Supporting Information

Structural basis for plant lutein biosynthesis from α-carotene

Guoqi Niu, Qi Guo, Jia Wang, Shun Zhao, Yikun He, Lin Liu

This Supporting Information includes:

Supplementary Methods

Supplementary References

Fig. S1: Purification and crystallization of CYP97A3, CYP97A3m, and CYP97C1.

Fig. S2: HPLC analysis of non-natural ligands of CYP97A3 and CYP97C1.

Fig. S3: Omit map of the bound ligands contoured at 3σ.

Fig. S4: CYP97A3m structures.

Fig. S5: Characterization of CYP97A3m.

Fig. S6: Heme in the CYP97A3 and CYP97C1 structures.

Table S1: Crystal diffraction and refinement data.

www.pnas.org/cgi/doi/10.1073/pnas.2001806117

Supplementary Methods

Construction of Constructs

To generate the pCW-MBP expression vector, we moved the *malE* gene from the pETMALc-H vector (1) to the pCWori+ vector. The *malE* gene (encoding an MBP-His_6-tag) was amplified by PCR with the sense primer 5'-GTCATATGAAAATCGAAGAAGGTA AAC-3' (underlined: the *Nde*I restriction site) and the antisense primer 5'-GCCTCGAGTGCGGCCGCAAGCTTG-3' (underlined: the *Xho*I restriction site) using the pETMALc-H vector as template. The PCR product was then ligated with the *Nde*I/*Xho*I-cleaved pCWori+ vector.

The *CYP97A3* gene sequence encoding residues 78–595 was amplified by PCR with the sense primer 5'-CGGAATTCGGAGAACCTGTACTTCCAAGGCAGTTTTCCATCTACAGTGAAG-3' (underlined: the *Eco*RI restriction site; underlined by wave line: the TEV cleavage site) and the antisense primer 5'-GCTCTAGATTACCTTTGGCTCACCTTCAT-3' (underlined: the *Sal*I restriction site) using *Arabidopsis thaliana* cDNA library as template. The PCR product was then ligated with the *Eco*RI/*Sal*I-cleaved pCW-MBP expression vector. The resulting construct encodes the fusion protein MBP-His_6-TEV-CYP97A3 (residues 78–595). The amino-acid sequence after TEV protease cleavage is shown below. The recombinant CYP97A3 (residues 78–595) contains an extra glycine residue at the N-terminal end (numbered as 77) introduced during cloning.

```
 77 GSFPSTVKNG LSKIGIPSNV LDFMFDWTGS DQDYPKVPEA KGSIQAVRNE 126
127 AFFIPLYELF LTYGGIFRLT FGPKSFLIVS DPSIAKHILK DNAKAYSKGI 176
177 LAEILDFVMG KGLIPADGEI WRRRRRAIVP ALHQKYVAAM ISLFGEASDR 226
227 LCQKLDAAAL KGEEVEMESL FSRLTLDIIG KAVFNYDFDS LTNDTGVIEA 276
277 VYTVLREAED RSVSPIPVWD IPIWKDISPR QRKVATSLKL INDTLDDLIA 326
327 TCKRMVEEEE LQFHEEYMNE RDPSILHFLL ASGDDVSSKQ LRDDLMTMLI 376
377 AGHETSAAVL TWTFYLLTTE PSVVAKLQEE VDSVIGDRFP TIQDMKKLKY 426
427 TTRVMNESLR LYPQPPVLIR RSIDNDILGE YPIKRGEDIF ISVWNLHRSP 476
477 LHWDDAEKFN PERWPLDGPN PNETNQNFSY LPFGGGPRKC IGDMFASFEN 526
527 VVAIAMLIRR FNFQIAPGAP PVKMTTGATI HTTEGLKLTV TKRTKPLDIP 576
577 SVPILPMDTS RDEVSSALS 595
```

The *CYP97C1* gene sequence (residues 70–539) was amplified by PCR with the sense primer 5'-CGGAATTCGGAGAACCTGTACTTCCAAGGCTCCTCAGGAAAAAACGACGAG-3' (underlined: the *Eco*RI restriction site; underlined by wave line: the TEV cleavage site) and the antisense primer 5'-GCTCTAGATTACCTTTGGCTCACCTTCAT-3' (underlined: the *Sal*I restriction site) using *A. thaliana* cDNA library as template. The recombinant CYP97C1 (residues 70–539) contains an extra glycine residue (numbered as 69).

```
 69 GSSGKNDESG IPIANAKLDD VADLLGGALF LPLYKWMNEY GPIYRLAAGP 118
119 RNFVIVSDPA IAKHVLRNYP KYAKGLVAEV SEFLFGSGFA IAEGPLWTAR 168
169 RRAVVPSLHR RYLSVIVERV FCKCAERLVE KLQPYAEDGS AVNMEAKFSQ 218
219 MTLDVIGLSL FNYNFDSLTT DSPVIEAVYT ALKEAELRST DLLPYWKIDA 268
269 LCKIVPRQVK AEKAVTLIRE TVEDLIAKCK EIVEREGERI NDEEYVNDAD 318
319 PSILRFLLAS REEVSSVQLR DDLLSMLVAG HETTGSVLTW TLYLLSKNSS 368
369 ALRKAQEEVD RVLEGRNPAF EDIKELKYIT RCINESMRLY PHPPVLIRRA 418
419 QVPDILPGNY KVNTGQDIMI SVYNIHRSSE VWEKAEEFLP ERFDIDGAIP 468
469 NETNTDFKFI PFSGGPRKCV GDQFALMEAI VALAVFLQRL NVELVPDQTI 518
519 SMTTGATIHT TNGLYMKVSQ R 539
```

All plasmids were sequenced to verify their fidelity. The CYP97A3 triple mutant S290D/W300L/S304V (CYP97A3m) was generated with a Fast Mutagenesis System Kit

(TransGen Biotech) using the MBP-His_6-TEV-CYP97A3 plasmid as template. Procedure for purification of the mutant proteins was the same as that of the wild type.

Recombinant Protein Purification

The fusion protein MBP-His_6-TEV-CYP97A3 was purified by nickel affinity chromatography and size exclusion chromatography. Buffer A (200 mM NaCl and 20 mM Tris-HCl, pH 7.5) was used throughout the purification procedure. The cleared lysate (in buffer A plus 5 mM imidazole) was loaded onto an Ni-NTA agarose (QIAGEN, Shanghai, China) column (3 ml), and washed with 10-column volume of buffer A plus 5 mM imidazole. The fusion protein was eluted with 2-column volume of 200 mM imidazole in buffer A. The eluate (6 ml) was concentrated by ultrafiltration using Amicon Ultra-15 filter (Merck Millipore, Beijing, China). The concentrate (~2 ml) was then loaded onto a HiLoad 16/60 Superdex 200 column (GE Healthcare, Shanghai, China), which had a column volume of 120.6 ml, and eluted with buffer A at a flow rate of 1.5 ml min^{-1}. The peak fractions corresponding to the fusion protein were collected and pooled (~12 ml). To cleave the MBP-His_6-tag, TEV protease with a His_6-tag (2) was added in the presence of 0.5 mM (final) EDTA and 1 mM (final) DTT, and the mixture was incubated overnight at 4°C. Then the mixture was dialyzed against buffer A to remove EDTA and DTT before loaded onto the Ni-NTA column (3 ml). The flow-through (~12 ml) was collected and concentrated by ultrafiltration. The concentrate (~2 ml) was applied onto the HiLoad 16/60 Superdex 200 column and eluted with buffer A plus 2 mM DTT. The peak fractions were collected and subjected to SDS–PAGE (5% stacking gel, 12% separating gel), which showed the purity of the recombinant CYP97A3.

Purifications of the recombinant CYP97A3m and CYP97C1 were the same as described for CYP97A3.

Supplementary References

1. Pryor KD, Leiting B (1997) High-level expression of soluble protein in Escherichia coli using a His6-tag and maltose-binding-protein double-affinity fusion system. *Protein Expr Purif* 10:309−319.
2. Tropea JE, Cherry S, Waugh DS (2009) Expression and purification of soluble His6-tagged TEV protease. *Methods Mol Biol* 498:297−307.

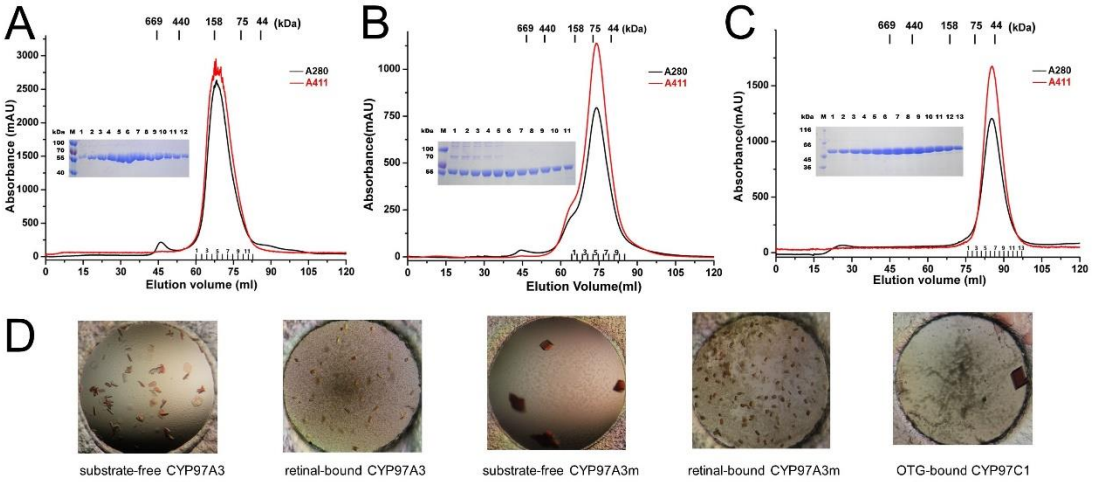

Fig. S1. Purification and crystallization of CYP97A3, CYP97A3m, and CYP97C1. (*A*–*C*) SDS–PAGE and SEC analyses of the purified CYP97A3 (*A*), the S290D/W300L/S304V mutant of CYP97A3 (CYP97A3m) (*B*), and CYP97C1 (*C*). GE Healthcare HiLoad 16/60 Superdex 200 column (1 × 120 ml) was used. Protein concentration was approximate 30 mg ml^{-1}. (*D*) Crystals of five forms.

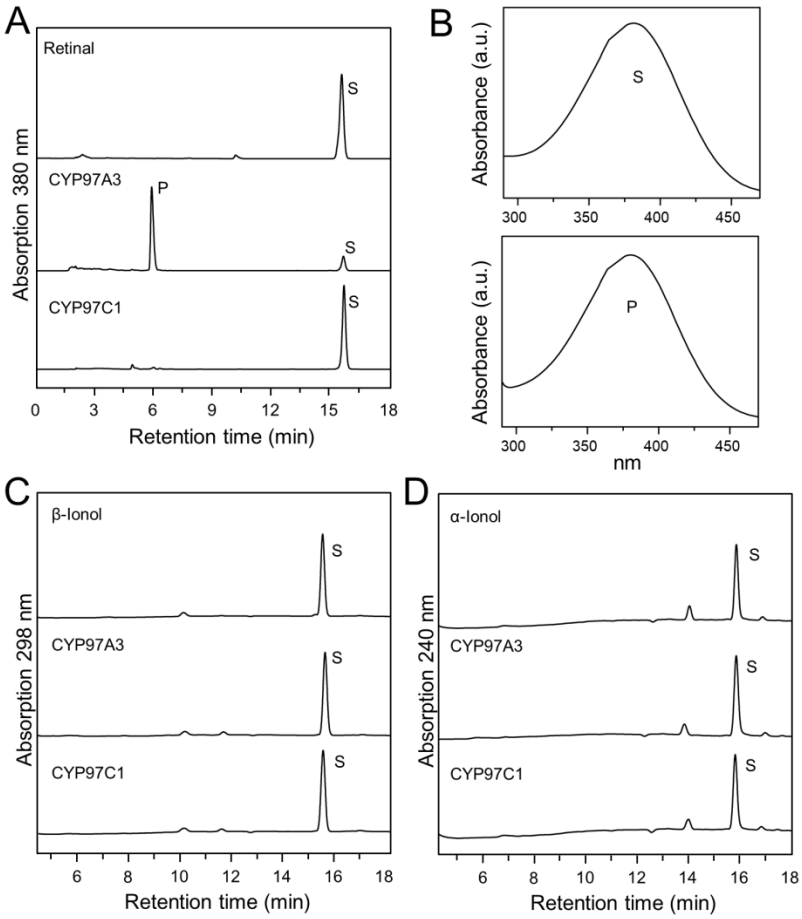

Fig. S2. HPLC analysis of the non-natural ligands of CYP97A3 and CYP97C1. (*A*) Incubation of all-*trans* retinal. The profiles (from top to bottom) are: the substrate (S) only; incubation of substrate, CYP97A3, NADPH, FNR and ferredoxin; incubation of substrate, CYP97C1, NADPH, FNR and ferredoxin. Arbitrary units were recorded. (*B*) The UV–visible spectra of the substrate all-*trans* retinal and the product (P). (*C*) Incubation of β-ionol. (*D*) Incubation of α-ionol. Concentrations of reagents: NADPH, 1 mM; glucose 6-phosphate, 10 mM; glucose 6-phosphatedehydrogenase, 5 units; substrate, 0.5 mM, ATR, 4 μM, or FNR, 2 μM and FDX1, 12 μM, and recombinant protein, 2 μM.

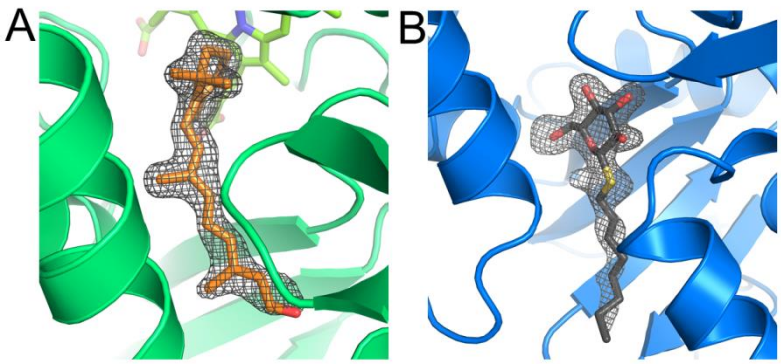

Fig. S3. Omit map of the bound ligands contoured at 3σ. (*A*) The retinal-bound CYP97A3 structure. The $|F_o| - |F_c|$ map of retinal is shown in *gray* mesh. (*B*) The OTG-bound CYP97C1 structure. The $|F_o| - |F_c|$ map of OTG is shown in *gray* mesh.

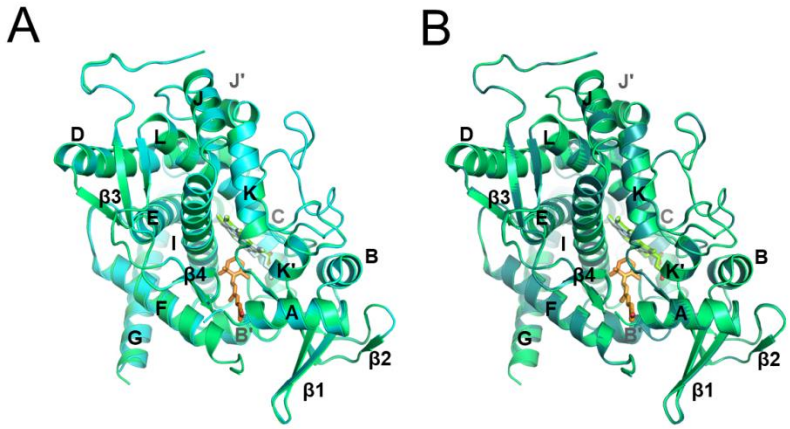

Fig. S4. CYP97A3m structures. (*A*) Superimposed structures of the substrate-free CYP97A3m (*cyan*) and retinal-bound CYP97A3 (*green*). The protein is shown as ribbon with ligands in stick representation. (*B*) Superimposed structures of the retinal-bound CYP97A3m (*teal*) and retinal-bound CYP97A3.

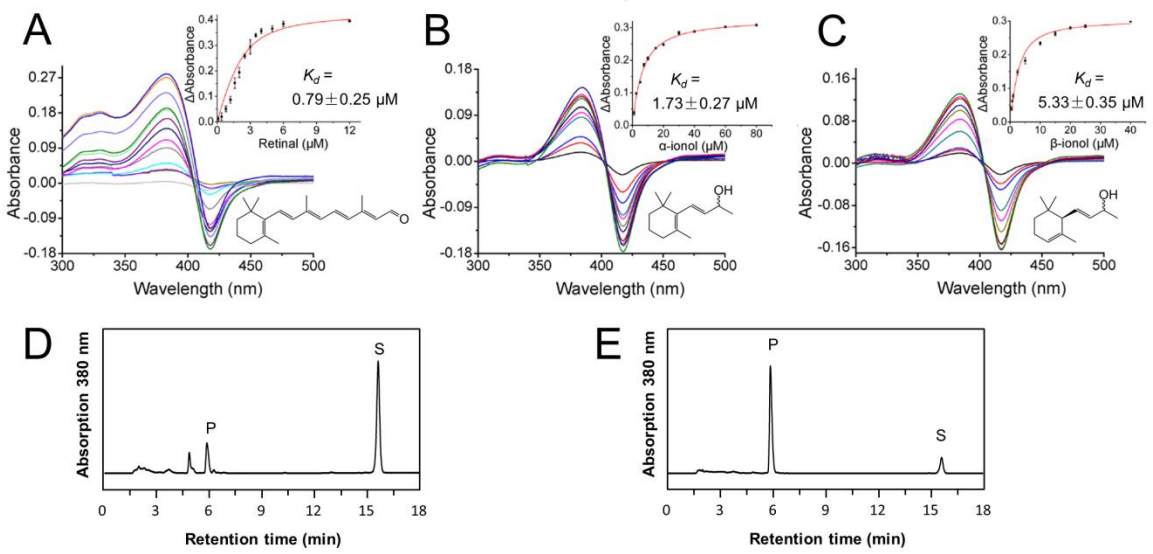

Fig. S5. Characterization of CYP97A3m. (*A*–*C*) Differential spectral titration of CYP97A3m with increasing concentrations of non-natural ligand. Sequential additions of (*A*) retinal, (*B*) β-ionol, and (*C*) α-ionol are individually colored. The K_d values are obtained from a nonlinear regression of the data. (*D*–*E*) HPLC analysis of CYP97A3m-catalyzed retinal hydroxylation. (*D*) Incubation of retinal with CYP97A3m, NADPH and the CPR ATR2. (*E*) Incubation of retinal with CYP97A3m, NADPH and the FNR–ferredoxin pair. P and S stand for product and substrate. Concentrations of reagents: NADPH, 1 mM; glucose 6-phosphate, 10 mM; glucose 6-phosphatedehydrogenase, 5 units; substrate, 0.5 mM, ATR, 4 μM, or FNR, 2 μM and FDX1, 12 μM, and recombinant protein, 2 μM.

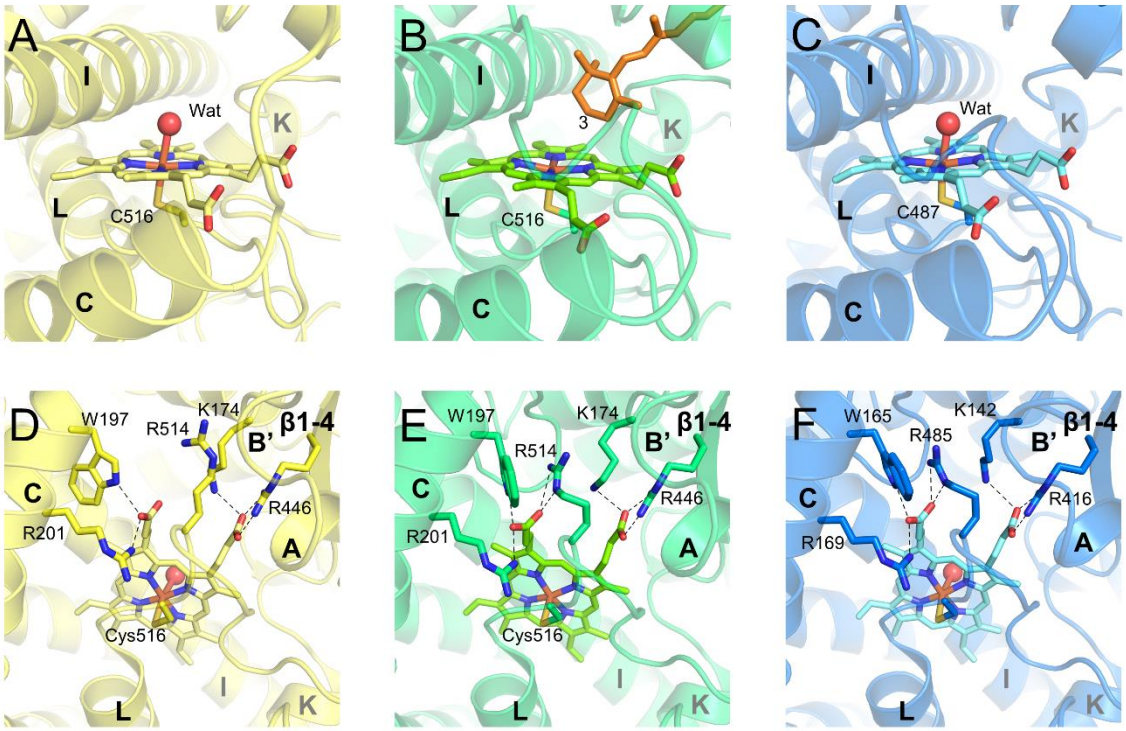

Fig. S6. Heme in the CYP97A3 and CYP97C1 structures. (*A*–*C*) The heme distal side in substrate-free CYP97A3 (*A*), retinal-bound CYP97A3 (*B*), and OTG-bound CYP97C1 (*C*). The water molecular is in sphere. (*D*–*F*) The heme proximal side in in substrate-free CYP97A3 (*D*), retinal-bound CYP97A3 (*E*), and OTG-bound CYP97C1 (*F*). The hydrogen bonds are in dashed lines.

Table S1. Crystal diffraction and refinement data

	substrate-free CYP97A3	retinal-bound CYP97A3	substrate-free CYP97A3m	retinal-bound CYP97A3m	OTG-bound CYP97C1
PDB ID	6J94	6J95	6L8I	6L8J	6L8H
Data collection					
Space group	C2	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_1$
Resolution (Å)[a]	50-2.40 (2.49-2.40)	50-2.00 (2.07-2.00)	50-1.70 (1.76-1.70)	50-2.40 (2.49-2.40)	50-2.00 (2.07-2.00)
Cell dimensions					
a, *b*, *c* (Å)	164.0,70.4, 133.2	58.5, 82.1, 110.3	58.1, 81.5, 110.2	58.2, 82.1, 110.5	61.7, 127.3, 143.7
α, β, γ (°)	90, 124.6, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 92.2, 90
No. of unique reflections	48368(4767)	36341(3557)	58443(5770)	21298(2077)	148875(14803)
$CC_{1/2}$[b]	0.937(0.752)	0.968(0.875)	0.993(0.993)	0.933(0.793)	0.934(0.745)
R_{merge}	0.090(0.695)	0.115(0.703)	0.053(0.521)	0.136(0.876)	0.094(0.749)
R_{pim}	0.054(0.409)	0.036(0.223)	0.015(0.153)	0.054(0.347)	0.056(0.450)
$I/\sigma I$	15.0(1.8)	23.2(2.8)	55.8(4.7)	13.0(1.8)	13.6(1.8)
Completeness (%)	98.5(97.8)	99.5(99.2)	99.9(100)	99.4(99.7)	99.9(100)
Redundancy	3.8(3.9)	11.1(10.6)	13.1(12.3)	7.2(7.3)	3.8(3.8)
Refinement					
Resolution (Å)	41.96-2.40 (2.49-2.40)	38.47-2.00 (2.07-2.00)	29.03-1.70 (1.76-1.70)	43.63-2.40 (2.49-2.40)	38.25-2.00 (2.07-2.00)
R_{work}/R_{free} (%)	18.6/21.6	16.1/19.8	17.7/19.7	22.3/26.8	17.3/22.4
No. of Proteins	2	1	1	1	4
No. of atoms	8080	4066	3898	3765	16954
Protein	7276	3486	3481	3449	14738
Ligand/ion	86	64	43	64	312
Water	648	516	374	252	1904
RMSD Bond (Å)	0.006	0.007	0.007	0.005	0.008
RMSD Angels (°)	0.957	1.160	1.04	1.00	0.918
Average B-factor	20.87	30.26	35.90	19.58	30.49
Protein	20.78	28.71	35.17	19.62	29.61
Ligand	6.64	20.33	25.30	16.17	32.70
Solvent	23.73	41.99	43.96	19.89	36.95
Ramachandran plot					
Favored (%)	97.03	98.18	98.85	97.71	98.54
Allowed (%)	2.97	1.82	1.15	2.29	1.46

[a]Numbers in parentheses are for highest-resolution shell.

[b]The values for $CC_{1/2}$ are for the highest-resolution shell.