Supplementary information: WHAT_CHECK analysis of AeJHAMT and 1M6E structures.

To check the quality of a model we used WHAT_CHECK, a program that evaluates the model structure in comparison to the Protein Data Bank database of crystal structures (Hooft *et al*., 1996). It calculates the deviation of parameters such as the amino acid dihedral angles or the amino acid environment with respect to what have been determined for all other crystal structures. The WHAT_CHECK analysis showed that the AeJHAMT model has a reasonable good conformation. In order to evaluate if in the instances where a deviation was observed for AeJHAMT was due to the modeling or was something intrinsic to the protein FOLD, we also evaluated the 1M6E experimental structure using WHAT_CHECK. Interestingly most of the deviations detected for AeJHAMT were also observed for the 1M6E experimental structure. The WHAT_CHECK information for both proteins is now provided as supplementary information.

References: Hooft, R.W., Vriend, G., Sander, C., Abola, E.E., 1996. Errors in protein structures, Nature. 381, 272.

Report of protein analysis

By the WHAT IF program

2010-10-07*

1 Introduction

WHAT_CHECK is the name of the validation option in WHAT IF. It doesn't matter whether you use the WHAT_CHECK program or the WHAT IF program for validation. Both produce exactly the same WHAT_CHECK-report.

This document is a WHAT_CHECK-report that holds the findings of the WHAT IF program during the analysis of a PDB-file. Each reported fact has an assigned severity, one of:

error : severe errors encountered during the analyses. 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 resid 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.

*This report was created by WHAT IF version WHATCHECK 8.0

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 JHAMT.pdb

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

2.1.1 Error: Missing unit cell information

No SCALE matrix is given in the PDB file.

2.1.2 Error: Missing symmetry information

Problem: No CRYST1 card is given in the PDB file.

2.1.3 Note: Ligand topologies OK

The topology could be determined for all ligands (or there are no ligands for which a topology is needed, in which case there is absolutely no problem, of course). That is good because it means that all ligands can be included in the hydrogen bond optimization and related options.

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 seem adequately unique.

2.2.3 Note: No mixed usage of alternate atom problems detected

Either this structure does not contain alternate atoms, or they have not been mixed up, or the errors have remained unnoticed.

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

WHAT IF saw no need to make any alternate atom corrections (which means they are all correct, or there aren't any).

2.2.5 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).

2.2.6 Note: No attached groups interfere with hydrogen bond calculations

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

2.2.7 Warning: Plausible side chain atoms detected with zero occupancy

Plausible side chain atoms were detected with (near) zero occupancy

When crystallographers don't see an atom they either leave it out completely, or give it an occupancy of zero or a very high B-factor. WHAT IF neglects these atoms. In this case some atoms were found with zero occupancy, but with coordinates that place them at a plausible position. Although WHAT IF knows how to deal with missing side chain atoms, validation will go more reliable if all atoms are presnt. So, please consider manually setting the occupancy of the listed atoms at 1.0.

	Residue		Atom
1	MET (1-)	-	CB
1	MET (1-)	-	CG
1	MET (1-)	-	SD
1	MET (1-)	-	CE
2	ASN (2-)	-	CB
2	ASN (2-)	-	CG
2	ASN (2-)	-	OD1
2	ASN (2-)	-	ND2
3	LYS (3-)	-	CB
3	LYS (3-)	-	CG
3	LYS (3-)	-	CD
3	LYS (3-)	-	CE
3	LYS (3-)	-	NZ
4	PRO (4-)	-	CB
4	PRO (4-)	-	CG
4	PRO (4-)	-	CD
5	ASN (5-)	-	CB
5	ASN (5-)	-	CG
5	ASN (5-)	-	OD1
5	ASN (5-)	-	ND2
6	LEU (6-)	-	CB
6	LEU (6-)	-	CG
6	LEU (6-)	-	CD1
6	LEU (6-)	-	CD2
7	TYR (7-)	-	CB

And so on for a total of 1147 lines.

2.2.8 Warning: Plausible backbone atoms detected with zero occupancy

Plausible backbone atoms were detected with (near) zero occupancy

When crystallographers don't see an atom they either leave it out completely, or give it an occupancy of zero or a very high B-factor. WHAT IF neglects these atoms. However, if a backbone atom is present in

the PDB file, and its position seems 'logical' (i.e. normal bond lengths with all atoms it should be bound to, and those atoms exist normally) WHAT IF will set the occupancy to 1.0 if it believes that the full presence of this atom will be beneficial to the rest of the validation process. If you get weird errors at, or near, these atoms, please check by hand what is going on, and repair things intelligently before running this validation again.

	Residue		Atom
1	MET (1-)	-	N
1	MET (1-)	-	CA
1	MET (1-)	-	C
1	MET (1-)	-	O
2	ASN (2-)	-	N
2	ASN (2-)	-	CA
2	ASN (2-)	-	C
2	ASN (2-)	-	O
3	LYS (3-)	-	N
3	LYS (3-)	-	CA
3	LYS (3-)	-	C
3	LYS (3-)	-	O
4	PRO (4-)	-	N
4	PRO (4-)	-	CA
4	PRO (4-)	-	C
4	PRO (4-)	-	O
5	ASN (5-)	-	N
5	ASN (5-)	-	CA
5	ASN (5-)	-	C
5	ASN (5-)	-	O
6	LEU (6-)	-	N
6	LEU (6-)	-	CA
6	LEU (6-)	-	C
6	LEU (6-)	-	O
7	TYR (7-)	-	N

And so on for a total of 1112 lines.

2.2.9 Note: All residues have a complete backbone.

No residues have missing backbone atoms.

2.2.10 Note: No C-alpha only residues

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

2.2.11 Note: Non-canonicals

WHAT IF has not detected any non-canonical residue that it doesn't understand, or there are no non-canonical residues in the PDB file.

2.3 Non-validating, descriptive output paragraph

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

Content of the PDB file as interpreted by WHAT IF. WHAT IF 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 IF output formats can be found at http://swift.cmbi.ru.nl/. Look under output formats. A course on reading this 'Molecules' table is part of the WHAT_CHECK web pages [REF].

'Molecules'

1	1 (1) 278 (278)	Protein		JHAMT.pdb
2	279 (278) 279 (278)	K O2 <-	278	JHAMT.pdb

2.3.2 Note: Some notes regarding the PDB file contents

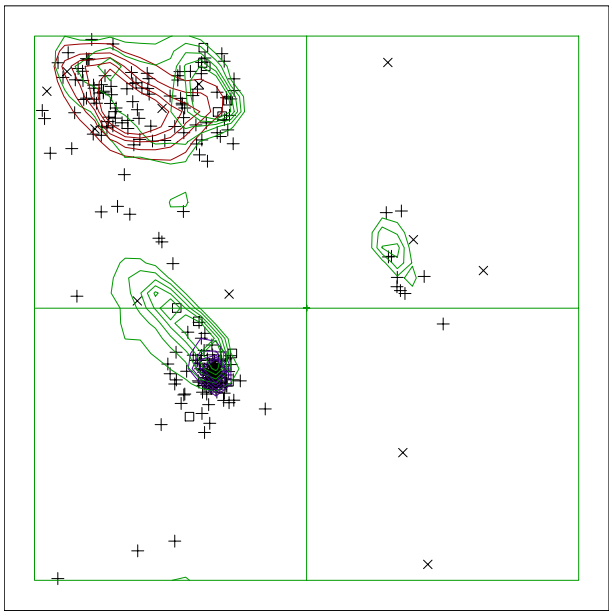
The numbers and remarks listed below have no explicit validation purpose, they are merely meant for the crystallographer or NMR spectroscopists to perhaps pinpoint something unexpected. See the WHAT_CHECK course [REF] for an explanation of terms like 'poor', 'missing', etcetera (in case those words pop up in the lines underneath this message).

Number of amino acids 278
of which 1 have poor or missing atoms.

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. "Allowed" 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 without chain identifier

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 http://swift.cmbi.ru.nl/gv 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
MNKPNLYHRANGVQRRDAKEILDEHGHLLRWKEENEDSLLDIGCGSGDVLIDFVIPMVPP
   HHHH           TTHHHHHHHHHHHHHTT      SST  T HHHHHHH     T
        70         80         90        100        110       120
KRARVLGTDVSEQMVRFARKVHSDVENLFFETLDIEGDISSFLNKWGCFDHITSFYCLHW
TT    SSST HHHHHHHHH TTT TTT SSS  TT TT     TT  SSSSSSST TTT
       130        140        150        160        170       180
VRSQRSAFSNIYNLMAPNGDCLLGFLARNPIFDIYDQLSNSAKWSMYMTDVDKYISPYQY
 T HHHHHHHHHHHSSSSSSSSSSSS TTHHHHHHHHHHHHHHHTT  TT 333TT   T
       190        200        210        220        230       240
CENPVGEIEEILSSVGFTKYKIHIADKIYVYEGIDSLKKAVQAVNPFSERMPLDLQEDFL
TT  HHHHHHHHHHHT    SSS    SS TTTHHHHHHHHHHTT  TTTTT HHHHHHH
       250        260        270
NDYIAVVRRMSLSENCCGNENDYKFITPYKLVVVYAVK
HHHHHHHHHHTTTTTTSSTTSSSS   SSSSSSS
```

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=0.00 or χ-1=180.00 have been detected.

2.4.3 Warning: Missing atoms

The atoms listed in the table below are missing from the entry. If many atoms are missing, the other checks can become less sensitive. Be aware that it often happens that groups at the termini of DNA or RNA are really missing, so that the absence of these atoms normally is neither an error nor the result of poor electron density. Some of the atoms listed here might also be listed by other checks, most noticeably by the options in the previous section that list missing atoms in several categories. The plausible atoms with zero occupancy are not listed here, as they already got assigned a non-zero occupancy, and thus are no longer 'missing'.

	Residue		Atom
56	PRO (56-)	-	N

2.4.4 Note: No C-terminal nitrogen detected

The PDB indicates that a residue is not the true C-terminus by including only the backbone N of the next residue. This has not been observed in this PDB file.

2.4.5 Note: Test capping of (pseudo) C-termini

No extra capping groups were found on pseudo C-termini. This can imply that no pseudo C-termini are present.

2.4.6 Note: Proper C-terminal capping groups found

All (presumably) real C-termini either contain a proper capping group (OXT, or something else), or they are followed by a single Nitrogen, indicating that the rest of the chain is invisible.

2.4.7 Note: No OXT found in the middle of chains

No OXT groups were found in the middle of protein chains.

2.4.8 Note: Weights checked OK

All atomic occupancy factors ('weights') fall in the 0.0–1.0 range.

2.4.9 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 IF'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.

2.4.10 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.

2.4.11 Warning: What type of B-factor?

WHAT IF 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 real, complete B-factors. When TLS refinement is used that assumption sometimes isn't correct. TLS seems not mentioned in the

header of the PDB file. But anyway, if WHAT IF complains about your B-factors, and you think that they are OK, then check for TLS related B-factor problems first.

Temperature not mentioned in PDB file
Room temperature assumed

2.4.12 Warning: Low M-factor

The B-factor flatness, the M-factor, is very low. This is very worrisome. I suggest you consult the WHAT_CHECK website and/or a seasoned crystallographer.

The M-factor = 0.003

2.4.13 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.

Percentage of buried atoms with B less than 5 : 0.00

2.4.14 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 : 0.361 over 2034 bonds
Average difference in B over a bond : 0.00
RMS difference in B over a bond : 0.00

2.4.15 Warning: B-factor plot useless

All average B-factors are equal. Plot suppressed.

Chain without chain identifier

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 δ 2 and δ 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 δ 1 and 2 and the other uses C δ 2 and 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 N-H-1 and N-H-2 swapped.

	Residue		
9	ARG (9-)	-	-
15	ARG (15-)	-	-
16	ARG (16-)	-	-
30	ARG (30-)	-	-
62	ARG (62-)	-	-
64	ARG (64-)	-	-
76	ARG (76-)	-	-
79	ARG (79-)	-	-
122	ARG (122-)	-	-
125	ARG (125-)	-	-
148	ARG (148-)	-	-
230	ARG (230-)	-	-
248	ARG (248-)	-	-
249	ARG (249-)	-	-

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		
116	TYR (116-)	-	-
167	TYR (167-)	-	-
174	TYR (174-)	-	-
178	TYR (178-)	-	-
180	TYR (180-)	-	-
209	TYR (209-)	-	-
211	TYR (211-)	-	-

2.5.8 Warning: Phenylalanine convention problem

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

	Residue	
89	PHE (89-) -	-
102	PHE (102-) -	-
115	PHE (115-) -	-
145	PHE (145-) -	-
197	PHE (197-) -	-

2.5.9 Note: Aspartic acid torsion conventions OK

No errors were detected in aspartic acid torsion angle conventions.

2.5.10 Note: Glutamic acid torsion conventions OK

No errors were detected in glutamic acid torsion angle conventions.

2.5.11 Note: Phosphate group names OK

No errors were detected in phosphate group naming conventions.

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: Chain names are OK

All chain names assigned to polymer molecules are unique, and 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 acid residues have been taken from Engh and Huber [REF], for DNA/RNA from Parkinson et al [REF])

2.6.2 Warning: Low bond length variability

Bond lengths were found to deviate less than normal from the mean Engh and Huber [REF] and/or Parkinson et al [REF] standard bond lengths. The RMS Z-score given below is expected to be around 1.0 for a normally restrained data set. The fact that it is lower than 0.667 in this structure might indicate that too-strong restraints have been used in the refinement. This can only be a problem for high resolution X-ray structures.

RMS Z-score for bond lengths: 0.599
RMS-deviation in bond distances: 0.013

2.6.3 Note: No bond length directionality

Comparison of bond distances with Engh and Huber [REF] standard values for protein residues and Parkinson et al [REF] values for DNA/RNA does not show significant systematic deviations.

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
5	ASN (5-)	-	CA	CB	CG	117.76	5.2
46	SER (46-)	-	N	CA	CB	118.88	4.9
46	SER (46-)	-	C	CA	CB	99.03	-5.8
47	GLY (47-)	-	-C	N	CA	128.26	4.5
48	ASP (48-)	-	-C	N	CA	133.92	6.8
55	ILE (55-)	-	CA	CB	CG2	103.60	-4.1
66	LEU (66-)	-	N	CA	CB	101.05	-5.6
69	ASP (69-)	-	CA	CB	CG	107.87	-4.7
73	GLN (73-)	-	NE2	CD	OE1	118.27	-4.3
88	LEU (88-)	-	CD1	CG	CD2	124.03	6.0
94	ASP (94-)	-	CA	CB	CG	117.91	5.3
103	LEU (103-)	-	CD1	CG	CD2	119.68	4.0
150	PRO (150-)	-	-CA	-C	N	124.57	5.1
151	ILE (151-)	-	CA	CB	CG1	118.07	4.5
172	ASP (172-)	-	CA	CB	CG	106.22	-6.4
179	GLN (179-)	-	NE2	CD	OE1	116.95	-5.7
183	ASN (183-)	-	ND2	CG	OD1	117.39	-5.2
197	PHE (197-)	-	CA	CB	CG	120.12	6.3
200	TYR (200-)	-	-C	N	CA	130.17	4.7
216	SER (216-)	-	CA	CB	OG	121.06	5.0
222	GLN (222-)	-	NE2	CD	OE1	118.35	-4.3
228	SER (228-)	-	-C	N	CA	130.11	4.7
239	PHE (239-)	-	C	CA	CB	119.28	4.8
239	PHE (239-)	-	CA	CB	CG	120.80	7.0
273	VAL (273-)	-	CA	CB	CG1	117.71	4.2

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 around 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: 1.241
RMS-deviation in bond angles: 2.277

2.6.6 Error: Nomenclature error(s)

You are asking for a hand-check. WHAT IF has over the course of this session already corrected the handedness of atoms in several residues. These residues are listed here.

Residue		
9	ARG (9-)	- -
15	ARG (15-)	- -
16	ARG (16-)	- -
30	ARG (30-)	- -
62	ARG (62-)	- -
64	ARG (64-)	- -
76	ARG (76-)	- -
79	ARG (79-)	- -
122	ARG (122-)	- -
125	ARG (125-)	- -
148	ARG (148-)	- -
230	ARG (230-)	- -
248	ARG (248-)	- -
249	ARG (249-)	- -

2.6.7 Note: Chirality OK

All protein atoms have proper chirality.

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 : 1.364

2.6.9 Note: Tau angles OK

All of the tau angles of amino acids that actually have a tau angle fall within expected RMS deviations.

2.6.10 Note: Normal tau angle deviations

The RMS Z-score for the tau angles in the structure falls within the normal rannge 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 : 1.023

2.6.11 Error: Side chain planarity problems

The side chains of the residues listed in the table below contain a planar group that was found to deviate from planarity by more than 4.0 times the expected value. For an amino acid residue that has a side chain with a planar group, the RMS deviation of the atoms to a least squares plane was determined. The number in the table is the number of standard deviations this RMS value deviates from the expected value. Not knowing better yet, we assume that planarity of the groups analyzed should be perfect.

	Residue		Z-score
269	TYR (269-)	-	8.87
178	TYR (178-)	-	8.67
120	TRP (120-)	-	6.28
243	TYR (243-)	-	5.92
197	PHE (197-)	-	5.55
180	TYR (180-)	-	5.53
2	ASN (2-)	-	5.45
116	TYR (116-)	-	5.04
31	TRP (31-)	-	4.64

2.6.12 Error: Connections to aromatic rings out of plane

The atoms listed in the table below are connected to a planar aromatic group in the sidechain of a protein residue but were found to deviate from the least squares plane.

For all atoms that are connected to an aromatic side chain in a protein residue the distance of the atom to the least squares plane through the aromatic system was determined. This value was divided by the standard deviation from a distribution of similar values from a database of small molecule structures.

	Residue		Atom	Z-score
90	PHE (90-)	-	CB	4.38
132	TYR (132-)	-	OH	4.34
132	TYR (132-)	-	CB	4.27

2.7 Torsion-related checks

2.7.1 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
56	PRO (56-)	-	0.00	LOW

2.7.2 Note: PRO puckering phases OK

Puckering phases for all PRO residues are normal

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 IF 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
176	SER (176-)	-	-2.5
271	LEU (271-)	-	-2.4
151	ILE (151-)	-	-2.2
226	PRO (226-)	-	-2.2
144	GLY (144-)	-	-2.1

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
8	HIS (8-)	- ω poor
37	ASP (37-)	- Poor ϕ/ψ
45	GLY (45-)	- ω poor
55	ILE (55-)	- Impossible ψ
56	PRO (56-)	- Impossible ϕ, ω poor
63	ALA (63-)	- ω poor
64	ARG (64-)	- ω poor
66	LEU (66-)	- ω poor
84	ASP (84-)	- Poor ϕ/ψ
85	VAL (85-)	- ω poor
96	GLU (96-)	- ω poor
97	GLY (97-)	- Poor ϕ/ψ, ω poor
98	ASP (98-)	- ω poor
102	PHE (102-)	- ω poor
106	TRP (106-)	- ω poor
116	TYR (116-)	- Poor ϕ/ψ
117	CYS (117-)	- ω poor
123	SER (123-)	- Poor ϕ/ψ
135	MET (135-)	- ω poor
138	ASN (138-)	- Poor ϕ/ψ
146	LEU (146-)	- ω poor
147	ALA (147-)	- Poor ϕ/ψ
148	ARG (148-)	- ω poor
149	ASN (149-)	- Poor ϕ/ψ
167	TYR (167-)	- Poor ϕ/ψ
168	MET (168-)	- ω poor
179	GLN (179-)	- ω poor
183	ASN (183-)	- PRO ω poor
196	GLY (196-)	- Poor ϕ/ψ
198	THR (198-)	- Poor ϕ/ψ
199	LYS (199-)	- ω poor
200	TYR (200-)	- ω poor
212	GLU (212-)	- Poor ϕ/ψ, ω poor
222	GLN (222-)	- ω poor
224	VAL (224-)	- ω poor
225	ASN (225-)	- Poor ϕ/ψ

226	PRO (226-) -	- Poor PRO-ϕ
230	ARG (230-) -	- Poor ϕ/ψ
233	LEU (233-) -	- ω poor
251	SER (251-) -	- ω poor
255	ASN (255-) -	- Poor ϕ/ψ
260	GLU (260-) -	- Poor ϕ/ψ
268	PRO (268-) -	- ω poor

2.7.5 Note: Ramachandran Z-score OK

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

Ramachandran Z-score : -1.849

2.7.6 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.

Standard deviation of ω values : 9.080

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

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

χ-1/χ-2 correlation Z-score : 3.534

2.7.8 Warning: Backbone oxygen evaluation

The residues listed in the table below have an unusual backbone oxygen position.

For each of the residues in the structure, a search was performed to find 5-residue stretches in the WHAT IF database with superposable C-α coordinates, and some restraining on the neighbouring backbone oxygens.

In the following table the RMS distance between the backbone oxygen positions of these matching structures in the database and the position of the backbone oxygen atom in the current residue is given. If this number is larger than 1.5 a significant number of structures in the database show an alternative position for the backbone oxygen. If the number is larger than 2.0 most matching backbone fragments in the database have the peptide plane flipped. A manual check needs to be performed to assess whether the experimental data can support that alternative as well. The number in the last column is the number of database hits (maximum 80) used in the calculation. It is "normal" that some glycine residues show up in this list, but they are still worth checking!

Residue			Distance (Å)	# hits
56	PRO (56-) -	-	3.96	14

2.7.9 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
83	SER (83-) - -	0.35

2.7.10 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!

	Residue	# hits
8	HIS (8-) - -	0
9	ARG (9-) - -	0
11	ASN (11-) - -	0
13	VAL (13-) - -	0
14	GLN (14-) - -	0
15	ARG (15-) - -	0
16	ARG (16-) - -	0
35	ASN (35-) - -	0
36	GLU (36-) - -	0
37	ASP (37-) - -	0
38	SER (38-) - -	0
42	ILE (42-) - -	0
44	CYS (44-) - -	0
45	GLY (45-) - -	0
46	SER (46-) - -	0
47	GLY (47-) - -	0
48	ASP (48-) - -	0
56	PRO (56-) - -	0
58	VAL (58-) - -	0
59	PRO (59-) - -	0
61	LYS (61-) - -	0
62	ARG (62-) - -	0
63	ALA (63-) - -	0
70	VAL (70-) - -	0
82	HIS (82-) - -	0

And so on for a total of 130 lines.

2.7.11 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 : -1.660

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 distance.

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.

The last text-item on each line represents the status of the atom pair. The text 'INTRA' means that the bump is between atoms that are explicitly listed in the PDB file. 'INTER' means it is an inter-symmetry bump. 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 aren't 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. In any case, each bump is listed in only one direction. If the MODEL number doesn't exist (like in most X-ray files), a minus sign is printed instead.

	Atom 1				Atom 2				Bump	Dist	Status	
135	MET (135-)	-	-	SD	ASP (140-)	-	-	C	0.31	3.09	INTRA	BL
146	LEU (146-)	-	-	CB	TYR (180-)	-	-	CE1	0.31	2.89	INTRA	BL
46	SER (46-)	-	-	N	MET (74-)	-	-	SD	0.29	3.01	INTRA	BL
151	ILE (151-)	-	-	CG2	TYR (180-)	-	-	CE1	0.26	2.94	INTRA	BL
135	MET (135-)	-	-	CE	PHE (197-)	-	-	CZ	0.26	2.94	INTRA	BL
192	LEU (192-)	-	-	CD1	LYS (201-)	-	-	NZ	0.25	2.85	INTRA	BL
10	ALA (10-)	-	-	N	MET (74-)	-	-	CE	0.25	2.85	INTRA	BL
95	ILE (95-)	-	-	CD1	CYS (117-)	-	-	SG	0.24	3.16	INTRA	BL
120	TRP (120-)	-	-	CH2	ASN (225-)	-	-	ND2	0.24	2.86	INTRA	BL
120	TRP (120-)	-	-	CZ3	TYR (178-)	-	-	CE1	0.23	2.97	INTRA	BL
146	LEU (146-)	-	-	CD2	ILE (266-)	-	-	CG2	0.22	2.98	INTRA	BL
55	ILE (55-)	-	-	CG2	VAL (65-)	-	-	CG2	0.22	2.98	INTRA	BL
120	TRP (120-)	-	-	CZ2	ASN (225-)	-	-	ND2	0.21	2.89	INTRA	BL
166	MET (166-)	-	-	CE	TYR (243-)	-	-	CD2	0.21	2.99	INTRA	BL
14	GLN (14-)	-	-	NE2	TYR (116-)	-	-	CD1	0.21	2.89	INTRA	BL
158	LEU (158-)	-	-	CD1	TYR (243-)	-	-	CZ	0.20	3.00	INTRA	BL
31	TRP (31-)	-	-	CZ2	LYS (61-)	-	-	NZ	0.19	2.91	INTRA	BL
135	MET (135-)	-	-	CE	PHE (197-)	-	-	CE1	0.19	3.01	INTRA	BL
256	CYS (256-)	-	-	SG	TYR (263-)	-	-	N	0.19	3.11	INTRA	BL
132	TYR (132-)	-	-	CE1	TYR (275-)	-	-	CE1	0.18	3.02	INTRA	BL
44	CYS (44-)	-	-	SG	THR (68-)	-	-	C	0.18	3.22	INTRA	BL

145	PHE (145-)	-	-	CE2 <–> 183	ASN (183-)	-	-	ND2	0.18	2.92	INTRA	BL
55	ILE (55-)	-	-	CG2 <–> 88	LEU (88-)	-	-	CD2	0.16	3.04	INTRA	BL
40	LEU (40-)	-	-	CD2 <–> 42	ILE (42-)	-	-	CD1	0.16	3.04	INTRA	BL
42	ILE (42-)	-	-	CG2 <–> 95	ILE (95-)	-	-	CG1	0.16	3.04	INTRA	BL

And so on for a total of 83 lines.

2.9 Packing, accessibility and threading

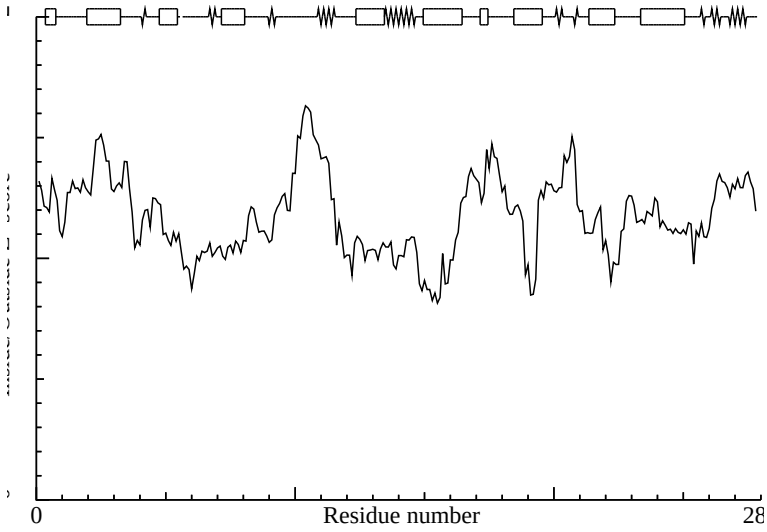
2.9.1 Warning: Inside/Outside residue distribution unusual

The distribution of residue types over the inside and the outside of the protein is unusual. Normal values for the RMS Z-score below are between 0.84 and 1.16. The fact that it is higher in this structure could be caused by transmembrane helices, by the fact that it is part of a multimeric active unit, or by mistraced segments in the density.

inside/outside RMS Z-score : 1.187

2.9.2 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 without chain identifier

2.9.3 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	Qualty value
230	ARG (230-) -	- -7.78
211	TYR (211-) -	- -7.13
148	ARG (148-) -	- -6.49
61	LYS (61-) -	- -6.08
62	ARG (62-) -	- -6.07
105	LYS (105-) -	- -5.95
199	LYS (199-) -	- -5.72
11	ASN (11-) -	- -5.54
227	PHE (227-) -	- -5.49
15	ARG (15-) -	- -5.48
13	VAL (13-) -	- -5.32
2	ASN (2-) -	- -5.28
9	ARG (9-) -	- -5.24
222	GLN (222-) -	- -5.22
209	TYR (209-) -	- -5.15
14	GLN (14-) -	- -5.14
277	VAL (277-) -	- -5.12

2.9.4 Warning: Abnormal packing environment for sequential residues

A stretch of at least three sequential residues with a questionable packing environment was found. This could indicate that these residues are part of a strange loop. It might also be an indication of misthreading in the density. However, it can also indicate that one or more residues in this stretch have other problems such as, for example, missing atoms, very weird angles or bond lengths, etc.

The table below lists the first and last residue in each stretch found, as well as the average residue score of the series.

	Start residue		End residue	Av. Qualty
7	TYR (7-) -	- 9 —	ARG 9- (-) -	-4.86
13	VAL (13-) -	- 15 —	ARG 15- (-) -	-5.31

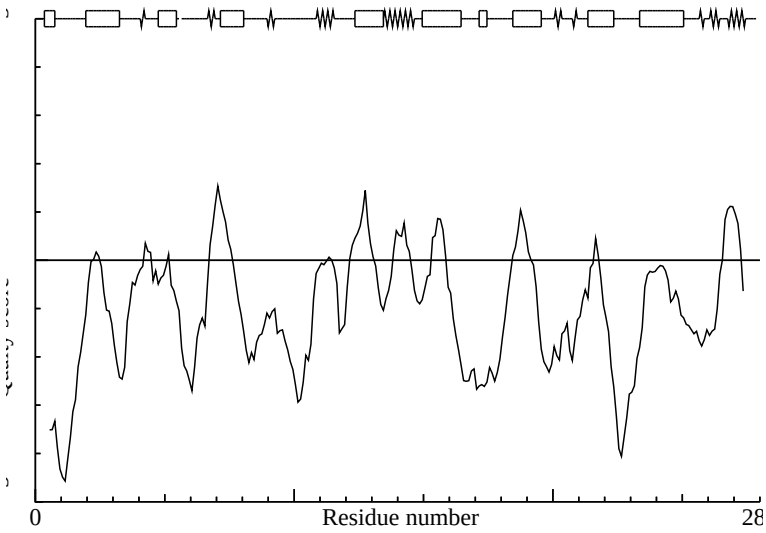
2.9.5 Note: Structural average packing environment OK

The structural average quality control value is within normal ranges.

Average for range 1 - 278 : -1.252

2.9.6 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 without chain identifier

2.9.7 Note: Second generation packing environment OK

None of the individual amino acid residues has a bad packing environment.

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

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

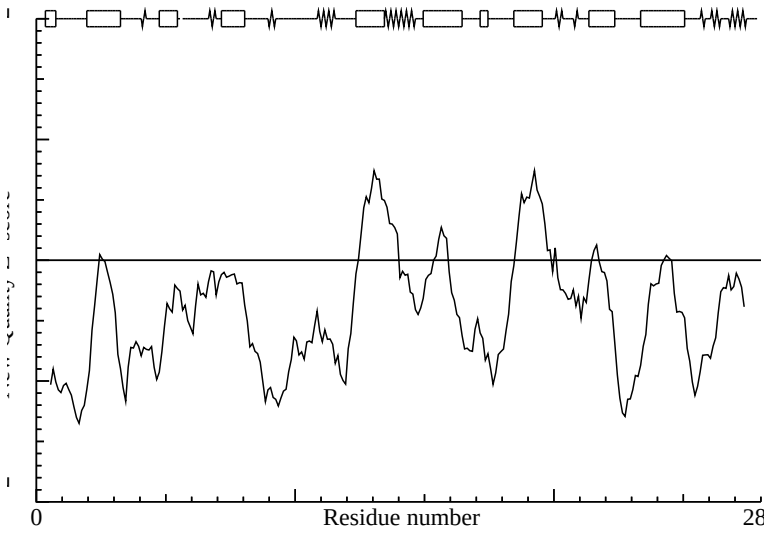
2.9.9 Note: Structural average packing Z-score OK

The structural average for the second generation quality control value is within normal ranges.

All contacts : Average = -0.455 Z-score = -2.57
BB-BB contacts : Average = -0.167 Z-score = -1.07
BB-SC contacts : Average = -0.396 Z-score = -2.80
SC-BB contacts : Average = -0.411 Z-score = -2.39
SC-SC contacts : Average = -0.460 Z-score = -2.88

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 without chain identifier

2.10 Water, ion, and hydrogenbond related checks

2.10.1 Note: HIS, ASN, GLN side chains OK

All of the side chain conformations of Histidine, Asparagine and Glutamine residues were found to be optimal for hydrogen bonding.

2.10.2 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.

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
8	HIS (8-) -	- HIS-E	1.12		
25	HIS (25-) -	- HIS-E	1.19		
27	HIS (27-) -	- HIS-E	1.25		
82	HIS (82-) -	- HIS-E	1.13		
111	HIS (111-) -	- HIS-E	0.67		
119	HIS (119-) -	- HIS-E	0.48		
203	HIS (203-) -	- HIS-D	0.82	HIS-E	0.93

2.10.3 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
5	ASN (5-)	-	N
8	HIS (8-)	-	NE2
9	ARG (9-)	-	N
9	ARG (9-)	-	NE
10	ALA (10-)	-	N
11	ASN (11-)	-	N
30	ARG (30-)	-	NE
31	TRP (31-)	-	NE1
33	GLU (33-)	-	N
37	ASP (37-)	-	N
38	SER (38-)	-	N
46	SER (46-)	-	N
48	ASP (48-)	-	N
66	LEU (66-)	-	N
84	ASP (84-)	-	N
87	ASN (87-)	-	N
88	LEU (88-)	-	N
89	PHE (89-)	-	N
111	HIS (111-)	-	NE2
115	PHE (115-)	-	N
117	CYS (117-)	-	N
118	LEU (118-)	-	N
151	ILE (151-)	-	N
152	PHE (152-)	-	N
173	LYS (173-)	-	N
180	TYR (180-)	-	N
182	GLU (182-)	-	N
183	ASN (183-)	-	ND2
199	LYS (199-)	-	N
201	LYS (201-)	-	NZ
211	TYR (211-)	-	N

213	GLY (213-)	-	-	N
216	SER (216-)	-	-	N
225	ASN (225-)	-	-	N
235	LEU (235-)	-	-	N
243	TYR (243-)	-	-	OH
255	ASN (255-)	-	-	N
262	ASP (262-)	-	-	N
267	THR (267-)	-	-	N
274	VAL (274-)	-	-	N
276	ALA (276-)	-	-	N

2.10.4 Warning: Buried unsatisfied hydrogen bond acceptors

The buried side-chain hydrogen bond acceptors listed in the table below are not involved in a hydrogen bond in the optimized hydrogen bond network.

Side-chain hydrogen bond acceptors that are buried inside the protein normally form hydrogen bonds within the protein. If there are any not hydrogen bonded in the optimized hydrogen bond network they will be listed here.

Waters are not listed by this option.

	Residue			Atom
27	HIS (27-)	-	-	ND1
41	ASP (41-)	-	-	OD2
48	ASP (48-)	-	-	OD1
52	ASP (52-)	-	-	OD1
52	ASP (52-)	-	-	OD2
111	HIS (111-)	-	-	ND1
172	ASP (172-)	-	-	OD1
183	ASN (183-)	-	-	OD1
203	HIS (203-)	-	-	ND1
262	ASP (262-)	-	-	OD2

2.10.5 Note: Content of the PDB file as interpreted by WHAT IF

Content of the PDB file as interpreted by WHAT IF. WHAT IF 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 IF output formats can be found at http://swift.cmbi.ru.nl/. Look under output formats. A course on reading this 'Molecules' table is part of the WHAT_CHECK web pages [REF].

'Molecules'				
1	1 (1) 278 (278)	Protein		JHAMT.pdb
2	279 (278) 279 (278)	K O2 <-	278	JHAMT.pdb

2.10.6 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.

2.10.7 Note: No ions (of a type we can validate) in structure

Since there are no ions in the structure of a type we can validate, this check will not be executed.

2.11 Final summary

2.11.1 Note: Summary report for users of a structure

This is an overall summary of the quality of the structure as compared with current reliable structures. This summary is most useful for biologists seeking a good structure to use for modelling calculations.

The second part of the table mostly gives an impression of how well the model conforms to common refinement restraint values. The first part of the table shows a number of restraint-independent quality indicators.

Structure Z-scores, positive is better than average:

1st generation packing quality :	-1.880	
2nd generation packing quality :	-2.568	
Ramachandran plot appearance :	-1.849	
χ-1/χ-2 rotamer normality :	3.534	
Backbone conformation :	-1.660	

RMS Z-scores, should be close to 1.0:

Bond lengths :	0.599	(tight)
Bond angles :	1.241	
Omega angle restraints :	1.651	(loose)
Side chain planarity :	2.161	(loose)
Improper dihedral distribution :	1.364	
B-factor distribution :	0.361	
Inside/Outside distribution :	1.187	(unusual)

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)

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).

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).

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).

Quality Control 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).

Symmetry Checks 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).

Ion Checks 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).

Report of protein analysis

By the WHAT IF program

2010-10-06*

1 Introduction

WHAT_CHECK is the name of the validation option in WHAT IF. It doesn't matter whether you use the WHAT_CHECK program or the WHAT IF program for validation. Both produce exactly the same WHAT_CHECK-report.

This document is a WHAT_CHECK-report that holds the findings of the WHAT IF program during the analysis of a PDB-file. Each reported fact has an assigned severity, one of:

error : severe errors encountered during the analyses. 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 resid 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.

*This report was created by WHAT IF version WHATCHECK 8.0

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 1M6E.pdb

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

2.1.1 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: 40958.824
Volume of the Unit Cell V= 1285507.9
Cell multiplicity: 8
Matthews coefficient for observed atoms Vm= 3.923

2.1.2 Note: No atoms with high occupancy detected at special positions

Either there were no atoms at special positions, or all atoms at special positions have adequately reduced occupancies. An atom is considered to be located at a special position if it is within 0.3 Å from one of its own symmetry copies. See also the next check...

2.1.3 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.

2.1.4 Warning: Ligands for which topology could not be determined

The ligands in the table below are too complicated for the automatic topology determination. WHAT IF uses a local copy of Daan van Aalten's Dundee PRODRG server to automatically generate topology information for ligands. Some molecules are too complicated for this software. If that happens, WHAT IF / WHAT-CHECK continue with a simplified topology that lacks certain information. Ligands with a simplified topology can, for example, not form hydrogen bonds, and that reduces the accuracy of all hydrogen bond related checking facilities.

	Ligand	
362	SAH (2814-) X	-
363	SAL (2000-) X	-

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 seem adequately unique.

2.2.3 Note: No mixed usage of alternate atom problems detected

Either this structure does not contain alternate atoms, or they have not been mixed up, or the errors have remained unnoticed.

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

WHAT IF saw no need to make any alternate atom corrections (which means they are all correct, or there aren't any).

2.2.5 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).

2.2.6 Note: No attached groups interfere with hydrogen bond calculations

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

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

Either there are no atoms with zero occupancy, or they are not present in the file, or their positions are sufficiently improbable to warrant a zero occupancy.

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

Either there are no backbone atoms with zero occupancy, or they are not present in the file, or their positions are sufficiently improbable to warrant a zero occupancy.

2.2.9 Note: All residues have a complete backbone.

No residues have missing backbone atoms.

2.2.10 Note: No C-alpha only residues

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

2.2.11 Note: Non-canonicals

WHAT IF has not detected any non-canonical residue that it doesn't understand, or there are no non-canonical residues in the PDB file.

2.3 Non-validating, descriptive output paragraph

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

Content of the PDB file as interpreted by WHAT IF. WHAT IF 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 IF output formats can be found at http://swift.cmbi.ru.nl/. Look under output formats. A course on reading this 'Molecules' table is part of the WHAT_CHECK web pages [REF].

'Molecules'

1	1 (1)	359 (359)	X	Protein		1M6E.pdb
2	360 (359)	360 (359)	X	D O2 <-	359	1M6E.pdb
3	361 (999)	361 (999)	X	LU		1M6E.pdb
4	362 (2814)	362 (2814)	X	SAH		1M6E.pdb
5	363 (2000)	363 (2000)	X	SAL		1M6E.pdb
6	364 (HOH)	364 (HOH)	X	water	(14)	1M6E.pdb

2.3.2 Note: Some notes regarding the PDB file contents

The numbers and remarks listed below have no explicit validation purpose, they are merely meant for the crystallographer or NMR spectroscopists to perhaps pinpoint something unexpected. See the WHAT_CHECK course [REF] for an explanation of terms like 'poor', 'missing', etcetera (in case those words pop up in the lines underneath this message).

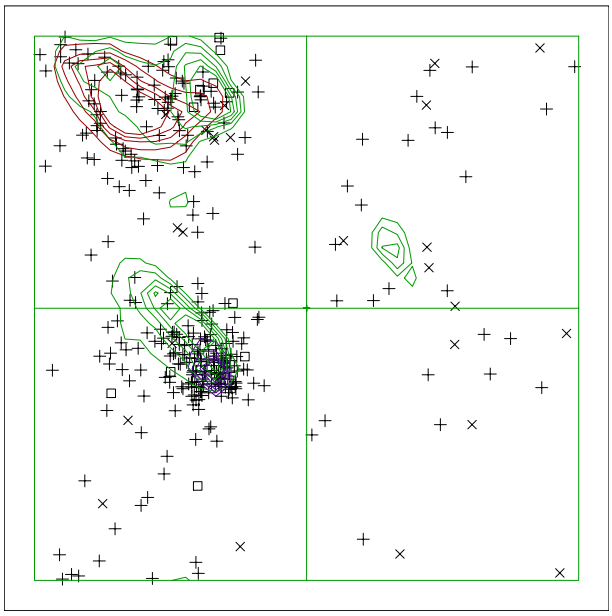
Number of amino acids 359
of which 70 have poor or missing atoms.
Number of water molecules 14

2.3.3 Note: All chain connections seem OK

2.3.4 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. "Allowed" 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: X

2.3.5 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 http://swift.cmbi.ru.nl/gv 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
MDVRQVLHMKGGAGENSYAMNSFIQRQVISITKPITEAAITALYSGDTVTTRLAIADLGC
   HHHH     TTTTTTTTT HHHHHHHHHHTHHHHHHHHHHHHTTTTTTTSS  SSST
        70         80         90        100        110        120
SSGPNALFAVTELIKTVEELRKKMGRENSPEYQIFLNDLPGNDFNAIFRSLPIENDVDGV
 TTTTTTT333TTHHHHHHHHHHTTT TT  SSSSSSSS TTT HHHHHTTTTTT T TT
       130        140        150        160        170        180
CFINGVPGSFYGRLFPRNTLHFIHSSYSLMWLSQVPIGIESNKGNIYMANTCPQSVLNAY
SSSSSSST TTT  T TT  T SSSST TT  TT  T     TTTTTT TTT  TTT  T
       190        200        210        220        230        240
YKQFQEDHALFLRCRAQEVVPGGRMVLTILGRRSEDRASTECCLIWQLLAMALNQMVSEG
HHHHHHHHHHHHHHHHHH  TT SSSSSSSS TTTTTTTTTTTTTTHHHHHHHHHHHHTT
       250        260        270        280        290        300
LIEEEKMDKFNIPQYTPSPTEVEAEILKEGSFLIDHIEASEIYWSSCTKDGDGGGSVEEE
 T  TTT333        THHHHHHHHHTTT   SSSSSSSSSTT  TT TT  TTTTTT
       310        320        330        340        350
GYNVARCMRAVAEPLLLDHFGEAIIEDVFHRYKLLIIERMSKEKTKFINVIVSLIRKSD
TTHHHHHHHHHHHHHHHHHH HHHHHHHHHHHHHHHHHHHTT  SSSSSSSSSS
```

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=0.00 or χ-1=180.00 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 elemantary building blocks (amino acids, nucleotides). Even the C-terminal oxygens are treated separately.

2.4.4 Note: No C-terminal nitrogen detected

The PDB indicates that a residue is not the true C-terminus by including only the backbone N of the next residue. This has not been observed in this PDB file.

2.4.5 Note: Test capping of (pseudo) C-termini

No extra capping groups were found on pseudo C-termini. This can imply that no pseudo C-termini are present.

2.4.6 Note: Proper C-terminal capping groups found

All (presumably) real C-termini either contain a proper capping group (OXT, or something else), or they are followed by a single Nitrogen, indicating that the rest of the chain is invisible.

2.4.7 Note: No OXT found in the middle of chains

No OXT groups were found in the middle of protein chains.

2.4.8 Note: Weights checked OK

All atomic occupancy factors ('weights') fall in the 0.0–1.0 range.

2.4.9 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 IF'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.

2.4.10 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.

2.4.11 Warning: What type of B-factor?

WHAT IF 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 real, complete B-factors. When TLS refinement is used that assumption sometimes isn't correct. TLS seems not mentioned in the header of the PDB file. But anyway, if WHAT IF complains about your B-factors, and you think that they are OK, then check for TLS related B-factor problems first.

Temperature cannot be read from PDB file

2.4.12 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.

Percentage of buried atoms with B less than 5 : 0.00

2.4.13 Error: The B-factors of bonded atoms show signs of over-refinement

For each of the bond types in a protein a distribution was derived for the difference between the square roots of the B-factors of the two atoms. All bonds in the current protein were scored against these distributions. The number given below is the RMS Z-score over the structure. For a structure with completely restrained B-factors within residues, this value will be around 0.35, for extremely high resolution structures refined with free isotropic B-factors this number is expected to be near 1.0. Any value over 1.5 is sign of severe over-refinement of B-factors.

RMS Z-score : 1.720 over 2223 bonds
Average difference in B over a bond : 6.03
RMS difference in B over a bond : 7.91

2.4.14 Note: B-factor plot

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

Chain identifier: X

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 δ 2 and δ 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 δ 1 and 2 and the other uses C δ 2 and 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 Note: Arginine nomenclature OK

No errors were detected in arginine nomenclature.

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	
18	TYR (18-) X	-
131	TYR (131-) X	-
147	TYR (147-) X	-
255	TYR (255-) X	-
283	TYR (283-) X	-

2.5.8 Warning: Phenylalanine convention problem

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

	Residue	
95	PHE (95-) X	-
104	PHE (104-) X	-
108	PHE (108-) X	-
122	PHE (122-) X	-
142	PHE (142-) X	-
250	PHE (250-) X	-
320	PHE (320-) X	-
329	PHE (329-) X	-
347	PHE (347-) X	-

2.5.9 Warning: Aspartic acid convention problem

The aspartic acid residues listed in the table below have their χ-2 not between -90.0 and 90.0, or their proton on OD1 instead of OD2.

	Residue	
318	ASP (318-) X	-

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		
79	GLU (79-) X	-
313	GLU (313-) X	-
338	GLU (338-) X	-

2.5.11 Note: Phosphate group names OK

No errors were detected in phosphate group naming conventions.

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: Chain names are OK

All chain names assigned to polymer molecules are unique, and all residue numbers are strictly increasing within each chain.

2.6 Geometric checks

2.6.1 Warning: Unusual bond lengths

The bond lengths listed in the table below were found to deviate more than 4 sigma from standard bond lengths (both standard values and sigmas for amino acid residues have been taken from Engh and Huber [REF], for DNA they were taken from Parkinson et al [REF]). In the table below for each unusual bond the bond length and the number of standard deviations it differs from the normal value is given.

Atom names starting with "-" belong to the previous residue in the chain. If the second atom name is "-SG*", the disulphide bridge has a deviating length.

Residue			Atom pair		Distance	Z-value
256	THR (256-) X	-	CA	CB	1.65	6.0

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: 0.611
RMS-deviation in bond distances: 0.014

2.6.3 Note: No bond length directionality

Comparison of bond distances with Engh and Huber [REF] standard values for protein residues and Parkinson et al [REF] values for DNA/RNA does not show significant systematic deviations.

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
18	TYR (18-) X -	-C	N	CA	131.60	5.5
19	ALA (19-) X -	N	CA	C	92.98	-6.5
47	ASP (47-) X -	N	CA	C	123.11	4.3
48	THR (48-) X -	-C	N	CA	129.77	4.5
48	THR (48-) X -	N	CA	C	93.23	-6.4
85	GLY (85-) X -	N	CA	C	100.42	-4.2
101	GLY (101-) X -	-C	N	CA	113.44	-4.2
102	ASN (102-) X -	N	CA	C	96.09	-5.4
109	ARG (109-) X -	CG	CD	NE	117.62	4.2
131	TYR (131-) X -	N	CA	C	99.22	-4.3
140	LEU (140-) X -	N	CA	C	123.85	4.5
149	LEU (149-) X -	CA	CB	CG	101.19	-4.3
179	ALA (179-) X -	-C	N	CA	129.23	4.2
179	ALA (179-) X -	N	CA	C	97.23	-5.0
243	GLU (243-) X -	N	CA	C	97.38	-4.9
245	GLU (245-) X -	N	CA	C	93.58	-6.3
269	GLU (269-) X -	N	CA	C	97.66	-4.8
356	ARG (356-) X -	N	CA	C	96.27	-5.3

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 around 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: 0.986
RMS-deviation in bond angles: 2.239

2.6.6 Error: Nomenclature error(s)

You are asking for a hand-check. WHAT IF has over the course of this session already corrected the handedness of atoms in several residues. These residues are listed here.

Residue	
79	GLU (79-) X -
313	GLU (313-) X -
318	ASP (318-) X -
338	GLU (338-) X -

2.6.7 Note: Chirality OK

All protein atoms have proper chirality.

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 : 0.932

2.6.9 Error: Tau angle problems

The side chains of the residues listed in the table below contain a tau angle that was found to deviate from te expected value by more than 4.0 times the expected standard deviation. The number in the table is the number of standard deviations this RMS value deviates from the expected value.

	Residue		Z-score
19	ALA (19-) X	-	7.64
245	GLU (245-) X	-	6.09
48	THR (48-) X	-	6.03
7	LEU (7-) X	-	5.96
179	ALA (179-) X	-	5.88
325	ILE (325-) X	-	5.83
356	ARG (356-) X	-	5.63
322	GLU (322-) X	-	5.16
102	ASN (102-) X	-	5.12
140	LEU (140-) X	-	5.10
243	GLU (243-) X	-	4.80
145	SER (145-) X	-	4.75
85	GLY (85-) X	-	4.72
269	GLU (269-) X	-	4.70
94	ILE (94-) X	-	4.60
131	TYR (131-) X	-	4.57
276	HIS (276-) X	-	4.54
237	VAL (237-) X	-	4.43
74	ILE (74-) X	-	4.40
353	SER (353-) X	-	4.25
242	ILE (242-) X	-	4.23
227	GLN (227-) X	-	4.17
268	LYS (268-) X	-	4.14
285	SER (285-) X	-	4.07
75	LYS (75-) X	-	4.05
47	ASP (47-) X	-	4.01

2.6.10 Warning: High tau angle deviations

The RMS Z-score for the tau angles in the structure is too high. For well refined structures this number is expected to be around 1.0. The fact that it is higher than 1.5 worries us. However, 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 : 2.185

2.6.11 Note: Side chain planarity OK

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

2.6.12 Error: Connections to aromatic rings out of plane

The atoms listed in the table below are connected to a planar aromatic group in the sidechain of a protein residue but were found to deviate from the least squares plane.

For all atoms that are connected to an aromatic side chain in a protein residue the distance of the atom to the least squares plane through the aromatic system was determined. This value was divided by the standard deviation from a distribution of similar values from a database of small molecule structures.

	Residue		Atom	Z-score
144	HIS (144-) X	-	CB	5.43
255	TYR (255-) X	-	OH	4.73
272	PHE (272-) X	-	CB	4.57
18	TYR (18-) X	-	CB	4.20

2.7 Torsion-related checks

2.7.1 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
136	PRO (136-) X	-	0.46	HIGH
253	PRO (253-) X	-	0.45	HIGH
257	PRO (257-) X	-	0.50	HIGH

2.7.2 Warning: Unusual PRO puckering phases

The proline residues listed in the table below have a puckering phase that is not expected to occur in protein structures. Puckering parameters were calculated by the method of Cremer and Pople [REF]. Normal PRO rings approximately show a so-called envelope conformation with the C-γ atom above the plane of the ring (ϕ=+72 degrees), or a half-chair conformation with C-γ below and C-β above the plane of the ring (ϕ=-90 degrees). If ϕ deviates strongly from these values, this is indicative of a very strange conformation for a PRO residue, and definitely requires a manual check of the data. 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 Phase	Conformation
112	PRO (112-) X	-	101.1	envelop C-β (108 degrees)
253	PRO (253-) X	-	111.2	envelop C-β (108 degrees)
259	PRO (259-) X	-	-18.2	half-chair C-α/N (-18 degrees)

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 IF 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
302	TYR (302-) X	-	-3.8
288	THR (288-) X	-	-3.5
259	PRO (259-) X	-	-3.0
147	TYR (147-) X	-	-3.0
18	TYR (18-) X	-	-2.8
274	ILE (274-) X	-	-2.8
255	TYR (255-) X	-	-2.6
130	PHE (130-) X	-	-2.6
32	THR (32-) X	-	-2.5
286	SER (286-) X	-	-2.5
358	SER (358-) X	-	-2.5
90	PRO (90-) X	-	-2.4
86	ARG (86-) X	-	-2.4
117	VAL (117-) X	-	-2.4
215	GLU (215-) X	-	-2.4
154	GLN (154-) X	-	-2.3
120	VAL (120-) X	-	-2.3
9	MET (9-) X	-	-2.3
167	TYR (167-) X	-	-2.2
298	GLU (298-) X	-	-2.2
61	SER (61-) X	-	-2.2
113	ILE (113-) X	-	-2.2
142	PHE (142-) X	-	-2.2
293	GLY (293-) X	-	-2.2
3	VAL (3-) X	-	-2.2
40	ILE (40-) X	-	-2.2
84	MET (84-) X	-	-2.2
225	ILE (225-) X	-	-2.2
254	GLN (254-) X	-	-2.1
326	GLU (326-) X	-	-2.1
273	LEU (273-) X	-	-2.1
281	GLU (281-) X	-	-2.1
252	ILE (252-) X	-	-2.1
276	HIS (276-) X	-	-2.1
45	SER (45-) X	-	-2.1
295	GLY (295-) X	-	-2.1
348	ILE (348-) X	-	-2.1
71	THR (71-) X	-	-2.1
214	SER (214-) X	-	-2.0
48	THR (48-) X	-	-2.0
275	ASP (275-) X	-	-2.0
6	VAL (6-) X	-	-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	ASP (2-) X	- Poor ϕ/ψ
13	ALA (13-) X	- Poor ϕ/ψ
17	SER (17-) X	- Poor ϕ/ψ
18	TYR (18-) X	- Poor ϕ/ψ
19	ALA (19-) X	- Poor ϕ/ψ
50	THR (50-) X	- Poor ϕ/ψ
61	SER (61-) X	- Poor ϕ/ψ
65	ASN (65-) X	- Poor ϕ/ψ
75	LYS (75-) X	- Poor ϕ/ψ
84	MET (84-) X	- Poor ϕ/ψ
85	GLY (85-) X	- Poor ϕ/ψ
86	ARG (86-) X	- Poor ϕ/ψ
90	PRO (90-) X	- Poor PRO-ϕ
109	ARG (109-) X	- Poor ϕ/ψ
111	LEU (111-) X	- Poor ϕ/ψ
112	PRO (112-) X	- Poor ϕ/ψ
114	GLU (114-) X	- Poor ϕ/ψ
115	ASN (115-) X	- Poor ϕ/ψ
116	ASP (116-) X	- Poor ϕ/ψ
120	VAL (120-) X	- Poor ϕ/ψ
142	PHE (142-) X	- Poor ϕ/ψ
147	TYR (147-) X	- Poor ϕ/ψ
148	SER (148-) X	- Poor ϕ/ψ
157	ILE (157-) X	- Poor ϕ/ψ
163	LYS (163-) X	- Poor ϕ/ψ

And so on for a total of 58 lines.

2.7.5 Error: Ramachandran Z-score very low

The score expressing how well the backbone conformations of all residues are corresponding to the known allowed areas in the Ramachandran plot is very low.

Ramachandran Z-score : -6.169

2.7.6 Warning: Omega angles too tightly restrained

The ω angles for trans-peptide bonds in a structure are expected to give a gaussian distribution with the average around +178 degrees and a standard deviation around 5.5 degrees. These expected values were obtained from very accurately determined structures. Many protein structures are too tightly restrained. This seems to be the case with the current structure too, as the observed standard deviation is below 4.0 degrees.

Standard deviation of ω values : 1.856

2.7.7 Error: chi-1/chi-2 angle correlation Z-score very low

The score expressing how well the χ-1/χ-2 angles of all residues are corresponding to the populated areas in the database is very low.

χ-1/χ-2 correlation Z-score : -5.749

2.7.8 Warning: Backbone oxygen evaluation

The residues listed in the table below have an unusual backbone oxygen position.

For each of the residues in the structure, a search was performed to find 5-residue stretches in the WHAT IF database with superposable C-α coordinates, and some restraining on the neighbouring backbone oxygens.

In the following table the RMS distance between the backbone oxygen positions of these matching structures in the database and the position of the backbone oxygen atom in the current residue is given. If this number is larger than 1.5 a significant number of structures in the database show an alternative position for the backbone oxygen. If the number is larger than 2.0 most matching backbone fragments in the database have the peptide plane flipped. A manual check needs to be performed to assess whether the experimental data can support that alternative as well. The number in the last column is the number of database hits (maximum 80) used in the calculation. It is "normal" that some glycine residues show up in this list, but they are still worth checking!

	Residue		Distance (Å)	# hits
12	GLY (12-) X	-	2.34	11
101	GLY (101-) X	-	1.63	80

2.7.9 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
30	SER (30-) X	-	0.40

2.7.10 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!

	Residue		# hits
7	LEU (7-) X	-	0
8	HIS (8-) X	-	0
9	MET (9-) X	-	0
13	ALA (13-) X	-	0
15	GLU (15-) X	-	0
16	ASN (16-) X	-	0
17	SER (17-) X	-	0
18	TYR (18-) X	-	0
19	ALA (19-) X	-	0
21	ASN (21-) X	-	0
22	SER (22-) X	-	0
23	PHE (23-) X	-	0
45	SER (45-) X	-	0
47	ASP (47-) X	-	0
48	THR (48-) X	-	0
50	THR (50-) X	-	0
51	THR (51-) X	-	0
58	LEU (58-) X	-	0
60	CYS (60-) X	-	0
61	SER (61-) X	-	0
62	SER (62-) X	-	0
64	PRO (64-) X	-	0
65	ASN (65-) X	-	0
67	LEU (67-) X	-	0
83	LYS (83-) X	-	0

And so on for a total of 162 lines.

2.7.11 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 : -0.869

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 distance.

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.

The last text-item on each line represents the status of the atom pair. The text 'INTRA' means that the bump is between atoms that are explicitly listed in the PDB file. 'INTER' means it is an inter-symmetry bump. 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 aren't 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. In any case, each bump is listed in only one direction. If the MODEL number doesn't exist (like in most X-ray files), a minus sign is printed instead.

Atom 1					Atom 2				Bump	Dist	Status	
275	ASP (275-) X	-	O	<–> 277	ILE (277-) X	-	N		0.68	2.02	INTRA	BF
195	ARG (195-) X	-	NH1	<–> 198	GLU (198-) X	-	OE1		0.62	2.08	INTRA	BL
296	SER (296-) X	-	OG	<–> 340	MET (340-) X	-	SD		0.53	2.47	INTRA	BF
296	SER (296-) X	-	CB	<–> 340	MET (340-) X	-	SD		0.52	2.88	INTRA	BF
112	PRO (112-) X	-	C	<–> 114	GLU (114-) X	-	N		0.49	2.41	INTRA	BF
241	LEU (241-) X	-	O	<–> 243	GLU (243-) X	-	N		0.49	2.21	INTRA	BF
111	LEU (111-) X	-	O	<–> 113	ILE (113-) X	-	N		0.48	2.22	INTRA	BF
214	SER (214-) X	-	O	<–> 216	ASP (216-) X	-	N		0.45	2.25	INTRA	BF
216	ASP (216-) X	-	O	<–> 218	ALA (218-) X	-	N		0.45	2.25	INTRA	BF
112	PRO (112-) X	-	O	<–> 114	GLU (114-) X	-	N		0.42	2.28	INTRA	BF
117	VAL (117-) X	-	CB	<–> 121	CYS (121-) X	-	SG		0.42	2.98	INTRA	BF
213	ARG (213-) X	-	NH2	<–> 281	GLU (281-) X	-	CG		0.41	2.69	INTRA	BF
15	GLU (15-) X	-	CG	<–> 16	ASN (16-) X	-	N		0.41	2.59	INTRA	BF
170	ASN (170-) X	-	CG	<–> 171	THR (171-) X	-	N		0.40	2.60	INTRA	BF
245	GLU (245-) X	-	N	<–> 246	LYS (246-) X	-	N		0.39	2.21	INTRA	BF
24	ILE (24-) X	-	CD1	<–> 286	SER (286-) X	-	N		0.39	2.71	INTRA	BF
226	TRP (226-) X	-	NE1	<–> 255	TYR (255-) X	-	CZ		0.39	2.71	INTRA	BF
226	TRP (226-) X	-	NE1	<–> 255	TYR (255-) X	-	OH		0.38	2.32	INTRA	BF
167	TYR (167-) X	-	OH	<–> 257	PRO (257-) X	-	CD		0.38	2.42	INTRA	BL
179	ALA (179-) X	-	N	<–> 180	TYR (180-) X	-	N		0.38	2.22	INTRA	BF
153	SER (153-) X	-	CB	<–> 187	ASP (187-) X	-	OD1		0.36	2.44	INTRA	BL
339	ARG (339-) X	-	NH2	<–> 343	GLU (343-) X	-	OE1		0.35	2.35	INTRA	BF
2	ASP (2-) X	-	CB	<–> 4	ARG (4-) X	-	NH2		0.34	2.76	INTRA	BF
177	LEU (177-) X	-	O	<–> 179	ALA (179-) X	-	N		0.34	2.36	INTRA	BF
116	ASP (116-) X	-	C	<–> 117	VAL (117-) X	-	CG2		0.33	2.77	INTRA	BF

And so on for a total of 262 lines.

2.9 Packing, accessibility and threading

2.9.1 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 : 1.050

2.9.2 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: X

2.9.3 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	Qualty value
9	MET (9-) X	- -7.18
289	LYS (289-) X	- -6.40
319	HIS (319-) X	- -5.50
154	GLN (154-) X	- -5.30
297	VAL (297-) X	- -5.08
157	ILE (157-) X	- -5.08

2.9.4 Warning: Abnormal packing environment for sequential residues

A stretch of at least three sequential residues with a questionable packing environment was found. This could indicate that these residues are part of a strange loop. It might also be an indication of misthreading in the density. However, it can also indicate that one or more residues in this stretch have other problems such as, for example, missing atoms, very weird angles or bond lengths, etc.

The table below lists the first and last residue in each stretch found, as well as the average residue score of the series.