Supporting Information

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

Geoffrey Liou[1,2], Ying-Chih Chiang[3], Yi Wang[3], and Jing-Ke Weng[1,2,*]

From the [1]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA; [2]Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA; [3]Department 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

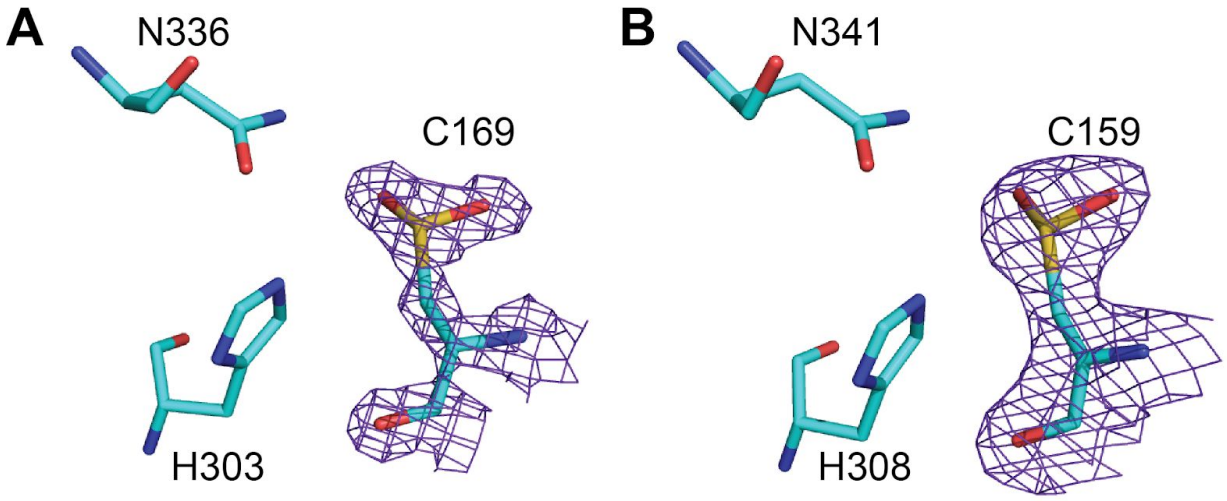

Figure S1. Active site structures of *Medicago sativa* CHS (**A**) (PDB ID 1BI5) and *Gerbera hybrida* 2-pyrone synthase (**B**) (PDB ID 1QLV) showing catalytic cysteine oxidized to sulfinic acid. The $2F_o$-F_c composite map contoured at 1.5σ is shown around the catalytic cysteine.

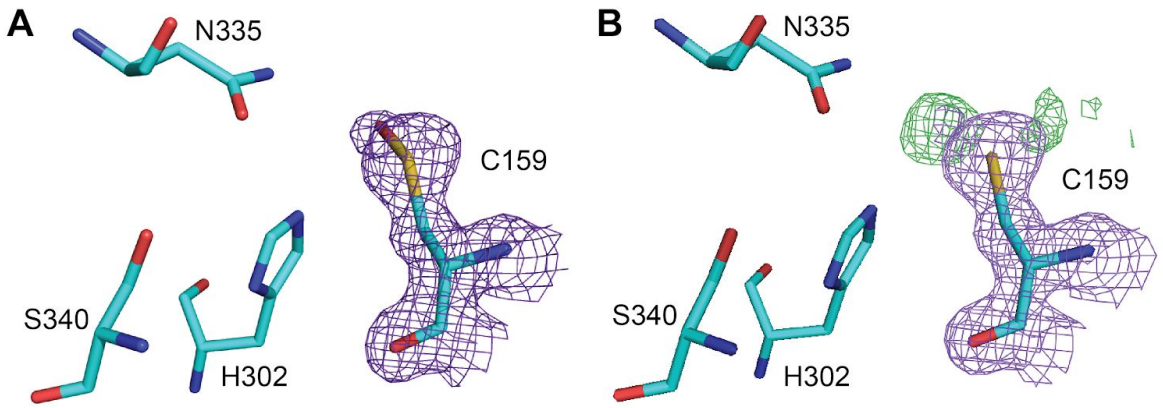

Figure S2. Active site structure of SmCHS crystals soaked in 1 mM hydrogen peroxide for 75 min. **A**, The $2F_o$-F_c composite map to 1.55 Å resolution and contoured at 1.5σ is shown around the catalytic cysteine, modeled as oxidized to sulfenic acid. **B**, The $2F_o$-F_c composite map to 1.55 Å resolution 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_2O_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.

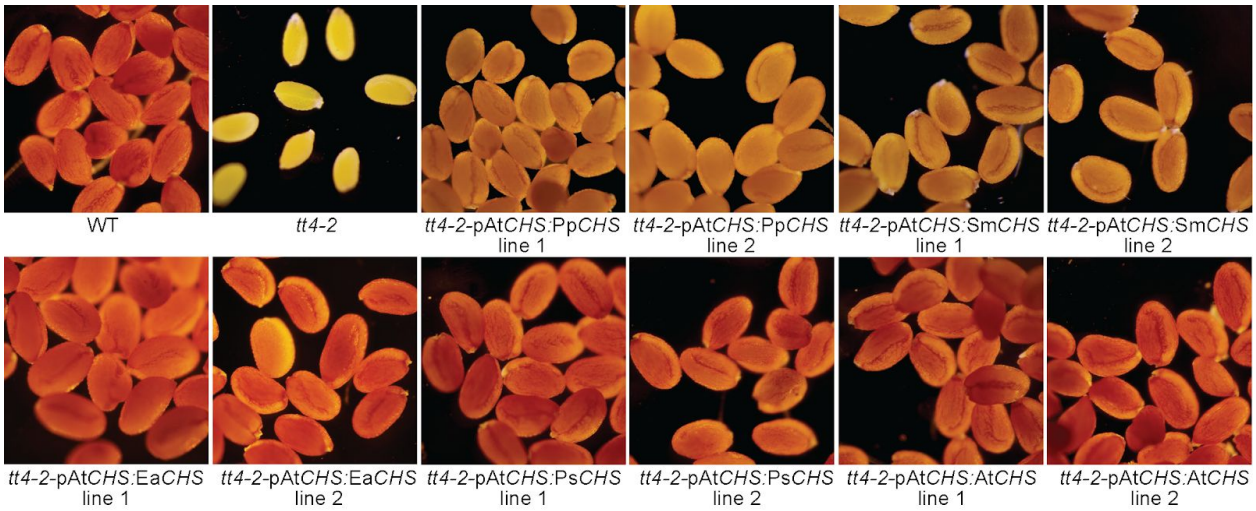

Figure S3. Complementation of the *transparent testa* seed phenotype of *tt4-2* mutant *Arabidopsis thaliana*. CHS orthologs were expressed under the At*CHS* promoter. CHS from euphyllophytes (AtCHS, PsCHS, EaCHS) fully complement the mutant phenotype, whereas CHS from basal land plants (SmCHS, PpCHS)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

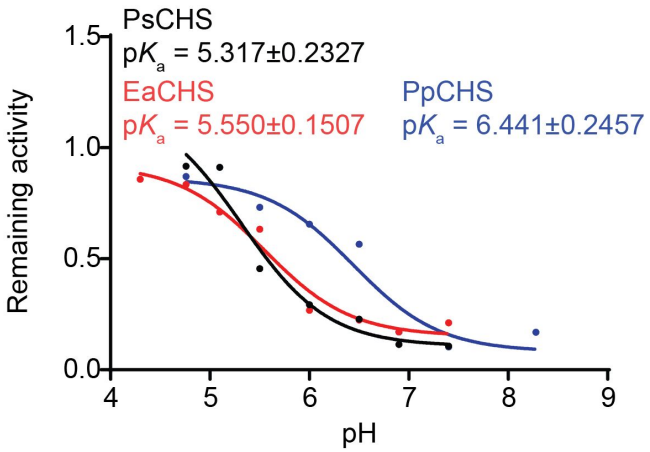

Figure S4. pK_a measurement of PsCHS, EaCHS, and PpCHS wild type enzymes. CHS enzyme was pre-incubated at various pH in the 25 µM iodoacetamide inhibitor or water (control) for 30 s, and an aliquot was taken to run in a CHS activity assay. The ratio of naringenin product produced in the iodoacetamide 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 euphyllophyte CHSs, whereas the pK_a of PpCHS is over 1 pH unit higher, similar to that of SmCHS.

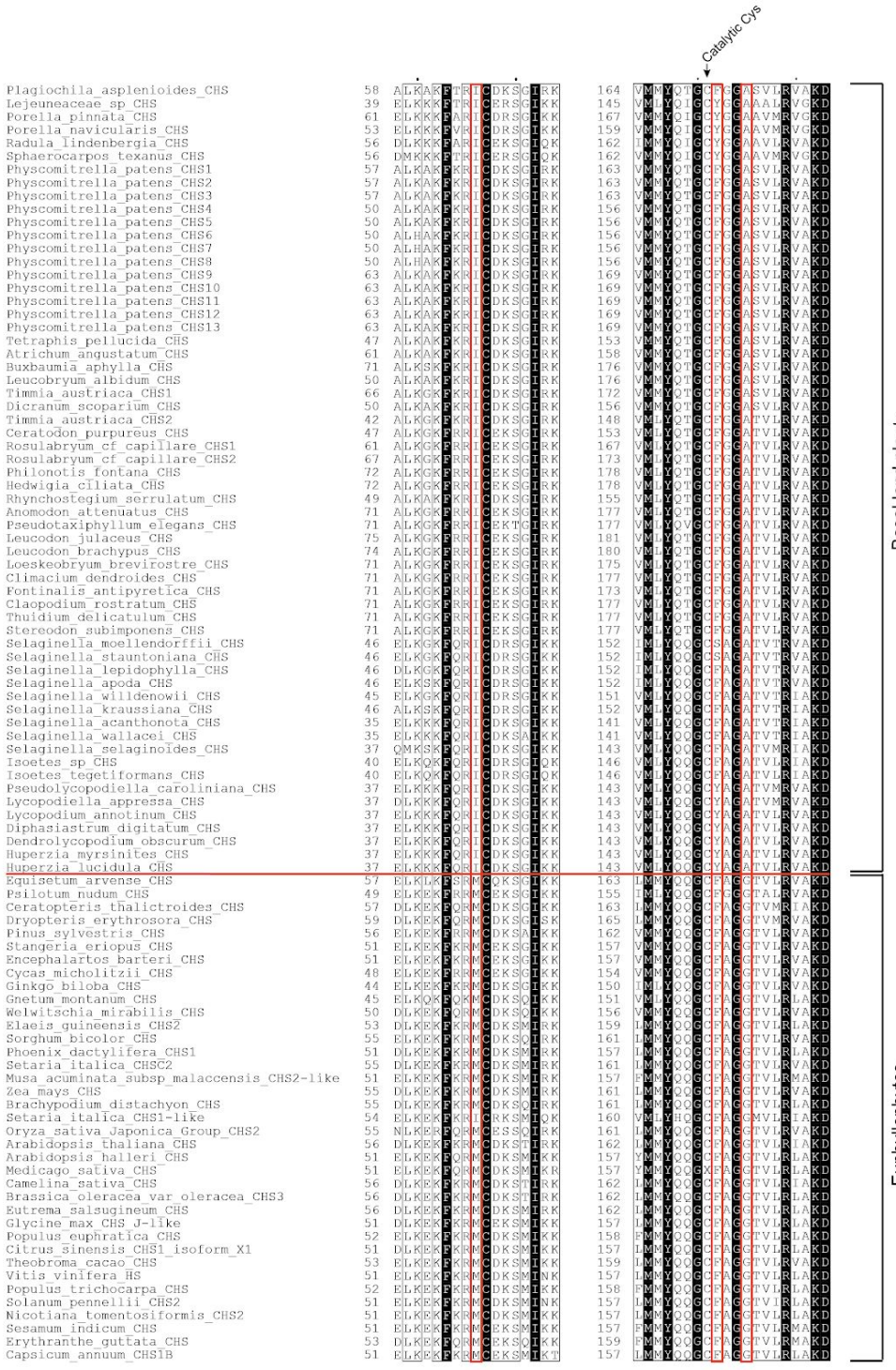

Figure S5. Multiple sequence alignment of CHSs. Sequence numbers of the beginning of each block for each CHS sequence are indicated. Residues outlined in thin black boxes are conserved with > 70% similarity across all sequences. Residues with 100% conservation are in white text with a black background.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

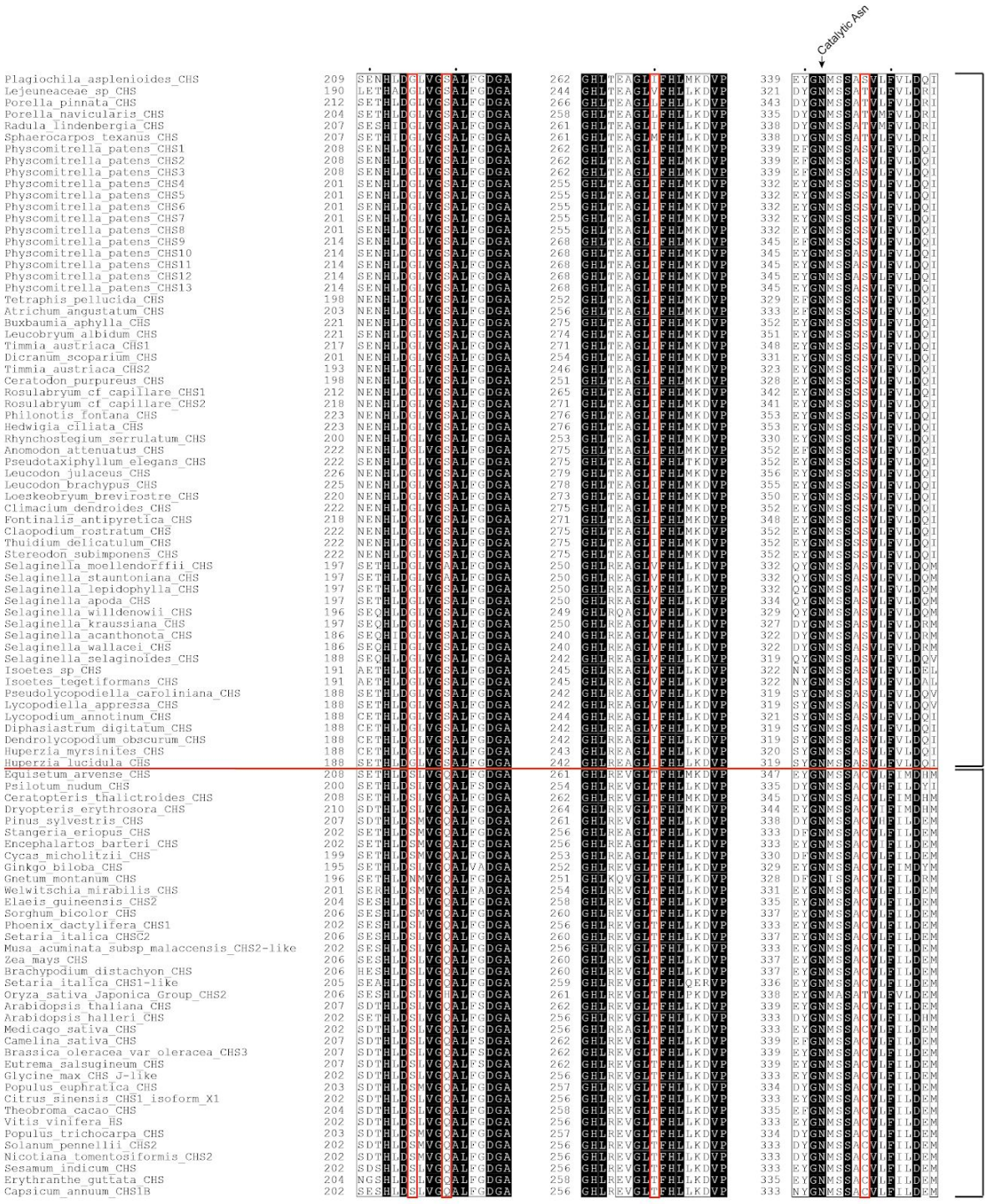

Figure S5 continued

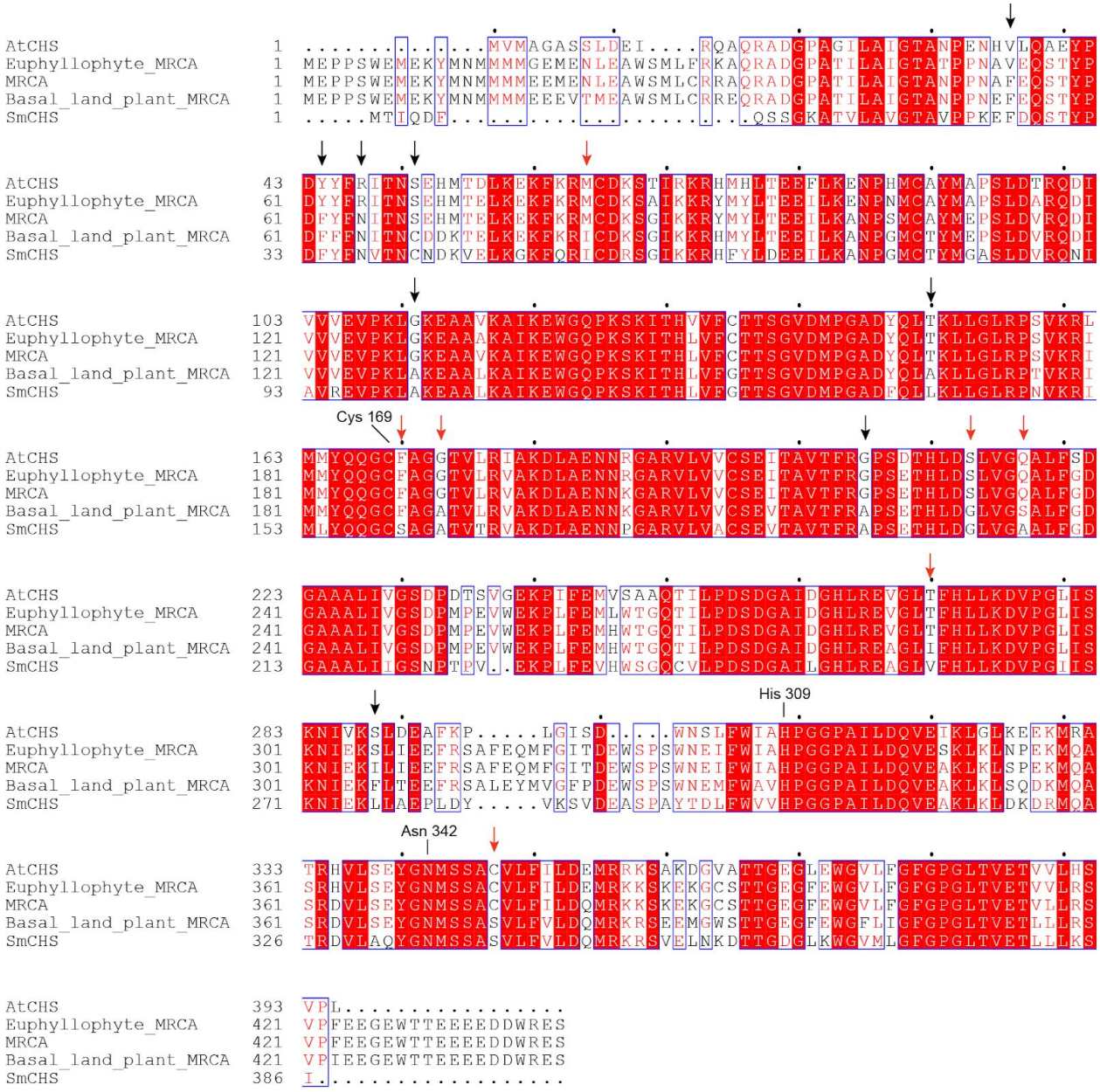

Figure S6. CHS ancestral sequence reconstruction. Sequences and phylogenetic tree of CHSs shown in Figure 1 were used to perform ancestral sequence construction with FastML. The most recent common ancestor (MRCA) sequences of all branches, euphyllophyte, and basal land plant clades are compared to AtCHS 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 euphyllophyte 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.

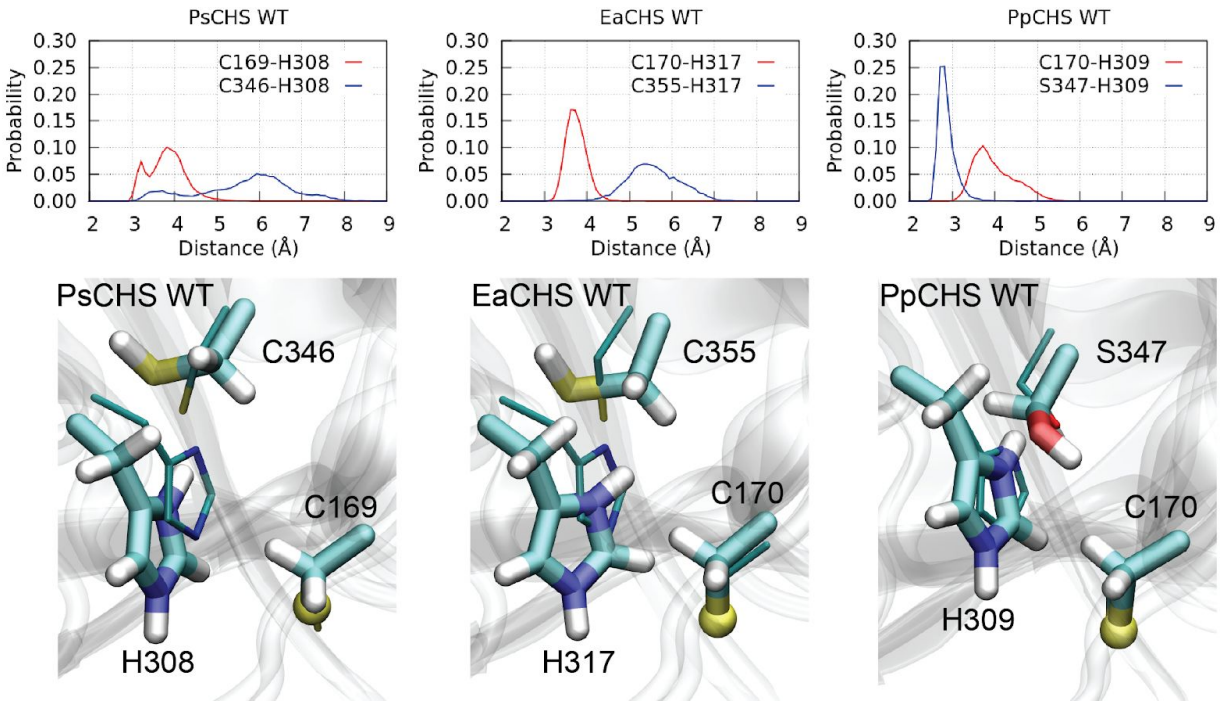

Figure S7. Distributions of inter-residue distances and the largest cluster conformations of EaCHS, PpCHS, PsCHS obtained from MD simulations. The observation of a serine forming a more stable hydrogen bond interaction than cysteine with the catalytic histidine is similar to the AtCHS and SmCHS wild-type 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).

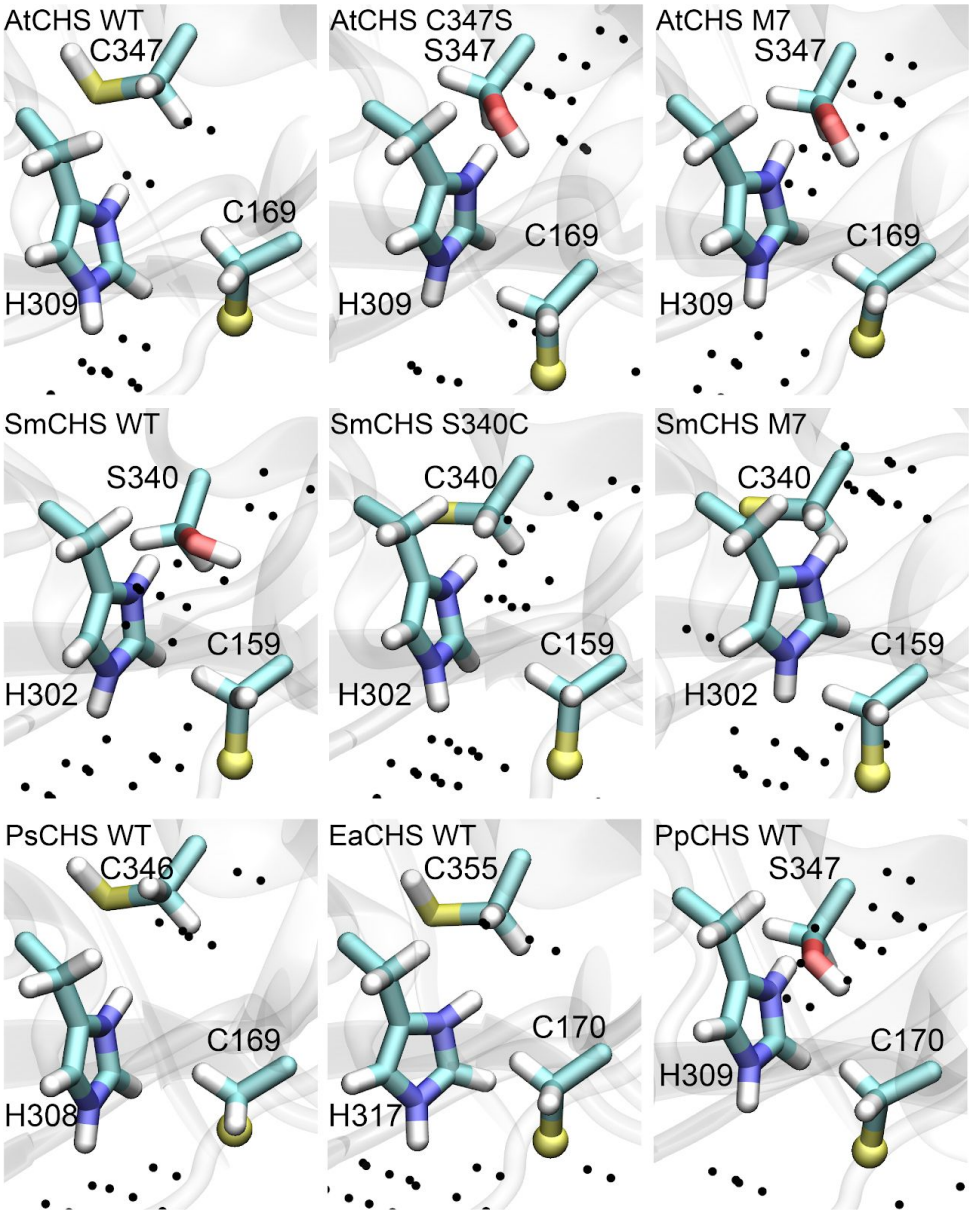

Figure S8. Average occupancy of water molecules obtained from MD simulations. Black dots represent grid points with an average water occupancy greater than 0.2. SmCHS in general has more water inside the active site, while the wild-type AtCHS has fewer water molecules. AtCHS mutants gradually attract 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.

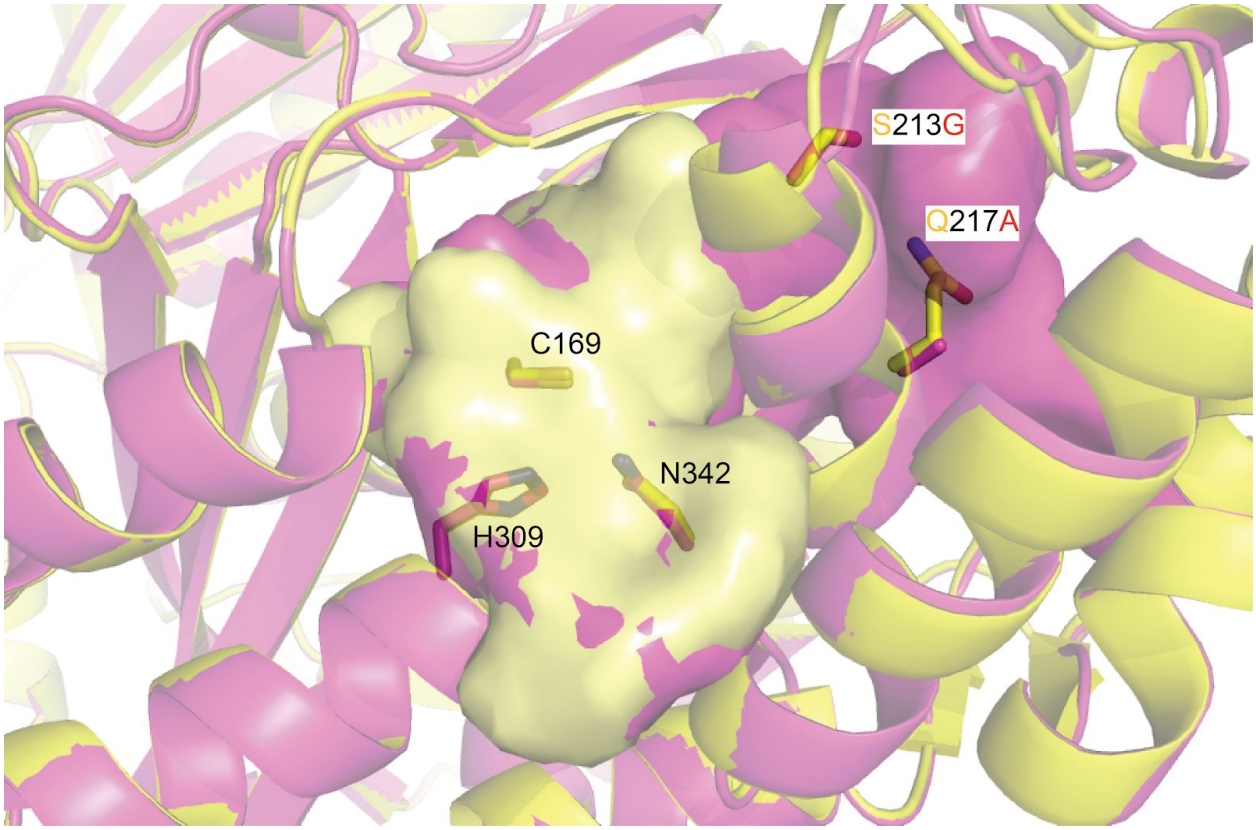

Figure S9. Comparison of wild-type AtCHS (yellow) and AtCHS M7 (yellow) crystal structures. The catalytic triad residues and two of the seven mutations from wild-type to M7 are modelled as sticks and labeled. The yellow and magenta surfaces represent the solvent-accessible cavities measured using the 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 $Å^3$ (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 Å and a height of 13 Å 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 pKa values of catalytic groups in enzyme active sites. *IUBMB Life*. **53**, 85–98