wwPDB NMR Structure Validation Summary Report (i)

May 28, 2020 – 10:49 pm BST

PDB ID	:	2L3P
Title	:	Structure of the prolyl cis isomer of the Crk Protein
Authors	:	Kalodimos, C.G.; Sarkar, P.; Saleh, T.; Tzeng, S.R.; Birge, R.
Deposited on	:	2010-09-21

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
https://www.wwpdb.org/validation/2017/NMRValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

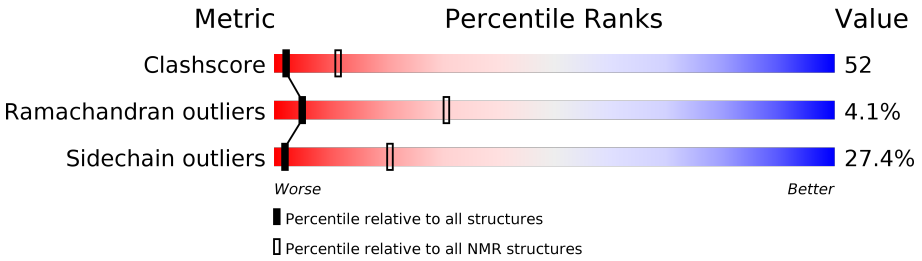
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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	158937	12864
Ramachandran outliers	154571	11451
Sidechain outliers	154315	11428

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	78	17% 42% 12% 29%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 16 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *lowest energy*.

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:238-A:270, A:276-A:297 (55)	0.16	16

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 4 clusters and 2 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1234 atoms, of which 618 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Cis isomer of Crk protein.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	78	1234	392	618	111	111	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: Cis isomer of Crk protein

Chain A: 17% 42% 12% 29%

G220 P221 Y222 A223 Q224 P225 S226 I227 N228 T229 P230 L231 P232 N233 L234 Q235 N236 G237 P238 F239 Y240 A241 R242 V243 I244 Q245 K246 R247 V248 P249 N250 D253 K254 T255 A256 L257 A258 L259 E260 V261 G262 E263 L264 V265 K266 V267 T268 K269 I270 N271 M272 S273 G274 Q275 W276 E277 G278 E279 C280

K283 R284 G285 H286 F287 P288 F289 T290 H291 V292 R293 L294 L295 D296 Q297

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

The representative model is number 16. Colouring as in section 4.1 above.

- Molecule 1: Cis isomer of Crk protein

Chain A: 18% 42% 10% 29%

G220 P221 Y222 A223 Q224 P225 S226 I227 N228 T229 P230 L231 P232 N233 L234 Q235 N236 G237 P238 F239 Y240 A241 R242 V243 I244 Q245 K246 R247 V248 P249 D253 K254 T255 A256 L257 A258 L259 E260 V261 G262 E263 L264 V265 K266 V267 T268 K269 I270 N271 M272 S273 G274 Q275 W276 E277 G278 E279 C280

K283 R284 G285 H286 F287 P288 F289 T290 H291 V292 R293 L294 L295 D296 Q297

5 Refinement protocol and experimental data overview (i)

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

Of the 20 calculated structures, 20 were deposited, based on the following criterion: *structures with the lowest energy*.

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

Software name	Classification	Version
ARIA	refinement	1.2

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

COVALENT-GEOMETRY INFOmissingINFO

5.1 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	449	458	456	47±4
All	All	8980	9160	9120	948

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 52.

5 of 140 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:265:VAL:HG23	1:A:280:CYS:SG	1.02	1.94	2	14
1:A:257:LEU:HD12	1:A:280:CYS:SG	0.92	2.04	20	15
1:A:270:ILE:H	1:A:270:ILE:HD12	0.89	1.27	10	8
1:A:257:LEU:HB3	1:A:259:LEU:HD21	0.82	1.50	10	12
1:A:247:ARG:O	1:A:249:PRO:HD3	0.79	1.78	17	20

5.2 Torsion angles (i)

5.2.1 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	54/78 (69%)	43±2 (79±3%)	9±1 (17±2%)	2±1 (4±1%)	5 31
All	All	1080/1560 (69%)	857 (79%)	179 (17%)	44 (4%)	5 31

All 3 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	249	PRO	20
1	A	290	THR	17
1	A	296	ASP	7

5.2.2 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	47/66 (71%)	34±2 (73±3%)	13±2 (27±3%)	2 21
All	All	940/1320 (71%)	682 (73%)	258 (27%)	2 21

5 of 27 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	246	LYS	20
1	A	269	LYS	20
1	A	259	LEU	20
1	A	291	HIS	18
1	A	270	ILE	17

5.2.3 RNA (i)

There are no RNA molecules in this entry.

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

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

5.4 Carbohydrates (i)

There are no carbohydrates in this entry.

5.5 Ligand geometry (i)

There are no ligands in this entry.

5.6 Other polymers (i)

There are no such molecules in this entry.

5.7 Polymer linkage issues (i)

There are no chain breaks in this entry.

6 Chemical shift validation (i)

No chemical shift data were provided