hydrolyase work shows that functional proteins designed using AI may look nothing like those found in nature, thus reducing the strength of current best match screening approaches [24, 43]. A red teaming study by Wittmann *et al.* demonstrated that the BSS used by many companies *circa* 2023 failed to detect a significant portion of SOCs deliberately obfuscated via an AIPD pipeline [44].

Despite mitigation of the vulnerability reported in Wittmann *et al.*, concerns about its severity remain due to the lack of experimental validation for AI-generated protein designs. This highlights a key question: Can AI accurately predict protein activity from sequence or structure? Powerful protein structure prediction models, such as AlphaFold and RoseTTAFold, excel precisely because they are trained using experimental structural data—the very same type of information they aim to predict [1, 4]. In contrast, protein activity prediction faces a fundamental data gap: AIPD models must extrapolate functional properties from sequence and structural training data encoding complex biochemical interactions. To begin to address this gap, we employed Wittmann *et al.*'s AIPD pipeline and developed an experimental validation framework, enabling us to investigate how *in silico* metrics correspond with empirical activity measurements of synthetic protein homologs.

2 Study Considerations and Process

To perform experimental validation safely and responsibly, we chose safe proxy protein targets rather than SOCs. Our testing, evaluation, validation, and verification (TEVV) framework used proteins of different length and classes of functional complexity to validate AI prediction through measurements of protein activity. These measurements and our validation framework can inform both future AIPD models and biosecurity strategy, and allowed us to begin to answer the outstanding question: **Is it feasible to reliably rewrite the sequence of a known protein beyond identification by BSS using open-source AIPD models, while preserving protein activity in an experimental context?**

2.1 Target Proposal and Rationale

To define a broad benchmark for rewriting protein sequences using AIPD, we selected three initial protein targets assigned to basic, moderate, or advanced "difficulty classes". We consider a protein active if it expresses, folds correctly, and interacts appropriately. We assume the difficulty of preserving protein activity scales with increasing sequence length and functional repertoire. Because the AIPD models we used generate sequences one amino acid at a time, the chance that a generated amino acid could result in an inactive final protein increases with sequence length. Similarly, as a protein's functional repertoire expands and its complexity grows, the biophysical constraints that govern its behavior become increasingly stringent, making the generation process more susceptible to errors. For example, SH3 and PDZ domains, which facilitate binding, are known for their general tolerance to substitutions in ligand recognition regions [25, 27]. In contrast, the activity of metabolic decarboxylase enzymes, which coordinate multiple interaction steps, is more acutely sensitive to amino acid changes [30]. Therefore, we anticipated that preserving a protein's active state after rewriting the sequence using AI becomes increasingly challenging, for longer or more functionally diverse protein sequences.

To select basic, moderate, and advanced protein targets, we considered candidates with known key residues and published activity assays. For the basic class, we selected the third PDZ (PDZ3) domain from the human postsynaptic density protein 95 (PSD95/DLG4, UniProt: P78352). Binding between PDZ3 and cysteine-rich PDZ-binding protein (CRIPT, UniProt: Q9P021), one of its protein ligands, can be approximated as a classical lock-and-key interaction [32]. In the moderate class, we selected orotidine 5'-phosphate decarboxylase from *Saccharomyces cerevisiae* (baker's yeast), known more commonly by its encoding gene URA3 (UniProt: P03962). As an enzyme, URA3 adds the challenge of mechanically or chemically lowering the activation energy of a chemical reaction [14]. To represent the advanced class, we selected the T7 bacteriophage RNA polymerase (RNAP) (UniProt: P00573). T7 RNAP is an enzyme often used for *in vitro* transcription in cell-free expression (CFE) systems, due to its monomeric nature, efficiency, and fast rate of messenger RNA (mRNA) production [39]. Active RNAP synthetic homologs must not only fold with the proper geometry and surface charge distribution to localize to a specific DNA promoter sequence, but also coordinate binding of multiple substrates and motion along DNA. These three protein targets offer a suitable scale of increasing sequence length