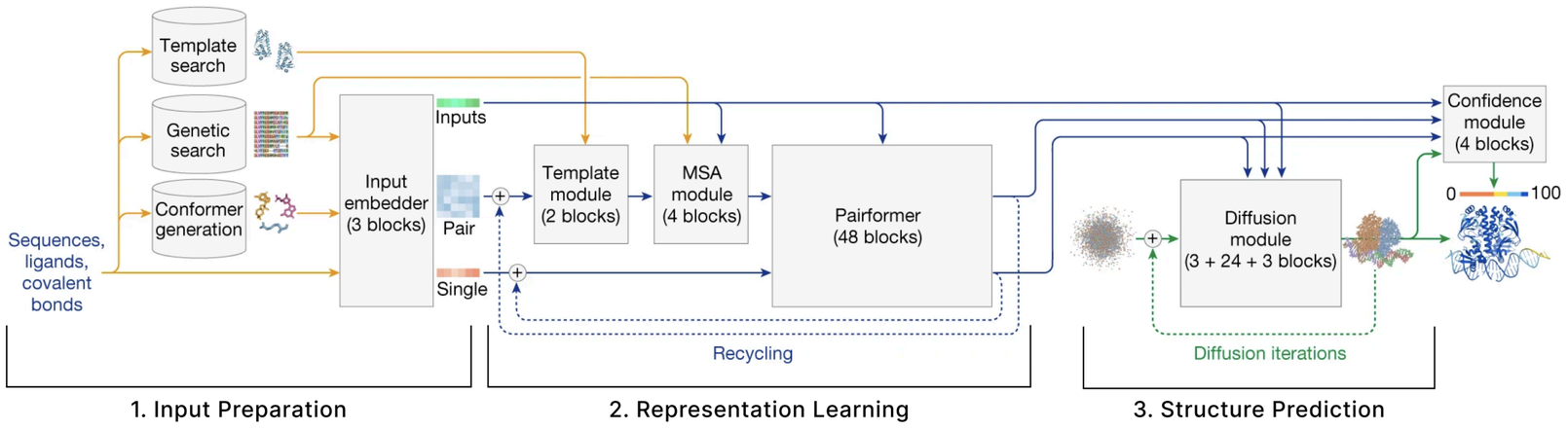
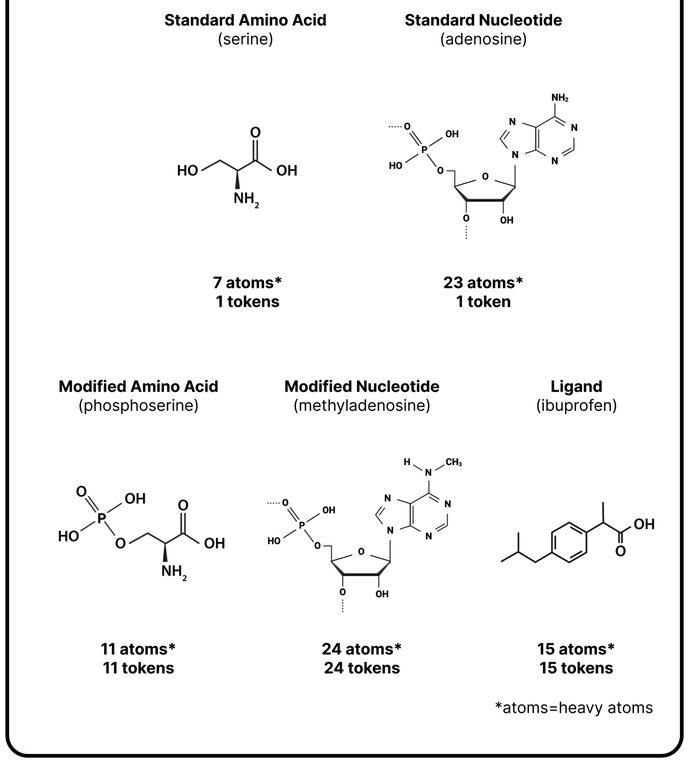
# Accurate structure prediction of biomolecular interactions with AlphaFold 3

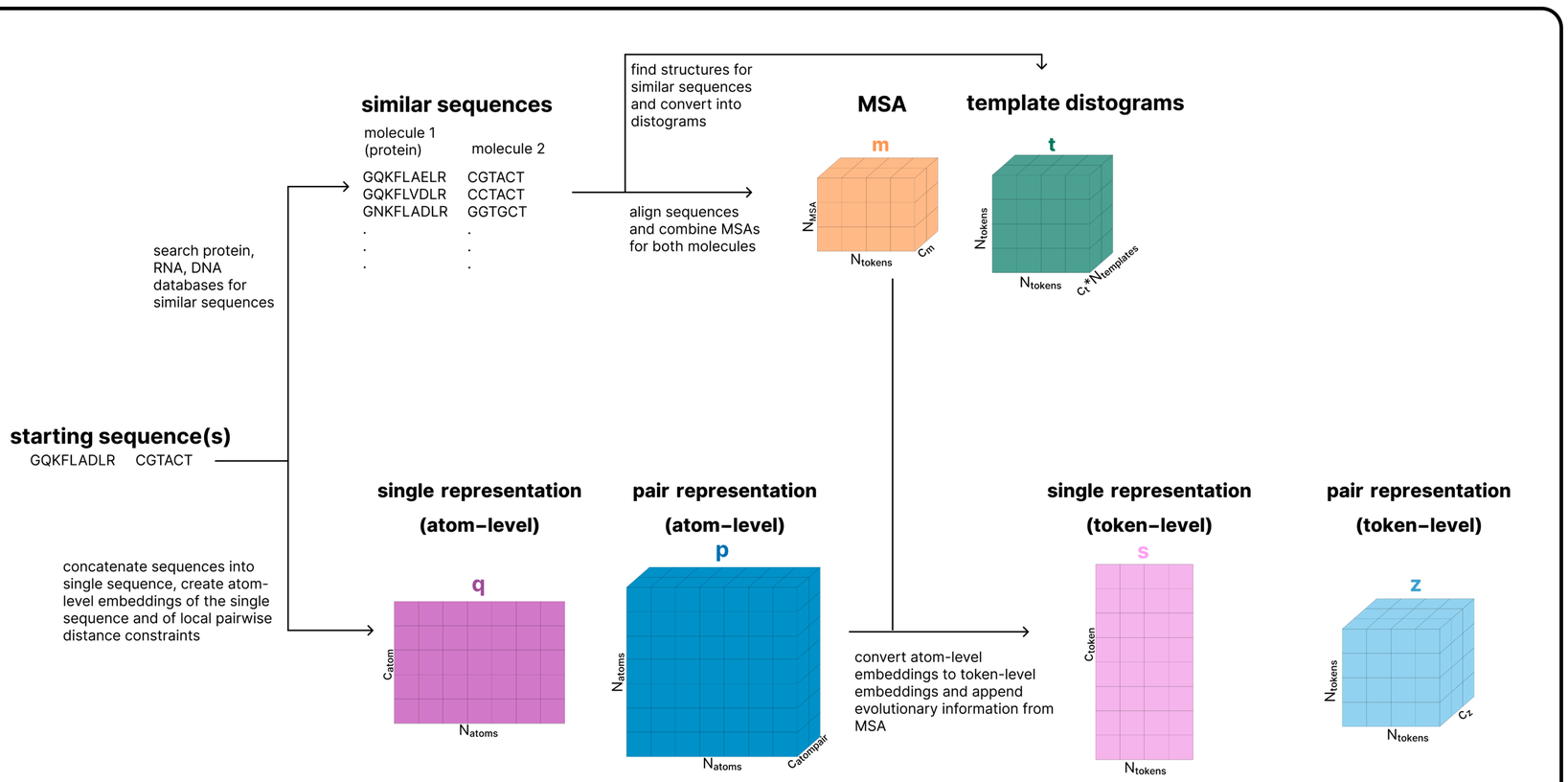
## Architecture Overview



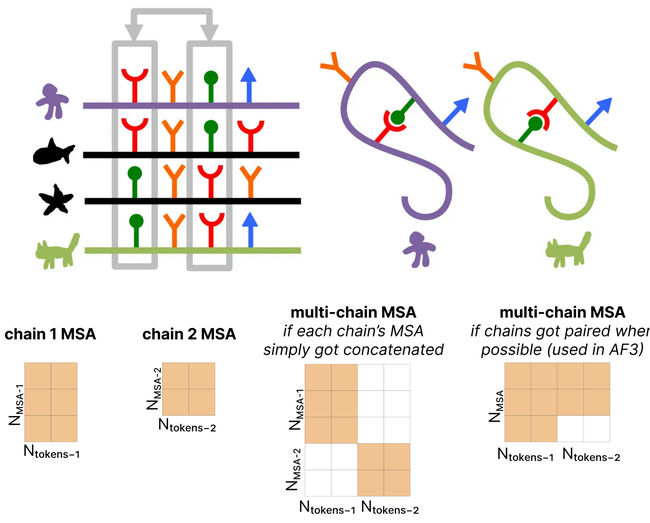

## Tokenization



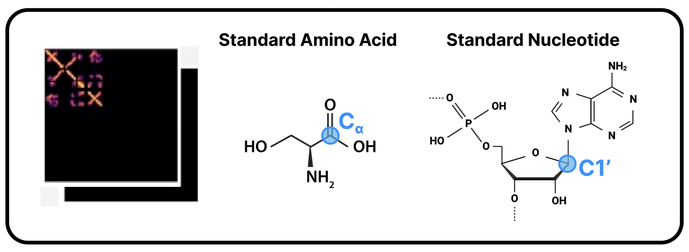

## Input Preparation Pipeline



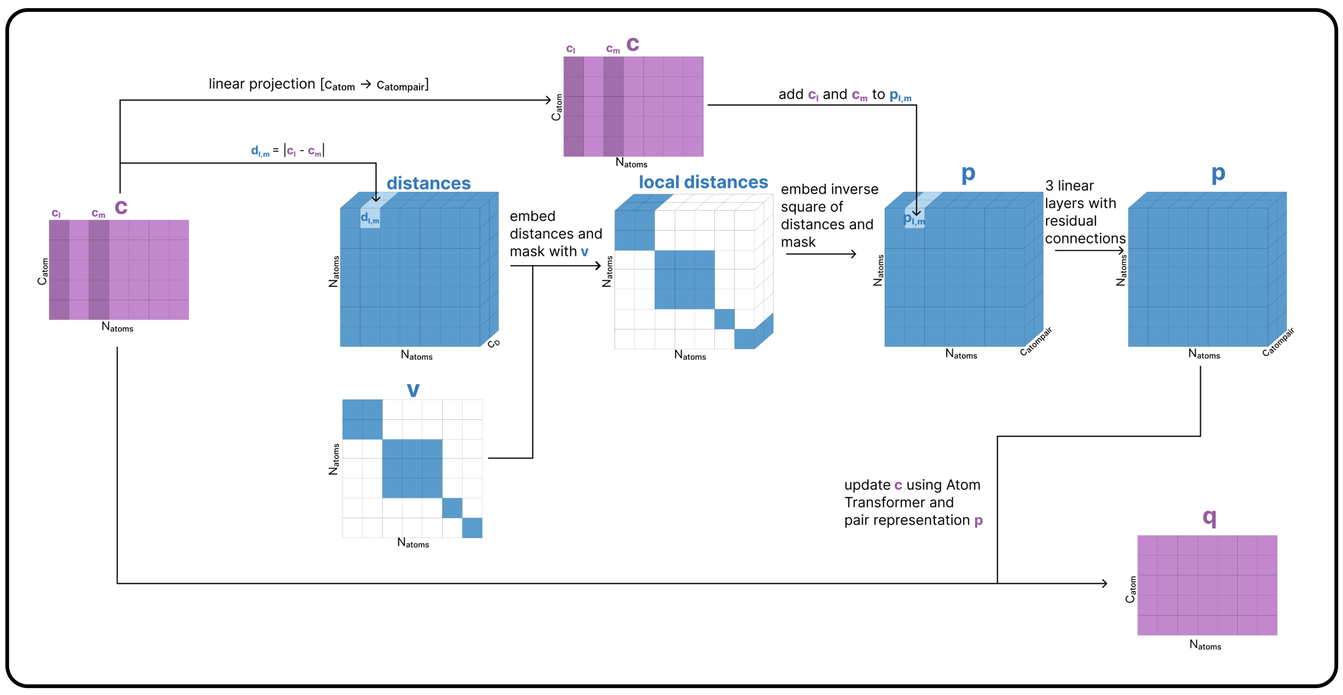

## Multiple Sequence Alignment (MSA)



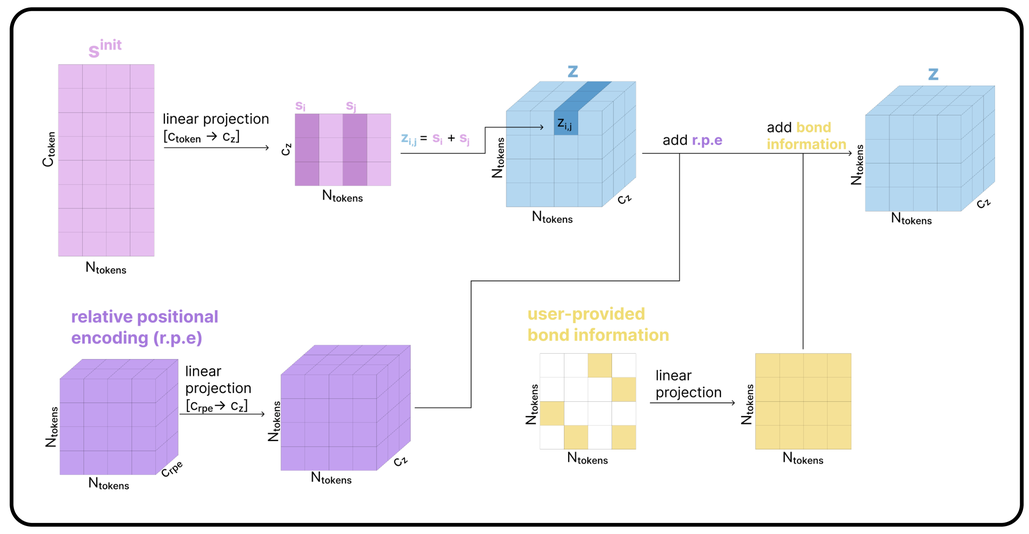

## Templates



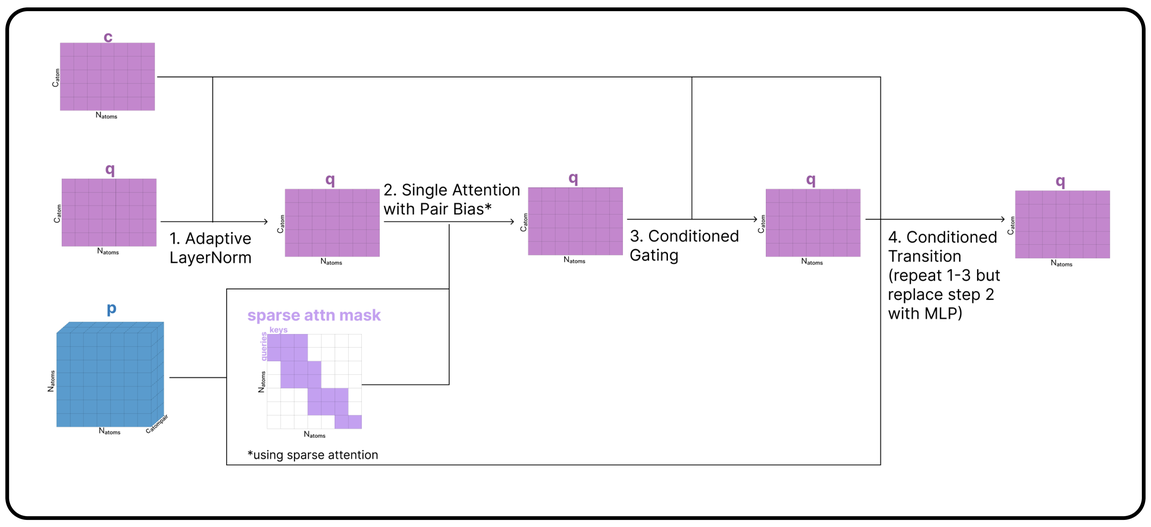

## Create Atom-Level Representations



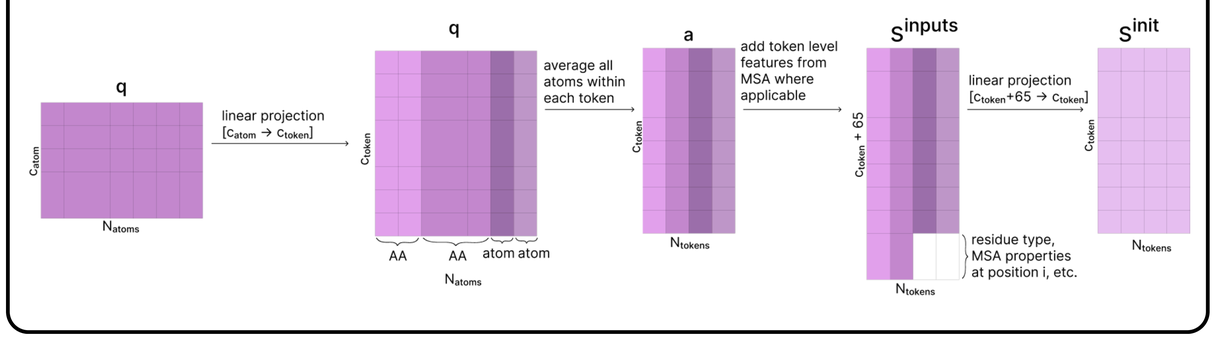

## Create Token-Level Representations




## Atom-Level Transformer

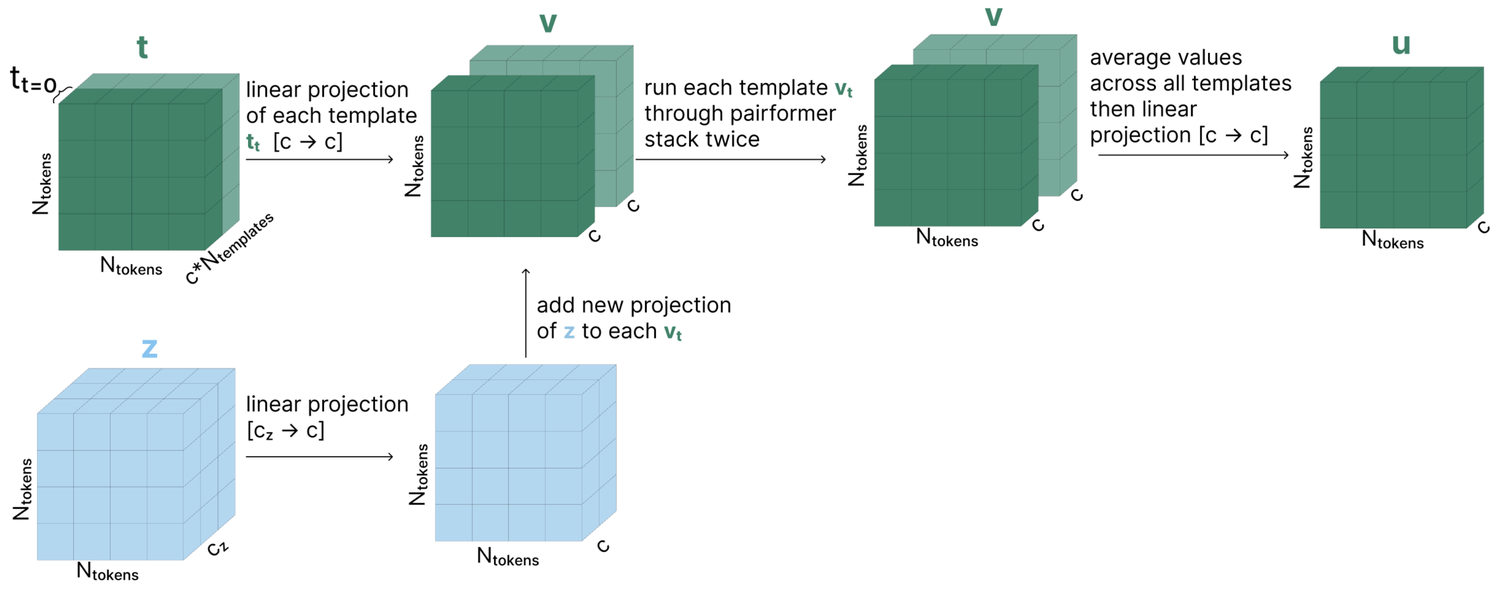



## Create Token-Level Single Sequence Representations

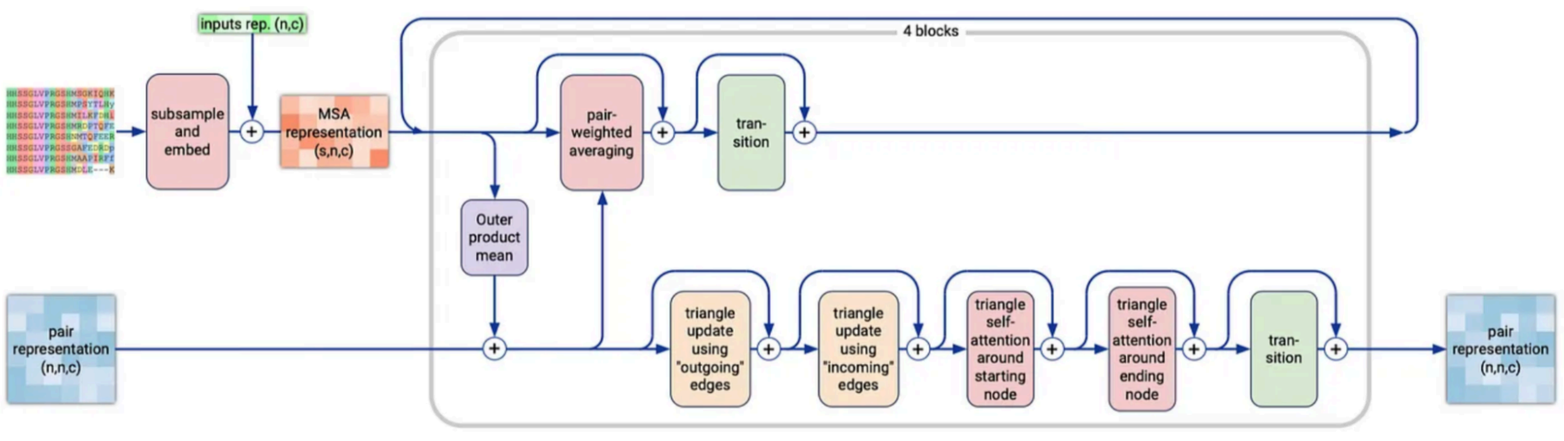


# Accurate structure prediction of biomolecular interactions with AlphaFold 3

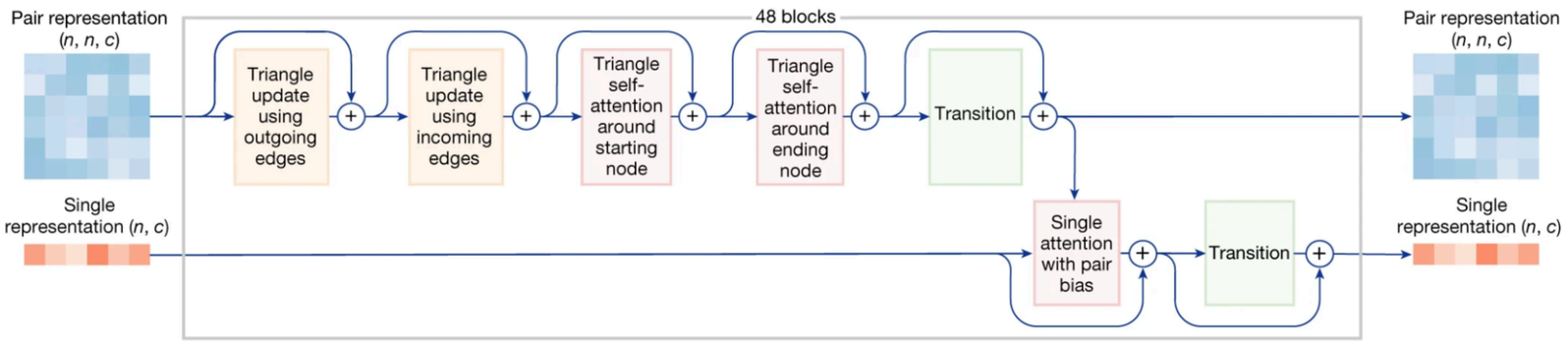
## Template Module



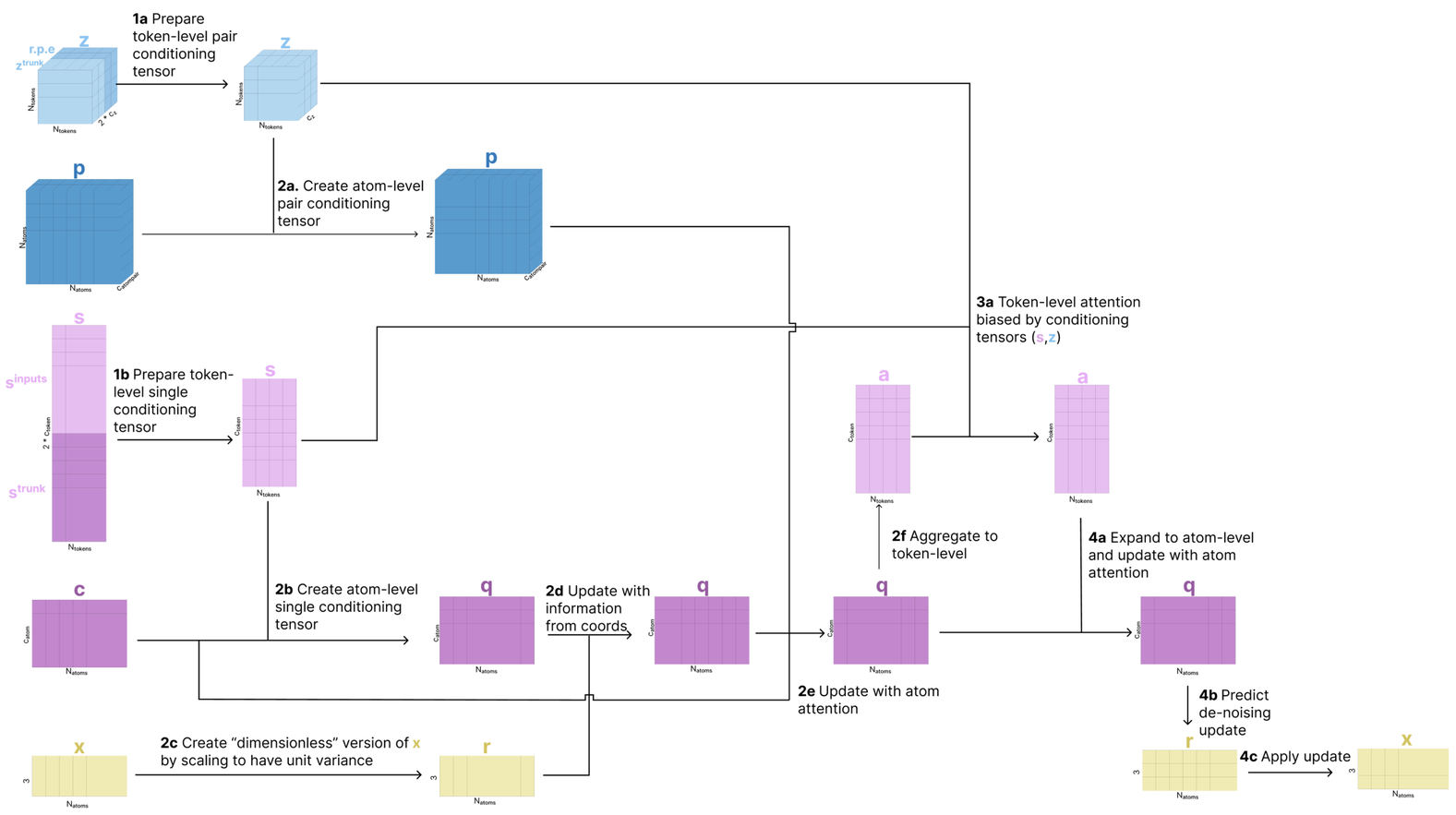

## MSA Module



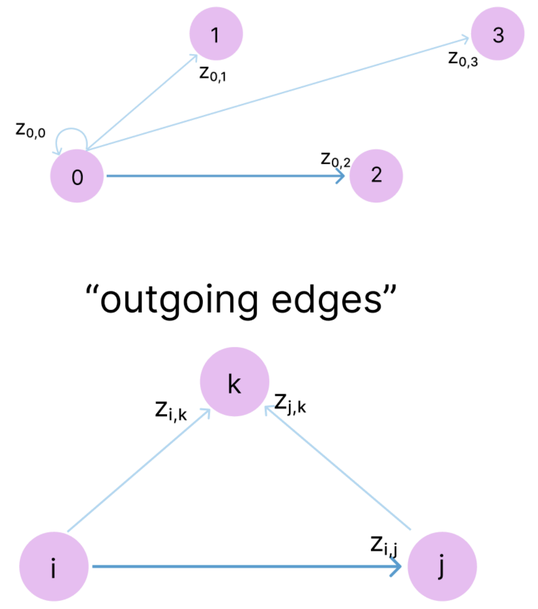

## Pairformer module




## Diffusion Module




## Triangle Attention and Update




## Loss Function

$L_{loss} = L_{distogram} * \alpha_{distogram} + L_{diffusion} * \alpha_{diffusion} + L_{confidence} * \alpha_{confidence}$

Where,

- $L_{distogram}$: Measures how accurately the model predicts token-level distance categories between residues using cross-entropy on the distogram.
- $L_{diffusion}$: Penalizes errors in atom-level pairwise distances, giving extra weight to distances between nearby atoms and protein-ligand interacting atoms.
- $L_{confidence}$: Trains the model to estimate its own uncertainty, rewarding correct confidence in accurate or inaccurate structural predictions.

## Diffusion Loss

$L_{diffusion} = (L_{MSE} + L_{MSE} * \alpha_{bond}) * (\hat{t}^2 + \sigma_{data}{}^2)/(\hat{t}+\sigma_{data})^2 + L_{smooth\_lddt}$

Where,

- $L_{MSE}$: Measures atom-level coordinate accuracy using MSE on all atom positions, with DNA/RNA/ligand atoms upweighted instead of using distogram bins.
- $L_{MSE}$: Adds extra MSE penalty on protein-ligand bond atom pairs to ensure predicted bond lengths match ground-truth distances.
- $L_{smooth\_lddt}$: Evaluates local structural accuracy by computing a smooth, differentiable probability that nearby atom-pair distances fall within multiple distance thresholds.

## Confidence Loss

$L_{confidence} = L_{plDDT} + L_{PDE} + L_{resolved} + L_{PAE} * \alpha_{PAE}$

Where,

- $L_{MSE}$: Measures atom-level coordinate accuracy using MSE on all atom positions, with DNA/RNA/ligand atoms upweighted instead of using distogram bins.
- $L_{MSE}$: Adds extra MSE penalty on protein-ligand bond atom pairs to ensure predicted bond lengths match ground-truth distances.
- $L_{smooth\_lddt}$: Evaluates local structural accuracy by computing a smooth, differentiable probability that nearby atom-pair distances fall within multiple distance thresholds.