

# A Grid-enabled web portal for NMR structure refinement with AMBER

Ivano Bertini<sup>1,2,\*</sup>, David A. Case<sup>3,4</sup>, Lucio Ferella<sup>1</sup>, Andrea Giachetti<sup>1</sup>  
and Antonio Rosato<sup>1,2</sup>

<sup>1</sup>Magnetic Resonance Center (CERM), University of Florence, Via L. Sacconi 6, <sup>2</sup>Department of Chemistry, University of Florence, Via della Lastruccia 3, 50019 Sesto Fiorentino, Italy, <sup>3</sup>BioMaPS Institute and <sup>4</sup>Department of Chemistry and Chemical Biology, Rutgers University, 610 Taylor Rd., Piscataway NJ 08854-8087, USA

Associate Editor: Martin Bishop

## ABSTRACT

**Motivation:** The typical workflow for NMR structure determination involves collecting thousands of conformational restraints, calculating a bundle of 20–40 conformers in agreement with them and refining the energetics of these conformers. The structure calculation step employs simulated annealing based on molecular dynamics (MD) simulations with very simplified force fields. The value of refining the calculated conformers using restrained MD (rMD) simulations with state-of-art force fields is documented. This refinement however presents various subtleties, from the proper formatting of conformational restraints to the definition of suitable protocols.

**Results:** We describe a web interface to set up and run calculations with the AMBER package, which we called AMPS-NMR (AMBER-based Portal Server for NMR structures). The interface allows the refinement of NMR structures through rMD. Some predefined protocols are provided for this purpose, which can be personalized; it is also possible to create an entirely new protocol. AMPS-NMR can handle various restraint types. Standard rMD refinement in explicit water of the structures of three different proteins are shown as examples. AMPS-NMR additionally includes a workspace for the user to store different calculations. As an ancillary service, a web interface to AnteChamber is available, enabling the calculation of force field parameters for organic molecules such as ligands in protein–ligand adducts.

**Availability and Implementation:** AMPS-NMR is embedded within the NMR services of the WeNMR project and is available at <http://py-enmr.cerm.unifi.it/access/index/amps-nmr>; its use requires registration with a digital certificate.

**Contact:** [ivanobertini@cerm.unifi.it](mailto:ivanobertini@cerm.unifi.it)

**Supplementary information:** Supplementary data are available at *Bioinformatics* online.

Received on May 18, 2011; revised on July 4, 2011; accepted on July 6, 2011

## 1 INTRODUCTION

Since the 1980's NMR spectroscopy has been used as a technique complementary to X-ray to determine the solution structure of proteins (Cavanagh *et al.*, 1996; Wüthrich 1986). Indeed, NMR has

contributed ~12% of the total content of protein structures in the PDB (Berman *et al.*, 2000), a share that has remained relatively steady in the last few years. The usefulness of applying NMR as a tool complementary to X-ray in protein structure determination efforts has been extensively documented in the context of Structural Genomics initiatives (Synder *et al.*, 2005; Yee *et al.*, 2005). A unique aspect of NMR data is their capability to provide a view of protein structural features that includes unfolded proteins and the residue-specific characterization of disordered protein segments, including functionally important flexible surface loops (Bernadó *et al.*, 2005; Mittermaier and Kay, 2009).

Standard protocols to determine protein structures by NMR involve three basic steps: (i) determining the chemical shift assignments of the target protein; (ii) measuring as many as possible intra- and, particularly (Doreleijers *et al.*, 1999), inter-residue NOEs (nuclear Overhauser enhancements) among <sup>1</sup>H nuclei to generate proton–proton distance restraints (typically in the form of upper distance bounds); and (iii) using the NOE-derived restraints to perform simulated annealing calculations providing the 3D structure of the protein. Additional restraints can be used, such as dihedral angle restraints [which often are derived from J-couplings (Karplus, 1959) or from chemical shifts (Shen *et al.*, 2009)], residual dipolar couplings (rdc's) (Tjandra *et al.*, 1997), and paramagnetism-based restraints (Bertini *et al.*, 1996). More recently, alternative protocols that are based on the use of chemical shift data (Shen *et al.*, 2008) or chemical shift and rdc's (Raman *et al.*, 2010) without any NOE information have been described. However, the adoption of these innovative protocols by the NMR community is still limited. Within standard protocols, an additional step that may have a dramatic impact on the quality of the final NMR structure is structure refinement, which typically involves the application of molecular dynamics (MD) simulations using a force field more accurate than that employed to generate the structural models to be refined. Indeed, it has long been realized that, due to the sparseness of NMR restraints, the force field used for refinement can have a large impact on the quality, and possibly the accuracy, of NMR structures (Chen *et al.*, 2004; Linge and Nilges 1999; Linge *et al.*, 2003). Refinement of protein structures using state-of-the-art force fields in MD calculations with explicit solvent molecules can afford a significant improvement of various stereochemical parameters that are often used as indicators of structural quality and can also have a positive impact on the accuracy of the structure, as measured from the agreement with the NMR constraints (Nabuurs *et al.*,

\*To whom correspondence should be addressed.

2004; Nederveen *et al.*, 2005). On the other hand, these very same refinement procedures may mask errors in the NMR structures, at least as far as the normality of stereochemical parameters is concerned (Nabuurs *et al.*, 2006; Saccenti and Rosato, 2008).

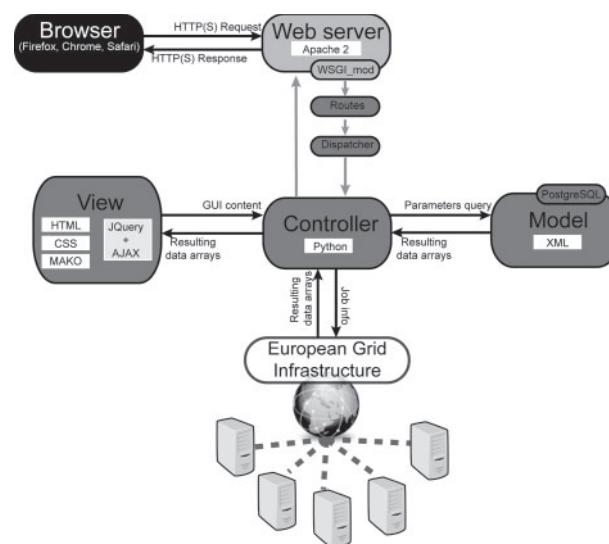
Despite the last caveat, there is large consensus in the bio-NMR community that structure refinement in explicit solvent should be included in standard protocols for the solution structure determination of proteins. To this end, one can use specialized MD programs, such as CNS (Brunger, 2007) or OPAL (Luginbuhl *et al.*, 1996) (which is no longer distributed), or more general packages such as AMBER (Case *et al.*, 2005), CHARMM (Brooks *et al.*, 2009) or GROMACS (Van Der *et al.*, 2005). It is important to note that not all of these programs comprise their own force field (i.e. a third-party force field may be needed) and that often a program can be used with more than one force field. As a complementary approach to refinement in explicit solvent, the use of implicit solvent models may also be warranted. For example, the AMBER package allows the use of the Generalized Born model for this purpose (Xia *et al.*, 2002); this model might perform even better than explicit solvent where large amounts of conformational space needs to be explored. General MD programs rely on conversion tools to generate the correct input data from the restraints used in the upstream structure calculation program [such as Xplor-NIH (Schwieters *et al.*, 2003), CNS (Brunger, 2007) or CYANA (Herrmann *et al.*, 2002)], whose success may sometimes be tricky e.g. because of different atom naming rules used by different programs. This creates a barrier to the general use of these programs, thereby effectively restricting the array of options available to bio-NMR spectroscopists. Another general aspect is that the refinement strategies to be implemented require a wall time for calculations that is significantly longer than that required for the calculation of the structural models to be refined (hours versus minutes). To address these two points, we developed a web interface that enables a very easy and quick set up of structure refinement from the restraint data used in common structural calculation programs and distributes the calculation over a computational Grid infrastructure comprising several thousands CPUs. This resource, which we called AMPS-NMR (AMBER-based Portal Server for NMR structures), is free for academia and non-profit users; access to the server requires prior registration with the eNMR.eu Virtual Organization (Bonvin *et al.*, 2010; Loureiro-Ferreira *et al.* 2010).

## 2 METHODS

### 2.1 AMPS-NMR

To support the development of AMPS-NMR we have decided to use a Web Application Framework, Pylons 1.0 (<http://pylonshq.com/>), a lightweight web framework that emphasizes flexibility and re-use (one of the first projects to leverage the emerging WSGI standard, <http://wsgi.org/wsgi/>) as well as rapid development. As with many frameworks, Pylons follow the Model View Controller (MVC) (Gamma *et al.*, 1994) architectural pattern to separate the data model with business rules from user interfaces (Fig. 1). This is generally considered a good practice as it modularizes code, promotes code reuse and allows multiple interfaces to be applied.

AMPS-NMR is based on an ad hoc mixture of HTML (<http://www.w3.org/>), CSS (<http://www.w3.org/>) and Mako (<http://www.makotemplates.org/>) (Fig. 1). The latter is the default template language for the Pylons web framework, a totally Python-centric system that is suited to the current generation of WSGI frameworks, and that offer features that are not found in any other template language. The web



**Fig. 1.** Architectural schema of the AMPS-NMR web portal. This representation outlines the different components (rounded rectangular boxes) and messages (solid black arrows with text) that constitute the software system of AMPS-NMR. All interactions with the user take place via a web browser (rounded black box) interacting with an Apache 2 Web Server (rounded light gray box). This is the only part of the process visible to the user. The Web Server exchanges messages with the Pylons Framework (rounded dark gray boxes), and in particular with the Controller component. The latter modifies the status of the Model and View components, which, in turn, signal back to the Web Server via the Controller component to send new contents to the web browser for user visualization. The software technologies used are indicated in the white rectangular boxes.

page's interactivity, speed and usability have been increased through Ajax, shorthand for Asynchronous JavaScript, XML (<http://www.w3.org/XML>) and JQuery (<http://jquery.com/>), which is a cross-browser JavaScript library designed to simplify the client-side scripting of HTML; both are web development techniques for creating interactive web applications. The intent is to make web pages more responsive by exchanging small amounts of data with the server behind the scenes, so that the entire web page does not have to be reloaded each time the user requests a change. With this approach we made AMPS-NMR easy to use for an end-user, without requiring any specialized training.

AMPS-NMR enables the users to create registered accounts through their own Digital Certificate (DC) (as detailed in the text in Supplementary Material), which is strictly required to be able to submit calculations on the Grid Infrastructure. The DC also allows user data and jobs to be maintained and managed across multiple work sessions. This functionality has implemented through the combined use of the HTTPS authentication protocol and a specifically designed database built on version 8.1 of the PostgreSQL Object-Relational DataBase Management System (<http://www.postgresql.org/>).

Running a calculation in AMBER through AMPS-NMR is a process logically split in four macro steps:

- (1) input data preparation;
- (2) setting MD parameters;
- (3) sending the calculation to the Grid infrastructure; and
- (4) retrieving the output from the Grid Infrastructure.

In Step (1) the user must upload to her/his AMPS-NMR workspace (Supplementary Fig. S1, see also later) a file containing cartesian coordinates

for a macromolecular structure in PDB format, parameters and topology file for any residues not present in the standard AMBER library, NMR restraints files (upper distance limits, dihedral angles, pseudocontact shifts (pcs) and rdc's). To cope with different atom naming rules, custom numbering and the possible presence of pseudo atoms, AMPS-NMR exploits newly developed python-based utilities that automatically convert/adapt all the input data from the native format of commonly used programs or databases (CYANA, XPLOR, CNS, TALOS, PDB), to the AMBER format. This process is completely transparent to the user. For Step (2), AMPS-NMR provides pre-configured sets of MD parameters, defined in standard XML. These range from a simple energy-minimization in water or *in vacuo* to a multi-step protocol for the refinement of an NMR structure. AMPS-NMR additionally allows users to set manually all MD parameters; for this purpose, an interactive guide, extracted from the AMBER manual, to parameters setting is available. In Step (3), the user defines a name for the new calculation. Then, a series of python routines collect all information about the new calculation, build all batch commands necessary to submit it, thereby creating a so-called job, and submit the calculation to the Grid Infrastructure. The submission process involves defining the user identity. Finally, in Step (4), the portal allows the user to check the status of the job (i.e. Scheduled, Running and Completed) and, upon completion, retrieves all output data, making them available through the AMPS-NMR Filemanager (Supplementary Fig. S2).

Finally, another relevant feature of AMPS-NMR is its persistent user workspace, where users' data and jobs are maintained and managed across multiple work sessions (Supplementary Figs S1 and S2). Users can exploit a simple, user-friendly interface to organize their projects and calculations within the workspace. This functionality has been implemented through the combined use of the HTTPS authentication protocol and of the SQLAlchemy (<http://www.sqlalchemy.org/>) Python Database Abstraction Library. SQLAlchemy provides a high-level interface to relational databases such as Oracle, DB2, MySQL, PostgreSQL and SQLite. We have adopted SQLAlchemy because it is unobtrusive to our Python code, allowing us to map plain old Python objects to database tables without substantially changing our existing Python code. Further, SQLAlchemy includes both a database server-independent SQL expression language and an Object-Relational Mapper (ORM) that allows SQL to persist our application objects automatically. The entity relationship diagram, a specialized graphic that illustrates the inter-relationships between entities of the database underlying the user workspace is shown in Supplementary Figure S3.

## 2.2 Antechamber

Antechamber is a set of programs for molecular mechanics (MM) whose purpose is to solve the following problems that are encountered during MM calculations: (i) recognizing the atom type; (ii) recognizing bond type and judging the atomic equivalence; (iii) generating residue topology file; and (iv) finding missing force field parameters and supplying reasonable and similar substitutes. In the context of a combined use with AMBER, Antechamber can automatically generate input force field parameters for most organic molecules, thereby allowing users to use AMPS-NMR to refine the structure of protein-ligand adducts. To facilitate this, we developed a web interface for the use of Antechamber, in the same style of the AMPS-NMR interface (Supplementary Fig. S4).

## 2.3 Examples

As examples for the use of AMPS-NMR we took two structures from structural genomics projects: 2KT0, an all-helical protein and 2LOA, an all-beta protein. Restraints were taken from the BMRB (Ulrich *et al.*, 2007) and used without modifications. They consisted of 889 distances and 382 dihedral angles for 2KT0, and of 1115 distances and 92 dihedral angles for 2LOA. We used the CYANA web server on the WeNMR page to calculate 100 conformers using the standard protocol and retained the 20 conformers in best agreement with the restraints. The output pdb file from CYANA was submitted to AMPS-NMR with all the available restraints, using



Fig. 2. The welcome page of AMPS-NMR.

the 'restrainedMD.xml' protocol. All conformers were kept for analysis. To demonstrate the use of AMPS-NMR with a dataset including also rdc's, we used structure 2KMY, containing 185 amino acids, which was determined in our laboratory. We took the final bundle of the original CYANA calculation and refined it with AMPS-NMR, with 2059 distances, 180 dihedral angles and 84 rdc's, using the 'restrainedMDrdc.xml' protocol. The latter performs an initial optimization of the alignment tensor parameters and then proceeds with a calculation scheme identical to 'restrainedMD.xml'.

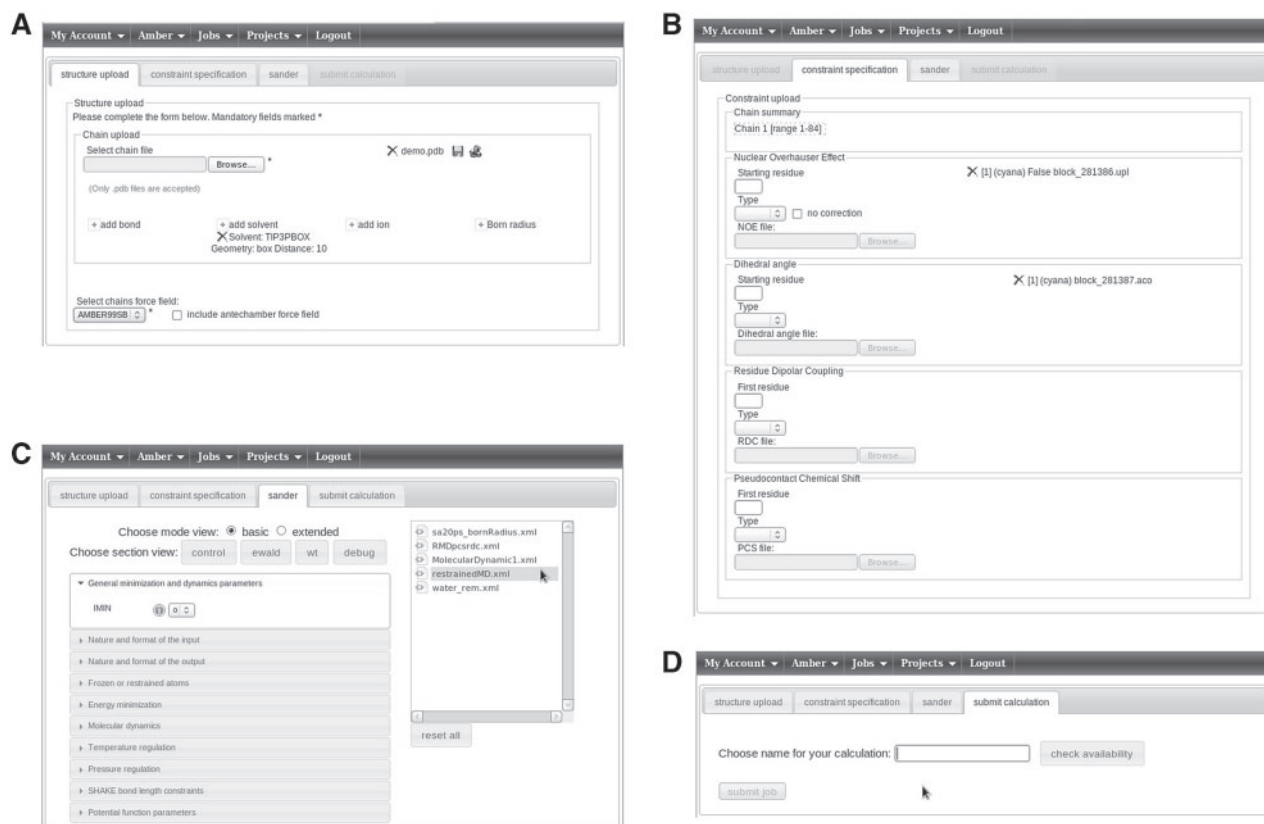
## 3 RESULTS AND DISCUSSION

### 3.1 RMD refinement with AMPS-NMR

The Welcome page to AMPS-NMR is shown in Figure 2. The portal integrates a project manager, which allows the creation of new projects (i.e. new top-level folders in the available disk space) as well as the addition of new calculations to existing projects (corresponding to sub-folders in the project folder, Supplementary Figs S1 and S2). This organization of the data allows users to store all their calculations relevant for an individual structure determination project in a single folder. This can be particularly advantageous when working on the structural characterization of two or more different systems, such as the complexes of a protein with two different ligands, or when trying various calculations schemes.

Alternatively, the 'New calculation' button automatically creates a new folder with a standard name (new\_project) followed by a 10-digit random number; this folder will contain the results of the calculation in a user-defined sub-folder (see next paragraph). The protocol for the refinement of a protein structure through restrained molecular dynamics (Clare *et al.*, 1985; Kaptein, R. *et al.*, 1985) with AMBER comprises three preparation steps and a final submission step (Fig. 3). The first step is the upload of the bundle of conformers (typically 20–40) to be refined. For input to AMPS-NMR, the entire bundle should be contained in a single PDB file, which can be normally produced by the structure calculation program.

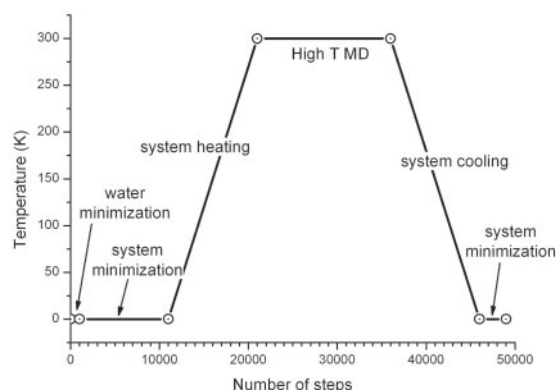




**Fig. 3.** Set-up of a RMD refinement. (A) structure upload and solvation; (B) upload of restraints; (C) selection of refinement protocol; and (D) submission of the calculation to the Grid.

The server will automatically prompt the number of conformers contained in the file and will notify the user if non-standard amino acids or ligands are identified. In the latter case suitable parameter files can be generated using the Antechamber web interface also provided on the website (Supplementary Fig. S4). After uploading the PDB, the user can add bonds, ions or solvent molecules to the system (Fig. 3A). Periodic boundary conditions are applied. Two standard AMBER force fields can be selected, after which the second step becomes available. In this (optional) step, the available NMR-derived restraints can be included in the calculation (Fig. 3B). The accepted restraint types are: upper distance limits (as derived from NOESY spectra or defining hydrogen bonds); dihedral angle restraints; rdc's; and pcs. The two latter types of restraints can be obtained from systems containing a paramagnetic metal ion (Bertini *et al.* 1996; Bertini *et al.*, 2005), or, only for rdc's, by dissolving the protein sample in orienting media, such as dilute liquid crystals (Tjandra *et al.*, 1997). The extent of partial orientation of the protein (for rdc's) or the anisotropy of the magnetic susceptibility of the metal ion (for pcs) can be obtained by fitting to an initial structural model (Banci *et al.*, 1997), using a separate web interface, AnisoFIT, on the WeNMR site (<http://www.wenmr.eu>). After uploading the NMR restraints, the user can proceed to the next step to define the calculation protocol. Skipping the optional second step will result in an unrestrained molecular dynamics simulation being set up. In the third step, the user can choose to create

her/his own calculation protocol by using the panel on the left side of the web frame or can upload a predefined protocol using the 'load protocol' button below the panel on the right side (Fig. 3C). To refine an NMR structure of a protein (or protein complex), we recommend the users to select the predefined 'restrainedMD.xml' protocol. The latter implements a calculation strategy similar to that used for the construction of the DRESS database (Nabuurs *et al.*, 2004), which has been shown to significantly improve the quality of NMR structures in terms of their compliance to standard stereochemical features (Saccenti and Rosato, 2008). The refinement protocol implemented in AMPS-NMR consists of the following steps: (i) minimization of the hydration shell (whose thickness can be defined by the user; a layer of 10 Å is suggested); (ii) minimization of the entire system (protein, water and counterions if present); (iii) heating to 300 K in 10 000 rMD steps; (iv) 30 ps (15 000 steps with a time step of 2.0 fs) of rMD at constant temperature and constant pressure (1.0 bar); (v) cooling to 0 K in 10 000 MD steps; (vi) a final energy minimization (Fig. 4). For calculations including also rdc's, a preliminary minimization with all the atoms frozen is carried out to optimize the alignment tensor parameters ('restrainedMDrdc.xml'). After choosing the protocol, the user can proceed to the submission of the calculation. This is done by selecting the name of the sub-folder in which the results will be stored (Fig. 3D). The chosen refinement protocol is applied to each conformer in the bundle. Thanks to the availability of the

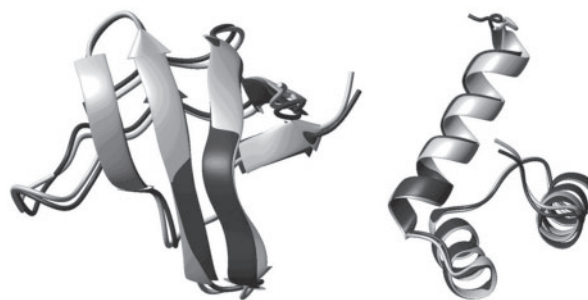


**Fig. 4.** Scheme of the rMD refinement protocol (restrainedMD.xml). The NMR restraints are always active. Output files are produced at the end of each segment.

WeNMR Grid computational infrastructure, the calculations for each conformer in the bundle are run in parallel, thereby minimizing the wall time needed for the overall procedure. Tutorials on how to setup a calculation with example data are provided on the WeNMR site (at [www.wenmr.eu/wenmr/tutorials/nmr-tutorials/amber](http://www.wenmr.eu/wenmr/tutorials/nmr-tutorials/amber)).

The Jobs menu allows users to monitor and update the status of the various jobs. For practical reasons, the Jobs status is not updated automatically; users must thus use the links at the top of the job list for this purpose. When the result of an update is that a job has been completed, the data are automatically retrieved and saved in the user's storage space, within the selected folder structure. In particular, the results for each conformer of the bundle are stored in a separate folder within the corresponding calculation sub-folder (Supplementary Fig. S1). From there they can be downloaded through the Project-> Manage menu, as entire sub-folders (compressed) or as individual files. A PDB together with AMBER output files are generated at the end of each step of the RMD protocol of Figure 4; however a user will typically be interested only in the files produced after the last step, i.e. after the final energy minimization. To this end, a button allows users to extract the bundle of refined conformers from a given calculation; energy and violation statistics can be obtained similarly (Supplementary Fig. S2). Calculations are typically uploaded to the grid in a few minutes after the 'Submit' button has been pressed. At this point they will appear in the S status in the Jobs list and will be queuing at one of the grid sites. We have implemented a mechanism that allows jobs to be submitted only to sites with a waiting time shorter than 12 h. Among these sites, the one with the largest number of available CPU's is selected. The average queue time is < 1 h. When calculations are running, their status in the Jobs list becomes R. For the examples described in the next paragraphs, the run time was 4–6 h. Thus, from submission to completion the RMD refinement of a protein structure normally takes between 6 and 18 h.

The results of two example calculations run with the standard strategy are shown in Figure 5. The re-calculated structures are essentially indistinguishable, in terms of coordinate deviation, from the deposited ones. Indeed, for both structures the backbone RMSD deviation (over ordered residues, Table 1) between the two mean conformers is less than the sum of the intra-bundle RMSD values (0.7–0.8 Å versus 1.1–1.7 Å). Thus, there are no significant global



**Fig. 5.** Recalculation and refinement of two protein structures (left: 2LOA, right 2KT0). The mean refined structure (dark gray) is superimposed to the mean deposited structure (light gray). This figure was prepared with MOLMOL (Koradi *et al.*, 1996).

rearrangements. This is also in agreement with the fact that restraint violations are only marginally affected by the refinement procedure (Table 1). Further insights on the performance/usefulness of AMPS-NMR can be obtained by looking at stereochemical parameters. The Z-scores as well as the residue distribution in the Ramachandran plot are for the AMBER-refined bundle always very similar to or better than both the deposited structure and the CYANA structure, with only a couple of non-significant exceptions. In some cases, significant improvements of the Z-scores are observed (highlighted in Table 1). This is particularly true when comparing to the CYANA structure, which is not unexpected due to the simplified force field used by CYANA (Nederveen *et al.*, 2005). The beneficial effects of refinement with AMPS-NMR are observed in particular for the MolProbity clashscore (Davis *et al.*, 2007) and the Procheck-all Z-score. In the latter case, the major contribution is probably due to improved side chain packing after refinement (Saccetti and Rosato, 2008). The same trends are observed also for 2KMV, which additionally includes rdc data.

### 3.2 Long MD simulations with AMPS-NMR

The most common application of software packages such as AMBER is presumably the simulation of molecular motions in the absence of experimental restraints. Given the typical computational power of the majority of the computing nodes of the current Grid infrastructure, simulations of this kind for medium-sized proteins may easily require wall times up to several weeks to reach the range of simulated time scales that are currently published in the literature. On the other hand, current Grid management policies stop calculations that take longer than a predefined wall time (of the order of one week). To allow long unrestrained MD simulations to be easily run on the Grid, our interface to set up AMBER simulations that was described in the preceding section (typically without adding any NMR restraint file) can be combined with a web interface that allows users to rapidly restart the calculations after their interruption by the Grid middleware. The latter interface implements a three-step procedure (Supplementary Fig. S5) to: (i) select a calculation present in the Projects catalog, where the files containing the data (coordinates and velocities) necessary to restart the AMBER simulation are located (the files will be automatically identified); (ii) select (and, if necessary, modify) a saved set of AMBER parameters to continue the simulation (from a list of either

**Table 1.** Comparison of various parameters characterizing the protein structures taken as examples for the use of AMPS-NMR

	PDB	CYANA	AMPS-NMR
<b>2KT0<sup>a</sup></b>			
Ordered residues <sup>b</sup>	27–78	27–78	27–78
BB RMSD to the mean (Å)	0.8 ± 0.2	0.8 ± 0.3	0.9 ± 0.2
Ramachandran plot most favored (%) <sup>c</sup>	96.1	92.3	97.7
Verify3D (Z-score)	−4.2 ± 0.5	−4.7 ± 0.6	−4.0 ± 0.7
Procheck $\phi$ – $\psi$ (Z-score)	2.0 ± 0.1	1.0 ± 0.3	<b>2.1 ± 0.2</b>
Procheck all (Z-score)	0.8 ± 0.3	−2.8 ± 0.4	<b>0.4 ± 0.2</b>
MolProbity Clashscore (Z-score)	−1.5 ± 0.8	<b>−0.1 ± 0.5</b>	<b>1.5 ± 0.1</b>
Distance violations >0.2 Å	NA	2 ± 1	2 ± 1
Angle violations > 10°	NA	0 ± 0	0 ± 0
<b>2L0A<sup>a</sup></b>			
Ordered residues <sup>b</sup>	18–71	17–31, 33–71	17–71
BB RMSD to the mean (Å)	0.5 ± 0.1	0.5 ± 0.1	0.6 ± 0.2
Ramachandran plot most favored (%) <sup>c</sup>	86.7	84.7	87.4
Verify3D (Z-score)	−3.4 ± 0.4	−2.9 ± 0.5	<b>−1.4 ± 0.6</b>
Procheck $\phi$ – $\psi$ (Z-score)	−2.2 ± 0.3	−2.3 ± 0.3	−1.8 ± 0.4
Procheck all (Z-score)	−3.0 ± 0.4	−4.7 ± 0.5	<b>−2.7 ± 0.3</b>
MolProbity Clashscore (Z-score)	−1.1 ± 0.4	<b>0.9 ± 0.5</b>	<b>1.5 ± 0.1</b>
Distance violations >0.2 Å	NA	0 ± 0	0 ± 0
Angle violations > 10°	NA	0 ± 0	1 ± 1
<b>2KMV<sup>a</sup></b>			
Ordered residues <sup>b</sup>	11–78, 131–184	11–78, 131–184	11–78, 131–184
BB RMSD to the mean (Å)	0.9 ± 0.2	0.9 ± 0.2	1.1 ± 0.3
Ramachandran plot most favored (%) <sup>c</sup>	87.4	84.9	90.6
Verify3D (Z-score)	−1.8 ± 0.4	−2.7 ± 0.5	−2.2 ± 0.4
Procheck $\phi$ – $\psi$ (Z-score)	−0.4 ± 0.2	−1.8 ± 0.2	<b>−0.5 ± 0.2</b>
Procheck all (Z-score)	−2.1 ± 0.4	−4.6 ± 0.6	<b>−1.7 ± 0.4</b>
MolProbity Clashscore (Z-score)	1.3 ± 0.1	−1.7 ± 0.7	<b>1.4 ± 0.1</b>
Distance violations >0.2 Å	NA	1 ± 1	2 ± 1
Angle violations > 10°	NA	1 ± 1	0 ± 0
Rdc deviation (Hz)	NA	0.8 ± 0.6	0.9 ± 0.7

A positive Z-score indicates that the parameter is better than the average value observed in the PSVS reference dataset (composed by high-resolution X-ray structures). The parameters in bold in the 'CYANA' column are at least one sigma value better than in the PDB structure. The parameters in bold in the 'AMPS-NMR' column are at least one sigma value better than in either the PDB or the CYANA structures. Standard deviations calculated over the bundle are also given. PDB, deposited structure; CYANA, structure recalculated from the original restraints using the CYANA server; AMPS-NMR, refined CYANA structure; NA, not applicable.

<sup>a</sup>All parameters refer only to ordered residues.

<sup>b</sup>As defined by PSVS (Bhattacharya *et al.*, 2007).

<sup>c</sup>Cumulative, calculated by PSVS using Procheck-NMR (Laskowski *et al.*, 1993).

standard or personal protocols); (iii) restart the simulations, which will be stored in a new user-defined folder. This procedure can be repeated several times in order to reach the desired total length of the simulation. The user will then need to use the Project manager to download locally all the parts of the trajectory, which can be input to standard software tools for its visualization and/or analysis

[e.g. VMD (Humphrey *et al.*, 1996) or Chimera (Pettersen *et al.*, 2004)].

## 4 CONCLUDING REMARKS

We have set up a publicly available, free of charge web portal (AMPS-NMR) facilitating the use of the AMBER package. The service exploits a Grid computational infrastructure in order to guarantee that there is always adequate CPU availability, provided by the European Grid Initiative, for calculations. In addition, the portal integrates a workspace where users can store their data and organize their projects and calculations. This service has been set up initially within the e-NMR collaborative project (<http://www.e-nmr.eu>) and it is now maintained in the context of the WeNMR initiative (<http://www.wenmr.eu>), which will last until 31 October 2013.

The main intended use of the portal is the refinement of NMR structures. For this purpose a predefined multi-step protocol (Fig. 4) is available, effectively removing any need for the user to know in detail the subtleties of the many parameters needed to set up a relatively complex MD calculation. The Grid infrastructure easily allows each conformer of the NMR bundle to be refined separately, yielding a significant saving in time. The portal automatically handles the native data formats of the most popular structure calculation programs, making the four-step procedure to start a calculation seamless from the user's point of view. The NMR structure refinement protocol provides appreciable improvements in various aspects of NMR structures, as indicated by different quality parameters (Table 1 and Fig. 5). Furthermore, the web interface can be used also to perform long unrestrained MD simulations. To comply with Grid policies, these simulations are split into shorter segments, which the user must manually restart via a specific web page.

It is important to underline that even though the present article and the current implementation of AMPS-NMR mainly focus on the refinement of protein solution structures, the excellent performance of AMBER in the refinement of the structure of nucleic acids is well documented (Tolbert *et al.*, 2010). Actually, the use of AMBER could be even more useful for these systems than for proteins due to the greater importance of electrostatics. The extension of AMPS-NMR to specifically address also nucleic acid structures is already underway.

AMPS-NMR has been available to users external to the WeNMR project since July 2010. At present, there are 42 registered users. Approximately 700 jobs have been already submitted (including test jobs), consuming a total of more than 10 000 CPU hours.

## ACKNOWLEDGEMENTS

Enrico Morelli is thanked for technical help and many useful suggestions. The national GRID Initiatives of Belgium, Italy, Germany, the Netherlands (via the Dutch BiG Grid project), Portugal, UK, South Africa and the Latin America GRID infrastructure via the Gisela project are acknowledged for the use of computing and storage facilities.

*Funding:* European Commission (contracts no. 213010 and 261572).

*Conflict of Interest:* none declared.

## REFERENCES

- Banci, L. et al. (1997) Pseudocontact shifts as constraints for energy minimization and molecular dynamic calculations on solution structures of paramagnetic metalloproteins. *Proteins Struct. Funct. Genet.*, **29**, 68–76.
- Berman, H.M. et al. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
- Bernadó, P. et al. (2005) A structural model for unfolded proteins from residual dipolar couplings and small-angle X-ray scattering. *Proc. Natl Acad. Sci. USA*, **102**, 17002–17007.
- Bertini, I. et al. (1996) The solution structure of paramagnetic metalloproteins. *Progr. Biophys. Mol. Biol.*, **66**, 43–80.
- Bertini, I. et al. (2005) NMR of paramagnetic metalloproteins. *ChemBioChem*, **6**, 1536–1549.
- Bhattacharya, A. et al. (2007) Evaluating protein structures determined by structural genomics consortia. *Proteins*, **66**, 778–795.
- Bonvin, A.M.J.J. et al. (2010) The eNMR platform for structural biology. *J. Struct. Funct. Genomics*, **11**, 1–8.
- Brooks, B.R. et al. (2009) CHARMM: the biomolecular simulation program. *J. Comput. Chem.*, **30**, 1545–1614.
- Brunger, A.T. (2007) Version 1.2 of the Crystallography and NMR system. *Nat. Protoc.*, **2**, 2728–2733.
- Case, D.A. et al. (2005) The Amber biomolecular simulation programs. *J. Comput. Chem.*, **26**, 1668–1688.
- Cavanagh, J. et al. (1996). *Protein NMR Spectroscopy. Principles and Practice*. San Diego, Academic Press.
- Chen, J. et al. (2004) Refinement of NMR structures using implicit solvent and advanced sampling techniques. *J. Am. Chem. Soc.*, **126**, 16038–16047.
- Clare, G.M. et al. (1985) The solution conformation of a heptadecapeptide comprising the DNA binding helix F of the cyclic AMP receptor protein of *Escherichia Coli*: Combined use of <sup>1</sup>H-nuclear magnetic resonance and restrained molecular dynamics. *J. Mol. Biol.*, **186**, 435–455.
- Davis, I.W. et al. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.*, **35**, W375–W383.
- Doreleijers, J.F. et al. (1999) Completeness of NOEs in protein structure: a statistical analysis of NMR. *J. Biomol. NMR*, **14**, 123–132.
- Gamma, E. et al. (1994). *Design Patterns: Elements of Reusable Object-Oriented Software*, 36th edn. Boston, Addison-Wesley, p. 1.
- Herrmann, T. et al. (2002) 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.
- Humphrey, W. et al. (1996) VMD: visual molecular dynamics. *J. Mol. Graph.*, **14**, 33–38.
- Kaptein, R. et al. (1985) A protein structure from nuclear magnetic resonance data lac repressor headpiece. *J. Mol. Biol.*, **182**, 179–182.
- Karplus, M. (1959) Contact Electron-Spin coupling of nuclear magnetic moments. *J. Chem. Phys.*, **30**, 11–15.
- Koradi, R. et al. (1996) MOLMOL: a program for display and analysis of macromolecular structure. *J. Mol. Graphics*, **14**, 51–55.
- Laskowski, R.A. et al. (1996) AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR*, **8**, 477–486.
- Linge, J.P. and Nilges, M. (1999) Influence of non-bonded parameters on the quality of NMR structures: a new force field for NMR structure calculation. *J. Biomol. NMR*, **13**, 51–59.
- Linge, J.P. et al. (2003) Refinement of protein structures in explicit solvent. *Proteins*, **50**, 496–506.
- Loureiro-Ferreira, N. et al. (2010) e-NMR gLite grid enabled infrastructure. In: Proença A, Pina A, Garcia Tobio J, Ribeiro L (eds.). *IBERGRID 4th Iberian Grid infrastructure Conference Proceedings*. Netbiblio, La Coruna, Spain, pp. 360–382.
- Luginbuhl, P. et al. (1996) The new program OPAL for molecular dynamics simulations and energy refinements of biological macromolecules. *J. Biomol. NMR*, **8**, 136–146.
- Mittermaier, A. and Kay, L.E. (2009) Observing biological dynamics at atomic resolution using NMR. *Trends Biochem. Sci.*, **34**, 601–611.
- Nabuurs, S.B. et al. (2004) DRESS: a database of Refined solution NMR structures. *Proteins*, **55**, 483–486.
- Nabuurs, S.B. et al. (2006) Traditional biomolecular structure determination by NMR spectroscopy allows for major errors. *PLoS Comput. Biol.*, **2**, e9.
- Nederveen, A.J. et al. (2005) RECOORD: a recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank. *Proteins*, **59**, 662–672.
- Pettersen, E.F. et al. (2004) UCSF Chimera - a visualization system for exploratory research and analysis. *J. Comp. Chem.*, **25**, 1605–1612.
- Raman, S. et al. (2010) NMR structure determination for larger proteins using backbone-only data. *Science*, **327**, 1014–1018.
- Saccetti, E. and Rosato, A. (2008) The war of tools: how can NMR spectroscopists detect errors in their structures? *J. Biomol. NMR*, **40**, 251–261.
- Schwieters, C.D. et al. (2003) The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.*, **160**, 65–73.
- Shen, Y. et al. (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR*, **44**, 213–223.
- Shen, Y. et al. (2008) Consistent blind protein structure generation from NMR chemical shift data. *Proc. Natl Acad. Sci. USA*, **105**, 4685–4690.
- Snyder, D.A. et al. (2005) Comparisons of NMR spectral quality and success in crystallization demonstrate that NMR and X-ray crystallography are complementary methods for small protein structure determination. *J. Am. Chem. Soc.*, **127**, 16505–16511.
- Tjandra, N. et al. (1997) Defining long range order in NMR structure determination from the dependence of heteronuclear relaxation times on rotational diffusion anisotropy. *Nature Struct Biol*, **4**, 443–449.
- Tolbert, B.S. et al. (2010) Major groove width variations in RNA structures determined by NMR and impact of <sup>13</sup>C residual chemical shift anisotropy and <sup>1</sup>H-<sup>13</sup>C residual dipolar coupling on refinement. *J. Biomol. NMR*, **47**, 205–219.
- Ulrich, E.L. et al. (2007) BioMagResBank. *Nucleic Acid Res.*, **36**, D402–D406.
- Van Der, S.D. et al. (2005) GROMACS: fast, flexible, and free. *J. Comput. Chem.*, **26**, 1701–1718.
- Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acids*. New York, Wiley.
- Xia, B. et al. (2002) Comparison of protein solution structures refined by molecular dynamics simulation in vacuum, with a generalized Born model, and with explicit water. *J. Biomol. NMR*, **22**, 317–331.
- Yee, A.A. et al. (2005) NMR and X-ray crystallography, complementary tools in structural proteomics of small proteins. *J. Am. Chem. Soc.*, **127**, 16512–16517.