Protein degradation: a novel computational approach to design protein degrader probes for main protease of SARS-CoV-2

Muhammed Shaheer, Ravi Singh and M. Elizabeth Sobhia

Department of Pharmacoinformatics, National Institute of Pharmaceutical Education and Research (NIPER), Mohali, Punjab, India

Communicated by Ramaswamy H. Sarma

ABSTRACT
Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has afflicted many lives and led to approvals of drugs and vaccines for emergency use. Even though vaccines have emerged, the high mortality of COVID-19 and its insurgent proliferation throughout the masses commands an innovative therapeutic proposition for the treatment. Targeted protein degradation has been applied to various disease domains and we propose that it could be incredibly beneficial to tackle the current pandemic. In this study, we have attempted to furnish insights on the design of suitable PROTACs for the main protease (Mpro) of SARS-CoV-2, a protein that is considered to be an essential target for viral replication. We have employed protein-protein docking to predict the possible complementarity between a cereblon E3 ligase and Mpro of SARS-CoV-2, and estimate possible linker length. Molecular Dynamic simulation and analysis on generated ternary complexes demonstrated stable interactions that suggested that designed PROTAC has a potential to cause degradation. The superior characteristics rendered by PROTACs led us to propose them as possibly the next-generation antiviral drugs for SARS-CoV-2.

ARTICLE HISTORY
Received 11 February 2021
Accepted 2 July 2021

KEYWORDS
PROTAC; SARS-CoV-2; main protease; protein-protein docking; MD simulation

Introduction
The current coronavirus disease (COVID-19) pandemic that began a year ago with the outbreak of SARS-CoV-2 has infected more than 181 million individuals worldwide as of June 2021, and the death toll continues to rise to over 3.9 million (WHO, 2021). However, a few drugs have emerged as candidates for clinical use for SARS-CoV-2 (Lamb, 2020; Wu et al., 2020). Several publications have appeared in recent years documenting repurposing studies to tackle the pandemic (Bhardwaj et al., 2021; Gordon et al., 2020; Riva et al., 2020; Sang et al., 2020; Schlagenhauf et al., 2020; Trezza et al., 2020; Zhou et al., 2020). Even though repurposing is the fastest method to identify potential drugs already available in the market, other domains should be explored with the vision of improved therapeutic efficacy.

Conventional drug therapy modifies the functions of disease-related proteins by using small molecules that occupy the protein’s active site to arrest or enhance its action – a strategy called occupancy-driven pharmacology. However, less than 20% of the approximately 20,300 known human proteins are considered drug targets which can be explored therapeutically by conventional drug discovery processes (Paik et al., 2012; Wishart et al., 2006). In the past few decades, several novel pharmacological approaches have emerged to overcome this limitation. These include monoclonal antibodies (Adams & Weiner, 2005) and genetic engineering techniques like clustered regularly interspaced short palindromic repeats (CRISPR)-cas9 (Cong et al., 2013) and RNA interference (RNAi) (Hannon, 2002) based approaches. Nevertheless, these advancements have their limitations (Neklesa et al., 2017). Hence, a better molecule with better pharmacokinetic
properties is the need of the hour. Concerning this, many new methodologies have been developed so far.

The Nobel Prize in Chemistry 2004 was awarded jointly to Aaron Ciechanover, Avram Hershko, and Irwin Rose for ‘The Discovery of ubiquitin-mediated protein degradation’. That event amplified the development and recognition of targeted protein degradation (TPD) utilizing the ubiquitin-proteasome system (UPS). The UPS functions by a series of enzymes that covalently link ubiquitin (Ub) to Lys residue on the target protein. Ubiquitin fundamentally has Lys residues which further leads to polyubiquitination. When the chain is at least 4 Ub long, the protein gets recognized by the 26S proteasome, a large proteolytic complex that degrades ubiquitinated proteins into smaller peptides. Three enzymes are involved in the ubiquitination process – E1 (Ub-activating enzyme), E2 (Ub-carrier), and E3 (ubiquitin ligase). E1 activates Ub in ATP dependent process, which is then transferred to E2, where E3 catalyzes the transfer of Ub to Lys residue of a target protein. E3 ligase is the substrate recognizing unit with 600–700 isoforms in mammals (Hershko & Ciechanover, 1998).

Targeted protein degradation (TPD) is an approach that deploys proteolysis targeting chimera (PROTAC or degrader), a promising technology that utilizes small molecule-based heterobifunctional entities, to modulate the level of protein of interest (POI) by hijacking the ubiquitin-proteasome system (UPS) to induce selective degradation. This strategy is called event-driven pharmacology. PROTAC brings two proteins that do not naturally interact into proximity – POI and an E3 ligase. This induced interaction ultimately leads to polyubiquitination of the POI, followed by its degradation (Figure 1). The basic skeleton of the degrader consists of two ligands that bind to the POI and E3 ligase recruiting moiety, which get connected with a linker molecule. They possess relatively higher molecular weight and polar surface area, typically in the range of 800–1000 Da and ~200 Å², respectively (Neklesa et al., 2017; Pettersson & Crews, 2019).

TPD depletes the targeted protein rapidly (in minutes to hours) in contrast to genetic techniques requiring days to weeks to manipulate mRNA or genomic DNA. TPD allows specific protein variants to be selectively degraded and offers rapid reversibility with superior control compared to genetic engineering. The therapeutic advantage of PROTAC is considered to be a combination of small molecular inhibitors and gene-based approaches. PROTACs are superior in providing a fast and reversible chemical strategy acting on protein directly rather than at the genome level. Gene-based tools usually have a longer cycle, irreversible action, and higher cost, making PROTAC an effective supplement for available genetic tools (Wu et al., 2020).

The main protease (also known as nsp5, 3CLpro, or Mpro) of SARS-CoV-2 is a promising target because of its prominent role in the early stage of viral replication. Once the viral genome enters the host cell, the two open reading frames (ORFs) encoded by this positive-stranded RNA genome undergoes translation into two polyproteins, 1a/1ab (pp1a/pp1ab) by ribosomal frameshifting (Chen et al., 2020). The polyproteins are auto-catalytically processed by two viral proteases, Mpro and papain-like protease (PLpro), into 15 to 16 non-structural proteins (NSPs). The NSPs are necessary for viral RNA synthesis (Gildenhuys, 2020). Due to their indispensable role, both proteases are relevant from a biological perspective. Unlike PLpro, Mpro does not have human homologous protein, making it a prime target for selective therapeutic intervention (Liu et al., 2020). PROTAC mediated degradation of disease-related protein is an exciting domain to be explored because of its advantages over inhibition, which could prove extremely valuable in the antiviral domain. This study focuses on providing insights to help design degrader molecules targeting the Mpro of SARS-CoV-2 using computational approaches.

**Background of the study**

Crews and Deshaies laboratories reported the first PROTAC in 2001, which showed broad therapeutic applicability for
macromolecule degradation and had better pharmacokinetic properties than earlier protein/gene knockdown approaches (Sakamoto et al., 2001). This approach introduced the utilization of UPS to the scientific community for drug discovery, and the domain has been growing exponentially. At present, two degrader molecules developed by Arvinas, Inc. have entered clinical trials. ARV-110, currently in Phase 1/2, is an orally bioavailable androgen receptor (AR) targeted degrader used to treat patients with metastatic castrate-resistant prostate cancer (Neklesa et al., 2019). ARV-471 is currently in Phase 1 clinical trials designed to specifically target advanced or metastatic ER-positive/HER2 negative breast cancer in women (Flanagan et al., 2018).

Out of many advantages of degraders, the ability to overcome drug resistance deserves special attention. Target upregulation is a case that can lead to resistance and gets countered by PROTACs. This effect is because it works in a catalytic manner providing efficacy that is not limited by equilibrium occupancy. This sub-stoichiometric catalytic nature helps achieve the target protein arrest to a level that is not possible with inhibition (Bondeson et al., 2015). This characteristic makes PROTACs useful in cancer and other diseases where specific proteins are involved and overexpressed, which may selectively be knocked down to achieve homeostasis. Many times, point mutations at the active sites result in drug resistance towards inhibitors (Rajendran et al., 2016; Rajendran et al., 2018). A degrader molecule derived from the inhibitor, for which resistance has developed, is proven effective in degrading the target protein. For instance, Bruton’s tyrosine kinase (BTK) degrader MT-802 can degrade the C481S mutant, which is resistant to ibrutinib (Buhimschi et al., 2018). Another example is Telaprevir resistant viral strains of Hepatitis C Virus (HCV) NS3/4A protease that was degraded efficiently by the degrader DGY-08-097 (de Wispelaere et al., 2019).

In some instances, target proteins exhibit multi-domain functions (enzymatic and scaffolding functions), which are challenging to regulate only by small molecule inhibitors. Protein knockdown should elicit better results here as it degrades and abolishes all functions than mere inhibition of one particular binding site. This applicability widens the therapeutic target protein domain since now scaffolding proteins can also be considered. One such example is focal adhesion kinase (FAK), an essential protein in tumor invasion and metastasis, showing kinase-dependent and kinase-independent signaling of which inhibition only hinders the former (Cromm et al., 2018). For instance, STAT3 protein (signal transducer and activator of transcription), which is considered undruggable and has a role in tumor growth and metastasis, has successfully been degraded by PROTAC technology (Zhou et al., 2019). These results are further evidence of a potential role in drug therapy by protein-knockdown even when there is no ligand known to inhibit a particular protein.

Recent advancements in TPD have modified certain parameters of degraders to address specific issues associated with PROTAC technology. PROTACs are catalytic and have the potential to cause uncontrolled degradation that can result in systemic toxicity. Trying to circumvent this barrier by efficiently controlling the degradation, multiple unrelated groups introduced a light-inducible switch on the degraders (Jin et al., 2020; Konde et al., 2020; Liu et al., 2020; Reynders et al., 2020; Xue et al., 2019). Testa et al., successfully demonstrated macrocyclization of PROTACs for an enhanced selectivity between homologous targets (Testa et al., 2020). The in-cell click-formed proteolysis targeting chimeras (CLIP-642TACs) were designed to overcome the limitation of cell permeation, solubility, and other physicochemical properties. This strategy exploits click chemistry to form the linker between respective ligand warheads to produce PROTAC molecules inside the cell (Lebraud et al., 2016).

**PROTACs in the antiviral domain**

In antiviral research, there is a desideratum for developing potent and specific antivirals to circumvent resistance challenges. de Wispelaere et al., provided proof of concept that targeted protein degradation could be the new paradigm for antiviral drug development with superior resistance profiles. They chose Telaprevir, an approved first-generation protease inhibitor, and linked it with CRBN ligand to form a PROTAC molecule that inhibited and induced degradation of the HCV NS3/4A protease (de Wispelaere et al., 2019).

**Available modeling techniques**

Conventional PROTAC design steps include recognizing the solvent-exposed part of the ligands bound to E3 ligase and target protein to determine the anchor atoms (to which the linker molecule is attached) PROTACs with varying linker length is synthesized and tested for their potency in degradation. This process is laborious work that requires enormous resources and time. Structural knowledge of the ternary complex consisting of E3 ligase, PROTAC and the target protein can help design optimal degrader molecules. Utilizing structural information for PROTAC design is highlighted as it has been applied to design selective degraders (Gadd et al., 2017). Computational modeling techniques for PROTAC design are still in their infancy yet able to produce significant results. Utilizing protein-protein docking algorithm for predicting possible interacting poses has given a new life to the computational design of degrader molecules.

In the pioneering work of *in silico* PROTAC design, Drummond et al., introduced four distinct methods and validated them to mimic already published crystal structures of ternary complexes (Drummond & Williams, 2019). Recently, it has been updated with follow-up studies that could correlate the potency of PROTAC with modeling techniques that paved the way for computational studies (Drummond et al., 2020). There are other methodologies revealed recently utilizing the Rosetta modeling suite (Bai et al., 2021; Zaidman et al., 2020), of which PRosettaC provides a server-based tool for ternary model generation. An absolute protocol to design PROTACs computationally is not established yet due to distinct modes of action compared to small molecular inhibitors. Nevertheless, a degrader’s ability to ensure the
formation of a stable ternary complex is regarded as a crucial step for efficient degradation (Gadd et al., 2017).

**PROTAC for COVID-19**

The applicability of PROTAC technology on SARS-CoV-2 has been discussed recently by different groups. One suggestive literature provides detailed requirements to be considered for designing a good degrader for COVID-19 therapy (Martinez-Ortiz & Zhou, 2020). Another group has synthesized reversible covalent degraders, and further screening is being conducted in their laboratory (Liu et al., 2020). Our lab has commenced computational modeling to design PROTACs.

**Computational methodology**

**Sequence and structural analysis**

Global pairwise sequence alignment (PSA) was performed using EMBoss needle (Madeira et al., 2019) with default settings, extracting protein sequences from UniProt (UniProt Consortium, 2019). Protein crystal structures obtained from Protein Data Bank (PDB) were subjected to structure alignment and analysis in Maestro v12.5 Package (Schrödinger, LLC, New York, NY).

**Protein–protein docking**

The protein–protein docking was executed in a Linux system using the Rosetta modeling suite (3.12 release) (Elsevier Leaver-Fay et al., 2011). The protein crystal structures obtained from PDB were prepared with Rosetta Relax protocol incorporating ref2015 forcefield (Alford et al., 2017). Relax refines and minimizes the protein structures to their nearest local minimum and optimizes side-chain conformation, and prepares them to be subjected to use with different Rosetta binaries. General options were used for the refinement. Minimized proteins were manually oriented to roughly face bound-ligands in proximity using Maestro and exported in .pdb format. This structure file was then prepacked for docking to refine the side chains outside the docking interface for proper scoring.

E3 ligase was specified as the receptor and main protease structure as the ligand with their chain IDs in the prepacked structure. The docking was performed to get 500 decoys using default docking protocol with 5 Å translation and 20° rotation perturbation to the second partner before every simulation. Protein with more residues was chosen as the receptor in protein-protein docking for computational efficiency. Randomizing docking partners was omitted because we do not need the protein orientation to change very much. The top 25 results, as per the best interface score since it represents the energy of interaction across the interface, were drawn out of 500 decoys, and the distance between the anchor atoms was measured using Maestro.

**Generation of ternary models**

Linker chains with varying lengths were introduced between two ligands to form heterobifunctional molecules. As input, PDB files of ligand-bound structures, the unique IDs of ligands, and the SMILES of molecules were submitted to the ProsettaC server (Zaidman et al., 2020), which can be accessed through this link https://prosettac.weizmann.ac.il/pacb/steps. The server provides ternary models as output clustered into ranked folders according to the number of members. More details on the working protocol of the server can be found in their literature.

**Molecular dynamic simulations**

One representative structure from the top clusters of generated models for PROTAC P3 was selected for molecular dynamics (MD) simulation to investigate the behavior of the ternary system in solution. The energy-minimized model of CRBN: P3: Mpro was used as a starting point for MD simulation, and simulation was carried out using the Desmond-v6.1 module (Bowers et al., 2006) of the Schrödinger software package (Schrödinger, LLC, New York, NY) and the OPLS3e force field. The system was solvated using the TIP3P (transferable intermolecular potential with 3 points) water model with padding of 10 Å from the edge of the box to any solute atom by the system builder tool of Desmond and neutralized by adding appropriate counter ions. The solvated system was minimized for 2000 steps with all protein and PROTAC atoms restrained to eliminate residual unfavorable interaction between the solute and the solvent, followed by another 2000 steps with unrestrained PROTAC and solvent atoms, and finally followed by another 5000 steps with all atoms free to move. The equilibration phase consisted of an initial 100ps Brownian dynamics simulation at 10 K with restraints on solute heavy atoms in the NVT ensemble, followed by one run of 120ps of MD at the same temperature, restraints, and ensemble. The restrained system was then subjected to two runs of 120ps in the NPT ensemble, first at 10 K and then at 300 K, and the final unrestrained simulation at the final temperature (300 K) for 240ps. Short-range non-bonded interactions were cut-off at 9 Å. A production run of 100 ns MD simulations was carried out under the NPT ensemble. Trajectories were recorded at each 50ps interval to generate 2000 frames. Trajectory analysis was done using simulation interaction diagram tools (SID) in Desmond. A schematic representation of the methodology followed is given (Figure 2).

**Results and discussion**

Telaprevir-based degraders are already reported for HCV NS3/4A protease (e.g., DGY-04-035, DGY-08-097) (de Wispelaere et al., 2019). The idea for this study was conceived when Telaprevir was shown to be bound with Mpro of SARS-CoV-2 in multiple crystal structures available in the Protein data bank (PDB IDs: 6ZRT, 7K6D, 7K6E, 7C7P, and 6XQS). Potency in inhibition of Mpro by Telaprevir is in the micromolar range (Anson et al., 2020; Baker et al., 2021;
Kneller et al., 2020). However, it is safe to assume that the drug can bind to the protease irrespective of its inhibitory potency. Therefore, Telaprevir was picked as the Mpro binding moiety since there are reported PROTACs using the same warhead with which the results can be compared computationally. On the other hand, pomalidomide was selected as the CRBN E3 ligase recruiting part, which is frequently utilized due to the ease of degrader synthesis. Besides, thalidomide ligands are comparatively smaller than other E3 ligase ligands, which finally will result in smaller-sized PROTACs.

**Sequence alignment and structural analysis**

We have assessed the similarity in residues since Telaprevir can bind with both NS3/4A protease of HCV (UniProt: P27958/1027-1657) and Mpro of SARS-CoV-2 (UniProt: P0DTD1/3264-3569). Sequence alignment results obtained were with the identity of 12.8%, similarity of 19.6%, and a total score of 37.5. They are indicating different amino acid residual compositions of the respective proteases. Then structural alignment performed using Maestro showed that the main protease (PDB ID: 6ZRT) has completely different folds compared to HCV NS3/4A protease (PDB ID: 3SV6). Upon inspecting bound structures, Telaprevir was displayed to be buried more profound in the case of Mpro. NS3/4A protease and Mpro show 8 and 7 hydrogen bonds, respectively, with Telaprevir which are interestingly formed with six amino acid residues. One charged (Glu166), four Polar (His41, Cys145, His164, Gln189) and one hydrophobic (Gly143) residue in case of Mpro, one charged (Arg1155), three polar (His1057, Ser1139, Ser1159), and two hydrophobic (Ala1157 and Gly1137) for NS3/4A protease. The similarity observed among these residues is that both have one charged residue in the interaction, though of opposite charges. Non-bonded contacts were 94 in Mpro and 75 in NS3/4A protease, as one can expect from more buried Telaprevir in the former (Figure 3(a)). Furthermore, structural inspection revealed that the solvent-exposed pyrazine ring should be suitable for linker attachment (Figure 3(b)). Moreover, Telaprevir interacts with Mpro residues which are considered the most attractive ones to form hydrogen bonds with ligands (Nguyen et al., 2020). These details strengthened the idea that Telaprevir can bind to the Mpro of SARS-CoV-2 and be used for degrader design.

**Protein–protein docking and linker length estimation**

As the first step, protein-protein docking of Mpro and CRBN E3 ligase was performed to see possible interacting poses between them. PDB ID: 6ZRT was chosen as the main protease structure bound with Telaprevir and PDB ID: 4CI3 as CRBN E3 ligase structure with pomalidomide. The protein-protein docking results are predicted based on shape complementarity and surface charges between the two proteins (Chaudhury et al., 2011). It means the results could be one of the possible interacting poses. We could exploit this and apply it to predict optimal linker length between the anchor.
atoms. From the resulting docking solutions, the top 25 with the best interface score were selected for further investigation (Supplementary Table S1). Rosetta suggests relying to an interface score for protein-protein docking results which has been utilized before (Zaidman et al., 2020).

The distances between the anchor atoms were estimated using Schrödinger Maestro for the above 25 complexes. The distances ranged from 4.7 Å to 38.7 Å (Figure 4). A linker could connect the shortest straight-line distance with as short as 5 to 6 alkyl chains. We were able to generate ternary complexes with such short linkers. Nowak et al., in their study of plasticity in PROTAC-mediated ternary complex, showed estimating linker length using Rosetta protein docking, which helped them in designing PROTAC molecule with high selectivity among homologous proteins (Nowak et al., 2018). They also observed that a short linker might improve selectivity among homologous proteins but decreases potency by high magnitude. For Mpro degradation, a longer linker can be chosen because no homologous host protein is there to cause off-target effects.

Moreover, Rosetta docking runs on Monte Carlo sampling; thus, higher iterations could have provided better results. An extensive study like that could give more insights into linker length. However, for an initial idea, a small number of iterations seems to be sufficient. The protein-protein interactions among the top results varied drastically. This kind of behavior was expected of global docking of proteins. The primary aim of protein-protein docking was merely to estimate linker length rather than predicting interaction poses.

**Ternary model generation**

The PRosettaC protocol was benchmarked against five available CRBN-PROTAC-Target complexes (Zaidman et al., 2020). It was successful in predicting near-native structure in the top generated clusters for four of them. This predictability makes it a reliable tool for ternary model generation. This protocol was deployed in this study to generate ternary models. We wanted to see how the ternary model generation deviates if the same degrader is used among two different target proteins. DGY-04-035 (Figure 5), a Telaprevir-based degrader that is already reported for HCV NS3/4A protease, is used as reference degrader, even though DGY-08-097 was reported as the most potent degrader with

---

**Figure 3.** (A) Telaprevir buried in the catalytic pocket of Mpro. (B) Ligand interaction diagram of telaprevir-main protease complex (6ZRT) with suitable linker attachment point marked with the red arrow.

**Figure 4.** The dotted straight-line (Green) distance measured between two anchor atoms from the protein-protein docked structure. Main protease (Red) with Telaprevir (Cyan) and CRBN E3 ligase (Orange) with Pomalidomide (Pink).
telaprevir warhead. The reason being the latter has a different E3 ligase ligand for which complex structure was not available in PDB that is required as input for PRosettaC. Since Telaprevir can bind to M\textsuperscript{pro} SARS-CoV-2, ternary complex modeling by PRosettaC protocol using the same degrader molecule was performed.

The procedure was repeated for HCV NS3/4A protease, taking PDB ID: 3SV6 as the respective crystal structure. The first observation was the lower number of models generated for M\textsuperscript{pro} (Table 1). In the literature of the protocol, they were able to correlate the number of models generated to the activity of PROTAC in one of the case studies (Zaidman et al., 2020). Since it follows constrained docking, many conformational solutions would be rejected if it appears to have steric clashes. Thus, a high number of results indicate more low energy solutions for PROTAC ternary ensembles. The clustering method compiles generated models within 4 Å RMSD to one cluster and ranks clusters based on the number of models in it. Therefore, the top cluster represents repeating docked poses which could be energetically minimal.

On repeating the procedure, there was only a slight variation in the number of local docking solutions, and the structures in top clusters had similarly docked poses, indicates repeatability of model generation and possible conserved interactions. The PEG linker in DGY-04-035 is of a dimer length. The distance constraints are calculated from the user-provided SMILES of the degrader molecule. The lower number of local docking solutions could be because the linker limits the possible low-energy conformational space for protein-protein docking. As observed in the crystal structure, Telaprevir is somewhat buried deep in the cavity of M\textsuperscript{pro}. It could further limit the conformational space. The next step was to increase the linker length (Figure 5) of the PROTAC and redo the PRosettaC job. As predicted, increasing linker length resulted in a higher number of local docking solutions (Table 2). It is essential to consider that long linkers may result in a greater number of solutions, but this can be attributed to the allowance of larger conformational space, not necessarily indicative of a stable docking solution. But generally, the linker is likely to fold on its own to give additional contacts with the proteins. Therefore, instead of a higher number of solutions generated, the number of solutions in the highest-ranked clusters was scrutinized since it represents repeating low energy models. Another exciting factor was that only models in top clusters rendered by DGY-04-035, P3, and P4 had similar poses, which we could repeat by redoing the ternary model generation procedure. PROTACs P5 and P6 deviated drastically, probably because higher linker length in them allowed vast search space resulting in models of varying poses.

The optimal linker length could be P3 or P4 according to the number of members in the highest-ranked clusters. Moreover, the linkers would allow some flexibility to elicit contact with the best interfaces for stable complex formation. A longer linker should be suitable to achieve that.

When the top clusters developed for these PROTACs were compared, the protein-protein poses were similar, which strengthens the possibility of conserved interactions. We observed that some of the residues have common protein-protein interactions (PPIs) among the generated models. Chain A in the generated model represents the CRBN E3 ligase, and Chain B represents the M\textsuperscript{pro} of SARS-CoV-2. A: His286 with B: Glu527 and A: Glu310 with B: Lys498 were the most common interactions among the models. Various other amino acid residues were also involved, but the interactions tend to shift to adjacent residues among models. Some common residues are A: Ile100, Gly305, Arg306, B: Asn503, Thr530, Asp558 (Supplementary Figure S1). We assumed that some of these interactions would stabilize dynamically by MD simulation. PPIs of generated models were compared, taking PDB ID: 6BOY as the reference, a CRBN-PROTAC-BRD4 ternary complex. We could not find any common interactions among them. However, this could be because protein-protein docking is based on shape complementarity and

| Protein of interest (PDB ID) | Local docking solutions | Low energy models | Members in the largest cluster |
|-----------------------------|--------------------------|-------------------|-------------------|
| 6ZRT                        | 823                      | 151               | 20                |
| 3SV6                        | 38746                    | 7951              | 37                |

Figure 5. 2D representation of PROTAC molecules.
interface charges, which are different for BRD4 and Mpro. Further, Nowak et al., showed that CRBN can interact by a different surface contact yet successfully elicit the degradation of the target protein (Nowak et al., 2018). One can interpret that the degradation signaling is not mediated by specific residual contacts, but rather these PPIs are helpful only in stabilizing the ternary complex. We should consider the fact that PROTAC induced PPIs do not naturally occur; hence it is not obligatory to display the same residue interaction among all proteins, i.e. the surface complementarity can vary with different proteins, as one may notice with another target of CRBN, i.e. CK1α (PDB ID: 5FQD). Lack of PROTAC ternary complex crystal structures of different proteins hinders validation of these objections (CRBN-PROTAC-PROTEIN crystal structures available for BRD4 only). Interestingly the interface area was comparatively higher in the generated model, which should stabilize the interactions to elicit degradation. In the ternary models generated, the ligands conserve the interactions with their respective proteins (Supplementary Figure S2). The effect on these interactions by introducing a linker between them will be explored by MD simulation.

The PRosettaC server generates results as PDB structures in a clustered fashion. Each cluster consists of models that are within 4 Å RMSD. Choosing a model pose from the cluster was subjected to several considerations. CAPRI allows 10 Å RMSD deviation from the experimental structure for protein-protein docking. Hence, a model generated using PRosettaC is more constrained. Secondly, we calculated the MMGBSA value of top generated models using the HawkDock server tool (Weng et al., 2019). The results showed only a slight difference in free energy change between each model, which does not support the idea of choosing a model based on MMGBSA calculation/scores for further evaluation. Free energy calculations are a valuable tool in ligand binding affinity prediction or even for small peptide ligand pose selection. However, to choose one among a series of similar poses in a cluster, these values seem insignificant. Scoring values and energetic values were indifferent among correct and incorrect poses, making us rely on the cluster size to choose the MD simulation model. In this study, we assumed top generated ternary complexes are among the many low energy conformations possible in the solution phase. At present, there are very few publications regarding MD simulation studies of PROTAC ternary complexes. We believe these MD studies could provide a glimpse of how a stable ternary complex is formed and how they behave dynamically. We have taken one model from each of the top 2 largest clusters and performed MD simulation on both to see if the behavior changes drastically between them. We assumed that minor deviations observed between poses in a cluster would stabilize dynamically through MD simulation.

### MD simulation analysis

RMSD change of the protein system indicated that the simulation attained equilibrium after 30 ns. The changes were of the order of 3.0 Å to 4.8 Å. The analysis of the ligand fit on the protein showed an RMSD maintained below 4 Å, which indicates a stable bound ligand in the pockets (Figure 6).

For most of the residues that are part of helices and sheets, the RMSF values were below 3 Å. Other than the N and C terminal of the proteins, residues such as Leu144-Leu147, Pro157-Ser159, Asp204, Asn638, and Gly639 exhibited high fluctuations since they are part of the loop region of the proteins. However, they were not involved as part of the binding interface (Figure 7).

Analysis of protein-ligand contacts throughout the simulation revealed consistently strong hydrogen bond interactions with binding site residues on both warheads. Ligand-protein contacts with more than 30% strength are shown (Figure 8). Trp313 and His311 can be visualized as bonded to the pomalidomide part for almost the entire simulation. Gly504, His525, Glu527, and Thr551 showed the most robust interactions on the telaprevir side. Glu527, a critical residue shown to interact with most of the reported inhibitors, was having multiple hydrogen bonds retained throughout the simulation. Additionally, consistent bonding with the carbonyl oxygen of the linker to His290 was observed. These strong hydrogen bonding contacts reveal that the degrader molecule can form a stable bridge between the two proteins.

Degrader interacted with A: Met41, Pro285, His286, Tyr288, His290, Trp319, Trp333, B: Met410, Pro413, Cys506, Leu528, Pro529, Phe546 and Ala552 residues of protein forming hydrophobic interactions, and with A: Val283, Tyr288, His290, Glu291, Ser308, Glu310, His311, His330, Thr351, Arg352, Ser353, B: His402, Asn503, Gly507, His524, His525, Pro529, and Gln550 formed H-bond-water interactions. There was no ionic interaction detected (Supplementary Figure S3). Another depiction of these contacts in a time-dependent manner is shown in Supplementary Figure S4.

These analyses suggest that the simulations produce stable trajectories of the protein structures, showing potential for further investigation. The radius of gyration (Rg) scores as a measure of the compactness of ligand were high as anticipated owing to the long linker. The value ranged from 7.164 Å to 7.831 Å for the ligand. Solvent accessible surface area (SASA) as the simulation proceeded came lower from >200 Å to <100 Å as the system equilibrated (Supplementary Figure S5), indicating that the degrader molecule got further buried between the two proteins correlating the ligand RMSD profiles. MolSA and PSA plots also supported the stability of the ligand during the simulation process.

| Table 2. Difference in models generated with varying linker length. |
| PROTAC used for model generation | Local docking solutions | Low energy models | Members in the largest cluster |
| DGY-04-035 | 823 | 151 | 20 |
| P3 | 24144 | 8453 | 82 |
| P4 | 18390 | 7600 | 91 |
| P5 | 35517 | 15664 | 37 |
| P6 | 36109 | 12179 | 10 |
Throughout the simulation trajectory, a significant number of intramolecular hydrogen bonds were observed/formed. Analysis of the trajectory revealed that throughout the simulation, the two proteins retained the interacting pose, which is aided by degrader-protein and protein-protein contacts. Few MD stabilized interactions were observed. Backbone hydrogen bonding between A: His330 (chain A = CRBN) and B: Ala552 (chain B = Mpro) was consistent for 98.65% of the time, where A: Gln39 and B: Met410 interact for 70.25% of simulation time. Interestingly these contacts were not shown strongly in the initial structure. A salt bridge between A: Arg306 and B: Asp558 was shown for 98.2% of

Figure 6. RMSD of Protein Cs (Blue) and Ligand (PROTAC) fit on protein (Red).

Figure 7. Protein RMSF graph. The green lines represent residues that have interaction with the ligand. The red region represents alpha-helices, the blue region represents beta-sheets, and the white region represents loops.
the time, conserved from the initial structure. The complex remained stable throughout the simulation, with retention of binding poses of respective warheads of the degrader. The significant implication of the simulation was that the PROTAC molecule could mediate a stable interaction between the two proteins. Both pomalidomide and telaprevir parts maintain specific interaction more than 90% of the simulation time. The linker retained the folded form, which in turn maximized the interaction between the two proteins.

The ligand molecule undergoes conformational strain to preserve a protein-bound conformation. This insight can be grasped from the torsion potential relationship and histogram (Supplementary Figure S6). The graph represents the number of rotatable bonds and freedom of rotation accompanied by the energy consumed. Additionally, it summarizes the simulation trajectory for the rotatable bonds. The linker has rotatable bonds that exhibit a high degree of freedom. Nevertheless, the ternary complex remained stable throughout the simulation. Overall, the system reached a plateau after 30 ns and maintained that state till 100 ns. These results indicate a cooperative ternary complex formation, and PROTAC can form a stable bridge between the two proteins. The stable interactions of degrader molecules with these proteins should keep them nearby for sufficient time. This induced proximity leads to ubiquitination of target protein, i.e., Mpro, and subsequent processes of degradation. MD simulation studies provided indications for forming stable ternary complexes, which could ultimately lead to efficient degradation.

However, there are several intrinsic complications associated with PROTAC and other setbacks that need to be taken into consideration. One such issue is that the degrader itself is a ‘big’ small molecule. Therefore, it often avoids Lipinski’s rules of five. High molecular weight could lead to decreased cell permeability, but the results elicited in the HCV case are encouraging. Other issues involve differential cell expression of E3 ligases, Ubiquitin availability, reach of Lysine residue on the surface of the target protein for possible ubiquitination, the action of deubiquitinase enzymes (DUBs), etc.

The most crucial hurdle associated specifically with SARS-CoV-2 is the native DUBs, PLpro, which is absent in HCV (Freitas et al., 2020). DUBs act inversely to E3 ligases, i.e. they break the ubiquitin from ubiquitinated protein (Mevissten & Komander, 2017). However, the extent of their activity may require investigation. The human cell contains native DUBs as well, which does not significantly affect degradation mediated by PROTAC. There is a paucity of information to explore this area since PROTAC technology has not been significantly applied in antiviral domains. The major limitation of this study is that it is purely dependent on computational tools and evidence-based reasoning. Even though we have valid reasons for performing each particular task, it still needs to be validated experimentally. The impact of changes in ligands and linker compositions can also be experimentally explored. The computational tools and practices followed in the PROTAC design domain are in their early stage and will likely take time to be refined and wholly matured. The distinct mode of action of PROTACs requires a different approach in molecular modeling studies.
The domain is advancing rapidly, and hopefully, more specific tools will be adopted soon.

**Conclusion**

PROTAC could be the future generation of antiviral agents to counter SARS-CoV-2. This study developed strong indications that TPD might be possible for the M\textsuperscript{pro} of SARS-CoV-2, and this paradigm could be employed for advanced therapeutics. This study shows that the linker length selection can be efficiently aided by computational ternary model generation techniques and protein-protein docking. From the study, it is clear that a longer linker results in a higher number of models generated, which previously has been used as a scoring tool to correlate the degradation efficiency of PROTACs. We found that qualitative analysis of the ternary complex is also required to get insights into PROTAC design. MD simulation studies confirmed the predicted behavior of the PROTAC mediated ternary complex. The degrader acted as a bridge between two proteins to elicit favorable stable contacts, which could finally lead to degradation. We believe further studies like steered-MD can be implemented to get insights on the general mechanism of PROTAC action, which we plan to do as a future project. Finally, possible conflicts that may arise with this mode of intervention are given. The primary concern is the DUBs activity of PL\textsuperscript{pro}, which needs to be investigated further.

**Acknowledgments**

The authors thank the Director, National Institute of Pharmaceutical Education and Research (NIPER), SAS. Nagar, Ministry of Chemicals and fertilizers, Gov. of India for providing facility to carry out this work. The authors also thank DST (Dept. of Science and Technology) and CSIR (Council of Scientific and Industrial Research), New Delhi for the computational infrastructure.

**Disclosure statement**

No potential conflict of interest was reported by the authors.

**Author contributions**

M. E. Sobhia (MES) designed the study. All the authors contributed to the computations, analysis, and preparation of the manuscript.

**References**

Adams, G. P., & Weiner, L. M. (2005). Monoclonal antibody therapy of cancer. *Nature Biotechnology*, 23(9), 1147–1157. https://doi.org/10.1038/nbt1137

Alford, R. F., Leaver-Fay, A., Jeliazkov, J. R., O'Meara, M. J., DiMaio, F. P., Park, H., Shapovalov, M. V., Renfrew, P. D., Mulligan, V. K., Kappel, K., Labonte, J. W., Pacella, M. S., Bonneau, R., Bradley, P., Dunbrack, R. L., Das, R., Baker, D., Kuhlman, B., Kortemme, T., & Gray, J. J. (2017). The Rosetta all-atom energy function for macromolecular modeling and design. *Journal of Chemical Theory and Computation*, 13(8), 3031–3048. https://doi.org/10.1021/acs.jctc.7b00125

Anson, B. J., Chapman, M. E., Lendy, E. K., Pshenichnyi, S., Richard, T., Satchell, K. J., & Mesecar, A. D. (2020). Broad-spectrum inhibition of coronavirus main and papain-like proteases by HCV drugs.

Bai, N., Miller, S. A., Andrianov, G. V., Yates, M., Kirubakaran, P., & Karanikolas, J. (2021). Rationalizing PROTAC-mediated ternary complex formation using Rosetta. *Journal of Chemical Information and Modeling*, 61(3), 1368–1382. https://doi.org/10.1021/acs.jcim.0c01451

Baker, J. J., Uhrich, R. L., Kraemer, G. C., Love, J. E., & Kraemer, B. C. (2021). A drug repurposing screen identifies hepatitis C antivirals as inhibitors of the SARS-CoV-2 main protease. *PLoS One*, 16(2), e0245962. https://doi.org/10.1371/journal.pone.0245962

Bhardwaj, V. K., Singh, R., Das, P., & Purohit, R. (2021). Evaluation of acridinedione analogs as potential SARS-CoV-2 main protease inhibitors and their comparison with repurposed anti-viral drugs. *Computers in Biology and Medicine*, 128, 104117. https://doi.org/10.1016/j.compbiomedi.2020.104117.

Bondeson, D. P., Mares, A., Smith, I. E. D., Ko, E., Campos, S., Miah, A. H., Mulholland, K. E., Routly, N., Buckley, D. L., Gustafson, J. L., Zinn, N., Grandi, P., Shimamura, S., Bergamini, G., Faeth-Savitski, M., Bantscheff, M., Cox, G., Gordon, D. A., Willard, R. R., ..., Crews, C. M. (2015). Catalytic in vivo protein knockdown by small-molecule PROTACs. *Nature Chemical Biology*, 11(8), 611–617. https://doi.org/10.1038/nchembio.1858

Bowers, K. J., Chow, D. E., Xu, H., Dror, R. O., Eastwood, M. P., Gregersen, B. A., Klepeis, J. L., Kolossvary, I., Moraes, M. A., Sacerdoti, F. D. (2006). Scalable algorithms for molecular dynamics simulations on commodity clusters. *Proceedings of the 2006 ACM/IEEE Conference on Supercomputing.*

Buhimschi, A. D., Armstrong, H. A., Toure, M., Jaime-Figueroa, S., Chen, T. L., Lehman, A. M., Woyach, J. A., Johnson, A. J., Byrd, J. C., & Crews, C. M. (2018). Targeting the C481S ibrutinib-resistance mutation in Bruton’s tyrosine kinase using PROTAC-mediated degradation. *Biochemistry*, 57(26), 3564–3575. https://doi.org/10.1021/acs.biochem.8b00391

Chaudhury, S., Berondo, M., Weitzner, B. D., Muthu, P., Bergman, H., & Gray, J. J. (2011). Benchmarking and analysis of protein docking performance in Rosetta v3.2. *PLoS One*, 6(8), e22477. https://doi.org/10.1371/journal.pone.0022477.

Chen, Y., Liu, Q., & Guo, D. (2020). Emerging coronaviruses: Genome structure, replication, and pathogenesis. *Journal of Medical Virology*, 92(4), 418–423. https://doi.org/10.1002/jmv.25681

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., & Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science (New York, N.Y).* 339(6121), 819–823. https://doi.org/10.1126/science.1231143

Cromm, P. M., Samarasinha, K. T., Hines, J., & Crews, C. M. (2018). Addressing kinase-independent functions of Fak via PROTAC-mediated degradation. *Journal of the American Chemical Society*, 140(49), 17019–17026. https://doi.org/10.1021/jacs.8b08008

de Zeeuw, M. D., Zhipakeeva, M., Du, G., Donovan, K. A., Zhang, T., Eleuteri, N. A., Yuan, J. C., Kalabathula, J., Nowak, R. P., Fischer, E. S., & Gray, N. S. (2019). Small molecule degraders of the hepatitis C virus protease reduce susceptibility to resistance mutations. *Nature Communications*, 10, 1–11.

Drummond, M. L., Henry, A., Li, H., & Williams, C. I. (2020). Improved accuracy for modeling PROTAC-mediated ternary complex formation and targeted protein degradation via new in silico methodologies. *Journal of Chemical Information and Modeling*, 60(10), 5234–5254. https://doi.org/10.1021/acs.jcim.0c00897

Drummond, M. L., & Williams, C. I. (2019). In silico modeling of PROTAC-mediated ternary complexes: Validation and application. *Journal of Chemical Information and Modeling*, 59(4), 1634–1644. https://doi.org/10.1021/acs.jcim.0c00872

Flanagan, J., Qian, Y., Gough, S., Andreoli, M., Bookbinder, M., Cadelina, G., Bradley, J., Rousseau, E., Chandler, J., & Willard, R. (2018). ARV-471, an oral estrogen receptor PROTAC™ protein degrader for breast cancer. *ARV*, 1000, 3.

Freitas, B. T., Durie, I. A., Murray, J., Longo, J. E., Miller, H. C., Chich, D., Hogan, R. J., Tripp, R. A., & Pegan, S. D. (2020). Characterization and noncovalent inhibition of the deubiquitinase and deISGylase activity of SARS-CoV-2 papain-like protease. *ACS Infectious Diseases*, 6(8), 2099–2109. https://doi.org/10.1021/acsinfecdis.0c0168/
WHO. (2021). Weekly Epidemiological Update – 29 June 2021. Retrieved June 30, from https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19--29-june-2021.

Wishart, D. S., Knox, C., Guo, A. C., Shrivastava, S., Hassanali, M., Stothard, P., Chang, Z., & Woolsey, J. (2006). DrugBank: A comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Research*, 34(Database issue), D668–D672. https://doi.org/10.1093/nar/gkj067.

Wu, R., Wang, L., Kuo, H.-C D., Shannar, A., Peter, R., Chou, P. J., Li, S., Hudlikar, R., Liu, X., Liu, Z., Poiani, G. J., Amorosa, L., Brunetti, L., & Kong, A.-N. (2020). An update on current therapeutic drugs treating COVID-19. *Current Pharmacology Reports*, 6(3), 56–70. https://doi.org/10.1007/s40495-020-00216-7

Xue, G., Wang, K., Zhou, D., Zhong, H., & Pan, Z. (2019). Light-induced protein degradation with photocaged PROTACs. *Journal of the American Chemical Society*, 141(46), 18370–18374. https://doi.org/10.1021/jacs.9b06422.

Zaidman, D., Prilusky, J., & London, N. (2020). PRosettaC: Rosetta based modeling of PROTAC mediated ternary complexes. *Journal of Chemical Information and Modeling*, 60(10), 4894–4903. https://doi.org/10.1021/acs.jcim.0c00589

Zhou, H., Bai, L., Xu, R., Zhao, Y., Chen, J., McEachern, D., Chinnaswamy, K., Wen, B., Dai, L., Kumar, P., Yang, C.-Y., Liu, Z., Wang, M., Liu, L., Meagher, J. L., Yi, H., Sun, D., Stuckey, J. A., & Wang, S. (2019). Structure-based discovery of SD-36 as a potent, selective, and efficacious PROTAC degrader of STAT3 protein. *Journal of Medicinal Chemistry*, 62(24), 11280–11300. https://doi.org/10.1021/acs.jmedchem.9b01530

Zhou, Y., Hou, Y., Shen, J., Huang, Y., Martin, W., & Cheng, F. (2020). Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Discovery*, 6(1), 1–18. https://doi.org/10.1038/s41421-020-0153-3