Published online 11 May 2010 *Nucleic Acids Research, 2010, Vol. 38, Web Server issue* ***W633–W640***
doi:10.1093/nar/gkq375

PROSESS: a protein structure evaluation suite and server

Mark Berjanskii[1], Yongjie Liang[1], Jianjun Zhou[1], Peter Tang[1], Paul Stothard[2], You Zhou[3], Joseph Cruz[1], Cam MacDonell[1], Guohui Lin[1], Paul Lu[1] and David S. Wishart[1,3,4,*]

[1]Department of Computing Science, [2]Faculty of Agriculture, Life and Environmental Sciences, [3]Department of Biological Sciences, University of Alberta and [4]National Research Council, National Institute for Nanotechnology (NINT), Edmonton, AB, Canada T6G 2E8

Received February 15, 2010; Revised April 19, 2010; Accepted April 26, 2010

ABSTRACT

PROSESS (PROtein Structure Evaluation Suite and Server) is a web server designed to evaluate and validate protein structures generated by X-ray crystallography, NMR spectroscopy or computational modeling. While many structure evaluation packages have been developed over the past 20 years, PROSESS is unique in its comprehensiveness, its capacity to evaluate X-ray, NMR and predicted structures as well as its ability to evaluate a variety of experimental NMR data. PROSESS integrates a variety of previously developed, well-known and thoroughly tested methods to evaluate both global and residue specific: (i) covalent and geometric quality; (ii) non-bonded/packing quality; (iii) torsion angle quality; (iv) chemical shift quality and (v) NOE quality. In particular, PROSESS uses VADAR for coordinate, packing, H-bond, secondary structure and geometric analysis, GeNMR for calculating folding, threading and solvent energetics, ShiftX for calculating chemical shift correlations, RCI for correlating structure mobility to chemical shift and PREDITOR for calculating torsion angle-chemical shifts agreement. PROSESS also incorporates several other programs including MolProbity to assess atomic clashes, Xplor-NIH to identify and quantify NOE restraint violations and NAMD to assess structure energetics. PROSESS produces detailed tables, explanations, structural images and graphs that summarize the results and compare them to values observed in high-quality or high-resolution protein structures. Using a simplified red–amber–green coloring scheme PROSESS also alerts users about both general and residue-specific structural problems. PROSESS is intended to serve as a tool that can be used by structure biologists as well as database curators to assess and validate newly determined protein structures. PROSESS is freely available at http://www.prosess.ca.

INTRODUCTION

Protein structure determination is still one of the most challenging tasks in chemistry and biology. Indeed, it is not uncommon for a complete structure determination (from cloning to solving) to take several person-years of intense effort. While substantial technical and computational strides have made both X-ray crystallography and NMR spectroscopy much more 'automated', there is still a large element of human intervention and human interpretation that is required to complete the process. Indeed, human expertise is often needed to address ambiguities or inconsistencies in the raw diffraction data (for X-ray) or NOE measurements (for NMR). Human intervention also plays an equally important role in the prediction of protein structures (via homology or *ab initio* modeling). This human element can also lead to errors, some of which can have profound consequences (1–4). However, the complexity of protein structures makes the visual or manual detection of these errors exceedingly difficult. As a result, a number of computer programs have been developed to help both 'consumers' and 'producers' of protein structures identify these errors. PROCHECK (5) and PROCHECK-NMR (6) were among the first programs to offer comprehensive geometrical and stereochemical analysis of X-ray, NMR and computationally modeled structures. The insightful concepts behind these

*To whom correspondence should be addressed. Tel: +1 780 492 0383; Fax: +1 780 492 5305; Email: david.wishart@ualberta.ca

two programs, along with their rich graphical output have largely set the standard for all subsequent structure validation tools. More recent stand-alone programs and web servers such as WHATCHECK (4), OOPS (7), VADAR (8), MolProbity (9), PROVE (10), Verify3D (11) and ProSA (12) have either built upon PROCHECK or added newer and better measurement concepts to the mix. All of these programs are excellent and many offer important global and residue-specific insights into structural errors or ambiguities. However, the abundance of programs, the diversity of outputs along with the growing obsolescence of certain reference parameters has made it increasingly difficult for users to take full advantage of what can—and should—be done in protein structure validation.

In an effort to consolidate some of the better structure validation concepts and to update many of the reference parameters first proposed in the early 1990s, we have developed a 'one-stop' protein structure validation server called PROSESS (PROtein Structure Evaluation Suite and Server). In creating PROSESS, we also decided to add new and improved NMR structure validation capabilities, since this is an area that has been neglected for nearly 15 years. PROSESS integrates a variety of previously developed and thoroughly tested methods to evaluate protein structures at both a global and a residue-specific level. Using more than 100 measurement criteria, PROSESS assesses covalent and geometric quality, non-bonded/packing quality, torsion angle quality—and if NMR data are available—chemical shift and NOE quality. PROSESS uses a number of locally developed programs including VADAR (8), GeNMR (13), RCI (14), ShiftX (15), SuperPose (16) and PREDITOR (17) along with several other external programs including MolProbity (9), Xplor-NIH (18) and NAMD (19) to perform its calculations. PROSESS produces detailed tables, explanations, structural images and graphs that summarize the results and compare them to values observed in high-quality or high-resolution protein structures. Using a simplified red–amber–green (RAG) coloring scheme, PROSESS also alerts users about both general and residue-specific structural problems. Relative to other structure validation tools, PROSESS appears to be unique in the breadth and depth of its structural assessments (Table 1). A more detailed description of PROSESS follows.

PROGRAM DESCRIPTION

PROSESS is composed of two parts, a front-end web-interface (written in Perl and HTML) and a back-end for calculation (written in Java, Python, C and Fortran). As with most servers, PROSESS has a data entry page (Home), a Help page, a Tutorial page, an Input Format page, an Output Format page and a Contact page, each of which can be accessed through a menu bar located at the top of each page. The PROSESS server requires either a PDB formatted file (for newly determined structures) or a PDB accession number (for previously determined structures) as input. The PDB files may consist of a single protein structure or chain or an ensemble of structures (up to 100) from an NMR structure calculation. The maximum number of residues is 10 000. Users may optionally add or paste the protein's sequence (in FASTA format), a chemical shift file (in BMRB or Shifty format), an NOE data file (in Xplor/CNS format) or any combination of the above. Detailed descriptions, along with examples of the allowable formats are given through hyperlinks to the PROSESS Input Format page. The back-end for PROSESS consists of more than a dozen different programs, many of which were developed and extensively tested in our laboratory over the past 10 years. These include VADAR (8) for coordinate, atomic packing, H-bond, secondary structure and geometric analysis, GeNMR (13) for calculating non-covalent, threading and solvent energetics, ShiftX (15) for calculating chemical shift correlations, RCI (14) for correlating structure mobility to chemical shifts, PREDITOR (17) for calculating torsion angle-chemical shifts agreement and SuperPose (16) for evaluating structure similarities to known homologues. A number of other programs for calculating and comparing bond lengths, bond angles, H-bond planarity, volume variability and B-factor quality were also developed locally and added to the PROSESS back end. PROSESS also incorporates several other externally developed programs including MolProbity (9) to assess atomic clashes, REDUCE (20) to identify His/Asn/Gln flips, Xplor-NIH (18) to identify and quantify NOE restraint violations and NAMD (19) to assess various energetic parameters. PROSESS is hosted on an Apache server (version 2.2.14) using a Linux operating system (Fedora Core 10). The server is equipped with two Intel Pentium 4 processors (2.8 GHz each) and 4 GB of physical memory. PROSESS is platform independent and has been tested successfully on Internet Explorer 8.0, Mozilla Firefox 3.0 and Safari 4.0.

PROSESS OUTPUT

Once the appropriate data files have been submitted, PROSESS returns an access hyperlink so that users may retrieve their output at a later time (data is securely stored on the site for up to 2 weeks). Alternately users can wait for the results to be presented on their computer screen. A typical PROSESS run takes 3–5 min. A screen-shot montage illustrating the typical output from a PROSESS run is shown in Figure 1. Every PROSESS output is divided into four 'clickable' pages: (i) Global Structure Assessment (GSA); (ii) Local (Per-residue) Structure Assessment; (iii) Graphs and Figures and (iv) Similarity Assessment. At the top of each output page is a summary of the protein structure providing the date of submission, name of the protein, number of residues, secondary structure content and other data. Below this summary is a set of graphs or tables that is specific to each of the four assessment pages.

The GSA page contains both images and tables. At the top of the GSA page are a set (4–6, depending on the input) of colored bars with numerical (0 for worst, 10

Table 1. Comparison of different protein structure evaluation programs and servers

	ProCheck	ProCheck NMR	WhatCheck (Whatif)	MolProbity	VADAR	OOPS	PROSESS
Program or Server	Program	Program	Program & Server	Server & Program	Server & Program	Program	Server
Protein evaluation	Yes	Yes	Yes	Yes	Yes	Yes	Yes
DNA/RNA evaluation	No	No	Partial	Yes	No	No	No
Handles NMR (NOE/δ) data	No	Yes	No	No	No	No	Yes
Secondary structure calc.	Yes	Yes	Yes	No	Yes	No	Yes
Bond length check	Yes	Yes	Yes	Yes	No	No	Yes
Bond angle check	Yes	Yes	Yes	Yes	No	No	Yes
Planarity and Chiral check	Yes	Yes	Yes	No	No	Yes	No
H-bond check	Yes	Yes	Yes	Yes	Yes	No	Yes
Volume check & calculation	No	No	No	No	Yes	No	Yes
Surface area check & calc	No	No	Yes	No	Yes	No	Yes
Heavy atom bump check	No	No	Yes	Yes	No	No	Yes
H-atom bump check (clash)	No	No	No	Yes	No	No	Yes
His/Asn/Gln flip check	No	No	No	Yes	No	No	Yes
VDW energy calculation	No	No	No	No	No	No	Yes
Threading energy calc.	No	No	No	No	Yes	No	Yes
B-factor check or correlation	No	No	Yes	Yes	No	Yes	Yes
Ramachandran check (global)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ramachandran check (res spc)	No	No	No	Yes	No	Yes/No	Yes
Side chain torsion check	Yes	Yes	Yes	Yes	Yes	Yes	Yes
NOE statistics calculations	No	Yes	No	No	No	No	Yes
NOE violation check	No	Yes	No	No	No	No	Yes
Chemical shift check	No	No	No	No	No	No	Yes
Ensemble RMSF check	No	No	No	No	No	No	Yes
Graphs & plots	Yes	Yes	Yes	No	Yes	Yes	Yes
Structure images/maps	No	No	Yes	Yes	No	No	Yes

for best) and color-coded assessments of the global structure quality. These RAG color bars are intended to provide users with a quick overview of the protein's structure from the perspective of its (i) overall quality; (ii) covalent and geometric quality; (iii) non-covalent/packing quality; (iv) torsion angle quality; (v) chemical shift quality and (vi) NOE quality. While it is difficult to synthesize dozens of different global scores into a single 'quality' value, we have attempted to do so using the following protocol. A protein's score for any given category (i.e. covalent, non-covalent, torsion, chemical shift and NOE) is determined relative to what has been measured for a set of 850 high-quality (<2.0 Å resolution) X-ray structures and, if chemical shifts or NOEs are provided, a set of 250 high-quality NMR structures (i.e. a *Z*-score). This *Z*-score is then scaled so that it can be expressed as a value from 0 to 10. Up to five different category scores can be calculated for a given a protein. These category scores are then used to calculate the overall score. To calculate an 'unscaled' overall score, the lowest category score (from the covalent, non-covalent, torsion, chemical shift and NOE category scores) is always more heavily weighted than the other category scores. Specifically, the 'unscaled' overall score = 0.5*(lowest score) + 0.5*(average of all other scores). This 'unscaled' overall score is then compared against the distribution of 'unscaled' overall scores previously calculated for PROSESS's set of high-quality structures to determine a *Z*-score. This *Z*-score is then scaled so that it can be expressed as a value from 0 to 10. Hence, if the 'unscaled' overall score of a given protein is 7.6, its 'scaled' overall score will be 9.5, since the 'unscaled' score falls within 0.25 standard deviations of the 'unscaled' overall score calculated for other high-quality structures. This weighting and calibration scheme allows structures that are uniformly good to have high overall scores, structures that are uniformly bad to have low overall scores and structures with one or two bad category scores to be scored somewhat lower than they might be via simple averaging.

Below each of the RAG color bars are hyperlinks that provide additional details, additional RAG graphs and more detailed explanations about PROSESS's scoring schema. The GSA page also provides a series of tables listing more than 90 calculated parameters that are broadly grouped into five general categories (covalent, non-covalent, torsion, chemical shift and NOE). Each parameter is hyperlinked to a brief description of that parameter that includes explanations of how this parameter may be used to detect an error or problem, how to find it, whether it may be important (or not) and potential methods for correcting it. The name of the program used to calculate that parameter is also provided. The value for the protein of interest is provided along with an expected value and a standard deviation determined from a set of 850 non-redundant, high-resolution (<2.0 Å resolution) X-ray structures and/or a set of 250 non-redundant, high quality NMR structures. If the calculated value is >2 SDs larger (or worse) than the ideal value it is flagged with red comment. Values that are within acceptable limits (<2 SD) are colored black. If an ensemble of NMR structures is provided, the GSA page provides averages and standard deviations calculated over the full set of structures.

The Local Structure Assessment page provides tables that assess the residue-specific properties of the protein.

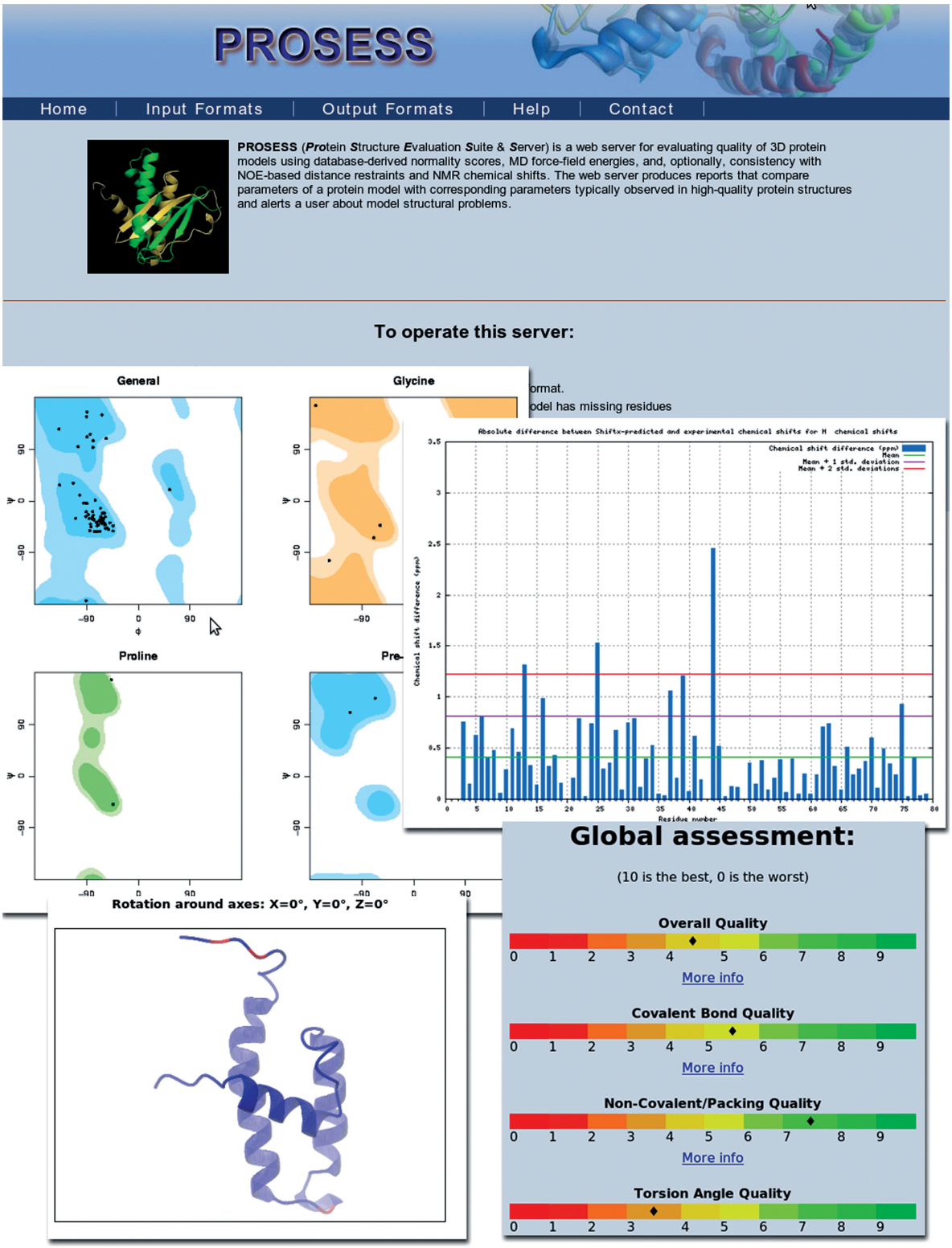

Figure 1. A screenshot montage of the PROSESS server showing examples of the different kinds of output that can be generated from a single run. Tables, charts, graphs and structural image-maps are all generated from a single PDB file.

Each residue is listed in a row and each property assessment is listed in a column. As with the GSA tables, descriptions for each property or parameter are hyperlinked to the name of that parameter. Each column is generally hyperlinked to a corresponding graph (particularly if outliers are identified). Several sets of local structure assessments are provided including: residue-specific aggregated or combined outliers, residue-specific main

chain or backbone evaluations, residue-specific side chain evaluations, residue-specific bond length/angle evaluations, residue-specific energies (H-bond, threading, covalent and non-covalent clashes); residue-specific chemical shift agreement(s) and residue-specific NOE violations. Values that exceed normally allowable limits (as previously described by the programs that calculate these values) are colored red. If an ensemble of NMR structures is provided, the Local Structure Assessment page calculates averages and standard deviations calculated over the full set of structures.

The Graphs and Figures (G & F) page provides a variety of visual output that summarizes the results from both the Global and Local Structure Assessment pages. At the top of each G & F page is an 'Aggregate Outlier' table displaying a set of summary plots and figures that map the protein's residue-specific problems. These expandable thumbnail images show both a bar graph and a set of colored 3D images of the protein. Each problem class for each problem residue in the bar graph is color-coded and annotated by the accompanying figure legend. This residue-specific information is also plotted on to the protein's 3D ribbon structure using a Yellow/Orange/Red coloring scheme. Residues highlighted in yellow have 1–2 problems, residues in orange have 3–4 problems and residues highlighted in red have 5 or more problems. Below this Aggregate Outlier table is another set of tables displaying thumbnail images and short titles (hyperlinked to explanations) so that users can navigate to different images, graphs or plots. Once clicked, the images expand to colorful, full-screen PNG images. These PNG images can also be viewed or downloaded as Postscript or PDF images so that users can paste these results directly into papers or reports. The first set of thumbnails is a collection of Covalent Quality plots. These are followed by the Packing Quality, Non-covalent Quality and Torsion Quality plots. These local structure assessment plots, are typically shown as two kinds of bar graphs (an 'outlier' plot and a 'parameter value' plot), with the parameter or outlier value on the *Y*-axis and the residue number displayed on the *X*-axis. Each graph and axis is titled to allow for easy identification. In addition to these bar graphs are a series of static ribbon diagrams generated via MolMol (21) that highlight the structural location of any local torsion, bond, packing, shift or NOE violations. Interactive images of the color-coded structures are also available using the JMol (22) applet. In addition to the standard bar graphs or histograms, a set of Ramachandran plots (with the Torsion Quality plots) is also provided to map the location of backbone torsion angles for all residues, for glycine-only residues, for proline-only residues and for pre-proline residues. These plots highlight the residues in the core, allowed and disallowed regions of Ramachandran space.

The Similarity Assessment page summarizes the results of BLAST searches of the protein sequence against the PDB. Those structures with Expect values $<10^{-7}$ are listed, along with their resolution/R_{free} values (if available). The calculated RMSD between the input structure and the related structures is calculated and displayed. Those structures that are significantly different from related structures (according to their RMSD and sequence identity) are flagged. The purpose of the Similarity Assessment page is to help users identify if their structure is already similar to something already solved and if it is, whether there may be structural differences that may be cause for concern.

CRITERIA FOR STRUCTURE ASSESSMENT

The computational assessment and validation of protein structures is an evolving process. Many of the most widely used systems such as PROCHECK, WHATCHECK, VADAR and OOPS were developed in the early 1990s using assessment criteria that seemed to be very good at the time. As the PDB has grown, as the quality of structures has improved and as structure solving/prediction methods have gotten better, a number of these early 1990s evaluation criteria have become obsolete or dated. This is particularly true regarding Ramachandran statistics of allowed and disallowed backbone torsion angles (9). Likewise a number of newer and better methods for assessing and validating structures have been discovered, such as the evaluation of threading energies, the measurement of hydrogen atom overlaps, the assessment of side chain stereochemistry, the use of B-factor assessment as well as atomic packing and cavity assessment. At the same time, new NOE-independent methods for solving NMR structures have been developed (23,24), leading to new challenges with regard to assessing and validating NMR structures.

In developing PROSESS, we conducted several broad surveys of the literature and evaluated many of the programs and validation criteria ourselves. From these assessments, we identified a number of parameters and evaluation criteria that seemed to be particularly robust and relatively up-to-date. These were incorporated into PROSESS so that it could robustly handle three kinds of structures (X-ray, NMR and predicted) as well as three kinds of structural problems: (i) misfolded structures; (ii) structures with generally poor stereochemistry; and (iii) good structures with localized problems.

To identify seriously misfolded structures (*ab initio* predicted structures, reversed chains, wrong space group, large topological errors and fraudulent structures) it is often necessary to use threading energy evaluations or homologous structure comparison methods. PROSESS uses a threading energy calculated from VADAR (8) as well as one calculated using GeNMR (13). Both methods have been assessed previously (8,13) and both are able to consistently distinguish misfolded structures from properly folded structures. PROSESS also uses a structure comparison method based on SuperPose (16) to identify structures that appear to be statistically unlike their close homologues. The criteria for flagging these structures are based on the well-known sequence identity-RMSD equation calculated by Chothia and Lesk (25).

To identify structures with generally poor stereochemistry or geometry PROSESS relies on nearly 70 global measurement criteria. These criteria are well-known and are generally widely used, although most individual

structure validation programs only use a fraction (~15–20) of those used by PROSESS. PROSESS assesses and identifies 'poor' structures by calculating the numbers of bond length violations, bond angle violations, Ramachandran outliers, H-bond energy violations, H-bond geometry inconsistencies, total van der Waals energy, numbers of bumps, side-chain flips or clashes, etc. To ensure consistency and improve the assessment criteria, we have collected updated statistics on a set of 850 non-redundant, high-resolution X-ray structures (Res. <2.0 Å) and used updated assessment criteria from a number of recent publications (4,8–10,13,18). This includes the use of updated covalent and non-covalent bond geometry statistics as well as updated Ramachandran statistics for non-glycine backbone torsion angles, glycine-only torsion angles, proline-only torsion angles and for pre-proline torsion angles (9).

Good structures with localized problems tend to 'fly under the radar' of most global structure assessment checks. Consequently, PROSESS uses local or residue-specific evaluations to identify these kinds of problems. Specifically, PROSESS identifies and flags individual bond length violations, bond angle violations, Ramachandran outliers, H-bond energy violations, H-bond geometry inconsistencies, disallowed pairwise contacts or clashes, volume or cavity violations, etc. Again, these kinds of local assessment criteria along with the appropriate cut-off values are generally well-known and widely used (4,8–10,13), although PROSESS uses a much more extensive set of criteria than most other programs. To simplify the identification of these local problems PROSESS uses color-coded tables, histograms and color-coded images of the protein structure under evaluation.

PROSESS and NMR STRUCTURES

A key strength of PROSESS lies in its ability to handle and interpret NMR data. Unlike X-ray crystallography where structures can be assessed directly against experimental diffraction data (as measured by the R_{free} value), NMR structures cannot be so easily assessed. Many constraints, prior knowledge and other parameters go into the construction of an NMR structure, making direct assessment to experimental NOE data quite difficult (6,18,26). Of the existing set of 'dedicated' protein structure validation programs only PROCHECK-NMR handles experimental NMR (i.e. NOE) data. Several other 'niche' programs such as RPF (26) or dedicated structure generation programs such as CYANA (27) and Xplor-NIH (18) can also analyze NOE data and identify problem NOEs. However, RPF, CYANA and Xplor-NIH do not perform most other kinds of structure validation checks.

As with other NMR structure validators, PROSESS not only uses NOEs to validate and assess protein structures, but also uses chemical shifts. Chemical shifts are much more precisely measurable and far more reproducibly measured than NOEs (15). Furthermore, several recently developed programs now allow protein structures to be determined independently of NOEs (23,24). These facts suggest that chemical shifts are a better and potentially more universal way of assessing and validating protein structures than NOEs (15,28). PROSESS uses SHIFTX (15) to calculate the correlation between observed and calculated chemical shifts as a measure of global structure agreement. It also uses SHIFTX to identify problem residues that have chemical shifts that appear to be inconsistent with the local structure. Criteria previously developed for the SHIFTCOR program (28) are used to flag problematic structures, shifts or residues. Another advantage of using chemical shifts lies in their utility for assessing ensembles of NMR structures. It has been previously shown that the chemical shift derived 'Random Coil Index' or RCI is strongly correlated with the RMSF (root mean square fluctuation) of NMR structure ensembles (14). It has also been shown that inconsistencies between the RCI value and the RMSF values are good indicators of local conformational sampling (i.e. over-sampling or under-sampling) problems that occurred in generating NMR ensembles (14). As a result, PROSESS uses the RCI method to identify and flag these local sampling problems.

STRENGTHS AND LIMITATIONS

PROSESS is unique at a number of levels. First, it provides a much more comprehensive assessment of protein structure quality than other structure validation programs or servers (Table 1). Second, PROSESS is capable of identifying and comprehensively assessing three types of structures: (i) misfolded structures; (ii) structures with generally poor stereochemistry; and (iii) good structures with localized problems. Most other structure validation servers/programs are only capable of handling one or two of these types of problem structures. Third, PROSESS incorporates BLAST searches and structure comparisons to homologous structures to identify potential structural disagreements or problems. This is not found in any structure assessment/validation structure we are aware of. Fourth, PROSESS uses a number of unique structure assessment criteria including hydrogen bond planarity, packing defect detection, B-factor analysis, RCI-ensemble analysis, threading energy evaluation, chemical shift correlation and others. While some of these criteria have been used or tested in certain specialized applications, they have not been routinely incorporated into any structure validation tools. Fifth, PROSESS uniquely provides detailed, hyperlinked explanations and/or expected values for each feature or parameter. In addition to these unique features, PROSESS also borrows some of the better ideas from existing structure validation tools, including the use of updated Ramachandran statistics and plots (8,9), the use of H-atom clash scores (9), the extensive use of graphs, charts and tables (5,6,9), and the generation of color-coded as well as color-mapped structure images (4,6,9).

Of course PROSESS is not without some shortcomings. Being a protein-only analysis system, PROSESS currently ignores DNA, RNA and small molecule ligands. Likewise,

it does not perform nomenclature, file formatting or labeling checks. Additionally, PROSESS does not evaluate the quality of packing among solvent molecules nor is it able to process multi-chain protein complexes as a single entry. This latter issue is being fixed and should be resolved by May 2010. It is also notable that PROSESS does not (yet) provide individual structure assessments for each structure in an NMR ensemble. This issue is also being addressed. One potential complaint is that PROSESS provides users with too much information. This is certainly a fair criticism. Nevertheless, when looking at something as complex and potentially important as a protein structure, we believe that it is generally better to provide too much information rather than too little.

CONCLUSION

The recent identification and withdrawal of 12 fraudulent X-ray structures from the PDB (2), has highlighted the need to develop better and more powerful structure validation tools. Certainly, the prevention of scientific fraud is an important goal for structure validation, but so too is the prevention of innocent mistakes or sloppy research. Over the years, through the use of programs like PROCHECK and Verify3D, dozens of erroneous structures have been identified, fixed or withdrawn from the PDB (1–6). Clearly structure validation software is needed not only by database curators, but also by the 'producers' and 'consumers' of structural data as well. The 'producers' are the X-ray crystallographers, NMR spectroscopists and computational modelers who are trying to generate the highest quality structures they can. Certainly structure validation software can prevent errors from creeping in during the refinement process or prevent the embarrassment of publishing or depositing a faulty structure. The 'consumers' are modelers, enzymologists, structural biologists and drug designers who need the best or most correct structures available. Identifying the highest quality structures gives them greater certainty that the trends they discover are real and reproducible. While PROSESS is certainly not the first structure validation tool to be produced, we believe that it offers curators, consumers and producers of structural data a better, more integrated and more informative approach to identify and prevent costly scientific errors.

ACKNOWLEDGEMENTS

The authors wish to thank Savita Shrivastava for her help in designing the PROSESS web page.

FUNDING

Alberta Prion Research Institute; PrioNet; Natural Sciences and Engineering Research Council; Genome Alberta. Funding for open access charge: Alberta Prion Research Institute.

Conflict of interest statement. None declared.

REFERENCES

1. Jeffrey,P.D. (2009) Analysis of errors in the structure determination of MsbA. *Acta Crystallogr. D Biol. Crystallogr.*, **65(Pt 2)**, 193–199.
2. UAB Statement on Protein Databank Issues. http://main.uab.edu/Sites/reporter/articles/71570/ (accessed 11February 2010).
3. Janssen,B.J., Read,R.J., Brünger,A.T. and Gros,P. (2007) Crystallography: crystallographic evidence for deviating C3b structure. *Nature*, **448**, E1–E2.
4. Hooft,R.W.W., Vriend,G., Sander,C. and Abola,E.E. (1996) Errors in protein structures. *Nature*, **381**, 272.
5. Laskowski,R.A., MacArthur,M.W., Moss,D.S. and Thornton,J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283–291.
6. Laskowski,R.A., Rullmannn,J.A., MacArthur,M.W., Kaptein,R. and Thornton,J.M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR*, **8**, 477–486.
7. Kleywegt,G.J. and Jones,T.A. (1996) Efficient rebuilding of protein structures. *Acta Crystallogr. D Biol. Crystallogr.*, **52(Pt 4)**, 829–832.
8. Willard,L., Ranjan,A., Zhang,H., Monzavi,H., Boyko,R.F., Sykes,B.D. and Wishart,D.S. (2003) VADAR: a web server for quantitative evaluation of protein structure quality. *Nucleic Acids Res.*, **31**, 3316–3319.
9. Davis,I.W., Leaver-Fay,A., Chen,V.B., Block,J.N., Kapral,G.J., Wang,X., Murray,L.W., Arendall,W.B. III, Snoeyink,J., Richardson,J.S. *et al.* (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.*, **35(Web Server issue)**, W375–W383.
10. Pontius,J., Richelle,J. and Wodak,S.J. (1996) Deviations from standard atomic volumes as a quality measure for protein crystal structures. *J. Mol. Biol.*, **264**, 121–136.
11. Lüthy,R., Bowie,J.U. and Eisenberg,D. (1992) Assessment of protein models with three-dimensional profiles. *Nature*, **356**, 83–85.
12. Wiederstein,M. and Sippl,M.J. (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.*, **35(Web Server issue)**, W407–W410.
13. Berjanskii,M., Tang,P., Liang,J., Cruz,J.A., Zhou,J., Zhou,Y., Bassett,E., MacDonell,C., Lu,P., Lin,G. *et al.* (2009) GeNMR: a web server for rapid NMR-based protein structure determination. *Nucleic Acids Res.*, **37(Web Server issue)**, W670–W677.
14. Berjanskii,M.V. and Wishart,D.S. (2008) Application of the random coil index to studying protein flexibility. *J. Biomol. NMR*, **40**, 31–48.
15. Neal,S., Nip,A.M., Zhang,H. and Wishart,D.S. (2003) Rapid and accurate calculation of protein 1H, 13C and 15N chemical shifts. *J. Biomol. NMR*, **26**, 215–240.
16. Maiti,R., Van Domselaar,G.H., Zhang,H. and Wishart,D.S. (2004) SuperPose: a simple server for sophisticated structural superposition. *Nucleic Acids Res.*, **32(Web Server issue)**, W590–W594.
17. Berjanskii,M.V., Neal,S. and Wishart,D.S. (2006) PREDITOR: a web server for predicting protein torsion angle restraints. *Nucleic Acids Res.*, **34**, W63–W69.
18. Schwieters,C.D., Kuszewski,J.J., Tjandra,N. and Clore,G.M. (2003) The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.*, **60**, 65–73.
19. Phillips,J.C., Braun,R., Wang,W., Gumbart,J., Tajkhorshid,E., Villa,E., Chipot,C., Skeel,R.D., Kalé,L. and Schulten,K. (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem.*, **26**, 1781–1802.
20. Word,J.M., Lovell,S.C., Richardson,J.S. and Richardson,D.C. (1999) Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J. Mol. Biol.*, **285**, 1735–1747.
21. Koradi,R., Billeter,M. and Wüthrich,K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.*, **14**, 51–55.
22. Jmol: an open-source Java viewer for chemical structures in 3D. http://www.jmol.org/ (11 February 2010, date last accessed).

23. Wishart,D.S., Arndt,D., Berjanskii,M., Tang,P., Zhou,J. and Lin,G. (2008) CS23D: a web server for rapid protein structure generation using NMR chemical shifts and sequence data. *Nucleic Acids Res.*, **36(Web Server issue)**, W496–W502.
24. Shen,Y., Lange,O., Delaglio,F., Rossi,P., Aramini,J.M., Liu,G., Eletsky,A., Wu,Y., Singarapu,K.K., Lemak,A. *et al.* (2008) Consistent blind protein structure generation from NMR chemical shift data. *Proc. Natl Acad. Sci. USA*, **105**, 4685–4690.
25. Chothia,C. and Lesk,A.M. (1986) The relation between the divergence of sequence and structure in proteins. *EMBO J.*, **5**, 823–826.
26. Huang,Y.J., Powers,R. and Montelione,G.T. (2005) Protein NMR recall, precision, and F-measure scores (RPF scores): structure quality assessment measures based on information retrieval statistics. *J. Am. Chem. Soc.*, **127**, 1665–1674.
27. Güntert,P. (2004) Automated NMR structure calculation with CYANA. *Methods Mol. Biol.*, **278**, 353–378.
28. Zhang,H., Neal,S. and Wishart,D.S. (2003) RefDB: a database of uniformly referenced protein chemical shifts. *J. Biomol. NMR*, **25**, 173–195.