

Supporting Information

FoldExplorer: Fast and Accurate Protein Structure Search with Sequence-Enhanced Graph Embedding

Yuan Liu, Ying Zhang, Zhen Zhou, Hong-Bin Shen *

Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, and Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai, 200240, China

* Address correspondence to H.B. Shen at hbshen@sjtu.edu.cn

Tel: +86-21-34205320

Fax: +86-21-34204022

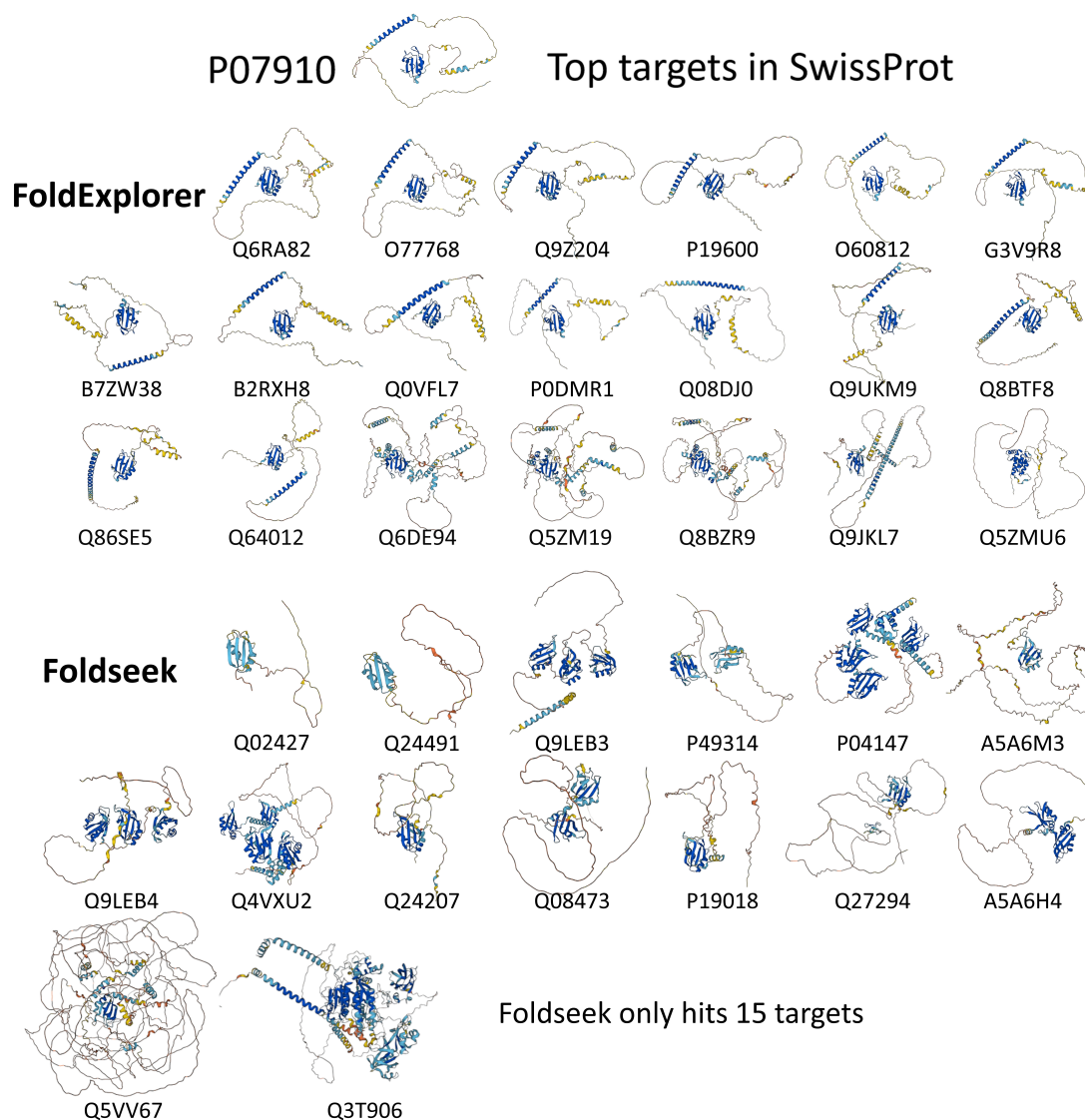
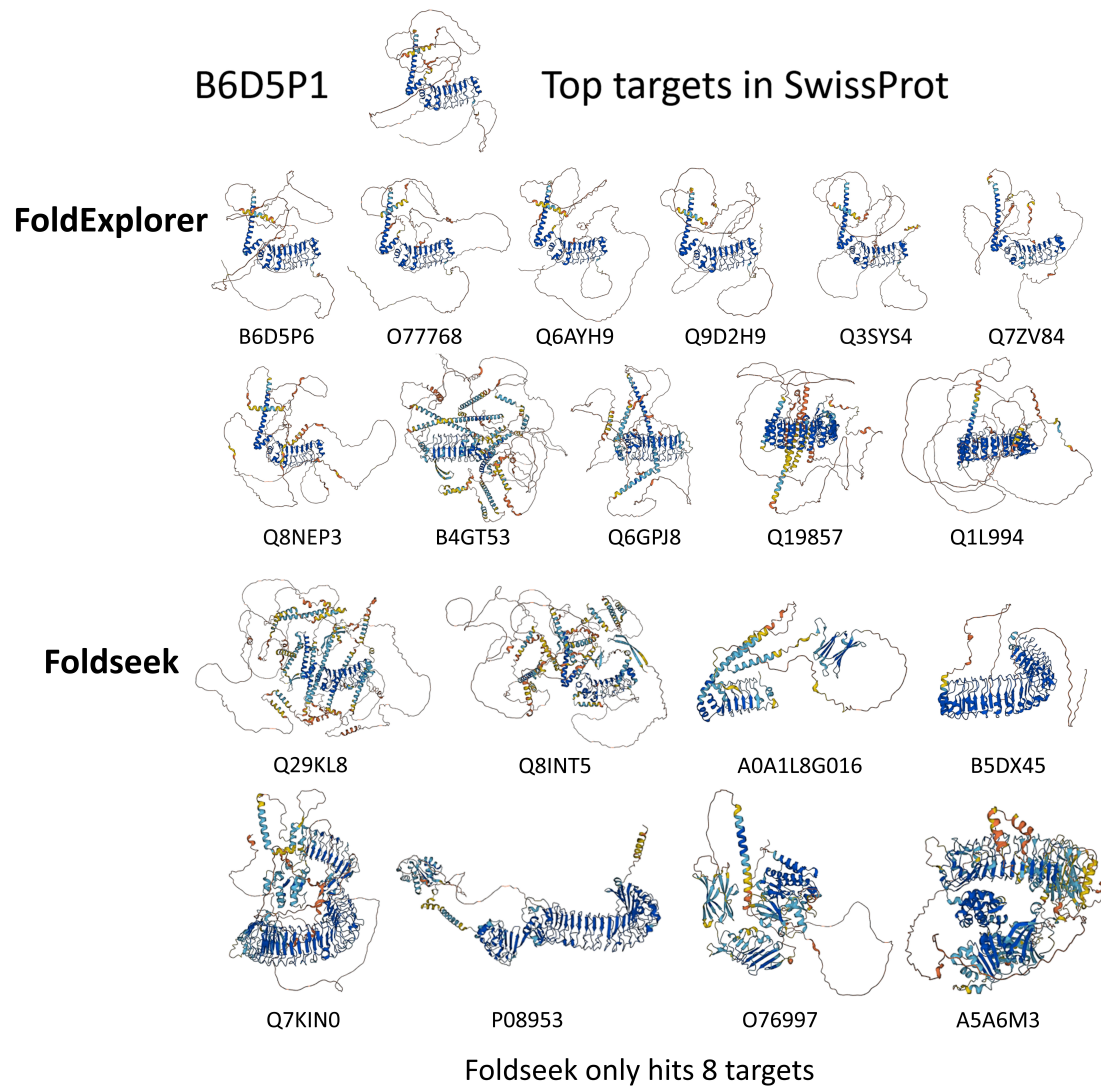


Figure S1. Search results of UniProt entry: P07910 in SwissProt (AlphaFold predicted structures) database using FoldExplorer website and Foldseek website. FoldExplorer incorporates sequence features, so it is less susceptible to interference from disturbed regions and more robust. The searched targets by FoldExplorer have a high overlap with the query in high-confidence regions. However, Foldseek is disturbed by low pLDDT regions, so the targets hit by Foldseek are only locally similar to the query. More information can be seen on our website <http://www.csbio.sjtu.edu.cn/bioinf/FoldExplorer/> and Foldseek website <https://search.foldseek.com/>.



28

29 **Figure S2. Search results of UniProt entry: P6D5P1 in SwissProt (AlphaFold**
 30 **predicted structures) database.**

31

32

33 Table S1. AUPRC of FoldExplorer and other methods on Fold, superfamily, and family
34 classification tasks.

Method	level		
	fold	superfamily	family
SGM	0.3335	0.3805	0.4894
SSEF	0.1279	0.0130	0.0169
DeepFold	0.3865	0.4386	0.5160
GraSR	<u>0.6079</u>	0.6616	0.6070
Foldseek	0.5591	0.6310	0.6143
FoldExplorer	0.5924	<u>0.6741</u>	<u>0.6212</u>
TM-align	0.6438	0.6781	0.6223