Report of protein analysis

By the WHAT IF program

2020-03-21

1 Introduction

This document is a WHAT_CHECK 14.0 report for a PDB-file. Each reported fact has an assigned severity, one of:

error : Items marked as errors are considered severe problems requiring immediate attention.

warning : Either less severe problems or uncommon structural features. These still need special attention.

note : Statistical values, plots, or other verbose results of tests and analyses that have been performed.

If alternate conformations are present, only the first is evaluated. Hydrogen atoms are only included if explicitly requested, and even then they are not used in all checks. The software functions less well for non-canonical amino acids and exotic ligands than for the 20 canonical residues and canonical nucleic acids.

1.1 Some remarks regarding the output:

Residue. Residues/atoms in tables are normally given in a few parts:

- A number. This is the internal sequence number of the residue used by WHAT IF. The first residues in the file get number 1, 2, etc.
- The residue type. Normally this is a three letter amino acid type.
- The sequence number, between brackets. This is the residue number as it was given in the input file. It can be followed by the insertion code.
- The chain identifier. A single character. If no chain identifier was given in the input file, this will be a minus sign or a blank.
- A model number. If no model number exists, like in most X-ray files, this will be a blank or occasionally a minus sign.
- In case an atom is part of the output, the atom will be listed using the PDB nomenclature for type and identifier.

Z-Value. To indicate the normality of a score, the score may be expressed as a Z-value or Z-score. This is just the number of standard deviations that the score deviates from the expected value. A property of Z-values is that the root-mean-square of a group of Z-values (the RMS Z-value) is expected to be 1.0. Z-values above 4.0 and below -4.0 are very uncommon. If a Z-score is used in WHAT IF, the accompanying text will explain how the expected value and standard deviation were obtained.

Nucleic acids. The names of nucleic acids are DGUA, DTHY, OCYT, OADE, etc. The first character is a D or O for DNA or RNA respectively. This is done to circumvent ambiguities in the many old PDB files in which DNA and RNA were both called A, C, G, and T.

2 6y2e.pdb

2.1 Checks that need to be done early-on in validation

2.1.1 Note: Introduction

WHAT CHECK needs to read a PDB file before it can check it. It does a series of checks upon reading the file. The results of these checks are reported in this section (section 2.1). The rest of the report will be more systematic in that section 2.2 reports on administrative problems. Section 2.3 gives descriptive output that is not directly validating things but more telling you how WHAT CHECK interpreted the input file. Section 2.4 looks at B-factors, occupancies, and the presence/absence of (spurious) atoms. Section 2.5 deals with nomenclature problems. Section 2.6 deals with geometric problems like bond lengths and bond angles. Section 2.7 deals with torsion angle issues. Section 2.8 looks at atomic clashes. Section 2.9 deals with packing, accessibility, etc, issues. Section 2.10 deals with hydrogen bonds, ion packing, and other things that can be summarized under the common name charge-charge interactions. Section 2.11 gives a summary of whole report and tells you (if applicable) which symmetry matrices were used. Section 2.12 tells the crystallographer which are the things most in need of manual correction. And the last section, section 2.13, lists all residues sorted by their need for visual inspection in light of the electron density.

2.1.2 Note: Header records from PDB file

Header records from PDB file.

PDB header info
VIRAL PROTEIN

2.1.3 Note: Counting molecules and matrices

The parameter Z as given on the CRYST card represents the molecular multiplicity in the crystallographic cell. Z equals the number of matrices of the space group multiplied by the number of NCS relations. These numbers seem to be consistent.

Space group as read from CRYST card: C 1 2 1
Number of matrices in space group: 4
Highest polymer chain multiplicity in structure: 1
Highest polymer chain multiplicity according to SEQRES: 1
No explicit MTRIX NCS matrices found in the input file
Value of Z as found on the CRYST1 card: 4
Z, spacegroup, and NCS seem to agree administratively

2.1.4 Note: Matthews coefficient OK

The Matthews coefficient [REF] is defined as the density of the protein structure in cubic Angstroms per Dalton. Normal values are between 1.5 (tightly packed, little room for solvent) and 4.0 (loosely packed, much space for solvent). Some very loosely packed structures can get values a bit higher than that.

Molecular weight of all polymer chains: 33779.035
Volume of the Unit Cell V= 271478.500
Space group multiplicity: 4
No NCS symmetry matrices (MTRIX records) found in PDB file
Matthews coefficient for observed atoms and Z: Vm= 2.009
No Matthews coefficient given in REMARK 280

2.1.5 Note: All atoms are sufficiently far away from symmetry axes

None of the atoms in the structure is closer than 0.77 Å to a proper symmetry axis.

2.1.6 Note: Chain identifiers OK

WHAT CHECK has not detected any serious chain identifier problems. But be aware that WHAT CHECK doesn't care about the chain identifiers of waters.

2.2 Administrative problems that can generate validation failures

2.2.1 Note: No strange inter-chain connections detected

No covalent bonds have been detected between molecules with non-identical chain identifiers.

2.2.2 Note: No duplicate atom names in ligands

All atom names in ligands (if any) seem adequately unique.

2.2.3 Note: In all cases the primary alternate atom was used

WHAT CHECK saw no need to make any alternate atom corrections (which means they either are all correct, or there are none).

2.2.4 Note: No residues detected inside ligands

Either this structure does not contain ligands with amino acid groups inside it, or their naming is proper (enough).

2.2.5 Note: No attached groups interfere with hydrogen bond calculations

It seems there are no sugars, lipids, etc., bound (or very close) to atoms that otherwise could form hydrogen bonds.

2.2.6 Note: No probable side chain atoms with zero occupancy detected.

Either there are no side chain atoms with zero occupancy, or the side chain atoms with zero occupancy were not present in the input PDB file (in which case they are listed as missing atoms), or their positions are sufficiently improbable to warrant a zero occupancy.

2.2.7 Note: No probable backbone atoms with zero occupancy detected.

Either there are no backbone atoms with zero occupancy, or the backbone atoms with zero occupancy were left out of the input PDB file (in which case they are listed as missing atoms), or their positions are sufficiently improbable to warrant a zero occupancy.

2.2.8 Note: All residues have a complete backbone.

No residues have missing backbone atoms.

2.2.9 Note: No C-alpha only residues

There are no residues that consist of only an α carbon atom.

2.3 Non-validating, descriptive output paragraph

2.3.1 Note: Content of the PDB file as interpreted by WHAT CHECK

Content of the PDB file as interpreted by WHAT CHECK. WHAT CHECK has read your PDB file, and stored it internally in what is called 'the soup'. The content of this soup is listed here. An extensive explanation of all frequently used WHAT CHECK output formats can be found at swift.cmbi.ru.nl. Look under output formats. A course on reading this 'Molecules' table is part of the WHAT CHECK website.

'Molecules'

1	1 (1)	306 (306)	A	Protein		To check
2	307 (HOH)	307 (HOH)	A	water	(313)	To check

2.3.2 Some numbers...

MODELs skipped upon reading PDB file: 0

X-ray structure. No MODELs found

The total number of amino acids found is 306.

No nucleic acids observed in input file

No sugars recognized in input file

Number of water molecules: 313

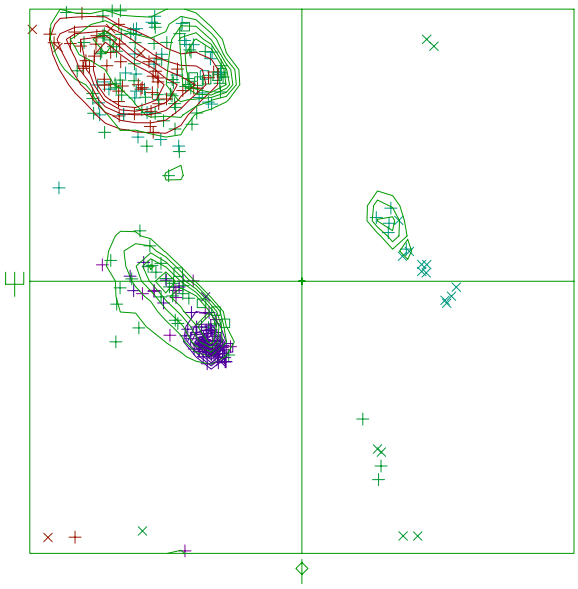
of which one is poor

Residue numbers increase monotonously OK

2.3.3 Note: Ramachandran plot

In this Ramachandran plot x-signs represent glycines, squares represent prolines, and plus-signs represent the other residues. If too many plus-signs fall outside the contoured areas then the molecule is poorly refined (or worse). Proline can only occur in the narrow region around ϕ=-60 that also falls within the other contour islands.

In a colour picture, the residues that are part of a helix are shown in blue, strand residues in red. Preferred regions for helical residues are drawn in blue, for strand residues in red, and for all other residues in green. A full explanation of the Ramachandran plot together with a series of examples can be found at the WHAT CHECK website [REF].



Chain identifier: A

2.3.4 Note: Secondary structure

This is the secondary structure according to DSSP. Only helix (H), overwound or 3/10-helix (3), strand (S), turn (T) and coil (blank) are shown [REF]. All DSSP related information can be found at swift.cmbi.ru.nl/gv/dssp/ This is not really a structure validation option, but a very scattered secondary structure (i.e. many strands of only a few residues length, many Ts inside helices, etc) tends to indicate a poor structure. A full explanation of the DSSP secondary structure determination program together with a series of examples can be found at the WHAT CHECK website [REF].

Secondary structure assignment

```
        10        20        30        40        50        60
SGFRKMAFPSGKVEGCMVQVTCGTTTLNGLWLDDVVYCPRHVICTSEDMLNPNYEDLLIR
         HH 33H3TSSSST  T SSSS T  TSSST3333TT 33H3T THHHHHHH
        70        80        90       100       110       120
KSNHNFLVQAGNVQLRVIGHSMQNCVLKLKVDTANPKTPKYKFVRIQPGQTFSVLACYNG
TT3 3TSSST  T     SSS T  TSSSS    T  T SSSSS  T TT SSSSS TTT
       130       140       150       160       170       180
SPSGVYQCAMRPNFTIKGSFLNGSCGSVGFNIDYDCVSFCYMHHMELPTGVHAGTDLEGN
T  SSSSS  T  T S    T T3 TT SSS T  TSSS SS  S T  T SSSS3  T
       190       200       210       220       230       240
FYGPFVDRQTAQAAGTDTTITVNVLAWLYAAVINGDRWFLNRFTTTLNDFNLVAMKYNYE
                  STHHHHHHHHHHHHHHT 3  T     THHHHHHHHHHHTSS
       250       260       270       280       290       300
PLTQDHVDILGPLSAQTGIAVLDMCASLKELLQNGMNGRTILGSALLEDEFTPFDVVRQC
  THHHHHHHHHHHHHHT THHHHHHHHHHHHHH 3TTT TTTT       THHHHHHHH

SGVTFQ
HT
```

2.4 Coordinate problems, unexpected atoms, B-factor and occupancy checks

2.4.1 Note: No rounded coordinates detected

No significant rounding of atom coordinates has been detected.

2.4.2 Note: No artificial side chains detected

No artificial side-chain positions characterized by χ-1=0.0 or χ-1=180.0 have been detected.

2.4.3 Note: No missing atoms detected in residues

All expected atoms are present in residues. This validation option has not looked at 'things' that can or should be attached to the elementary building blocks (amino acids, nucleotides). Even the C-terminal oxygens are treated separately.

2.4.4 Note: All B-factors fall in the range 0.0 - 100.0

All B-factors are larger than zero, and none are observed above 100.0.

2.4.5 Note: C-terminus capping

The residues listed in the table below are either C-terminal or pseudo C-terminal (i.e. last residue before a missing residue). In X-ray the coordinates must be located in density. Mobility or disorder sometimes cause this density to be so poor that the positions of the atoms cannot be determined. Crystallographers tend to leave out the atoms in such cases. In many cases the N- or C-terminal residues are too disordered to see. In case of the N-terminus, you can often see from the residue numbers if there are missing residues; at the C-terminus this is impossible. Therefore, often the position of the backbone nitrogen of the first residue missing at the C-terminal end is calculated and added to indicate that there are missing residues. As a single N causes validation trouble, we remove these single-N-residues before doing the validation. If this happened, the label -N is added to the pseudo C-terminus. Other labels can be +X in case something weird is bound to the backbone C, or +OXT if a spurious OXT atom is found. -OXT indicates that an expected OXT is missing. 'Swap' means that the O' and O" (O and OXT in PDB files) have been swapped in terms of nomenclature. 'Bad' means that something bad happened that WHAT IF does not understand. In such cases you might get three residue numbers in square brackets; one of those might be what WHAT IF had expected to find, but then it also might not). In case of chain breaks the number of missing residues is listen in round brackets. OK means what it suggests...

Be aware that we cannot easily see the difference between these errors and errors in the chain and residue numbering schemes. So do not blindly trust the table below. If you get weird errors at, or near, the left-over incomplete C-terminal residue, please check by hand if a missing Oxt or a removed single N is the cause. Also, many peptidic ligands get the same chain identifier as the larger protein they are bound to. In such cases there are more than one C-termini and OXTs with the same ID. WHAT IF gives some random warnings about these cases. So, don't take everything at face value, but think for yourself.

	Residue		Comment
306	GLN (306-) A	-	-OXT

2.4.6 Note: Weights administratively correct

All atomic occupancy factors ('weights') fall in the 0.0–1.0 range, which makes them administratively correct.

2.4.7 Note: Normal distribution of occupancy values

The distribution of the occupancy values in this file seems 'normal'.

Be aware that this evaluation is merely the result of comparing this file with about 500 well-refined high-resolution files in the PDB. If this file has much higher or much lower resolution than the PDB files used in WHAT CHECK's training set, non-normal values might very well be perfectly fine, or normal values might actually be not so normal. So, this check is actually more an indicator and certainly not a check in which I have great confidence.

2.4.8 Note: All occupancies seem to add up to 0.0 - 1.0.

In principle, the occupancy of all alternates of one atom should add up till 0.0 - 1.0. 0.0 is used for the missing atom (i.e. an atom not seen in the electron density). Obviously, there is nothing terribly wrong when a few occupancies add up to a bit more than 1.0, because the mathematics of refinement allow for that. However, if it happens often, it seems worth evaluating this in light of the refinement protocol used.

2.4.9 Warning: What type of B-factor?

WHAT CHECK does not yet know well how to cope with B-factors in case TLS has been used. It simply assumes that the B-factor listed on the ATOM and HETATM cards are the total B-factors. When TLS refinement is used that assumption sometimes is not correct. TLS seems not mentioned in the header of the PDB file. But anyway, if WHAT CHECK complains about your B-factors, and you think that they are OK, then check for TLS related B-factor problems first. In case this is not your own PDB file, but one downloaded from the PDB, then check the BDB (http://swift.cmbi.ru.nl/gv/facilities/) first.

Temperature not mentioned in PDB file. This most likely means that the temperature record is absent.
Room temperature assumed

2.4.10 Note: Number of buried atoms with low B-factor is OK

For protein structures determined at room temperature, no more than about 1 percent of the B factors of buried atoms is below 5.0. In liquid nitrogen this percentage is allowed to be higher, of course.

Percentage of buried atoms with B less than 5 : 0.00

2.4.11 Note: B-factor distribution normal

The distribution of B-factors within residues is within expected ranges. A value over 1.5 here would mean that the B-factors show signs of over-refinement.

RMS Z-score : 0.677 over 2115 bonds
Average difference in B over a bond : 1.82
RMS difference in B over a bond : 2.54

2.4.12 Note: B-factor plot

The average atomic B-factor per residue is plotted as function of the residue number.

Chain identifier: A

2.5 Nomenclature related problems

2.5.1 Note: Introduction to the nomenclature section.

Nomenclature problems seem, at first, rather unimportant. After all who cares if we call the δ atoms in leucine $\delta 2$ and $\delta 1$ rather than the other way around. Chemically speaking that is correct. But structures have not been solved and deposited just for chemists to look at them. Most times a structure is used, it is by software in a bioinformatics lab. And if they compare structures in which the one used C $\delta 1$ and $\delta 2$ and the other uses C $\delta 2$ and $\delta 1$, then that comparison will fail. Also, we recalculate all structures every so many years to make sure that everybody always can get access to the best coordinates that can be obtained from the (your?) experimental data. These recalculations will be troublesome if there are nomenclature problems.

Several nomenclature problems actually are worse than that. At the WHAT CHECK website [REF] you can get an overview of the importance of all nomenclature problems that we list.

2.5.2 Note: Valine nomenclature OK

No errors were detected in valine nomenclature.

2.5.3 Note: Threonine nomenclature OK

No errors were detected in threonine nomenclature.

2.5.4 Note: Isoleucine nomenclature OK

No errors were detected in isoleucine nomenclature.

2.5.5 Note: Leucine nomenclature OK

No errors were detected in leucine nomenclature.

2.5.6 Warning: Arginine nomenclature problem

The arginine residues listed in the table below have their NH1 and NH2 swapped.

	Residue	
4	ARG (4-) A	-
60	ARG (60-) A	-
217	ARG (217-) A	-
222	ARG (222-) A	-

2.5.7 Warning: Tyrosine convention problem

The tyrosine residues listed in the table below have their χ-2 not between -90.0 and 90.0

	Residue	
101	TYR (101-) A	-

2.5.8 Note: Phenylalanine torsion conventions OK

No errors were detected in phenylalanine torsion angle conventions.

2.5.9 Note: Aspartic acid torsion conventions OK

No errors were detected in aspartic acid torsion angle conventions.

2.5.10 Warning: Glutamic acid convention problem

The glutamic acid residues listed in the table below have their χ-3 outside the -90.0 to 90.0 range, or their proton on OE1 instead of OE2.

	Residue	
47	GLU (47-) A	-
270	GLU (270-) A	-

2.5.11 Note: Phosphate group names OK in DNA/RNA

No errors were detected in nucleic acid phosphate group naming conventions (or this structure contains no nucleic acids).

2.5.12 Note: Heavy atom naming OK

No errors were detected in the atom names for non-hydrogen atoms. Please be aware that the PDB wants us to deliberately make some nomenclature errors; especially in non-canonical amino acids.

2.5.13 Note: No decreasing residue numbers

All residue numbers are strictly increasing within each chain.

2.6 Geometric checks

2.6.1 Note: All bond lengths OK

All bond lengths are in agreement with standard bond lengths using a tolerance of 4 sigma (both standard values and sigma for amino acids have been taken from Engh and Huber [REF], for DNA/RNA from Parkinson et al [REF]).

2.6.2 Note: Normal bond length variability

Bond lengths were found to deviate normally from the standard bond lengths (values for Protein residues were taken from Engh and Huber [REF], for DNA/RNA from Parkinson et al [REF]).

RMS Z-score for bond lengths: 0.719
RMS-deviation in bond distances: 0.014

2.6.3 Warning: Possible cell scaling problem

Comparison of bond distances with Engh and Huber [REF] standard values for protein residues and Parkinson et al [REF] values for DNA/RNA shows a significant systematic deviation. It could be that the unit cell used in refinement was not accurate enough. The deformation matrix given below gives the deviations found: the three numbers on the diagonal represent the relative corrections needed along the A, B and C cell axis. These values are 1.000 in a normal case, but have significant deviations here (significant at the 99.99 percent confidence level)

There are a number of different possible causes for the discrepancy. First the cell used in refinement can be different from the best cell calculated. Second, the value of the wavelength used for a synchrotron data set can be miscalibrated. Finally, the discrepancy can be caused by a dataset that has not been corrected for significant anisotropic thermal motion.

Please note that the proposed scale matrix has NOT been restrained to obey the space group symmetry. This is done on purpose. The distortions can give you an indication of the accuracy of the determination.

If you intend to use the result of this check to change the cell dimension of your crystal, please read the extensive literature on this topic first. This check depends on the wavelength, the cell dimensions, and on the standard bond lengths and bond angles used by your refinement software.

SCALE matrix obtained from PDB file

$$\begin{bmatrix} 0.008697 & 0.000000 & 0.001729 \\ 0.000000 & 0.018600 & 0.000000 \\ 0.000000 & 0.000000 & 0.022771 \end{bmatrix}$$

Unit Cell deformation matrix

$$\begin{bmatrix} 0.995885 & -0.000027 & -0.000007 \\ -0.000027 & 0.994596 & -0.001239 \\ -0.000007 & -0.001239 & 0.995278 \end{bmatrix}$$

Proposed new scale matrix

$$\begin{bmatrix} 0.008733 & 0.000002 & 0.001737 \\ 0.000000 & 0.018701 & 0.000023 \\ 0.000000 & 0.000028 & 0.022879 \end{bmatrix}$$

With corresponding cell

$A =$	114.509	$B =$	53.473	$C =$	44.565
$\alpha =$	90.000	$\beta =$	101.251	$\gamma =$	90.000

The CRYST1 cell dimensions

$A =$	114.983	$B =$	53.763	$C =$	44.774
$\alpha =$	90.000	$\beta =$	101.240	$\gamma =$	90.000

Variance: 224.808
(Under-)estimated Z-score: 11.050

2.6.4 Warning: Unusual bond angles

The bond angles listed in the table below were found to deviate more than 4 sigma from standard bond angles (both standard values and sigma for protein residues have been taken from Engh and Huber [REF], for DNA/RNA from Parkinson et al [REF]). In the table below for each strange angle the bond angle and the number of standard deviations it differs from the standard values is given. Please note that disulphide bridges are neglected. Atoms starting with "-" belong to the previous residue in the sequence.

Residue			Atom Triplet			Bond Angle	Z-value
64	HIS (64-) A	-	CG	ND1	CE1	109.68	4.1
163	HIS (163-) A	-	CA	CB	CG	108.18	-5.6
164	HIS (164-) A	-	CG	ND1	CE1	109.79	4.2
304	THR (304-) A	-	C	CA	CB	118.91	4.6

2.6.5 Note: Normal bond angle variability

Bond angles were found to deviate normally from the mean standard bond angles (normal values for protein residues were taken from Engh and Huber [REF], for DNA/RNA from Parkinson et al [REF]). The RMS Z-score given below is expected to be near 1.0 for a normally restrained data set, and this is indeed observed for very high resolution X-ray structures.

RMS Z-score for bond angles: 1.029
RMS-deviation in bond angles: 1.882

2.6.6 Error: Nomenclature error(s)

Checking for a hand-check. WHAT CHECK has over the course of this session already corrected the handedness of atoms in several residues. These were administrative corrections. These residues are listed here.

Residue		
4	ARG (4-) A	-
47	GLU (47-) A	-
60	ARG (60-) A	-
217	ARG (217-) A	-
222	ARG (222-) A	-
270	GLU (270-) A	-

2.6.7 Note: Chirality OK

All protein atoms have proper chirality. But, look at the previous table to see a series of administrative chirality problems that were corrected automatically upon reading-in the PDB file.

2.6.8 Note: Improper dihedral angle distribution OK

The RMS Z-score for all improper dihedrals in the structure is within normal ranges.

Improper dihedral RMS Z-score : 1.067

2.6.9 Note: Tau angles OK

All of the tau angles (N-C-α-C) of amino acids fall within expected RMS deviations.

2.6.10 Note: Normal tau angle deviations

The RMS Z-score for the tau angles (N-C-α-C) in the structure falls within the normal range that we guess to be 0.5 - 1.5. Be aware, we determined the tau normal distributions from 500 high-resolution X-ray structures, rather than from CSD data, so we cannot be 100 percent certain about these numbers.

Tau angle RMS Z-score : 1.022

2.6.11 Note: Side chain planarity OK

All of the side chains of residues that have an intact planar group are planar within expected RMS deviations.

2.6.12 Note: Atoms connected to aromatic rings OK

All of the atoms that are connected to planar aromatic rings in side chains of amino-acid residues are in the plane within expected RMS deviations.

2.7 Torsion-related checks

2.7.1 Note: Ramachandran Z-score OK

The score expressing how well the backbone conformations of all residues correspond to the known allowed areas in the Ramachandran plot is within expected ranges for well-refined structures.

Ramachandran Z-score : -1.706

2.7.2 Note: Ramachandran check

The list contains per-residue Z-scores describing how well each residue fits into the allowed areas of the Ramachandran plot will not be printed because WHAT CHECK found no reason to cry.

2.7.3 Warning: Torsion angle evaluation shows unusual residues

The residues listed in the table below contain bad or abnormal torsion angles.

These scores give an impression of how 'normal' the torsion angles in protein residues are. All torsion angles except ω are used for calculating a 'normality' score. Average values and standard deviations were obtained from the residues in the WHAT CHECK database. These are used to calculate Z-scores. A residue with a Z-score of below -2.0 is poor, and a score of less than -3.0 is worrying. For such residues more than one torsion angle is in a highly unlikely position.

	Residue	Z-score
225	THR (225-) A	- -2.6
154	TYR (154-) A	- -2.6
286	LEU (286-) A	- -2.4
200	ILE (200-) A	- -2.3
27	LEU (27-) A	- -2.2
177	LEU (177-) A	- -2.1
23	GLY (23-) A	- -2.1
43	ILE (43-) A	- -2.0
139	SER (139-) A	- -2.0

2.7.4 Warning: Backbone evaluation reveals unusual conformations

The residues listed in the table below have abnormal backbone torsion angles.

Residues with 'forbidden' ϕ-ψ combinations are listed, as well as residues with unusual ω angles (deviating by more than 3 sigma from the normal value). Please note that it is normal if about 5 percent of the residues is listed here as having unusual ϕ-ψ combinations.

	Residue	Description
2	GLY (2-) A	- Poor ϕ/ψ
8	PHE (8-) A	- Omega to (next) Pro poor
9	PRO (9-) A	- ω poor
23	GLY (23-) A	- Poor ϕ/ψ
33	ASP (33-) A	- Poor ϕ/ψ
38	CYS (38-) A	- Omega to (next) Pro poor
44	CYS (44-) A	- ω poor
51	ASN (51-) A	- Omega to (next) Pro poor
52	PRO (52-) A	- ω poor
70	ALA (70-) A	- ω poor
71	GLY (71-) A	- Poor ϕ/ψ
84	ASN (84-) A	- Poor ϕ/ψ
95	ASN (95-) A	- Omega to (next) Pro poor
98	THR (98-) A	- Omega to (next) Pro poor
99	PRO (99-) A	- ω poor

And so on for a total of 52 lines.

2.7.5 Error: Chi-1/chi-2 rotamer problems

List of residues with a poor χ-1/χ-2 combination. Be aware that for this validation option the individual scores are far less important than the overall score that is given below the table.

	Residue		Z-score
50	LEU (50-) A	-	-1.43
227	LEU (227-) A	-	-1.44
232	LEU (232-) A	-	-1.44
27	LEU (27-) A	-	-1.35
67	LEU (67-) A	-	-1.35
87	LEU (87-) A	-	-1.35
177	LEU (177-) A	-	-1.31
286	LEU (286-) A	-	-1.32
41	HIS (41-) A	-	-1.21
60	ARG (60-) A	-	-1.21
200	ILE (200-) A	-	-1.25
216	ASP (216-) A	-	-1.23
217	ARG (217-) A	-	-1.22
276	MET (276-) A	-	-1.23
47	GLU (47-) A	-	-1.18

And so on for a total of 114 lines.

2.7.6 Note: chi-1/chi-2 angle correlation Z-score OK

The score expressing how well the χ-1/χ-2 angles of all residues correspond to the populated areas in the database is within expected ranges for well-refined structures.

χ-1/χ-2 correlation Z-score : -2.947

2.7.7 Warning: Unusual rotamers

The residues listed in the table below have a rotamer that is not seen very often in the database of solved protein structures. This option determines for every residue the position specific χ-1 rotamer distribution. Thereafter it verified whether the actual residue in the molecule has the most preferred rotamer or not. If the actual rotamer is the preferred one, the score is 1.0. If the actual rotamer is unique, the score is 0.0. If there are two preferred rotamers, with a population distribution of 3:2 and your rotamer sits in the lesser populated rotamer, the score will be 0.667. No value will be given if insufficient hits are found in the database.

It is not necessarily an error if a few residues have rotamer values below 0.3, but careful inspection of all residues with these low values could be worth it.

	Residue		Fraction
42	VAL (42-) A	-	0.36
102	LYS (102-) A	-	0.37
162	MET (162-) A	-	0.37
80	HIS (80-) A	-	0.39

2.7.8 Warning: Unusual backbone conformations

For the residues listed in the table below, the backbone formed by itself and two neighbouring residues on either side is in a conformation that is not seen very often in the database of solved protein structures. The

number given in the table is the number of similar backbone conformations in the database with the same amino acid in the centre.

For this check, backbone conformations are compared with database structures using C-α superpositions with some restraints on the backbone oxygen positions.

A residue mentioned in the table can be part of a strange loop, or there might be something wrong with it or its directly surrounding residues. There are a few of these in every protein, but in any case it is worth looking at, especially if a regular DSSP secondary structure (H or S for helix or strand, respectively) is indicated!

	Residue		DSSP	# hits
182	TYR (182-) A	-		0
276	MET (276-) A	-	H	0
85	CYS (85-) A	-		1
154	TYR (154-) A	-		1
222	ARG (222-) A	-		1
145	CYS (145-) A	-		2
165	MET (165-) A	-	S	2

2.7.9 Note: Backbone conformation Z-score OK

The backbone conformation analysis gives a score that is normal for well refined protein structures.

Backbone conformation Z-score : -1.280

2.7.10 Warning: Omega angle restraints not strong enough

The ω angles for trans-peptide bonds in a structure is expected to give a gaussian distribution with the average around +178 degrees, and a standard deviation around 5.5. In the current structure the standard deviation of this distribution is above 7.0, which indicates that the ω values have been under-restrained.

Omega average and std. deviation= 178.750 7.579

2.7.11 Warning: Unusual PRO puckering amplitudes

The proline residues listed in the table below have a puckering amplitude that is outside of normal ranges. Puckering parameters were calculated by the method of Cremer and Pople [REF]. Normal PRO rings have a puckering amplitude Q between 0.20 and 0.45 Å. If Q is lower than 0.20 Å for a PRO residue, this could indicate disorder between the two different normal ring forms (with C-γ below and above the ring, respectively). If Q is higher than 0.45 Å something could have gone wrong during the refinement. Be aware that this is a warning with a low confidence level. See: Who checks the checkers? Four validation tools applied to eight atomic resolution structures [REF]

	Residue		Pucker Amplitude	Qualifier
96	PRO (96-) A	-	0.18	LOW

2.7.12 Note: PRO puckering phases OK

Puckering phases for all PRO residues are normal

2.7.13 Note: Backbone oxygen evaluation OK

All residues for which similar local backbone conformations could be found in the WHAT CHECK database

have a backbone oxygen position that has been observed at least a few times in that database.

2.7.14 Warning: Possible peptide flips

For the residues listed in the table below, the backbone formed by the residue mentioned and the one N-terminal of it show systematic deviations from normality that are consistent with a peptide flip. This can either be a 180 degree flip of the entire peptide plane or a trans to cis flip. (Cis to trans flips cannot be detected yet). The type can be TT+, TC-, or TC+: TT+ indicates a 180 degree flip of the entire peptide plane. TC- indicates a trans to cis conversion that requires a flip of the N atom. TC+ indicates a trans to cis conversion that requires a flip of the O atom. Note that the method will only work correctly for PDB files with full isotropic B-factors.

	Residue		Type	Comment
155	ASP (155-) A	-	TT+	Likely
195	GLY (195-) A	-	TT+	Likely

2.8 Bump checks

2.8.1 Error: Abnormally short interatomic distances

The pairs of atoms listed in the table below have an unusually short interactomic distance; each bump is listed in only one direction.

The contact distances of all atom pairs have been checked. Two atoms are said to 'bump' if they are closer than the sum of their Van der Waals radii minus 0.40 Å. For hydrogen bonded pairs a tolerance of 0.55 Å is used. The first number in the table tells you how much shorter that specific contact is than the acceptable limit. The second distance is the distance between the centres of the two atoms. Although we believe that two water atoms at 2.4 A distance are too close, we only report water pairs that are closer than this rather short distance.

INTRA and INTER indicate whether the clashes are between atoms in the same asymmetric unit, or atoms in symmetry related asymmetric units, respectively. The last text-item on each line represents the status of the atom pair. If the final column contains the text 'HB', the bump criterion was relaxed because there could be a hydrogen bond. Similarly relaxed criteria are used for 1–3 and 1–4 interactions (listed as 'B2' and 'B3', respectively). If the last column is 'BF', the sum of the B-factors of the atoms is higher than 80, which makes the appearance of the bump somewhat less severe because the atoms probably are not there anyway. BL, on the other hand, indicates that the bumping atoms both have a low B-factor, and that makes the bumps more worrisome.

Bumps between atoms for which the sum of their occupancies is lower than one are not reported. If the MODEL number does not exist (as is the case in most X-ray files), a minus sign is printed instead.

	Atom 1				Atom 2			Bump	Dist	Status	
218	TRP (218-) A	-	CE3	<–> 279	ARG (279-) A	-	CZ	0.24	2.96	INTRA	BF
221	ASN (221-) A	-	ND2	<–> 270	GLU (270-) A	-	CG	0.12	2.98	INTRA	BF
70	ALA (70-) A	-	N	<–> 73	VAL (73-) A	-	O	0.08	2.62	INTRA	BF
198	THR (198-) A	-	OG1	<–> 240	GLU (240-) A	-	OE2	0.06	2.34	INTRA	
40	ARG (40-) A	-	NE	<–> 187	ASP (187-) A	-	OD2	0.04	2.66	INTRA	BL
49	MET (49-) A	-	O	<–> 189	GLN (189-) A	-	N	0.04	2.66	INTRA	BL
218	TRP (218-) A	-	CE3	<–> 279	ARG (279-) A	-	NH1	0.03	3.07	INTRA	BF
100	LYS (100-) A	-	NZ	<–> 155	ASP (155-) A	-	OD2	0.02	2.68	INTRA	BF
138	GLY (138-) A	-	N	<–> 172	HIS (172-) A	-	ND1	0.01	2.99	INTRA	BL
251	GLY (251-) A	-	N	<–> 252	PRO (252-) A	-	CD	0.01	2.99	INTRA	BL

	Atom 1		Atom 2		Bump	Dist	Status	
4	ARG (4-) A	- N <–> 299	GLN (299-) A	- OE1	0.01	2.69	INTRA	BL

2.8.2 Note: Some notes regarding these bumps

The bumps have been binned in 5 categories ranging from 'please look at' till 'must fix'. Additionally, the integrated sum of all bumps, the squared sum of all bumps, and these latter two values normalized by the number of contacts are listed too for comparison purposes between, for example, small and large proteins.

Total bump value: 0.658
Total bump value per residue: 0.036
Total number of bumps: 11
Total squared bump value: 0.089
Total number of bumps in the mildest bin: 11
Total number of bumps in the second bin: 0
Total number of bumps in the middle bin: 0
Total number of bumps in the fourth bin: 0
Total number of bumps in the worst bin: 0

2.9 Packing, accessibility and threading

2.9.1 Note: Inside/outside distribution check

The following list contains per-residue Z-scores describing how well the residue's observed accessibility fits the expected one. A positive Z-score indicates "more exposure than usual", whereas a negative Z-score means "more buried than usual". The absolute value of the Z-score must be used to judge the quality. Today WHAT CHECK saw no reason to complain.

2.9.2 Note: Inside/Outside residue distribution normal

The distribution of residue types over the inside and the outside of the protein is normal.

inside/outside RMS Z-score : 0.978

2.9.3 Note: Inside/Outside RMS Z-score plot

The Inside/Outside distribution normality RMS Z-score over a 15 residue window is plotted as function of the residue number. High areas in the plot (above 1.5) indicate unusual inside/outside patterns.

Chain identifier: A

2.9.4 Warning: Abnormal packing environment for some residues

The residues listed in the table below have an unusual packing environment.

The packing environment of the residues is compared with the average packing environment for all residues of the same type in good PDB files. A low packing score can indicate one of several things: Poor packing, misthreading of the sequence through the density, crystal contacts, contacts with a co-factor, or the residue is part of the active site. It is not uncommon to see a few of these, but in any case this requires further inspection of the residue.

	Residue		Quality value
222	ARG (222-) A	-	-7.24
97	LYS (97-) A	-	-5.70
256	GLN (256-) A	-	-5.54
60	ARG (60-) A	-	-5.53
273	GLN (273-) A	-	-5.51
274	ASN (274-) A	-	-5.41
189	GLN (189-) A	-	-5.33
107	GLN (107-) A	-	-5.26
137	LYS (137-) A	-	-5.26
214	ASN (214-) A	-	-5.13
141	LEU (141-) A	-	-5.09
105	ARG (105-) A	-	-5.03

2.9.5 Note: No series of residues with bad packing environment

There are no stretches of three or more residues each having a packing score worse than -4.0.

2.9.6 Note: Structural average packing environment OK

The structural average packing score is within normal ranges.

Average for range 1 - 306 : -0.817

2.9.7 Note: Quality value plot

The quality value smoothed over a 10 residue window is plotted as function of the residue number. Low areas in the plot (below -2.0) indicate unusual packing.

Chain identifier: A

2.9.8 Warning: Low packing Z-score for some residues

The residues listed in the table below have an unusual packing environment according to the 2nd generation packing check. The score listed in the table is a packing normality Z-score: positive means better than average, negative means worse than average. Only residues scoring less than -2.50 are listed here. These are the unusual residues in the structure, so it will be interesting to take a special look at them.

	Residue		Z-score
276	MET (276-) A	-	-2.74

2.9.9 Note: No series of residues with abnormal new packing environment

There are no stretches of four or more residues each having a packing Z-score worse than -1.75.

2.9.10 Note: Second generation quality Z-score plot

The second generation quality Z-score smoothed over a 10 residue window is plotted as function of the residue number. Low areas in the plot (below -1.3) indicate unusual packing.

Chain identifier: A

2.10 Water, ion, and hydrogen bond related checks

2.10.1 Warning: No crystallisation information

No, or very inadequate, crystallisation information was observed upon reading the PDB file header records. This information should be available in the form of a series of REMARK 280 lines. Without this information a few things, such as checking ions in the structure, cannot be performed optimally.

2.10.2 Error: Water clusters without contacts with non-water atoms

The water molecules listed in the table below are part of water molecule clusters that do not make contacts with non-waters. These water molecules are part of clusters that have a distance at least 1 Å larger than the sum of the Van der Waals radii to the nearest non-solvent atom. Because these kinds of water clusters usually are not observed with X-ray diffraction their presence could indicate a refinement artifact. The number in brackets is the identifier of the water molecule in the input file.

	Residue		Atom
307	HOH (667) A	-	O
307	HOH (711) A	-	O

2.10.3 Note: No waters need moving

All water molecules are sufficiently close to the asymmetric unit given in the input file.

2.10.4 Error: Water molecules without hydrogen bonds

The water molecules listed in the table below do not form any hydrogen bonds, neither with the protein or DNA/RNA, nor with other water molecules. This is a strong indication of a refinement problem.

	Residue		Atom
307	HOH (652) A	-	O
307	HOH (667) A	-	O
307	HOH (669) A	-	O
307	HOH (673) A	-	O
307	HOH (678) A	-	O
307	HOH (685) A	-	O
307	HOH (690) A	-	O
307	HOH (698) A	-	O
307	HOH (705) A	-	O
307	HOH (708) A	-	O
307	HOH (711) A	-	O
307	HOH (713) A	-	O

2.10.5 Error: His, Asn, Gln side chain flips

Listed here are Histidine, Asparagine or Glutamine residues for which the orientation determined from hydrogen bonding analysis are different from the assignment given in the input. Either they could form energetically more favourable hydrogen bonds if the terminal group was rotated by 180 degrees, or there is no assignment in the input file (atom type 'A') but an assignment could be made. Be aware, though, that if the topology could not be determined for one or more ligands, then this option will make errors.

	Residue	
19	GLN (19-) A	-
127	GLN (127-) A	-
180	ASN (180-) A	-
203	ASN (203-) A	-
246	HIS (246-) A	-
273	GLN (273-) A	-
274	ASN (274-) A	-

2.10.6 Note: Histidine type assignments

For all complete HIS residues in the structure a tentative assignment to HIS-D (protonated on ND1), HIS-E (protonated on NE2), or HIS-H (protonated on both ND1 and NE2, positively charged) is made based on the hydrogen bond network. A second assignment is made based on which of the Engh and Huber [REF] histidine geometries fits best to the structure.

In the table below all normal histidine residues are listed. The assignment based on the geometry of the residue is listed first, together with the RMS Z-score for the fit to the Engh and Huber parameters. For

all residues where the H-bond assignment is different, the assignment is listed in the last columns, together with its RMS Z-score to the Engh and Huber parameters.

As always, the RMS Z-scores should be close to 1.0 if the residues were restrained to the Engh and Huber parameters during refinement, and if enough (high resolution) data is available.

Please note that because the differences between the geometries of the different types are small it is possible that the geometric assignment given here does not correspond to the type used in refinement. This is especially true if the RMS Z-scores are much higher than 1.0.

If the two assignments differ, or the 'geometry' RMS Z-score is high, it is advisable to verify the hydrogen bond assignment, check the HIS type used during the refinement and possibly adjust it.

	Residue	Geometry	Z-score	H-bond	Z-score
41	HIS (41-) A	- HIS-E	0.78	HIS-D	1.59
64	HIS (64-) A	- HIS-H	0.74		
80	HIS (80-) A	- HIS-E	0.43	HIS-H	0.90
163	HIS (163-) A	- HIS-E	0.63	HIS-D	1.03
164	HIS (164-) A	- HIS-E	0.72	HIS-H	0.73
172	HIS (172-) A	- HIS-E	0.65		
246	HIS (246-) A	- HIS-E	0.55		

2.10.7 Warning: Buried unsatisfied hydrogen bond donors

The buried hydrogen bond donors listed in the table below have a hydrogen atom that is not involved in a hydrogen bond in the optimized hydrogen bond network.

Hydrogen bond donors that are buried inside the protein normally use all of their hydrogens to form hydrogen bonds within the protein. If there are any non hydrogen bonded buried hydrogen bond donors in the structure they will be listed here. In very good structures the number of listed atoms will tend to zero.

Waters are not listed by this option.

	Residue		Atom
26	THR (26-) A	-	OG1
41	HIS (41-) A	-	N
93	THR (93-) A	-	N
94	ALA (94-) A	-	N
126	TYR (126-) A	-	OH
152	ILE (152-) A	-	N
161	TYR (161-) A	-	OH
185	PHE (185-) A	-	N
198	THR (198-) A	-	N
219	PHE (219-) A	-	N
221	ASN (221-) A	-	ND2
288	GLU (288-) A	-	N

2.10.8 Note: Buried hydrogen bond acceptors OK

All buried polar side-chain hydrogen bond acceptors are involved in a hydrogen bond in the optimized hydrogen bond network.

2.10.9 Note: Some notes regarding these donors and acceptors

The donors and acceptors have been counted, also as function of their accessibility. The buried donors and acceptors have been binned in five categories ranging from not forming any hydrogen bond till forming a

poor till perfect hydrogen bond. Obviously, the buried donors and acceptors with no or just a poor hydrogen bond should be a topic of concern. As every protein contains more acceptors than donors, unsatisfied donors are more in need of attention than unsatisfied acceptors.

Total number of donors: 436
- of which buried: 226
Total number of acceptors: 470
- of which buried: 180
Total number of donor+acceptors: 51 (e.g. the Ser Ogamma that can donate and accept)
- of which buried: 14
Buried donors: 226
- without H-bond: 12
- essentially without H-bond: 0
- with only a very poor H-bond: 0
- with a poor H-bond: 1
- with a H-bond: 213
Buried acceptors: 180
- without H-bond: 20
- essentially without H-bond: 0
- with only a very poor H-bond: 3
- with a poor H-bond: 2
- with a H-bond: 155

2.10.10 Note: Content of the PDB file as interpreted by WHAT CHECK

Content of the PDB file as interpreted by WHAT CHECK. WHAT CHECK has read your PDB file, and stored it internally in what is called 'the soup'. The content of this soup is listed here. An extensive explanation of all frequently used WHAT CHECK output formats can be found at swift.cmbi.ru.nl. Look under output formats. A course on reading this 'Molecules' table is part of the WHAT CHECK website.

'Molecules'

1	1 (1)	306 (306) A	Protein		To check
2	307 (HOH)	307 (HOH) A	water	(310)	To check

2.11 Final summary

2.11.1 Note: Summary report

This is an overall summary of the quality of the structure as compared with current reliable structures. Numbers in brackets are the average and standard deviation observed for a large number of files determined with a similar resolution.

The second table mostly gives an impression of how well the model conforms to common refinement restraint values. These numbers are less than 1.0 if the spread in data is too little, and larger than 1.0 when the spread is too large. The former does not need to be a problem, the latter always is bad.

Structure Z-scores, positive is better than average:

Resolution read from PDB file :	-1.000
1st generation packing quality :	-0.792
2nd generation packing quality :	-1.632
Ramachandran plot appearance :	-1.706
χ-1/χ-2 rotamer normality :	-2.947
Backbone conformation :	-1.280
Inside/Outside distribution :	0.978

RMS Z-scores, should be close to 1.0:

Bond lengths :	0.719	
Bond angles :	1.029	
Omega angle restraints :	1.378	(loose)
Side chain planarity :	1.168	
Improper dihedral distribution :	1.067	
B-factor distribution :	0.677	

2.12 Suggestions for the refinement process

2.12.1 Note: Introduction to refinement recommendations

First, be aware that the recommendations for crystallographers listed below are produced by a computer program that was written by a guy who got his PhD in NMR...

We have tried to convert the messages written in this report into a small set of things you can do with your refinement software to get a better structure. The things you should do first are listed first. And in some cases you should first fix that problem, then refine a bit further, and then run WHAT CHECK again before looking at other problems. If, for example, WHAT CHECK has found a problem with the SCALE and CRYST cards, then you must first fix that problem, refine the structure a bit further, and run WHAT CHECK again because errors in the SCALE and or CRYST card can lead to many problems elsewhere in the validation process.

It is also important to keep in mind that WHAT CHECK is software and that it occasionally totally misunderstands what is the cause of a problem. But, if WHAT CHECK lists a problem there normally is a problem albeit that it not always is the actual problem that gets listed.

2.12.2 Note: No crippling problems detected

Some problems can be so crippling that they negatively influence the validity of other validation steps. If such a problem is detected, it must be solved and some further refinemnet must be done before you can continue working with a new WHAT CHECK report. In this file such problems were not detected. You can therefore try to fix as many problems in one go as you want.

2.12.3 Note: No resolution information detected

WHAT CHECK needs to know the resolution of your data to provide advice for the refinement process. This resolution information is needed because a Z-score that is very good at 1.0 Å resolution might actually be a sign of over-refinement at 3.5 Å, etcetera. So, take a look at the formats of REMARK 2 and/or REMARK 3 and put the resolution in ether of those 2. An example of a REMARK 2 card is:

REMARK 2 RESOLUTION. 4.50 ANGSTROMS.

A References

WHAT IF G.Vriend, *WHAT IF: a molecular modelling and drug design program* J. Mol. Graph. **8** 52–56 (1990).

WHAT_CHECK (verification routines from WHAT IF) R.W.W.Hooft, G.Vriend, C.Sander and E.E.Abola, *Errors in protein structures* Nature **381**, 272 (1996). (see http://swift.cmbi.ru.nl/gv/whatcheck for a course and extra information)

PDB facilities Touw WG, Baakman C, Black J, te Beek TA, Krieger E, Joosten RP, Vriend G. *A series of PDB-related databanks for everyday needs* Nucleic Acids Research **Database issue**, D364–368 (2015).

Bond lengths and angles, protein residues R.Engh and R.Huber, *Accurate bond and angle parameters for X-ray protein structure refinement* Acta Crystallogr. **A47**, 392–400 (1991) and International Tables for Crystallography (2001).

Bond lengths and angles, DNA/RNA G.Parkinson, J.Voitechovsky, L.Clowney, A.T.Brünger and H.Berman, *New parameters for the refinement of nucleic acid-containing structures* Acta Crystallogr. **D52**, 57–64 (1996).

DSSP W.Kabsch and C.Sander, *Dictionary of protein secondary structure: pattern recognition of hydrogen bond and geometrical features* Biopolymers **22**, 2577–2637 (1983).

Hydrogen bond networks R.W.W.Hooft, C.Sander and G.Vriend, *Positioning hydrogen atoms by optimizing hydrogen bond networks in protein structures* PROTEINS, **26**, 363–376 (1996).

Matthews' Coefficient B.W.Matthews, *Solvent Content of Protein Crystals* J. Mol. Biol, **33**, 491–497 (1968).

Peptide flips Touw WG, Joosten RP, Vriend G. *Detection of trans-cis flips and peptide-plane flips in protein structures* Acta Crystallogr D Biological Crystallograhy **71**, 1604–1614 (2015).

Protein side chain planarity R.W.W. Hooft, C. Sander and G. Vriend, *Verification of protein structures: side-chain planarity* J. Appl. Cryst, **29**, 714–716 (1996).

Puckering parameters D.Cremer and J.A.Pople, *A general definition of ring puckering coordinates* J. Am. Chem. Soc. **97**, 1354–1358 (1975).

Packing G.Vriend and C.Sander, *Quality control of protein models: directional atomic contact analysis* J. Appl. Cryst. **26**, 47–60 (1993).

Ramachandran plot G.N.Ramachandran, C.Ramakrishnan and V.Sasisekharan. *Stereochemistry of Polypeptide Chain Conformations* J. Mol. Biol. **7** 95–99 (1963).

R.W.W. Hooft, C.Sander and G.Vriend. *Objectively judging the quality of a protein structure from a Ramachandran plot* CABIOS **13** 425–430 (1997).

Symmetry R.W.W.Hooft, C.Sander and G.Vriend, *Reconstruction of symmetry related molecules from protein data bank (PDB) files* J. Appl. Cryst. **27**, 1006–1009 (1994).

Tau angle W.G.Touw and G.Vriend, *On the complexity of Engh and Huber refinement restraints: the angle tau as example* Acta Crystallogr D **66**, 1341–1350 (2010).

Ions I.D.Brown and K.K.Wu, *Empirical parameters for calculating cation-oxygen bond valences* Acta Cryst. B **32**, 1957–1959 (1975).

M.Nayal and E.Di Cera, *Valence Screening of Water in Protein Crystals Reveals Potential Na+ Binding Sites* J. Mol. Biol. **256** 228–234 (1996).

P.Mueller, S.Koepke and G.M.Sheldrick, *Is the bond-valence method able to identify metal atoms in protein structures?* Acta Cryst. D **59** 32–37 (2003).

Checking checks K.Wilson, C.Sander, R.W.W.Hooft, G.Vriend, et al. *Who checks the checkers* J. Mol. Biol. **276** 417–436 (1998).