TARGETING AGGREGATING PROTEINS WITH LANGUAGE MODEL-DESIGNED DEGRADERS

Rio Watson[1], **Kishan Patel**[1], **Tong Chen**[2], **Pranam Chatterjee**[1,2,3,†]
[1]Department of Biomedical Engineering, Duke University
[2]Department of Computer Science, Duke University
[3]Department of Biostatistics and Bioinformatics, Duke University

[†]Corresponding author: pranam.chatterjee@duke.edu

ABSTRACT

Protein aggregation drives several neurological diseases and pediatric cancers, yet current inhibitors fail to directly target aggregating proteins or provide long-term disease modification. Advances in generative artificial intelligence (AI), particularly protein language models (pLMs), have enabled the design of peptide binders for disordered and oncogenic targets. Using these models, we designed peptide binders for mutant GFAP (Alexander Disease) and PAX3::FOXO1 (Alveolar Rhabdomyosarcoma). When fused to E3 ubiquitin ligase domains, these binders selectively degrade their targets, and in the case of GFAP, reducing aggregation. Our results demonstrate that pLM-designed peptide-guided degraders provide a powerful strategy for treating aggregation-driven diseases.

1 INTRODUCTION

A significant number of diseases are driven by pathological protein aggregation (Aguzzi & O'Connor, 2010). Several neurodegenerative disorders, including Alzheimer's disease (AD) and Alexander disease (AxD), are characterized by protein aggregation in the central nervous system (Tiwari et al., 2019; Battaglia et al., 2019; Tang et al., 2010). Additionally, cancer-driving fusion oncoproteins, such as PAX3::FOXO1 in alveolar rhabdomyosarcoma (ARMS), have been shown to aggregate within cancer cells (Vincoff et al., 2024). These aggregates often result from genetic alterations, including chromosomal translocations that create oncogenic fusion proteins (e.g., PAX3::FOXO1 in ARMS) (Vincoff et al., 2024) or missense mutations (e.g., GFAP in AxD) (Hampel et al., 2021).

PAX3::FOXO1 is a transcription factor that aberrantly activates genes involved in cell proliferation, promoting tumorigenesis in ARMS (Kikuchi et al., 2008). In AxD, gain-of-function missense mutations, particularly in arginine residues such as R79 in glial fibrillary acidic protein (GFAP), lead to its aggregation into Rosenthal fibers (RFs) (Battaglia et al., 2019). GFAP, an intermediate filament protein essential for astrocyte structure and function, forms these pathological aggregates, disrupting normal astrocytic activity and driving disease progression (Battaglia et al., 2019; Tang et al., 2010).

A growing class of targeted protein degradation (TPD) strategies has emerged to selectively eliminate pathogenic proteins. One such approach involves ubiquibodies (uAbs), which are engineered by fusing peptide binders with E3 ubiquitin ligase domains, directing target proteins for degradation (Brixi et al., 2023; Bhat et al., 2025). Unlike small-molecule degraders that require cryptic binding pockets, uAbs can be designed to target proteins lacking structured domains, including disordered proteins and fusion oncoproteins (Vincoff et al., 2025).

Advances in protein language models (pLMs) have further enabled the rational design of peptide binders directly from sequence data. Models such as PepMLM, PepPrCLIP, SaLT&PepPr, and moPPIt leverage learned sequence representations to generate binders with high specificity, even in the absence of structural information (Brixi et al., 2023; Bhat et al., 2025; Chen et al., 2023b;a). These AI-driven approaches allow for the rapid identification and optimization of binders, expanding the range of targetable disease-associated proteins.

To counteract the pathological effects of AxD and ARMS, we applied this TPD strategy to degrade mutant GFAP and PAX3::FOXO1. Existing degradation approaches, such as molecular glues (Dong et al., 2021) and PROTACs (Békés et al., 2022), leverage the ubiquitin-proteasome system (UPS) but are limited by the absence of accessible or cryptic binding pockets (Zhao et al., 2022). Using the generative protein language models PepMLM (Chen et al., 2023a), moPPit(Chen et al., 2024), and muPPit, we designed peptide binders targeting mutant GFAP and PAX3::FOXO1. When genetically integrated into uAbs, these binders facilitated selective degradation of PAX3::FOXO1 and GFAP, including the aggregate-prone isoforms of GFAP.

In total, our study demonstrates that AI-designed peptide-guided degraders provide a promising strategy for targeting aggregating proteins in both pediatric cancers and neurological diseases, offering a novel therapeutic avenue for conditions driven by aberrant protein aggregation.

2 METHODS

2.1 GENERATION OF PLASMIDS

The binders used to engineer the uAbs were generated by the previously described PepMLM algorithm(Chen et al., 2023a). The uAb vectors were generated from the standard pcDNA3 vector, harboring a cytomegalovirus (CMV) promoter. An Esp3I restriction site was introduced immediately upstream of the E3 ligase CHIP catalytic domain and flexible GSGSG linker. The oligonucleotides corresponding to the computationally-derived binders were annealed and ligated via T4 DNA ligase into the Esp3I-digested vector.

2.2 CELL CULTURE

RH4, RH30, and SW13 Vim(-) cells were maintained in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 100 units/ml penicillin, 100 mg/ml streptomycin, and 10% fetal bovine serum (FBS). The PepMLM peptides (800 ng) plasmids were transfected into cells (4x105/well in a 12-well plate) with Lipofectamine 2000 (Invitrogen) in Opti-MEM (Gibco).

2.3 IMMUNOBLOTTING

The above-mentioned cell lines were all collected 72 hours post-transfection, and subsequently lysed using RIPA buffer (Thermo) on ice. The lysed cells incubated on ice for 30 minutes followed by centrifugation at 15,000 rpm for 10 minutes at 4C. The supernatant was collected and quantified using the BCA Protein Assay Kit (Sigma). Immunoblotting was performed using precast 8% gradient gels (Thermo) and transferred onto polyvinylidene difluoride membranes (PVDF) membranes using the iBlot™ 2 Transfer Stacks and iBlot2 transfer system (Thermo). For FOXO1 blots, the membranes were blocked in 1% BSA in TBS-T (Thermo) and probed with rabbit anti-FOXO1 antibody (CST, 1:1000), and GFAP blots were blocked in 5% skim milk powder in 1x TBS-T 4C overnight, then probed with rabbit GFAP antibody (DAKO, 1:5000). The membranes were washed three times with 1x TBS-T, then probed with HRP-conjugated anti-rabbit antibody (abcam, 1:5000) for 1h at room temperature, then washed as previously described. The membranes were then incubated with PICO Western Chemiluminescent HRP Substrate (Thermo) and imaged with iBright5000 chemiluminescence detector (Thermo).

2.4 COOMASSIE STAINING

Transfected cells were collected then lysed with Triton X buffer (Millipore), then centrifuged at 15,000rpm for 10 minutes at 4C. The supernatant was removed, and the remaining insoluble pellet was resuspended in Tris-Glycine SDS buffer (Thermo). The resuspended insoluble fraction was next resolved on 4–20% gradient gels (Thermo) transferred onto PVDF membranes using the iBlot2 transfer system described above. The transferred gels were stained with SimplyBlue™ SafeStain (Thermo) and imaged with the ChemiDoc imaging system (BioRad).

2.5 Statistical Analysis

All experiments were performed in independent biological triplicates ($n = 3$). Image analysis was conducted offline using ImageJ. Statistical significance was determined using one-way ANOVA, with significance thresholds defined as $*p \leq 0.05$, $**p \leq 0.01$, $***p \leq 0.001$, $****p \leq 0.0001$, and $*****p \leq 0.00001$.

3 Results

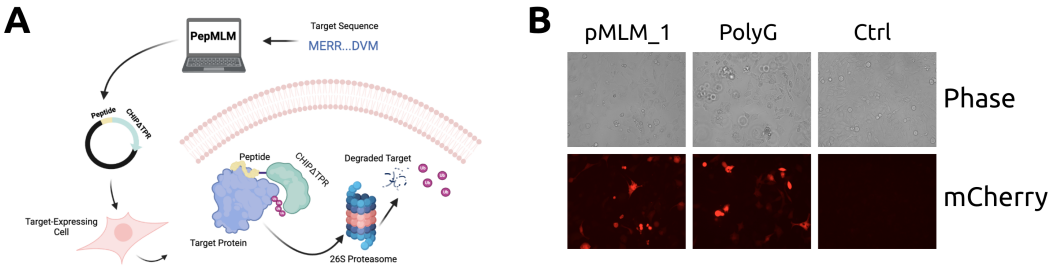

Figure 1: uAb transfection and testing. A) pLM-derived binders are genetically cloned into uAb vectors and transfected into cells to test degradation. B) Transfection of uAbs into RH30 cells as depicted by the mCherry fluorescence.

3.1 uAb Engineering and Delivery

To design our uAbs, we first generated peptide binders predicted to bind our target proteins using a variety of algorithms. We first used PepMLM, a protein language model that employs a novel span masking strategy, positioning cognate peptide sequences at the C-terminus of target protein sequences (Chen et al., 2023a). The resulting peptide sequences, chosen for their low pseudo-perplexities (Supplementary Table 1), were converted into oligonucleotides and cloned into uAb vectors, replacing an Esp3I restriction site with the oligonucleotide inserts. In addition, we utilized moPPit and muPPit, algorithms which respectively generate motif-specific and mutant-specific peptide binders(Chen et al., 2024), to target various isoforms of GFAP.

Each uAb was engineered to position the peptide sequence immediately upstream of the CHIP E3 ligase catalytic domain, followed by an mCherry reporter gene (Figure 1A). We then transiently transfected these uAbs into disease-relevant cell lines, such as the RH30 ARMS cell line. After 72 hours of incubation, mCherry fluorescence was detected via microscopy, confirming successful transfection (Figure 1B).

3.2 uAbs degrade aggregating proteins

To evaluate whether our uAbs can degrade aggregating proteins, we first tested their ability to degrade PAX3::FOXO1 by liposomally transfecting uAbs into RH4 cells, which endogenously express the fusion protein. Immunoblotting analysis revealed a decrease in PAX3::FOXO1 expression in ARMS cells treated with PepMLM-derived uAbs compared to controls (Figure 2A).

Next, we examined whether our uAbs could also degrade mutant GFAP. SW13 cells lacking vimentin (Vim), another intermediate filament (IF) protein, were co-transfected with a vector expressing R79H GFAP along with our uAbs. Immunoblotting showed a significant reduction in GFAP expression in uAb-treated cells compared to the PolyG control, which contains a poly-glycine sequence cloned into the uAb vector as a negative control peptide (Figure 2B). Since R79H GFAP oligomerization is a key marker of GFAP aggregation severity (Battaglia et al., 2019) and has been shown to inhibit proteasomal degradation (Tang et al., 2010), we assessed whether uAbs could degrade GFAP oligomers. Coomassie staining of the insoluble fraction of uAb-treated SW13 Vim(-) cells confirmed a decrease in oligomeric GFAP in cells treated with two of our uAbs (Figure 2C). In addition to oligomerization, it has been shown that Caspase-6-induced cleavage of mutant GFAP

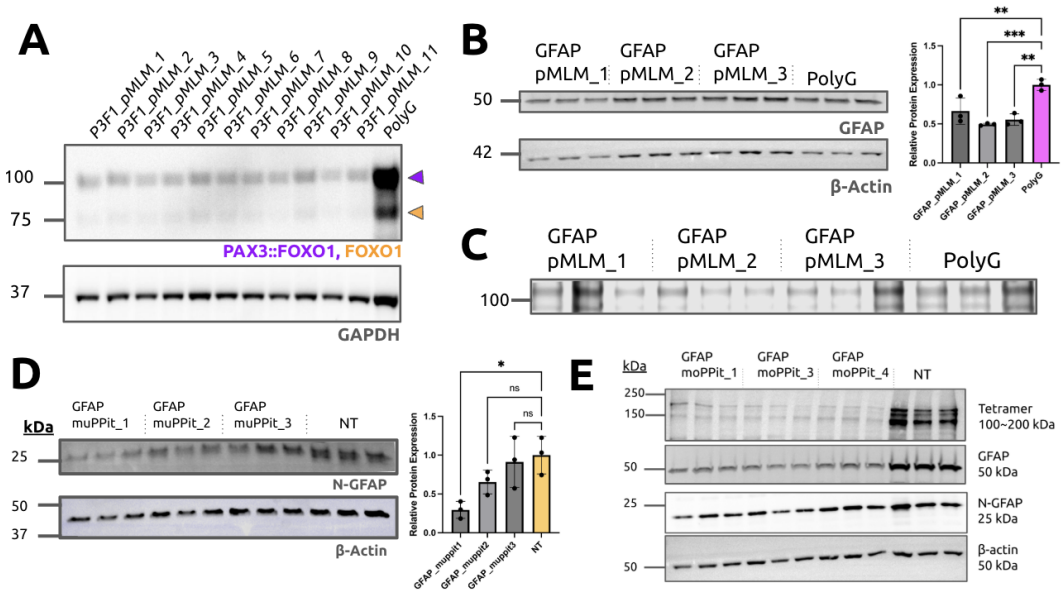

Figure 2: uAb-treated cells mitigate disease phenotypes. A) PAX3::FOXO1 protein expression in uAb-treated RH4 cells measured via Western blotting. B) GFAP protein expression in uAb-treated SW13 Vim(-) cells measured via Western blotting. C) GFAP oligomerization measured via Coomassie staining. D) D) N-GFAP protein expression in muPPit-derived uAb-treated SW13 Vim(-) cells measured via Western blotting. E) Expression of tetrameric GFAP, monomeric GFAP, and N-GFAP in moPPit-derived uAb-treated cells measured via Western blotting.

leads to an aggregation-prone N-GFAP cleavage product(Chen et al., 2013). Thus, we assessed the capabilities of muPPit and moPPit-derived uAbs through co-transfection of SW13 Vim(-) cells with R79H and the respective uAbs. Both muPPit-and moPPit-derived uAbs were able to significantly degrade N-GFAP. Furthermore, all three of the transfected moPPit-derived uAbs resulted in significant degradation of monomeric and tetrameric mutant GFAP. These results suggest that uAbs can facilitate the degradation of aggregated proteins such as PAX3::FOXO1 and mutant GFAP.

4 Discussion

Our findings demonstrate that uAbs effectively degrade disease-associated aggregating proteins, including the oncogenic fusion protein PAX3::FOXO1 and the mutant intermediate filament GFAP. By leveraging peptide binders designed using generative pLMs, we achieve potent degradation of these targets, leading to reduced protein aggregation and, in the case of ARMS, decreased tumor cell viability. These results highlight the potential of AI-driven peptide-guided degradation as a therapeutic strategy for diseases previously considered undruggable due to the lack of defined binding pockets.

Our future efforts will focus on ensuring aggregate-selective or fusion-specific degradation, either by using aggregation-focused degrader platforms Benn et al. (2024) or by designing new pLM algorithms that capture fusion-specific features Vincoff et al. (2025). Furthermore, we hope to extend this approach to other aggregating proteins implicated in neurodegenerative disorders, such as Alzheimer's and Parkinson's disease, as well as pediatric cancers, such as Ewing sarcoma and synovial sarcoma. Our findings in this work thus lay the foundation for AI-driven, peptide-based targeted protein degradation as a promising strategy for treating protein aggregation-associated diseases.

REFERENCES

Adriano Aguzzi and Tracy O'Connor. Protein aggregation diseases: pathogenicity and therapeutic perspectives. *Nature Reviews Drug Discovery*, 9:237–248, 2010. URL https://api.semanticscholar.org/CorpusID:5756683.

Rachel A. Battaglia, Adriana S. Beltran, Samed Delic, Raluca Ana Maria Dumitru, Jasmine Robinson, Parijat Kabiraj, Laura E. Herring, Victoria J. Madden, Namritha Ravinder, Erik Willems, Rhonda A. Newman, Roy A. Quinlan, James E. Goldman, Ming Der Perng, Masaki Inagaki, and Natasha T. Snider. Site-specific phosphorylation and caspase cleavage of gfap are new markers of alexander disease severity. *eLife*, 8, 2019. URL https://api.semanticscholar.org/CorpusID:207897946.

Miklós Békés, David R. Langley, and Craig M. Crews. Protac targeted protein degraders: the past is prologue. *Nature Reviews. Drug Discovery*, 21:181 – 200, 2022. URL https://api.semanticscholar.org/CorpusID:246034014.

Jonathan Benn, Shi Cheng, Sophie Keeling, Annabel E. Smith, Marina J. Vaysburd, Dorothea Böken, Lauren V. C. Miller, Taxiarchis Katsinelos, Catarina Franco, Elian Dupré, Clément Danis, Isabelle Landrieu, Luc Buée, David Klenerman, Leo C. James, and William A. McEwan. Aggregate-selective removal of pathological tau by clustering-activated degraders. *Science*, 385(6712):1009–1016, August 2024. ISSN 1095-9203. doi: 10.1126/science.adp5186. URL http://dx.doi.org/10.1126/science.adp5186.

Suhaas Bhat, Kalyan Palepu, Lauren Hong, Joey Mao, Tianzheng Ye, Rema Iyer, Lin Zhao, Tianlai Chen, Sophia Vincoff, Rio Watson, et al. De novo design of peptide binders to conformationally diverse targets with contrastive language modeling. *Science Advances*, 11(4):eadr8638, 2025.

Garyk Brixi, Tianzheng Ye, Lauren Hong, Tian Wang, Connor Monticello, Natalia Lopez-Barbosa, Sophia Vincoff, Vivian Yudistyra, Lin Zhao, Elena Haarer, et al. Salt&peppr is an interface-predicting language model for designing peptide-guided protein degraders. *Communications Biology*, 6(1):1081, 2023.

Mei-Hsuan Chen, Tracy L. Hagemann, Roy A. Quinlan, Albee Messing, and Ming Der Perng. Caspase cleavage of gfap produces an assembly-compromised proteolytic fragment that promotes filament aggregation. *ASN NEURO*, 5, 2013. URL https://api.semanticscholar.org/CorpusID:15193394.

Tianlai Chen, Madeleine Dumas, Rio Watson, Sophia Vincoff, Christina Peng, Lin Zhao, Lauren Hong, Sarah Pertsemlidis, Mayumi Shaepers-Cheu, Tian Wang, Divya Srijay, Connor Monticello, Pranay Vure, Rishab Pulugurta, Kseniia Kholina, Shrey Goel, Matthew P. DeLisa, Ray Truant, Hector C. Aguilar, and Pranam Chatterjee. Pepmlm: Target sequence-conditioned generation of therapeutic peptide binders via span masked language modeling. *ArXiv*, 2023a. URL https://api.semanticscholar.org/CorpusID:263831645.

Tianlai Chen, Madeleine Dumas, Rio Watson, Sophia Vincoff, Christina Peng, Lin Zhao, Lauren Hong, Sarah Pertsemlidis, Mayumi Shaepers-Cheu, Tian Zi Wang, Divya Srijay, Connor Monticello, Pranay Vure, Rishab Pulugurta, Kseniia Kholina, Shrey Goel, Matthew P. DeLisa, Ray Truant, Hector C. Aguilar, and Pranam Chatterjee. Pepmlm: Target sequence-conditioned generation of therapeutic peptide binders via span masked language modeling. *arXiv*, 2023b. doi: 10.48550/ARXIV.2310.03842. URL https://arxiv.org/abs/2310.03842.

Tong Chen, Yinuo Zhang, and Pranam Chatterjee. moppit: De novo generation of motif-specific binders with protein language models. *bioRxiv*, 2024. URL https://api.semanticscholar.org/CorpusID:271645405.

Guoqiang Dong, Yu Ding, Shipeng He, and Chunquan Sheng. Molecular glues for targeted protein degradation: From serendipity to rational discovery. *Journal of medicinal chemistry*, 2021. URL https://api.semanticscholar.org/CorpusID:236471135.

Harald Hampel, John Hardy, Kaj Blennow, Christopher Chen, George Perry, Seung Hyun Kim, Victor L. Villemagne, Paul Aisen, Michele Vendruscolo, Takeshi Iwatsubo, Colin L. Masters,

Min Cho, Lars Lannfelt, Jeffrey L. Cummings, and Andrea Vergallo. The amyloid-β pathway in alzheimer's disease. *Molecular Psychiatry*, 26:5481 – 5503, 2021. URL https://api.semanticscholar.org/CorpusID:237349387.

Ken Kikuchi, Kunihiko Tsuchiya, Osamu Otabe, Takahiro Gotoh, Shinichi Tamura, Yoshiki Katsumi, Shigeki Yagyu, Satoko Tsubai-Shimizu, Mitsuru Miyachi, Tomoko Iehara, and Hajime Hosoi. Effects of pax3-fkhr on malignant phenotypes in alveolar rhabdomyosarcoma. *Biochemical and biophysical research communications*, 365 3:568–74, 2008. URL https://api.semanticscholar.org/CorpusID:12218977.

Guomei Tang, Ming Der Perng, Sherwin Wilk, Roy A. Quinlan, and James E. Goldman. Oligomers of mutant glial fibrillary acidic protein (gfap) inhibit the proteasome system in alexander disease astrocytes, and the small heat shock protein αb-crystallin reverses the inhibition*. *The Journal of Biological Chemistry*, 285:10527 – 10537, 2010. URL https://api.semanticscholar.org/CorpusID:22937459.

Sneham Tiwari, Venkata Atluri, Ajeet Kumar Kaushik, Adriana Yndart, and Madhavan Nair. Alzheimer's disease: pathogenesis, diagnostics, and therapeutics. *International Journal of Nanomedicine*, 14:5541 – 5554, 2019. URL https://api.semanticscholar.org/CorpusID:199545358.

Sophia Vincoff, Shrey Goel, Kseniia Kholina, Rishab Pulugurta, Pranay Vure, and Pranam Chatterjee. Fuson-plm: A fusion oncoprotein-specific language model via focused probabilistic masking. *bioRxiv*, 2024. URL https://api.semanticscholar.org/CorpusID:270286369.

Sophia Vincoff, Shrey Goel, Kseniia Kholina, Rishab Pulugurta, Pranay Vure, and Pranam Chatterjee. Fuson-plm: a fusion oncoprotein-specific language model via adjusted rate masking. *Nature Communications*, 16(1), February 2025. ISSN 2041-1723. doi: 10.1038/s41467-025-56745-6. URL http://dx.doi.org/10.1038/s41467-025-56745-6.

Lin Zhao, Jia Zhao, Kunhong Zhong, Aiping Tong, and Da Jia. Targeted protein degradation: mechanisms, strategies and application. *Signal Transduction and Targeted Therapy*, 7, 2022. URL https://api.semanticscholar.org/CorpusID:247924534.

A SUPPLEMENTARY TABLES

Binder Name	Target	Binder Sequence	Pseudo Perplexity
P3F1_pMLM_1	PAX3::FOXO1	ALTSSSSSSYSYSLDLPK	8.3283
P3F1_pMLM_2	PAX3::FOXO1	NLVDSTSVSSSSKDDLSN	9.5369
P3F1_pMLM_3	PAX3::FOXO1	ALSSLTSDLSSYKDVLKN	10.661
P3F1_pMLM_4	PAX3::FOXO1	AVTQLDSVSYNSKDDCPK	12.835
P3F1_pMLM_5	PAX3::FOXO1	AIVSLSSDSSNYSLVISS	13.7282
P3F1_pMLM_6	PAX3::FOXO1	NVTSSSDSDGTSYDDISK	13.7607
P3F1_pMLM_7	PAX3::FOXO1	NVVDSTSVSSNTYDVCSN	14.292
P3F1_pMLM_8	PAX3::FOXO1	NVTQLTSDLYNYYLDIKK	16.589
P3F1_pMLM_9	PAX3::FOXO1	NVVDLSVDLYNSKSLCPK	17.7314
P3F1_pMLM_10	PAX3::FOXO1	NLVQLDSSLYTYSSLCPS	18.8698
P3F1_pMLM_11	PAX3::FOXO1	NITQDSSVDYTYSLVLPN	11.996
GFAP_pMLM_1	R79H GFAP	KPLLLPPGLPKK	7.01859
GFAP_pMLM_2	R79H GFAP	SPLPDPYLIPLL	12.4299
GFAP_pMLM_3	R79H GFAP	SDLLLPGGLGKK	6.03670
GFAP_moPPit_1	R79H GFAP	GAVFGL	9.54417
GFAP_moPPit_3	R79H GFAP	KRLIMGMK	11.32052
GFAP_moPPit_4	R79H GFAP	ERRLLISALELLL	11.52996
GFAP_muPPit_1	R79H GFAP	LPTSKKRKRVRN	28.59711
GFAP_muPPit_2	R79H GFAP	ATKKKKRRSKPP	25.86114
GFAP_muPPit_3	R79H GFAP	RFTYPRGKGHPD	59.15612

Table 1: Peptide binders generated using PepMLM, moPPit, and muPPit, and their pseudo-perplexity scores.