Protein Knowledge Report — P01116

1. Target summary

- Input type: uniprot_accession
- Label: P01116
- Length: 189 aa
- UniProt accession: P01116
- Submitted: 2026-07-15T12:56:53Z
- Sequence:

MTEYKLVVVGAGGVGKSALTIQLIQNHFVDEYDPTIEDSYRKQVVIDGETCLLDILDTAGQEEYSAMRDQYMRTGEGFLCVFAINNTKSFEDIHHYREQIKR
VKDSEDVPMVLVGNKCDLPSRTVDTKQAQDLARSYGIPFIETSAKTRQRVEDAFYTLVREIRQYRLKKISKEEKTPGCVKIKKCIIM

2. Family classification

- Family: Ras (PF00071) — Ras family
- E-value: 6.80e-61 Bit score: 205.3
- Model span: 1-161 of model (5-164 of query)
- Passes family's curated gathering threshold: yes

Source: HMMER hmmscan vs Pfam-A (local)

3. Known ligand precedent

- Total structures in this family: 1566
- Distinct ligand scaffolds curated below: 10

PDB ID	Ligand(s)	Resolution (Å)	Why selected
9IAY	GDP, WYU	0.95	highest-resolution structure...
9IAW	GDP, A1I1R	1.00	highest-resolution structure...
2CE2	XY2, GDP	1.00	highest-resolution structure...
3X1X	CD, GNP	1.00	highest-resolution structure...
8ONV	VU6, GDP	1.01	highest-resolution structure...
6P0Z	GDP, ACE	1.01	highest-resolution structure...
8AZZ	GDP, OFU	1.02	highest-resolution structure...
2EVW	XY2, CAG	1.05	highest-resolution structure...

scanned the 100 highest-resolution of 1566 total structures in this family (RCSB result capped for query cost) — the curated examples below are drawn from that subset, not from the full family

Source: RCSB PDB REST API, curated top-N by distinct ligand scaffold + resolution

Interaction fingerprints (Evidence Integration Layer, Phase 1)

Representative complex: PDB 9IAY (GDP, 0.95 Å)

Residue (PDB)	Residue (pipeline)	Interaction	Partner
11	11	hydrogen bond	ALA11 (hydrogen bond) with l...
30	30	hydrogen bond	ASP30 (hydrogen bond) with l...
30	30	hydrogen bond	ASP30 (hydrogen bond) with l...
32	32	hydrogen bond	TYR32 (hydrogen bond) with l...
13	13	hydrogen bond	GLY13 (hydrogen bond) with l...
16	16	hydrogen bond	LYS16 (hydrogen bond) with l...
17	17	hydrogen bond	SER17 (hydrogen bond) with l...
14	14	hydrogen bond	VAL14 (hydrogen bond) with l...
15	15	hydrogen bond	GLY15 (hydrogen bond) with l...
18	18	hydrogen bond	ALA18 (hydrogen bond) with l...

117	117	hydrogen bond	LYS117 (hydrogen bond) with ...
116	116	hydrogen bond	ASN116 (hydrogen bond) with ...
117	117	hydrogen bond	LYS117 (hydrogen bond) with ...
146	146	hydrogen bond	ALA146 (hydrogen bond) with ...
147	147	hydrogen bond	LYS147 (hydrogen bond) with ...
28	28	π-stacking	PHE28 (pi stacking) with lig...
28	28	π-stacking	PHE28 (pi stacking) with lig...
16	16	salt bridge	LYS16 (salt bridge) with lig...
119	119	salt bridge	ASP119 (salt bridge) with li...
34	34	water bridge	PRO34 (water bridge) with li...
34	34	water bridge	PRO34 (water bridge) with li...
32	32	water bridge	TYR32 (water bridge) with li...
60	60	water bridge	GLY60 (water bridge) with li...

Representative complex: PDB 2CE2 (XY2, 1.00 Å)

Residue (PDB)	Residue (pipeline)	Interaction	Partner
33	33	hydrogen bond	ASP33 (hydrogen bond) with l...
25	25	hydrogen bond	GLN25 (hydrogen bond) with l...
40	40	π-stacking	TYR40 (pi stacking) with lig...
40	40	π-stacking	TYR40 (pi stacking) with lig...
57	57	water bridge	ASP57 (water bridge) with li...

Representative complex: PDB 3X1X (CD, 1.00 Å)

no interactions detected in reference structure 3X1X for ligand CD — check structure selection

2/3 representative structure(s) produced interaction data

Source: PLIP (Protein-Ligand Interaction Profiler), run on representative structures from known ligand precedent

4. Conservation

- Family: PF00071
- Seed alignment size: 60 sequences
- Mean pairwise identity (seed alignment): 39.3%
- Note: computed over a sample of 40/60 seed sequences

Source: Pfam seed alignment (via InterPro API), mean pairwise identity across aligned members at matched-region columns

5. Structural analysis

- Structure provider: AlphaFold DB (model created 2025-08-01T00:00:00Z)
- Pocket detector: fpocket (default parameters)
- Mean per-residue confidence: 0.9152
- Conformational ensemble: 5 conformers (ANM, 5 frames)
- Static (single-frame) pocket count: 10
- Residue numbering: pipeline-sequential, with literature (author-deposited PDB 9IAY) numbering shown in parentheses — 161/189 residues cross-walked
- Ensemble-informed ranked pocket clusters (customer-visible top): 3

Rank (cluster id)	Frames present	Persistence	Residues
1 (cluster 0)	5/5	1.0	13(13), 15(15), 16(16), 17(1...
2 (cluster 7)	5/5	1.0	97(97), 107(107), 108(108), ...
3 (cluster 6)	5/5	1.0	73(73), 74(74), 75(75), 76(7...

Pocket residues overlapping known interaction sites (from Interaction fingerprints, section 3):

- Pocket 1 (cluster 0): residues 13(13), 15(15), 16(16), 17(17), 18(18), 28(28), 30(30), 32(32), 33(33), 34(34), 40(40), 60(60), 116(116), 117(117), 119(119), 146(146), 147(147) (17 of 27 pocket residues match a known interaction site)

- Pocket 2 (cluster 7): no overlap with known interaction sites (13 pocket residues checked)
- Pocket 3 (cluster 6): no overlap with known interaction sites (12 pocket residues checked)

Pocket functional context (UniProt features + ClinVar variants):

- Pocket 1 (cluster 0):
  - Residue 13: Binding site — Binding site
  - Residue 13: ClinVar variant [Pathogenic] — G13R (NM_004985.5(KRAS):c.36_37delinsGC (p.Gly13Arg))
  - Residue 15: Binding site — Binding site
  - Residue 16: Binding site — Binding site
  - Residue 17: Binding site — Binding site
  - Residue 18: Binding site — Binding site
  - Residue 18: ClinVar variant [Pathogenic] — A18V (NM_004985.5(KRAS):c.53C>T (p.Ala18Val))
  - Residue 28: Binding site — Binding site
  - Residue 29: Binding site — Binding site
  - Residue 30: Binding site — Binding site
  - Residue 34: ClinVar variant [Likely pathogenic] — P34Q (NM_004985.5(KRAS):c.101C>A (p.Pro34Gln))
  - Residue 35: Glycosylation — (Microbial infection) O-linked (Glc) threonine; by P.sordellii toxin TcsL
  - Residue 61: ClinVar variant [Pathogenic] — Q61R (NM_004985.5(KRAS):c.182_183delinsGT (p.Gln61Arg))
  - Residue 116: Binding site — Binding site
  - Residue 117: Binding site — Binding site
  - Residue 117: ClinVar variant [Likely pathogenic] — K117R (NM_004985.5(KRAS):c.350A>G (p.Lys117Arg))
  - Residue 117: ClinVar variant [Likely pathogenic] — K117T (NM_004985.5(KRAS):c.350A>C (p.Lys117Thr))
  - Residue 119: Binding site — Binding site
  - Residue 146: Binding site — Binding site
  - Residue 147: Binding site — Binding site
  - Residue 147: ClinVar variant [Likely pathogenic] — K147N (NM_004985.5(KRAS):c.441G>T (p.Lys147Asn))
  - Residue 147: ClinVar variant [Likely pathogenic] — K147M (NM_004985.5(KRAS):c.440A>T (p.Lys147Met))
- Pocket 2 (cluster 7):
  - Residue 166: Region — Hypervariable region
  - Residue 169: Region — Hypervariable region
- Pocket 3 (cluster 6):
  - Residue 104: Modified residue — N6-acetyllysine
  - Residue 166: Region — Hypervariable region
  - Residue 167: Region — Hypervariable region
  - Residue 170: Region — Hypervariable region

Source: UniProt REST API (features) + ClinVar via NCBI E-utilities (variants, filtered by clinical significance)

3 single-frame-only cluster(s) absent from the static structure were found but filtered out as likely ANM sampling noise (present in only 1/5 frames) — not shown.

3 pocket cluster(s) in the customer-visible top ranking; static (first-frame-only) fpocket found 10 pocket(s); the ensemble-informed ranker considered 13 cluster(s) across 5 conformers — compare static_pocket_count against pocket_ranking.clusters to see whether the ensemble surfaces anything the static structure alone misses; no binding-site ground truth supplied, so no recovery/overlap metrics are reported; residue numbers above are pipeline-sequential, NOT necessarily literature convention — see numbering_crosswalk (ref: 9IAY) for author-deposited numbering

Source: StructureProvider + PocketDetector abstraction (see api/knowledge_protein/structural.py)

6. Confidence flags

No confidence flags triggered — none of the rules-based checks below found a condition worth discounting. This is not itself a confidence score; it means the specific checks run did not fire, not that every aspect of this report is high-confidence.

Source: rules-based checks in api/knowledge_protein/confidence_flags.py — fixed thresholds on numbers already computed elsewhere in this report, no learned calibration, no combined/weighted score.

7. Similar known proteins

Sequence-based (Pfam family co-membership, top 5 by identity):

PDB ID	Sequence identity to query	Ligand(s)
8EDY	86.8%	GDP
8EER	86.8%	GDP
9G4B	82.3%	A1IIU, GDP
9GGV	82.3%	A1IK9, GDP
8TVK	81.0%	GDP

Structure-based (Foldseek vs. full PDB structural index, Evidence Integration Layer Phase 2):

PDB ID	TM-score	Sequence identity	In sequence-based list…	Description
6V6F	1.000	96.4%	no	Crystal structure of Q…
6ZIZ	1.000	92.2%	no	CRYSTAL STRUCTURE OF N…
2UZI	1.000	93.9%	no	Crystal structure of H…
8ELU	1.000	93.9%	no	HRAS R97G Crystal form…
8EM0	1.000	93.9%	no	HRAS R97V Crystal Form…

5/5 top structural neighbor(s) — 6V6F, 6ZIZ, 2UZI, 8ELU, 8EM0 — are not in the small cached sequence-based sample above (similar_protein.py surveys only the ~60 highest-resolution structures per family, itself capped for query cost, while Foldseek searches the full PDB structural index) — this does not mean they're unrelated, only that they weren't in that capped sample; treat both lists as complementary, partial views, not a contradiction

Source: sequence-based: Pfam family co-membership among cached PDB reference structures (sequence-based); Foldseek vs local PDB structural database (structure-based)

8. Limitations & data provenance

Auto-populated from each section's own status — not hand-maintained.

Section	Status	Detail
Family classification	returned data	
Known ligand precedent	returned data	scanned the 100 highest-resolution of …
Interaction fingerprints	returned data	2/3 representative structure(s) produc…
Functional context (UniProt + ClinVar)	returned data	
Conservation	returned data	computed over a sample of 40/60 seed s…
Structural analysis	returned data	3 pocket cluster(s) in the customer-vi…
Similar known proteins	returned data	
Similar known proteins (structure-base…	returned data	

All sections returned data this run.