

J Grid Computing (2012) 10:743–767
DOI 10.1007/s10723-012-9246-z

WeNMR: Structural Biology on the Grid

Tsjerk A. Wassenaar · Marc van Dijk · Nuno Loureiro-Ferreira · Gijs van der Schot · Sjoerd J. de Vries · Christophe Schmitz · Johan van der Zwan · Rolf Boelens · Andrea Giachetti · Lucio Ferella · Antonio Rosato · Ivano Bertini · Torsten Herrmann · Hendrik R. A. Jonker · Anurag Bagaria · Victor Jaravine · Peter Güntert · Harald Schwalbe · Wim F. Vranken · Jurgen F. Doreleijers · Gert Vriend · Geerten W. Vuister · Daniel Franke · Alexey Kikhney · Dmitri I. Svergun · Rasmus H. Fogh · John Ionides · Ernest D. Laue · Chris Spronk · Simonas Jurkša · Marco Verlato · Simone Badoer · Stefano Dal Pra · Mirco Mazzucato · Eric Frizziero · Alexandre M. J. J. Bonvin

Received: 12 December 2011 / Accepted: 18 September 2012 / Published online: 30 November 2012
© The Author(s) 2012. This article is published with open access at Springerlink.com

Abstract The WeNMR (<http://www.wenmr.eu>) project is a European Union funded international effort to streamline and automate analy-

sis of Nuclear Magnetic Resonance (NMR) and Small Angle X-Ray scattering (SAXS) imaging data for atomic and near-atomic resolution

T. A. Wassenaar · M. van Dijk · N. Loureiro-Ferreira · G. van der Schot · S. J. de Vries · C. Schmitz · J. van der Zwan · R. Boelens · A. M. J. J. Bonvin (✉)
Bijvoet Center for Biomolecular Research, Faculty of Science, Utrecht University, Padualaan 8, 3584 CH, Utrecht, The Netherlands
e-mail: a.m.j.j.bonvin@uu.nl

A. Giachetti · L. Ferella · A. Rosato · I. Bertini
Magnetic Resonance Center, University of Florence, 50019 Sesto Fiorentino, Italy

T. Herrmann
Centre de RMN à très Hauts Champs, Institut des Sciences Analytiques, Université de Lyon, UMR-5280 CNRS, ENS Lyon, UCB Lyon 1, 5 rue de la Doua, 69100 Villeurbanne, France

H. R. A. Jonker · H. Schwalbe
Institute of Organic Chemistry and Chemical Biology and Biomolecular Magnetic Resonance Center, Goethe University Frankfurt, 60438 Frankfurt am Main, Germany

A. Bagaria · V. Jaravine · P. Güntert
Institute of Biophysical Chemistry and Biomolecular Magnetic Resonance Center, Goethe University Frankfurt, 60438 Frankfurt am Main, Germany

W. F. Vranken
European Bioinformatics Institute, Hinxton, Cambridge, CB10 1SD, UK

J. F. Doreleijers
Protein Biophysics/IMM, Radboud University Nijmegen, Geert Grooteplein 26-28, Nijmegen, The Netherlands

G. Vriend
CMBI, Radboud University Nijmegen Medical Centre, Geert Grooteplein 26-28, Nijmegen, The Netherlands

G. W. Vuister
Department of Biochemistry, School of Biological Sciences, Henry Wellcome Building, University of Leicester, Lancaster Road, Leicester LE1 9HN, UK

D. Franke · A. Kikhney · D. I. Svergun
European Molecular Biology Laboratory, Hamburg Outstation, Notkestrasse 85, D22603 Hamburg, Germany

R. H. Fogh · J. Ionides · E. D. Laue
Department of Biochemistry, University of Cambridge, 80 Tennis Court Road, CB2 1GA, Cambridge, UK

C. Spronk · S. Jurkša
UAB “Spronk NMR Consultancy”, Palangos gatvė 4, 01402, Vilnius, Lithuania

M. Verlato · S. Badoer · S. Dal Pra · M. Mazzucato · E. Frizziero
Istituto Nazionale di Fisica Nucleare, Sez. di Padova, 35131 Padova, Italy

molecular structures. Conventional calculation of structure requires the use of various software packages, considerable user expertise and ample computational resources. To facilitate the use of NMR spectroscopy and SAXS in life sciences the WeNMR consortium has established standard computational workflows and services through easy-to-use web interfaces, while still retaining sufficient flexibility to handle more specific requests. Thus far, a number of programs often used in structural biology have been made available through application portals. The implementation of these services, in particular the distribution of calculations to a Grid computing infrastructure, involves a novel mechanism for submission and handling of jobs that is independent of the type of job being run. With over 450 registered users (September 2012), WeNMR is currently the largest Virtual Organization (VO) in life sciences. With its large and worldwide user community, WeNMR has become the first Virtual Research Community officially recognized by the European Grid Infrastructure (EGI).

Keywords Web portals · Nuclear magnetic resonance · Small angle x-ray scattering · Structural biology · Proteins · Virtual research community

Present Address:

T. A. Wassenaar
Biocomputing Group, Department of Biological Sciences, University of Calgary, 2500 University Drive NW, AB T2N 1N4 Calgary, Canada

Present Address:

N. Loureiro-Ferreira
European Grid Infrastructure (EGI),
140 Science Park, 1098 XG Amsterdam,
The Netherlands

Present Address:

W. F. Vranken
Department of Structural Biology, VIB, and
Structural Biology Brussels, Vrije Universiteit Brussel,
Pleinlaan 2, 1050 Brussels, Belgium

Present Address:

S. Dal Pra
Istituto Nazionale di Fisica Nucleare, CNAF,
40127 Bologna, Italy

1 Introduction

1.1 NMR Spectroscopy

NMR Spectroscopy is one of two main techniques that allow determining three dimensional (3D) structures of biomacromolecules, such as proteins, RNA, DNA, and their complexes, at atomic resolution. Knowledge of their 3D structures is vital for understanding functions and mechanisms of action of macromolecules, and for elucidating and predicting the effect of mutations. 3D structures are also important as guides for the design of new experimental studies and as starting points for rational drug design. An advantage of NMR over X-ray crystallography is that it also allows investigation of time-dependent chemical and conformational phenomena, including reaction and folding kinetics and intramolecular dynamics. For these reasons, NMR plays an important role within the life sciences.

The principles underlying NMR are modulation of the natural magnetic moment of atomic nuclei, and measurements of how the system relaxes back to the initial state [1, 2]. The signal thus obtained is a fading wave consisting of many individual frequency contributions: the Free Induction Decay, FID. Typically, up to 27000 different frequencies can be resolved at the highest magnetic fields that are nowadays available. To investigate the frequency contributions and their decays, such measurements have to be repeated many times, due to the low signal-to-noise ratio. To obtain structural information from NMR data, many more, but also more complex measurements have to be run, yielding substantial amounts of data that need processing.

Processing data from NMR to obtain a 3D structure typically involves the following steps, summarized graphically in Fig. 1. First the raw data have to be processed, more specifically Fourier-transformed, to obtain spectra revealing the different frequency contributions and their relations. These frequencies are the resonances of the atoms measured, but to infer structural information from them, these resonances subsequently have to be assigned to individual contributors (atoms/residues). If the assignment is sufficiently complete, structural restraints can be determined

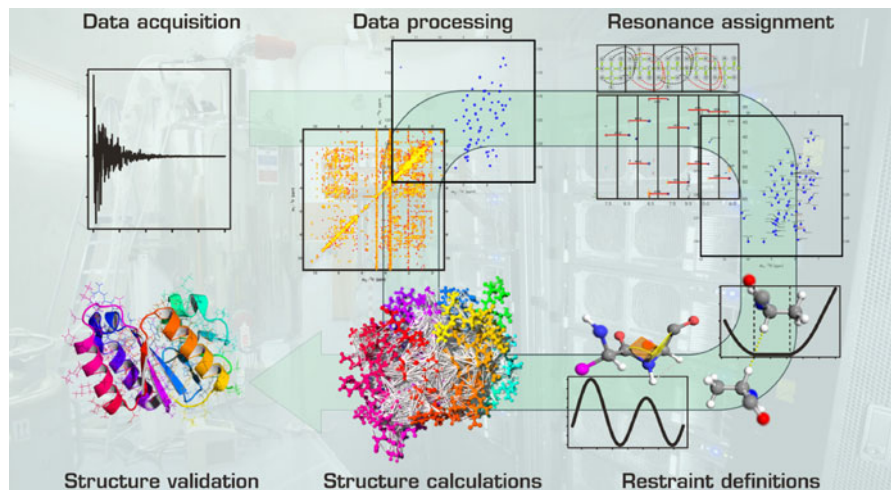


Fig. 1 NMR data processing from signal to 3D structure. After acquisition of the primary NMR data, these are Fourier transformed to obtain spectra in which the individual frequency contributions or resonances of spin systems, and their relations, are revealed. The resonances subsequently have to be assigned to individual atoms. If sufficient resonances have been assigned, restraints can be inferred from the data, pertaining to distances between

atoms, dihedral angles, domain orientations, etc. When an adequate number of restraints is available, these can be used to calculate a set of three-dimensional structures optimally satisfying these restraints. The resulting structures represent the structure of the protein in solution, which is validated against the available experimental data. Although the process is here depicted linearly, intermediate stages may involve iterative cycles of refinement

from the spectra, including inter-atomic distance restraints, dihedral angle restraints, and orientation restraints. These structural restraints are then used to calculate a number of structures using a variety of molecular modeling approaches, after which structure validation checks are performed to assert the quality of the results.

For each of the steps involved, specialized computer programs are available, each with its own characteristics and often with its own data format. Processing of NMR data has thus become a task for specialists, who can understand the data and their formats, as well as the programs, with installation requirements and usage details. Furthermore, NMR data processing requires considerable data storage and computational resources. These factors together currently represent a barrier for groups in life sciences to employ the full power of NMR. Against this background, the eNMR project was run as a European initiative funded under the Framework 7 e-Infrastructure programme to considerably facilitate this process [3]. It is now carried on by the WeNMR (a Worldwide e-Infrastructure for NMR and struc-

tural biology) project since November 2010. The project aims at allowing groups lacking the resources to add NMR to their toolbox, as well as allowing dedicated NMR groups to improve their standard from basic practice towards cutting-edge research.

1.2 Small Angle X-Ray Scattering

Small Angle X-Ray Scattering (SAXS) is a widely used method for the low resolution structural analysis of biological macromolecules in solution [4, 5]. The sample is exposed to a collimated and possibly focused beam of X-rays, where the elastic scattering of the illuminated sample volume is recorded. The observed scattering is anisotropic function of the scattering angle thus yielding a one-dimensional scattering pattern after radially averaging the scattering data (typically recorded by a two-dimensional detector) around the beam center. From these scattering patterns, a variety of important overall parameters, e.g. information about size, volume or shape, can readily be derived. Although sample requirements of SAXS

are mostly similar to those of NMR or crystallography, i.e. the need of highly pure, monodisperse particles that remain solvable at high (starting from ca 1 mg/ml) concentration, there are distinct advantages of SAXS over the other methods. Firstly, no additional sample preparation step, e.g. crystallization, is required; secondly, on modern synchrotrons data collection and initial processing can be performed within a few seconds or minutes, allowing for an almost instant assessment of sample quality and its invariant parameters such as the radius of gyration, molecular mass and volume. Further, *ab initio* modeling can often be performed directly at the beamline, so that preliminary shapes of the measured protein are available within minutes, not hours or even days as required with other methods.

SAXS is highly complementary to NMR. Indeed, SAXS covers a broad range of macromolecular sizes but has a limited resolution (typically, 10–20 Å), whereas NMR has a limited applicability to large objects. For multisubunit complexes or multidomain proteins SAXS provides the information on intersubunit distances, but might be less sensitive to rotations; NMR data can yield local distances and orientations but does not probe the global arrangement of the components. The atomic models obtained by NMR can be cross-validated using SAXS or employed as rigid bodies in quaternary structure analysis of multicomponent particles (e.g. [6, 7]). In addition to the hybrid approaches, methods have already been developed to directly incorporate SAXS constraints during NMR structure refinement [8–10]. The two methods can also be jointly used to analyze flexible systems like partially unfolded proteins providing an ensemble description of such systems which provides more insight than a single conformation solution in the rigid case [11].

1.3 Objectives of WeNMR

The main objectives of the WeNMR project thus are:

- to provide integrated protocols for NMR and SAXS data processing
- to provide access to end users through user-friendly web interfaces
- to exploit Grid technology for computationally demanding tasks in structural biology
- to lower the barriers for access to Grid resources in life sciences, notably in structural biology
- to build a Virtual Research Community around a web portal

Considering the background sketched, these objectives set the challenges to be met within the project. The first of these has been the implementation of a new NMR Grid infrastructure. Historically, due to the requirements for processing of large amounts of data, NMR spectroscopy has always been intimately linked with high performance computing. Therefore, sites with high-end facilities for performing NMR measurements commonly also have considerable computational resources. For the WeNMR partners it thus came as a natural first step to integrate the existing resources into a Grid, offering a single standard for deployment and use of applications across the contributing sites, as well as a natural mechanism to share resources. Currently, the WeNMR project involves an operational Grid that was initially based on the gLite middleware, developed in the context of the EGEE project series [12]. Since May 2010 gLite has been supported by the EMI project (European Middleware Initiative), aimed to deliver a consolidated set of middleware components based on the four major middleware providers in Europe: ARC, dCache, gLite and UNICORE. The WeNMR Grid infrastructure (latter referred as ‘the Grid’) consists of the computing and mass storage resources of the four WeNMR core partner sites operating a number of central EMI Grid services like WMS [13], LB, Top-BDII, LFC and locally the EMI Grid services CREAM-CE and SE, plus a number of Grid sites part of the EGI infrastructure operated by those National Grid Initiatives that agreed to support our Virtual Organization. During 2011 WeNMR, in collaboration with the SBGrid Consortium, carried out a program to achieve interoperability with the Open Science Grid (OSG) infrastructure in US, that is based on the Globus middleware (see Section 4.3). Having an operational Grid, the programs involved in the different steps, which often require direct user interaction, have to be

interfaced in such a way that they can be run automatically. Focus has been initially placed on the CPU intensive programs, which have to be operated remotely as Grid enabled applications. This has to be done in such a way that they can be combined in automated workflows for protocolized processing of data, raising the issue of interoperability. In addition, web interfaces should be set up to be easy to use, yet sufficiently flexible for expert users. At the same time a mechanism is required to handle job traffic to and from the Grid. In the following sections, these different aspects are discussed in more detail, providing an account of the state of the project thus far. But before discussing the more technical details regarding the implementation, the application portals that are available are discussed in more detail.

1.4 Related Work

WeNMR is clearly not the first and only project offering web portals and science gateways. The latter are defined as a community-specific set of tools, applications, and data collections that are integrated together via a web portal or a desktop application, providing access to resources and services [14]. The concept introduced by the former TeraGrid project, XSEDE nowadays, is gaining momentum as several projects and e-infrastructures like WeNMR, Gisela and the European Grid Infrastructure promote the same gateway model to their communities (<https://gisela-gw.ct.infn.it/science-gateways>, <http://go.egi.eu/sciencegateways>). The gateways support a variety of capabilities including workflows, virtualisation software and hardware, visualisation as well as resource discovery, job execution services, access to data collections, applications, and tools for data analysis. The science gateway paradigm has been introduced to overcome the difficulties met by virtual research communities to access distributed resources. Their main goal is to hide the underlying complexity of a system, which gathers together heterogeneous resources, while enforcing access security to distributed computing infrastructures (DCIs) [15]. When it comes to developing gateways, there is no unique solution but typically various philosophies and implementations are followed. For instance,

next to the WeNMR portals described here, gUSE/WS-PGRADE [16, 17] and Genius [18] are two Grid portal technologies with a rich set of features, well known for their ability to build complex application workflows. Other examples of workflow-enabled Grid portals are OGCE [19] and Vine Toolkit [20]. All these tools offer features for managing the whole life cycle of generic workflows on diverse DCIs, like Grids and clouds. Collaborative environments, like InSilicoLab (<http://insilicolab.Grid.cyfronet.pl/>) and SOMA2 (<http://www.csc.fi/english/pages/soma>) offer modeling environments for researchers.

In the case of WeNMR most of the web portals have been developed prior to the WeNMR project, and independently of the use of the Grid. This explains the heterogeneity of the web portals. WeNMR main focus was not to define a perfect model to apply it to all our web portals, but to provide solutions that are attractive to the end user. In that context, allowing some flexibility for the WeNMR developers seems reasonable. Maybe more important than the choice of an architecture, a model, and its implementation, we consider that the success of the WeNMR project should be measured by its number of users, and their satisfaction.

2 The WeNMR Web Portals for Structural Biology

The web portals developed within WeNMR are among the most important elements of the project, as these form the points of entry for the end users. The ultimate goal is to offer to users registered with the eNMR/WeNMR Virtual Organization (VO) complete online protocols for processing NMR and SAXS data. The steps of a typical NMR data analysis are shown in Fig. 1. It is to note that each of the steps shown, and every program involved, has value by itself as web based service. For this reason, a piece-wise implementation has been adopted and programs that are ported to the Grid are simultaneously being made available as a web portal. Currently, nineteen application portals are operational and can be accessed on <http://www.wenmr.eu>. The term ‘application portals’ in this text is used

to describe the access to a specific software package via a web interface giving access in some instance to various pre-defined workflows or protocols settings within the application. The workflow are thus integral part of the software package and not implemented at a higher level using workflow managers like Taverna [21] or Shiwa (<http://www.shiwa-workflow.eu/>) for example. The WeNMR website, next to offering a large variety of documentation and a user support center, presents an aggregation of several applications portals (or simply ‘portals’). These portals provide access, among other services, to HADDOCK [22, 23] for the prediction of biomolecular complexes, Xplor-NIH [24], CYANA [25, 26] and CS-ROSETTA [27, 28] for calculating structures from NMR data, AMBER [29] and GROMACS [30] for structure refinement and molecular dynamics (MD) simulations, CcpNmr [31] for data conversion, MARS [32] for backbone assignment, TALOS+ [33] for torsion angle prediction, MDDNMR [34] for NUS (Non-Uniform Sampling) spectral processing, and selected applications from the SAXS data processing suite ATSAS [35]. Next to these available portals, several new ones for various NMR applications have been recently introduced, including the UNIO program [36, 37] that provides computational routines for each individual step depicted in Fig. 1. The main WeNMR portal for NMR applications is shown in Fig. 2. The central access point for all WeNMR portals is: <http://www.wenmr.eu/wenmr/nmr-services>.

2.1 HADDOCK

HADDOCK [22, 23] is an acronym for High Ambiguity Driven DOCKing and is a program to predict structures of biomolecular complexes from individual components. As the full name indicates, this approach in docking of biomolecules distinguishes itself from other methods by using external information to guide the docking process. Such information can be empirical, theoretical or both, pertaining to the residues or atoms involved in the binding interface. From this information ambiguous restraints are derived that are used to drive the docking. HADDOCK is particularly useful in predicting complexes from known experimental

structures of the partners using NMR data, such as chemical shift perturbations and residual dipolar couplings (RDCs). Chemical shift perturbations and RDCs can be obtained relatively easily and also for macromolecules of increasing size, making the large applicability of HADDOCK as a tool for cutting edge structural biology apparent. HADDOCK has proven its value within the CAPRI (Critical Assessment of PRediction of Interactions) experiment, a blind evaluation of the performance of current docking methods [22, 38–40].

The initial rigid-body docking process generates typically a large number of structures, typically in the order of thousands. From these a number of structures, typically several hundred, are selected for further flexible refinement. The results are scored, analyzed and returned to the user. The structure calculations are the CPU intensive part of the process typically resulting in 400–1,000 individual jobs being sent to the Grid, depending on the exact parameter settings.

Offering the full functionality of HADDOCK through a web portal requires putting forth a complicated form, contrasting with the objective of having a simple interface. To avoid compromises regarding user friendliness and functionality the portal is divided in four interfaces, corresponding to different levels of control and user experience:

- The *Easy Interface* requires no more than providing the two components of a complex and the residues of each that are involved in the interaction.
- The *Expert Interface* allows the user to provide his own customized restraints to be included in the docking process and to specify certain aspects of the system setup, sampling and analysis.
- The *Guru Interface* offers almost full control of parameters, allowing e.g. specification of symmetry and relaxation anisotropy restraints and RDCs as well as of parameters pertaining to the energy, the scoring and the analysis of results.
- Finally, for complete control a *File Upload Interface* is available, where a HADDOCK run parameter file can be provided. This is particularly useful for those who have their

Fig. 2 The WeNMR web portal (<http://www.wenmr.eu/wenmr/nmr-services>)

The screenshot displays the WeNMR web portal interface. At the top, the logo 'we-nmr' is accompanied by the tagline 'A worldwide e-Infrastructure for NMR and structural biology'. The user 'Christophe' is logged in. A navigation menu includes 'About Us', 'NMR Services', 'Support', 'News & Events', and 'Access'. Below this, a secondary menu lists 'Processing', 'Assignment', 'Analysis', 'Structure Calculation', 'Molecular Dynamics', 'Modelling', and 'Tools'. The main content area is titled 'NMR Services' and provides an overview of the available software packages. It is organized into several sections:

- Processing:** Includes MDD NMR, a service for processing non-uniformly sampled (NUS) 2D NMR spectra using Multi-Dimensional Decomposition.
- Assignment:** Includes AutoAssign (constraint-based expert system for backbone assignments), MARS (program for robust automatic backbone assignment of 13C/15N labeled proteins), and Xplor-NIH (allows running NMR structure calculations with NOE, dihedral angle, and paramagnetic restraints).
- Structure Calculation:** Includes CS-ROSETTA (generates 3D models of proteins using 13CA, 13CB, 13C, 15N, 1HA and 1HN NMR chemical shifts) and CYANA (allows automated nuclear magnetic resonance (NMR) structure calculation).
- Analysis:** Includes TALOS+ (predicts torsion angles using chemical shift, amino acid type, and Ramachandran map), AnisoFIT (fits pseudocontact shift and/or residual dipolar couplings), MaxOcc (graphical user interface for making rigorous numerical assessments), and iCing (webinterface to the NMR structure and data validation software suite CING).
- Molecular Dynamics Simulation:** Includes AMBER (suite of programs for molecular dynamics simulations) and GROMACS (versatile package for molecular dynamics simulation).
- Modelling:** Includes HADDOCK (flexible docking approach) and 3D-DART (3D-Driven DNA Analysis and Rebuilding Tool).
- Various Software Tools:** Includes Format Converter (facilitates conversions between NMR data formats), SHIFTX2 (predicts backbone and side chain 1H, 13C and 15N chemical shifts), and Antechamber (automates the process of developing a charge distribution and assigning atom types).

At the bottom of the page, there are three informational boxes:

- Cite WeNMR:** Usage of the WeNMR portals should be acknowledged in any publication. *The WeNMR project (European FP7 e-Infrastructure grant, contract no. 261572, www.wenmr.eu), supported by the national GRID Initiatives of Belgium, Italy, Germany, the Netherlands (via the Dutch BIG Grid project), Portugal, UK, South Africa, Taiwan and the Latin America GRID infrastructure via the Gisela project is acknowledged for the use of web portals, computing and storage facilities. *
- e-NMR:** WeNMR builds on the established eNMR e-Infrastructure project that is tightly integrated into the EGEE. eNMR is a steadily growing virtual organization currently representing the second largest VO in the life science area. [Read more ...](#)
- European Union:** WeNMR is an e-Infrastructure project funded under the 7th framework of the EU. Contract no. 261572. Includes the logo for the SEVENTH FRAMEWORK PROGRAMME.

The footer contains the copyright notice: © WeNMR 2010 | [Contact](#) | [Home](#) | [Colophon](#) | [RSS](#)

own standard protocol or who want to replicate a previous run with minor modifications. This option also offers a simple way to build pipelines from other applications.

To facilitate the user's task and keep the forms manageable, foldable menus were introduced that group related parameters under a single header. In this way, users only need to unfold groups of options that should be changed from their default values. Except for the *File Upload Interface*, the various HADDOCK portals share the data structure, albeit that part of the variables is fixed to predefined values for the *Easy* and *Expert interfaces*. This has the advantage that they can all couple to a single back end CGI (Common Gateway Interface) script to handle the request.

After issuing a request, the user is presented a link to a site where the progress can be followed. After the run is finished, the results can be viewed online and selected complexes or the complete output data of the run can be downloaded to a local machine.

The use of the HADDOCK portal requires registration with a valid Grid certificate, giving a username and password. These are thereafter used to sign service requests. The requests themselves are handled using an eToken-based robot certificate, as is explained in more detail in the implementation details.

2.2 Xplor-NIH

Xplor-NIH [24, 41] is one of the programs for structure calculations that have been ported to the Grid and are available through the WeNMR web portal. It is a versatile program that can be operated through a command line interface or with scripts in the specific Xplor language. It can incorporate several experimental and empirical information, such as NMR or SAXS data, to be used as restraints. These restraints are added to the topological description of the system and used to drive it to a folded state using simulated annealing. This procedure is typically repeated many times to obtain sufficient statistics regarding the goodness of fit of the structures determined against the experimental data. Since the different

annealing runs are independent of each other, they can be easily sent to the Grid as individual Grid jobs.

The portal uses a design that is different from the HADDOCK portal, and is aimed at more direct user interaction during the process. Users log in with their Grid certificate loaded in the web browser, gaining access to an environment where projects can be started, stored and managed. Structure calculation projects are initiated by filling in a form and providing files for the structures and topological descriptions of the molecules, as well as for the different restraints to be included in the calculations.

When the structure calculations have finished, the user can view and download the results. In addition, it is possible to select a number of structures for further energy refinement using the AMBER package [29] for MD simulations.

2.3 CYANA

Another widely used program for calculating structures from conformational restraints is CYANA (Combined Assignment and dYnamics Algorithm for NMR Applications) [25, 26]. Its main characteristics are the ability for iterative assignment of NOE peaks, and structure calculations through simulated annealing in torsion angle space. Like with Xplor-NIH, the structure calculations involve many simulated annealing runs, divided over several iterative cycles.

The design of the web portal for CYANA is similar to that of HADDOCK. Foldable menus are used to hide optional sets of parameters, by default presenting an intuitive menu offering a standard structure calculation protocol. The portal allows three modes of invocation of the service. Users can request structure calculation (i) using a set of upper distance bound restraints, (ii) providing a list of assigned peaks, or (iii) providing a list of unassigned peaks, in which case the automated peak assignment will be performed.

Use of the service requires having a license for CYANA and registering for use of the portal, presenting a valid Grid certificate. This will give a username and password that can be used to sign service requests.

2.4 CS-ROSETTA

Chemical-Shift ROSETTA or CS-ROSETTA [27, 28] unlike Xplor-NIH and CYANA, allows structure determination of proteins, based on chemical shift information alone. It thus bypasses the need for NOE (nuclear Overhauser effect) based distance restraints, which usually require considerable time to obtain. Further advantages of using chemical shifts are that these are among the most reliable parameters that can be obtained from NMR spectroscopy and that they can potentially be obtained for larger macromolecules for which NOEs become impractical. Chemical shift-based structure determination is computationally much more expensive than structure calculations using distance restraints. The most time consuming part of a CS-ROSETTA run consists of a large number (500 to 2,500) of independent Monte Carlo calculations that can be easily distributed over the Grid to calculate in the order of 10,000 to 50,000 structures.

Structure determination using CS-ROSETTA requires as only input the amino acid sequence and a list of chemical shifts and a number of parameters to control the process that can be changed from the default values. The chemical shifts are first used to select a set of protein fragments from a structure database, e.g. the Protein Data Bank (PDB) [42], based on the list of chemical shifts as predicted with SPARTA. Then the regular ROSETTA protocol [43] for Monte Carlo assembly and relaxation is used to reassemble the protein from the fragments. For the resulting models the chemical shifts are again predicted using SPARTA [44] and the deviations between the predicted and target values are used as a pseudo-energy term in the scoring of the models, yielding a ranking based on both overall structural quality as well as on the match with the experimental data.

The computational cost involved in chemical shift based structure determination makes CS-ROSETTA a typical example of a program that is beyond the capacity of most local sites. Here, the access to Grid resources through a web-portal, combining computational power and ease of use, clearly demonstrates its added value.

To use the service, users have to register with a valid Grid certificate to obtain a user name and password that can subsequently be used to sign service requests. The web interface for CS-ROSETTA itself is straightforward and only requires uploading a file containing chemical shifts in TALOS [45] format, and modifying the parameters to control the calculations. When the job has finished, the user receives an e-mail containing a link to the result page that gives an overview of the run, including some statistics and images to assess the overall quality of the results. The user can then select a number of structures to view in more detail or to download, or choose to download the whole set of results as an archive.

2.5 MARS

The fifth portal, MARS performs automatic backbone assignment of $^{13}\text{C}/^{15}\text{N}$ labeled proteins and is applicable to a wide variety of NMR data, including RDCs [32].

Its advanced features compare favorably with other assignment tools and include:

- simultaneous optimization of the local and global quality of assignment to minimize propagation of initial assignment errors and thus providing robustness against missing chemical shift information; applicable to proteins above 15 kDa using only C^{α} and C^{β} chemical shift information with connectivity thresholds as high as 0.5 ppm;
- applicable to proteins with very high degeneracy such as partially or fully unfolded proteins;
- combination of the secondary structure prediction program PSIPRED [46] with statistical chemical shift distributions, which were corrected for neighboring residue effects [47], to improve identification of likely positions in the primary sequence;
- assessment of the reliability of fragment mapping by performing multiple assignment runs with noise-disturbed chemical shifts.

Registration and a valid Grid certificate are needed to use the service. Its interface is similar to CYANA: after uploading peak lists a file containing assigned chemical shifts is downloaded after job has finished.

2.6 TALOS+

TALOS+ [33], like its predecessor TALOS [45], is a program to predict torsion angles for amino acids given information regarding chemical shift and the probable regions in the Ramachandran plot for each type of amino acid. TALOS+ distinguishes itself by the inclusion of a neural network component, the output of which is added as an empirical term in the conventional TALOS data base search. To prevent assignment of torsion angles to the backbone of flexible regions, TALOS+ first identifies such regions using the flexibility prediction program RCI developed by Berjanskii and Wishart [48].

The TALOS+ portal offers a simple interface where an input file in either TALOS or BMRB (Biological Magnetic Resonance Data Bank [49]) format can be uploaded. In addition a number of PDB ID codes can be given, indicating structures that should be excluded from the calculations. Unlike most of the other portals, the TALOS+ portal can be used without a Grid certificate, as the calculations are run on a local server.

2.7 MDDNMR

The MDDNMR [34] portal can process individual NUS (Non-Uniform Sampling) multidimensional NMR spectra. The interface supports the NUS data recorded by two major spectrometer brands: Varian and Bruker. The main advantage of usage of NUS data is the substantially higher resolution in the indirect spectral dimensions. The NUS acquisition mode of both Vnmr and TopSpin spectrometers controller software makes use of standard NMR experiments, except that only a fraction of a full data set is recorded. This means that it can be used with virtually any pulse sequence available. After acquisition, MDDNMR replenishes the missing data points in the full matrix. The resulting regular spectra are then processed conventionally with FFT (fast Fourier transform), LP (Linear Programming), window functions etc. The current portal allows this to be repeated for each experiment in the dataset; several such high-resolution experiments are processed sequentially as single matrices, and the resulting high-resolution FT-domain spectra,

after peak-picking, are amenable for analysis, e.g. automated assignment using MARS. Most of the experimental data types supported by MDDNMR can be processed via the portal, including constant-time acquisition (CT), J-coupling splitting etc. Use of the service requires registering for use of the portal, presenting a valid Grid certificate.

2.8 CcpNmr

The CcpNmr portal, for CcpNmr [31] based data conversions, is not directly related to NMR data processing, but is an important element within the WeNMR project. The reason for this is the fact that the programs already ported to the Grid often have their own data formats, making it impossible to combine these as steps in a direct pipeline. Establishing interoperability of such programs requires conversion of output from one step to match the input of a next step. This is what CcpNmr FormatConverter was designed for. It has a comprehensive internal data model [50] that can contain all different types of NMR related data. These data can be imported from files in a large number of formats. Likewise, the data stored in a CcpNmr project can be exported in any of the file formats required, provided that the data are present in the model. In this way, CcpNmr provides a straightforward approach in meeting one of the challenges of the WeNMR project, namely establishing interoperability of the programs involved in NMR data processing and building automated protocols for complex tasks.

The portal for CcpNmr FormatConverter was developed to offer the WeNMR members an easy solution for matching program output and input during the steps involved in processing of their data. At present, the portal allows conversion between several file formats, aimed at facilitating the use of the other portals. The file conversions for the portal are performed locally, as these are not computationally intensive. Use of the CcpNmr portal thus does not require a Grid certificate.

The CcpNmr FormatConverter portal greatly simplifies data transfer between programs, but it is still a separate, user-initiated step. Work is currently in progress towards truly seamless interoperability, where data conversion happens in

the background without the user being aware of it. This requires storing data in the more comprehensive and precise CcpNmr internal data model and integrating the conversion to and from program-specific data formats with the calculation launch. Once the input data for a calculation are known and precisely defined, it is much easier to interpret the calculation results in a fully automatic manner. In particularly difficult cases, additional data necessary for interpreting the results can be generated as part of the program launch. The use of a single data format along a processing pipeline also makes it possible to create an alternative, semi-standardized pipeline interface, and simplifies the handling of workflows where a single set of starting data is used to launch different calculations in parallel.

The first result along these lines is the alternative HADDOCK interface created as part of the Extend-NMR project (www.extend-nmr.eu). The interface allows complete control over HADDOCK input data and parameters, and works entirely within the CcpNmr data model. The resulting set-up is exported as a HADDOCK parameter file, which can be executed on the HADDOCK portal through the file upload interface (2.2). Another example is the facility for running the Cing validation server on CcpNmr project files directly from the CcpNmr Analysis program.

2.9 AnisoFit

The AnisoFIT web portal allows the determination of the magnetic susceptibility tensor anisotropy of a paramagnetic center (using pseudocontact shift data or RDCs) or of the anisotropy of the diffusion tensor (for RDCs induced by orienting media). This is done by fitting the input experimental data to the protein coordinates, which the user must provide in the form of a PDB file. In addition, it is necessary for the user to provide information on the experimental conditions (temperature, magnetic field) and to define a value for the experimental error (if not specified in the input data file). The results are provided in a graphical as well as tabular form, which can be saved locally. Unlike most of the other portals, the AnisoFIT portal can be used

without a Grid certificate, as the calculations are run on a local server.

2.10 AMBER

AMBER [29] is a suite of programs for use in molecular modeling and molecular simulations of proteins and nucleic acids that is now available through the WeNMR web portal. The AMBER-based Portal Server for NMR structures (AMPS-NMR) has been developed to provide an user-friendly entry point mainly designed for the energy refinement of NMR structures based on restrained molecular dynamics (rMD), but permitting also free MD simulations [51].

AMPS-NMR offers a personal workspace for the creation and management of calculations. A new user with her/his personal certificate installed in the web browser can access to the portal from WeNMR web page and straightforwardly create a new user login and password. The portal allows users to create a new single calculation or a so-called project in which it is possible to save a number of different calculations that belong to a common project (e.g. MD refinements of protein structures generated with different ensembles of restraints).

The setup of a rMD refinement with AMPS-NMR comprises four steps:

- In the first step the protein structure to be optimized is uploaded. Here it is possible to upload one PDB file, whose format will be automatically validated and converted into the format recognized by AMBER. During this step the user can add explicit water molecules to solvate the protein, add counter ions, insert new bonds for selected atoms (e.g. protein to metal, disulfide bridges). For NMR structures, which are typically represented by a bundle of 20–40 different conformers, this first step is carried out only for the first conformer and then automatically applied to all the structures in the bundle. For each structure in the bundle, an individual job is sent to the Grid.
- The second step manages the NMR restraints. Four types of restraint are allowed: NOE, dihedral angles, RDCs and pseudocontact shifts (PCS). The last restraints are the so-called

paramagnetic restraints. For all restraints it is possible to upload Xplor, Dyana, or Cyana files. For paramagnetic restraints, it is possible to fit the anisotropy tensor directly in the web site, since all the functionalities of ANISOFIT are embedded in AMPS-NMR.

- The third step manages the setting of rMD calculations. The user can set manually all MD parameters or select a pre-configured set of parameters, ranging from a simple energy-minimization to a multi-step MD refinement protocol.

- The fourth and final step allows the user to give a name to the calculation and submit it to the Grid.

After the job has been submitted, the user can easily check the status of the job (i.e. Scheduled, Running and Completed) through the user-friendly jobs management web page, as shown in Fig. 3. Each completed job is automatically retrieved from the Grid. All output data are stored in the personal workspace of the user. Within this workspace, the user can browse a completed

Fig. 3 AMPS-NMR jobs management. This webpage allows users to manage their jobs submitted to/running on the Grid (box A). In particular, it is possible to check the status (e.g. scheduled, running, completed) of either individual or all calculations (box B), and to erase calculations either individually or for all in a user-selected status (box C). Completed jobs are automatically retrieved from the Grid and their status is changed to 'E'.

The screenshot shows the AMPS-NMR web portal interface. At the top, there is a navigation bar with links for 'WeNMR home', 'NMR services', 'SAXS services', and 'WeNMR Support Center'. Below this, a user greeting 'WELCOME LUCIO FERELLA' is displayed. A menu bar contains 'My Account', 'Amber', 'Jobs', 'Projects', and 'Logout'. A warning message states: 'The job status does not update automatically (even if you log out and log in again). To check if the job status has changed, please use the Check all or selected jobs status link above the Job List.' Below the warning, a section titled 'A quick explanation about some Status column values:' lists:

- **S** (Scheduled/Submitted) -- the job has been correctly submitted and is waiting to be processed;
- **R** (Running) -- the job is currently running;
- **E** (Completed/Retrieved) -- the job has been completed and automatically retrieved from the Grid.

 A note follows: 'If your job has spent more than 12 hours in the S status, we advise you to kill and re-submit it.' The main 'Job List' section features a table with columns: Project Name, Calculation Name, Job Guid, Start Date, Status, Log, and Kill Job. Above the table are three highlighted boxes: 'B' (containing 'Check all or selected jobs status'), 'A' (containing 'remove all jobs'), and 'C' (containing 'Select status' and 'remove'). The table contains 12 rows of job data, with statuses alternating between 'S' and 'R'.

<input checked="" type="checkbox"/>	Project Name	Calculation Name	Job Guid	Start Date	Status	Log	Kill Job
<input type="checkbox"/>	protRefinement	reTIP3P20models	click here to view	2012-04-18 11:47:47	S	Read	
<input type="checkbox"/>	protRefinement	reTIP3P20models	click here to view	2012-04-18 11:47:43	S	Read	
<input type="checkbox"/>	protRefinement	reTIP3P20models	click here to view	2012-04-18 11:47:38	R	Read	
<input type="checkbox"/>	protRefinement	reTIP3P20models	click here to view	2012-04-18 11:47:34	R	Read	
<input type="checkbox"/>	protRefinement	reTIP3P20models	click here to view	2012-04-18 11:47:30	S	Read	
<input type="checkbox"/>	protRefinement	reTIP3P20models	click here to view	2012-04-18 11:47:27	R	Read	
<input type="checkbox"/>	protRefinement	reTIP3P20models	click here to view	2012-04-18 11:47:23	S	Read	
<input type="checkbox"/>	protRefinement	reTIP3P20models	click here to view	2012-04-18 11:47:19	S	Read	
<input type="checkbox"/>	protRefinement	reTIP3P20models	click here to view	2012-04-18 11:47:15	R	Read	
<input type="checkbox"/>	protRefinement	reTIP3P20models	click here to view	2012-04-18 11:47:11	R	Read	
<input type="checkbox"/>	protRefinement	reTIP3P20models	click here to view	2012-04-18 11:47:08	S	Read	

calculation and download all its various files, or just extract the bundle of refined conformers. Energy and violation statistics can also be obtained automatically.

AMPS-NMR computed more than 4,500 jobs in the last year.

2.11 GROMACS

Like AMBER, GROMACS [30, 52] is a suite of programs for molecular simulations. GROMACS is freely available, open source and known for its versatility and performance. Unlike most other packages, GROMACS is not linked to a specific force field, but can perform simulations with any force field that has been ported to the file format required. Available force fields include recent versions from GROMOS [53–56], AMBER [57], CHARMM [58] and OPLS [59].

The GROMACS portal [60] now made operational within WeNMR interfaces to a protocolized, yet flexible GROMACS workflow. This workflow has originally been developed to facilitate setting up and running simulations for parametric studies, including statistical comparison of sets of simulations [61, 62]. Such studies typically involve many simulations, requiring robust, automated processing, yet with sufficient flexibility to specify particular settings. The workflow starts with generating a topology for protein and nucleic acid chains, according to the force field specified. The system is then solvated and equilibrated in various steps, followed by a short run production run. A second interface allows users to upload pre-equilibrated systems for production runs. For a more detailed description of the workflow and its various options please refer to [60].

The user front end of the portal resembles other WeNMR portals, using foldable menus to organize the form. However, this portal, shown schematically in Fig. 4 represents a new class, together with the UNIO portal discussed next. The new design aims to provide a more dynamic and user-friendly experience, both for the end-user, as well as for the server administrator. To this end, the portal comprises of three front-end pages, two for regular users and one for admin-

istrators, which are derived from a single portal description. From the user perspective, the first line of entry is the submission page, where a file can be indicated for uploading and the simulation conditions can be specified. Following submission, the user is redirected to the results page. This is a PHP enabled page that updates every minute, dynamically loading data representing the actual state of the simulation, allowing the user staying up-to-date. Major events are also communicated through e-mail.

The dynamic nature of the user and administrator front-ends results from a combination of PHP code and Python CGI scripts. The scripts interface to a global server configuration file or to a specific job configuration file, which are also actively used by the server processes. This allows an active flow of data between the front-end pages and the server as well as individual processes.

A new job is processed initially by a single python CGI script, based on the Spyder data model also used in the HADDOCK server [63]. First the user credentials are validated, the input file format is checked and the input is inspected for the presence of non-standard amino acids and ligands. Then a project directory is created, together with a job description file that is placed in the server job pool, which is under control of the jobmanager and Gridmanager daemon processes.

The jobmanager process regularly checks each job in the server job pool. New jobs are submitted to the Grid by placing them in the Grid pool, and active jobs are inspected for a change in status, which will invoke a specified action. The jobmanager also performs sanity checks to assert consistency between the files in the project directory and those in the Grid pool. If an irregularity is detected a recovery is attempted, but if that fails the job will be terminated. When a job finishes, a post-processing function is invoked that extracts statistics to display on the results page, cleans the job and makes the job archive available for download. The jobmanager is also responsible for keeping the user informed by e-mail about major changes in the status.

The jobs put in the Grid pool come under control of the Gridmanager, which takes care

Fig. 4 The GROMACS web portal, accessible via <http://www.wenmr.eu/wenmr/molecular-dynamics-software>. Three categories of input form are displayed to the end-user: (i) the required parameters, to supply a PDB file containing the coordinate of the structure, (ii) the optional parameters, to control the simulation time, the forcefield, the solvent model and the electrostatic treatment, and (iii) the advanced parameters to control additional expert settings

GROMACS
e-NMR (GRID-enabled) web portal

[Return to the WeNMR website](#)

WELCOME TO THE WENMR WEB PORTAL >>

Welcome to the GROMACS web server your entry point for molecular dynamics on the GRID. New molecular dynamics simulations are started by filling out the form below. Once submitted you will be redirected to the results page for your run where you will be informed on its progress and will be able to retrieve the results when the job is finished.

Required parameters

PDB file
Please upload a PDB file
[Help](#)

Optional parameters

Simulation time
Please specify the simulation time in ns
Output resolution ns
Forcefield
Please specify the forcefield to use
Solvent model
Solvent model (default according to force field)
Treatment of electrostatic interactions
Please choose how to treat electrostatics
[Help](#)

Advanced parameters

Salt concentration
Please specify the salt concentration (mol/l)
Temperature and Pressure
Specify temperature (K)
Specify pressure (Bar)
Minimal distance between periodic images
Specify minimal distance
Define specific seed
Define specific seed for random number generation (velocity distribution)
Various settings
Perform basic set of analysis on the MD results
Perform simulations in a Near-Densest Lattice Packing simulation cell. Using roto-translational constraints is advisable
Use roto-translational constraints
Use virtual sites
[Help](#)

of trafficking to and from the Grid: submitting jobs, querying statuses, and retrieving results from finished runs. The jdl files submitted to the Grid by the Gridmanager use functionality from the new EMI CREAM-CE enabled sites on the Grid.

This involves the CPUNumber argument which allows GROMACS to run on 6 CPU's using multithreading, speeding up the simulation significantly. The presence of this argument instructs, the workload management system (WMS) to

select only those sites. The Gridmanager will attempt submitting, querying or retrieving a job a preset number of times, before reporting the job as failed to the user.

The GROMACS server can be used for protocolized single runs, but also allows simultaneous execution of multiple molecular dynamics simulations on Grid infrastructure for parametric studies. A more detailed description of this server and the GROMACS workflow involved, expanding on the various options available is now published [60].

2.12 UNIO

UNIO (latin: to combine into one, large precious pearl) is a comprehensive NMR data analysis suite that assembles expert systems for all sequentially applied parts of the conventional NOE-based NMR structure determination workflow. UNIO interconnects the MATCH algorithm for automated sequence-specific backbone assignment [37], the ASCAN algorithm for automated side-chain assignment [36], the CANDID algorithm for automated NOE assignment [26], and the ATNOS algorithm for automated NMR signal identification [64]. These four major data analysis components form the core of a UNIO NMR structure determination. Crucial key elements in order to seamlessly integrate the four principal data analysis components are routines for automated chemical shift referencing and automated chemical shift adaptation between the different input NMR spectra used. The standard UNIO data analysis protocol requires only a minimal set of six NMR spectra, namely 3 APSY and 3 NOESY. Notably, UNIO promotes the use of projection NMR spectroscopy of high-dimensional spectra. Conventional triple-resonance NMR experiments can alternatively be used instead of APSY data for obtaining the backbone resonance assignments using MATCH. The UNIO application suite is thoroughly road-tested as attested by more than 30 *de novo* protein structures determined using the entire four-step unsupervised protocol (with minor human intervention), and by more than hundreds to thousands of NMR structures determined using individual components

implemented in the UNIO NMR analysis suite (as for the NOE assignment routine in CYANA ported also to WeNMR). The power of UNIO for unsupervised protein NMR structure determination is well documented as proven by the results obtained after the first round of the eNMR-organized NMR community-wide blind testing initiative CASD-NMR (first ever performed objective NMR software comparison) for assessing the potential of different computational approaches claiming to determine three-dimensional protein structures in an unattended manner. Notably, it is worth mentioning that all the programs ported to the WeNMR Grid infrastructure performed very well, confirming that the WeNMR portals (CYANA, CS-ROSETTA, UNIO) present indeed cutting-edge research for protein NMR structure determination.

The UNIO portal is designed as a multiple component web portal, similar to the HADDOCK and GROMACS portals. The UNIO portal consists of four principal web sub-interfaces, corresponding to the essential parts of a NMR structure determination project. It offers expert systems for the computationally demanding tasks of NMR signal identification, sequence-specific backbone and side-chain resonance assignment, and comprehensive collection of distance restraints with the latter task focusing on the primary source of NMR-based protein modeling. UNIO is compatible with the most commonly used and extremely powerful NMR structure calculation programs, such as CYANA and CNS, which are also operational on the WeNMR Grid infrastructure. This interfacing to other programs represents a template example of smoothly interconnecting different web portals in WeNMR that is provided “under the hood” and do not require any intervention by the web portal user. The developed interfacing to the two other abovementioned WeNMR portals conveniently equips the structural biology community with all computational processing tools necessary for pursuing their own, complete protein structure determination by NMR. All UNIO sub-portals support all commonly used file formats for different input data (AMBER, ANSIG, BMRB, BRUKER, CARA, CNS, CYANA, FASTA, NMRVIEW, PDB, SPARKY, XEASY,

Xplor, Xplor-NIH) assuring the ease of use of the individual UNIO portals. In addition, the entrance UNIO portal provides a newly developed expert system for detecting errors in chemical shift referencing and sequence-specific resonance assignments (UNIO-ShiftInspector available as fifth sub-portal). Again, the interplay between the result of this UNIO subtask and other WeNMR portals (CS-Rosetta, TALOS+) shows the tight interconnection between different WeNMR-developed portals.

Concerning the practical use of the UNIO portal, for each of the UNIO sub-interfaces (MATCH, ASCAN, ATNOS/CANDID, CANDID, and UNIO-ShiftInspector), folded menus are provided, in which standard default values for controlling the calculation process are already filled in. After job submission, the user is presented a link to a web site that enables monitoring the progress of the calculation. After the calculation of a subtask has finished, the user will be informed by email providing a web address in order to store the complete output data at a local computer. The use of the UNIO portal requires registration with a valid Grid certificate, and subsequently the user will be provided with a login name and password for use of the UNIO portal. The UNIO portal is free-of-charge for non-profit organizations.

2.13 iCing

Validation of the structural models constitutes an essential step in the whole process of structure determination, not only by NMR but also all other methods, such as X-ray crystallography. Its importance is underscored by the recent establishment by the wwPDB of validation taskforces, that will draft recommendations for the scientific community. The WeNMR Grid is designed to incorporate a validation server as integral part of the flow of data, allowing for users to examine their structural results prior to final submission to wwPDB databases.

The iCing web portal (<https://nmr.cmbi.ru.nl/icing/>) presents a front-end to the CING validation suite. The software tools implemented in

CING provide for a comprehensive validation analysis, which includes the results of as many as eleven [65] different programs and integration of all results into a single, comprehensive and interactive HTML/Web2.0 report.

In addition to its own CING project files, the iCing server also accepts as input 'plain' PDB files, Cyana, or CcpNmr project files containing all data generated in the typical course of a structure calculation project, e.g. shifts, peaks, restraints, coordinate data, etc. The iCing server also allows direct communication with programs in the WeNMR environment as a service. The CcpNmr program Analysis for example, directly uploads its data and retrieves the iCing results and a similar mechanism is currently under implementation for the UNIO and Xplor-NIH structure calculation programs. The iCing front-end is based on the popular Google Web Toolkit (GWT) technology. The iCing server routinely processes on a weekly basis all new NMR entries released by the wwPDB, which are accessible via the NRG-CING portal [66]. Use of the iCing portal does not require a Grid certificate.

2.14 ATSAS

Since its initial release in 2003, the ATSAS software suite for Small Angle X-Ray and Neutron Scattering (SAS), became the de-facto standard in SAS data analysis of biological solutions. The package provides tools for data reduction and analysis, for ab initio shape determination and rigid body modeling, it provides means to add missing fragments to complexes, can provide information on oligomeric mixtures and hierarchical and flexible systems. Since 2003 it was downloaded over 30,000 times by over 5,500 users from 1,500 institutions worldwide. It helped to facilitate about 1,100 publications in the time from 2007 to 2011, of which roughly 700 were devoted to biological solution SAS. This represents more than 50 % of all publications in this field.

ATSAS is a suite containing over 20 major programs, and each individual application of the suite was ported to be usable on Grid nodes. They are as such available in a package for download and

individual use, for direct usage on Grid CEs associated with the enmr.eu VO where the package was installed on, and through the associated Grid portals. Although development of the Grid portals for ATSAS began only recently, progress was quickly made due to the experiences accumulated with the ATSAS Online service, a community service that allowed running of ATSAS programs on a small-scale cluster facility of EMBL Hamburg. With its rapidly growing user base, relocation to the Grid is the logical next step.

The first available ATSAS Grid portal provides an easy to use interface for the ab initio bead modeling program DAMMIF [67]. Contrary to NMR methods that allow the reconstruction of models at atomic resolution, bead modeling with SAS is meant to provide information of the overall shape of the macromolecule. DAMMIF achieves this goal by filling a search space with densely packed dummy atoms, or beads, which initially are randomly assigned to be part of the molecule space or part of the solvent space. Employing a simulated annealing search procedure, DAMMIF then attempts to find a compact confirmation of interconnected beads whose simulated scattering pattern fits the recorded scattering data well; other conditions besides compactness can be enforced simultaneously. It is to note that the dummy atom positions in the final ab initio model do not represent atoms as they do in NMR results, but only scattering centers at defined locations. These must not be interpreted in any other way but as “filled space”.

Although DAMMIF is not too computationally demanding (the execution time of on a typical PC computer, depending on the search parameters, generally varies from 30 s to a couple of hours), one needs to run DAMMIF multiple times to verify the stability of the solution [68]. This requirement, which holds for any ab initio shape determination application, makes ab initio modeling a perfect use case for the scalability of the Grid.

Besides DAMMIF it is planned to provide portals for DAMMIN, MONSA [69] and GASBOR [70] within a short time frame. Other parts of the ATSAS suite, including rigid body modeling programs utilizing also the NMR information, will be made available on portals in due course.

2.15 Accessing the WeNMR Web Portals

Access to the WeNMR portals is granted to a user upon completion of the following steps:

- The user needs to obtain a Grid certificate. The procedure is country-dependent, it usually involve identification (passport, ID card) with the registration authority. The WeNMR website provides a support page to guide as much as possible users to the proper registration authorities and procedures (see <http://wenmr.eu/wenmr/access/registration>).
- The Grid certificate needs to be installed in a web browser.
- The user has to join the enmr.eu Virtual Organization (<https://voms2.cnaf.infn.it:8443/voms/enmr.eu>) using the Grid certificate enabled web browser. The ‘Acceptable Usage Policy’ of EGI must be accepted.
- Because each software package used by the WeNMR portals have different license policies, the user need to register separately for each of the portals to be used. All web portals are accessible free of charge for users working in non-profit academic institutions.
- The user will receive a username and password for each portal to be used, once a WeNMR admin has checked the validity of the enmr.eu membership. The WeNMR team is in a process to adopt a user-friendlier authentication process such as Single Sign On, allowing users to manage their registration with the various portals from a central location.

Every run launched via one of the portals is effectively submitted using an eToken-based robot certificate. In this way, users do not need to have their Grid certificate loaded into a web browser to access the application portals. The application portals keep track of all user submissions.

Most of the portals are documented with tutorials (to explain the use of the web interface) and with ‘command line use cases’ for advanced users who want to submit directly (i.e. not through the web portals) their job to the Grid, using

the software packages that are already deployed. WeNMR also offers a help center (<http://www.wenmr.eu/wenmr/support/help-center-forums>) to help users with specific questions.

3 Structural Biology on the Grid: Design Strategies and Implementation

Successfully running web portals requires a proper machinery to handle requests. This machinery involves various steps that can be categorized in three layers of operation: The server level involves handling of service requests, either by direct human interaction or through requests from another machine. This stage includes input type checking. The next level involves preparation, trafficking and monitoring of jobs between the server and the Grid. The third layer is the core layer involving the process(es) to be run on a worker node. The tasks associated with these different levels are conceptually unrelated and allow for a component based development approach, in which distinct tasks are programmed in a most generic form. This has the

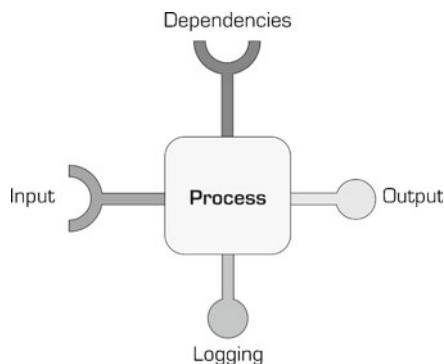


Fig. 5 Process model. To facilitate implementation and management, processes are represented and, if needed, rewritten in a manner adhering to a simple five node model, with the process itself as the central block that has four connectors: an input and an output connector, a connector for dependencies and one for logging. The output of one process can be connected to the input of another to build larger sequences or workflows. Obviously, each connection can involve several components, e.g. a process' input can consist of several files and/or option settings. The process itself may be regarded a black box, as long as the connections are well-defined

advantage that such building blocks can be easily maintained, adapted and reused.

To facilitate the component-based implementation, a single, simple model, illustrated in Fig. 5, was designed within the WeNMR consortium for the representation of processes. This model characterizes any process as a block with four connectors: input, output, dependencies and logging. The input and output connectors allow building larger sequences or complex workflows. The dependencies are considered static to a process and the logging is for messaging and provenance. Logging information can also be used to check the status and react to errors. Processes are all shaped to adhere to this simple model, which can be achieved by rewriting programs or by wrapping them inside a script. Doing so, a set of process modules is obtained that can easily be combined in a pipeline.

Process pipelines are often built imperatively, using one of the standard scripting languages. But an imperative approach has the drawback that it is inflexible: e.g. a failure will cause the whole pipeline to fail and a process has to be started anew. For this reason, a partial declarative approach was designed, in which direct communication between processes from the different layers of operation are eliminated. Rather, output from one level that forms the input for another is 'pooled' on disk. Processes from the next layer that depend on these data are run periodically, scanning the pool for data matching the input requirements. This has the advantage that the state of all processes is naturally check pointed and that the use of computational resources can be better controlled. How this approach is used to connect the web portals to Grid calculations is illustrated in Fig. 6 and explained in more detail below. Note that this description is relatively general: while the GROMACS, HADDOCK, UNIO and CS-ROSETTA portals strictly adhere to this model, the remaining portals currently only partially follow the separation into the three layers.

3.1 The First Level: Request and Data Handling, Invocation of the Service, Reporting

The first level of operation involves the interaction with the user, both processing the request, as well as presenting the results. A service request is

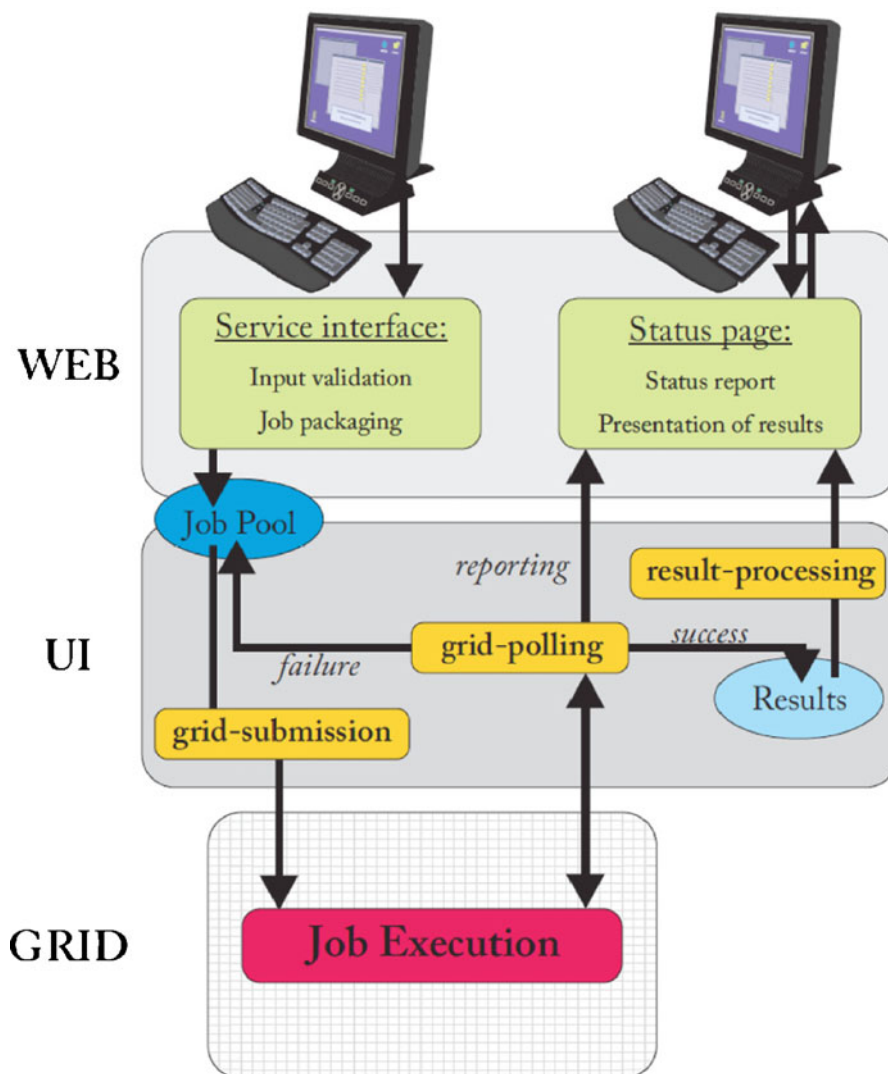


Fig. 6 Grid job submission management using job pooling. The figure shows a general scheme for managing job trafficking to and from the Grid, using server side job pooling. This scheme is characterized by a separation of three layers of operation, between which there is no direct communication. Green boxes indicate user interaction, whereas yellow boxes indicate jobs that are running periodically as daemon jobs and that use an eToken-based robot certificate for generating a Grid proxy. The blue ellipses represent 'pools', which are used for storage of job or result packages. User service requests are processed on the server, up to the point of generating a job package

that is stored on disk. On the Grid UI (User Interface) a daemon job (Grid-submission) is running on a scheduled base scanning the 'job pool' for job packages and submitting these to the Grid when found. Another daemon job (Grid-polling) is periodically checking running jobs for their status, retrieving the results when ready and placing these in a result pool. Finally, results are presented back to the user, possibly after post-processing (results-processing). Currently the HADDOCK, CS-ROSETTA, UNIO, CYANA, MARS, GROMACS and MDD-NMR portals, which all send jobs to the Grid, are implemented following this model

made by filling in a form that is parsed by a CGI script. This script also performs type checking and validation of the user input and presents the user a unique ID with which the results can be retrieved.

Both the web form and the CGI script depend primarily on the data to be provided, and it is possible to generate these automatically from a description of these data. To this purpose, the

Spyder framework (S.J. de Vries, unpublished) was designed, initially to facilitate setting up and managing the portals for HADDOCK. Next to the generation of web forms, Spyder natively supports data validation for known types and can convert between data types if all intermediary conversions are defined. Thus, Spyder offers a single framework for managing most of the elements involved in the first stage of processing of requests.

The data parsed and validated are then processed by a request specific script that combines all data and control parameters required for further processing into a self-contained job package. This package is subsequently placed into a job pool directory, which ends processing of the request at the first level.

3.2 The Second Level: Grid Job Preparation, Trafficking and Monitoring

At the second level, a daemon job is running periodically, scanning the job pool for jobs that are ready to be run on the Grid and submitting these when found. In principle, this daemon job does not require information regarding the nature of the job, although in practice different instances are run, each linked to one type of job to better control the work load associated with the different tasks.

A separate daemon job is running, also periodically, checking the status of the jobs running on the Grid and retrieving the results when finished. Alternatively, this process can resubmit the job when it has failed. The results are put back, after validation, in a place where they can be accessed through a web page. Like the submission process, the polling and retrieval process is in principle independent, since all information regarding the job, such as the directory to place the results in, are contained in the job package.

Submission, polling and retrieval of output are handled using a standard toolbox for Grid operation, which, in the case of WeNMR, is the EMI middleware suite (using standard commands such as *glite-wms-job-submit*, *glite-wms-job-status* and *glite-wms-job-output*). Accordingly, the jobs that operate at this level require the use of a valid proxy. To facilitate proxy management, all of the processes at the second level of

operation are running using an eToken-based robot certificate, in accordance with the security requirements for data portals formulated by the Joint Security Policy Group (https://www.jspg.org/wiki/VO_Portal_Policy). The submission and polling daemons are documented and can be downloaded at <http://www.wenmr.org/wenmr/automated-Grid-submission-and-polling-daemons>.

3.3 The Third Level: Primary Tasks

The third level of operation involves the tasks running on the Grid. This requires programs to be ported to the Grid, but that process is relatively straightforward. The only aspect that is different from more common strategies is that the processes have to adhere to the process model discussed previously, facilitating automation and provenance.

3.4 Software Deployment

Software required by the various portals is managed by the WeNMR team. Jobs sent to the Grid by portals are not including the executables. Instead, they use pre-deployed software on the various sites supporting the enmr.eu VO. Software deployment is done using standard Grid jobs with software manager role. Once an application has been deployed and shown to give proper results from test Grid jobs, a software tag with the application name is added to the particular site (using the *lcg-tags* command). The web portals, on their side, create jdl scripts that use requirements to target the sites offering a particular software. For example, to target a site supporting the CNS software (version1.2) the following requirement is added the the jdl script:

```
Member("VO-enmr.eu-CNS1.2",other.  
GlueHostApplicationSoftwareRunTime  
Environment)
```

This mechanism allows flexibility in software deployment and site tagging for specific applications and has the advantage that no specific site needs to be defined in the jdl script at submission time.

4 Results

4.1 Status and Statistics

Since the start of the eNMR project in fall 2007, considerable progress has been made, both in the deployment and the utilization of an infrastructure for structural biology. Currently, the WeNMR-dedicated infrastructure is distributed over four partner sites, which together provide a body of 320 dedicated CPUs, and 3.4 TB of dedicated storage. Resources are shared with 29 other sites, giving access to over **42,000 shared** CPU cores and over 200 TB of shared storage. While the storage space is available for all enmr.eu members, it is barely used by the web portals application, for two reasons: Firstly because the various applications used are deployed on the different Grid sites by WeNMR administrators. This means that the statically compiled binaries with the necessary auxiliary data needed for execution are installed in a software directory of each Grid site. Once a site has been validated for a specific application, it is “tagged” with the appropriate software tag. WeNMR jobs JDL’s include a reference to the tag so that the WMS consider only the relevant Grid sites for job execution. Secondly, most of the input and output data transferred via network are relatively small (several MB). Finally, the results are retrieved locally on the specific site where a given portal is running where they are further analyzed and post-processed. Upon successful completion of their job, users are notified by email. Results are stored for a limited period of time so that the user can retrieve it.

About 99 % of the CPU resources available to the enmr.eu VO are shared with other VO’s. However, there are no assigned CPU usage limits, implying that potentially there could be competition among different VO’s for resources. In fact, WeNMR already accounts for more than 30 % of Grid usage in the Life Sciences domain. Over the last year, around 2 million jobs have been run on the Grid, corresponding to about 1,000 years of normalized CPU time. The overall CPU efficiency, the total CPU time divided by the total wall time, of all jobs (excluding software manager role) was around 89.3 %

(statistics taken from the EGI accounting portal, <https://accounting.egi.eu/egi.php>). In terms of consumption of Grid resources, WeNMR has access to 42 k CPU-cores and has reached a maximum of 4.1 k jobs running simultaneously, i.e. ca. 10 % of the total capacity.

Including the nineteen applications that have already been made available through a web portal, twenty programs have been ported to the Grid so far and are accessible for advanced users with a computer configured with a EMI User Interface service. In addition, we have successfully tested a new setup based on cloud computing for software that is hard, if not impossible to install on the Grid. For the NRG-CING portal [66], the CING (G. Vuister, et al., CING; an integrated residue-based structure validation program suite, *manuscript in preparation*) structure validation program suite was run on Virtual Machine instances called VirtualCing, on all 9,000+ NMR projects in the PDB.

Currently there are more than 450 users registered, several of which use the portals on a regular basis. All portals are collecting Grid accounting information (how many runs submitted by a user, how many jobs submitted to the Grid etc...). The most active portals are the ones for HADDOCK and for CS-ROSETTA, which have processed over 3,750 and 775 requests thus far. Together these requests account for over 85 % of the jobs that have been run on the Grid, as a result of the task farming (delegation by a master computer of a large numbers of independent jobs on various computers/clusters/Grid working nodes, followed by a collection and integration of the results by the master) approach involved.

The WeNMR webportals are accessible to any user with a valid Grid certificate. Although the WeNMR project is founded by the European community, anyone with a valid Grid certificate, and attached to a public research institution, can join the enmr.eu Virtual Organization and access our web portal to run their calculation on the Grid. Currently, our user base spans 41 countries, 11 of them outside Europe, representing 27 % of our users (for user statistics and worldwide distribution see <http://wenmr.eu/wenmr/wenmr-Grid-statistics>). This implies that the three Grid sites operated by the WeNMR

partners, and all other Grid sites supporting our VO, are allowing jobs on their facilities independently of the nationality of the users, as long as they have been approved for the use of a specific application portal (most of which using a robot certificate).

4.2 The WeNMR Virtual Research Community (VRC)

WeNMR groups different research teams into a worldwide Virtual Research Community. The WeNMR VRC represents the largest Virtual Organization in the life sciences, accounting for about 33 % of the CPU time on the Grid within the life science area. It has been the first VRC officially recognized by the EGI. To serve its various communities, from the general public to experienced users and softwares developers, WeNMR operates a VRC portal (www.wenmr.eu) that offers a variety of services, information channels and communication tools. User support is provided by documentation pages and tutorials, maintained by the WeNMR partners, as well as Wiki pages and a forum for which any registered user can contribute to. The WeNMR web site also provides a chat and a video conferencing tools. Important events (courses, workshops and conferences) are announced on the web sites, and event reports are available via a blog. Altogether, those community services are aimed to facilitate the communication between all WeNMR members, life sciences or structural biologist experts, e-Science developers, and the general public.

4.3 Collaboration with OSG and SBGrid

WeNMR has been in contact since its beginning with the SBGrid Consortium (<http://sbGrid.org>) which is serving the structural biology user community in the US. SBGrid is also a Virtual Organization of the Open Science Grid (OSG, <http://www.openscienceGrid.org>), the US national distributed computing Grid for data-intensive research, which is based on the Globus middleware and is interoperable with the European EGI Grid. At the beginning of 2011 it has been agreed between WeNMR and SBGrid that

every US user registering with SBGrid is given the option to register with the enmr.eu VO too, allowing them to use the WeNMR services.

Moreover, OSG, SBGrid and WeNMR representatives designed a technical plan aiming at setting up a test-bed for allowing the submission of WeNMR Grid jobs towards the OSG resource centers supporting the SBGrid VO.

The first proof of concept has been achieved through the use of dedicated EMI middleware services components deployed in Europe and configured to submit WeNMR jobs to OSG. A further step has started to enable the job submission from Europe making use of the Condor based GlideinWMS [71] system services deployed in OSG and properly configured to receive WeNMR jobs. A detailed description of the current achievements is available in [72].

The testing phase is expected to last until the end of 2012. The submission and polling daemons developed by WeNMR will be adapted to interact with the GlideinWMS system. After that, if successful, the WeNMR production portals will have the technical possibility to route their workload in US, exploiting potentially a few additional tens of thousands of CPU-cores provided by OSG resource centers and making the extension of WeNMR Grid infrastructure to OSG Grid a reality.

5 Conclusions

Since the beginning of the eNMR project, the eNMR/WeNMR consortium has managed to set up an operational Grid (<http://www.wenmr.eu/wenmr/wenmr-Grid-statistics>), to port over twenty applications and bring up nineteen web portals (<http://www.wenmr.eu/wenmr/nmr-services>), with several others being finalized. At the time of writing, WeNMR has already grown to be the largest virtual organization within the life sciences. This successful start has been underlined by the award for the best demonstration of an application, received at the EGEE (Enabling Grids for E-science) 2009 User Forum. With the present-day momentum, the WeNMR project is rapidly evolving into a factor of importance within structural biology, and life sciences in general. It

has been the first Virtual Research Community officially recognized by the EGI. WeNMR also aims at serving the ESFRI INSTRUMENT structural biology community and as such has been recognized by INSTRUMENT (<http://www.structuralbiology.eu/support/funding/wenmr-facilities>).

Currently, efforts include writing WSDL (Web Service Definition Language) definitions for the portals that will allow calling services remotely, e.g. from a workflow-manager. At the next stage of the project the different elements will be combined, providing comprehensive, yet easy-to-use tools for integrated analysis of NMR data. Incorporation of SAXS services has already begun with the first pilot project offering an ab initio shape determination, and a number of other SAXS methods are presently being added to support a wider user community in structural biology. They also include an innovative Grid-based computational approach that integrates NMR and SAXS data which has recently been developed by some of us [73].

Up-to-date information, regarding the state of the project, the available services, and how to join WeNMR and its virtual organization [enmr.eu](http://www.wenmr.eu), can be found on the project web page at <http://www.wenmr.eu>.

Acknowledgements The WeNMR project is funded by the European Commission under an FP7 e-Infrastructure grant, contract no. 261572 and builds on the previous FP7 e-Infrastructure project e-NMR, contract no. 213010. Support from the former EGEE and the current EGI in terms of expertise and recognition is also acknowledged. The national Grid Initiatives of Belgium, France, Germany, Italy, the Latin America Grid infrastructure via the GISELA project, the Netherlands (via the Dutch Big Grid project), Portugal, Spain, South Africa, Taiwan and UK is acknowledged for the use of web portals, computing and storage facilities. Finally, the authors like to thank those that have expressed their interest in and support to the project.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Bloch, F.: Nuclear induction. *Phys. Rev.* **70**, 460 (1946)
- Purcell, E.M., Torrey, H.C., Pound, R.V.: Resonance absorption by nuclear magnetic moments in a solid. *Phys. Rev.* **69**, 37 (1946)
- Bonvin, A.M.J.J., Rosato, A., Wassenaar, T.A.: The eNMR platform for structural biology. *J. Struct. Funct. Genomics* **11**, 1–8 (2010)
- Feigin, L.A., Svergun, D.I.: *Structure Analysis by Small-angle X-ray and Neutron Scattering*. Plenum Press, New York (1987)
- Mertens, H.D.T., Svergun, D.I.: Structural characterization of proteins and complexes using small-angle X-ray solution scattering. *J. Struct. Biol.* **172**, 128–141 (2010)
- Morgan, H.P., et al.: Structural basis for engagement by complement factor H of C3b on a self surface. *Nat. Struct. Mol. Biol.* **18**, 463–470 (2011)
- Prischi, F., et al.: Structural bases for the interaction of frataxin with the central components of iron-sulphur cluster assembly. *Nat. Commun.* **1**, 95 (2010)
- Gabel, F., et al.: A structure refinement protocol combining NMR residual dipolar couplings and small angle scattering restraints. *J. Biomol. NMR* **41**, 199–208 (2008)
- Grishaev, A., Tugarinov, V., Kay, L.E., Trehella, J., Bax, A.: Refined solution structure of the 82-kDa enzyme malate synthase G from joint NMR and synchrotron SAXS restraints. *J. Biomol. NMR* **40**, 95–106 (2008)
- Grishaev, A., Wu, J., Trehella, J., Bax, A.: Refinement of multidomain protein structures by combination of solution small-angle X-ray scattering and NMR data. *J. Am. Chem. Soc.* **127**, 16621–16628 (2005)
- Bernado, P., Svergun, D.I.: Structural insights into intrinsically disordered proteins by small-angle X-ray scattering. In: Uversky, V.N., Longhi, S. (eds.) *Instrumental Analysis of Intrinsically Disordered Proteins: Assessing Structure and Conformation*, pp. 451–476 Wiley, Hoboken (2010)
- Ferrari, T., Gaido, L.: Resources and services of the EGEE production infrastructure. *J. Grid Computing* **9**, 119–133 (2011)
- Avellino, G., et al.: The dataGrid workload management system: challenges and results. *J. Grid Computing* **2**, 353–367 (2004)
- Wilkins-Diehr, N.: Special issue: science gateways—common community interfaces to Grid resources: editorials. *Concurr. Comput.: Pract. Exp.* **19**, 743–749 (2007)
- Fargette, M., Barbera, R., Rotondo, R.: A simplified access to Grid resources by science gateways. In: *International Symposium on Grids and Clouds and the Open Grid Forum*. Taipei (2011)
- Farkas, Z., Kacsuk, P.: P-GRADE portal: a generic workflow system to support user communities. *Future Gener. Comput. Syst.* **27**, 454–465 (2011)
- Kacsuk, P.: P-GRADE portal family for Grid infrastructures. *Concurr. Comput.: Pract. Exp.* **23**, 235–245 (2011)
- Barbera, R., Falzone, A., Ardizzone, V., Scardaci, D.: The GENIUS Grid portal: its architecture,

- improvements of features, and new implementations about authentication and authorization. In: Proceedings of the 16th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises, pp. 279–283. IEEE Computer Society, (2007)
19. Zhang, C., Kelley, I., Allen, G.: Grid portal solutions: a comparison of GridPortlets and OGCE. *Concurr. Comput.: Pract. Exp.* **19**, 1739–1748 (2007)
 20. Szejnfeld, D., et al.: Vine toolkit—towards portal based production solutions for scientific and engineering communities with Grid-enabled resources support. *Scalable Comput.: Pract. Exp.* **11**, 161–172 (2010)
 21. Hull, D., et al.: Taverna: a tool for building and running workflows of services. *Nucleic Acids Res.* **34**, W729–W732 (2006)
 22. De Vries, S.J., et al.: HADDOCK versus HADDOCK: New features and performance of HADDOCK2.0 on the CAPRI targets. *Proteins* **69**, 726–733 (2007)
 23. Dominguez, C., Boelens, R., Bonvin, A.M.J.J.: HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **125**, 1731–1737 (2003)
 24. Schwieters, C.D., Kuszewski, J.J., Tjandra, N., Clore, G.M.: The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* **160**, 65–73 (2003)
 25. Güntert, P., Mumenthaler, C., Wüthrich, K.: Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **273**, 283–298 (1997)
 26. Herrmann, T., Güntert, P., Wüthrich, K.: Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.* **319**, 209–227 (2002)
 27. Shen, Y., et al.: Consistent blind protein structure generation from NMR chemical shift data. *Proc. Natl. Acad. Sci. USA.* **105**, 4685–4690 (2008)
 28. Shen, Y., Vernon, R., Baker, D., Bax, A.: De novo protein structure generation from incomplete chemical shift assignments. *J. Biomol. NMR* **43**, 63–78 (2009)
 29. Case, D.A., et al.: The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668–1688 (2005)
 30. Van Der Spoel, D., et al.: GROMACS: fast, flexible, and free. *J. Comput. Chem.* **26**, 1701–1718 (2005)
 31. Vranken, W.F., et al.: The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**, 687–696 (2005)
 32. Jung, Y.S., Zweckstetter, M.: Mars—robust automatic backbone assignment of proteins. *J. Biomol. NMR* **30**, 11–23 (2004)
 33. Shen, Y., Delaglio, F., Cornilescu, G., Bax, A.: TALOS plus: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* **44**, 213–223 (2009)
 34. Jaravine, V.A., Zhuravleva, A.V., Permi, P., Ibraghimov, I., Orekhov, V.Y.: Hyperdimensional NMR spectroscopy with nonlinear sampling. *J. Am. Chem. Soc.* **130**, 3927–3936 (2008)
 35. Petoukhov, M.V., Konarev, P.V., Kikhney, A.G., Svergun, D.I.: ATSAS 2.1—towards automated and web-supported small-angle scattering data analysis. *J. Appl. Crystallogr.* **40**, S223–S228 (2007)
 36. Fiorito, F., Herrmann, T., Damberger, F.F., Wüthrich, K.: Automated amino acid side-chain NMR assignment of proteins using ^{13}C - and ^{15}N -resolved 3D [^1H , ^1H]-NOESY. *J. Biomol. NMR* **42**, 23–33 (2008)
 37. Volk, J., Herrmann, T., Wüthrich, K.: Automated sequence-specific protein NMR assignment using the memetic algorithm MATCH. *J. Biomol. NMR* **41**, 127–138 (2008)
 38. Lensink, M.F., Mendez, R., Wodak, S.J.: Docking and scoring protein complexes: CAPRI 3rd edition. *Proteins* **69**, 704–718 (2007)
 39. Mendez, R., Leplae, R., Lensink, M.F., Wodak, S.J.: Assessment of CAPRI predictions in rounds 3–5 shows progress in docking procedures. *Proteins* **60**, 150–169 (2005)
 40. van Dijk, A.D., et al.: Data-driven docking: HADDOCK's adventures in CAPRI. *Proteins* **60**, 232–238 (2005)
 41. Schwieters, C.D., Kuszewski, J.J., Clore, G.M.: Using Xplor-NIH for NMR molecular structure determination. *Prog. NMR Spectrosc.* **48**, 47–62 (2006)
 42. Berman, H.M., et al.: The protein data bank. *Nucleic Acids Res.* **28**, 235–242 (2000)
 43. Rohl, C.A., Strauss, C.E., Misura, K.M., Baker, D.: Protein structure prediction using Rosetta. *Methods Enzymol.* **383**, 66–93 (2004)
 44. Shen, Y., Bax, A.: Protein backbone chemical shifts predicted from searching a database for torsion angle and sequence homology. *J. Biomol. NMR* **38**, 289–302 (2007)
 45. Cornilescu, G., Delaglio, F., Bax, A.: Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* **13**, 289–302 (1999)
 46. McGuffin, L.J., Bryson, K., Jones, D.T.: The PSIPRED protein structure prediction server. *Bioinformatics* **16**, 404–405 (2000)
 47. Wang, Y.J., Jardetzky, O.: Investigation of the neighboring residue effects on protein chemical shifts. *J. Am. Chem. Soc.* **124**, 14075–14084 (2002)
 48. Berjanskii, M.V., Wishart, D.S.: A simple method to predict protein flexibility using secondary chemical shifts. *J. Am. Chem. Soc.* **127**, 14970–14971 (2005)
 49. Ulrich, E.L., et al.: BioMagResBank. *Nucleic Acids Res.* **36**, D402–D408 (2008)
 50. Fogh, R.H., et al.: MEMOPS: data modelling and automatic code generation. *J. Integr. Bioinform.* **7**, 123 (2010)
 51. Bertini, I., Case, D.A., Ferella, L., Giachetti, A., Rosato, A.: A Grid-enabled web portal for NMR structure refinement with AMBER. *Bioinformatics* **27**, 2384–2390 (2011)
 52. Lindahl, E., Hess, B., van der Spoel, D.: GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model.* **7**, 306–317 (2001)
 53. Oostenbrink, C., Villa, A., Mark, A.E., Van Gunsteren, W.F.: A biomolecular force field based on the free

- enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* **25**, 1656–1676 (2004)
54. Schuler, L.D., Daura, X., Van Gunsteren, W.F.: An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. *J. Comput. Chem.* **22**, 1205–1218 (2001)
55. Scott, W.R.P., et al.: The GROMOS biomolecular simulation program package. *J. Phys. Chem. A* **103**, 3596–3607 (1999)
56. van Gunsteren, W.F., et al.: *Biomolecular Simulation: The Gromos 96 Manual and User Guide*. BIOMOS BV, Zürich, Groningen (1996)
57. Ponder, J.W., Case, D.A.: Force fields for protein simulations. *Adv. Protein Chem.* **66**, 27–85 (2003)
58. Brooks, B.R., et al.: Charmm—a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187–217 (1983)
59. Jorgensen, W.L., Maxwell, D.S., TiradoRives, J.: Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **118**, 11225–11236 (1996)
60. Van Dijk, M., Wassenaar, T., Bonvin, A.M.J.J.: A flexible, Grid-enabled web portal for GROMACS molecular dynamics simulations. *J. Chem. Theory Comput.* (2012). doi:10.1021/ct300102d
61. Villa, A., Fan, H., Wassenaar, T., Mark, A.E.: How sensitive are nanosecond molecular dynamics simulations of proteins to changes in the force field? *J. Phys. Chem. B* **111**, 6015–6025 (2007)
62. Wassenaar, T.A., Mark, A.E.: The effect of box shape on the dynamic properties of proteins simulated under periodic boundary conditions. *J. Comput. Chem.* **27**, 316–325 (2006)
63. De Vries, S.J., van Dijk, M., Bonvin, A.M.J.J.: The HADDOCK web server for data-driven biomolecular docking. *Nat. Protoc.* **5**, 883–897 (2010)
64. Herrmann, T., Güntert, P., Wüthrich, K.: Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. *J. Biomol. NMR* **24**, 171–189 (2002)
65. Doreleijers, J.F., Sousa da Silva, A.W., Krieger, E., Nabuurs, S.B., Spronk, C.A.E.M., Stevens, T.J., Vranken, W.F., Vriend, G., Vuister, G.W.: CING: an integrated residue-based structure validation program suite. *J. Biomol. NMR* **54**, 267–283 (2012)
66. Doreleijers, J.F., et al.: NRG-CING: integrated validation reports of remediated experimental biomolecular NMR data and coordinates in wwPDB. *Nucleic Acids Res.* (2011). doi:10.1093/nar/gkr1134
67. Franke, D., Svergun, D.I.: DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* **42**, 342–346 (2009)
68. Volkov, V.V., Svergun, D.I.: Uniqueness of ab initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* **36**, 860–864 (2003)
69. Svergun, D.I.: Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76**, 2879–2886 (1999)
70. Svergun, D.I., Petoukhov, M.V., Koch, M.H.: Determination of domain structure of proteins from X-ray solution scattering. *Biophys. J.* **80**, 2946–2953 (2001)
71. Bradley, D., Sfiligoi, I., Padhi, S., Frey, J., Tannenbaum, T.: Scalability and interoperability within glideinWMS. *J. Phys.: Conf. Ser.* **219**, 062036 (2010)
72. Verlato, M.: Extending WeNMR e-Infrastructure outside Europe. In: EGI Community Forum 2012/EMI Second Technical Conference, Munich (2012)
73. Bertini, I., et al.: Conformational space of flexible biological macromolecules from average data. *J. Am. Chem. Soc.* **132**, 13553–13558 (2010)