Improving Stereochemical Limitations in Protein-Ligand Complex Structure Prediction

Ryuichiro Ishitani*[a,b], Yoshitaka Moriwaki[a,c]

a. Department of Computational Drug Discovery and Design, Medical Research Laboratory, Institute of Integrated Research, Institute of Science Tokyo, 1-5-45 Yushima, Bunkyo-ku, Tokyo 113-8510, Japan.

b. Department of Biological Sciences, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan.

c. Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan.

* Corresponding author.

Abstract

AlphaFold3 has revolutionized biology by enabling the prediction of protein complexes with various biomolecules, including small molecular ligands. However, the ligand structures predicted by the model often contain stereochemical errors. In this study, our comprehensive evaluation of AlphaFold3 and its clone model, Boltz-1, demonstrated significant limitations in ligand stereochemistry of their predicted structures, including chirality, bond, and angle geometries. To address the issue, we developed the restraint-guided inference method that applies stereochemical restraints during the reverse diffusion process. This approach perfectly reproduces the chirality specified in the input chemical structure and improves bond and angle geometries of ligands, while maintaining comparable performance in binding pose prediction. Our method provides a practical solution to the stereochemical errors in predicting protein-ligand complexes, thereby enhancing applications of structure prediction in structural biology and drug discovery.

Introduction

The field of protein structure prediction has experienced revolutionary advancements in recent years. Notably, AlphaFold2, introduced by DeepMind in 2021, demonstrated unprecedented accuracy in predicting protein structures[1]. Following this, AlphaFold3 has expanded its capabilities beyond traditional protein structure prediction to include the structural determination of heterogeneous complexes involving RNA, DNA, glycans, and small molecular ligands[2]. These remarkable contributions have received significant recognition within the scientific community and have been awarded the 2024 Nobel Prize in Chemistry. Several derivative models, such as Boltz-1[3], Protenix[4], Chai-1[5], and HelixFold3[6], have also been developed.

AlphaFold3 employs a diffusion model approach, allowing it to predict structures for various biological molecules. In particular, concerning the prediction of protein-ligand complexes, the AlphaFold3 paper claims that it can achieve greater accuracy than conventional molecular docking methods[2]. However, several issues remain to be addressed before it can be used as a substitute for conventional molecular docking methods. For example, it has been reported that the predicted ligand structure sometimes contains stereochemistry violations[2], including errors with chirality, despite the correct reference structure given to the input features. These challenges could pose significant limitations for applications in biological research, including drug design and biosynthesis.

In this study, we first evaluate the structure prediction of protein-ligand complexes by AlphaFold3[2] and its clone Boltz-1[3], mainly focusing on the stereochemistry of ligands. Our analysis reveals that both models often failed to reproduce the local conformations and chirality of ligands specified by the input. Then, we propose a novel method to

address the issues of chirality and local conformation prediction without retraining the model. By applying this approach, we achieve a 100% success rate in chirality reproduction. This research elucidates the limitations of structure prediction models and provides a practical solution to overcome these problems.

Methods

Benchmark dataset

We evaluated the accuracy of protein-ligand complex structure prediction using AlphaFold3[2] and its clone model, Boltz-1[3]. Both prediction models utilize a "reference conformer" as one of their input features. When the chemical structure of the ligand is provided in SMILES notation or Chemical Component Dictionary (CCD) code, its three-dimensional structure (conformer) is generated using ETKDGv3 algorithm[7] implemented in RDKit and then used as the reference conformer. This study systematically assessed how accurately the stereochemistry—specifically the chiralities, bond lengths, and bond angles—of the reference conformer is reproduced in the predicted structures.

To evaluate the stereochemistry of the ligands in the predicted structures, we utilized entries extracted from the PLINDER[8] and PDBBind2020 dataset[9]. The criteria for selecting entries to construct the evaluation dataset were as follows:

1. Entries where proteins do not form multimers.
2. Entries containing one ligand that is not a polymer and thus can be represented by a single CCD code.
3. Entries with a ligand containing at least one chiral center at an sp^3-hybridized carbon atom.
4. Entries where the non-hydrogen atoms of the ligand consist solely of the elements C, N, O, F, P, S, Cl, Br, and I.
5. Entries containing a ligand whose reference conformer can be generated by RDKit.

For the PLINDER dataset, we employed clustering based on a Tanimoto similarity threshold of 0.95 to exclude identical compounds. By applying these selection criteria,

we extracted 6,600 and 4,500 entries from the PLINDER and PDBBind2020 datasets, respectively. The PLINDER dataset was further divided into "Before" and "After" based on whether the PDB entry's release date is before or after the cut-off date of 2021-09-30. Thus, the "Before" split comprises structures included in the training data for AlphaFold3 and Boltz-1, while the "After" split serves as a test set comprising structures not used during model training. Although the PDBBind-based dataset shares many entries with PLINDER's "Before" split, it also includes distinct compounds that are absent in the PLINDER dataset.

Evaluation metrics

We created RDKit Mol objects from the SMILES notation found in the CCD and assigned conformers, including chirality, based on the three-dimensional structures of the compounds in the respective dataset. Using these RDKit Mol objects, we enumerated chiral centers and examined the R/S chirality of each atom according to the Cahn–Ingold–Prelog (CIP) priority rules implemented in RDKit[10]. We then assessed the chirality of the corresponding atoms in the predicted structures and evaluated their consistency. If even one atom exhibited a different chirality, we recorded it as inconsistent and evaluated the fraction of consistent molecules against the total entries. To simplify the problem, we focused solely on the chirality of sp^3 carbon atoms, which are commonly found in biomolecules.

We further evaluated the deviation of the predicted structures from the reference conformer in terms of their bond lengths and angles. Specifically, we calculated the root mean square of the deviations (RMSDs) from the ideal bond lengths ($\hat{r}_i$) and angles ($\hat{\theta}_i$)

of the reference conformers generated by ETKDGv3 algorithm of RDKit[7].

$$\text{BondRMSD} = \sqrt{\frac{1}{N_b}\sum_{i}^{N_b}(r_i - \hat{r}_i)^2}$$

$$\text{AngleRMSD} = \sqrt{\frac{1}{N_a}\sum_{i}^{N_a}\left(\theta_i - \hat{\theta}_i\right)^2}$$

To assess the accuracy of ligand binding site predictions, we performed a least-squares superposition of protein residues within 10 Å of the ligand binding site and calculated the RMSD of the corresponding ligand atoms, as described in the AlphaFold3 paper[2].

Improvement of geometry through restraints

To overcome the stereochemistry violation problems described in the AlphaFold3 paper[2], we developed a novel approach in this study. Instead of modifying or retraining the neural network model, we attempted to alter the diffusion model's inference process to utilize chirality information, bond lengths, and bond angles as restraints. Our method was inspired by the symmetry-guided generation approach used in RFDiffusion[11]. In RFDiffusion, symmetric multimer protein structures are generated by applying hard constraints to protein Cα atoms at each step to enforce the desired symmetry. Similarly, we adopted an approach that applies restraints related to compound stereochemistry to the structures at each step of the reverse diffusion process. This method aims to accurately predict compound stereochemistry without requiring model retraining.

The diffusion model used in AlphaFold3 is the EDM[12]. In each inference step i of

EDM, the next step sample $\boldsymbol{x}_{i+1}$ is calculated using the difference $\boldsymbol{d}_i$ computed from the output of the denoiser function D_θ and the current step sample $\hat{\boldsymbol{x}}_i$:

$$\hat{t}_i = t_i + \gamma_i t_i$$

$$\hat{\boldsymbol{x}}_i = \boldsymbol{x}_i + \lambda\sqrt{\hat{t}_i^2 - t_i^2} \cdot \mathcal{N}(\mathbf{0}, \mathbf{I}_3)$$

$$\boldsymbol{y}_i = D(\hat{\boldsymbol{x}}_i; \hat{t}_i)$$

$$\boldsymbol{x}_{i+1} = \hat{\boldsymbol{x}}_i + \eta(t_{i+1} - \hat{t}_i)\,(\hat{\boldsymbol{x}}_i - \boldsymbol{y}_i)/\hat{t}_i$$

where, γ_i, λ, and η are hyperparameters to control the stochasticity levels.

To achieve denoising based on the training dataset, D is implemented through a neural network D_θ with a parameter set θ. This $\boldsymbol{x}_{i+1}$ is influenced by the noise scale and exhibits a distribution significantly different from the actual coordinate scale. This discrepancy makes applying restraints directly to $\boldsymbol{x}_{i+1}$ impractical. In contrast, it was shown that the optimal denoising function D, which analytically minimizes the training objective, converges toward the mean of the dataset as the noise level increases. Conversely, as the noise level decreases, it approaches the ground truth of the dataset (see Equation B.3, eq. 57 in ref[12]). Consequently, the output of D_θ is expected to remain consistent with the scale of the samples (i.e., experimental structures) present in the training dataset. Therefore, we implemented a method where stereochemical restraints based on the input reference conformer are applied to $\boldsymbol{y}_i$ to obtain $\tilde{\boldsymbol{y}}_i$, and then the difference between $\hat{\boldsymbol{x}}_i$ and $\tilde{\boldsymbol{y}}_i$ is used for calculating $\boldsymbol{x}_{i+1}$ for the subsequent step. These constrained reverse diffusion steps are applied when the condition $t_{i+1} < \sigma_{\text{start}}$ is satisfied, where σ_{start} is an input parameter.

Implementation of the restraints

In this study, we implemented stereochemical restraints targeting chirality of sp^3 carbon atoms, bond lengths, and bond angles. These parameters typically do not deviate significantly from their ideal values calculated from the chemical structures when compounds are bound to the target protein molecules. In contrast, parameters such as dihedral angles can undergo substantial changes during protein-ligand binding processes and are therefore excluded from our restraint targets.

To maintain the correct chirality, we constrained the chiral volume to its ideal value. The chiral volume ($V_i(j,k,l)$) for the chiral center atom (i) was calculated by the following formula:

$$V_i(j,k,l) = \left(\vec{\boldsymbol{x}}_i - \vec{\boldsymbol{x}}_j\right) \cdot \left((\vec{\boldsymbol{x}}_i - \vec{\boldsymbol{x}}_k) \times (\vec{\boldsymbol{x}}_i - \vec{\boldsymbol{x}}_l)\right)$$

where the atoms j, k, l are adjucent atoms directly connected to atom i by covalent bonds. Then, the chirality restraints are achieved by minimizing the following term:

$$L_{\text{chiral}}(\{\vec{\boldsymbol{x}}\}) = \sum_{i \in \text{chiral centers}} \left(V_i - V_i^{\text{ref}}\right)^2$$

where V_i represents the calculated chiral volume of the chiral center i, and V_i^{ref} is the reference value obtained from the input reference conformer. It is important to maintain the same order of j, k, and l in the above formula for both the current structure (V_i) and the reference conformer (V_i^{ref}) to preserve the same chirality.

For constraining bond lengths and bond angles within the compound, we minimized the following terms:

$$L_{\text{bond}}(\{\vec{\boldsymbol{x}}\}) = \sum_{i,j \in \text{bonds}} \left(d_{ij} - d_{ij}^{\text{ref}}\right)^2$$

$$L_{\text{angle}}(\{\vec{\boldsymbol{x}}\}) = \sum_{i,j,k \in \text{angles}} \left(\theta_{ijk} - \theta_{ijk}^{\text{ref}}\right)^2$$

where d_{ij} denotes the bond length between atoms i and j, and θ_{ijk} represents the bond angle formed by atoms i, j, and k. The reference values d_{ij}^{ref} and $\theta_{ijk}^{\text{ref}}$ were derived from the input reference conformer.

The final restraint implementation was achieved by minimizing a weighted sum of the individual terms:

$$L_{\text{total}}(\{\vec{\boldsymbol{x}}\}) = w_{\text{chiral}} \cdot L_{\text{chiral}} + w_{\text{geom}} \cdot (L_{\text{bond}} + L_{\text{chiral}})$$

$$\tilde{\boldsymbol{y}}_i = \underset{\vec{\boldsymbol{x}} \in \mathcal{N}(\boldsymbol{y}_i)}{\operatorname{argmin}} L_{\text{total}}(\{\vec{\boldsymbol{x}}\})$$

where w_{chiral} and w_{geom} are weighting coefficients for each corresponding term. We implemented this method in Boltz-1[3], a PyTorch-based clone of AlphaFold3. The minimization was performed using the conjugate gradient method implemented in the SciPy optimization library[13].

Multiple Sequence Alignment and Structure Prediction

Multiple sequence alignment (MSA) is a crucial component of the structure prediction pipeline of AlphaFold3 and Boltz-1, as evolutionary information significantly influences their structure prediction accuracy. In this study, we used the same pre-computed MSAs for all approaches, AlphaFold3, Boltz-1, and Boltz-1 with restraint-guided inference, to ensure fair comparison conditions. The MSAs were calculated using MMSeqs2 v15[14] through colabfold_search implemented in LocalColabFold[15]. MMSeqs2 provides sufficient MSAs for most proteins while being significantly faster than the AlphaFold3's pipeline that uses HMMER3[16,17]. For AlphaFold3 predictions, we used the default input parameters. For Boltz-1, to maintain consistency with AlphaFold3 conditions, we performed predictions with settings of recycle=10 and nsamples=5. All

calculations were performed using NVIDIA GeForce RTX 4090 and H100 GPUs.

Results

Benchmark of protein-ligand complex structure prediction

In this study, we evaluated the protein-ligand complex prediction capabilities of AlphaFold3 (AF3)[2] and Boltz-1[3] using datasets derived from PLINDER[8] and PDBbind2020[9]. Regarding the protein components of the predicted structures, both models exhibited excellent performance in structure prediction. For the PLINDER Before split, the Protein RMSD was ~0.6 Å (Table 1), and similar values of ~0.6 Å were obtained for the PDBbind dataset (Table S1). Even for the After split, which contains entries unseen by the models during training, both models achieved favorable performance with Protein RMSD values of ~1.0 Å (Table 1, Figure 1A). Although performance declined in the After split compared to the Before split, both models consistently demonstrated excellent performance in protein structure prediction. In contrast, different trends were observed for the ligand structure prediction. For the PLINDER Before split, which consists of entries seen by the models, the Ligand RMSD was 3–4 Å (Figure 1B). Similarly, for the PDBbind dataset, also composed of entries seen by the models, the Ligand RMSD showed slightly lower RMSD values of ~2.0 Å (Table S1). However, for the PLINDER After split, consisting of entries unseen by the models, Ligand RMSD deteriorated significantly, exceeding 7 Å (Figure 1B). These results are consistent with a previous report, which showed overfitting tendencies in AF3's ligand binding pose predictions[18].

Regarding the stereochemistry of the predicted ligand structures, we found an issue in reproducing the chirality of input molecules (Table 1, Figure 1C, Table S1). A

substantial number of entries exhibited changes in chirality from the input structure. While the original AF3 paper reported error rates of only a few percent[2], our dataset, which specifically focuses on chiral molecules, showed remarkably high error rates of 30–40%. One possible reason for this stereochemistry issue may stem from the ligand atom names used to create the input features. The atom names differ between the inputs generated using CCD code and those in SMILES notation. As a result, the input features can vary even for identical ligands, which might impact performance. To explore this hypothesis, we also compared the prediction results using CCD codes and found that chirality reproduction improved slightly (from 70.77% to 76.50% in PLINDER dataset). This aligns with the fact that the atom names from CCD are used during model training. However, a significant number of ligands still exhibited stereochemical errors.

A comparison of the results from AF3 and Boltz-1 revealed that AF3 showed superior performance in chirality reproduction, Bond, Angle, and Ligand RMSDs (Table 1, Figure 1). Overall, while AF3 generates higher-quality predicted structures than Boltz-1, both models still face challenges regarding ligand stereochemistry.

Structure prediction using the restraint-guided inference

We also evaluated the prediction accuracy of Boltz-1 with the restraint-guided inference proposed in this study. This inference method involves multiple hyperparameters (σ_{start}, w_{chiral}, and w_{geom}), so we assessed the model's performance under several conditions (Table 1). Specifically, we conducted experiments where restraints were applied at the final step ($\sigma_{\mathrm{start}} = 0.005$; Boltz R1) and during the last 55 steps ($\sigma_{\mathrm{start}} = 1$; Boltz R) of the 200-step reverse diffusion process. Furthermore, we

compared experiments that applied restraints only to the chiral volume ($w_{\text{chiral}} = 1$ and $w_{\text{geom}} = 0$; Boltz Rc) against those that applied restraints to all parameters: chiral volume, bond length, and bond angle ($w_{\text{chiral}} = w_{\text{geom}} = 1$; Boltz R).

The results showed that the chirality of the input molecule was reproduced 100% across all conditions with restraints except for Boltz R1, in which a small fraction of the chirality inversion was observed (Table 1, Figure 1C). These findings emphasize the importance of applying restraints iteratively within the reverse diffusion process. Additionally, results from Boltz Rc, which applied restraints solely on the chiral volumes, demonstrated that accurate chirality could be achieved by constraining only the chiral volume (Table 1, Figure 1C).

We then evaluated the Bond and Angle RMSD values for each prediction result. Under the Boltz R1 conditions, these values became worse compared to those from the model without restraints (Table 1, Figures 1D, 1E, 2A, 2B). This finding is consistent with the reproducibility of chirality, further showing the importance of applying multiple restraint steps. In contrast, under Boltz Rc conditions where restraints were applied only to the chiral volumes, the Bond and Angle RMSD values were comparable to those from the model without restraints (Table 1, Figures 1D, 1E, 2A, 2B). Thus, applying restraints on the chiral volumes does not adversely affect the bond length and angle geometries. One particularly notable finding was that under Boltz R conditions, where restraints were applied to all parameters (i.e., chiral volume, bond length, and angle), the Bond and Angle RMSD values showed significant improvement. They exceeded the values achieved by AF3 (Table 1, Figures 1D, 1E, 2A, 2B). This result again underscores that our method can accurately reproduce the stereochemical properties of the input reference conformers.

Furthermore, we evaluated steric clashes between predicted protein and ligand

structures by calculating the minimum distance (d_{min}) between ligand atoms and protein atoms. We here defined a serious steric clash as the case where the d_{min} value fell below 2.0 Å, and evaluated the proportion of entries with such serious clashes over all entries. In Boltz R1, where restraints were applied only to the final step of the reverse diffusion process, the highest percentage of serious steric clashes were observed (Table 1, Figure 1F). In contrast, in Boltz R and Boltz Rc, no significant increase in steric clashes was observed compared to Boltz-1 (Table 1, Figure 1F). These results show that multi-step restraints throughout the reverse diffusion process are necessary to prevent steric clashes.

Case studies for chirality reproducibility of ligands

Next, we explored specific instances where the chirality of ligands differed from that in the input reference conformer. In the angiotensin-converting enzyme and inhibitor complex (Figure 3) (PDB ID: 4BXK)[19], the chirality of one carbon atom was predicted to be different from that in the reference conformer (Figure 3C, D). We also analyzed the case where the ligand was specified by the CCD code. The resulting output structure also exhibited a different chirality from that in the input, exemplifying that the reproducibility of chirality was not improved using the atom name assignments derived from CCD (Figure 3E). In contrast, the Boltz R result with restraint-guided inference produced a structure with the correct chirality (Figure 3F). A comparison of the predicted structures showed that the Boltz-R result superposes better with the experimental structure than those by Boltz-1 or AF3 (Figure 3C, D, F). However, when measured by Ligand RMSD, AF3 and Boltz R exhibited similar values of 0.38 Å and 0.39 Å, respectively. The changes in the atomic positions due to the chirality change are not substantial, which are averaged

out when evaluating RMSD. This result may explain why there were no significant changes in the distribution of Ligand RMSD despite improvements in chirality by the restraint-guided inference (Figure 2C).

We also investigated the specific example of Cα carbon atom chirality in amino acid ligands. In the case of the glutamate transporter and D-aspartate complex (Figure 4) (PDB ID: 6R7R)[20], both AF3 and Boltz-1 predicted conformations corresponding to the L-form amino acid (Figure 4C, D). Using CCD code to conform to the atom name nomenclature of amino acids (*i.e.*, CA, CB, etc.) did not improve the results (Figure 4E). In contrast, with the restraint-guided inference, Boltz-R predicted the structure with the correct chirality (Figure 4F). Similar results were also observed in other complex structures with D-form amino acids. These findings demonstrate that when the ligands resemble the monomers of biopolymers, such as L-amino acids in proteins or riboses in nucleic acids, AF3 and Boltz-1 tend to alter their chirality to match that of these biopolymers. While the input atom embeddings of AF3 include information on the atom names, elements, and formal charges, they lack specific information about the chiral centers. Consequently, the model cannot learn the relationship between the atom representations and the arrangement of substituents around the sp^3-carbon atoms. Further examples demonstrating the chirality problem are shown in Figures S1–S4.

Case studies for bond and angle geometries of ligands

Next, we investigated specific cases where the bond and angle RMSDs of the ligand structures predicted by AF3 and Boltz-1 showed significant deterioration. In these cases, we observed several inconsistencies between the predicted ligand conformers and

the reference conformers, including:

1. Structural inconsistencies in aromatic and non-aromatic rings
2. Hybridization state inconsistencies in non-cyclic structures

For the first case, we examined the ectoine-binding protein UehA complexed with ectoine (Figure 5) (PDB ID: 3FXB)[21]. In the structures predicted by AF3 and Boltz-1, the ligand's ring structure was incorrectly modeled as planar, while the reference conformer contains a non-aromatic saturated ring (Figure 5). This error likely occurs because the chemical structure of the ring resembles that of pyrimidine nucleobase, which is commonly found in the training data. This similarity causes the model to predict it as an aromatic structure incorrectly (Figure 5D, E). For the second case, we examined the FimH lectin domain complexed with a small molecule (Figure 6) (PDB ID: 4AV4)[22]. In the prediction structure by AF3 and Boltz-1, the sp^1 carbon atoms of the alkyne group in the ligand, which should form a linear structure, is instead predicted as an sp^3 alkyl chain-like structure (Figure 6). This observation suggests that structures rarely found in natural biomolecules, such as sp^1 carbons, are biased toward frequently occurring structures like sp^3 carbons, resulting in predicted ligand structures with incorrect stereochemistry.

These problems may stem from fundamental limitations in the input feature representation of the AF3 model. Specifically, the atom and bond embeddings in AF3 do not include explicit information about hybridization states, as well as bond orders, such as single or double bonds. As a result, chemically distinct atoms and bonds are treated as similar input features. This leads the model to skew predictions toward structural patterns frequently occurring in the training dataset. The cases discussed here illustrate stereochemical errors likely caused by this input embedding issues. Even in such cases, our restraint-guided inference strategy successfully generated structures with correct

stereochemical properties (Figures 5C, 6C). These results underscore that the proposed method effectively overcame the limitations of AF3 and Boltz-1 in ligand stereochemistry. Further examples demonstrating the bond angle geometry problem are shown in Figures S5 and S6.

Discussion

In this study, we comprehensively investigated stereochemical accuracy in protein-ligand complex structure predictions by AlphaFold3 (AF3) and its clone model Boltz-1, using unified assessment criteria against the same benchmark dataset. Previous protein-ligand complex prediction methods, such as conventional molecular docking approaches[23] and neural network-based methods like DiffDock[24], typically treat dihedral angles as the target degrees of freedom for structure prediction. AlphaFold2 also uses dihedral angles as prediction targets for protein side-chain structures[1]. These methods can generate molecules with the desired stereochemical properties without problems. In contrast, AF3 employs a diffusion model approach and directly predicts the Cartesian coordinates of each atom in the target molecules. This methodology enabled the prediction of biomolecule and ligand structures within the unified and straightforward framework. However, the direct prediction of Cartesian coordinates creates a disadvantage in that it can generate molecules with incorrect stereochemistry. Our results confirmed that the ligands of the predicted structures contain significant levels of stereochemical errors in their chiralities, bond lengths, and angles. Then, we developed the restraint-guided inference method to address these stereochemical errors. Our evaluation confirmed that this method effectively improves the stereochemical validity of

the ligand structures. Importantly, we demonstrated that the restraint-guided inference method does not adversely affect the performance of the binding pose prediction.

We focused on the method to address the stereochemical errors without retraining the model in this study. However, as discussed in the Results section, one potential cause of the ligand stereochemistry violations may arise from using feature representations, such as atom names, which can vary depending on the input notations. Therefore, a fundamental solution would require a model design that directly incorporates invariant features of compound structures, including hybridization states, chirality, and bond orders.

In the current models, natural biomolecules such as protein amino acids, nucleic acids, and glycans rarely experience those stereochemical issues. This is likely because these molecules are abundant in their training dataset. Therefore, a similar solution could theoretically be possible for small organic compounds by including sufficient numbers and diversity in the training dataset. However, the chemical space of small molecules is orders of magnitude larger than the limited set of biomolecular building blocks, making it impractical to train the neural network model to correctly preserve stereochemistry across arbitrary compounds. From this perspective, our approach provides a practical solution for the ligand stereochemistry problems.

Author contributions

R.I. designed the study and analyzed the data. R.I., and Y.M. performed computational experiments and wrote the manuscript. R.I. directed and supervised this study.

Acknowledgements

The authors thank Naruki Yoshikawa for constructive comments. This research was partially supported by the Research Support Project for Life Science and Drug Discovery (Basis for Supporting Innovative Drug Discovery and Life Science Research (BINDS)) from AMED under Grant Numbers JP25ama121027 for Y.M. and JP25ama121012 for R.I. This work was also supported by JSPS KAKENHI Grant-in-Aid for Transformative Research Areas (A) under Grant Numbers 25H01570 and 25H02250 for R.I. Additional support was provided by the Medical Research Center Initiative for High Depth Omics at the Institute of Science Tokyo, Nanken-Kyoten at the Institute of Science Tokyo 2025, and the Multilayered Stress Diseases project (JPMXP1323015483) at the Institute of Science Tokyo. This work utilized computational resources from the supercomputer “Flow” provided by the Information Technology Center at Nagoya University through the HPCI System Research Project (Project ID: hp240040) and the TSUBAME4.0 supercomputer at the Institute of Science Tokyo.

Data and Software Availability

The materials supporting this article, including the source code and Colab notebook are available at github (https://github.com/cddlab/boltz_ext and https://github.com/cddlab/colabfold_boltz_restr/blob/main/Boltz1.ipynb).

Figures

Figure 1

Performance evaluation of protein-ligand complex structure prediction methods on the PLINDER dataset.

Bar charts comparing AlphaFold3 (AF3), Boltz-1, and Boltz-1 with restraint-guided inference variants (Boltz R, Boltz Rc, and Boltz R1). (A) Protein RMSD, (B) Ligand RMSD, (C) chirality reproduction rate, (D) Bond RMSD, (E) Angle RMSD, and (F) fraction of structures with steric clashes ($d_{\min} < 2$ Å). Results are stratified by entry release dates: "Before" (green) represents entries released before the cut-off date (2021-09-30), while "After" (orange) represents subsequent entries.

Figure 2

Distribution of ligand evaluation metrics on the PLINDER dataset.

Histograms showing the distribution of prediction accuracy metrics for AlphaFold3 (AF3), Boltz-1, and Boltz-1 with restraint-guided inference variants (Boltz R, Boltz Rc, Boltz R1). (A) Bond RMSD, (B) Angle RMSD, and (C) Ligand RMSD distributions. All *x*-axes are displayed on logarithmic scales.

Figure 3

Chirality error in small molecule ligand prediction.

Case study of angiotensin-converting enzyme and inhibitor complex (PDB ID:

4BXK). (A) Chemical structure of the ligand, (B) experimental structure, (C) Boltz-1 prediction, (D) AlphaFold3 prediction, (E) AlphaFold3 prediction using CCD code (1IU), and (F) Boltz-1 with restraint-guided inference (Boltz R). Proteins are depicted as ribbon models, and ligands are shown as ball-and-stick models. In panels (C)-(F), predicted ligand structures are overlaid with the semi-transparent experimental ligand structures for comparison. Arrows indicate chiral centers with stereochemical errors, and *R*/*S* designations denote their stereochemical configurations. Only the restraint-guided method (F) correctly reproduces the experimental chirality.

Figure 4

Chirality error in D-amino acid ligand prediction.

Case study of glutamate transporter in complex with D-aspartate (PDB ID: 6R7R). (A) Chemical structure of the ligand, (B) experimental structure, (C) Boltz-1 prediction, (D) AlphaFold3 prediction, (E) AlphaFold3 prediction using CCD code (DAS), and (F) Boltz-1 with restraint-guided inference (Boltz R). Proteins are shown as ribbon models and ligands as ball-and-stick models. In panels (C)-(F), predicted ligand structures are overlaid with semi-transparent experimental ligand structure for comparison. Arrows indicate chiral centers with stereochemical errors, and *R*/*S* designations denote their stereochemical configurations. Only the restraint-guided method (F) correctly reproduces the chirality of D-aspartate.

Figure 5

Ring conformation errors in ligand structure prediction.

Case study of ectoine-binding protein UehA in complex with ectoine (PDB ID: 3FXB). (A) Chemical structure of the ligand, (B) experimental structure, (C) Boltz-1 with restraint-guided inference (Boltz R), (D) Boltz-1 prediction, and (E) AlphaFold3 prediction. Proteins are shown as ribbon models and ligands as ball-and-stick models. The ring structure is incorrectly predicted as planar in baseline methods (D, E), while the restraint-guided method (C) correctly reproduces the non-aromatic saturated ring conformation.

Figure 6

Bond and angle geometry errors in ligand structure prediction.

Case study of FimH lectin domain in complex with a small molecule (PDB ID: 4AV4). (A) Chemical structure of the ligand, (B) experimental structure, (C) Boltz-1 with restraint-guided inference (Boltz R), (D) Boltz-1 prediction, and (E) AlphaFold3 prediction. Proteins are shown as ribbon models and ligands as ball-and-stick models. Arrows indicate the regions where the bond and angle geometry errors occur.

Tables

Table 1

Evaluation metrics for proteins and ligands in the predicted structures from the PLINDER dataset used in this study.

	Conditions	Protein RMSD (Å)	Chirality (%)	Bond RMSD (Å)	Angle RMSD (°)	Ligand RMSD (Å)	Min dist<2Å (%)
AF3		**0.7178**	70.77	0.0209	0.328	**3.694**	**9.71**
Boltz-1		0.7487	60.58	0.0313	0.780	4.511	14.50
Boltz R1	σ_{start}=0.005, w_{chiral}=w_{geom}=1	0.7421	99.77	0.0524	1.558	4.607	21.93
Boltz RC	σ_{start}=1, w_{chiral}=1, w_{geom}=0	0.7400	**100**	0.0239	0.609	4.544	14.95
Boltz R	σ_{start}=1, w_{chiral}=w_{geom}=1	0.7386	**100**	**0.0005**	**0.090**	4.559	16.74

References

(1) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583–589.

(2) Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A. J.; Bambrick, J.; Bodenstein, S. W.; Evans, D. A.; Hung, C.-C.; O'Neill, M.; Reiman, D.; Tunyasuvunakool, K.; Wu, Z.; Žemgulytė, A.; Arvaniti, E.; Beattie, C.; Bertolli, O.; Bridgland, A.; Cherepanov, A.; Congreve, M.; Cowen-Rivers, A. I.; Cowie, A.; Figurnov, M.; Fuchs, F. B.; Gladman, H.; Jain, R.; Khan, Y. A.; Low, C. M. R.; Perlin, K.; Potapenko, A.; Savy, P.; Singh, S.; Stecula, A.; Thillaisundaram, A.; Tong, C.; Yakneen, S.; Zhong, E. D.; Zielinski, M.; Žídek, A.; Bapst, V.; Kohli, P.; Jaderberg, M.; Hassabis, D.; Jumper, J. M. Accurate Structure Prediction of Biomolecular Interactions with AlphaFold 3. *Nature* **2024**, *630* (8016), 493–500.

(3) Wohlwend, J.; Corso, G.; Passaro, S.; Reveiz, M.; Leidal, K.; Swiderski, W.; Portnoi, T.; Chinn, I.; Silterra, J.; Jaakkola, T.; Barzilay, R. Boltz-1 Democratizing Biomolecular Interaction Modeling. *bioRxivorg*, 2024, 2024.11.19.624167. https://doi.org/10.1101/2024.11.19.624167.

(4) Chen, X.; Zhang, Y.; Lu, C.; Ma, W.; Guan, J.; Gong, C.; Yang, J.; Zhang, H.; Zhang,

K.; Wu, S.; Zhou, K.; Yang, Y.; Liu, Z.; Wang, L.; Shi, B.; Shi, S.; Xiao, W. Protenix - Advancing Structure Prediction through a Comprehensive AlphaFold3 Reproduction. *bioRxiv*, 2025, 2025.01.08.631967. https://doi.org/10.1101/2025.01.08.631967.

(5) Boitreaud, J.; Dent, J.; McPartlon, M.; Meier, J.; Reis, V.; Rogozhnikov, A.; Wu, K. Chai-1: Decoding the Molecular Interactions of Life. *Synthetic Biology*, 2024. https://www.biorxiv.org/content/10.1101/2024.10.10.615955v1.full.pdf.

(6) Liu, L.; Zhang, S.; Xue, Y.; Ye, X.; Zhu, K.; Li, Y.; Liu, Y.; Gao, J.; Zhao, W.; Yu, H.; Wu, Z.; Zhang, X.; Fang, X. Technical Report of HelixFold3 for Biomolecular Structure Prediction. *arXiv [q-bio.BM]*, 2024. http://arxiv.org/abs/2408.16975 (accessed 2025-03-14).

(7) Wang, S.; Witek, J.; Landrum, G. A.; Riniker, S. Improving Conformer Generation for Small Rings and Macrocycles Based on Distance Geometry and Experimental Torsional-Angle Preferences. *J. Chem. Inf. Model.* **2020**, *60* (4), 2044–2058.

(8) Durairaj, J.; Adeshina, Y.; Cao, Z.; Zhang, X.; Oleinikovas, V.; Duignan, T.; McClure, Z.; Robin, X.; Studer, G.; Kovtun, D.; Rossi, E.; Zhou, G.; Veccham, S.; Isert, C.; Peng, Y.; Sundareson, P.; Akdel, M.; Corso, G.; Stärk, H.; Tauriello, G.; Carpenter, Z.; Bronstein, M.; Kucukbenli, E.; Schwede, T.; Naef, L. PLINDER: The Protein-Ligand Interactions Dataset and Evaluation Resource. *bioRxiv*, 2024, 2024.07.17.603955. https://doi.org/10.1101/2024.07.17.603955.

(9) Liu, Z.; Li, Y.; Han, L.; Li, J.; Liu, J.; Zhao, Z.; Nie, W.; Liu, Y.; Wang, R. PDB-Wide Collection of Binding Data: Current Status of the PDBbind Database. *Bioinformatics* **2015**, *31* (3), 405–412.

(10) Hanson, R. M.; Musacchio, S.; Mayfield, J. W.; Vainio, M. J.; Yerin, A.; Redkin, D.

Algorithmic Analysis of Cahn-Ingold-Prelog Rules of Stereochemistry: Proposals for Revised Rules and a Guide for Machine Implementation. *J. Chem. Inf. Model.* **2018**, *58* (9), 1755–1765.

(11) Watson, J. L.; Juergens, D.; Bennett, N. R.; Trippe, B. L.; Yim, J.; Eisenach, H. E.; Ahern, W.; Borst, A. J.; Ragotte, R. J.; Milles, L. F.; Wicky, B. I. M.; Hanikel, N.; Pellock, S. J.; Courbet, A.; Sheffler, W.; Wang, J.; Venkatesh, P.; Sappington, I.; Torres, S. V.; Lauko, A.; De Bortoli, V.; Mathieu, E.; Ovchinnikov, S.; Barzilay, R.; Jaakkola, T. S.; DiMaio, F.; Baek, M.; Baker, D. De Novo Design of Protein Structure and Function with RFdiffusion. *Nature* **2023**, *620* (7976), 1089–1100.

(12) Karras, T.; Aittala, M.; Aila, T.; Laine, S. Elucidating the Design Space of Diffusion-Based Generative Models. *arXiv [cs.CV]*, 2022. http://arxiv.org/abs/2206.00364 (accessed 2025-03-14).

(13) Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; van der Walt, S. J.; Brett, M.; Wilson, J.; Millman, K. J.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C. J.; Polat, İ.; Feng, Y.; Moore, E. W.; VanderPlas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; van Mulbregt, P.; SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17* (3), 261–272.

(14) Steinegger, M.; Söding, J. MMseqs2 Enables Sensitive Protein Sequence Searching for the Analysis of Massive Data Sets. *Nat. Biotechnol.* **2017**, *35* (11), 1026–1028.

(15) Mirdita, M.; Schütze, K.; Moriwaki, Y.; Heo, L.; Ovchinnikov, S.; Steinegger, M. ColabFold: Making Protein Folding Accessible to All. *Nat. Methods* **2022**, *19* (6),

679–682.

(16) Johnson, L. S.; Eddy, S. R.; Portugaly, E. Hidden Markov Model Speed Heuristic and Iterative HMM Search Procedure. *BMC Bioinformatics* **2010**, *11* (1), 431.

(17) Eddy, S. R. Accelerated Profile HMM Searches. *PLoS Comput. Biol.* **2011**, *7* (10), e1002195.

(18) Zheng, H.; Lin, H.; Alade, A. A.; Chen, J.; Monroy, E. Y.; Zhang, M.; Wang, J. AlphaFold3 in Drug Discovery: A Comprehensive Assessment of Capabilities, Limitations, and Applications. *bioRxiv*, 2025, 2025.04.07.647682. https://doi.org/10.1101/2025.04.07.647682.

(19) Douglas, R. G.; Sharma, R. K.; Masuyer, G.; Lubbe, L.; Zamora, I.; Acharya, K. R.; Chibale, K.; Sturrock, E. D. Fragment-Based Design for the Development of N-Domain-Selective Angiotensin-1-Converting Enzyme Inhibitors. *Clin. Sci. (Lond.)* **2014**, *126* (4), 305–313.

(20) Arkhipova, V.; Trinco, G.; Ettema, T. W.; Jensen, S.; Slotboom, D. J.; Guskov, A. Binding and Transport of D-Aspartate by the Glutamate Transporter Homolog GltTk. *Elife* **2019**, *8*. https://doi.org/10.7554/eLife.45286.

(21) Lecher, J.; Pittelkow, M.; Zobel, S.; Bursy, J.; Bönig, T.; Smits, S. H. J.; Schmitt, L.; Bremer, E. The Crystal Structure of UehA in Complex with Ectoine-A Comparison with Other TRAP-T Binding Proteins. *J. Mol. Biol.* **2009**, *389* (1), 58–73.

(22) Wellens, A.; Lahmann, M.; Touaibia, M.; Vaucher, J.; Oscarson, S.; Roy, R.; Remaut, H.; Bouckaert, J. The Tyrosine Gate as a Potential Entropic Lever in the Receptor-Binding Site of the Bacterial Adhesin FimH. *Biochemistry* **2012**, *51* (24), 4790–4799.

(23) Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading.

J. Comput. Chem. **2010**, *31* (2), 455–461.

(24) Corso, G.; Stärk, H.; Jing, B.; Barzilay, R.; Jaakkola, T. DiffDock: Diffusion Steps, Twists, and Turns for Molecular Docking. *arXiv [q-bio.BM]*, 2022. http://arxiv.org/abs/2210.01776.

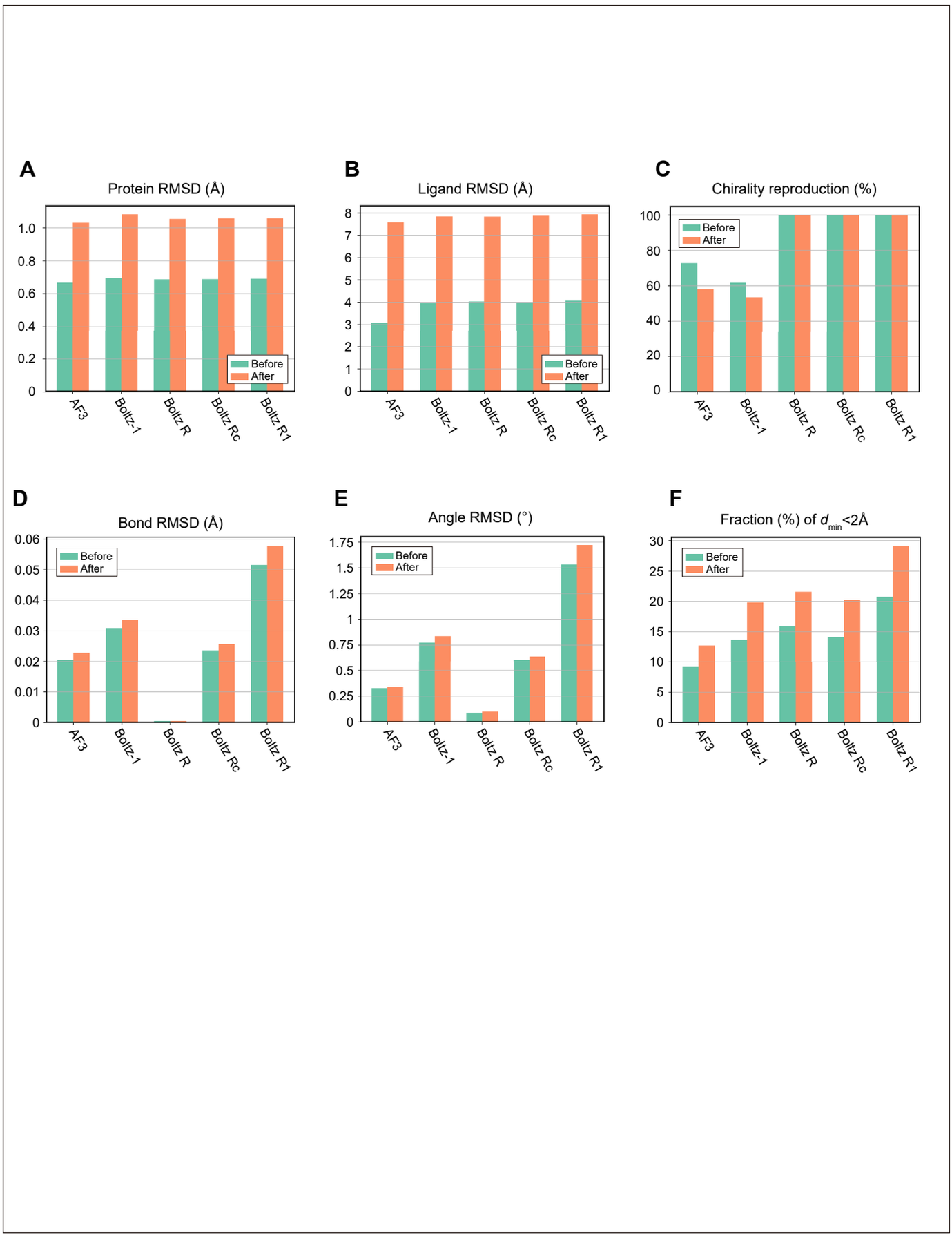
A
Protein RMSD (Å)
Before
After
AF3
Boltz-1
Boltz R
Boltz Rc
Boltz R1
B
Ligand RMSD (Å)
C
Chirality reproduction (%)
D
Bond RMSD (Å)
E
Angle RMSD (°)
F
Fraction (%) of d_{min}<2Å

Figure 1

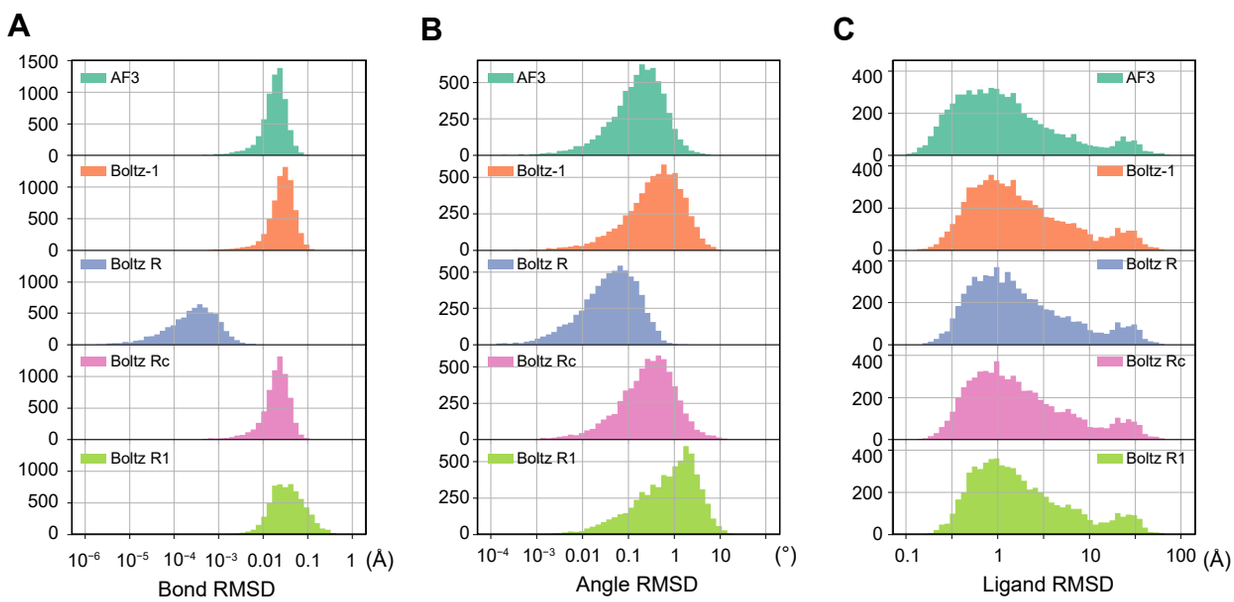
A
B
C
AF3
Boltz-1
Boltz R
Boltz Rc
Boltz R1
Bond RMSD
Angle RMSD
Ligand RMSD
(Å)
(°)

Figure 2

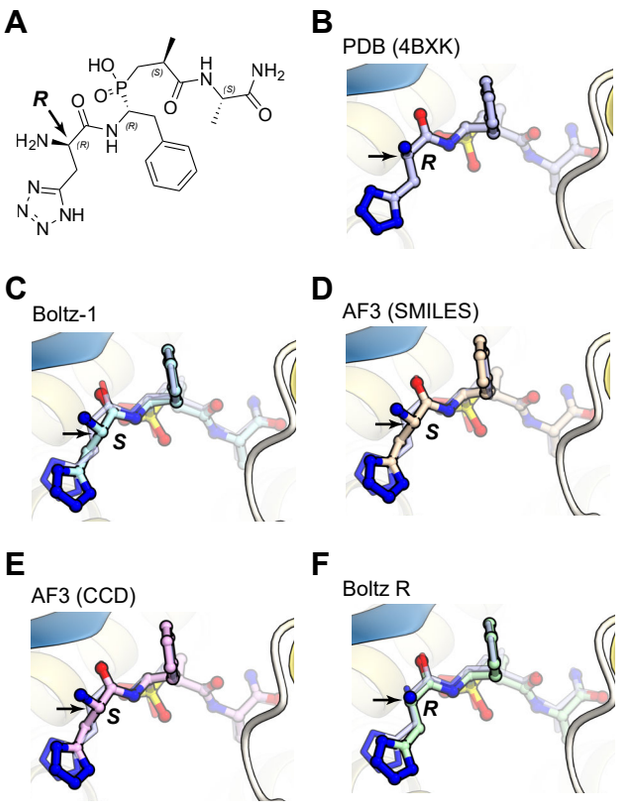

Figure 3

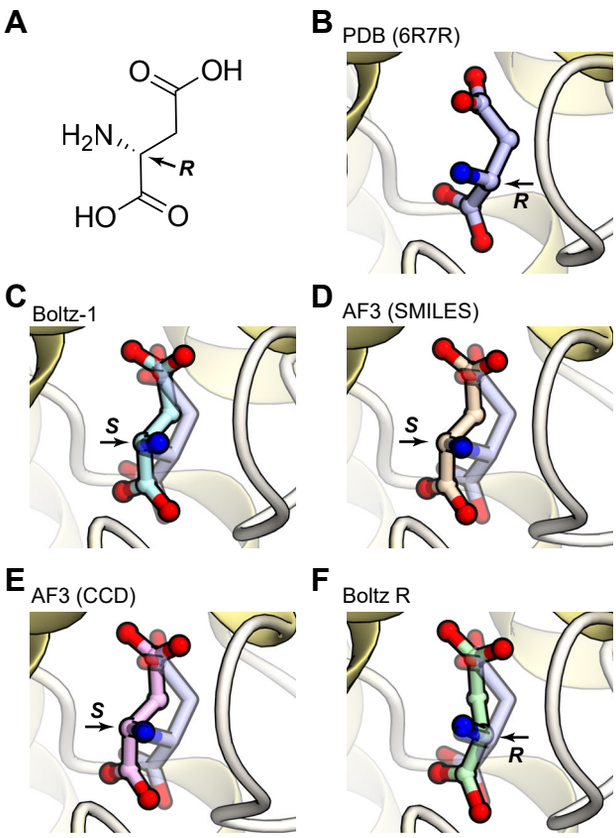
A
O
OH
H2N
R
HO
O
B
PDB (6R7R)
R
C
Boltz-1
S
D
AF3 (SMILES)
S
E
AF3 (CCD)
S
F
Boltz R
R

Figure 4

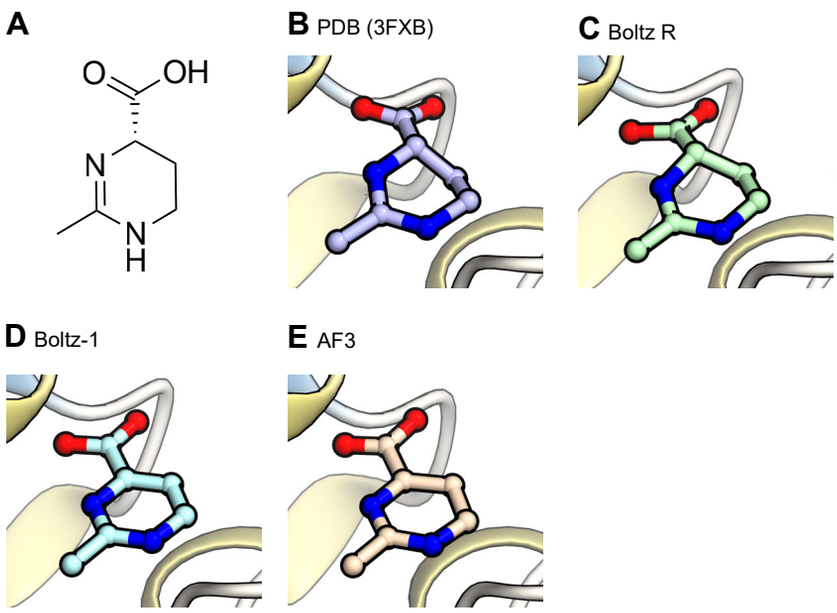

Figure 5

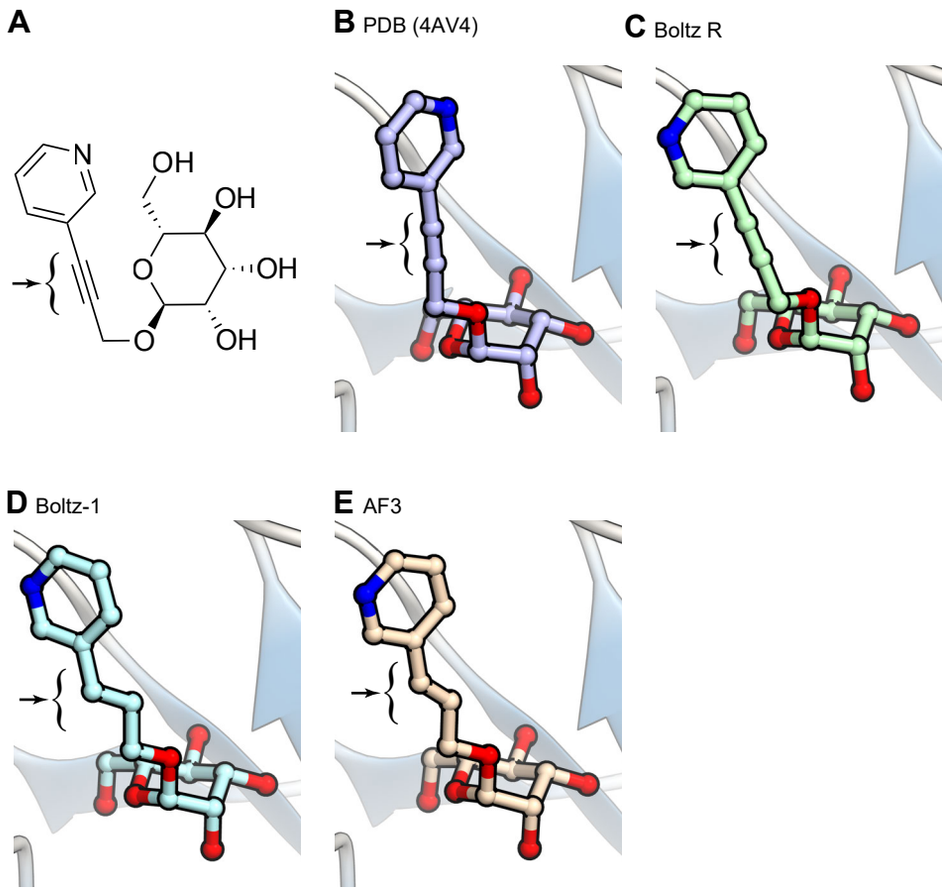

Figure 6