In this section, we provide additional data analysis on our training and test data upon the reviewer's request.

Training Data

We pre-train AMOLE with the PubChem database, which is one of the most extensive public molecular databases available. PubChem database consists of multiple data sources including DrugBank, CTD, PharmGKB, and more. Please refer to the following URL for more details: https://pubchem.ncbi.nlm.nih.gov/sources/.

The PubChem database we used during training comprises a total of 299K unique molecules and 336K molecule-text pairs. During preprocessing, we consolidate each expertise into a unified description. Thus, each description for an individual molecule originates from a distinct database (expertise). On average, molecules are associated with 1.115 descriptions, with a maximum of 17 descriptions and a minimum of one. We provide a histogram and boxplot on the number of descriptions per molecule in Figure 1.

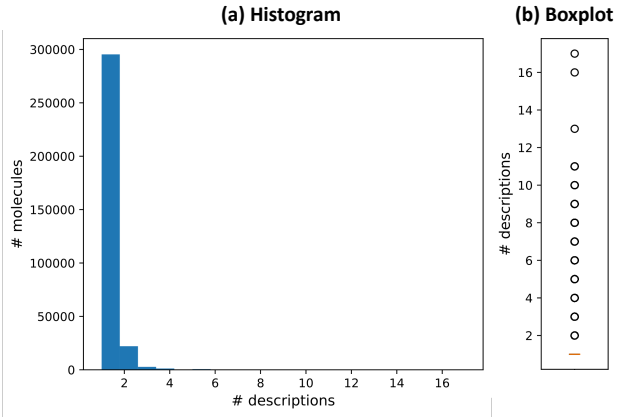

Figure 1: (a) Histogram and (b) Boxplot on the number of descriptions per molecule.

Each description in training data consists of 17.62 words on average, with a maximum of 874 words and a minimum of one. We also provide a histogram on the number of words per description in Figure 2.

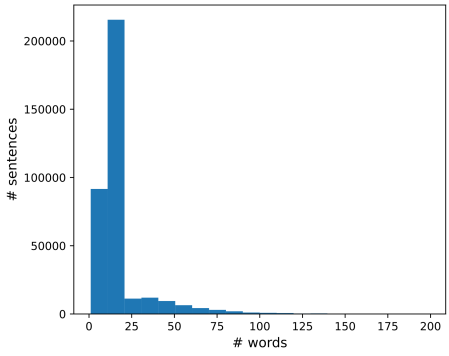

Figure 2: (a) Histogram on the number of words per description.

Test Data

In this study, we undertake experiments on four downstream tasks: zero-shot cross-modal retrieval, zero-shot question and answering, molecular property prediction, and zero-shot virtual screening. Detailed statistics for the molecular property prediction and zero-shot virtual screening datasets are available in Appendices C.3 and C.4, respectively. However, we realize that a comprehensive explanation for the zero-shot cross-modal retrieval and zero-shot question and answering tasks was omitted.

We employ testing data from a prior study, i.e., MoleculeSTM, ensuring no overlap between training and testing datasets. Specifically, molecules are excluded from the testing dataset if they share the same canonical SMILES as those in the training dataset. Moreover, for the ATC dataset, exclusion criteria also consider high similarity between textual descriptions in addition to identical canonical SMILES. For further details, please see Appendix C.1 in the MoleculeSTM documentation.

Field	Data
Description	1,154
Pharmacodynamics	1,005
ATC	3,007

Table 1: Number of paired data for zero-shot cross-modal retrieval and zero-shot question and answering task.

For the zero-shot question and answering task, we create one question per given molecule-text pair, as detailed in Appendix C.2. As a result, the total number of questions generated matches the number of pairs utilized in the zero-shot retrieval task.