

Supporting Information

Mechanistic basis for the evolution of chalcone synthase catalytic cysteine reactivity in land plants

Geoffrey Liou^{1,2}, Ying-Chih Chiang³, Yi Wang³, and Jing-Ke Weng^{1,2,*}

From the ¹Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA;
²Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA; ³Department of Physics,
The Chinese University of Hong Kong, Shatin, NT, Hong Kong

Running title: *Evolution of chalcone synthase cysteine reactivity*

*To whom correspondence should be addressed:

Jing-Ke Weng: Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA;
wengj@wi.mit.edu, Tel. (617) 324-4921

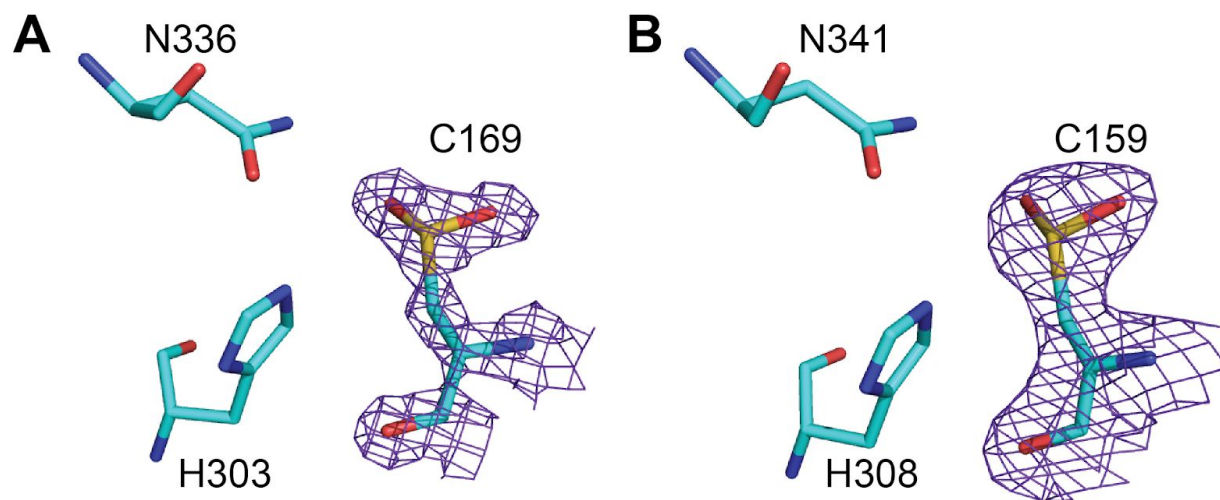


Figure S1. Active site structures of *Medicago sativa* CHS (**A**) (PDB ID 1BI5) and *Gerbera hybrida* 2-pyrone synthase (**B**) (PDB ID 1QLV) showing catalytic cysteine oxidized to sulfinic acid. The 2F_o-F_c composite map contoured at 1.5σ is shown around the catalytic cysteine.

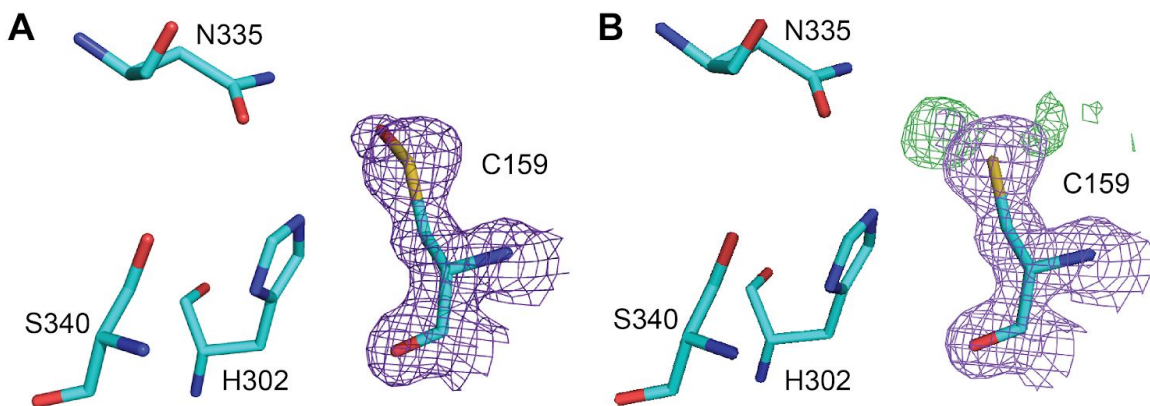


Figure S2. Active site structure of SmCHS crystals soaked in 1 mM hydrogen peroxide for 75 min. **A**, The 2F_o-F_c composite map to 1.55 Å resolution and contoured at 1.5σ is shown around the catalytic cysteine, modeled as oxidized to sulfenic acid. **B**, The 2F_o-F_c composite map to 1.55 Å resolution and contoured at 1.5σ is shown as purple and the F_o-F_c difference map contoured at 3.0σ is shown as green around the catalytic cysteine, modeled as reduced cysteine, indicating clear residual electron density for the oxidized sulfenic acid.

	SmCHS H ₂ O ₂ 75 min
PDB ID	6DXF
Data collection	
Total reflections	404281 (37428)
Unique reflections	108309 (10712)
Multiplicity	3.7 (3.5)
Completeness (%)	98.71 (98.24)
Mean I/sigma(I)	11.08 (1.66)
R-merge	0.08912 (0.825)
CC1/2	0.996 (0.532)
Refinement	
Resolution range (Å)	102.9 - 1.55 (1.605 - 1.55)
Space group	P 1 21 1
Unit cell (Å)	55.54 67.064 102.993
Unit cell (°)	90 91.719 90
R-work	0.1550 (0.2771)
R-free	0.1834 (0.3058)
Non-hydrogen protein atoms	5807
Water molecules	686
RMSD bonds (Å)	0.01
RMSD angles (°)	1.25
Ramachandran favored (%)	97.32
Ramachandran allowed (%)	2.54
Ramachandran outliers (%)	0.13
Average B-factor	24.03

Table S1. Statistics for crystal structure of SmCHS crystals soaked in 1 mM hydrogen peroxide for 75 minutes. The highest-resolution shell values are given in parentheses.

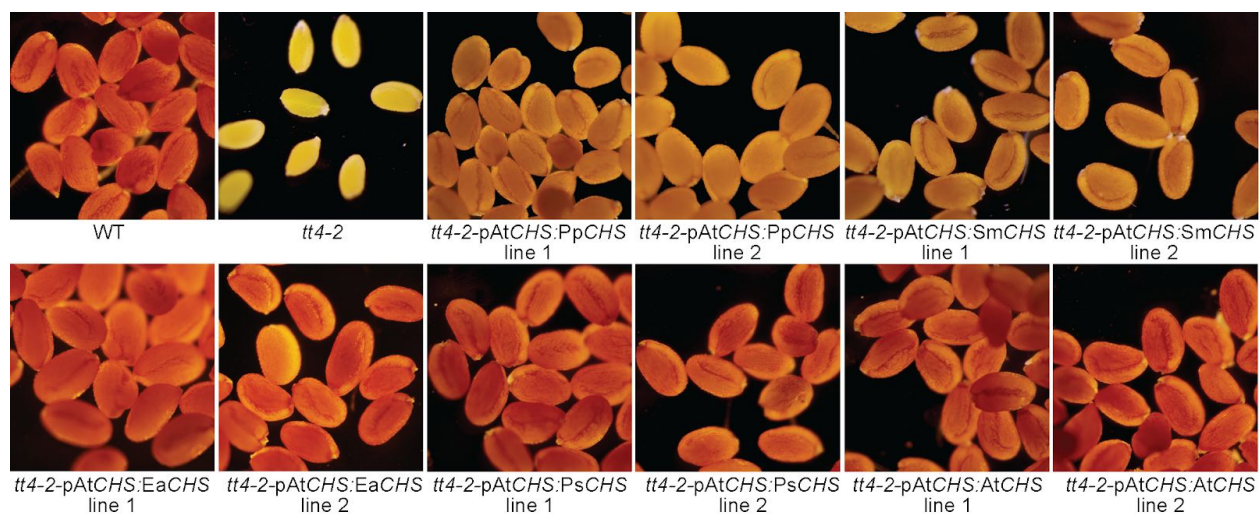


Figure S3. Complementation of the *transparent testa* seed phenotype of *tt4-2* mutant *Arabidopsis thaliana*. CHS orthologs were expressed under the *AtCHS* promoter. CHS from ephylllophytes (*AtCHS*, *PsCHS*, *EaCHS*) fully complement the mutant phenotype, whereas CHS from basal land plants (*SmCHS*, *PpCHS*) only partially complement.

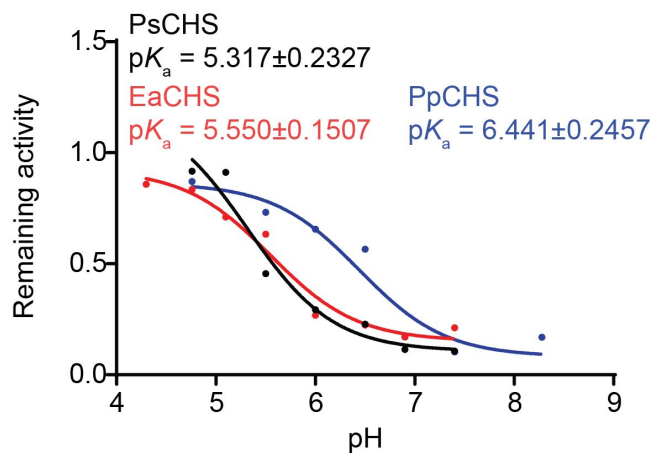


Figure S4. pK_a measurement of PsCHS, EaCHS, and PpCHS wild type enzymes. CHS enzyme was pre-incubated at various pH in the 25 μ M iodoacetamide inhibitor or water (control) for 30 s, and an aliquot was taken to run in a CHS activity assay. The ratio of naringenin product produced in the iodoacetamide treatment divided by the control treatment was calculated for each pH point. A nonlinear regression was performed to fit a log(inhibitor) vs. response curve to determine the pH at which 50% of maximal inhibition was achieved, which was determined to be the pK_a of the catalytic cysteine residue. The pK_a of PsCHS and EaCHS are close to the 5.5 determined for other euphylllophyte CHSs, whereas the pK_a of PpCHS is over 1 pH unit higher, similar to that of SmCHS.

Catalytic Cys

Plagiocchila asplenoides_CHS	58	ALRAAKFRFRD	OKDSGLTKK	164	VMMWQGGCFPGAGSVTRVAKD
Lejeuneaceae_sp_CHS	39	ELKFKKFRFRD	ERSSGTKK	145	VMMWQGGCFPGAAAVTRVAKD
Porella pinnata_CHS	61	ELKFKKFRFRD	OKDSGLTKK	167	VMMWQGGCFPGAAVTRVAKD
Porella navicularis_CHS	53	ELKFKKFRFRD	OKDSGLTKK	159	VMMWQGGCFPGAAVTRVAKD
Radula lindenbergia_CHS	56	DLKFKKFRFRD	EKSGTKK	162	VMMWQGGCFPGAAVTRVAKD
Sphaerocarpos texanus_CHS	56	DMKKKFRFRD	EKSGTKK	162	VMMWQGGCFPGAAVTRVAKD
Physcomitrella patens_CHS1	57	ALRAAKFRFRD	OKDSGLTKK	163	VMMWQGGCFPGAGSVTRVAKD
Physcomitrella patens_CHS2	57	ALRAAKFRFRD	OKDSGLTKK	163	VMMWQGGCFPGAGSVTRVAKD
Physcomitrella patens_CHS3	57	ALRAAKFRFRD	OKDSGLTKK	163	VMMWQGGCFPGAGSVTRVAKD
Physcomitrella patens_CHS4	50	ALRAAKFRFRD	OKDSGLTKK	156	VMMWQGGCFPGAGSVTRVAKD
Physcomitrella patens_CHS5	50	ALRAAKFRFRD	OKDSGLTKK	156	VMMWQGGCFPGAGSVTRVAKD
Physcomitrella patens_CHS6	50	ALHAAKFRFRD	OKDSGLTKK	156	VMMWQGGCFPGAGSVTRVAKD
Physcomitrella patens_CHS7	50	ALHAAKFRFRD	OKDSGLTKK	156	VMMWQGGCFPGAGSVTRVAKD
Physcomitrella patens_CHS8	50	ALHAAKFRFRD	OKDSGLTKK	156	VMMWQGGCFPGAGSVTRVAKD
Physcomitrella patens_CHS9	63	ALRAAKFRFRD	OKDSGLTKK	169	VMMWQGGCFPGAGSVTRVAKD
Physcomitrella patens_CHS10	63	ALRAAKFRFRD	OKDSGLTKK	169	VMMWQGGCFPGAGSVTRVAKD
Physcomitrella patens_CHS11	63	ALRAAKFRFRD	OKDSGLTKK	169	VMMWQGGCFPGAGSVTRVAKD
Physcomitrella patens_CHS12	63	ALRAAKFRFRD	OKDSGLTKK	169	VMMWQGGCFPGAGSVTRVAKD
Physcomitrella patens_CHS13	63	ALRAAKFRFRD	OKDSGLTKK	169	VMMWQGGCFPGAGSVTRVAKD
Tetraphis pellucida_CHS	47	ALRAAKFRFRD	OKDSGLTKK	153	VMMWQGGCFPGAGSVTRVAKD
Trichum angustatum_CHS	61	ALRAAKFRFRD	OKDSGLTKK	158	VMMWQGGCFPGAGSVTRVAKD
Buxbaumia aphylla_CHS	71	ALKSKFRFRD	OKDSGLTKK	176	VMMWQGGCFPGAGSVTRVAKD
Leucobryum albidum_CHS	50	ALRAAKFRFRD	OKDSGLTKK	176	VMMWQGGCFPGAGSVTRVAKD
Timmia austriaca_CHS1	66	ALKGGFRFRD	OKDSGLTKK	172	VMMWQGGCFPGAGSVTRVAKD
Dicranum scoparium_CHS	50	ALRAAKFRFRD	OKDSGLTKK	156	VMMWQGGCFPGAGSVTRVAKD
Timmia austriaca_CHS2	42	ALKGGFRFRD	OKDSGLTKK	148	VMMWQGGCFPGATVTRVAKD
Ceratodon purpureus_CHS	47	ALKGGFRFRD	OKDSGLTKK	153	VMMWQGGCFPGATVTRVAKD
Rosulabryum cf capillare_CHS1	61	ALKGGFRFRD	EKSGTKK	167	VMMWQGGCFPGATVTRVAKD
Rosulabryum cf capillare_CHS2	67	ALKGGFRFRD	EKSGTKK	173	VMMWQGGCFPGATVTRVAKD
Philonotis fontana_CHS	72	ALKGGFRFRD	EKSGTKK	178	VMMWQGGCFPGATVTRVAKD
Hedwigia ciliata_CHS	72	ALKGGFRFRD	EKSGTKK	178	VMMWQGGCFPGATVTRVAKD
Rhynchostegium serrulatum_CHS	49	ALRAAKFRFRD	OKDSGLTKK	155	VMMWQGGCFPGATVTRVAKD
Anomodon attenuatus_CHS	71	ALKGGFRFRD	EKSGTKK	177	VMMWQGGCFPGATVTRVAKD
Pseudotaxiphylum elegans_CHS	71	ALKGGFRFRD	EKSGTKK	177	VMMWQGGCFPGATVTRVAKD
Leucodon julaceus_CHS	75	ALKGGFRFRD	EKSGTKK	181	VMMWQGGCFPGATVTRVAKD
Leucodon brachyopus_CHS	74	ALKGGFRFRD	EKSGTKK	180	VMMWQGGCFPGATVTRVAKD
Loeskeobryum brevifrostre_CHS	71	ALKGGFRFRD	EKSGTKK	175	VMMWQGGCFPGATVTRVAKD
Climacium dendroides_CHS	71	ALKGGFRFRD	EKSGTKK	177	VMMWQGGCFPGATVTRVAKD
Fontinalis antipyretica_CHS	71	ALKGGFRFRD	EKSGTKK	173	VMMWQGGCFPGATVTRVAKD
Claopodium rostratum_CHS	71	ALKGGFRFRD	EKSGTKK	177	VMMWQGGCFPGATVTRVAKD
Thuidium delicatulum_CHS	71	ALKGGFRFRD	EKSGTKK	177	VMMWQGGCFPGATVTRVAKD
Stereodon subimponens_CHS	71	ALKGGFRFRD	EKSGTKK	177	VMMWQGGCFPGATVTRVAKD
Selaginella moellendorffii_CHS	46	ELKFKKFRFRD	OKDSGLTKK	152	VMMWQGGCFPGATVTRVAKD
Selaginella stantoniiana_CHS	46	ELKFKKFRFRD	OKDSGLTKK	152	VMMWQGGCFPGATVTRVAKD
Selaginella lepidophylla_CHS	46	DLKFKKFRFRD	OKDSGLTKK	152	VMMWQGGCFPGATVTRVAKD
Selaginella apoda_CHS	46	ELKSKFRFRD	OKDSGLTKK	152	VMMWQGGCFPGATVTRVAKD
Selaginella willdenowii_CHS	45	ELKGGFRFRD	OKDSGLTKK	151	VMMWQGGCFPGATVTRVAKD
Selaginella kraussiana_CHS	46	ALKSKFRFRD	OKDSGLTKK	152	VMMWQGGCFPGATVTRVAKD
Selaginella acanthonota_CHS	35	ELKFKKFRFRD	OKDSGLTKK	141	VMMWQGGCFPGATVTRVAKD
Selaginella wallacei_CHS	35	ELKFKKFRFRD	OKDSGLTKK	141	VMMWQGGCFPGATVTRVAKD
Selaginella selaginoides_CHS	37	DMKSKFRFRD	OKDSGLTKK	143	VMMWQGGCFPGATVTRVAKD
Isoetes sp CHS	40	ELKGGFRFRD	OKDSGLTKK	146	VMMWQGGCFPGATVTRVAKD
Isoetes tegetiformans_CHS	40	ELKGGFRFRD	OKDSGLTKK	146	VMMWQGGCFPGATVTRVAKD
Pseudolycopodiella caroliniana_CHS	37	ELKFKKFRFRD	OKDSGLTKK	143	VMMWQGGCFPGATVTRVAKD
Lycopodiella appressa_CHS	37	DLKFKKFRFRD	OKDSGLTKK	143	VMMWQGGCFPGATVTRVAKD
Lycopodium annotinum_CHS	37	ELKFKKFRFRD	OKDSGLTKK	143	VMMWQGGCFPGATVTRVAKD
Diphasiastrum digitatum_CHS	37	ELKFKKFRFRD	OKDSGLTKK	143	VMMWQGGCFPGATVTRVAKD
Dendrolycopodium obscurum_CHS	37	ELKFKKFRFRD	OKDSGLTKK	143	VMMWQGGCFPGATVTRVAKD
Huperzia myrsinoides_CHS	37	ELKFKKFRFRD	OKDSGLTKK	143	VMMWQGGCFPGATVTRVAKD
Huperzia lucidula_CHS	37	ELKFKKFRFRD	OKDSGLTKK	143	VMMWQGGCFPGATVTRVAKD
Equisetum arvense_CHS	57	ELKFKKFRFRD	OKDSGLTKK	163	VMMWQGGCFPGATVTRVAKD
Psilotum nudum_CHS	49	ELKFKKFRFRD	EKSGTKK	155	VMMWQGGCFPGCTAVTRVAKD
Ceratopteris thalictroides_CHS	57	DLKFKKFRFRD	OKDSGLTKK	163	VMMWQGGCFPGCTAVTRVAKD
Dryopteris erythrosora_CHS	59	DLKFKKFRFRD	OKDSGLTKK	165	VMMWQGGCFPGCTAVTRVAKD
Pinus sylvestris_CHS	56	ELKFKKFRFRD	OKDSGLTKK	162	VMMWQGGCFPGCTAVTRVAKD
Stangeria eriopus_CHS	51	ELKFKKFRFRD	EKSGTKK	157	VMMWQGGCFPGCTAVTRVAKD
Eucephalartos barberi_CHS	51	ELKFKKFRFRD	EKSGTKK	157	VMMWQGGCFPGCTAVTRVAKD
Cycas micholitzii_CHS	48	ELKFKKFRFRD	EKSGTKK	154	VMMWQGGCFPGCTAVTRVAKD
Ginkgo biloba_CHS	44	ELKFKKFRFRD	EKSGTKK	150	VMMWQGGCFPGCTAVTRVAKD
Gnetum montanum_CHS	45	ELKFKKFRFRD	OKDSGLTKK	151	VMMWQGGCFPGCTAVTRVAKD
Welwitschia mirabilis_CHS	50	DLKFKKFRFRD	OKDSGLTKK	156	VMMWQGGCFPGCTAVTRVAKD
Elaeis guineensis_CHS2	53	DLKFKKFRFRD	OKDSGLTKK	159	VMMWQGGCFPGCTAVTRVAKD
Sorghum bicolor_CHS	55	ELKFKKFRFRD	OKDSGLTKK	161	VMMWQGGCFPGCTAVTRVAKD
Phoenix dactylifera_CHS1	51	DLKFKKFRFRD	OKDSGLTKK	157	VMMWQGGCFPGCTAVTRVAKD
Setaria italica_CHS2	55	DLKFKKFRFRD	OKDSGLTKK	161	VMMWQGGCFPGCTAVTRVAKD
Musa acuminata_subsp_malaccensis_CHS2-like	51	ELKFKKFRFRD	OKDSGLTKK	154	VMMWQGGCFPGCTAVTRVAKD
Zea mays_CHS	55	DLKFKKFRFRD	OKDSGLTKK	161	VMMWQGGCFPGCTAVTRVAKD
Brachypodium distachyon_CHS	55	DLKFKKFRFRD	OKDSGLTKK	161	VMMWQGGCFPGCTAVTRVAKD
Setaria italica_CHS1-like	54	ELKFKKFRFRD	OKDSGLTKK	160	VMMWQGGCFPGCTAVTRVAKD
Oryza sativa_Japonica_Group_CHS2	55	NLKFKKFRFRD	OKDSGLTKK	161	VMMWQGGCFPGCTAVTRVAKD
Arabidopsis thaliana_CHS	56	DLKFKKFRFRD	OKDSGLTKK	162	VMMWQGGCFPGCTAVTRVAKD
Arabidopsis halleri_CHS	51	ELKFKKFRFRD	OKDSGLTKK	157	VMMWQGGCFPGCTAVTRVAKD
Medicago sativa_CHS	51	ELKFKKFRFRD	OKDSGLTKK	157	VMMWQGGCFPGCTAVTRVAKD
Camelina sativa_CHS	56	DLKFKKFRFRD	OKDSGLTKK	162	VMMWQGGCFPGCTAVTRVAKD
Brassica oleracea_var_oleracea_CHS3	56	DLKFKKFRFRD	OKDSGLTKK	162	VMMWQGGCFPGCTAVTRVAKD
Eutrema salsugineum_CHS	56	DLKFKKFRFRD	OKDSGLTKK	162	VMMWQGGCFPGCTAVTRVAKD
Glycine max_CHS_J-like	51	DLKFKKFRFRD	OKDSGLTKK	157	VMMWQGGCFPGCTAVTRVAKD
Populus euphratica_CHS	52	ELKFKKFRFRD	OKDSGLTKK	158	VMMWQGGCFPGCTAVTRVAKD
Citrus sinensis_CHS1_isoform_X1	51	DLKFKKFRFRD	OKDSGLTKK	157	VMMWQGGCFPGCTAVTRVAKD
Theobroma cacao_CHS	53	ELKFKKFRFRD	OKDSGLTKK	159	VMMWQGGCFPGCTAVTRVAKD
Vitis vinifera_CHS	51	ELKFKKFRFRD	OKDSGLTKK	157	VMMWQGGCFPGCTAVTRVAKD
Populus trichocarpa_CHS	52	ELKFKKFRFRD	OKDSGLTKK	158	VMMWQGGCFPGCTAVTRVAKD
Solanum pennellii_CHS2	51	ELKFKKFRFRD	OKDSGLTKK	157	VMMWQGGCFPGCTAVTRVAKD
Nicotiana tomentosiformis_CHS2	51	ELKFKKFRFRD	OKDSGLTKK	157	VMMWQGGCFPGCTAVTRVAKD
Sesamum indicum_CHS	51	ELKFKKFRFRD	OKDSGLTKK	157	VMMWQGGCFPGCTAVTRVAKD
Erythranthe guttata_CHS	53	DLKFKKFRFRD	EKSGTKK	159	VMMWQGGCFPGCTAVTRVAKD
Capsicum annuum_CHS1B	51	ELKFKKFRFRD	EKSGTKK	157	VMMWQGGCFPGCTAVTRVAKD

Basal land plants

Euphyllphytes

Figure S5. Multiple sequence alignment of CHSs. Sequence numbers of the beginning of each block for each CHS sequence are indicated. Residues outlined in thin black boxes are conserved with > 70% similarity across all sequences. Residues with 100% conservation are in white text with a black background. Red boxes indicate the seven positions mutated in the AtCHS M7 and SmCHS M7 constructs; these positions are differentially conserved between euphyllphyte and basal-plant CHSs, which are divided by the horizontal red line.

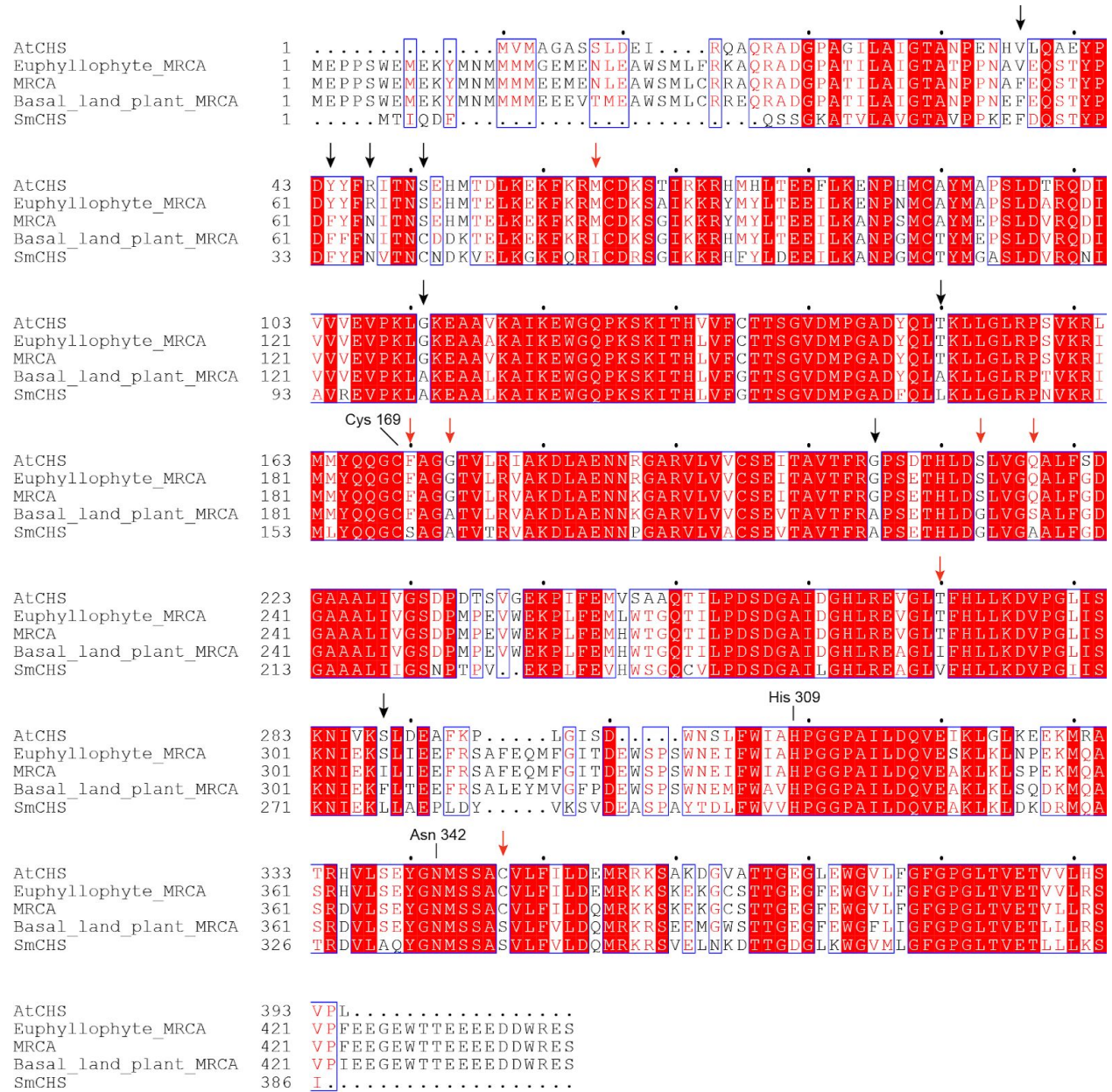


Figure S6. CHS ancestral sequence reconstruction. Sequences and phylogenetic tree of CHSs shown in Figure 1 were used to perform ancestral sequence construction with FastML. The most recent common ancestor (MRCA) sequences of all branches, euphylllophyte, and basal land plant clades are compared to AtCHS and SmCHS. Among the five sequences shown, absolutely conserved residues are shown in white text with red background. Residues with > 70% similarity are shown in red text and white background and blue outline. Other residues are shown in black text. Red arrows indicate the seven differentially conserved positions previously identified and mutated in the M7 CHS constructs. Black arrows indicate additional residue positions that are differentially conserved between euphylllophyte and basal-plant CHSs and determined to have possible functional impact based on their position in the CHS crystal structure. The catalytic triad residues are also labeled.

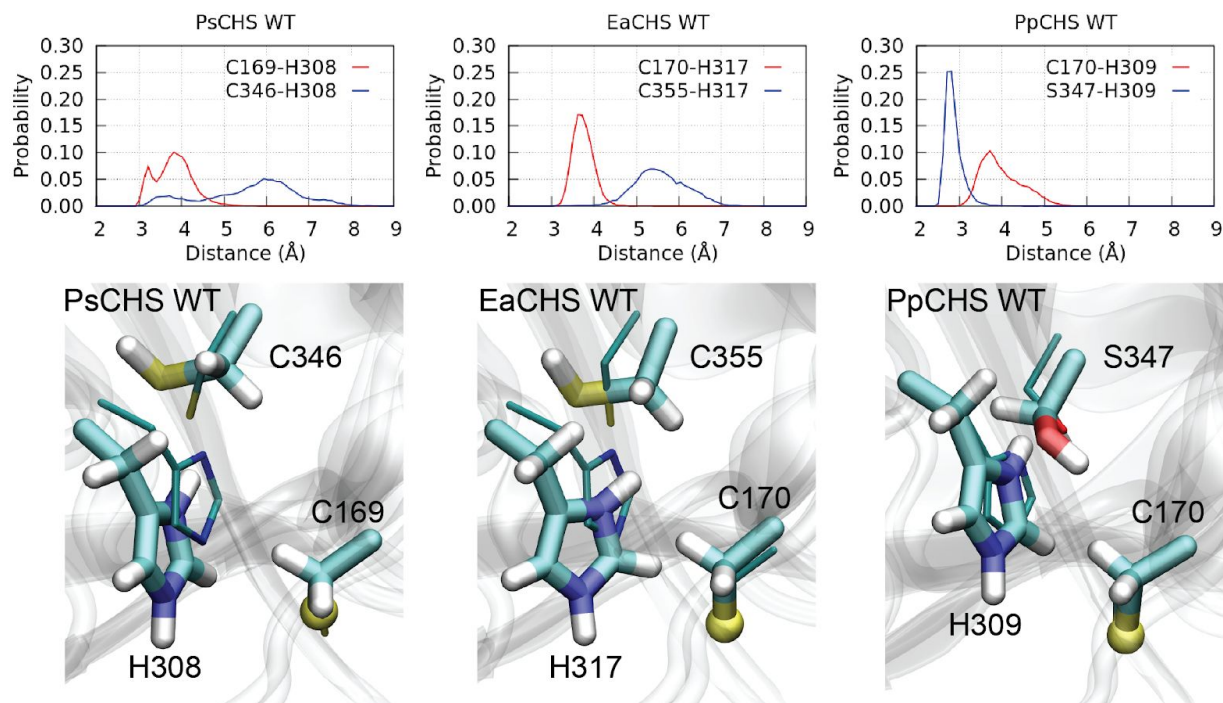


Figure S7. Distributions of inter-residue distances and the largest cluster conformations of EaCHS, PpCHS, PsCHS obtained from MD simulations. The observation of a serine forming a more stable hydrogen bond interaction than cysteine with the catalytic histidine is similar to the AtCHS and SmCHS wild-type and mutant simulations (Figure 5). Notably, with the rather weak interaction between the cysteine C346/C355 and the catalytic histidine, the latter moves more freely and often shows a much larger displacement from the corresponding position in the crystal structure (thin sticks).

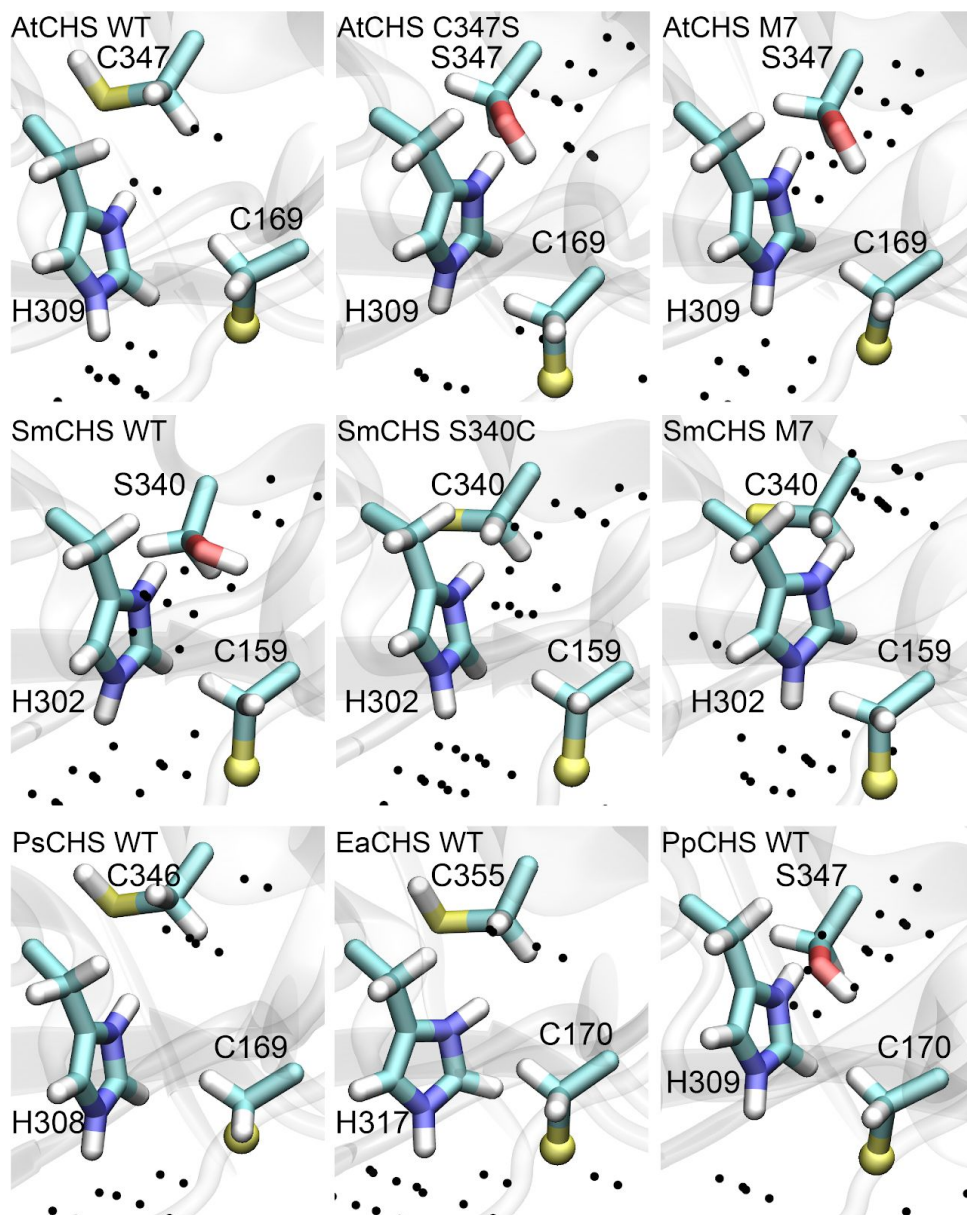


Figure S8. Average occupancy of water molecules obtained from MD simulations. Black dots represent grid points with an average water occupancy greater than 0.2. SmCHS in general has more water inside the active site, while the wild-type AtCHS has fewer water molecules. AtCHS mutants gradually attract more water around S347. This pattern is also observed in PpCHS, which also attracts more water around its serine than CHS where the serine is replaced by a cysteine (EaCHS, PsCHS). See also Supporting Note below.

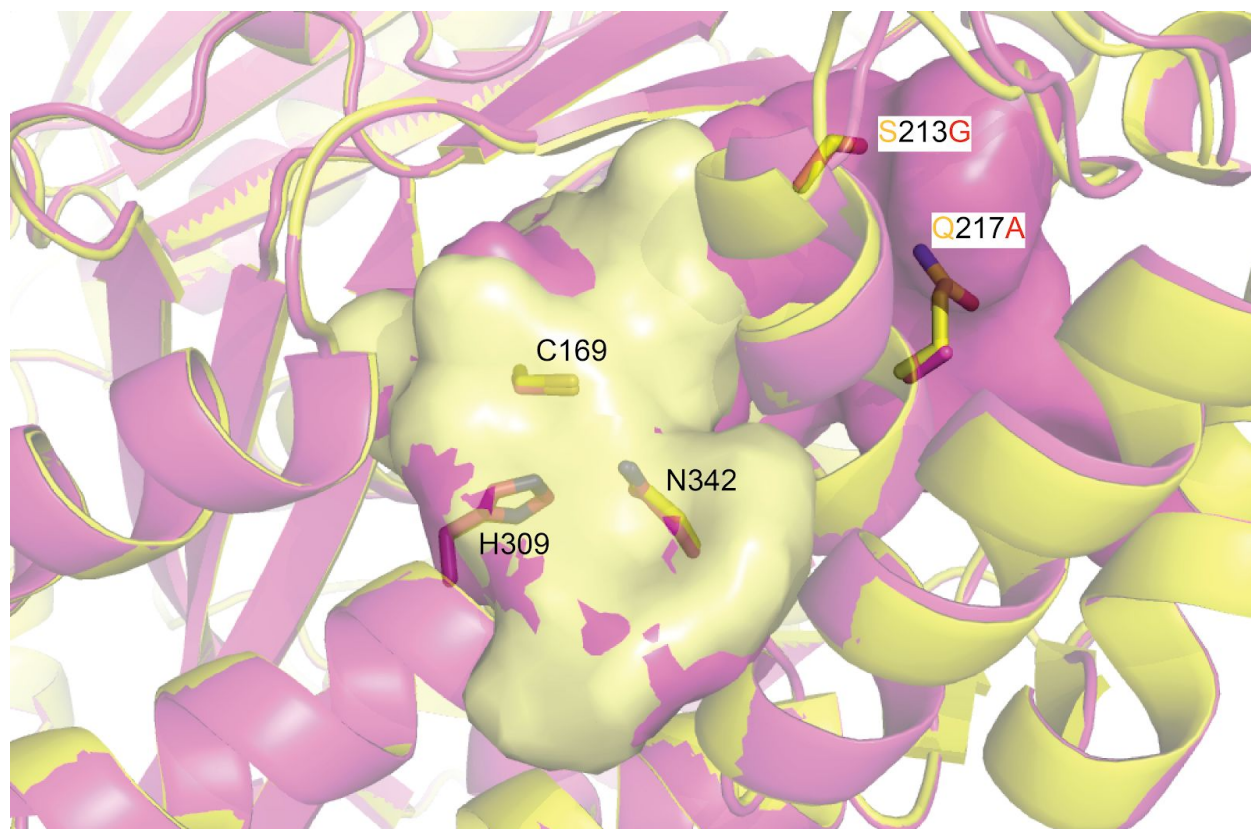


Figure S9. Comparison of wild-type AtCHS (yellow) and AtCHS M7 (yellow) crystal structures. The catalytic triad residues and two of the seven mutations from wild-type to M7 are modelled as sticks and labeled. The yellow and magenta surfaces represent the solvent-accessible cavities measured using the cavity-finding program KVFinder. The helix containing the two marked mutations is shifted in AtCHS M7 compared to wild type, leading to a larger active-site cavity.

Supporting Note

Our MD calculations show that the C347S substitution (AtCHS numbering) can significantly affect active-site solvation. The occupancy of water molecules within the active site was measured with a resolution of 1 \AA^3 (Figure S8). Interestingly, S347 in AtCHS C347S and M7 mutants attracts more water toward itself and H309. Similarly, the wild-type SmCHS is also considerably wetter than the wild-type AtCHS: employing a cylinder with a radius of 9 \AA and a height of 13 \AA to enclose the catalytic residues, we found that the average number of water molecules enclosed was 40.0 for SmCHS and 31.4 for AtCHS. The ability of a serine to attract more water is also observed in simulations of EaCHS, PpCHS, and PsCHS, although in SmCHS mutants the active site remains rather wet despite the mutation of serine to cysteine (Figure S8).

AtCHS M7 also showed a wider active-site opening than wild-type AtCHS, which may also affect solvent access to the active site, as shown by the large cavity found in cavity analysis. In addition to changing the hydrogen bonding network, the decreased solvation in euphyllophyte CHSs would enhance the pK_a -lowering effect of the histidine, because ionic effects are enhanced as the dielectric constant decreases along with solvent polarity (1).

Supporting References

1. Harris, T. K., and Turner, G. J. (2002) Structural basis of perturbed pK_a values of catalytic groups in enzyme active sites. *IUBMB Life*. **53**, 85–98