Check for updates

Structural Biology in the Clouds: The WeNMR-EOSC Ecosystem

Rodrigo V. Honorato[1], Panagiotis I. Koukos[1], Brian Jiménez-García[1], Andrei Tsaregorodtsev[2], Marco Verlato[3], Andrea Giachetti[4], Antonio Rosato[4] and Alexandre M. J. J. Bonvin[1]*

[1]Bijvoet Centre for Biomolecular Research, Faculty of Science, Department of Chemistry, Utrecht University, Utrecht, Netherlands, [2]Aix Marseille University, CNRS/IN2P3, CPPM, Marseille, France, [3]INFN, Sezione di Padova, Padova, Italy, [4]Department of Chemistry and Magnetic Resonance Center, University of Florence, and C.I.R.M.M.P, Fiorentino, Italy

Structural biology aims at characterizing the structural and dynamic properties of biological macromolecules at atomic details. Gaining insight into three dimensional structures of biomolecules and their interactions is critical for understanding the vast majority of cellular processes, with direct applications in health and food sciences. Since 2010, the WeNMR project (www.wenmr.eu) has implemented numerous web-based services to facilitate the use of advanced computational tools by researchers in the field, using the high throughput computing infrastructure provided by EGI. These services have been further developed in subsequent initiatives under H2020 projects and are now operating as Thematic Services in the European Open Science Cloud portal (www.eosc-portal.eu), sending >12 millions of jobs and using around 4,000 CPU-years per year. Here we review 10 years of successful e-infrastructure solutions serving a large worldwide community of over 23,000 users to date, providing them with user-friendly, web-based solutions that run complex workflows in structural biology. The current set of active WeNMR portals are described, together with the complex backend machinery that allows distributed computing resources to be harvested efficiently.

Keywords: structural biology, distributed computing, web portal, e-infrastructure, web services

INTRODUCTION

Proteins and nucleic acids are the main biological macromolecules responsible for function and maintenance of most of the machinery of life. The study of these macromolecules, their three-dimensional (3D) structure, dynamical behavior and interactions are crucial to gain a better understanding of relevant biological processes both related to biotechnological applications as well as health related such as the development of new drugs.

Since many years, in the context of WeNMR (www.wenmr.eu) now operating as a Thematic service provider in the European Open Science Cloud (EOSC), we have provided the community with relevant and specialized software that make up a valuable toolkit for structural biology researchers. One critical aspect aside from providing the community with valuable tools is to also provide the means of executing them in a user-friendly, efficient and cost-effective distributed manner, as done *via* the DIRAC Workload Manager (Tsaregorodtsev and Project, 2014).

Here we present an overview of the structural biology services provided *via* the WeNMR-EOSC ecosystem, what are the technological challenges and solutions that need to be implemented for an efficient use of distributed computing resources and how those are used by an active community worldwide.

	AMPS-NMR	DISVIS	FANTEN	HADDOCK	MetalPDB	PDB-tools	Powerfit	proABC-2	Prodigy	SpotON
Input data										
Single structures (PDB)	■		■	■	■	■	■			
Complex (PDB)	■	■	■	■	■	■			■	■
Sequence					■			■		
Restraints										
Distances[1]	■	■		■						
Interface residues[2]				■						
NMR pseudo contact shifts	■		■	■						
NMR residual dipolar couplings	■		■	■						
NMR relaxation anisotropy				■						
Dihedral angles	■			■						
Radius of gyration				■						
Symmetry				■						
Non-crystallographic symmetry				■						
Cryo-EM density				■			■			
Shapes				■						
Execution[3]										
Local	■	■	■	■	■	■	■	■	■	■
HTC	■	■	■	■			■			
GPU	■	■					■			
Approximate execution time	minutes to hours	minutes to hours	minutes	tens of minutes to days[5]	n.a.	seconds	minutes to hours	seconds	seconds	minutes to hours
Registration Required	■	■		■			■			■
Robot Certificate handling[4]	■	■		■			■			

[1] These distances can be obtained from NMR, MS cross-linking, FRET, EPR, etc or from bioinformatic predictions (e.g. co-evolution restraints). They can be unambiguous (defined between specific pairs of atoms) or ambiguous (defined between multiple atoms)

[2] Information about residue involved in the binding such as obtained for example from mutagenesis, NMR titration, H/D exchange experiments or bioinformatic predictions.

[3] Local: Runs on local resources; HTC: Makes use of the EOSC/EGI distributed high-throughput computing resources; GPU: Can make use of local or EOSC/EGI distributed GPU resources

[4] Portals accessing EOSC/EGI HTC resources using X509 robot certificates. No certificate required from users

[5] The execution time depends on many variables such as number of size of the molecules to be docked, type and amount of restraints, docking settings (number of models, number of refinement steps, explicit solvent refinement) and the load on the server

FIGURE 1 | Schematic panel displaying which type of input data and restraint are supported by each WeNMR-EOSC service as well as execution infrastructure used, registration policy and certificate handling.

SERVICES

The study of macromolecular structures is a multifaceted challenge in which researchers must apply different approaches and tools to gain a better understanding of their function, behavior and dynamics. To this end, within the WeNMR-EOSC project, we have been developing different tools, making them freely available to the community as webservices. The following paragraphs provide a short overview of currently active services under the EOSC portal.

AMBER-Based Portal Server for-NMR

The AMPS-NMR (AMBER-based Portal Server for NMR structures) (Bertini et al., 2011) provides a user-friendly interface to run restrained molecular dynamics simulations (rMD) to refine experimental NMR structures, and to perform molecular dynamics simulations of biomacromolecular systems in general. Calculations can be run on CPUs or GPGPUs. The latter provide significantly faster performance (Andreetto et al., 2017) but are limited in availability. For NMR structures, the use of the AMPS-NMR portal results in a consistent improvement of features such as rotamer distributions, backbone normality and occurrence of steric clashes, without affecting agreement with the experimental data. The effect of rMD refinement is especially relevant for protein regions where the amount of experimental information is scarce. The rMD protocol has been implemented as a predefined multi-step procedure, so that the user does not need to know in detail how to tune the many parameters involved in a relatively complex MD calculation; the portal also handles automatically the most commonly used formats for experimental data. The AMPS-NMR webserver is available at http://py-enmr. cerm.unifi.it/access/index.

TABLE 1 | Usage statistics and satisfactions of the WeNMR services as of July 2021.

Service	Satisfaction rating [scale 1–5] (#of respondents)	Total number of users to date	Total number of served requests to date	Web portal Release year
AMPS-NMR	-	657	27144	2011
DisVis	4.76 (76)	3545	3947	2016
HADDOCK	4.9 (5448)	23254	354910	2008
MetalPDB	-	158930*	225425**	2013
PDB-tools web	4.91 (74)	-	5012	2020
Powerfit	4.78 (18)	2834	747	2016
proABC-2	4.84 (73)	-	8017	2019
Prodigy	4.79 (2191)	-	158981	2016
SpotON	4.7 (82)	3253	6829	2017

*-Unique Ips.
**-Pages served.

TABLE 2 | Tutorials and user support material for the WeNMR-EOSC services.

Service	Tutorials/Manuals	Support
AMPS-NMR	https://www.wenmr.eu/tutorials/#rmd-amber	*Via* e-mail
DISVIS	https://www.bonvinlab.org/education/Others/disvis-webserver/	https://ask.bioexcel.eu
FANTEN	http://abs.cerm.unifi.it:8080/manual/FANTEN_manual.pdf	*Via* e-mail
HADDOCK	https://www.bonvinlab.org/education/HADDOCK24/	https://ask.bioexcel.eu
MetalPDB	https://metalpdb.cerm.unifi.it/glossary	*Via* e-mail
PDB-tools web	https://wenmr.science.uu.nl/pdbtools/manual	https://ask.bioexcel.eu
Powerfit	https://www.bonvinlab.org/education/Others/powerfit-webserver	https://ask.bioexcel.eu
proABC-2	https://wenmr.science.uu.nl/proabc2/manual	*Via* e-mail
Prodigy	https://wenmr.science.uu.nl/prodigy/manual	https://ask.bioexcel.eu
SpotON	https://alcazar.science.uu.nl/cgi/services/SPOTON/spoton/help	https://ask.bioexcel.eu

DISVIS

The interaction between molecules can be experimentally characterized with methods such as cross-linking mass spectrometry (XL-MS) and FRET (Okamoto and Sako, 2017; Yu and Huang, 2017) which can provide information about which residues are participating in the interaction as well as the distance (or an upper limit to it) between the reacting groups of the residues. DISVIS is a tool that allows researchers to visualize the three-dimensional distance-restrained possible interaction space between the interacting proteins partners. It can be used to investigate if the restraints are mutually exclusive and identify possible false positives in the input data. DISVIS is freely available at https://wenmr.science.uu.nl/disvis.

FANTEN

FANTEN (Rinaldelli et al., 2015) aims at the determination of the anisotropy tensors associated with NMR paramagnetic pseudocontact shifts (pcs) and/or residual dipolar couplings (rdc's). FANTEN also permits the pcs-driven determination of the 3D structure of protein-protein adducts using a rigid body approach; the user can then submit these models directly to HADDOCK for flexible refinement. The FANTEN web server is available at http://abs.cerm.unifi.it:8080/.

HADDOCK

The pioneer of integrative docking, HADDOCK (*High Ambiguity Driven biomolecular DOCKing*) (Dominguez et al., 2003) allows users to make use of a variety of experimental and theoretical information to drive the docking of up to 20 macromolecules. The latest version (v2.4) supports the docking of proteins, small-molecules, nucleic acids, peptides, cyclic peptides, glycans, and glycosylated proteins. To accommodate the complexity that naturally arises from >18 years of development of new features and docking modes without compromising usability and accessibility, HADDOCK has been made freely available (with registration) to the community since 2008 *via* an intuitive webserver (Vries et al., 2010; Zundert et al., 2016). Since exposing 400+ parameters could hinder user experience (and usability), users are separated in access tiers: Users belonging to the "Easy" tier can access and modify a limited number of parameters whereas users belonging to the "Guru" tier (granted upon specific request from users in their registration page) have full access and control of the docking protocol. The HADDOCK webserver is available at https://wenmr.science.uu.nl/haddock2.4.

MetalPDB

MetalPDB (Andreini et al., 2013a; Putignano et al., 2017) is a database collecting information on the metal-binding sites present in the 3D structures of biological macromolecules (Putignano et al., 2017). In particular, metal sites are represented in MetalPDB, as 3D templates describing the local environment around the metal ion(s) independently of the overall macromolecular fold. Among its features, MetalPDB provides detailed statistical analyses regarding metal usage in proteins of

known 3D structure, encompassing protein families, and in catalysis; it also includes tools for the structural comparison of metal sites (Andreini et al., 2013b). MetalPDB is available at https://metalpdb.cerm.unifi.it/

Protein Data Bank-Tools Web

The classic Protein Data Bank (PDB) file format is a flat text file that is still used by many structural biology software to represent the spatial coordinates of macromolecular structures, even though the official format of the worldwide PDB (wwPDB) (Berman et al., 2003) has now shifted to the Crystallographic Information Framework (mmCIF) dictionary (Bourne et al., 1997). The simple flat text PDB format, despite its apparent simplicity, follows strict formatting rules which can be easily overlooked during editing resulting in various errors. Such errors are usually of syntactic nature (e.g., shifting columns or missing fields) due to typos introduced by users after manual-editing PDB files, although semantic errors are also very common (e.g., incorrect atom naming, overlapping numbering, duplications, etc.). PDB-tools web (Jiménez-García et al., 2021) is a fully configurable, user-friendly web interface for the command-line application *pdb-tools* (Rodrigues et al., 2018). Using the portal users can, in a few clicks, build a complex pipeline which can then be saved (and uploaded) for future use and reproducibility. This pipeline is composed by different processing blocks with atomic tasks on the PDB provided data, to name a few, removing certain atom types, renaming chains, renumbering residues, removing multiple occupancies or detecting gaps among many others. The webserver is available at https://wenmr.science.uu.nl/pdbtools/ and the list of all available processing blocks is described on the online manual at https://wenmr.science.uu.nl/pdbtools/manual.

Powerfit

Cryo-electron microscopy (cryo-EM) is an experimental technique that can be used to obtain direct images of large macromolecular complexes (Fernandez-Leiro and Scheres, 2016). When the resolution of the resulting EM data is not sufficient to build models directly in the maps, it is common to fit existing 3D structures into those maps. To this end, Powerfit has been developed as a Python package for fast and sensitive rigid body fitting of molecular structures into EM density maps with a core-weighted local cross correlation scoring function (Zundert and Bonvin, 2015b). It is freely available as a webserver (Zundert et al., 2017) running on both local and GPU-accelerated distributed infrastructures at https://alcazar.science.uu.nl/services/POWERFIT.

proABC-2

Monoclonal antibodies have been consolidated as therapeutic tools due to their high affinity and specificity towards their target. proABC-2 (Ambrosetti et al., 2020), the updated version of proABC (Olimpieri et al., 2013), is a valuable tool that predicts which antibody residues can form intermolecular contacts with its antigen and further give insight into the chemical components of such interactions, differentiating between hydrophobic and hydrophilic interactions. The latest version replaced the previously used random forest paratope predictor by a convolutional neural network algorithm. Predictions obtained by proABC-2 can be used for example, to guide molecular docking, leading to improvements of success rate

and quality of the predicted models. proABC-2 is available at https://wenmr.science.uu.nl/proabc2/.

Prodigy

In addition to elucidating the structural aspects of a biomolecular complex and its interacting components, it is of crucial importance to be able to estimate the strength of the interaction, or, in other words, to be able to predict the binding affinity of the complex. PRODIGY, the *PROtein binDIng enerGY Prediction tool* (Vangone and Bonvin, 2015) allows to predict the binding affinity of protein-protein complexes from their three-dimensional structures. It does so by calculating features derived both from interfacial residue contacts as well as those from the non-interface residues (Kastritis et al., 2014). The same rationale has been expanded to protein-ligand complexes (PRODIGY-LIGAND) (Kurkcuoglu et al., 2018; Vangone et al., 2018) and to the classification of interfaces in crystal structures as biological or crystallographic (PRODIGY-CRYSTAL) (Elez et al., 2018; Jiménez-García et al., 2019). PRODIGY services are available at https://wenmr.science.uu.nl/prodigy/.
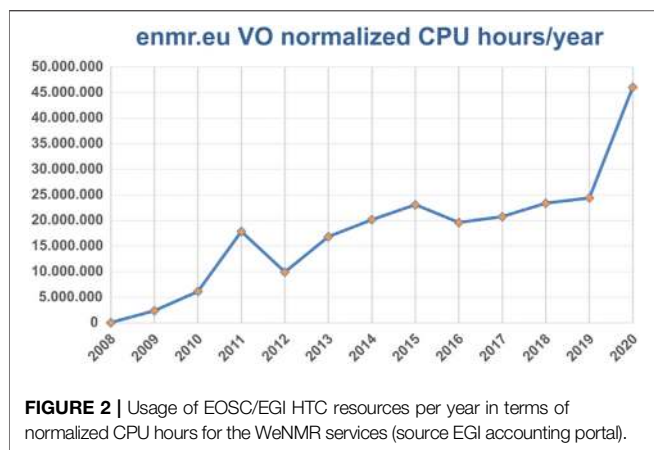
SpotON

Once an interaction has been characterized, it is valuable to know which parts of the interface contribute more to the binding, namely what are interaction "hot spots"—residues that upon mutation to alanine confer a difference in the binding free energy of the complex greater than 2.0 kcal/mol. The experimental identification of these regions is typically low throughput, laborious and expensive since it involves multiple rounds of point mutations and binding energy evaluations. SpotON was developed a computational approach to identify and classify interface residues as either hot spots, with positive contribution to the interface or null spots, with little to no significant contribution to the interface (Melo et al., 2016; Moreira et al., 2017). Its predictor combines different machine-learning algorithms, and it is available as a free and easy-to-use webserver https://alcazar.science.uu.nl/services/SPOTON.

INFRASTRUCTURE

The WeNMR portals use either local or distributed computing resources, while harvesting GPU resources in some cases to speed up the computations. Some portals, especially the ones making use of the EOSC-EGI high-throughput computational resources (see below) do require registration for use (**Figure 1**).

European Open Science Cloud-EGI

The EGI Federation is an international e-Infrastructure providing advanced computing and data analytics for research and innovation. In the last decade it evolved from the high energy physics compute grid (WLCG) towards a multi-disciplinary, multi-technology infrastructure federating hundreds of resource centers worldwide and delivering large-scale data analysis capabilities (>1 millions of CPU-cores and >900 PB of disk and tape storage) to more than 70,000 researchers covering many scientific disciplines. During 2020 it delivered 13 billions of CPU hours of High Throughput Computing (HTC) and 18 millions of CPU hours of cloud computing (**Figure 2**).

FIGURE 2 | Usage of EOSC/EGI HTC resources per year in terms of normalized CPU hours for the WeNMR services (source EGI accounting portal).
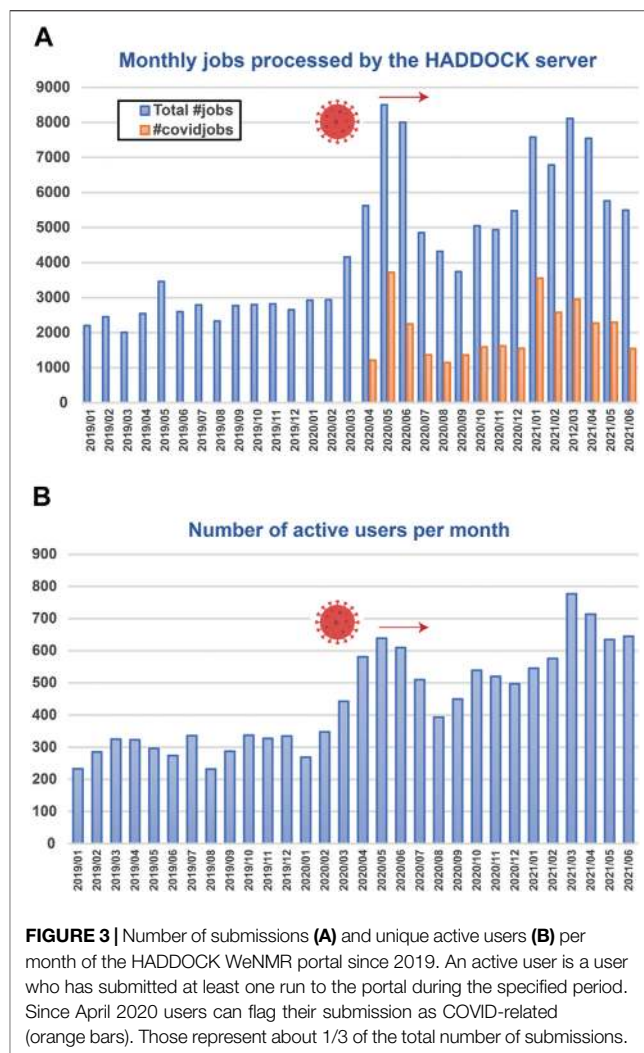
EGI was one of the main actors and the coordinator of EOSC-Hub, a three-year European project that ended in March 2021, aimed at starting the design and implementation of the European Open Science Cloud. EGI operates several core elements of EOSC and contributes to EOSC service portfolio offering services like e.g., HTC, Cloud Compute, Workload Manager, and Check-in which are relevant for WeNMR. Moreover, with the support of the recently started EGI-ACE project, under which WeNMR services are further operated, EGI will deliver the EOSC Compute Platform and will contribute to the EOSC Data Commons through a federation of cloud compute and storage facilities, Platform as a Service (PaaS) services and data spaces with analytics tools and federated access services.

The collaboration between EGI and WeNMR dates back to their early days, when in 2011 WeNMR became the first Virtual Research Community officially recognized by EGI. A formal Service Level Agreement (SLA) was established in 2016 and has since been periodically updated and recently extended until June 2023. The SLA between EGI and WeNMR is a document summarizing ten Operational Level Agreements (OLAs) between EGI and ten resource centers committed to providing HTC, cloud compute and storage capacity with availability and reliability above well-defined targets constantly monitored by the EGI operation team. These ten resource centers are located in Czech Republic, Italy, Portugal, Spain, Taiwan, Netherlands, and Ukraine. They ensure the availability of 6050 HEPSPEC′06 normalized CPU-years per year of HTC computing time, 540 virtual CPU-cores of cloud compute capacity, and 59 TB of online storage capacity. Furthermore, a grid site hosted at the *Consorzio Interuniversitario Risonanze Magnetiche di Metalloproteine Paramagnetiche* (CIRMMP) provides several GPGPU cards that greatly enhanced the performance of the DISVIS, POWERFIT, and AMPS-NMR applications. Next to the SLA sites, additional resources are also available on an opportunistic use basis. During 2020 the HTC capacity has further increased by adding some Open Science Grid resources hosted in the US and a few WLCG sites hosted in France, Germany and Spain that decided to support WeNMR COVID-19 related jobs.

Single Sign-On

The WeNMR community has since the beginning recognized the importance of having a Single Sign-On (SSO) mechanism to access their web portals with unique login credentials. The first

FIGURE 3 | Number of submissions **(A)** and unique active users **(B)** per month of the HADDOCK WeNMR portal since 2019. An active user is a user who has submitted at least one run to the portal during the specified period. Since April 2020 users can flag their submission as COVID-related (orange bars). Those represent about 1/3 of the total number of submissions.

SSO was implemented in 2013 under the WeNMR European project, allowing users to subscribe to portals, validate personal grid certificates and manage job submissions all from one web page, collecting at the same time accounting records and user's geographic information. A few years later, with the support of the West-Life H2020 European project (2015–2018), the WeNMR SSO was gradually replaced with the West-Life SSO. In recent years the SSO mechanisms typically have relied on an Authentication and Authorization Infrastructure (AAI) based on services like the IdP-SP-Proxy and standard protocols like SAML, OpenID Connect and OAuth, and have been designed following the guidelines of the *Authentication and Authorization for Research and Collaboration* (AARC) initiative launched in May 2015. With the support of the EOSC-Hub project (2018–2021) the WeNMR services transitioned to the current AAI solution based on the EGI Check-in service. Developed from the same AARC Blueprint Architecture[1], EGI Check-in solution

---

[1]https://aarc-project.eu/wp-content/uploads/2019/05/AARC2-DJRA1.4_v2-FINAL.pdf.

is part of EOSC service catalogue, and has also been adopted by the Instruct-ERIC infrastructure. Nowadays all the WeNMR portals integrate EGI Check-in as SSO mechanism. Users can register to the EGI Check-in at https://aai.egi.eu and use those credentials to register to WeNMR services. All portals are compliant with the European General Data Protection Regulation (GDPR) with clear condition of use (https://wenmr.science.uu.nl/conditions) and privacy measures (https://wenmr.science.uu.nl/privacy).

Dirac Workload Manager

The EGI Workload Manager is a service enabling users to access distributed computing resources of various types through a single user-friendly interface. It is one of the services offered by the EGI e-infrastructure project and it is also available *via* the EOSC Compute Platform. The service is built upon the software from the DIRAC Interware project (http://diracgrid.org) (Tsaregorodtsev and Project, 2014). The project is providing a development framework and a set of ready-to-use components to build distributed computing systems of arbitrary complexity. It provides a complete solution for communities needing access to computing and storage resources distributed geographically, integrated in different grid and cloud infrastructures or standalone computing clusters and supercomputers. The DIRAC Workload Manager serves multiple scientific communities–partners of the EGI e-infrastructure and the WeNMR Collaboration as one of its most active users. User tasks prepared by the WeNMR application portals are submitted to the Workload Manager service. The service performs reservation of computing resources by means of so-called pilot jobs which are submitted to various computing centers with appropriate access protocols. Once deployed on worker nodes, pilot jobs verify the execution environment and then request user payloads from the central DIRAC Task Queue. Altogether, the pilot jobs and the central Task Queue form a dynamic virtual batch system that overcomes the heterogeneity of the underlying computing infrastructures. This workload scheduling architecture allows to easily add new resources transparently for the users. It also makes it easy to apply usage policies by defining fine grained priorities to certain activities. For example, during the COVID pandemic it allowed to quickly make available resources of the sites willing to contribute to the COVID-related studies and ensure high priority of these jobs compared to other regular WeNMR activities.

COMMUNITY AND USAGE

The WeNMR services combined have had over 150.000+ users (**Table 1**). User satisfaction is monitored and the services have various support mechanisms summarized at https://www.wenmr.eu/support allowing users to fill in requests and ask specific questions. This large userbase is provided with tutorials to ensure the tools are being used to their best (**Table 2**). The user's feedback and issues are one of the driving forces towards the improvement of the services and implementation of new features.

The services discussed here have been making use of EGI HTC resources since 2009. **Figure 1**, plots the normalized CPU usage per month, clearly showing the increasing use of the EGI HTC resources over time, culminating in 2020 to the equivalent of ~5300 CPU years' worth of computing.

During the COVID pandemic, one of the most used WeNMR service, HADDOCK, has seen a significant increase of both the number of submissions and single users, with about 1/3 of all submissions since April 2021 being COVID-related (**Figure 3**).

CONCLUSION

Over more than a decade the WeNMR project has facilitated the use of advanced computational tools. Taking advantage of the EGI high throughput compute infrastructure, it has developed as a Thematic Services in the European Open Science Cloud. Its large worldwide user community of over 23,000 users distributed over 125 countries, submitted in 2020 >12 million jobs that accounted for ~4,000 CPU-years. In tight collaboration with EGI, WeNMR has constantly improved its services, for example, piloting the use of distributed GPU resources and facilitating user access through the implementation of a single sign one mechanism. Its valuable services are offered *via* user-friendly web-based solutions that allows researchers to execute complex workflows at the click of their mouse without having to deal with the complexity of managing and distributing computations. The impact is evident from both usage of the services in research and education and from the number of citations of the various tools (>800 since 2020). The active WeNMR user community, in turn, acts a catalyst for further development and improvement of the services.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

RH and AB conceived the structure of the manuscript and all authors contributed to the writing.

REFERENCES

Ambrosetti, F., Olsen, T. H., Olimpieri, P. P., Jiménez-García, B., Milanetti, E., Marcatilli, P., et al. (2020). proABC-2: PRediction of AntiBody Contacts V2 and its Application to Information-Driven Docking. *Bioinformatics* 36, 5107–5108. doi:10.1093/bioinformatics/btaa644

Andreetto, P., Astalos, J., Dobrucky, M., Giachetti, A., Rebatto, D., Rosato, A., et al. (2017). EGI Federated Platforms Supporting Accelerated Computing. *Proc. Int. Symp. Grids Clouds Isgc* 2017, Pos Isgc2017. doi:10.22323/1.293.0020

Andreini, C., Cavallaro, G., Lorenzini, S., and Rosato, A. (2013a). MetalPDB: a Database of Metal Sites in Biological Macromolecular Structures. *Nucleic Acids Res.* 41, D312–D319. doi:10.1093/nar/gks1063

Andreini, C., Cavallaro, G., Rosato, A., and Valasatava, Y. (2013b). MetalS2: A Tool for the Structural Alignment of Minimal Functional Sites in Metal-Binding Proteins and Nucleic Acids. *J. Chem. Inf. Model.* 53, 3064–3075. doi:10.1021/ci400459w

Berman, H., Henrick, K., and Nakamura, H. (2003). Announcing the Worldwide Protein Data Bank. *Nat. Struct. Mol. Biol.* 10, 980. doi:10.1038/nsb1203-980

Bertini, I., Case, D. A., Ferella, L., Giachetti, A., and Rosato, A. (2011). A Grid-Enabled Web portal for NMR Structure Refinement with AMBER. *Bioinformatics* 27, 2384–2390. doi:10.1093/bioinformatics/btr415

Bourne, P. E., Berman, H. M., McMahon, B., Watenpaugh, K. D., Westbrook, J. D., and Fitzgerald, P. M. D. (1997). [30] Macromolecular Crystallographic Information File. *Methods Enzymol.* 277, 571–590. doi:10.1016/s0076-6879(97)77032-0

Dominguez, C., Boelens, R., and Bonvin, A. M. J. J. (2003). HADDOCK: A Protein–Protein Docking Approach Based on Biochemical or Biophysical Information. *J. Am. Chem. Soc.* 125, 1731–1737. doi:10.1021/ja026939x

Elez, K., Bonvin, A. M. J. J., and Vangone, A. (2018). Distinguishing Crystallographic from Biological Interfaces in Protein Complexes: Role of Intermolecular Contacts and Energetics for Classification. *Bmc Bioinformatics* 19, 438. doi:10.1186/s12859-018-2414-9

Fernandez-Leiro, R., and Scheres, S. H. W. (2016). Unravelling Biological Macromolecules with Cryo-Electron Microscopy. *Nature* 537, 339–346. doi:10.1038/nature19948

Jiménez-García, B., Teixeira, J. M. C., Trellet, M., Rodrigues, J. P. G. L. M., and Bonvin, A. M. J. J. (2021). PDB-tools Web: A User-friendly Interface for the Manipulation of PDB Files. *Proteins Struct. Funct. Bioinform* 89, 330–335. doi:10.1002/prot.26018

Jiménez-García, B., Elez, K., Koukos, P. I., Bonvin, A. M., and Vangone, A. (2019). PRODIGY-crystal: a Web-Tool for Classification of Biological Interfaces in Protein Complexes. *Bioinformatics* 35, 4821–4823. doi:10.1093/bioinformatics/btz437

Kastritis, P. L., Rodrigues, J. P. G. L. M., Folkers, G. E., Boelens, R., and Bonvin, A. M. J. J. (2014). Proteins Feel More Than They See: Fine-Tuning of Binding Affinity by Properties of the Non-interacting Surface. *J. Mol. Biol.* 426, 2632–2652. doi:10.1016/j.jmb.2014.04.017

Kurkcuoglu, Z., Koukos, P. I., Citro, N., Trellet, M. E., Rodrigues, J. P. G. L. M., Moreira, I. S., et al. (2018). Performance of HADDOCK and a Simple Contact-Based Protein-Ligand Binding Affinity Predictor in the D3R Grand Challenge 2. *J. Comput. Aided Mol. Des.* 32, 175–185. doi:10.1007/s10822-017-0049-y

Melo, R., Fieldhouse, R., Melo, A., Correia, J., Cordeiro, M., Gümüş, Z., et al. (2016). A Machine Learning Approach for Hot-Spot Detection at Protein-Protein Interfaces. *Ijms* 17, 1215. doi:10.3390/ijms17081215

Moreira, I. S., Koukos, P. I., Melo, R., Almeida, J. G., Preto, A. J., Schaarschmidt, J., et al. (2017). SpotOn: High Accuracy Identification of Protein-Protein Interface Hot-Spots. *Sci. Rep.* 7, 8007. doi:10.1038/s41598-017-08321-2

Okamoto, K., and Sako, Y. (2017). Recent Advances in FRET for the Study of Protein Interactions and Dynamics. *Curr. Opin. Struct. Biol.* 46, 16–23. doi:10.1016/j.sbi.2017.03.010

Olimpieri, P. P., Chailyan, A., Tramontano, A., and Marcatili, P. (2013). Prediction of Site-specific Interactions in Antibody-Antigen Complexes: the proABC Method and Server. *Bioinformatics* 29, 2285–2291. doi:10.1093/bioinformatics/btt369

Putignano, V., Rosato, A., Banci, L., and Andreini, C. (2017). MetalPDB in 2018: a Database of Metal Sites in Biological Macromolecular Structures. *Nucleic Acids Res.* 46, D459–D464. doi:10.1093/nar/gkx989

Rinaldelli, M., Carlon, A., Ravera, E., Parigi, G., and Luchinat, C. (2015). FANTEN: a New Web-Based Interface for the Analysis of Magnetic Anisotropy-Induced NMR Data. *J. Biomol. Nmr* 61, 21–34. doi:10.1007/s10858-014-9877-4

Rodrigues, J. P. G. L. M., Teixeira, J. M. C., Trellet, M., and Bonvin, A. M. J. J. (2018). Pdb-Tools: a Swiss Army Knife for Molecular Structures. *F1000Res* 7, 1961. doi:10.12688/f1000research.17456.1

Tsaregorodtsev, A., and Project, the. D. (2014). DIRAC Distributed Computing Services. *J. Phys. Conf. Ser.* 513, 032096. doi:10.1088/1742-6596/513/3/032096

van Zundert, G. C., Trellet, M., Schaarschmidt, J., Kurkcuoglu, Z., David, M., Verlato, M., et al. (2017). The DisVis and PowerFit Web Servers: Explorative and Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* 429, 399–407. doi:10.1016/j.jmb.2016.11.032

van Zundert, G. C. P., Rodrigues, J. P. G. L. M., Trellet, M., Schmitz, C., Kastritis, P. L., Karaca, E., et al. (2016). The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* 428, 720–725. doi:10.1016/j.jmb.2015.09.014

Vangone, A., and Bonvin, A. M. (2015). Contacts-based Prediction of Binding Affinity in Protein-Protein Complexes. *Elife* 4, e07454. doi:10.7554/elife.07454

Vangone, A., Schaarschmidt, J., Koukos, P., Geng, C., Citro, N., Trellet, M. E., et al. (2018). Large-scale Prediction of Binding Affinity in Protein-Small Ligand Complexes: the PRODIGY-LIG Web Server. *Bioinformatics* 35, 1585–1587. doi:10.1093/bioinformatics/bty816

Vries, S. J. de., Dijk, M. van., and Bonvin, A. M. J. J. (2010). The HADDOCK Web Server for Data-Driven Biomolecular Docking. *Nat. Protoc.* 5, 883–897. doi:10.1038/nprot.2010.32

Xue, L. C., Rodrigues, J. P., Kastritis, P. L., Bonvin, A. M., and Vangone, A. (2016). PRODIGY: a Web Server for Predicting the Binding Affinity of Protein-Protein Complexes. *Bioinformatics* 32, btw514–3678. doi:10.1093/bioinformatics/btw514

Yu, C., and Huang, L. (2017). Cross-Linking Mass Spectrometry: An Emerging Technology for Interactomics and Structural Biology. *Anal. Chem.* 90, 144–165. doi:10.1021/acs.analchem.7b04431

Zundert, G. C. P. van., and Bonvin, A. M. J. J. (2015a). DisVis: Quantifying and Visualizing Accessible Interaction Space of Distance-Restrained Biomolecular Complexes. *Bioinformatics* 31, 3222–3224. doi:10.1093/bioinformatics/btv333

Zundert, G. C. P. van., and Bonvin, A. M. J. J. (2015b). Fast and Sensitive Rigid-Body Fitting into Cryo-EM Density Maps with PowerFit. *Aims Biophys.* 2, 73–87. doi:10.3934/biophy.2015.2.73