Supporting Information

Rationalizing PROTAC-mediated Ternary Complex Formation using Rosetta

Nan Bai[1,2], Sven Miller[1], Grigorii Andrianov[1,3], Max Yates[1], Palani Kirubakaran[1], and John Karanicolas[1*]

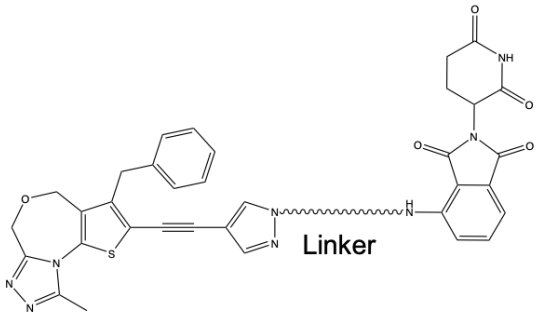
[1] Program in Molecular Therapeutics, Fox Chase Cancer Center, Philadelphia, PA 19111

[2] Department of Molecular Biosciences, University of Kansas, Lawrence, KS 66045

[3] Institute of Fundamental Medicine and Biology, Kazan Federal University, Kazan, Russia, 420008

[*] To whom correspondence should be addressed. E-mail: john.karanicolas@fccc.edu, 215-728-7067

Supporting Tables



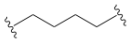
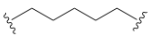
PROTAC	Linker	# conformers generated	# geometrically compatible binding modes	FFC	Cellular degradation of Brd4^{BD1} (from Western blotting)
28		46	86	0.37	Degrades effectively at 10 nM
29		102	493	0.97	Degrades effectively at 1 nM
30		241	2404	2.00	Degrades effectively at 0.1 nM
31		618	429	0.14	Degrades effectively at 0.1 nM

Table S1: Evaluation of the "28-31" PROTAC series. This series uses JQ1 and pomalidomide to recruit Brd4BD1 for degradation by CRBN. Cellular degradation data are from Qin et al, 2018.

Supporting Figures

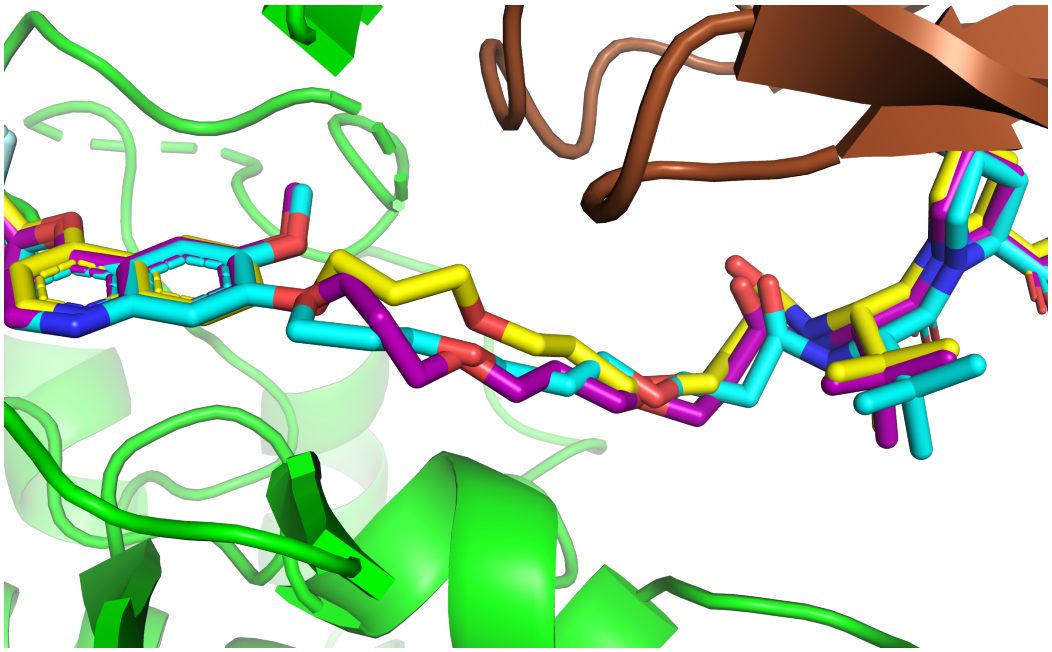


Figure S1: Multiple linker conformers can match to a single docked binding mode, leading to multiple models of the ternary complex from this binding mode. This example is drawn from PROTAC1 bound to c-Met and VHL, the case study presented in Figure 3.

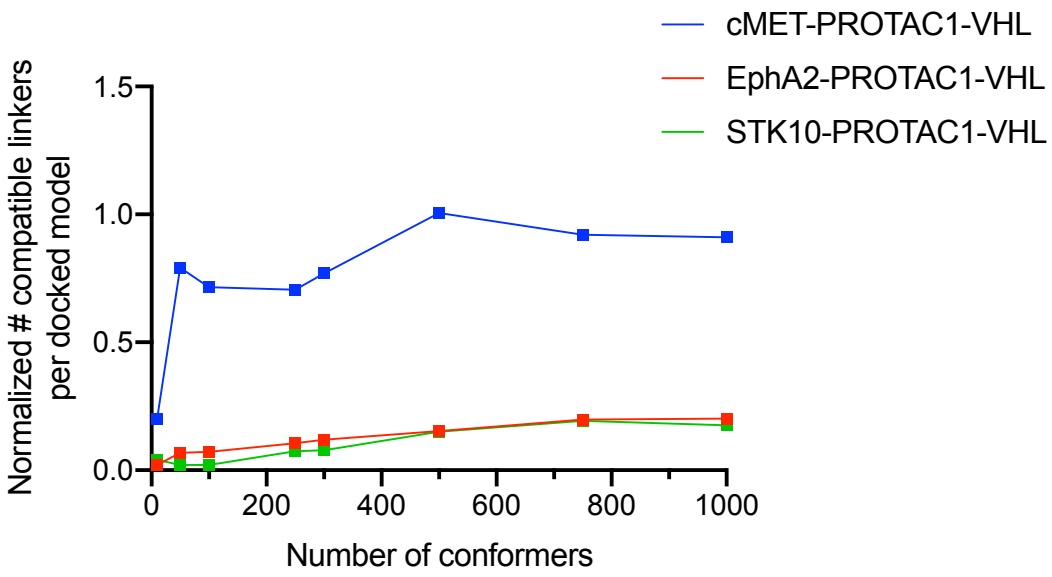


Figure S2: Dependence on the number of linker conformers generated. Using the PROTAC1 case study (from Figure 3), the number of linker conformers was systematically reduced. The number of geometrically-compatible models was calculated using each conformer set. Naturally, more models are generated when more conformers are provided; thus, the number of geometrically-compatible ternary complex models was normalized to the number of conformers. With 1000 conformers, the 5000 docked models in the c-Met case yielded a total of 4552 ternary models, or 0.91 ternary models for each of the input docked models (this is the value plotted above). A large difference is observed between the different kinases (c-Met/EphA2/STK10), as discussed in the context of Figure 3. However, the results are relatively insensitive to whether 500 or 1000 conformers are used.

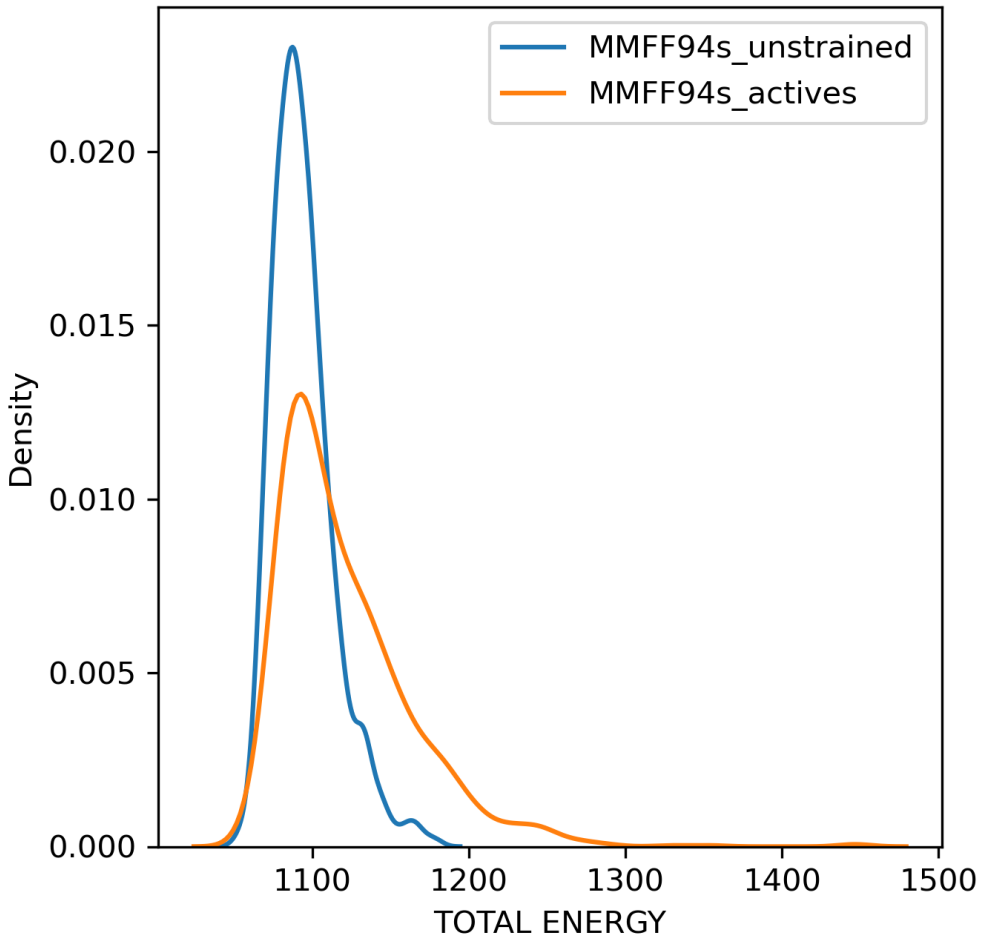
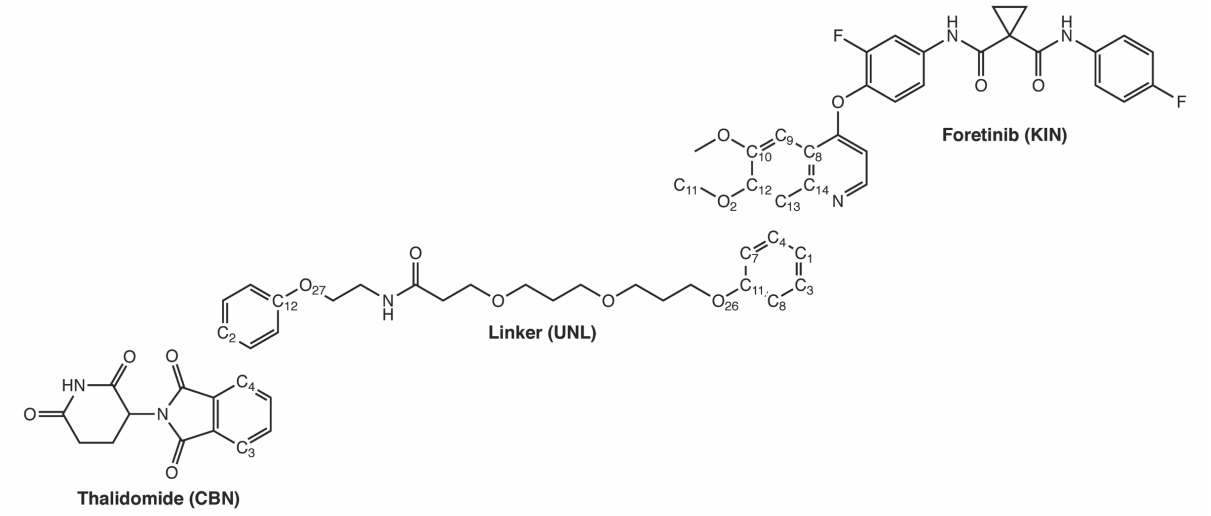


Figure S3: PROTAC conformations in the ternary models have little energetic strain. Using PROTAC2 (a CRBN-recruiting moiety linked to the kinase inhibitor foretinib, shown in **Table 2**), we sought to evaluate how the conformations in the ternary complexes compare energetically to conformations in which the linker is built in energetically optimal conformations. For the latter set, we generated 1000 conformers of PROTAC2 alone, with the only requirement that the foretinib warhead be fixed in the conformation found in its c-Met bound crystal structure. We refer to these as the "unstrained" set of conformers, since the linker has been built from scratch with no functional requirements. We then extracted each of the PROTAC2 conformations from each of the 1168 models of the c-Met / PROTAC2 / CRBN ternary complexes, and called these the "active" set. To evaluate these different conformations' internal energy, we scored the members of each set using the MMFF94s force field, in the absence of any protein partner. While a small number of "active" conformations score slightly worse than the corresponding "unstrained" set, the two distributions are largely overlapping: this suggests that the refined ternary complexes do not build the PROTAC into strained conformations.



Decoy Atom List:	Linker Atom List:	Linker Atom Delete List:	Decoy Atom Delete List:
CBN C3 ⟷	UNL C2	UNL C2	KIN C11
CBN C4 ⟷	UNL C12	UNL C12	
KIN O2 ⟷	UNL O26	UNL O26	
KIN C12 ⟷	UNL C11	UNL C11	
KIN C13 C10 C14 C9 C8 ⟷	UNL C8 C7 C3 C4 C1	UNL C8 C7 C3 C4 C1	

Figure S4: Defining shared warhead/linker atoms. This example uses PROTAC2, from the case study in Figure 3. Each of the atom lists correspond directly to the input files provided to the Python script used for building ternary models. The leftmost two files provide the correspondence between atoms on the warhead and the matched atom on the linker (atoms are listed in the same order, to provide this correspondence). Atoms in these two lists are aligned in the RMSD calculation. Importantly, atoms that do not interact with the protein should be sampled as part of the linker wherever possible: for example, foretinib's methoxy carbon (C_{11}) is simply deleted and replaced with the corresponding linker atom (it is included on the "decoy atom delete list" to achieve this). Conversely, atoms that engage the protein should be held at their original locations from the warhead, rather than replaced with those of the linker: for this reason, linker atoms that overlay with the foretinib ring are not used in building the ternary complex (these atoms are included on the "linker atom delete list" to achieve this). Further instructions on the use of the Python script for building ternary models is included in the README alongside the code itself.