wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 02:14 AM GMT

PDB ID	:	2G67
Title	:	E. Coli Pyruvate Dehydrogenase E1 Component (Apoenzyme)
Authors	:	Furey, W.; Chandrasekhar, K.; Arjunan, P.
Deposited on	:	2006-02-24
Resolution	:	2.32 Å(reported)

This is a wwPDB X-ray 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/XrayValidationReportHelp
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:

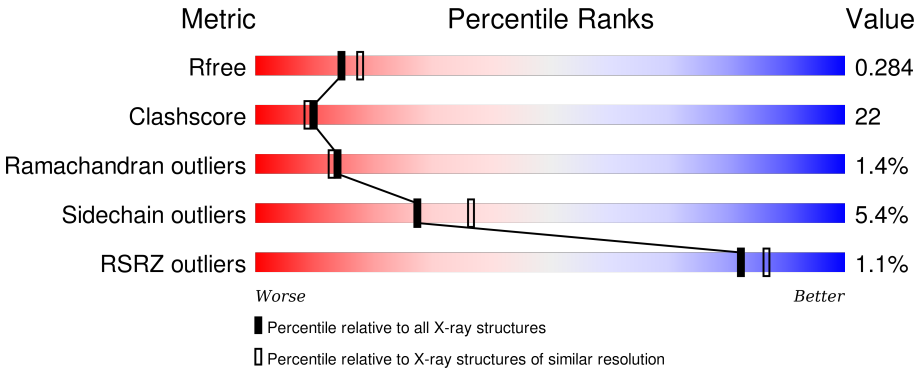
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.32 Å.

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)	Similar resolution (#Entries, resolution range(Å))
R_{free}	91344	4425 (2.34-2.30)
Clashscore	102246	5057 (2.34-2.30)
Ramachandran outliers	100387	5008 (2.34-2.30)
Sidechain outliers	100360	5007 (2.34-2.30)
RSRZ outliers	91569	4432 (2.34-2.30)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. 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% The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	886	% 52% 34% · 10%
1	B	886	% 56% 31% · 10%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 13150 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Pyruvate dehydrogenase E1 component.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	801	Total	C	N	O	S	0	1	0
			6351	4024	1096	1205	26			
1	B	801	Total	C	N	O	S	0	1	0
			6351	4024	1096	1205	26			

- Molecule 2 is water.

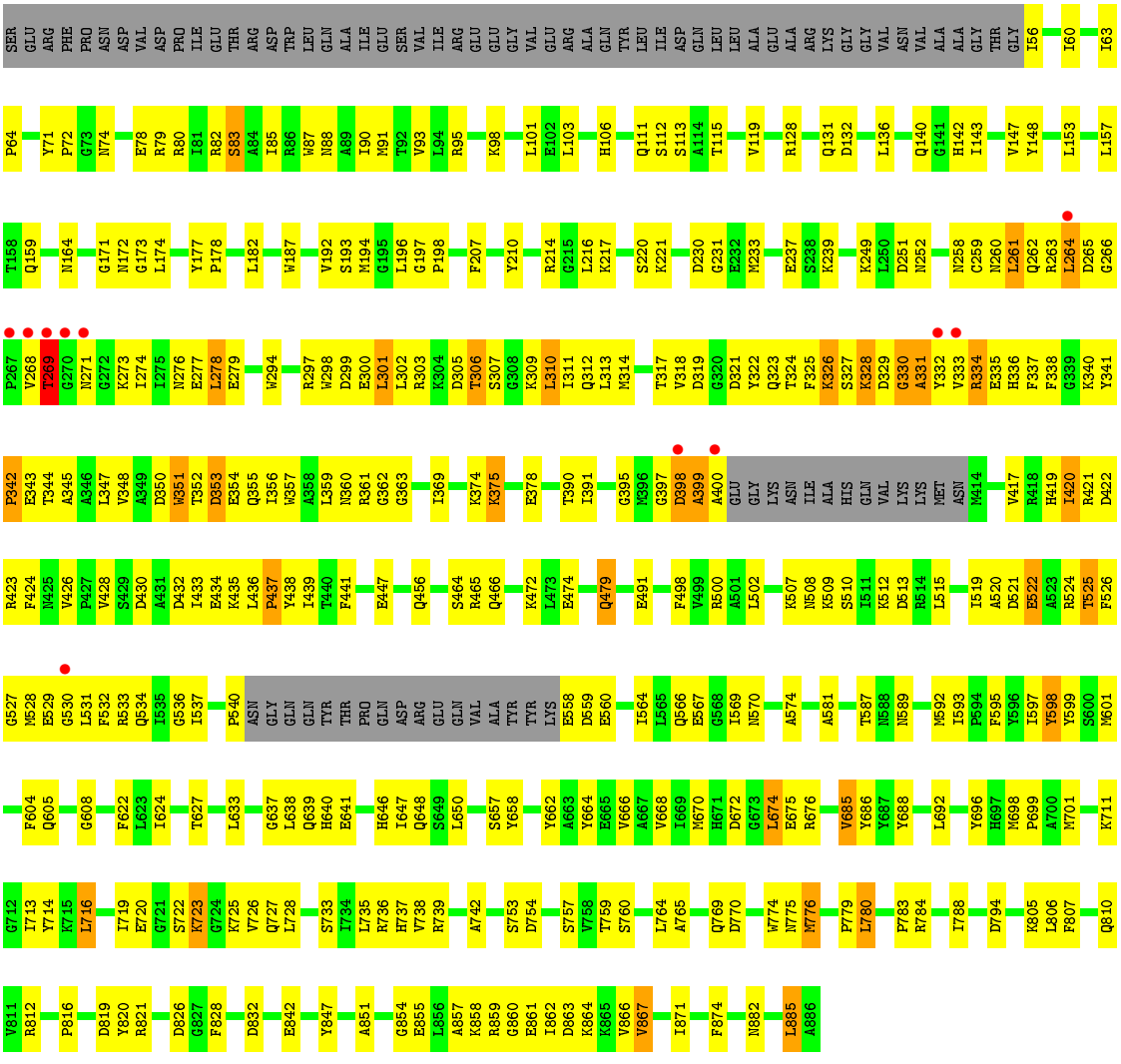
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	241	Total	O	0	0
			241	241		
2	B	207	Total	O	0	0
			207	207		

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-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. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

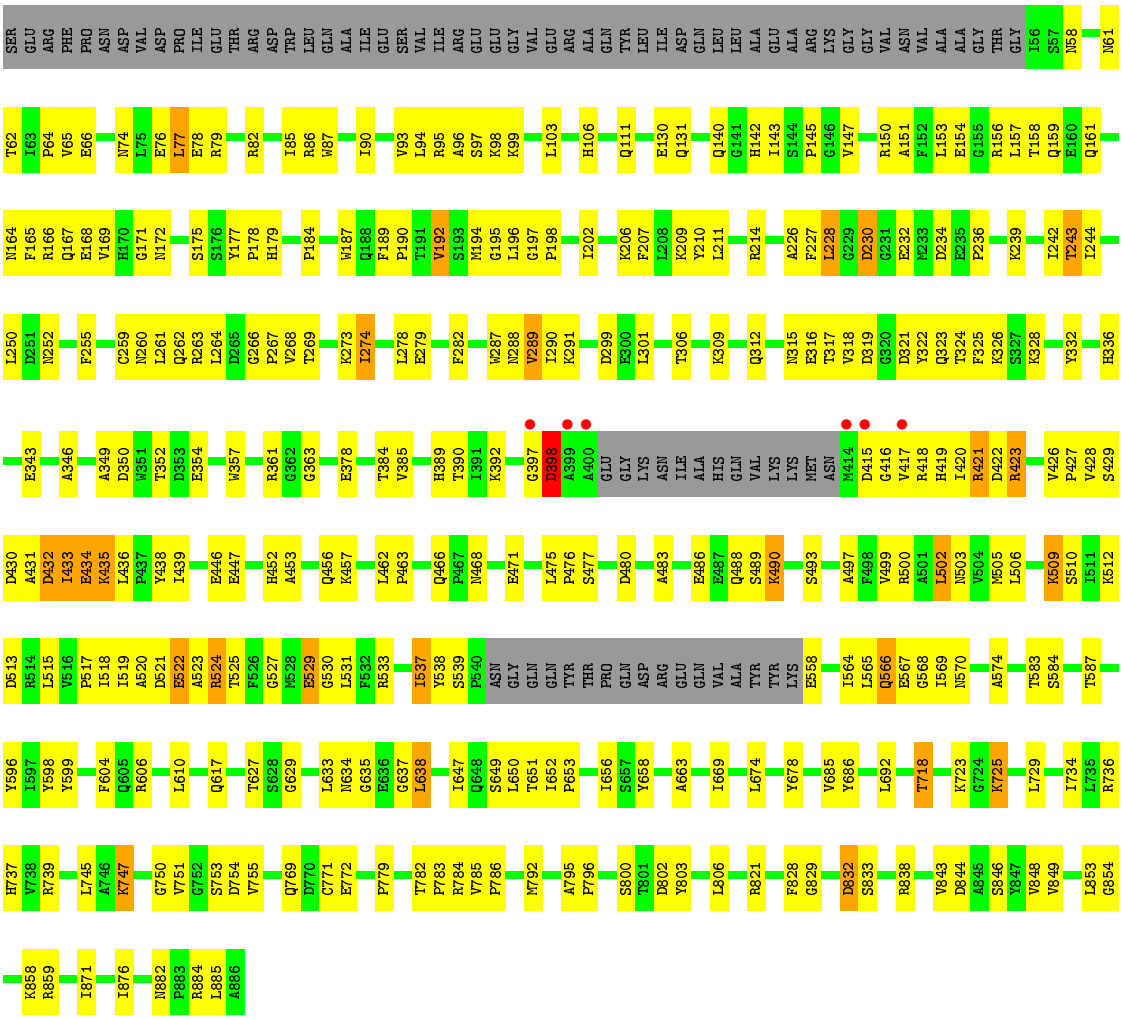
- Molecule 1: Pyruvate dehydrogenase E1 component

Chain A:



- Molecule 1: Pyruvate dehydrogenase E1 component

Chain B:


SER GLU ARG PHE PRO ASN ASP VAL ASP PRO ILE GLU THR ARG ASP TRP LEU GLN ALA ILE GLU SER VAL ILE ARG GLU GLU GLY VAL GLU ARG ALA GLN TYR LEU ILE ASP GLN LEU LEU ALA GLU ALA ARG LYS GLY GLY VAL ASN VAL ALA ALA GLY THR GLY I56 S57 N58 M61
T62 I63 P64 V65 E66 N74 L75 E76 L77 E78 R79 R82 I85 R86 W87 I90 V93 L94 R95 A96 S97 K98 K99 L103 H106 Q111 E130 Q131 Q140 G141 H142 I143 S144 P145 G146 V147 R150 A151 F152 L153 E154 G155 R156 L157 T158 Q159 E160 Q161
N164 F165 R166 Q167 E168 V169 H170 G171 N172 S175 S176 Y177 P178 H179 P184 W187 Q188 F189 P190 T191 V192 S193 M194 G195 L196 G197 P198 I202 K206 F207 L208 K209 Y210 L211 R214 A226 F227 L228 G229 D230 G231 E232 M233 D234 E235 P236 K239 I242 T243 I244
L250 D251 M252 F255 C259 N260 L261 Q262 R263 L264 D265 G266 P267 V268 T269 K273 I274 L278 E279 F282 W287 N288 Y289 I290 K291 D299 E300 L301 T306 K309 Q312 N315 E316 T317 Y318 D319 G320 D321 Y322 Q323 T324 F325 K326 S327 K328 Y332 H336
E343 A346 A349 D350 W351 T352 D353 E354 W357 R361 G362 G363 E378 T384 V385 H389 T390 I391 K392 G397 D398 A399 A400 GLU GLY LYS ASN ILE ALA HIS GLN VAL LYS LYS MET ASN M414 D415 G416 V417 R418 H419 I420 R421 D422 R423 V426 P427 V428 S429
D430 A431 D432 I433 E434 K435 L436 P437 Y438 I439 E446 E447 H452 A453 Q456 K457 L462 P463 Q466 P467 N468 E471 L475 P476 S477 D480 A483 E486 E487 Q488 S489 K490 S493 A497 F498 V499 R500 A501 L502 N503 V504 M505 L506 K509 S510 I511 K512
D513 R514 L515 V516 P517 I518 I519 A520 D521 D522 A523 R524 T525 F526 G527 M528 E529 G530 L531 F532 R533 I537 Y538 S539 P540 ASN GLY GLN GLN TYR THR PRO GLN ASP ARG GLU GLN VAL ALA TYR TYR LYS E558 I564 L565 Q566 E567 G568 I569 N570 A574 T583 S584 T587
Y596 I597 Y598 Y599 F604 Q605 R606 L610 Q617 T627 S628 G629 L633 N634 G635 E636 G637 L638 I647 Q648 S649 L650 T651 I652 P653 I656 S657 Y658 A663 I669 L674 Y678 V685 Y686 L692 T718 K723 G724 K725 L729 I734 L735 R736
H737 V738 R739 L745 A746 K747 G750 V751 G752 S753 D754 V755 Q769 D770 C771 E772 F779 T782 F783 R784 V785 P786 N792 A795 P796 S800 T801 D802 Y803 L806 R821 F828 G829 D832 S833 R838 V843 D844 A845 S846 Y847 V848 V849 L853 G854
K858 R859 I871 I876 N882 P883 R884 L885 A886

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	81.61Å 140.63Å 81.91Å 90.00° 102.87° 90.00°	Depositor
Resolution (Å)	34.62 – 2.32 34.72 – 2.31	Depositor EDS
% Data completeness (in resolution range)	91.7 (34.62-2.32) 90.3 (34.72-2.31)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	7.39 (at 2.31Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.212 , 0.281 0.213 , 0.284	Depositor DCC
R_{free} test set	3615 reflections (5.07%)	DCC
Wilson B-factor (Å^2)	14.2	Xtriage
Anisotropy	0.170	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 21.0	EDS
Estimated twinning fraction	0.057 for l,-k,h	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.42$, $< L^2 > = 0.25$	Xtriage
Outliers	0 of 71362 reflections	Xtriage
F_o,F_c correlation	0.89	EDS
Total number of atoms	13150	wwPDB-VP
Average B, all atoms (Å^2)	15.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 10.14% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

5 Model quality

5.1 Standard geometry

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 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	A	0.52	0/6495	0.72	1/8781 (0.0%)
1	B	0.50	0/6495	0.72	2/8781 (0.0%)
All	All	0.51	0/12990	0.72	3/17562 (0.0%)

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 outliers	#Planarity outliers
1	A	0	1
1	B	0	3
All	All	0	4

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	171	GLY	N-CA-C	5.40	126.60	113.10
1	A	171	GLY	N-CA-C	5.37	126.52	113.10
1	B	637	GLY	N-CA-C	5.16	126.00	113.10

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	598	TYR	Sidechain
1	B	598	TYR	Sidechain
1	B	678	TYR	Sidechain
1	B	803	TYR	Sidechain

5.2 Too-close contacts

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the 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 within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	6351	0	6185	315	0
1	B	6351	0	6185	259	0
2	A	241	0	0	22	0
2	B	207	0	0	10	0
All	All	13150	0	12370	553	0

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 22.

The worst 5 of 553 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:502:LEU:HD21	1:B:531:LEU:HD11	1.19	1.14
1:B:490:LYS:HE2	1:B:500:ARG:HH22	1.15	1.07
1:B:638:LEU:HD21	1:B:828:PHE:HB3	1.41	1.02
1:A:177:TYR:CG	1:A:192:VAL:HG11	1.96	0.99
1:A:855:GLU:O	1:A:859:ARG:HG3	1.65	0.97

There are no symmetry-related clashes.

5.3 Torsion angles

5.3.1 Protein backbone

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

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	796/886 (90%)	712 (89%)	68 (8%)	16 (2%)	9 7
1	B	796/886 (90%)	722 (91%)	67 (8%)	7 (1%)	21 23

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	1592/1772 (90%)	1434 (90%)	135 (8%)	23 (1%)	14 13

5 of 23 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	261	LEU
1	A	326	LYS
1	A	330	GLY
1	A	351	TRP
1	A	398	ASP

5.3.2 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 X-ray entries followed by that with respect to entries of similar resolution.

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	666/735 (91%)	637 (96%)	29 (4%)	35 47
1	B	666/735 (91%)	623 (94%)	43 (6%)	21 27
All	All	1332/1470 (91%)	1260 (95%)	72 (5%)	27 36

5 of 72 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	192	VAL
1	B	234	ASP
1	B	638	LEU
1	B	196	LEU
1	B	211	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 13 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	737	HIS
1	B	106	HIS
1	B	377	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	466	GLN
1	B	288	ASN

5.3.3 RNA

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains

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

5.5 Carbohydrates

There are no carbohydrates in this entry.

5.6 Ligand geometry

There are no ligands in this entry.

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

There are no chain breaks in this entry.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95^{th} percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å^2)	Q<0.9
1	A	801/886 (90%)	-0.16	11 (1%) 78 83	2, 13, 35, 44	0
1	B	801/886 (90%)	-0.23	6 (0%) 89 92	3, 14, 31, 45	0
All	All	1602/1772 (90%)	-0.19	17 (1%) 82 87	2, 13, 34, 45	0

The worst 5 of 17 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	267	PRO	4.9
1	B	400	ALA	4.4
1	A	268	VAL	4.2
1	B	399	ALA	3.3
1	A	269	THR	3.3

6.2 Non-standard residues in protein, DNA, RNA chains

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

6.3 Carbohydrates

There are no carbohydrates in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.