Supporting Information for

Homology models of Wuhan Coronavirus 3CL^{pro} Protease

Dr. Martin J. Stoermer
Division of Chemistry and Structural Biology, Institute for Molecular Bioscience
The University of Queensland
St. Lucia, 4072, Queensland, Australia
E-mail: martin.stoermer@uq.edu.au
Orcid: https://orcid.org/0000-0003-3445-2104
Twitter: @MartinStoermer

Nomenclature

Original Swissmodel homology models were numbered 1-8. However due to redundancy monomer Models 1,2,4,6,7,8 were taken through to refinement. Model 8 was only obtained as a monomer, so just 5 models were refined as their homodimers. For clarity and to define their point of origin, these were named Models11,12,14,16,17&18.

E.g.:

Raw Model	PDB Template	Refined monomer model	Refined dimer model
1 (dimer)	2z9j	Model 1	Model 11
2 (dimer)	3vb3	Model 2	Model 12
3 (dimer)	2z9j		
4 (dimer)	1uk3	Model 4	Model 14
5 (dimer)	1uk3		
6 (dimer)	2a5i	Model 6	Model 16
7 (dimer)	1uj1	Model 7	Model 17
8 (monomer)	1z1i	Model 8	

Table S1. Naming system used for the models at each stage of refinement.

```
#=======================================
#
# Aligned_sequences: 2
# 1: WH-Human_1pr
# 2: BatHKU4_pr (2ynb)
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 314
# Identity:     154/314 (49.0%)
# Similarity:   203/314 (64.6%)
# Gaps:          16/314 ( 5.1%)
# Score: 789.0
#
#=======================================

WH-Human_1pr       1 SGFRKMAFPSGKVEGCMVQVTCGTTTLNGLWLDDVVYCPRHVICTSEDML     50
                     ||..||:.|||.||.|:||||||:.|||||||:.|:||||::|.::.:.
BatHKU4_pr         1 SGLVKMSAPSGAVENCIVQVTCGSMTLNGLWLDNTVWCPRHIMCPADQLT     50

WH-Human_1pr      51 NPNYEDLLIRKSNHNFLVQ---AGNVQLRVIGHSMQNCVLKLKVDTANPK     97
                     :|||:.|||.|:||:|:||   .....|||:.|||...:|||.||.|||.
BatHKU4_pr        51 DPNYDALLISKTNHSFIVQKHIGAQANLRVVAHSMVGVLLKLTVDVANPS    100

WH-Human_1pr      98 TPKYKFVRIQPGQTFSVLACYNGSPSGVYQCAMRPNFTIKGSFLNGSCGS    147
                     ||.|.|..::||.:|||||||||.|:||:...:|.|.|||||||.|||||
BatHKU4_pr       101 TPAYTFSTVKPGASFSVLACYNGKPTGVFTVNLRHNSTIKGSFLCGSCGS    150

WH-Human_1pr     148 VGFNIDYDCVSFCYMHHMELPTGVHAGTDLEGNFYGPFVDRQTAQAAGTD    197
                     ||:..:...::|.|||.|||..|.|.|:..:|..||.|.|:||.|...||
BatHKU4_pr       151 VGYTENGGVINFVYMHQMELSNGTHTGSSFDGVMYGAFEDKQTHQLQLTD    200

WH-Human_1pr     198 TTITVNVLAWLYAAVINGDRWFLNRFTTTLNDFNLVAMKYNYEPLTQDHV    247
                     ...|:||:|||||||:||.:||:......:       :.||...|:....
BatHKU4_pr       201 KYCTINVVAWLYAAVLNGCKWFVKPTRVGI-------VTYNEWALSNQFT    243

WH-Human_1pr     248 DILGP-----LSAQTGIAVLDMCASLKELLQNGMNGRTILGSALLEDEFT    292
                     :.:|.     |:.:||::|..|.|:::. |..|..|:||||.:.||||||
SEQUENCE         244 EFVGTQSIDMLAHRTGVSVEQMLAAIQS-LHAGFQGKTILGQSTLEDEFT    292

WH-Human_1pr     293 PFDVVRQCSGVTFQ    306
                     |.||..|..||..|
BatHKU4_pr       293 PDDVNMQVMGVVMQ    306
```

Figure S1. EMBOSS Needle alignment of the extracted protease sequence of Wuhan coronavirus against Bat coronavirus (PDB:2YNB) Catalytic dyads (His41, Cys145) shown in red bold.

```
#=======================================
#
# Aligned_sequences: 2
# 1: WH-Human_1pr
# 2: SARS_pr (2z9j_A)
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 306
# Identity:     294/306 (96.1%)
# Similarity:   302/306 (98.7%)
# Gaps:           0/306 ( 0.0%)
# Score: 1600.0
#
#=======================================

WH-Human_1pr       1 SGFRKMAFPSGKVEGCMVQVTCGTTTLNGLWLDDVVYCPRHVICTSEDML     50
                     |||||||||||||||||||||||||||||||||.||||||||||:||||
SARS_pr            1 SGFRKMAFPSGKVEGCMVQVTCGTTTLNGLWLDDTVYCPRHVICTAEDML     50

WH-Human_1pr      51 NPNYEDLLIRKSNHNFLVQAGNVQLRVIGHSMQNCVLKLKVDTANPKTPK    100
                     ||||||||||||||:||||||||||||||||||:|:|||||:|||||||
SARS_pr           51 NPNYEDLLIRKSNHSFLVQAGNVQLRVIGHSMQNCLLRLKVDTSNPKTPK    100

WH-Human_1pr     101 YKFVRIQPGQTFSVLACYNGSPSGVYQCAMRPNFTIKGSFLNGSCGSVGF    150
                     ||||||||||||||||||||||||||||||||.||||||||||||||||
SARS_pr          101 YKFVRIQPGQTFSVLACYNGSPSGVYQCAMRPNHTIKGSFLNGSCGSVGF    150

WH-Human_1pr     151 NIDYDCVSFCYMHHMELPTGVHAGTDLEGNFYGPFVDRQTAQAAGTDTTI    200
                     ||||||||||||||||||||||||||||.||||||||||||||||||||
SARS_pr          151 NIDYDCVSFCYMHHMELPTGVHAGTDLEGKFYGPFVDRQTAQAAGTDTTI    200

WH-Human_1pr     201 TVNVLAWLYAAVINGDRWFLNRFTTTLNDFNLVAMKYNYEPLTQDHVDIL    250
                     |:|||||||||||||||||||||||||||||||||||||||||||||||||
SARS_pr          201 TLNVLAWLYAAVINGDRWFLNRFTTTLNDFNLVAMKYNYEPLTQDHVDIL    250

WH-Human_1pr     251 GPLSAQTGIAVLDMCASLKELLQNGMNGRTILGSALLEDEFTPFDVVRQC    300
                     ||||||||||||||:||||||||||||||||.:|||||||||||||||
SARS_pr          251 GPLSAQTGIAVLDMCAALKELLQNGMNGRTILGSTILEDEFTPFDVVRQC    300

WH-Human_1pr     301 SGVTFQ    306
                     ||||||
SARS_pr          301 SGVTFQ    306
```

Figure S2. EMBOSS Needle alignment of the extracted protease sequence of Wuhan coronavirus against SARS coronavirus (PDB: 2z9j) Catalytic dyads shown in red bold.

MolProbity scoring and statistics for refined monomer models 1,2,4,6,7,8 and dimer models

Monomer models.

Model	Template name	Ramachandran favoured (%)	Ramachandran allowed (%)	Clashscore (percentile)	MolProbity score (percentile)	CA Geometry outliers (%)	Cβ deviations (%)
1	2z9j.1.B	96.33	100	1.52 (99th)	1.36 (98th)	0.0	0.0
2	3vb3.1.A	96.66	100	1.30 (99th)	1.12 (100th)	0.0	0.0
4	1uk3.1.B	93.98	99.67	1.52 (99th)	1.74 (88th)	0	0
6	2a5i.1.A	96.38	99.67	0.63 (99th)	0.95 (100th)	0	0
7	1uj1.1.B	96.66	99.67	1.52 (99th)	1.43 (96th)	0	0
8	1z1i.1.A	91.30	99.33	3.04 (98th)	2.02 (75th)	0.34	0.36

Table S2. Statistics via MolProbity with no His/Asn/Gln flips allowed.

Model	Template name	Ramachandran favoured (%)	Ramachandran allowed (%)	Clashscore (percentile)	MolProbity score (percentile)	CA Geometry outliers (%)	Cβ deviations (%)
1 (1 flip)	2z9j.1.B	96.33	100	1.30 (99th)	1.32 (98th)	0.0	0.0
2 (2 flips)	3vb3.1.A	96.66	100	1.30 (99th)	1.12 (100th)	0.0	0.0
4 (0 flips)	1uk3.1.B	93.98	99.67	1.52 (99th)	1.74 (88th)	0	0
6 (2 flips)	2a5i.1.A	96.38	99.67	0.63 (99th)	0.95 (100th)	0	0
7 (1 flip)	1uj1.1.B	96.66	99.67	1.52 (99th)	1.43 (96th)	0	0
8 (3 flips)	1z1i.1.A	91.30	99.33	3.04 (98th)	2.02 (75th)	0.34	0.36

Table S3. Statistics via MolProbity with His/Asn/Gln flips allowed. 5/6 models had the possibility of suggested hydrogen bonding flips, but only the top model1 resulted in any improvement to the clash and MolProbity scores.

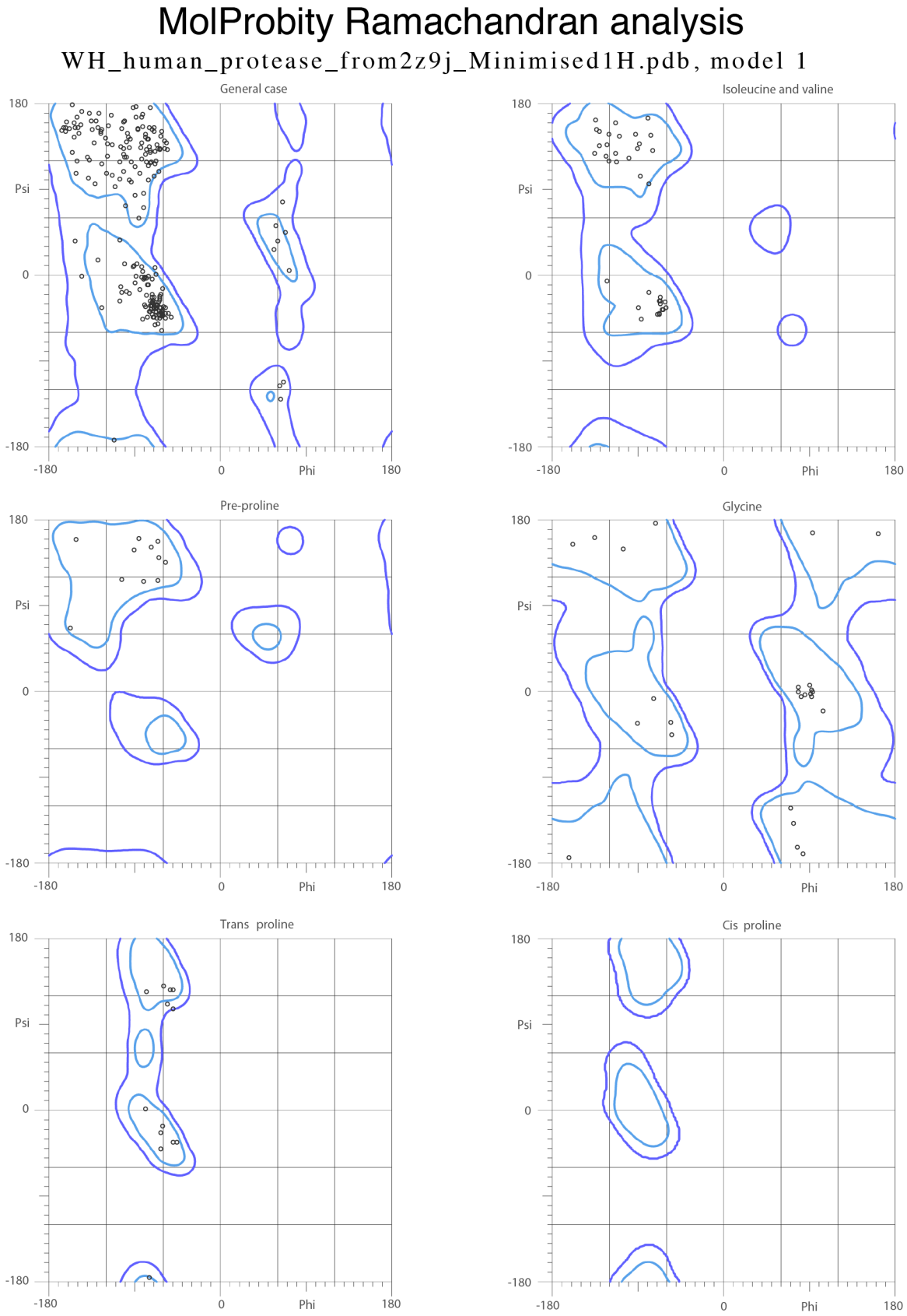
Dimer models:

Dimer	Template name	Ramachandran favoured (%)	Ramachandran allowed (%)	Clashscore (percentile)	MolProbity score (percentile)	CA Geometry outliers (%)	Cβ deviations (%)
11	2z9j.1.B	96.51	99.83	1.62 (99th)	1.41 (97th)	0.17	0.0
12	3vb3.1.A	96.35	100	1.18 (99th)	1.12 (100th)	0.0	0.0
14	1uk3.1.B	94.16	99.67	1.84 (99th)	1.73 (88th)	0.34	0
16	2a5i.1.A	95.72	99.67	0.73 (99th)	1.03 (100th)	0	0
17	1uj1.1.B	94.82	99.83	1.30 (99th)	1.58 (93rd)	0	0.18

Table S4. Statistics via MolProbity with no His/Asn/Gln flips allowed.

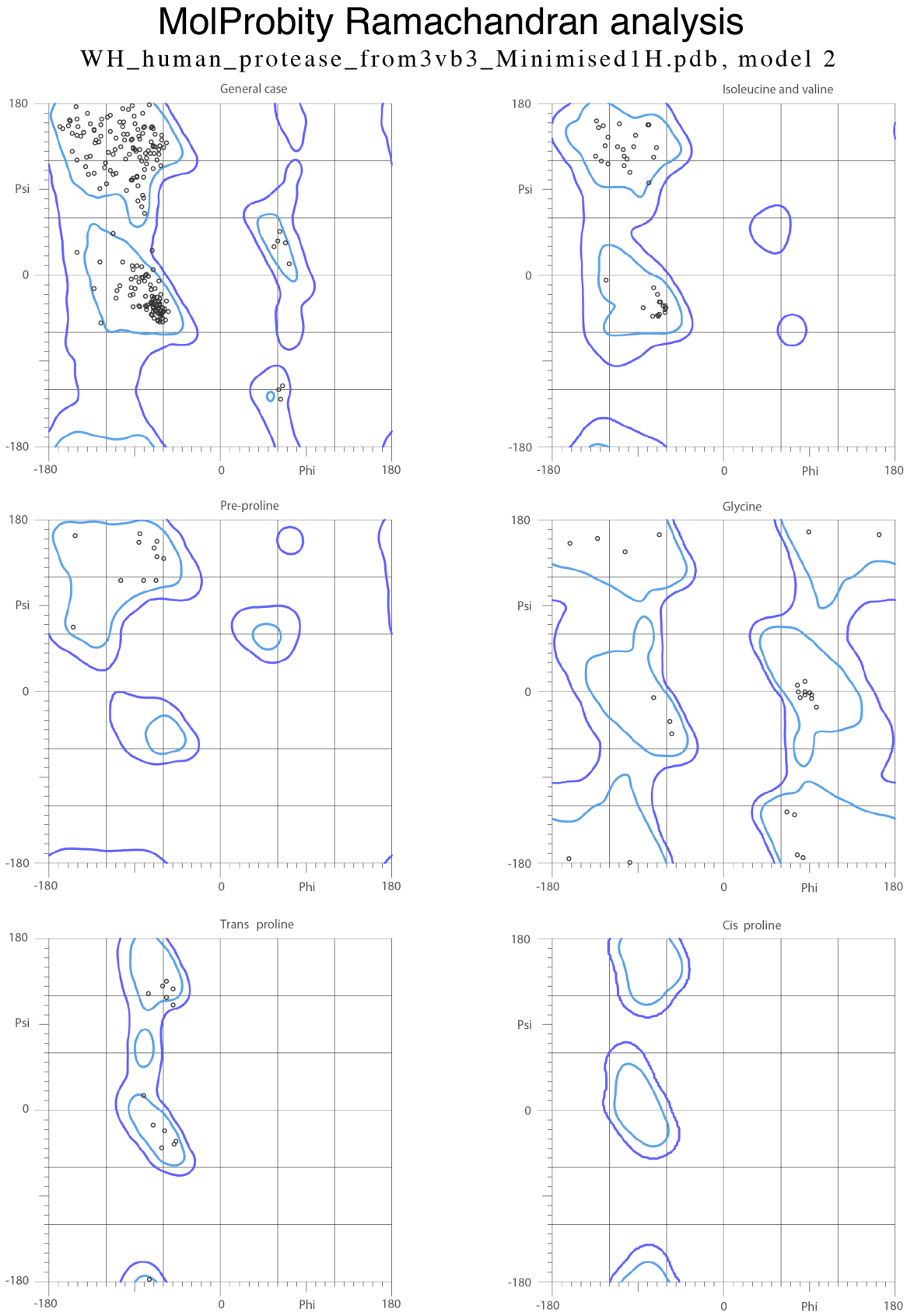
Dimer	Template name	Ramachandran favoured (%)	Ramachandran allowed (%)	Clashscore (percentile)	MolProbity score (percentile)	CA Geometry outliers (%)	Cβ deviations (%)
11 (2 flips)	2z9j.1.B	96.51	99.83	1.51 (99th)	1.40 (97th)	0.17	0.0
12 (4 flips)	3vb3.1.A	96.35	100	1.18 (99th)	1.12 (100th)	0.0	0.0
14 (1 flip)	1uk3.1.B	94.16	99.67	1.84 (99th)	1.73 (88th)	0.34	0
16 (4 flips)	2a5i.1.A	95.72	99.67	0.73 (99th)	1.03 (100th)	0	0
17 (3 flips)	1uj1.1.B	94.82	99.83	1.30 (99th)	1.58 (93rd)	0	0.18

Table S5. Statistics via MolProbity with His/Asn/Gln flips allowed.



96.3% (289/300) of all residues were in favored (98%) regions.
100.0% (300/300) of all residues were in allowed (>99.8%) regions.

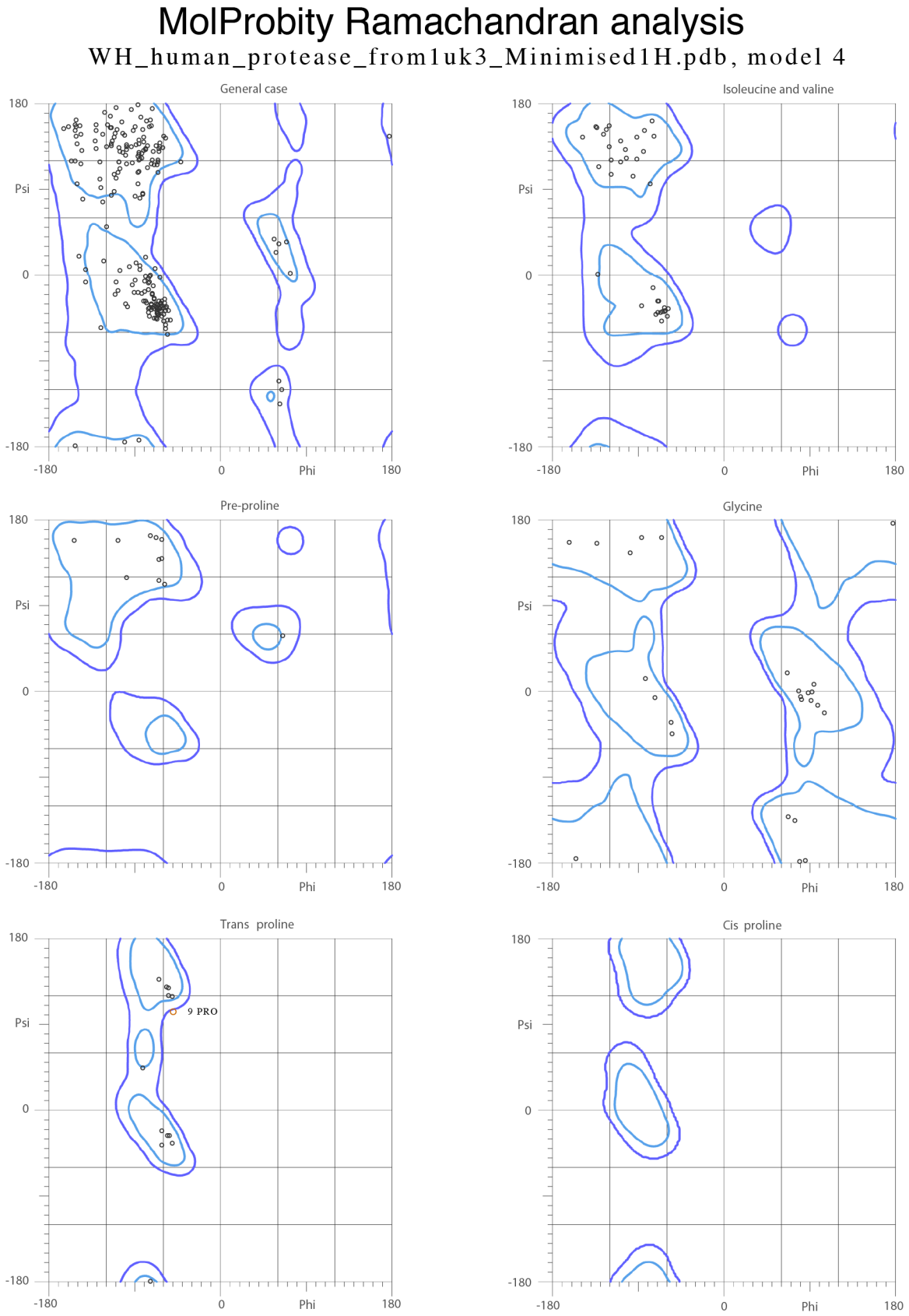
Figure S2. Ramachandran plots for Wuhan CoV 3CL^{pro} homology monomer model1



96.7% (289/299) of all residues were in favored (98%) regions.

100.0% (299/299) of all residues were in allowed (>99.8%) regions.

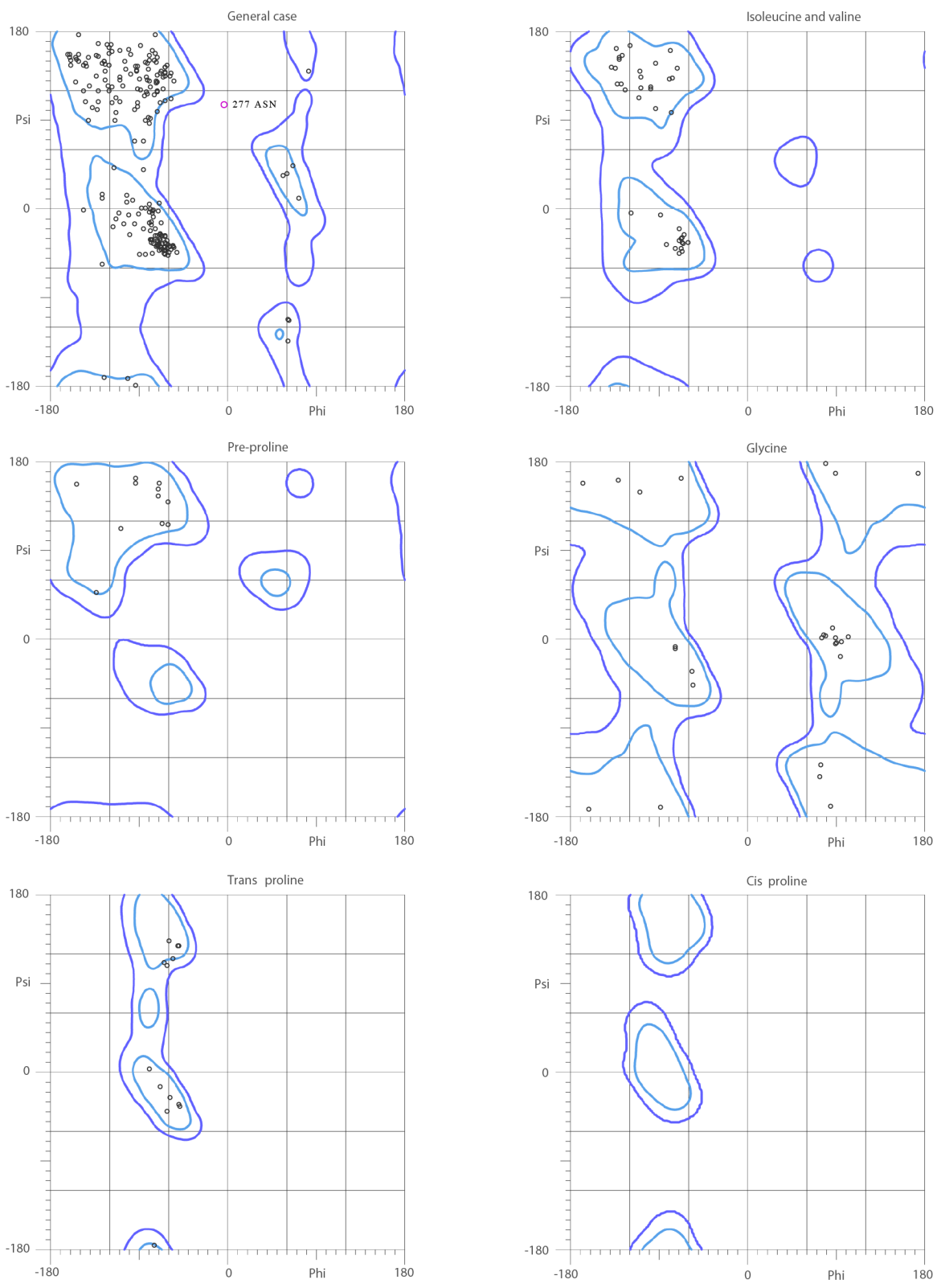
Figure S3. Ramachandran plots for Wuhan CoV 3CL^{pro} homology monomer model2



94.0% (281/299) of all residues were in favored (98%) regions.

99.7% (298/299) of all residues were in allowed (>99.8%) regions.

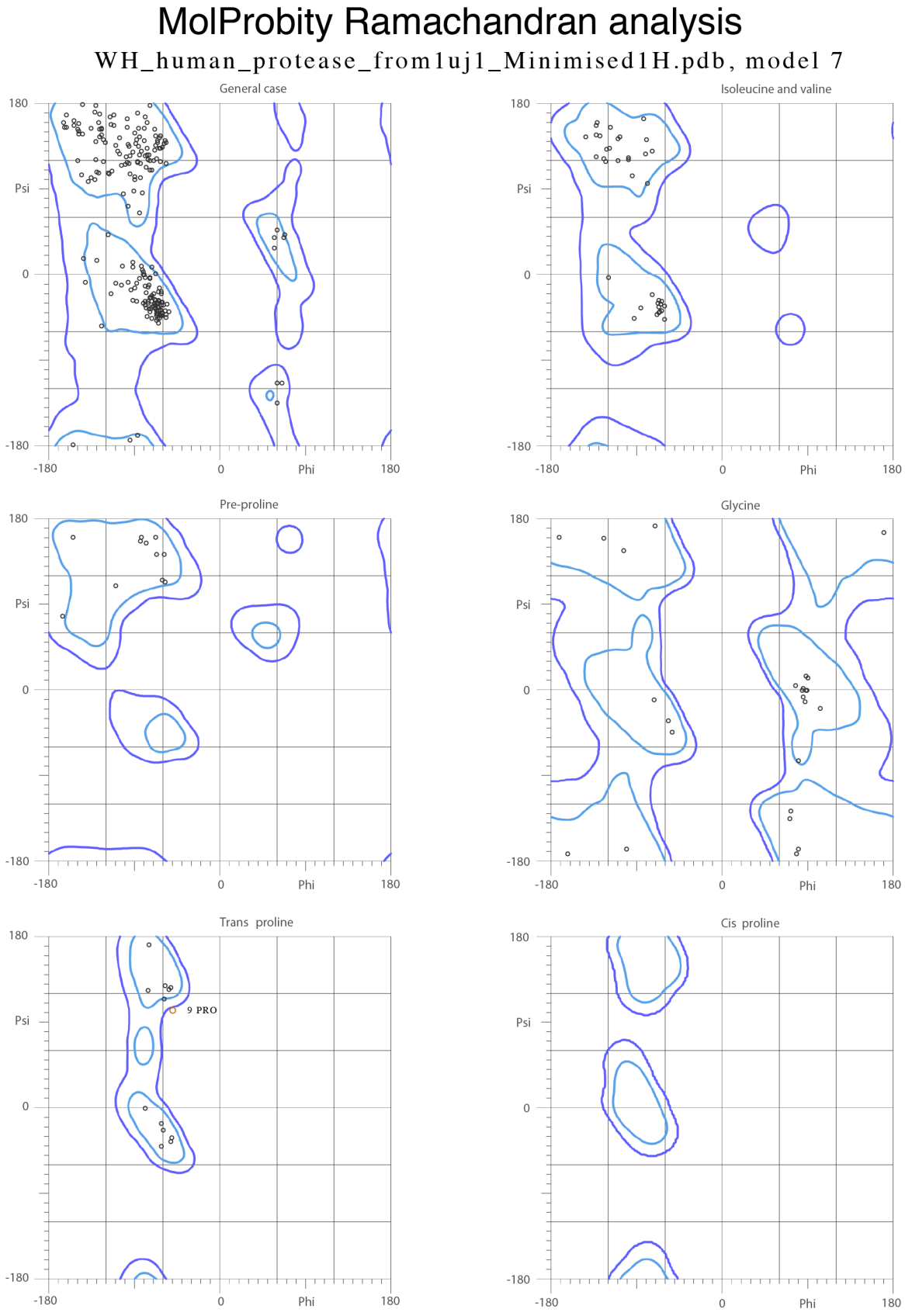
Figure S4. Ramachandran plots for Wuhan CoV 3CL^{pro} homology monomer model4



96.4% (293/304) of all residues were in favored (98%) regions.

99.7% (303/304) of all residues were in allowed (>99.8%) regions.

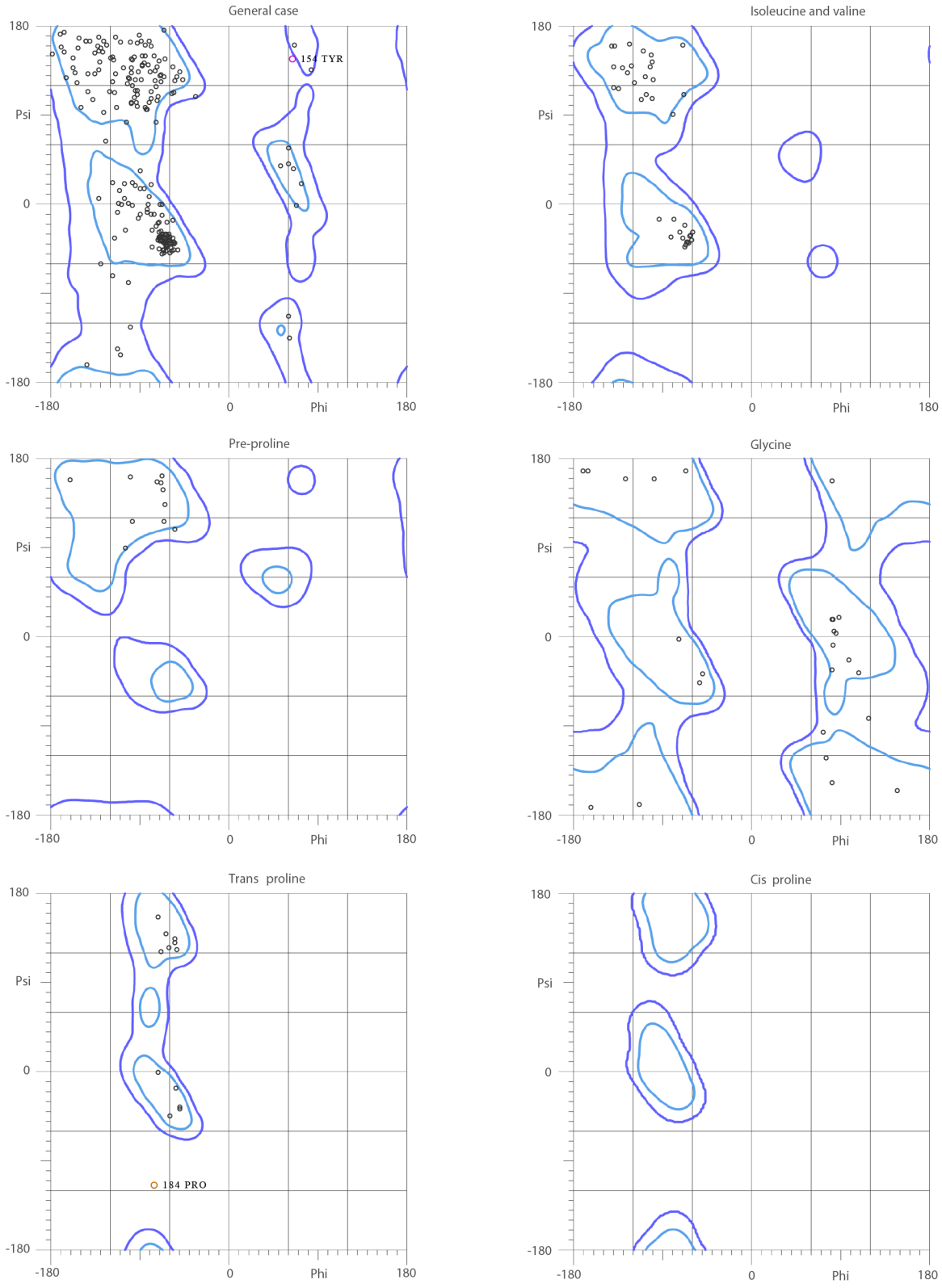
Figure S5. Ramachandran plots for Wuhan CoV 3CL^{pro} homology monomer model6



96.7% (289/299) of all residues were in favored (98%) regions.

99.7% (298/299) of all residues were in allowed (>99.8%) regions.

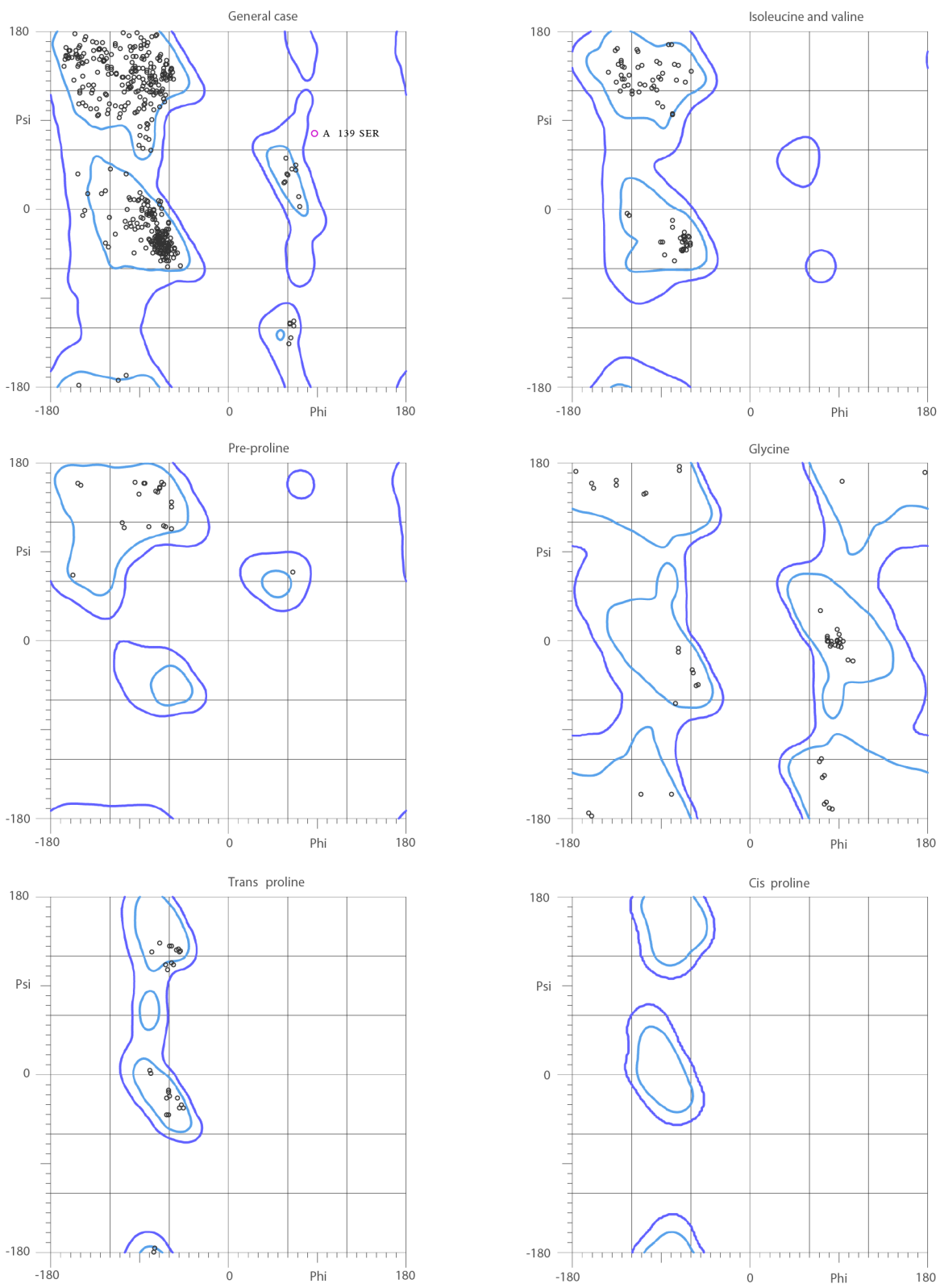
Figure S6. Ramachandran plots for Wuhan CoV 3CL^{pro} homology monomer model7



91.3% (273/299) of all residues were in favored (98%) regions.

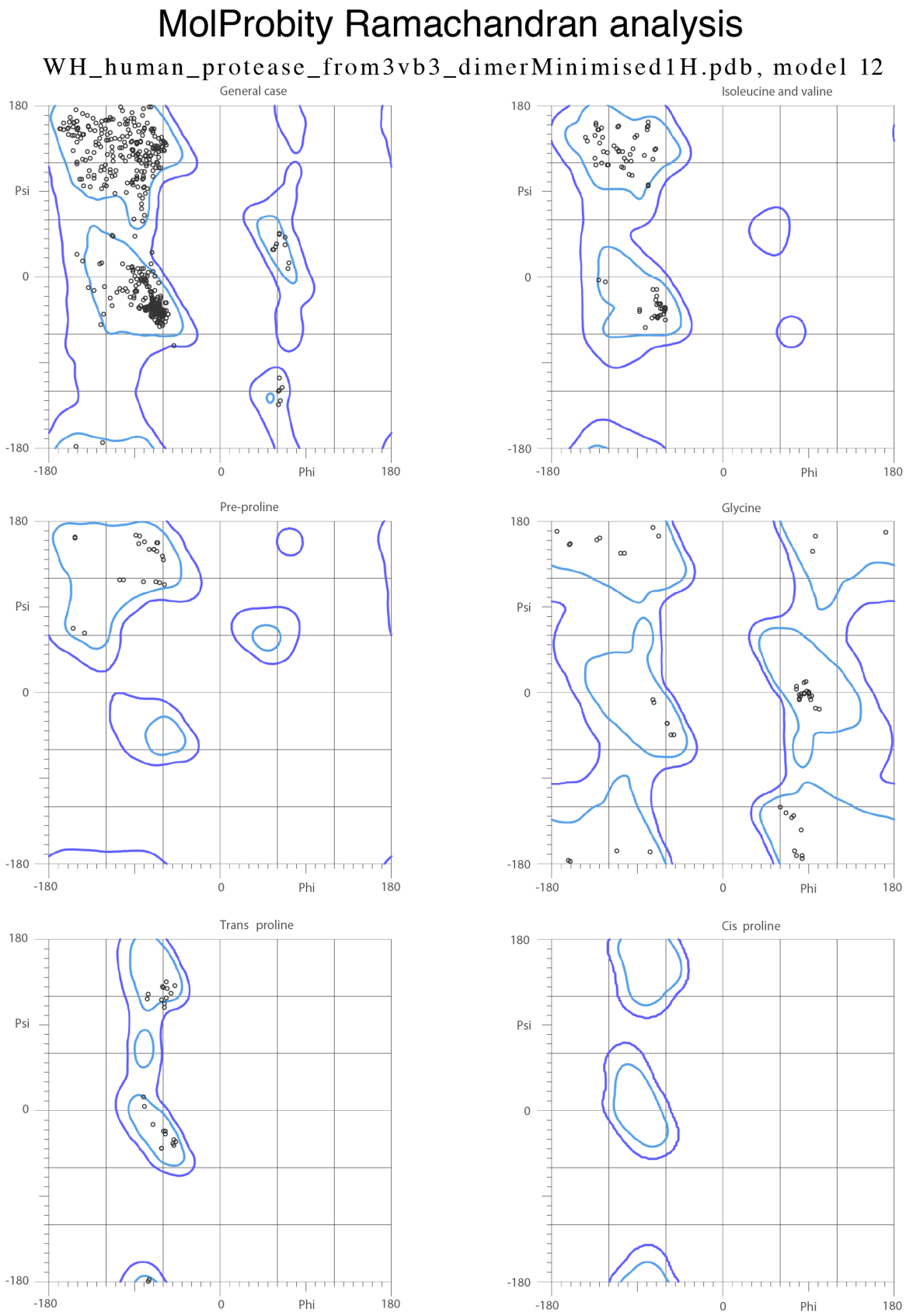
99.3% (297/299) of all residues were in allowed (>99.8%) regions.

Figure S7. Ramachandran plots for Wuhan CoV 3CL^{pro} homology monomer model8



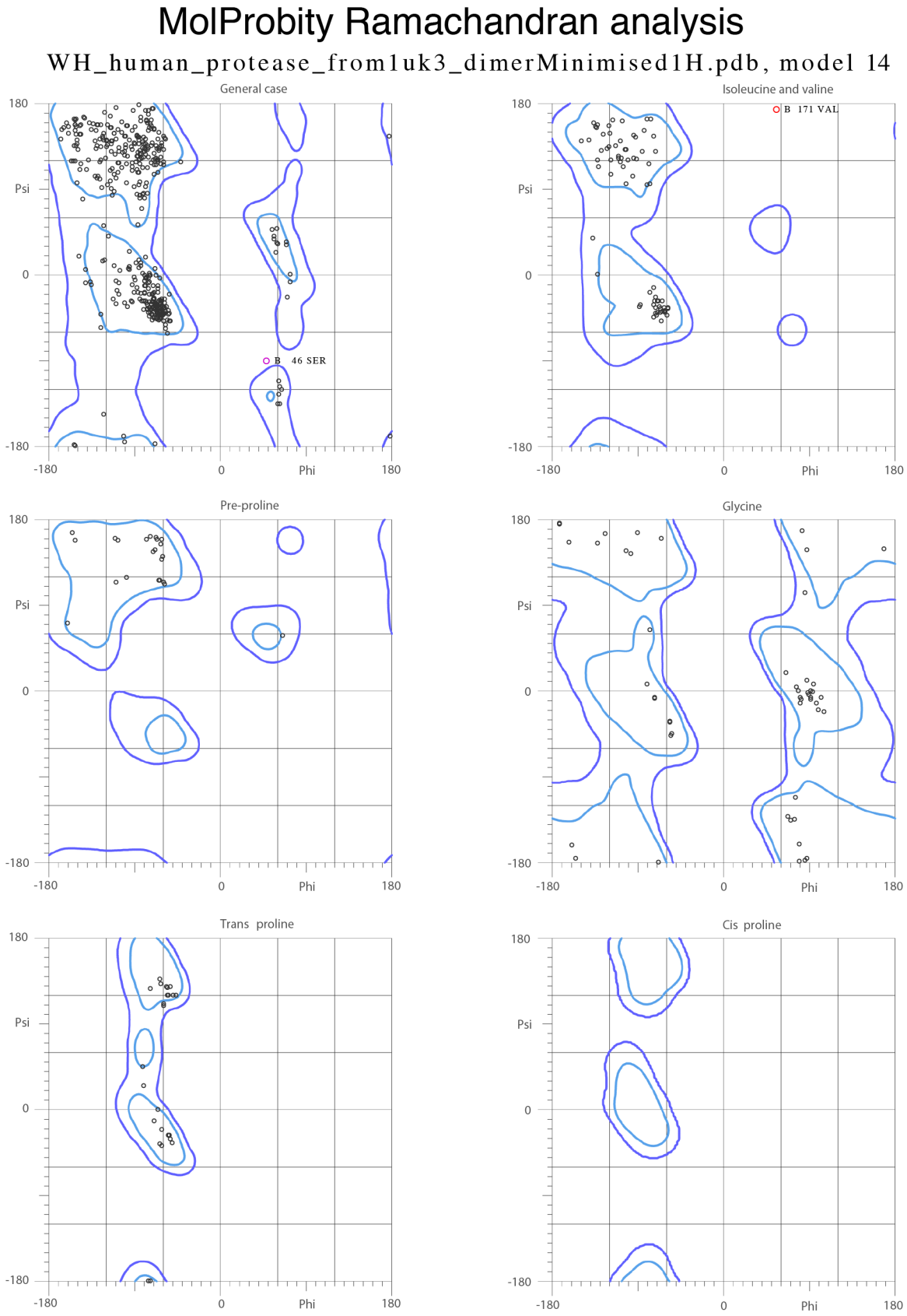
96.5% (581/602) of all residues were in favored (98%) regions.
99.8% (601/602) of all residues were in allowed (>99.8%) regions.

Figure S8. Ramachandran plots for Wuhan CoV 3CL^{pro} homology dimer model11



96.4% (581/603) of all residues were in favored (98%) regions.
100.0% (603/603) of all residues were in allowed (>99.8%) regions.

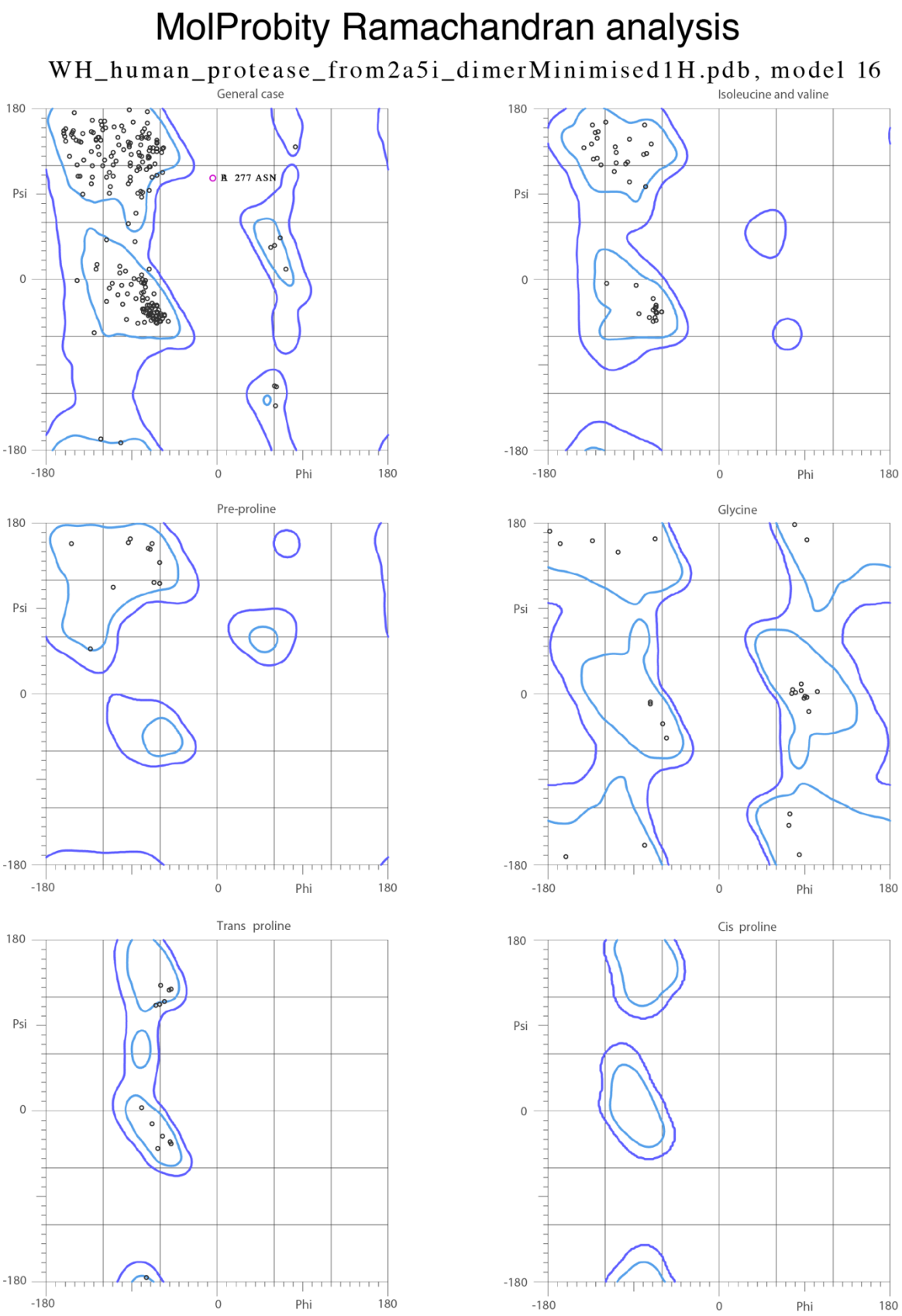
Figure S9. Ramachandran plots for Wuhan CoV 3CL^{pro} homology dimer model12



94.2% (564/599) of all residues were in favored (98%) regions.

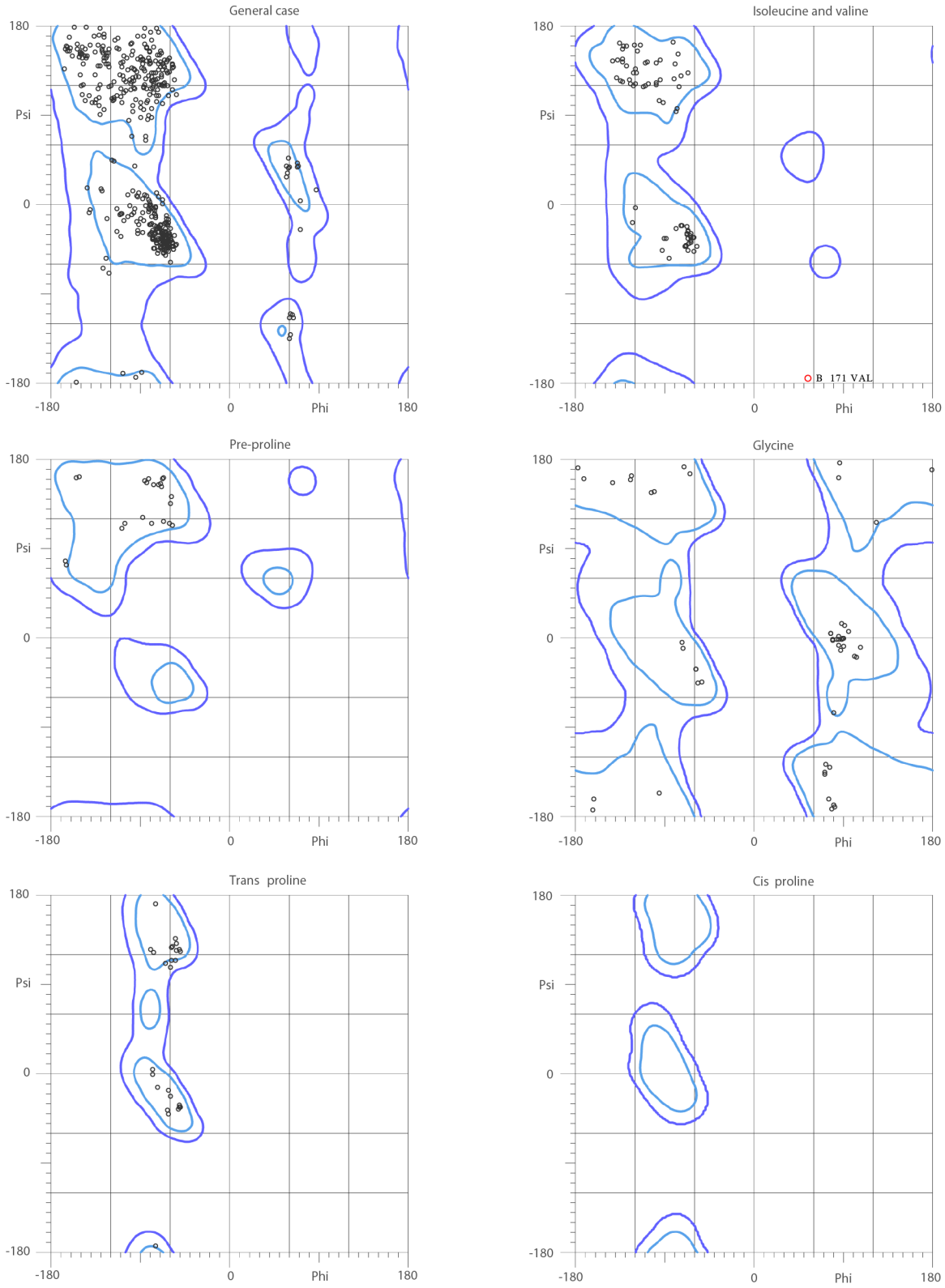
99.7% (597/599) of all residues were in allowed (>99.8%) regions.

Figure S10. Ramachandran plots for Wuhan CoV 3CL^{pro} homology dimer model14



95.7% (582/608) of all residues were in favored (98%) regions.
99.7% (606/608) of all residues were in allowed (>99.8%) regions.

Figure S11. Ramachandran plots for Wuhan CoV 3CL^{pro} homology dimer model16



94.8% (568/599) of all residues were in favored (98%) regions.

99.8% (598/599) of all residues were in allowed (>99.8%) regions.

Figure S12. Ramachandran plots for Wuhan CoV 3CL^{pro} homology dimer model17

Molecular Dynamics Method

Molecular dynamics simulations were performed with the Desmond Molecular Dynamics System (D. E. Shaw Research, New York, NY) by using the tools incorporated in the Schrödinger Suite 2019-2. The model system was set up by using the "System Setup" utility. The OPLS2005 force field parameters were implemented during simulations. Each monomer, dimer or ternary complex was placed in an orthorhombic box with solvent buffers up to 10Å along each dimension. The TIP3P solvent model was used to describe water molecules. Overall neutralisation of the system was achieved by adding sodium and chloride ions at the physiological concentration of 0.15M. The prepared model systems were then relaxed by using the multi-step default protocol in Desmond. After relaxation, the whole system was subjected to a 100ns simulation. The cut-off distance for computing short-range electrostatics and Lennard–Jones interaction was set to 9.0Å. The trajectories and energies were recorded every 100 and 1.2ps, respectively. MD simulations were performed at the IMB cluster from the University of Queensland on NVIDIA P100 GPUs or on a CentOS Linux Desktop with a Quadro RTX 5000 GPU. Simulation analysis was performed on Linux or Macintosh desktops using the Simulation Interactions Diagrams module in Schrödinger Suite 2019-2.

Model2 from SARS 3vb3 WH_human_protease_from3vb3_Minimised.pdb

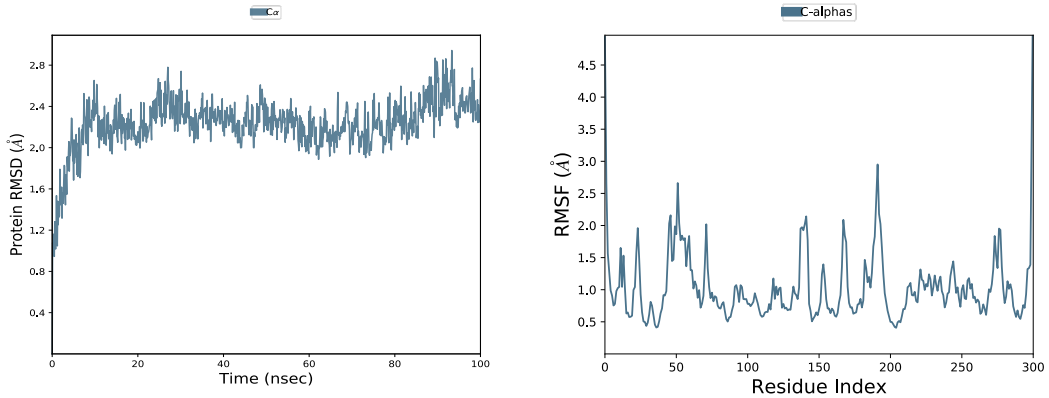


Figure S1-13. Protein C-alpha RMSD over the 100 ns duration molecular dynamics simulation. **Figure S1-14**. Protein C-alpha RMSF plot. Per residue variability in Cα position.

Flexible residues

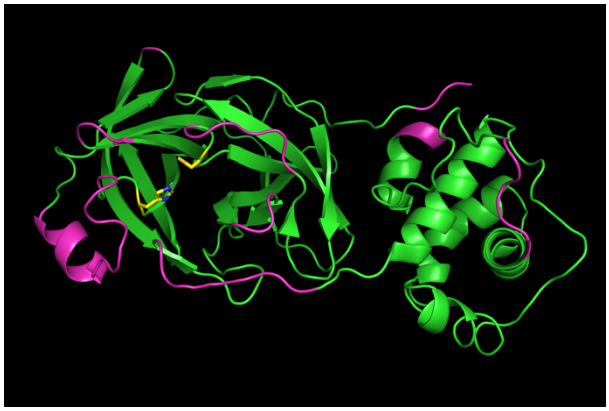


Figure S1-15. WH CoV 3CL^{pro} Model 2 shown as green ribbons, except catalytic dyad (His41, Cys145 – yellow sticks), and magenta ribbons for regions found to be mobile in 100 ns MD simulation.

Moderately mobile regions (>1.5 RMSF) N-terminal residues Ser1-Phe3, β-hairpin loop Gly23-Thr25, α-helix and loop Thr45-Arg60, loop tip Asn72, strand Gly138-Gly143, β-hairpin loop Pro168-Gly170, long strand Gln189-Gly195, strand Asn274-Gly278, C-terminus Gln299-Ser301.

Model6 from SARS 2a5i WH_human_protease_from2a5i_Minimised.pdb

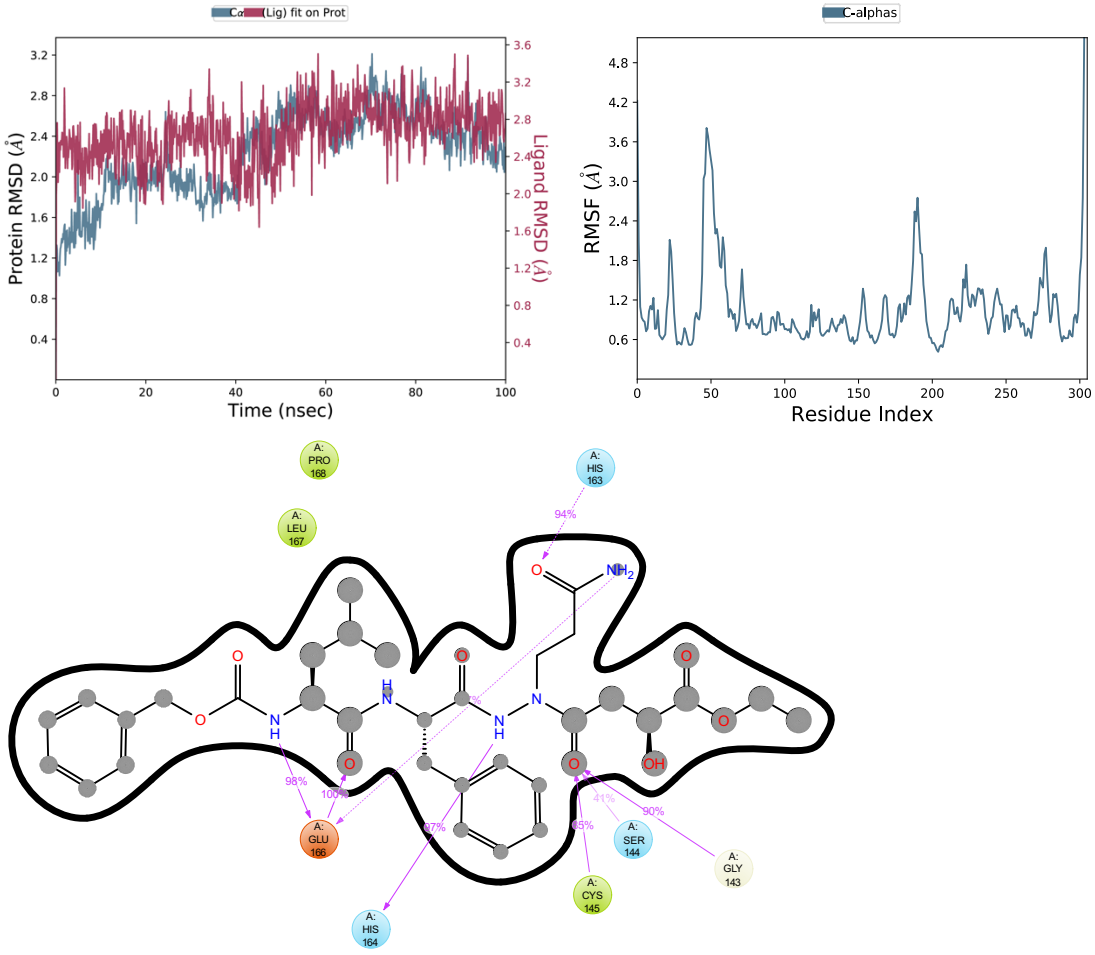


Figure S1-16. Protein C-alpha (blue line) and ligand (red line) RMSD over the 100 ns duration MD run. **Figure S1-17** Protein RMSF plot. Per residue variability in Cα position. **Figure S1-18**. Protein- ligand interaction summary diagram. Orange: negative charge interaction, blue: polar interactions, green: hydrophobic interactions, grey circles: solvent exposed atoms.

Over the first 10 ns of the simulation the mobile loop ^{44}CTSEDMLNPN^{53} defining the boundary of the S2 subsite moves away from the bound ligand, and remains distant during the remaining 80ns.

Flexible residues

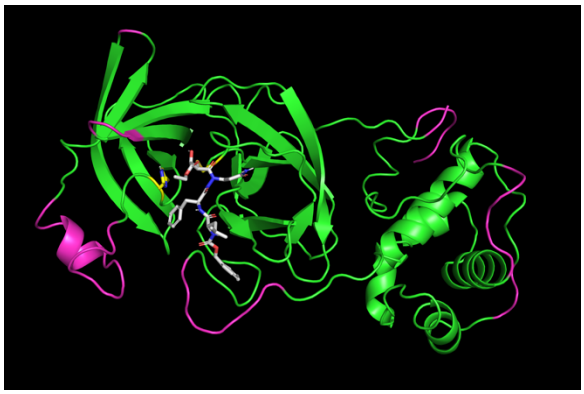
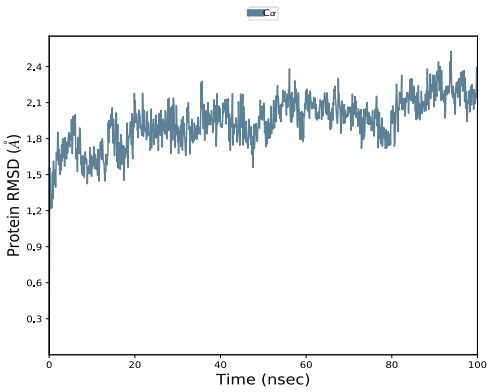


Figure S1-19. WH CoV 3CL^{pro} Model 6 shown as green ribbons, except catalytic dyad (His41, Cys145 – yellow sticks), and magenta ribbons for regions found to be mobile in the 100 ns molecular dynamics simulation. Covalently bound SARS inhibitor carried through in model from template PDB:2a5i shown as grey sticks.

Moderately mobile regions (>1.5 RMSD) N-terminal residues Ser1-Gly2, β-hairpin loop Gly23-Thr25, very mobile α-helix and loop Thr45-Arg60, loop tip Asn72, long strand Arg188-Ala194, strand Arg222-Thr224, strand Asn274-Arg279, C-terminus Ser301-Gln306.

Model12 from SARS 3vb3 WH_human_protease_from3vb3_dimerMinimised.pdb



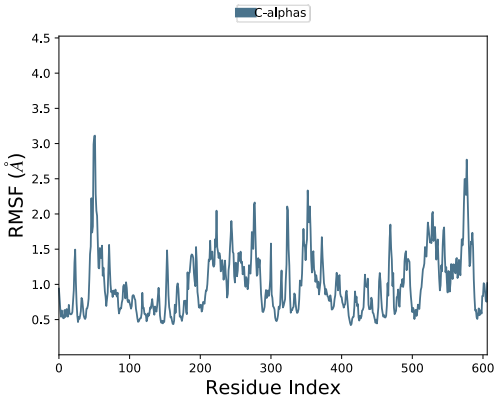


Figure S1-20. Protein C-alpha (blue line) RMSD over the 100 ns duration MD run. **Figure S1-21**. Protein RMSF plot. Per residue variability in Cα position.

Flexible residues

Figure S1-21. WH CoV 3CL^{pro} Model 12 Chain A shown as green ribbons, Chain B cyan ribbons, except catalytic dyads (His41, Cys145 – green and cyan sticks), and magenta ribbons for regions found to be mobile in 100 ns MD simulation.

In comparison to the MD simulation of the monomer (Model 2), the dimer shows stable N- and C-termini for both subunits, as these form part of the dimer interface which remains quite stable during the simulation. The very mobile helix and loop region show slight differences in mobility (Chain A Ser46-Ser62, Chain B Thr45-Asp56). In addition there are differences in the N-terminal domains, the Chain B helices Leu227-Met235 and Glu270-Met276 being slightly more mobile than those in Chain A.

Moderately mobile regions (>1.5 RMSD) Chain A: α-helix and loop Ser46-Ser62, loop tip Asn72, long strand 215-Thr225, helical turn Thr243-His246, large loop Asn274-Arg279, C-terminus Ser301. Chain B: β-hairpin loop Gly23-Thr25, α-helix and loop Thr45-Asp56, loop tip Asn72, loop Pro168-Gly170, long strand and helix Trp218-Met235, loop Thr243-Gln244, loop and helical turn Glu270-Thr280, loop Gly283-Leu286.

Model16 from SARS 2a5i WH_human_protease_from2a5i_dimerMinimised.pdb

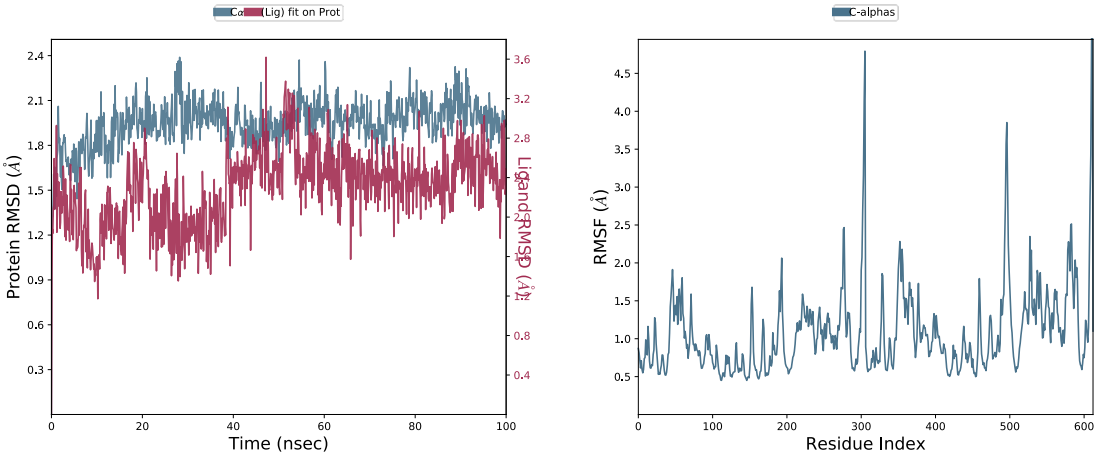


Figure S1-22. Protein C-alpha (blue line) and ligand (red line) RMSD over the 100 ns duration MD run. **Figure S1-23**. Protein RMSF plot. Per residue variability in Cα position.

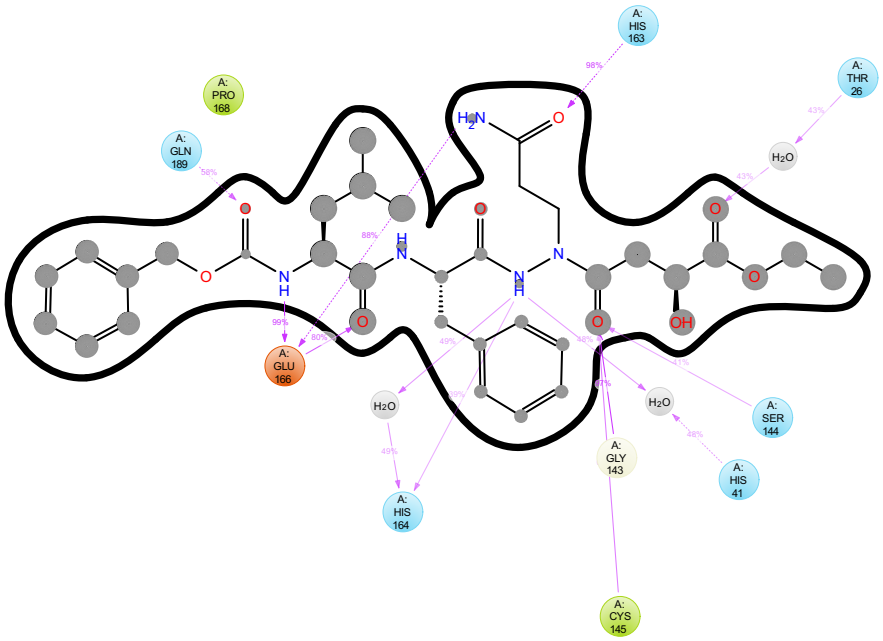


Figure S1-24. Protein- ligand interaction summary diagram. Orange: negative charge interaction, blue: polar interactions, green: hydrophobic interactions, grey circles: solvent exposed atoms.

Flexible residues

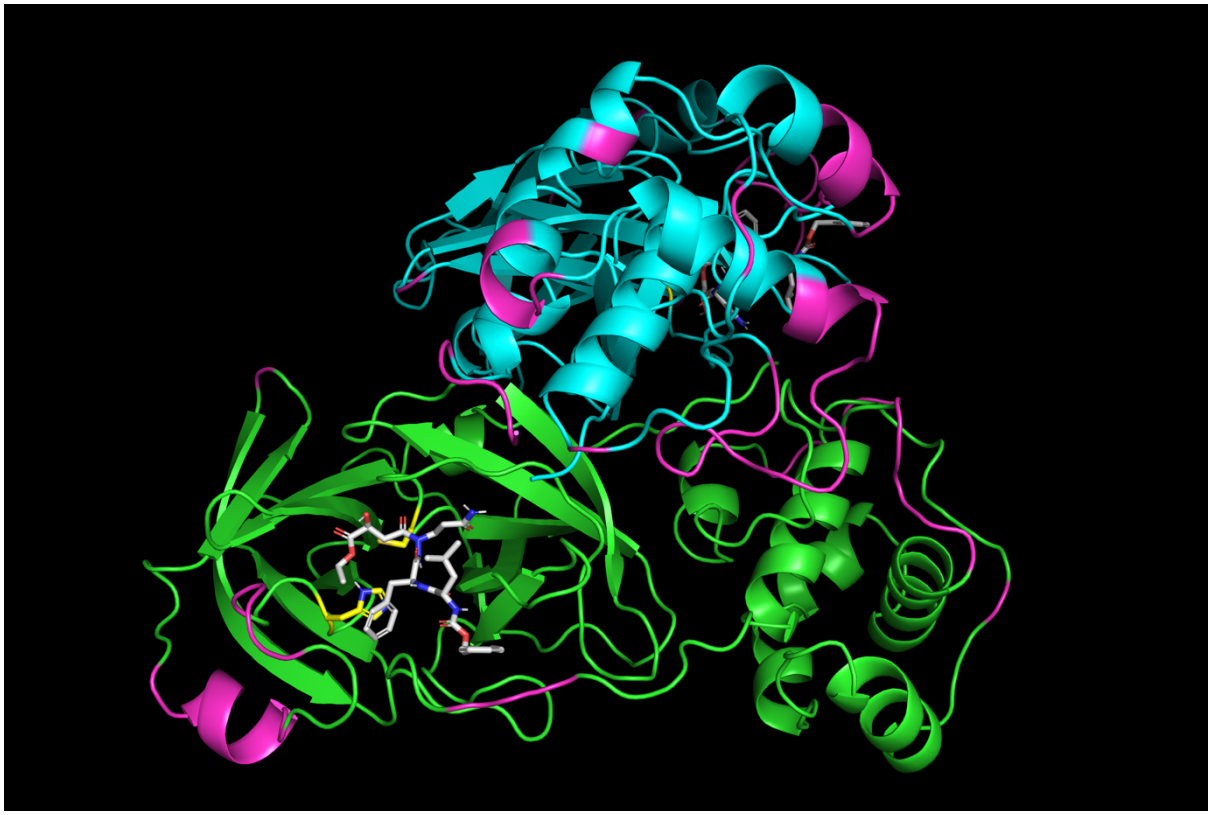


Figure S1-14. WH CoV 3CL^{pro} Model 16 Chain A shown as green ribbons, Chain B cyan ribbons, except catalytic dyads (His41, Cys145 – yellow sticks), and magenta ribbons for regions found to be mobile in 100 ns MD simulation. SARS 3CL^{pro} inhibitor shown as grey sticks.

Apart from the more stable N- and C-termini for both subunits which form part of the dimer interface the remainder of the dimer remains of similar overall flexibility. One major difference is that in the dimeric form the loop ^{44}CTSEDMLNPN^{53} which forms the boundary of the S2-subsite which binds bulky substrate residues Met/Phe/Leu is much less mobile than was seen in the 100ns simulation of the monomer.

Moderately mobile regions (>1.5 RMSD) Chain A: loop Thr45-Asp48, α-helix Asn53-Arg60, loop tip Asn72, loop tip Tyr154, strand Ala191-Ala193, strand residue Phe223, strand Asn274-Gly278, C-terminus Gly302-Gln306. Chain B: β-hairpin loop Gly23-Thr24, large loop Cys44-Asp53, strand Ile59-His64, loop tip Asn72, loop tip Tyr154, long strand and loop Val186-Gly195, strand residue Gly215, strand Phe223-Thr224, α-helical turn Leu232-Tyr237, helix residue Val247, helical turn Ala255-Gly258, very mobile loop and helical turn Glu270-Leu286, C-terminus Gly302-Gln306