wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 02:32 PM BST

PDB ID	:	1EQK
Title	:	SOLUTION STRUCTURE OF ORYZACYSTATIN-I, A CYSTEINE PROTEINASE INHIBITOR OF THE RICE, ORYZA SATIVA L. JAPONICA
Authors	:	Nagata, K.; Kudo, N.; Abe, K.; Arai, S.; Tanokura, M.
Deposited on	:	2000-04-05

This is a wwPDB NMR Structure Validation Summary Report for a publicly released PDB entry.
We welcome your comments at validation@mail.wwpdb.org
A user guide is available at
http://wwpdb.org/validation/2016/NMRValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (i)) were used in the production of this report:

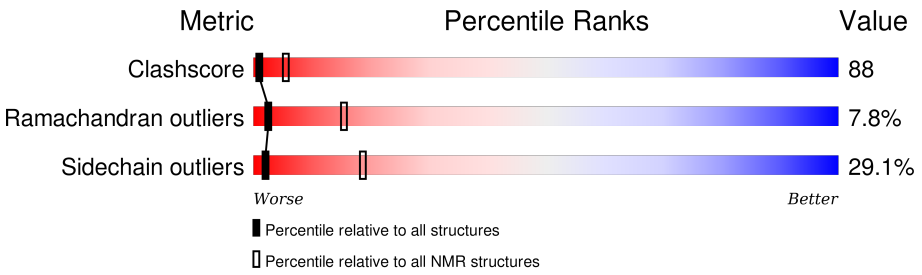
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027457

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
SOLUTION NMR

The overall completeness of chemical shifts assignment was not calculated.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	NMR archive (#Entries)
Clashscore	114402	11133
Ramachandran outliers	111179	9975
Sidechain outliers	111093	9958

The table below summarises the geometric issues observed across the polymeric chains and their fit to the experimental data. The red, orange, yellow and green segments indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. A cyan segment indicates the fraction of residues that are not part of the well-defined cores, and a grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5%

Mol	Chain	Length	Quality of chain
1	A	102	11% 56% 16% 18%

2 Ensemble composition and analysis (i)

This entry contains 21 models. Model 15 is the overall representative, medoid model (most similar to other models). The authors have identified model 21 as representative, based on the following criterion: *minimized average structure*.

The following residues are included in the computation of the global validation metrics.

Well-defined (core) protein residues			
Well-defined core	**Residue range (total)**	**Backbone RMSD (Å)**	**Medoid model**
1	A:13-A:96 (84)	0.27	15

Ill-defined regions of proteins are excluded from the global statistics.

Ligands and non-protein polymers are included in the analysis.

The models can be grouped into 1 clusters and 2 single-model clusters were found.

Cluster number	Models
1	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 19, 21
Single-model clusters	18; 20

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1601 atoms, of which 797 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called ORYZACYSTATIN-I.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	102	1601	516	797	130	156	2	0

4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA and DNA chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. Stretches of 2 or more consecutive residues without any outliers are shown as green connectors. Residues which are classified as ill-defined in the NMR ensemble, are shown in cyan with an underline colour-coded according to the previous scheme. Residues which were present in the experimental sample, but not modelled in the final structure are shown in grey.

- Molecule 1: ORYZACYSTATIN-I

Chain A: 11% 56% 16% 18%

M1 S2 S3 D4 G5 G6 P7 V8 L9 G10 G11 V12 E13 P14 V15 G16 N17 E18 N19 D20 L21 H22 L23 L26 A27 R28 F29 A30 V31 T32 E33 H34 N35 K36 K37 A38 N39 S40 L41 L42 E43 F44 E45 K46 L47 V50 K51 Q52 Q53 V54 V55 A56 G57 T58 L59 Y60 Y61 F62

T63 I64 E65 V66 K67 E68 K72 K73 L74 Y75 E76 A77 K78 V79 W80 E81 K82 P83 W84 M85 D86 F87 K88 E89 L90 Q91 E92 F93 K94 P95 V96 D97 A98 S99 A100 N101 A102

4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 15. Colouring as in section 4.1 above.

- Molecule 1: ORYZACYSTATIN-I

Chain A: 17% 44% 22% 18%

M1 S2 S3 D4 G5 G6 P7 V8 L9 G10 G11 V12 E13 P14 V15 G16 N17 E18 N19 D20 L21 H22 L23 L26 A27 R28 F29 A30 V31 T32 E33 H34 N35 S40 L41 L42 E43 F44 E45 K46 L47 V50 K51 Q52 Q53 V54 V55 A56 G57 T58 L59 Y60 Y61 F62 T63 I64 E65

V66 K67 E68 G69 K72 K73 L74 Y75 E76 A77 K78 V79 W80 E81 K82 P83 W84 M85 D86 F87 K88 E89 L90 Q91 E92 F93 K94 P95 V96 D97 A98 S99 A100 N101 A102

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *torsion angle dynamics*.

Of the 100 calculated structures, 21 were deposited, based on the following criterion: *structures with favorable non-bond energy*.

The following table shows the software used for structure solution, optimisation and refinement.

Software name	Classification	Version
DYANA	structure solution	1.4
CNS	refinement	0.9

No chemical shift data was provided. No validations of the models with respect to experimental NMR restraints is performed at this time.

6 Model quality (i)

6.1 Standard geometry (i)

There are no covalent bond-length or bond-angle outliers.

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

There are no planarity outliers.

6.2 Too-close contacts (i)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in each chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes averaged over the ensemble.

Mol	Chain	Non-H	H(model)	H(added)	Clashes
1	A	693	694	692	121±21
All	All	14553	14574	14532	2546

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 88.

5 of 607 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:63:THR:HG23	1:A:74:LEU:HD21	1.08	1.21	7	4
1:A:63:THR:HG22	1:A:74:LEU:HD11	1.08	1.10	2	10
1:A:74:LEU:HD12	1:A:74:LEU:N	1.01	1.69	8	1
1:A:42:LEU:HD22	1:A:66:VAL:HG23	1.00	1.24	19	2
1:A:23:LEU:HD12	1:A:50:VAL:HG11	0.99	1.32	17	13

6.3 Torsion angles (i)

6.3.1 Protein backbone (i)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all NMR entries. The Analysed column shows the number of residues for which the backbone conformation was analysed and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	A	84/102 (82%)	59±3 (70±3%)	18±2 (22±3%)	7±1 (8±2%)	2 15
All	All	1764/2142 (82%)	1238 (70%)	388 (22%)	138 (8%)	2 15

5 of 18 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	A	55	VAL	21
1	A	92	GLU	20
1	A	96	VAL	19
1	A	84	TRP	16
1	A	17	ASN	13

6.3.2 Protein sidechains (i)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all NMR entries. The Analysed column shows the number of residues for which the sidechain conformation was analysed and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	A	75/86 (87%)	53±3 (71±4%)	22±3 (29±4%)	2 19
All	All	1575/1806 (87%)	1117 (71%)	458 (29%)	2 19

5 of 50 unique residues with a non-rotameric sidechain are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	A	59	LEU	21
1	A	26	LEU	20
1	A	90	LEU	19
1	A	23	LEU	19
1	A	64	ILE	18

6.3.3 RNA (i)

There are no RNA molecules in this entry.

6.4 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

6.5 Carbohydrates (i)

There are no carbohydrates in this entry.

6.6 Ligand geometry (i)

There are no ligands in this entry.

6.7 Other polymers (i)

There are no such molecules in this entry.

6.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

7 Chemical shift validation (i)

No chemical shift data were provided