

Gromacs - where are we now, what's new, what's coming in the future

Christian Blau

blau@kth.se

GROMACS origins

- Originally a hardware project at the University of Groningen in the early 1990s
 - custom-designed 32-processor ring using PVM
- Design and simulation methodology largely adopted from existing GROMOS molecular simulation package
- Key decision: develop in ANSI C, rather than FORTRAN
- Early adopters and core developers:
 - David van der Spoel, Berk Hess, Anton Feenstra, Rudi van Drunen

GROMACS origins

- First public release – 2.0 some time in the 1990s
- First paper – “GROMACS: a message-passing parallel molecular dynamics implementation” Comp. Phys. Comm. 1995
- first CVS commit in 1997, already 750 files and 145k LOC
- first FOSS release – 3.0 in August 2001, GPL 2.0
- provide preparation and analysis tools along with the core simulation engine
- aim for performance portability x86 assembly for key kernels

Who develops GROMACS

- Developers in Stockholm / Uppsala
 - Mark Abraham (project manager)
 - Paul Bauer (project/release manager)
 - Szilard Pall (GPU acceleration)
 - Artem Zhmurov (GPU acceleration)
 - Berk Hess (enhanced sampling, domain decomposition, algorithm development)
 - Erik Lindahl (SIMD, PME, large scale project coordination)
 - David van der Spoel (Energy analysis, analysis tools, force field development)
 - Joe Jordan (NB-LIB)
 - Christian Blau (external forces)

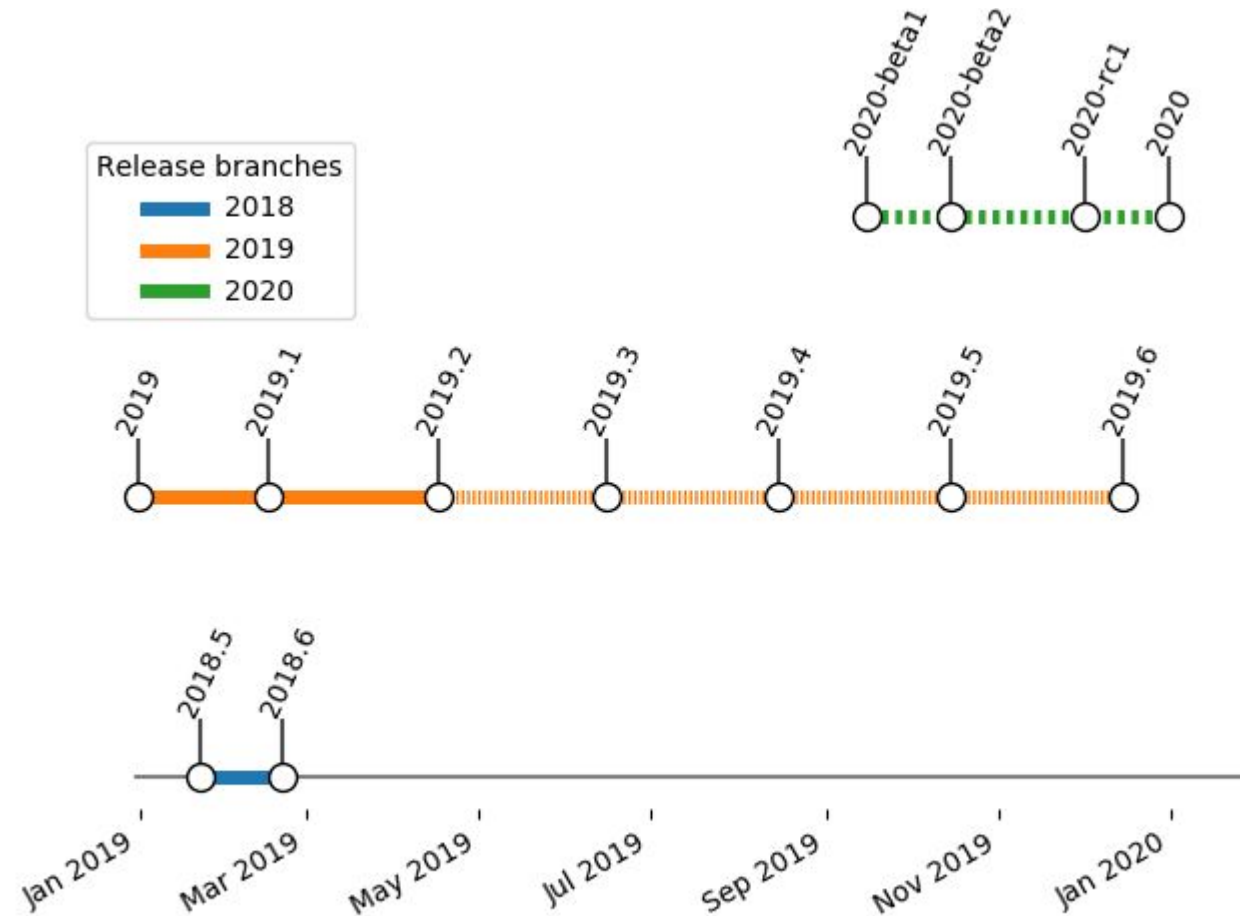
Who develops GROMACS

- Developers in Finland
 - JYU Jyväskylä (Gerrit Groenhof, Dmitry Morozov, Noora Aho, Emmi Pohjolainen)
 - constant PH, QM/MM
 - Teemu Mortula (options, selections, analysis tools, C++ transition)
- Developers in Germany
 - Carsten Kutzner (MPI-BPC Göttingen, ion swapping, enforced rotation)
 - Thomas Ullman (MPI-BPC Göttingen, lambda dynamics)
 - Vytautas Gapsys (MPI-BPC Göttingen, pmx, free energy)
 - Markus Rampp group (MPI-Computing and Data Facility)

Who develops GROMACS

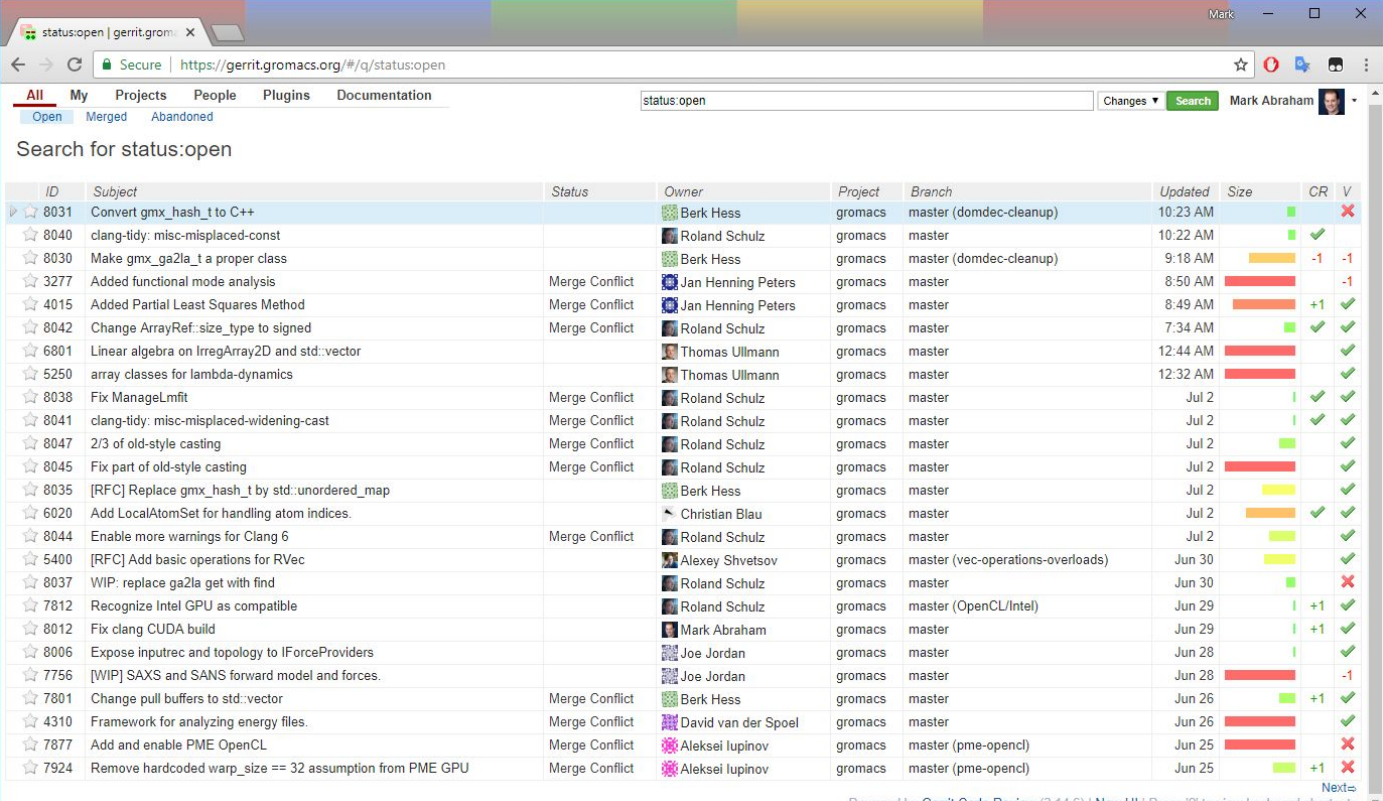
- US and others
 - Roland Schulz (@intel, parallelisation and acceleration, C++ transition)
 - Alan Gray, Jonathan Vincent (@NVidia, GPU acceleration)
 - Eric Irrgang, Peter Kasson (python api, ensemble simulations)
 - Pascal Merz, Michael Shirts (general integrator framework)
 - Kevin Boyd (code refactoring)

GROMACS development cycles



How to contribute

- Know what you want
- Ask developers
 - personally
 - mailing list
 - redmine
- Describe and document your feature
- Have proof of concept
- Upload to Gerrit
- Make small commits



The screenshot shows the Gerrit Code Review interface for the 'status:open' query. The table lists various pull requests with their IDs, subjects, statuses, owners, projects, branches, update times, sizes, and CR (Change Review) status. The table is sorted by update time, with the most recent changes at the top.

ID	Subject	Status	Owner	Project	Branch	Updated	Size	CR	V
8031	Convert gmx_hash_t to C++		Berk Hess	gromacs	master (domdec-cleanup)	10:23 AM			
8040	clang-tidy: misc-misplaced-const		Roland Schulz	gromacs	master	10:22 AM			
8030	Make gmx_ga2la_t a proper class		Berk Hess	gromacs	master (domdec-cleanup)	9:18 AM		-1	-1
3277	Added functional mode analysis	Merge Conflict	Jan Henning Peters	gromacs	master	8:50 AM			-1
4015	Added Partial Least Squares Method	Merge Conflict	Jan Henning Peters	gromacs	master	8:49 AM		+1	
8042	Change ArrayRef::size_type to signed	Merge Conflict	Roland Schulz	gromacs	master	7:34 AM			
6801	Linear algebra on IrregArray2D and std::vector		Thomas Ullmann	gromacs	master	12:44 AM			
5250	array classes for lambda-dynamics		Thomas Ullmann	gromacs	master	12:32 AM			
8038	Fix ManageLmfit	Merge Conflict	Roland Schulz	gromacs	master	Jul 2			
8041	clang-tidy: misc-misplaced-widening-cast	Merge Conflict	Roland Schulz	gromacs	master	Jul 2			
8047	2/3 of old-style casting	Merge Conflict	Roland Schulz	gromacs	master	Jul 2			
8045	Fix part of old-style casting	Merge Conflict	Roland Schulz	gromacs	master	Jul 2			
8035	[RFC] Replace gmx_hash_t by std::unordered_map		Berk Hess	gromacs	master	Jul 2			
6020	Add LocalAtomSet for handling atom indices.		Christian Blau	gromacs	master	Jul 2			
8044	Enable more warnings for Clang 6	Merge Conflict	Roland Schulz	gromacs	master	Jul 2			
5400	[RFC] Add basic operations for RVec		Alexey Shvetsov	gromacs	master (vec-operations-overloads)	Jun 30			
8037	WIP: replace ga2la get with find		Roland Schulz	gromacs	master	Jun 30			
7812	Recognize Intel GPU as compatible		Roland Schulz	gromacs	master (OpenCL/Intel)	Jun 29		+1	
8012	Fix clang CUDA build		Mark Abraham	gromacs	master	Jun 29		+1	
8006	Expose inputrec and topology to IForceProviders		Joe Jordan	gromacs	master	Jun 28			
7756	[WIP] SAXS and SANS forward model and forces.		Joe Jordan	gromacs	master	Jun 28			-1
7801	Change pull buffers to std::vector	Merge Conflict	Berk Hess	gromacs	master	Jun 26		+1	
4310	Framework for analyzing energy files.	Merge Conflict	David van der Spoel	gromacs	master	Jun 26			
7877	Add and enable PME OpenCL	Merge Conflict	Aleksei lupinov	gromacs	master (pme-opencil)	Jun 25			
7924	Remove hardcoded warp_size == 32 assumption from PME GPU	Merge Conflict	Aleksei lupinov	gromacs	master (pme-opencil)	Jun 25		+1	

Documentation is a must

- All code has to be documented using Doxygen
 - Allows automatic generation of webpage content
 - Can be cross referenced in other parts
 - Will be used to determine public and internal API
- General documentation using ReStructured Text (rst)
 - Generates all remaining webpages and PDF manual
 - Makes it possible to have cross referencing between the sections
- If you use or develop for GROMACS, the manual is the best way to get started

GROMACS user and developer mailing lists

- General questions should be asked on user mailing list
 - gmx-users@gromacs.org
- Check first the archive if the questions have been asked
- Developer questions better suited at
 - gmx-developers@gromacs.org
- Plan to move both to a BioExcel organized forum
 - ask.bioexcel.eu

New features in GROMACS 2020

- Density-guided simulations
- Python api to gromacs: gmxapi 0.1
- Modular simulator
- Parrinello-Rahman pressure for the md-vv integrator.
- Performance enhancements
 - Speed-up for free energy simulations
 - More SIMD acceleration
 - own-FFTW configured better
 - Faster bonded interactions on GPUs

Development challenges (why don't you just...)

New code needs to be

- Maintainable
 - Understanding and working on a piece of code may not be tied to a single person
 - everything will come to bite us eventually
- Portable
- Scientifically correct
- Performant

How to add good new code

- Have a case
 - Start with the minimal unit that is useful
 - Well cited papers help
- Make a case
 - Write a careful complete documentation of what you want
 - discuss with other developers
- Have tests
 - Write small tested functional units
 - Write a regression test, where all of your functionality plays together
- Avoid complex interactions
 - Touch as little code as possible
- Provide generally useful infrastructure
 - Start with for necessary code that will benefit the rest of GROMACS

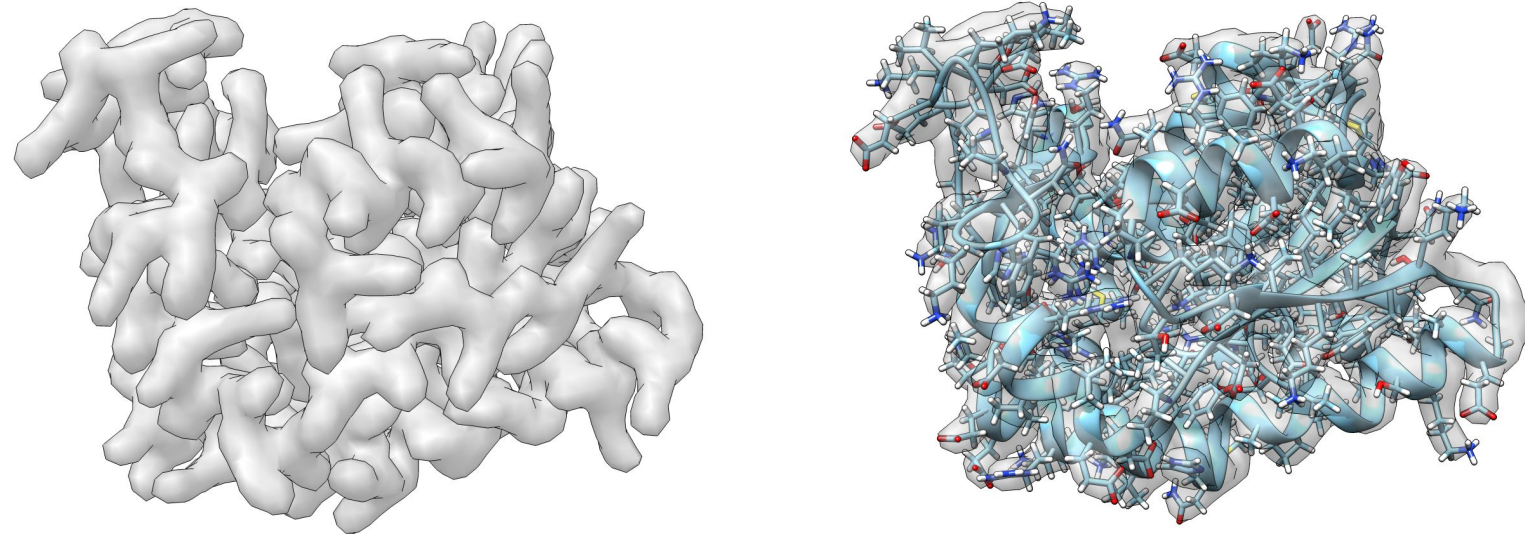
New in GROMACS 2020 - under the hood

Extensive code refactoring

- More tests
- Removal of the “group” scheme - loads of lines of code removed :)
- Domain-decomposition refactoring
- GPU code and communication refactoring
- extending the MDModule functionality
- modern Multidimensional arrays (mdspan)

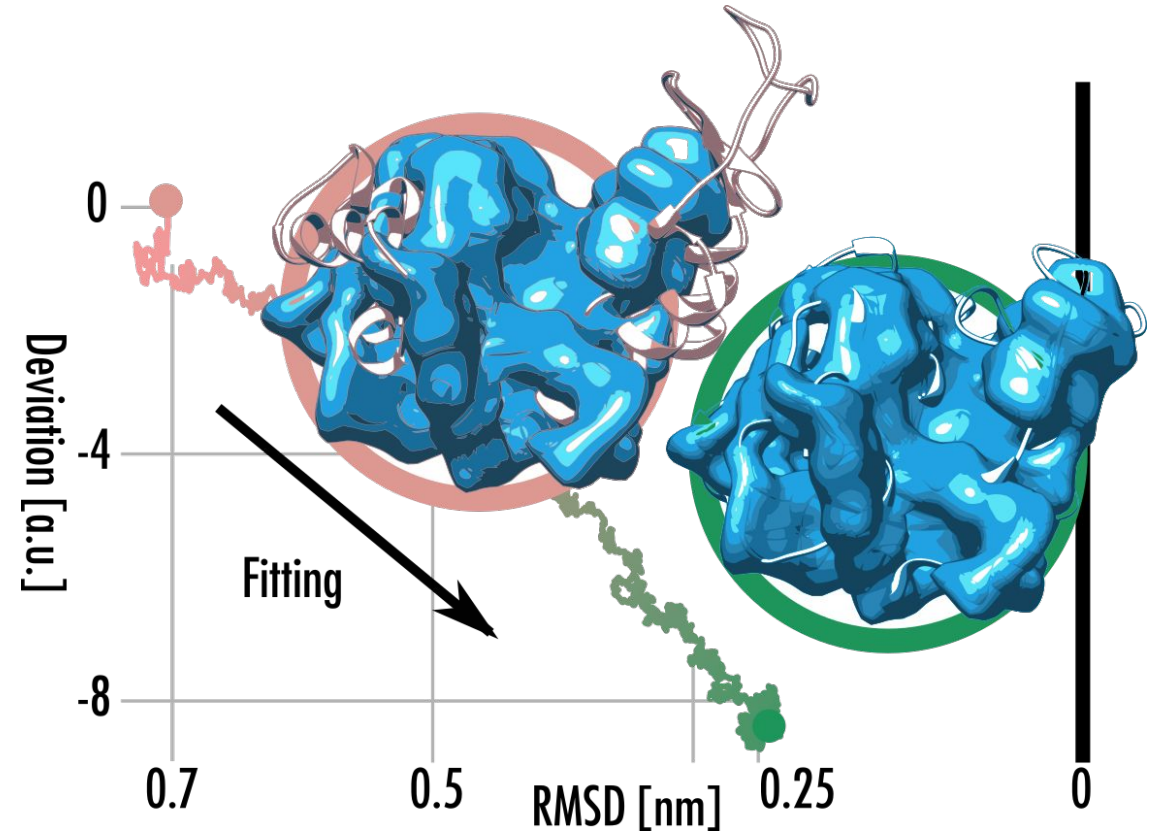
GROMACS 2020 - density-guided simulations

- Making the most use of cryo-EM densities
 - Automated density fitting to speed up time-to-result in the modeling process
 - Explore all the structures that fit a given cryo-EM density
 - Allowing to balance between bio-chemistry and cryo-EM data influence
 - See structural transitions between states



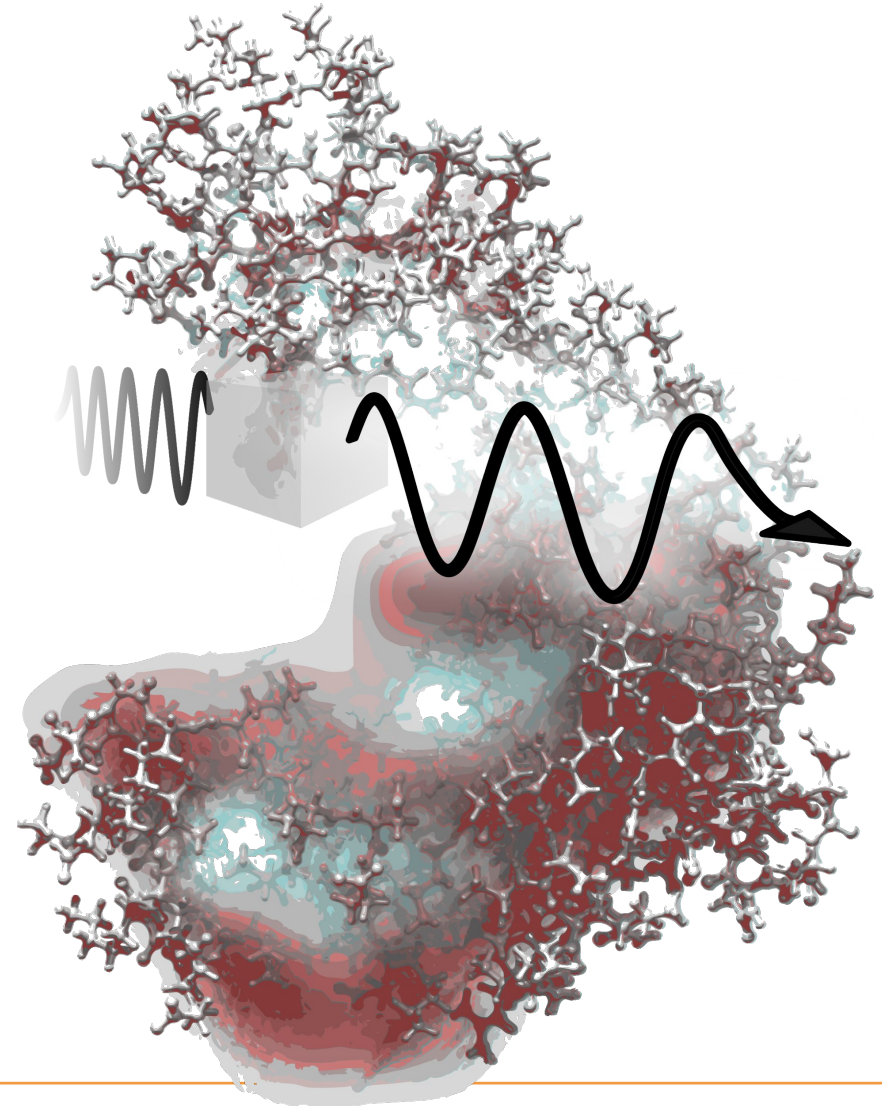
GROMACS 2020 - density-guided simulations

- The pipeline
 - set up a vanilla molecular dynamics simulation
 - add a reference density (.mrc or .ccp4 file)
 - choose balance between force field and density information
 - convergence usually within a day of walltime running



GROMACS 2020 - density-guided simulations

- Further features
 - force calculation derived from modeling underlying physics
 - different models of comparing densities (relative entropy, cross-correlation,...) allows fine-tuning depending on the research question
 - not only limited to cryo-EM, but applicable to any density / shape information



GROMACS API

- Running simulations from a python interface
- “Simple things should be simple, hard things not impossible”
- Vision: More features with less effort, shared in a community at maximum performance
- Requirement: Extensive code refactoring

Python command

Work graph

(a)

```
>>> md = gmx.from_file([filename1, filename2, filename3, ...])
```

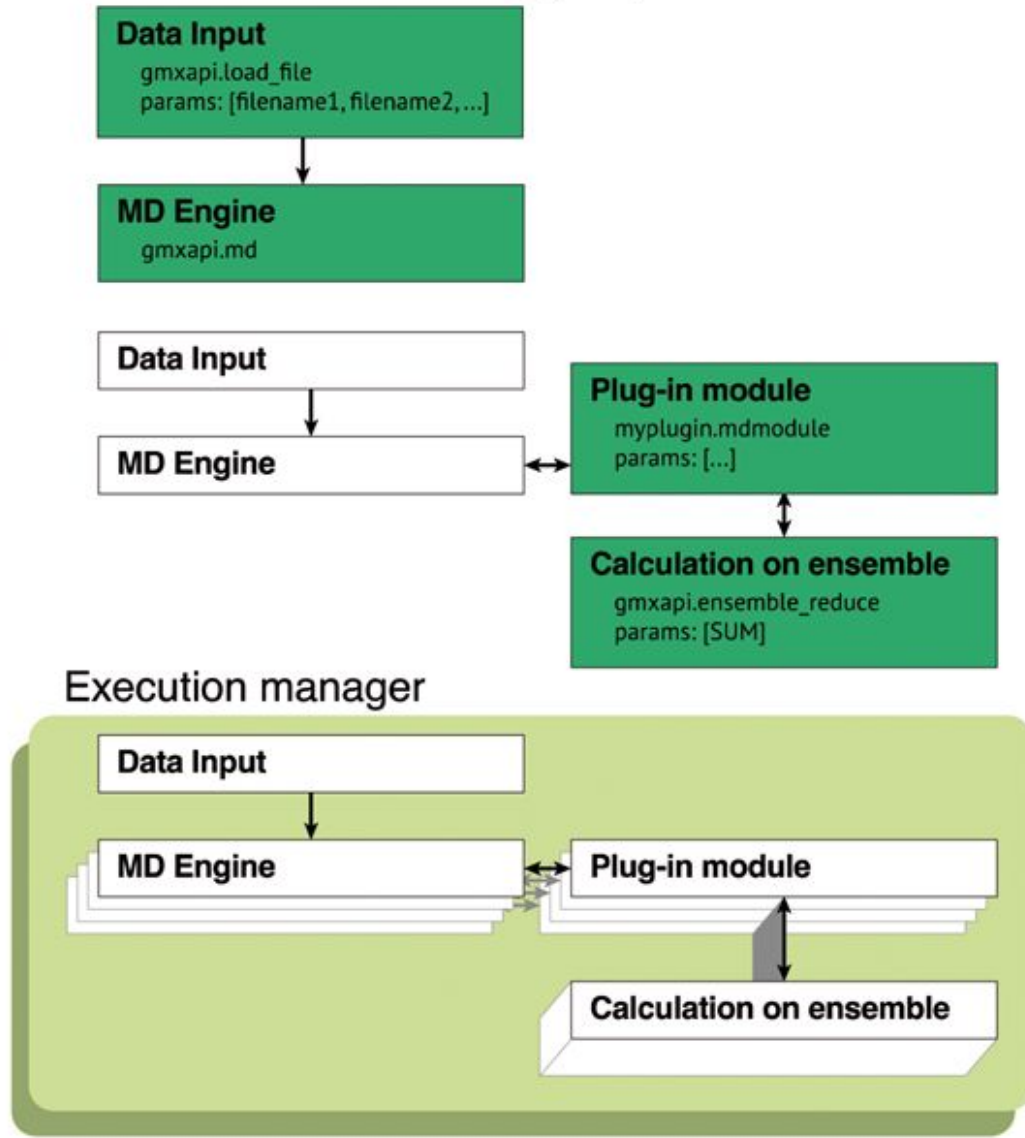
(b)

```
>>> potential = myplugin.EnsembleRestraint(sites, *args, **kwargs)
```

```
>>> md.add_dependency(potential)
```

(c)

```
>>> gmx.run()
```



GROMACS interfaces workflows

- Wrapping gromacs is (comparably) easy, maintaining performance and sanity is hard
- Reproducible science workflows require reproducible workflows
- Challenges
 - Define interfaces
 - Be general, but use specific hardware features
 - Manage data

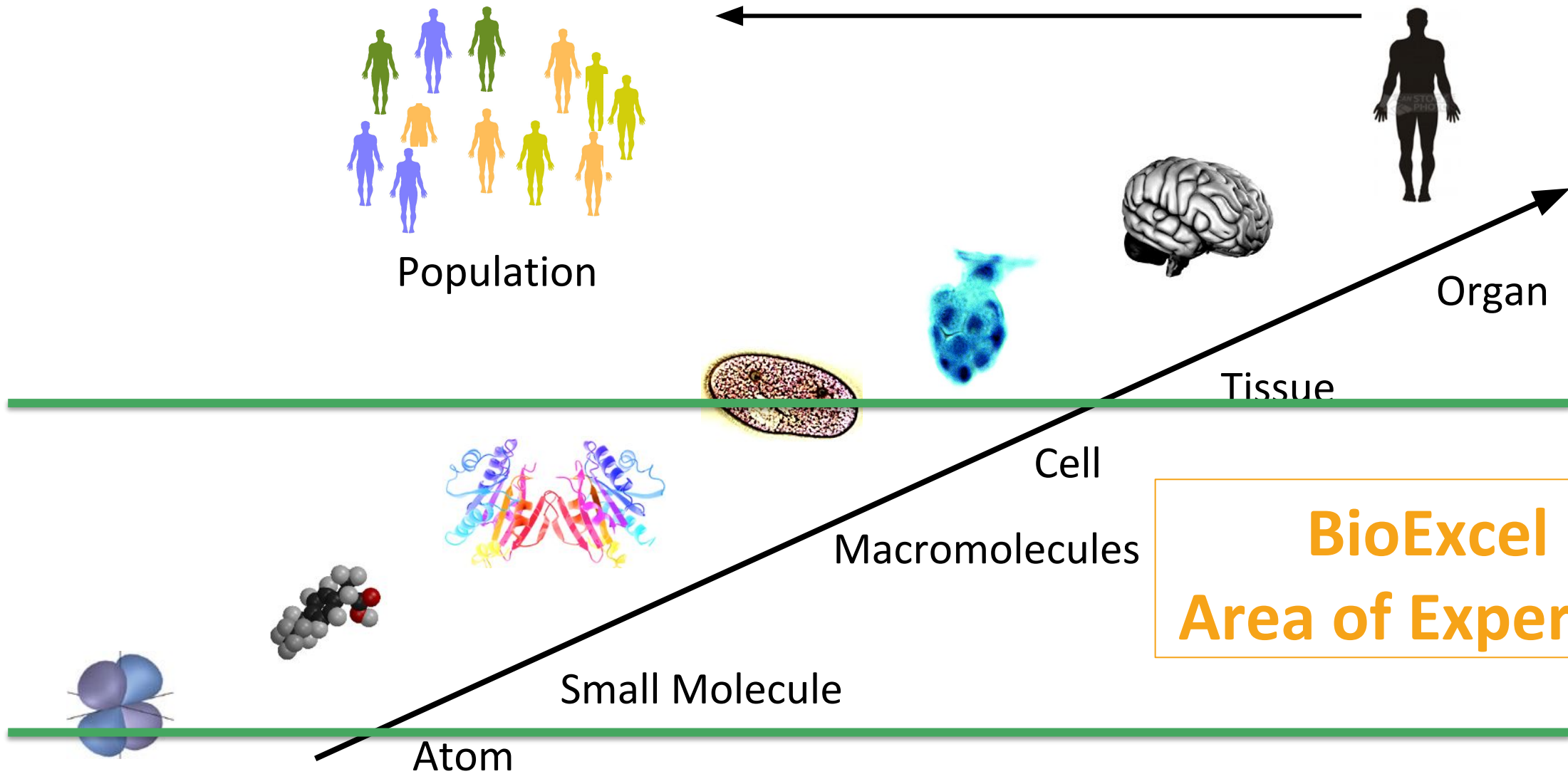
Who enables GROMACS

- Funding agencies - BioExcel, PRACE, ...
- GROMACS community - research groups, “Freelancers”
- Compute centers
- ...

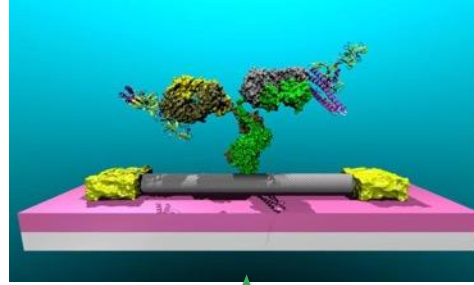
Future of GROMACS hinges on

- Code and Research
- Training and Dissemination
- Qualified personnel and funding

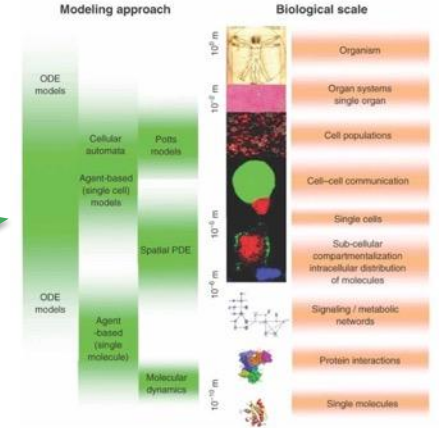
Life Science and HPC



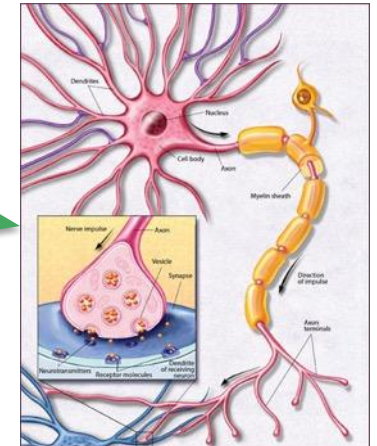
Biomarkers design



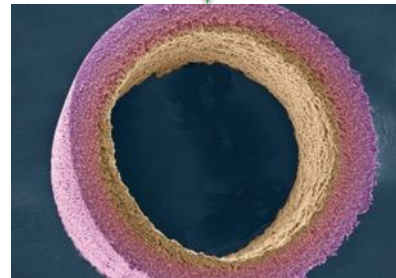
Physiology



Biomolecular Modeling and Simulations

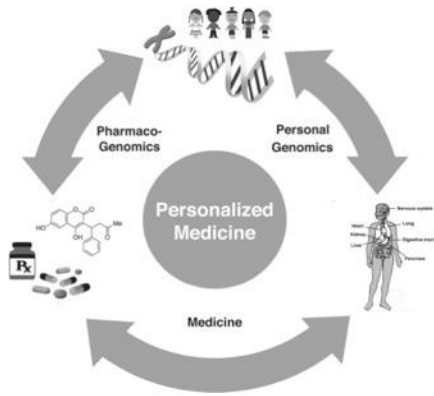
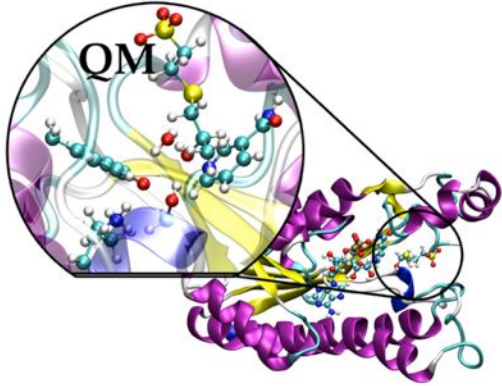


Neuroinformatics



Biomaterials science and

Electronic structure



Drug Design Personalized Medicine

BioExcel's Vision and Mission

A central hub for biomolecular modeling
and simulations

Enabling better science by **improving** the most popular biomolecular **software** and spreading best practices and expertise among the communities through **consultancy** and **training**

**Dissemination &
Training**

Excellence in Software

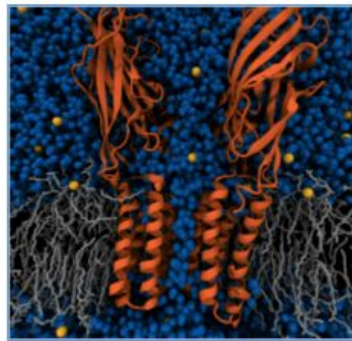
Excellence in Usability

Excellence in Support

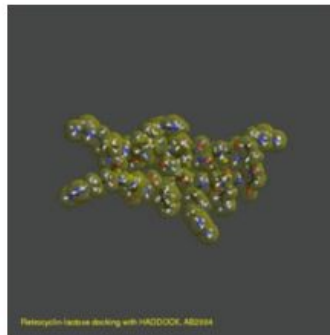
**Scientific Excellence &
User Driven Governance**

Improve the performance, efficiency and scalability of key codes

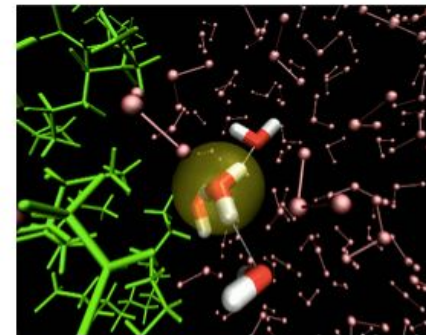
- GROMACS (Molecular Dynamics Simulations)
- HADDOCK (Integrative modeling of macro-assemblies)
- CP2K (hybrid QM/MM code for enzymatic reactions, photochemistry and electron transfer processes)



MD simulations
/GROMACS/



Docking
/HADDOCK/



QM/MM
/CP2K/

Training and workshops

- Closing the “knowledge gap”
 - show users what is available
 - transferring new software developments to users
 - enable users to apply the newest and most advanced features
 - ensure software is used in the best possible way
 - scientific correctness
 - high performance
 - ease-of-use
 - follow up on user questions
 - feed-back into the development process

Training and workshops

- BioExcel workshops

- focused events for 15-50 participants from beginners to advanced levels
- industry participation is encouraged
- combination of BioExcel tools
- selection of events organized or supported by BioExcel
 - Minisymposium: Biomolecular simulation in drug discovery (with Scilifelab), Stockholm
 - Alchemical Free Energy Workshop 2019, Göttingen, Germany
 - BioExcel/PRACE Seasonal School, Stockholm
 - BioExcel Summer School on Biomolecular Simulations, Pula, Italy
 - Advanced Gromacs, HADDOCK + PMX workshop, Espoo, Finland



Training and workshops

- Webinars

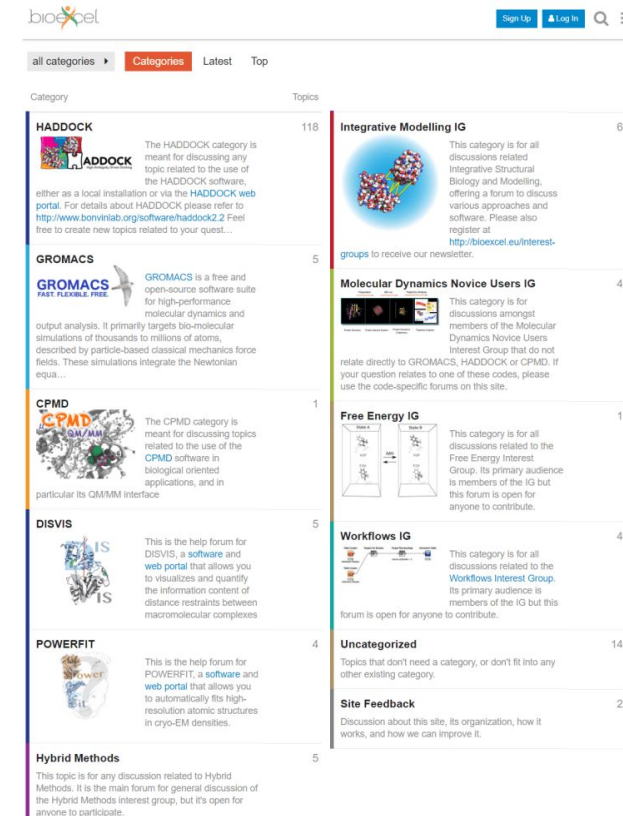
- dozens of sessions on all topics related to BioExcel from free energy calculations to software development at <https://bioexcel.eu/category/webinar/>
- Session every other week
 - enables discovery new aspects of BioExcel tools
 - deepen understanding of use

- BioExcel forum

- ask.bioexcel.eu

- Knowledge resource center with tutorials

- krc.bioexcel.eu



The screenshot shows the BioExcel forum website. At the top, there is a navigation bar with 'all categories', 'Categories', 'Latest', and 'Top' tabs. Below this is a table listing various categories and their corresponding topic counts. The categories include HADDOCK (118), GROMACS (5), CPMD (1), DISVIS (5), POWERFIT (4), Hybrid Methods (5), Integrative Modelling IG (6), Molecular Dynamics Novice Users IG (4), Free Energy IG (1), Workflows IG (4), Uncategorized (14), and Site Feedback (2). Each category entry includes a small icon and a brief description of the category's focus.

Category	Topics
HADDOCK The HADDOCK category is meant for discussing any topic related to the use of the HADDOCK software, either as a local installation or via the HADDOCK web portal. For details about HADDOCK please refer to http://www.bonvinlab.org/software/haddock2.2 . Feel free to create new topics related to your quest...	118
GROMACS GROMACS is a free and open-source software suite for high-performance molecular dynamics and output analysis. It primarily targets bio-molecular simulations of thousands to millions of atoms, described by particle-based classical mechanics force fields. These simulations integrate the Newtonian equa...	5
CPMD The CPMD category is meant for discussing topics related to the use of the CPMD software in biological oriented applications, and in particular its QMMM interface	1
DISVIS This is the help forum for DISVIS, a software and web portal that allows you to visualize and quantify the information content of distance restraints between macromolecular complexes	5
POWERFIT This is the help forum for POWERFIT, a software and web portal that allows you to automatically fit high-resolution atomic structures in cryo-EM densities.	4
Hybrid Methods This topic is for any discussion related to Hybrid Methods. It is the main forum for general discussion of the Hybrid Methods interest group, but it's open for anyone to participate.	5
Integrative Modelling IG This category is for all discussions related Integrative Structural Biology and Modelling, offering a forum to discuss various approaches and software. Please also register at http://bioexcel.eu/interest-groups to receive our newsletter.	6
Molecular Dynamics Novice Users IG This category is for discussions amongst members of the Molecular Dynamics Novice Users Interest Group that do not relate directly to GROMACS, HADDOCK or CPMD. If your question relates to one of these codes, please use the code-specific forums on this site.	4
Free Energy IG This category is for all discussions related to the Free Energy Interest Group. Its primary audience is members of the IG but this forum is open for anyone to contribute.	1
Workflows IG This category is for all discussions related to the Workflows Interest Group. Its primary audience is members of the IG but this forum is open for anyone to contribute.	4
Uncategorized Topics that don't need a category, or don't fit into any other existing category.	14
Site Feedback Discussion about this site, its organization, how it works, and how we can improve it.	2



BioExcel Partners 2019



Horizon 2020
European Union Funding
for Research & Innovation

BioExcel is funded by the European Union
Horizon 2020 program under grant
agreements 675728 and 823830.