BoltzGen-ADC: Generative Design of Universal Protease-Cleavable Peptide Linkers for Antibody–Drug Conjugates

Mahda Ghasemiagreh, Alireza Noroozi
Koç University

July 2026

Abstract

Peptide linkers have emerged as the preferred cleavable linker class in modern ADCs owing to their biocompatibility, tunable protease specificity, and favorable pharmacological properties. Recent generative biomolecular models can now design high-quality proteins and peptides *de novo* with structural fidelity. We present **BoltzGen-ADC**, a pipeline built on the generative models **ADFlip** and **BoltzGen**, that designs short protease-cleavable peptide linkers for **any** antibody variable-region sequence and **any** small-molecule payload specified by isomeric SMILES, without retraining or program-specific templates. BoltzGen co-generates linker backbone and sequence within a covalently constrained Fv–linker–payload assembly; ranked conjugates are then refined with **ADFLIP** all-atom inverse folding on the designed peptide in structural context. Downstream filters enforce **Fc preservation** (no spurious contacts to constant-region surfaces), **protease-cleavage liability** scoring, and Boltz-based validation of peptide–protease complexes. We demonstrate the workflow on Adcetris-inspired brentuximab–MMAE conjugates. BoltzGen-ADC compresses empirical linker iteration into a GPU-accelerated *in silico* campaign applicable to arbitrary antibody–payload pairs.

1 Introduction

1.1 Antibody–drug conjugates and the role of linkers

Cytotoxic chemotherapeutics remain a cornerstone of cancer treatment but are frequently limited by poor tumor selectivity and dose-limiting toxicity toward healthy proliferating tissues [2]. Antibody–drug conjugates (ADCs) were developed to overcome these limitations by combining the targeting specificity of monoclonal antibodies with highly potent cytotoxic agents [3, 4]. Structurally, an ADC comprises three essential components: a tumor-targeting antibody, a cytotoxic payload, and a chemical linker that covalently connects the two [4, 5]. Although the antibody and payload determine target specificity and cytotoxic mechanism, respectively, the linker is critical for maintaining systemic stability and controlling payload release, thereby directly influencing both the efficacy and safety of the conjugate [5, 6].

ADC linkers are broadly classified as non-cleavable or cleavable, the latter including acid-sensitive (hydrazone), reduction-sensitive (disulfide), and enzyme-cleavable linkers [7, 8]. Although hydrazone linkers were among the first clinically used linker chemistries, their susceptibility to premature hydrolysis can result in systemic payload release and off-target toxicity [9]. Likewise, disulfide linkers may undergo undesired extracellular reduction, making it difficult to simultaneously achieve high plasma stability and efficient intracellular payload release [7]. In contrast, non-cleavable thioether linkers exhibit excellent circulatory stability but rely on complete lysosomal degradation of the antibody for drug release and generally lack a bystander effect because the released lysine–linker–payload metabolite is poorly membrane permeable and therefore cannot efficiently diffuse into neighboring cells [7]. These limitations have driven the development of enzyme-cleavable peptide linkers, which have become the predominant linker class in clinically approved ADCs owing to their favorable balance between stability and tumor-selective payload release [8].

1.2 Peptide linkers as a platform for next-generation ADCs

Peptide linkers possess excellent biocompatibility, high plasma stability, and efficient intracellular payload release following protease-mediated cleavage [8, 10]. Their amino acid sequences can be readily engineered to modulate cleavage kinetics, stability, and enzyme specificity, while numerous tumor-associated proteases—including cathepsins, legumain, fibroblast activation protein (FAP), and plasmin—provide diverse and well-characterized biological triggers for selective drug release [7, 8, 11]. Moreover, extensive peptide sequence repositories and protease–substrate databases provide rich biochemical knowledge for rational linker engineering [12]. Recent advances in machine learning and generative artificial intelligence have enabled the computational design and optimization of peptide sequences with tailored biochemical and functional properties [13], making peptide linkers an attractive platform for next-generation ADC engineering.

The challenge is not whether to use peptide linkers, but rather to identify peptide sequences that simultaneously preserve antibody function, remain stable during systemic circulation, and undergo efficient cleavage within the tumor microenvironment. Achieving these competing objectives requires integrating structural compatibility with protease specificity, motivating the development of generative computational approaches for peptide linker design.

1.3 Generative models for linker design

Recent advances in generative AI have begun to automate linker discovery. For example, LinkerGPT [1] generates chemically valid ADC linker scaffolds using language models and reinforcement learning, but it is limited to small-molecule linkers and does not account for the three-dimensional structure of the antibody, payload, or conjugate. In contrast, recent structure-aware generative models such as **Boltz** and **BoltzGen** jointly model biomolecular structure and sequence, enabling the design of peptide linkers conditioned on a specific antibody–payload pair. This motivates us to treat ADC linker discovery as a structure-aware generative design problem rather than a purely chemical optimization task.

1.4 Hypothesis

We hypothesize that structure-aware generative models can design short peptide linkers tailored to a given antibody and payload while satisfying the structural and functional constraints required for antibody–drug conjugates. To investigate this hypothesis, we propose a computational pipeline that combines BoltzGen with inverse folding, Fc-preservation filtering, protease-cleavage prediction, and structural validation. As a proof of concept, we evaluate the pipeline *in silico* on the brentuximab–MMAE system and demonstrate that it can generate multiple candidate peptide linkers that satisfy these computational design criteria.

2 Methods

2.1 Pipeline overview

BoltzGen-ADC proceeds in six stages (Figure 1):

1. **Specification** — antibody heavy/light Fv sequences and payload isomeric SMILES; conjugation atom indices.
2. **Generative design (BoltzGen)** — sample linker peptide backbones and sequences in a covalently constrained Fv–linker–payload assembly.
3. **Inverse folding (BoltzGen + ADFLIP)** — assign sequences during design; re-sequence designed peptides in final structures with ADFLIP all-atom inverse folding.
4. **Structural filtering (BoltzGen)** — hard thresholds on refolding RMSD, composition bias, interface confidence, and diversity selection.
5. **Fc filtering** — reject designs whose linker peptide contacts antibody constant-region (non-Fv) surfaces.

6. **Protease-cleavage filtering (Boltz)** — rank by liability motifs and validate peptide–cathepsin B complexes with Boltz structure prediction.

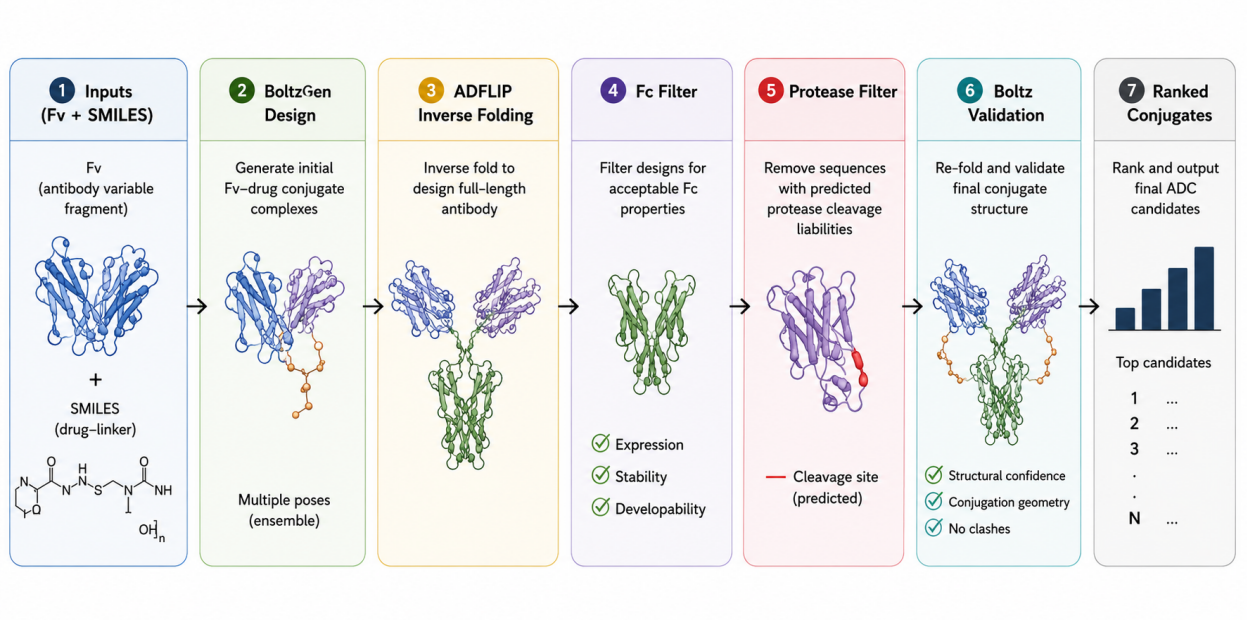

Figure 1: BoltzGen-ADC computational pipeline.

2.2 Generative models: Boltz and BoltzGen

Boltz is a structure-prediction model for proteins, nucleic acids, and small-molecule ligands in complex [15]. We use Boltz in the protease-validation stage to predict cathepsin B–peptide complex structures and confidence metrics (`ptm`, `iptm`, `complex_plddt`).

BoltzGen extends the Boltz architecture with a generative design module that jointly samples binder backbone geometry and target context, followed by inverse folding and refolding to assess designability [14]. The standard `boltzgen run` workflow comprises design, inverse fold, folding, analysis, and filtering.

2.3 Universal antibody–linker–payload conjugate design

We model the ADC as a covalent assembly of four entities:

- **H, L** — user-supplied Antibody sequences (heavy and light chains).
- **P** — designed peptide linker, length 2–6 residues.
- **D** — payload specified by isomeric SMILES (e.g., MMAE).

Two covalent `bond` constraints define topology:

1. Antibody heavy-chain C-terminus → linker N-terminus (`[H, C_term, C] - [P, 1, N]`).
2. Linker C-terminus → payload attachment atom (`[P, L, C] - [D, 1, N5]` for MMAE).

No antibody PDB is required; BoltzGen builds that from sequence. Swapping antibody or payload requires only new sequences/SMILES in the YAML specification.

2.4 ADFLIP all-atom inverse folding on final structures

After BoltzGen produces ranked conjugate structures, we run **ADFLIP** [16] on the **linker peptide residues only** within each final designed structure. ADFLIP conditions on the full atomic context containing antibody Fv, payload atoms, and linker backbone geometry, to propose alternative sequences compatible with the designed three-dimensional pose. ADFLIP complements BoltzGen's built-in inverse folding by providing an independent, all-atom sequence proposal on the final refolded conjugate geometry.

2.5 Fc filtering

ADC linkers must not engage antibody surfaces outside the intended attachment corridor, especially the constant region (Fc), where off-target contacts could impair effector functions (ADCC, CDC) or accelerate clearance. We implement an **Fc filter** using heavy-atom contact analysis.Peptide–antibody contacts within 4.5 Å were quantified for Fv and constant-region residues, and designs were retained only if all peptide contacts were confined to the Fv region. This enforces the challenge objective to preserve antibody binding and Fc functions at the structural level.

2.6 Protease-cleavage filtering

To ensure payload release, candidate linker peptides are evaluated against a user-specified target protease. For each design problem, the corresponding protease structure is provided (e.g., cathepsin B, PDB 3PBH), and **Boltz** predicts the peptide–protease complex. Confidence metrics (`ptm`, `iptm`, and `complex_plddt`) are used as *in silico* evidence of a plausible cleavage-compatible binding mode. 2a

2.7 Peptide physicochemical properties

Linker candidacy also depends on solution properties that influence conjugate solubility, aggregation risk, and charge state across physiological pH. We compute peptide-level descriptors with BioPython `ProteinAnalysis` for each BoltzGen and ADFLIP sequence:

- **GRAVY** (grand average of hydropathy) — hydrophobicity proxy; lower values favor aqueous solubility.
- **Instability index** — sequences below 40 are classed as stable in the ProtParam model.
- **Isoelectric point (pI)** and **net charge** at pH 7.4 (circulation), 6.0, and 5.0 (endosomal/lysosomal) — relevant to off-target binding and protease accessibility.
- **Molecular weight** and **aromaticity** — manufacturing and hydrophobicity indicators.

The results are summarized in Table 4.

2.8 Exemplar: brentuximab–MMAE

We instantiate the pipeline for Adcetris (brentuximab vedotin): cAC10 Fv sequences, MMAE payload SMILES, and `peptide-anything` protocol with 50 intermediate designs and a final budget of 10 ranked candidates.

(a) Protease validation structures from Boltz.

(b) Best-ranked design for the peptide sequence “MR”.

Figure 2: Computational validation of the designed peptide linker. Left: predicted protease–linker complex. Right: highest-ranked linker design generated by the pipeline.

3 Results

3.1 BoltzGen design outputs

From 50 sampled conjugates, BoltzGen filtering retained 10 diverse final ranked brentuximab–MMAE linker designs (Table 1). Sequences span 2–6 residues; all pass refolding RMSD thresholds (< 2 Å). Mean design-to-target iPTM is 0.40 (range 0.37–0.46). Liability scoring flags MetOx at N-terminal methionine in 7/10 designs; ranks 8 and 10 carry ProtTryp motifs (cathepsin-favorable); rank 9 carries a DPP4 circulation liability.

Table 1: BoltzGen outputs for 10 final ranked brentuximab–MMAE linker designs.

Rank	Seq	Len	iPTM	PAE	RMSD (Å)	Filters	Liability
1	MR	2	0.460	8.19	1.29	9/9	MetOx
2	MHH	3	0.437	8.39	1.47	9/9	MetOx
3	MSPPA	5	0.396	9.53	0.73	9/9	MetOx
4	GSPP	4	0.406	9.55	1.53	9/9	—
5	MSPPT	5	0.396	9.75	0.77	9/9	MetOx
6	MSPIIV	6	0.385	9.44	1.61	9/9	MetOx; HydroPatch
7	MGPLS	5	0.375	10.21	1.69	9/9	MetOx
8	MRPE	4	0.418	9.35	1.37	9/9	ProtTryp; MetOx
9	ASPTFS	6	0.375	10.76	1.85	9/9	DPP4
10	GSKAP	5	0.373	10.41	1.35	9/9	ProtTryp

3.2 Fc filtering results

All 10 final designs pass the Fc filter: linker peptides contact only Fv-region residues (8–17 heavy-atom Fv contacts per design; 0 non-Fv contacts at 4.5 Å for every candidate), consistent with preservation of constant-region integrity.

3.3 ADFLIP inverse folding on final structures

ADFLIP was run on all 10 final ranked conjugate mmCIF files with peptide-only design masks (Table 2). Native sequences are those assigned by BoltzGen; ADFLIP sequences are all-atom inverse-folded proposals conditioned on the full conjugate structure. Sequence recovery ranges from 0.00 (ranks 1–2) to 0.75 (rank 4); ADFLIP frequently proposes glycine- and proline-rich alternatives, consistent with backbone flexibility at the antibody–payload junction.

Table 2: ADFLIP inverse folding on linker peptides in final BoltzGen structures (10 designs).

Rank	Native (BoltzGen)	ADFLIP	Recovery	Len
1	MR	GG	0.00	2
2	MHH	SDD	0.00	3
3	MSPPA	LAPLL	0.20	5
4	GSPP	GGPP	0.75	4
5	MSPPT	MGPDD	0.40	5
6	MSPIIV	GGPHHH	0.17	6
7	MGPLS	GPPEV	0.20	5
8	MRPE	GLPP	0.25	4
9	ASPTFS	PPPPPF	0.17	6
10	GSKAP	GGPPP	0.40	5

3.4 Boltz protease-validation outputs

Table 3 reports Boltz confidence metrics for cathepsin B–peptide complexes built from all 10 unique linker sequences.

Table 3: Boltz structure-prediction outputs for cathepsin B–linker peptide complexes (10 designs).

Peptide	Confidence	ptm	iptm	complex_pldt
MRPE	0.758	0.924	0.770	0.755
ASPTFS	0.744	0.878	0.724	0.749
MSPIIV	0.742	0.859	0.743	0.742
GSKAP	0.724	0.911	0.602	0.754
MHH	0.703	0.947	0.598	0.729
MGPLS	0.697	0.958	0.583	0.725
GSPP	0.694	0.959	0.513	0.740
MSPPA	0.686	0.870	0.478	0.738
MSPPT	0.676	0.847	0.363	0.755
MR	0.667	0.847	0.321	0.753

3.5 Peptide physicochemical properties

Table 4 compares representative BoltzGen-native, ADFLIP-refined, and Adcetris Val–Cit reference sequences across the 10-design panel.

Table 4: Representative peptide physicochemical properties. Instability index < 40 = stable (ProtParam).

Source	Seq	GRAVY	Instab.	pI	Charge (pH 7.4)	MW (Da)
BoltzGen rank 1	MR	−1.30	−32.7	9.50	+0.28	305
BoltzGen rank 4	GSPP	−1.10	165.5	5.53	−0.44	356
BoltzGen rank 7	MGPLS	+0.58	8.0	5.28	−0.72	504
BoltzGen rank 8	MRPE	−1.93	80.3	5.97	−0.71	532
ADFLIP rank 5	MGPDD	−1.42	−7.1	4.05	−2.71	534
Adcetris (Val–Cit)	VCIT	+2.63	7.5	5.49	−0.50	435

4 Discussion

With BoltzGen-ADC By treating peptide linkers as **the** preferred cleavable trigger class and designing them *in context* of arbitrary antibody–payload pairs, the pipeline aligns computational design with clinical linker biology (cathepsin cleavage, circulation stability, Fc preservation).

The multi-model architecture—BoltzGen for joint structure–sequence generation, ADFLIP for all-atom sequence refinement, Boltz for protease validation—provides complementary *in silico* proof points rather than relying on a single score. Fc filtering ensures that linker placement does not spuriously engage constant-region surfaces, directly addressing translational risk to effector function.

Data and Code Availability

All code, scripts, and curated outputs are available at https://github.com/Alirzeanoroozi/Merck-Challenge.

References

[1] A. Su, Y. Luo, C. Zhang, and H. Duan, "Linker-GPT: Design of Antibody–Drug Conjugate Linkers with Molecular Generators and Reinforcement Learning." *Scientific Reports*, vol. 15, no. 1, p. 20525, 2025.

[2] B. A. Chabner and T. G. Roberts Jr., Chemotherapy and the war on cancer. *Nature Reviews Cancer*. 2005;5:65–72.

[3] J. M. Lambert and A. Berkenblit, Antibody–drug conjugates for cancer treatment. *Annual Review of Medicine*. 2018;69:191–207.

[4] P. A. Trail, Antibody drug conjugates as cancer therapeutics. *Antibodies*. 2013;2:113–129.

[5] A. Beck, L. Goetsch, C. Dumontet, and N. Corvaïa, Strategies and challenges for the next generation of antibody–drug conjugates. *Nature Reviews Drug Discovery*. 2017;16:315–337.

[6] R. P. Lyon, T. D. Bovee, S. O. Doronina, *et al.*, Reducing hydrophobicity of homogeneous antibody–drug conjugates improves pharmacokinetics and therapeutic index. *Nature Biotechnology*. 2015;33:733–735.

[7] J. D. Bargh, A. Isidro-Llobet, J. S. Parker, and D. R. Spring, Cleavable linkers in antibody–drug conjugates. *Chemical Society Reviews*. 2019;48:4361–4374.

[8] L. Yang, J. Ma, B. Liu, *et al.*, Recent advances in peptide linkers of antibody–drug conjugates. *Journal of Medicinal Chemistry*. 2025;68:18099–18113.

[9] P. R. Hamann, L. M. Hinman, I. Hollander, *et al.*, Gemtuzumab ozogamicin, a potent and selective anti-CD33 antibody–calicheamicin conjugate for treatment of acute myeloid leukemia. *Bioconjugate Chemistry*. 2002;13:47–58.

[10] G. M. Dubowchik, R. A. Firestone, L. Padilla, *et al.*, Cathepsin B-labile dipeptide linkers for lysosomal release of doxorubicin from internalizing immunoconjugates: model studies of enzymatic drug release and antitumor activity. *Bioconjugate Chemistry*. 2002;13:855–869.

[11] B. Turk, Targeting proteases: successes, failures and future prospects. *Nature Reviews Drug Discovery*. 2006;5:785–799.

[12] N. D. Rawlings, A. J. Barrett, P. D. Thomas, *et al.*, MEROPS: the database of proteolytic enzymes, their substrates and inhibitors in 2017. *Nucleic Acids Research*. 2018;46(D1):D624–D632.

[13] J. Wang, S. Lisanza, D. Juergens, *et al.*, Scaffolding protein functional sites using deep learning. *Science*. 2022;377:387–394.

[14] H. Stark *et al.*, BoltzGen: Generative protein design. *bioRxiv*, 2025. Preprint. Available at: https://github.com/HannesStark/boltzgen.

[15] J. Wohlwend *et al.*, Boltz-1: Democratizing biomolecular interaction modeling. *bioRxiv*, 2024. doi: 10.1101/2024.11.19.624167.

[16] K. Yi, K. Jamali, and S. H. W. Scheres, All-atom inverse protein folding through discrete flow matching. *Proceedings of the International Conference on Machine Learning (ICML)*, 2025. Available at: https://openreview.net/forum?id=8tQdwSCJmA

[17] J. A. Francisco, C. G. Cerveny, D. L. Meyer, B. J. Mixan, K. Klussman, D. F. Chace, S. X. Rejniak, K. A. Gordon, R. DeBlanc, B. E. Toki, C. L. Law, S. O. Doronina, C. B. Siegall, P. D. Senter, and A. F. Wahl, cAC10-vcMMAE, an anti-CD30–monomethyl auristatin E conjugate with potent and selective antitumor activity. *Blood*, 102(4):1458–1465, 2003. doi: 10.1182/blood-2003-01-0034.

[18] G. M. Dubowchik *et al.*, Cathepsin B-labile dipeptide linkers for lysosomal release of doxorubicin from internalizing immunoconjugates: model studies of enzymatic drug release and antitumor activity. *Bioconjugate Chemistry*, 13(4):855–869, 2002. doi: 10.1021/bc025506x.