Supplementary Information for:

# LLM Drug Discovery Challenge: A Contest as a Feasibility Study on the Utilization of Large Language Models in Medicinal Chemistry

**Kusuri Murakumo**
The University of Tokyo
SHaLX Inc.
souyakuchan@shalx.co.jp

**Naruki Yoshikawa**
University of Toronto

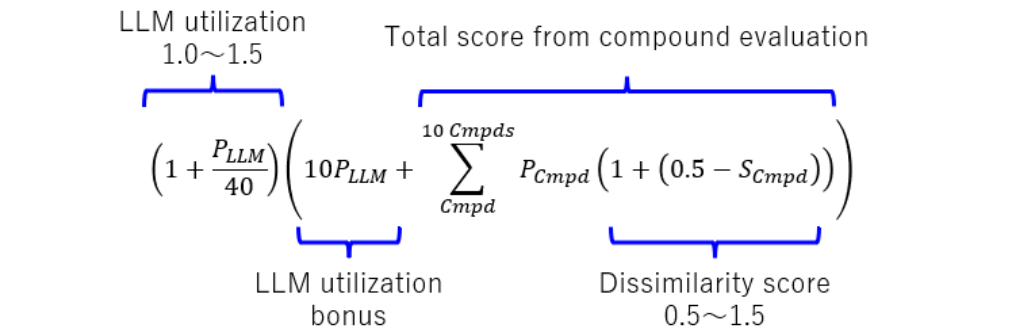
**Kentaro Rikimaru**
Exawizards Inc.

**Shogo Nakamura, Kairi Furui, Takamasa Suzuki, Masahito Ohue**
Tokyo Institute of Technology

**Hioyuki Yamasaki**
Cykinso, Inc.

**Yuki Nishigaya**
AgroDesign Studios Co. Ltd.

**Yuzo Takagi**
iSiP, Inc.




$P_{LLM}$: Sum of LLM utilization scores awarded by each judge on a five-point scale.

$P_{Cmpd}$: Points acquired in individual compound evaluation.

$S_{Cmpd}$: Similarity to known active compounds.

※ The value was obtained by summing the Tanimoto coefficients of the entire molecular structure's Morgan Fingerprints (Morgan FP) and the Murcko scaffold's Morgan Fingerprints, and then dividing by two.

Fig. S1: The overall scoring formulation in the LLM Drug Discovery Challenge.

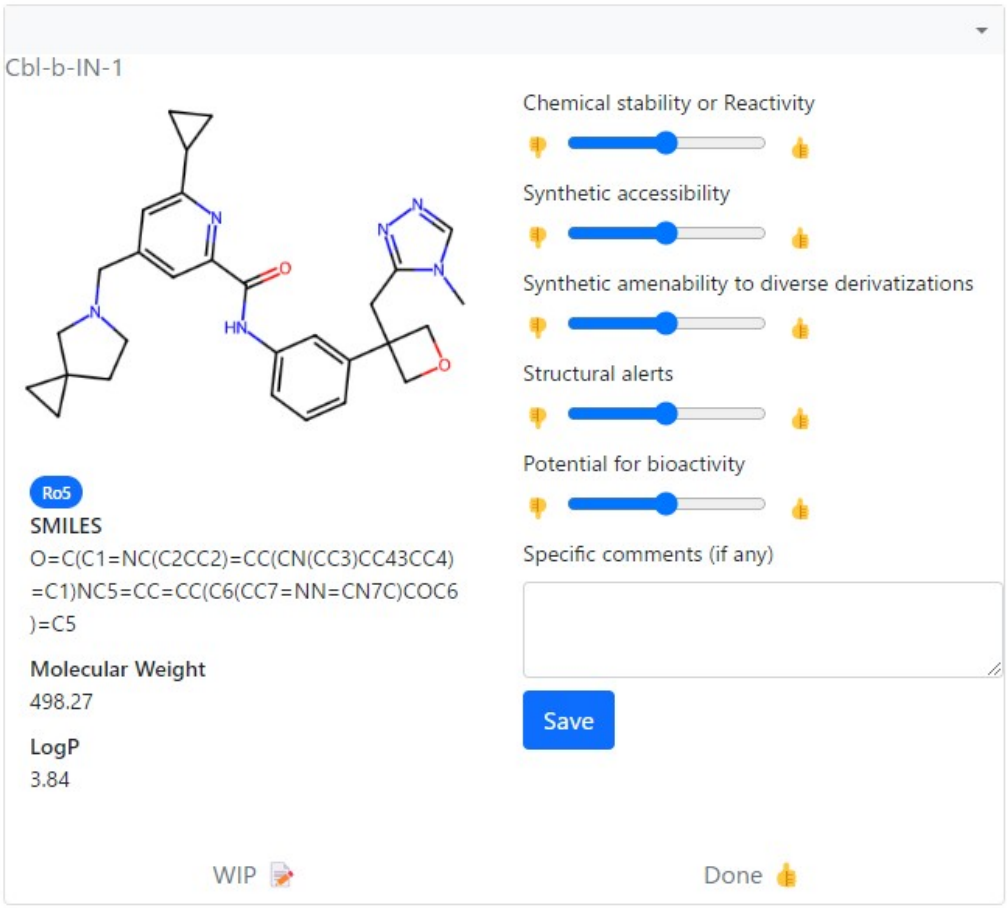



Fig. S2: A web application for the scoring process. Available at wisdom.chemical.space [a].

[a] https://doi.org/10.26434/chemrxiv-2021-9nhm1-v2

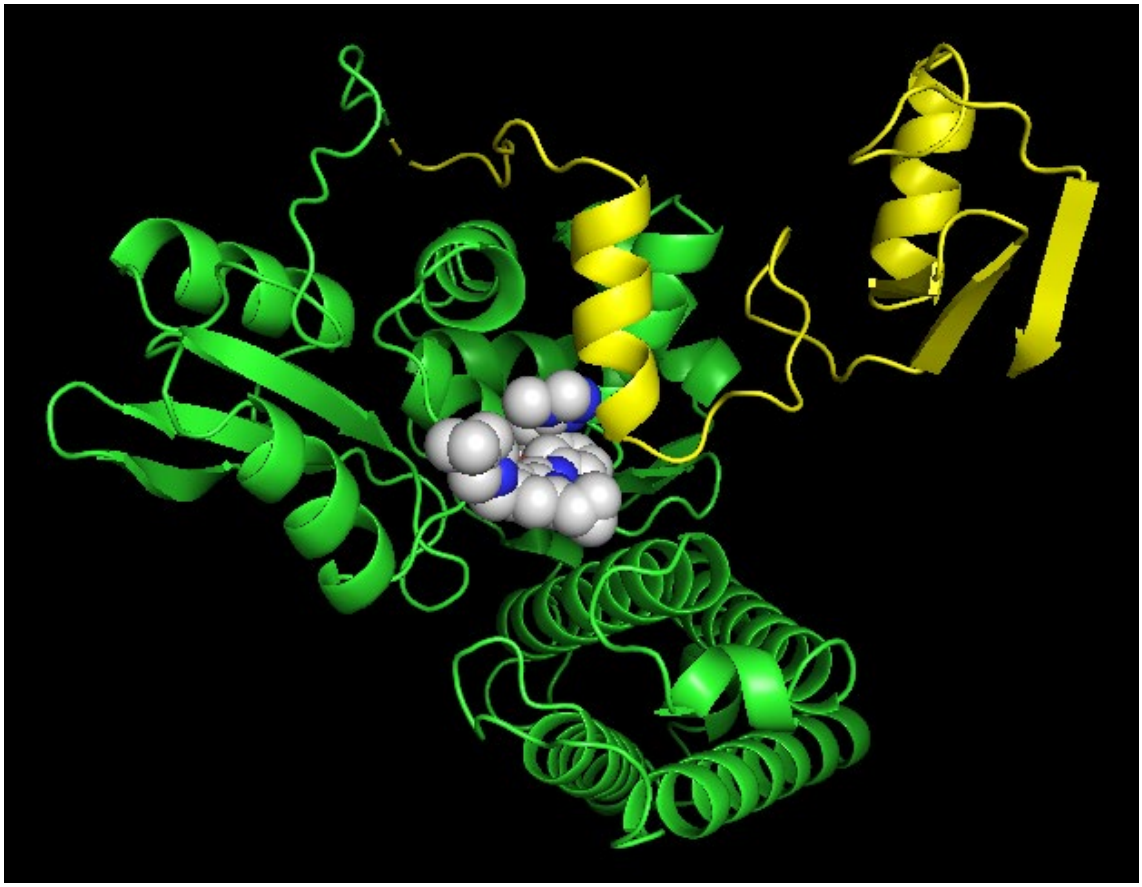


Fig. S3: The crystal structure of tyrosine kinase binding (TKB) domain of E3 ubiquitin-protein ligase CBL-B (Casitas B-lineage lymphoma proto-oncogene b) with its inhibitor which works as an intramolecular glue molecule. This complex structure was experimentally determined by courtesy of AgroDesign Studios Co., Ltd. after the LLM Drug Discovery Challenge.

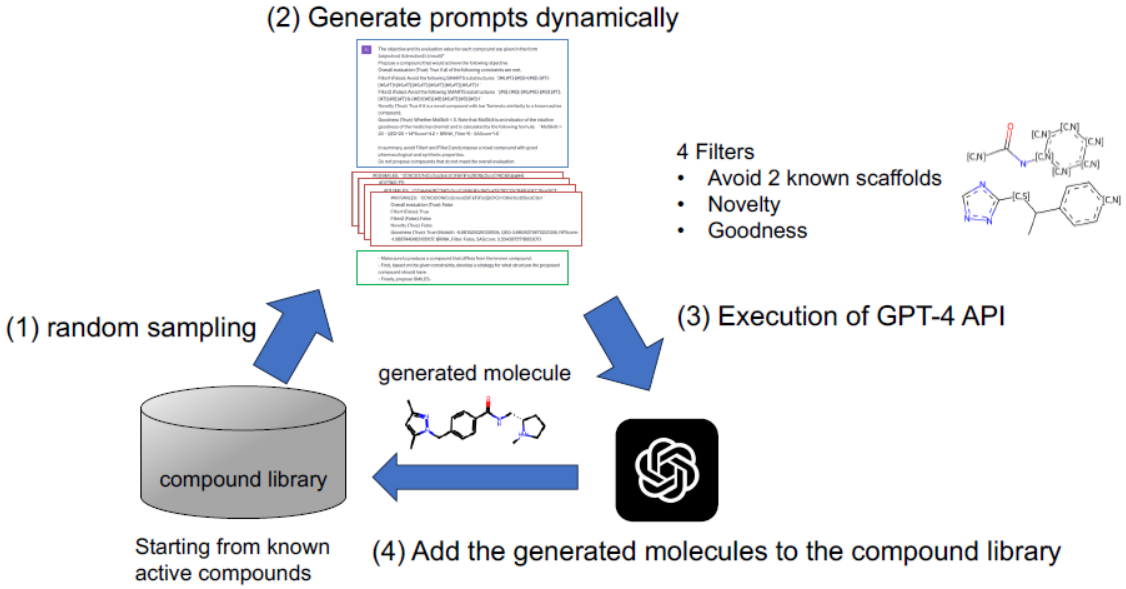



Fig. S4: Summary of the methodology employed by the second-place winner.

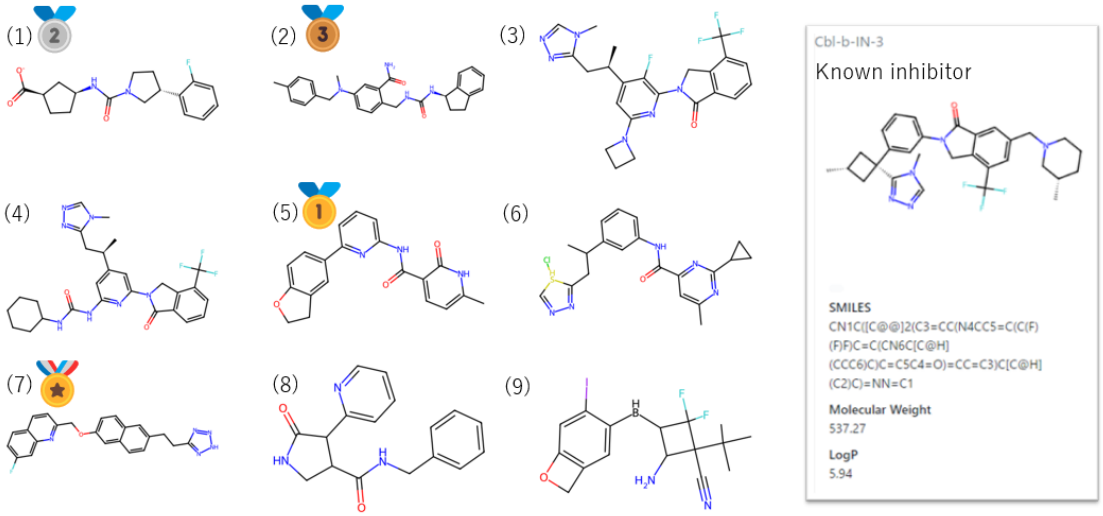



Fig. S5: Representative compounds from each participant. The numbers in parentheses correspond to the Player number in Table 1. The compound on the right is the most active among the known inhibitors.