Electronic supplementary materials *Mendeleev Commun.*, 2022, **32**, 334–335

Identification of natural compounds targeting SARS-CoV-2 Mpro by virtual screening and molecular dynamics simulations

Chuanming Zhang, Chao Zhang, Yanli Meng, Tai Li, Zhe Jin, Shicheng Hou and Chun Hu

1. RESULTS AND DISSCUSSION

1.1 Virtual screening hits

All natural compounds were docked into the recognized binding pocket of the SARS-CoV-2 Mpro to identify compounds with lower or similar docking scores than the control drug, remdesivir (Glide score: -10.437 kcal mol^{-1}). Then, the three compounds (pentagalloylglucose, malonylawobanin, gnetin E dihydride with Glide score value of -13.387 kcal mol^{-1}, -12.020 kcal mol^{-1} and -10.003 kcal mol^{-1}, respectively) with the similar binding affinity with the active sites of the SARS-CoV-2 Mpro were selected for the further analysis.

Molecular docking analysis revealed that the best docked compounds had similar interactions with any of the two key amino acids (His 41 and His 164) (Figure S1). Pentagalloylglucose had H-bond interaction with Thr 26, His 41, Ser 46, His 164, Asp 187 and Thr 190. Malonylawobanin formed H-bonds with His 41, Gly 143, His 163, His 164, Glu 166 and Thr 190. The other natural compound, gnetin E dihydride, generated H-bond interactions with four amino acids (Thr 26, Cys 145, His 164 and Glu 166) and π-π interaction with His 41. Meanwhile, remdesivir formed H-bonds with Thr 26, Thr 45 and Gly 143. Three natural compounds generated more H-bonds with lower or similar Glide scores compared with remdesivir, and hydroxy may be a main group to make protein-ligand bind more firmly. Altogether, these three natural compounds developed strongly interaction with one or both catalytic residues (His 41 and Cys 145) of Mpro, which was similar to Das's and Ghosh's study [S1,S2].

1.2 Molecular dynamics simulations analysis

Molecular dynamics could provide information about inner motions of time with the protein-ligand complex. To study the steady nature and conformations stability of SARS-CoV-2 Mpro, 100 ns molecular dynamics simulations were performed on the apo-state of Mpro, and their complexes (Mpro-ligand: Mpro-pentagalloylglucose, Mpro-malonylawobanin, and Mpro-gnetin E dihydride). The behaviors and interactions of proteins and ligands during the course of simulation could be conveyed graphically in Simulation Interactions Diagram panel of Desmond package.

The ligand torsions plot (Figure S2, pentagalloylglucose; Figure S3, malonylawobanin; Figure S4, gnetin E dihydride) summarized the conformational evolution of every rotatable bond (RB) in the ligand throughout the simulation trajectory (0 through 100 ns). Radial plots on the left described the angle of each bond at a given time during simulations. The beginning of the simulation was in the center of the radial plot and the time evolution is plotted radially outwards. The simulation started at the center of the radial plot, and the time evolution was plotted radially outwards. The histogram summarized the data on the radial plots by showing the probability density of the torsion, which might give insights into the conformational strain the ligand undergone to maintain a protein-bound conformation[S3]. In combination with Figure S1, Figure S2, Figure S3 and Figure S4, the relationship between the torsional angle changes and the interactions could be analyzed.

Depending on Figure S1 and Figure S2, the torsional angle changes of pentagalloylglucose in p promoted the π-π interaction between the aromatic ring and residue His 41. Meanwhile, the torsional angle changes in a, l, m, s, r, w and x were corresponding to the H-bond interaction between the hydroxyl groups and residue Thr 26, Ser 46, His 164, Asp 187, Arg 188 and Thr 190. It was obvious to see that the bonds r and s had stronger rigidity, and bonds a, l, m, w and x were more flexible. In Figure S1 and Figure S3, the torsional angle changes of malonylawobanin in g, h, m, s, t, u and a1 were promoted the H-bond interaction between the hydroxyl groups and residue Thr 190, His 164, Glu 166, His 163 and His 41. Bonds a1 showed stronger rigidity, and bonds g, h, m, s, t and u seem to be more flexible according to the radial plots and the histogram. As for gnetin E dihydride (Figure S1 and Figure S4), the torsional angle changes in a, g, i, m and n corresponded to the H-bond interaction between the hydroxyl groups and residue Glu 166, Phe 140, His 164, Cys 145 and Thr 26. Radial plots and the histogram showed that the bonds a, g, i, m and n were flexible. It was interesting to see that the three compounds with relatively more flexible bonds, such as gnetin E dihydride, showed higher Glide score. This might be a coincidence, but guiding our design of the structural skeleton about on novel and effective virus inhibitors targeting SARS-CoV-2 Mpro, such as which parts of compounds should be rigid and which parts should be flexible.

The Root Mean Square fluctuation (RMSF) of alpha-carbon atoms of the four systems (Mpro, Mpro-pentagalloylglucose, Mpro-malonylawobanin, and Mpro-gnetin E dihydride) were represented in Figure S5. The curves numbers 1, 2, 3 and 4 of the RMSF diagrams represented proteins and ligands features of the free protein, Mpro-pentagalloylglucose, Mpro-malonylawobanin, and Mpro-gnetin E dihydride).All these showed a similar kind of fluctuation pattern. It was clear that all these three complexes (Mpro-pentagalloylglucose, Mpro-malonylawobanin, and Mpro-gnetin E dihydride) exhibited relatively less conformational fluctuation than Mpro. Meanwhile, the fluctuation of the crucial amino acids (Thr 26, His 41, His 164 and Glu 166) in these three complexes were also less than that in Mpro, indicating that these amino acids within the active site of Mpro experienced interaction with pentagalloylglucose, malonylawobanin, and gnetin E dihydride.

The protein–ligand (Mpro-pentagalloylglucose, Mpro-malonylawobanin, and Mpro-gnetin E dihydride) contacts and contacts over the course of trajectory were showed in Figure S6 and supported in supplemental Table S1, Table S2 and Table S3. Glu 166 formed water bridges and hydrogen bonds with pentagalloylglucose, which maintaining approximately by 330% of the interactions fraction (Supplemental Table S1). There were

mainly hydrophobic interactions between pentagalloylglucose and these seven amino acids (Thr 25, Thr 26, His 41, Glu 47, Phe 140, Asn 142 and Glu 166), and formed water bridges with the other key amino acid (Ser 46, Gly 143, Cys 145, His 163 and His 164). The total interactions fraction maintaining over 70% was Thr 26, His 41, Glu 47, Asn 142, Ser 144, Glu 166 and Gln189 (Supplemental Table S2). Malonylawobanin generated hydrogen bonds interactions with Thr 25, Thr 26, Tyr 54, Gly 143, Cys 145, Glu 166, Arg 188, Gln 189 and Gln 192. Glu 166 formed water bridges and hydrogen bonds with malonylawobanin, which maintaining approximately by 240% of the interactions fraction (Supplemental Table S2). It was obvious that Glu 166 was forming a crucial interaction between malonylawobanin and protein. Some key amino acid formed water bridges with malonylawobanin showing in Figure S5 were His 41, Gly 143, His 164, Gln 189 and Thr 190. These amino acid (Thr 25, Thr 26, His 41, Met 49, Tyr 54, Gly 143, Glu 166, Arg 188, Gln 189 and Thr 190) maintained approximately at 70% of the interactions fraction (Supplemental Table S2). It was obvious that two residues (Thr 26 and Glu 166) interacted with pentagalloylglucose and malonylawobanin almost in each trajectory frame (Figure S5, Supplemental Table S1 and Table S2). As for gnetin E dihydride, these seven amino acids (Thr 24, Thr 26, Ser 46, Asn 142, Gly 143, Cys 145 and Asp 187) had hydrogen bonds integration. Thr 24, Thr 26, His 41, Cys 44, Thr 45, Tyr 54, Asn 142, Glu 166 and Gln 189 generated water bridges with gnetin E dihydride. Malonylawobanin generated hydrophobic with His 41, Met 49, His 164 and Met 165. The interactions fraction of gnetin E dihydride with Thr 26, His 41, Asn 142, Gly 143 and Asp 187 were all up to 70%. Thr 26 and Asp 187 interacts with gnetin E dihydride almost in each trajectory frame (Figure S5, Supplemental Table S3). In conclusion, Thr 26, His 41, His 164 and Glu 166 were crucial amino acids in forming water bridges, hydrophobic or hydrogen bonds between protein and ligands (pentagalloylglucose, malonylawobanin and gnetin E dihydride).

1.3 The calculation for prime molecular mechanics / generalized born surface area (MM-GBSA)

The total binding free energy consisted of the Coulomb energy (Coulomb), the covalent binding (Covalent), the hydrogen bonding (Hbond), the lipophilic binding (Lipo), the π-π packing interaction (Packing), the solvent generalized binding (SolvGB) and the binding from the van der Waals energy (VDW). Compounds with better binding score in Glide XP were further selected to calculate binding energy using the MM/GBSA method. The binding free energy values of Mpro-pentagalloylglucoseMpro-malonylawobanin and Mpro-gnetin E dihydride complexes were -85.63, -59.59 and -53.80 kcal mol^{-1}, respectively. It was clear that Mpro-pentagalloylglucose showed higher free binding energy compared with other complexes. Coulomb energy and van der Waals energy play the key component in MM-GBSA value for Mpro-pentagalloylglucose complex system.

2. MATERIALS AND METHODS

2.1 Structure-based virtual screening

We use Glide in Schrödinger Maestro Suite 2016 to explore protein-ligand interactions, calculate the corresponding binding free energy and elucidate their binding mechanisms between natural compounds and SARS-CoV-2 Mpro[S4,S5]. Remdesivir is being used as control drug, which has been screened for the treatment of COVID-19 and strongly binds to both RNA-dependent RNA polymerase and SARS-CoV-2 Mpro[S6,S7].

The crystal structure of SARS-CoV-2 Mpro (PDB ID: 6LU7)[S8] was obtained from RCSB Protein Data Bank (RCSB PDB, http://www.rcsb.org/). The structure was cleaned up via the Protein Preparation Wizard in Schrödinger Suite[S9]. This step removed co-crystallized water in the protein structure, created disulfide bonds, corrected the bond orders and added hydrogen atoms before docking[S10-S12]. Afterwards, protein restrained minimization was performed with the selected OPLS_2005 (optimized potential for liquid simulations) force field, and the root means square deviation (RMSD) was set to 0.50Å[S13]. The original ligand N3, which located in the center and docked in the receptor, helped us to determin the binding cavity and generate the grid box of SARS-CoV-2 Mpro cavity. Ligand docking was used for glide docking of the screened compounds. The parameters in the docking process were set as defaults. 5,600 natural compounds were docked through HTVS mode, and then the top 10 % ranking ligands were re-docked with Glide SP docking. Subsequently, Glide XP docking was performed to the top 10 % natural compounds. After three steps (HTVS, SP and XP) of virtual screening to natural compounds, we obtained three compounds (pentagalloylglucose[S14], malonylawobanin[S15] and gnetin E dihydride[S16] with well docking score through this rigorous docking based in-silico filtering.

2.2 Molecular dynamics simulations

We proceeded the MD on Desmond v4.3 suite software in M40 GPU processor with 128 GB RAM, running on Red Hat 6.10 Linux operating system. The protein-ligand complexes were generated by the protein preparation wizard of Desmond module with the explicit solvent model and SPC and the orthorhombic box shape. And we set the size of the box at a distance of 10 Å from the outermost atom of the protein-ligand complex. Sequentially, to set the ionic strength, we generated the default environmental solution (NaCl) with the approximately physiological concentration of 0.15M into the simulation box to set the ionic strength. We established this biological system composed of the receptor, small molecules, explicit solvent, counter ions, and the salt. Ultimately, the minimization jobs would be performed to relax the system into a local energy minimization, and then submit this model system. The specific steps were operated according to the previous work[S17-S21].

The Simulation Quality Analysis tools and the Simulation Event Analysis tool were used to assess and predict the quality of MD simulations, respectively. In addition, the Simulation Interaction Diagram tool was used to identify protein-ligand interactions.

2.3 Binding free energy calculation

The conformation obtained from MD simulations was energy-minimized by the Prime version 3.5 module. The Molecular Mechanics Generalized Born Surface Area (MM-GBSA) method was used to calculate the binding-free energy of the complexes, using the following equation[S22-S27].

$$\Delta G_{GB} = \Delta E_{MM} + \Delta G_{solv} + \Delta G_{SA}$$

where the electrostatic solvation energy (ΔG_{GB}) was based on generalizing Born (GB) model; ΔE_{MM} included the van der Waals energy, internal energy and electrostatic energy contributions and was the difference in the minimized energy between the minimized energy of protein-ligand complexes and the total energy of protein and ligand in free form; ΔG_{solv} was the sum of the non-polar solvation energy and polar (electrostatic) solvation energy contributions; ΔG_{SA} was calculated using the solvent-accessible surface area (SASA).

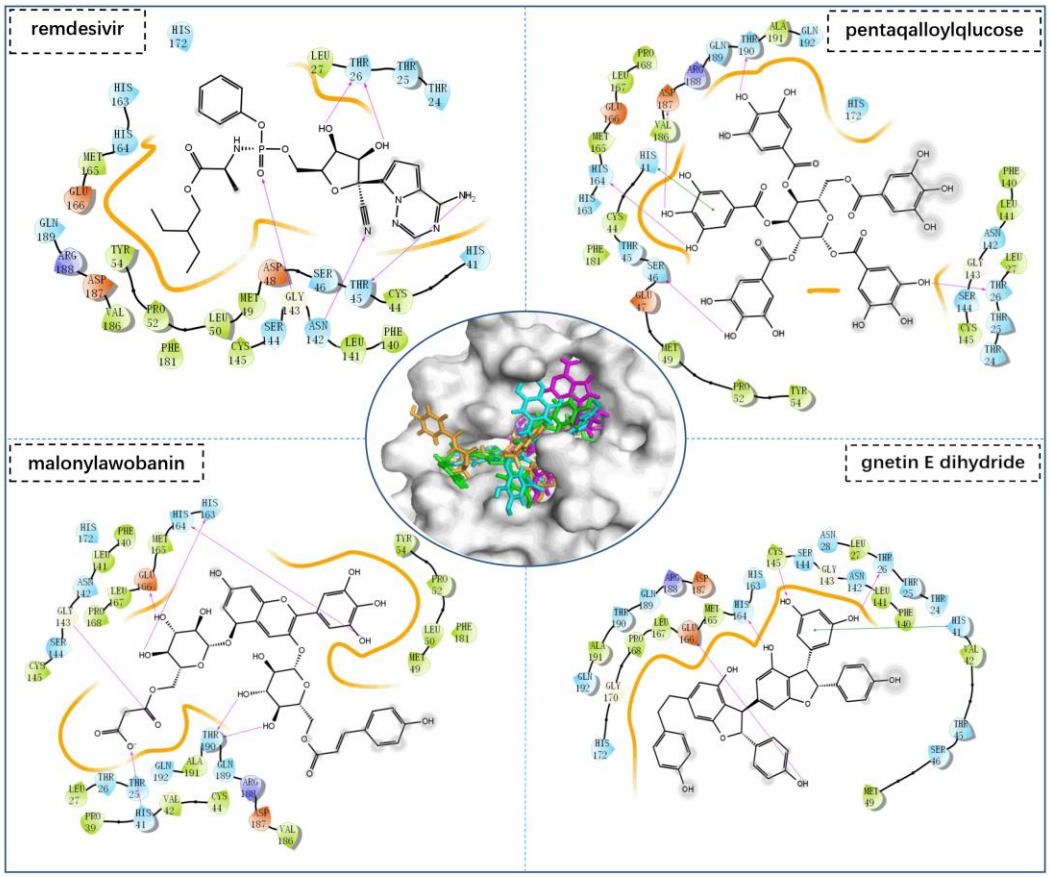

Figure S1 Two-dimensional (2D) diagrams of protein-ligand interactions in the active sites of the Mpro. The H-bond inter-actions with residues are represented by a purple dashed arrow directed towards the electron donor. The π-π interactions are represented by a green line.

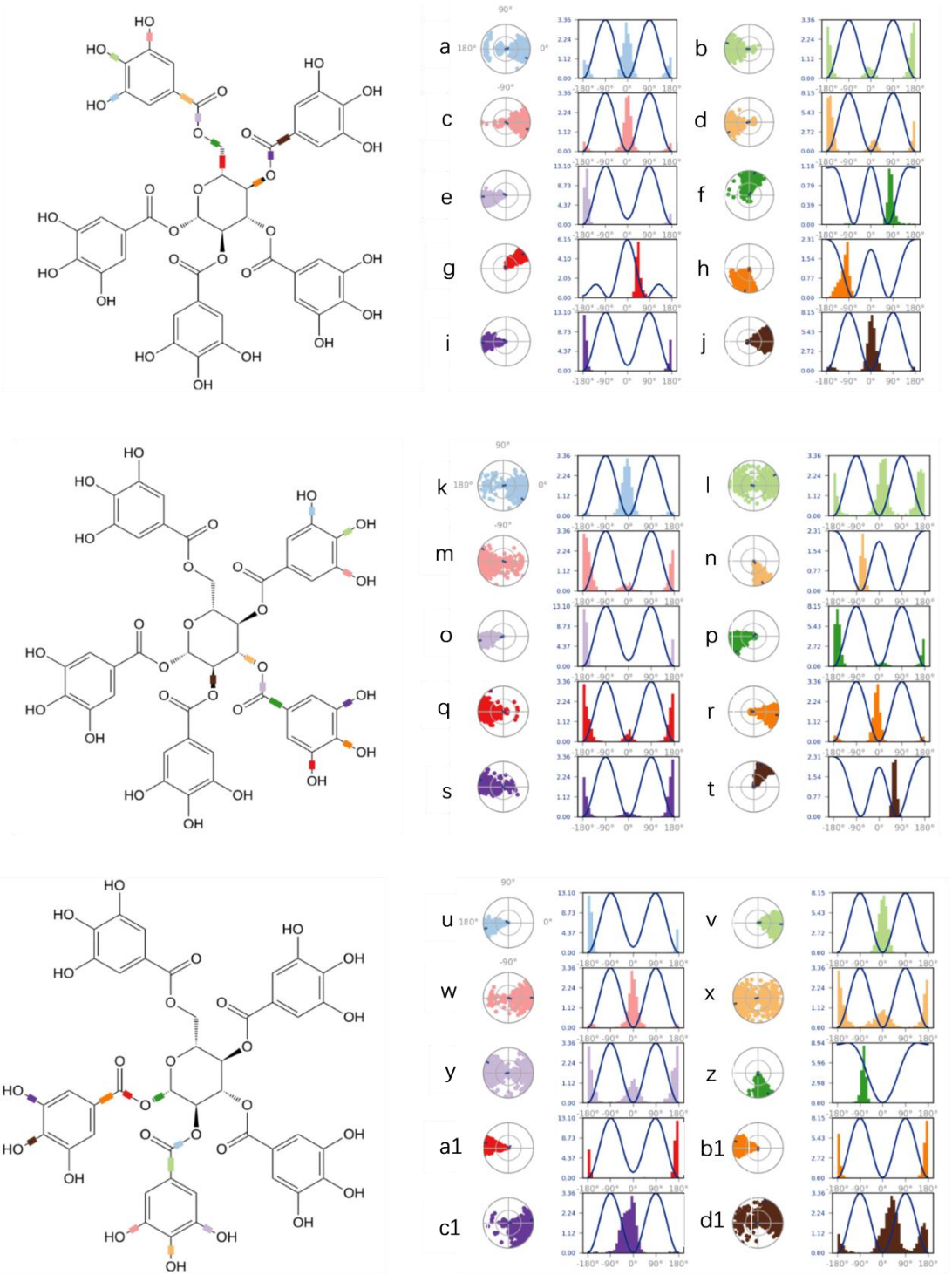

Figure S2. Graphics of the torsional conformation of each rotatable bond of the ligand pentagalloylglucose in SARS-CoV-2 Mpro pocket during MD simulations.

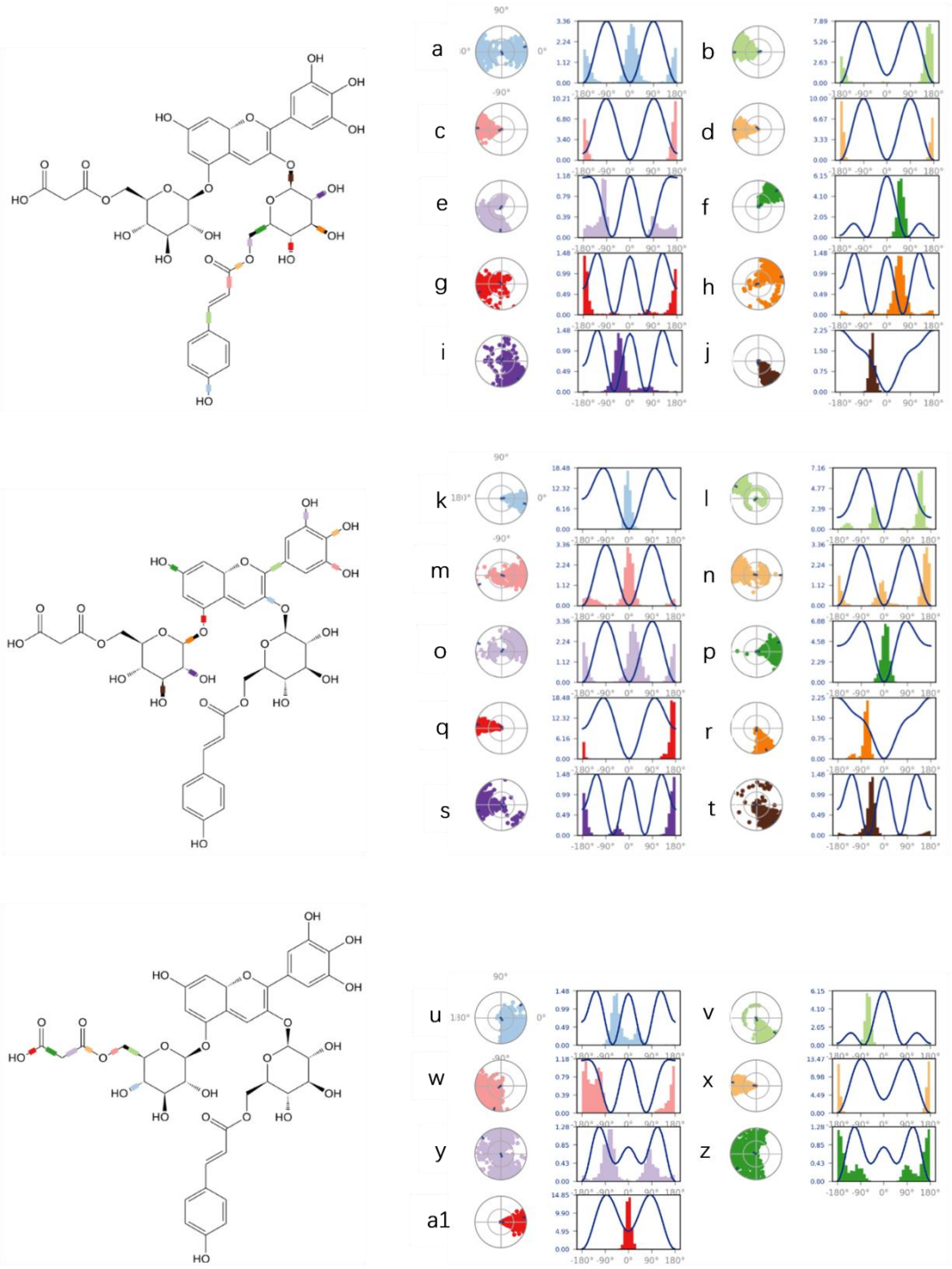

Figure S3. Graphics of the torsional conformation of each rotatable bond of the ligand malonylawobanin in SARS-CoV-2 Mpro pocket during MD simulations.

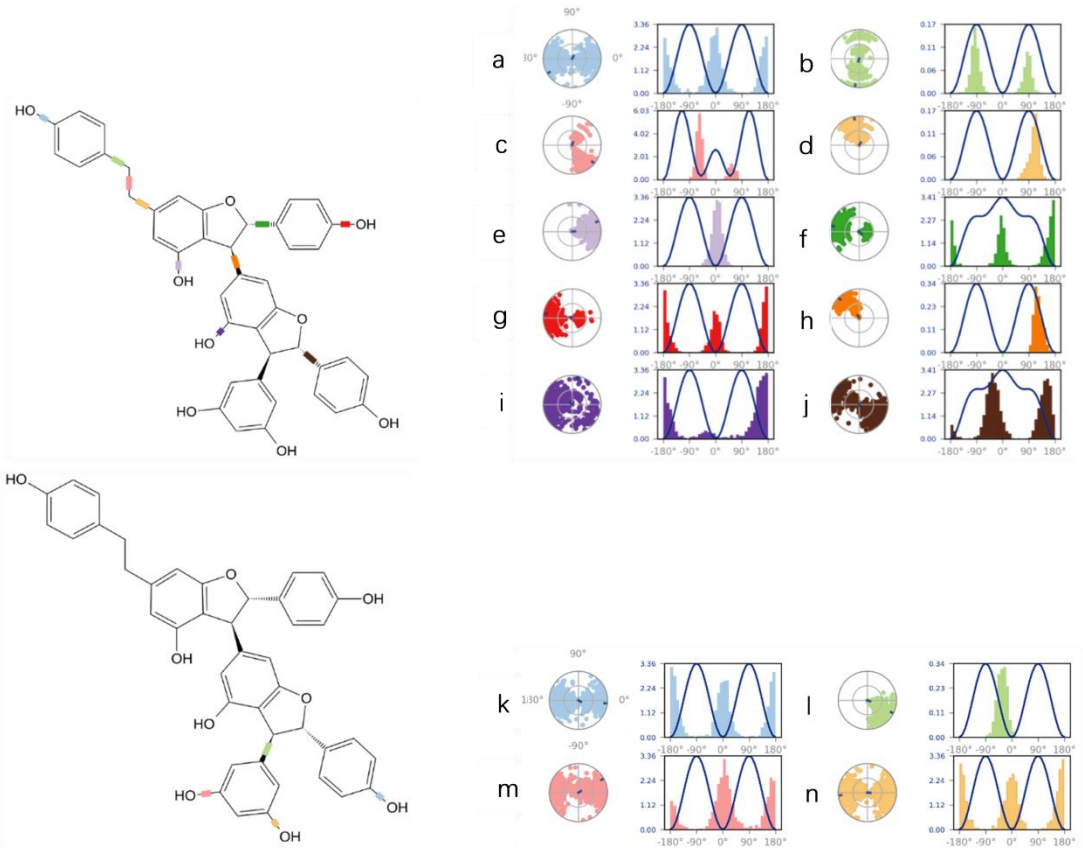

Figure S4. Graphics of the torsional conformation of each rotatable bond of the ligand gnetin E dihydride in SARS-CoV-2 Mpro pocket during MD simulations.

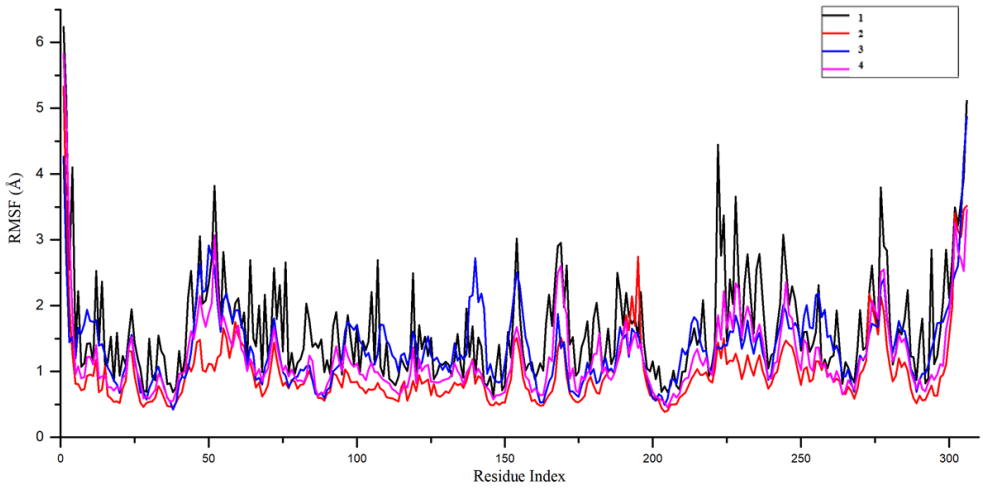

Figure S5 The RMSF curves of MPro, Mpro-pentagalloylglucose, Mpro-malonylawobanin, and Mpro-gnetin E dihydride during 100 ns simulations

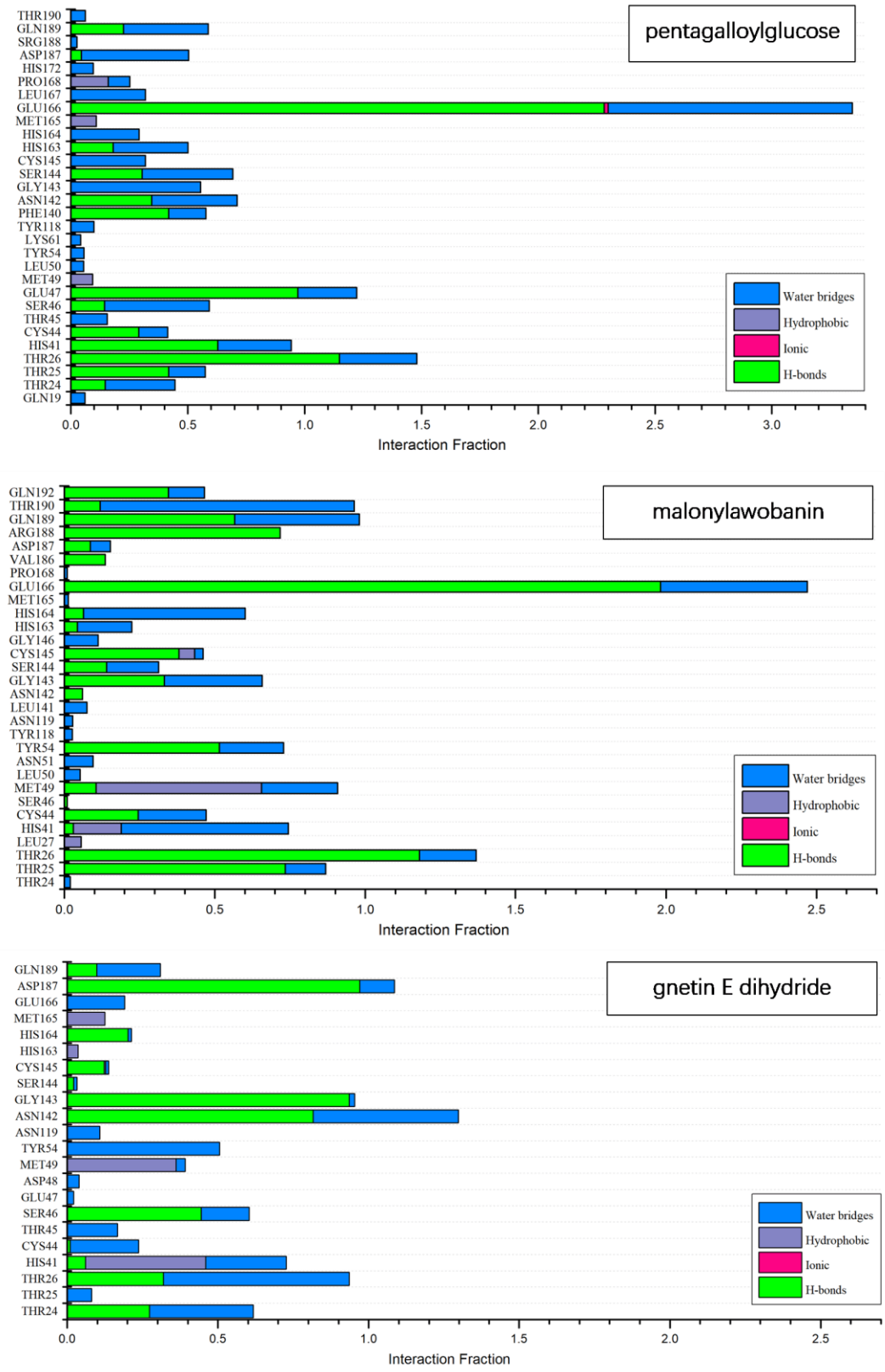

Figure S6. The bar charts of protein–ligand contacts and contacts over the course of trajectory

Table S1 The specfic Mpro-pentagalloylglucose contacts and contacts over the course of trajectory

Amino acid	H-bonds	Ionic	Hydrophobic	Water bridges
GLN19	0	0	0	0.061
THR24	0.148	0	0	0.298
THR25	0.419	0	0	0.156
THR26	1.149	0	0	0.332
HIS41	0.63	0	0	0.314
CYS44	0.291	0	0	0.124
THR45	0	0	0	0.156
SER46	0.145	0	0	0.448
GLU47	0.972	0	0	0.251
MET49	0	0	0.094	0
LEU50	0	0	0	0.055
TYR54	0	0	0	0.057
LYS61	0	0	0	0.042
TYR118	0	0	0	0.099
PHE140	0.419	0	0	0.159
ASN142	0.347	0	0	0.365
GLY143	0	0	0	0.556
SER144	0.305	0	0	0.389
CYS145	0	0	0	0.32
HIS163	0.182	0	0	0.319
HIS164	0	0	0	0.292
MET165	0	0	0.109	0
GLU166	2.283	0.018	0	1.043
LEU167	0	0	0	0.319
PRO168	0	0	0.161	0.092
HIS172	0	0	0	0.097
ASP187	0.045	0	0	0.46
SRG188	0	0	0	0.027
GLN189	0.226	0	0	0.362
THR190	0	0	0	0.063
GLN192	0	0	0	0.022

Table S2 The specfic Mpro-malonylawobanin contacts and contacts over the course of trajectory

Amino acid	H-bonds	Ionic	Hydrophobic	Water bridges
THR24	0	0	0	0.02
THR25	0.734	0	0	0.135
THR26	1.18	0	0	0.189
LEU27	0	0	0.056	0
HIS41	0.031	0	0.159	0.555
CYS44	0.246	0	0	0.226
SER46	0.01	0	0	0
MET49	0.106	0	0.549	0.253
LEU50	0	0	0	0.053
ASN51	0	0	0	0.096
TYR54	0.516	0	0	0.213
TYR118	0	0	0	0.027
ASN119	0	0	0	0.028
LEU141	0	0	0	0.076
ASN142	0.061	0	0	0
GLY143	0.333	0	0	0.325
SER144	0.141	0	0	0.173
CYS145	0.381	0	0.052	0.028
GLY146	0	0	0	0.113
HIS163	0.044	0	0	0.181
HIS164	0.064	0	0	0.537
MET165	0	0	0	0.014
GLU166	1.982	0	0	0.488
PRO168	0	0	0	0.01
VAL186	0.136	0	0	0
ASP187	0.087	0	0	0.067
ARG188	0.718	0	0	0
GLN189	0.566	0	0	0.414
THR190	0.12	0	0	0.844
GLN192	0.346	0	0	0.12

Table S3 The specfic Mpro-gnetin E dihydride contacts and contacts over the course of trajectory

Amino acid	H-bonds	Ionic	Hydrophobic	Water bridges
THR24	0.273	0	0	0.344
THR25	0	0	0	0.081
THR26	0.319	0	0	0.616
HIS41	0.061	0	0.398	0.268
CYS44	0.01	0	0	0.227
THR45	0	0	0	0.167
SER46	0.445	0	0	0.159
GLU47	0	0	0	0.021
ASP48	0	0	0	0.04
MET49	0	0	0.361	0.03
TYR54	0	0	0	0.506
ASN119	0	0	0	0.108
ASN142	0.816	0	0	0.481
GLY143	0.935	0	0	0.018
SER144	0.022	0	0	0.011
CYS145	0.123	0	0.005	0.01
HIS163	0	0	0.036	0
HIS164	0.202	0	0	0.011
MET165	0	0	0.125	0
GLU166	0	0	0	0.191
ASP187	0.97	0	0	0.115
GLN189	0.098	0	0	0.211

Table S4. The predicted binding free energy about protein-ligand using MM-GBSA method.

Energy (kcal mol^{-1})	Mpro-pentagalloylglucose	Mpro-malonylawobanin	Mpro-gnetin E dihydride
Total binding free energy	-85.63	-59.59	-53.80
Coulomb	-58.14	-14.7	-43.22
Covalent	14.49	6.75	10.11
Hbond	-6.41	-6.18	-3.47
Lipo	-24.52	-22.04	-20.41
Packing	-5.12	-1.49	-4.91
SolvGB	61.88	34.41	54.80
VDW	-67.81	-56.34	-45.98

REFERENCES

S1. S. Das, S. Sarmah, S. Lyndem and A. S. Roy, *J Biomol Struct Dyn*, 2021, 39(9), 3347-3357 .
S2. R. Ghosh, A. Chakraborty, A. Biswas and S. Chowdhuri, *J Biomol Struct Dyn*. 2021, 39(12), 4362-4374.
S3. J. Zhang, X. Liu, S. Q. Wang, G. Y. Liu, W. R. Xu, X. C. Cheng and R. L. Wang, *J Biomol Struct Dyn*, 2017, 35 (12), 2665-2680.
S4. R. A. Friesner, R. B. Murphy, M. P. Repasky, L. L. Frye, J. R. Greenwood, T. A. Halgren, P. C. Sanschagrin and D. T. Mainz, *J Med Chem*, 2006, 49 (21), 6177-6196.
S5. M. Wang, W. Li, Y. Wang, Y. Song, J. Wang and M. Cheng, *J Mol Graph Model*,, 2018, 84, 18-28.
S6. H. L. Nguyen, N. Q. Thai, D. T. Truong and M. S. Li, *J Phys Chem B*, 2020, 124(50), 11337-11348.
S7. M. Costanzo, M. A. Rachele De Giglio and G. N. Roviello, *Curr Med Chem*, 2020, 27(27), 4536-4541.
S8. Z. Jin, X. Du, Y. Xu, Y. Deng, M. Liu, Y. Zhao, B. Zhang, X. Li, L. Zhang, C. Peng, Y. Duan, J. Yu, L. Wang, K. Yang, F. Liu, R. Jiang, X. Yang, T. You, X. Liu, X. Yang, F. Bai, H. Liu, X. Liu, L.W. Guddat, W. Xu, G. Xiao, C. Qin, Z. Shi, H. Jiang, Z. Rao and H. Yang, *Nature*, 2020, 582(7811), 289-293.
S9. G. M. Sastry, M. Adzhigirey, T. Day, R. Annabhimoju and W. Sherman, *J Compu Aid Mol Des*, 2013, 27 (3), 221-234.
S10. Z. Jin, Y. Wang, X. F. Yu, Q. Q. Tan, S. S. Liang, T. Li, H. Zhang, P. C. Shaw, J. Wang and C. Hu, *Comput Biol Chem*, 2020, 85, 107241.
S11. P. Mehta and R. Malik, *J Biomol Struct Dyn*, 2020, 38(18), 5320-5337.
S12. D. Shivakumar, J. Williams, Y. Wu, W. Damm, J. Shelley and W. J. Sherman, *J Chem Theory Comput*, 2010, 6 (5), 1509-1519.
S13. R. Itteboina, S. Ballu, S. Sivan and V. Manga, *J Recept Sig Transd*, 2017, 37 (5), 453-469.

S14. K. Ono, T. Sawada, Y. Murata, E. Saito, A. Iwasaki, Y. Arakawa, K. Kurokawa and Y. Hashimoto, *Clin Chim Acta*, 2000, 290 (2), 159-167.
S15. T. Goto, T. Kondo, H. Tamura and S. Takase, *Tetrahedron Lett*, 1983, 24 (44), 4863-4866.
S16. A. P. Lins, M. N. S. Ribeiro, O. R. Gottlieb and H. E. Gottlieb, *J Nat Prod*, 1982, 45 (6), 754-761.
S17. X. Liu, Z. Jing, W. Q. Jia, S. Q. Wang, Y. Ma, W. R. Xu, J. W. Liu and X. C. Cheng, *J Biomol Struct Dyn*, 2018, 36 (11), 2988-3002.
S18. W. Q. Jia, Z. Jing, X. Liu, X. Y. Feng, Y. Y. Liu, S. Q. Wang, W. R. Xu, J. W. Liu and X. C. Cheng, *J Biomol Struct Dyn*, 2018, 36 (13), 3496-3512.
S19. Y. Y. Liu, X. Y. Feng, W. Q. Jia, Z. Jing, W. R. Xu and X. C. Cheng, *Comput Biol Chem*, 2019, 78, 190-204.
S20. M. Wang, Y. Wang, D. Kong, H. Jiang, J. Wang and M. Cheng, *Comput Biol Chem*, 2018, 77, 214-225.
S21. Y. Wang, S. Feng, H. Gao and J. Wang, *J Biomol Struct Dyn*, 2020, 38(5), 1435-1447.
S22. J. Du, H. Sun, L. Xi, J. Li, Y. Yang, H. Liu and X. Yao, *J Comput Chem*, 2011, 32 (13), 2800-2809.
S23. Y. Wang, B. Hu, Y. Peng, X. Xiong, W. Jing, J. Wang and H. Gao, *J Chem Inf Model*, 2019, 59 (5), 2309-2323.
S24. Z. Z. Wang, C. Y. Ma, J. Yang, Q. B. Gao, X. D. Sun, L. Ding and H. M. Liu, *J Mol Struct*, 2019, 1175, 698-707.
S25. Z. Z. Wang, J. Yang, X. D. Sun, C. Y. Ma, Q. B. Gao, L. Ding and H. M. Liu, *J Biomol Struct Dyn*, 2019, 37 (13), 3482-3495.
S26. Y. Xu, J. Lee, Y. D. Park, J. M. Yang, J. Zheng and Q. Zhang, *J Biomol Struct Dyn*, 2017, 36 (4), 830-840.
S27. C. Mulakala and V. N. Viswanadhan, *J Mol Graph Model*, 2013, 46, 41-51.