Supplemental information for the manuscript:

A non-canonical monovalent zinc finger stabilizes the integration of Cfp1 into SET1 complexes within COMPASS.

Yidai Yang[1], Monika Joshi[1], Yoh-hei Takahashi[2], Zhibin Ning[1], Qianhui Qu[3], Joseph S. Brunzelle[4], Giorgios Skiniotis[3], Daniel Figeys[1], Ali Shilatifard[2] and Jean-François Couture[1]

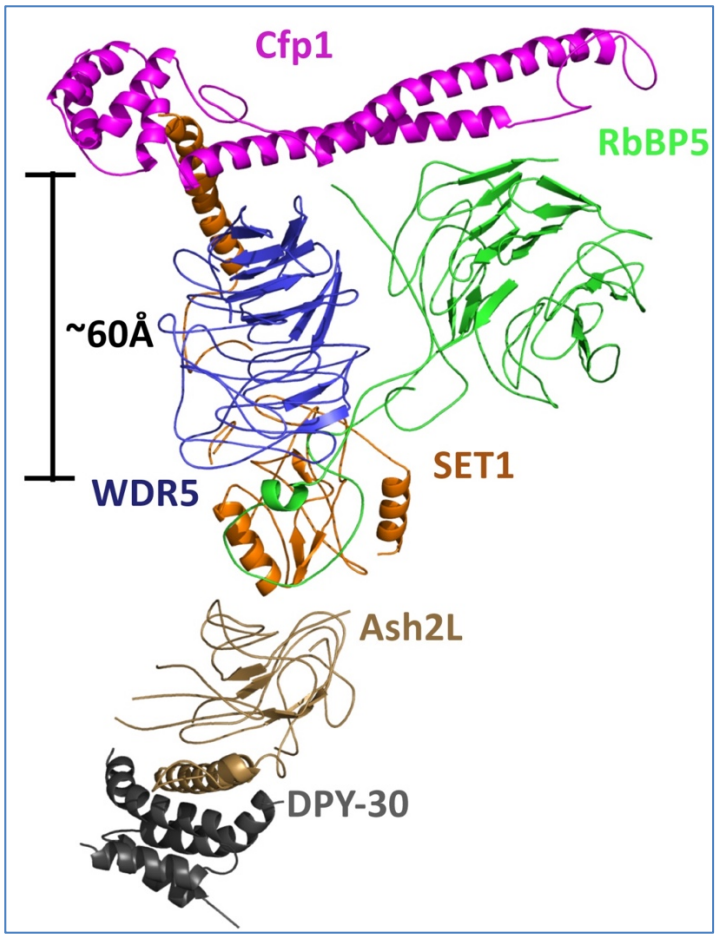

Figure S1. The Cryo-EM model of COMPASS. Colors are rendered related subunits: Cfp1 (pink), RbBp5 (green), WDR5 (blue), SET (orange), Ash2L (sand), DPY-30 (grey). Cfp1 is about 60 Å away from SET domain.

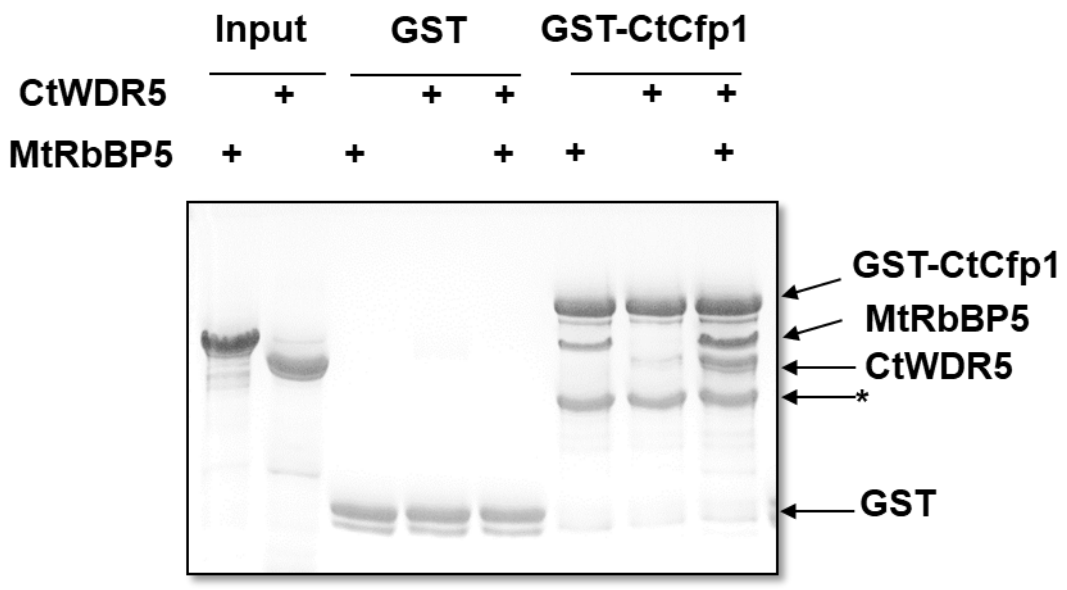

Figure S2. Cfp1 tightly interacts with RbBP5 but not WDR5. GST-tagged Cfp1 were incubated with RbBP5, WDR5 or RbBP5/WDR5 complex. Bound proteins were separate on a 15% SDS-PAGE and detected by Coomassie staining. Star (*) indicates a degradation product of GST-tagged CtCfp1.

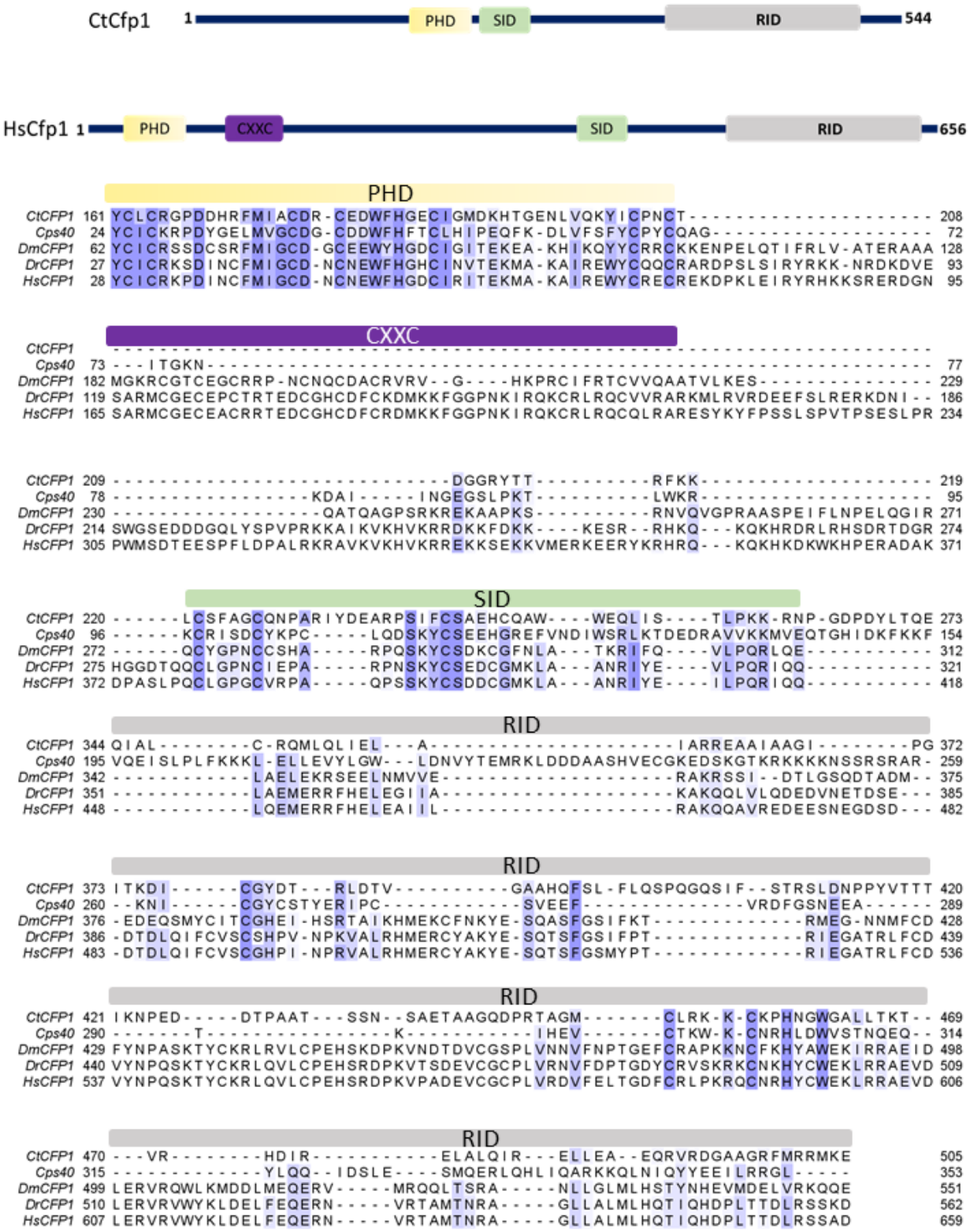
CtCfp1
1
PHD
SID
RID
544
HsCfp1
1
PHD
CXXC
SID
RID
656
PHD
CtCFP1 161 YCLCRGPDDHRFMIACDR-CEDWFHGECIGMDKHTGENLVQKYICPNCT 208
Cps40 24 YCICKRPDYGELMVGCDG-CDDWFHFTCLHIPEQFK-DLVFSFYCPYCQAG 72
DmCFP1 62 YCICRSSDCSRFMIGCD-GCEEWYHGDCIGITEKEA-KHIKQYYCRRCKKENPELQTIFRLV-ATERAAA 128
DrCFP1 27 YCICRKSDINCFMIGCD-NCNEWFHGHCINVTEKMA-KAIREWYCQQCRARDPSLSIRYRKK-NRDKDVE 93
HsCFP1 28 YCICRKPDINCFMIGCD-NCNEWFHGDCIRITEKMA-KAIREWYCRECREKDPKLEIRYRHKKSRERDGN 95
CXXC
CtCFP1
Cps40 73 ---ITGKN 77
DmCFP1 182 MGKRCGTCEGCRRP-NCNQCDACRVRV--G----HKPRCIFRTCVVQAATVLKES 229
DrCFP1 119 SARMCGECEPCTRTEDCGHCDFCKDMKKFGGPNKIRQKCRLRQCVVRARKMLRVRDEEFSLRERKDNI-- 186
HsCFP1 165 SARMCGECEACRRTEDCGHCDFCRDMKKFGGPNKIRQKCRLRQCQLRARESYKYFPSSLSPVTPSESLPR 234
CtCFP1 209 DGGRYTT RFKK 219
Cps40 78 KDAI INGEGSLPKT LWKR 95
DmCFP1 230 QATQAGPSRKREKAAPKS RNVQVGPRAASPEIFLNPELQGIR 271
DrCFP1 214 SWGSEDDDGQLYSPVPRKKAIKVKHVKRRDKKFDKK----KESR--RHKQ---KQKHRDRLRHSDRTDGR 274
HsCFP1 305 PWMSDTEESPFLDPALRKRAVKVKHVKRREKKSEKKVMERKEERYKRHRQ---KQKHKDKWKHPERADAK 371
SID
CtCFP1 220 ------LCSFAGCQNPARIYDEARPSIFCSAEHCQAW----WEQLIS----TLPKK-RNP-GDPDYLTQE 273
Cps40 96 ------KCRISDCYKPC-----LQDSKYCSEEHGREFVNDIWSRLKTDEDRAVVKKMVEQTGHIDKFKKF 154
DmCFP1 272 ------QCYGPNCCSHA-----RPQSKYCSDKCGFNLA---TKRIFQ----VLPQRLQE 312
DrCFP1 275 HGGDTQQCLGPNCIEPA-----RPNSKYCSEDCGMKLA---ANRIYE----VLPQRIQQ 321
HsCFP1 372 DPASLPQCLGPGCVRPA-----QPSSKYCSDDCGMKLA---ANRIYE----ILPQRIQQ 418
RID
CtCFP1 344 QIAL--------C-RQMLQLIEL---A----------IARREAAIAAGI--------PG 372
Cps40 195 VQEISLPLFKKKL-ELLEVYLGW---LDNVYTEMRKLDDDAASHVECGKEDSKGTKRKKKKNSSRSRAR- 259
DmCFP1 342 ----------LAELEKRSEELNMVVE----------RAKRSSI--DTLGSQDTADM-- 375
DrCFP1 351 ----------LAEMERRFHELEGIIA----------KAKQQLVLQDEDVNETDSE--- 385
HsCFP1 448 ----------LQEMERRFHELEAIIL----------RAKQQAVREDEESNEGDSD--- 482
RID
CtCFP1 373 ITKDI------CGYDT---RLDTV----------GAAHQFSL-FLQSPQGQSIF--STRSLDNPPYVTTT 420
Cps40 260 --KNI------CGYCSTYERIPC----------SVEEF----------VRDFGSNEEA------ 289
DmCFP1 376 -EDEQSMYCITCGHEI-HSRTAIKHMEKCFNKYE-SQASFGSIFKT----------RMEG-NNMFCD 428
DrCFP1 386 -DTDLQIFCVSCSHPV-NPKVALRHMERCYAKYE-SQTSFGSIFPT----------RIEGATRLFCD 439
HsCFP1 483 -DTDLQIFCVSCGHPI-NPRVALRHMERCYAKYE-SQTSFGSMYPT----------RIEGATRLFCD 536
RID
CtCFP1 421 IKNPED-----DTPAAT---SSN--SAETAAGQDPRTAGM------CLRK-K-CKPHNGWGALLTKT-- 469
Cps40 290 -------T----------K----------IHEV-------CTKW-K-CNRHLDWVSTNQEQ-- 314
DmCFP1 429 FYNPASKTYCKRLRVLCPEHSKDPKVNDTDVCGSPLVNNVFNPTGEFCRAPKKNCFKHYAWEKIRRAEID 498
DrCFP1 440 VYNPQSKTYCKRLQVLCPEHSRDPKVTSDEVCGCPLVRNVFDPTGDYCRVSKRKCNKHYCWEKLRRAEVD 509
HsCFP1 537 VYNPQSKTYCKRLQVLCPEHSRDPKVPADEVCGCPLVRDVFELTGDFCRLPKRQCNRHYCWEKLRRAEVD 606
RID
CtCFP1 470 ---VR--------HDIR----------ELALQIR---ELLEA--EQRVRDGAAGRFMRRMKE 505
Cps40 315 ----------YLQQ--IDSLE----SMQERLQHLIQARKKQLNIQYYEEILRRGL----- 353
DmCFP1 499 LERVRQWLKMDDLMEQERV-----MRQQLTSRA-----NLLGLMLHSTYNHEVMDELVRKQQE 551
DrCFP1 510 LERVRVWYKLDELFEQERN-----VRTAMTNRA-----GLLALMLHQTIQHDPLTTDLRSSKD 562
HsCFP1 607 LERVRVWYKLDELFEQERN-----VRTAMTNRA-----GLLALMLHQTIQHDPLTTDLRSSAD 659

Figure S3. Sequence alignment of Cfp1. Top: domain structure of Cfp1. Bottom: A protein sequence alignment of full-length Cfp1 from *Chaetomium thermophilum (Ct), Saccharomyces cerevisiae (Cps40), Danio rerio (Dr), Drosophila melanogaster (Dm), and Homo sapiens (Hs).* The PHD (Yellow), CxxC (Purple), SID (green), and RID (grey) domains are shown on the top of the corresponding sequence. Evolutionary conserved residues are highlighted with Blosum62 color scheme. Residues matching the consensus sequence at that position are colored in dark blue while less conserved residues are rendered using different shades of blue. CtCfp1 residues 1-160, 274-343, and506-544 were omitted from the alignment.

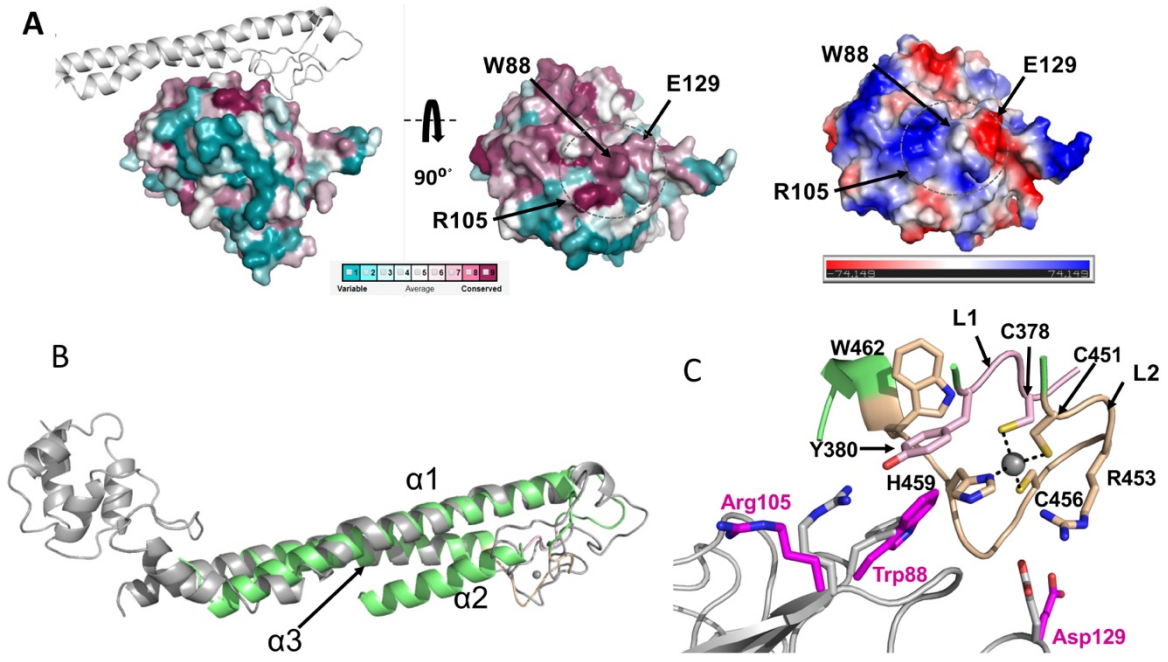

Figure S4. Interaction between Cfp1 RID and RbBP5 β-propeller. A) ZnF binding site on RbBP5 is conserved. Crystal structure of MtRbBP5 was docked into the cryo-EM structure of COMPASS. Cfp1 is rendered in grey and conservation of RbBP5 β-propeller surface residues are colored as in Figure 2. The middle panel shows an orthogonal view of RbBP5 β-propeller. Electrostatic potential surface of RbBP5 β-propeller is shown in the right panel. Electrostatic potentials are contoured from +10kbTe^{-1} (blue) to -10 kbTe^{-1} (red). B) Overlay of Cfp1 crystal structure (green) on top of the Cps40 cryo-EM structure (grey). Helix α1 and α3 of Cfp1 and Cps40 aligns with r.m.s.d. of ~1A. C) A zoomed view of Cfp1 RID ZnF. Cfp1 RID ZnF is colored as in Figure 2. The residues found to impact the interaction between RbBP5 and Cfp1 are highlighted. The residues on Cfp1 are labeled as one-letter code, while residues on MtRbBP5 in the cryo-EM structure are labeled as a 3-letter code. The residues from MtRbBP5 crystal structure are were overlaid on the top of Cps50 (yeast homolog in RbBP5, rendered in grey) from the COMPASS cryo-EM structure and rendered in magenta.

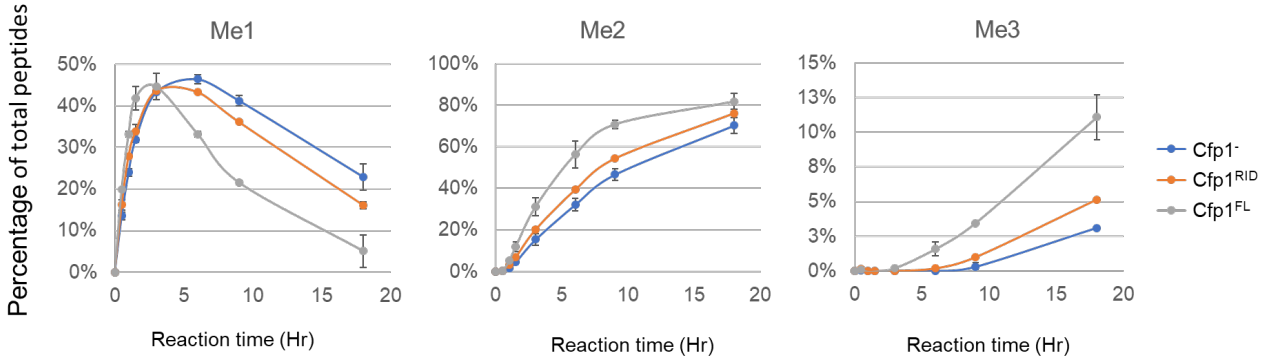

Figure S5. RID of Cfp1 alone is not sufficient to fully activate COMPASS. Methyltransferase assays were performed in presence of 0.3 µM of purified COMPASS with or without CtCfp1, H3 peptide substrates corresponding to the first 40 residues of histone H3 and S-adenosyl-L-methionine in the assay buffer. The proportion of H3K4me1 (left), H3K4me2 (middle) and H3K4me3 (right) peptides were reported as a percentage of the total amount of peptide used in the assay and reported on the Y-axis. Error bars indicate the s.d. of two independent experiments performed in triplicate.

TABLES

Supplementary Table S1 : ICP-MS analysis of purified Cfp1 RID domain concentrated at 100µM.	
Sample	***Zn concentration (µM)***
protein	119.41 ± 5.31
Buffer	0.04 ± 0.01

Supplementary table S2: Data collection and refinement statistics for the CtCfp1 RID domain	
PDB Accession number	
Data collection	
Space group	I222
Cell dimensions	
a, b, c (Å)	67.3, 71.6, 78.1
Resolution (Å)	51.98 – 2.31 (2.44 – 2.31)
R_{meas}	0.043 (0.377)
$I / \sigma I$	15.0 (2.7)
Completeness (%)	99.2 (96.7)
Redundancy	13.3 (13.0)
CC1/2	0.998
Refinement	

Resolution (Å)	49.0 – 2.3
No. reflections	8459
R_{work} / R_{free}	23.5/27.7
No. Atoms	
CtCfp1	1075
Zn^{2+}	1
Water	22
B-factors (Å^2)	
Protein	52.2
Zn^{2+}	75.6
Water	47.9
R.m.s. deviations	
Bond lengths (Å)	0.008
Bond angles (°)	0.904
Molprobity scores	1.24
Ramachandran favored (%)	97.8
Ramachandran allowed (%)	1.2

! Highest resolution shell is shown in parentheses.