

# MolProbity for the masses—of data

Vincent B. Chen<sup>1</sup> · Jonathan R. Wedell<sup>2</sup> · R. Kent Wenger<sup>2</sup> · Eldon L. Ulrich<sup>2</sup> · John L. Markley<sup>1</sup>

Received: 5 April 2015 / Accepted: 13 July 2015 / Published online: 21 July 2015  
© Springer Science+Business Media Dordrecht 2015

**Abstract** MolProbity is a powerful software program for validating structures of proteins and nucleic acids. Although MolProbity includes scripts for batch analysis of structures, because these scripts analyze structures one at a time, they are not well suited for the validation of a large dataset of structures. We have created a version of MolProbity (MolProbity-HTC) that circumvents these limitations and takes advantage of a high-throughput computing cluster by using the HTCCondor software. MolProbity-HTC enables the longitudinal analysis of large sets of structures, such as those deposited in the PDB or generated through theoretical computation—tasks that would have been extremely time-consuming using previous versions of MolProbity. We have used MolProbity-HTC to validate the entire PDB, and have developed a new visual chart for the BioMagResBank website that enables users to easily ascertain the quality of each model in an NMR ensemble and to compare the quality of those models to the rest of the PDB.

**Keywords** Structure validation · MolProbity · BioMagResBank (BMRB) · High-throughput computing · Protein Data Bank · HTCCondor

## Introduction

The three-dimensional structures of macromolecules, such as proteins and RNA molecules, provide important information about their evolution and function. All methods for structure determination provide models of those structures that require validation to assess their quality and to discover errors that can be corrected through comparison with underlying data. A number of freely available computational tools check for mistakes such as steric clashes between non-bonded atoms, non-standard bond lengths and angles, problems with dihedral angles, non-rotameric protein sidechains, poor packing, and other issues. Examples of these tools include WHAT\_CHECK (Hooft et al. 1996), the iCing validation package (Doreleijers et al. 2012), and the MolProbity web server (Chen et al. 2010).

Validation of structures and underlying data is an issue currently being addressed by the Worldwide Protein Data Bank (wwPDB). The wwPDB formed validation task forces (VTFs) for each of the main structural determination techniques (X-ray, NMR, EM) and charged them with preparing recommendations for making validation more accessible to users. The initial reports of these groups have been published (Henderson et al. 2012; Montelione et al. 2013; Read et al. 2011). Recommendations of the NMR VTF (Montelione et al. 2013) included the development of approaches to simplify the validation process and to make it more accessible to both structural biologists and general users of structural data.

We describe a new version of the MolProbity structure validation software suite (MolProbity-HTC) that has been modified to take advantage of the HTCCondor high-throughput computing software (Tannenbaum et al. 2001). HTCCondor works by allowing users to harness the computational power of a large number of machines (dedicated

✉ John L. Markley  
markley@nmrfam.wisc.edu

<sup>1</sup> National Magnetic Resonance Facility at Madison, Biochemistry Department, University of Wisconsin-Madison, 433 Babcock Drive, Madison, WI 53706, USA

<sup>2</sup> BioMagResBank, Biochemistry Department, University of Wisconsin-Madison, 433 Babcock Drive, Madison, WI 53706, USA

compute servers or desktop workstations that are used when they are idle). It has flexible job distribution and monitoring systems, and is tolerant to faults, including computer crashes and network outages. By making use of HTCondor, we have greatly reduced the amount of time needed to run MolProbity validation on a large set of macromolecular 3D PDB structures.

MolProbity-HTC is able to more quickly validate large sets of structures, such as those generated using homology modeling methods, and allows us to compare the quality of those structures. A faster MolProbity also allows us to more easily develop and test new validation methods. We have run MolProbity-HTC on the entire set of protein and nucleic acid structures deposited at BMRB/PDB, and have set up an automated system which automatically updates the MolProbity statistics weekly in order to incorporate data from newly released structures. Results generated by MolProbity-HTC include both “online” model-level summary scores and residue-level scores. These results are freely available on the BioMagResBank (BMRB) (Ulrich et al. 2008) website at [http://www.bmrwisc.edu/ftp/pub/bmrwisc/pdb\\_molprobity/](http://www.bmrwisc.edu/ftp/pub/bmrwisc/pdb_molprobity/).

To help present these results to a wider audience, we also created an online interactive chart for visualizing the overall quality of a structure or ensemble. We created a page (<http://www.bmrwisc.edu/validate/molprobity/>) on the BMRB website that allows users to enter any PDB code and retrieve the corresponding MolProbity quality chart. The chart allows users to see the overall quality of a structure, both for the “well-defined” core region and for the structure as a whole, to compare different models of a structure ensemble, and to compare a model with the validation results for the rest of the PDB. Although also compatible with X-ray structures, this chart is particularly geared toward visualizing validation data for NMR ensembles.

## Materials and methods

We added several new command-line Python scripts to MolProbity (version 4.02), which, when run on a directory of PDB files, automatically generate everything needed to use MolProbity to validate the structures using a preexisting HTCondor computing cluster. We used the Directed Acyclic Graph Manager (DAGMan) (Couvares et al. 2007) tool in HTCondor to organize the HTCondor jobs and to ensure they run in the correct order. Unlike previously available command-line scripts, which needed to be run sequentially, the DAGMan feature allows our new script to combine multiple steps of MolProbity analysis.

The main script we added, “make\_condor\_files”, takes as input a directory of PDB files, divides up the input PDB

files for analysis in separate HTCondor jobs, creates all needed HTCondor submit and DAGMan files, and creates all the needed output directories. At its core, MolProbity contains multiple programs written in multiple languages, as well as a whole website interface, making it impractical to send out the entire MolProbity package to HTCondor compute nodes. To address this issue, we rewrote some of the primary analysis code of MolProbity in Python so that only the Python code and several executable files need to be sent to compute nodes rather than the entire MolProbity suite. The `make_condor_files` script divides up the input PDB files into separate jobs to give each job the optimal HTCondor job run time of approximately 30 min–1 h (runtime depends mainly on the size of structures in the job). MolProbity-HTC requires compute nodes with Python and Java. Each of these HTCondor jobs encompasses MolProbity analysis of several PDB files, the number of which depends on the size of the files. The first set of HTCondor jobs runs “Reduce” for trimming and addition of hydrogens (Word et al. 1999). Hydrogens are added with and without Asn/Gln/His flip corrections. Nuclear bond lengths or electron cloud bond lengths can be specified, depending on the type of input PDB files (NMR vs. X-ray structures). Ultimately, MolProbity scores are calculated for three versions of a PDB file: the original file, the file with hydrogens added with flip corrections, and the file with hydrogens added without flip corrections. The second set of HTCondor jobs performs “online” analysis to provide model-by-model summary scores. The original files and both “reduce” files are analyzed. The third set of HTCondor jobs performs residue analysis to provide residue-level validation scores. A final HTCondor job compiles the results into text and NMR-STAR formatted output. Because the analysis jobs cannot run without the results from the Reduce jobs, DAGMan is used to enforce the job dependencies. All of the job monitoring and error logging is handled with built-in HTCondor features. At the BMRB, we use the “croncondor” feature of HTCondor to run this script automatically each week on any newly released files in the PDB archive.

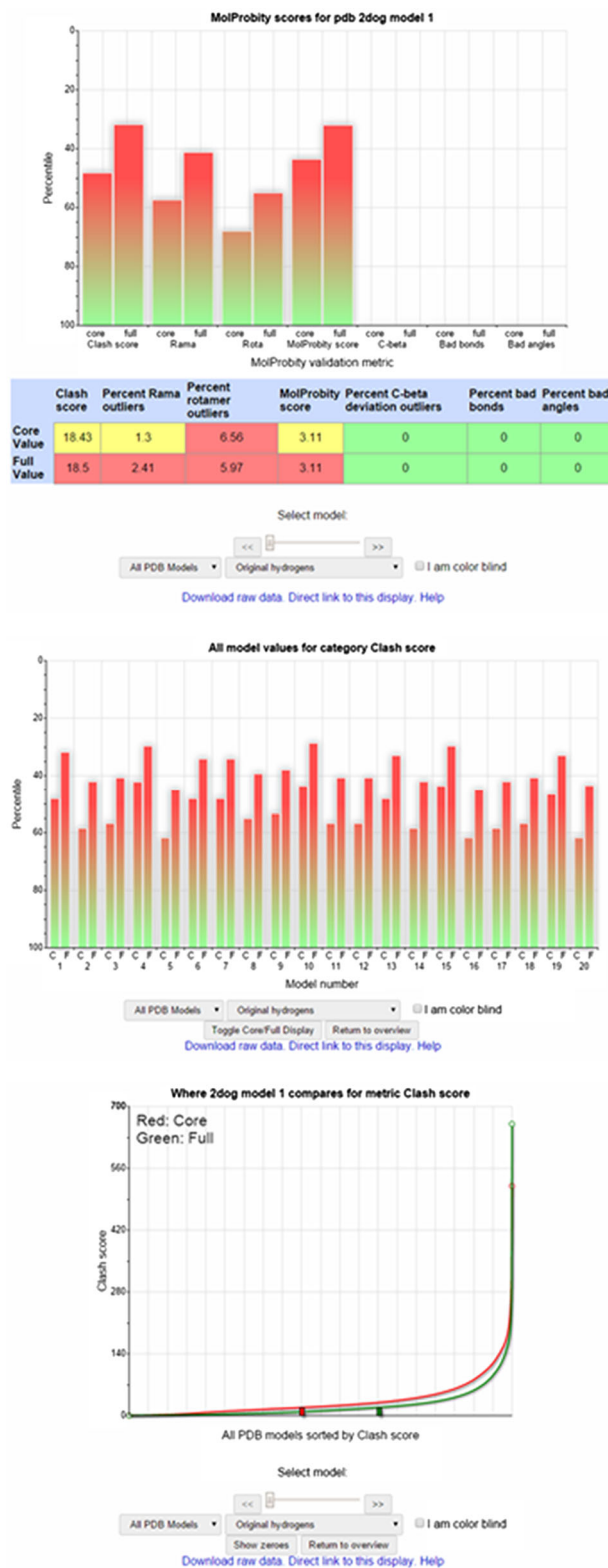
We added several analyses to the pre-existing residue analysis script. These new features include a report on the tau (N–CA–C) angle and omega (peptide planarity) dihedral angle, and a report on the bond length, angles, and dihedral angles of disulfide bonds. These results could form the basis for new quality measures, such as disulfide rotamers.

NMR ensembles often contain a well-packed core region, with disordered tails and loops. These loops often contain more errors than the well packed core regions. This issue was recognized by the recent NMR validation task force, which recommended that validation of the core regions be carried out separately from validation of the

whole structures. Per their recommendations, we incorporated an option to run the CYRANGE (Kirchner and Guntert 2011) program for calculating the core residues of protein structures. This was accomplished by creating a new script “run\_condor\_mp\_analysis” which runs CYRANGE on the input files in order to calculate the core residues for any PDB file containing multiple models. This script uses those residues to generate a new “core” PDB file. The run\_condor\_mp\_analysis script then automatically calls the make\_condor\_files script with the core and full PDB files in order to set up the HTCondor files.

One objective of the VTF was the development of a visual validation “slider” image that could be generated for a given PDB structure that would provide users with a quick way to see the overall quality of a structure. The X-ray validation task force (Read et al. 2011) developed a prototype slider image that was adopted by the PDB in April 2014. One issue with this slider representation is that it is limited to showing the data for one model at a time. For the majority of NMR structures, which are typically deposited as ensembles of models, this image would need to be modified in order to show the quality of each model, as well as how they compare to one another and to the rest of the PDB.

In order to address this issue, we developed an interactive chart that can show validation statistics for single models, multiple models, and a comparison with the entire PDB (Fig. 1). This interactive chart makes use of the JavaScript charts library RGraph (www.rgraph.net) and HTML5, which do not require users to install any web browser plugins, and is compatible with viewing on mobile devices. This interactive chart has three different modes. The first mode shows a bar graph for each quality metric with the height of the bar indicating the corresponding percentile against all structures. This mode also shows a table of the raw scores and percentages from which the percentiles were calculated. The second mode shows bar graphs for all the models for a single quality metric. The third mode shows a graph of the distribution of data for a single quality metric and where a given model falls compared to other structures. The third mode also has a toggle for displaying or hiding zero scores in the distribution graph; this is useful for structures having many models that score “perfectly”, resulting in skewed distributions. In the first and third modes, controls are provided to allow users to easily step through the different models. In the two bar-chart modes, the bar height shows the percentile, with taller bars indicating a worse percentile. Percentiles are calculated from either the percent outliers or from the raw score (i.e. for clashscore and MolProbity score). This chart has been deployed on the BMRB website and is available for



all X-ray and NMR structures deposited in the PDB. The distribution data displayed on the chart are also automatically updated each week to include any new results.

**Fig. 1** Screenshots of the interactive validation chart at the BMRB, showing the results for PDB ID: 2dog (Suzuki et al. 2007). **a** MolProbity quality measures for a single model; both core and full scores are shown. **b** Percentile scores for a single MolProbity quality measure for all models of an ensemble. **c** Comparison of a single model for a given quality measure to the entire PDB along with the distribution of the data for that quality measure

## Results

MolProbity-HTC allows larger sets of structures to be validated using MolProbity far more efficiently than was previously possible. Several command-line options are available in MolProbity for running batch-mode analyses, but all of these options operate on models only one at a time. Also, these command-line options are constrained to analyzing a limited number of files. This is particularly problematic when analyzing NMR structures, because most of those consist of multiple models. These limitations mean that MolProbity analysis of a large dataset of structures such as the entire PDB requires dividing the structures into several different directories, running the analysis scripts on each of those directories one at a time, and additional manual processing to compile the results. This labor-intensive process would require months to complete on a single machine. Currently, MolProbity-HTC, running on the small HTCCondor cluster of approximately 150 compute nodes at the BMRB, takes approximately one week to run the entire suite of MolProbity analyses, including addition of hydrogens, oneline and residue analyses, on both core and full structures of the entire PDB.

We also performed a brief comparison of the MolProbity quality scores, sorted by time, for structures determined by X-ray crystallography and NMR spectroscopy (Table 1). In general, NMR structures scored worse than X-ray structures according to clashscore, Ramachandran, and rotamer outliers, but NMR structures scored equally as well as X-ray structures on bond geometric quality measures. As shown by the large standard deviations (given in parentheses in Table 1) of the scores, there is a wide range in the quality of structures; certainly, as noted previously (Bhattacharya et al. 2007), some NMR structures are of high quality and some X-ray structures are of low quality, which reemphasizes the need to check the quality of any structure, regardless of the determination method or when it was solved, before using that structure for any purpose. Scores have improved over time for both X-ray and NMR structures, but structures are still being published with high percentages of outliers. Interestingly, the average clash scores for the NMR core structures are higher than for the full structures, but the percent Ramachandran outliers are lower. This is probably due to the removal of the more dynamic (and less well packed) loops and flexible tails (see

Fig. 2). Also, according to the RNA-specific quality measure “perpendicular phosphate distance,” the quality of RNA structures has stayed roughly the same over time, probably due to fewer well-developed tools for calculating RNA structures than for protein structures.

In the course of analyzing the PDB, we discovered several structures that gave unexpected MolProbity results. First, several NMR structures were deposited without hydrogen atoms, which led to incorrect clashscores and MolProbity scores for those structures. In order to consistently address this issue, our script adds hydrogens to all structures prior to analysis. In addition, several structures contained no standard protein or nucleic acid residues. Minimized ensemble-averaged models also yielded exceptionally poor MolProbity scores. Finally, in some depositions, the authors modified the models in unusual ways. An example is PDB ID: 1gjj in which two subdomains of a protein were superimposed to show their similarity, but were deposited in the PDB with both subdomains in the same chain. This is not necessarily an incorrect structure, but yielded a poor clashscore owing to many overlapping atoms.

The NMR VTF emphasized that the validation report for NMR structures should define the well-defined core and the ill-defined regions of the structures for separate structure validation. The bar chart in Fig. 2 uses the MolProbity score to show that the well-defined regions are higher quality than the whole structure for the majority of NMR structures. The chart also reveals that a significant fraction of structures have well-defined regions that score worse than the structure as a whole. One example (Fig. 2, lower left panel, PDB ID: 2ku7) shows the well-defined core as having more steric clashes than the more disordered (and presumably less restrained) tails. As noted earlier, in general core structures have higher clashscores than full structures, but because the MolProbity score combines clashscore, Ramachandran outliers, and sidechain rotamer outliers into a single score, an individual core structure with a higher clashscore than the full version could still score better by MolProbity score if it has a lower percentage of Ramachandran and sidechain rotamer outliers. These results re-emphasize that it is important to validate both the well-defined and ill-defined regions for individual structures.

## Conclusions and future prospects

This work is a further demonstration of the power of high-throughput computing for tackling problems in structural biology. We were able to apply HTCCondor to a software package containing a diverse set of components and significantly reduce the amount of time required to use that

**Table 1** Average and standard deviation for five year periods of various MolProbity quality measures (lower averages indicate better quality)

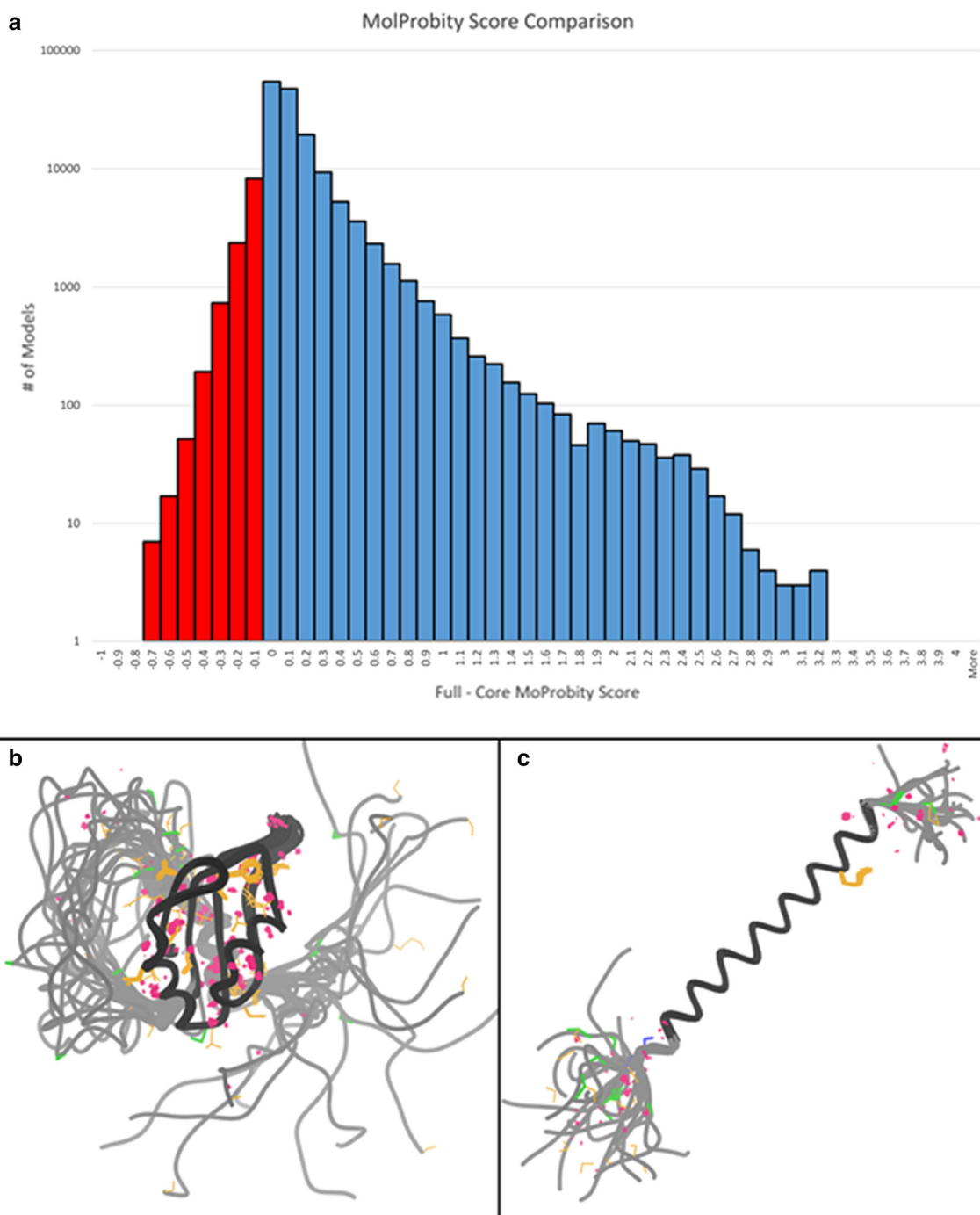
	Clashscore	% Rama outliers	% Rotamer outliers	% Bad bonds	% Bad angles	% RNA pperp outliers
NMR-core						
1994–1998	56.5 (58.4)	5.8 (5.9)	20.3 (13.2)	0.05 (0.49)	0.22 (1.38)	n/a
1999–2003	58.1 (60.0)	5.2 (6.8)	22.0 (14.1)	0.03 (0.60)	0.10 (0.72)	n/a
2004–2008	35.4 (35.3)	3.1 (5.3)	13.8 (12.4)	0.02 (0.47)	0.06 (0.59)	n/a
2009–2013	27.7 (30.8)	2.2 (3.9)	12.7 (11.7)	0.03 (0.67)	0.05 (0.35)	n/a
NMR-full						
1994–1998	51.2 (54.6)	7.1 (6.1)	20.5 (13.5)	0.18 (1.47)	0.66 (2.40)	5.9 (9.3)
1999–2003	51.8 (55.3)	6.6 (6.9)	22.1 (13.7)	0.14 (1.46)	0.41 (1.73)	7.5 (10.9)
2004–2008	32.7 (33.3)	4.2 (4.9)	14.1 (12.3)	0.08 (0.94)	0.18 (1.08)	8.0 (10.6)
2009–2013	26.6 (29.6)	3.2 (4.3)	12.7 (11.3)	0.09 (1.24)	0.18 (0.99)	8.5 (12.8)
X-ray						
1994–1998	14.8 (16.9)	1.1 (2.0)	7.5 (6.1)	0.34 (2.24)	0.89 (2.64)	3.7 (9.5)
1999–2003	14.2 (14.6)	1.0 (2.0)	4.9 (4.8)	0.12 (1.38)	0.33 (1.27)	5.1 (9.5)
2004–2008	12.6 (17.2)	0.9 (1.8)	4.4 (4.3)	0.07 (0.86)	0.17 (0.80)	4.8 (8.6)
2009–2013	9.2 (12.0)	0.6 (1.5)	3.6 (3.9)	0.05 (0.67)	0.10 (0.75)	3.6 (6.6)

software to analyze a large set of structures. HTCCondor is straightforward to use and is a powerful tool for large computational problems requiring large numbers of repetitive calculations. Additionally, in principle, the part of the `make_condor_files` script that divides up the input PDB files into smaller jobs and the rewritten MolProbity analysis Python scripts could be extracted from MolProbity-HTC and modified for use with other parallel computing software environments. Planned improvements to MolProbity-HTC will make it more powerful. Currently, the system requires that the PDB files to be analyzed reside on a shared file system accessible by all of the HTCCondor compute machines. If we enable our script to transfer the PDB files to remote compute clusters, we could use non-local compute resources to further speed up the analysis. Also, we are exploring how to apply HTCCondor to the recently rewritten version 4.1 of MolProbity, which switched from using a set of small, diverse programs written in a number of different languages to a single program, the Computational Crystallography Toolbox (CCTBX) (Grosse-Kunstleve et al. 2002). This version of MolProbity is more complex to install and makes it impractical to distribute the entire CCTBX package to the compute nodes; we are currently investigating whether use of a caching system, such as a Squid proxy, as an intermediate host for CCTBX in the MolProbity-HTC scripts could make version 4.1 compatible with HTCCondor. Whereas this work has focused on MolProbity, we plan to incorporate validation from other software packages into our system and to display those results in our validation chart.

Although significant progress has been made in improving the quality of structures, it is clear that still more can be done.

Tools such as MolProbity-HTC are an essential first step to finding errors within structures, but correction of those errors remains a tricky task, particularly for NMR. Improvements to the force fields used for NMR structure calculation have helped the overall quality of NMR structures, especially the development of refinement methods including explicit solvent (Linge et al. 2003; Ramelot et al. 2009), but these methods can still produce models with local errors. Ideally, an NMR structural biologist would diagnose these local errors in a structure, and examine the corresponding restraints and primary data to correct misassignments or misplaced peaks in the spectra during several rounds of refinement. Development of improved “semi-automated” software tools, such as NMRFAM-SPARKY (Lee et al. 2015) and Ponderosa-C/S (Lee et al. 2014), for working with the restraints, assignments, and data can help with this task. Also, there are experimental causes of errors. For example, spin diffusion is known have an effect on the number and volume of NOESY cross-peaks, which leads to incorrect distance restraints (Hoogstraten et al. 1993). Spin diffusion effects can be diagnosed experimentally, but issues such as these are not always possible to correct. Structures can legitimately contain “errors” if there are compensating factors (such as hydrogen bonds) to offset the energetic penalty of those errors, so striving for completely perfect MolProbity scores should not be the ultimate goal. Ultimately it is vital that structural biologists and general users identify local errors in the region of interest of their structures before making scientific conclusions about those parts of the structures.

The current version of the MolProbity-HTC scripts are available as part of the main source code repository of MolProbity at <https://github.com/rlabduke/MolProbity>.



**Fig. 2** Comparison of the overall quality of full versus well-defined (“core”) NMR structures. **a** Histogram showing the distribution of NMR structures, binned by the difference of the MolProbity score for the full structure and the core structure. A lower MolProbity score indicates better quality. Negative bins indicate structures in which the core structure is of lower quality than the full structure. **b**, **c** Overall structural views of examples of cases where either (**b**) the core was lower quality than the full structure (PDB ID: 2ku7) (Wang et al.

2010), or (**c**) the full structure was of lower quality than the core (PDB ID: 2rmz) (Lau et al. 2008). In both **b**, **c**, the core region is colored in *black*, with the rest of the structure in *grey*. Problems as diagnosed by MolProbity are shown with *pink*, *gold*, or *green* indicators. The structure in (**b**) has most of the problems located in the core region, whereas the structure in (**c**) has more problems in the disordered tail regions

**Acknowledgments** We thank Miron Livny for helpful and lively discussions about the use of HTCCondor. We also thank Dimitri Maziuk for technical assistance and help with systems administration at the BMRB, and W. Milo Westler for discussions about NMR

methodology. This work was supported by NIH Grants R01 GM109046 and P41GM103399. VBC has received partial support from NLM Grant 5T15LM007359.

## References

- Bhattacharya A, Tejero R, Montelione GT (2007) Evaluating protein structures determined by structural genomics consortia. *Proteins* 66:778–795. doi:[10.1002/prot.21165](https://doi.org/10.1002/prot.21165)
- Chen VB et al (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr Sect D: Biol Crystallogr* 66:12–21. doi:[10.1107/S0907444909042073](https://doi.org/10.1107/S0907444909042073)
- Couvares P, Kosar T, Roy A, Weber J, Wenger K (2007) Workflow management in condor. In: Taylor IJ, Deelman E, Gannon DB, Shields M (eds) *Workflows for e-science*. Springer, London. doi:[10.1007/978-1-84628-757-2](https://doi.org/10.1007/978-1-84628-757-2)
- Doreleijers JF et al (2012) CING: an integrated residue-based structure validation program suite. *J Biomol NMR* 54:267–283. doi:[10.1007/s10858-012-9669-7](https://doi.org/10.1007/s10858-012-9669-7)
- Grosse-Kunstleve RW, Sauter NK, Moriarty NW, Adams PD (2002) The computational crystallography toolbox: crystallographic algorithms in a reusable software framework. *J Appl Crystallogr* 35:126–136. doi:[10.1107/S0021889801017824](https://doi.org/10.1107/S0021889801017824)
- Henderson R et al (2012) Outcome of the first electron microscopy validation task force meeting. *Structure* 20:205–214. doi:[10.1016/j.str.2011.12.014](https://doi.org/10.1016/j.str.2011.12.014)
- Hooft RW, Vriend G, Sander C, Abola EE (1996) Errors in protein structures. *Nature* 381:272. doi:[10.1038/381272a0](https://doi.org/10.1038/381272a0)
- Hoogstraten CG, Westler WM, Macura S, Markley JL (1993) improved measurement of longer proton-proton distances in proteins by relaxation network editing. *J Mag Reson Ser B* 102:232–235. doi:[10.1006/jmrb.1993.1090](https://doi.org/10.1006/jmrb.1993.1090)
- Kirchner DK, Guntert P (2011) Objective identification of residue ranges for the superposition of protein structures. *BMC Bioinf* 12:170. doi:[10.1186/1471-2105-12-170](https://doi.org/10.1186/1471-2105-12-170)
- Lau TL, Partridge AW, Ginsberg MH, Ulmer TS (2008) Structure of the integrin beta3 transmembrane segment in phospholipid bicelles and detergent micelles. *Biochemistry* 47:4008–4016. doi:[10.1021/bi800107a](https://doi.org/10.1021/bi800107a)
- Lee W, Stark JL, Markley JL (2014) PONDEROSA-C/S: client-server based software package for automated protein 3D structure determination. *J Biomol NMR* 60:73–75. doi:[10.1007/s10858-014-9855-x](https://doi.org/10.1007/s10858-014-9855-x)
- Lee W, Tonelli M, Markley JL (2015) NRMFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* 31:1325–1327. doi:[10.1093/bioinformatics/btu830](https://doi.org/10.1093/bioinformatics/btu830)
- Linge JP, Williams MA, Spronk CA, Bonvin AM, Nilges M (2003) Refinement of protein structures in explicit solvent. *Proteins* 50:496–506. doi:[10.1002/prot.10299](https://doi.org/10.1002/prot.10299)
- Montelione GT et al (2013) Recommendations of the wwPDB NMR validation task force. *Structure* 21:1563–1570. doi:[10.1016/j.str.2013.07.021](https://doi.org/10.1016/j.str.2013.07.021)
- Ramelot TA et al (2009) Improving NMR protein structure quality by Rosetta refinement: a molecular replacement study. *Proteins* 75:147–167. doi:[10.1002/prot.22229](https://doi.org/10.1002/prot.22229)
- Read RJ et al (2011) A new generation of crystallographic validation tools for the protein data bank. *Structure* 19:1395–1412. doi:[10.1016/j.str.2011.08.006](https://doi.org/10.1016/j.str.2011.08.006)
- Suzuki S et al (2007) Structural characterization of the ribosome maturation protein RimM. *J Bacteriol* 189:6397–6406. doi:[10.1128/JB.00024-07](https://doi.org/10.1128/JB.00024-07)
- Tannenbaum T, Wright D, Miller K, Livny M (2001) Condor: a distributed job scheduler. In: Sterling T (ed) *Beowulf cluster computing with Linux*. MIT Press, Cambridge
- Ulrich EL et al (2008) BioMagResBank. *Nucleic Acids Res* 36:D402–D408. doi:[10.1093/nar/gkm957](https://doi.org/10.1093/nar/gkm957)
- Wang Z, Song J, Milne TA, Wang GG, Li H, Allis CD, Patel DJ (2010) Pro isomerization in MLL1 PHD3-bromo cassette connects H3K4me readout to Cyp33 and HDAC-mediated repression. *Cell* 141:1183–1194. doi:[10.1016/j.cell.2010.05.016](https://doi.org/10.1016/j.cell.2010.05.016)
- Word JM, Lovell SC, Richardson JS, Richardson DC (1999) Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J Mol Biol* 285:1735–1747. doi:[10.1006/jmbi.1998.2401](https://doi.org/10.1006/jmbi.1998.2401)