

# Boltz-2

## 简介

Boltz 是一系列用于生物分子相互作用预测的模型。早期版本 Boltz-1 是首个在准确性上接近 AlphaFold3 的完全开源模型，标志着结构预测的一个重要突破。

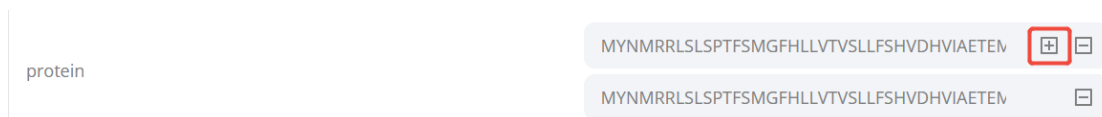
最新推出的 **Boltz-2** 是一个全新的生物分子基础模型，能够**同时建模分子复合物结构与结合亲和力**，这对于实现高精度的分子设计至关重要。相比前代产品和 AlphaFold3，Boltz-2 在建模能力上实现了质的飞跃。

更重要的是，**Boltz-2 是首个准确性可媲美基于物理的自由能微扰（FEP）方法的深度学习模型**，但其计算速度快了 **1000 倍**。这使得高精度的计算筛选（in silico screening）首次在早期药物发现阶段变得切实可行，极大提高了研发效率和成本效益。

## 参数说明

### ● protein

- 说明：输入蛋白质的氨基酸序列。
- 格式：支持多条序列，适用于多链复合物建模，可通过“+”或“-”控制多序列的输入。
- 注意：每条序列具有唯一的编号，从上至下，分别是 P1、P2 等。
- 示例：



### ● protein\_MSA\_method

- 说明：指定蛋白质多序列比对的方法
- 可选值：
  - ◆ **Default Search**: 启用 MSA，使用 mmseqs2 方法自动生成 MSA 文件作为序列搜索数据库。

- ◆ Upload File: 启用 MSA, 使用用户提供的预计算的 MSA 文件。
- ◆ No MSA (默认值): 不启用 MSA。
- 注意:
  - ◆ MSA 当前只针对蛋白序列。
  - ◆ 使用 mmseqs2 方法后将根据目标序列搜索数据库。
  - ◆ 启用 MSA 后将提取同源序列信息作为模型输入特征。
- **protein\_MSA\_file**
  - 说明: 提供预计算的蛋白质 MSA 文件。
  - 格式: a3m 文件。
  - 适用场景: 当 protein\_MSA\_method 设置为 “Upload File” 时使用。
- **rna**
  - 说明: 输入 RNA 分子的核苷酸序列。
  - 格式: 字符串, 支持多条序列。可通过 “+” 或 “-” 控制多序列的输入。
  - 注意: 每条序列具有唯一的编号, 从上至下, 分别是 R1、R2 等。
- **dna**
  - 说明: 输入 DNA 分子的核苷酸序列。
  - 格式: 字符串, 支持多条序列。可通过 “+” 或 “-” 控制多序列的输入。
  - 注意: 每条序列具有唯一的编号, 从上至下, 分别是 D1、D2 等。
- **modifications**
  - 说明: 包括翻译后修饰 (PTM) 信息的文本。
  - 格式:
    - ◆ 每个输入框定义一个修饰信息, 若需定义多条, 可通过 “+” 或 “-” 按钮添加或删除输入框。
    - ◆ 每个修饰信息包含 3 个部分, 使用符号 “-” 连接:
      - EntityID: 要修饰的链编号。如, 第 2 条蛋白序列为 P2。
      - ptmType: 修饰后残基的 CCD 编号。
      - ptmPosition: 被修饰残基的位置。从 1 开始的残基序号。
  - 注意: 可以通过访问 RCSB Chemical Component Dictionary 查询各种 CCD 修饰代码。例如, MSE 表示硒代蛋氨酸, TPO 表示磷酸化苏氨酸。

- 示例: P1-HY3-1 表示第一条蛋白上第 1 位氨基酸上发生了类型为 HY3 的 PTM。

### ● ligand\_smiles

- 说明: 输入小分子的 SMILES 表达式 Boltz-1 将自动生成其三维结构并用于结构预测。

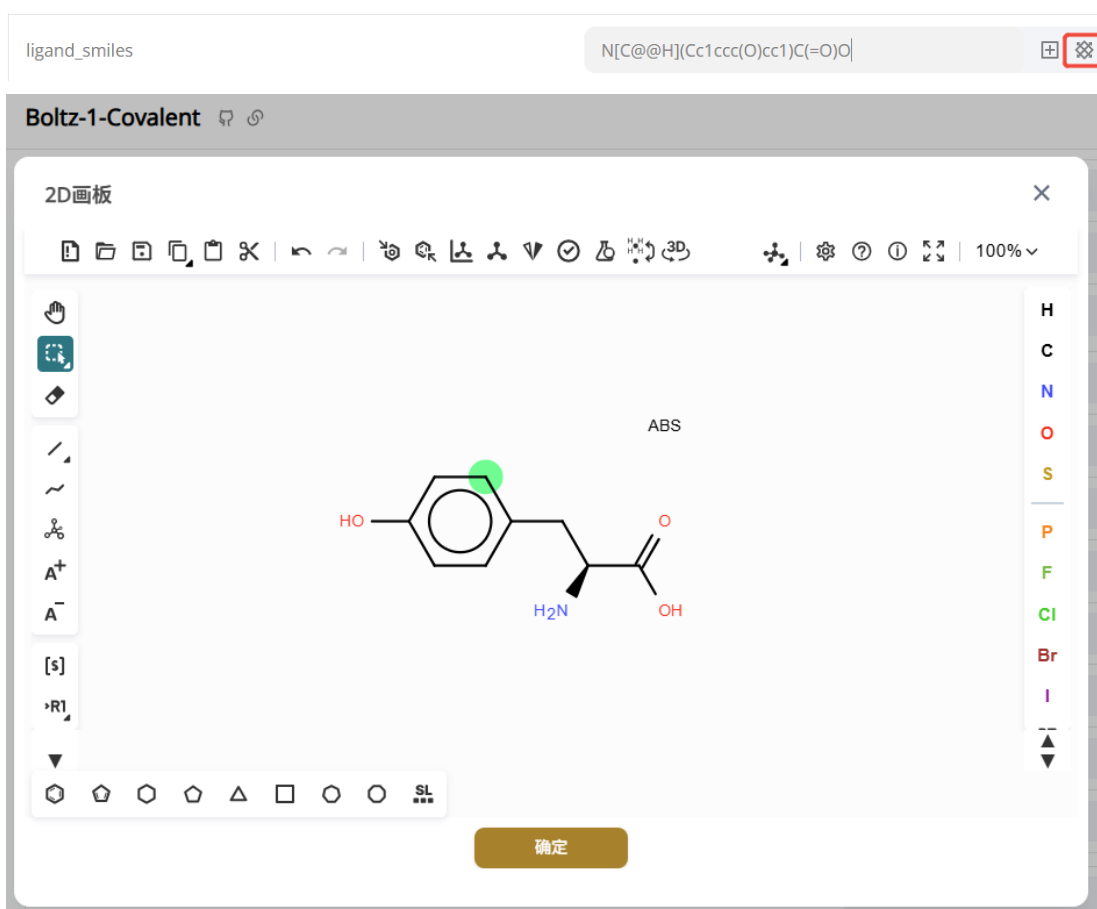
- 格式:

- ◆ 支持多个小分子, 可通过“+”或“-”控制多个小分子的输入。
- ◆ 输入格式有 2 种:

- 当前小分子参与共价键连接:

点击“+”右边的插件, 复制小分子 SMILES 进入插件, 在画板上粘贴后点击选择参与共价键连接的原子。

示例图如下:



注释: 选中后的结果分为3部分: 小分子的 SMILES 表达式; 选中的原子; 当前小分子的 SDF 文件内容。

□ 当前小分子不参与共价键连接：直接将小分子的 SMILES 表达式填写入输入框即可。

○ 注意：

◆ 每个小分子具有唯一编号，从上至下依次为 L1、L2 等。

◆ 多个小分子，表示一起进行结构预测，而非分别与输入序列做结构预测。

### ● **ligand\_ccd**

○ 说明：填写 RCSB 化学组件数据库的三字母配体编码(如 ATP、FAD)，用于调用已有的标准配体结构。

○ 格式：

◆ 支持多个 CCD 编码的配体小分子，可通过“+”或“-”控制多个 CCD 编码的配体小分子的输入。

◆ 输入格式有 2 种：

□ 当前 CCD 编码配体参与共价键连接：CCD code,AtomName。(使用“,”分割信息)。

□ 当前 CCD 编码配体不参与共价键连接：CCD code。

○ 注意：每个小分子具有唯一编号，此部分的小分子编号从 ligand\_smiles 编号之后继续递增，如 L3、L4... (若前者为 L1、L2)。

### ● **covalent\_bonds**

○ 说明：用于指定共价键连接的原子对 (例如蛋白与小分子之间)。

○ 格式：

◆ 每个输入框定义一条共价键信息，若需定义多条，可通过“+”或“-”按钮添加或删除输入框。

◆ 每条共价键定义涉及 2 个原子 (一个来自生物大分子，一个来自配体)，包含 4 个部分，并使用：“-”符号连接，格式如下：

**EntityID\_A-ResidueID\_A-AtomName\_A-Ligand\_B**

各部分说明如下：

□ EntityID\_A：生物大分子中原子 A 所在的序列编号(顺序编号)。  
如，第 2 条蛋白序列为 P2。

- ResidueID\_A: 生物大分子中原子 A 所在残基的编号，格式为编号（如 145 表示 145 号位的氨基酸）。
- AtomName\_A: 生物大分子中原子 A 的原子名称，如 SG（半胱氨酸的硫原子）。
- Ligand\_B: 配体中原子 B 所在的序列编号。

○ 示例:

covalent\_bonds P1-63-SG-L1 +

注释: 在参数 ligand\_smiles 中选择了要做共价键的原子后，当前定义了链 P1 中第 63 位残基 SG 与小分子链 L1 选中原子形成共价键。

● **pocket\_restraints**

○ 说明: 指定一个**特定残基与另一条链**之间的结合位点限制，常用于分子建模或蛋白质-配体对接等场景。

○ 格式: 每条限制信息由以下 3 个部分组成，并使用“-”连接，格式如下:

**BinderChainID-ContactsChainID-ContactsRESIDX**

各字段含义如下:

- ◆ **BinderChainID:** 与结合位点结合的链的序列编号(顺序编号)。例如，该链位于多个小分子序列中的第 2 条序列时，编号为 L2。
- ◆ **ContactsChainID:** 结合位点所在的链的序列编号(顺序编号)。例如，该结合位点位于多个 DNA 序列中的第 3 条序列时，编号名为 D3。
- ◆ **ContactsRESIDX:** 结合位点所在的链中与结合位点关联的残基编号，从 1 开始。

○ 注意: 在当前一次提交中，如果有多条限制信息，结构预测过程中将同时考虑它们。

○ 示例:

pocket\_restraints L1-P1-829 + -  
L1-P1-139 -

注释: 当前定义了小分子 L1 与 链 P1 中的 829、139 位置的残基结合。

○ 说明:

- ◆ 当前模型只支持指定 1 个主绑定链 **Binder** 。
- ◆ 可以指定任意数量的来自其他链的与 **Binder** 链结合的残基。

● **contact\_restraints**

○ 说明: 用于指定两个残基之间的空间距离限制, 常用于结构建模、模拟或约束优化等场景。

○ 格式: 每条距离约束信息由以下 6 个部分组成, 并使用 - 符号连接, 格式如下:

EntityID\_A-RES\_IDX/ATOM\_NAME-EntityID\_B-RES\_IDX/ATOM\_NAME-MAX\_DISTANCE

各字段说明如下:

- ◆ **EntityID\_A**: 残基 A 所在的序列编号(顺序编号)。如, 第 2 条蛋白链为 P2。
- ◆ **RES\_IDX/ATOM\_NAME**: 残基 A 的位置 (如果是小分子位置则为 1, 则应输入原子名)。
- ◆ **EntityID\_B**: 残基 B 所在的序列编号(顺序编号)。如, 第 3 条 dna 链为 D3。
- ◆ **RES\_IDX/ATOM\_NAME**: 残基 B 的位置 (如果是小分子位置则为 1, 则应输入原子名)。
- ◆ **MAX\_DISTANCE**: 残基间的最大距离 (单位: Å, 埃)。

○ 示例:

contact\_restraints②

P1-387-P2-101-5.5



注释: 当前定义了 P1 链中第 387 位残基 C 与 P2 链中第 101 位残基 Y 之间的距离约束, 距离小于 5.5Å。

● **template\_file**

○ 说明: 结构预测时蛋白的参考模板文件。

- 格式：CIF 文件。
- 注意：当前上传的 CIF 文件只能用于蛋白链做模板文件。

- **template\_info**

- 说明：模板文件的使用信息。
- 格式：每个模板文件的信息由 3 个部分组成，并使用符号“-”连接，格式如下：

template\_file\_id-chain\_id-template\_id

各字段说明如下：

- ◆ **template\_file\_id**: 当前使用的模板文件编号。例如，使用上传的第一个模板文件，则模板文件编号为“temp1”。
- ◆ **chain\_id**: 结构预测时使用模板参考的链的序列编号。例如，结构预测时，第 2 个蛋白链使用模板参考文件，则序列编号为“P2”。
- ◆ **template\_id**: 模板文件中被参考的链编号。例如，若参考模板文件中的 A 链，则填写“A”。

- **affinity\_binder**

- 说明：选择对小分子链中的哪个链计算亲和力。
- 格式：输入小分子的编号。例如，计算第 2 个小分子的亲和力，输入“L2”。

- **diffusion\_sample**

- 说明：设置扩散采样数量（如 5）。采样次数越多，可获得更多结构候选，增强结构多样性与可信度。

## 结果说明

输出结果分为 **原始输出** 和 **可视化结果** 两部分。

- **原始输出**: 工具直接返回的内容。
- **可视化结果**: 指满足平台可视化规范的输出格式，系统会自动将其渲染为

图表、结构图、表格等形式，使任务结果更加直观、清晰，便于理解与展示。

**Boltz-2** 的可视化结果包含 3 个部分：

● Boltz-2 Result

列名	说明
Entry	当前结果的唯一编号 ID。
Confidence	模型对该结构预测结果文件的置信度评分。
ptm	衡量整体折叠结构准确性的一种预测 TM-score (Template Modeling Score)，考虑整体拓扑关系。值范围为 0-1，越大结构准确度越高，大于 0.7 认为结构准确度越高。
iptm	评估蛋白质-蛋白质接口的可信度，仅在链与链（例如两个蛋白）接口之间聚合的 TM-score。值范围为 0-1，越大可信度越高。
ligand_iptm	仅在蛋白-配体接口计算得到的 TM-score。值范围为 0-1。
protein_iptm	仅针对蛋白-蛋白界面聚合的 TM-score，和 iptm 值相同（若系统中没有配体）。
complex_plddt	整个复合体的平均 pLDDT 值，表示每个氨基酸的预测准确性。取值范围为 0-1，大于 0.7 认为预测的结构比较准确。
complex_iplddt	仅在接口区域计算的平均 pLDDT。值范围为 0-1。
complex_pde	衡量预测中残基对之间距离的平均误差的置信度，用于交叉验证结构质量。值范围为 0-1，值越接近 1，表示模型预测残基之间距离的置信度越高。
complex_ipde	接口区域的 PDE 分数，单位为埃 (Å)，值越大，预测误差越大。
complex FileName	结构预测的复合物结果文件。

○ 可视化示例图：

Boltz-2 Result 📄

Entry	Confidence	ptm	iptm	ligand_iptm	protein_iptm	complex_plddt	complex_iplddt	cor
0	0.7268	0.6342	0.9201	0.9201	0.0	0.6785	0.636	1.
1	0.725	0.6631	0.8833	0.8833	0.0	0.6854	0.6317	1.
2	0.7242	0.6734	0.8283	0.8283	0.0	0.6982	0.6372	1.
3	0.7083	0.6405	0.8784	0.8784	0.0	0.6658	0.604	1.
4	0.6914	0.6775	0.7416	0.7416	0.0	0.6788	0.5747	1.

iptm	ligand_iptm	protein_iptm	complex_pld dt	complex_ipld dt	complex_pde	complex_ipde	Complex FileName
0.9201	0.9201	0.0	0.6785	0.636	1.6796	2.5624	Entry_0.cif
0.8833	0.8833	0.0	0.6854	0.6317	1.5816	2.5172	Entry_1.cif
0.8283	0.8283	0.0	0.6982	0.6372	1.5473	2.7416	Entry_2.cif
0.8784	0.8784	0.0	0.6658	0.604	1.7077	2.7229	Entry_3.cif
0.7416	0.7416	0.0	0.6788	0.5747	1.563	3.2276	Entry_4.cif

### ● Boltz-2 Affinity Result

列名	说明
affinity_pred_value	模型预测的结合强度, IC50 值(单位 $\mu\text{M}$ )。
affinity_probability_binary	模型预测为 binder 的概率。

○ 可视化示例图:

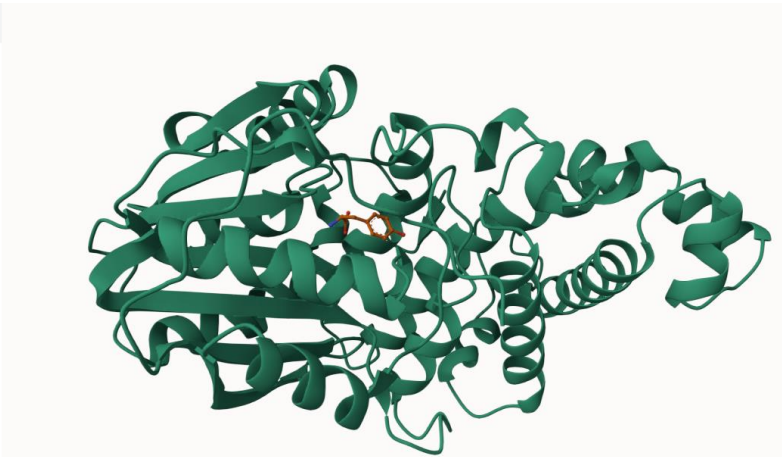
affinity_pred_value	affinity_probability_binary
362.7469	0.4153

### ● 一系列 3D 的结果文件。

○ 可视化示例图:

模式选择 单选 ▾

- Entry\_0.cif
- Entry\_1.cif
- Entry\_2.cif
- Entry\_3.cif
- Entry\_4.cif



## 参考文献

Boltz-2: Towards Accurate and Efficient Binding Affinity Prediction. Passaro, Saro and Corso, Gabriele and Wohlwend, Jeremy and Reveiz, Mateo and Thaler, Stephan and Somnath, Vignesh Ram and Getz, Noah and Portnoi, Tally and Roy, Julien and Stark, Hannes and Kwabi-Addo, David and Beaini, Dominique and Jaakkola, Tommi and Barzilay, Regina;

**Github:** <https://github.com/jwohlwend/boltz>

**Article:** <https://bit.ly/boltz2-pdf>