wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 08:33 PM BST

PDB ID	:	2FCI
Title	:	Structural basis for the requirement of two phosphotyrosines in signaling mediated by Syk tyrosine kinase
Authors	:	Groesch, T.D.; Zhou, F.; Mattila, S.; Geahlen, R.L.; Post, C.B.
Deposited on	:	2005-12-12

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:

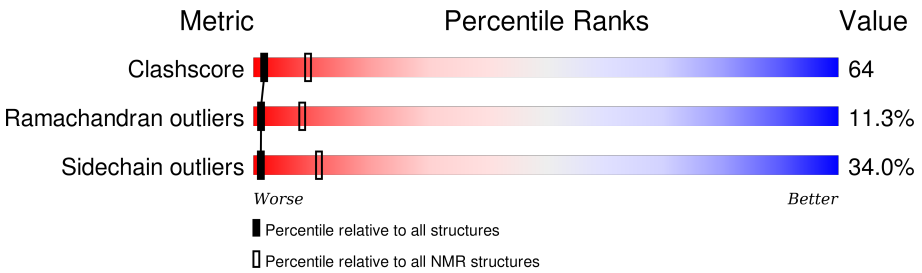
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.7.1 (RC1), CSD as537be (2016)
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	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	B	14	50% 7% 43%
2	A	105	5% 59% 27% • 7%

2 Ensemble composition and analysis (i)

This entry contains 16 models. The atoms present in the NMR models are not consistent. Some calculations may have failed as a result. All residues are included in the validation scores. Model 1 is the overall representative, medoid model (most similar to other models).

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	B:4-B:5, B:7-B:9, B:11-B:13, A:4-A:101 (106)	0.54	1

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 2 clusters and 1 single-model cluster was found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 1915 atoms, of which 935 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Doubly phosphorylated peptide derived from Syk kinase comprising residues 338-350.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
1	B	14	207	68	89	14	34	2	0

- Molecule 2 is a protein called C-termainl SH2 domain from phospholipase C-gamma-1 comprising residues 663-759.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
2	A	105	1708	539	846	159	159	5	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	CLONING ARTIFACT	UNP P08487
A	2	SER	-	CLONING ARTIFACT	UNP P08487
A	3	PRO	-	CLONING ARTIFACT	UNP P08487
A	4	GLY	-	CLONING ARTIFACT	UNP P08487
A	5	ILE	-	CLONING ARTIFACT	UNP P08487
A	103	ASN	-	CLONING ARTIFACT	UNP P08487
A	104	SER	-	CLONING ARTIFACT	UNP P08487
A	105	SER	-	CLONING ARTIFACT	UNP P08487

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: Doubly phosphorylated peptide derived from Syk kinase comprising residues 338-350

Chain B:

• Molecule 2: C-termainl SH2 domain from phospholipase C-gamma-1 comprising residues 663-759

Chain A:

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

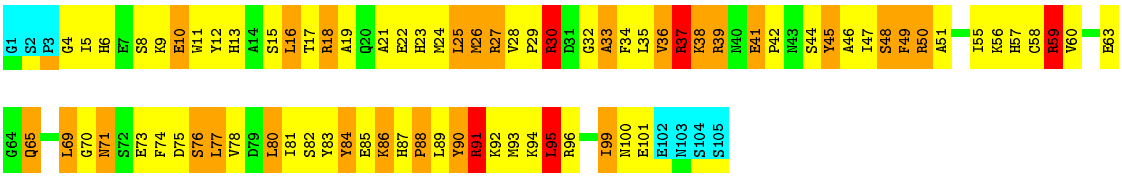
The representative model is number 1. Colouring as in section 4.1 above.

• Molecule 1: Doubly phosphorylated peptide derived from Syk kinase comprising residues 338-350

Chain B:

• Molecule 2: C-termainl SH2 domain from phospholipase C-gamma-1 comprising residues 663-759

Chain A:


G1 S2 P3 G4 I5 H6 E7 S8 K9 E10 W11 Y12 H13 A14 S15 L16 T17 R18 A19 Q20 A21 E22 H23 M24 L25 M26 R27 V28 P29 R30 D31 G32 A33 F34 L35 V36 R37 K38 R39 N40 E41 P42 N43 S44 Y45 A46 I47 S48 F49 R50 A51 I55 K56 H57 C58 H59 V60 E63
G64 Q65 L69 G70 N71 S72 E73 F74 D75 S76 L77 V78 D79 L80 I81 S82 Y83 Y84 E85 K86 H87 P88 L89 Y90 R91 K92 M93 K94 L95 R96 I99 N100 E101 E102 N103 S104 S105

5 Refinement protocol and experimental data overview (i)

Of the 50 calculated structures, 16 were deposited, based on the following criterion: *Lack of NOE violations & correct main chain geometry.*

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

Software name	Classification	Version
X-PLOR	structure solution	NIH
X-PLOR	refinement	NIH

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)

Bond lengths and bond angles in the following residue types are not validated in this section: GMA, PTR, ACE

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the (average) root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#Z>5	RMSZ	#Z>5
1	B	1.23±0.03	0±0/57 (0.0±0.0%)	1.12±0.07	0±0/75 (0.0±0.0%)
2	A	1.16±0.01	0±0/835 (0.0±0.0%)	1.26±0.01	1±0/1122 (0.1±0.0%)
All	All	1.16	0/14272 (0.0%)	1.25	14/19152 (0.1%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	Chirality	Planarity
2	A	0.0±0.0	8.9±0.2
All	All	0	143

There are no bond-length outliers.

All unique angle outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
2	A	44	SER	N-CA-CB	-6.41	100.88	110.50	11	13
2	A	68	MET	N-CA-CB	-5.14	101.35	110.60	6	1

There are no chirality outliers.

5 of 9 unique planar outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Group	Models (Total)
2	A	30	ARG	Sidechain	16
2	A	91	ARG	Sidechain	16
2	A	27	ARG	Sidechain	16

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group	Models (Total)
2	A	18	ARG	Sidechain	16
2	A	96	ARG	Sidechain	16

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	B	58	49	49	10±3
2	A	815	807	807	110±11
All	All	13968	13696	13696	1772

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 64.

5 of 909 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
2:A:28:VAL:HG11	2:A:99:ILE:HG21	1.04	1.29	15	12
1:B:11:ALA:HB2	2:A:69:LEU:HD22	0.97	1.35	4	4
2:A:69:LEU:HD22	2:A:80:LEU:HD22	0.96	1.37	2	1
2:A:21:ALA:HB1	2:A:35:LEU:HD21	0.96	1.38	8	12
1:B:11:ALA:HB2	2:A:69:LEU:CD2	0.95	1.90	8	3

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	B	8/14 (57%)	3±1 (42±12%)	3±1 (34±13%)	2±1 (24±13%)	0 1
2	A	98/105 (93%)	66±2 (67±2%)	22±3 (22±3%)	10±2 (10±2%)	1 10

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	1696/1904 (89%)	1112 (66%)	393 (23%)	191 (11%)	1 8

5 of 39 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
2	A	65	GLN	16
2	A	5	ILE	15
2	A	71	ASN	15
2	A	88	PRO	15
1	B	9	PRO	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	B	7/9 (78%)	5±1 (77±11%)	2±1 (23±11%)	3 29
2	A	87/93 (94%)	57±3 (65±4%)	30±3 (35±4%)	1 10
All	All	1504/1632 (92%)	992 (66%)	512 (34%)	1 11

5 of 81 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)
2	A	16	LEU	16
2	A	27	ARG	16
2	A	80	LEU	16
2	A	56	LYS	16
2	A	37	ARG	15

6.3.3 RNA (i)

There are no RNA molecules in this entry.

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

3 non-standard protein/DNA/RNA residues are modelled in this entry.

In the following table, the Counts columns list the number of bonds for which Mogul statistics could be retrieved, the number of bonds that are observed in the model and the number of bonds that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length is the number of standard deviations the observed value is removed from the expected value. A bond length with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond lengths.

Mol	Type	Chain	Res	Link	Bond lengths		
					Counts	RMSZ	#Z>2
1	PTR	B	10	1	13,16,17	0.82±0.03	0±0 (0±0%)
1	GMA	B	14	1	6,9,9	0.65±0.04	0±0 (0±0%)
1	PTR	B	6	1	13,16,17	0.82±0.04	0±0 (0±0%)

In the following table, the Counts columns list the number of angles for which Mogul statistics could be retrieved, the number of angles that are observed in the model and the number of angles that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond angle is the number of standard deviations the observed value is removed from the expected value. A bond angle with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond angles.

Mol	Type	Chain	Res	Link	Bond angles		
					Counts	RMSZ	#Z>2
1	PTR	B	10	1	19,22,24	1.12±0.04	0±0 (0±0%)
1	GMA	B	14	1	7,11,11	1.61±0.06	0±0 (0±0%)
1	PTR	B	6	1	19,22,24	1.05±0.06	0±0 (0±0%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
1	PTR	B	10	1	-	0±0,9,11,13	0±0,1,1,1
1	GMA	B	14	1	-	0±0,6,9,9	0±0,0,0,0
1	PTR	B	6	1	-	0±0,9,11,13	0±0,1,1,1

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

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