Swiss Institute of Bioinformatics

DAY 2, PART 2

AlphaFold3

Diana Rapota, Rok Breznikar, Janani Durairaj

23-24 June 2026

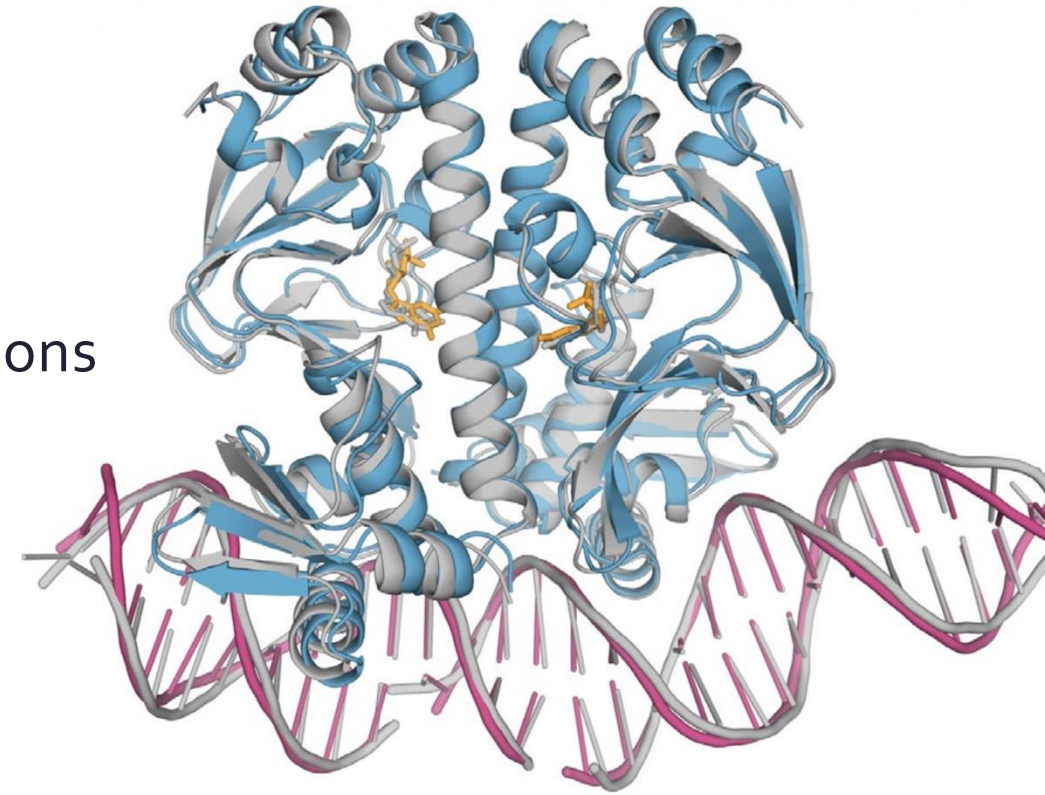
elixir SWITZERLAND

What more you can do with it

- Proteins
- DNA, RNA
- Small molecules
- Ions
- Modified residues, post-translational modifications

- Complexes of all of these

- Doesn't model:
  - Water
  - Hydrogens

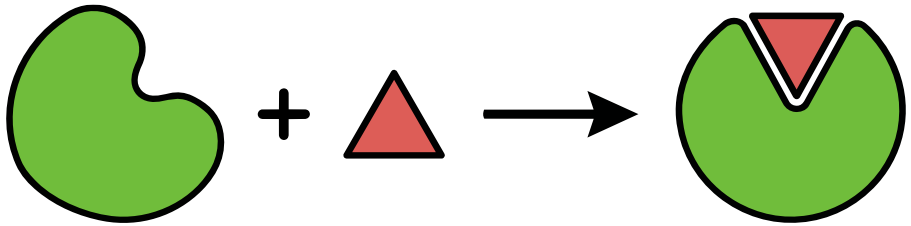


Abramson, J., Adler, J., Dunger, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*

Protein-ligand complex

- Taking AlphaFold2 structure (protein on its own) and adding the ligand doesn't work that well (docking)

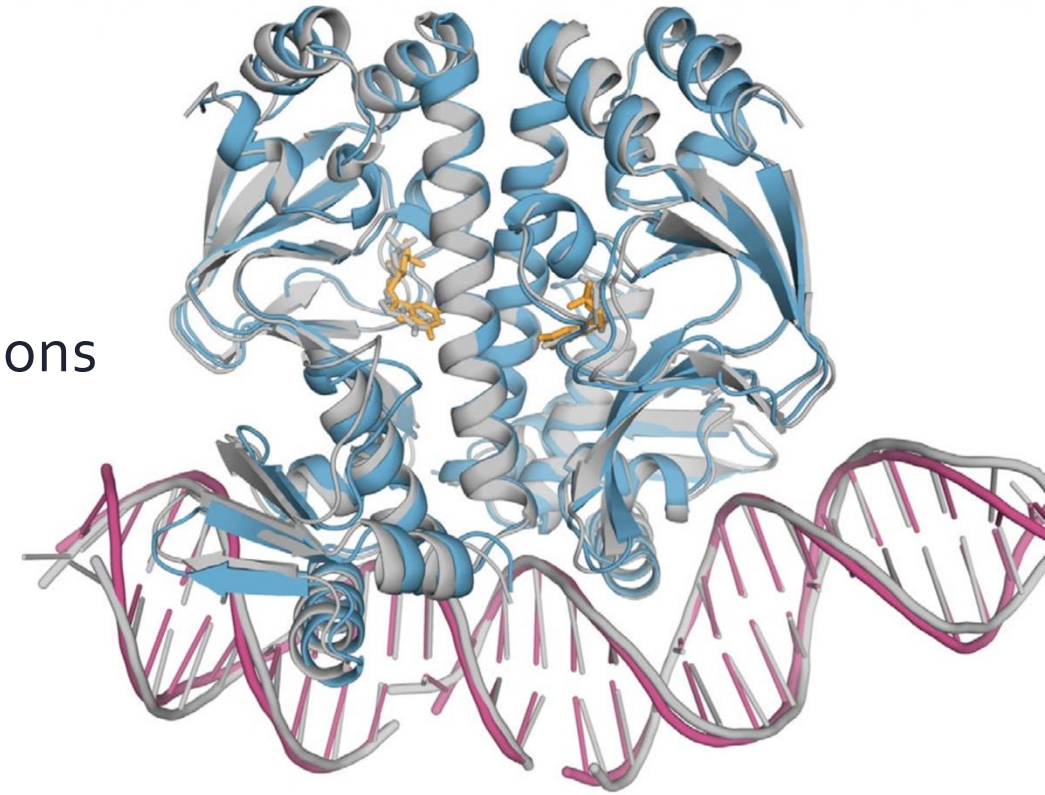
- Induced fit
(protein is not in the correct conformation for the ligand)



What more you can do with it

- Proteins
- DNA, RNA
- Small molecules
- Ions
- Modified residues, post-translational modifications

- Aims to model the whole complex together

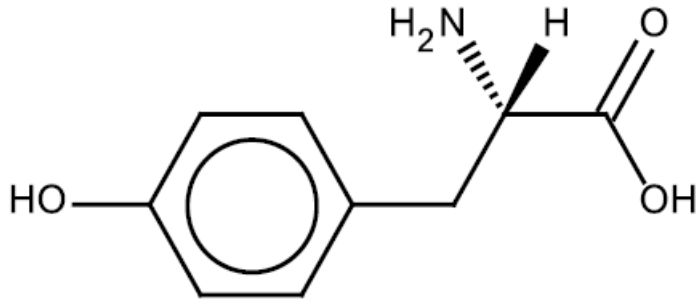


Abramson, J., Adler, J., Dunger, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*

SMILES

- String notation for small molecules (ligands)

- Describes a molecular graph with 4 basic rules
  - Atoms
  - Bonds
  - Branching
  - Ring closures



- Hydrogens can be included or omitted (typically omitted)

A. Varnek, I. Baskin, T. Madzhidov: Fundamentals of Chemoinformatics

SMILES

Atoms

C

O

N

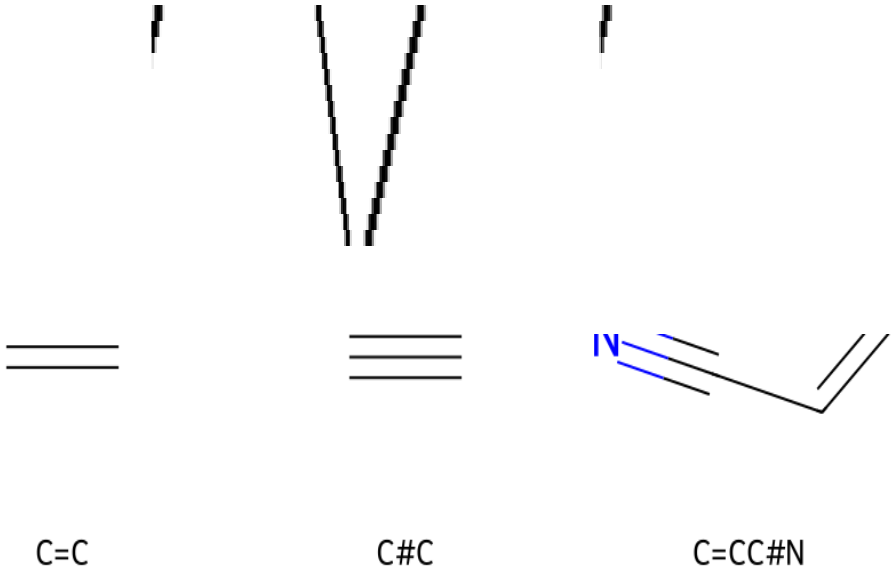
…

Bonds

- (single)
(usually omitted)

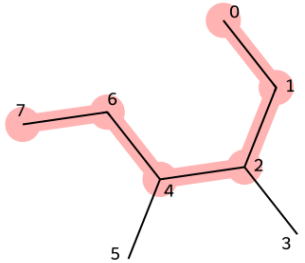
= (double)

(triple)



Branching

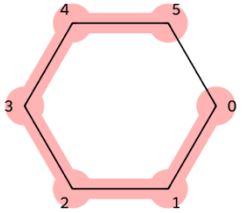
( )



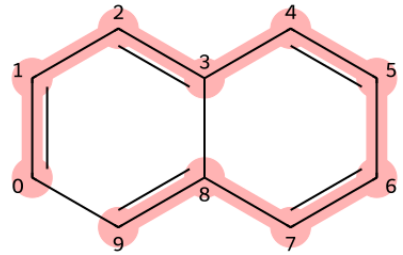
CCC(C)C(C)CC

Ring closure

numbers



C1CCCCC1



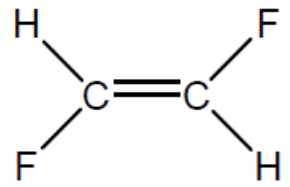
c1ccc2ccccc2c1

https://matlantis.com/en/resources/blog/how-to-write-smiles/
A. Varnek, I. Baskin, T. Madzhidov: Fundamentals of Chemoinformatics

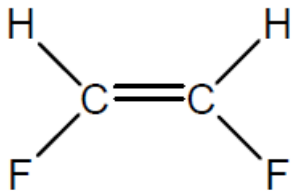
SMILES

Orientation around double bonds

/ and \



F/C=C/F

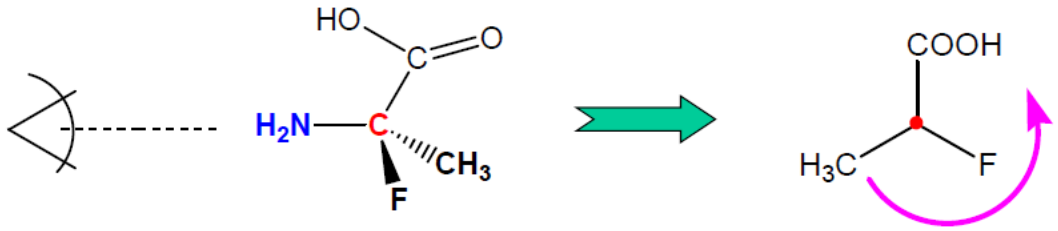


F/C=C\F

Chiral centers

@: ↺

@@: ↻



N[C@](C)(F)C(=O)O

Tokenization

AlphaFold2 only dealt with proteins

→ divided into residues

https://www.ebi.ac.uk/training/online/courses/alphafold/alphafold-3-and-alphafold-server/introducing-alphafold-3/how-does-alphafold-3-work/

Tokenization

AlphaFold3 deals with more

Proteins → residues

DNA, RNA → nucleotides

Small molecules

Ions

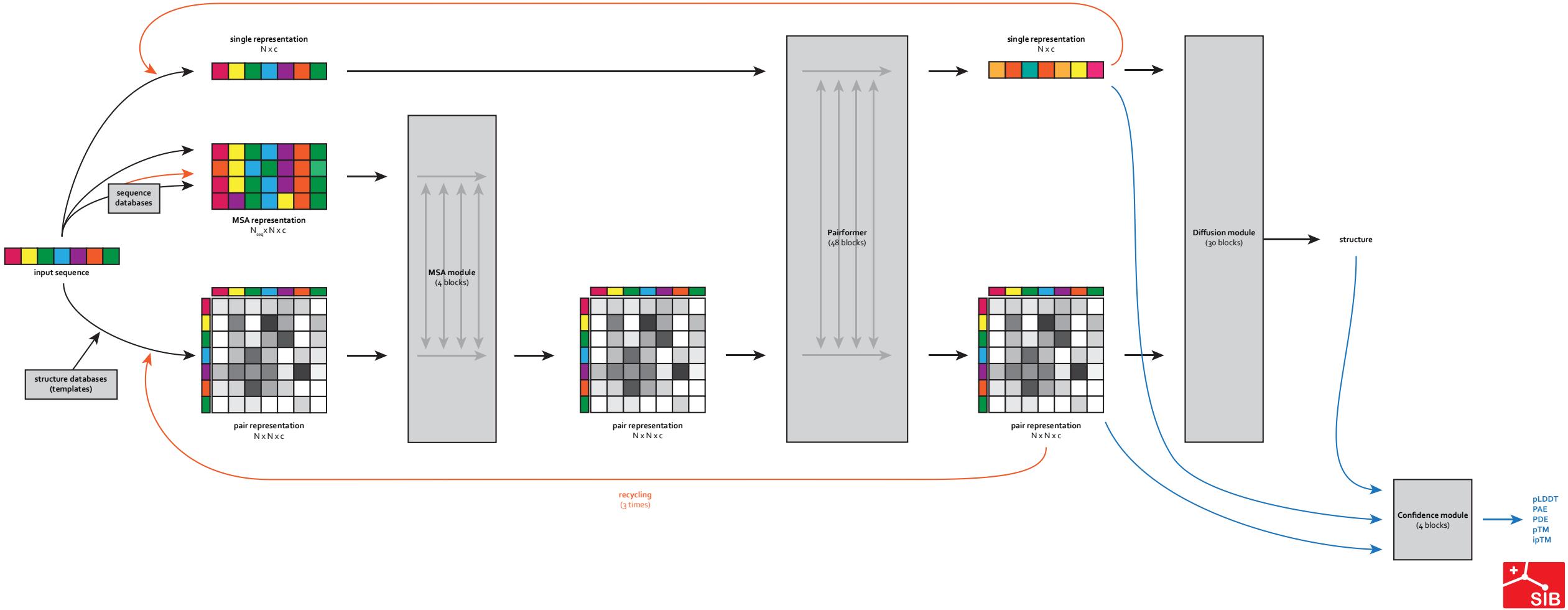
Modified residues

} → atoms

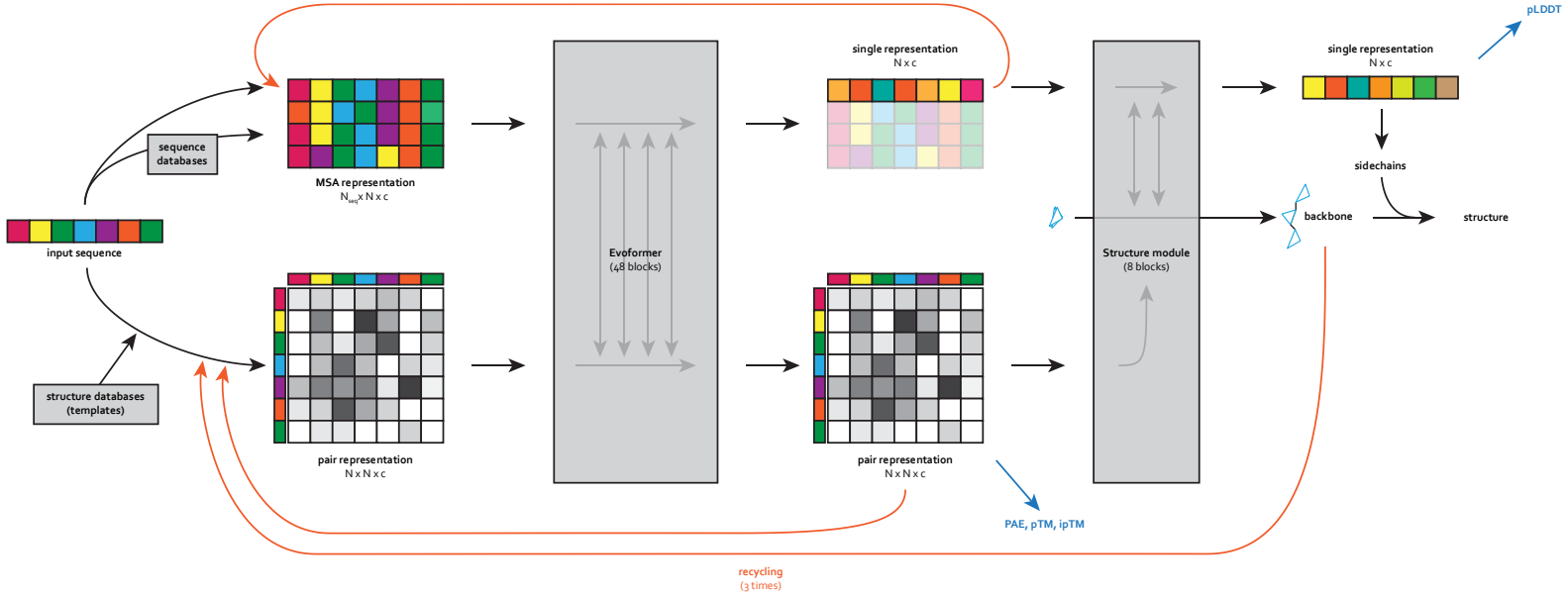
https://www.ebi.ac.uk/training/online/courses/alphafold/alphafold-3-and-alphafold-server/introducing-alphafold-3/how-does-alphafold-3-work/

Architecture

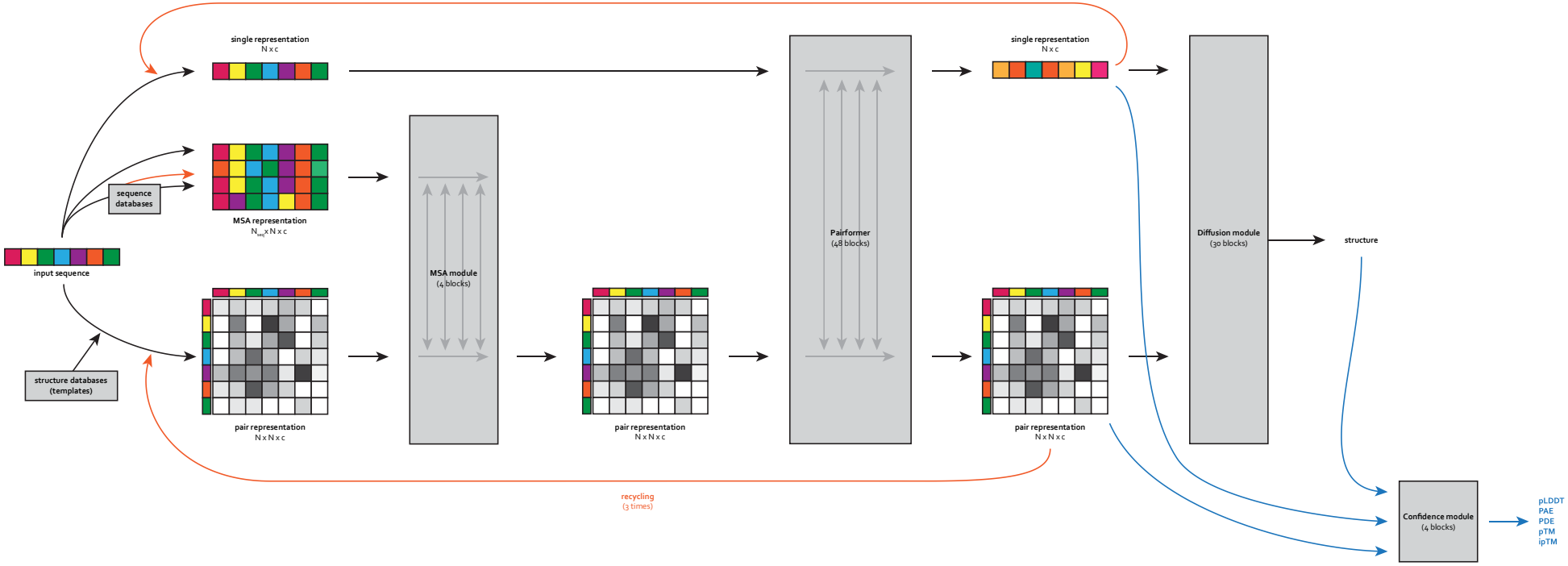
- Simplified scheme (conversions between atom and token level representations)



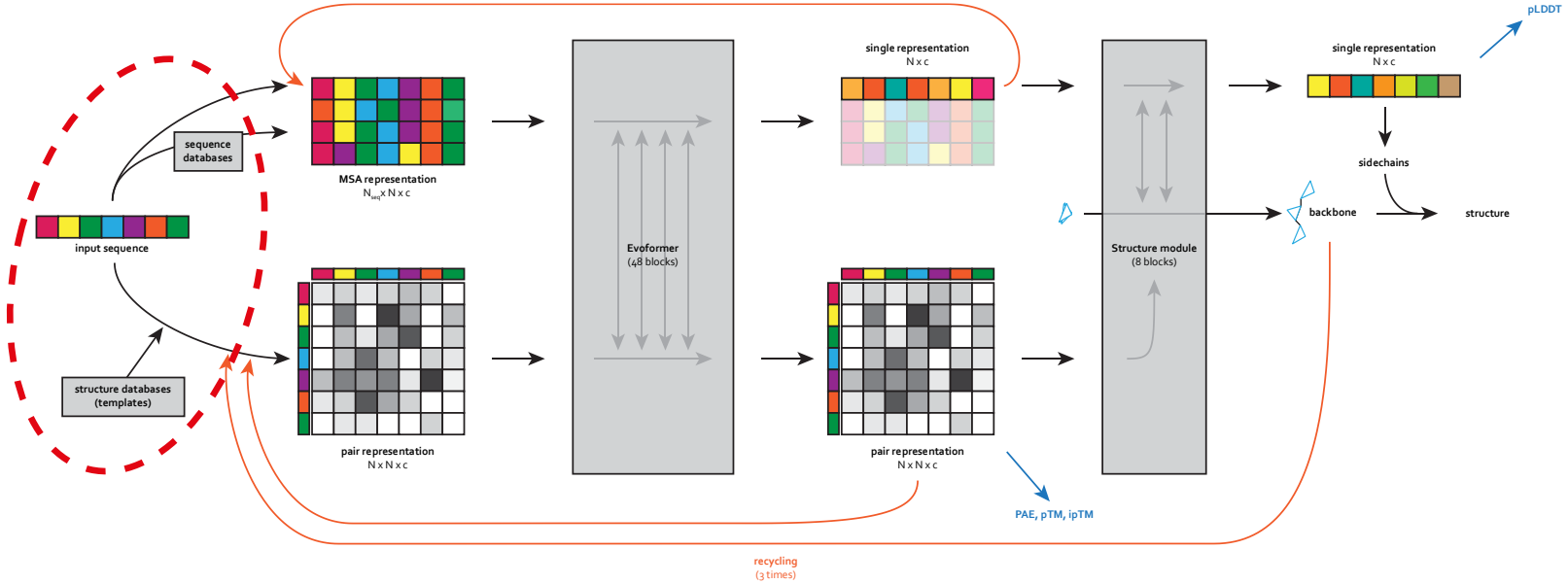
AlphaFold2



AlphaFold3

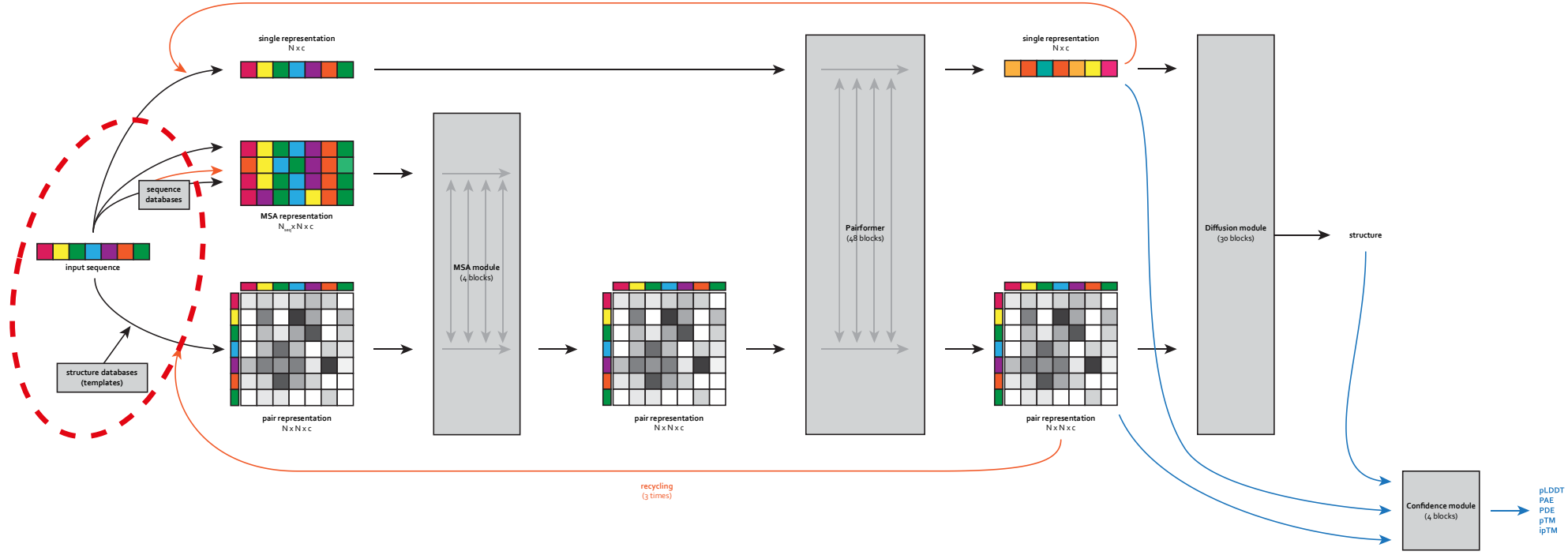


AlphaFold2

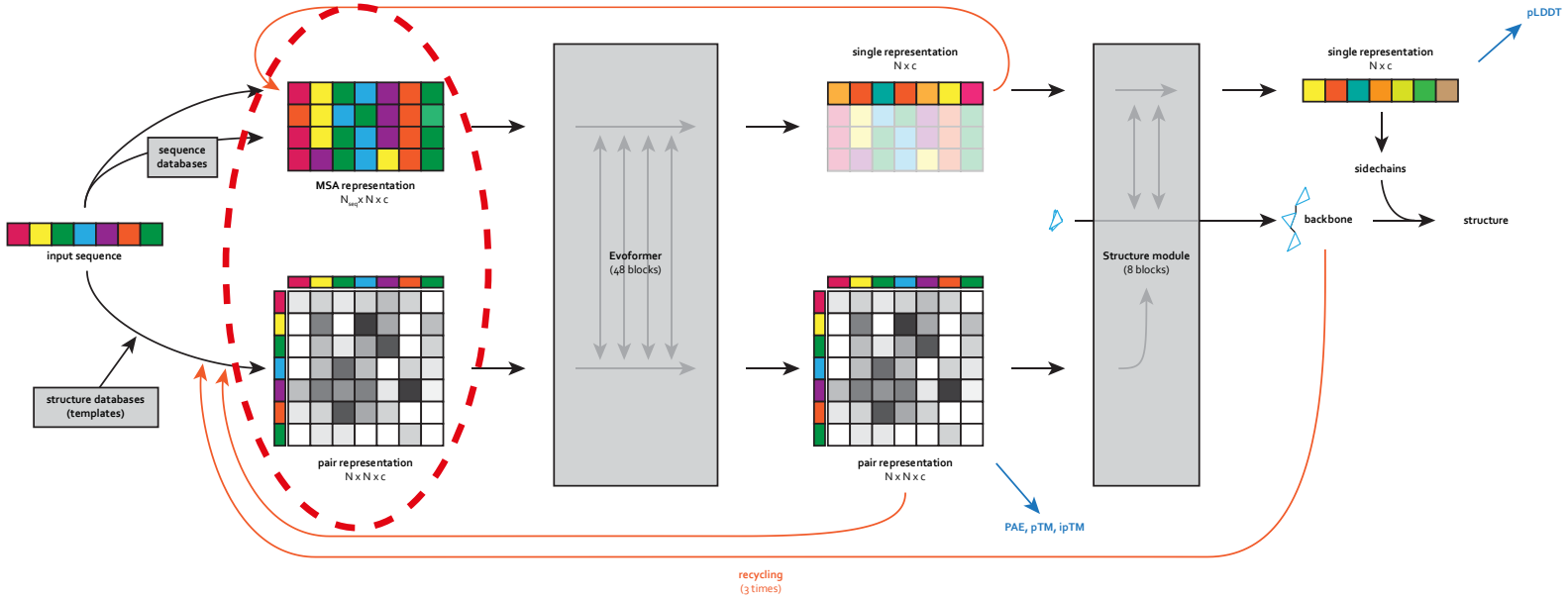


Similar inputs

AlphaFold3

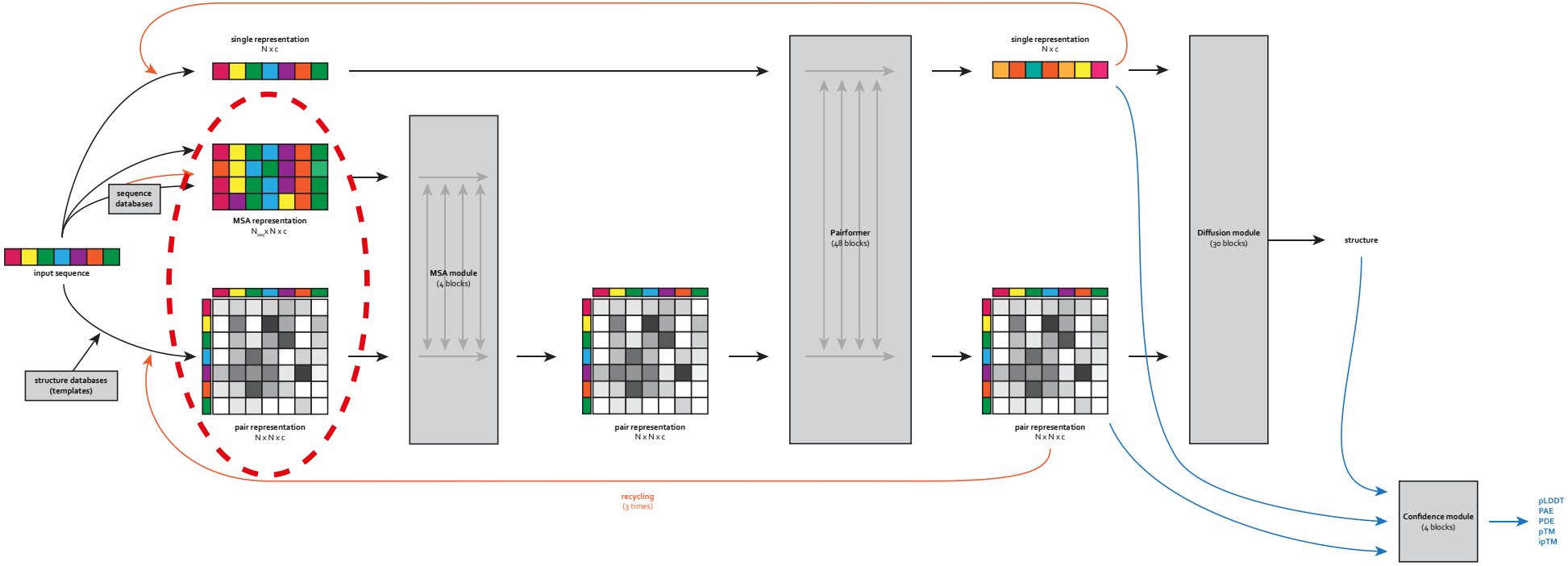


AlphaFold2

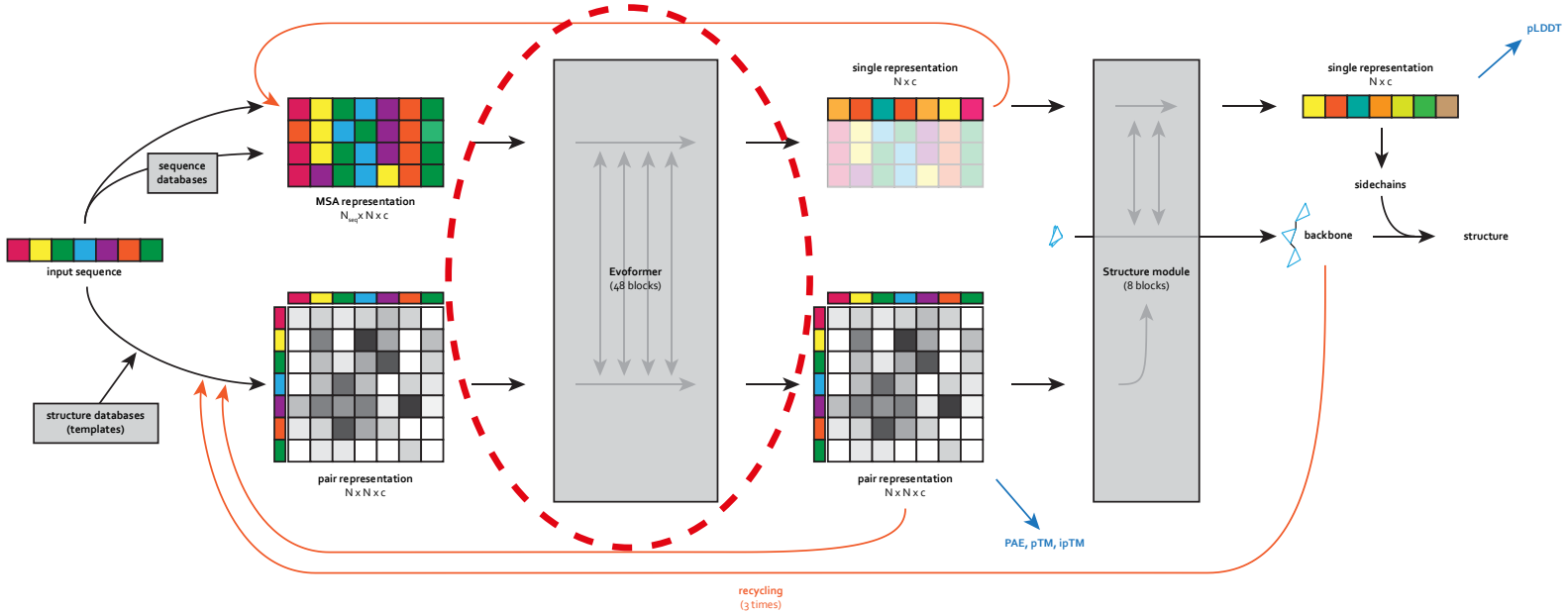


Similar representations

AlphaFold3

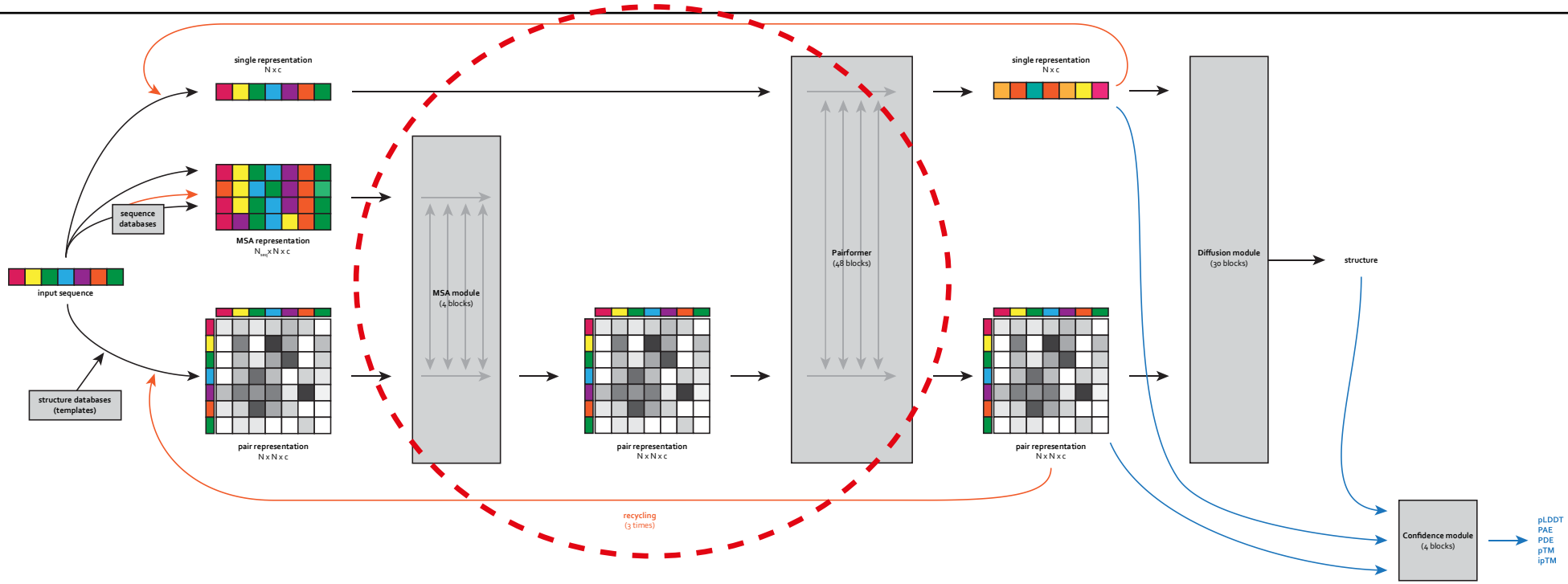


AlphaFold2

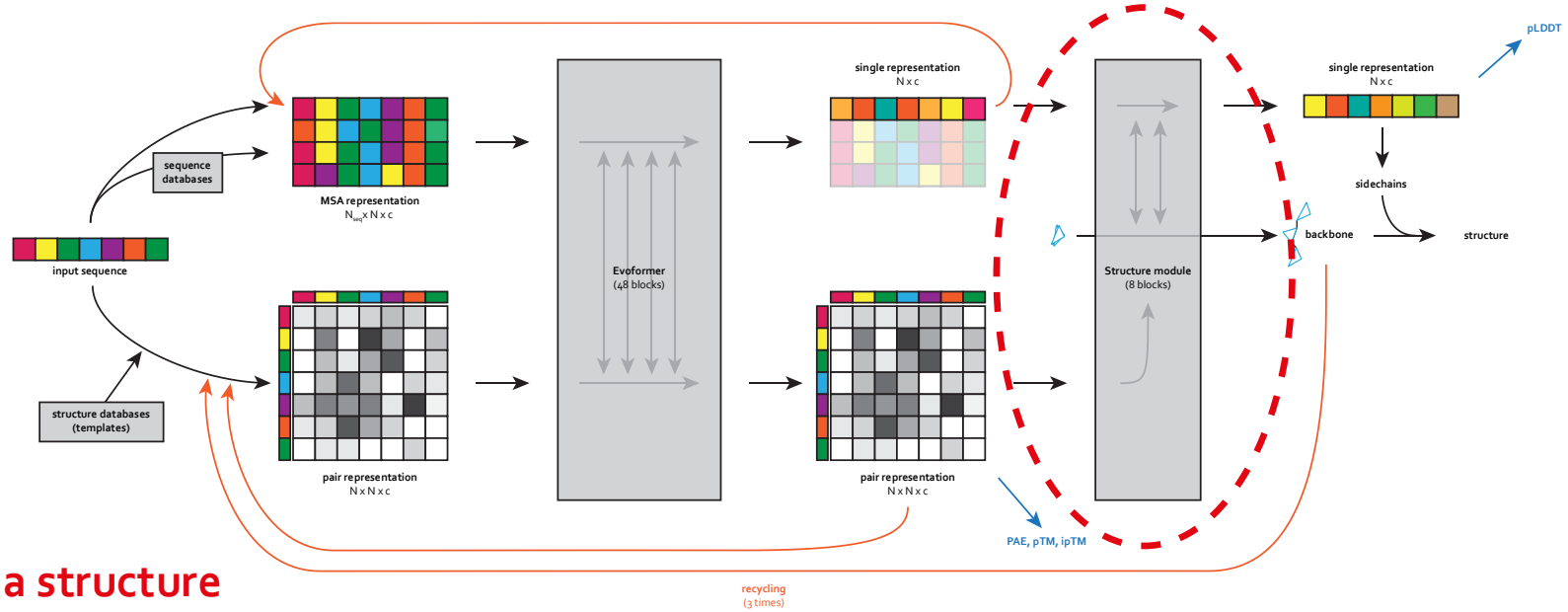


Different MSA processing

AlphaFold3

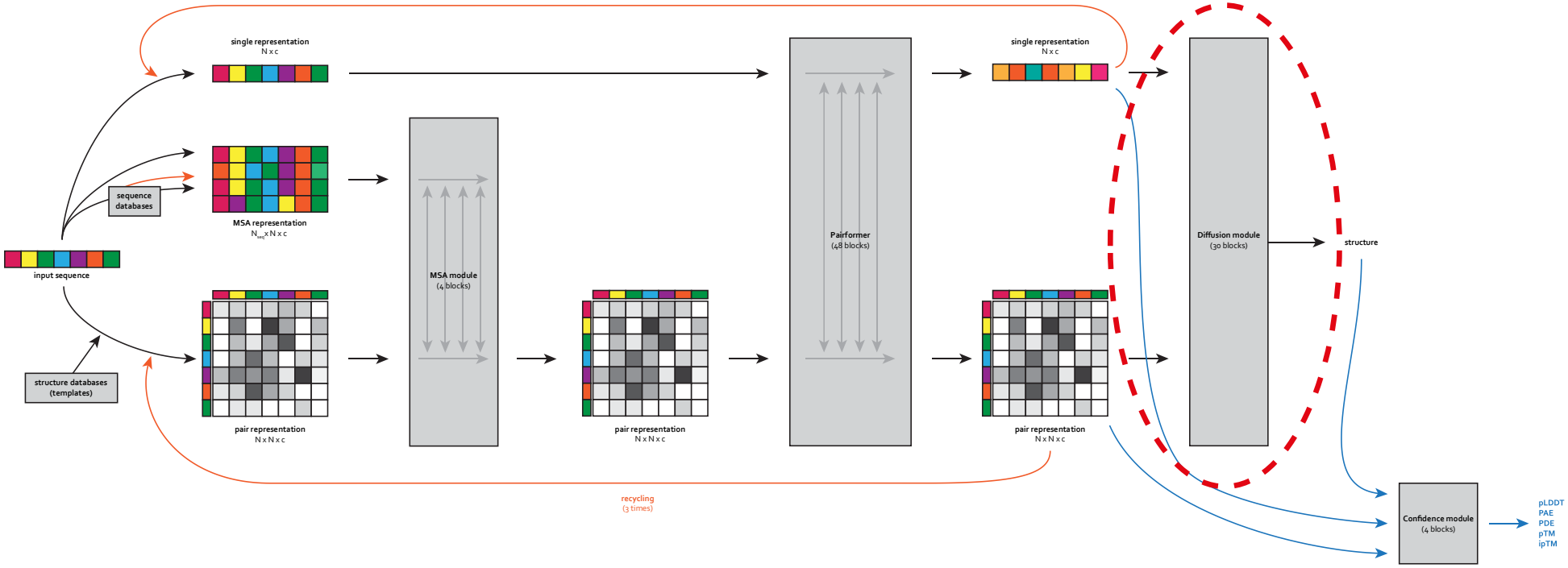


AlphaFold2



Different way of generating a structure

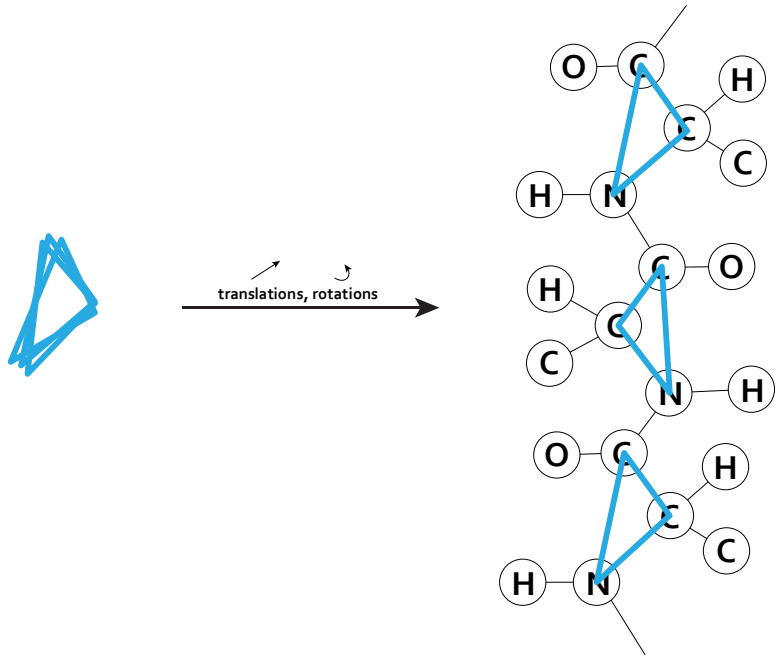
AlphaFold3



Structure/Diffusion module

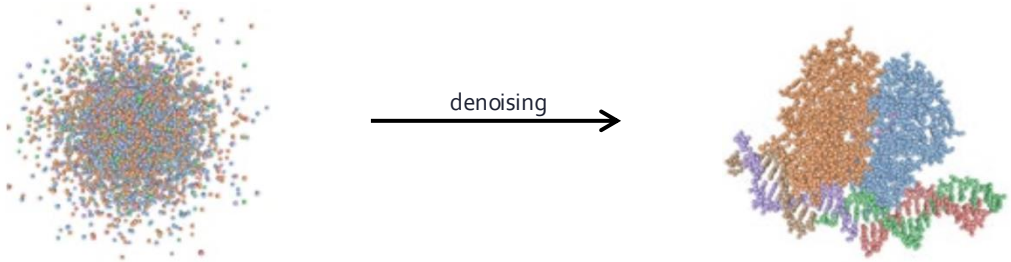
AlphaFold2
(Structure module)

- Backbone frames
- Starting at o
- Rotations, translations



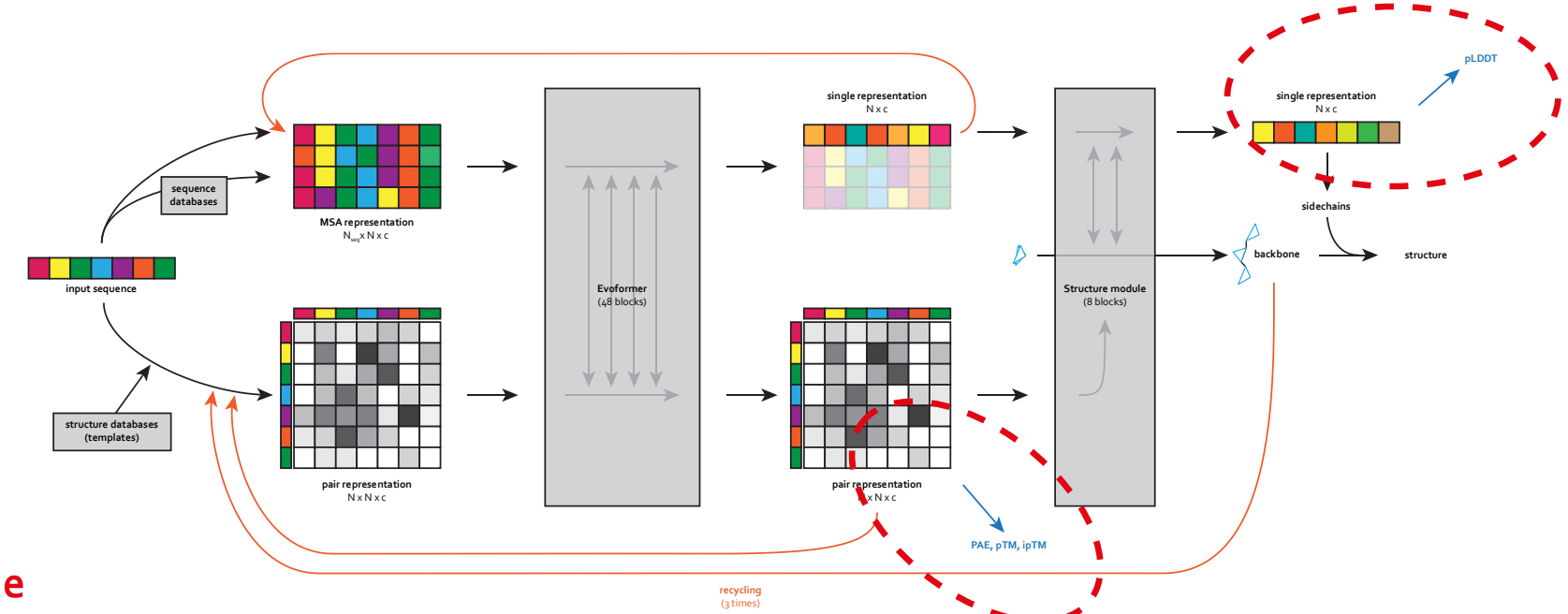
AlphaFold3
(Diffusion module)

- Coordinates of all atoms
- Starting from random coordinates
- Denoising task



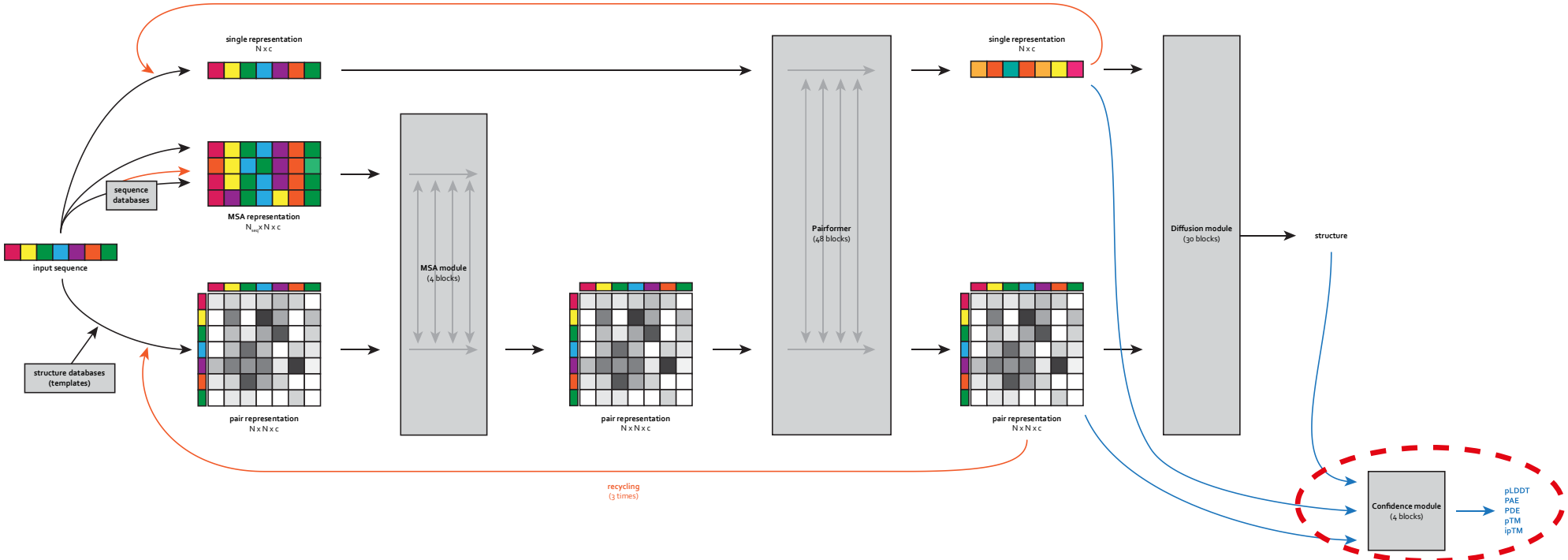
Abramson, J., Adler, J., Dunger, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*

AlphaFold2



Different confidence module

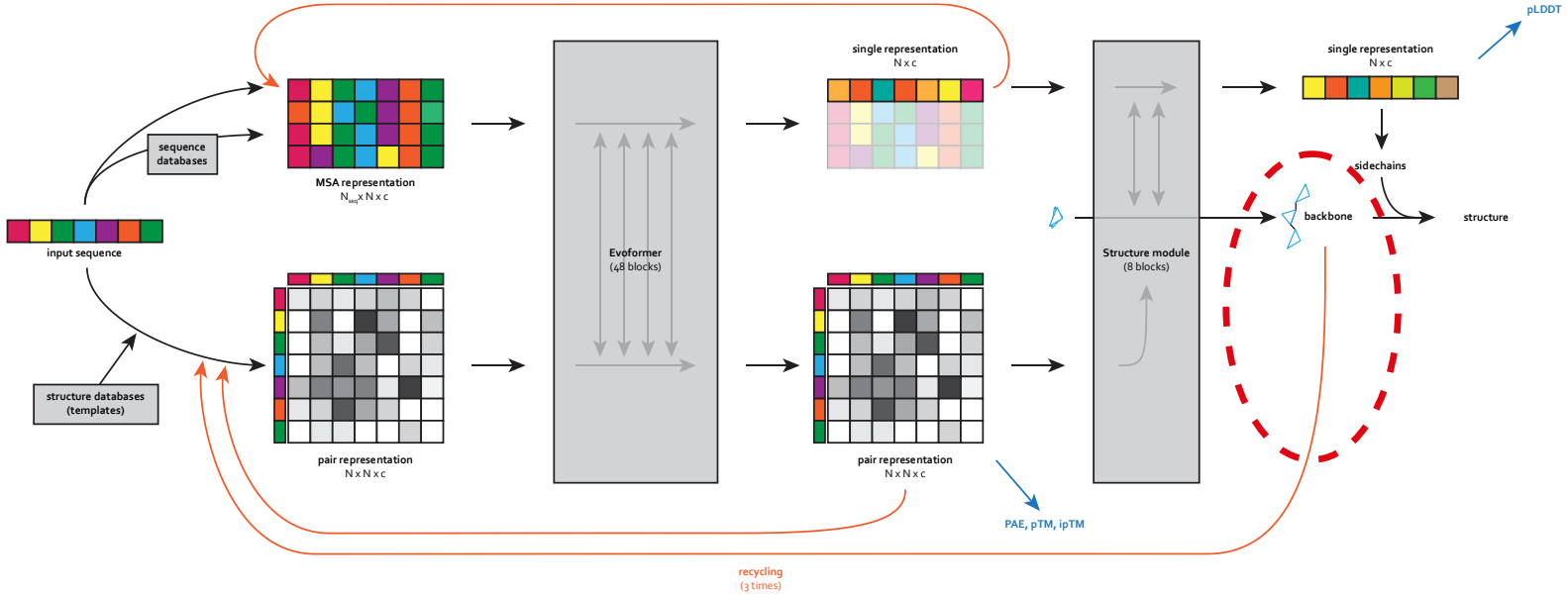
AlphaFold3



Confidence scores

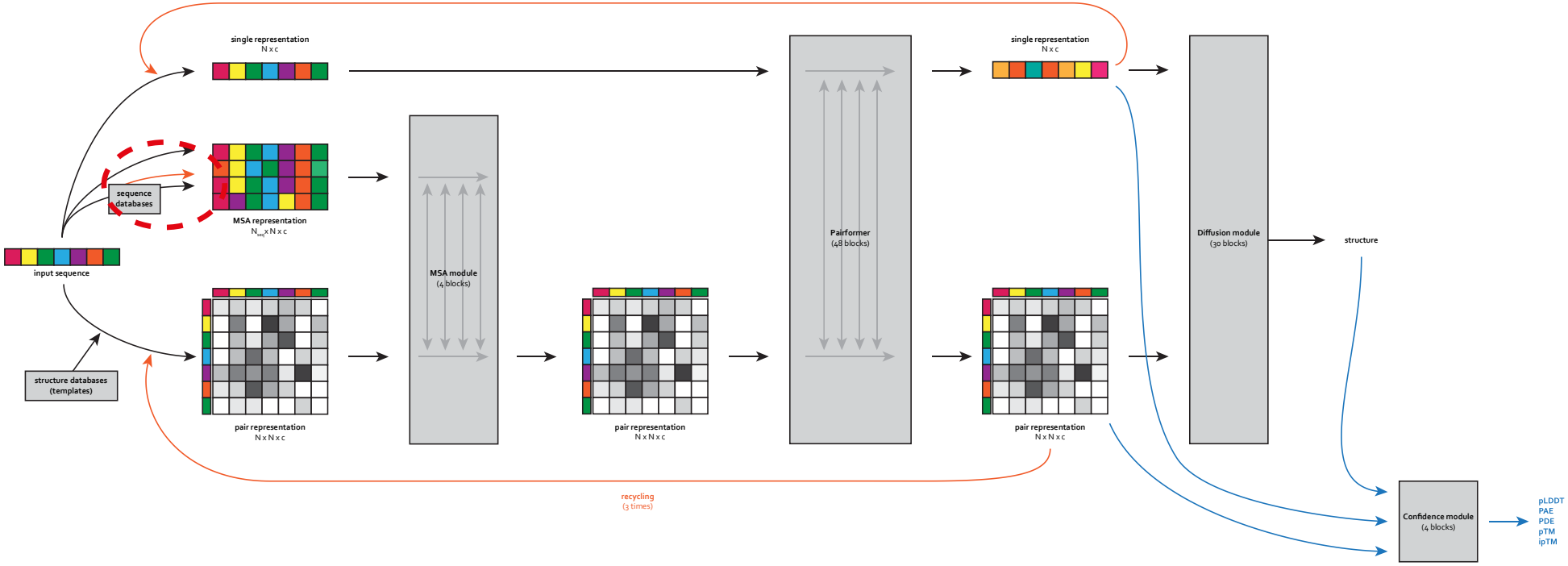
- The same metrics
- Calculated per-token
- PDE – predicted distance error
  - Predicted error in distance between two tokens (slightly different than PAE)
- Chain pair ipTM

AlphaFold2



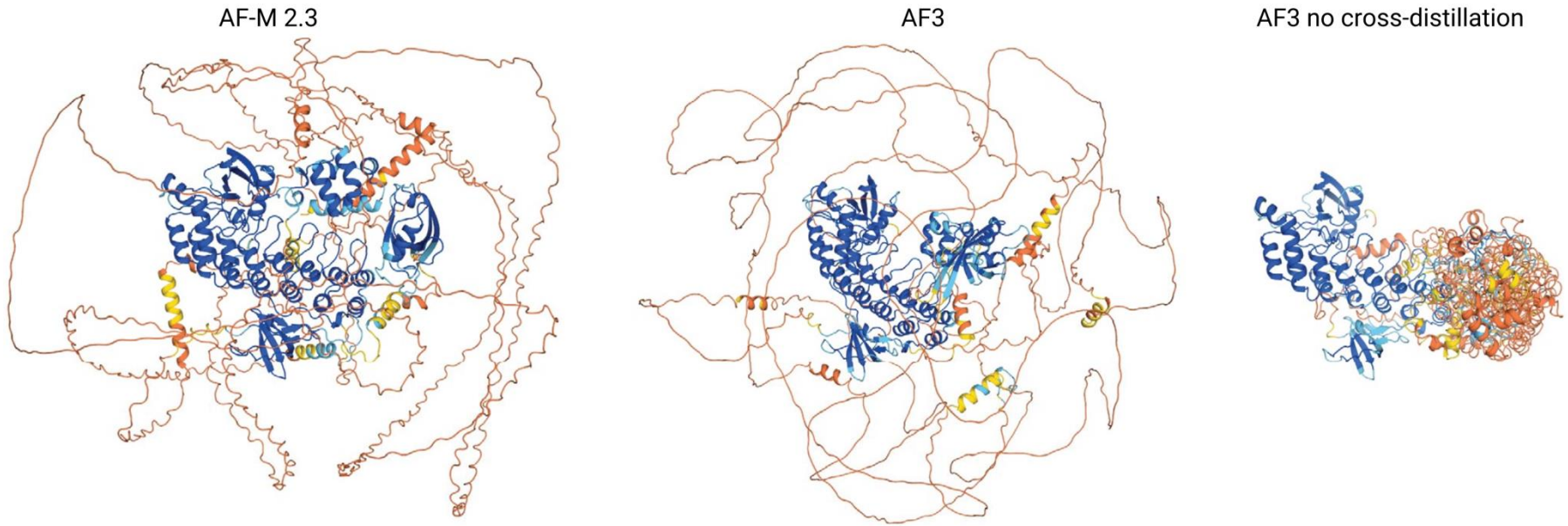
Slightly different recycling

AlphaFold3



Hallucinations - structure in unstructured regions

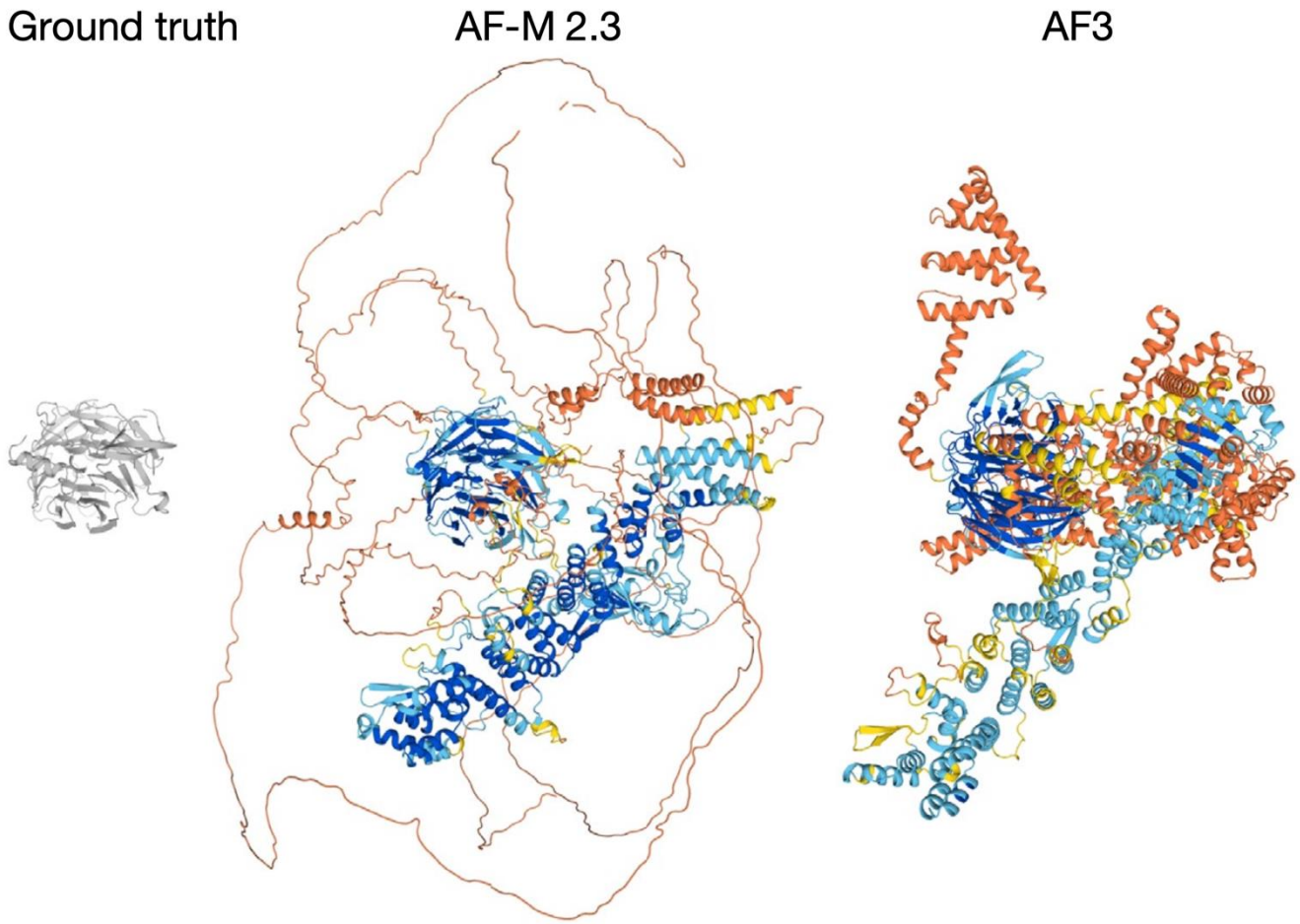
- AlphaFold2 tends to represent these regions by long extended loops (low confidence + visual)
- AlphaFold3 might invent plausible-looking structure in these regions (still with low confidence, but people didn't like it visually)
- Cross-distillation: AF3 trained on AF2 structures to teach this behaviour



Abramson, J., Adler, J., Dunger, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*

Hallucinations - structure in unstructured regions

- Hallucinations can still appear



Abramson, J., Adler, J., Dunger, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*

How good is it?

Proteins:

similar to AlphaFold2

Protein-protein complexes:

similar to AlphaFold2

DNA:

not great, but the best for now

RNA:

there are now better tools

How good is it?

Peptides:

work well

Orphan proteins:

not great (similar as AlphaFold2)

Antibodies:

work well, but require many samples (~500)

→ also check ESMFold2

Viruses:

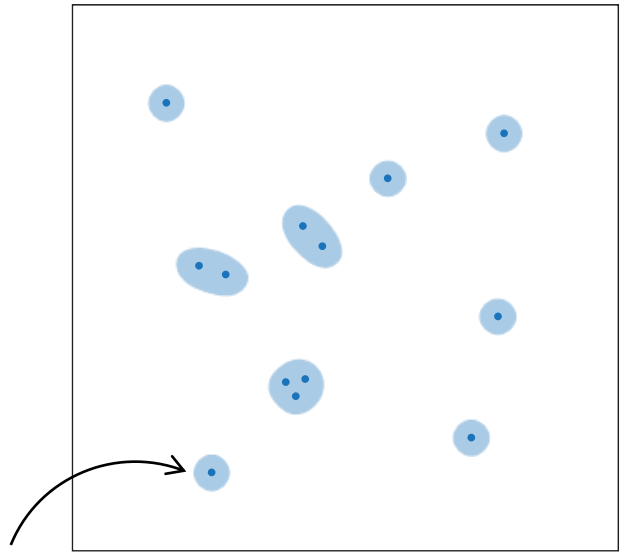
reasonable results for host–pathogen interactions (including viruses)[1]

[1] https://pmc.ncbi.nlm.nih.gov/articles/PMC12258915/

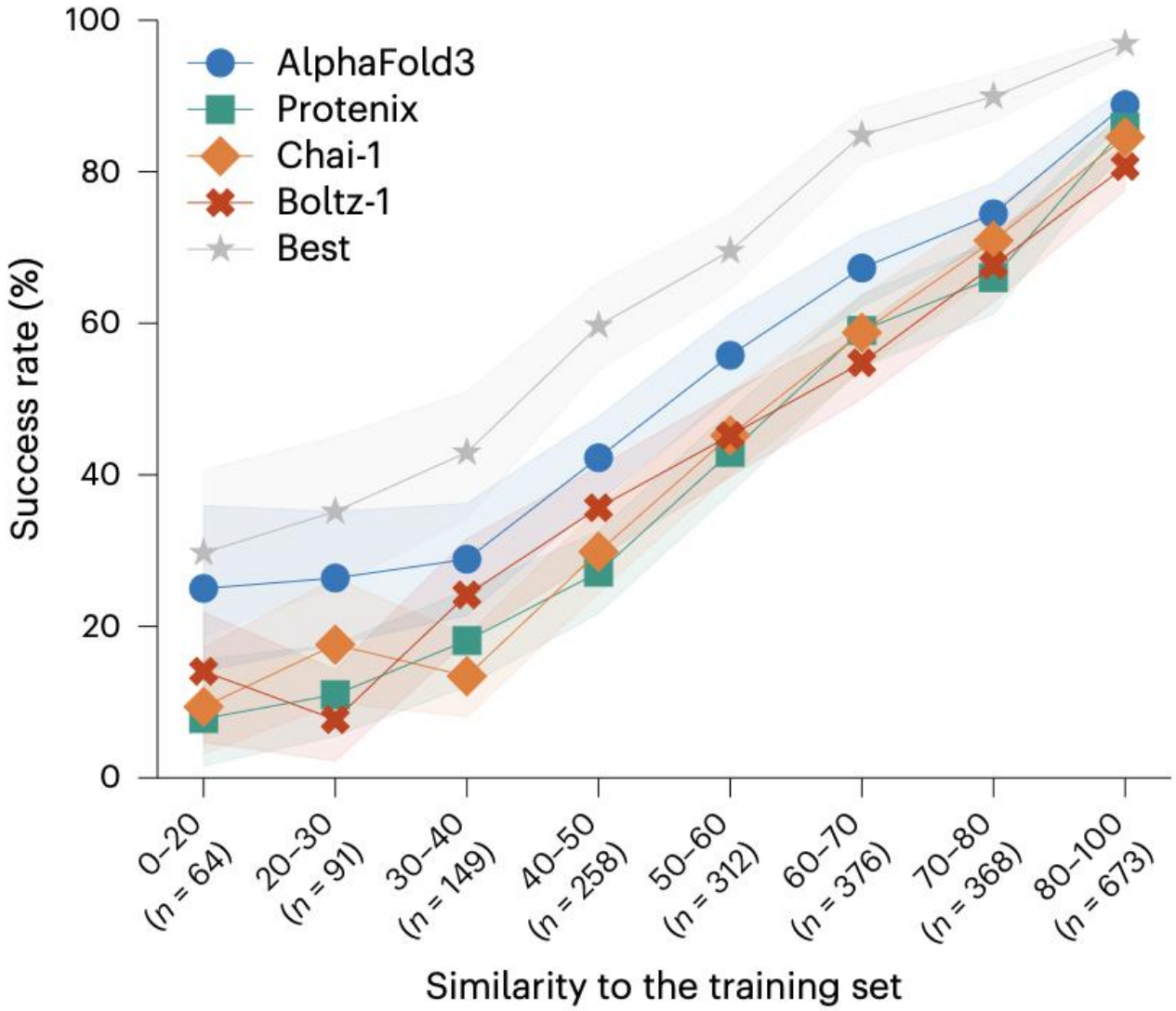
[2] https://colab.research.google.com/github/Biohub/esm/blob/main/cookbook/tutorials/esmc_mutation_scoring.ipynb

Protein-ligand complexes

- Poor performance
- No such thing as MSA for protein-ligand interaction



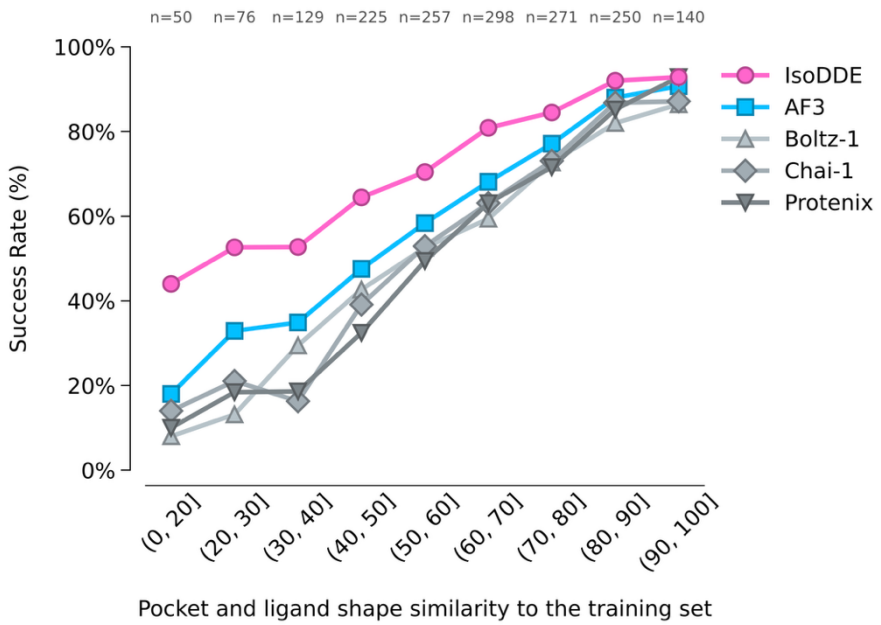
Known **structures**

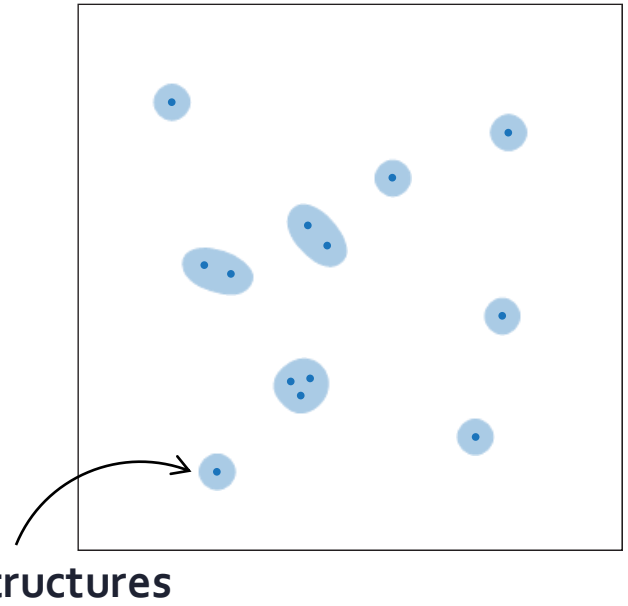


Škrinjar, P., Eberhardt, J., ..., Durairaj, J. Evaluating generalization in protein–ligand cofolding methods. *Nature Structure & Molecular Biology*

Improving the performance on protein-ligand complexes

- More experimental structure data to train on[1]
  - Will take years
- Synthetic data[2]
  - Generated using physics-based methods

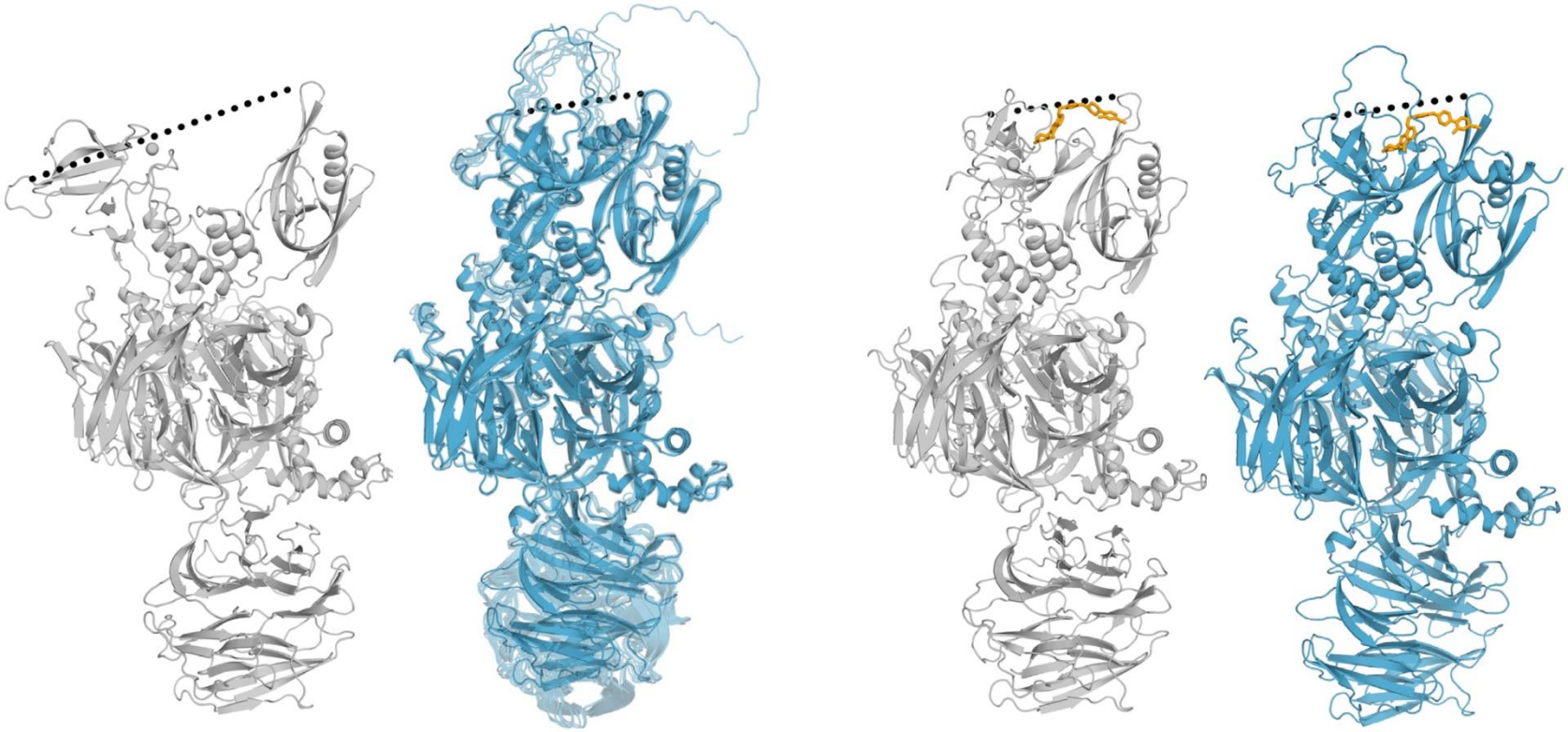




[1] https://openbind.uk
[2] https://arxiv.org/abs/2510.24670v2
Figure: https://storage.googleapis.com/isomorphiclabs-website-public-artifacts/isodde_technical_report.pdf

Limitations

- Conformation coverage is limited

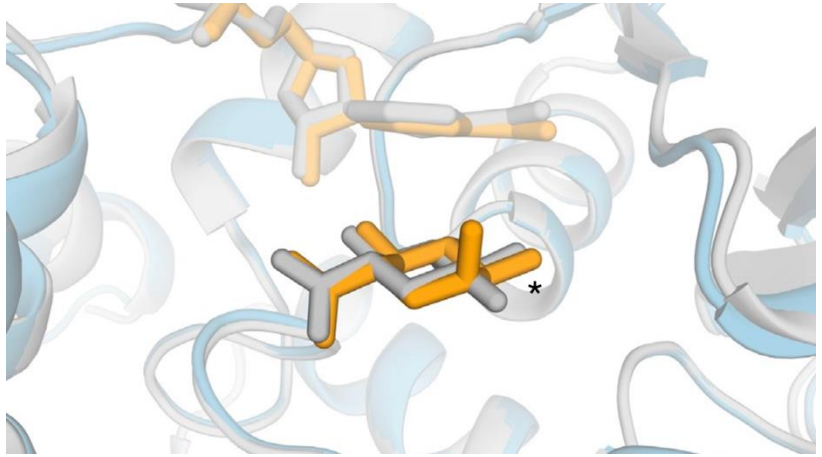


Ground-truth (no ligand) | AF3 prediction (no ligand) | Ground-truth (ligand-bound) | AF3 prediction (ligand-bound)

Abramson, J., Adler, J., Dunger, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*

Limitations

- Stereochemistry (chirality) violations
  - Not the one we specified as the input



- Confidence scores are overestimated
  - Users prefer AlphaFold2 metrics for some applications

Abramson, J., Adler, J., Dunger, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*

Which one should I use? (AF2 or AF3)

- AF2 is freely available for both academic and commercial use
  AF3 is available for non-commercial projects only
- Custom MSA/template mode is easier in AF3 compared to AF2 (for the local version)
- Online versions:
  ColabFold (AlphaFold 2) allows you to adjust parameters;
  the AlphaFold Server does not offer as much flexibility;
- AF2 is better for de novo proteins (due to better confidence metrics);
- Use AF3 if you want to model ligands, DNA, RNA, or Post-translational modification (PTM) – possible but not great
- AF3 is better for antibody-antigen complex prediction

How to access

- Source code
  - Hard
  - Licence limitations

- AlphaFold Server
  - Online
  - No SMILES input

- A lot of “clones” exist, some include extra features:
  - Boltz (Boltz-1, Boltz-1x, Boltz-2, Boltz-2x)
  - Chai-1
  - …

Swiss Institute of Bioinformatics

DAY 2, PART 3

AlphaFold Server tutorial

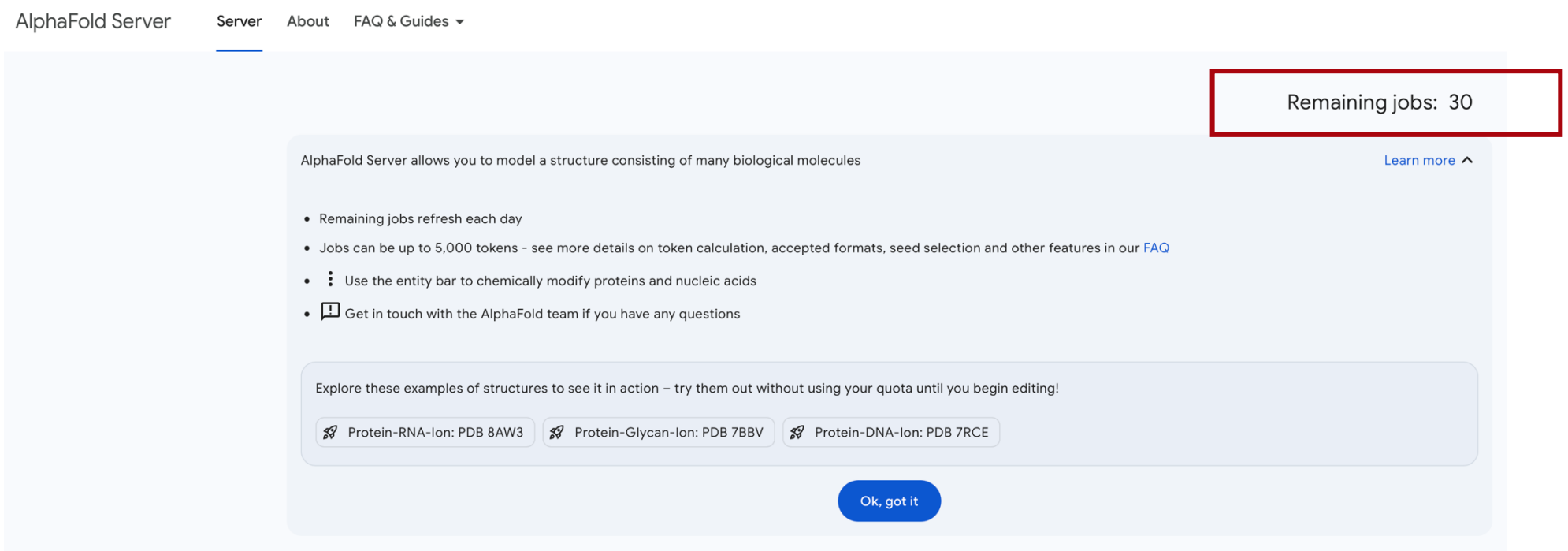
Diana Rapota, Rok Breznikar, Janani Durairaj

23-24 June 2026

elixir SWITZERLAND

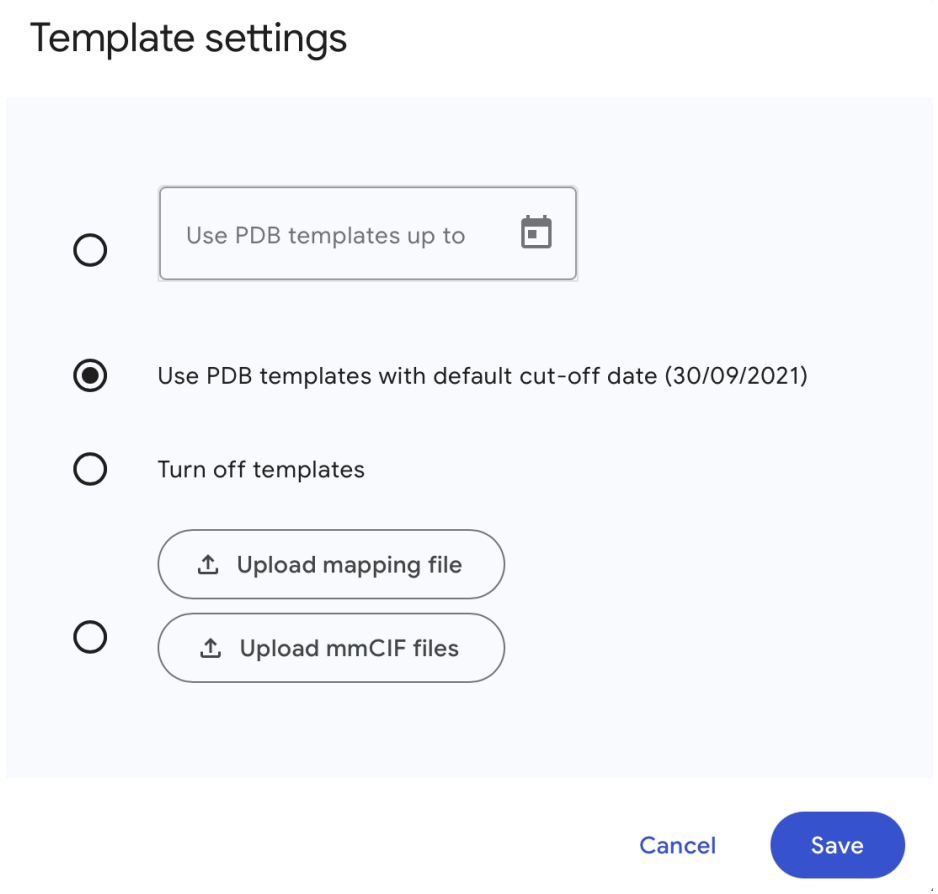
AlphaFold Server

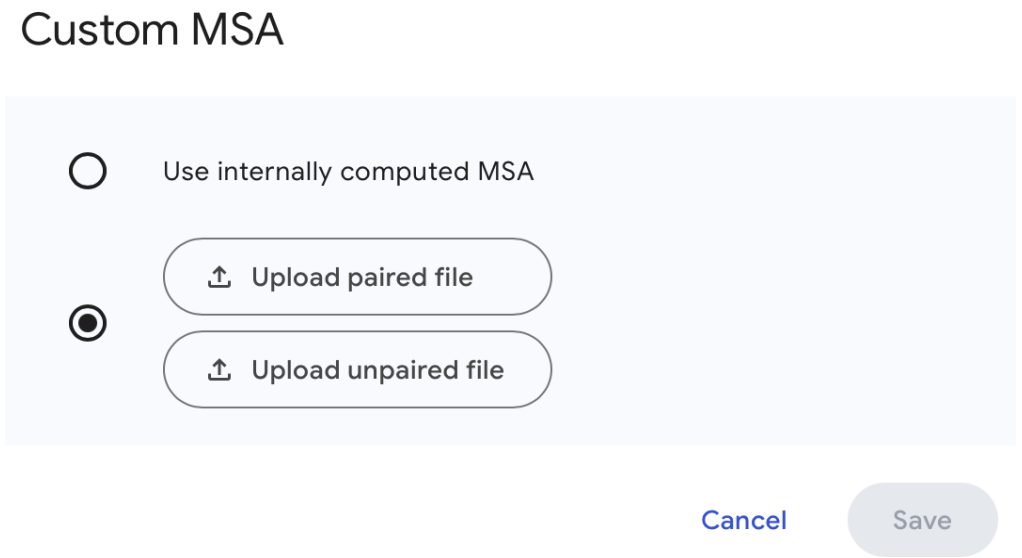
- Free and you do not need to install any software on your computer
- Supports modelling of multi-entity complexes (proteins, DNA, RNA, ligands, ions), but exposes only a limited subset of the full AlphaFold 3 (check FAQs)
- The minimum sequence length is four amino acids or four nucleotides
- Each modelling job is limited to 5,000 tokens
- Each user is limited to 30 jobs per day



AlphaFold Server: Settings

- There are options for custom templates and MSA
- You can specify a cut-off date for the PDB templates to be used. The cut-off date for the template is usually used if you don't want templates that AF3 was trained on (training cut-off date: 2021-09-30).



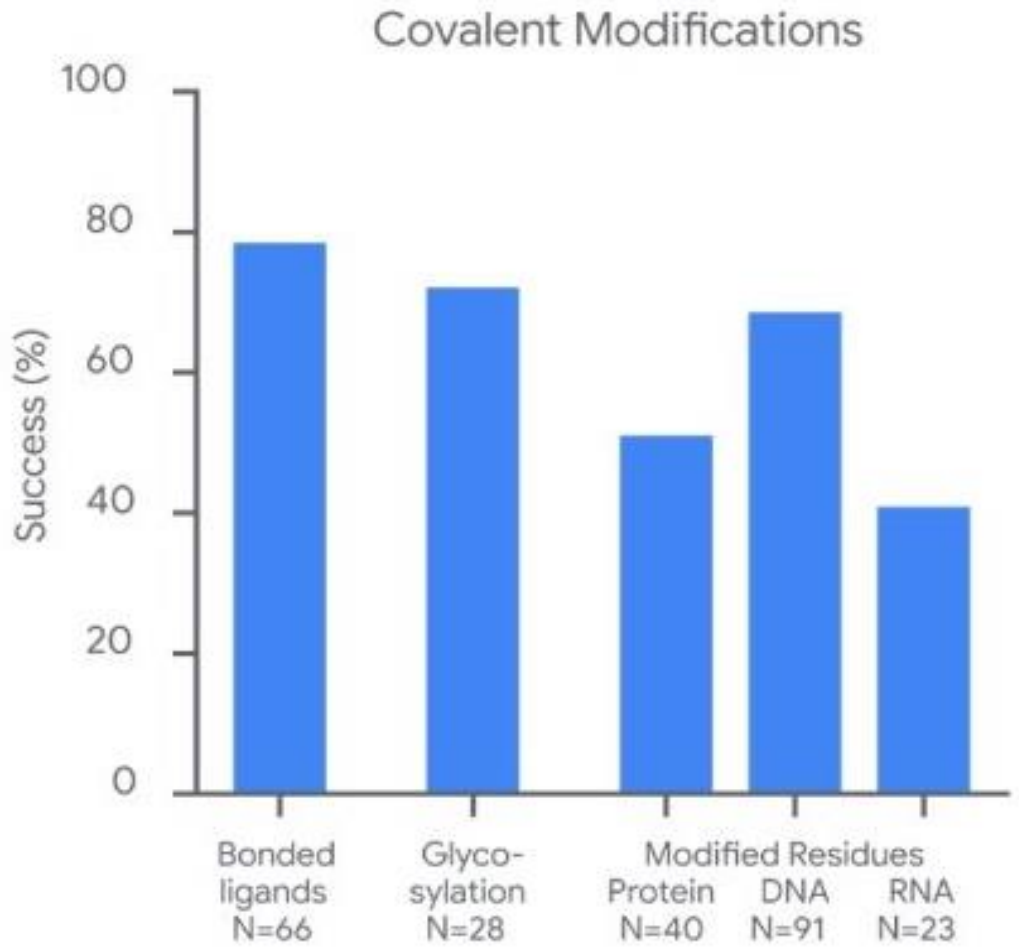


AlphaFold Server: Settings

Adding modifications

- Multiple modifications to different residues one at a time (for Glycosylated proteins see AlphaFold Server FAQ)

- Modifications for DNA and RNA



Abramson et al, 2024

AlphaFold Server: Advanced Settings

- **JSON job submission** for automatic job generation (useful for screening experiments, up to 100 jobs per file; you are allowed up to 500 saved drafts)

- **Sampling multiple seeds** (20 seeds is usually enough to get a better prediction)

- **Reproducing jobs** (just use the same input and seed number, which can be found in fold_<job_name>_job_request.json - contains all the information required for reproduction)

AlphaFold Server: Output

- Five .cif files named **fold_<job_name>_model_<N>cif** contain predicted structures in the mmCIF format (can be viewed in any molecular viewer);
- Five .json files named **fold_<job_name>_summary_confidences_<N>.json** contain summaries of the confidence metrics for the predictions;
- Five .json files named **fold_<job_name>_full_data_<N>.json** contain detailed confidence metrics (per-atom pLDDT, per-token contact prob matrix, per-token pae etc.)
- A file named **fold_<job_name>_job_request.json** contains the inputs of the modelling job and could be used to re-run the job;
- A file named **terms_of_use.md**. This is a legal document detailing the terms of use for the predictions.

where N is the rank of the predicted structure, structures are ranked from 0 to 4, where 0 has the highest confidence

AlphaFold Server: Confidence metrics

JSON files with **summary** contain:

- **chain_iptm:** the average confidence (ipTM) in the interfaces between each chain and all other chains, can be used for ranking predicted structures for a specific chain;
- **chain_pair_iptm:** pairwise ipTM scores, can be used for ranking predictions of a structure by the accuracy of a specific interface between two chains;
- **chain_pair_pae_min:** element (i, j) of the array contains the lowest PAE value across rows restricted to chain i and columns restricted to chain j. This has been found to correlate with whether or not two chains interact;
- **chain_ptm:** pTM restricted to chain i;
- **fraction_disordered:** range 0-1 that indicates what fraction of the prediction structure is disordered;
- **has_clash:** 1.0 if more than 50% of a chain's atoms clash, or if a chain has more than 100 clashing atoms; 0.0 otherwise.
- **iptm:** ipTM for all interfaces in the structure;
- **ptm:** pTM-score for the full structure;
- **ranking_score:** 0.8 × ipTM + 0.2 × pTM + 0.5 × disorder – 100 × has_clash, range from –100 to 1.5

AlphaFold Server: Confidence metrics

JSON files with **full outputs** contain:

- **atom_chain_ids:** the chain IDs corresponding to each atom in the prediction;
- **atom_plddts:** A [num_atoms] array, element i indicates the pLDDT **for atom** i in the prediction;
- **contact_probs:** A square [num_tokens, num_tokens] array. Element (i, j) indicates the predicted probability that **token i** and **token** j are in contact;
- **pae:** A square [num_tokens, num_tokens] array. Element (i, j) indicates the PAE in the position of **token** j, when the prediction is aligned to the ground truth using the frame of **token** I;
- **token_chain_ids:** A [num_tokens] array indicating the chain IDs corresponding to each token in the prediction;
- **token_res_ids:** A [num_res] array.