Supporting Text

Materials and Methods

Crystallization. Human coproporphyrinogen oxidase (CPO) homodimer (subunit M_r 39,248) underwent time-dependent proteolytic cleavage at residue K230 (Fig. 1*B*, indicated by a ⊗), resulting in two fragments of M_r 13,000 and 26,000 and thus precluding its crystallization. To overcome this problem, we devised a cross-seeding strategy. First, we grew crystals of a bacterial CPO (*Chloroflexus aurantiacus*, a thermophilic phototroph). These crystals were obtained via sitting drop vapor diffusion setups at 22°C from a reservoir buffer containing 30% methylene propanediol (MPD) and 100 mM Tris·HCl, pH 6.5. Crystals obtained under these conditions belong to the hexagonal space group with cell dimensions 205.53 × 205.53 × 85.92 Å, $\alpha = 90^o$, $\beta = 90^o$, $\gamma = 120^o$, and easily diffract X-rays to a Bragg spacing of 1.9 Å. Second, seed stocks of *C. aurantiacus* CPO crystals were prepared and used to streak seed into preequilibrated solutions containing fresh human CPO (40 mg·ml^{-1}), 20% MPD, 0.05 M Tris·HCl pH 7.5, and 10 mM sodium citrate as an additive. Cubic-shaped crystals appeared after 48 h. We used the same strategy to grow crystals of human selenomethionine (Se-Met)-substituted CPO that, on their own, were incapable of nucleation. Mature crystals were stabilized in a glycerol-containing cryoprotectant before flash freezing in liquid nitrogen. The crystals belong to space group P23 with unit cell dimensions of $a = b = c = 112.72$ Å. There is one molecule in the asymmetric unit corresponding to a solvent content of ≈60%.

Structure Determination. Data from a native crystal were collected to a Bragg spacing of 1.5Å by using an ADSC (Poway, CA) Quantum-315 detector at beam line 9-2 of the Stanford Synchrotron Radiation Laboratory. Multiwavelength data on Se-Met human CPO crystals were collected on an ADSC Quantum-4 CCD detector at beamline 5.0.2 of the Advanced Light Source, Berkeley, CA. All data sets were integrated and scaled by using the HKL2000 package, and the statistics are reported in Table 1. Human CPO contains seven methionine residues, and we were able to identify four of these in native

Bijvoet Patterson maps before solving the structure. All seven selenium sites, however, were readily interpreted by using $F_{PH} - F_P$ coefficients as input to the direct methods option of SHELXS. Heavy-atom parameters were refined and phases were calculated at 1.9 Å resolution by using SHARP. Solvent flattening with SOLOMON and phase extension to 1.58 Å resolution against structure factor amplitudes from the native crystal produced an electron density map into which majority of residues could be built unambiguously by using the program O. The first couple of key refinement steps were performed by using the program BUSTER. This program was a *sine qua non* for modeling four loop regions that had no electron density when refined with CNS. All subsequent refinement calculations, however, were performed with CNS. After the addition and verification of water molecules, the model was further refined with REFMAC5 by using maximum-likelihood target and the translation-liberation-screw (TLS) refinement option. Statistics are reported in Table 1. The current model includes residues 119–454, one molecule of citrate, and 364 water molecules.

Equilibrium Analytical Ultracentrifugation. All data were analyzed with the ULTRASCAN software (10). Monte Carlo analysis was performed on a Linux Beowulf cluster and was used to determine 95% confidence intervals of all parameter estimates as described in ref. 11. All scans were fitted to a global model describing either a single ideal species, or a reversibly self-associating monomer-dimer or monomer-dimer-tetramer system. The generalized model is described by Eq. **1**:

$$A_{280,r} = B + \sum_{i=1}^{n} \frac{i C_{ref,m} K_i}{(\epsilon_{280} l)^{i-1}} \exp[i\sigma] \quad \textbf{[1]}$$

where A_{280} is the observed optical density at 280 nm at some radius r in the cell, B is a baseline offset, i is the association state, n is the maximum association state, $C_{ref,m}$ is the concentration of the monomer at a reference radius r_{ref}, K_i is the equilibrium constant for the association of the *ith* state, ε_{280} is the molar extinction coefficient of CPO at 1 cm pathlength, l is the pathlength of the epon-filled centerpiece, and σ is given by:

$$\sigma = \frac{M\omega^2(1-\bar{\nu}\rho)(r^2 - r_{ref}^2)}{2RT} \quad \textbf{[2]}$$

Here, M is the monomer molecular weight, ω is the radial velocity, $\bar{\nu}$ is the partial specific volume of CPO, ρ is the density of the buffer, R is the gas constant, and T is the temperature. Models with n = 1, 2, and 4 were fitted, and the variance, molecular weight, and equilibrium constants were determined. Hydrodynamic corrections for buffer conditions were made in ULTRASCAN according to data published by Laue *et al.* (12). The partial specific volume of CPO was estimated according to the method by Cohn and Edsall (13), and was found to be 0.7272 ccm/g. The molar extinction coefficient was estimated for 280 nm from the sequence of the denatured protein by the method of Gill and von Hippel (14).

Results

The results of the fitted equilibrium experiments are shown in Table 2. All three models suggest that in the examined concentration range, the protein is present almost exclusively in the dimeric form of CPO. The best variance, random residuals, and most accurate monomer molecular weight was observed when the data were fitted to a monomer-dimer-tetramer model. In this model, the monomer molecular weight was in excellent agreement with the molecular weight predicted from the protein sequence. The concentration distributions of the equilibrium scans and the monomer-dimer-tetramer fit and the combined residuals are shown in Fig. 7. The equilibrium constant for the monomer-dimer association suggested monomer only present in the low nanomolar range, whereas a slight amount of tetramer was predicted only for the highest concentrations examined in the equilibrium experiment. This small signal contribution of the tetramer species not unexpectedly translated into a low confidence for the monomer-tetramer equilibrium constant, which was confirmed by the Monte Carlo analysis.

1. Stura, E. A. & Wilson, I. A. (1990) *Methods* **1,** 38–49.

2. Otwinowski, Z. & Minor, W. (1997) *Methods Enzymol.* **276,** 307–325.

3. Sheldrick, G. M. (1997) *Methods Enzymol.* **276,** 628–641.

4. de La Fortelle, E. & Bricogne, G. (1997) *Methods Enzymol.* **276,** 472–494.

5. Abrahams, J. P. & Leslie, A. G. W. (1996) *Acta Crystallogr. D* **52,** 30–42.

6. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991) *Acta Crystallogr. A* **47,** 110–119.

7. Bricogne, G. (1997) *Methods Enzymol.* **276,** 361–423.

8. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., *et al.* (1998) *Acta Crystallogr. D* **54,** 905–921.

9. Winn, M. D., Isupov, M. N. & Murshudov, G. N. (2001) *Acta Crystallogr. D* **57,** 122–133.

10. Demeler, B. (2004) ULTRASCAN (Department of Biochemistry, Univ. of Texas Health Science Center at San Antonio), Version 6.2.

11. Lambert, L. J., Schirf, V., Demeler, B., Cadene, M. & Werner, M. H. (2001) *EMBO J.* **20,** 7149–7159.

12. Laue, T. M., Shah, B. D., Ridgeway, T. M. & Pelletier, S. L. (1992) *Computer-Aided Interpretation of Analytical Sedimentation Data for Proteins: Analytical Ultracentrifugation in Biochemistry and Polymer Science*, eds. Harding, S. E., Rowe, A. J. & Horton, J. C. (R. Soc. Chem., Cambridge, U.K.), pp. 90–125.

13. Cohn, E. J. & Edsall, J. T. (1943) *Proteins, Amino Acids, and Peptides as Ions and Dipolar Ions* (Reinhold, New York).

14. Gill, S. C. & von Hippel, P. H. (1989) *Anal. Biochem.* **182,** 319–326.