

BioExcel Building Blocks Workflows (BioBB-Wfs), an integrated web-based platform for biomolecular simulations

Genís Bayarri¹, Pau Andrio², Adam Hospital^{1,*}, Modesto Orozco^{1,3,*} and Josep Lluís Gelpí^{2,3,*}

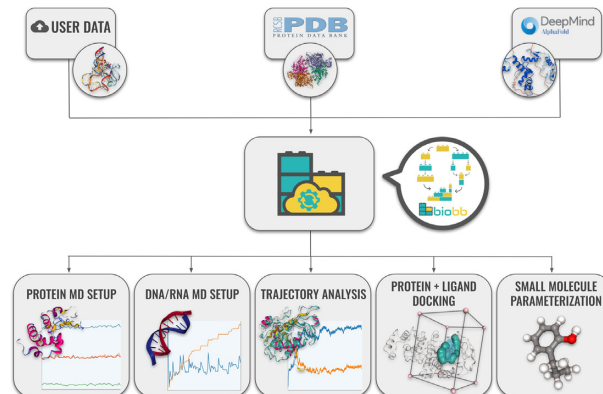
¹Institute for Research in Biomedicine (IRB Barcelona). The Barcelona Institute of Science and Technology. Baldri Reixac 10-12, 08028 Barcelona, Spain, ²Barcelona Supercomputing Center (BSC), Jordi Girona 29, 08034, Barcelona, Spain and ³Department of Biochemistry and Molecular Biology, University of Barcelona, 08028 Barcelona, Spain

Received March 29, 2022; Revised April 19, 2022; Editorial Decision April 28, 2022; Accepted May 02, 2022

ABSTRACT

We present BioExcel Building Blocks Workflows, a web-based graphical user interface (GUI) offering access to a collection of transversal pre-configured biomolecular simulation workflows assembled with the BioExcel Building Blocks library. Available workflows include Molecular Dynamics setup, protein-ligand docking, trajectory analyses and small molecule parameterization. Workflows can be launched in the platform or downloaded to be run in the users' own premises. Remote launching of long executions to user's available High-Performance computers is possible, only requiring configuration of the appropriate access credentials. The web-based graphical user interface offers a high level of interactivity, with integration with the NGL viewer to visualize and check 3D structures, MDsrv to visualize trajectories, and Plotly to explore 2D plots. The server requires no login but is recommended to store the users' projects and manage sensitive information such as remote credentials. Private projects can be made public and shared with colleagues with a simple URL. The tool will help biomolecular simulation users with the most common and repetitive processes by means of a very intuitive and interactive graphical user interface. The server is accessible at <https://mmb.irbbarcelona.org/biobb-wfs>.

GRAPHICAL ABSTRACT



INTRODUCTION

Experimental techniques have traditionally provided information about the dynamics of proteins and other biomolecules. Molecular dynamics (MD), originated in the late 1970s (1), were designed to be their computational counterpart, building a molecular-level understanding of biology. Theoretical methods were designed to analyse, model and predict biomolecular structures, mechanisms, dynamics and flexibility properties (2–5). Claimed now to be the ‘*computational microscope for molecular biology*’, the field has become a central piece for the design of biological molecules and their interactions, making increasingly significant contributions to biology in the recent years (6). The 21st century, in particular, has left a succession of breakthroughs in biomolecular simulations (7), including the first millisecond all-atom simulation of a protein folding (8), the study of entire virus (HIV-1) from all-atom simulations (9), the recent SARS-CoV-2 theoretical works (10–13) or the

*To whom correspondence should be addressed. Tel: +34 93 403 7155; Email: adam.hospital@irbbarcelona.org
Correspondence may also be addressed to Josep Lluís Gelpí. Tel: +34 934137730; Email: gelpi@ub.edu
Correspondence may also be addressed to Modesto Orozco. Email: modesto.orozco@irbbarcelona.org

coupling of MD simulations with artificial intelligence techniques (14).

Continuum improvements in force-fields, simulation protocols and software has converted MD into a mature technique that can be used effectively to understand macromolecular structure-to-function relationships (3), reaching simulation times close to biologically relevant ones. However, its practical use has been always hindered by its steep learning curve and the tedious series of actions required to prepare a reliable setup. As an example, the process of setting up a system requires a complex pipeline of operations, together with a non-negligible number of decisions, which require a significant degree of expertise. The problems are amplified as setups notably change depending on the MD software package used. Integrated software platforms, usually coupled to standalone or web-based Graphical User Interfaces (GUIs), have been designed to help in this endeavor. Although commercial licenses are required for the most complete packages (e.g. Schrödinger, BioVia, Acellera (15)), some free academic tools are also available.

CHARMM-GUI, with >9000 citations for the different related publications (16–20) at the point of writing this paper, deserves credit for being the most popular of these tools. Created in 2006, it offers interactive building of complex systems and inputs preparation with well-established and reproducible simulation protocols for state-of-the-art molecular simulations and widely used simulation packages through a web-based GUI; The Visual Molecular Dynamics (VMD) (21) tool, a popular 3D/4D molecule visualization program for MD trajectories, has a collection of plugins to setup and run MD simulations, including QwikMD (22), an integrative MD toolkit for novices and experts; Similarly, the PyMol visualizer, also has a plugin offering a GUI to setup GROMACS MD simulations: Dynamics (23,24); WebGRO for Macromolecular Simulations is a recent web-based tool developed to ease the setup process of protein and protein-ligand complexes with GROMACS; and the most recent tool, Making-it-rain (25), uses the power of Google Colab notebooks to prepare and run MD simulations using OpenMM engine and AMBER and CHARMM force fields. Finally MDWeb (26), developed in our group 10 years ago and still with 9000 unique users registered, is a web server providing a friendly environment to setup new systems, run test simulations and perform analysis within a guided interface, compatible with AMBER (27), NAMD (28,29) and GROMACS (30). All these platforms have facilitated the use of MD, but they fail to provide a complete and integrated single interface from which to build, launch, and control biomolecular workflows and associated analyses. They also fail to integrate the MD simulations into more complex working pipeline (interoperability), they lack portability and fail to launch in an automatic manner the calculation to HPC clusters (damaging scalability).

We present here BioBB-Wfs, a web-based GUI designed with modern web technology and powered by workflows built using the BioExcel Building Blocks (BioBB) library (31), developed under the framework of the BioExcel Centre of Excellence (<https://bioexcel.eu/>). BioBB is a collection of interoperable building blocks built as portable wrappers on top of common biomolecular simulation tools.

The interoperable building blocks can be easily joined together assembling complex computational biomolecular workflows, and can be shared thanks to its integration with the Conda Packaging system, ensuring reproducibility. Besides, scripts can be exported to Common Workflow Language (CWL) using the available CWL specifications. BioBB-Wfs smooths the learning curve of MD, and facilitates the access to interoperable and reproducible computational biomolecular simulations. The tool facilitates the integration of MD simulation protocols in more complex pipelines like those used in medicinal chemistry, biophysics or computational biology pipelines, including automatic small molecule parameterization, protein-ligand docking or protein MD analyses.

MATERIALS AND METHODS

BioExcel Building Blocks library (BioBB)

BioBB-Wfs is powered by workflows built using the BioExcel Building Blocks library (31). BioBB is a collection of portable wrappers of common biomolecular simulation tools. The BioBB library is designed to (i) increase the *interoperability* between the tools wrapped; (ii) *ease the implementation* of biomolecular simulation workflows and (iii) increase the *reusability and reproducibility* of the generated workflows. To achieve these main goals, the library was designed following the best practices for applying the FAIR (Findable, accessible, interoperable and reusable) principles to research software (32,33). The result is a collection of building block modules, divided in sets of tool wrappers focused on similar functionalities (e.g. Molecular Dynamics, Virtual Screening). BioBB is supported by known frameworks and tools: (i) software packaging (Pip, BioConda (34), BioContainers (35)), (ii) documentation (ReadTheDocs, <https://readthedocs.org/>), (iii) interactive tools (Jupyter Notebooks (36), myBinder, <https://mybinder.org/>), (iv) registry & findability (bio.tools (37), BioSchemas, WorkflowHub (38)), (v) workflow management systems (CWL, Galaxy (39,40), PyCOMPSs (41)), (vi) source code (GitHub, <https://github.com/>) and (vii) REST APIs (OpenAPI, <https://swagger.io/specification/>). Notably all building blocks follow the same pattern of installation, configuration and user interaction which facilitate their integrated use in complex workflows.

Workflows

Templates were built using the BioBB library as Python scripts and automatically converted to CWL workflow scripts thanks to the available BioBB CWL adapters. Templates are then used to generate the final workflow scripts with user's input files and parameters. Workflow scripts are divided in two files: Python/CWL file, with the code logic (building blocks, loops, conditionals) and a yaml-formatted file, with input parameters and dependencies between steps. Both workflow types are relying on software packaging for easy deployment and reproducibility, Conda packages for the Python ones, and Docker containers for the CWL. Instructions on how to install and run them in the users' own premises are included in the downloadable files. Workflows included in the current version of the server are listed in

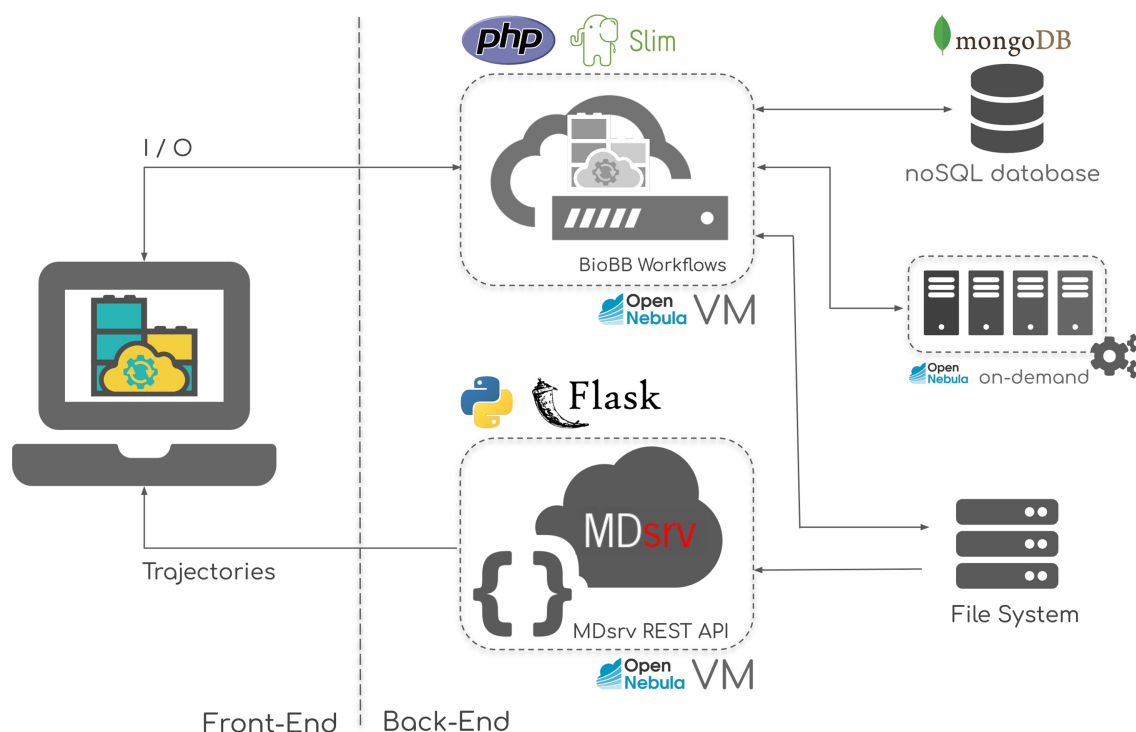


Figure 1. BioBB-Wfs Internal schema. The server is divided into two main blocks: front-end (left) and back-end (right).

Supplementary Table S1. All workflows were uploaded to the WorkflowHub repository (38) (<https://workflowhub.eu/>) and available from a unique DOI.

Remote execution

Remote execution of BioBB-Wfs is available for production MDs. MD setup processes are run in the server infrastructure, but the length is limited to 500 ps. Production (long) MDs using the files generated by the setup pipelines can then be launched through remote executions. Connection with external computing clusters is achieved through a module of the BioBB library (`biobb_remote`). The connection is established with a Secure Shell (SSH) protocol, with the server generating a specific key pair for the remote connection and saving them in the user's profile. Limitations of this feature must be noted: only servers accepting direct SSH sessions can be integrated, and only users with currently active access to the particular HPC clusters are allowed to launch jobs (the server is not giving free access to the HPC resources, but connecting to them). The current implementation of the server allows to users with credentials a direct connection with the Barcelona Supercomputing Center (BSC) supercomputers MareNostrum (CPU-based) and Minotauro (GPU-based). `Biobb_remote` library can be easily configured to integrate additional host computers or queuing systems.

SERVER DESIGN AND IMPLEMENTATION

Infrastructure

The server is divided in two main blocks: the front-end, embodied by the online accessible web server; and the back-

end, where user's data is stored and workflows are run (Figure 1).

The main BioBB-Wfs GUI web portal is implemented in PHP with the help of the Slim framework (<https://www.slimframework.com>). User interactivity is given by a set of JavaScript libraries and tools, including NGL viewer (42) to visualize and check 3D structures and Plotly to explore 2D plots. Backend resources are deployed on an OpenNebula private cloud infrastructure. MD trajectories are saved in a common storage device and streamed to the web server thanks to the MDsrv tool (43) and its associated REST API. Workflows are built and launched to a queue system connected to an on-demand configuration, where new VMs are dynamically deployed when needed. All data used by the server is stored in a noSQL distributed MongoDB database.

Internal design

The internal organization of the web server follows a modular design (Supplementary Figure S1). Workflow descriptions are the core of the structure, and from them, a list of components (e.g. input form, workflow schema, output interface) are generated. Some of the components are automatically generated, whereas others (like specific results interfaces) were exclusively designed. Results interfaces take advantage of the web interactivity to graphically display the generated data, following the new-generation web-based GUIs recommendations for biomolecular data described elsewhere (44). Key information for the workflows such as steps (building blocks), input files needed, accepted formats, etc. is organized in the MongoDB database and is queried by the PHP frontend. This structure allows easy integration of new workflows.

biobb

Home Workflows Help Contact My Projects

List of projects

Home • My Projects

5 GB

1AKI GROMACS Set Up 2.12% ABC MD Setup 1.42% 3HTB - JZ4 GROMACS Complex Set Up 2.85% Structural DNA helical parameters 16.37%

List of projects Filter by status: View All

20 records Search:

Name	Status	Creation date	Expiration date	Size	Actions
Structural DNA helical parameters	WF FINISHED	13/01/2022 12:37	PERSISTENT	838.28 MB	Project actions
3HTB - JZ4 GROMACS Complex Set Up	WF FINISHED	13/01/2022 12:27	PERSISTENT	145.80 MB	Project actions
ABC MD Setup	WF FINISHED	13/01/2022 12:25	PERSISTENT	72.63 MB	Project actions
1AKI GROMACS Set Up	WF FINISHED	13/01/2022 12:23	PERSISTENT	108.59 MB	Project actions

Showing 1 to 4 of 4 entries

Figure 2. BioBB-Wfs user's workspace. (A) List of projects; (B) user profile; (C) my projects instant access.

Personal workspace

The server provides users with a personal workspace where project's metadata and results of analysis can be stored. Although access to the server is free and most of the functionalities are available without a user account, registration is required to maintain a permanent workspace. Also, launching simulations to external machines is only available to registered users, due to the need to store the appropriate connection details and credentials. More advantages of having a user account are the possibility to have an organized list of projects listed in the workspace (Figure 2), create private (only visible by the registered user) and persistent projects (with no expiration date), rerun or clone existing projects. In the present implementation, registered users are provided with 5 GB of storage space, expandable upon request.

For each of the listed projects in the workspace (Figure 2A), useful information such as the status of the project (queued, running, finished), the size, and the creation and expiration dates are shown. A dropdown button gives access to a list of utilities including links to the summary and results page, to the downloadable results and workflow scripts, and to the project settings and log info. Used and available storage is displayed at the top of the page.

User profile can be accessed from the circle button placed in the top-right part of the interface (Figure 2B). Personal information, password and SSH keys to access to external HPC hosts can be managed here. A 'My projects' tab gives instant access to the user workspace (Figure 2C).

SERVER USAGE

BioBB-Wfs uses a uniform way of working for the whole collection of workflows. The process is always divided into three main steps: (i) project creation; (ii) biomolecular workflows settings configuration and launch and (iii) workflow results.

Project creation

Users can start a new project either by (i) choosing a desired workflow (see Supplementary Table S1) or (ii) selecting an input data type (macromolecular structure, DNA/RNA sequence, small molecule, MD trajectory and protein + ligand structures).

When selecting an input type, a set of specific intermediate steps to extract required or useful information is presented. The provided information can be displayed and confirmed before using it in the actual workflow execution. As an example, the intermediate steps for the Protein-Ligand Docking workflow allows to choose between different pockets computed on the fly in the input structure (Supplementary Figure S2). Another example is an intermediate step relevant to all workflows working from a macromolecular structure: the structure-checking process (Figure 3). The interface, that incorporates an interactive NGL viewer, on one side, allows to select the components of the structures to be included in the simulation (e.g. chains, alternative conformations), and on the other, shows a comprehensive series of possible issues (e.g. wrong chirality or amide atoms as-

CHECK STRUCTURE

System Configuration

- Detect/Select Models: Single one
- Detect/Select Chains: Select chain(s)
- Detect/Select Alt Locations
- Detect/Remove Metals
- Detect/Remove Ligands
- This structure doesn't contain DNA / RNA

Fix Structure Errors

- Detect/Fix Amide Assignment

Structure Warnings

- Detect SS Bonds
- Unusual cis/trans backbone
- Modified residues
- Backbone breaks
- Missing atoms
- Severe clashes
- Polar donor clashes
- Polar acceptor clashes
- Apolar clashes
- Positive clashes
- Negative clashes

Background: Grey background
Style: Cartoon
Color: By Secondary Structure
Ligand: Ball and Stick
 Water Ions

Figure 3. BioBB-Wfs project creation. Example of intermediate step: macromolecule structure checking.

signments, or VdW clashes) in the chosen structure that can affect MD simulation. When available, solutions to such issues are offered to the user and can be confirmed on the interactive display.

After the intermediate steps, a final workflow settings input form permits the user to change specific parameters for the selected workflow. As an example, parameters such as force field, solvent water type and box type/size can be changed in all workflows performing an MD setup. If more than one workflow is compatible with the input type of the project, a dropdown selector will give the possibility to choose the one desired.

Biomolecular workflows

Pre-configured workflows currently integrated in the web server cover the basic needs of a newcomer in the field. They include MD simulation setup and analysis, and protein-ligand docking. The collection of current available workflows is listed in Supplementary Table S1. Examples are: (i) complete setups for GROMACS or AMBER pack-

ages, generating a system with the molecule immersed in a box of water and counterions, energetically minimized and equilibrated; (ii) modelling protein mutations, adding to the setup pipeline the possibility to mutate one or more residues/nucleotides from the protein/nucleic acid structure; (iii) automatic ligand parameterization, generating force-field library files for small molecules to be included in MD simulations and (iv) protein MD analyses, performing quality control (QC) analysis on an uploaded trajectory, including root mean square deviations (RMSd), radius of gyration (Rgyr), atomic fluctuations and trajectory clustering.

Nucleic acids structures have their own set of specific workflows. The *ABC DNA/RNA MD setup* pipeline, provides a reliable workflow for Nucleic Acids setup, following the expert recommendations by the ABC consortium (45). The *Structural DNA helical parameters* performs an exhaustive flexibility exploration of a DNA/RNA trajectory, including helical parameters correlations and bimodality analysis, in line with the latest ABC studies (46–49).

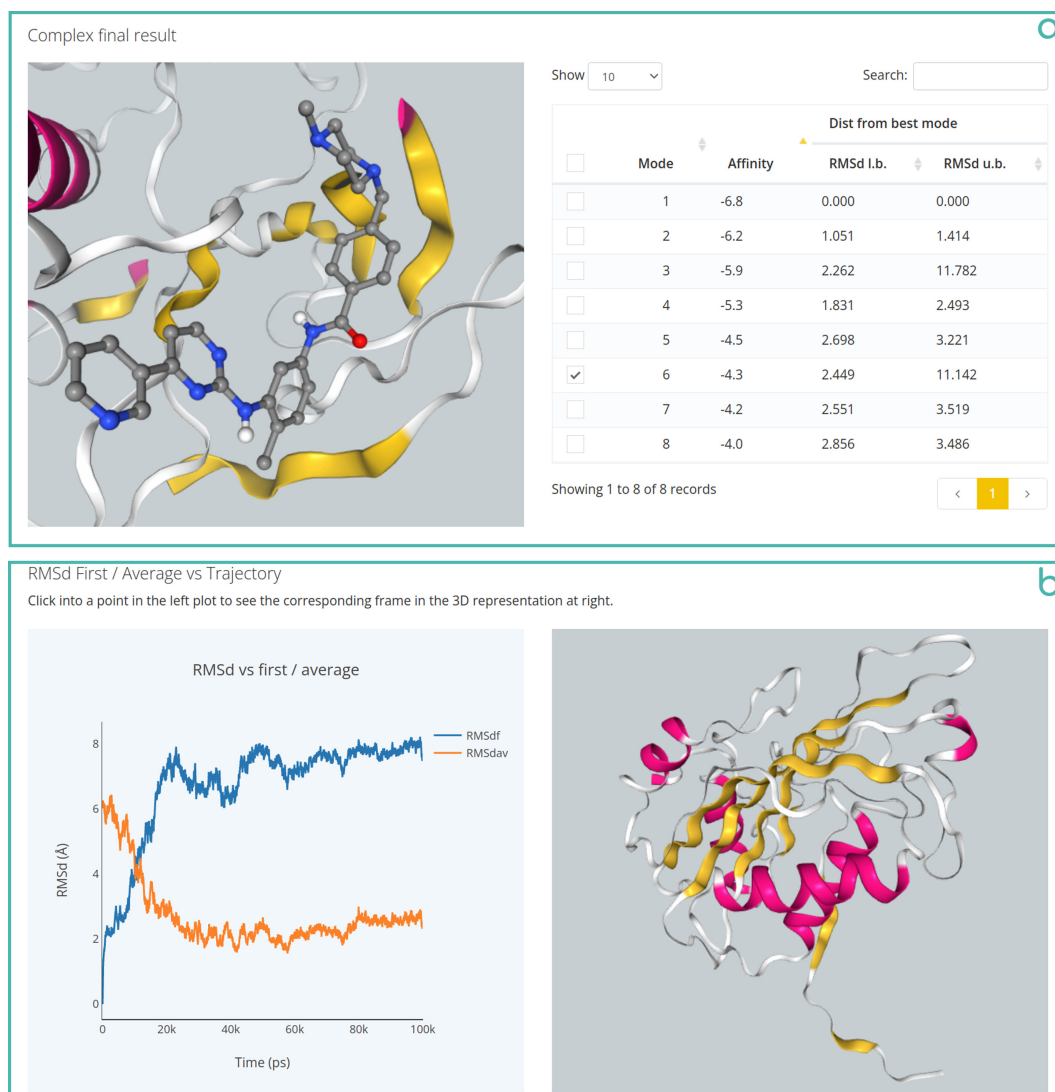


Figure 4. Workflow results examples. (A) Protein–ligand docking outputs: best ligand poses; docking experiments were done with Autodock Vina. (B) Protein MD analysis: RMSd values against first and average structure. An associated NGL viewer presents the results in 3D representation.

For each of the workflows, a list of the main steps and default input parameters are documented. Besides, an interactive schema gives a quick and graphical idea on the pipeline steps and its dependencies (Supplementary Figure S3). The diagram is also displaying the BioBB's used for each of the steps linked to their official documentation.

Workflow results

Results of the workflows executions are summarized in a final common section. Information provided includes, besides of the execution results, a summary of the project settings, possible project actions (such as downloading files, re-running, cloning, making it persistent or sharing the project), and the workflow execution log with provenance data. Projects can be shared with other users through a specific URL link as view-only or public (project can be forked).

The central section of the workflows output page is the analysis section that is represented in a graphical and in-

teractive way. Different custom analysis interfaces are integrated, specific for each of the workflows. Results are shown with the graphical support of NGL 3D visualizer and plotly online graphing. As examples, results of protein–ligand docking and MD trajectory analysis are shown in Figure 4. Results of the docking procedure display a table with the best ligand poses, which includes affinities and distances from the best mode (Figure 4A). The table is connected to an NGL visualizer, allowing the representation of individual or groups of selected poses. Protein MD trajectory analysis shows RMSd, Rgyr, atomic fluctuation and cluster families with 2D plots and 3D interactive representations. 2D plots are linked to the 3D representations, where snapshots with interesting analysis values can be explored with just a click on the 2D plot lines (Figure 4B).

Examples of use are included in the supplementary material (use cases, Supplementary Figures S4 and S5), tutorials section of the help pages, and the corresponding projects are available through the demo user workspace.

DISCUSSION

MD simulations are widely used in a repertoire of fields from biophysics to biotechnology and from medicinal chemistry to personalized medicine. The consequence of this wide use is a number of scientists applying the technique with just a superficial training in the method. The biomolecular field has been developing freely available online tools for the last 15 years, trying to help this community. However, these automatic tools have raised a controversial discussion in the field: Is the use of the hidden, black box workflows behind these tools, a good practice? The question is not trivial, taking into consideration the complex machinery behind MD simulations and the difference in treatment depending on the system being studied. BioBB-Wfs is designed to facilitate the use of MD to non-experts, facilitate the integration of MD into more complex research pipelines and allows tailored configurations to escape from the ‘black-box’ paradigm implicit to most of automatic setup procedures. There are pre-configured biomolecular workflows, but they are extensively described, step-by-step, with diagrams showing the connections between them. Workflow scripts can be checked and downloaded, in two different versions, a Python script and a CWL definition. The possibility of downloading the workflow scripts opens the opportunity to: (i) reproduce the result launching the execution in the user’s own premises; and (ii) modify/change/extend the workflow until it fits the user’s preferences.

Another point of common misunderstanding has been how an on-line server could offer computationally expensive MD simulations for free. This deserves careful attention. Online servers offering almost on-the-fly macromolecular flexibility properties are mostly using Normal Mode analysis or Coarse-Grained methods, which are more efficient in terms of time, at a cost of reducing the molecule resolution (50–52). Servers offering true atomistic simulations, including BioBB-Wfs, cannot give immediate results and can hardly afford expensive production state-of-the-art simulation runs. These platforms have been always limited to the preparation of the system (MD setup) and the generation of the configuration files needed in a production MD run. Where this final simulation is run is up to the user and his/her possibility to access to HPC resources. BioBB-Wfs is filling this gap with an additional feature: to perform the simulation on user’s behalf in a selected HPC facility, in a transparent manner. Hence, BioBB-Wfs allows executing a complete MD experiment (MD setup, run, analysis) in the same GUI, reaching, if user’s credentials allow it, to reach Exascale machines.

This integration of different biomolecular simulation workflows and the possibility to run the whole production pipeline, including the connection with an external HPC cluster, brings our platform closer to the available commercial tools. Pharmaceutical companies stated their interest in such platforms, as a free alternative to the extremely expensive integrated packages. However, the concerns here are the data privacy. An on-line server is not an acceptable option, but a containerized version of the server running in local premises is feasible. The whole machinery behind the BioBB-Wfs server can be installed in a single Virtual Ma-

chine (VM) or container (Docker/Singularity) to be then deployed in a private infrastructure.

One of the most critical issues that biomolecular simulation tools must tackle is the quick and constant update of the MD software packages (for instance GROMACS publishes a new release every few months), challenging the update of automatic setup tools. Newer versions are not always back compatible, and they might not recognize files produced by older versions. Moreover, newer versions often come with new functionalities increasing for example the efficiency, which are recommended by the developers and should be used accordingly. Flexibility is then crucial for tools offering biomolecular simulation workflows. BioBB-Wfs is relying on the BioBB library, a project that is considered now mature (recently recognized by the European Commission as an Excellent Science Innovation). The packaging and reproducibility aspects of the BioBB building blocks ensure its consistency and compatibility with the corresponding tools versions, and new releases every 4 months ensure workflows can use the most updated software. At the same time, provenance data included in workflow results always assure a clear reference to the software version used on each invocation. Crucial for a widespread use out of the MD community, the flexibility of the BioBB-Wfs server allows easy addition of new biomolecular workflows, opening a plethora of possibilities for biomolecular simulations, with the combination of the different available modules (machine learning, virtual screening, molecular interaction potentials, etc.).

CONCLUSION

BioBB-Wfs platform takes advantage of the power of new hardware, web technologies and network bandwidth to bring biomolecular simulations closer to the scientific community. The platform provides to non-expert users a set of trustable pre-configured workflows performing transversal biomolecular studies without requiring a deep knowledge of the details involved. At the same time, the availability of detailed documentation allows experienced users to adapt the workflows at their specific needs. The possibility to remotely launch computationally demanding processes to external clusters opens the door to perform complete simulation experiments from a free of use sophisticated interactive graphical interface, which helps performing their tasks efficiently and ensures reproducibility. The growing number of modules being integrated in the BioBB library behind the server admits regular updates and addition of new workflows and functionalities.

DATA AVAILABILITY

This website is free and open to all users. Upload and analysis of sensitive information requires a login.

BioExcel Building Blocks (BioBB) library: <https://mmb.ibbarcelona.org/biobb/>

BioBB remote: https://github.com/bioexcel/biobb_remote

BioBB-Wfs Workflows: https://github.com/bioexcel/biobb_workflows

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We are indebted to the BioExcel partners UNIMAN and EMBL-EBI for many discussions about BioBB and CWL workflows. We thank the BSC-INB Computational Group for useful discussions about BioBB workflows. We are also grateful to the developers of all the open source software tools wrapped by the BioBB library (Gromacs, AmberTools, ACPype, AutoDock Vina, OpenBabel, Curves+, Canal), which are being used in the server.

FUNDING

Spanish Ministry of Science [RTI2018-096704-B-100]; Catalan SGR [2017SGR1110]; European Union's Horizon 2020 research and innovation program [BioExcel-2 project, 823830]; SIFECAT Ris3Cat Clúster Emergente del Cerebro Humano (CECH) [001-P-001682]; Fondo Europeo de Desarrollo Regional (FEDER); Instituto Nacional de Bioinformática, ISCIII, Biomolecular and Bioinformatics Resources Platform [ISCIII PT17/0009/0007 and PT17/0008/0001]; Fondo Europeo de Desarrollo Regional (FEDER); MINECO Severo Ochoa Award of Excellence (Government of Spain) (to IRB Barcelona). Funding for open access charge: BioExcel Centre of Excellence for Computational Biomolecular Research, EU-funded project [823830].

Conflict of interest statement. None declared.

REFERENCES

- McCammon, J.A., Gelin, B.R. and Karplus, M. (1977) Dynamics of folded proteins. *Nature*, **267**, 585–590.
- Mulholland, A.J. (2008) Introduction. Biomolecular simulation. *J. R. Soc. Interface*, **5**(Suppl. 3), S169–S172.
- Hospital, A., Goñi, J.R., Orozco, M. and Gelpi, J.L. (2015) Molecular dynamics simulations: advances and applications. *Adv. Appl. Bioinform. Chem.*, **10**, 37–47.
- Dror, R.O., Dirks, R.M., Grossman, J.P., Xu, H. and Shaw, D.E. (2012) Biomolecular simulation: a computational microscope for molecular biology. *Annu. Rev. Biophys.*, **41**, 429–452.
- Mura, C. and McAnany, C.E. (2014) An introduction to biomolecular simulations and docking. *Mol. Simul.*, **40**, 732–764.
- Huggins, D.J., Biggin, P.C., Dämgen, M.A., Essex, J.W., Harris, S.A., Henchman, R.H., Khalid, S., Kuzmanic, A., Laughton, C.A., Michel, J. *et al.* (2019) Biomolecular simulations: from dynamics and mechanisms to computational assays of biological activity. *WIREs Comput. Mol. Sci.*, **9**, e1393.
- Schlick, T. and Portillo-Ledesma, S. (2021) Biomolecular modeling thrives in the age of technology. *Nat. Comput. Sci.*, **1**, 321–331.
- Lindorff-Larsen, K., Piana, S., Dror, R.O. and Shaw, D.E. (2011) How fast-folding proteins fold. *Science*, **334**, 517–520.
- Perilla, J.R. and Schulten, K. (2017) Physical properties of the HIV-1 capsid from all-atom molecular dynamics simulations. *Nat. Commun.*, **8**, 15959.
- Koehler, M., Ray, A., Moreira, R.A., Juniku, B., Poma, A.B. and Alsteens, D. (2021) Molecular insights into receptor binding energetics and neutralization of SARS-CoV-2 variants. *Nat. Commun.*, **12**, 6977.
- Pedebos, C. and Khalid, S. (2022) Simulations of the spike: molecular dynamics and SARS-CoV-2. *Nat. Rev. Microbiol.*, **20**, 192.
- Ali, A. and Vijayan, R. (2020) Dynamics of the ACE2-SARS-CoV-2/SARS-CoV spike protein interface reveal unique mechanisms. *Sci. Rep.*, **10**, 14214.
- Sztain, T., Ahn, S.H., Bogetti, A.T., Casalino, L., Goldsmith, J.A., Seitz, E., McCool, R.S., Kearns, F.L., Acosta-Reyes, F., Maji, S. *et al.* (2021) A glycan gate controls opening of the SARS-CoV-2 spike protein. *Nat. Chem.*, **13**, 963–968.
- Glielmo, A., Husic, B.E., Rodriguez, A., Clementi, C., Noé, F. and Laio, A. (2021) Unsupervised learning methods for molecular simulation data. *Chem. Rev.*, **121**, 9722–9758.
- Doerr, S., Harvey, M.J., Noé, F. and De Fabritiis, G. (2016) HTMD: high-throughput molecular dynamics for molecular discovery. *J. Chem. Theory Comput.*, **12**, 1845–1852.
- Lee, J., Cheng, X., Swails, J.M., Yeom, M.S., Eastman, P.K., Lemkul, J.A., Wei, S., Buckner, J., Jeong, J.C., Qi, Y. *et al.* (2016) CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. *J. Chem. Theory Comput.*, **12**, 405–413.
- Kim, S., Lee, J., Jo, S., Brooks, C.L., Lee, H.S. and Im, W. (2017) CHARMM-GUI ligand reader and modeler for CHARMM force field generation of small molecules. *J. Comput. Chem.*, **38**, 1879–1886.
- Jo, S., Cheng, X., Lee, J., Kim, S., Park, S.J., Patel, D.S., Beaven, A.H., Lee, K.I., Rui, H., Park, S. *et al.* (2017) CHARMM-GUI 10 years for biomolecular modeling and simulation. *J. Comput. Chem.*, **38**, 1114–1124.
- Park, S., Choi, Y.K., Kim, S., Lee, J. and Im, W. (2021) CHARMM-GUI membrane builder for lipid nanoparticles with ionizable cationic lipids and PEGylated lipids. *J. Chem. Inf. Model.*, **61**, 5192–5202.
- Choi, Y.K., Park, S.J., Park, S., Kim, S., Kern, N.R., Lee, J. and Im, W. (2021) CHARMM-GUI polymer builder for modeling and simulation of synthetic polymers. *J. Chem. Theory Comput.*, **17**, 2431–2443.
- Humphrey, W., Dalke, A. and Schulten, K. (1996) VMD: visual molecular dynamics. *J. Mol. Graph.*, **14**, 33–38.
- Ribeiro, J.V., Bernardi, R.C., Rudack, T., Stone, J.E., Phillips, J.C., Freddolino, P.L. and Schulten, K. (2016) QwikMD - Integrative molecular dynamics toolkit for novices and experts. *Sci. Rep.*, **6**, 26536.
- Makarewicz, T. and Kazmierkiewicz, R. (2013) Molecular dynamics simulation by GROMACS using GUI plugin for PyMOL. *J. Chem. Inf. Model.*, **53**, 1229–1234.
- Makarewicz, T. and Kazmierkiewicz, R. (2016) Improvements in GROMACS plugin for PyMOL including implicit solvent simulations and displaying results of PCA analysis. *J. Mol. Model.*, **22**, 109.
- Arantes, P.R., Polêto, M.D., Pedebos, C. and Ligabue-Braun, R. (2021) Making it rain: cloud-based molecular simulations for everyone. *J. Chem. Inf. Model.*, **61**, 4852–4856.
- Hospital, A., Andrio, P., Fenollosa, C., Cicin-Sain, D., Orozco, M. and Gelpi, J.L. (2012) MDWeb and MDMoby: an integrated web-based platform for molecular dynamics simulations. *Bioinformatics*, **28**, 1278–1279.
- Case, D.A., Cheatham, T.E., Darden, T., Gohlke, H., Luo, R., Merz, K.M., Onufriev, A., Simmerling, C., Wang, B. and Woods, R.J. (2005) The Amber biomolecular simulation programs. *J. Comput. Chem.*, **26**, 1668–1688.
- Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kalé, L. and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem.*, **26**, 1781–1802.
- Phillips, J.C., Hardy, D.J., Maia, J.D.C., Stone, J.E., Ribeiro, J.V., Bernardi, R.C., Buch, R., Fiorin, G., Hénin, J., Jiang, W. *et al.* (2020) Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J. Chem. Phys.*, **153**, 044130.
- Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B. and Lindahl, E. (2015) GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, **1–2**, 19–25.
- Andrio, P., Hospital, A., Conejero, J., Jordá, L., Del Pino, M., Codo, L., Soiland-Reyes, S., Goble, C., Lezzi, D., Badiá, R.M. *et al.* (2019) BioExcel Building Blocks, a software library for interoperable biomolecular simulation workflows. *Sci Data*, **6**, 169.
- Lamprecht, A.-L., Garcia, L., Kuzak, M., Martinez, C., Arcila, R., Martin Del Pico, E., Dominguez Del Angel, V., van de Sandt, S., Ison, J., Martinez, P.A. *et al.* (2019) Towards FAIR principles for research software. *Data Sci.*, **3**, 37–59.
- Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B.,

- Bourne, P.E. *et al.* (2016) The FAIR guiding principles for scientific data management and stewardship. *Sci. Data*, **3**, 160018.
34. Grüning, B., Dale, R., Sjödin, A., Chapman, B.A., Rowe, J., Tomkins-Tinch, C.H., Valieris, R., Köster, J. and Team, B. (2018) Bioconda: sustainable and comprehensive software distribution for the life sciences. *Nat. Methods*, **15**, 475–476.
 35. da Veiga Leprevost, F., Grüning, B.A., Alves Afifitos, S., Röst, H.L., Uszkoreit, J., Barsnes, H., Vaudel, M., Moreno, P., Gatto, L., Weber, J. *et al.* (2017) BioContainers: an open-source and community-driven framework for software standardization. *Bioinformatics*, **33**, 2580–2582.
 36. Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S. *et al.* (2016) In: Loizides, F. and Schmidt, B. (eds) *Positioning and Power in Academic Publishing: Players, Agents and Agendas*. pp. 87–90.
 37. Ison, J., Rapacki, K., Ménager, H., Kalaš, M., Rydza, E., Chmura, P., Anthon, C., Beard, N., Berka, K., Bolser, D. *et al.* (2016) Tools and data services registry: a community effort to document bioinformatics resources. *Nucleic Acids Res.*, **44**, D38–D47.
 38. Goble, C., Soiland-Reyes, S., Bacall, F., Owen, S., Williams, A., Eguinoa, I., Drosesbeke, B., Leo, S., Pireddu, L., Rodríguez-Navas, L. *et al.* (2021) Implementing FAIR Digital Objects in the EOSC-Life Workflow Collaboratory. *Zenodo*, <https://zenodo.org/record/4605654>.
 39. Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Eberhard, C. *et al.* (2016) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res.*, **44**, W3–W10.
 40. Goecks, J., Nekrutenko, A., Taylor, J. and Team, G. (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.*, **11**, R86.
 41. Tejedor, E., Becerra, Y., Alomar, G., Queralt, A., Badia, R.M., Torres, J., Cortes, T. and Labarta, J. (2017) PyCOMPS: parallel computational workflows in Python. *Int. J. High Perform. Comput. Appl.*, **31**, 66–82.
 42. Rose, A.S., Bradley, A.R., Valasatava, Y., Duarte, J.M., Prlić, A. and Rose, P.W. (2018) NGL viewer: web-based molecular graphics for large complexes. *Bioinformatics*, **34**, 3755–3758.
 43. Tiemann, J.K.S., Guixà-González, R., Hildebrand, P.W. and Rose, A.S. (2017) MDsrv: viewing and sharing molecular dynamics simulations on the web. *Nat. Methods*, **14**, 1123–1124.
 44. Hospital, A., Battistini, F., Soliva, R., Gelpí, J.L. and Orozco, M. (2020) Surviving the deluge of biosimulation data. *WIREs Comput. Mol. Sci.*, **10**, e1449.
 45. Roe, D.R. and Brooks, B.R. (2020) A protocol for preparing explicitly solvated systems for stable molecular dynamics simulations. *J. Chem. Phys.*, **153**, 054123.
 46. Dans, P.D., Balaceanu, A., Pasi, M., Patelli, A.S., Petkevičiūtė, D., Walther, J., Hospital, A., Bayarri, G., Lavery, R., Maddocks, J.H. *et al.* (2019) The static and dynamic structural heterogeneities of B-DNA: extending Calladine-Dickerson rules. *Nucleic Acids Res.*, **47**, 11090–11102.
 47. Pasi, M., Maddocks, J.H., Beveridge, D., Bishop, T.C., Case, D.A., Cheatham, T., Dans, P.D., Jayaram, B., Lankas, F., Laughton, C. *et al.* (2014) μ ABC: a systematic microsecond molecular dynamics study of tetranucleotide sequence effects in B-DNA. *Nucleic Acids Res.*, **42**, 12272–12283.
 48. Balaceanu, A., Buitrago, D., Walther, J., Hospital, A., Dans, P.D. and Orozco, M. (2019) Modulation of the helical properties of DNA: next-to-nearest neighbour effects and beyond. *Nucleic Acids Res.*, **47**, 4418–4430.
 49. da Rosa, G., Grille, L., Calzada, V., Ahmad, K., Arcon, J.P., Battistini, F., Bayarri, G., Bishop, T., Carloni, P., Cheatham III, T. *et al.* (2021) Sequence-dependent structural properties of B-DNA: what have we learned in 40 years? *Biophys. Rev.*, **13**, 995–1005.
 50. Marchetto, A., Si Chaib, Z., Rossi, C.A., Ribeiro, R., Pantano, S., Rossetti, G. and Giorgetti, A. (2020) CGMD platform: integrated web servers for the preparation, running, and analysis of coarse-grained molecular dynamics simulations. *Molecules*, **25**, 5934.
 51. Tiwari, S.P., Fuglebakk, E., Hollup, S.M., Skjærven, L., Cragnolini, T., Grindhaug, S.H., Tekle, K.M. and Reuter, N. (2014) WEBnm@ v2.0: web server and services for comparing protein flexibility. *BMC Bioinf.*, **15**, 427.
 52. Camps, J., Carrillo, O., Emperador, A., Orellana, L., Hospital, A., Rueda, M., Cicin-Sain, D., D’Abramo, M., Gelpí, J.L. and Orozco, M. (2009) FlexServ: an integrated tool for the analysis of protein flexibility. *Bioinformatics*, **25**, 1709–1710.