Report of protein analysis

By the WHAT IF program

2010-07-21*

1 Introduction

This document contains a report of findings by 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.

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.

*This report was created by WHAT IF version WHATCHECK 8.0

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 pdb1crn.ent

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

2.1.1 Warning: Class of conventional cell differs from CRYST1 cell

The crystal class of the conventional cell is different from the crystal class of the cell given on the CRYST1 card. If the new class is supported by the coordinates this is an indication of a wrong space group assignment.

The CRYST1 cell dimensions

$A =$	40.960	$B =$	18.650	$C =$	22.520
$\alpha =$	90.000	$\beta =$	90.770	$\gamma =$	90.000

Dimensions of a reduced cell

$A =$	18.650	$B =$	22.520	$C =$	40.960
$\alpha =$	89.230	$\beta =$	90.000	$\gamma =$	90.000

Dimensions of the conventional cell

$A =$	18.650	$B =$	22.520	$C =$	40.960
$\alpha =$	90.770	$\beta =$	90.000	$\gamma =$	90.000

Transformation to conventional cell

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

Crystal class of the cell: MONOCLINIC

Crystal class of the conventional CELL: ORTHORHOMBIC

Space group name: P 1 21 1

Bravais type of conventional cell is: P

2.1.2 Error: Negated value in scale matrix

One or more of the values of the scale matrix are wrong.

Possible cause: Comparison with the matrix derived from the CRYST1 card reveals that values have been inverted in sign.

SCALE matrix

$$\begin{bmatrix} 0.024414 & 0.000000 & -0.000328 \\ 0.000000 & 0.053619 & 0.000000 \\ 0.000000 & 0.000000 & 0.044409 \end{bmatrix}$$

Calculated from CRYST1

$$\begin{bmatrix} 0.024414 & 0.000000 & 0.000328 \\ 0.000000 & 0.053619 & 0.000000 \\ 0.000000 & 0.000000 & 0.044409 \end{bmatrix}$$

2.1.3 Warning: Conventional cell is pseudo-cell

The extra symmetry that would be implied by the transition to the previously mentioned conventional cell has not been observed. It must be concluded that the crystal lattice has pseudo-symmetry.

2.1.4 Note: Matthews coefficient OK

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

Molecular weight of all polymer chains: 4722.412
Volume of the Unit Cell V= 17201.699
Cell multiplicity: 2
Matthews coefficient for observed atoms Vm= 1.821

2.1.5 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...

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

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

2.1.7 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 overlapping non-alternates detected

Either this structure does not contain overlapping non-alternate atoms, or they are all correct, or the errors have remained unnoticed.

2.2.6 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).

2.2.7 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.

2.2.8 Note: All residues have a complete backbone.

No residues have missing backbone atoms.

2.2.9 Note: No probable 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.

2.2.10 Note: No crippling errors.

Problems can exist that make it impossible to continue the validation. WHAT IF seems not to have encountered any of these.

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].

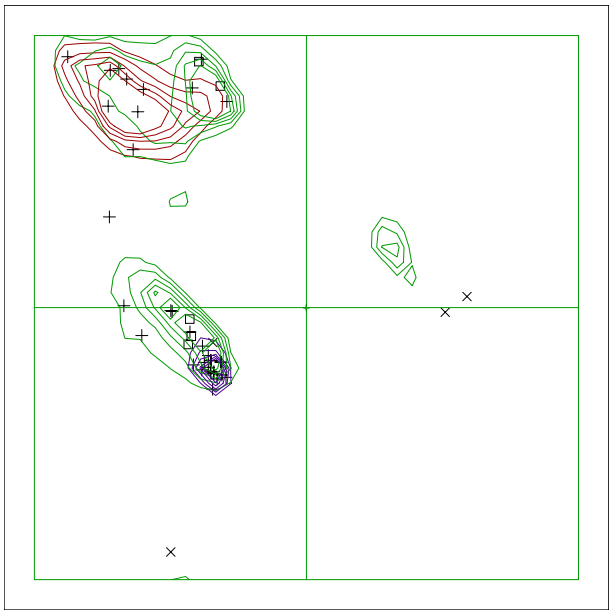
'Molecules'

1	1 (1) 46 (46) A	Protein		pdb1crn.ent
2	47 (46) 47 (46) A	N O2 <-	46	pdb1crn.ent

2.3.2 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

2.3.3 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/ds 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
TTCCPSIVARSNFNVCRLPGTPEAICATYTGCIIIPGATCPGDYAN
 SS TTHHHHHHHHHHHTTT  HHHHHHHHT SS TTT   333
```

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 Note: Average B-factor OK

The average B-factor of buried atoms is within expected values for a room-temperature X-ray study.

Average B-factor for buried atoms : 5.509

Crystal temperature : -1.000

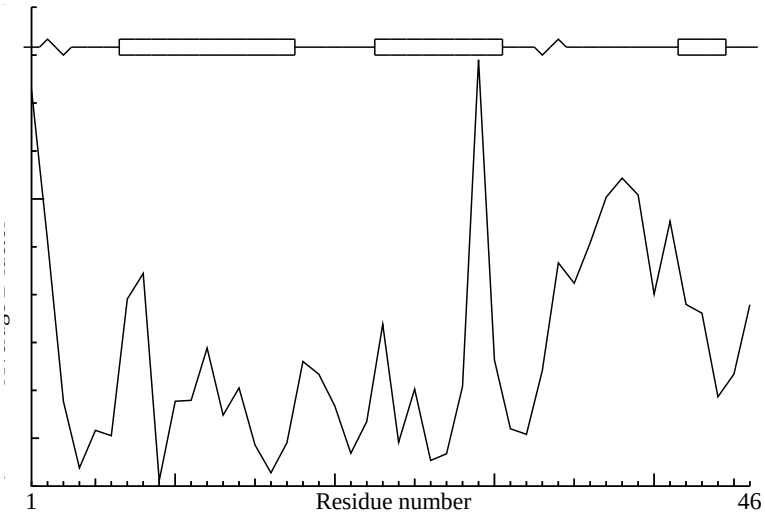
2.4.12 Warning: More than 5 percent of buried atoms has low B-factor

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 : 49.12

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 WHTA_CHECK website [REF] you can get an overview of the importance of all nomenclature problems that we list.

2.5.2 Note: Valine nomenclature OK

No errors were detected in valine nomenclature.

2.5.3 Note: Threonine nomenclature OK

No errors were detected in threonine nomenclature.

2.5.4 Note: Isoleucine nomenclature OK

No errors were detected in isoleucine nomenclature.

2.5.5 Note: Leucine nomenclature OK

No errors were detected in leucine nomenclature.

2.5.6 Note: Arginine nomenclature OK

No errors were detected in arginine nomenclature.

2.5.7 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
37	GLY (37-) A	-	N	CA	1.52	4.2

2.6.2 Note: Normal bond length variability

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

RMS Z-score for bond lengths: 1.124
RMS-deviation in bond distances: 0.024

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

$$\begin{bmatrix} 0.987188 & 0.000293 & -0.008811 \\ 0.000293 & 0.991730 & -0.001286 \\ -0.008811 & -0.001286 & 0.992544 \end{bmatrix}$$

Proposed new scale matrix

$$\begin{bmatrix} 0.024730 & -0.000007 & -0.000111 \\ -0.000015 & 0.054066 & 0.000070 \\ 0.000399 & 0.000058 & 0.044746 \end{bmatrix}$$

With corresponding cell

$A =$ 40.437 $B =$ 18.496 $C =$ 22.348
$\alpha =$ 90.148 $\beta =$ 90.254 $\gamma =$ 89.965

The CRYST1 cell dimensions

$A =$ 40.960 $B =$ 18.650 $C =$ 22.520
$\alpha =$ 90.000 $\beta =$ 89.230 $\gamma =$ 90.000

Variance: 144.753
(Under-)estimated Z-score: 8.867

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	THR (1-) A	-	CA	CB	OG1	103.06	-4.4
1	THR (1-) A	-	CG2	CB	OG1	117.41	4.1
12	ASN (12-) A	-	ND2	CG	OD1	127.61	5.0
14	ASN (14-) A	-	ND2	CG	OD1	128.63	6.0
39	THR (39-) A	-	CA	CB	OG1	103.59	-4.0
45	ALA (45-) A	-	N	CA	CB	103.98	-4.3

2.6.5 Note: Normal bond angle variability

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

RMS Z-score for bond angles: 1.285
RMS-deviation in bond angles: 2.346

2.6.6 Note: Residue hand error(s)

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

2.6.7 Note: Chirality OK

All protein atoms have proper chirality.

2.6.8 Note: Improper dihedral angle distribution OK

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

Improper dihedral RMS Z-score : 1.110

2.6.9 Note: Tau angles OK

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

2.6.10 Note: Normal tau angle deviations

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

Tau angle RMS Z-score : 1.287

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
19	PRO (19-) A	-	0.13	LOW
36	PRO (36-) A	-	0.04	LOW

2.7.2 Note: PRO puckering phases OK

Puckering phases for all PRO residues are normal

2.7.3 Note: Torsion angles OK

All individual residues have normal overall torsion angle scores.

2.7.4 Note: Backbone torsion angles OK

All individual residues have normal backbone torsion angles.

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 : -0.307

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 : 3.769

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

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

χ-1/χ-2 correlation Z-score : -0.900

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: Rotamers checked OK

None of the residues that have a normal backbone environment have abnormal rotamers.

2.7.10 Warning: Unusual backbone conformations

For the residues listed in the table below, the backbone formed by itself and two neighbouring residues on either side is in a conformation that is not seen very often in the database of solved protein structures. The number given in the table is the number of similar backbone conformations in the database with the same amino acid in the centre.

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

A residue mentioned in the table can be part of a strange loop, or there might be something wrong with it or its directly surrounding residues. There are a few of these in every protein, but in any case it is worth looking at!

	Residue		# hits
19	PRO (19-) A	-	0
5	PRO (5-) A	-	1
30	THR (30-) A	-	1
44	TYR (44-) A	-	1
4	CYS (4-) A	-	2
32	CYS (32-) A	-	2
36	PRO (36-) A	-	2
38	ALA (38-) A	-	2

2.7.11 Note: Backbone conformation Z-score OK

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

Backbone conformation Z-score : 0.109

2.8 Bump checks

2.8.1 Note: No Van der Waals overlaps

All interatomic distances (including symmetry transformations) have been verified. No unusual contacts were found. No pair of atoms has an unusual short contact distance.

2.9 Packing, accessibility and threading

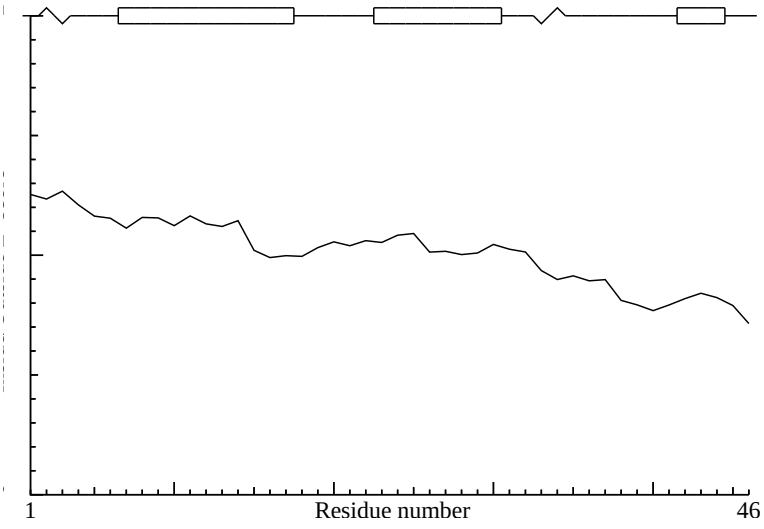
2.9.1 Note: Inside/Outside residue distribution normal

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

inside/outside RMS Z-score : 1.012

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 Note: Packing environment OK

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

2.9.4 Note: No series of residues with bad packing environment

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

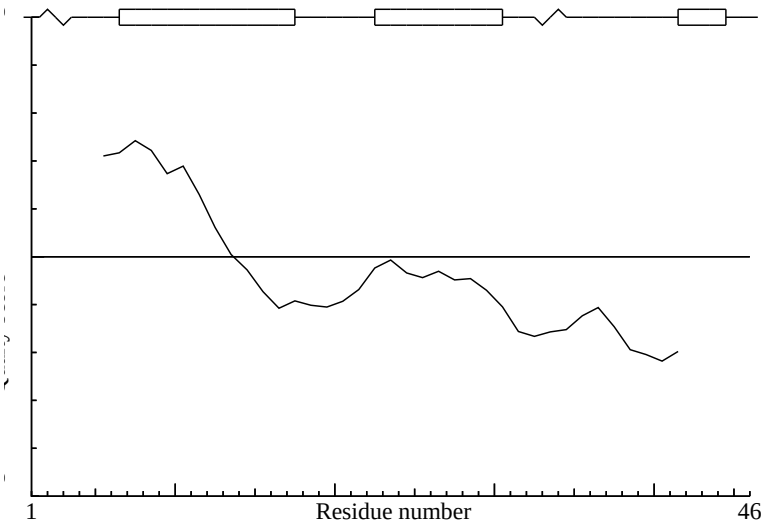
2.9.5 Note: Structural average packing environment OK

The structural average quality control value is within normal ranges.

Average for range 1 - 46 : -0.435

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.

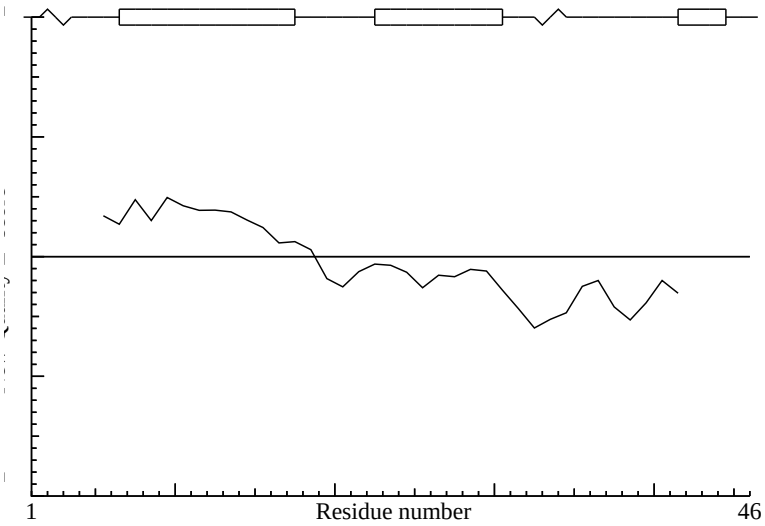
2.9.9 Note: Structural average packing Z-score OK

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

All contacts : Average = -0.057 Z-score = -0.34
BB-BB contacts : Average = 0.038 Z-score = 0.19
BB-SC contacts : Average = -0.110 Z-score = -0.91
SC-BB contacts : Average = -0.036 Z-score = -0.17
SC-SC contacts : Average = -0.135 Z-score = -0.62

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: HIS, ASN, GLN side chains OK

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

2.10.2 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
6	SER (6-) A	-	N
8	VAL (8-) A	-	N
45	ALA (45-) A	-	N

2.10.3 Note: Buried hydrogen bond acceptors OK

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

2.10.4 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) 46 (46) A	Protein		pdb1crn.ent
2	47 (46) 47 (46) A	N O2 <-	46	pdb1crn.ent

2.10.5 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.

2.10.6 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 :	0.163	
2nd generation packing quality :	-0.344	
Ramachandran plot appearance :	-0.307	
χ-1/χ-2 rotamer normality :	-0.900	
Backbone conformation :	0.109	

RMS Z-scores, should be close to 1.0:

Bond lengths :	1.124	
Bond angles :	1.285	
Omega angle restraints :	0.685	(tight)
Side chain planarity :	0.842	
Improper dihedral distribution :	1.110	
Inside/Outside distribution :	1.012	

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 : 1.50

Structure Z-scores, positive is better than average:

1st generation packing quality :	0.5	
2nd generation packing quality :	-0.8	
Ramachandran plot appearance :	-0.8	
χ-1/χ-2 rotamer normality :	-1.4	
Backbone conformation :	-0.4	

RMS Z-scores, should be close to 1.0:

Bond lengths :	1.124	
Bond angles :	1.285	
Omega angle restraints :	0.685	(tight)
Side chain planarity :	0.842	
Improper dihedral distribution :	1.110	
Inside/Outside distribution :	1.012	

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).