The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes

G.C.P. van Zundert [1,†], **J.P.G.L.M. Rodrigues** [1,†], **M. Trellet** [2], **C. Schmitz** [3],
P.L. Kastritis [4], **E. Karaca** [4], **A.S.J. Melquiond** [5], **M. van Dijk** [6], **S.J. de Vries** [7] and
A.M.J.J. Bonvin [1]

1 - Bijvoet Center for Biomolecular Research, Faculty of Science Department of Chemistry, Utrecht University, Domplein 29, 3512 JE
Utrecht, the Netherlands
2 - Centre National de la Recherche Scientifique Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur,
rue John Von Neumann, 91403 Orsay, France
3 - Instaclustr Level 5, 1 Moore Street, Canberra ACT 2600, Australia
4 - European Molecular Biology Laboratory Heidelberg, Meyerhofstrasse 1, 69117 Heidelberg, Germany
5 - Hubrecht Institute, Uppsalalaan 8, 3584 CT Utrecht, the Netherlands
6 - Faculty of Sciences, VU University Amsterdam, De Boelelaan 1083, 1081 HV Amsterdam, the Netherlands
7 - Physik-Department, Technische Universität München, James-Franck-Strasse 1, 85748 Garching, Germany

Correspondence to A.M.J.J. Bonvin:
http://dx.doi.org/10.1016/j.jmb.2015.09.014
Edited by M. Sternberg

Abstract

The prediction of the quaternary structure of biomolecular macromolecules is of paramount importance for fundamental understanding of cellular processes and drug design. In the era of integrative structural biology, one way of increasing the accuracy of modeling methods used to predict the structure of biomolecular complexes is to include as much experimental or predictive information as possible in the process. This has been at the core of our information-driven docking approach HADDOCK. We present here the updated version 2.2 of the HADDOCK portal, which offers new features such as support for mixed molecule types, additional experimental restraints and improved protocols, all of this in a user-friendly interface. With well over 6000 registered users and 108,000 jobs served, an increasing fraction of which on grid resources, we hope that this timely upgrade will help the community to solve important biological questions and further advance the field. The HADDOCK2.2 Web server is freely accessible to non-profit users at http://haddock.science.uu.nl/services/HADDOCK2.2.

Introduction

Cellular metabolism is a highly regulated and adaptive system where proteins, the main participants, form a vast network of interactions collectively known as the interactome. Knowledge of the three-dimensional atomic structure of protein–protein interactions is therefore critical for a fundamental understanding of cellular and molecular biology, as well as for rational drug design. Unfortunately, solving such structures using classical high-resolution methods (X-ray crystallography and NMR spectroscopy) is not trivial, as each has its own limitations (e.g., protein flexibility, size and strength of the interaction).

Considering the magnitude of the interactome, complementary high-throughput methods such as computational docking are necessary if we aim to close the structure gap [1]. The goal of protein–protein docking is to predict the structure of a complex starting from the individual structures of its components [2], which can be either experimentally determined or predicted [3].

Despite continuous advances in the field, the accuracy of *ab initio* docking—without using any experimental restraints—remains generally low [4]. Data-driven approaches such as HADDOCK [5,6], which integrate information derived from biochemical, biophysical or bioinformatics methods to enhance sampling, scoring or both [2], perform remarkably

better. The information that can be integrated is quite diverse: interface restraints from NMR, mutagenesis experiments or bioinformatics predictions [7,8]; shape data from small-angle X-ray scattering [9] and cryo-electron microscopy experiments [10]; and orientations of the individual structures in the complex from NMR residual dipolar couplings [11], relaxation anisotropy [12] and pseudocontact shifts experiments [13]. The potential of data-driven docking is reflected in the success of the HADDOCK server and software in recent CAPRI experiments (Critical Assessment of Protein Interaction) [14,15], as well as in the number of structures deposited (>120) in the Protein Data Bank (PDB), which were calculated using our software.

Five years ago, we introduced the HADDOCK Web server to provide a user-friendly interface to the software and streamline its usage by non-expert users in the structural biology field [16]. Shortly after, it was updated to handle multi-body docking [17]. The development of new and improved protocols and the inclusion of additional sources of restraints culminated in the recently released version 2.2 of the software, followed by an update of the Web server interfaces. Throughout the next section, we will provide an overview of the newly updated HADDOCK Web server and discuss the most relevant additions[‡]. We conclude by presenting usage statistics of the server to demonstrate the usefulness and power of providing easy and free access to scientific software.

Overview and advances

The HADDOCK Web server was created to facilitate the use of our docking software, by removing the burden of its installation and setup, as well as by providing validation routines for input data and options. In addition, since HADDOCK runs are computationally demanding, the Web server offers the users access to sufficient resources—our local clusters—to complete their runs within a few hours. A grid-enabled version of the server can be accessed via the WeNMR Web site[§] [18], which uses resources provided by the European Grid Initiative (EGI) and the associated National Grid Initiatives. This setup, which currently handles most submissions, provides more than 110,000 CPU cores distributed over 41 sites worldwide[‖].

The HADDOCK Web server aggregates seven different interfaces, each associated with a different level of control over the docking protocol reflected by the number of parameters that can be changed. New users are granted access to the *Easy* and the associated *Prediction Interface* only, but they can request access to the *Expert* and *Guru* levels and their associated interfaces, if necessary.

The *Easy* interface provides the most basic level of control. It allows the user to either upload two structures in PDB format or download them directly from the Research Collaboratory for Structural

Bioinformatics PDB, and it allows the user to define sets of active and passive residues that represent the (putative) interface. Unlike previous versions, HADDOCK 2.2 supports single (protein, small molecule, RNA or DNA) and mixed (protein–DNA, protein–RNA) molecule types. This was implemented to handle the docking of proteins onto a nucleosome complex—a recent CAPRI target.

The *Expert* interface builds on the *Easy* interface and allows the user to manually specify the protonation state of each histidine residue in the proteins, which is otherwise determined automatically with MolProbity [19]. Also, it offers control over which regions of the molecules are semi-flexible and fully flexible segments, which has an impact during the refinement stage of the docking. Lastly, the user is given the option to define the charge state of the N- and C-terminus of the protein. The *Expert* interface also provides a *Distance Restraints* section, where the user has the option not only to upload user-defined ambiguous and unambiguous restraint files and/or use center-of-mass restraints, useful for blind or *ab initio* docking when no other information is available, but also to ensure compactness of the generated models. The center-of-mass restraints are automatically generated by calculating the dimensions of each molecule along the *x*-axis, *y*-axis and *z*-axis (d_x, d_y, d_z) and summing the average of the two smallest components per molecule. The resulting distance is used to define a restraint between the center of mass of each subunit with an additional upper bound correction of 1 Å [9]. In addition, the *Expert* interface gives control over the *Sampling Parameters*, including namely the number of structures to generate at each stage and whether or not to perform solvated docking [20,21,23]. Finally, it exposes the *Clustering Parameters* that define the clustering algorithm and cutoff. In version 2.2, in addition to RMSD-based clustering, there is the option of using the Fraction of Common Contacts clustering algorithm [23], which is significantly faster and especially useful for symmetric complexes.

The *Guru* interface gives full access and control to ~500 parameters, nearly all that are available in HADDOCK. The *Distance Restraints* section now offers a new radius of gyration restraint, information that can be extracted, for example, from small-angle X-ray scattering experiments. *Non-crystallographic Symmetry Restraints* and *Symmetry Restraints* are also available at this level and have been extended to handle *C*4 and *D*2 symmetries in addition to the already available *C*2, *C*3 and *C*5 symmetries. There are additional sections for other types of NMR-based restraints, such as *Residual Dipolar Couplings* [11], *Relaxation Anisotropy* [12] and the recently added *Pseudo Contact Shifts* [13]. These latter require a tensor distance restraint file and the definition of the rhombic and axial components of the anisotropic

tensor. Besides the restraints, all the energy evaluations, scoring functions and analysis parameters can be tweaked; advanced parameters for the sampling protocols are also available at this level, offering a greater degree of control, for example, on the extent of each refinement stage. There are also dedicated options to the solvated docking protocol, which now uses by default propensities based on the Kyte–Doolittle hydrophobicity scale, as these have been shown to improve the protocol [21]. The original statistical-based propensities [20], recently expanded to include nucleotides [22], can still be selected via a dropdown menu.

The remaining four interfaces consist of the *Prediction Interface*, which is similar to the *Easy* interface but with settings geared toward using bioinformatics interface predictors such as CPORT [7]; the *Refinement Interface* (expert-level access), which runs only the water refinement stage on the uploaded structures and can be used for scoring purposes; the *Multi-body Interface*, based on the *Guru* interface, supporting upload of up to six molecules that will be docked simultaneously [17] and also featuring the *Molecule Interaction Matrix* section. This new addition displays a table with scaling factors to adjust the interaction forces between different subunits,

allowing molecules to become invisible to each other during the docking, which is useful in cases where multiple binding modes are required to satisfy the experiment data (see, e.g., Ref. [24]). The use of ambiguous interaction restraints within this interface requires the user to upload a restraints table file in the *Distance Restraints* section This requirement of uploading distance restraint files instead of supplying residue lists as in the other interfaces was meant to make users think carefully about their system since, in multi-body docking, multiple interfaces will be defined that might not all be supposed to interact. To facilitate the creation of custom ambiguous interaction restraint files between any numbers of molecules, we created an interface called *Gentbl*. Finally, the Web server also offers a *File Upload Interface* to allow the user to upload a run parameter file, created upon successful validation and submission to the queue, and thus easily redo a docking run or re-run it with slight changes in the parameters.

At submission time, once the input data have been properly validated, the server offers the option to download a parameter file and provides a link to the results page, which is also e-mailed to the user. Users are encouraged to save the parameter file since it contains all required input data and settings
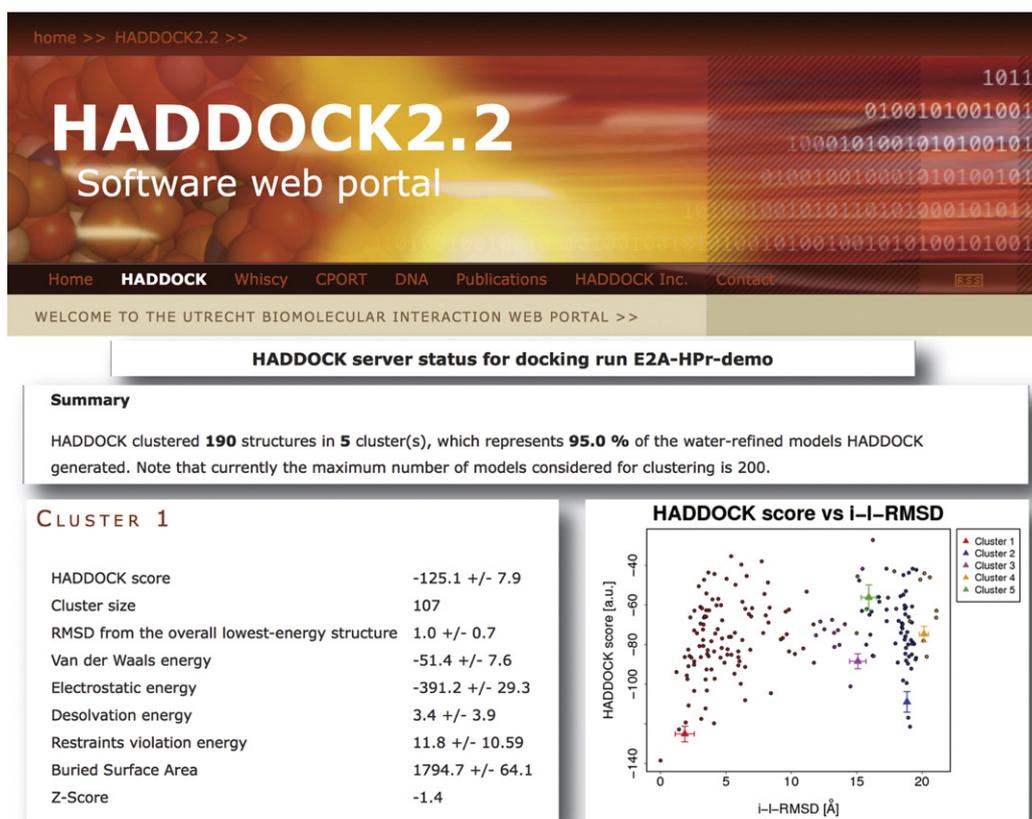

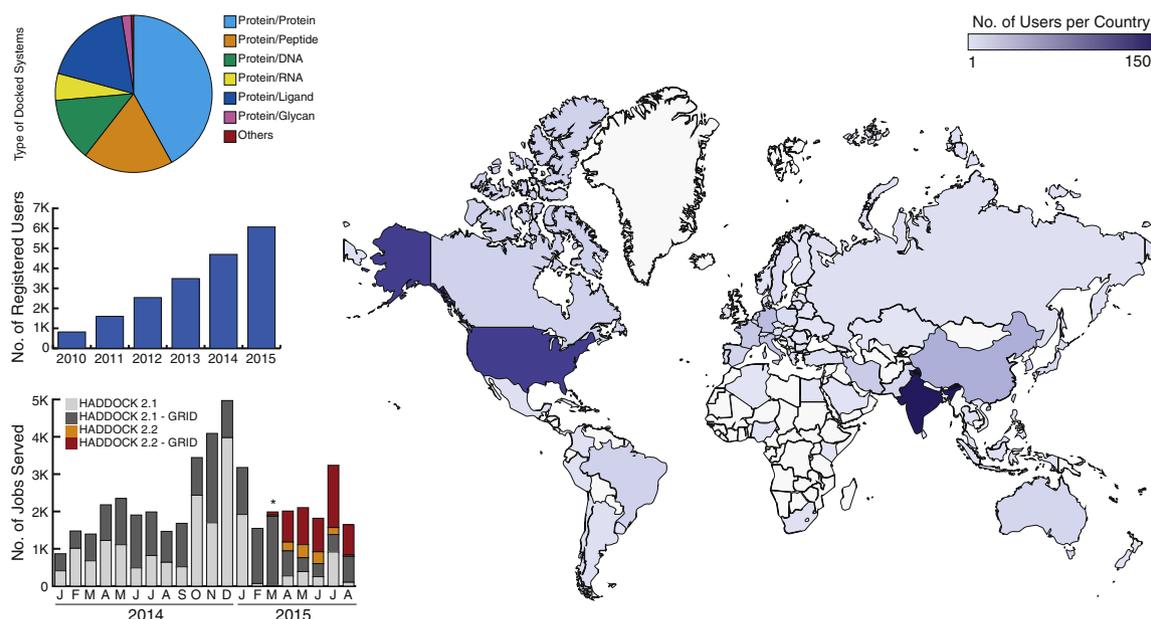**Fig. 1.** Excerpts from an example result page of a HADDOCK2.2 docking run.

Fig. 2. HADDOCK Web server statistics (as of September 1, 2015). The worldwide distribution of registered users (>6000) is shown with a blue-scale colorcoding. On the top right, the distribution of the type of systems docked using HADDOCK is shown (taken from a HADDOCK online survey with >325 respondents). The middle bar graph shows the yearly increase in the number of registered users since 2010 and the bottom plot shows the number of processed user runs per month since last year (2014), distinguishing between runs that have been processed on our local resources or on the EGI grid resources and between the previous and current (2.2) version of the server.

to reproduce the docking, as recommended in the "Outcome of the First wwPDB Hybrid/Integrative Methods Task Force Workshop" (Recommendation 1) [25]. The results page allows monitoring of the progress of the docking run. After a successful docking run, the user will receive another e-mail redirecting him/her to the updated results page (see Fig. 1 for an excerpt of presented results). The page indicates how many structures of the water-refined models could be clustered and lists the clusters in the order of their HADDOCK score. For each cluster, detailed statistics are displayed, representing the average values calculated over the top four best-scoring structures within each cluster. Besides the HADDOCK score and other standard energies (van der Waals, etc.), a *z*-score has been added. The *z*-score represents how many standard deviations the HADDOCK score of a given cluster is separated from the mean of all clusters, that is, the lower the *z*-score is, the better. To visualize the results, we display plots at the bottom of the results page, showing the HADDOCK score of all solutions against the interface–ligand–RMSD compared to the best-scoring structure for example, together with cluster averages and their spreads.

Usage statistics

Since its opening in June 2008, the HADDOCK Web server has seen a sustained increase in the number of registrations to reach over 6000 registered users to date distributed all over the world (Fig. 2). More than 108,000 runs have been processed, 28% of which have run on EGI grid resources. This percentage has increased to 75% for the HADDOCK2.2 server submission. An overview of the number of runs processed per month with their distribution over local and grid resources is shown in Fig. 2. Since the launch of the HADDOCK2.2 Web server in March 2015, an increased fraction of runs is handled by the new 2.2 portal. Statistics over the last year (since January 2014) indicate that the portal is processing on average 75 docking runs per day. The execution wall time averaged over both local cluster and grid resources is around 16 h per HADDOCK run, as approximately 75 runs are handled per day with up to 50 jobs running in parallel (and a maximum of 5 concurrent jobs per user). The exact run time depends on the size of the system being docked and the parameter settings and can vary between half an hour and several days. The server home page reports both the number of running and pending jobs, allowing users to get an estimate of the waiting time. A majority of the docking runs are dealing with protein–protein and protein–peptide docking (~61%), ~19% correspond to protein–nucleic acids systems and quite a significant fraction (~20%) is dealing with protein–small molecule docking (both small ligand and oligosaccharides). These numbers demonstrate the

popularity and widespread usage (in terms of both geographic distribution and type of systems being studied) of our HADDOCK Web server.

References

[1] D. Petrey, B. Honig, Structural bioinformatics of the interactome, Annu. Rev. Biophys. 43 (2014) 193–210.

[2] J.P.G.L.M. Rodrigues, A.M.J.J. Bonvin, Integrative computational modeling of protein–protein interactions, FEBS J. 281 (2014) 1988–2003.

[3] J.P.G.L.M. Rodrigues, A.S.J. Melquiond, E. Karaca, M. Trellet, M. van Dijk, G.C.P. van Zundert, C. Schmitz, S.J. de Vries, A. Bordogna, L. Bonati, P.L. Kastritis, A.M.J.J. Bonvin, Defining the limits of homology modeling in information-driven protein docking, Proteins 81 (2013) 2119–2128.

[4] S.-Y. Huang, Exploring the potential of global protein–protein docking: An overview and critical assessment of current programs for automatic *ab initio* docking, Drug Discov. Today 8 (2015) 969–977.

[5] C. Dominguez, R. Boelens, A.M.J.J. Bonvin, HADDOCK: A protein–protein docking approach on biochemical or biophysical information, J. Am. Chem. Soc. 125 (2003) 1731–1737.

[6] S.J. De Vries, A.D.J. van Dijk, M. Krzeminski, M. van Dijk, A. Thureau, V. Hsu, T. Wassenaar, A.M.J.J. Bonvin, HADDOCK *versus* HADDOCK: New features and performance of HADDOCK2.0 on the CAPRI targets, Proteins 69 (2007) 726–733.

[7] S.J. De Vries, A.M.J.J. Bonvin, CPORT: A consensus interface predictor and its performance in prediction-driven docking with HADDOCK, PLoS One 6 (2011)http://dx.doi.org/10.1371/journal.pone.0017695.

[8] T.A. Hopf, C.P.I. Schärfe, J.P.G.L.M. Rodrigues, A.G. Green, O. Kohlbacher, C. Sander, A.M.J.J. Bonvin, D.S. Marks, Sequence co-evolution gives 3D contacts and structures of protein complexes, Elife 3 (2014)http://dx.doi.org/10.7554/eLife.03430.

[9] E. Karaca, A.M.J.J. Bonvin, On the usefulness of ion-mobility mass spectrometry and SAXS data in scoring docking decoys, Acta Crystallogr. D Biol. Crystallogr. 69 (2013) 683–694.

[10] G.C.P. Van Zundert, A.S.J. Melquiond, A.M.J.J. Bonvin, Integrative modeling of biomolecular complexes: HADDOCKing with cryo-electron microscopy data, Structure 23 (2015) 949–960.

[11] A.D.J. Van Dijk, D. Fushman, A.M.J.J. Bonvin, Various strategies of using residual dipolar couplings in NMR-driven protein docking: Application to Lys48-linked di-ubiquitin and validation against ^{15}N-relaxation data, Proteins 60 (2005) 367–381.

[12] A.D.J. Van Dijk, R. Kaptein, R. Boelens, A.M.J.J. Bonvin, Combining NMR relaxation with chemical shift perturbation data to drive protein–protein docking, J. Biomol. NMR 34 (2006) 237–244.

[13] C. Schmitz, A.M.J.J. Bonvin, Protein–protein HADdocking using exclusively pseudocontact shifts, J. Biomol. NMR 50 (2011) 263–266.

[14] J. Janin, Assessing predictions of protein–protein interaction: The CAPRI experiment, Protein Sci. 14 (2005) 278–283.

[15] M.F. Lensink, S.J. Wodak, Docking, scoring and affinity prediction in CAPRI, Proteins 81 (2013) 2082–2095.

[16] S.J. De Vries, M. van Dijk, A.M.J.J. Bonvin, The HADDOCK Web server for data-driven biomolecular docking, Nat. Protoc. 5 (2010) 883–897.

[17] E. Karaca, A.S.J. Melquiond, S.J. de Vries, P.L. Kastritis, A.M.J.J. Bonvin, Building macromolecular assemblies by information-driven docking: Introducing the HADDOCK multibody docking server. Moll, Cell. Proteomics 9 (2010) 1784–1794.

[18] T. Wassenaar, M. van Dijk, N. Loureiro-Ferreira, G. van der Schot, S. de Vries, C. Schmitz, J. van der Zwan, R. Boelens, A. Giachetti, L. Ferella, A. Rosato, I. Bertini, T. Herrmann, H. Jonker, A. Bagaria, V. Jaravine, P. Güntert, H. Schwalbe, W.

Vranken, J. Doreleijers, G. Vriend, G. Vuister, D. Franke, A. Kikhney, D. Svergun, R. Fogh, J. Ionides, E. Laue, C. Spronk, Jurkša, M. Verlato, S. Badoer, S. Dal Pra, M. Mazzucato, E. Frizziero, A.M.J.J. Bonvin, WeNMR: Structural Biology on the Grid, J. Grid Comput. 10 (2012) 743–767.

[19] V.B. Chen, W.B. Arendall 3rd, J.J. Headd, D.A. Keedy, R.M. Immormino, G.J. Kapral, L.W. Murray, J.S. Richardson, D.C. Richardson, MolProbity: All-atom structure validation for macromolecular crystallography, Acta Crystallogr. D Biol. Crystallogr. 66 (2010) 12–21.

[20] A.D.J. Van Dijk, A.M.J.J. Bonvin, Solvated docking: Introducing water into the modelling of biomolecular complexes, Bioinformatics 22 (2006) 2340–2347.

[21] P.L. Kastritis, K.M. Visscher, A.D.J. van Dijk, A.M.J.J. Bonvin, Solvated protein–protein docking using Kyte-Doolittle-based water preferences, Proteins 81 (2013) 510–518.

[22] M. Van Dijk, K. Visscher, P.L. Kastritis, A.M.J.J. Bonvin, Solvated protein–DNA docking using HADDOCK, J. Biomol. NMR 56 (2013) 51–63.

[23] J.P.G.L.M. Rodrigues, M. Trellet, C. Schmitz, P. Kastritis, E. Karaca, A.S.J. Melquiond, A.M.J.J. Bonvin, Clustering biomolecular complexes by residue contacts similarity, Proteins 80 (2012) 1810–1817.

[24] E. Escobar-Cabrera, M. Okon, D.K.W. Lau, C.F. Dart, A.M.J.J. Bonvin, L.P. McIntosh, Characterizing the N- and C-terminal small ubiquitin-like modifier (SUMO)-interacting motifs of the scaffold protein DAXX, J. Biol. Chem. 286 (2011) 19816–19829.

[25] A. Sali, H.M. Bergman, T. Schwede, J. Trewhella, G. Kleywegt, S.K. Burley, J. Markley, H. Nakamura, P. Adams, A.M.J.J. Bonvin, W. Chiu, M.D. Peraro, F. Di Maio, T.E. Ferrin, K. Grünewald, A. Gutmanas, R. Henderson, G. Hummer, K. Iwasaki, G. Johnson, C.L. Lawson, J. Meiler, M.A. Marti-Renom, G.T. Montelione, M. Nilges, R. Nussinov, A. Patwardhan, J. Rappsilber, R.J. Read, H. Saibil, G.F. Schröder, C.D. Schwieters, C.A.M. Seidel, D. Svergun, M. Topf, E.L. Ulrich, S. Velankar, J.D. Westbrook, Outcome of the First wwPDB Hybrid/Integrative Methods Task Force Workshop, Structure 7 (2015) 1156–1167.