

All-Atom Structural Knowledge for Omni Biomolecular Design

AAA^{†1} Han Liu^{‡§2}

[†] Center for Foundation Models and Generative AI, Northwestern University
Department of Computer Science, Northwestern University

[‡] Department of Statistics and Data Science, Northwestern University

Omni biomolecular design models aim to support diverse molecular modalities within a single generative model. A representative approach, AnewOmni, performs geometric diffusion in a shared block-level latent space. This unified latent space supports diverse molecular modalities, but may not fully retain the fine-grained structural information captured by all-atom representations. We first compare AnewOmni with BoltzGen, a strong all-atom binder generator, on the LNR peptide-design benchmark and observe a clear performance gap. This comparison motivates a simple question: whether pretrained all-atom structural knowledge can improve an omni generator without replacing its native unified latent space. We use a frozen BoltzGen trunk to extract PairFormer single and pair representations and transfer them into AnewOmni through a lightweight cross-attention adapter. We train only the adapter and leave the AnewOmni latent space, geometric denoiser, reverse diffusion process, and all-atom decoder unchanged. We further prove that, conditional on the extracted PairFormer representations, the adapter preserves the native rotation-equivariant coordinate prediction of AnewOmni. The same property holds under CFG-style representation guidance. Experiments on LNR show that PairFormer conditioning outperforms AnewOmni. Our results suggest a simple route to combine the broad modeling scope of omni generation with structural knowledge from pretrained all-atom models.

¹aaa@u.northwestern.edu

²hanliu@northwestern.edu

*Code is available at [GitHub](#); future updates will be posted on the [project page](#).

1 Introduction

Generative models have achieved strong results in biomolecular design, including protein and peptide generation [Watson et al., 2023, Kong et al., 2024]. A broader goal is to support small molecules, peptides, antibodies, nucleic acids, and other molecular classes within a single model. These modalities differ substantially in chemistry, topology, size, and structural organization. An omni biomolecular design model therefore requires a shared generative interface across heterogeneous molecular systems. Recent models such as UniMoMo and AnewOmni pursue this goal through unified latent representations [Kong et al., 2025, 2026].

AnewOmni represents each chemical block with a latent feature and a three-dimensional point, and performs geometric diffusion in this shared block-level space [Kong et al., 2026]. The same latent format applies to amino acids, molecular fragments, and other chemical units. This unified representation supports diverse molecular modalities within one generator. At the same time, the compact block-level state may not fully retain the fine-grained structural information captured by all-atom representations. Modern all-atom models instead maintain rich token-level and pairwise representations before coordinate generation. BoltzGen, for example, uses PairFormer single and pair representations to encode individual tokens and their structural relations [Stark et al., 2025].

We first examine this difference empirically. We compare AnewOmni with BoltzGen, a strong all-atom binder generator, on the LNR peptide-design benchmark [Tsaban et al., 2022, Kong et al., 2024]. We observe a clear performance gap between the two models on this shared task. We do not attribute this gap to a single architectural choice, since AnewOmni and BoltzGen also differ in model scale, pretraining data, training objectives, and generation mechanisms. Instead, the comparison motivates a simple question:

Can pretrained all-atom structural knowledge improve an omni biomolecular generator without replacing its native unified latent space?

We address this question through lightweight representation transfer. We use a frozen BoltzGen trunk to extract final-layer PairFormer single and pair representations from the visible design condition. These representations are computed once and reused throughout AnewOmni sampling. A cross-attention adapter transfers them into the native scalar conditioning stream of AnewOmni. We train only the adapter and leave the AnewOmni latent space, geometric denoiser, reverse diffusion process, and all-atom decoder unchanged. This design preserves the unified generator while enriching its conditioning with structural knowledge learned by an all-atom model. At inference time, CFG-style representation guidance controls the strength of the transferred representation.

The geometric structure of AnewOmni introduces a basic correctness requirement. Its latent state combines invariant scalar features with equivariant three-dimensional coordinates. Our adapter modifies only the scalar conditioning stream and does not introduce an external coordinate update. We prove that, conditional on the extracted PairFormer representations, the augmented model preserves the native rotation-equivariant coordinate prediction of AnewOmni. The same property holds under CFG-style representation guidance. This result establishes geometric compatibility

of the transfer interface; the usefulness of the transferred representation is evaluated empirically.

We evaluate the resulting model on LNR. PairFormer conditioning improves several sequence and structural metrics over the original AnewOmni model. We further separate the effects of single and pair representations, compare against a parameter-matched conditioning control, study representation guidance, and measure computational overhead. Together, these experiments test both whether pretrained all-atom representations improve omni generation and which components of the transfer are responsible for the observed gains.

Contributions. Our contributions are three-fold.

- **Method.** We introduce a lightweight all-atom representation-transfer framework for omni biomolecular design. Frozen PairFormer single and pair representations are injected into the scalar conditioning stream of AnewOmni through cross-attention, while all pretrained AnewOmni components remain unchanged (Sections 3.1, 3.2 and 3.4).
- **Geometry compatibility.** We establish a correctness property for the transfer interface. Conditional on the extracted PairFormer representations, the augmented denoiser preserves the native rotation-equivariant coordinate prediction of AnewOmni (Theorem 3.1). CFG-style representation guidance preserves the same property (Corollary 3.1.1).
- **Experiments.** We first compare AnewOmni with a strong all-atom generator on LNR and use the observed performance gap to motivate representation transfer (Section 4.2). We then evaluate PairFormer conditioning, single/pair representations, a parameter-matched conditioning control, representation guidance, and computational cost (Sections 4.3 to 4.6).

2 Preliminaries

Biomolecular Design. Let c denote the design condition, including the target, design region, requested binder type, and prescribed constraints. Let x denote the generated binder, including atom identities, coordinates, and chemical bonds. We seek a conditional generator $p_\theta(x|c)$. An omni biomolecular model uses one generative interface across different molecular modalities. The hidden binder sequence and geometry are excluded from c .

Block-Level Geometric Representation. AnewOmni maps an all-atom structure to a block-level latent representation [Kong et al., 2026]. Each chemical block is represented by

$$z_i = (h_i, q_i), \quad h_i \in \mathbb{R}^d, \quad q_i \in \mathbb{R}^3, \quad (2.1)$$

where h_i is an invariant latent feature and q_i is a latent spatial point. For n blocks, we write $z = (h, q)$ with $h \in \mathbb{R}^{n \times d}$ and $q \in \mathbb{R}^{n \times 3}$. AnewOmni uses the same latent format for amino acids, molecular fragments, and other chemical blocks. A pretrained decoder D maps the generated latent state back to an all-atom binder:

$$p_\theta(x|c) = \int D(x|z_0, c) p_\theta(z_0|c) dz_0. \quad (2.2)$$

Conditional Latent Diffusion. AnewOmni performs diffusion over the block-level latent state [Kong et al., 2026]. We separate generated binder blocks from the fixed target context and write $z_t^G = (h_t^G, q_t^G)$ and $z^C = (h^C, q^C)$. Noise is added only to the generated state:

$$h_t^G = \alpha_t h_0^G + \sigma_t \epsilon_h, \quad (2.3)$$

$$q_t^G = \alpha_t q_0^G + \sigma_t \epsilon_q, \quad \epsilon_h, \epsilon_q \sim \mathcal{N}(0, I), \quad (2.4)$$

with $\alpha_t^2 + \sigma_t^2 = 1$. For each block i , AnewOmni constructs an invariant conditioning embedding a_i from its sequence-position embedding and amino-acid indicator. Before geometric message passing, a_i is combined with the noisy latent feature $h_{t,i}^G$ and the diffusion-time embedding to form the input feature of the equivariant denoiser. We write $a = \{a_i\}$ for these block-conditioning embeddings. The denoiser predicts

$$\epsilon_\theta(z_t^G, t \mid z^C, a) = (\epsilon_{\theta,h}, \epsilon_{\theta,q}), \quad (2.5)$$

and minimizes

$$\mathcal{L}_{\text{diff}}(\theta) = \mathbb{E} \left[\frac{\lambda}{n_G} \|\epsilon_{\theta,h} - \epsilon_h\|_F^2 + \frac{1}{n_G} \|\epsilon_{\theta,q} - \epsilon_q\|_F^2 \right]. \quad (2.6)$$

Sampling follows the native reverse process from T to 0. The pretrained decoder then reconstructs the all-atom binder.

Geometric Symmetry. AnewOmni uses an equivariant geometric denoiser [Kong et al., 2026]. Since the latent coordinates are centered before diffusion, we consider global rotations $R \in \text{SO}(3)$. A rotation acts on the latent state as

$$R \cdot (h, q) = (h, Rq). \quad (2.7)$$

The block-conditioning features a are invariant under rotation. Native geometric prompts, when present, are processed by the original equivariant conditioning mechanism and are omitted from the notation. The native denoiser therefore satisfies

$$\epsilon_{\theta,h}(R \cdot z_t^G, t \mid R \cdot z^C, a) = \epsilon_{\theta,h}(z_t^G, t \mid z^C, a), \quad (2.8)$$

$$\epsilon_{\theta,q}(R \cdot z_t^G, t \mid R \cdot z^C, a) = R \epsilon_{\theta,q}(z_t^G, t \mid z^C, a). \quad (2.9)$$

All-Atom Structural Representations. BoltzGen maintains pretrained structural representations through its PairFormer trunk [Stark et al., 2025]. Given the visible design condition c , PairFormer produces

$$r(c) = (s, p), \quad s = \{s_i\}, \quad p = \{p_{ij}\}, \quad (2.10)$$

where s_i is the single representation associated with token i and p_{ij} is the pair representation associated with tokens i and j . The single representation encodes token-level structural information,

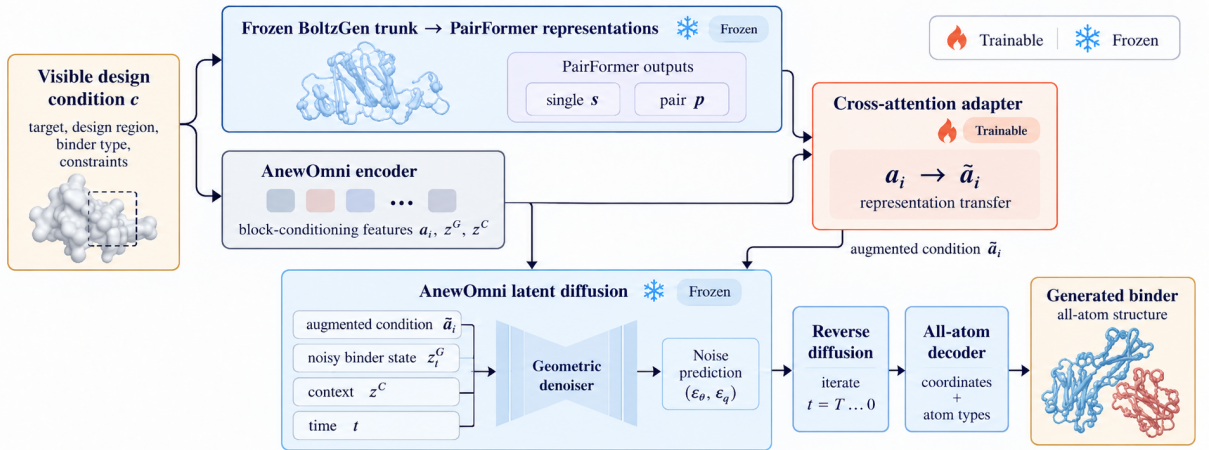


Figure 1: Overview of all-atom representation transfer for omni biomolecular design. Given the visible design condition c , a frozen BoltzGen trunk extracts PairFormer single and pair representations. In parallel, AnewOmni encodes the design input and produces the native block-conditioning features a_i . A lightweight cross-attention adapter transfers the PairFormer representations into the AnewOmni conditioning stream and produces the augmented features $\tilde{a}_i$. The frozen AnewOmni latent diffusion model then takes $\tilde{a}_i$, the noisy binder state z_t^G , the fixed context z^C , and the diffusion time t as input. The denoiser predicts the latent noise, the reverse process runs from $t = T$ to 0, and the all-atom decoder produces the final binder structure. Only the cross-attention adapter is trained.

while the pair representation encodes learned relations between tokens.

3 All-Atom Representation Transfer for Omni Design

We study whether pretrained all-atom structural representations improve an omni biomolecular generator without replacing its native unified latent space. We retain the pretrained AnewOmni generator and use a frozen BoltzGen trunk to provide PairFormer representations. A lightweight cross-attention adapter enriches the native invariant block condition of AnewOmni with these representations. Only the adapter is trained. The latent state, geometric denoiser, native graph prompts, reverse diffusion process, and all-atom decoder remain unchanged.

3.1 Frozen PairFormer Representations

Given the visible design condition c , a frozen BoltzGen trunk produces the final PairFormer representations

$$r(c) = (s, p), \quad s = \{s_j\}, \quad p = \{p_{ij}\}. \quad (3.1)$$

Here $s_j \in \mathbb{R}^{d_s}$ is a single representation and $p_{ij} \in \mathbb{R}^{d_p}$ is a pair representation. The PairFormer input contains the target, design slots, and supplied task constraints. It does not contain the hidden binder identity or geometry.

The two models use different internal representations. We therefore align PairFormer tokens and design slots with the corresponding AnewOmni blocks using molecular correspondence. For notation simplicity, s_j and p_{ij} below denote the aligned representations. We compute (s, p) once for each design task and reuse them throughout AnewOmni sampling. Thus, PairFormer provides fixed pretrained structural representations rather than a second generated state.

3.2 Cross-Attention Representation Transfer

For each generated block i , let a_i denote the native invariant block-conditioning feature used by AnewOmni. In the original model, a_i is combined with the noisy latent feature $h_{t,i}^G$ and the diffusion-time embedding before the resulting invariant feature enters the equivariant denoiser. Our modification is inserted immediately before this native input transformation.

We use a_i as the query and read the PairFormer representations through cross-attention. For one attention head,

$$\ell_{ij} = \frac{(W_Q a_i)^\top W_K s_j}{\sqrt{d}} + b_\phi(p_{ij}), \quad (3.2)$$

$$\alpha_{ij} = \text{softmax}_j(\ell_{ij}), \quad (3.3)$$

$$m_i = \sum_j \alpha_{ij} (W_V s_j + W_P p_{ij}), \quad (3.4)$$

$$\tilde{a}_i = a_i + W_O m_i. \quad (3.5)$$

Here d is the attention width. The function b_ϕ maps the pair representation to an attention bias. All projections belong to the trainable adapter.

The single representation supplies token-level structural information. The pair representation affects both the attention weights and the transferred messages. The residual form in (3.5) recovers the native condition when the adapter output is zero.

The original AnewOmni denoiser uses a_i together with $h_{t,i}^G$ and the diffusion time to construct the invariant feature entering its geometric transformer. We replace only a_i by $\tilde{a}_i$. The generated latent coordinates q_t^G and context state $z^C = (h^C, q^C)$ are unchanged. The native equivariant interaction between generated and context blocks is also unchanged. In particular, PairFormer does not predict a new three-dimensional displacement. It only modifies the invariant information used by the native geometric denoiser. Let $\tilde{a} = \{\tilde{a}_i\}$. The augmented prediction is

$$\epsilon_{\text{aug}} = \epsilon_\theta(z_t^G, t \mid z^C, \tilde{a}), \quad (3.6)$$

where the frozen parameters θ are exactly those of the pretrained AnewOmni denoiser. The dependence on the trainable parameters ϕ enters only through $\tilde{a}$.

3.3 Geometry Compatibility

We next verify that the transfer interface preserves the native geometric transformation law of AnewOmni. This is a correctness property of the interface rather than an explanation of its empirical benefit. Recall that a_i is an invariant feature. The cross-attention adapter in [Equations \(3.2\) to \(3.5\)](#) also produces only invariant features. All geometric variables remain inside the original AnewOmni denoiser. Native geometric prompts, when present, are unchanged by our method and continue to be processed by the original equivariant conditioning mechanism.

Theorem 3.1 (Geometry-compatible representation transfer). Fix the extracted PairFormer representations s and p . Suppose the native block-conditioning features a are rotation invariant, and the native AnewOmni denoiser satisfies [\(2.9\)](#). Then the cross-attention transfer in [Equations \(3.2\) to \(3.5\)](#) has the following properties.

Then, first, the transferred condition remains rotation invariant:

$$\tilde{a} \mapsto \tilde{a} \quad \text{under any } R \in \text{SO}(3). \quad (3.7)$$

Second, the augmented denoiser preserves the native geometric symmetry:

$$\epsilon_{\text{aug},h} (R \cdot z_t^G, t \mid R \cdot z^C, \tilde{a}) = \epsilon_{\text{aug},h} (z_t^G, t \mid z^C, \tilde{a}), \quad (3.8)$$

$$\epsilon_{\text{aug},q} (R \cdot z_t^G, t \mid R \cdot z^C, \tilde{a}) = R \epsilon_{\text{aug},q} (z_t^G, t \mid z^C, \tilde{a}). \quad (3.9)$$

Proof. Please see [Appendix B](#) for a detailed proof. $\square$

The theorem is conditional on the extracted PairFormer representations. It does not require the PairFormer extractor itself to be rotation invariant. It therefore characterizes the geometry compatibility of the transfer interface, rather than end-to-end equivariance under recomputation of PairFormer features after rotating the raw input.

3.4 Training and Representation Guidance

We freeze the BoltzGen trunk and all pretrained AnewOmni components. Only the cross-attention parameters ϕ are optimized. Let (ϵ_h, ϵ_q) denote the noise added to the clean generated latent state. We use the native AnewOmni noise-prediction objective:

$$\mathcal{L}_{\text{fusion}}(\phi) = \mathbb{E} \left[\frac{\lambda}{n_G} \|\epsilon_{\text{aug},h} - \epsilon_h\|_F^2 + \frac{1}{n_G} \|\epsilon_{\text{aug},q} - \epsilon_q\|_F^2 \right]. \quad (3.10)$$

No additional structural supervision is introduced.

At inference time, the original AnewOmni model gives

$$\epsilon_{\text{base}} = \epsilon_{\theta} (z_t^G, t \mid z^C, a), \quad (3.11)$$

while the transferred model gives ϵ_{aug} from [\(3.6\)](#). Motivated by classifier-free guidance [\[Ho and](#)

Salimans, 2022], we use

$$\epsilon_{\gamma} = \epsilon_{\text{base}} + \gamma (\epsilon_{\text{aug}} - \epsilon_{\text{base}}), \quad \gamma \geq 0. \quad (3.12)$$

The choice $\gamma = 0$ recovers the original AnewOmni prediction. The choice $\gamma = 1$ uses the learned PairFormer-conditioned prediction. Values above one extrapolate the transferred correction.

Corollary 3.1.1 (Geometry compatibility of representation guidance). Under the conditions of [Theorem 3.1](#), $\epsilon_{\gamma,q}$ remains rotation equivariant for every $\gamma \geq 0$.

Both $\epsilon_{\text{base},q}$ and $\epsilon_{\text{aug},q}$ obey the same rotation transformation law. Their linear combination therefore preserves equivariance. The reverse diffusion transitions and the final all-atom decoder remain identical to those of AnewOmni.

4 Experimental Studies

We evaluate whether pretrained all-atom structural representations improve a frozen omni biomolecular generator through the representation-transfer interface in [Section 3](#). Our experiments address four questions. First, does PairFormer representation transfer improve AnewOmni on peptide design? Second, which components of the PairFormer representation contribute to the effect? Third, does the improvement come from pretrained structural representations rather than additional trainable capacity? Finally, how does inference-time representation guidance affect generation quality and computational cost?

4.1 Experimental Setup

Benchmark and protocol. We evaluate on the LNR peptide-design benchmark [[Tsaban et al., 2022](#), [Kong et al., 2024](#)]. Each task specifies a protein target, a binding region, and a peptide length. The native peptide sequence and geometry are excluded from the input. All compared methods receive the same visible target information and design constraints. We reserve LNR exclusively for evaluation and use a target-disjoint validation set for checkpoint and guidance selection. Each method generates the same number of candidates for each target. Any additional ranking or refinement procedure is reported separately.

Compared methods. We compare four main configurations. *AnewOmni* is the original pretrained generator. *PairFormer transfer* uses the learned cross-attention interface in [Equations \(3.2\) to \(3.5\)](#) with $\gamma = 1$. *PairFormer transfer + guidance* applies the inference-time scaling in [\(3.12\)](#). *BoltzGen* serves as a strong all-atom reference on the same peptide-design task. We further evaluate single-only and pair-only transfer, together with a parameter-matched control that does not use pretrained PairFormer representations.

Training. We freeze the AnewOmni VAE, denoiser, and decoder. We also freeze the BoltzGen trunk. Only the cross-attention parameters ϕ are optimized using the native AnewOmni objective in [\(3.10\)](#). PairFormer representations are computed once for each design task and cached

Table 1: Peptide-design results on LNR. All methods use the same visible task inputs and candidate budget. PairFormer transfer uses the learned representation interface with $\gamma = 1$.

Method	AAR (%) $\uparrow$	Complex RMSD (Å) $\downarrow$	Peptide RMSD (Å) $\downarrow$	Peptide collision (%) $\downarrow$	Interface collision (%) $\downarrow$	Rosetta ΔG $\downarrow$
AnewOmni	6.484	8.550	2.892	0.914	0.908	-21.89
PairFormer transfer ($\gamma = 1$)	6.664	8.248	2.849	0.911	1.113	-21.96
PairFormer transfer + guidance	TBD	TBD	TBD	TBD	TBD	TBD
BoltzGen	TBD	TBD	TBD	TBD	TBD	TBD

throughout training and sampling. Checkpoint selection and guidance-scale selection use only validation targets. Comparisons among AnewOmni variants use matched latent initialization, diffusion noise, and decoder randomness.

Metrics and statistical reporting. We report complementary sequence, structural, steric, and energetic metrics. Amino-acid recovery (AAR) measures sequence agreement with the native peptide. Complex RMSD measures peptide deviation after alignment on the target. Peptide RMSD measures internal peptide geometry after peptide alignment. Peptide collision and interface collision measure steric conflicts within the peptide and across the peptide–target interface, respectively. We additionally report Rosetta ΔG . Higher AAR is better; lower values are better for the remaining metrics. Candidate-level metrics are first averaged within each target, and targets receive equal weight. For the final evaluation, we report paired bootstrap confidence intervals over targets and variation across adapter-training seeds.

4.2 Main Results on LNR

Overall performance. Table 1 reports the current LNR comparison. At $\gamma = 1$, PairFormer transfer improves five of the six reported point estimates over AnewOmni. AAR increases from 6.484% to 6.664%. Complex RMSD decreases from 8.550 Å to 8.248 Å, and Peptide RMSD decreases from 2.892 Å to 2.849 Å. Peptide collision changes from 0.914% to 0.911%, and Rosetta ΔG changes from -21.89 to -21.96 . These results show that modifying only the invariant block-conditioning features through pretrained PairFormer representations changes the behavior of an otherwise frozen AnewOmni generator.

The improvement is not uniform. Interface collision increases from 0.908% to 1.113%. We therefore evaluate the method using complementary metrics rather than interpreting the current results as a uniform improvement in geometric quality. The final evaluation reports target-level uncertainty and determines which differences are statistically reliable. The matched BoltzGen comparison further measures the remaining gap between the omni generator and the all-atom reference.

Target-level consistency. Aggregate averages may hide heterogeneous effects across targets. We therefore compute the paired per-target change relative to AnewOmni and report the fraction of targets that improve under each metric. We additionally stratify results by peptide length. This

Table 2: Ablation of transferred PairFormer representations on LNR.

Representation	AAR (%) $\uparrow$	Complex RMSD (Å) $\downarrow$	Peptide RMSD (Å) $\downarrow$	Interface collision (%) $\downarrow$
Single only	TBD	TBD	TBD	TBD
Pair only	TBD	TBD	TBD	TBD
Single + pair	TBD	TBD	TBD	TBD

Table 3: Control for additional trainable conditioning capacity.

Method	AAR (%) $\uparrow$	Complex RMSD (Å) $\downarrow$	Interface collision (%) $\downarrow$
AnewOmni	6.484	8.550	0.908
Matched adapter	TBD	TBD	TBD
PairFormer transfer	6.664	8.248	1.113

analysis tests whether the aggregate change is broadly distributed across targets or concentrated on a small subset.

4.3 Which PairFormer Representations Matter?

PairFormer provides single representations s and pair representations p . Our interface uses the single representation in the transferred message and the pair representation in both the attention bias and message. We therefore evaluate three variants at $\gamma = 1$: single only, pair only, and single plus pair. All variants use the same frozen AnewOmni checkpoint, training data, optimization budget, and insertion point. Projection widths are adjusted to keep trainable parameter counts comparable.

The single-only variant tests the effect of pretrained token-level information. The pair-only variant isolates the contribution of learned relational information. The full model tests whether the two representation types provide complementary signals. These experiments are interpreted empirically; we do not assume that either representation type must dominate.

4.4 Do Pretrained Representations Matter?

The cross-attention interface introduces additional trainable parameters. An improvement over AnewOmni could therefore arise from additional adaptation capacity rather than from the pretrained PairFormer representation itself. We test this alternative with a parameter-matched control. The control uses the same visible task information and the same conditioning insertion point, but does not receive pretrained PairFormer features. It uses the same training examples, optimization budget, checkpoint-selection rule, and frozen AnewOmni backbone.

This comparison separates the effect of additional trainable capacity from the effect of the pretrained PairFormer representation. A difference between the matched adapter and PairFormer

Table 4: Effect of representation guidance on LNR.

γ	AAR (%) $\uparrow$	Complex RMSD (Å) $\downarrow$	Interface collision (%) $\downarrow$	Rosetta ΔG $\downarrow$
0	6.484	8.550	0.908	-21.89
1	6.664	8.248	1.113	-21.96
Selected > 1	TBD	TBD	TBD	TBD

transfer provides evidence about whether pretrained structural information contributes beyond the adapter architecture itself. The comparison does not control for the pretraining compute or data used to obtain the PairFormer representation.

4.5 Representation Guidance

We next study the inference-time scale γ in (3.12). The cross-attention parameters remain fixed throughout this experiment. The choice $\gamma = 0$ recovers the original AnewOmni prediction. The choice $\gamma = 1$ uses the learned PairFormer-transferred prediction. Values $\gamma > 1$ extrapolate the difference between the transferred and original predictions. We select the reported guidance scale using validation targets and fix it before evaluation on LNR.

The sweep tests whether the transferred correction remains useful when amplified beyond its training strength. For $\gamma \neq 0, 1$, our implementation evaluates both the original and transferred denoiser predictions. We therefore report quality together with the corresponding inference overhead. If different metrics favor different guidance scales, the selection rule is fixed on the validation set rather than chosen retrospectively on LNR.

4.6 Computational Cost

Our method trains only the cross-attention interface, but inference additionally requires PairFormer feature extraction. We therefore report trainable parameter count, feature-extraction time, cache size, peak memory, and end-to-end generation time under identical hardware. PairFormer representations are computed once for each design task and reused across all diffusion steps and generated candidates.

At $\gamma = 1$, each diffusion step evaluates only the PairFormer-transferred AnewOmni branch. At $\gamma = 0$, PairFormer extraction and the transfer interface are unnecessary. For $\gamma \neq 0, 1$, our implementation evaluates both the original and transferred predictions at each diffusion step.

We report PairFormer extraction separately from the amortized per-candidate sampling cost. This distinction matters when multiple candidates share the same design task.

Table 5: Training and inference cost. Generation time is measured using the same hardware and candidate budget.

Method	Trainable params.	Time / candidate	Peak memory
AnewOmni	0	TBD	TBD
PairFormer transfer ($\gamma = 1$)	TBD	TBD	TBD
PairFormer transfer + guidance	TBD	TBD	TBD

5 Discussion and Conclusion

In this work, we study whether pretrained all-atom structural representations improve an omni biomolecular generator without replacing its native generative space. We focus on AnewOmni, which performs geometric diffusion in a shared block-level latent space, and use the performance gap to a strong all-atom generator on LNR as motivation rather than as evidence for a specific failure mechanism. Our method extracts frozen PairFormer single and pair representations from the visible design condition and transfers them into the native scalar conditioning stream through a lightweight cross-attention interface. The AnewOmni latent state, geometric denoiser, reverse diffusion process, and all-atom decoder remain unchanged. The interface therefore separates structural transfer from geometric generation: PairFormer supplies additional pretrained structural knowledge, while AnewOmni retains responsibility for all coordinate updates. Our symmetry analysis establishes the corresponding correctness property. Conditional on the extracted donor memory, scalar representation transfer preserves the native rotation-equivariant coordinate prediction. Experiments then determine whether this additional structural information improves generation rather than attributing the gain to the symmetry argument itself.

Our study has several limitations. The current evaluation focuses on canonical peptide design, and other molecular modalities require a more general alignment between donor tokens and AnewOmni blocks. PairFormer representations are extracted once and remain fixed during sampling, so they do not directly observe the evolving generated binder. Guidance extrapolation also requires an additional denoiser evaluation, and comparisons with BoltzGen retain differences in architecture, scale, pretraining data, and objectives. Computational structural metrics further do not establish experimental binding activity. Despite these limitations, the results support a simple direction for improving omni biomolecular generators: preserve the unified latent generator and augment it with structural knowledge learned by stronger pretrained models. This modular design separates advances in structural representation learning from the choice of generative state and provides a lightweight path toward stronger unified biomolecular design models.

Appendix

A	Related Work	13
B	Proof of Theorem 3.1	14

Impact Statement

This work develops computational methods for biomolecular design. Such models may support applications in protein engineering, therapeutic discovery, and molecular research. At the same time, increasingly capable biomolecular generation systems may also be applied to undesirable biological design tasks. Our experiments are computational and do not involve wet-lab validation or the synthesis of generated molecules. Future deployment of such systems should consider appropriate screening, access controls, and experimental safety procedures.

LLM Usage Disclosure

We used large language models (LLMs) to aid and polish writing, such as improving clarity, grammar, and conciseness. We also used LLMs for retrieval and discovery, for example exhausting literature to identify potential missing related work. All technical content, proofs, experiments, and results are original contributions by the authors.

A Related Work

Unified biomolecular generation. Structure-based generative models have achieved strong results in specialized settings, including protein design with RFdiffusion [Watson et al., 2023] and full-atom peptide design with PepGLAD [Kong et al., 2024]. More recent work seeks a shared generative interface across molecular classes. UniMoMo and AnewOmni represent heterogeneous molecules through common block-level latent states and perform geometric generation in this shared space [Kong et al., 2025, 2026]. PocketXMol instead adopts an atom-level representation for multiple molecular interaction tasks [Peng et al., 2024], while ODesign develops an all-atom generative world model for broad biomolecular interaction design [Zhang et al., 2025]. BoltzGen focuses on all-atom protein and peptide binder generation across diverse target modalities [Stark et al., 2025]. Our work is complementary to these efforts. Rather than introducing another base generator, we study whether structural knowledge from a pretrained all-atom model can improve an existing omni latent generator without changing its native state space or decoder.

All-atom structural representations. Modern biomolecular models increasingly maintain rich token-level and pairwise representations before coordinate generation. AlphaFold 3 uses PairFormer representations to reason over biomolecular entities and their relations before its diffusion-based structure module [Abramson et al., 2024]. BoltzGen builds on related all-atom structural modeling and jointly supports structure reasoning and generative binder design [Stark et al., 2025]. These models motivate the use of single and pair representations as reusable structural knowledge. In contrast to running the donor generator itself, we stop at the frozen PairFormer trunk and transfer its final representations into AnewOmni. The donor therefore provides structural memory rather than a second coordinate trajectory.

Lightweight conditioning and representation transfer. A related line of work augments large pretrained generative models with lightweight conditioning modules while keeping the backbone frozen. ControlNet adds external spatial control to pretrained diffusion models through trainable side networks [Zhang et al., 2023], while T2I-Adapter learns compact adapters for additional structural conditions [Mou et al., 2024]. IP-Adapter uses a lightweight cross-attention mechanism to inject representations from an external image encoder into a frozen text-to-image diffusion model [Ye et al., 2023]. Our method follows this modular principle but differs in two aspects. The donor and receiver are independently pretrained biomolecular models, and the receiver is an equivariant geometric generator. We therefore transfer PairFormer features only through the scalar conditioning stream and leave all coordinate updates to the native AnewOmni denoiser.

B Proof of Theorem 3.1

Theorem B.1 (Restate: Geometry-compatible representation transfer). Fix the extracted PairFormer representations s and p . Suppose the native block-conditioning features a are rotation invariant, and the native AnewOmni denoiser satisfies (2.9). Then the cross-attention transfer in Equations (3.2) to (3.5) has the following properties.

Then, first, the transferred condition remains rotation invariant:

$$\tilde{a} \mapsto \tilde{a} \quad \text{under any } R \in \text{SO}(3). \quad (\text{B.1})$$

Second, the augmented denoiser preserves the native geometric symmetry:

$$\epsilon_{\text{aug},h} (R \cdot z_t^G, t \mid R \cdot z^C, \tilde{a}) = \epsilon_{\text{aug},h} (z_t^G, t \mid z^C, \tilde{a}), \quad (\text{B.2})$$

$$\epsilon_{\text{aug},q} (R \cdot z_t^G, t \mid R \cdot z^C, \tilde{a}) = R \epsilon_{\text{aug},q} (z_t^G, t \mid z^C, \tilde{a}). \quad (\text{B.3})$$

Proof. For fixed s and p , a joint rotation leaves a_i , s_j , and p_{ij} unchanged. Hence the attention logits

$$\ell_{ij} = \frac{(W_Q a_i)^\top W_K s_j}{\sqrt{d}} + b_\phi(p_{ij}) \quad (\text{B.4})$$

remain unchanged. Therefore, the attention weights α_{ij} and transferred messages m_i are unchanged. By (3.5),

$$\tilde{a}_i = a_i + W_O m_i \quad (\text{B.5})$$

is also unchanged under rotation. This proves (3.7).

The augmented model uses the original AnewOmni geometric denoiser with $\tilde{a}$ in place of a . Since $\tilde{a}$ is invariant, the native symmetry in (2.9) directly gives the invariance of the latent-feature prediction and the equivariance of the coordinate prediction in (3.9). $\square$

References

- Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J. Ballard, Joshua Babrick, et al. Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature*, 630:493–500, 2024.
- Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*, 2022.
- Xiangzhe Kong, Yinjun Jia, Wenbing Huang, and Yang Liu. Full-atom peptide design with geometric latent diffusion. In *Advances in Neural Information Processing Systems*, volume 37, pages 74808–74839, 2024.
- Xiangzhe Kong, Zishen Zhang, Ziting Zhang, Rui Jiao, Jianzhu Ma, Wenbing Huang, Kai Liu, and Yang Liu. Unimomo: Unified generative modeling of 3d molecules for de novo binder design. In *Proceedings of the 42nd International Conference on Machine Learning*, pages 31397–31418, 2025.
- Xiangzhe Kong, Junwei Chen, Ziting Zhang, Gaodeng Li, Qingyuan Zhu, Lei Wei, Mingyu Li, Yan Shi, Weiyang Dai, Zishen Zhang, Wenjuan Tan, Rui Jiao, Xiaolun Wang, Jiqing Zheng, Ziyang Yu, Qilong Wu, Zhiye Guo, Li Zhang, Wentao Li, Qiaojing Huang, Tian Zhu, Xiaowo Wang, Wenbing Huang, Yuli She, Jian Zhang, Yang Liu, Kai Liu, and Jianzhu Ma. Programming biomolecular interactions with all-atom generative model. *bioRxiv*, 2026.
- Chong Mou, Xintao Wang, Liangbin Xie, Yanze Wu, Jian Zhang, Zhongang Qi, and Ying Shan. T2i-adapter: Learning adapters to dig out more controllable ability for text-to-image diffusion models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 4296–4304, 2024.
- Xingang Peng, Fenglin Guo, Ruihan Guo, Jiayu Sun, Jiaqi Guan, Yinjun Jia, Yan Xu, Yanwen Huang, Muhan Zhang, Jian Peng, Xinquan Wang, Chuanhui Han, Zihua Wang, and Jianzhu Ma. Atom-level generative foundation model for molecular interaction with pockets. *bioRxiv*, 2024.
- Hannes Stark, Felix Faltings, MinGyu Choi, Yuxin Xie, Eunsu Hur, Timothy O’Donnell, Anton Bushuiev, Talip Uçar, Saro Passaro, Weian Mao, Mateo Reveiz, Roman Bushuiev, Tally Portnoi, Tomáš Pluskal, Josef Sivic, Karsten Kreis, Arash Vahdat, Shamayeeta Ray, Jonathan T. Goldstein, Andrew Savinov, Jacob A. Hambalek, Anshika Gupta, Diego A. Taquiri-Diaz, Yaotian Zhang, Samuel J. Snyder, A. Katherine Hatstat, Angelika Arada, Nam Hyeong Kim, Haoyu Fan, Ethel Tackie-Yarboi, Dylan Boselli, Lee Schnaider, Chang C. Liu, Gene-Wei Li, Denes Hnisz, David M. Sabatini, William F. DeGrado, Jeremy Wohlwend, Gabriele Corso, Regina Barzilay, and Tommi Jaakkola. Boltzgen: Toward universal binder design. *bioRxiv*, 2025.
- Tomer Tsaban, Julia K. Varga, Orly Avraham, Ziv Ben-Aharon, Alisa Khramushin, and Ora

- Schueler-Furman. Harnessing protein folding neural networks for peptide–protein docking. *Nature Communications*, 13:176, 2022.
- Joseph L. Watson, David Juergens, Nathaniel R. Bennett, Brian L. Trippe, Jason Yim, Helen E. Eisenach, Woody Ahern, Andrew J. Borst, Robert J. Ragotte, Lukas Milles, et al. De novo design of protein structure and function with rfdiffusion. *Nature*, 620:1089–1100, 2023.
- Hu Ye, Jun Zhang, Sibor Liu, Xiao Han, and Wei Yang. Ip-adapter: Text compatible image prompt adapter for text-to-image diffusion models. *arXiv preprint arXiv:2308.06721*, 2023.
- Lymin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3836–3847, 2023.
- Odin Zhang, Xujun Zhang, Haitao Lin, Cheng Tan, Qinghan Wang, Yuanle Mo, Qiantai Feng, Gang Du, Yuntao Yu, Zichang Jin, et al. Odesign: A world model for biomolecular interaction design. *arXiv preprint arXiv:2510.22304*, 2025.