SUPPORTING INFORMATION

High-level production and properties of the cysteine-depleted cytochrome P450 3A4

Irina F. Sevrioukova*

Department of Molecular Biology and Biochemistry, University of California, Irvine, CA 92697-3900, United States

*Corresponding author: University of California Irvine, Department of Molecular Biology and Biochemistry, 3205 McGaugh Hall, Irvine, California 92697-3900, USA. Tel: (949) 824-1953, Email: sevrioui@uci.edu

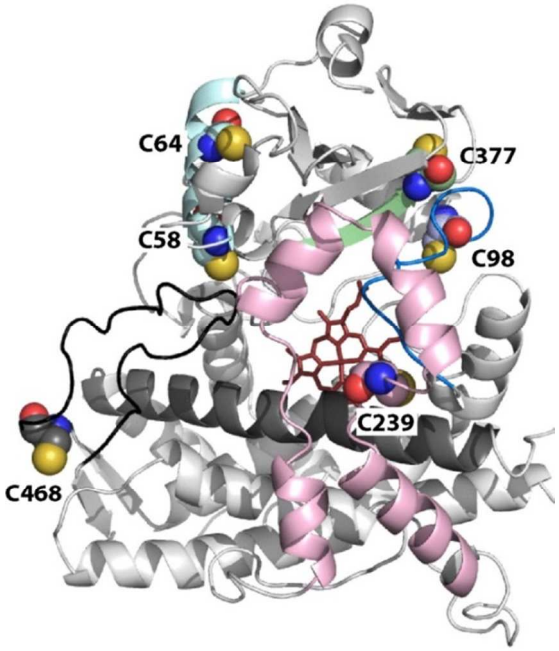

Figure S1. Location of non-heme-ligating cysteines in CYP3A4. Cys58 and Cys64 are part of the A-helix (light blue), Cys377 is at the end of the β-strand comprising the β-domain (green), whereas cysteines 98, 239, and 468 are located in the flexible B-B' (blue), G'-G (pink), and C-terminal loops (black), respectively. The F-G fragment and the I-helix are rendered in pink and dark gray, respectively.

```
ATGGCCCTGATTCCGGATCTGGCAATGGAGACATGGTTACTGCTGGCAGTGAGCCTGGTGCTGCTGTATCTGTATG
GTACCCATAGCCACGGTCTGTTCAAGAAACTGGGCATTCCGGGTCCTACACCGTTACCGTTTCTGGGTAACATCCTGAG
TTACCACAAAGGCTTCTGCATGTTTGATATGGAGTGCCACAAAAAGTACGGCAAGGTGTGGGGCTTCTATGACGGCCAA
CAGCCGGTGCTGGCCATCACCGACCCTGATATGATTAAAACCGTGCTGGTGAAGGAGTGCTACAGTGTGTTCACCAATC
GTCGCCCTTTTGGTCCGGTGGGTTTCATGAAGAGCGCCATTAGCATCGCCGAAGATGAGGAGTGGAAGCGCTTACGCAG
CCTGTTAAGCCCTACCTTCACCAGCGGCAAGCTGAAGGAGATGGTGCCGATCATTGCCCAGTACGGTGATGTGCTGGTT
CGCAATCTGCGTCGCGAGGCCGAAACAGGCAAGCCGGTTACACTGAAGGACGTGTTCGGTGCCTATAGCATGGACGTGA
TTACCAGCACCAGCTTCGGCGTGAATATCGATAGCCTGAACAACCCGCAGGACCCGTTTGTTGAAAACACAAAGAAATT
ACTGCGCTTCGATTTCCTGGACCCGTTTTTCCTGAGCATCACCGTGTTCCCTTTCCTGATTCCGATCCTGGAAGTGCTG
AACATCTGTGTGTTCCCGCGCGAAGTGACCAACTTTCTGCGTAAAAGTGTTAAACGCATGAAGGAGAGCCGTCTGGAGG
ACACACAGAAACACCGCGTGGACTTCCTGCAGCTGATGATTGACAGCCAGAACAGTAAAGAGACCGAAAGCCATAAAGC
CCTGAGCGACCTGGAGCTGGTTGCCCAGAGCATCATCTTCATCTTCGCCGGCTACGAGACCACCAGCAGCGTTTTAAGC
TTCATTATGTACGAGCTGGCCACCCACCCGGATGTTCAGCAGAAGCTGCAAGAGGAGATCGATGCCGTGCTGCCTAATA
AGGCCCCTCCTACCTATGACACCGTTCTGCAGATGGAATACCTGGATATGGTTGTGAACGAGACCCTGCGCCTGTTCCC
GATTGCAATGCGCCTGGAGCGCGTGTGCAAAAAGGACGTGGAGATCAATGGTATGTTTATTCCGAAAGGCGTTGTGGTT
ATGATCCCGAGCTACGCACTGCATCGCGATCCGAAGTACTGGACAGAGCCGGAGAAGTTCTTACCGGAGCGTTTTAGCA
AGAAGAACAAGGACAACATTGACCCGTACATCTACACACCGTTCGGCAGTGGTCCTCGCAACTGTATCGGTATGCGCTT
CGCCCTGATGAACATGAAGCTGGCACTGATCCGCGTGCTGCAGAACTTTAGCTTCAAGCCGTGCAAGGAGACCCAAATC
CCGCTGAAGCTGAGCCTGGGTGGTTTATTACAGCCTGAGAAGCCGGTGGTGCTGAAAGTTGAAAGCCGTGACGGCACAG
TTAGTGGCGCACACCATCATCATTAA
```

Figure S2. cDNA sequence of the full-length WT CYP3A4 optimized for expression in *Escherichia coli*. Codon optimization and gene synthesis was conducted by Genewiz (USA). Blue and red nucleotides correspond to the C-terminal 4-histidine tag and a stop codon, respectively.

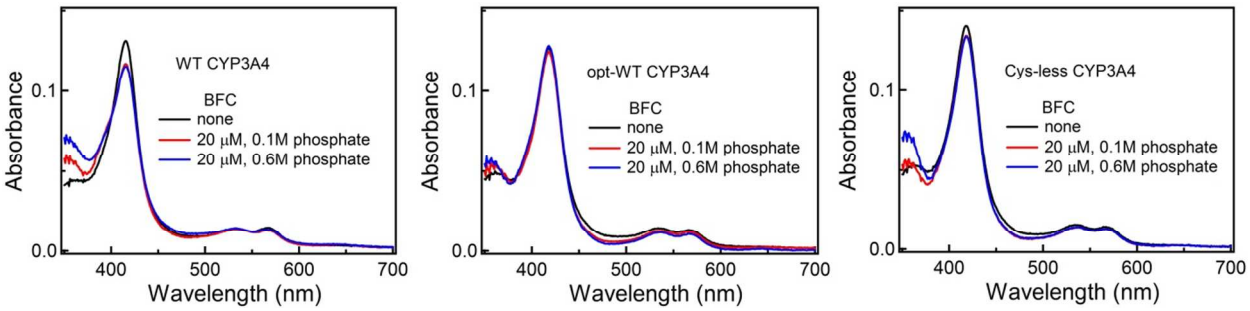

Figure S3. Absorbance spectra of 1.3 μM WT, opt-WT and Cys-less CYP3A4 recorded in the absence and presence of 30 μM 7-benzyloxy-(4-trifluoromethyl)coumarin (BFC) in 0.1 or 0.6M phosphate buffer pH 7.4. BFC induces a small high spin shift in WT but not in opt-WT or Cys-less CYP3A4.

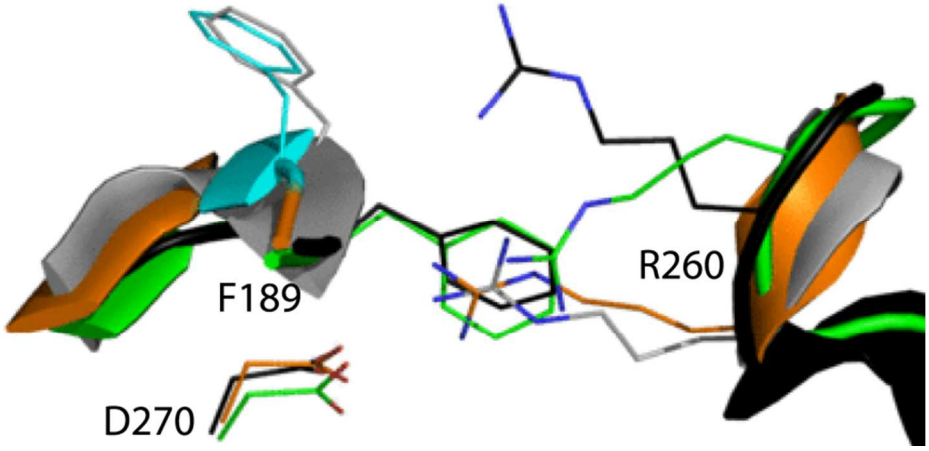

Figure S4. Relative orientation of Phe189, Arg260 and Asp260 in opt-WT (PDB ID: 5VCC, 1.70 Å; orange/cyan) and the ligand-free structures of WT CYP3A4: 1TQN (2.05 Å; black), 1W0E (2.80 Å; gray), and 4I3Q (2.6 Å; green). Asp270 was disordered and not modeled in the 1W0E structure. Naturally occurring heterogeneity in the Phe189-region could complicate the ligand binding and catalytic behavior of CYP3A4.

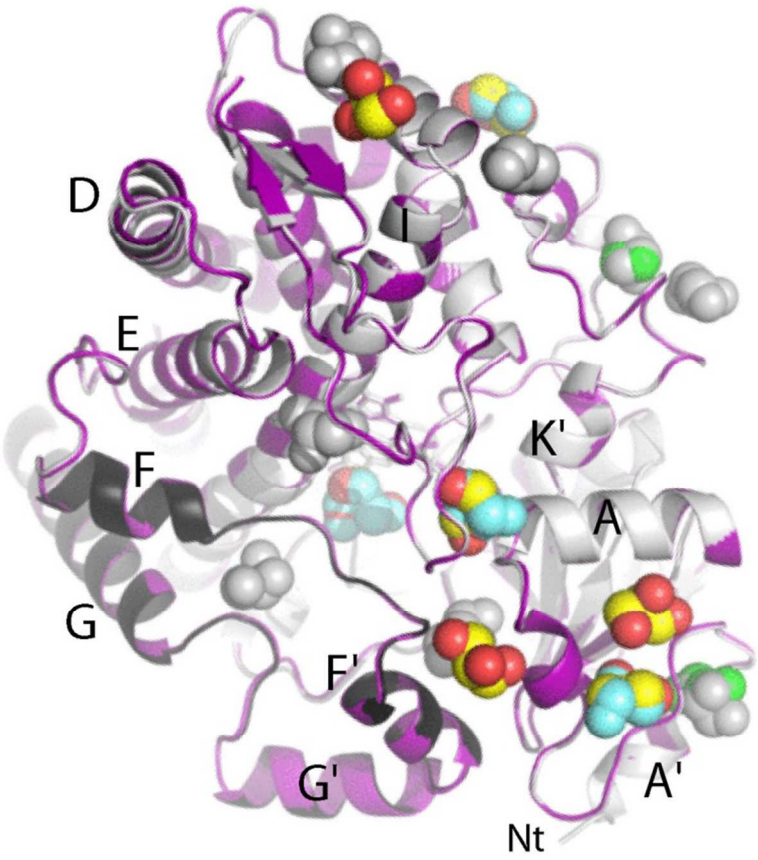

Figure S5. Superposition of Cys-less and opt-WT CYP3A4 (gray/black and purple, respectively). Glycerol molecules bound to the corresponding structures are shown in red/yellow and gray/cyan, respectively. Four coinciding glycerol binding sites are shown in cyan, whereas those occupied by ethylene glycol are in green.

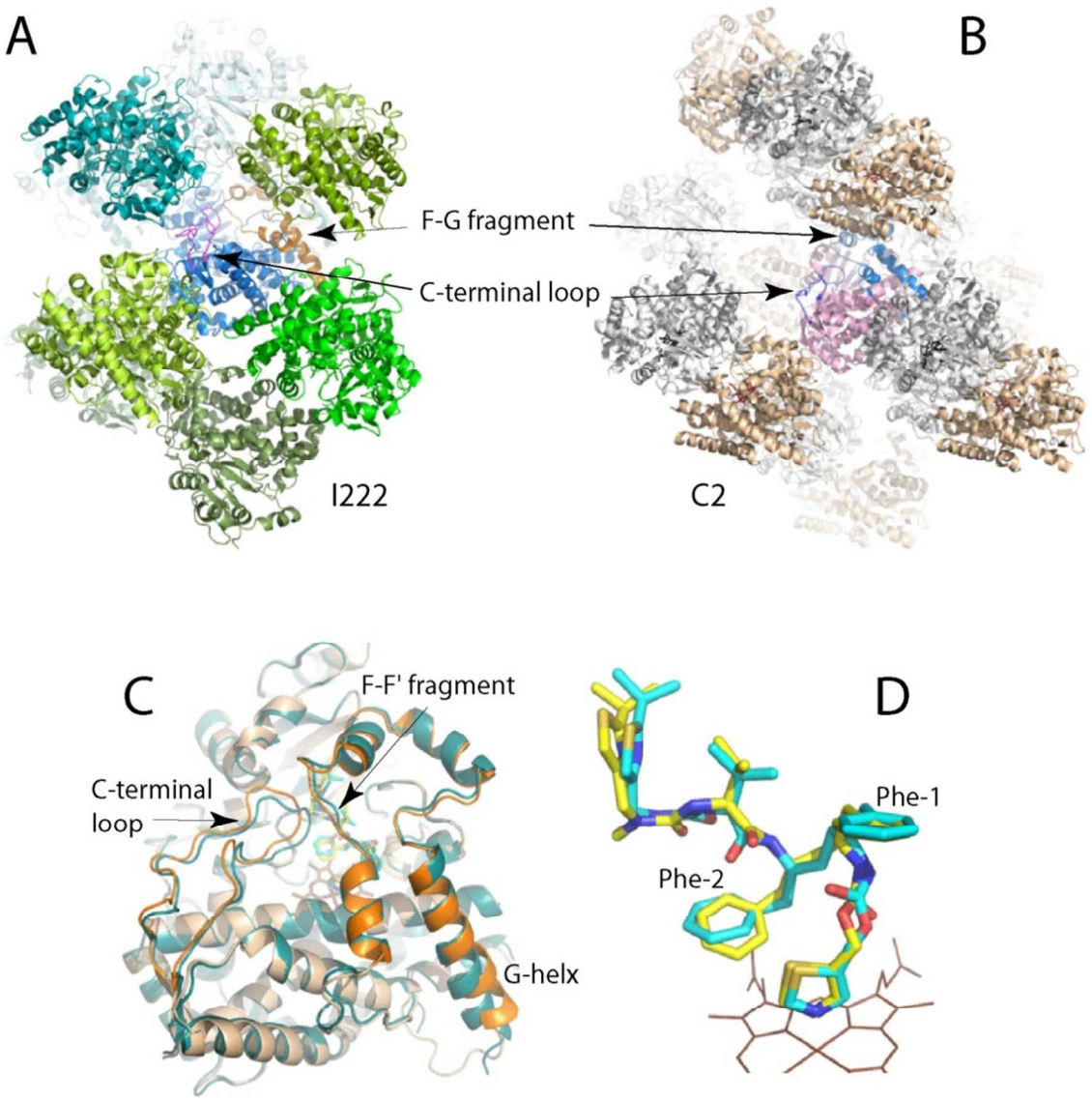

Figure S6. ***A*** and ***B***, Packing of the RIT-bound WT CYP3A4 in the I222 and C2 crystal lattice (PDB ID: 5VCO and 3NXU, respectively). The latter crystal form has two molecules in the asymmetric unit, which are packed tighter and form different intermolecular contacts. ***C***, Superposition of the 5VCO model (light and dark orange) and molecule A of the 3NXU structure (in cyan) shows distinct positioning of the G-helix, F-F' fragment and C-terminal loop. ***D***, The RIT binding mode observed in the 5VCO and 3NXU structures (in yellow and cyan, respectively). Differences in the backbone conformation, carbonyl oxygen orientation and placement of Phe-2 indicate that crystal packing restraints can alter the ligand binding mode in CYP3A4.

Table S1. Data collection and refinement statistics

	WT	opt-WT	Cys-less		
Ligand	RIT	water, glycerol	water, glycerol	RIT	BEC
PDB ID	5VCO	5VCC	5VCD	5VCE	5VCG
Data statistics					
Space group	I222	I222	I222	I222	I222
Unit cell parameters	*a* = 77 Å, *b* = 100 Å, *c* = 125 Å; α, β, γ = 90°	*a* = 76 Å, *b* = 99 Å, *c* = 134 Å; α, β, γ = 90°	*a* = 76 Å, *b* = 100 Å, *c* = 133 Å; α, β, γ = 90°	*a* = 77 Å, *b* = 101 Å, *c* = 129 Å; α, β, γ = 90°	*a* = 77 Å, *b* = 100 Å, *c* = 134 Å; α, β, γ = 90°
Molecules per asymmetric unit	1	1	1	1	1
Resolution range (Å)	78.12-2.70 (2.85-2.70)	44.82-1.70 (1.79-1.70)	79.73-1.95 (2.06-1.95)[a]	79.43-2.20 (2.32-2.20)	80.13-2.20 (2.32-2.20)
Total reflections	81,594	263,425	203,249	110,030	114,395
Unique reflections	13,476	55,451	37,249	25,852	25,847
Redundancy	6.1 (6.0)	4.8 (4.7)	5.5 (5.5)	4.3 (4.3)	4.4 (4.5)
Completeness	99.5 (99.6)	99.3 (100)	99.8 (100)	99.4 (99.7)	97.8 (99.2)
Average $I/\sigma I$	9.1 (1.2)	12.2 (1.6)	11.8 (1.5)	9.4 (1.1)	2.9 (0.50)
R_{merge}	0.096 (1.798)	0.051 (0.951)	0.074 (1.209)	0.071 (1.247)	0.199 (1.811)
R_{pim}	0.042 (0.790)	0.026 (0.481)	0.035 (0.562)	0.040 (0.694)	0.099 (0.936)
CC ½	0.996 (0.340)	0.998 (0.514)	0.996 (0.532)	0.996 (0.549)	0.976 (0.472)
Refinement statistics					
R/R_{free}[b]	21.2/ 28.3	16.6/ 20.4	19.3/ 24.5	24.2/ 27.5	23.5/ 27.2
Number of atoms:					
Protein	3731	3798	3715	3708	3717
Solvent	none	331	106	17	76
R.m.s. deviations:					
Bond lengths, Å	0.007	0.007	0.008	0.009	0.008
Bond angles, °	1.189	1.114	1.149	1.381	1.34

[a]Values in brackets are for the highest resolution shell.

[b]R_{free} was calculated from a subset of 5% of the data that were excluded during refinement.