# wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 02:01 PM BST

| | | |
|---|---|---|
| PDB ID | : | 1BXD |
| Title | : | NMR STRUCTURE OF THE HISTIDINE KINASE DOMAIN OF THE E. COLI OSMOSENSOR ENVZ |
| Authors | : | Tanaka, T.; Saha, S.K.; Tomomori, C.; Ishima, R.; Liu, D.; Tong, K.I.; Park, H.; Dutta, R.; Qin, L.; Swindells, M.B.; Yamazaki, T.; Ono, A.M.; Kainosho, M.; Inouye, M.; Ikura, M. |
| Deposited on | : | 1998-10-02 |

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:

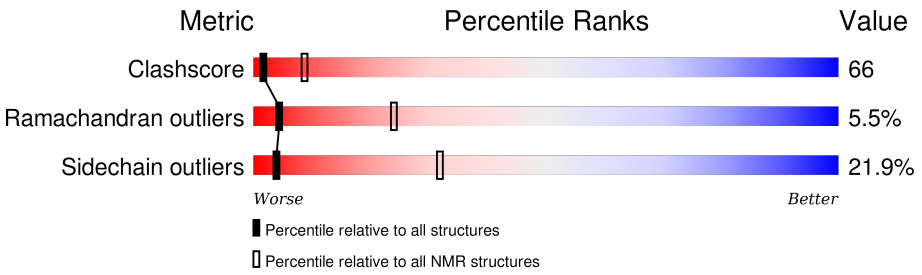
| | | |
|---|---|---|
| Cyrange | : | Kirchner and Güntert (2011) |
| NmrClust | : | Kelley et al. (1996) |
| MolProbity | : | 4.02b-467 |
| Mogul | : | 1.7.1 (RC1), CSD as537be (2016) |
| Percentile statistics | : | 20151230.v01 (using entries in the PDB archive December 30th 2015) |
| RCI | : | v_1n_11_5_13_A (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 | A | 161 | 12% 52% 9% • 25% |

# 2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 9 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 | A:295-A:386, A:408-A:424, A:429-A:439 (120) | 0.70 | 9 |

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 3 clusters and 4 single-model clusters were found.

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

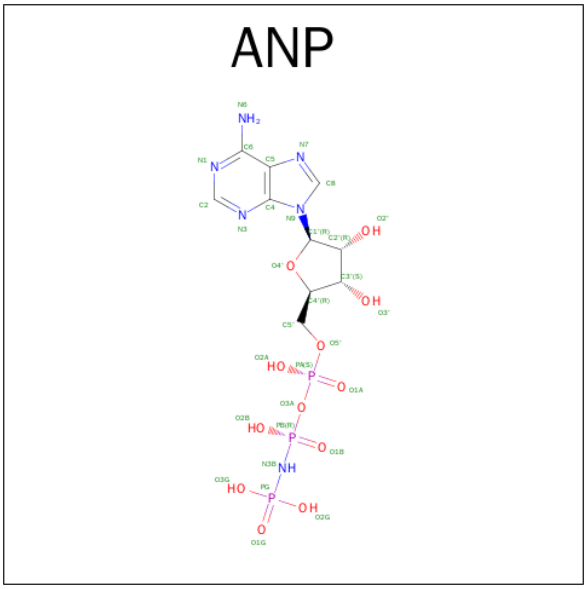
# 3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 2481 atoms, of which 1228 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called PROTEIN (OSMOLARITY SENSOR PROTEIN (ENVZ)).

| Mol | Chain | Residues | Atoms | | | | | | Trace |
|---|---|---|---|---|---|---|---|---|---|
| | | | Total | C | H | N | O | S | |
| 1 | A | 161 | 2436 | 761 | 1214 | 221 | 234 | 6 | 0 |

- Molecule 2 is PHOSPHOAMINOPHOSPHONIC ACID-ADENYLATE ESTER (three-letter code: ANP) (formula: $C_{10}H_{17}N_6O_{12}P_3$).




| Mol | Chain | Residues | Atoms | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | Total | C | H | N | O | P |
| 2 | A | 1 | 45 | 10 | 14 | 6 | 12 | 3 |

# 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: PROTEIN (OSMOLARITY SENSOR PROTEIN (ENVZ))

Chain A: 12% 52% 9% • 25%

T290 G291 Q292 E293 M294 P295 M296 E297 M298 A299 D300 L301 N302 A303 V304 L305 G306 E307 V308 I309 S313 G314 Y315 E316 R317 E318 I319 E320 T321 A322 L323 Y324 P325 G326 S327 I328 E329 V330 K331 M332 H333 F334 L335 S336 I337 K338 R339 A340 V341 A342 N343 N344 V345 V346 N347 A348 A349 R350 Y351

G352 N353 G354 W355 I356 K357 V358 S359 S360 G361 T362 E363 P364 N365 R366 A367 W368 F369 Q370 V371 E372 D373 D374 I378 A379 Q382 R383 K384 H385 L386 F387 Q388 P389 F390 V391 R392 G393 D394 S395 A396 R397 T398 I399 S400 G401 T402 G403 L404 G405 L406 A407 I408 V409 Q410 R411 I412 V413 D414

N415 H416 N417 G418 M419 L420 E421 L422 G423 T424 S425 E426 R427 G428 G429 L430 S431 I432 R433 A434 W435 L436 P437 V438 P439 V440 T441 R442 A443 Q444 G445 T446 T447 K448 E449 G450

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

The representative model is number 9. Colouring as in section 4.1 above.

- Molecule 1: PROTEIN (OSMOLARITY SENSOR PROTEIN (ENVZ))

Chain A: 19% 43% 11% • 25%

T290 G291 Q292 E293 M294 P295 M296 E297 M298 A299 D300 L301 N302 A303 V304 L305 V308 I309 S313 G314 Y315 E316 R317 E318 I319 L323 Y324 P325 G326 S327 I328 E329 V330 K331 M332 H333 F334 L335 S336 I337 K338 R339 A340 V341 A342 N343 N344 V345 V346 N347 A348 A349 R350 Y351 G352 W355

I356 K357 V358 S359 S360 G361 T362 E363 P364 N365 R366 A367 W368 F369 Q370 D374 G377 I378 A379 Q382 R383 K384 H385 L386 F387 Q388 P389 F390 V391 R392 G393 D394 S395 A396 R397 T398 I399 S400 G401 T402 G403 L404 G405 L406 A407 I408 V409 Q410 R411 I412 V413 D414 N415 H416 N417 G418 M419

L420 E421 L422 G423 T424 S425 E426 R427 G428 G429 L430 S431 I432 R433 A434 W435 L436 P437 V438 P439 V440 T441 R442 A443 Q444 G445 T446 T447 K448 E449 G450

# 5 Refinement protocol and experimental data overview (i)

Of the ? calculated structures, 20 were deposited, based on the following criterion: *?*.

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

| Software name | Classification | Version |
|---|---|---|
| X-PLOR | refinement | 3.8.5.1 |

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: ANP

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

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 |
|---|---|---|---|
| 1 | A | 0.0±0.0 | 6.7±0.6 |
| All | All | 0 | 134 |

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

5 of 7 unique planar outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

| Mol | Chain | Res | Type | Group | Models (Total) |
|---|---|---|---|---|---|
| 1 | A | 366 | ARG | Sidechain | 20 |
| 1 | A | 433 | ARG | Sidechain | 20 |
| 1 | A | 350 | ARG | Sidechain | 19 |
| 1 | A | 383 | ARG | Sidechain | 19 |
| 1 | A | 411 | ARG | Sidechain | 19 |

## 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 | 924 | 920 | 920 | 125±13 |
| 2 | A | 31 | 14 | 13 | 10±3 |
| All | All | 19100 | 18680 | 18660 | 2509 |

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 66.

5 of 1011 unique clashes are listed below, sorted by their clash magnitude.

| Atom-1 | Atom-2 | Clash(Å) | Distance(Å) | Models | |
|---|---|---|---|---|---|
| | | | | Worst | Total |
| 1:A:309:ILE:HD11 | 1:A:358:VAL:HG11 | 1.07 | 1.15 | 5 | 8 |
| 1:A:340:ALA:HB2 | 1:A:412:ILE:HG21 | 1.03 | 1.31 | 3 | 9 |
| 1:A:386:LEU:HD11 | 2:A:451:ANP:O4' | 1.01 | 1.55 | 1 | 6 |
| 1:A:340:ALA:HB2 | 1:A:412:ILE:HD13 | 1.00 | 1.33 | 11 | 1 |
| 1:A:330:VAL:HG12 | 1:A:438:VAL:HG22 | 0.99 | 1.32 | 19 | 7 |

## 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 | 120/161 (75%) | 89±3 (74±2%) | 24±3 (20±3%) | 7±2 (6±2%) | 4 | 24 |
| All | All | 2400/3220 (75%) | 1783 (74%) | 485 (20%) | 132 (6%) | 4 | 24 |

5 of 30 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 | 327 | SER | 19 |
| 1 | A | 326 | GLY | 15 |
| 1 | A | 317 | ARG | 12 |
| 1 | A | 366 | ARG | 12 |
| 1 | A | 354 | GLY | 7 |

### 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 | 97/127 (76%) | 76±3 (78±3%) | 21±3 (22±3%) | 4 32 |
| All | All | 1940/2540 (76%) | 1516 (78%) | 424 (22%) | 4 32 |

5 of 67 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 | 362 | THR | 20 |
| 1 | A | 296 | MET | 20 |
| 1 | A | 411 | ARG | 17 |
| 1 | A | 331 | LYS | 15 |
| 1 | A | 298 | MET | 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)

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)

1 ligand is 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 |
| 2 | ANP | A | 451 | - | 29,33,33 | 1.40±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 |
| 2 | ANP | A | 451 | - | 26,52,52 | 1.30±0.03 | 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 |
|---|---|---|---|---|---|---|---|
| 2 | ANP | A | 451 | - | - | 0±0,13,38,38 | 0±0,3,3,3 |

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

All unique torsion outliers are listed below.

| Mol | Chain | Res | Type | Atoms | Models (Total) |
|---|---|---|---|---|---|
| 2 | A | 451 | ANP | O1G-PG-N3B-PB | 1 |

There are no ring outliers.

## 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