Supplemental Materials

Biophysical characterization of full-length human phenylalanine hydroxylase provides a deeper understanding of its quaternary structure equilibrium

Emilia Arturo[1,2], Kushol Gupta[3,] Michael R. Hansen[1], Elias T. Borne[1], Eileen K. Jaffe[1,*]

From [1]The Department of Molecular Therapeutics, Fox Chase Cancer Center, Temple University Health Systems, Philadelphia, P.A. 19111, USA, [2]The Department of Biochemistry and Molecular Biology, Drexel University College of Medicine, Philadelphia, P.A. 19102, USA, and [3]The Department of Biochemistry and Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, P.A. 19104, USA.

Supplemental Methods.

Protein Expression and Purification. *E. coli* BL21(DE3) containing plasmids pGro7 (1) and pMRH161 (encoding 6His-PAH-SUMO-hPAH-C29S) was grown overnight at 37 °C in 2xYT medium supplemented with 100 µg/mL ampicillin, 25 µg/mL chloramphenicol and 1% glucose. The overnight culture was used to inoculate 6 L 2x YT medium supplemented with 100 µg/mL ampicillin, 25 µg/mL chloramphenicol, 3.33 mM arabinose, $FeCl_3$ (10 or 100 µM), 4 mM $MgSO_4$ and grown at 25 °C. Protein expression was induced by addition of 0.5 mM IPTG at OD_{600} of 0.5 and grown overnight at 18 °C. Cells were harvest after 22 h and stored at -80 °C. The cell pellet (~20 g) was thawed and resuspended in (120 ml) 20 mM sodium phosphate (pH 7.4), 0.5 M NaCl, 20 mM imidazole (Buffer Ni-20). Supernatant of lysed cells was applied onto a Ni Sepharose HP (GE) 25 mL column pre-equilibrated with Buffer Ni-20. After extensively washing with Buffer Ni-20, the fusion protein was eluted in 20 mM sodium phosphate (pH 7.4), 0.5 M NaCl, 500 mM imidazole (Buffer Ni-500). Fractions containing the fusion protein (~85 mg) were pooled and cleaved in the elution buffer by the addition of 85 µg of the catalytic domain of the *S. cerevisiae* SUMO hydrolase (2) overnight (16 h) at 4 °C during dialysis (12-14 kDa cut-off range) against 20 mM sodium phosphate (pH 7.4), 0.5 M NaCl (buffer Ni-0). The digested dialyzed protein was reloaded onto the Ni Sepharose HP column pre-equilibrated with Buffer Ni-20. De-tagged protein came out in the flow-through and Buffer Ni-20 wash. The pooled de-tagged protein was dialyzed overnight at 4 °C against Buffer Q-0 (30 mM Tris-HCl pH 7.4, 15% glycerol). Dialyzed protein was applied on a HiTrap Q HP column (1 mL) and protein was eluted using a linear gradient of elution buffer (0 mM to 400 mM KCl over a 30-column volume) and 1.5 mL fractions were collected. Fractions containing C29S were estimated by SDS-PAGE to be >95% pure. We note that a more reproducible yield of high activity C29S was obtained when the Ni Sepharose steps were carried out using Tris buffer in place of the phosphate buffer. Native C29S was expressed in *E. coli* Rosetta(DE3) cells transformed with plasmid pAHhum-C29S. Expression method was identical as described above except that growth media containing 10 uM $FeCl_3$ instead of 0.1 mM $FeCl_3$. The protein was purified using the Shiman method (3) followed by purification on a HiTrap Q HP column (1 mL) as described above.

WT hPAH was expressed in BL21(DE3) transformed with plasmid pMRH160 (encoding 6His-SUMO-hPAH) and purified as previously described (4). C29S was expressed in BL21(DE3) transformed with plasmid pMRH174 (encoding 6His-TEV-C29S) and purified as previously described (4) except that 6His-tagged TEV protease (SIGMA) in a ratio 1:70 was used in the digestion mixture instead of 6His-tagged SUMO protease.

Note that the preparation of C29S that yielded this structure used the phosphate buffer Ni-NTA affinity purification scheme described above, and resulted in protein with reduced specific activity, a normal K_M value, and overall identical chromatographic retention profiles and intrinsic fluorescence with and without preincubation with Phe (data not shown). For this reason, it is notable that the *CheckMyMetal* (CMM) server (5) suggests alternative metals (cobalt and zinc) to the modeled iron at all but one active site of the tetramer. Although "experimental", CMM suggests cobalt with highest confidence as an alternative to iron. We found that dialysis against the metal chelator EDTA followed by addition of iron restored activity to similar preparations of rPAH, consistent with the notion that native iron was displaced by some other metal during purification using phosphate buffer. As the nature of the metal is unknown, and iron at this resolution is an equally acceptable metal candidate to any of the CMM predicted alternative metals, iron remains a part of the

model.
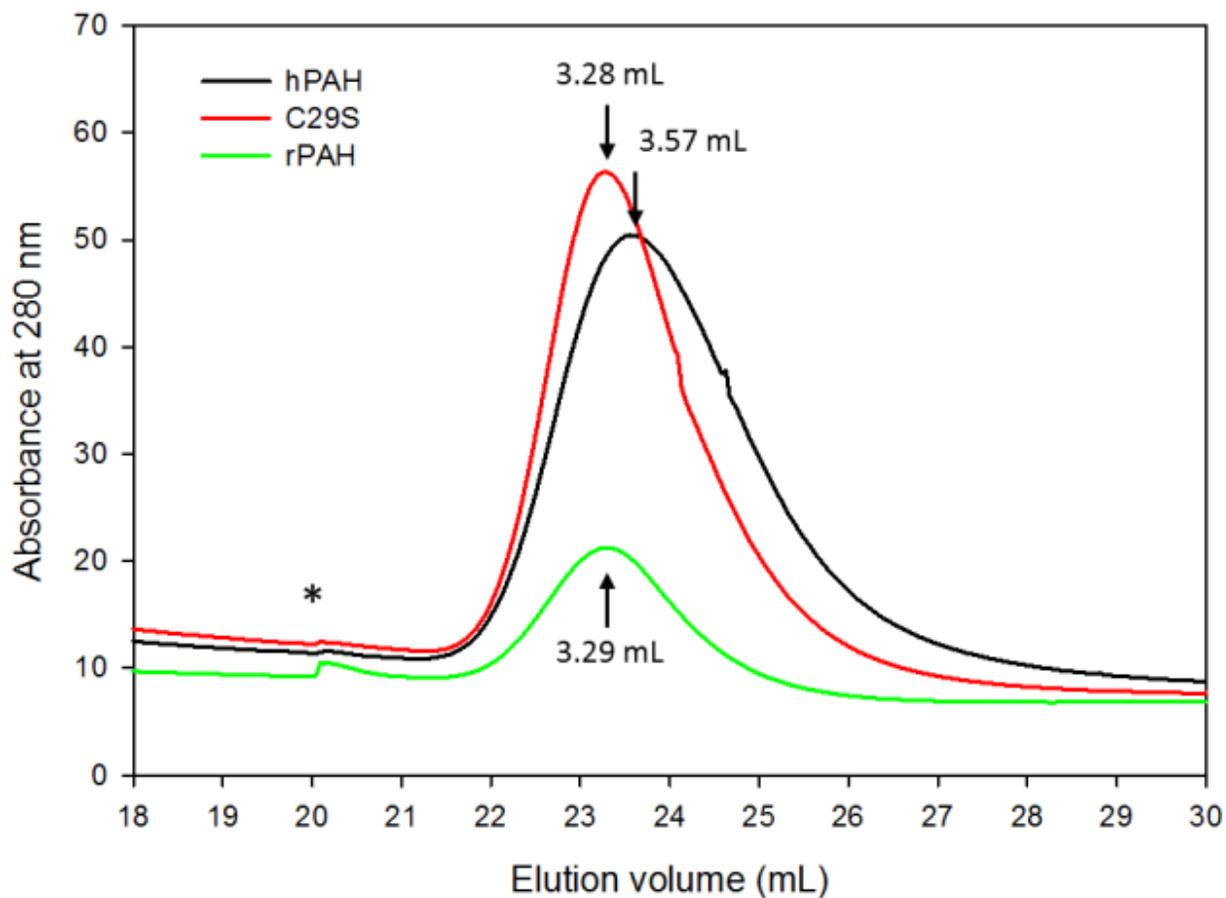
Figure S1. *Analytical elution profiles of hPAH, C29S and rPAH in complex with 5 mM Phe on a 1 mL Phenyl-Sepharose column*. Protein is loaded and washed in 30 mM Tris-HCl, pH 7.25, 0.2 M KCl, 5 mM Phe. Column is then washed with this same buffer, without the KCl but including + 0.037% Tween-80. Finally, protein is eluted with the Tris/tween buffer without Phe.

A

No Phe
Dimer-Tetramer (4°C)
K_{d2-4}=18.8 μM ± 0.91 μM

B

1 mM Phe
Dimer-Tetramer (4°C)
K_{d2-4}=0.46 μM ± 0.82 μM
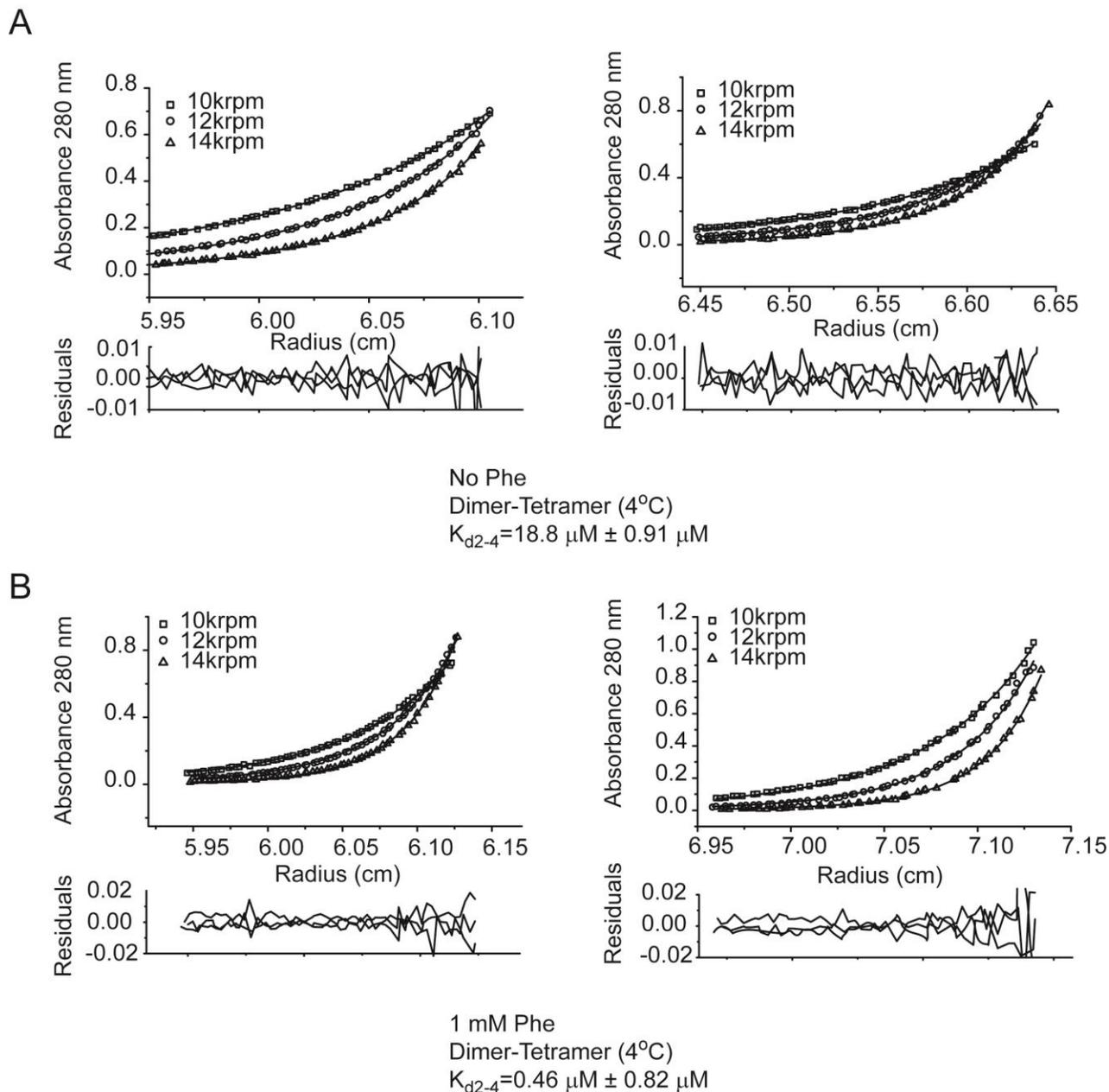
Figure S2. Global fit analysis of sedimentation equilibrium data forC29S indicates a dimer-tetramer equilibrium in solution that changes in response to the addition of 1 mM Phe at 4°C. Data in the absence **(A)** and presence of 1 mM Phe **(B)** were recorded at 10, 12, and 14 krpm at both 2.5 (left) and 5 μM (right) monomer concentrations and globally fit to derive an apparent dissociation constant. Errors for the dissociation constants reported were estimated using 1000 iterations of Monte Carlo analysis as implemented in the program SEDPHAT (6). Residuals shown reflect the discrepancy between model fits and the experimental data, with RMSD figures below 0.010.
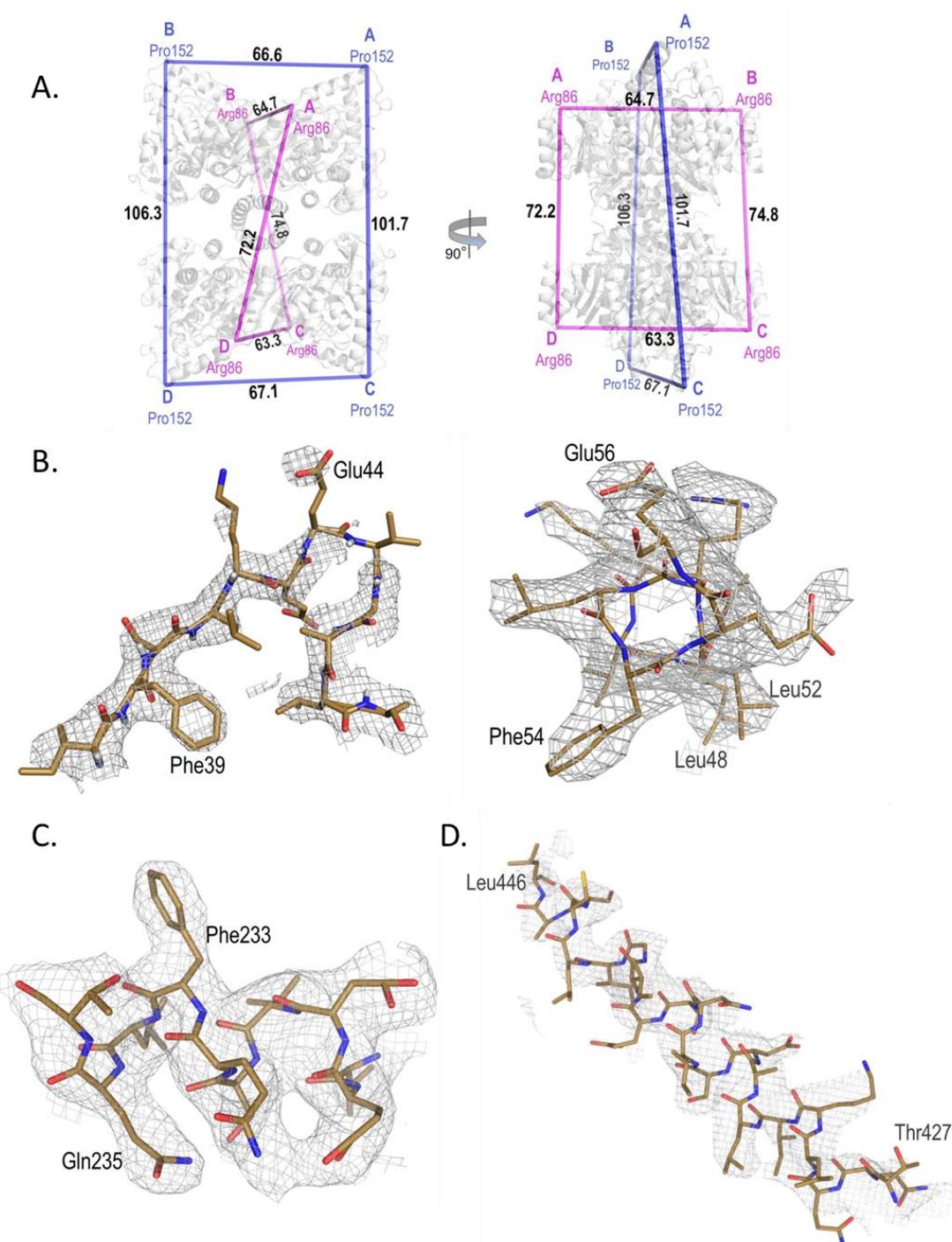
Figure S3. Characteristics of the crystal structure of C29S. (**A**) C29S is an asymmetric tetramer. All distances are shown in angstroms. (**B – D**) Representative electron density omit maps are contoured at 0.75σ, carving 1.6-1.8 Å about the residue selection. (**B**) The ACT subdomain is shown. *Left,* shown is *c*hain A, residues 28-49, which constitute a portion of β1 and all of L1. *Right,* shown is chain A, residues 48-56, which constitute α1. (**C**) Shown is chain A, residues 227-236, which constitute one helix interior to the catalytic domain. (**D**) Shown is chain C, residues 427-446, which is the final residue that could be modeled at this chain.
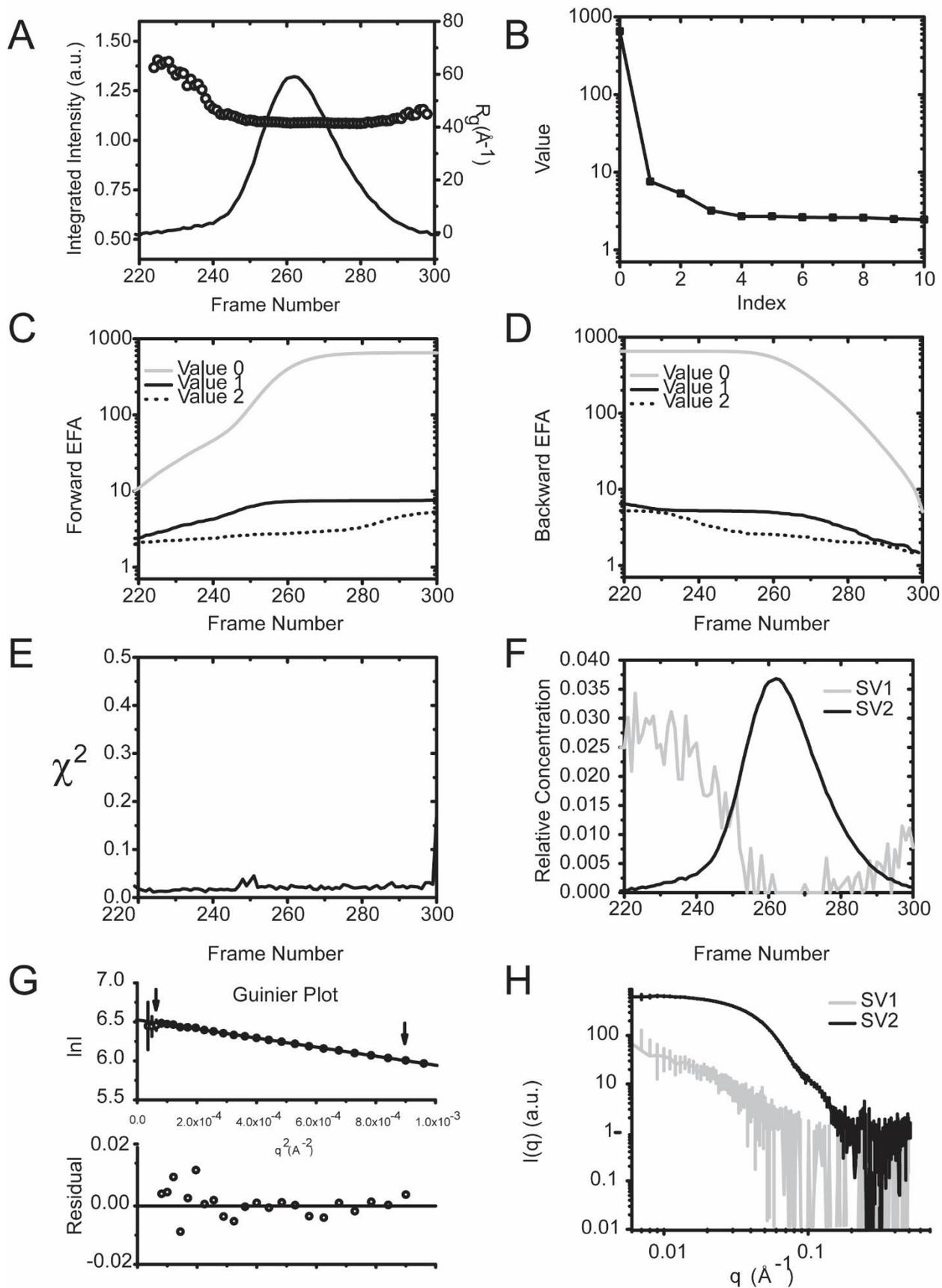
Figure S4. SEC-SAXS analysis of C29S in the resting state (-Phe). **(A)** Shown as a line is the integrated intensity of recorded small angle scattering as a function of frames recorded off a Superdex 200 10/300 column at 0.5 ml/min (1 s exposures). The intensity trace mirrors the traces obtained by UV and refractive index (RI) detection in-line (data not shown). Plotted as circles are the derived radii of gyration (R_g) for the background subtracted recorded profiles. R_g analysis suggests the presence of multiple species, as evidenced by the rising slope at the leading edge of the peak. **(B)** Conventional singular value decomposition (SVD) of SEC-SAXS data the existence of three significant components within the redundant dataset. Shown in panels **(C and D)** are the forward and backward evolving factor analyses of the SEC-SAXS dataset, showing the changes of singular values for each component as a function of frame number, corresponding to two species. **(E)** χ^2 is plotted as a function of frame number for the resulting model fits from the rotation method employed in the analysis (Hybrid, 1000 iterations, convergence threshold of $1e^{-12}$), showing uniformity in the solution. **(F)** Concentration plot shows the distribution of the two species (a higher order aggregate and a tetramer species). **(G)** Guinier plot analysis of the tetramer profile reveals linearity (top panel) and well distributed residuals (bottom panel). Arrows denote the region used for the analysis and the residuals shown reflect the agreement between the experimental data and Guinier fit. Error bars shown intensity shown here and in (H) reflect the instrumental uncertainty in recorded intensity. See **Table 3** for parameters derived from the analysis. **(H)** Resulting deconvoluted profiles are shown from the SVD-EFA analysis. The first component is a higher order aggregate and the second species corresponds to a tetramer of C29S, as indicated by mass calculations using Q_r (7) and Porod volume (8) relationships and correlation with atomic models (see **Fig 6**).
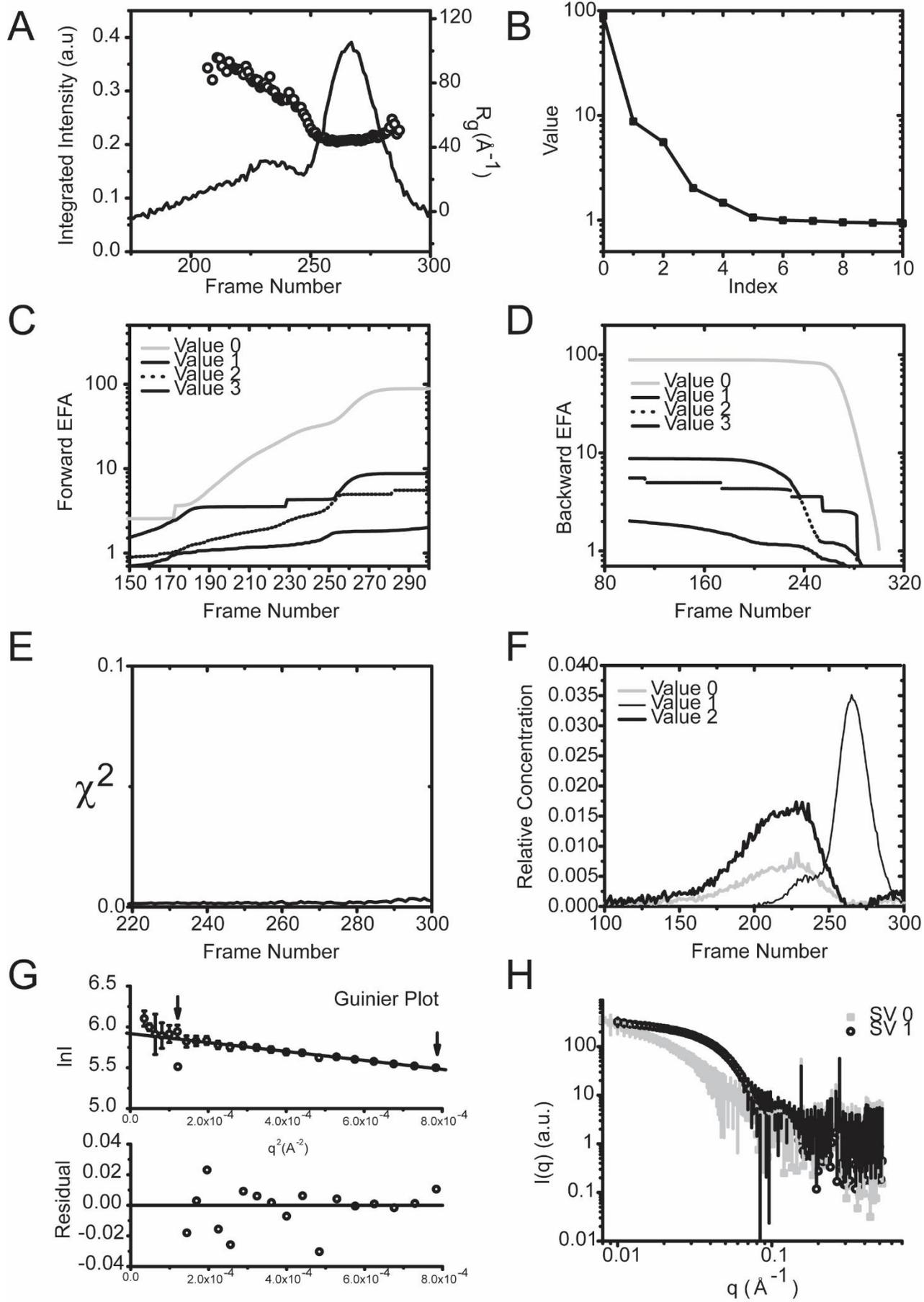
Figure S5. SEC-SAXS analysis of C29S in the activated state (+Phe). **(A)** Shown as a line is the integrated intensity of recorded small angle scattering as a function of frames recorded from a Superdex 200 10/300 column at 0.5 ml/min (1 s exposures). Plotted as circles are the derived radii of gyration (R_g) for the background subtracted recorded profiles. The intensity trace and R_g analysis suggests the presence of multiple species in a polydisperse sample, as evidenced by multiple peak species and the rising slope at the leading edge of the peak, mirroring traces obtained by UV and refractive index (RI) detection in-line (data not shown). **(B)** Conventional singular value decomposition (SVD) of SEC-SAXS data shows the existence of three significant components within the redundant dataset. Shown in panels **(C and D)** are the forward and backward evolving factor analyses of the SEC-SAXS dataset, showing the changes of singular values for each component as a function of frame number, corresponding to three species and a buffer mismatch. **(E)** χ^2 is plotted as a function of frame number for the resulting model fits from the rotation method employed in the analysis (Hybrid, 1000 iterations, convergence threshold of $1e^{-12}$), showing uniformity in the solution. **(F)** Concentration plot shows the distribution of the three species (a higher order aggregate, a buffer mismatch, and a tetramer species). **(G)** Guinier plot analysis of the tetramer profile reveals linearity (top panel) and well distributed residuals (bottom panel). Arrows denote the region used for the analysis and the residuals shown reflect the agreement between the experimental data and Guinier fit. Error bars shown intensity shown here and in (H) reflect the instrumental uncertainty in recorded intensity. See **Table 3** for parameters derived from the analysis. **(H)** Shown are resulting deconvoluted profiles from the SVD-EFA analysis for protein species. The first component is a higher order aggregate and the second species corresponds to a tetramer of hPAH, as indicated by the numerical analysis provided in Table 3 and results presented herein.

Table S1: Primers used in this study

Primer	Purpose
5'-AGATTGGTGGCTCCACTGCGGTCCTGGAAAAC-3'	PCR of hPAH from plasmid pEPAH1
5'-GAGGAGAGTTTAGACTTACTTTATTTTCTGGAGGGC-3'	PCR of hPAH from plasmid pEPAH1
5'-AACAAGCTATATTGAAGACAACAGCA-3'	QuikChange mutagenesis to created hPAH C29S (pMRH161)
5'-GATATGGCACCATTTTGATTGGTGTTG-3'	QuikChange mutagenesis to created hPAH C29S (pMRH161)
5'-AACAAGCTATATTGAAGACAACAGCA-3'	QuikChange mutagenesis to created hPAH C29S (pAHhum-C29S)
5'-GATATGGCACCATTTTGATTGGTGTTG-3'.	QuikChange mutagenesis to created hPAH C29S (pAHhum-C29S)
5'- CGAGAATCTTTATTTTCAAaGCATGTGAGACCGGATCC-3'	Gly-to-Ser mutation in TEV recognition site using QuikChange mutagenesis (pMRH173)
5'- GGATCCGGTCTCACATGCtTTGAAAATAAAGATTCTCG-3'	Gly-to-Ser mutation in TEV recognition site using QuikChange mutagenesis (pMRH173)
5'-phos-CACTGCGGTCCTGGAAAACCCAGGC-3'	PCR of hPAH from pMRH161
5'-agtGTCgacTTACTTTATTTTCTGGAGGGCACTgc-3'	PCR of hPAH from pMRH161

All oligonucleotides were purchased from IDT DNA.

References:
1. Nishihara, K., Kanemori, M., Kitagawa, M., Yanagi, H., and Yura, T. (1998) Chaperone coexpression plasmids: differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2, in Escherichia coli. *Appl Environ Microbiol* **64**, 1694-1699
2. Weeks, S. D., Drinker, M., and Loll, P. J. (2007) Ligation independent cloning vectors for expression of SUMO fusions. *Protein Expr Purif* **53**, 40-50
3. Shiman, R., Gray, D. W., and Pater, A. (1979) A simple purification of phenylalanine hydroxylase by substrate-induced hydrophobic chromatography. *J Biol Chem* **254**, 11300-11306
4. Ge, Y., Borne, E., Stewart, S., Hansen, M. R., Arturo, E. C., Jaffe, E. K., and Voelz, V. A. (2018) Simulations of the regulatory ACT domain of human phenylalanine hydroxylase (PAH) unveil its mechanism of phenylalanine binding. *J Biol Chem* **293**, 19532-19543
5. Zheng, H., Cooper, D. R., Porebski, P. J., Shabalin, I. G., Handing, K. B., and Minor, W. (2017) CheckMyMetal: a macromolecular metal-binding validation tool. *Acta Crystallogr D Struct Biol* **73**, 223-233
6. Vistica, J., Dam, J., Balbo, A., Yikilmaz, E., Mariuzza, R. A., Rouault, T. A., and Schuck, P. (2004) Sedimentation equilibrium analysis of protein interactions with global implicit mass conservation constraints and systematic noise decomposition. *Anal Biochem* **326**, 234-256
7. Rambo, R. P., and Tainer, J. A. (2013) Accurate assessment of mass, models and resolution by small-angle scattering. *Nature* **496**, 477-481
8. Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A. G., Gajda, M., Gorba, C., Mertens, H. D., Konarev, P. V., and Svergun, D. I. (2012) New developments in the ATSAS program package for small-angle scattering data analysis. *J Appl Crystallogr* **45**, 342-350