A Virtual Research Environment for Integrative Modelling of Biomolecular Complexes with the new Modular Version of HADDOCK

Prof. Alexandre M.J.J. Bonvin Bijvoet Center for Biomolecular Research Faculty of Science, Utrecht University the Netherlands a.m.j.j.bonvin@uu.nl @amjjbonvin



Overview

General introduction Introducing the new modular HADDOCK3 Towards a virtual research environment for HADDOCK3 – i-VRESSE

Conclusions



The social network of proteins



Majority of 'life' depends on interactions, particularly protein-protein



HADDOCK: An integrative modeling platform



netherlands

Sciencecenter





Universiteit Utrecht

Bioinformatics, Volume 39, Issue 1, January 2023, btac759, https://doi.org/10.1093/bioinformatics/btac759









- ADDOCK High-Ambiguity Driven Docking
- > 35000 registered users
- > 530000 served runs since June 2008
- 65% on the EOSC HTC resources (>80% for 2.4)
- > Integrated in the EOSC marketplace

De Vries et al. Nature Prot. 2010

Van Zundert et al. J.Mol.Biol. 2016







https://wenmr.science.uu.nl





HADDOCK: Meeting the increased demand

The HADDOCK workflow machinery was modified to improve its efficiency and meet the increased demand (and more resources were allocated thanks to EGI/EOSC/OSG).



COVID19 effect



Overview

General introduction Introducing the new modular HADDOCK3 Towards a virtual research environment for HADDOCK3 – i-VRESSE

Conclusions



Introducing HADDOCK v3





Rodrigo Vargas Honorato



Brian Jimenez



João MC Teixeira



Marco Giulini





From static workflow ...













Combining the different pieces







Combining the different pieces



Universiteit Utrecht



refinement Semi-Flexible ref. **Rigid Body Energy minimization** lightdock Short MD w/ explicit water openMM 1 **All-atom scoring** All-atom scoring w/ MDexp. water

Protein-protein docking example with # NMR-derived ambiguous interaction restraints # directory in which the scoring will be done run dir = "run1" # molecules to be docked molecules = "data/e2aP 1F3G.pdb", "data/hpr ensemble.pdb" [topoaa] [rigidbody] ambig fname = "data/e2a-hpr air.tbl" sampling = 1000[seletop] select = 200[flexref] ambig fname = "data/e2a-hpr air.tbl" [emref] ambig fname = "data/e2a-hpr air.tbl"

[caprieval] [clustfcc]

[seletopclusts]

[caprieval]

HADDOCK3's open-source repository

💡 main 👻 🐉 24 branches 🕟 6 tag	gs	Go to file Add fi	le - <> Code -	About
mgiulini Merge pull request #639 from	n i-VRESSE/add_mol_viewer	✓ 0996f0c last week	(1) 2,377 commits	The official repo of the new modular BioExcel2 version of HADDOCK
github	change codecov fail_ci_if_error		last week	
evtools	Merge pull request #520 from haddocking/	module_template	10 months ago	bioinformatics proteins python3
b docs	Update INSTALL.md sementic on various ha	addock3	last month	modelling docking workflows complexes integrative-modeling
examples	Merge pull request #632 from haddocking/	adjust_clustrmsd	3 months ago	utrecht-university
src src	Merge pull request #639 from i-VRESSE/ad	d_mol_viewer	last week	🛱 Readme
tests	use cluster rank instead of id, replace struc	ture with model, fix test	3 weeks ago	শ্রু Apache-2.0 license
🖿 varia	Update README.md		7 months ago	-^ Activity
bumpversion.cfg	name		last year	 ☆ 55 stars ◆ 14 watching
🗅 .gitignore	add haddock3-docs to .gitignore		last year	父 22 forks
🗋 .gitmodules	add FCC submodule		2 years ago	Report repository





https://github.com/haddocking/haddock3

Universiteit Utrecht



HADDOCK3 development time line

13/03/2019









Universiteit Utrecht



Current execution modes

• Local

compute mode
mode = "local"
ncores = 40

• **HPC** (slurm & torque currently supported)

```
# compute mode
mode = "hpc"
# Batch queue to use
queue = "short"
# Concatenate models inside each job
concat = 5
# Limit the number of concurrent submissions to the queue
queue_limit = 250
```





Current execution modes



compute mode
mode = "mpi"
5 nodes x 50 tasks = ncores = 250

ncores = 250

Antibody-antigen surface example Sampling: 10240/2048/2048 rigidbody/flexref/emref









netherlands



Examples of runner config files

HADDOCK2.4

HADDOCK3.0

scenarios: scenarios: - name: true-interface name: true-interface parameters: parameters: general: run_cns: noecv: false mode: hpc queue: short structures_0: 1000 queue_limit: 100 structures_1: 200 concat: 5 waterrefine: 200 restraints: modules: ambig: ambig order: [topoda, rigidbody, seletop, flexref, emref, clustfcc, seletopclusts] unambig: restraint-bodies topoaa: hbonds: hbonds autohis: true custom_toppar: rigidbody: topology: _ligand.top ambig_fname: _ambig.tbl param: _ligand.param unambig_fname: _restraint-bodies.tbl ligand_param_fname: _ligand.param name: center-of-mass ligand_top_fname: _ligand.top parameters: seletop: select: 200 run_cns: flexref: cmrest: true ambig_fname: _ambig.tbl structures 0: 10000 unambig_fname: _restraint-bodies.tbl structures 1: 400 ligand_param_fname: _ligand.param waterrefine: 400 ligand_top_fname: _ligand.top anastruc_1: 400 emref: custom_toppar: ambig_fname: _ambig topology: _ligand.top clustfcc: param: _ligand.param seletopclusts:









Antibody-antigen modelling with only CDR info and full surface on the antigen

- 1. Build topo
- **2. Rigid-body** (10000)
- 3. CAPRIeval



4. FCC clustering



5. Select clusters (min size 4) (max. 20 models per cluster)



6. Flexref





8. EMref



9. CAPRIeval

7. CAPRIeval



10. FCC clustering



11. Select clusters (min size 4)









Antibody-antigen modelling with only CDR info and full surface on the antigen



Final refinement clusters



 Cluster 1 Cluster 2

 Cluster 3 Cluster 4

Cluster 5

Cluster 6

 Cluster 7 Cluster 8

Cluster 9

Cluster 10

Other

Overview

General introduction Introducing the new modular HADDOCK3 Towards a virtual research environment for HADDOCK3 – i-VRESSE

Conclusions

The plan...





netherlands Science center

The plan...



netherlands Science center

The plan...

Data analysis, storage and sharing infrastructure









Toward a new interface for HADDOCK3



Web application to build and execute HADDOCK3 workflows and manage, view and analyze results





Build Use the workflow builder to create and submit a job.



Upload Upload a workflow and submit as job. Data analysis, storage and sharing infrastructure Data storage/sharing Analysis Workspace Analysis Workspace Analysis tools Motocutar Visualization Cellaborative Interactive piols Import/Export

Manage

Explore and analyse the results of completed jobs.

Universiteit Utrecht

github.com/i-VRESSE/bartended-haddock3



HA	DOCK Workflow Builder
Community	Input Data
Workflows	Parameters

The workflow builder





https://github.com/i-VRESSE/workflow-builder







Workflow builder

Haddock3 Manage Build

About Help Bonvin Lab

Workflow builder is dynamically configured based on HADDOCK3's modules definitions and parameters

Drag and drop workflow builder Catalog Nodes ▼topology topoaa ▼ sampling gdock rigidbody lightdock ▼refinement mdref flexref

Workflov Global para	V ameters	Upload	Node Calculate HADDOCK
Visual	Text		
1. topoaa			🔺 ana
2. rigidbody			Refe
3. caprieval			Cho
4. seletop			Struc
5. flexref			Parfe
6. caprieval			rend
Submit Do	ownload	archive	Perfo

e CAPRI metrics. (caprieval) (3 module to calculate the CAPRI metrics. alysis erence structure oose File No file chosen ta/target.pdb (chemical/x-pdb, 229768 bytes) ture to be used when calculating the CAPRI metrics. Calculate I-RMSD orms Interface RMSD calculations. Delete Save Cancel Calculate FNAT orms FNAT calculations

Universiteit Utrecht



i-VRESSE workflow builder: Haddock 3 on easy level

naudocr	(Sedsy ·		Workfle	2010/			Node
Nodes			VUINIC	/ • •			No node or clobal parameters selected for configuration
▼ topol	ogy	I	Global p	arameters	Upload		no nous of global parameters selected for configuration.
0	topoaa		Visual	Text			
▼ samp	bling		Append	node to wor	rkflow by c	licking node in catalog	
0	gdock		or	by dragging	node from	n catalog to here.	
0	rigidbody						
0	lightdock						
▼ refine	ement						
0	mdref						
0	flexref						
0	emref						
▼ scori	ng						
0	mdscoring	· •					
0	emscoring						
▼ analy	vsis						
0	seletopclusts						
0	clustrmsd						
0	caprieval						
0	rmsdmatrix						
0	clustfcc		Downloa	ad archive	Clear		-







Easy integration: E.g. Galaxy

🗧 Galaxy	*	Workflow Visualize Shared Data ▼ Admin I	Help 🕶 User 🕶 🚖 🥖 🔍 🔚 🎆	
Tools 🏠 🔹	i-VRESSE wo		Noted successfully On duru level	
search tools X				
🏦 Upload Data	Catalog	(opens in new wi	ndow)	
Get Data	haddock3guru 🗸	Global param		
Send Data	Nodes	Visual Text	Bath to the OND successful	
Collection Operations	▼ topology			
Text Manipulation	∘ topoaa	molecules = [Choose File No file chosen	
Docking	T compling	'data/e2aP_1F3G.pdb', 'data/bpr_ensemble.pdb',	If not provided, HADDOCK3 will use the cns path configured during the installation.	
Filter and Sort	✓ sampling]	Number of CPU cores	
Join, Subtract and Group	 gdock 	<pre>ncores = 12 run dir = 'run1-test'</pre>	12	
Built-in Converters	 rigidbody 		Number of CPU cores to use for the CNS calculations. It is truncated to max available	
WORKFLOWS	 lightdock 	autohis = false	CPUs minus 1.	
All workflows		[topoaa.mol1]	Mode of execution	
	✓ reinement	nhisd = 0	local ¢	
	 mdref 	nhise = 1 hise 1 = 75		
	 flexref 	[topoaa.mol2]	Mode of execution of the jobs, either local or using a batch system.	
	 emref 	nhisd = 1	Batch system	
	T appring	hisd_1 = 76 nhise = 1	slurm 🗢	
	✓ sconing	hise_1 = 15	Type of batch system running on your server	
	 mdscoring 	[rigidbody]		
	 emscoring 	tolerance = 20 ambig_fname = 'data/e2a-hpr_air.	Queue name	
	▼ analysis	sampling = 20		
	 seletopclusts 	[caprieval]	Name of the batch queue to which jobs will be submitted	
		reference_fname = 'data/e2a-hpr_	Number of jobs to submit to the batch system	
	 clustrmsd 	[seletop]	100	
	 caprieval 	select = 5	Number of jobs to submit to the batch system	
	• rmsdmatrix	[TLEXTET]	Number of models to produce per job.	
	 clustfcc 	Submit - Clear	Submit Cancel	



Using 196.7 MB

+ = -× ×

1

Q2 😅

0/1

-

History

scratch

2 : index.html

1 : docking-protein-ligand.zip 🛛 👁 🧪 🧃

🛢 23.6 MB

M 42

search datasets Unnamed history



<





Easy integration: E.g. JupyterLab

💭 File Edit View Run Kernel Tabs Settings Help

+	± C	Z Launcher X	I untitled1.cfg × +		Z Launcher X +
Filter files by name	e Q	Catalog	Workflow	Global parameters	Notebook
Name	Last Modified	Nodes	Global parameters	Run directory*	
haddock3 config	urator 22 days ano	▼ topology		х	
haddock3_config	urator 6 minutes ago	t topology	Visual Text	Folder to store the HADDOCK3 run	
	an hour ago	∘ topoaa			Python 3 (ipykernel)
lite	an hour ago	▼ sampling	run_dir = 'x'	Input Molecules	
node modules	5 days ago	, samping	molecules = [·	Console
schema	22 days ago	 gdock]	The input molecules that will be used for docking.	
src	a day ago	o rigidhody			
style	6 days ago		[gdock]	O Choose file 200b.pdb	-
ui-tests	6 days ago	 lightdock 	ambig_fname = 'pyproject	.toml'	Python 3 (ipykernel)
200b.pdb	6 days ago	▼ refinement			
🗅 babel.config.js	22 days ago	▼ rennement	Copy to clipboard		\$_ Other
CHANGELOG.md	d 2 minutes ago	 mdref 	Files		
CITATION.cff	21 minutes ago	a flauraf	• 200b.pdb		Ś
CODE_OF_CON	DUCT 2 minutes ago	o llexrei			* -
CONTRIBUTING.	.md 2 minutes ago	∘ emref	pyproject.toml	If true, evaluates and tries to correct the input PDB before the workflow.	Terminal
docking-protein-li	gand 22 days ago				
(:) install.json	22 days ago	▼ scoring			
jest.config.js	22 days ago	 mdscoring 			·
LICENSE	22 days ago				Haddock3 config
MANIFEST.in	a minute ago	 emscoring 			
 (:) package.json 	10 minutes ago	▼ analysis			
pyproject.toml	22 days ago				
README.md	2 minutes ago	 seletopclusts 			Text File
RELEASE.md	22 days ago	 clustrmsd 			
screenshot.png	22 days ago				M
 setup.py tsconfig ison 	an hour ago	 caprieval 			•
tsconfig.jaon	o 7 minutes ago	 rmsdmatrix 			Markdown File
u2.cfg	a dav ano				
untitled1.cfg	a day ago	 clustfcc 			2
yarn-error.log	22 days ago	o seleton	Save Clear	Submit C	Cancel
C varn lock	6 days ago	o seletop			Python File



U U

Simple 0 🔬 0 🏟

Launcher





Backend to schedule jobs on various infrastructures and user management with social logins

Generic: Can be configured as scheduler for any kind of of service (requires CLI or container)

The bartender







Universiteit Utrecht

github.com/i-VRESSE/bartender





RESTful application with dynamic documentation of its APIs

Needs to be configured on some (remote) server(s) default \wedge /api/health Health Check \sim job $\overline{}$ \checkmark /api/job/ Retrieve Jobs \checkmark /api/job/{jobid} Retrieve Job \checkmark /api/job/{jobid}/files/{path} Retrieve Job Files \sim /api/job/{jobid}/stdout Retrieve Job Stdout \checkmark /api/job/{jobid}/stderr Retrieve Job Stderr application $\overline{}$ /api/application/ List Applications \sim /api/application/{application} Get Application \sim ✓ 🔒 /api/application/{application}/job Upload Job users \wedge V 🔒 🖻 /api/users/profile Profile \sim /users/me Users:Current User \checkmark /users/me Users:Patch Current User \checkmark /users/{id} Users:User









Universiteit Utrecht





Management and analysis platform





Will provide data management, analysis and visualization tools







github.com/i-VRESSE/...



HADDOCK3 analyse report



HADDOCK3 can automatically generate interactive analysis plots (similar to what the 2.4 server gives)

Will form the basis for the analysis interface







HADDOCK3 analyse report











Bartended HADDOCK3 overall workflow





Universiteit Utrecht



Universiteit Utrecht





https://github.com/i-VRESSE





Overview

General introduction Introducing the new modular HADDOCK3 Towards a virtual research environment for HADDOCK3 – i-VRESSE

Conclusions

Conclusions

- Finalizing the 1st production release of HADDOCK3
 - Modularity and custom workflows
 - Integrated analysis tools that can be leveraged by the VRE (only offered until now via the web server for HADDOCK2.X)
 - Both HTC and HPC oriented (EuroHPC BioExcel CoE core software)
- New VRE software for integrative modelling (i-VRESSE)
 - Will provide the user-friendly interface and integrated backend to build/execute/analyze HADDOCK3 workflows
 - Easy off-site deployment with docker compose (next to centralized WeNMR service)
 - Generic technology that can be reused for other projects



netherlands

Science center

https://github.com/i-VRESSE

WeNMR West-Life EGI-Engage INDIGO-Datacloud BioExcel CoE EOSC-Hub EGI-Ace



Acknowledgments:

former and current

CSB group@UU

bonvinlab.org/people1010000101

Acknowledgments: the eScience Center team

Stefan Verhoeven eScience Research Engineer Netherlands eScience Center



Peter Kalverla Netherlands eScience Center D 0000-0002-5025-7862

WeNMR West-Life EGI-Engage INDIGO-Datacloud BioExcel CoE EOSC-Hub EGI-Ace

€€



T

Fakhereh (Sarah) Alidoost Netherlands eScience Center 0 0000-0001-8407-6472 netherlands Science center

Thank you for your attention!





HADDOCK online:

- https://wenmr.science.uu.nl
- https://bonvinlab.org/software
- https://github.org/haddocking
- https://github.org/i-VRESSE

https://ask.bioexcel.eu



Center

Utrecht • https://ask.bioexcel.eu