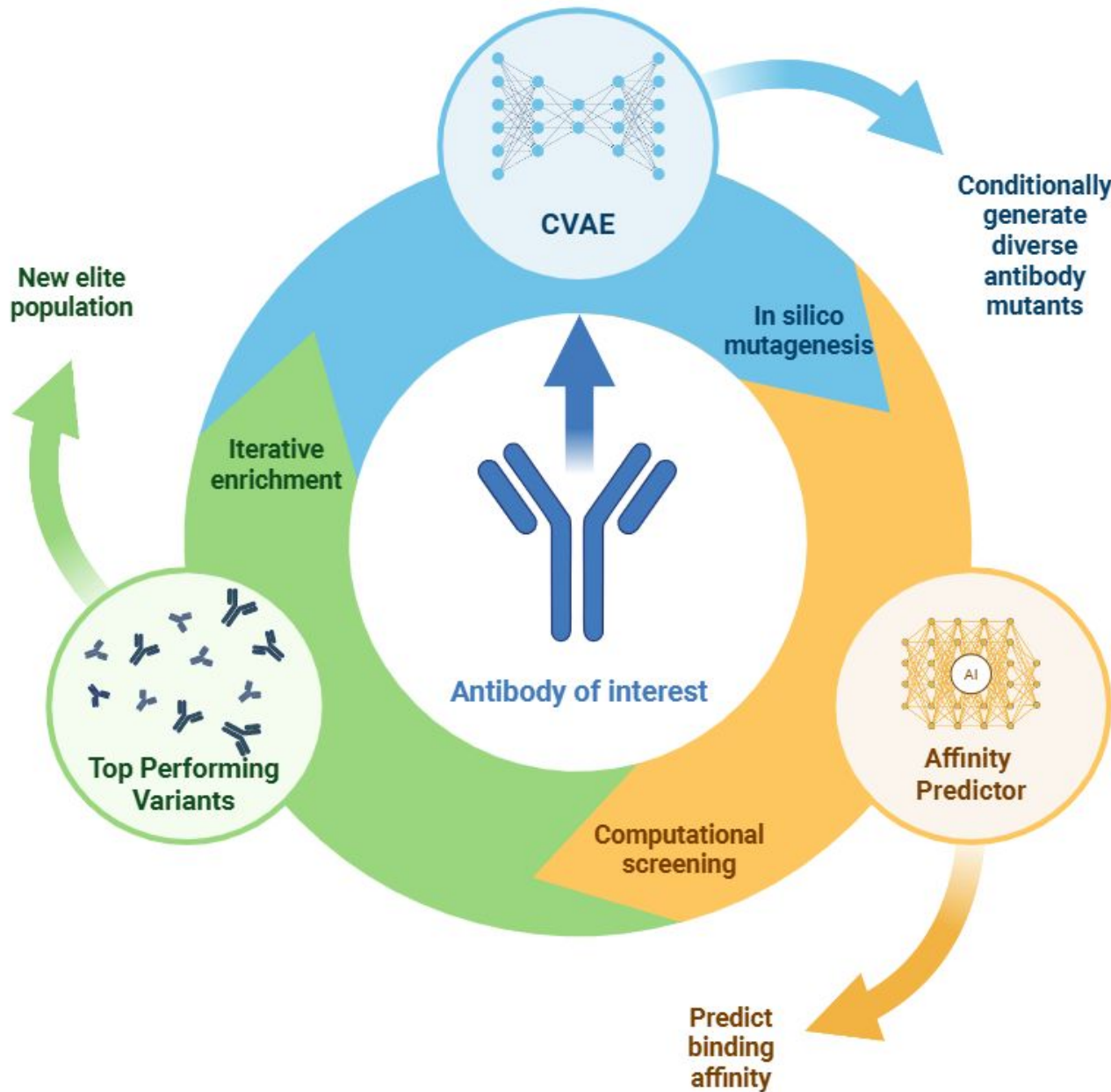# TeBaAb: <u>Te</u>xt-<u>Ba</u>sed Antigen-Conditioned <u>A</u>nti<u>b</u>ody Redesign via Directed Evolution

Cuong Manh Nguyen, Huy-Hoang Do-Huu, Viet Thanh Duy Nguyen, Truong-Son Hy

University of Alabama at Birmingham
Birmingham, AL 35294, United States

Introduction

The TeBaAb pipeline: a Conditional Variational Autoencoder (CVAE) generates candidate sequences conditioned on antigen sequences and textual descriptions, a two-stage affinity predictor evaluates binding strength, top variants are iteratively fed back into the CVAE, forming an in-silico directed evolution loop. This process yields antibodies with enhanced binding affinity while maintaining structural integrity.

Dataset: AbDes - 7,684 samples

A novel dataset of antibody–antigen complexes with descriptive annotations, structures, and binding affinities.

Antibody Sequence	Antigen	Description	Binding Affinity (ΔG, kJ/mol)
QVQLV...|QS ALT...	VVKFMDVY...	Vascular endothelial growth factor in complex with a neutralizing antibody, classified as an immune system, derived from mus musculus and expressed in escherichia coli, forms a Hetero 6-mer with Cyclic - C2 symmetry.	-11.55
QVQLQ...|QV QLQ...	KVFGRCEL...	Hen egg white lysozyme, d18a mutant, in complex with mouse monoclonal antibody d1.3, classified as a complex (immunoglobulin/hydrolase), derived from mus musculus and expressed in escherichia coli, forms a Hetero 3-mer with Asymmetric - C1 symmetry, and has pseudo-symmetry of Asymmetric - C1 with Hetero 3-mer stoichiometry.	-10.45
QIQLVQ...|DI VMT...	IRDFNNLT...	Refined crystal structure of the influenza virus n9 neuraminidase-nc41 fab complex, classified as a hydrolase(o-glycosyl), derived from influenza a virus (a/tern/australia/g70c/1975(h11n9)), forms a Hetero 12-mer with Cyclic - C4 symmetry.	-11.02

Methodology

Feature Extraction   **CVAE**
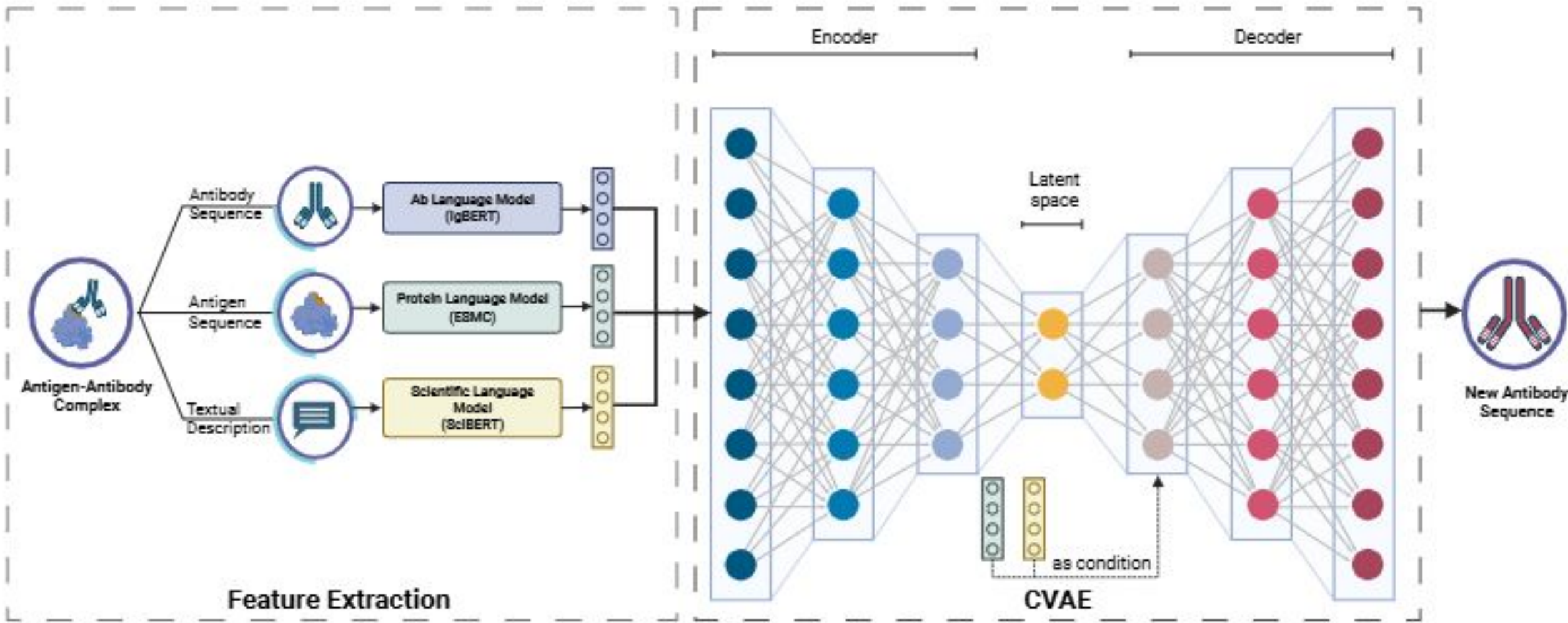
The CVAE integrates antibody embeddings (IgBERT), antigen embeddings (ESMC), and textual description embeddings (SciBERT) to generate antibody sequences. The encoder produces a latent distribution, and the transformer-based decoder reconstructs new antibody sequences conditioned on both antigen sequences and textual descriptions, enabling antigen-specific and text-guided antibody design.

Algorithm 1 TeBaAb In Silico Directed Evolution

Input: Initial antibody set S_0, antigen sequence A, text description D
  Number of generations G, top-K selection size, mutation batch size B
Output: Optimized antibody set S_G

1 Initialize generation counter $g \leftarrow 0$ Initialize population $S_0 = \{s_1, s_2, \ldots, s_K\}$
2 **while** $g < G$ **do**
3   Initialize empty candidate pool $C_g \leftarrow \emptyset$
4   **foreach** *antibody* $s \in S_g$ **do**
5   Generate B variants $\{s'_1, \ldots, s'_B\}$ via CVAE conditioned on (s, A, D) **foreach** *variant* s'_i **do**
6   Predict binding affinity $\hat{y}(s'_i, A)$ using affinity predictor Add $(s'_i, \hat{y})$ to C_g
7   **end**
8   **end**
9   Combine S_g and C_g into one candidate pool Select top K sequences with lowest predicted $\hat{y}$ to form S_{g+1} $g \leftarrow g + 1$
10 **end**
11 **return** final optimized set S_G

Results

	Average Predicted Binding Affinity (ΔG) (kJ/mol)	Improvement (%)	Diversity	Novelty
Original	-10.04 + 0.13			
TeBaAb (with description)	-11.60 + 0.14	15.5	87.73	28.95
TeBaAb (without description)	-10.93 + 0.12	8.9	86.70	26.11

	Average Structural Confidence							
	Framework H-chain	CDR-H1	CDR-H2	CDR-H3	Framework L-chain	CDR-L1	CDR-L2	CDR-L3
Original	0.315	0.303	0.190	0.188	0.238	0.252	0.182	0.235
TeBaAb (with description)	0.362	0.290	0.191	0.196	0.260	0.307	0.192	0.238

Conclusion

- **TeBaAb** enables text-guided, antigen-specific, and structurally robust antibody design.
- Opens new directions for **controllable protein engineering**.
- Future work will extend textual conditioning to a **wider range of properties**, incorporate **structural constraints**,