6wrh – Papain-like Protease Domain of nsp3

Can an excellent PDB model be improved?

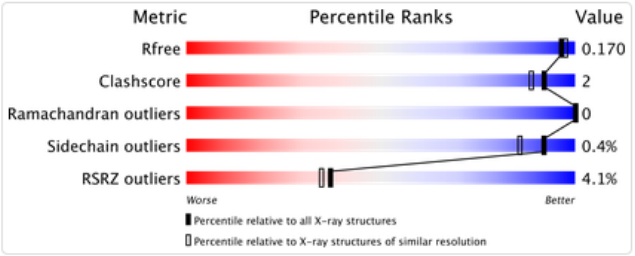
Author: Dale E. Tronrud

Initial Impressions

Experimental Data Snapshot

Method: X-RAY DIFFRACTION
Resolution: 1.60 Å
R-Value Free: 0.164
R-Value Work: 0.123
R-Value Observed: 0.126

wwPDB Validation 3D Report | Full Report



|

All-Atom Contacts	Clashscore, all atoms:	1.72		99th percentile* (N=718, 1.60Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	1	0.34%	Goal: <0.3%
	Favored rotamers	285	97.60%	Goal: >98%
	Ramachandran outliers	0	0.00%	Goal: <0.05%
	Ramachandran favored	306	97.45%	Goal: >98%
	Rama distribution Z-score	0.22 ± 0.43		Goal: abs(Z score) < 2
	MolProbity score^	1.06		99th percentile* (N=7200, 1.60Å ± 0.25Å)
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	2 / 2697	0.07%	Goal: 0%
	Bad angles:	3 / 3682	0.08%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 12	0.00%	Expected: ≤1 per chain, or ≤5%
Additional validations	Chiral volume outliers	0/407		
	Waters with clashes	10/380	2.63%	See UnDowser table for details

Property	Value	Source
Space group	P 32 2 1	Depositor
Cell constants a, b, c, α, β, γ	82.40Å 82.40Å 134.50Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	48.99 – 1.60 48.94 – 1.60	Depositor EDS
% Data completeness (in resolution range)	99.0 (48.99-1.60) 99.0 (48.94-1.60)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.79 (at 1.60Å)	Xtriage
Refinement program	REFMAC 5.8.0258	Depositor
R, R_{free}	0.123 , 0.164 0.135 , 0.170	Depositor DCC
R_{free} test set	3516 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å^2)	26.0	Xtriage
Anisotropy	0.151	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 45.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.024 for -h,-k,l	Xtriage
F_o,F_c correlation	0.98	EDS
Total number of atoms	3009	wwPDB-VP
Average B, all atoms (Å^2)	31.0	wwPDB-VP

This model is rated as excellent by nearly all the validation measures. It is middle of the road for the RSRZ outliers, but when I checked the details in the full validation report on the PDB I found that most of the misfits are at the N and C terminus. These regions of a protein are often disordered and have poor fit. Many people would have simply not build amino acids into such poor density, and achieved a better validation score as a reward for not trying. The fact that these residues fit their density poorly does not indicate that the rest of the model has problems.

While it is unusual with 1.6 Å data, this model has been refined with full anisotropic B factors. I will check the validity of this choice later.

Reprocessing the Images

Before attempting any serious refinement I wanted to see if the data could be improved by reprocessing the images. Anything that could improve the data would help. Kay_Diederichs was recruited to the project and his help was very useful.

The images can be found at https://proteindiffraction.org/project/IDP51000_6wrh/.The citation is https://doi.org/10.18430/m36wrh.

Kay integrated the images using xdsgui and sent the results to me. I used CCP4 to merge the free flags from the depositors structure factor file with Kay's data to create the file named 6wrh_xds_free.mtz. The report Kay received from XDS is in the table on the next page.

Inspection of the Model

The deposited model was created via Molecular Replacement (HKL-3000) using the SARS-CoV-2 model 6w9c and refined with REFMAC to the R Values 12.6%/16.4%.

The crystal contains a single protein molecule in the asymmetric unit, in space group P 3_2 2 1. Other than the N and C termini and a few side chains on the protein's surface, the protein is well ordered throughout. The electron density map is clear and gives good indication for the locations of the main chain carbonyl bumps and the side chain atom locations. There are a few places where side chains appear to have a small number of alternative conformations. The depositors have built both conformations in most cases.

The Zinc-Cys_4 binding site is well ordered and the model fits the density well.

The model also contains three Phosphate radicals, two Chlorine atoms, and two Glycerol cryoprotectant molecules. The occupancy of the Zinc atom, two of the Phosphates, and the Chlorine atoms are not equal to one. Refmac cannot refine occupancies and the practice is to manually change occupancies to remove difference map peaks. This appears to have been done in this model since most of the occupancies have values with "0" in the hundredths digit.

Yes, I said "Chlorine" atoms. The atoms are defined to have a neutral charge, despite the lack of covalent bonds.

6wrh contains 380 water molecules. This is just more than one water molecule per residue, which is low for this resolution. I, also, am conservative when building water so I am happy with this choice.

SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE >= -3.0 AS FUNCTION OF RESOLUTION

RESOLUTION LIMIT	NUMBER OF REFLECTIONS OBSERVED	UNIQUE	POSSIBLE	COMPLETENESS OF DATA	R-FACTOR observed	R-FACTOR expected	COMPARED	I/SIGMA	R-meas	CC(1/2)	Anomal Corr	SigAno	Nano
4.42	48802	6425	6434	99.9%	3.4%	3.8%	48793	49.03	3.7%	99.9*	61*	1.478	2834
3.13	84833	11586	11592	99.9%	4.2%	4.0%	84824	44.63	4.5%	99.9*	52*	1.341	5415
2.56	110950	14998	15006	99.9%	5.6%	5.0%	110921	32.55	6.1%	99.8*	29*	1.167	7100
2.22	134306	17758	17768	99.9%	8.0%	7.4%	134283	22.63	8.6%	99.6*	10*	0.953	8485
1.98	159082	20168	20170	100.0%	12.3%	12.3%	159079	14.95	13.2%	99.3*	1	0.804	9710
1.81	169599	22311	22319	100.0%	23.4%	28.2%	169571	7.00	25.1%	97.9*	-4	0.669	10749
1.68	187166	24134	24161	99.9%	51.2%	71.0%	187136	3.02	54.9%	91.2*	-1	0.601	11671
1.57	122580	23147	26031	88.9%	86.4%	125.7%	121548	1.26	95.7%	67.4*	1	0.562	10565
1.48	65712	18303	27687	66.1%	157.7%	233.2%	62979	0.50	182.0%	27.0*	-1	0.523	6964
total	1083030	158830	171168	92.8%	6.0%	6.3%	1079134	14.43	6.4%	99.9*	14*	0.808	73493

Interesting Things about nsp3/Papain-like Protease

The principal catalytic residue for the protease functionality is Cys 111. This residue is assisted by His 272 and Trp 108. In this particular protein residue 111 has been mutated to Serine. This change is very conservative and should eliminate the catalytic activity without significant disruption of the structure.

The only other model of the Papain-like Protease that I've examined closely is 6w9c. In that protein a Chloride ion appears to bind in a pocket "behind" the catalytic Cystine residue. While 6wrh contains two "Chlorine atoms" neither bind in the location where one binds in 6w9c. 6w9c is based on a very poor diffraction dataset and is rather unreliable.

The Zinc site is located at the distal end of the protein in the C terminal domain. The literature says that this region forms a "zinc finger". The zinc is held in place by four cystine residues – 189, 192, 224, and 226. You will note that the first pair of Cystine residues form a tight circle when bound to the Zinc, and an even more constrained cycle for the last pair. It is difficult to imagine much flexibility in the entire binding site.

Initial Refinement Goals

My goal is to see if any improvement can be made to this carefully refined model.

Refinement

I started by refining the deposited model against Kay's new data set using Buster/TNT. I like this package because I believe it produces the best maps for model building and ligand identification. In this case it is limited because it cannot refine anisotropic B factors. It will refine B factors using the TLS (Translation-Libration-Screw) model. My tests here showed that TLS refinement was little better than isotropic B factors.

In all my refinements I asked Buster to place Hydrogen atoms in riding locations at bond lengths consistent with X-ray scattering and with full occupancy. I have found that the presence of Hydrogen atoms has little effect on the R values but does improve the MolProbity assessment.

Following this refinement I walked through the entire model and map using Coot. I deleted residue "0" and most of residue "1". I also added alternative conformations for six amino acids – Ser 170, Thr 197, Met 206, Glu 238, Try 268, and Thr 301. I also deleted

a number of water molecules because either their 2Fo-Fc density wasn't strong enough or they didn't make proper hydrogen bonds. Most of these were water molecules created by Buster.

This model was passed back to Buster and then I made another pass in Coot. In this round I flipped several side chains to improve Hydrogen bonding (Asn 60, Gln 174, and Asn 186). I also created an alternative conformation for Ser 262 to relive a bad contact with Thr 115 and fit a small peak in the difference map. I changed rotomers for the B conformation of residue Ser 70, and Arg 138.

Gln 236 showed a bad contact with Glu 238. Both of these side chains have alternative conformations. Swapping the designation of "A" and "B" for 236 relived the conflict without moving any atoms.

I found that Phosphate 503 had no links to any part of the structure. I deleted it.

Glycerol 507 had a bad contact with a neighboring side chain. The density for this Glycerol does not cover the entire molecule so I assumed the creators of 6wrh believed it was disordered to some extent. I could not come up with a model that would relieve the clash and fit the density that was there. A check of the crystal growth conditions revealed that the buffer contained Acetate. I could build a model into this density that consisted of two alternative conformations of Acetate that fit the density, avoided the clash, and made a Hydrogen-bond.

Finally there was a difference density peak near the Oxygen atom of one of the partially occupied Phosphate radicals. I placed a water molecule (W|105) in that peak with the occupancy consistent with the Phosphate and made them both alternative conformations.

Since my plan was to finish refinement in Refmac and that program will not refine occupancies, I did a refinement in Phenix with isotropic B factors and copied the occupancies it produced back into my model-built coordinate file.

The model now seemed to be in good shape. I took it to Refmac and refinement with anisotropic B-factors.

Model		Rwrk	Rfree	Bond	Angle	#H2O	
Deposited		12.3%	16.4%	0.009	1.433	380	Against their data
Phase 2r	Buster	16.13%	18.82%	0.008	0.933	463	
Phase 3r	Refmac	12.1%	15.9%	0.017	1.829	480	

One matter that came up after the Refmac refinement: Refmac produces an anomalous scattering difference map. When I examined it I found that I could see peaks for the Zinc ion and most of the Sulfur atoms and Chloride ions. I also found a peak on a water molecule! Looking at this molecule more carefully than before I could see that its interactions with its neighbors were more like that of a Chloride ion. I redesignated the "water molecule" as a Chloride ion and reran the Refmac refinement. (Which is the run reported above.)

Refinement with anisotropic B factors dropped the free R by nearly three percentage points and all of my changes resulted in a model with a half percentage point drop from the deposited model. The stereochemistry rmsd's are larger, particularly that for bond lengths, but my values are not unreasonable for a quality high resolution model.

All-Atom Contacts	Clashscore, all atoms:	1.76		99th percentile* (N=718, 1.60Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	1	0.34%	Goal: <0.3%
	Favored rotamers	289	96.98%	Goal: >98%
	Ramachandran outliers	0	0.00%	Goal: <0.05%
	Ramachandran favored	305	97.44%	Goal: >98%
	Rama distribution Z-score	0.23 ± 0.42		Goal: abs(Z score) < 2
	MolProbity score^	1.06		99th percentile* (N=7200, 1.60Å ± 0.25Å)
	Cβ deviations >0.25Å	2	0.63%	Goal: 0
	Bad bonds:	10 / 2726	0.37%	Goal: 0%
	Bad angles:	10 / 3731	0.27%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 12	0.00%	Expected: ≤1 per chain, or ≤5%
Additional validations	Chiral volume outliers	0/414		
	Waters with clashes	23/439	5.24%	See UnDowser table for details

The MolProbity report for my model is nearly identical to that of the deposited model, with the exception of a significant increase in the number of outliers for bond lengths and angles. I do not understand this result and will have to look into it further. I have looked at some of these instances and can see nothing outstanding about them that could cause a problem.

Summary

It is possible to improve even upon a well refined model in the PDB. The changes I made are small, and do not affect the fold of the protein or cause large movements of atoms. These changes do improve the Hydrogen bonding network in their local areas. In addition the removal of a Phosphate radical and the inclusion of an additional Chloride ion will affect any electrostatic calculation based on this model.

The models are, however, nearly identical in the active and substrate binding sites.

Acknowledgements

Refinement was performed using computer access donated by the Guillemin Lab at the University of Oregon. The principal software packages used were Coot, Buster/TNT, Refmac, and CCP4.

.