

Report of protein analysis

By the WHAT IF program

2010-09-19*

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 20100906-1130

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 pdb1bfx.ent

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

2.1.1 Note: Cell is 1 Angstrom cube

The unit cell in the CRYST1 card of the PDB file is given as a cube with vertices of 1 Å. This is the convention for structures obtained using a method other than crystallography.

2.1.2 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 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)	94 (94)	A	Protein	SET.1bfx
2	95	(94)	95 (94)	A	L O2 <- 94	SET.1bfx
3	96	(96)	96 (96)	A	HEM	SET.1bfx

2.3.2 Note: Some notes regarding the PDB file contents

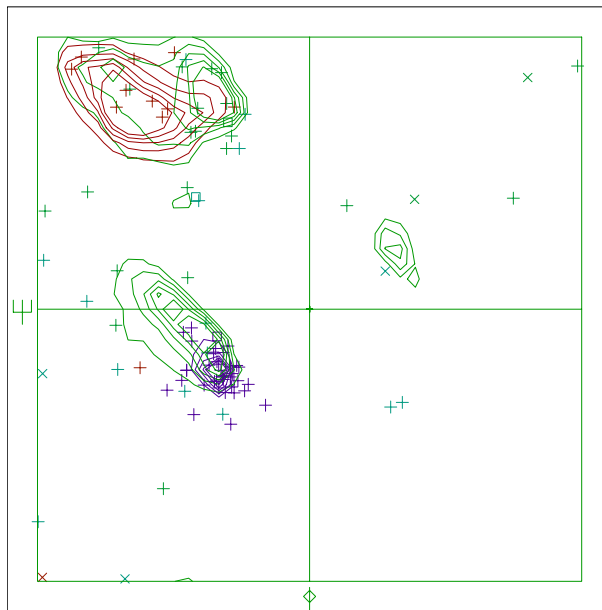
The numbers and remarks listed below have no explicit validation purpose, they are merely meant for the crystallographer or NMR spectroscopists to perhaps pinpoint something unexpected. See the WHAT_CHECK course [REF] for an explanation of terms like 'poor', 'missing', etcetera (in case those words pop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

The total number of amino acids found is 94.

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 $\phi=-60$ 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: 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 <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
DKDVKYYTLEEIQKHKDSKSTWVILHHKVYDLTKFLEEHPGGEVLRQAGGDATENFED
      HHHHHHTTTT  SSSSTSSSS TTHHHHTTT HHHHHHTTT HHHHHH
              70      80      90
VGHSTDARELSKTYIIGELHPDDRSKIAPSETL
H  TTHHHHHHHSSSS HHHHHH  T

```

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 $\chi^2=0.00$ or $\chi^2=180.00$ 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 elementary 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 not mentioned in PDB file
Room temperature assumed

2.4.12 Warning: More than 5 percent of buried atoms has low B-factor

For normal protein structures, no more than about 1 percent of the B factors of buried atoms is below 5.0. The fact that this value is much higher in the current structure could be a signal of overrefined B-factors, restraints or constraints to too-low values, misuse of the B-factor field in the PDB file, or a scaling problem. If the average B factor is low too, it is probably a low temperature structure determination.

Percentage of buried atoms with B less than 5 : 100.00

2.4.13 Warning: B-factor plot impossible

All average B-factors are zero. Plot suppressed.

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 2 and the other uses C $\delta 2$ 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 Error: Threonine nomenclature problem

The threonine residues listed in the table below have their O- γ -1 and C- γ -2 swapped.

Residue	
93	THR (93-) A -

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 Note: Tyrosine torsion conventions OK

No errors were detected in tyrosine torsion angle conventions.

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 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 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
93	THR (93-) A	- CB OG1	1.53	6.3

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: 0.581

RMS-deviation in bond distances: 0.012

2.6.3 Warning: Directionality in bond lengths and no X-ray cell

Comparison of bond distances with Engh and Huber [REF] standard values for protein residues and Parkinson et al [REF] standard values for DNA/RNA shows a significant systematic deviation.

You have most probably seen symmetry problems earlier. Please correct these and rerun this check to see the possible implications on the X-ray cell axes.

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
3	ASP (3-) A	- -C N CA	129.00	4.1
4	VAL (4-) A	- C CA CB	118.71	4.5
5	LYS (5-) A	- -C N CA	130.81	5.1
18	SER (18-) A	- -C N CA	128.96	4.0
26	HIS (26-) A	- -C N CA	128.93	4.0
44	GLU (44-) A	- CG CD OE2	79.16	-17.1
44	GLU (44-) A	- CG CD OE1	164.05	19.8
44	GLU (44-) A	- OE2 CD OE1	84.89	-15.8
58	PHE (58-) A	- CA CB CG	121.04	7.2
92	GLU (92-) A	- -C N CA	130.32	4.8

92	GLU (92-)	A	-	C	CA	CB	119.34	4.9
92	GLU (92-)	A	-	CA	CB	CG	126.80	6.3
93	THR (93-)	A	-	-CA	-C	N	126.79	5.3
93	THR (93-)	A	-	-C	N	CA	132.53	6.0

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: 1.230

RMS-deviation in bond angles: 2.292

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 were administrative corrections. These residues are listed here.

Residue	
93	THR (93-) A -

2.6.7 Warning: Chirality deviations detected

The atoms listed in the table below have an improper dihedral value that is deviating from expected values. As the improper dihedral values are all getting very close to ideal values in recent X-ray structures, and as we actually don't know how big the spread around these values should be, this check only warns for 6 sigma deviations.

Improper dihedrals are a measure of the chirality/planarity of the structure at a specific atom. Values around -35 or +35 are expected for chiral atoms, and values around 0 for planar atoms. Planar side chains are left out of the calculations, these are better handled by the planarity checks.

Three numbers are given for each atom in the table. The first is the Z-score for the improper dihedral. The second number is the measured improper dihedral. The third number is the expected value for this atom type. A final column contains an extra warning if the chirality for an atom is opposite to the expected value.

Please also see the previous table that lists a series of administrative chirality problems that were corrected automatically upon reading-in the PDB file.

	Residue	Atom	Z-value	Value	Expected	Notes
1	ASP (1-) A	- CA	-33.7	-33.35	33.73	Wrong hand
2	LYS (2-) A	- CA	-42.0	-35.83	33.92	Wrong hand
93	THR (93-) A	- CA	-40.3	-33.41	33.84	Wrong hand

2.6.8 Error: High improper dihedral angle deviations

The RMS Z-score for the improper dihedrals 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 2.5 worries us. However, we determined the improper normal distribution from 500 high-resolution X-ray structures, rather than from CSD data, so we cannot be 100 percent certain about these numbers.

Improper dihedral RMS Z-score : 6.744

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 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 : 1.226

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
30	TYR (30-) A	5.72

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
6	TYR (6-) A	OH	8.36
58	PHE (58-) A	CB	6.42
74	TYR (74-) A	CB	4.34

2.7 Torsion-related checks

2.7.1 Note: PRO puckering amplitude OK

Puckering amplitudes for all PRO residues are within normal ranges.

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
93	THR (93-) A -	-3.6
27	HIS (27-) A -	-3.2
28	LYS (28-) A -	-2.5
92	GLU (92-) A -	-2.5
23	VAL (23-) A -	-2.4
75	ILE (75-) A -	-2.2
18	SER (18-) A -	-2.2
19	LYS (19-) A -	-2.2
76	ILE (76-) 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	LYS (2-) A -	Poor ϕ/ψ
4	VAL (4-) A -	Poor ϕ/ψ
5	LYS (5-) A -	Poor ϕ/ψ
15	HIS (15-) A -	Poor ϕ/ψ
18	SER (18-) A -	Poor ϕ/ψ , ω poor
26	HIS (26-) A -	Poor ϕ/ψ
27	HIS (27-) A -	Poor ϕ/ψ
30	TYR (30-) A -	ω poor
43	GLU (43-) A -	ω poor
82	ASP (82-) A -	Poor ϕ/ψ
86	LYS (86-) A -	ω poor
87	ILE (87-) A -	Poor ϕ/ψ
88	ALA (88-) A -	ω poor
90	PRO (90-) A -	ω poor
91	SER (91-) A -	ω poor
92	GLU (92-) A -	Poor ϕ/ψ , ω poor
93	THR (93-) A -	ω poor

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 : -5.232

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 : 12.738

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 : -5.937

2.7.8 Note: Backbone oxygen evaluation OK

All residues for which the local backbone conformation could be found in the WHAT IF database have a normal backbone oxygen position.

2.7.9 Note: Peptide bond conformations

We could not find any peptide bonds that are likely to actually be a cis bond.

2.7.10 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
23	VAL (23-) A -	0.38
85	SER (85-) A -	0.38

2.7.11 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
3	ASP (3-) A	-	0
4	VAL (4-) A	-	0
5	LYS (5-) A	-	0
6	TYR (6-) A	-	0
15	HIS (15-) A	-	0
16	LYS (16-) A	-	0
17	ASP (17-) A	-	0
18	SER (18-) A	-	0
19	LYS (19-) A	-	0
20	SER (20-) A	-	0
21	THR (21-) A	-	0
22	TRP (22-) A	-	0
25	LEU (25-) A	-	0
26	HIS (26-) A	-	0
27	HIS (27-) A	-	0
32	LEU (32-) A	-	0
33	THR (33-) A	-	0
35	PHE (35-) A	-	0
36	LEU (36-) A	-	0
40	PRO (40-) A	-	0
43	GLU (43-) A	-	0
49	GLN (49-) A	-	0
50	ALA (50-) A	-	0
53	ASP (53-) A	-	0
54	ALA (54-) A	-	0
61	VAL (61-) A	-	0
63	HIS (63-) A	-	0
76	ILE (76-) A	-	0
78	GLU (78-) A	-	0
86	LYS (86-) A	-	0
87	ILE (87-) A	-	0
88	ALA (88-) A	-	0
89	LYS (89-) A	-	0
90	PRO (90-) A	-	0
91	SER (91-) A	-	0
92	GLU (92-) A	-	0
39	HIS (39-) A	-	1
75	ILE (75-) A	-	1
80	HIS (80-) A	-	2

2.7.12 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 : -1.686

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; 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 Å 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. 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).

	Atom 1				Atom 2				Bump Dist		Status		
92	GLU (92-)	A	-	CA <->	93	THR (93-)	A	-	CG2	0.29	2.81	INTRA	BL
92	GLU (92-)	A	-	C <->	93	THR (93-)	A	-	CG2	0.26	2.84	INTRA	BL
27	HIS (27-)	A	-	NE2 <->	87	ILE (87-)	A	-	CD1	0.22	2.88	INTRA	BL
39	HIS (39-)	A	-	NE2 <->	96	HEM (96-)	A	-	NA	0.18	2.82	INTRA	BL
63	HIS (63-)	A	-	NE2 <->	96	HEM (96-)	A	-	NB	0.17	2.83	INTRA	BL
63	HIS (63-)	A	-	NE2 <->	96	HEM (96-)	A	-	ND	0.16	2.84	INTRA	BL
39	HIS (39-)	A	-	NE2 <->	96	HEM (96-)	A	-	NB	0.15	2.85	INTRA	BL
63	HIS (63-)	A	-	NE2 <->	96	HEM (96-)	A	-	NC	0.14	2.86	INTRA	BL
92	GLU (92-)	A	-	CB <->	93	THR (93-)	A	-	CG2	0.11	2.99	INTRA	BL
39	HIS (39-)	A	-	NE2 <->	96	HEM (96-)	A	-	NC	0.11	2.89	INTRA	BL
63	HIS (63-)	A	-	NE2 <->	96	HEM (96-)	A	-	NA	0.11	2.89	INTRA	BL
89	LYS (89-)	A	-	NZ <->	92	GLU (92-)	A	-	O	0.09	2.61	INTRA	BL
91	SER (91-)	A	-	C <->	92	GLU (92-)	A	-	CG	0.09	3.01	INTRA	BL
17	ASP (17-)	A	-	O <->	19	LYS (19-)	A	-	N	0.07	2.63	INTRA	BL
89	LYS (89-)	A	-	C <->	90	PRO (90-)	A	-	C	0.06	2.74	INTRA	BL
39	HIS (39-)	A	-	NE2 <->	96	HEM (96-)	A	-	ND	0.05	2.95	INTRA	BL
54	ALA (54-)	A	-	CB <->	96	HEM (96-)	A	-	CBB	0.05	3.15	INTRA	BL
25	LEU (25-)	A	-	O <->	27	HIS (27-)	A	-	N	0.02	2.68	INTRA	BL
2	LYS (2-)	A	-	CG <->	3	ASP (3-)	A	-	N	0.02	2.98	INTRA	BL
86	LYS (86-)	A	-	C <->	87	ILE (87-)	A	-	C	0.02	2.78	INTRA	BL
26	HIS (26-)	A	-	N <->	55	THR (55-)	A	-	OG1	0.01	2.69	INTRA	BL
39	HIS (39-)	A	-	O <->	41	GLY (41-)	A	-	N	0.01	2.69	INTRA	BL
35	PHE (35-)	A	-	CD1 <->	74	TYR (74-)	A	-	CZ	0.01	3.19	INTRA	BL
31	ASP (31-)	A	-	OD1 <->	33	THR (33-)	A	-	N	0.00	2.70	INTRA	BL
91	SER (91-)	A	-	N <->	92	GLU (92-)	A	-	OE2	0.00	2.60	INTRA	BL

2.9 Packing, accessibility and threading

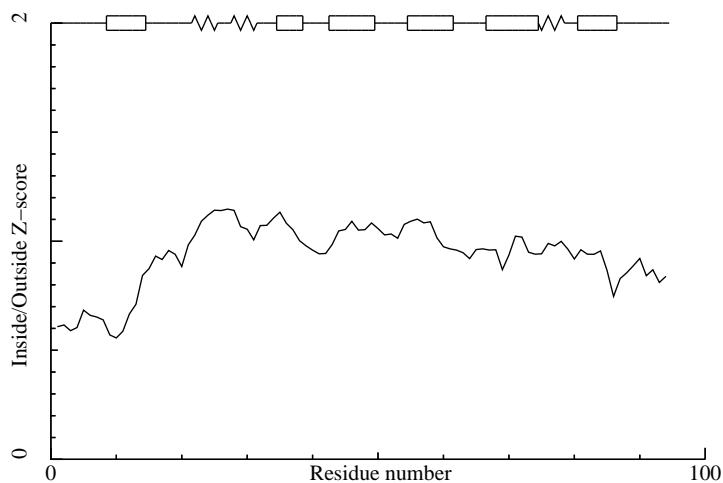
2.9.1 Note: Inside/Outside residue distribution normal

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

inside/outside RMS Z-score : 0.954

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: A

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	Quality value
2	LYS (2-) A	-6.13
49	GLN (49-) A	-5.32
92	GLU (92-) A	-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.

	Start residue	End residue	Av. Quality
91	SER (91-) A	93 — THR 93- (A)	-4.55

2.9.5 Warning: Structural average packing environment a bit worrysome

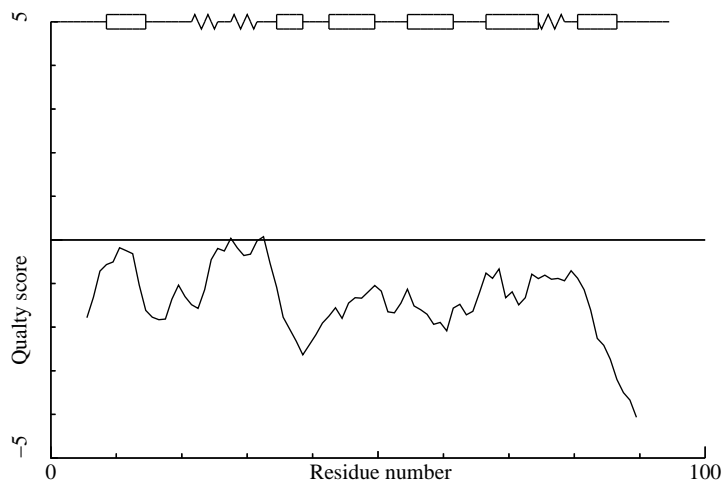
The structural average quality control value is a bit low.

The protein is probably threaded correctly, but either poorly refined, or it is just a protein with an unusual (but correct) structure. The average quality of 200 highly refined Xray structures was -0.5 ± 0.4 [REF].

Average for range 1 - 94 : -1.685

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 identifier: A

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.

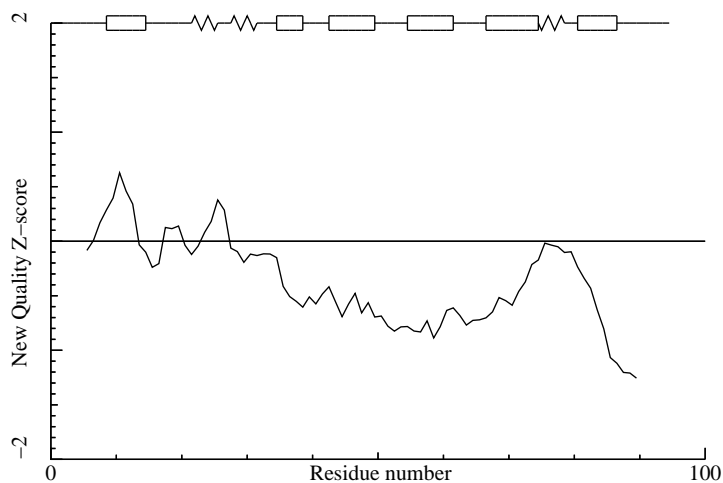
2.9.9 Note: Structural average packing Z-score OK

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

All contacts : Average = -0.352 Z-score = -2.02
BB-BB contacts : Average = -0.153 Z-score = -1.09
BB-SC contacts : Average = -0.335 Z-score = -2.54
SC-BB contacts : Average = -0.269 Z-score = -1.63
SC-SC contacts : Average = -0.302 Z-score = -1.65

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 hydrogenbond related checks

2.10.1 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	
27	HIS (27-) A -

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
15	HIS (15-) A	- HIS-D	0.97		
26	HIS (26-) A	- HIS-D	1.03	HIS-E	1.12
27	HIS (27-) A	- HIS-D	0.88		
39	HIS (39-) A	- HIS-E	1.12	HIS-D	1.17
63	HIS (63-) A	- HIS-D	1.07		
80	HIS (80-) A	- HIS-D	0.99	HIS-E	1.13

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
17	ASP (17-) A	- N
34	LYS (34-) A	- N
55	THR (55-) A	- OG1
66	ASP (66-) A	- N
67	ALA (67-) A	- N
78	GLU (78-) A	- 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
49	GLN (49-) A	- OE1

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)	94 (94)	A	Protein	SET.1bfx
2	95	(94)	95 (94)	A	L O2 <- 94	SET.1bfx
3	96	(96)	96 (96)	A	HEM	SET.1bfx

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 : -2.964
 2nd generation packing quality : -2.022
 Ramachandran plot appearance : -5.232 (bad)
 χ -1/ χ -2 rotamer normality : -5.937 (bad)
 Backbone conformation : -1.686

RMS Z-scores, should be close to 1.0:

Bond lengths : 0.581 (tight)
 Bond angles : 1.230
 Omega angle restraints : 2.316 (loose)
 Side chain planarity : 1.658
 Improper dihedral distribution : 6.744 (loose)
 Inside/Outside distribution : 0.954

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.Engl 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).