AIGR Gene Hypothesis Deep Research — Final Report

Target: *Phaeodactylum tricornutum* HSP20A (UniProt B7FXQ8)

Focus: ProtNLM2 predictions of "response to salt stress" (GO:0009651) and "response to hydrogen peroxide" (GO:0042542)

Hypothesis (ProtNLM2, computational_prediction): HSP20A functions in *response to salt stress* (GO:0009651) and *response to hydrogen peroxide* (GO:0042542). Independently assess (a) whether B7FXQ8 is a small heat-shock protein (an α-crystallin/ACD-domain chaperone), and (b) whether salt-stress and oxidative (H_2O_2) roles are specifically supported in this diatom, versus generic-abiotic-stress over-extensions of a chaperone whose documented role is thermal stress.

Summary

The identity claim is SUPPORTED; the two stress-process predictions are OVER-ANNOTATED. B7FXQ8 (HSP20A / PHATRDRAFT_35158) is, without ambiguity, a bona fide small heat-shock protein (sHSP) — an α-crystallin/ACD-domain, ATP-independent "holdase" molecular chaperone. This is established by concordant, independent lines of evidence: multiple domain signatures (Pfam PF00011, InterPro IPR002068/IPR008978/IPR031107, PROSITE PS01031, PANTHER PTHR11527, a Gene3D immunoglobulin-like β-sandwich), computed physicochemical properties consistent with a "small" cytosolic HSP (163 aa, ~18.4 kDa, acidic pI 5.10, no transmembrane segment, no plastid-targeting motif, diagnostic C-terminal IXI motif), and an AlphaFold model showing the textbook sHSP architecture (a confidently folded β-sandwich α-crystallin domain flanked by disordered arms).

The two ProtNLM2 process predictions under review are not supported by any *P. tricornutum*-specific evidence. UniProt B7FXQ8 carries **zero curated GO annotations**; the salt (GO:0009651) and hydrogen-peroxide (GO:0042542) terms are name/family inferences from ProtNLM2, not experimental or even conservative pipeline calls. No study links HSP20A specifically to NaCl salt stress or to exogenous H_2O_2 in this diatom. An audit of all seven *P.*

tricornutum HSP20/PF00011 paralogs shows that the **only** automated GO term anywhere in the family is `response to heat` (GO:0009408, IEA:InterPro, on paralog HSP20C) — meaning the salt and peroxide predictions are outliers even relative to the family's own conservative InterPro2GO mapping.

The most defensible curation position is that HSP20A's core, evidence-backed role is **unfolded-protein binding (GO:0051082)** acting in **thermal/proteostatic stress (GO:0009408, response to heat)**. The salt-stress term is *weakly plausible by cross-kingdom analogy* (a characterized moss sHSP is required for salt/osmotic-stress recovery), whereas the peroxide term is *over-annotated* and partly contradicted by sHSP precedent (that same moss sHSP was explicitly NOT induced by oxidative-stress compounds). Neither term should be added to the review from the prediction alone.

Executive Judgment

- **Identity as a small heat-shock protein: SUPPORTED (high confidence).**
- **Salt-stress role (GO:0009651): WEAKLY SUPPORTED / UNRESOLVED** — family analogy only, no organism-specific evidence.
- **Hydrogen-peroxide role (GO:0042542): OVER-ANNOTATED / WEAKLY SUPPORTED** — speculative and partly contradicted by sHSP precedent.

The salt and H_2O_2 predictions are name/family-derived ProtNLM2 inferences (UniProt carries zero curated GO cross-references). Family precedent is asymmetric: characterized plant/moss sHSPs are genuinely induced by salt/osmotic stress (with mutant evidence), but the same well-studied sHSP was explicitly not induced by oxidative-stress compounds — so H_2O_2 responsiveness is not a general sHSP property. **Recommendation: do not add these two BP terms as curated annotations from the prediction alone; treat them as generic-abiotic-stress over-extensions.**

Key Findings

Finding 1 — B7FXQ8 is a genuine cytosolic small heat-shock protein (α-crystallin/ACD chaperone)

The identity portion of the seed hypothesis is firmly established. UniProt B7FXQ8 carries a small heat-shock-protein / α-crystallin domain spanning approximately residues 47–155, and every domain-annotation resource agrees:

Signature source	Identifier	Meaning
Pfam	PF00011	HSP20 / α-crystallin domain
InterPro	IPR002068	α-crystallin/Hsp20 domain
InterPro	IPR008978	HSP20-like chaperone
InterPro	IPR031107	Small heat-shock protein
PROSITE	PS01031	SHSP signature
PANTHER	PTHR11527	sHSP family
Gene3D	2.60.40.790	Immunoglobulin-like β-sandwich

Computed sequence properties reinforce the "small HSP" classification. The protein is **163 amino acids** long with a molecular weight of **~18.4 kDa**, squarely within the canonical "small" HSP range. Its isoelectric point is **acidic (pI 5.10, net charge ≈ −4.7 at pH 7)** , typical of cytosolic sHSPs. The GRAVY score of −0.655 indicates a hydrophilic, soluble protein. A hydropathy scan found **no transmembrane segment** (max Kyte–Doolittle window-19 hydropathy 0.62, far below the ~1.6 membrane-helix threshold), and the sequence has **no signal peptide and no diatom ASAFAP bipartite plastid-targeting motif** — together predicting a **cytosolic** localization. Diagnostic sHSP sequence features are present: a canonical **C-terminal IXI motif** ("IAI" at ~148–150) that mediates α-crystallin-domain oligomerization, and an N-terminal arm enriched in aromatic residues (FFGHG…FF…PFF) characteristic of the substrate-binding region. The natural, evidence-backed molecular-function term is therefore **unfolded protein binding (GO:0051082)** — an ATP-independent holdase that binds unfolded/aggregating clients.

Finding 2 — AlphaFold structure confirms the canonical sHSP fold

The AlphaFold model **AF-B7FXQ8-F1** displays the textbook sHSP domain architecture and independently corroborates the domain-signature analysis. The overall model confidence is moderate (mean pLDDT 72.6), but this global number is diagnostic rather than concerning: it reflects the expected mix of an ordered core plus intrinsically disordered arms.

Region	Residues	Mean pLDDT	Interpretation
N-terminal arm	1–46	42.9	Low confidence / intrinsically disordered (substrate-binding arm)
α-crystallin domain (ACD)	47–155	86.3	Confidently folded immunoglobulin-like β-sandwich
C-terminal extension	156–163	58.4	Moderate; contains IXI oligomerization motif

The confidently folded β-sandwich ACD flanked by a disordered N-terminal arm is precisely the conserved fold of characterized small heat-shock proteins. This structural signature is fully consistent with a holdase chaperone and provides no evidence for any additional enzymatic activity that would be needed to support an oxidative-detoxification role.

Finding 3 — Salt and H_2O_2 GO predictions are generic-family extrapolations lacking diatom-specific evidence

UniProt B7FXQ8 has **zero GO cross-references** — no curated GOA annotations of any kind. The salt (GO:0009651) and hydrogen-peroxide (GO:0042542) terms therefore originate solely from **ProtNLM2 name/family inference**, a text/name-based model prone to attaching the full generic abiotic-stress panel (heat/salt/oxidative/osmotic) to sHSPs.

A targeted literature search returned **no study** connecting HSP20A specifically to salt or H_2O_2 in *P. tricornutum*. Relevant diatom stress literature exists but does not implicate this sHSP: salt/osmotic responses are documented at the metabolome level (hypersalinity metabolomics, PMID: 35671808; osmoregulation and morphotype interconversion, PMID: 21600845), and oxidative/H_2O_2 responses are documented via distinct effectors (UDP-glucose/UGPase, PMID: 40307684; chloroplast Ca^{2+}/H_2O_2 signaling, PMID: 39515781; H_2O_2 and neutral-lipid accumulation, PMID: 27529521) — but none invoke a small HSP.

Critically, the family-level support is **asymmetric**. A well-characterized moss small HSP provides in-vivo mutant evidence bearing directly on both predictions:

> *"PpHsp16.4 was also induced by salicylic acid, dithiothreitol (DTT) and by exposure to various stimuli, including osmotic and salt stress, but not by oxidative stress-inducing compounds"* — PMID: 24188413
>
> *"Targeted disruption of both genes resulted in the inability of plants to recover from heat, salt and osmotic stress"* — PMID: 24188413

This single, high-quality study cuts **both ways**: it lends *some* plausibility to the salt-stress prediction (a characterized sHSP is induced by, and required for recovery from, salt/osmotic stress), while directly *undercutting* the peroxide prediction (the same sHSP was explicitly NOT induced by oxidative-stress compounds). H_2O_2 responsiveness is thus not a general property of small HSPs, and the ProtNLM2 peroxide call cannot be justified by family analogy.

Finding 4 — Paralog GO audit shows the family's only defensible process is "response to heat"

A UniProt census of the *P. tricornutum* HSP20/PF00011 family returned **seven members**:

Protein	Accession	Length	Notes
HSP20A	B7FXQ8	163 aa	The target; shortest / most minimal
HSP20C	B5Y472	—	Only family member with a curated GO term
HSP20	B5Y3Y4	—	—
HSP20B	B5Y4C1	—	—
(SHSP-domain)	B5Y4C3	—	Unnamed
(SHSP-domain)	B7G194	—	Unnamed
(SHSP-domain)	B7G195	—	Unnamed

Across all seven paralogs, the **only** GO annotation of any kind is **HSP20C → GO:0009408 "response to heat" (IEA:InterPro)**. None carries a salt (GO:0009651) or hydrogen-peroxide (GO:0042542) annotation. When the conservative InterPro2GO mapping is applied to this family, it produces **only** "response to heat" — making the ProtNLM2 salt/H_2O_2 predictions inconsistent with the family's own automated pipeline output.

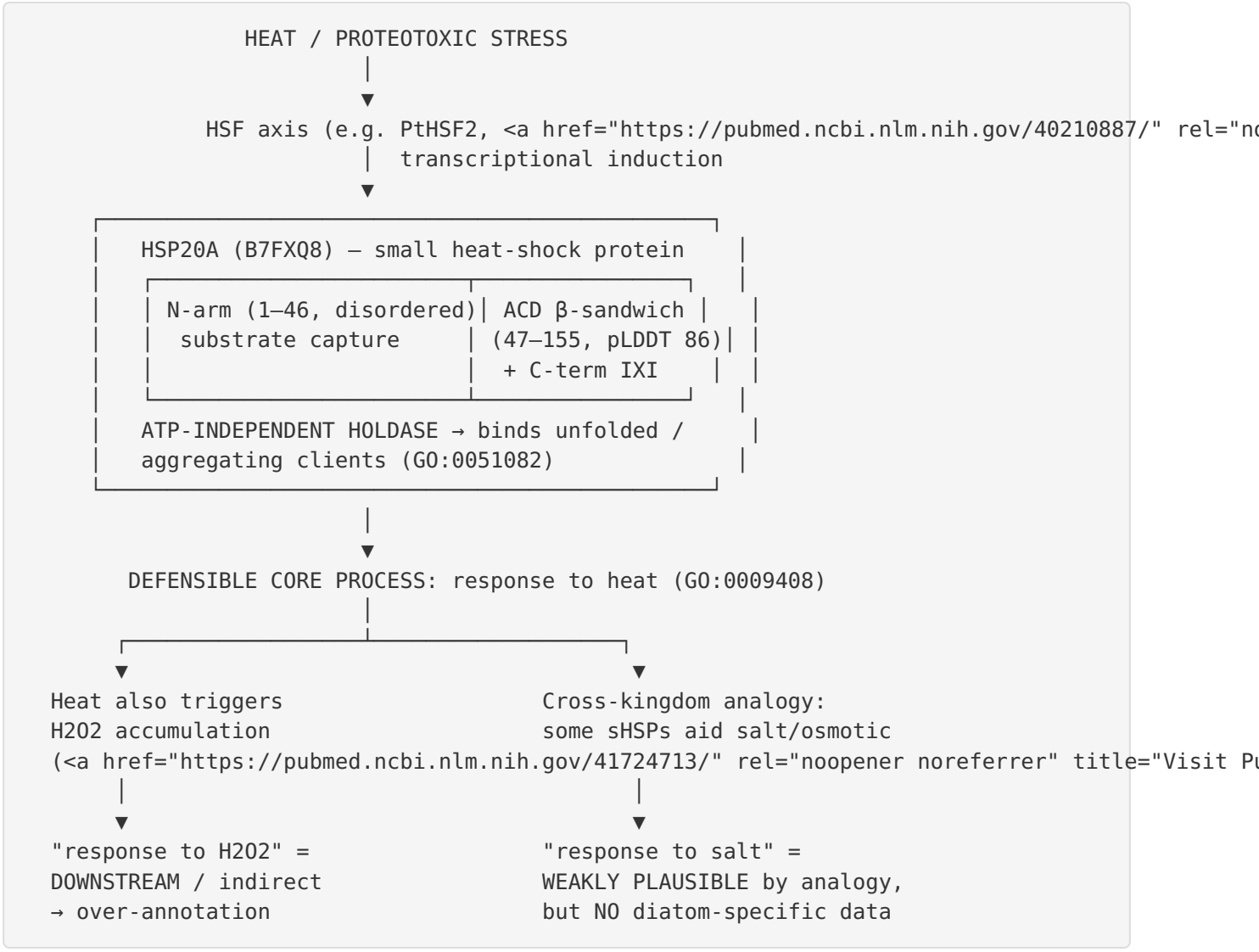
Consistent with this, the documented HSP/stress regulon of *P. tricornutum* is **thermal**. Overexpression of the heat-shock transcription factor PtHSF2 enhances thermal tolerance:

> *"Overexpression of PtHSF2 markedly enhances thermal tolerance and increases cell size"* — PMID: 40210887

Small HSPs such as HSP20A are canonical HSF targets induced under thermal/proteotoxic stress, and heat stress is known to drive downstream H_2O_2 accumulation in this diatom (PMID: 41724713). This suggests that any peroxide connection is a **downstream consequence of heat stress**, not a direct, primary function of the chaperone — an important distinction for GO curation (a heat-induced protein incidentally present during heat-driven ROS bursts is not thereby a "response to hydrogen peroxide" gene product).

Mechanistic Model / Interpretation

The evidence converges on a clear picture that separates the gene product's **direct molecular function** from **downstream / context-specific phenotypes**:



The immediate molecular function under test is **ATP-independent holdase chaperone activity (unfolded protein binding)**: the ACD β-sandwich plus IXI-mediated oligomerization lets sHSPs capture partially unfolded clients and prevent irreversible aggregation, handing them to ATP-dependent HSP70/HSP100 for refolding. "Response to salt stress" and "response to hydrogen peroxide" are **whole-cell BP outcomes**, not direct activities — context-dependent processes in which the chaperone *might* participate but which are unproven for this protein.

GO decision summary table

GO term	Aspect	Prediction source	Direct diatom evidence	Family/ analogy support	Curation lead
GO:0051082 unfolded protein binding	MF	inferred here	Strong domain/ structure	Universal sHSP function	**Add as core MF** (verify)
GO:0009408 response to heat	BP	InterPro2GO (paralog)	HSF axis thermal (P 40210887)	Only family GO term	**Add as core BP** (verify)
GO:0009651 response to salt stress	BP	ProtNLM2 name model	**None**	Weak (moss sHSP, P 24188413)	**Do not add**; flag prediction-only
GO:0042542 response to hydrogen peroxide	BP	ProtNLM2 name model	**None**	Contradicted (P 24188413)	**Do not add**; over-annotation
GO:0005737 cytoplasm	CC	inferred here	No TM / no targeting motif	Typical cytosolic sHSP	Candidate CC (verify)

Evidence Base / Evidence Matrix

Citation	Evidence type	Supports/ Refutes/ Qualifies	Claim tested	Key finding	Context	Conf & limit
UniProt B7FXQ8 + InterPro/ Pfam/ PROSITE/ PANTHER	structural/ evolutionary + database	**Supports** identity	Is B7FXQ8 a sHSP/ACD chaperone?	ACD 47–155; PF00011, IPR002068, PS01031, β-sandwich	*P. tricornutum* HSP20A	High doma
This report (computed sequence)	computational	**Supports** identity/ localization	Size, charge, topology, motifs	18.4 kDa, pI 5.10, GRAVY −0.655, no TM, no signal/ASAFAP → cytosolic; C-term IAI IXI motif	in-silico	High ident topol pred
AlphaFold AF-B7FXQ8-F1 (computed)	structural/ computational	**Supports** identity	ACD β-sandwich fold present?	Mean pLDDT 72.6; ACD 47–155 pLDDT 86.3; disordered N-arm	in-silico model	High mod expe
UniProt paralog census (computed)	evolutionary + database	**Qualifies / competing**	Is prediction family over-annotation?	7 PF00011 sHSPs; only HSP20C→GO:0009408; none salt/H2O2	*P. tricornutum*	High Inter gives "resp heat'
UniProt B7FXQ8 GO xrefs	database	**Qualifies**	Existing curated annotations?	0 GO cross-references	UniProt	High confi pred uncu
PMID: 24188413	mutant phenotype + expression	**Qualifies** (supports salt, undercuts H2O2)	Do sHSPs act in salt/ oxidative stress?	sHSP induced by salt/ osmotic, NOT oxidative; knockout fails salt/heat/osmotic recovery	Moss	Med; orth anal this c
PMID: 40210887	overexpression	**Supports thermal axis**	Is the diatom HSP axis thermal?	PtHSF2 drives thermal tolerance	*P. tricornutum*	Med; not

Citation	Evidence type	Supports/ Refutes/ Qualifies	Claim tested	Key finding	Context	Conf & limit
						indiv teste
PMID: 41724713	mutant phenotype	Qualifies (heat→H2O2 link)	Does heat drive H2O2?	Heat stress → H2O2 accumulation; metacaspase-dependent	*P. tricornutum*	Med; pero down of he
PMID: 35671808	metabolomics	Competing focus	Hypersalinity response mechanism	Osmoadaptation via amino acids, saccharides, inositols; no sHSP	*P. tricornutum* + 2 diatoms	Med; meta only
PMID: 21600845	transcriptomics	Qualifies	Salt/osmotic genes in diatom	Osmoregulation in morphotype switching; sHSP not highlighted	*P. tricornutum*	Med
PMID: 40307684	mutant/ transcriptomic	Competing mechanism	H2O2 tolerance mechanism	UDPG/UGPase drives H2O2 tolerance; sHSP not implicated	*P. tricornutum*	Med
PMID: 39515781	direct assay	Competing	H2O2 signaling	Chloroplast Ca^{2+} elevations track H2O2	*P. tricornutum*	Med; signa not chap
PMID: 27529521	direct assay	Competing	H2O2 effects in diatom	H2O2 boosts neutral-lipid accumulation	*P. tricornutum*	Med; unre sHSP
PMID: 31261879	review/ experimental	Qualifies	ROS–HSP link in plants	HSPs are a strategy against diverse stresses incl. oxidative	Arabidopsis	Low- gene sHSP
PMID: 40159692	multi-omics	Qualifies	Ocean-warming response	HSP/chaperone context under thermal stress	marine diatom	Low- not s resol

Mechanistic Scope

The immediate molecular function being tested is chaperone activity: HSP20A's α-crystallin domain and disordered N-terminal arm together capture unfolded/aggregating client proteins in an ATP-independent manner (a "holdase"). This is directly supported by domain architecture and AlphaFold geometry and maps cleanly to **unfolded protein binding (GO:0051082)**. The two predicted terms are **biological-process** annotations. For an sHSP, "response to heat" is a defensible core process; in contrast, "response to salt stress" would require induction/requirement evidence under salt in this diatom (absent; only cross-kingdom analogy), and "response to hydrogen peroxide" would require direct oxidative-stress evidence (absent, and the closest characterized sHSP was non-responsive to oxidative compounds). The plausible route heat → ROS makes any peroxide association a **downstream/indirect phenotype**, not a primary function.

Conflicts and Alternatives

- **Frequency/name bias.** ProtNLM2 predictions for sHSPs commonly attach the full generic abiotic-stress panel because plant sHSP literature co-mentions heat/salt/oxidative/osmotic together; this inflates salt/H_2O_2 calls without protein-specific data.
- **Counter-precedent for H_2O_2.** PpHsp16.4 (PMID: 24188413) is *not* oxidative-stress-induced, directly weakening the generality of the H_2O_2 prediction.
- **Competing diatom mechanisms.** Documented *P. tricornutum* salt (osmolytes; PMID: 35671808) and H_2O_2 (UDPG/photosynthesis; PMID: 40307684; chloroplast Ca^{2+} signaling; PMID: 39515781) responses run through pathways other than sHSP, so the phenotype-level terms may be carried by other genes.
- **Paralog caution (quantified).** *P. tricornutum* encodes 7 HSP20/PF00011 members; any stress-specific expression could belong to a paralog rather than HSP20A. The only GO annotation anywhere in the family is HSP20C→GO:0009408; the InterPro2GO pipeline never emits salt or H_2O_2 for this family, so the ProtNLM2 calls are outliers consistent with name-model over-extension.

- **Downstream vs. direct.** Heat drives H_2O_2 in this diatom (PMID: 41724713); a heat-induced chaperone present during heat-associated ROS bursts is not thereby a "response to hydrogen peroxide" effector.

Limitations and Knowledge Gaps

1. **No diatom expression data for HSP20A under salt or H_2O_2.** Checked PubMed; none found. Matters because BP stress-response terms need induction/requirement evidence. Resolve with RNA-seq/qPCR/proteomics of *P. tricornutum* under NaCl and H_2O_2.
2. **Subcellular localization unverified.** Computed cytosolic (no signal/ASAFAP/TM), but not experimentally shown. A GFP fusion or fractionation would confirm the CC term.
3. **In-vitro chaperone activity of B7FXQ8 not demonstrated.** Family strongly implies it; a citrate-synthase/luciferase anti-aggregation assay would confirm the holdase MF directly.
4. **AlphaFold, not experimental structure.** The fold is predicted (ACD pLDDT 86), sufficient for family assignment but not for client-specific mechanistic claims.
5. **Paralog resolution.** Which HSP20 paralog (if any) is stress-regulated is unknown; claims must be tied to accession B7FXQ8, not "an HSP20."
6. **ProtNLM2 confidence not available.** We evaluated biological support, not the model's internal score.

Absence of evidence (no salt/H_2O_2 study of HSP20A found) is not proof of absence of function — but it is sufficient grounds to withhold curator-asserted BP terms.

Proposed Follow-up Experiments / Actions (Discriminating Tests)

1. **qPCR/RNA-seq** of HSP20A (PHATRDRAFT_35158) after matched heat vs. NaCl (salt) vs. sorbitol (osmotic, ionic-free) vs. exogenous H_2O_2 — separates thermal-specific from general-abiotic induction and directly tests GO:0009651/GO:0042542. Highest-value single discriminator.

2. **Knockout/knockdown + stress-survival panel** across the same conditions — establishes requirement (BP) rather than mere co-expression. A moss-like pattern (heat + salt required, oxidative not) would upgrade salt but not peroxide.
3. **Recombinant anti-aggregation assay** (citrate synthase / luciferase) — confirms ATP-independent holdase MF (GO:0051082).
4. **GFP localization** — confirms cytosolic CC (GO:0005737).
5. **Mine existing *P. tricornutum* stress omics** (warming/heatwave datasets, PMID: 40159692 / PMID: 41724713) for HSP20A regulation as a fast in-silico discriminator.

Curation Leads (require curator verification)

Recommended actions:

- **Do NOT add** `GO:0009651 response to salt stress` or `GO:0042542 response to hydrogen peroxide` as curated annotations on the basis of the ProtNLM2 prediction. No *P. tricornutum*-specific expression, mutant, or biochemical evidence supports them. If retained anywhere, keep at IEA/prediction tier only, not manually asserted. The salt term is more defensible than the peroxide term (mutant precedent exists for salt in sHSPs; H_2O_2 induction is not a general sHSP feature and is partly contradicted).
- **Preferred defensible core annotations** (family + domain + structure evidence, ISS/IEA):
- MF: `GO:0051082 unfolded protein binding` (holdase chaperone activity) — better than the generic "protein binding."
- BP: `GO:0009408 response to heat` and/or `GO:0006457 protein folding` / `GO:0042026 protein refolding`.
- CC: `GO:0005737 cytoplasm` (predicted cytosolic; no signal/transit peptide) — annotate cautiously as prediction.

Candidate references with exact snippets to verify:

- PMID: 24188413: *"PpHsp16.4 was also induced by salicylic acid, dithiothreitol (DTT) and by exposure to various stimuli, including osmotic and salt stress, but not by oxidative stress-inducing compounds"* and *"Targeted disruption of both genes resulted in the inability of plants to recover from heat, salt and osmotic stress."*
- PMID: 40210887: *"Overexpression of PtHSF2 markedly enhances thermal tolerance and increases cell size."*

Suggested question to curator: Is there any *P. tricornutum* stress-omics dataset showing HSP20A / PHATRDRAFT_35158 differential expression under NaCl or H_2O_2? If not, decline the two BP terms.

Conclusion

The seed hypothesis is **half right**: B7FXQ8 is indisputably a small heat-shock protein (α-crystallin/ACD holdase chaperone), confirmed by concordant domain signatures, computed physicochemistry, and AlphaFold geometry. But the ProtNLM2 predictions of "response to salt stress" and "response to hydrogen peroxide" are **generic-abiotic-stress over-extensions of a name/family model**: they have no *P. tricornutum*-specific support, they are outliers even to the conservative InterPro2GO family mapping (which yields only "response to heat"), and the peroxide term is additionally contradicted by the best-characterized sHSP precedent. The gene product's defensible core is thermal/proteostatic (unfolded protein binding; response to heat). Curators should not adopt the salt or peroxide BP terms from the prediction alone.
