We thank all the reviewers for their thoughtful and constructive feedback. We are encouraged to hear that the reviewers find the paper novel (**R2** , **R3**), tackles a critical problem (**R3**), well written (**R2** , **R3** , **R4**), has extensive experiments (**R3**), is appropriately positioned with respect to prior work (**R2** , **R3** , **R4**) and provides impressive results (**R3**). Moreover, **R4** acknowledges the completeness of our analysis and believes that we have worked hard to show the validity of the proposed results. **R2** acknowledges that we incorporate multiple aspects of the generated molecules like binding affinity, selectivity, toxicity, etc., in our pipeline. We answer specific questions posed by the reviewers below and will incorporate all feedback in the revision. We apologize for the terseness of the replies due to space limitations.

R1 : **Benchmarking of VAE with respect to a null model (random sampling of chemical space).** We use the MOSES benchmarking framework for evaluating the molecular VAE model and perform extensive analysis of the generated molecules in *Supp. Mat. Sec. B.* Comparison to other related molecular generative models (JT-VAE, AAE, etc) is in *Supp. Mat. Tables B.1 and B.2*. Comparison of controlled vs random sampling is in *Table 1*. Our objective is to show that controlled sampling on our VAE model can generate novel, diverse, targeted, drug-like molecules, when trained on ZINC+BindingDB data. Molecules selected through random sampling of the chemical space will not have these properties and hence are not included in the MOSES benchmarking framework as well.

R1 : **Why generating high-affinity ligands is more challenging for NSP9?** In *Supp. Mat. Sec A*, we perform a BLAST similarity search of MPro, RBD and NSP9 to the training data and show that NSP9 has the least sequence similarity to the training data and is therefore more novel compared to other targets and hence more challenging.

R1 : **Tanimoto similarity - specific algorithm and parameters used.** In *Supp. Mat. Sec E.*, we mention that we use MACCS keys [58] for fingerprint generation. MACCS uses fixed 166-bit keys and has no additional parameters.

R1 : **Details of docking program and protocol.** Details of the docking program (Autodock Vina) and the protocol are explained in *Supp. Mat. Sec. H.*

R1 : **Docking score for a random sample of ZINC ligands.** Our experiments reveal that CogMol consistently yields higher percentage of low binding energy (< -7 kcal/mol) molecules compared to random ZINC ligands – this difference is 37%, 39% and 22% for RBD, MPro and NSP9 respectively. We will add these results in Table 3 in final manuscript.

R1 : **Affinities, targets, and selectivity of all generated molecules that match exactly with PubChem.** Avg. Affinity/Selectivity/Fraction of matches for MPro (7.4/1.3/0.018), RBD (7.6/1.1/0.005), NSP9 (6.8/0.89/0.014).

R1 : **Limitations of docking scores.**We agree with **R1** that the docking scores may provide limited information due to simplification of various physical properties, such as the ignoring water and protein dynamics. However, recent work (Gaillard et al., J. Chem. Inf. Model. 2018, 58, 8, 1697–1706) shows Autodock Vina scoring function is in the first quarter (Vina) among all methods tested in CASF-2013 benchmark. Therefore, we use Vina scores as the first step towards evaluating generated molecules in terms of target structure binding.

R1 : **Some claims in the conclusions of the study require more supporting and experimental data to provide sufficient evidence to support them.** We have performed extensive in silico experiments to validate the generated molecules on multiple criteria (Parent Molecule and Metabolite Toxicity, Affinity based on an independent machine learning model, Selectivity, Synthetic Feasibility, Target Structure Docking, Number of retrosynthetic steps, etc). See *Section 5 and Supp. Materials*.

R2 : **authors have no choice but to rely on an unvalidated folded structure, or use docking methods that only take their sequence into account.** Docking was done using x-ray crystal structures of target proteins available in Protein Data Bank, whereas the generative model relied on protein sequence information only.

R2 : **The diverse set of methods used means that there is no underlying theoretical framework for the approach**. Our work is the first deep generative approach that generates novel, specific, and selective drug-like small molecules for an unseen target protein sequence without requiring protein-specific model retraining. The diverse set of screening methods provide a comprehensive analysis of generated molecules to validate the generated molecules and show the efficacy of the generative framework.

R2 : **Prioritization among the 3500 compounds that are generated by the approach.** Great suggestion! We will provide a discussion around different prioritization schemes. For example, one scheme will be based on number of synthesis steps, and the cost and availability of ingredients.

R3 : **Reproducibility.** All training datasets are publicly available. Also, the generated molecules and their properties are available for download using our molecule explorer tool. We did not provide a URL of the tool to keep our submission anonymous during review, but will add it upon acceptance. Some of the molecular evaluation components are already publicly available (*e.g.* rxn4chemistry package , Autodock Vina) We have also provided extensive details of the architecture and parameters of our models in the *Supp. Mat.* and will add any specific details we may have missed. We are also working on releasing the source code of some other components upon acceptance.