Protein Knowledge Report — P07550

1. Target summary

- Input type: uniprot_accession
- Label: P07550
- Length: 413 aa
- UniProt accession: P07550
- Submitted: 2026-07-15T13:00:48Z
- Sequence:

MGQPGNGSAFLLAPNGSHAPDHDVTQERDEVWVVGMGIVMSLIVLAIVFGNVLVITAIAKFERLQTVTNYFITSLACADLVMGLAVVPFGAAHILMKMWT
FGNFWCEFWTSIDVLCVTASIETLCVIAVDRYFAITSPFKYQSLLTKNKARVIILMVWIVSGLTSFLPIQMHWYRATHQEAINCYANETCCDFFTNQAYA

2. Family classification

- Family: 7tm_1 (PF00001) — 7 transmembrane receptor (rhodopsin family)
- E-value: 4.40e-70 Bit score: 236.8
- Model span: 1-260 of model (50-326 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: 1561
- Distinct ligand scaffolds curated below: 24

PDB ID	Ligand(s)	Resolution (Å)	Why selected
5NM4	OLA, ZMA, CLR	1.70	highest-resolution structure...
9H37	CLR, OLA, A1IR0	1.72	highest-resolution structure...
9H2X	CLR, A1IR1, OLA	1.75	highest-resolution structure...
4N6H	OLA, EJ4, TLA	1.80	highest-resolution structure...
6LPJ	ZMA, CLR, D12, MYS, HEX, 8K6...	1.80	highest-resolution structure...
7ZBC	ACE, RET, NAG, DAO, PLM	1.80	highest-resolution structure...
7IO5	TEP, CLR, OLA, A1CPA	1.84	highest-resolution structure...
7IPE	OLA, A1CQJ, CLR, TEP	1.84	highest-resolution structure...

scanned the 100 highest-resolution of 1561 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)

8 ligand-bound structure(s) exist for this target's Pfam family (PF00001), but none checked aligned to the query at >=50% sequence coverage — likely other members of the same broad family (e.g. related kinases), not this specific protein. Interaction fingerprints require a structure of the query protein itself.

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

4. Conservation

- Family: PF00001
- Seed alignment size: 63 sequences
- Mean pairwise identity (seed alignment): 17.8%
- Note: computed over a sample of 40/63 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: 0.7914
- Conformational ensemble: 5 conformers (ANM, 5 frames)
- Static (single-frame) pocket count: 37
- Residue numbering: pipeline-sequential, with literature (author-deposited PDB 1GQ4) numbering shown in parentheses — 5/413 residues cross-walked
- Ensemble-informed ranked pocket clusters (customer-visible top): 3

Rank (cluster id)	Frames present	Persistence	Residues
1 (cluster 18)	5/5	1.0	63(?), 64(?), 66(?), 68(?), ...
2 (cluster 36)	4/5	0.8	90(?), 93(?), 94(?), 97(?), ...
3 (cluster 35)	5/5	1.0	60(?), 61(?), 62(?), 63(?), ...

Pocket functional context (UniProt features + ClinVar variants):

- Pocket 1 (cluster 18):
  - Residue 131: Disulfide bond — Disulfide bond
  - Residue 134: Disulfide bond — Disulfide bond
  - Residue 135: Disulfide bond — Disulfide bond
  - Residue 138: Disulfide bond — Disulfide bond
- Pocket 2 (cluster 36):
  - Residue 109: Disulfide bond — Disulfide bond
  - Residue 110: Disulfide bond — Disulfide bond
  - Residue 113: Binding site — Binding site
  - Residue 113: Disulfide bond — Disulfide bond
  - Residue 114: Disulfide bond — Disulfide bond
  - Residue 117: Disulfide bond — Disulfide bond
  - Residue 118: Binding site — Binding site
  - Residue 118: Disulfide bond — Disulfide bond
  - Residue 178: Disulfide bond — Disulfide bond
  - Residue 180: Disulfide bond — Disulfide bond
  - Residue 191: Disulfide bond — Disulfide bond
  - Residue 203: Binding site — Binding site
  - Residue 293: Binding site — Binding site
  - Residue 312: Binding site — Binding site
  - Residue 316: Binding site — Binding site
- Pocket 3 (cluster 35):
  - Residue 345: Modified residue — Phosphoserine; by PKA
  - Residue 346: Modified residue — Phosphoserine; by PKA

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

7 pocket cluster(s) (of 19 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
39	4/5	0.8	136(?), 137(?), 138(?), 139(...
38	3/5	0.6	343(?), 345(?), 346(?), 347(...
37	2/5	0.4	79(?), 82(?), 116(?), 117(?)...
40	2/5	0.4	284(?), 287(?), 288(?), 291(...
44	2/5	0.4	240(?), 243(?), 244(?), 246(...
47	2/5	0.4	109(?), 110(?), 113(?), 114(...

50	2/5	0.4	23(?), 26(?), 27(?), 30(?)

3 pocket cluster(s) in the customer-visible top ranking; static (first-frame-only) fpocket found 37 pocket(s); the ensemble-informed ranker considered 56 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: 1GQ4) for author-deposited numbering

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

6. Confidence flags

1 caution flag(s) — discount confidence accordingly:

- [Structural analysis] numbering_low_coverage: Numbering cross-walk against 1GQ4 only mapped 5/413 residues (1%) — the reference structure found is a poor match for most of this sequence (e.g. a short peptide/fragment, or a construct covering only a small domain of a larger protein). Most residues will show as unmapped ('?'); treat any '(literature)' number that DOES appear as coincidental unless independently checked, not as evidence the cross-walk is reliable for this target.

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)
7MBX	17.7%	Y01
5WIU	17.2%	AQD, OLA
8CU7	17.0%	CLR, ETF, LJX, OLA
4EIY	16.8%	CLR, OLA, ZMA
6LPJ	16.8%	8K6, CLR, D10, D12, ER0, HEX, MYS, OCT…

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

PDB ID	TM-score	Sequence identity	In sequence-based list…	Description
2Y03	0.982	57.1%	no	TURKEY BETA1 ADRENERGI…
5F8U	0.982	57.7%	no	Ligand occupancy in cr…
4AMI	0.980	57.0%	no	Turkey beta1 adrenergi…
8W1V	0.978	89.4%	no	The beta2 adrenergic r…
3ZPR	0.977	55.8%	no	Thermostabilised turke…

5/5 top structural neighbor(s) — 2Y03, 5F8U, 4AMI, 8W1V, 3ZPR — 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	no self-matching ligand-bound structur…	8 ligand-bound structure(s) exist for …
Functional context (UniProt + ClinVar)	returned data	

Conservation	returned data	computed over a sample of 40/63 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	

1 of 8 sections did not return data this run: Interaction fingerprints.