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 pdb5tln.ent

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

2.1.1 Note: The SCALE matrix has a nonstandard orientation

The SCALE matrix represents the same cell as the CRYST1 card. However, the orientation of the SCALE matrix is different.

Possible cause: The SCALE matrix may represent the actual orientation of a crystal on a diffractometer or another convenient orthogonal system.

SCALE matrix

$$\begin{bmatrix} 0.012258 & 0.000000 & 0.000000 \\ 0.006129 & 0.010616 & 0.000000 \\ 0.000000 & 0.000000 & 0.007610 \end{bmatrix}$$

Calculated from CRYST1

$$\begin{bmatrix} 0.010616 & 0.006129 & 0.000000 \\ 0.000000 & 0.012258 & 0.000000 \\ 0.000000 & 0.000000 & 0.007610 \end{bmatrix}$$

The SCALE transformation matrix

$$\begin{bmatrix} 0.866023 & 0.499999 & 0.000000 \\ -0.499986 & 0.866002 & 0.000000 \\ 0.000000 & 0.000000 & 1.000046 \end{bmatrix}$$

(This is a rotation of 30.0 degrees)

2.1.2 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: 34878.902
Volume of the Unit Cell V= 1009798.7
Cell multiplicity: 12
Matthews coefficient for observed atoms Vm= 2.413

2.1.3 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...

2.1.4 Error: Atoms too close to symmetry axis

The atoms listed in the table below are closer than 0.77 Å to a proper symmetry axis. This creates a bump between the atom and its symmetry relative(s). It is likely that these represent refinement

artefacts. The number in the right-hand column is the number of the symmetry matrix that was applied when this problem was detected.

	Atom	Matrix	
324	HOH (485) Al-	O	7

2.1.5 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)	316 (	316)	A	Protein	SET.5TLN
2	317 (	316)	317 (	316)	A	K O2 <- 316	SET.5TLN
3	318 (	317)	318 (	317)	A	CA	SET.5TLN
4	319 (	318)	319 (	318)	A	CA	SET.5TLN
5	320 (	319)	320 (	319)	A	CA	SET.5TLN
6	321 (	320)	321 (	320)	A	CA	SET.5TLN
7	322 (	321)	322 (	321)	A	ZN	SET.5TLN
8	323 (	322)	323 (	322)	A	BAN	SET.5TLN
9	324 (HOH)		324 (HOH)		A	water (146)	SET.5TLN

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).

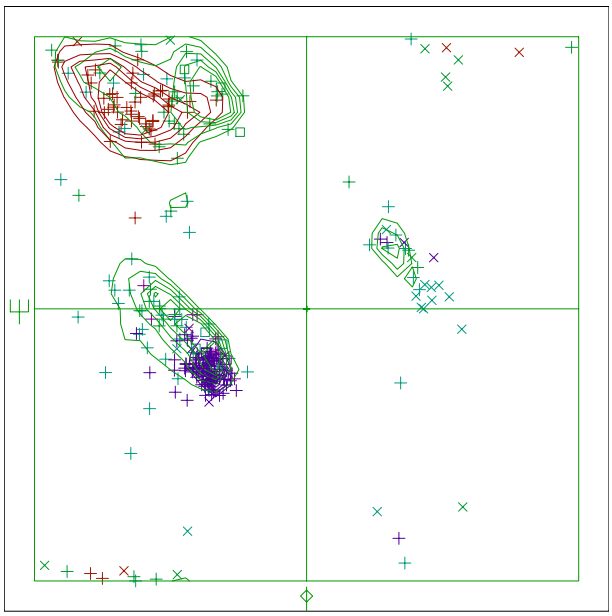
The total number of amino acids found is 316.
Number of water molecules 146

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: A

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
ITGTSTVGVGRGVLGDQKNINTTYSTYYYLQDNTRGDGIFTYDAKYRTTLPGSLWADADN
   SSSSSSSS TTT SSSSSSSSTTTSSTSS TTTT SSSSS TTTT TT SS SSTTT
        70        80        90       100       110       120
QFFASYDAPAVDAHYYAGVTYDYYKNVHNRLSYDGNNAAIRSSVHYSQGYNNAFWNGSEM
SS T333HHHHHHHHHHHHHHHHHHHHT   TTTTT   SSSSST TTT   SSS TT S
       130       140       150       160       170       180
VYGDGDGQTFIPLSGGIDVVAHELTHAVTDYTAGLIYQNESGAINEAISDIFGTLVEFYA
SS    TTT   333 HHHHHHHHHHHHHHHTT   TTHHHHHHHHHHHHHHHHHHHHHH
       190       200       210       220       230       240
NKNPDWEIGEDVYTPGISGDSLRSMSDPAKYGDPDHYSKRYTGTQDNGGVHINSGIINKA
TTT  TSSTTTT  TTTTT  TSSTT 3333T  TT333   TTTHHHHTTTTHHHHHHH
       250       260       270       280       290       300
AYLISQGGTHYGVSVVGIGRDKLGKIFYRALTQYLTPTSNFSQLRAAAVQSATDLYGSTS
HHHHHHSSSSTTSSSS   HHHHHHHHHHHHHHT  TT  HHHHHHHHHHHHHHHH TT
       310
QEVASVKQAFDAVGVK
HHHHHHHHHHHHTT
```

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.

Crystal temperature :285.000

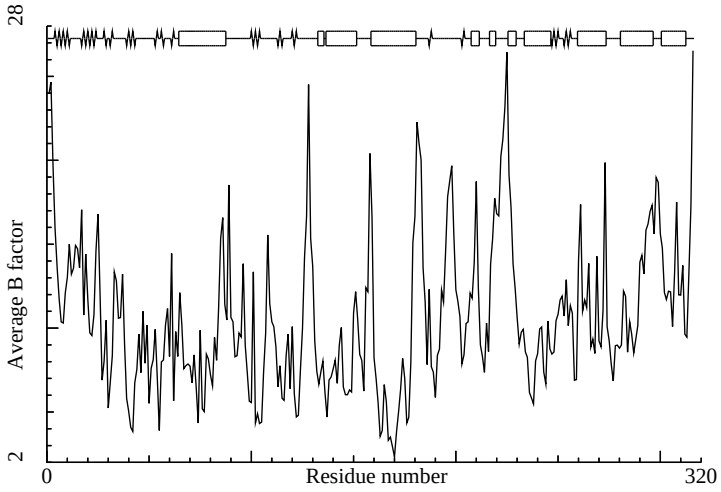
2.4.12 Warning: More than 5 percent of buried atoms has low B-factor

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 : 12.96

2.4.13 Note: B-factor plot

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



Chain identifier: A

2.5 Nomenclature related problems

2.5.1 Note: Introduction to the nomenclature section.

Nomenclature problems seem, at first, rather unimportant. After all who cares if we call the δ atoms in leucine 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	
47	ARG (47-) A	-

2.5.7 Warning: Tyrosine convention problem

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

	Residue	
27	TYR (27-) A	-
42	TYR (42-) A	-
46	TYR (46-) A	-
66	TYR (66-) A	-
76	TYR (76-) A	-
81	TYR (81-) A	-
84	TYR (84-) A	-
122	TYR (122-) A	-
157	TYR (157-) A	-
179	TYR (179-) A	-
211	TYR (211-) A	-
242	TYR (242-) A	-
251	TYR (251-) A	-
268	TYR (268-) A	-
296	TYR (296-) A	-

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	
40	PHE (40-) A	-
62	PHE (62-) A	-
63	PHE (63-) A	-
114	PHE (114-) A	-
130	PHE (130-) A	-
267	PHE (267-) A	-

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	
32	ASP (32-) A	-
37	ASP (37-) A	-
72	ASP (72-) A	-
94	ASP (94-) A	-
124	ASP (124-) A	-

126	ASP (126-) A	-
138	ASP (138-) A	-
170	ASP (170-) A	-
185	ASP (185-) A	-
261	ASP (261-) A	-

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	
166	GLU (166-) A	-
187	GLU (187-) A	-

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
1	ILE (1-) A	-	CA	CB	1.63	5.0
2	THR (2-) A	-	CB	OG1	1.52	5.5
3	GLY (3-) A	-	C	O	1.32	4.2
4	THR (4-) A	-	N	CA	1.57	6.0
4	THR (4-) A	-	C	O	1.33	5.1
4	THR (4-) A	-	CA	CB	1.63	5.2
4	THR (4-) A	-	CB	OG1	1.36	-4.3
5	SER (5-) A	-	C	O	1.14	-4.4

6	THR (6-) A	-	CB	OG1	1.51	4.6
8	GLY (8-) A	-	N	CA	1.57	7.4
9	VAL (9-) A	-	C	O	1.35	5.8
10	GLY (10-) A	-	N	CA	1.53	5.1
10	GLY (10-) A	-	C	O	1.33	4.8
10	GLY (10-) A	-	N	-C	1.24	-4.2
11	ARG (11-) A	-	NE	CZ	1.42	6.5
11	ARG (11-) A	-	CZ	NH2	1.40	4.1
11	ARG (11-) A	-	N	-C	1.16	-8.2
12	GLY (12-) A	-	N	-C	1.23	-4.9
17	GLN (17-) A	-	CD	OE1	1.32	4.7
20	ILE (20-) A	-	C	O	1.38	7.3
21	ASN (21-) A	-	CG	OD1	1.14	-4.6
21	ASN (21-) A	-	CG	ND2	1.22	-5.1
23	THR (23-) A	-	N	-C	1.24	-4.2
24	TYR (24-) A	-	C	O	1.33	4.8
25	SER (25-) A	-	CB	OG	1.54	5.9

And so on for a total of 189 lines.

2.6.2 Warning: High bond length deviations

Bond lengths were found to deviate more than normal from the mean standard bond lengths (standard values for protein residues were taken from Engh and Huber [REF], for DNA/RNA these values were taken from Parkinson et al [REF]). The RMS Z-score given below is expected to be around 1.0 for a normally restrained data set. The fact that it is higher than 1.5 in this structure might indicate that the restraints used in the refinement were not strong enough. This will also occur if a different bond length dictionary is used.

RMS Z-score for bond lengths: 2.190
RMS-deviation in bond distances: 0.045

2.6.3 Warning: Possible cell scaling problem

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

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

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

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

Unit Cell deformation matrix

$$\begin{bmatrix} 0.999223 & -0.000721 & 0.000166 \\ -0.000721 & 0.996954 & 0.000946 \\ 0.000166 & 0.000946 & 0.999267 \end{bmatrix}$$

Proposed new scale matrix

$$\begin{bmatrix} 0.012268 & 0.000009 & -0.000002 \\ 0.006141 & 0.010653 & -0.000011 \\ -0.000001 & -0.000007 & 0.007616 \end{bmatrix}$$

With corresponding cell

$A =$ 94.131 $B =$ 93.911 $C =$ 131.310
$\alpha =$ 89.891 $\beta =$ 90.038 $\gamma =$ 120.005

The CRYST1 cell dimensions

$A =$ 94.200 $B =$ 94.200 $C =$ 131.400
$\alpha =$ 90.000 $\beta =$ 90.000 $\gamma =$ 120.000

Variance: 30.496
(Under-)estimated Z-score: 4.070

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
1	ILE (1-) A	-	CA	CB	CG2	117.86	4.3
1	ILE (1-) A	-	CA	CB	CG1	99.62	-6.3
2	THR (2-) A	-	-O	-C	N	113.88	-5.7
2	THR (2-) A	-	CA	CB	OG1	116.89	4.9
3	GLY (3-) A	-	-O	-C	N	113.84	-5.7
3	GLY (3-) A	-	-CA	-C	N	130.96	7.4
4	THR (4-) A	-	-C	N	CA	113.65	-4.5
4	THR (4-) A	-	N	CA	CB	102.18	-4.9
4	THR (4-) A	-	C	CA	CB	118.33	4.3
4	THR (4-) A	-	CA	CB	CG2	120.40	5.8
5	SER (5-) A	-	N	CA	CB	118.93	5.0
5	SER (5-) A	-	CA	CB	OG	123.26	6.1
6	THR (6-) A	-	CA	CB	CG2	118.28	4.6
7	VAL (7-) A	-	CA	C	O	109.47	-6.7
8	GLY (8-) A	-	-O	-C	N	129.82	4.3
9	VAL (9-) A	-	-O	-C	N	113.48	-5.9
9	VAL (9-) A	-	-CA	-C	N	128.60	5.8
9	VAL (9-) A	-	C	CA	CB	97.26	-6.8
9	VAL (9-) A	-	CA	CB	CG1	121.27	6.3
10	GLY (10-) A	-	-CA	-C	N	128.09	5.9
11	ARG (11-) A	-	CA	C	O	112.91	-4.6
11	ARG (11-) A	-	CG	CD	NE	102.77	-4.5
11	ARG (11-) A	-	CD	NE	CZ	107.77	-7.7
11	ARG (11-) A	-	NH1	CZ	NH2	128.21	4.7

12 GLY (12-) A - -CA -C N 124.80 4.3

And so on for a total of 625 lines.

2.6.5 Warning: High bond angle deviations

Bond angles were found to deviate more than normal 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. The fact that it is higher than 2.0 in this structure might indicate that the restraints used in the refinement were not strong enough. This will also occur if a different bond angle dictionary is used.

RMS Z-score for bond angles: 3.033
RMS-deviation in bond angles: 5.531

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	
32	ASP (32-) A	-
37	ASP (37-) A	-
47	ARG (47-) A	-
72	ASP (72-) A	-
94	ASP (94-) A	-
124	ASP (124-) A	-
126	ASP (126-) A	-
138	ASP (138-) A	-
166	GLU (166-) A	-
170	ASP (170-) A	-
185	ASP (185-) A	-
187	GLU (187-) A	-
261	ASP (261-) 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
14	LEU (14-) A	-	CG	-6.1	-43.83	-33.01	
26	THR (26-) A	-	CA	-6.4	23.23	33.84	
79	VAL (79-) A	-	CB	6.3	-24.66	-32.96	
119	GLU (119-) A	-	CA	9.8	49.99	33.96	
144	LEU (144-) A	-	CG	-6.9	-45.09	-33.01	
156	ILE (156-) A	-	CB	-6.2	24.26	32.31	
220	ARG (220-) A	-	CA	7.2	45.74	33.91	
224	THR (224-) A	-	CA	6.6	44.79	33.84	
306	VAL (306-) A	-	CB	6.5	-24.41	-32.96	

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 : 1.766

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
260	ARG (260-) A	-	7.82
209	ALA (209-) A	-	7.22
164	ILE (164-) A	-	6.99
170	ASP (170-) A	-	6.80
229	GLY (229-) A	-	6.60
133	LEU (133-) A	-	6.53
224	THR (224-) A	-	6.04
301	GLN (301-) A	-	5.97
272	THR (272-) A	-	5.87
261	ASP (261-) A	-	5.74
295	LEU (295-) A	-	5.74
282	SER (282-) A	-	5.68
139	VAL (139-) A	-	5.37
270	ALA (270-) A	-	5.37
163	ALA (163-) A	-	5.09
38	GLY (38-) A	-	4.91
78	GLY (78-) A	-	4.87
267	PHE (267-) A	-	4.76
101	ARG (101-) A	-	4.65
115	TRP (115-) A	-	4.62
274	TYR (274-) A	-	4.60
263	LEU (263-) A	-	4.56
75	TYR (75-) A	-	4.52
256	VAL (256-) A	-	4.38
135	GLY (135-) A	-	4.38
161	SER (161-) A	-	4.38
147	ALA (147-) A	-	4.36
292	ALA (292-) A	-	4.30

187	GLU (187-) A	- 4.29
28	TYR (28-) A	- 4.28
287	ALA (287-) A	- 4.28
234	SER (234-) A	- 4.25
306	VAL (306-) A	- 4.25
1	ILE (1-) A	- 4.11
123	GLY (123-) A	- 4.02

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 : 2.479

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 Note: Atoms connected to aromatic rings OK

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

2.7 Torsion-related checks

2.7.1 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
51	PRO (51-) A	-	0.08	LOW
132	PRO (132-) A	-	0.52	HIGH
184	PRO (184-) A	-	0.17	LOW
195	PRO (195-) A	-	0.04	LOW
208	PRO (208-) A	-	0.45	HIGH
214	PRO (214-) A	-	0.47	HIGH

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
107	SER (107-) A	-2.6
46	TYR (46-) A	-2.6
157	TYR (157-) A	-2.5
20	ILE (20-) A	-2.4
197	ILE (197-) A	-2.4
132	PRO (132-) A	-2.4
26	THR (26-) A	-2.4
25	SER (25-) A	-2.4
18	LYS (18-) A	-2.3
313	VAL (313-) A	-2.2
92	SER (92-) A	-2.2
45	LYS (45-) A	-2.2
49	THR (49-) A	-2.2
47	ARG (47-) A	-2.2
182	LYS (182-) A	-2.1
185	ASP (185-) A	-2.1
89	ASN (89-) A	-2.0
219	LYS (219-) 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
26	THR (26-) A	- Poor ϕ/ψ
46	TYR (46-) A	- Poor ϕ/ψ
50	LEU (50-) A	- PRO ω poor
60	ASN (60-) A	- Poor ϕ/ψ
89	ASN (89-) A	- Poor ϕ/ψ
92	SER (92-) A	- Poor ϕ/ψ
97	ASN (97-) A	- Poor ϕ/ψ
105	HIS (105-) A	- Poor ϕ/ψ
107	SER (107-) A	- Poor ϕ/ψ
112	ASN (112-) A	- Poor ϕ/ψ
118	SER (118-) A	- Poor ϕ/ψ
152	THR (152-) A	- Poor ϕ/ψ
157	TYR (157-) A	- Poor ϕ/ψ
159	ASN (159-) A	- Poor ϕ/ψ

181	ASN (181-) A	- Poor ϕ/ψ
184	PRO (184-) A	- Poor PRO-ϕ
194	THR (194-) A	- Poor ϕ/ψ

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 : -2.010

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 : 1.382

2.7.7 Warning: chi-1/chi-2 angle correlation Z-score 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 a bit low.

χ-1/χ-2 correlation Z-score : -3.542

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
291	SER (291-) A	-	0.36

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
5	SER (5-) A	-	0
9	VAL (9-) A	-	0
14	LEU (14-) A	-	0
17	GLN (17-) A	-	0
19	ASN (19-) A	-	0
21	ASN (21-) A	-	0
24	TYR (24-) A	-	0
25	SER (25-) A	-	0
26	THR (26-) A	-	0
27	TYR (27-) A	-	0
29	TYR (29-) A	-	0
30	LEU (30-) A	-	0
31	GLN (31-) A	-	0
34	THR (34-) A	-	0
35	ARG (35-) A	-	0
37	ASP (37-) A	-	0
44	ALA (44-) A	-	0
45	LYS (45-) A	-	0
46	TYR (46-) A	-	0
51	PRO (51-) A	-	0
53	SER (53-) A	-	0
55	TRP (55-) A	-	0
58	ALA (58-) A	-	0
60	ASN (60-) A	-	0
61	GLN (61-) A	-	0

And so on for a total of 129 lines.

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 : -1.033

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 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. 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	
61	GLN (61-) A	-	C	<–> 324	HOH (421) A	-	O	1.85	0.95	INTRA	
62	PHE (62-) A	-	N	<–> 324	HOH (421) A	-	O	1.66	1.04	INTRA	
92	SER (92-) A	-	CB	<–> 324	HOH (451) A	-	O	0.92	1.88	INTRA	
118	SER (118-) A	-	N	<–> 324	HOH (442) A	-	O	0.92	1.78	INTRA	
228	GLY (228-) A	-	C	<–> 324	HOH (382) A	-	O	0.86	1.94	INTRA	
61	GLN (61-) A	-	CA	<–> 324	HOH (421) A	-	O	0.85	1.95	INTRA	
92	SER (92-) A	-	OG	<–> 324	HOH (451) A	-	O	0.75	1.65	INTRA	
92	SER (92-) A	-	CA	<–> 324	HOH (451) A	-	O	0.73	2.07	INTRA	
62	PHE (62-) A	-	CA	<–> 324	HOH (421) A	-	O	0.61	2.19	INTRA	
119	GLU (119-) A	-	OE1	<–> 324	HOH (447) A	-	O	0.55	1.85	INTRA	
228	GLY (228-) A	-	CA	<–> 324	HOH (382) A	-	O	0.54	2.26	INTRA	
219	LYS (219-) A	-	CE	<–> 324	HOH (391) A	-	O	0.54	2.26	INTRA	
216	HIS (216-) A	-	ND1	<–> 324	HOH (435) A	-	O	0.53	2.17	INTRA	
269	ARG (269-) A	-	NH1	<–> 294	ASP (294-) A	-	OD2	0.52	2.18	INTRA	
218	SER (218-) A	-	CB	<–> 324	HOH (435) A	-	O	0.50	2.30	INTRA	
308	GLN (308-) A	-	CG	<–> 324	HOH (388) A	-	O	0.49	2.31	INTRA	
61	GLN (61-) A	-	O	<–> 324	HOH (421) A	-	O	0.45	1.95	INTRA	
229	GLY (229-) A	-	N	<–> 324	HOH (382) A	-	O	0.44	2.26	INTRA	
273	GLN (273-) A	-	CG	<–> 324	HOH (372) A	-	O	0.42	2.38	INTRA	
126	ASP (126-) A	-	OD1	<–> 128	GLN (128-) A	-	N	0.39	2.31	INTRA	BL
116	ASN (116-) A	-	OD1	<–> 324	HOH (442) A	-	O	0.38	2.02	INTRA	
216	HIS (216-) A	-	CE1	<–> 324	HOH (435) A	-	O	0.34	2.46	INTRA	
33	ASN (33-) A	-	ND2	<–> 324	HOH (459) A	-	O	0.33	2.37	INTRA	
324	HOH (410) A	-	O	<–> 324	HOH (492) A	-	O	0.31	1.89	INTRA	
32	ASP (32-) A	-	OD2	<–> 35	ARG (35-) A	-	NH1	0.28	2.42	INTRA	BL

And so on for a total of 135 lines.

2.9 Packing, accessibility and threading

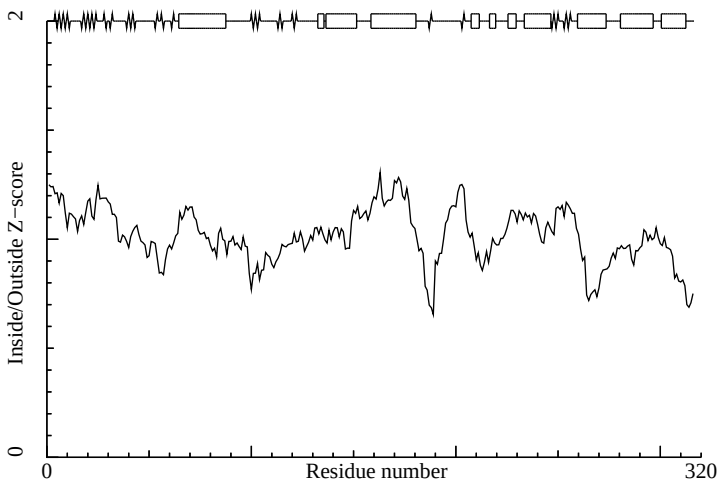
2.9.1 Note: Inside/Outside residue distribution normal

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

inside/outside RMS Z-score : 1.024

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	Qualty value
108	GLN (108-) A	- -5.78
225	GLN (225-) A	- -5.68
246	GLN (246-) A	- -5.57
157	TYR (157-) A	- -5.54
273	GLN (273-) A	- -5.17
251	TYR (251-) A	- -5.07
221	TYR (221-) A	- -5.05

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
225	GLN (225-) A	- 228 — GLY	228- (A) -	-4.60

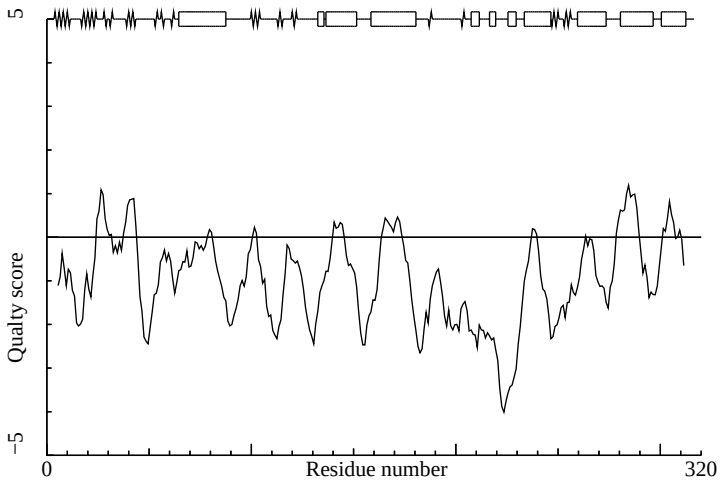
2.9.5 Note: Structural average packing environment OK

The structural average quality control value is within normal ranges.

Average for range 1 - 316 : -1.087

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 Warning: Low packing Z-score for some residues

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

	Residue		Z-score
211	TYR (211-) A	-	-2.72
47	ARG (47-) A	-	-2.52

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.

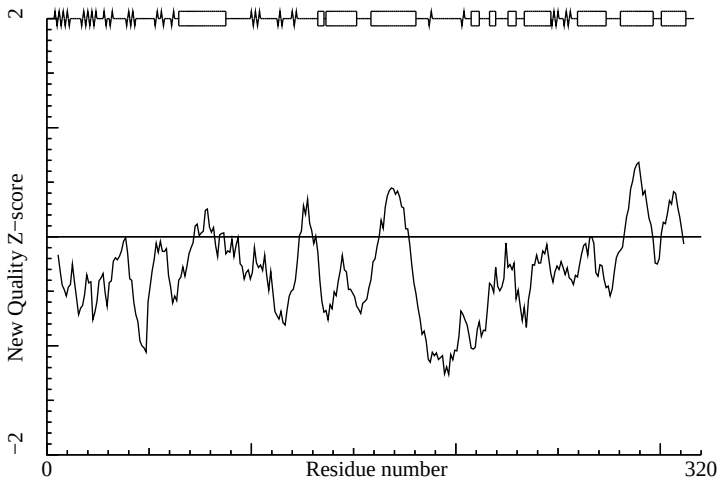
2.9.9 Note: Structural average packing Z-score OK

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

All contacts : Average = -0.329 Z-score = -1.89
BB-BB contacts : Average = -0.180 Z-score = -1.27
BB-SC contacts : Average = -0.329 Z-score = -2.50
SC-BB contacts : Average = -0.195 Z-score = -1.16
SC-SC contacts : Average = -0.327 Z-score = -1.80

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 Note: Water contacts OK

All water clusters make at least one contact with a non-water atom.

2.10.2 Warning: Water molecules need moving

The water molecules listed in the table below were found to be significantly closer to a symmetry related non-water molecule than to the ones given in the coordinate file. For optimal viewing convenience revised coordinates for these water molecules should be given.

The number in brackets is the identifier of the water molecule in the input file. Suggested coordinates are also given in the table. Please note that alternative conformations for protein residues are not taken into account for this calculation. If you are using WHAT IF / WHAT-CHECK interactively, then the moved waters can be found in PDB format in the file: MOVEDH2O.pdb.

Residue			Atom	Suggested coordinates		
324	HOH (416) A	-	O	66.53	5.33	-5.28
324	HOH (423) A	-	O	69.25	7.07	-5.32
324	HOH (436) A	-	O	38.56	29.93	-8.26
324	HOH (449) A	-	O	37.62	14.80	6.53
324	HOH (478) A	-	O	74.65	5.04	-11.70
324	HOH (495) A	-	O	29.73	38.77	12.91

2.10.3 Note: Water hydrogen bonds OK

All water molecules can form hydrogen bonds.

2.10.4 Error: HIS, ASN, GLN side chain flips

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

Residue		
19	ASN (19-) A	-
31	GLN (31-) A	-
33	ASN (33-) A	-
273	GLN (273-) A	-
280	ASN (280-) A	-
290	GLN (290-) A	-
301	GLN (301-) A	-

2.10.5 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
74	HIS (74-) A	- HIS-E	1.76		
88	HIS (88-) A	- HIS-E	2.30		
105	HIS (105-) A	- HIS-D	2.72	HIS-H	3.09
142	HIS (142-) A	- HIS-E	1.89	HIS-D	1.93
146	HIS (146-) A	- HIS-D	1.17		
216	HIS (216-) A	- HIS-E	2.96		
231	HIS (231-) A	- HIS-E	2.36	HIS-D	2.64
250	HIS (250-) A	- HIS-E	2.71		

2.10.6 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
35	ARG (35-) A	-	N
49	THR (49-) A	-	N
58	ALA (58-) A	-	N
90	ARG (90-) A	-	NE
97	ASN (97-) A	-	ND2
130	PHE (130-) A	-	N
155	LEU (155-) A	-	N
182	LYS (182-) A	-	NZ
203	ARG (203-) A	-	NH2
216	HIS (216-) A	-	N
234	SER (234-) A	-	N
261	ASP (261-) A	-	N
262	LYS (262-) A	-	N

2.10.7 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
119	GLU (119-) A	-	OE2
143	GLU (143-) A	-	OE1
143	GLU (143-) A	-	OE2
238	ASN (238-) A	-	OD1

2.10.8 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)	316 (316)	A	Protein		SET.5TLN
2	317 (316)	317 (316)	A	K O2 <-	316	SET.5TLN
3	318 (317)	318 (317)	A	CA		SET.5TLN
4	319 (318)	319 (318)	A	CA		SET.5TLN
5	320 (319)	320 (319)	A	CA		SET.5TLN
6	321 (320)	321 (320)	A	CA		SET.5TLN
7	322 (321)	322 (321)	A	ZN		SET.5TLN
8	323 (322)	323 (322)	A	BAN		SET.5TLN
9	324 (HOH)	324 (HOH)	A	water	(146)	SET.5TLN

2.10.9 Note: Crystallisation conditions from REMARK 280

Crystallisation conditions as found in the PDB file header.

REMARK 280

CRYSTAL
SOLVENT CONTENT, VS (%): 49.76
MATTHEWS COEFFICIENT, VM (ANGSTROMS**3/DA): 2.45
CRYSTALLIZATION CONDITIONS: PH 7.2

2.10.10 Note: Overview of ions

When ions are located at special positions, their occupancy should be reduce by a factor that is the same as the multiplicity of that special position. This seems to have been done OK in this PDB file.

2.10.11 Warning: Unusual ion packing

We implemented the ion valence determination method of Brown and Wu [REF] similar to Nayal and Di Cera [REF]. See also Mueller, Koepke and Sheldrick [REF]. It must be stated that the validation of ions in PDB files is very difficult. Ideal ion-ligand distances often differ no more than 0.1 Å, and in a 2.0 Å resolution structure 0.1 Å isn't very much. Nayal and Di Cera showed that this method has great potential, but the method has not been validated. Part of our implementation (comparing ion types) is even fully new and despite that we see it work well in the few cases that are trivial, we must emphasize that this validation method is untested. See: http://swift.cmbi.ru.nl/teach/theory/ for a detailed explanation.

The output gives the ion, the valency score for the ion itself, the valency score for the suggested alternative ion, and a series of possible comments *1 indicates that the suggested alternate atom type has been observed in the PDB file at another location in space. *2 indicates that WHAT IF thinks to have found this ion type in the crystallisation conditions as described in the REMARK 280 cards of the PDB file. *S Indicates that this ions is located at a special position (i.e. at a symmetry axis). N4 stands for NH4+.

	Ion / H2O		Certainty	Alternate ion
318	CA (317-) A	-	-.- -.-	Part of ionic cluster
318	CA (317-) A	-	0.64 0.87	Scores about as good as NA
319	CA (318-) A	-	-.- -.-	Part of ionic cluster

2.10.12 Warning: Unusual water packing

We implemented the ion valence determination method of Brown and Wu [REF] similar to Nayal and Di Cera [REF] and Mueller, Koepke and Sheldrick [REF]. It must be stated that the validation of ions in PDB files is very difficult. Ideal ion-ligand distances often differ no more than 0.1 Å, and in a 2.0 Å resolution structure 0.1 Å isn't very much. Nayal and Di Cera showed that this method nevertheless has great potential for detecting water molecules that actually should be metal ions. The method has not been extensively validated, though. Part of our implementation (comparing waters with multiple ion types) is even fully new and despite that we see it work well in the few cases that are trivial, we must emphasize that this method is untested.

The score listed is the valency score. This number should be close to (preferably a bit above) 1.0 for the suggested ion to be a likely alternative for the water molecule. Ions listed in brackets are good alternate choices. *1 indicates that the suggested ion-type has been observed elsewhere in the PDB file too. *2 indicates that the suggested ion-type has been observed in the REMARK 280 cards of the PDB file. Ion-B and ION-B indicate that the B-factor of this water is high, or very high, respectively. H2O-B indicates that the B-factors of atoms that surround this water/ion are suspicious. See: http://swift.cmbi.ru.nl/teach/theory/ for a detailed explanation.

	H2O			Certainty	Comments
324	HOH (333) A	-	O	0.95	CA *1 (or NA) Ion-B H2O-B
324	HOH (334) A	-	O	0.95	K
324	HOH (346) A	-	O	0.90	K
324	HOH (350) A	-	O	0.89	NA
324	HOH (353) A	-	O	1.15	K
324	HOH (382) A	-	O	1.01	K Ion-B
324	HOH (383) A	-	O	1.06	K
324	HOH (396) A	-	O	1.00	NA ION-B
324	HOH (410) A	-	O	1.03	CA *1
324	HOH (423) A	-	O	0.87	K
324	HOH (442) A	-	O	0.95	CA *1 (or NA) Ion-B
324	HOH (480) A	-	O	1.05	K
324	HOH (482) A	-	O	1.04	K Ion-B
324	HOH (492) A	-	O	0.98	NA ION-B

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 :	-1.467	
2nd generation packing quality :	-1.893	
Ramachandran plot appearance :	-2.010	
χ-1/χ-2 rotamer normality :	-3.542	(poor)
Backbone conformation :	-1.033	

RMS Z-scores, should be close to 1.0:

Bond lengths :	2.190	(loose)
Bond angles :	3.033	(loose)
Omega angle restraints :	0.251	(tight)
Side chain planarity :	1.221	
Improper dihedral distribution :	1.766	(loose)
Inside/Outside distribution :	1.024	

2.11.2 Note: Summary report for depositors of a structure

This is an overall summary of the quality of the X-ray structure as compared with structures solved at similar resolutions. This summary can be useful for a crystallographer to see if the structure makes the best possible use of the data. Warning. This table works well for structures solved in the resolution range of the structures in the WHAT IF database, which is presently (summer 2008) mainly 1.1 - 1.3 Å. The further the resolution of your file deviates from this range the more meaningless this table becomes.

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, which have been calibrated against structures of similar resolution.

Resolution found in PDB file : 2.30

Structure Z-scores, positive is better than average:

1st generation packing quality :	-1.0	
2nd generation packing quality :	-1.0	
Ramachandran plot appearance :	-0.5	
χ-1/χ-2 rotamer normality :	-1.8	
Backbone conformation :	-0.9	

RMS Z-scores, should be close to 1.0:

Bond lengths :	2.190	(loose)
Bond angles :	3.033	(loose)
Omega angle restraints :	0.251	(tight)
Side chain planarity :	1.221	
Improper dihedral distribution :	1.766	(loose)
Inside/Outside distribution :	1.024	

2.11.3 Note: Matrices used

When WHAT IF calculates interactions in the crystal, it normally uses more matrices then the space-group dictates because transformations that included a unit-cell translation become independent matrices. This table lists the matrices used.

Matrices used: 12

Symmetry operation number 1

$$\begin{bmatrix} 1.000000 & 0.000000 & 0.000000 & 0.000000 \\ 0.000000 & 1.000000 & 0.000000 & 0.000000 \\ 0.000000 & 0.000000 & 1.000000 & 0.000000 \end{bmatrix}$$

Symmetry operation number 2

$$\begin{bmatrix} 0.500000 & 0.866047 & 0.000000 & 0.000000 \\ -0.866004 & 0.500000 & 0.000000 & 0.000000 \\ 0.000000 & 0.000000 & 1.000000 & 109.505028 \end{bmatrix}$$

Symmetry operation number 3

$$\begin{bmatrix} -0.500000 & 0.866047 & 0.000000 & 0.000000 \\ -0.866004 & -0.500000 & 0.000000 & 0.000000 \\ 0.000000 & 0.000000 & 1.000000 & 87.604027 \end{bmatrix}$$

Symmetry operation number 4

$$\begin{bmatrix} -1.000000 & 0.000000 & 0.000000 & 0.000000 \\ 0.000000 & -1.000000 & 0.000000 & 0.000000 \\ 0.000000 & 0.000000 & 1.000000 & 65.703018 \end{bmatrix}$$

Symmetry operation number 5

$$\begin{bmatrix} -0.500000 & -0.866047 & 0.000000 & 0.000000 \\ 0.866004 & -0.500000 & 0.000000 & 0.000000 \\ 0.000000 & 0.000000 & 1.000000 & 43.802013 \end{bmatrix}$$

Symmetry operation number 6

$$\begin{bmatrix} 0.500000 & -0.866047 & 0.000000 & 0.000000 \\ 0.866004 & 0.500000 & 0.000000 & 0.000000 \\ 0.000000 & 0.000000 & 1.000000 & 21.901007 \end{bmatrix}$$

Symmetry operation number 7

$$\begin{bmatrix} 0.500000 & 0.866047 & 0.000000 & 0.000000 \\ 0.866004 & -0.500000 & 0.000000 & 0.000000 \\ 0.000000 & 0.000000 & -1.000000 & 43.802013 \end{bmatrix}$$

Symmetry operation number 8

$$\begin{bmatrix} 1.000000 & 0.000000 & 0.000000 & 0.000000 \\ 0.000000 & -1.000000 & 0.000000 & 0.000000 \\ 0.000000 & 0.000000 & -1.000000 & 21.901007 \end{bmatrix}$$

Symmetry operation number 9

$$\begin{bmatrix} 0.500000 & -0.866047 & 0.000000 & 0.000000 \\ -0.866004 & -0.500000 & 0.000000 & 0.000000 \\ 0.000000 & 0.000000 & -1.000000 & 0.000000 \end{bmatrix}$$

Symmetry operation number 10

$$\begin{bmatrix} -0.500000 & -0.866047 & 0.000000 & 0.000000 \\ -0.866004 & 0.500000 & 0.000000 & 0.000000 \\ 0.000000 & 0.000000 & -1.000000 & 109.505028 \end{bmatrix}$$

Symmetry operation number 11

$$\begin{bmatrix} -1.000000 & 0.000000 & 0.000000 & 0.000000 \\ 0.000000 & 1.000000 & 0.000000 & 0.000000 \\ 0.000000 & 0.000000 & -1.000000 & 87.604027 \end{bmatrix}$$

Symmetry operation number 12

$$\begin{bmatrix} -0.500000 & 0.866047 & 0.000000 & 0.000000 \\ 0.866004 & 0.500000 & 0.000000 & 0.000000 \\ 0.000000 & 0.000000 & -1.000000 & 65.703018 \end{bmatrix}$$

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).