Protein Knowledge Report — P24941

1. Target summary

- Input type: uniprot_accession
- Label: P24941
- Length: 298 aa
- UniProt accession: P24941
- Submitted: 2026-07-15T12:56:31Z
- Sequence:

MENFQKVEKIGEGTYGVVYKARNKLTGEVVALKKIRLDTETEGVPSTAIREISLLKELNHPNIVKLLDVIHTENKLYLVFEFLHQDLKKFMDASALTGIPLPLIK
SYLFQLLQGLAFCHSHRVLHRDLKPQNLLINTEGAIKLADFGLARAFGVPVRTYTHEVVTLWYRAPEILLGCKYYSTAVDIWSLGCIFAEMVTRR

2. Family classification

- Family: Pkinase (PF00069) — Protein kinase domain
- E-value: 1.90e-77 Bit score: 260.9
- Model span: 1-262 of model (4-286 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: 4850
- Distinct ligand scaffolds curated below: 9

PDB ID	Ligand(s)	Resolution (Å)	Why selected
6TGU	N92	0.83	highest-resolution structure...
9Q9D	3NG, R9G	0.89	highest-resolution structure...
9Q8C	T9Y	0.90	highest-resolution structure...
9Q9G	R9J	0.94	highest-resolution structure...
9QAB	RA7, 3NG	0.95	highest-resolution structure...
6HMQ	4B0	0.97	highest-resolution structure...
7ATV	RXE	0.98	highest-resolution structure...
9Q8Q	A1I4N	0.98	highest-resolution structure...

scanned the 100 highest-resolution of 4850 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 6Q49 (HGQ, 1.00 Å)

Residue (PDB)	Residue (pipeline)	Interaction	Partner
268	268	hydrogen bond	HIS268 (hydrogen bond) with ...
269	269	hydrogen bond	TYR269 (hydrogen bond) with ...
245	245	π-cation	ARG245 (pi cation) with liga...
226	226	hydrophobic contact	VAL226 (hydrophobic) with li...
270	270	water bridge	ASP270 (water bridge) with l...
245	245	water bridge	ARG245 (water bridge) with l...
264	264	water bridge	SER264 (water bridge) with l...
264	264	water bridge	SER264 (water bridge) with l...

Representative complex: PDB 6Q4H (HGH, 1.00 Å)

Residue (PDB)	Residue (pipeline)	Interaction	Partner
83	83	hydrogen bond	LEU83 (hydrogen bond) with l…
33	33	hydrogen bond	LYS33 (hydrogen bond) with l…
14	14	hydrogen bond	THR14 (hydrogen bond) with l…
14	14	hydrogen bond	THR14 (hydrogen bond) with l…
145	145	hydrogen bond	ASP145 (hydrogen bond) with …
83	83	hydrogen bond	LEU83 (hydrogen bond) with l…
129	129	salt bridge	LYS129 (salt bridge) with li…
18	18	hydrophobic contact	VAL18 (hydrophobic) with lig…
18	18	hydrophobic contact	VAL18 (hydrophobic) with lig…
131	131	hydrophobic contact	GLN131 (hydrophobic) with li…
10	10	hydrophobic contact	ILE10 (hydrophobic) with lig…
132	132	water bridge	ASN132 (water bridge) with l…
132	132	water bridge	ASN132 (water bridge) with l…

Representative complex: PDB 6Q48 (HHQ, 1.03 Å)

Residue (PDB)	Residue (pipeline)	Interaction	Partner
83	83	hydrogen bond	LEU83 (hydrogen bond) with l…

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

4. Conservation

- Family: PF00069
- Seed alignment size: 37 sequences
- Mean pairwise identity (seed alignment): 23.4%

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: 0.8846
- Conformational ensemble: 5 conformers (ANM, 5 frames)
- Static (single-frame) pocket count: 18
- Residue numbering: pipeline-sequential, with literature (author-deposited PDB 6Q4G) numbering shown in parentheses — 281/298 residues cross-walked
- Ensemble-informed ranked pocket clusters (customer-visible top): 3

Rank (cluster id)	Frames present	Persistence	Residues
1 (cluster 6)	5/5	1.0	8(8), 9(9), 10(10), 11(11), …
2 (cluster 0)	5/5	1.0	14(14), 15(15), 126(126), 12…
3 (cluster 8)	5/5	1.0	51(51), 54(54), 55(55), 57(5…

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

- Pocket 1 (cluster 6): residues 10(10), 14(14), 18(18), 33(33), 83(83), 129(129), 131(131), 132(132), 145(145) (9 of 34 pocket residues match a known interaction site)
- Pocket 2 (cluster 0): residues 14(14), 129(129), 131(131) (3 of 23 pocket residues match a known interaction site)
- Pocket 3 (cluster 8): no overlap with known interaction sites (15 pocket residues checked)

Pocket functional context (UniProt features + ClinVar variants):

- Pocket 1 (cluster 6):
  - Residue 9: Site — CDK7 binding
  - Residue 10: Binding site — Binding site
  - Residue 11: Binding site — Binding site

  - Residue 12: Binding site — Binding site
  - Residue 13: Binding site — Binding site
  - Residue 14: Binding site — Binding site
  - Residue 14: Modified residue — Phosphothreonine
  - Residue 15: Binding site — Binding site
  - Residue 15: Modified residue — Phosphotyrosine; by WEE1
  - Residue 16: Binding site — Binding site
  - Residue 18: Binding site — Binding site
  - Residue 33: Binding site — Binding site
  - Residue 81: Binding site — Binding site
  - Residue 82: Binding site — Binding site
  - Residue 83: Binding site — Binding site
  - Residue 86: Binding site — Binding site
  - Residue 89: Site — CDK7 binding
  - Residue 127: Active site — Proton acceptor
  - Residue 127: Binding site — Binding site
  - Residue 129: Binding site — Binding site
  - Residue 131: Binding site — Binding site
  - Residue 132: Binding site — Binding site
  - Residue 145: Binding site — Binding site
- Pocket 2 (cluster 0):
  - Residue 14: Binding site — Binding site
  - Residue 14: Modified residue — Phosphothreonine
  - Residue 15: Binding site — Binding site
  - Residue 15: Modified residue — Phosphotyrosine; by WEE1
  - Residue 127: Active site — Proton acceptor
  - Residue 127: Binding site — Binding site
  - Residue 129: Binding site — Binding site
  - Residue 131: Binding site — Binding site
  - Residue 160: Modified residue — Phosphothreonine; by CAK and CCRK
- Pocket 3 (cluster 8): no UniProt functional features or ClinVar variants overlap with pocket residues

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

8 pocket cluster(s) (of 28 total, filtered to those present in >=2 frames) appear only in later ensemble frames, absent from the static structure-provider output:

Cluster id	Frames present	Persistence	Residues
25	3/5	0.6	44(?), 46(46), 47(47), 48(48…
29	3/5	0.6	90(90), 100(100), 102(102), …
33	3/5	0.6	60(60), 113(113), 116(116), …
20	2/5	0.4	88(88), 91(91), 92(92), 94(9…
22	2/5	0.4	176(176), 177(177), 178(178)…
28	2/5	0.4	85(85), 90(90), 135(135), 13…
30	2/5	0.4	219(219), 220(220), 221(221)…
32	2/5	0.4	7(7), 8(8), 20(20), 21(21), …

3 pocket cluster(s) in the customer-visible top ranking; static (first-frame-only) fpocket found 18 pocket(s); the ensemble-informed ranker considered 46 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: 6Q4G) 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)
6Q49	95.9%	HGQ
6Q4H	95.9%	HGH
6Q48	95.9%	HHQ
6Q4J	95.9%	HHB
6Q4E	95.9%	HH5

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

PDB ID	TM-score	Sequence identity	In sequence-based list…	Description
2VTQ	0.989	91.9%	no	Identification of N-(4…
5AND	0.989	93.1%	no	Crystal structure of C…
4FKS	0.989	92.9%	no	Crystal structure of t…
4FKJ	0.988	93.2%	no	Crystal structure of t…
2W05	0.988	92.6%	no	Structure of CDK2 in c…

5/5 top structural neighbor(s) — 2VTQ, 5AND, 4FKS, 4FKJ, 2W05 — 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	
Functional context (UniProt + ClinVar)	returned data	
Conservation	returned data	
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.