INTRODUCTION

The Protein Data Bank (PDB)1,2 archive is a rich repository of data and information on the structure and function of biologically relevant macromolecules and their complexes. The archive currently contains over 84,500 entries (referencing over 28,000 unique UniProt3 accession codes), of which almost 10,000 NMR-derived structures (almost 5000 unique UniProt codes, Table I). The PDB archive is managed by the Worldwide Protein Data Bank organization (wwPDB),4 which consists of the Research Collaboratory for Structural Bioinformatics (RCSB)5 and the BioMagResBank (BMRB)6 in the USA, the Protein Data Bank Japan (PDBj),7 and the Protein Data Bank in Europe (PDBe; http://pdbe.org/).8 Atomic models deposited to the PDB are curated by RCSB, PDBe, and PDBj, while experimental NMR data are processed by BMRB. As one of the founding wwPDB partners, PDBe provides annotated data on three-dimensional (3D) structures of biomacromolecules to the scientific community as well as advanced services based on these structures.

As the PDB archive continues to grow, the provision of adequate validation tools to its users becomes increasingly important.9 In many cases, the archive contains multiple structures of the same or similar molecules, for example, solved by competing groups or by different experimental techniques, with different ligands or containing mutations. Choosing the most appropriate structure in such cases is not an easy task for experts, let alone for non-specialist users. Even if only a single structure is known of a molecule of interest, it is still essential to assess whether the structure is suitable for the intended use.

A number of validation software packages are available,10–12 but a single robust and widely-accepted validation score or standard set of validation criteria is yet to be defined for NMR-derived biomacromolecular structures. Therefore, a wwPDB NMR validation task force (VTF) has been convened to define standard validation criteria that will be applied by all wwPDB partners to all...
Table 1

Data Coverage

| Source         | PDB archive | NMR entries |
|----------------|-------------|-------------|
| Total number of entries | PDB 1,2     | 84,508   9616 |
| Unique UniProt accession codes referenced | SIFTS 49    | 29,104   4022 |
| Pfam 50 sequence families referenced | SIFTS       | 6143     2455 |
| CATH 51 domain architectures present | SIFTS       | 2549     660  |
| SCOP 52 domain architectures present | SIFTS       | 4191     1150 |
| Cluster analysis | OLDERADO 14 | –         7289 |
| Chemical shift analysis | VASCO 28    | –         3361 |
| Distance constraints | BMRB-NRG 38 | –         5978 |
| Dihedral constraints | BMRB-NRG    | –         3850 |
| RDCs | BMRB-NRG 660 |  602 |
| Validation scores | NRG-CING 47 | –         9491 |

Coverage of the protein universe in the PDB, and data available in Vivaldi for NMR entries (as of September 12, 2012). Up to date data coverage statistics can be found at http://pdbe.org/nmrsstats/.

depositions of NMR structures (http://www.wwpdb.org/workshop/2010/nmr_validation.html), in analogy with the wwPDB X-ray VTF 13 In an effort to make information about the quality of NMR structures in the PDB accessible to the wider scientific community, PDBe has developed an interactive web-based tool called Vivaldi (Visualization and Validation Display; http://pdbe.org/vivaldi; Fig. 1).

MATERIALS AND METHODS

Several types of information are accessible through the Vivaldi service, including information about the homogeneity of the modeled structure ensemble, validation of the available experimental NMR data, and knowledge-based validation of the modeled conformations. The decisions on which sources of data and which validation criteria to include in Vivaldi were based on three factors: (a) accessibility, (b) the sources being well-established and published, and (c) data coverage of the PDB archive. In the future, Vivaldi will be modified to accommodate the recommendations of the wwPDB NMR VTF.

Typically, an NMR entry in the PDB archive contains 20 models of a macromolecule or a complex. For all intents and purposes, these are deposited in an arbitrary order and therefore no single model can be assumed to be more important than the others without analyzing the structure ensemble in more detail. One such type of analysis is performed by OLDERADO 14 which is used by PDBe to determine core domains within proteins and to cluster together individual models of the ensemble. OLDERADO identifies the most representative model of each cluster and the most representative model of the entire ensemble. With the Vivaldi web service, users can interactively explore the results of this cluster and domain analysis, overlay different models and visualize similarities and differences within and between clusters. Furthermore, the most representative model as identified by OLDERADO is used throughout Vivaldi as the one displayed by default (rather than, for instance, the first model of the ensemble). Supporting Information Figure S1(a) shows representative cluster models as identified by OLDERADO for PDB entry 2k4v 15 while Supporting Information Figure S1(b) highlights the domains on the most representative model of the ensemble.

To help users assess the local variability in the deposited ensemble of structures, Vivaldi calculates a simple dihedral order parameter 16 $S^2$, for each residue in protein and nucleic acid molecules [Eq. (1)] and presents this information as a graph of $S^2$ vs. residue number [Supporting Information Fig. S1(c)]. Selecting a residue in the graph displays a table with the dihedral angle values, pie-charts for the relevant dihedral angles and, if appropriate, a Ramachandran plot [Supporting Information Fig. S1(d)].

$$S^2 = \frac{1}{MN^2} \sum_i \left( \sum_j \cos \gamma_{ij} \right)^2 + \left( \sum_j \sin \gamma_{ij} \right)^2,$$

where $\gamma$ are backbone dihedral angles ($\phi$ and $\psi$ for proteins and $\omega$- $\zeta$ for nucleic acids), index $i$ runs over the $M$ different dihedral angles and index $j$ runs over the $N$ conformers of the ensemble. Values of $S^2$ close to 1 indicate that the dihedral angles show little variation across the ensemble.

Chemical shifts are arguably the most studied NMR parameters with respect to the covalent structure, local conformations, and immediate spatial surroundings. The complex dependence of the chemical shift on an atom’s chemical environment has resulted in a wide variety of shift prediction and validation tools based on different approaches (e.g., quantum-chemical calculations and neural networks) and taking into account different aspects of the chemical environment (e.g., solvent effects, aromatic ring currents, and hydrogen bonds). Some of the more recent software packages include AVS 17, CamShift 18, LACS 19,20 TALOS 21,22,23 SPARTA+ 24,25 CheShift 26, DANGLE 27, and VASCO 28. As deposition of chemical shift information has been mandatory for NMR structures since December 2010, chemical shifts are becoming an increasingly important source of data for validating PDB structures.

For Vivaldi, VASCO 28 validates experimental chemical shifts of proteins in a two-step process. First, the optimal referencing offset for each nucleus type ($^1$H, $^{15}$N, $^{13}$C, $^{13}$C, $^{13}$C, $^{13}$C) is calculated by comparing the reported chemical shifts to a set of protein chemical shifts that were correctly referenced by hand. Second, for each atom a database of chemical shifts is queried to identify atoms in a similar environment, defined as a
Figure 1

Layout of a Vivaldi page showing the default view for protein PA1076 from Pseudomonas aeruginosa (PDB entry 2k4v).15 The page contains a header including PDBprints,9 an interactive 3D viewer (OpenAstexViewer),48 a graph and a textual information section. The most representative model according to OLDERADO14 cluster analysis is displayed in the 3D viewer and the graph shows the NRG-CING47 red-orange-green (ROG) scores for this protein.
combination of atom and residue type, secondary structure, and solvent accessibility. If a sufficient number of similar atoms are found, the $Z$-score of the reported chemical shift is calculated and if the absolute $Z$-score exceeds a value of 3 (99.7% confidence interval assuming a true normal distribution) the shift is labeled as an outlier.

Vivaldi shows these outliers as solid spheres in its 3D viewer and as bars in the interactive graph [Fig. 2(a,b)] with default coloring for both the spheres and the bars varying smoothly from green ($Z$-score < 2) via yellow ($Z$-score = 3.5) to red ($Z$-score > 5) depending on the degree of the deviation from the expected range. Aromatic ring currents, which can have a profound influence...
on chemical shifts, are not explicitly taken into account by VASCO. Therefore, aromatic rings in the structure are highlighted in the 3D viewer as stick models. Furthermore, the user can select a residue by clicking on a bar in the graph or on an atom in the 3D viewer and thereby obtain a table of the original chemical shift values and the VASCO scores for the nuclei in that residue.

Distance and dihedral angle constraints, together with residual dipolar couplings (RDCs) constitute the bulk of the experimental data used during structure calculation for the majority of NMR-derived macromolecular structures in the PDB. Although these types of experimental data are used directly for structure calculation (hence, the models are expected to fit them rather well), validation can still pinpoint problems. The Queen29 and PSVS10 validation suites provide analyses of distance constraints, whereas PALES30 REDCAT,31 and MOD-ULE232 can be used to analyze RDC data.

Vivaldi retrieves pre-processed CCPN projects,33 which include all the constraints, from the Database Of Converted Restraints (DOCR),34 which is part of the NMR Restraints Grid (NRG)35,36 at BMRB. Distance constraints are further processed to classify them into groups, that is, intra-residual, short (1 or 2 residues apart in sequence), medium (3 or 4 residues apart), and long-range (5 or more residues apart), as well as to identify constraint violations for each group. Vivaldi can show violated and satisfied constraints, graphically in charts and in the 3D viewer as well as in tables [Fig. 2(c,d)]. Any individual constraint can be selected to show the atoms involved, the actual inter-atomic distance(s) from the displayed model(s) and the constraint’s lower and upper limit. Weighted summations (\(d^{-6}\)) were used for constraints involving ambiguous assignments or magnetically equivalent groups (e.g., methyl groups).37

Dihedral-angle constraints are commonly obtained from an analysis of backbone chemical shifts by programs such as TALOS+21 and DANGLE27 or, less frequently today, from experimental measurements of scalar coupling constants. As with distance constraints, Vivaldi presents these constraints and their violations interactively in the 3D viewer and as graphs. Selecting a residue with reported dihedral angle constraints will display a table as well as a Ramachandran plot and/or pie-charts for each dihedral angle as applicable, highlighting the constraint limits and angles calculated from the displayed models in a fashion similar to that shown in Supporting Information Figure S1(d).

RDC38,39 constraints are obtained from the same resource at BMRB as the distance and torsion angle constraints. Alignment tensors, which are generally not deposited together with the RDC constraints, are calculated locally by fitting one tensor per model in the NMR ensemble per alignment medium. If multiple RDC types have been deposited for a single alignment medium (e.g., N–HN, Cα–Hα, N–Co), Vivaldi checks whether the RDC constraints were deposited as their original experimental values or as scaled values to match N–HN RDC values. The optimal alignment tensor is obtained by minimizing the RMSD between the experimental and recalculated RDC values for the alignment tensor.

The Vivaldi 3D viewer can show the principal axes of the fitted alignment tensor as three orthogonal bars tied to the molecule’s orientation in the viewer and the RDC constraints are visualized by showing the atoms involved as spheres [Fig. 2(e)]. Further, fitted and experimental RDC values are presented as bar and scatter plots [Fig. 2(f)]. Details about the fitted tensor, that is tensor magnitude (\(D_a\)) and rhombicity (\(R\)), as well as three values relating to the goodness-of-fit, that is RMS deviation (Hz), Pearson R-value, and Cornilescu Q-value,40 are shown in a separate information section. When a residue is selected from a bar chart or in the 3D viewer, experimental and fitted values of all corresponding RDC constraints are listed.

Empirical knowledge about the structure of biological macromolecules can be used to validate local conformations and the overall fold. A number of programs, such as WHATIF41 PROCHECK,42 WHATCHECK,43 and Verify3D44 can be used to assess the quality of macromolecular structures. More recent software packages including PSVS,10 ProSA-web,45 MolProbity,46 and CING12,47 combine and extend them to gain further insight into the quality of an atomic structure model.

Residue-based quality scores are extracted from the external NRG-CING database, including a general ROG score that labels residues as red, orange, or green depending on a series of quality measures. Several PROCHECK42 and WHATIF41 scores are also shown by Vivaldi. As for any type of information, residue-based scores can be visualized as a graph, in a table or in the 3D viewer. Figure 1 shows the CING ROG score for PDB entry 2k4v.15

For any NMR entry in the PDB, Vivaldi can be launched from the corresponding entry page at PDBee or from a direct shortcut URL (http://pdbe.org/vivaldi/NNNN where NNNN is the four-character PDB code of the entry). Alternatively, the homepage of the NMR resource at PDB (http://pdbe.org/nmr) provides a quick-access form into which the PDB code of interest can be entered. Finally, Vivaldi can be accessed from its own webpage (http://pdbe.org/vivaldi) and the information that is displayed can be customized by means of a user-friendly wizard. On the Vivaldi page of a PDB entry, the available information can be accessed, analyzed and visualized in a multitude of ways. High-quality images from the 3D viewer and the interactive graphs can be saved for use in publications, presentations, or teaching resources. The amount and type of information that is available for a particular entry depends on what type of experimental data (if any) was deposited, whether an ensemble or a single model was deposited, and so forth.
Table I lists all the types of information presented in Vivaldi, its sources, and the number of PDB entries for which it is available. All available data for each entry is assembled and presented in a consistent and intuitive way that makes it useful to both NMR experts and non-expert users.

Generating the information used by Vivaldi involves collection and format conversion of relevant data on the one hand and data presentation on the other. The first aspect is dealt with by automated Python scripts that carry out weekly checks for new, updated or removed NMR entries in the PDB and for updated validation reports. These scripts import, check, and match data in different formats and ensure their internal consistency. This is necessary as curation and annotation of coordinates and constraints is performed separately in current wwPDB curation practice. Finally, the gathered data are output as objects in JSON files, which are used for presentation to the end user. A modified version of the OpenAstexViewer 3.0 Java applet accommodating some of the specific features of Vivaldi is used as 3D display engine, while the Flotr JavaScript Plotting Library (http://solutoire.com/flotr) is used to generate charts and graphs.

RESULTS AND DISCUSSION

Table I summarizes the types of data that can be encountered for NMR structures and the number of PDB entries for which each type of data is available in Vivaldi at the time of manuscript submission (weekly updated statistics are available from the PDBe website http://pdbe.org/nmrstats). While all NMR entries can be viewed...
in Vivaldi and almost all have at least a geometric validation report in the NRG-CING database, only about 60% have associated experimental constraints data deposited and processed for further analysis or visualization and only about a third have VASCO reports on chemical shift outliers at present. These numbers are set to improve, since deposition of both the experimental chemical shifts and constraints data is now mandatory (http://wwpdb.org/policy.html).

The major motivation for developing Vivaldi is to provide the users of the PDB archive (expert NMR spectroscopists and non-experts alike) with an easily accessible way to view and analyze NMR structures, their associated experimental data, and validation-related information. To demonstrate its capabilities, we present two use cases that demonstrate how Vivaldi can help evaluate the suitability of a particular entry for a user’s needs.

**Example 1.** Analysis of RDCs in the context of other data for uncharacterized protein PA1076 from *Pseudomonas aeruginosa*, PDB entry 2k4v.15 The VASCO analysis of the chemical shifts [Fig. 2(a,b)] indicates that the 23 atoms with unusual chemical shift values are spread evenly throughout the structure and many of them are in the vicinity of aromatic rings. This information confirms that the structure and chemical shift values are compatible. Furthermore, OLDERADO analysis suggests that the majority of the protein residues form one rigid body domain [Fig. 2(c), orange], including residues 91–105 (UniProt3 sequence numbering), which form an α-helix. However, review of the long-range distance constraints indicates that there is little support for the exact positioning of this helix [Fig. 2(d)]. Indeed, solving the structure of PA1076 with only distance and dihedral angle constraints did not allow for a reliable placement of the helix (data not shown). Introducing RDC data [Fig. 2(e,f)] allowed the positioning of helix 91–105 with more confidence.

**Example 2.** Selecting a structure from a set of identical or similar proteins: comparing two structures for protein HP_0495 from *Helicobacter pylori* (PDB entries 2h9z33 and 2joq15), solved by different structural genomics projects. While in this case the two entries only differ by their purification tags, a similar investigation could be performed for homologous proteins. To compare two structures, one needs to launch two instances of Vivaldi in separate browser windows. In this case, the two structures are similar, but the mostly red and orange ROG scores from the NRG-CING database for entry 2h9z suggest that there may be problems with this structure [Fig. 3(c)], whereas the ROG scores for entry 2joq [Fig. 3(a)] are mostly green. Comparison of several knowledge-based scores available in Vivaldi confirms that both the Ramachandran statistics (not shown) and unusually short distance analysis [Fig. 3(d,e)] are slightly less favorable for 2h9z. However, analysis of the number of long-range distance constraints available for each residue reveals that there are significant local differences between the two structures in the residue range 24–40 (UniProt3 sequence numbering), which spans between Strands 1 and 2 of the central β-sheet and includes helix 1 (residues 27–34). Only one residue from this range (Leu 34) has a significant number of long-range distance constraints (highlighted in the graph by a single green bar) to other parts of the molecule in 2h9z, while 2joq contains 10 residues that each have 10 or more long-range distance constraints (green bars), which would presumably help position helix 1 and surrounding loops relative to the rest of the protein with more confidence [Fig. 3(f,g)]. While a detailed analysis of the differences between the two structures is beyond the scope of this paper, use of Vivaldi reveals these differences and may direct further investigation. Indeed, overlaying the representative models from 2h9z and 2joq [Fig. 3(b)] shows that the positioning of helix 1 and surrounding loops differs significantly and the PDBeFold55 service excludes this region (residues 23–42 and 45–51) from the structural alignment of the two models. While it is not trivial to pinpoint the exact reasons for these differences, it is interesting to note that BMRB entry 15,101 associated with 2h9z has only 485 assigned chemical shift values, of which 159 belong to 1H nuclei, whereas BMRB entry 15,190 associated with 2joq has 1157 assigned chemical shifts, of which 653 belong to 1H nuclei. Completeness of assignments and consequently the number and level of ambiguity of distance constraints strongly depend on the number, types, and resolutions of collected spectra, as well as on the usually labor-intensive step of spectral analysis. At this point one can only speculate if the observed differences are due to genuine differences in experimental conditions (spectra for 2h9z were collected at 35°C, and for 2joq at 25°C), to different types and resolutions of collected data, to variations in data analysis protocols or to some combination of these factors.

**CONCLUSION**

A powerful interactive web-based tool has been developed to facilitate the analysis, visualization and validation of NMR structure ensembles from the PDB. It combines structural, experimental, and validation data from a variety of sources and presents them in a consistent fashion. A user-friendly interface, including a 3D viewer, tightly coupled with charts and tables, enables both experts and non-experts in NMR structure determination to analyze NMR ensembles and to assess their quality.

**ACKNOWLEDGMENTS**

The authors wish to thank Dr W.F. Vranken for discussions on the concept and technical details of the Vivaldi service, Drs J.F. Doreleijers, Ch. Schulte and E. Ulrich for technical assistance in obtaining data from the BMRB...
NMR Restraints Grid, and Prof. M. Sattler for valuable comments on an early version of Vivaldi. Mrs P. Haslam’s help in preparation of the manuscript is greatly appreciated. We would also like to acknowledge our colleagues in the PDBe team for testing and constructive criticism, as well as technical assistance in implementing and maintaining the service. Finally, we wish to thank the structural biology community for depositing their macromolecular structures and experimental data in the PDB and BMRB archives.

REFERENCES

1. Berman HM. The Protein Data Bank: a historical perspective. Acta Crystallogr A 2008;64:88–95.
2. Bernstein FC, Koetzle TF, Williams GJ, Meyer EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. Protein Data Bank—computer-based archival file for macromolecular structures. J Mol Biol 1977;112:535–542.
3. UniProt Consortium. Reorganizing the protein space at the Universal Protein Resource (UniProt). Nucleic Acids Res 2012;40:D71–D75.
4. Berman H, Henrick K, Nakamura H, Markley JL. The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. Nucleic Acids Res 2007;35:D301–D303.
5. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res 2000;28:235–242.
6. Ulrich EL, Akutsu H, Doreleijers JF, Harano Y, Ioannidis YE, Lin J, Livny M, Mading S, Maziuk D, Miller Z, Nakatani E, Schulte CF, Tolmie DE, Wenger RK, Yao HY, Markley JL. BioMagResBank. Nucleic Acids Res 2008;36:D402–D408.
7. Kinjo AR, Suzuki H, Yamashita R, Igarashi R, Kuhara M, Lagerstedt I, Mir S, Newman LE, Oldfield TJ, Patwardhan A, Haslam P, Hendrickx PMS, Heuson E, Hirshberg M, John F, Fernandez Montecelo MA, van Ginkel G, Golovin A, Gore SP, Gutmanas A, Haslam P, Hendrickx PMS, Heuson E, Hirshberg M, John M, Labeuthe I, Mir S, Newman LE, Oldfield TJ, Patwardhan A, Rinaldi L, Sahni G, Sanz-Garcia E, Sen S, Slowey R, Suarez-Urueña A, Swaminathan GJ, Symmons MF, Vranken WF, Wainwright M, Kleywegt GJ. PDBe: Protein Data Bank in Europe. Nucleic Acids Res 2012;40:D453–D460.
8. Velankar S, Alhroub Y, Best C, Caboche S, Conroy MJ, Dana JM, Fernandez Montecelo MA, van Ginkel G, Golovin A, Gore SP, Gutmanas A, Haslam P, Hendrickx PMS, Heuson E, Hirshberg M, John M, Labeuthe I, Mir S, Newman LE, Oldfield TJ, Patwardhan A, Rinaldi L, Sahni G, Sanz-Garcia E, Sen S, Slowey R, Suarez-Urueña A, Swaminathan GJ, Symmons MF, Vranken WF, Wainwright M, Kleywegt GJ. PDBe: Protein Data Bank in Europe (PDBe): bringing structure to biology. Acta Crystallogr D 2011;67:324–330.
9. Bhattacharya A, Tejero R, Montelione GT. Evaluating protein structures determined by structural genomics consortia. Proteins 2007;66:778–795.
10. Berjanskii M, Liang YJ, Zhou JJ, Tang P, Stothard P, Zhou Y, Cruz PH. A new generation of crystallographic validation tools for the Protein Data Bank. Structure 2011;19:1395–1412.
11. Bhattacharya A, Tejero R, Montelione GT. Evaluating protein structures determined by structural genomics consortia. Proteins 2007;59:687–696.
12. Doreleijers JF, Sousa da Silva AW, Krieger E, Nabuurs SB, Spronk CAEM, Stevens TJ, Vranken WF, Vriend G, Vuister GW. CING: an ensemble structure evaluation suite and server. Nucleic Acids Res 2010;38:D402–D408.
13. Shen Y, Bax A. Protein backbone chemical shifts predicted from protein backbone torsion angles from NMR chemical shifts. J Biomol NMR 2009;44:95–99.
14. Kohlhoff KJ, Robustelli P, Cavalli A, Salvatella X, Vendruscolo M. Fast and accurate predictions of protein NMR chemical shifts from interatomic distances. J Am Chem Soc 2009;131:13894–13895.
15. Wang LY, Eghbalnia HR, Bharami A, Markley JL. Linear analysis of carbon-13 chemical shift differences and its application to the detection and correction of errors in referencing and spin system identifications. J Biomol NMR 2005;32:13–22.
16. Wang LY, Markley JL. Empirical correlation between protein backbone N-15 and C-13 secondary chemical shifts and its application to nitrogen chemical shift re-referencing. J Biomol NMR 2009;44:215–240.
17. Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res 2007;35:D401–D403.
18. Han B, Liu YF, Ginzinger SW, Wishart DS. SHIFTX2: significantly improved protein chemical shift prediction. J Biomol NMR 2011;50:43–57.
19. Neal S, Nip AM, Zhang HY, Wishart DS. Rapid and accurate calculation of protein H-1, C-13 and N-15 chemical shifts. J Biomol NMR 2003;26:215–240.
20. Shen Y, Bax A. Protein backbone chemical shifts predicted from searching a database for torsion angle and sequence homology. J Biomol NMR 2007;38:289–302.
21. Shen Y, Bax A. SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. J Biomol NMR 2010;44:13–22.
22. Shen Y, Bax A. SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. J Biomol NMR 2010;44:13–22.
23. Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res 2007;35:D301–D303.
24. Velankar S, Alhroub Y, Best C, Caboche S, Conroy MJ, Dana JM, Fernandez Montecelo MA, van Ginkel G, Golovin A, Gore SP, Gutmanas A, Haslam P, Hendrickx PMS, Heuson E, Hirshberg M, John M, Labeuthe I, Mir S, Newman LE, Oldfield TJ, Patwardhan A, Rinaldi L, Sahni G, Sanz-Garcia E, Sen S, Slowey R, Suarez-Urueña A, Swaminathan GJ, Symmons MF, Vranken WF, Wainwright M, Kleywegt GJ. PDBe: Protein Data Bank in Europe. Nucleic Acids Res 2012;40:D453–D460.
25. Bhattacharya A, Tejero R, Montelione GT. Evaluating protein structures determined by structural genomics consortia. Proteins 2007;66:778–795.
26. Berjanskii M, Liang YJ, Zhou JJ, Tang P, Stothard P, Zhou Y, Cruz PH. A new generation of crystallographic validation tools for the Protein Data Bank. Structure 2011;19:1395–1412.
27. Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res 2007;35:D401–D403.
28. Rieping W, Vranken WF. Validation of archived chemical shifts for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings and its comparison with X-ray structures. Protein Sci 2011;20:95–99.
29. Velankar S, Kleywegt GJ. The Protein Data Bank in Europe (PDBe): bringing structure to biology. Acta Crystallogr D 2011;67:324–330.
30. Bhattacharya A, Tejero R, Montelione GT. Evaluating protein structures determined by structural genomics consortia. Proteins 2007;66:778–795.
31. Berjanskii M, Liang YJ, Zhou JJ, Tang P, Stothard P, Zhou Y, Cruz PH. A new generation of crystallographic validation tools for the Protein Data Bank. Structure 2011;19:1395–1412.
32. Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res 2007;35:D401–D403.
33. Vranken WF, Boucher W, Stevens TJ, Vranken WF, Vriend G, Vuister GW. Quantitative evaluation of experimental NMR restraints. J Am Chem Soc 2003;125:12026–12034.
34. Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res 2007;35:D401–D403.
35. Velankar S, Alhroub Y, Best C, Caboche S, Conroy MJ, Dana JM, Fernandez Montecelo MA, van Ginkel G, Golovin A, Gore SP, Gutmanas A, Haslam P, Hendrickx PMS, Heuson E, Hirshberg M, John M, Labeuthe I, Mir S, Newman LE, Oldfield TJ, Patwardhan A, Rinaldi L, Sahni G, Sanz-Garcia E, Sen S, Slowey R, Suarez-Urueña A, Swaminathan GJ, Symmons MF, Vranken WF, Wainwright M, Kleywegt GJ. PDBe: Protein Data Bank in Europe. Nucleic Acids Res 2012;40:D453–D460.
36. Bhattacharya A, Tejero R, Montelione GT. Evaluating protein structures determined by structural genomics consortia. Proteins 2007;66:778–795.
35. Doreleijers JF, Mading S, Maziuk D, Sojourner K, Yin L, Zhu J, Markley JL, Ulrich EL. BioMagResBank database with sets of experimental NMR constraints corresponding to the structures of over 1400 biomolecules deposited in the Protein Data Bank. J Biomol NMR 2003;26:139–146.

36. Doreleijers JF, Vranken WF, Schulte C, Lin JD, Wedell JR, Penkett CJ, Vuister GW, Vriend G, Markley JL, Ulrich EL. The NMR restraints grid at BMRB for 5,266 protein and nucleic acid PDB entries. J Biomol NMR 2009;45:389–396.

37. Neuhaus D, Williamson MP. The nuclear Overhauser effect in structural and conformational analysis. NY: Wiley-VCH, 2000.

38. Bax A, Kontaxis G, Tjandra N. Dipolar couplings in macromolecular structure determination. Method Enzymol 2001;339:127–174.

39. Losonczi JA, Andre M, Fischer MWF, Prestegard JH. Order matrix analysis of residual dipolar couplings using singular value decomposition. J Magn Reson 1999;138:333–342.

40. Cornilescu G, Marquardt JL, Ottiger M, Bax A. Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. J Am Chem Soc 1998;120:6836–6837.

41. Vriend G. What IF—a molecular modeling and drug design program. J Mol Graphics 1990;8:52–56.

42. Laskowski RA, Rullmann JAC, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR 1996;8:477–486.

43. Hooft RWW, Vriend G, Sander C, Abola EE. Errors in protein structures. Nature 1996;381:272–272.

44. Eisenberg D, Lathy R, Bowie JU. VERIFY3D: assessment of protein models with three-dimensional profiles. Macromol Crystallogr B 1997;277:396–404.

45. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 2007;35:W407–W410.

46. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D 2010;66:12–21.

47. Doreleijers JF, Vranken WF, Schulte C, Markley JL, Ulrich EL, Vriend G, Vuister GW. NRG-CING: integrated validation reports of remediated experimental biomolecular NMR data and coordinates in wwPDB. Nucleic Acids Res 2012;40:D519–D524.

48. Hartshorn MJ. AstexViewer (TM): a visualisation aid for structure-based drug design. J Comput Aid Mol Des 2002;16:871–881.

49. Velankar S, McNeil P, Mittard-Runte V, Suarez A, Barrell D, Apweiler R, Henrick K. E-MSD: an integrated data resource for bioinformatics. Nucleic Acids Res 2005;33:D262–D265.

50. Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K, Holm L, Sonnhammer ELL, Eddy SR, Bateman A. The Pfam protein families database. Nucleic Acids Res 2010;38:D211–D222.

51. Cuff AL, Sillitoe I, Lewis T, Clegg AB, Rentzsch R, Furnham N, Pellegrini-Calace M, Jones D, Thornton J, Orengo CA. Extending CATH: increasing coverage of the protein structure universe and linking structure with function. Nucleic Acids Res 2011;39: D420–D426.

52. Andreeva A, Howorth D, Chandonia JM, Brenner SE, Hubbard TJF, Chothia C, Murzin AG. Data growth and its impact on the SCOP database new developments. Nucleic Acids Res 2008;36:D419–D425.

53. Seo MD, Park SJ, Kim HJ, Lee BJ. Solution structure of hypothetical protein, HP0495 (Y495_HELPPY) from Helicobacter pylori. Proteins 2007;67:1189–1192.

54. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. J Mol Graph Model 1996;14:33–38.

55. Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr D 2004;60:2256–2268.