Deliverable D5.2

Project Title:	World-wide E-infrastructure for structural biology	
Project Acronym:	West-Life	
Grant agreement no.:	**675858**	
Deliverable title:	Overview (baseline) of services and portals to be integrated into the new VRE	
WP No.	5	
Lead Beneficiary:	8: Utrecht University	
WP Title	Virtual Research Environment	
Contractual delivery date:	4	
Actual delivery date:	4	
WP leader:	Alexandre M.J.J. Bonvin	Utrecht University
Contributing partners:	STFC, NKI, EMBL, MU, CSIC, CIRMMP, Instruct, UU, Luna, INFN	

COPYRIGHT NOTICE

DELIVERY SLIP

	Name	***Partner/Activity***	***Date***
From:	J. Schaarschmidt / A. Bonvin	UU/WP5	23.02.2016
Approved by:			

DOCUMENT LOG

Issue	***Date***	***Comment***	***Author/Partner***
v.1	15.02.2016	First version	J. Schaarschmidt/A.Bonvin / UU
V.2	23.02.2016	Second version after internal review by project partners	J. Schaarschmidt/A.Bonvin / UU

Contents

1. Executive summary

— West-Life aims to bring complex data analysis in Structural Biology to a simple Web browser-based Virtual Research Environment (VRE). Capitalizing on European and National projects, such as Instruct, WeNMR, and CCP4, as well as other projects, like EGI-Engage, through the MoBrain Competence Center, West-Life thus starts from a leading position in the Structural Biology field, in that its partners are already providing valued services in specific disciplines. As foundation for future assessments of the performance of this project we therefore define the baseline by providing the current user base and usage statistics of 12 web services provided by 6 project partners, which are to be incorporated into the VRE.

2. Project objectives

This deliverable is contributing to the following objectives:

No.	Objective	Yes	No
1	**Provide analysis solutions for the different Structural Biology approaches**	X	
2	**Provide automated pipelines to handle multi-technique datasets in an integrative manner**		X
3	**Provide integrated data management for single and multi-technique projects, based on existing e-infrastructure**		X
4	**Foster best practices, collaboration and training of end users**		X

3. Detailed report on the deliverable

1. Background

The overarching objective of the West-Life project is to bring the world of complex data analysis in Structural Biology to a simple Web browser-based Virtual Research Environment (VRE), available to any laboratory involved in the experimental structural characterization of biomolecules and their complexes and assemblies. Capitalizing on European and National projects, such as Instruct, WeNMR, and CCP4, as well as other projects, like EGI-Engage, through the MoBrain Competence Center formed by several West-Life partners, a series of Web Services addressing specific pipelines in NMR, X-ray diffraction, SAXS and cryo Electron Microscopy data analysis are offered with direct impact on a large and worldwide user base. West-Life thus starts from a leading position in the Structural Biology field, in that its partners are already providing valued services in specific disciplines. In the following, a description is given of 19 existing portals operated by West-Life partners that form the initial basis for the services offered by the West-Life VRE. For each, a baseline in terms of users, number of jobs, etc. is defined. This defines the baseline to monitor the KPIs of the projects related to the services offered.

2. Description of existing portals

1. SCIPION

The SCIPION server offers access to 3D electron microscopy online processing workflows, to provide a first analysis of the data without any local installation. The portal currently offers 3 web tools:

- My first map, to obtain an initial 3D map from your averaged images.
- Movie alignment, to align the movies obtained in an Electron Microscope, correcting for global frame movements, and download the corrected averaged micrograph.
- My resmap, to run ResMap online to compute the local resolution of 3D density maps and download resmap charts results.

2. GROMACS

GROMACS (www.gromacs.org) is a versatile package to perform molecular dynamics, i.e. simulate the Newtonian equations of motion for systems with hundreds to millions of particles. GROMACS is able to work with many biochemical molecules like proteins, lipids and nucleic acids. The WeNMR GROMACS grid-enabled webportal combines the versatility of this molecular dynamics package with the calculation power of the WeNMR grid. This will enable the user to perform many simulations from the comfort of his/her internet browser anywhere in the world. The server is furthermore aimed to provide a user friendly and efficient MD experience by performing many preparation and optimization steps automatically.

The GROMACS web server, originally developed under the WeNMR e-Infrastructure project (www.wenmr.eu) uses resources provided by the EGI (www.egi.eu) and the associated National Grid Initiatives (NGIs).[1,2]

3. ViCi

ViCi is an innovative software for ligand-based drug design available free of charge to academic researchers via a webserver. ViCi uses a combination of mathematical descriptors of molecular size, shape and topology to describe small molecule structures. Following input of a template molecule, typically that of a known ligand in its bound conformation in a particular protein, the software will rapidly screen a database (currently 8 million compounds) and extract those predicted to have similar shape and electrostatic compositions and therefore to be possible ligands for the same protein. Results are typically obtained in a matter of hours and are returned to the user ranked by probability of binding.

4. HADDOCK

HADDOCK2.2 (High Ambiguity Driven protein-protein DOCKing) is an integrative, information-driven flexible docking approach for the modeling of biomolecular complexes. HADDOCK distinguishes itself from ab-initio docking methods in the fact that it encodes information from identified or predicted protein interfaces in ambiguous interaction restraints (AIRs) to drive the docking process. HADDOCK can deal with a large class of modeling problems including protein-protein, protein-nucleic acids and protein-ligand complexes.

The HADDOCK2.2 grid-enabled web server, originally developed under the WeNMR e-Infrastructure project (www.wenmr.eu) uses resources provided by the EGI (www.egi.eu) and the associated National Grid Initiatives (NGIs).[2-4]

5. AMPS-NMR

AMPS-NMR (AMBER-based Portal Server for NMR structures) is a web interface to set up and run calculations with the AMBER package. The interface allows the refinement of NMR structures of biological macromolecules through restrained Molecular Dynamics (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. As an ancillary service, it provides access to a web interface to AnteChamber, enabling the calculation of force field parameters for organic molecules such as ligands in protein–ligand adducts.

The AMPS-NMR grid-enabled web server, originally developed under the WeNMR e-Infrastructure project (www.wenmr.eu) uses resources provided by the EGI (www.egi.eu) and the associated National Grid Initiatives (NGIs).[2,5]

6. CS-Rosetta3

CS-Rosetta is a protocol which generates 3D models of proteins, using only the 13CA, 13CB, 13C', 15N, 1HA and 1HN NMR chemical shifts as input. Based on these parameters, CS ROSETTA uses a SPARTA-based selection procedure to select a set of fragments from a fragment-library (where the chemical shifts and the 3D structure of the fragments are known). The fragments are assembled using the Rosetta protocol. The generated models are rescored based on the difference between the back-calculated chemical shifts of the generated models and the input chemical shifts, and when available, with a post-scoring procedure based on unassigned NOE lists.

The CS-Rosetta3 grid-enabled web server, originally developed under the WeNMR e-Infrastructure project (www.wenmr.eu) uses resources provided by the EGI (www.egi.eu) and the associated National Grid Initiatives (NGIs).[2,6]

7. FANTEN

Pseudocontact shifts (PCSs) and residual dipolar couplings (RDCs) arising from the presence of paramagnetic metal ions in proteins as well as RDCs due to partial orientation induced by external orienting media are nowadays routinely measured as a part of the NMR characterization of biologically relevant systems. PCSs and RDCs can be used: 1) to determine and/or refine protein structures in solution, 2) to monitor the extent of conformational heterogeneity in systems composed of rigid domains which can reorient with respect to one another, and 3) to obtain structural information in protein-protein complexes. The use of both PCSs and RDCs proceeds through the determination of the anisotropy tensors which are at the origin of these NMR observables. A new user-friendly web tool, called FANTEN (Finding ANisotropy TENsors), has been developed for the determination of the anisotropy tensors related to PCSs and RDCs and has been made freely available through the WeNMR (http://fanten-enmr.cerm.unifi.it:8080) gateway. The program has many features not available in other existing programs, among which the possibility of a joint analysis of several sets of PCS and RDC data and the possibility to perform rigid body minimizations.[7]

8. UNIO

UNIO program enables users to perform automated NMR data analysis for 3D protein structure determination. UNIO represents the result of more than a decade of basic research performed in order to enable accurate, objective and highly automated protein structure determination by NMR. The UNIO program includes data analysis algorithms for all parts of an NMR structure determination process ranging from backbone and side-chain assignment to NOE assignment and structure calculation.

The UNIO web server, originally developed under the WeNMR e-Infrastructure project (www.wenmr.eu) uses resources provided by the EGI (www.egi.eu) and the associated National Grid Initiatives (NGIs).[2]

9. XPLOR-NIH

The Xplor-NIH program is a version of X-PLOR, one of the most popular programs to obtain protein solution structures through structural restraints, simulated annealing calculations and energy minimization. Xplor-NIH is based on torsion angle dynamics thereby allowing fast calculations of large protein structures. In addition, paramagnetism-based restraints have been introduced into Xplor-NIH in a uniform way and by properly considering all their interconnections. The whole set of modules which allows the use of paramagnetic restraint is called PARArestraints for Xplor-NIH.

The Xplor-NIH grid-enabled web server provides an easy interface to set up calculations using also the PARArestraints modules. It was originally developed under the WeNMR e-Infrastructure project (www.wenmr.eu) uses resources provided by the EGI (www.egi.eu) and the associated National Grid Initiatives (NGIs).[2,8]

10. ARP/wARP

ARP/wARP is a software project for automated protein model building and structure refinement in macromolecular crystallography. ARP/ wARP combines pattern recognition-based interpretation of electron density, its modelling as a hybrid model and a maximum likelihood parameter refinement with REFMAC. Typically, X-ray data to 2.7 Å resolution or better are required, although a considerable part of a protein model can sometimes be built at a resolution of 3.0 Å or worse. ARP/wARP builds proteins, RNA/DNA, secondary structure, side chains, loops, solvent and ligands.

The ARP/wARP portal is free for members of a public funded academic, education or research institution. Proprietary users are required to obtain a commercial license from EMBLE (http://webapps.embl-hamburg.de/ARPwARP/licence.htm)[9-11]

11. Auto-Rickshaw

The EMBL-Hamburg automated crystal structure determination platform Auto-Rickshaw is a software pipeline system, which contains several distinct decision-makers, and which executes a number of macromolecular crystallographic software programs to provide automated and efficient crystal structure determination. A large number of possible structure solution paths are encoded in the system, and the optimal path is selected by the decision-makers as the structure solution evolves. The processes have been optimised for speed so that the pipeline can be used effectively for validation of the X-ray experiment at a synchrotron beamline. The platform offers SAD, S-SAD, SIRAS, 2W-MAD, 3W-MAD or 4W-MAD phase determination, molecular replacement (MR) and MRSAD phasing. It also includes RIP and MRRIP phasing and the phasing protocols have been optimised for UV induced radiation damage X-ray data. Recently it has been extended to include MRSIRAS phasing.[12,13]

12. CCP4 online – Ample

The CCP4 Ample server allows to perform an automated search model generation and molecular replacement using decoys from ab initio modelling. The service requires structure factors and sequence for the target, and decoys from the Quark server. The decoys are clustered, truncated to common cores, and presented to MrBUMP for structure solution.[14,15]

13. CCP4 online – Balbes

An automated Molecular Replacement (MR) pipeline - Balbes integrates into one system all the components necessary for solving a crystal structure by Molecular Replacement. Given structure factors and a sequence for the target, Balbes will search for models from an internal database (derived from the PDB). Checking the ARP/wARP checkbox will send Balbes' results to the ARP/wARP server.[16,17]

14. CCP4 online – Crank2

The CCP4 Crank2 server offers an automated structure solution pipeline for experimental phasing using maximum likelihood methods. The service covers SAD, SIRAS and MAD techniques. Unlike the traditional stepwise approach, the combined function simultaneously uses the information from density modification, model building and from the data to provide the best estimate of the electron density.[18-20]

15. CCP4 online – MrBUMP

The CCP4 MrBUMP server offers an automation pipeline for macromolecular structure solution by molecular replacement. There is a special emphasis on the discovery and preparation of a large number of search models, all of which can be passed to the core molecular-replacement programs Phaser or Molrep. Given a target sequence and experimental structure factors, it will search for homologous structures, create a set of suitable search models from the template structures, do molecular replacement, and test the solutions with some rounds of restrained refinement.[17,21]

16. CCP4 online – Shelx

The CCP4 Shelx server offers an automated SHELXC/D/E structure solution pipeline for fast routine experimental phasing. Accepts data in XDS, Scalepack, SHELX hkl or mtz formats and outputs phases and a poly-Ala trace. If a protein sequence is provided, BUCCANEER and REFMAC complete the structure.[22]

17. CCP4 online – Zanuda

The CCP4 Zanuda server offers a space group and crystallographic origin validation. The program Zanuda was developed to automate the validation of space group assignment. In addition, the program can be used to restore the correct space group in structures which were intentionally solved in low symmetry space groups including P1. The validation is based on the results of a series of refinements in space groups, which are compatible with the observed unit cell parameters.[23]

18. PDB_REDO

The PDB_REDO server provides a fully automated procedure for optimizing crystallographic structure models. It is based on the PDB_REDO pipeline that combines standard crystallographic tools with state-of-the-art decision-making algorithms and dedicated model rebuilding programs. Extensive model validation is used to guide the decision-making and to report the results to the user. The pipeline is thoroughly tested by systematically applying it to all crystallographic structure models in the protein data bank (PDB). The resulting structure models are made available to the structural biology community through the PDB_REDO data bank.[24-26]

19. CCD

CCD is a metaserver that collects predictions of secondary structure, disorder, membrane topology, from several web services to allow users to make an informed decision making a construct of a protein. It also designs primers for PCR amplification of the construct.[27]

3. Portal Statistics - Baseline

Portal	Method	Service	grid/ cloud-enabled	Total no. of Users	No. of new users in 2015	No. of user submissions 2015	No. of grid/cloud jobs 2015
Scipion	Cryo-EM	3D electron microscopy online processing workflows	-/+	-	-	100	-
GROMACS	Modelling	Molecular dynamics simulations	+/-	112	26	173	588
ViCi	Modelling	In silico ligand-based drug design	-/-	292	92	93	-
HADDOCK	NMR/modelling	Docking of biomolecular complexes	+/-	6699	1450	25k	7,5M
AMPS-NMR	NMR	Molecular dynamics simulations with AMBER	+/-	300	50	-	8k
CS-Rosetta3	NMR	Structure prediction with chemical shifts from NMR	+/-	51	16	67	189k
FANTEN	NMR	Determination of anisotropy tensors (NMR)	-/-	-	-	-	-
UNIO	NMR	Structure calculations including NOE assignment from NMR data	+/-	59	31	62	2285
XPLOR-NIH	NMR	Protein solution structure determination through structural restraints, simulated annealing calculations and energy minimization	+/-	100	10	-	80k
ARP/wARP	X-ray	Crystallographic Macromolecular Model Building	-/-	4088	424	3,2k	-
Auto-Rickshaw	X-ray	Automated crystal structure determination platform	-/-	2319	205	3,5k	-
CCP4 - Ample*	X-ray	Automated search model generation and molecular replacement (MR)	-/-	20	-	-	-
CCP4 - Balbes	X-ray	Automated MR pipeline	-/-	1155	728	3,3k	-
CCP4 - Crank2*	X-ray	Structure solution pipeline for experimental phasing	-/-	10	-	-	-
CCP4 - MrBUMP	X-ray	Macromolecular structure solution by MR	-/-	580	473	1,3k	-
CCP4 - Shelx*	X-ray	SHELXC/D/E structure solution	-/-	19	-	-	-
CCP4 - Zanuda	X-ray	Space group and crystallographic origin validation	-/-	264	189	371	-
PDB_REDO	X-ray	Optimization of crystallographic structure models	-/-	700	500	3k	-
CCD	X-ray/Molecular Biology	Design of constructs for protein crystallography	-/-	-	-	~3k	-

** Portal operation started 2016*

References cited

1. van Dijk, M., Wassenaar, T. A. & Bonvin, A. M. J. J. A Flexible, Grid-Enabled Web Portal for GROMACS Molecular Dynamics Simulations. *J Chem Theory Comput* **8,** 3463–3472 (2012).
2. Wassenaar, T. A. *et al.* WeNMR: Structural Biology on the Grid. *J Grid Computing* **10,** 743–767 (2012).
3. van Zundert, G. C. P. *et al.* The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J Mol Biol* (2015). doi:10.1016/j.jmb.2015.09.014
4. 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).
5. 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).
6. Schot, G. & Bonvin, A. M. J. J. Performance of the WeNMR CS-Rosetta3 web server in CASD-NMR. *Journal of Biomolecular NMR* **62,** 497–502 (2015).
7. Rinaldelli, M., Carlon, A., Ravera, E., Parigi, G. & Luchinat, C. FANTEN: a new web-based interface for the analysis of magnetic anisotropy-induced NMR data. *J Biomol NMR* **61,** 21–34 (2014).
8. Banci, L. *et al.* Paramagnetism-Based Restraints for Xplor-NIH. *J Biomol NMR* **28,** 249–261 (2004).
9. Perrakis, A., Morris, R. & Lamzin, V. S. Automated protein model building combined with iterative structure refinement. *Nature Structural & Molecular Biology* **6,** 458–463 (1999).
10. Wiegels, T., Lamzin, V. S.IUCr. Use of noncrystallographic symmetry for automated model building at medium to low resolution. *Acta Crystallogr. D Biol. Crystallogr.* **68,** 446–453 (2012).
11. Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat Protoc* **3,** 1171–1179 (2008).
12. Panjikar, S. *et al.* Auto-Rickshaw: an automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment. *Acta Crystallogr. D Biol. Crystallogr.* **61,** 449–457 (2005).
13. Panjikar, S. *et al.* On the combination of molecular replacement and single-wavelength anomalous diffraction phasing for automated structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **65,** 1089–1097 (2009).
14. Bibby, J., Keegan, R. M., Mayans, O., Winn, M. D. & Rigden, D. J. AMPLE: a cluster-and-truncate approach to solve the crystal structures of small proteins using rapidly computed ab initio models. *Acta Crystallogr. D Biol. Crystallogr.* **68,** 1622–1631 (2012).
15. Keegan, R. M. *et al.* Exploring the speed and performance of molecular replacement with AMPLE using QUARK ab initio protein models. *Acta Crystallogr. D Biol. Crystallogr.* **71,** 338–343 (2015).
16. Long, F., Vagin, A. A., Young, P. & Murshudov, G. N. BALBES: a molecular-replacement pipeline. *Acta Crystallogr. D Biol. Crystallogr.* **64,** 125–132 (2008).
17. Keegan, R. M. *et al.* Evaluating the solution from MrBUMP and BALBES. *Acta Crystallogr. D Biol. Crystallogr.* **67,** 313–323 (2011).
18. Skubák, P. & Pannu, N. S. Automatic protein structure solution from weak X-ray data. *Nat Commun* **4,** 2777 (2013).
19. Pannu, N. S. *et al.* Recent advances in the CRANK software suite for experimental phasing. *Acta Crystallogr. D Biol. Crystallogr.* **67,** 331–337 (2011).
20. Ness, S. R., de Graaff, R. A. G., Abrahams, J. P. & Pannu, N. S. CRANK: new methods for automated macromolecular crystal structure solution. *Structure* **12,** 1753–1761 (2004).
21. Keegan, R. M. & Winn, M. D. MrBUMP: an automated pipeline for molecular replacement. *Acta Crystallogr. D Biol. Crystallogr.* **64,** 119–124 (2008).
22. Sheldrick, G. M. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 479–485 (2010).

23. Lebedev, A. A. & Isupov, M. N. Space-group and origin ambiguity in macromolecular structures with pseudo-symmetry and its treatment with the program Zanuda. *Acta Crystallogr. D Biol. Crystallogr.* **70,** 2430–2443 (2014).
24. Joosten, R. P. *et al.* PDB_REDO: automated re-refinement of X-ray structure models in the PDB. *J Appl Crystallogr* **42,** 376–384 (2009).
25. Joosten, R. P., Joosten, K., Murshudov, G. N., Perrakis, A.IUCr. PDB_REDO: constructive validation, more than just looking for errors. *Acta Crystallogr. D Biol. Crystallogr.* **68,** 484–496 (2012).
26. Joosten, R. P., Long, F., Murshudov, G. N. & Perrakis, A. The PDB_REDO server for macromolecular structure model optimization. *IUCrJ* **1,** 213–220 (2014).
27. Mooij, W. T. M., Mitsiki, E. & Perrakis, A. ProteinCCD: enabling the design of protein truncation constructs for expression and crystallization experiments. *Nucleic Acids Res* **37,** W402–5 (2009).

Appendix 1: Portal summary – Scipion Web Tools

Portal name	**Scipion Web Tools**
Short description	The SCIPION server offers access to 3D electron microscopy online processing workflows, to provide a first analysis of the data without any local installation. The portal currently offers 3 web tools: ● *My first map*, to obtain an initial 3D map from your averaged images. ● *Movie alignment*, to align the movies obtained in an Electron Microscope, correcting for global frame movements, and download the corrected averaged micrograph. ● *My resmap*, to run ResMap online to compute the local resolution of 3D density maps and download resmap charts results.
Keywords	3DEM, workflow, initial volume, movie alignment, resmap
URL	http://scipion.cnb.csic.es/m/services/
Grid-enabled	no
Cloud-enabled	Yes
Total number of registered users	N/A
Number of new users over 2015	N/A
Number of projects created	30
Number of user submissions processed over 2015	100
Number of grid/cloud jobs over 2015	0
Key references	Nothing yet.

Appendix 1: Portal summary – GROMACS

Portal name	**GROMACS**
Short description	GROMACS (www.gromacs.org) is a versatile package to perform molecular dynamics, i.e. simulate the Newtonian equations of motion for systems with hundreds to millions of particles. GROMACS is able to work with many biochemical molecules like proteins, lipids and nucleic acids. The WeNMR GROMACS grid-enabled webportal combines the versatility of this molecular dynamics package with the calculation power of the WeNMR grid. This will enable the user to perform many simulations from the comfort of his/her internet browser anywhere in the world. The server is furthermore aimed to provide a user friendly and efficient MD experience by performing many preparation and optimization steps automatically. The GROMACS web server, originally developed under the WeNMR e-Infrastructure project (www.wenmr.eu) uses resources provided by the EGI (www.egi.eu) and the associated National Grid Initiatives (NGIs).
Keywords	Molecular dynamics; proteins; simulations
URL	http://haddock.science.uu.nl/enmr/services/GROMACS
Grid-enabled	Yes (gLite submission system - multithreading)
Cloud-enabled	No
Total number of registered users	112 (February 2nd, 2016)
Number of new users over 2015	26
Number of user submissions processed over 2015	173
Number of grid/cloud jobs over 2015	588 (internal stats from gLite submissions)
Key references	• M. van Dijk, T.A. Wassenaar and A.M.J.J. Bonvin A flexible, grid-enabled web portal for GROMACS molecular dynamics simulations *J. Chem. Theo. Comput.*, B, 3463-3472 (2012). • 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. Giachetti, L. Ferella, A. Rosato, I. Bertini, T. Herrmann, H.R.A. Jonker, A. Bagaria, V. Jaravine, P. Guntert, H. Schwalbe, W.F. Vranken, J.F. Doreleijers, G. Vriend, G.W. Vuister, D. Franke, A. Kikhney, D.I. Svergun, R. Fogh, J. Ionides, E.D. Laue, C. Spronk, S. Jurka, M. Verlato, S. Badoer, S. Dal Pra, M. Mazzucato, E. Frizziero and A.M.J.J. Bonvin WeNMR: Structural Biology on the Grid. *J. Grid. Comp.*, **10**, 743-767 (2012).

Appendix 1: Portal summary – ViCi

Portal name	**ViCi**
Short description	ViCi is an innovative software for ligand-based drug design available free of charge to academic researchers via a webserver. ViCi uses a combination of mathematical descriptors of molecular size, shape and topology to describe small molecule structures. Following input of a template molecule, typically that of a known ligand in its bound conformation in a particular protein, the software will rapidly screen a database (currently 8 million compounds) and extract those predicted to have similar shape and electrostatic compositions and therefore to be possible ligands for the same protein. Results are typically obtained in a matter of hours and are returned to the user ranked by probability of binding.
Keywords	In silico ligand-based drug design, small molecule structures, scaffold hopping
URL	http://www.embl-hamburg.de/vici/index
Grid-enabled	no
Cloud-enabled	no
Total number of users	292 (based on unique Email addresses as of 31.12.2015).
Number of new users over 2015	92
Number of user submissions processed over 2015	93
Number of grid/cloud jobs over 2015	0
Key references	No publication yet; refer to website URL: http://www.embl-hamburg.de/vici/index

Appendix 1: Portal summary – HADDOCK

Portal name	**HADDOCK2.2**
Short description	HADDOCK2.2 (High Ambiguity Driven protein-protein DOCKing) is an integrative, information-driven flexible docking approach for the modeling of biomolecular complexes. HADDOCK distinguishes itself from ab-initio docking methods in the fact that it encodes information from identified or predicted protein interfaces in ambiguous interaction restraints (AIRs) to drive the docking process. HADDOCK can deal with a large class of modeling problems including protein-protein, protein-nucleic acids and protein-ligand complexes. The HADDOCK2.2 grid-enabled web server, originally developed under the WeNMR e-Infrastructure project (www.wenmr.eu) uses resources provided by the EGI (www.egi.eu) and the associated National Grid Initiatives (NGIs).
Keywords	Integrative modelling; biomolecular complexes; docking
URL	http://haddock.science.uu.nl/enmr/services/HADDOCK2.2
Grid-enabled	Yes (both gLite and DIRAC4EGI submission system)
Cloud-enabled	No
Total number of registered users	6699 (February 2nd, 2016)
Number of new users over 2015	1450
Number of user submissions processed over 2015	24790
Number of grid/cloud jobs over 2015	7.49 millions (internal stats from both gLite and DIRAC4EGI submissions)
Key references	• G.C.P van Zundert, J.P.G.L.M. Rodrigues, M. Trellet, C. Schmitz, P.L. Kastritis, E. Karaca, A.S.J. Melquiond, M. van Dijk, S.J. de Vries and A.M.J.J. Bonvin. The HADDOCK2.2 webserver: User-friendly integrative modeling of biomolecular complexes. J. Mol. Biol., Advanced Online Publication (2015). • 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. Giachetti, L. Ferella, A. Rosato, I. Bertini, T. Herrmann, H.R.A. Jonker, A. Bagaria, V. Jaravine, P. Guntert, H. Schwalbe, W.F. Vranken, J.F. Doreleijers, G. Vriend, G.W. Vuister, D. Franke, A. Kikhney, D.I. Svergun, R. Fogh, J. Ionides, E.D. Laue, C. Spronk, S. Jurka, M. Verlato, S. Badoer, S. Dal Pra, M. Mazzucato, E. Frizziero and A.M.J.J. Bonvin WeNMR: Structural Biology on the Grid. J. Grid. Comp., 10, 743-767 (2012). • S.J. de Vries, M. van Dijk and A.M.J.J. Bonvin The HADDOCK web server for data-driven biomolecular docking. *Nature Protocols*, **5**, 883-897 (2010).

Appendix 1: Portal summary – AMPS-NMR

Portal name	**AMPS-NMR**
Short description	AMPS-NMR (AMBER-based Portal Server for NMR structures) is a web interface to set up and run calculations with the AMBER package. The interface allows the refinement of NMR structures of biological macromolecules through restrained Molecular Dynamics (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. As an ancillary service, it provides access to a web interface to AnteChamber, enabling the calculation of force field parameters for organic molecules such as ligands in protein–ligand adducts. The AMPS-NMR grid-enabled web server, originally developed under the WeNMR e-Infrastructure project (www.wenmr.eu) uses resources provided by the European Grid Initiative EGI (www.egi.eu) and the associated National Grid Initiatives (NGIs).
Keywords	NMR; structure determination; structural biology; molecular dynamics
URL	http://py-enmr.cerm.unifi.it/access/index/amps-nmr
Grid-enabled	Yes
Cloud-enabled	No
Total number of registered users	300
Number of new users over 2015	50
Number of user submissions processed over 2015	n.a.
Number of grid/cloud jobs over 2015	8000
Key references	• Bertini I, Case DA, Ferella L, Giachetti A, Rosato A. A Grid-enabled web portal for NMR structure refinement with AMBER. Bioinformatics. 27:2384-2390, 2011 • 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. Giachetti, L. Ferella, A. Rosato, I. Bertini, T. Herrmann, H.R.A. Jonker, A. Bagaria, V. Jaravine, P. Guntert, H. Schwalbe, W.F. Vranken, J.F. Doreleijers, G. Vriend, G.W. Vuister, D. Franke, A. Kikhney, D.I. Svergun, R. Fogh, J. Ionides, E.D. Laue, C. Spronk, S. Jurka, M. Verlato, S. Badoer, S. Dal Pra, M. Mazzucato, E. Frizziero and A.M.J.J. Bonvin. WeNMR: Structural Biology on the Grid. J. Grid. Comp., 10, 743-767 (2012).

Appendix 1: Portal summary – CS-Rosetta3

Portal name	**CS-Rosetta3**
Short description	CS-Rosetta is a protocol which generates 3D models of proteins, using only the 13CA, 13CB, 13C', 15N, 1HA and 1HN NMR chemical shifts as input. Based on these parameters, CS ROSETTA uses a SPARTA-based selection procedure to select a set of fragments from a fragment-library (where the chemical shifts and the 3D structure of the fragments are known). The fragments are assembled using the Rosetta protocol. The generated models are rescored based on the difference between the back-calculated chemical shifts of the generated models and the input chemical shifts, and when available, with a post-scoring procedure based on unassigned NOE lists. The CS-Rosetta3 grid-enabled web server, originally developed under the WeNMR e-Infrastructure project (www.wenmr.eu) uses resources provided by the EGI (www.egi.eu) and the associated National Grid Initiatives (NGIs).
Keywords	NMR; chemical shifts; structure prediction
URL	http://haddock.science.uu.nl/enmr/services/CS-ROSETTA3/
Grid-enabled	Yes (gLite submission system)
Cloud-enabled	No
Total number of registered users	51 (February 2nd, 2016)
Number of new users over 2015	16
Number of user submissions processed over 2015	67
Number of grid/cloud jobs over 2015	189106 (internal stats from gLite submissions)
Key references	• G. van der Schot and A.M.J.J. Bonvin. Performance of the WeNMR CS-Rosetta3 web server in CASD-NMR. *J. Biomol. NMR*. **62**, 497-502 (2015). • 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. Giachetti, L. Ferella, A. Rosato, I. Bertini, T. Herrmann, H.R.A. Jonker, A. Bagaria, V. Jaravine, P. Guntert, H. Schwalbe, W.F. Vranken, J.F. Doreleijers, G. Vriend, G.W. Vuister, D. Franke, A. Kikhney, D.I. Svergun, R. Fogh, J. Ionides, E.D. Laue, C. Spronk, S. Jurka, M. Verlato, S. Badoer, S. Dal Pra, M. Mazzucato, E. Frizziero and A.M.J.J. Bonvin WeNMR: Structural Biology on the Grid. *J. Grid. Comp.*, **10**, 743-767 (2012).

Appendix 1: Portal summary – FANTEN

Portal name	**FANTEN**
Short description	Pseudocontact shifts (PCSs) and residual dipolar couplings (RDCs) arising from the presence of paramagnetic metal ions in proteins as well as RDCs due to partial orientation induced by external orienting media are nowadays routinely measured as a part of the NMR characterization of biologically relevant systems. PCSs and RDCs can be used: 1) to determine and/or refine protein structures in solution, 2) to monitor the extent of conformational heterogeneity in systems composed of rigid domains which can reorient with respect to one another, and 3) to obtain structural information in protein-protein complexes. The use of both PCSs and RDCs proceeds through the determination of the anisotropy tensors which are at the origin of these NMR observables. A new user-friendly web tool, called FANTEN (Finding ANisotropy TENsors), has been developed for the determination of the anisotropy tensors related to PCSs and RDCs and has been made freely available through the WeNMR (http://fanten-enmr.cerm.unifi.it:8080) gateway. The program has many features not available in other existing programs, among which the possibility of a joint analysis of several sets of PCS and RDC data and the possibility to perform rigid body minimizations
Keywords	Paramagnetic NMR; metalloprotein; structural biology
URL	http://fanten-enmr.cerm.unifi.it:8080/
Grid-enabled	No
Cloud-enabled	No
Total number of registered users	No registration required
Number of new users over 2015	n.a.
Number of user submissions processed over 2015	n.a.
Number of grid/cloud jobs over 2015	n.a.
Key references	• Rinaldelli M, Carlon A, Ravera E, Parigi G, Luchinat C. FANTEN: a new web-based interface for the analysis of magnetic anisotropy-induced NMR data. *J Biomol NMR* **61,** 21-34 (2015)

Appendix 1: Portal summary – UNIO

Portal name	**UNIO**
Short description	UNIO program enables users to perform automated NMR data analysis for 3D protein structure determination. UNIO represents the result of more than a decade of basic research performed in order to enable accurate, objective and highly automated protein structure determination by NMR. The UNIO program includes data analysis algorithms for all parts of an NMR structure determination process ranging from backbone and side-chain assignment to NOE assignment and structure calculation. The UNIO web server, originally developed under the WeNMR e-Infrastructure project (www.wenmr.eu) uses resources provided by the EGI (www.egi.eu) and the associated National Grid Initiatives (NGIs).
Keywords	NMR; NOE assignment; structure calculations
URL	http://haddock.science.uu.nl/enmr/services/UNIO
Grid-enabled	Yes (gLite submission system)
Cloud-enabled	No
Total number of registered users	59 (February 2nd, 2016)
Number of new users over 2015	31
Number of user submissions processed over 2015	62
Number of grid/cloud jobs over 2015	2285 (internal stats from gLite submissions)
Key references	• 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. Giachetti, L. Ferella, A. Rosato, I. Bertini, T. Herrmann, H.R.A. Jonker, A. Bagaria, V. Jaravine, P. Guntert, H. Schwalbe, W.F. Vranken, J.F. Doreleijers, G. Vriend, G.W. Vuister, D. Franke, A. Kikhney, D.I. Svergun, R. Fogh, J. Ionides, E.D. Laue, C. Spronk, S. Jurka, M. Verlato, S. Badoer, S. Dal Pra, M. Mazzucato, E. Frizziero and A.M.J.J. Bonvin WeNMR: Structural Biology on the Grid. *J. Grid. Comp.*, **10**, 743-767 (2012).

Appendix 1: Portal summary – XPLOR-NIH

Portal name	**XPLOR-NIH**
Short description	The Xplor-NIH program is a version of X-PLOR, one of the most popular programs to obtain protein solution structures through structural restraints, simulated annealing calculations and energy minimization. Xplor-NIH is based on torsion angle dynamics thereby allowing fast calculations of large protein structures. In addition, paramagnetism-based restraints have been introduced into Xplor-NIH in a uniform way and by properly considering all their interconnections. The whole set of modules which allows the use of paramagnetic restraint is called PARArestraints for Xplor-NIH. The Xplor-NIH grid-enabled web server provides an easy interface to set up calculations using also the PARArestraints modules. It was originally developed under the WeNMR e-Infrastructure project (www.wenmr.eu) uses resources provided by the EGI (www.egi.eu) and the associated National Grid Initiatives (NGIs).
Keywords	NMR; structure determination; structural biology; molecular dynamics
URL	http://py-enmr.cerm.unifi.it/access/index/xplor-nih
Grid-enabled	Yes
Cloud-enabled	No
Total number of registered users	100
Number of new users over 2015	10
Number of user submissions processed over 2015	n.a.
Number of grid/cloud jobs over 2015	80,000
Key references	• Banci L, Bertini I, Cavallaro G, Giachetti A, Luchinat C, Parigi G *J Biomol NMR*. Paramagnetism-based restraints for Xplor-NIH. **28**:249-61 (2004). • 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. Giachetti, L. Ferella, A. Rosato, I. Bertini, T. Herrmann, H.R.A. Jonker, A. Bagaria, V. Jaravine, P. Guntert, H. Schwalbe, W.F. Vranken, J.F. Doreleijers, G. Vriend, G.W. Vuister, D. Franke, A. Kikhney, D.I. Svergun, R. Fogh, J. Ionides, E.D. Laue, C. Spronk, S. Jurka, M. Verlato, S. Badoer, S. Dal Pra, M. Mazzucato, E. Frizziero and A.M.J.J. Bonvin WeNMR: Structural Biology on the Grid. *J. Grid. Comp.*, **10**, 743-767 (2012).

Appendix 1: Portal summary – ARP/wARP

Portal name	**ARP/wARP 7.6**
Short description	Crystallographic Macromolecular Model Building, Version 7.6 ARP/wARP is a software project for automated protein model building and structure refinement in macromolecular crystallography. ARP/ wARP combines pattern recognition-based interpretation of electron density, its modelling as a hybrid model and a maximum likelihood parameter refinement with REFMAC. Typically, X-ray data to 2.7 Å resolution or better are required, although a considerable part of a protein model can sometimes be built at a resolution of 3.0 Å or worse. ARP/wARP builds proteins, RNA/DNA, secondary structure, side chains, loops, solvent and ligands. The ARP/wARP portal is free for members of a public funded academic, education or research institution. Proprietary users are required to obtain a commercial license from EMBLE (http://webapps.embl-hamburg.de/ARPwARP/licence.htm)
Keywords	Crystallographic Macromolecular Model Building
URL	http://cluster.embl-hamburg.de/ARPwARP/remote-http.html
Grid-enabled	no
Cloud-enabled	no
Total number of registered users	4088 (based on unique Email addresses as of 31.12.2015).
Number of new users over 2015	424
Number of user submissions processed over 2015	3250
Number of grid/cloud jobs over 2015	0
Key references	• Langer G, Cohen SX, Lamzin VS, Perrakis A. (2008) Automated macromolecular model building for x-ray crystallography using ARP/wARP version 7. Nat. Protoc. 3, 1171-1179 • Perrakis A, Morris RM, Lamzin VS. (1999) Automated protein model building combined with iterative structure refinement. Nature Struct. Biol. 6, 458-463 • Wiegels T. & Lamzin, V.S. (2012) Use of noncrystallographic symmetry for automated model building at medium to low resolution. Acta Crystallogr D Biol Crystallogr. 68, 446-453

Appendix 1: Portal summary – Auto-Rickshaw

Portal name	**Auto-Rickshaw**
Short description	The EMBL-Hamburg automated crystal structure determination platform Auto-Rickshaw is a software pipeline system, which contains several distinct decision-makers, and which executes a number of macromolecular crystallographic software programs to provide automated and efficient crystal structure determination. A large number of possible structure solution paths are encoded in the system, and the optimal path is selected by the decision-makers as the structure solution evolves. The processes have been optimised for speed so that the pipeline can be used effectively for validation of the X-ray experiment at a synchrotron beamline. The platform offers SAD, S-SAD, SIRAS, 2W-MAD, 3W-MAD or 4W-MAD phase determination, molecular replacement (MR) and MRSAD phasing. It also includes RIP and MRRIP phasing and the phasing protocols have been optimised for UV induced radiation damage X-ray data. Recently it has been extended to include MRSIRAS phasing.
Keywords	Automated crystal structure determination, SAD, S-SAD, SIRAS, 2W-MAD, 3W-MAD or 4W-MAD, MRSAD, MRSIRAS, MR, molecular replacement, phase determination, model building
URL	http://webapps.embl-hamburg.de/cgi-bin/Auto-Rick/arinitAR1.cgi
Grid-enabled	no
Cloud-enabled	no
Total number of registered users	2319 (based on unique Email addresses as of 31.12.2015).
Number of new users over 2015	205
Number of user submissions processed over 2015	3569
Number of grid/cloud jobs over 2015	0
Key references	• Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S. & Tucker, P. A. (2005). Auto-Rickshaw - An automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment. Acta Cryst. D61, 449-457. • Panjikar S., Parthasarathy, V., Lamzin, V., Weiss, M.S., Tucker, P.A., (2009). On the combination of molecular replacement and single anomalous diffraction phasing for automated structure determination Acta Cryst. D65,1089-1097.

Appendix 1: Portal summary – CCP4 online - Ample

Portal name	**CCP4 online - Ample**
Short description	The CCP4 Ample server allows to perform an automated search model generation and molecular replacement using decoys from ab initio modelling. The service requires structure factors and sequence for the target, and decoys from the Quark server. The decoys are clustered, truncated to common cores, and presented to MrBUMP for structure solution.
Keywords	Crystallography; molecular replacement; de novo modelling
URL	http://www.ccp4.ac.uk/ccp4online
Grid-enabled	No
Cloud-enabled	No
Total number of registered users	20 (since 7th Jan 2016)
Number of new users over 2015	0 (service started 2016)
Number of user submissions processed over 2015	0 (77 jobs in 2016)
Number of grid/cloud jobs over 2015	0
Key references	• Jaclyn Bibby, Ronan M. Keegan, Olga Mayans, Martyn D. Winn and Daniel J. Rigden (2012) "AMPLE: a cluster-and-truncate approach to solve the crystal structures of small proteins using rapidly computed ab initio models" *Acta Cryst.* **D68**, 1622-1631 • R.M.Keegan, J. Bibby, J. Thomas, D. Xu, Y. Zhang, O. Mayans, M.D.Winn and D.J.Ridgen, *Acta Cryst*., **D71**, 338-43 (2015) "Exploring the speed and performance of molecular replacement with AMPLE using QUARK ab initio protein models"

Appendix 1: Portal summary – CCP4 online - Balbes

Portal name	**CCP4 online - Balbes**
Short description	An automated Molecular Replacement (MR) pipeline - Balbes integrates into one system all the components necessary for solving a crystal structure by Molecular Replacement. Given structure factors and a sequence for the target, Balbes will search for models from an internal database (derived from the PDB). Checking the ARP/wARP checkbox will send Balbes' results to the ARP/wARP server.
Keywords	Crystallography, molecular replacement; structure database
URL	http://www.ccp4.ac.uk/ccp4online
Grid-enabled	No
Cloud-enabled	No
Total number of registered users	1155 (since Sept 2013)
Number of new users over 2015	728
Number of user submissions processed over 2015	3313
Number of grid/cloud jobs over 2015	0
Key references	• Fei Long, Alexei A. Vagin, Paul Young, and Garib N. Murshudov (2008) “BALBES: a molecular-replacement pipeline” Acta Cryst D64, 125–132. • R. M. Keegan, F. Long, V. J. Fazio, M. D. Winn, G. N. Murshudov and A. A. Vagin *Acta. Cryst.* **D67** , 313-323 (2011) "Evaluating the solution from MrBUMP and BALBES"

Appendix 1: Portal summary – CCP4 online - Crank2

Portal name	**CCP4 online – Crank2**
Short description	The CCP4 Crank2 server offers an automated structure solution pipeline for experimental phasing using maximum likelihood methods. The service covers SAD, SIRAS and MAD techniques. Unlike the traditional stepwise approach, the combined function simultaneously uses the information from density modification, model building and from the data to provide the best estimate of the electron density.
Keywords	Crystallography; experimental phasing
URL	http://www.ccp4.ac.uk/ccp4online
Grid-enabled	No
Cloud-enabled	No
Total number of registered users	10 (since 9th Jan 2016)
Number of new users over 2015	0 (service started 2016)
Number of user submissions processed over 2015	0 (47 jobs in 2016)
Number of grid/cloud jobs over 2015	0
Key references	• Skubak and Pannu (2013) “Automatic protein structure solution from weak X-ray data” *Nature Communications* **4**, 2777 • Pannu, N.S., Waterreus, W.J., Skubak, P., Sikharulidze, I, Abrahams, J.P., de Graaff, R.A.G. , (2011) “Recent advances in the CRANK software suite for experimental phasing”, *Acta Cryst*. **D67**, 331-337. • Ness, S.R., de Graaff, R.A., Abrahams, J.P., Pannu, N.S. , (2004) “CRANK: new methods for automated macromolecular crystal structure solution”, *Structure*, **12**, 1753-61.

Appendix 1: Portal summary – CCP4 online - MrBUMP

Portal name	**CCP4 online - MrBUMP**
Short description	The CCP4 MrBUMP server offers an automation pipeline for macromolecular structure solution by molecular replacement. There is a special emphasis on the discovery and preparation of a large number of search models, all of which can be passed to the core molecular-replacement programs Phaser or Molrep. Given a target sequence and experimental structure factors, it will search for homologous structures, create a set of suitable search models from the template structures, do molecular replacement, and test the solutions with some rounds of restrained refinement.
Keywords	Crystallography; molecular replacement; search models
URL	http://www.ccp4.ac.uk/ccp4online
Grid-enabled	No
Cloud-enabled	No
Total number of registered users	580 (since Oct 2014)
Number of new users over 2015	473
Number of user submissions processed over 2015	1302
Number of grid/cloud jobs over 2015	0
Key references	• R. M. Keegan, F. Long, V. J. Fazio, M. D. Winn, G. N. Murshudov and A. A. Vagin *Acta. Cryst.* **D67** , 313-323 (2011) "Evaluating the solution from MrBUMP and BALBES" • Keegan, R.M. and Winn, M.D. *Acta Cryst.* **D64**, 119-124 (2008) "MrBUMP: An automated pipeline for molecular replacement"

Appendix 1: Portal summary – CCP4 online - Shelx

Portal name	**CCP4 online - Shelx**
Short description	The CCP4 Shelx server offers an automated SHELXC/D/E structure solution pipeline for fast routine experimental phasing. Accepts data in XDS, Scalepack, SHELX hkl or mtz formats and outputs phases and a poly-Ala trace. If a protein sequence is provided, BUCCANEER and REFMAC complete the structure.
Keywords	Crystallography; experimental phasing; structure refinement
URL	http://www.ccp4.ac.uk/ccp4online
Grid-enabled	No
Cloud-enabled	No
Total number of registered users	19 (since 7th Jan 2016)
Number of new users over 2015	0 (service started 2016)
Number of user submissions processed over 2015	0 (91 jobs in 2016)
Number of grid/cloud jobs over 2015	0
Key references	• G. M. Sheldrick (2010) “Experimental phasing with SHELXC/D/E: combining chain tracing with density modification” *Acta Cryst*. **D66**, 479-485

Appendix 1: Portal summary – CCP4 online - Zanuda

Portal name	**CCP4 online - Zanuda**
Short description	The CCP4 Zanuda server offers a space group and crystallographic origin validation. The program Zanuda was developed to automate the validation of space group assignment. In addition, the program can be used to restore the correct space group in structures which were intentionally solved in low symmetry space groups including P1. The validation is based on the results of a series of refinements in space groups, which are compatible with the observed unit cell parameters.
Keywords	Crystallography, molecular replacement
URL	http://www.ccp4.ac.uk/ccp4online
Grid-enabled	No
Cloud-enabled	No
Total number of registered users	265 (since Aug 2014)
Number of new users over 2015	189
Number of user submissions processed over 2015	371
Number of grid/cloud jobs over 2015	0
Key references	• Lebedev AA and Isupov MN (2014) “Space-group and origin ambiguity in macromolecular structures with pseudo-symmetry and its treatment with the program Zanuda” *Acta Cryst*. **D70** 2430-43.

Appendix 1: Portal summary – PDB_REDO

Portal name	**PDB_REDO server**
Short description	The PDB_REDO server provides a fully automated procedure for optimizing crystallographic structure models. It is based on the PDB_REDO pipeline that combines standard crystallographic tools with state-of-the-art decision-making algorithms and dedicated model rebuilding programs. Extensive model validation is used to guide the decision-making and to report the results to the user. The pipeline is thoroughly tested by systematically applying it to all crystallographic structure models in the protein data bank (PDB). The resulting structure models are made available to the structural biology community through the PDB_REDO data bank.
Keywords	X-ray crystallography, refinement, validation, protein structure
URL	xtal.nki.nl/PDB_REDO
Grid-enabled	no
Cloud-enabled	no
Total number of new users	700 (active in 2015)
Number of new users over 2015	500
Number of user submissions processed over 2015	3000
Number of grid/cloud jobs over 2015	NA
Key references	• Joosten RP, Long F, Murshudov GN, Perrakis A. The PDB_REDO server for macromolecular structure model optimization. IUCrJ. 2014 May 30;1(Pt 4):213-20. • Joosten RP, Joosten K, Murshudov GN, Perrakis A. PDB_REDO: constructive validation, more than just looking for errors. Acta Cryst. 2012; D68:484-496. • Joosten RP, et al and Vriend G. PDB_REDO: automated re-refinement of X-ray structure models in the PDB. J. Appl. Cryst. 2009; 42:376-384.

Appendix 1: Portal summary – CCD

Portal name	**Crystallographic construct designer (CCD)**
Short description	CCD is a metaserver that collects predictions of secondary structure, disorder, membrane topology, from several web services to allow users to make an informed decision making a construct of a protein. It also designs primers for PCR amplification of the construct.
Keywords	Protein crystallography, crystallization, PCR primers
URL	xtal.nki.nl/ccd
Grid-enabled	NA
Cloud-enabled	NA
Total number of registered users	0 (no registration required)
Number of new users over 2015	NA
Number of user submissions processed over 2015	~3,000
Number of grid/cloud jobs over 2015	0
Key references	• Mooij WTM, Mitsiki E, Perrakis A. ProteinCCD: enabling the design of protein truncation constructs for expression and crystallization experiments. Nucl. Acids Res. (2009); 37:W402-W405.

Background information

This deliverable relates to WP5; background information on this WP as originally indicated in the description of work (DOW) is included below.

WP5 Title: Virtual Research Environment
Lead: Alexandre M.J.J. Bonvin (UU)
Participants: STFC, NKI, EMBL, MU, CSIC, CIRMMP, Instruct, UU, Luna, INFN

Work package number	5		**Start date or starting event:**		0
Work package title	Virtual Research Environment				
Activity Type	Support				
Participant number	**1**	**2**	**3**	**4**	**5**
Person-months per participant:	6	3	22	9	27
Participant number	**6**	**7**	**8**	**9**	**10**
Person-months per participant:	24	9	22	22	15

Objectives
This WP is centered on building and operating the VRE web portal that will provide the entry point for users, developers and all other stakeholders. We will build a web portal integrating all already existing and operating services from the various partners and the WeNMR Virtual Research Community (O5.1), and expand it to include new portals, training material and knowledge, and a support center (O5.2, O5.3). In order to better serve the community, customized end-user VMs and/or application containers (e.g. via Docker) will be built for various scenarios (O5.4), to be used on local infrastructures (e.g. within a company) or on the EGI federated cloud resources. Additionally, portals for newly identified applications will be developed and put in production during the project to increase the service portfolio of the VRE (O5.5). The list of objectives is thus:
• **O5.1**: Deployment and operation of the West-Life-VRE portal, integrating all relevant existing services, training and support components (from WeNMR and other partner sites) and extending them.
• **O5.2**: Establishment and operation of the West-Life-VRE support and expertise center for users and software developers, covering all VRE areas. This task will cooperate closely with the

relevant EGI-Engage Competence Centers (e.g. MoBrain).
• **O5.3**: Provision of information and training material covering all VRE areas and offered services.
• **O5.4**: Development and integration of new service portals.
• **O5.5**: Provision of customized end-users VMs and/or containers for various applications.

Description of work and role of participants

The above objectives will be addressed through the following tasks:

Task 5.1 – Deployment and operation of the West-Life portal (**Luna**, all)**.**
This task will directly address **O5.1**. It will start by defining the baseline of existing services across all partners (such as X-ray crystallography from CCP4 and the corresponding ones for cryoEM from the CSIC) together with those of the WeNMR VRC. The CSIC will contribute with the Web Services developed at the Instruct Image Processing Center in Madrid, making use of the Web interface of the SCIPION platform for software integration. These will then be integrated into a new VRE portal which will provide end users with a friendly and dynamical entry point to all services, knowledge and support center. The portal will be built on innovative technology developed by LUNA and we aim to migrate when possible existing portals to make direct use of the technology solutions offered by LUNA. In this task, we will also investigate and harmonize user authentication and authorization mechanisms (AAI) (e.g. both the Instruct and the WeNMR sites have user registration mechanisms in place, and WeNMR has implemented a single-sign-on (SSO) mechanism connected to Edugain). The choice and implementation of AAI mechanism will be done in close collaboration with EGI-Engage to maximize compatibility and impact. The new VRE portal will also implement tools and services related to data discovery and access (see WP6).

Task 5.2 – Knowledge and support center (**Instruct**, all)**.** This task will directly address **O5.2 and O5.3**. We will integrate the existing knowledge and support center of WeNMR, covering NMR and SAXS services into the new VRE portal, and add all the missing components (tutorials, use cases, help center) to support X-ray crystallography, cryo-electron microscopy and the related integrative methods. A choice will have to be made early on in the project for technology platform to build this knowledge and support center, since various existing components currently use different solutions (e.g. the Instruct web site is based on php while WeNMR operates on Drupal). As in Task 5.1, this will be done in close collaboration with the related EGI-Engage Competence Centers to minimize heterogeneity and maximize impact. Again, in this task, we will as much as possible built on the integrated solutions developed by LUNA.

Task 5.3 – Development and integration of new service portals (**UU**, all)**.** This task will directly address **O5.5**. While most of the existing WeNMR portals are already making use of the EGI Grid infrastructure with support from several NGIs within and outside Europe, this VRE project will be adding several portals that are already in place but depend on local and possibly limited resources, as is currently the case for most services for X-ray crystallography

and cryoEM. This task will interface those portals (and newly identified ones during the projects) to the most suited e- Infrastructure solution(s), being it grid, CLOUD or HPC resources. Note that we will benefit here from the interaction with various Competence Centers under the new EGI-Engage project, specially the MoBrain Competence Center, to which several partners of West-Life VRE participate (UU, CSIC, CIRMMP and STFC) . Care will be taken to offer user-friendly interfaces, with a VRE- integrated AAI. The most suited submission mechanisms will be selected. For example, we might adopt the efficient DIRAC4EGI service, but could also build on CLOUD and desktop grid (crowd computing) resources offered by the International Desktop Grid Federation (IDGF). A commercial service will also be offered by LUNA for users (both for profit and non-profit) requesting priority access to resources.

Task 5.4 – **Customized end-users VMs** (**STFC**, all). This task will directly address **O5.5**. Structural biology research has been targeting increasingly larger macromolecular machinery of the cell. Consequently, researchers need access to a wide range of techniques and expertise in order to truly exploit structural biology data. In most cases, however they are expert in only one or a few techniques and associated software. In this task we will build custom VMs for different use cases, with all the necessary software, documentation and examples. Thanks to their suitably designed customization, these VMs will be useful not only to expert structural biologists but also to researchers who want to exploit structural biology as a tool to gain insight in their biological/biomedical research. Different VM types and/or application containers (e.g. via Docker) will be provided, to allow use on both the EGI Federated Cloud and OpenStack/Nebula resources for example, but also local installation on a user's laptop (e.g. with VirtualBox and VMware). This will also potentially be an attractive mechanism for offering commercial services to companies, on their own internal infrastructure when IP issues are preventing external use.

Deliverables

No.	Name	Due month
5.1	Project portal	3
5.2	Overview (baseline) of services and portals to be integrated into the new VRE	4
5.3	Prototype of the new VRE portal functionality	6
5.4	Report on activities of the Helpdesk	18
5.5	VRE-integrated PDBe search and query API's	18
5.6	Report on available VMs with associated documentation/use case for each of them	24
5.7	Report on access and usage statistics of the various services	24
5.8	Report on access and usage statistics of the various services	36
5.9	Update Report on activities of the Helpdesk	36