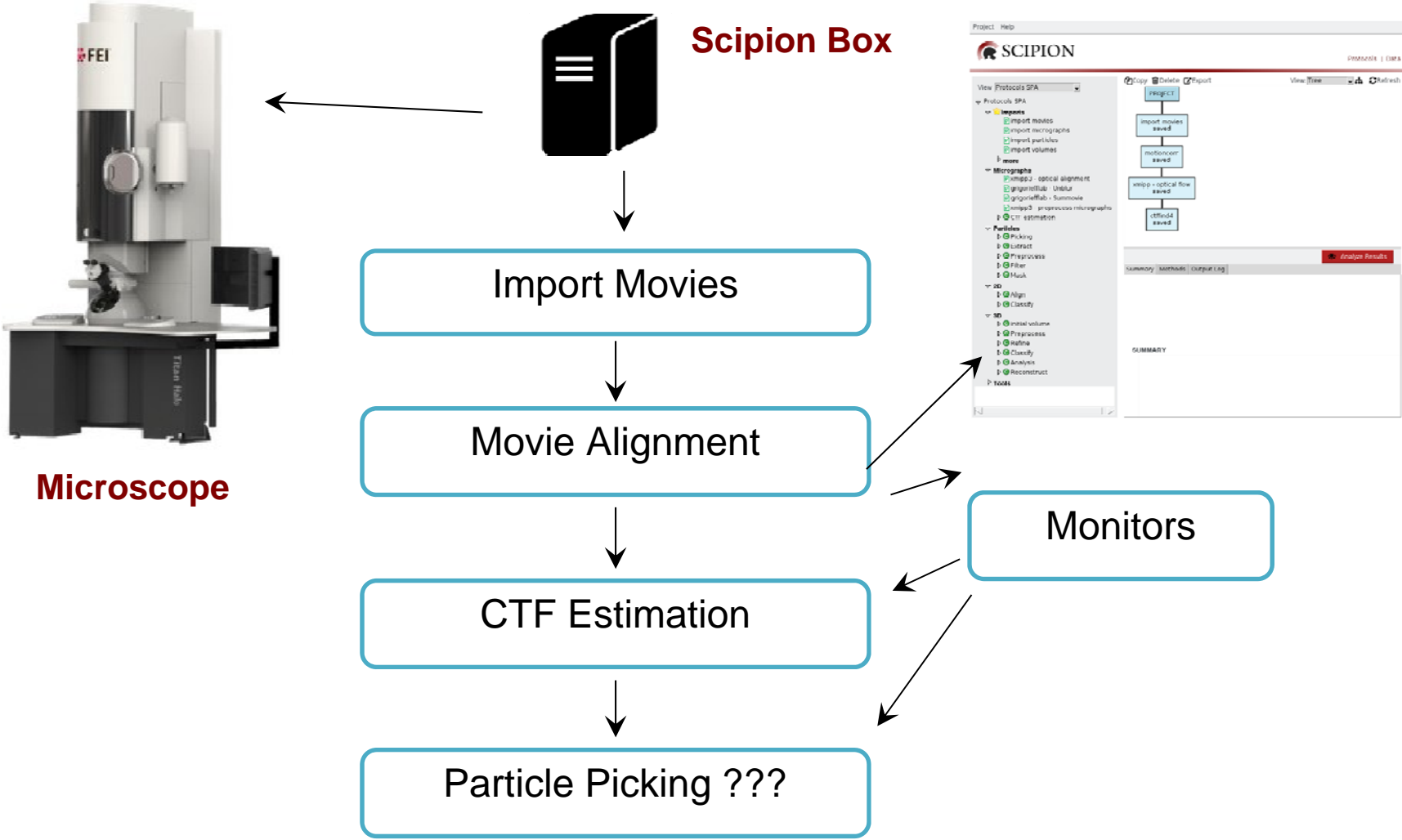
Input	Output
InputMicrographs (from xmipp3 - optical alignment -> outputMicro SetOfMicrographs (15 items, 1950 x 1950, 3.54 A/px))	grigoriefflab - ctffind -> outputCTF SetOfCTF (15 items)



Goal 3: Execute complete workflows in an automated manner.



Run workflows automatically and in streaming



Goal 4: Local and Cloud deployment

Fully deployed on Amazon cloud and on the European Federated Cloud!



Scipion on the Cloud



Goal 5: Easy to extend with new methods



Collaboration with other developers

- **Example 1:** Spider-MDA (with Tanvir Shaikh)
- **Example 2:** NMA and flexible fitting (with Slavica Jonic)
- **Example 3:** Refactoring of ResMap (with A.Kucukelbir and H.Tagare)
- **Example 5:** Localized Reconstruction (with S.Ilca and J.Huiskonen)
- **Example 6:** SIMPLE protocols (with H.Elmlund)



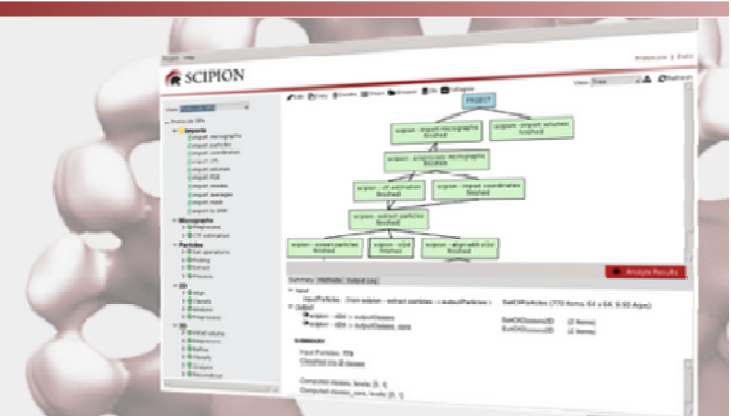
Scipion is open-source and freely available at <http://scipion.cnb.csic.es>



Integration, Reproducibility and Analysis

Scipion is an image processing framework to obtain 3D models of macromolecular complexes using Electron Microscopy.

 Download



Documentation

Find out more information about Scipion for both users and developers. Check how to install Scipion, the list of integrated protocols and some introductory tutorials. Learn how to extend Scipion with new protocols, its API or description of our development tools.

Web Tools

Give a try to our online processing workflows, which are a subset of the protocols in Scipion to provide a first try without any local installation. Current web-tools include initial volume estimation, movie alignment and local resolution (ResMap).

News and Events

Keep an eye on latest Scipion news. Check out about bugfixes, new features and release plan. Don't miss the next Scipion workshop or any related event. Don't hesitate to contact us for any feedback or if you want to organize a workshop at your institution.

© 2015 - Scipion is licensed under the GNU General Public License



Scipion web tools



In this context: What are “Web tools”

From our site (<http://scipion.cnb.csic.es/>):

“online processing workflows, which are a subset of the protocols in Scipion to provide a first try without any local installation.”



Web tools home page



Scipion Web Tools - **W**est-**L**ife

Single particle analysis tools

Align your movies

My movie alignment

Create your initial volume

My first map

Analyze your maps

My resolution map

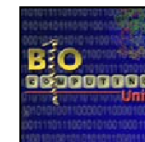
Reliability tools

Explore protein interactions



Check the reliability of your data

My reliability tool



Web tool page layout



Obtain your initial volume

To obtain an initial 3D map from your average images several methods are available from different packages (Xmipp3 and Eman2). At the end of the processing you will be able to compare the output volumes from the different methods, downloading the selected ones.

It is recommended to read the ["Initial model estimation" tutorial](#) before using this portal.

Choose your project

Create an empty project

Create

Create a project with Test data

Test data

Explore project with Test data (read-only)

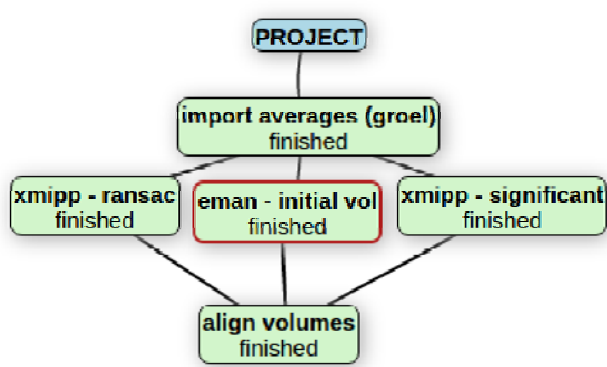
Explore



Example of a web tool

Workflow View: Initial_Volume ▾ Edit Copy Delete Small Tree List Refresh Help

- 1. Upload data
 - import averages
- 2. Create a 3D volume
 - xmipp3 - ransac
 - eman2 - initial volume
 - xmipp3 - significant
- 3. Align volumes.
 - xmipp3 - align volumes



```
graph TD; PROJECT[PROJECT] --> A[import averages (groel) finished]; A --> B[xmipp3 - ransac finished]; A --> C[eman - initial vol finished]; A --> D[xmipp3 - significant finished]; B --> E[align volumes finished]; C --> E; D --> E;
```

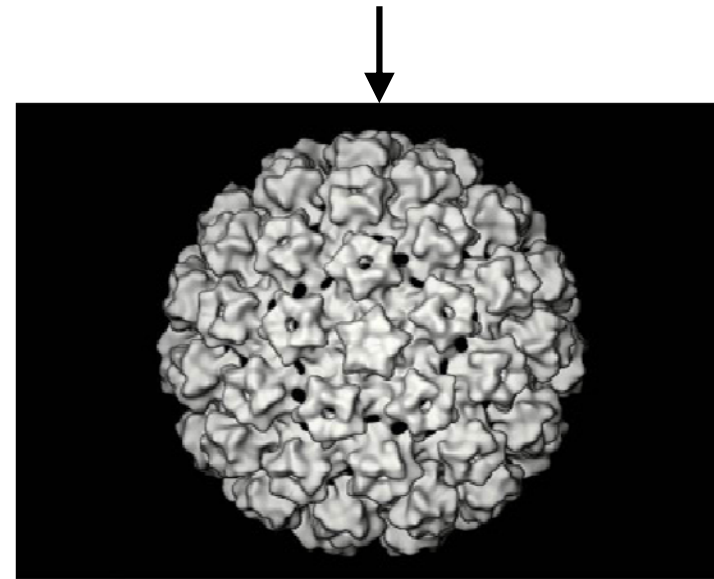
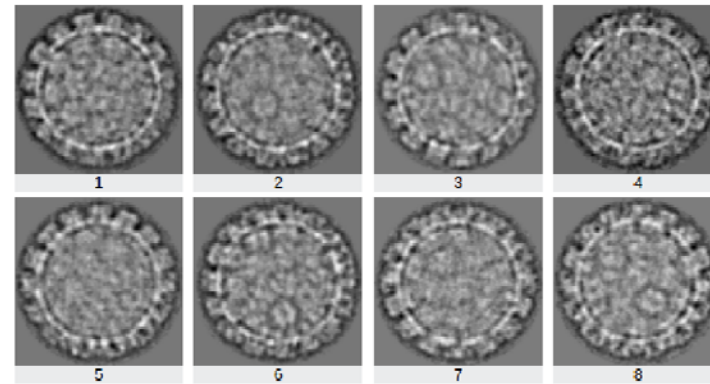
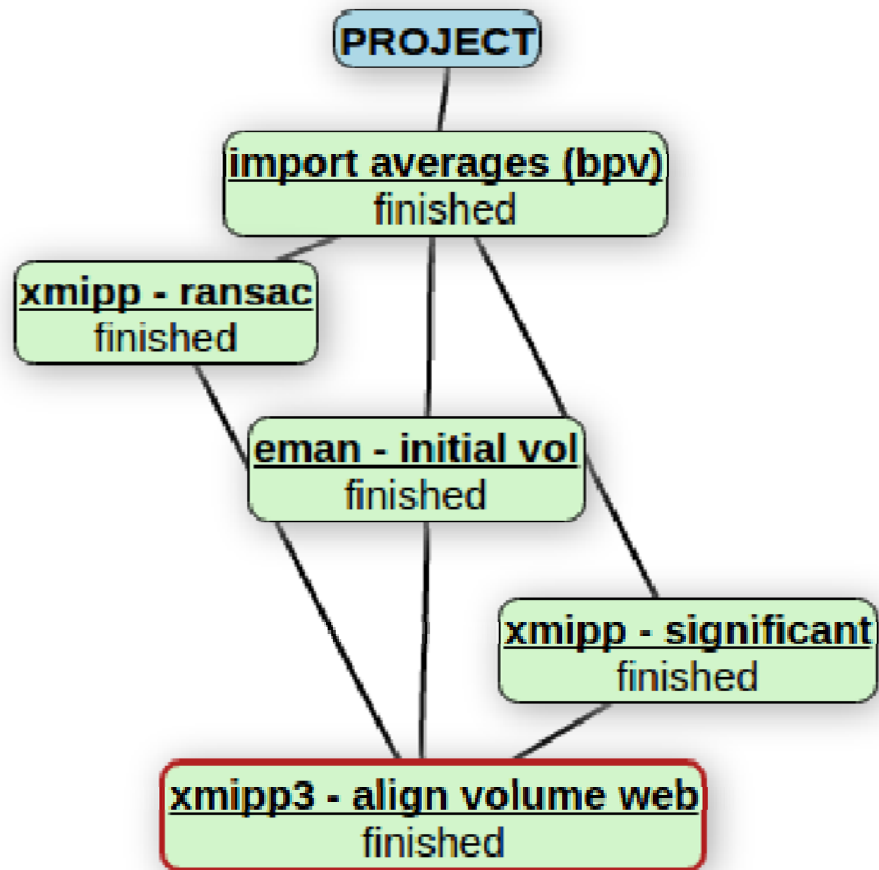
Summary | Methods | Output Logs | **Analyze Results** | Download Results

Input
inputSet (from 103.inputSet.116) SetOfAverages (31 items, 64 x 64, 1.00 A/px)

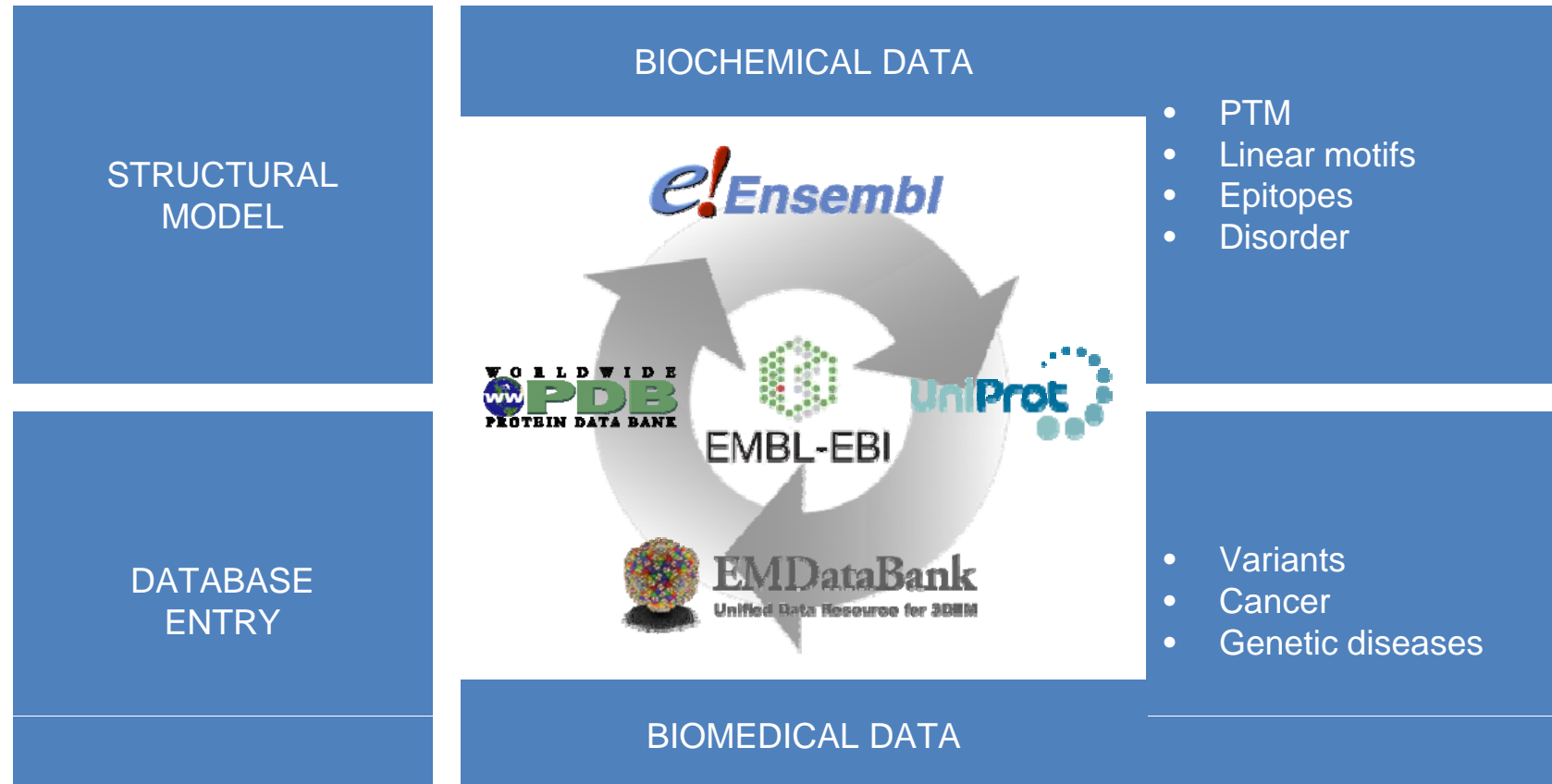
Output
outputVolumes SetOfVolumes (10 items, 64 x 64 x 64, 1.00 A/px)

Summary
Input Images: 2.outputAverages.47
Output initials volumes: None

My first map: Create your initial volume



3DBionotes: Enriching data



3DBionotes: Enriching data

- Single interface to display EMDB, PDB and UniProt entries
- Annotating structural models
- Protein annotations displayed at sequence and structural level
- Interactive annotation viewer
- Enriching data with other protein resources
- <http://3dbionotes.cnb.csic.es>

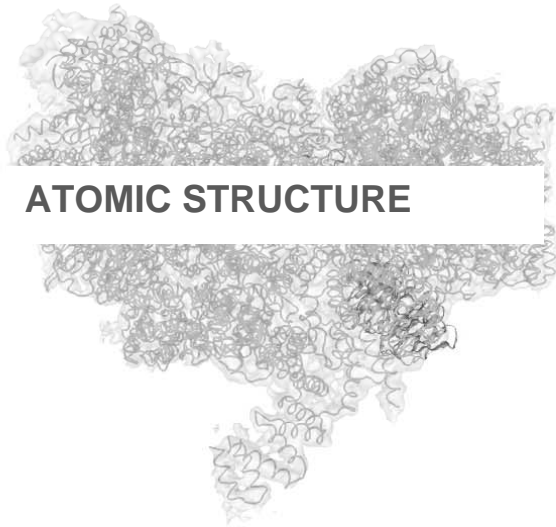


3DBionotes GUI

CRYO-EM STRUCTURE OF THE HUMAN APC/C-CDH1-EMI1 TERNARY COMPLEX AT 3.6 ANGSTROM RESOLUTION.

EMD-2924 PROTEINS IN THIS MODEL: PDB:4UI9 CH:R UniProt:Q9UM11 FIZZY-RELATED Pi

5



1 MDDDYERRLLRQIVIONENTHPRVTEMRRRLTPASSPVSSPSKHGDRFIPSRAGANWVNFHRRINENEKS
SKHGDRFIPSRAGANWVNFHRRINEN.....

71 PSQNRKAKDAT LFTYSLSTKRSSPD
LFTYSL.....

141 DGNVSPYSLSPVSNKSKQLLRSKPKPKIRKISKIPFKVLDAPELQDDFFYLNLDVWSSLNVLVSVGLGTCVY
SPYSLSPVSNKSKQLLRS.....SKIPFKVLDAPELQDDFFYLNLDVWSSLNVLVSVGLGTCVY

PROTEIN SEQUENCE



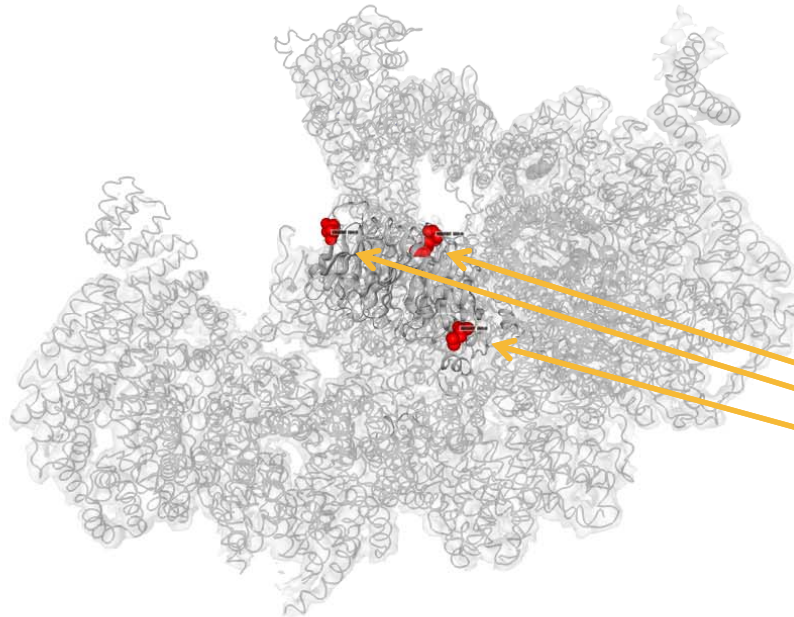
PROTEIN ANNOTATIONS



3DBionotes: Biomedical data

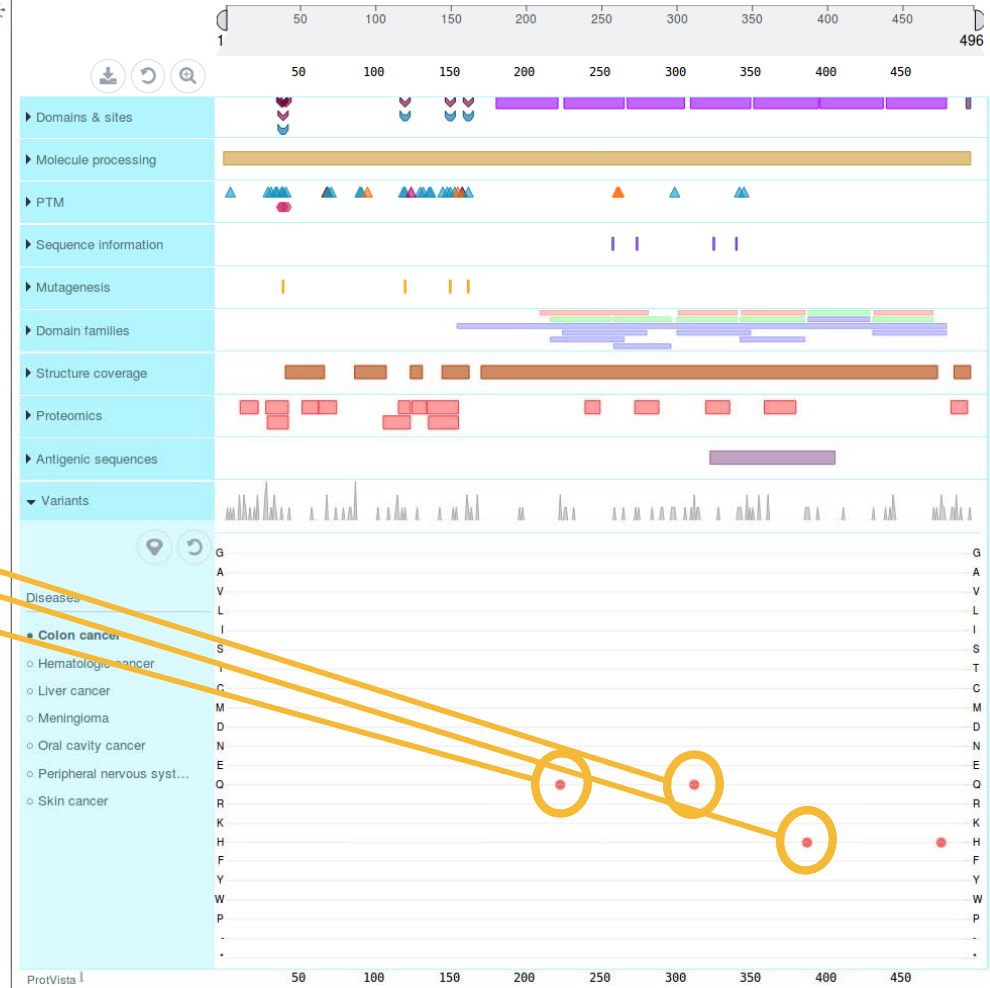
CRYO-EM STRUCTURE OF THE HUMAN APC/C-CDH1-EM1 TERNARY COMPLEX AT 3.6 ANGSTROM RESOLUTION.

EMD-2924 PROTEINS IN THIS MODEL: PDB:4UI9 CH-R Uniprot:Q9UM11 FIZZY-RELATED PI



```

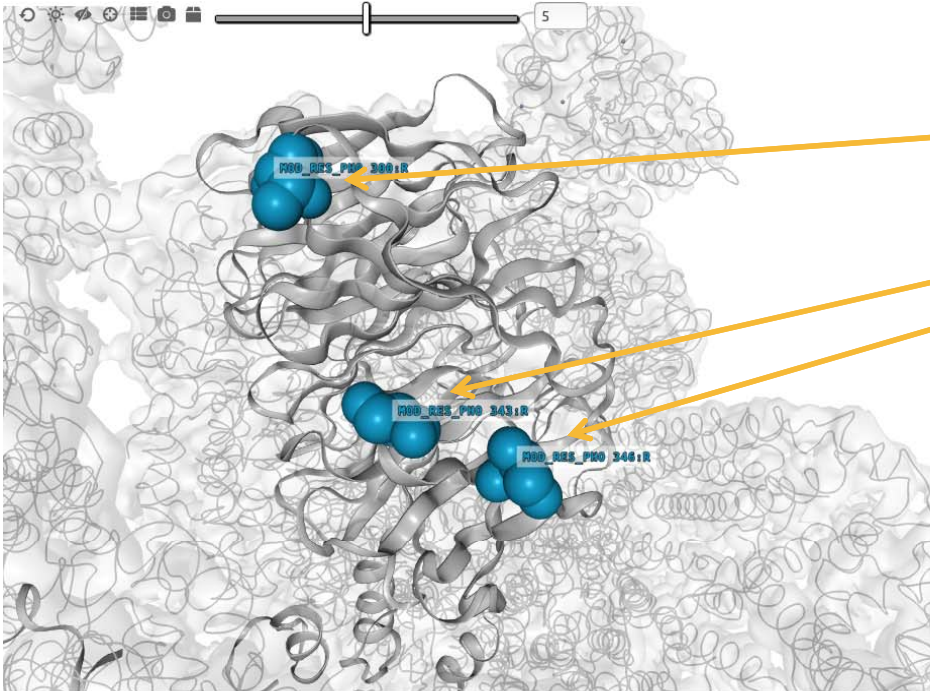
1  MDQDYERRLLRQIVIONENTMHRVTEMRRLTLPASSPVSSPSKHGDRFIPSRAGANWSVNFHRRINENEKS
   .....SKHGDRFIPSRAGANWSVNFHRRINEN...
71  PSQNRKAKDATSNGKDGGLAYSALLKNELLGAGIEKVQDPQTEDRRLQPSTPEKKGLFTYSLSLTKRSSPD
   .....GLAYSALLKNELLGAGIEKVQ.....KGLFTYSL.....
141  DGNQVSPYSLSPVSNKSQKLLRSRPRKPTRKISKIPFKVLDAPELQDDFYLNLDVWSSLNVLVSVGLGTCVY
   .....SPYSLSPVSNKSQKLLRS.....SKIPFKVLDAPELQDDFYLNLDVWSSLNVLVSVGLGTCVY
    
```



3DBionotes: Proteomics data

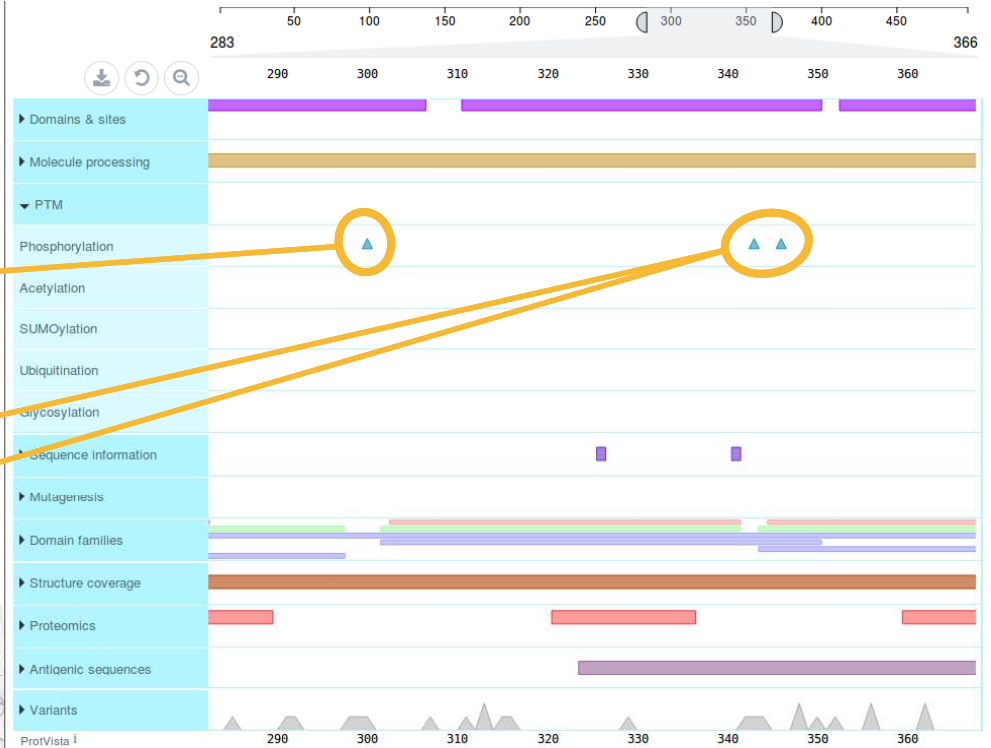
CRYO-EM STRUCTURE OF THE HUMAN APC/C-CDH1-EM1 TERNARY COMPLEX AT 3.6 ANGSTROM RESOLUTION.

EMD-2924 PROTEINS IN THIS MODEL: PDB:4UI9 CH-R UniProt:Q9UM11 FIZZY-RELATED P1

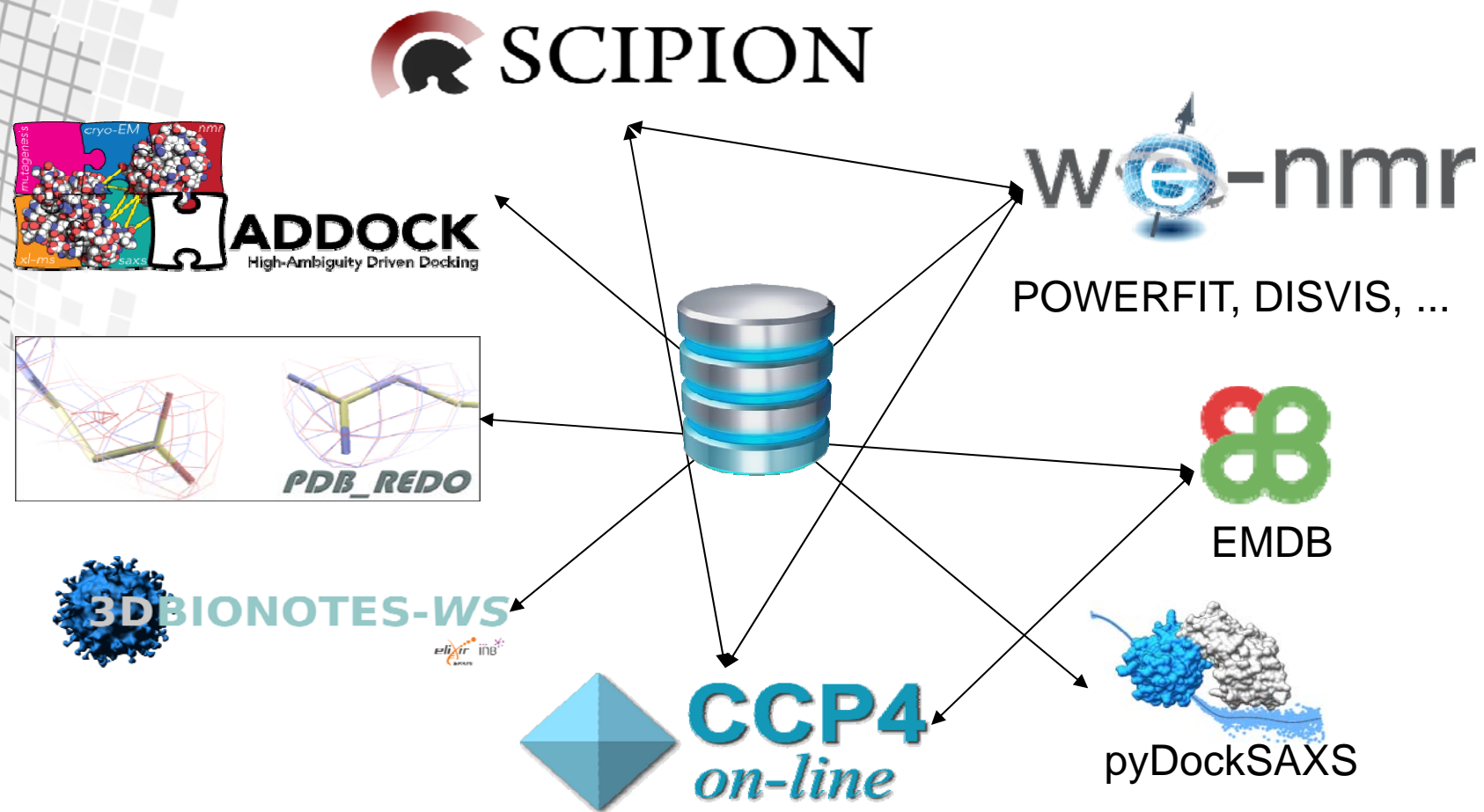


```

1  MDQDYERRLLRQIVIONENTHPRVTEMRRRLTPASSPVSSPSKHGDRFIPSRAGANWSVNFHRRINENEKS
   .....SKHGDRFIPSRAGANWSVNFHRRINEN...
71  PSQNRKAKDATS DNGKDGGLAYSALLKKNELLGAGIEKVQDDPQTEDRRLOPSTPEKKGLFTYSLSTRKSSPD
   .....GLAYSALLKKNELLGAGIEKVQ.....KGLFTYSL...
141 DGNVSPYSLSPVSNKSQKLLRSRPRKPTRKISKIPFKVLDAPELODDFYLNLDVMSSLNVLVSVGLGTCVY
   .....SPYSLSPVSNKSQKLLRS.....SKIPFKVLDAPELODDFYLNLDVMSSLNVLVSVGLGTCVY
    
```



Interconnecting Structural Biology resources: West-Life



And the people behind.....

- Carlos Oscar Sanchez-Sorzano and Roberto Marabini
- Javier Vargas
- Joan Segura
- Jose Luis Vilas and Ruben Sanchez
- Pablo Conesa, Jose Miguel de la Rosa and Scipion team
- And all the talented predocs, postdocs and engineers along the years!



And the people to come...

Always looking for talented
engineers, predocs and
postdoc fellows!

