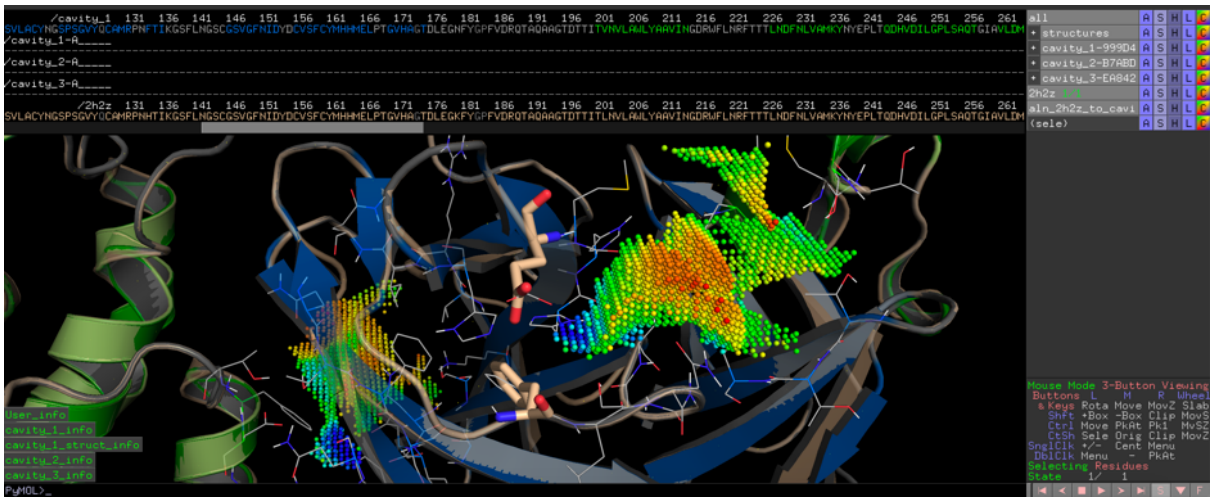
+43 316 269 205 office@innophore.com

Innophore
enzyme discovery

Wuhan coronavirus 2019-nCoV – what we can find out on a structural bioinformatics level



Christian Gruber, Georg Steinkellner – Innophore Enzyme Discovery 2020

2019-nCoV Background

As of January, 23rd 2020, the Wuhan coronavirus (WHO 2019-nCoV)[i], a positive-sense, single-stranded RNA coronavirus first reported in 2019 is spreading from Wuhan, China, the primary location outbreak.

The Chinese government placed the cities of Wuhan, Huanggang, and Ezhou with a combined population of approximately 15 million people, under lockdown in an attempt to contain the viral outbreak *e.g.* [ii],[iii]. The human-to-human transmission was confirmed in Guangdong, China, according to Zhong Nanshan, head of the health commission team investigating the outbreak.[iv] No specific treatment for the new virus is currently available, but existing anti-virals might be repurposed[v].

According to Wikipedia[vi], sequences of Wuhan betacoronavirus show similarities to beta coronaviruses found in bats; however, the virus is genetically distinct from other coronaviruses such as *Severe acute respiratory syndrome-related coronavirus* (SARS) and the *Middle East respiratory syndrome-related coronavirus* (MERS). Like SARS-CoV, it is a member of Beta-CoV lineage B (i. e. subgenus *Sarbecovirus*). Eighteen genomes of the novel coronavirus have been isolated and reported including BetaCoV/Wuhan/IVDC-HB-01/2019, BetaCoV/Wuhan/IVDC-HB-04/2020, BetaCoV/Wuhan/IVDC-HB-05/2019, BetaCoV/Wuhan/WIV04/2019, and BetaCoV/Wuhan/IPBCAMS-WH-01/2019 from the China CDC, Institute of Pathogen Biology, and Wuhan Jinyintan Hospital. Its RNA sequence is approximately 30 kb in length.

The new genome has led to several protein modeling experiments on the receptor-binding protein (RBD) of the nCoV spike (S) protein. Two Chinese groups, as of 23rd January 2020, believe that the S protein retains sufficient affinity to the SARS receptor (angiotensin-converting enzyme 2, ACE2) to use it as a mechanism of cell entry.

The RNA genome is replicated and a long polyprotein is formed, where all of the proteins are attached. Coronaviruses have a non-structural protein – a protease – which is able to separate the proteins in the chain. This is a form of genetic economy for the virus, allowing it to encode the greatest number of genes in a small number of nucleotides.[vii]

University of Hong Kong School of medicine, has previously said that SARS has been studied earlier and found that protease inhibitors and other drugs can effectively treat respiratory diseases such as SARS, middle respiratory syndrome and other coronaviruses.

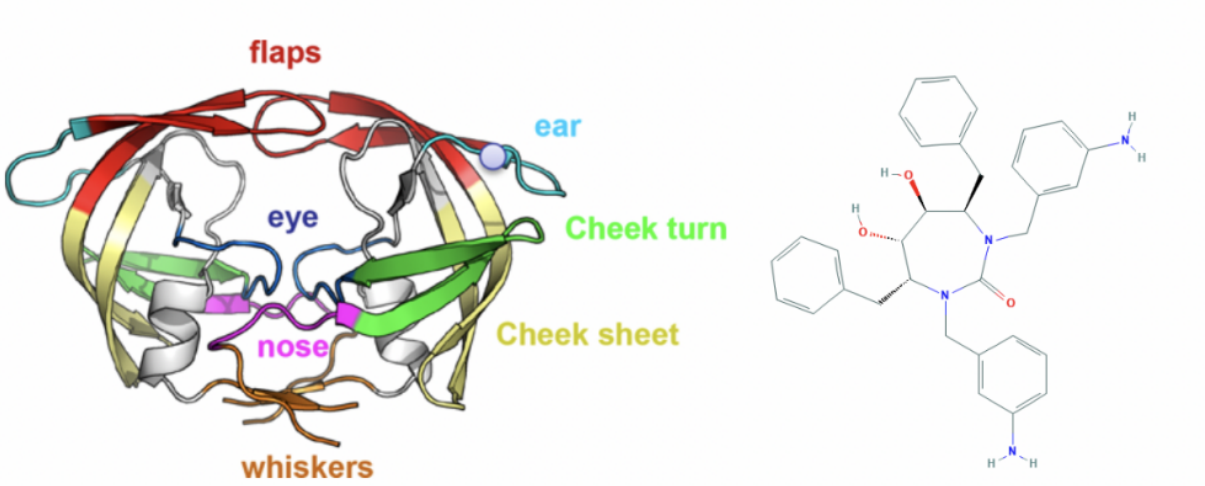
There were six kinds of coronaviruses that could infect humans, as well as 24 other kinds that could infect animals including bats, birds, rats, and cows. As most Wuhan patients had connections with the Huanan Seafood Market, there was a high chance the unknown coronavirus was transmitted to wild animals from bats and became mutated before it spread to humans, he said. Usually, a new disease would not be highly infectious between humans so only people who had very close contact with the patients could be infected, he said. If the Wuhan disease was similar to SARS, patients could be potentially cured by doses of ribavirin, protease inhibitor, and interferon.[viii]

Michael Mina, an epidemiologist at the Harvard School of Public Health yesterday said he has heard that some patients in China are being treated with protease inhibitors, antivirals that were developed to treat people with HIV and that were used “somewhat successfully” to treat SARS[ix].

Discovering the protease of coronavirus 2019-nCoV

Innophore decided to allocate significant human- and computational resources to support modelling efforts in this situation. Although Innophore is not active in the field of global epidemics, the fundamental principles of structural enzymology, our main expertise, are independent of the field of application. In the last years we had the chance to work with numerous academic- and corporate partners in the chemical, pharma, nutrition and agricultural industry on many different enzyme classes involving proteases. We worked on proteases used in consumer products, proteases for

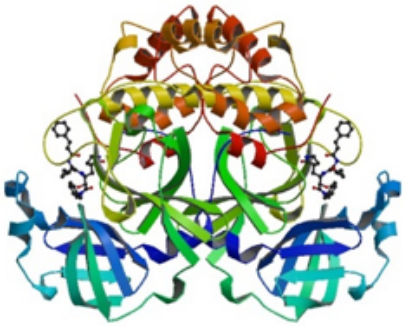
biocatalytic applications and human and simian IV proteases, studying the interaction dynamics, especially HIV1-protease[x] with various stereoisomers of classical inhibitors like Mozenavir.



Left: Topology of the HIV-1 protease[xi], Right: HIV-1 protease inhibitor DMP-450/Mozenavir[xii]

Validating the 2019-nCoV protease sequence

Although there are already modeling activities targeting this virus *e.g.* [xiii],[xiv], we decided to start from scratch to circumvent any potential biases and to focus on the protein class that our team is most familiar with: The viral protease of 2019-nCoV. Andrew Mesecar, Purdue's Walther Professor in Cancer Structural Biology and head of the Department of Biochemistry is also working on structure prediction of this target enzyme and the interaction with potential inhibitors. We are waiting for these structures to become publically available. In the meantime, analyzing the viral Wuhan seafood market pneumonia virus genome (NCBI genome ID MN908947[xv], GenBank: MN908947.3) published by Wu, F. Et al. today (LOCUS MN908947, 29903 bp, ss-RNA linear VRL 23-JAN-2020) we identified the potential protease sequence based on multiple sequences alignments with known SARS coronavirus proteases. The following figure shows the aligning sequence region of 2019-nCoV, "orf1ab polyprotein" with protein id QHD43415.1[xvi] with the sequence of PDB entry 5N5O[xvii], a structure of SARS coronavirus main protease deposited by Zhang, L., and Hilgenfeld, R. from the German Center for Infection Research in 2017 using Clustal O:



PDB entry 5N5O

```
5N5O:A|PDBID|CHAIN|SEQUENCE     ----------------------SGFRKMAFPSGKVEGCMVQVTCGTTTLNGLWLDDTVY   37
QHD43415.1                      FSNSGSDVLYQPPQTSITSAVLQSGFRKMAFPSGKVEGCMVQVTCGTTTLNGLWLDDVVY   3300
                                                      ********************************.**

5N5O:A|PDBID|CHAIN|SEQUENCE     CPRHVICTAEDMLNPNYEDLLIRKSNHSFLVQAGNVQLRVIGHSMQNCLLRLKVDTSNPK   97
QHD43415.1                      CPRHVICTSEDMLNPNYEDLLIRKSNHNFLVQAGNVQLRVIGHSMQNCVLKLKVDTANPK   3360
                                ********:******************.******************:*:*****:***

5N5O:A|PDBID|CHAIN|SEQUENCE     TPKYKFVRIQPGQTFSVLACYNGSPSGVYQCAMRPNHTIKGSFLNGSCGSVGFNIDYDCV   157
QHD43415.1                      TPKYKFVRIQPGQTFSVLACYNGSPSGVYQCAMRPNFTIKGSFLNGSCGSVGFNIDYDCV   3420
                                ************************************.**********************

5N5O:A|PDBID|CHAIN|SEQUENCE     SFCYMHHMELPTGVHAGTDLEGKFYGPFVDRQTAQAAGTDTTITLNVLAWLYAAVINGDR   217
QHD43415.1                      SFCYMHHMELPTGVHAGTDLEGNFYGPFVDRQTAQAAGTDTTITVNVLAWLYAAVINGDR   3480
                                **********************:**********************:***************

5N5O:A|PDBID|CHAIN|SEQUENCE     WFLNRFTTTLNDFNLVAMKYNYEPLTQDHVDILGPLSAQTGIAVLDMCAALKELLQNGMN   277
QHD43415.1                      WFLNRFTTTLNDFNLVAMKYNYEPLTQDHVDILGPLSAQTGIAVLDMCASLKELLQNGMN   3540
                                **************************************************:**********

5N5O:A|PDBID|CHAIN|SEQUENCE     GRTILGSTILEDEFTPFDVVRQCSGVTFQ------------------------------   306
QHD43415.1                      GRTILGSALLEDEFTPFDVVRQCSGVTFQSAVKRTIKGTHHWLLLTILTSLLVLVQSTQW   3600
                                *******:*:********************
```

Using EMBOSS Needle aligning the sequence the translated 2019-nCoV genome with another PDB entry 3TLO[xviii], a crystal structure of HCoV-NL63 3C-like protease, we get the same aligning region:

```
# Length:        7097
# Identity:      136/7097 ( 1.9%)
# Similarity:    192/7097 ( 2.7%)
# Gaps:         6795/7097 (95.7%)
# Score:        651.5
```

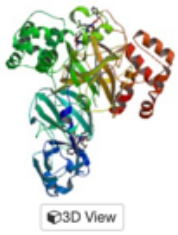
```
SEQUENCE          1 -------------SGLKKMAQPSGCVERCVVRVCYGSTVLNGVWLGDTVT     37
                                 ||.:|||.|||.||.|:|:|..|:|.|||:||.|.|.
QHD43415.1     3251 QPPQTSITSAVLQSGFRKMAFPSGKVEGCMVQVTCGTTTLNGLWLDDVVY   3300

SEQUENCE         38 CPRHVIAPSTTVL-IDYDHAYSTMRLHNFSVSHNGVFLGVVGVTMHGSVL     86
                    ||||||..|..:| .:|:........|||.|....|.|.|:|.:|...||
QHD43415.1     3301 CPRHVICTSEDMLNPNYEDLLIRKSNHNFLVQAGNVQLRVIGHSMQNCVL   3350

SEQUENCE         87 RIKVSQSNVHTPKHVFKTLKPGDSFNILACYEGIASGVFGVNLRTNFTIK    136
                    ::||..:|..|||:.|..::||.:|::|||||.|..|||:...:|.|||||
QHD43415.1     3351 KLKVDTANPKTPKYKFVRIQPGQTFSVLACYNGSPSGVYQCAMRPNFTIK   3400

SEQUENCE        137 GSFINGACGSPGYNVRNDGTVEFCYLHQIELGSGAHVGSDFTGSVYGNFD    186
                    |||:||:|||.|:|:..| .|.|||:|.:||.:|.|.|:|..|:.||.|.
QHD43415.1     3401 GSFLNGSCGSVGFNIDYD-CVSFCYMHHMELPTGVHAGTDLEGNFYGPFV   3449

SEQUENCE        187 DQPSLQVESANLMLSDNVVAFLYAALLNGCRWWLCSTRVNVDGFNEWAMA    236
                    |:.:.|....:..::.||:|:||||::||.||:|......::.||..||.
QHD43415.1     3450 DRQTAQAAGTDTTITVNVLAWLYAAVINGDRWFLNRFTTTLNDFNLVAMK   3499

SEQUENCE        237 NGYTSVSS--VECYSILAAKTGVSVEQLLASIQH-LHEGFGGKNILGYSS    283
                    ..|..::.  |:....|:|:||::|..:.||::. |..|..|:.|||.:.
QHD43415.1     3500 YNYEPLTQDHVDILGPLSAQTGIAVLDMCASLKELLQNGMNGRTILGSAL   3549

SEQUENCE        284 LCDEFTLAEVVKQMYGVNLQ----------------------------    303
                    |.||||..:||:|..||..|
QHD43415.1     3550 LEDEFTPFDVVRQCSGVTFQSAVKRTIKGTHHWLLLTILTSLLVLVQSTQ   3599
```

Extracting the putative protease sequence from position X to Y yields a putative protease sequence of 306 amino acids with a calculated protein weight of 33.8 kilodaltons, which is at the upper range of typical proteases.

```
>QHD43415.1 putative protease by 5N5O:A sequence alignment
SGFRKMAFPSGKVEGCMVQVTCGTTTLNGLWLDDVVYCPRHVICTSEDMLNPNYEDLLIRKSNHNFLVQ
AGNVQLRVIGHSMQNCVLKLKVDTANPKTPKYKFVRIQPGQTFSVLACYNGSPSGVYQCAMRPNFTIKG
SFLNGSCGSVGFNIDYDCVSFCYMHHMELPTGVHAGTDLEGNFYGPFVDRQTAQAAGTDTTITVNVLAW
LYAAVINGDRWFLNRFTTTLNDFNLVAMKYNYEPLTQDHVDILGPLSAQTGIAVLDMCASLKELLQNGM
NGRTILGSALLEDEFTPFDVVRQCSGVTFQ
```

Blasting this sequence again against the PDB revealed proteins with very high sequence similarity and sufficient resolution for subsequent homology modeling, e.g. PDB entry 2A5K[xix].

```
PDB Sequence Search:
SGFRKMAFPS GKVEGCMVQV TCGTTTLNGL WLDDVVYCPR HVICTSEDML NPNYEDLLIR KSNHNFLVQA GN
Expectation Value = 10.0, Sequence Identity = 0%, Search Tool = blast, Mask Low
```



2A5K: Entity 1 containing Chain A, B Download File View File

Crystal structures of SARS coronavirus main peptidase inhibited by an aza-peptide epoxide in space group P212121

Lee, T.W., Cherney, M.M., Huitema, C., Liu, J., James, K.E., Powers, J.C., Eltis, L.D., James, M.N.

(2005) J Mol Biol **353** 1137-1151

Released: 10/25/2005
Method: X-ray Diffraction
Resolution: 2.3 Å
Residue Count: 614

Macromolecule:
3C-like peptidase (protein)
Unique Ligands: AZP

Length: 306 **E-value:** 6.88846E-156 **Score:** 548.169bits (1412) **Identities:** 294/306 (96%) **Positives:** 302/306 (99%) **Gaps:** 0/306 (0%)

```
Query SGFRKMAFPSGKVEGCMVQVTCGTTTLNGLWLDDVVYCPRHVICTSEDMLNPNYEDLLIRKSNHNFLVQAGNVQLRVIGHSMQNCVLKLKVDTANPKTPKY
      SGFRKMAFPSGKVEGCMVQVTCGTTTLNGLWLDD VYCPRHVICT+EDMLNPNYEDLLIRKSNH+FLVQAGNVQLRVIGHSMQNC+L+LKVDT+NPKTPKY
Sbjct SGFRKMAFPSGKVEGCMVQVTCGTTTLNGLWLDDTVYCPRHVICTAEDMLNPNYEDLLIRKSNHSFLVQAGNVQLRVIGHSMQNCLLRLKVDTSNPKTPKY
```

In total, 138 structures were found meeting a generous expectation value of 10:

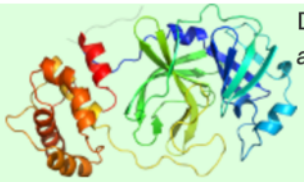
```
1LVO,1P9S,1P9U,1Q2W,1UJ1,1UK2,1UK3,1UK4,1WOF,1Z1I,1Z1J,2A5A,2A5I,2A5K,2ALV,2AMI
2BX3,2BX4,2C3S,2D2D,2DUC,2GT7,2GT8,2GTB,2GX4,2GZ7,2GZ8,2GZ9,2H2Z,2HOB,2K7X,2LIZ
2Q6D,2Q6F,2Q6G,2QC2,2QCY,2QIQ,2V6N,2VJ1,2YNA,2YNB,2Z3C,2Z3D,2Z3E,2Z94,2Z9G,2Z9J
2ZU2,2ZU4,2ZU5,3ATW,3AVZ,3AW0,3AW1,3D23,3D62,3E91,3EA7,3EA8,3EA9,3EAJ,3EBN,3F9E
3F9H,3FZD,3IWM,3J1Z,3M3S,3M3T,3M3V,3MOG,3SN8,3SNA,3SNB,3SNC,3SND,3SNE,3SZN,3TIT
3TNS,3TNT,3V3M,3VB3,3VB4,3VB5,3VB6,3VB7,4F49,4HI3,4MDS,4RSP,4TWW,4TWY,4WMD,4WME
4XFQ,4YLU,4YO9,4YOG,4YOI,4YOJ,4ZRO,4ZUH,5B6O,5C3N,5C5N,5C5O,5EU8,5GWY,5GWZ,5HYC
5NH0,5NH0,5NH0,5VRF,5WKJ,5WKK,5WKL,5WKM,5ZQG,6FV1,6FV2,6JIJ
```

For modeling the 3D structure of 2019-nCoV protease, we used our Catalophore^{TM} platform as well as the public Phyre2[xx] server to generate homology models. Both approaches yielded satisfying results

as expected given by the very high sequence similarity.

Phyre2 top model d2duca1 is based on the fold library id d2duca1[xxi], a trypsin-like serine protease oft the viral cysteine protease of trypsin fold from the SARS coronavirus main proteinase[xxii].

Lineage for d2duca1 (2duc A:2-301)

1. Root: SCOP 1.75
2. Class b: All beta proteins [48724] (174 folds)
3. Fold b.47: Trypsin-like serine proteases [50493] (1 superfamily)
   barrel, closed; n=6, S=8; greek-key
   duplication: consists of two domains of the same fold
4. Superfamily b.47.1: Trypsin-like serine proteases [50494] (4 families) $\mathcal{S}$
5. Family b.47.1.4: Viral cysteine protease of trypsin fold [50603] (3 proteins)
6. Protein Coronavirus main proteinase (3Cl-pro, putative coronavirus nsp2) [74979] (3 species)
   contains an extra alpha-helical domain
7. Species SARS coronavirus [TaxId:227859] [89349] (20 PDB entries)
8. 
   Domain d2duca1: 2duc A:2-301 [146586]
   *automatically matched to d1q2wb_*

Phyre2 confidence in the model is 100.0%, although we wouldn't go that far.

Top model

Image coloured by rainbow N → C terminus

Model dimensions (Å): **X**:37.952 **Y**:75.434 **Z**:46.236

Model (left) based on template d2duca1

Top template information

Fold: Trypsin-like serine proteases
Superfamily: Trypsin-like serine proteases
Family: Viral cysteine protease of trypsin fold

Confidence and coverage

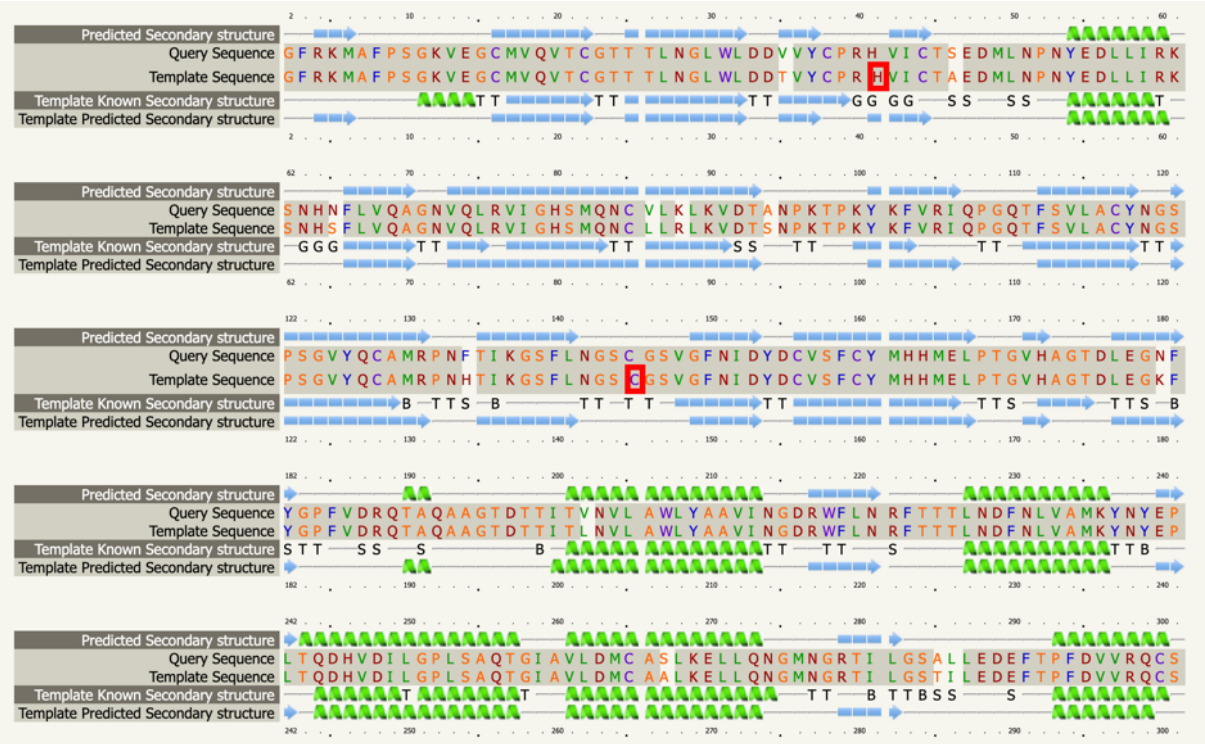
Confidence: **100.0%** Coverage: **98%**

300 residues (98% of your sequence) have been modelled with 100.0% confidence by the single highest scoring template.

3D viewing

Interactive 3D view in JSmol

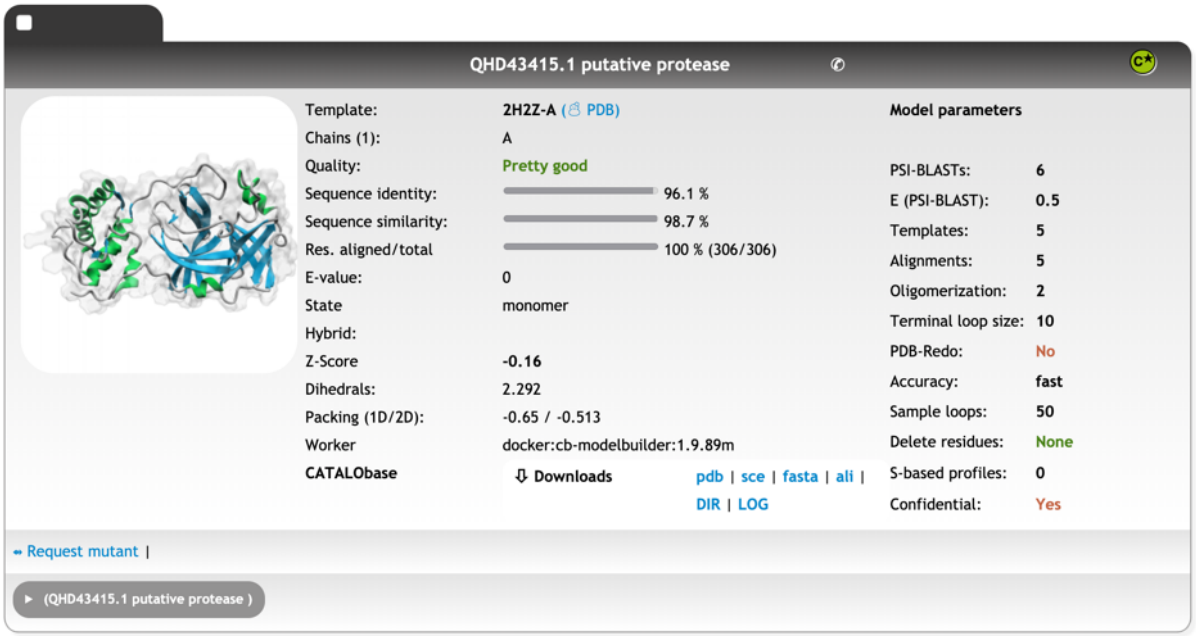
For other options to view your downloaded structure offline see the FAQ



You can download the complete Phyre2 run here: d06ff0dcb8400814.tar

In subsequent steps, we will identify cavities in the homology models, annotate them to generate point clouds.

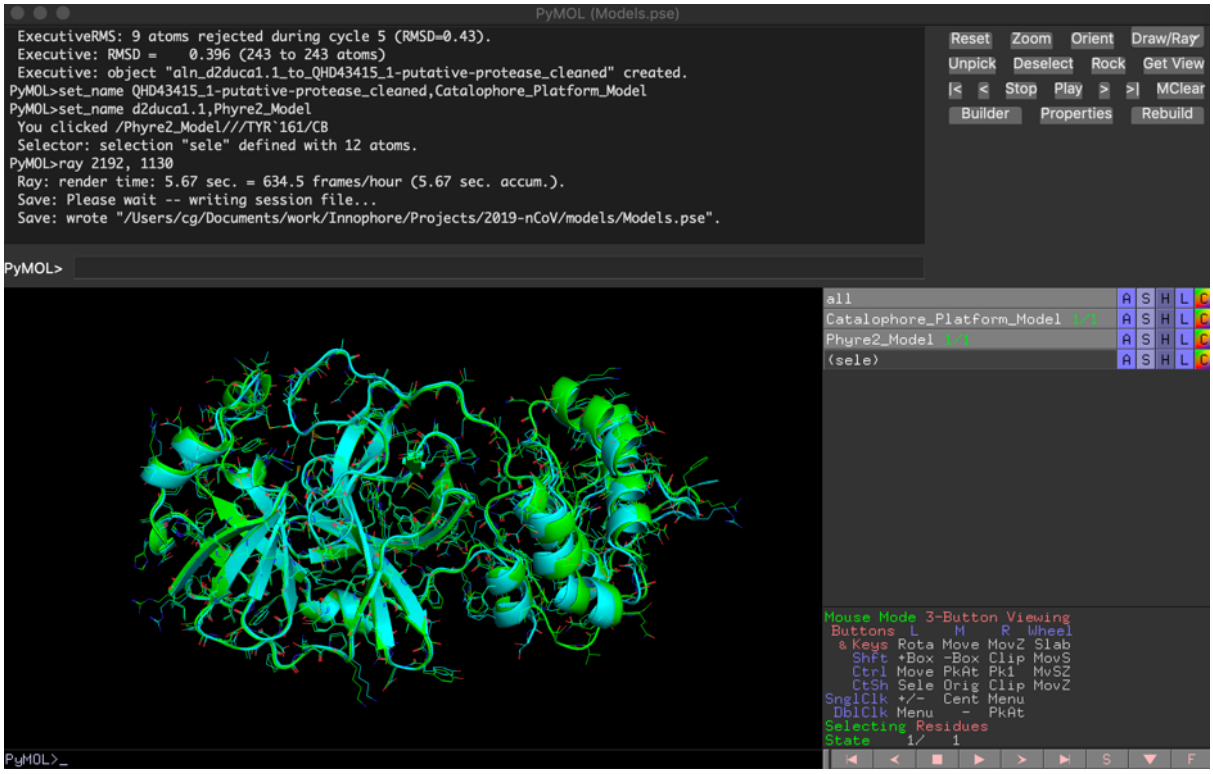
Innophore's CatalophoreTM platform predicated a homology model bases on the structural template 2H2Z[xxiii], chain A, the crystal structure of SARS-CoV main protease with authentic N and C-termini, with an overall quality "Pretty good". We expect the protein to be a monomer. You can download the model as compressed PDB file here: QHD43415_1-putative-protease_cleaned.pdb



> **Sequence identity: 96.1 %**
> **Sequence similarity: 98.7 %**
> **Res. aligned/total: 306/306**

The comparison of the two independent models showed high similarity, with a final RMSD of 0.396 Å:

```
Match: read scoring matrix.
Match: assigning 300 x 306 pairwise scores.
MatchAlign: aligning residues (300 vs 306)...
MatchAlign: score 1616.000
ExecutiveAlign: 300 atoms aligned.
ExecutiveRMS: 7 atoms rejected during cycle 1 (RMSD=0.89).
ExecutiveRMS: 14 atoms rejected during cycle 2 (RMSD=0.62).
ExecutiveRMS: 12 atoms rejected during cycle 3 (RMSD=0.54).
ExecutiveRMS: 15 atoms rejected during cycle 4 (RMSD=0.49).
ExecutiveRMS: 9 atoms rejected during cycle 5 (RMSD=0.43).
Executive: RMSD = 0.396 (243 to 243 atoms)
```

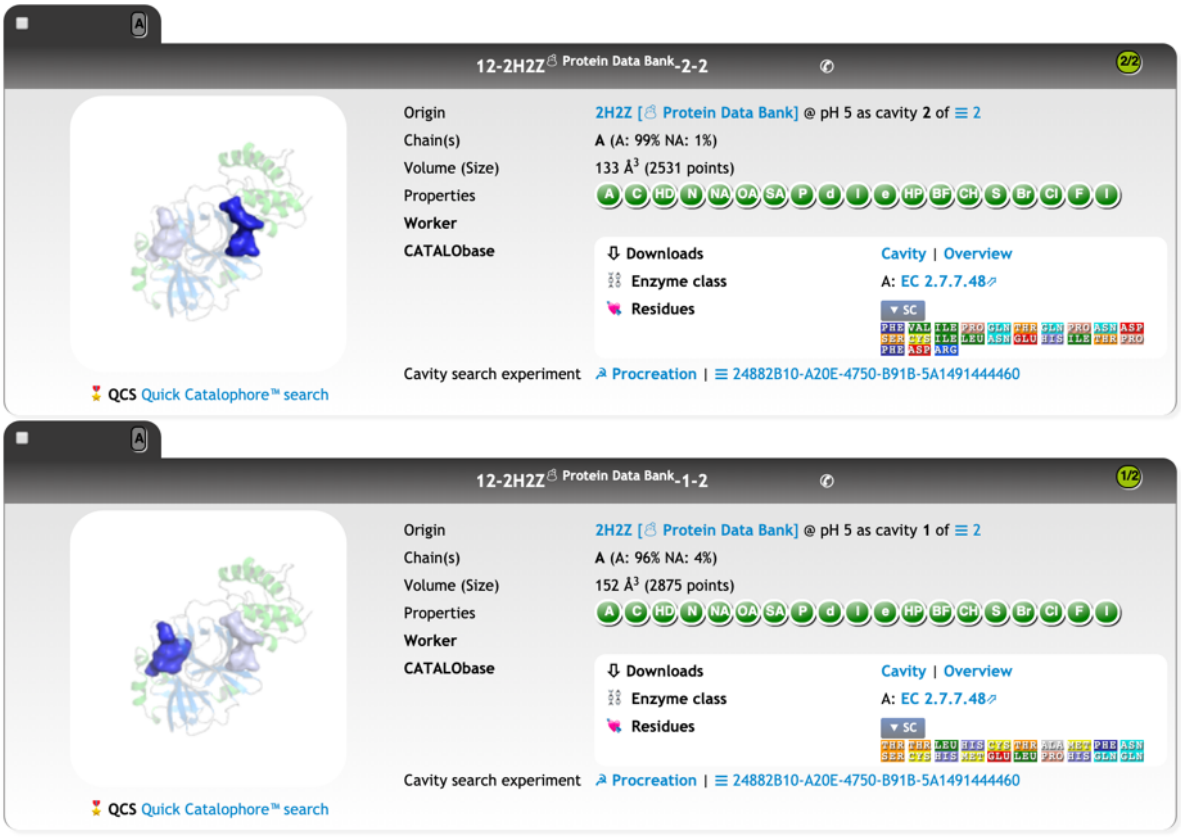


The PyMol session file containing both models can be downloaded here: **2019-nCoV_putative_protease-Models.pse**

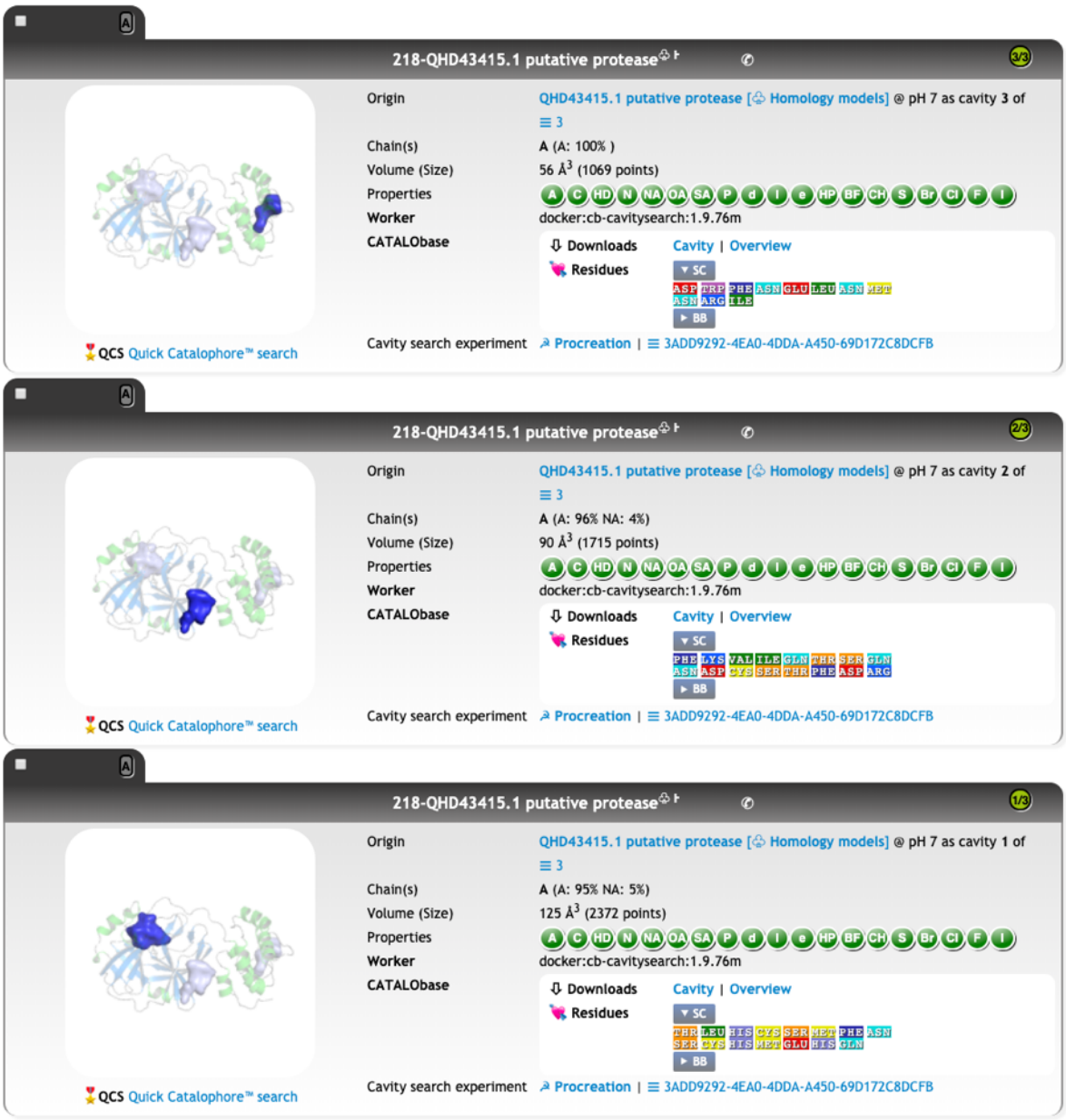
2019-nCoV active sites considerations

For the template structure 2H2Z, our Catalophore^{TM} database has point-cloud cavities of 6 cavity breeds available, calculated under different environments (e.g. pH):

At protonation state pH 7, we have two cavities for the template, both annotated as EC 2.7.48:



We calculated the active site CatalophoreTM point-cloud for the putative 2019-nCoV protease. Using standard setting, we obtained three cavities in the 2019-nCoV protease model:



We are currently fingerprinting the most likely candidate for the active site and re-checking our cavity procreation parameters. We will come back shortly with a downloadable version including the physicochemical parameter point-clouds and analysis of the differences to the proteases from other coronaviruses.

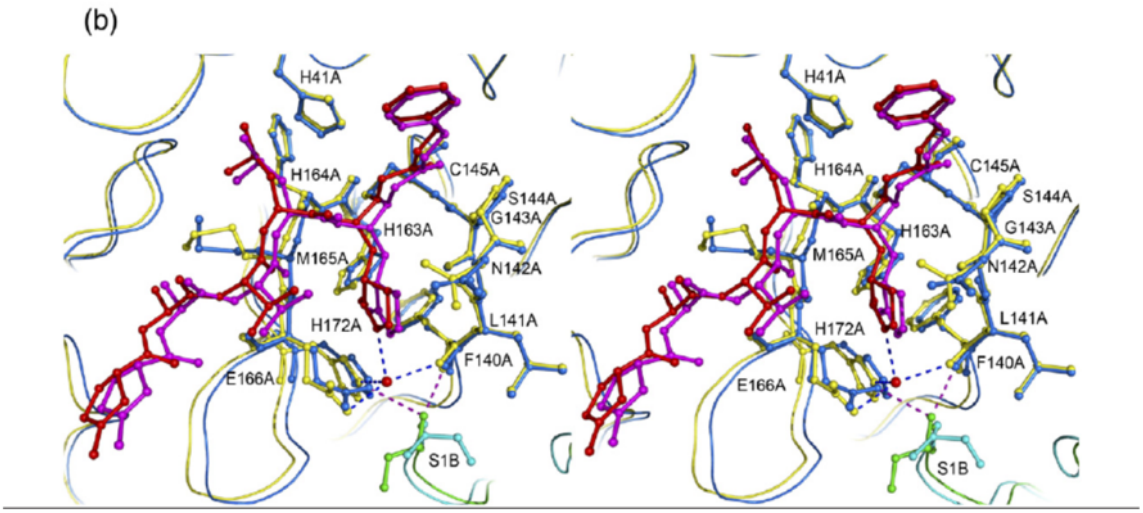
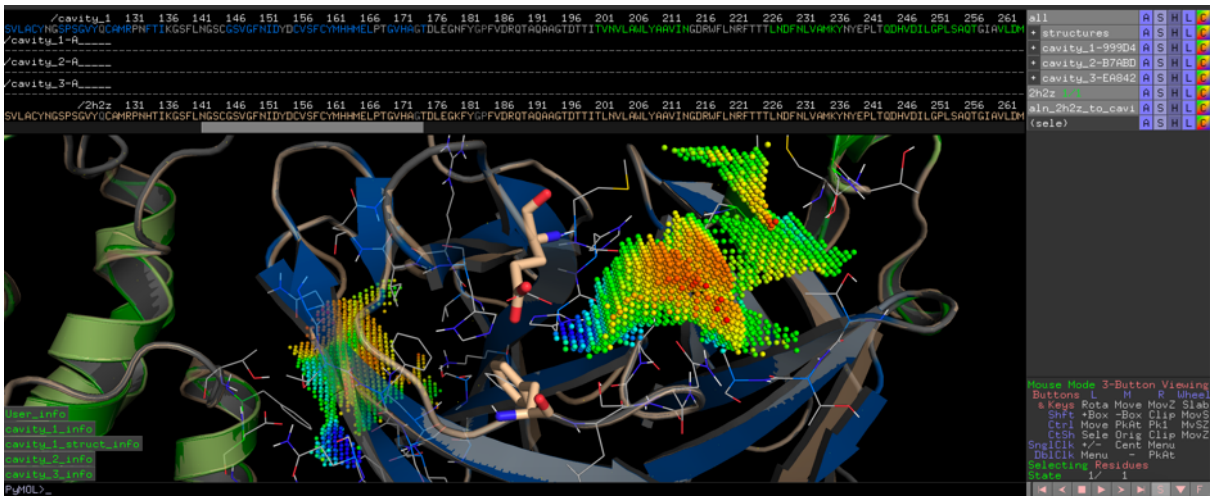
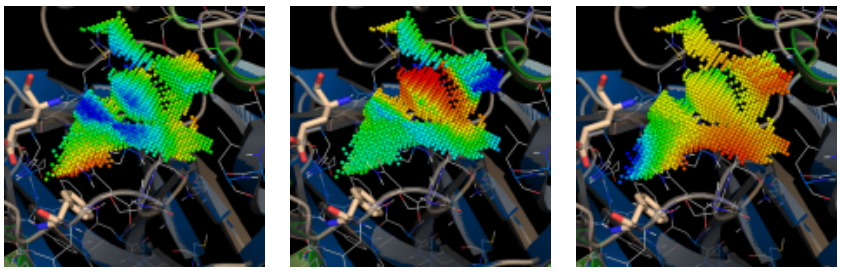


Figure 4. Differences between the complex structures of WT and GPLGS-WT. (a) Inhibitor N3. (b) Superposition of the substrate-binding pockets in protomer A of GPLGS-WT and that in protomer A* of WT. In the WT-N3 complex structure, the NH_2 group of Ser1 in protomer B* was still hydrogen-bonded to the carboxylate group of Glu166 and the carbonyl group of Phe140 in protomer A*, stabilizing the S1 pocket. In the GPLGS-WT-N3 complex structure, however, the two hydrogen bonds described above were not found. Instead, an ordered water molecule was observed in the S1 pocket. Protomer A* of WT is in blue; protomer A of GPLGS-WT is in yellow; inhibitor N3 (complexed with WT) is in magenta; inhibitor N3 (complexed with GPLGS-WT) is in red; protomer B* of WT is in green; protomer B of GPLGS-WT is in cyan.

By aligning PDB entry 2H2Z from Yang, H. Et. al 2006[xxiv] with our model and mapping the residues Glu166 and Phe140 (figure above) of the inhibition site to our point-cloud CatalophoresTM sites, we could identify cavity "1" to be a potential target site for inhibition.

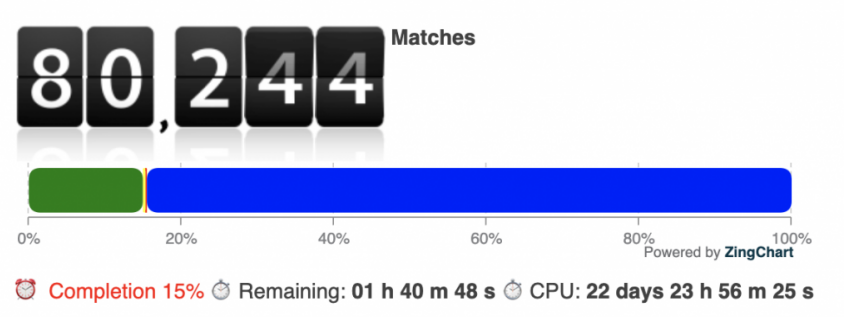


Suqsequentially we will search our CatalophoreTM databases for similar point clouds in hope of identifying proteins with similar distribution patterns in the physicochemical property space with known inhibitors to potentially find inhibitors that bind to the protease of the Wuhan virus as well.



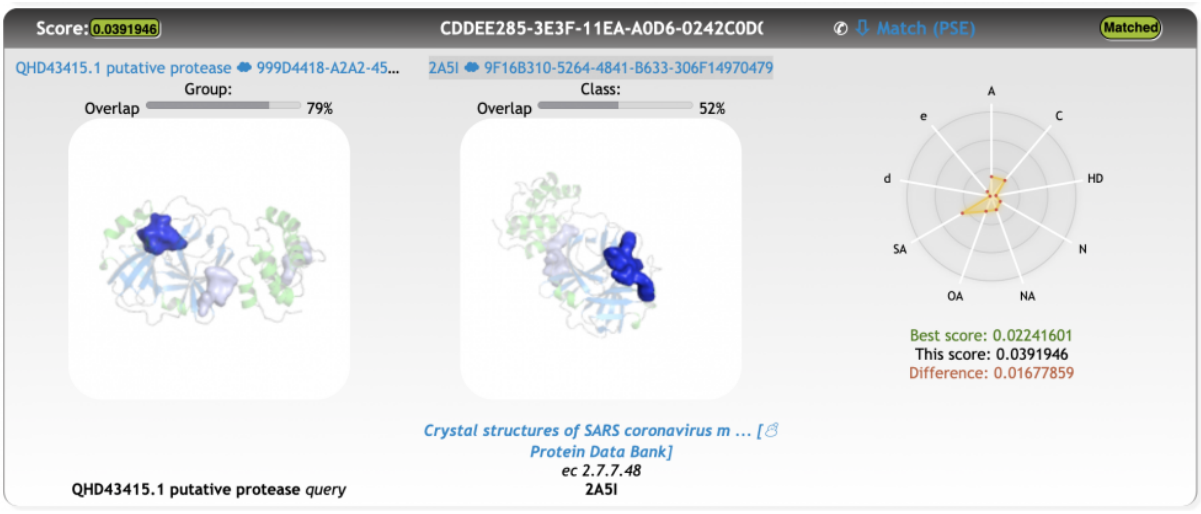
Update January, 24rd 2020 1:24 UTC: Catalophore search started

We reallocated 3/4 of our computational resources – several thousand cores – to screen 535.879 cavities derived from the PDB overnight. The estimated total CPU time for this screening is approximately 23 days.



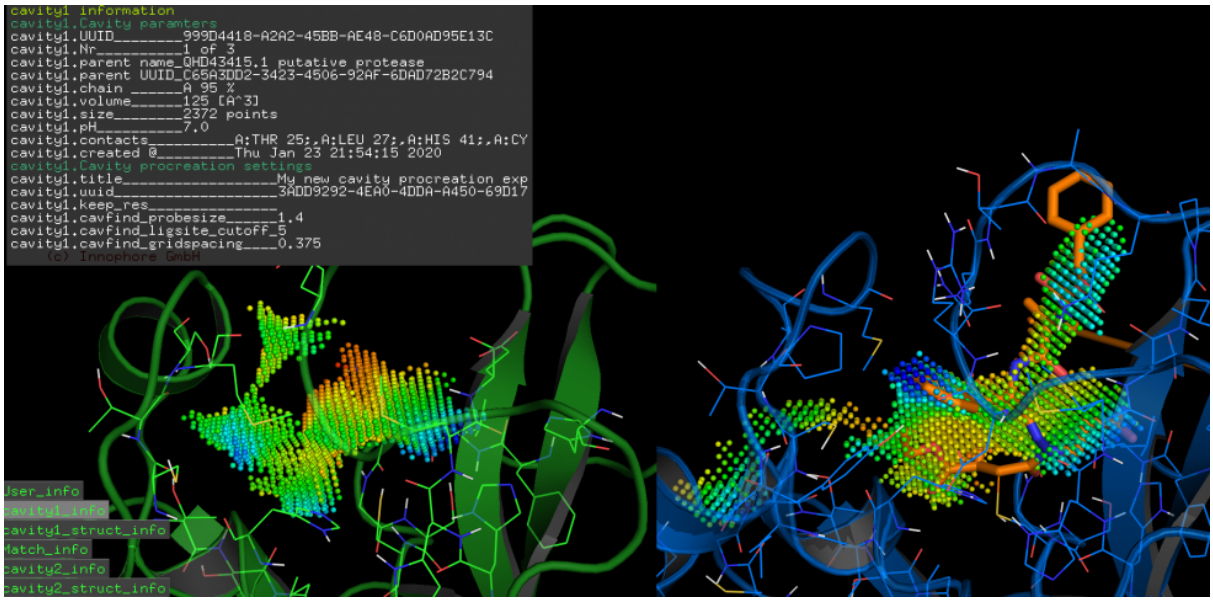
One of the first meaningful matches out of the 15% that were screened until now is a cavity match with a point-cloud from PDB entry 2A5I (https://www.rcsb.org/structure/2A5I, Crystal structures of SARS coronavirus main peptidase inhibited by an aza-peptide epoxide in the space group C2):

> **match-306-CDDEE285-3E3F-11EA-A0D6-0242C0D00007.zip**
> **Matching results**
>
Match Total score	0.03919460
Match Distance score	0.00555660
Match Overlap 1	79 %
Match Overlap 2	52 %

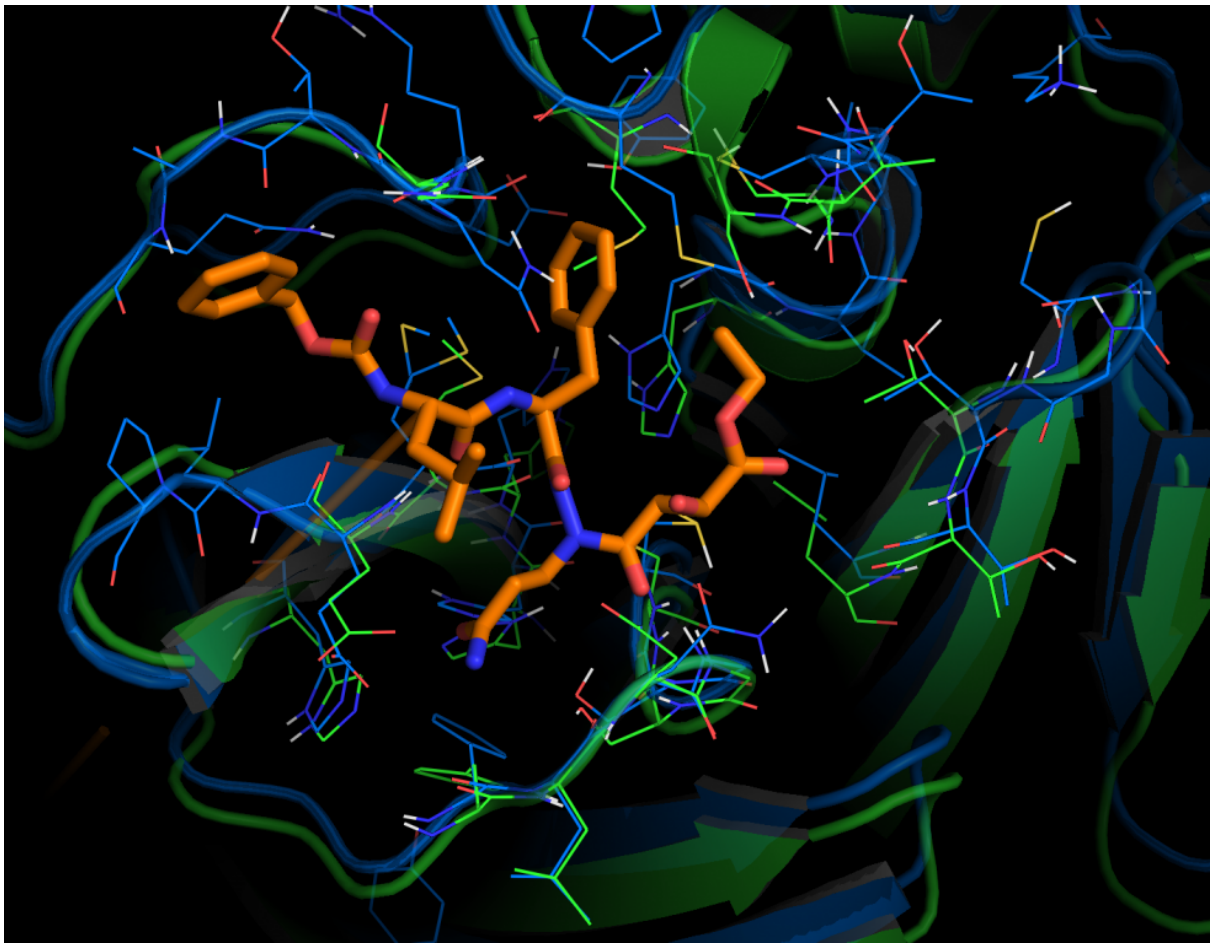


Although the cavity overlap is not perfect, the alignment of the protein structures solely based on the cavity rotation-translation matrix is satisfying.

Left: active site of the putative 2019-nCoV protease, right: database entry FDBB501F-1EF4-4AC9-8AA3-0CE76, the point-cloud of the SARS coronavirus peptidase inhibited by an aza-peptide epoxide:



Overlay of the compound AZP (https://www.rcsb.org/ligand/AZP) based on the cavity matching alignment. This is not a docking result – the coordinates of the ligand were transformed based on the cavity match and transferred onto the 2019-nCoV protease.



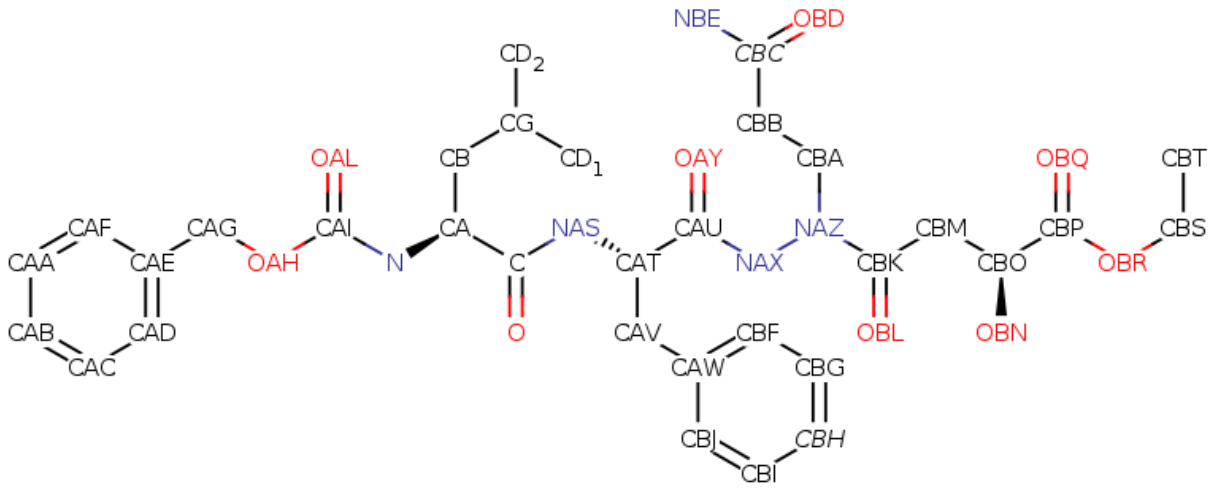
Ethyl (5S,8S,14S)-11-(3-amino-3-oxopropyl)-8-benzyl-14-hydroxy-5-isobutyl-3,6,9,12-tetraoxo-1-phenyl-2-oxa-4,7,10,11-tetraazapentadecan-15-oate

Molecular Formula: $C_{32}H_{43}N_5O_9$

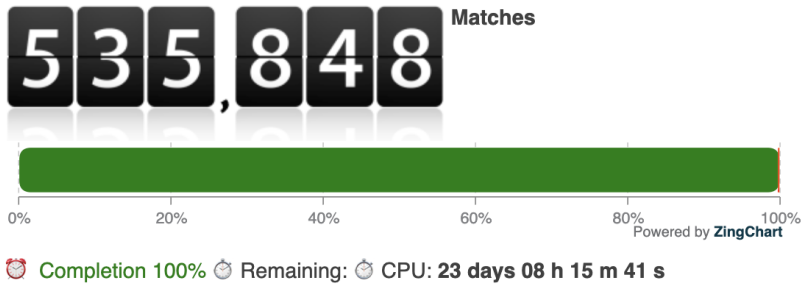
Average mass: 641.712 Da

Monoisotopic mass: 641.306091 Da

ChemSpider: ID4450034



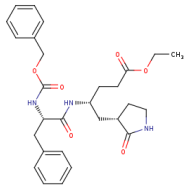
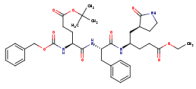
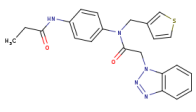
Update January, 24rd 2020 2:16 UTC: Catalophore search finished

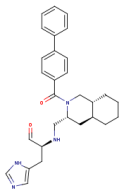


Our in-silico Catalophore screening completed after 01 h 38 m 21 s real-time, roughly the expected 23 CPU days. We now filtered the results for Catalophore point-clouds in the database with an overlap of more than 70% with the 2019-nCoV virus protease point-cloud and favored cavities were crystallographic ligands were bound before calculating the cavities. We limited the results to cavities larger than 150A^3 covering the complex ligand to enrich medium- to large-sized organic compounds in the ranked list.

Disclaimer: This following list does not take into account any pharmacological-, toxic- or side effect nor does it represent compounds directly suggested as potential drugs against 2019-nCoV. The list currently contains 148 organic compounds in total, that bind to protein cavities that share high physicochemical similarity to the 2019-nCoV protease active site cavity based on our multi-dimensional point-cloud matching.

The **preliminary top 5 hits** potentially binding to the putative 2019-nCoV protease are listed in the following table – after further inspection, the remaining candidates will be available too:

Catalophore Score	Compound	Formula	InChIKey
0.023	G75		PIZHLOUXQJUQHF-VXNXHJTFSA-N
0.026	G82		MMMLJIROCXIHMV-XJYHXZFBSA-N
0.027	DTZ		KEQKAMYELZXRRN-UHFFFAOYSA-L
0.030	R30		TWIVXCFEBRGEKY-UHFFFAOYSA-N

0.031	3X5		VZCULZJNALRGNB-DNZWLJDLSA-N

Update January 24th 2020, 16:00UTC

Since we were mentioned in the Wikipedia article today about the novel coronavirus 2019-nCoV having published comparative models and preliminary inhibitors of the #2019-nCoV protease we are in contact with several official bodies to further contribute to the field.

Update January 24th 2020, 22:00UTC

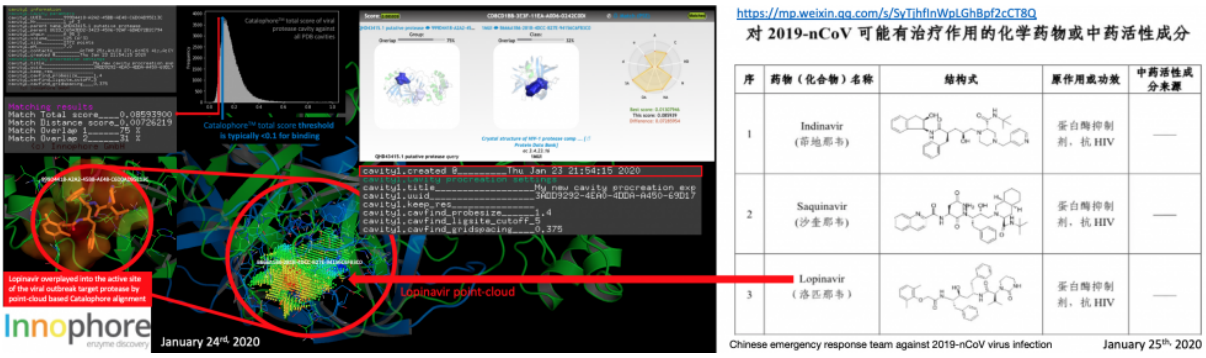
Thanks to our colleagues from the GISAID initiative, tonight we have gained access to 17 additional nCoV genome data sets recently derived from patients.

Update January 25th 2020, 00:30UTC

We are so proud to be working with a group of bioinformaticians from a major pharmaceutical company in Beijing and the Chinese CDC since 1 a.m. to search our Catalophore databases for potential experimental or approved drug targets that could bind to the #2019-nCoV protease and to review our data with our colleagues in China.

Update January 25th 2020 09:30UTC

More than 20 research groups incl. the Shanghai Institute of Materia Medica, Chinese Academy of Sciences and Shanghai University of Science and Technology's Institute of Immunochemistry joined the emergency response team against #2019-nCoV virus infection, using the accumulated anti-SARS drug research experience to conduct anti-2019-nCoV drug research. A list of potential compounds was published just now. The HIV-1 protease inhibitor DMP450 we mentioned yesterday night in our post is not in the list, however many of the entries are HIV inhibitors. We were informed by CDC, that Innophore's top 2 ranked candidate molecules from yesterday, G75 and G82 are missing drug status. So the search will be focused on approved drugs only now. Still, these molecules are supposed to be the potentially the best binders, derived from the analysis of the previously published crystal complexes e.g. by Lee Et. al (https://www.rcsb.org/structure/2A5I). Molecular dynamics analysis will be available in 4-5 hours. Some of the compounds listed by the emergency response team are found in crystal structures of complexes and are highly ranked in cavities of our yesterday Catalophore search, meaning having a total score under 0.1, e.g. **Lopinavir** (is the Top3 candidate of the Chinese emergency response team) and scored in our search with a Catalophore total score of 0.085939 by matching our cavity of 2019-nCoV protease cavity with the cavity of **PDB 1MUI (https://www.rcsb.org/structure/1MUI, published in 2002)**. So Lopinavir is potentially one of the better binders to the Wuhan coronavirus 2019-nCoV protease cavity – and a previously approved "old" drug (https://de.wikipedia.org/wiki/Lopinavir).

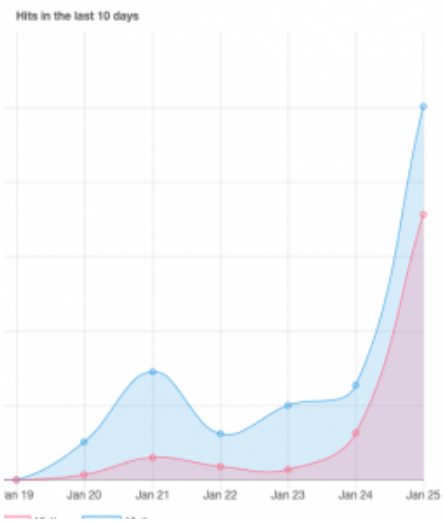


Update January 25th 2020 10:03UTC

As mentioned, we are working with a group of bioinformaticians from a major pharmaceutical company in Beijing and the Chinese CDC since 1 a.m. to search our Catalophore databases for potential experimental or approved drug targets that could bind to the #2019-nCoV protease and to review our data with our colleagues in China. We have to coordinate the file transfers and communication with our partners in China, therefore we stop to publish now to get the work done. Cross your fingers and if you have any suggestions, contact us anytime.

Update January 25th 2020 10:25UTC

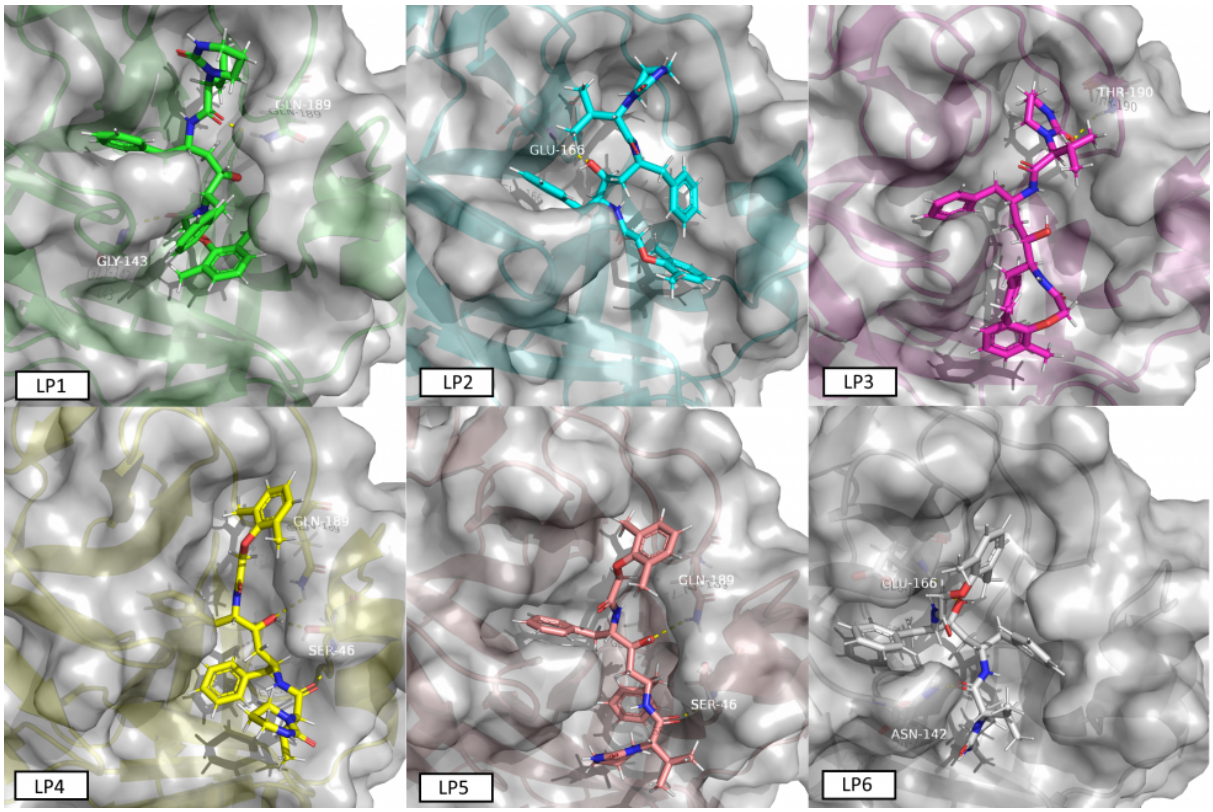
Since the last days, the requests for our webserver are continuously increasing. Since this night, it's even growing much faster. Our website is hosted at an external provider – if the server breaks under the load we will migrate to AWS or Google with the following updates – if any. We would post the links on LinkedIn.



Update January 29th 2020 16:10UTC

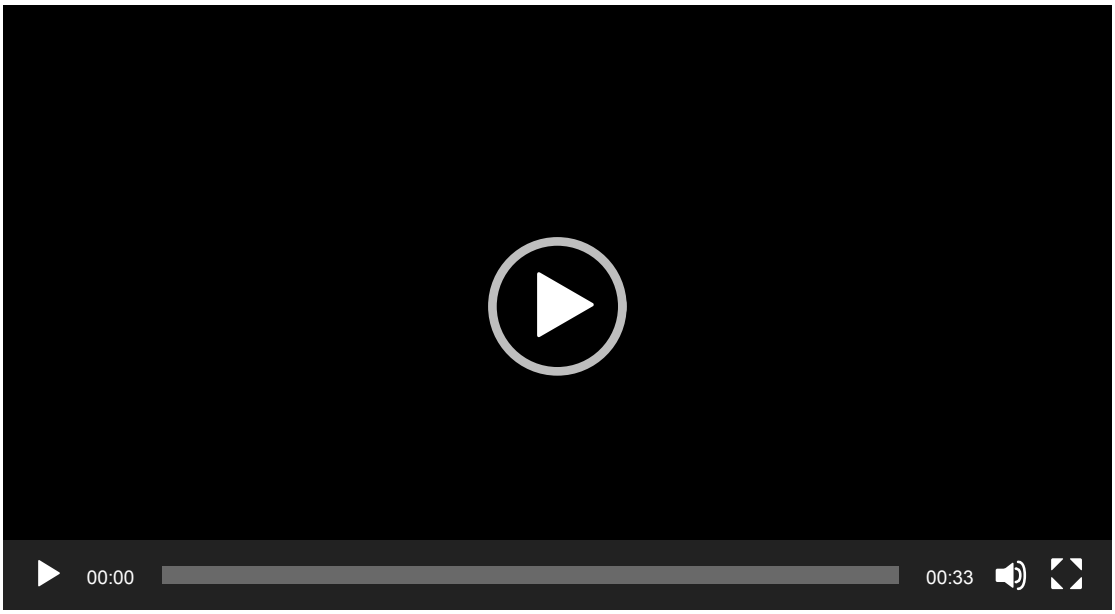
We were rather busy since Saturday to confirm, reevaluate, screen and communicate our data about ncov protease. Refering strictly only to public media here, f.i. NYT from yesterday **without any other governmental, industrial or scientific or medical source from any continent**, it's officially confirmed since yesterday that China tests Lopinavir for the treatment of the coronavirus in hospitals. Lopinavir was identified as one of the top targets by Innophore Catalophore search last week. This compound has already shown promising results in SARS that has a structurally very similar protease. We finished running the (short) MD simulations on different conformations based on the Catalophore point-cloud alignment and (re)- docking lopinavir into the 2019ncov virus protease model. The docking experiment produced 8 clusters of possible conformations, we chose 6 out of 8 conformers and ran an all-atom 300 ps MD at 310 K (=36.85°C) for each one of them. The conformers have been named LP1 (highest

binding energy) to LP6 (lowest binding energy) based on the docking clustering. We will publish the trajectories today here a little later.



.

The conformer in the video is the number LP1.



Update January 29th 2020 21:30UTC

A 2019-ncov protease model was published by Xu etl al. yesterday. Their sequence alignment is basically identical to ours from 23rd of January. However, we couldn't find the structure files for download. If anybody has access to a model that differs significantly from hours, please let us know.

Figure 1. Sequence alignment of 2019-nCov M^{pro} and SARS M^{pro}.

References

- [i] https://www.who.int/publications-detail/surveillance-case-definitions-for-human-infection-with-novel-coronavirus-(ncov)
- [ii] https://www.bbc.com/news/world-asia-china-51217455
- [iii] https://edition.cnn.com/2020/01/23/china/wuhan-coronavirus-update-intl-hnk/index.html
- [iv] https://www.cbc.ca/news/health/coronavirus-human-to-human-1.5433187
- [v] https://www.reuters.com/article/us-china-health-pneumonia-who-idUSKBN1ZD16J
- [vi] https://en.wikipedia.org/wiki/Novel_coronavirus_(2019-nCoV)
- [vii] https://link.springer.com/protocol/10.1007%2F978-1-4939-2438-7_1
- [viii] https://www.asiatimes.com/2020/01/article/wuhan-disease-spread-by-bats-animals-expert/
- [ix] https://www.webmd.com/lung/news/20200122/what-to-know-about-new-coronavirus-from-china
- [x] https://proteopedia.org/wiki/index.php/HIV-1_protease
- [xi] https://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-15-S7-S5
- [xii] https://pubchem.ncbi.nlm.nih.gov/compound/Mozenavir
- [xiii] https://www.biorxiv.org/content/10.1101/2020.01.20.913368v1
- [xiv] https://www.biorxiv.org/content/10.1101/2020.01.20.913368v1
- [xv] https://www.ncbi.nlm.nih.gov/nuccore/MN908947
- [xvi] https://www.ncbi.nlm.nih.gov/nuccore/MN908947.3?location=266:13468,13468:21555

- [xvii] https://www.rcsb.org/structure/5N5O
- [xviii] https://www.rcsb.org/structure/3TLO
- [xix] https://www.rcsb.org/structure/2A5K
- [xx] http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index
- [xxi] http://scop.berkeley.edu/sunid=146586&ver=1.75
- [xxii] https://www.rcsb.org/structure/2DUC
- [xxiii] https://www.rcsb.org/structure/2H2Z
- [xxiv http://europepmc.org/article/med/17189639

Related Posts:

1. **Biocatalysis: A Pharma Perspective**
2. **Innophore & SignalChem start partnership**
3. **Drivers of innovation Spin-offs & Start-ups: Enzymes for industry – JUST Business**
4. **Meet us at the EuropeCEO 2019 in Vienna!**

Search

Categories

Awards (6)

Conference (5)

Drug Discovery (1)

Enzyme Discovery (2)

Innophore (27)

Media (32)

News (25)

Online (7)

Paper (2)

Podcast, Movies (4)

Print (10)

Research (3)

Recent Comments

Archives

January 2020

November 2019

October 2019