Protein Knowledge Report — P00918

1. Target summary

- Input type: uniprot_accession
- Label: P00918
- Length: 260 aa
- UniProt accession: P00918
- Submitted: 2026-07-15T13:01:36Z
- Sequence:

MSHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAKYDPSLKPLSVSYDQATSLRILNNGHAFNVEFDDSQDKAVLKGGPLDGTYRLIQFHFHWG
SLDGQGSEHTVDKKKYAAELHLVHWNTKYGDFGKAVQQPDGLAVLGIFLKVGSAKPGLQKVVDVLDSIKTKGKSADFTNFDPRGLLPESLDYWTYPGSL
TTP

2. Family classification

- Family: Carb_anhydrase (PF00194) — Eukaryotic-type carbonic anhydrase
- E-value: 1.40e-94 Bit score: 317.2
- Model span: 1-255 of model (11-259 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: 1429
- Distinct ligand scaffolds curated below: 13

PDB ID	Ligand(s)	Resolution (Å)	Why selected
3K34	HGB, SUA	0.90	highest-resolution structure...
5Y2S	CO2	0.90	highest-resolution structure...
6KLZ	BCT	0.90	highest-resolution structure...
6ROB	HG, BE7, KBZ, GLC, BGC	0.93	highest-resolution structure...
6SBL	HG, L4Q, BE7	0.94	highest-resolution structure...
6RH4	BGC, 4NZ, GLC, BE7, HG	0.95	highest-resolution structure...
1LUG	SUA, MBO, HG	0.95	highest-resolution structure...
6RQI	BGC, GLC, FBW, BE7, HG	0.95	highest-resolution structure...

scanned the 100 highest-resolution of 1429 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 3K34 (HGB, 0.90 Å)

Residue (PDB)	Residue (pipeline)	Interaction	Partner
138	137	hydrophobic contact	PRO138 (hydrophobic) with li...

Representative complex: PDB 5Y2S (CO2, 0.90 Å)

Residue (PDB)	Residue (pipeline)	Interaction	Partner
199	198	hydrogen bond	THR199 (hydrogen bond) with ...
200	199	water bridge	THR200 (water bridge) with l...
200	199	water bridge	THR200 (water bridge) with l...
92	92	water bridge	GLN92 (water bridge) with li...

Representative complex: PDB 6KLZ (BCT, 0.90 Å)

Residue (PDB)	Residue (pipeline)	Interaction	Partner
199	198	hydrogen bond	THR199 (hydrogen bond) with ...
119	119	hydrogen bond	HIS119 (hydrogen bond) with ...
199	198	hydrogen bond	THR199 (hydrogen bond) with ...
94	94	water bridge	HIS94 (water bridge) with li...
200	199	water bridge	THR200 (water bridge) with l...
92	92	water bridge	GLN92 (water bridge) with li...
92	92	water bridge	GLN92 (water bridge) with li...
92	92	water bridge	GLN92 (water bridge) with li...
92	92	water bridge	GLN92 (water bridge) with li...

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

4. Conservation

- Family: PF00194
- Seed alignment size: 250 sequences
- Mean pairwise identity (seed alignment): 31.3%
- Note: computed over a sample of 40/250 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: 0.9737
- Conformational ensemble: 5 conformers (ANM, 5 frames)
- Static (single-frame) pocket count: 13
- Residue numbering: pipeline-sequential, with literature (author-deposited PDB 3K34) numbering shown in parentheses — 258/260 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	5(5), 7(7), 62(62), 64(64), ...
2 (cluster 2)	5/5	1.0	3(3), 4(4), 5(5), 6(6), 7(7)...
3 (cluster 7)	5/5	1.0	3(3), 4(4), 5(5), 9(9), 10(1...

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

- Pocket 1 (cluster 0): residues 92(92), 94(94), 119(119), 198(199), 199(200) (5 of 24 pocket residues match a known interaction site)
- Pocket 2 (cluster 2): no overlap with known interaction sites (18 pocket residues checked)
- Pocket 3 (cluster 7): no overlap with known interaction sites (11 pocket residues checked)

Pocket functional context (UniProt features + ClinVar variants):

- Pocket 1 (cluster 0):
  - Residue 7: Site — Fine-tunes the proton-transfer properties of H-64
  - Residue 7: ClinVar variant [Pathogenic] — Y7* (NM_000067.3(CA2):c.21C>A (p.Tyr7Ter))
  - Residue 62: Site — Fine-tunes the proton-transfer properties of H-64; involved in the binding of some activators, including histamine and L-histidine
  - Residue 64: Active site — Proton donor/acceptor
  - Residue 67: Site — Fine-tunes the proton-transfer properties of H-64; involved in the binding of some activators, including histamine and L-histidine
  - Residue 92: Site — Involved in the binding of some activators, including histamine and L-histidine
  - Residue 92: ClinVar variant [Likely pathogenic] — Q92P (NM_000067.3(CA2):c.275A>C (p.Gln92Pro))

  - Residue 92: ClinVar variant [Likely pathogenic] — Y92* (NM_000067.3(CA2):c.579C>G (p.Tyr193Ter))
  - Residue 94: Binding site — Binding site
  - Residue 96: Binding site — Binding site
  - Residue 119: Binding site — Binding site
  - Residue 198: Binding site — Binding site
  - Residue 199: Binding site — Binding site
  - Residue 208: ClinVar variant [Pathogenic] — W208fs (NM_000067.3(CA2):c.621del (p.Trp208fs))
- Pocket 2 (cluster 2):
  - Residue 7: Site — Fine-tunes the proton-transfer properties of H-64
  - Residue 7: ClinVar variant [Pathogenic] — Y7* (NM_000067.3(CA2):c.21C>A (p.Tyr7Ter))
  - Residue 64: Active site — Proton donor/acceptor
  - Residue 236: ClinVar variant [Pathogenic] — P236H (NM_000067.3(CA2):c.707C>A (p.Pro236His))
- Pocket 3 (cluster 7):
  - Residue 18: ClinVar variant [Pathogenic] — K18E (NM_000067.3(CA2):c.52A>G (p.Lys18Glu))

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

4 pocket cluster(s) (of 17 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
14	4/5	0.8	7(7), 13(13), 14(14), 17(17)...
20	3/5	0.6	51(51), 52(52), 54(54), 55(5...
19	2/5	0.4	152(153), 157(158), 160(161)...
25	2/5	0.4	41(41), 43(43), 44(44), 45(4...

3 pocket cluster(s) in the customer-visible top ranking; static (first-frame-only) fpocket found 13 pocket(s); the ensemble-informed ranker considered 30 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: 3K34) 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)
3K34	100.0%	HGB, SUA
3KS3	100.0%	(none recorded)
5Y2S	100.0%	CO2
4YXI	100.0%	4J8, MBO
4FPT	100.0%	0VZ, MBO

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

PDB ID	TM-score	Sequence identity	In sequence-based list...	Description
1ZSC	1.000	99.6%	no	CARBONIC ANHYDRASE II ...
5JE7	1.000	99.6%	no	Human carbonic anhydra...

3RYY	1.000	99.6%	no	Fluoroalkyl and Alkyl ...
2H4N	1.000	99.6%	no	H94N CARBONIC ANHYDRAS...
6QL3	1.000	98.0%	no	Crystal structure of c...

5/5 top structural neighbor(s) — 1ZSC, 5JE7, 3RYY, 2H4N, 6QL3 — 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	5 variant(s) listed more than one prot...
Conservation	returned data	computed over a sample of 40/250 seed ...
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.