Deciphering Features of Metalloprotease Cleavage Targets Using Protein Structure Prediction

Dae sun Chung[1,2,*], Jongkeun Park[1,*], Won Jong Choi[2], Do Young Seong[2] and Dongwan Hong[1,2,3,4,5,6,7,8,#]

Affiliation:

1. Department of Medical Sciences, Graduate School of the Catholic University of Korea, Seoul 06591, Republic of Korea
2. Department of Medical Informatics, The Catholic University of Korea, College of Medicine, Seoul 06591, Republic of Korea
3. Department of Precision Medicine and Big data, The Catholic University of Korea, College of Medicine, Seoul 06591, Republic of Korea
4. Department of Medical Informatics, The Catholic University of Korea, College of Medicine, Seoul 06591, Republic of Korea
5. Precision Medicine Research Center, The Catholic University of Korea, College of Medicine, Seoul 06591, Republic of Korea
6. Cancer Evolution Research Center, The Catholic University of Korea, College of Medicine, Seoul 06591, Republic of Korea
7. CMC Institute for Basic Medical Science, The Catholic University of Korea, College of Medicine, Seoul 06591, Republic of Korea
8. INNOONE, Seoul 06654, Republic of Korea

Abstract

Metalloproteases are a class of enzymes that utilize metal ions within their active sites to catalyze the hydrolysis of peptide bonds in proteins. Among these, ADAM10 (A Disintegrin and Metalloproteinase 10), a member of the ADAM family, plays a crucial role in mediating intracellular signaling by cleaving specific substrates, thereby influencing a variety of physiological and pathological processes. The mechanisms underlying the activity of ADAM10 present significant opportunities for the development of novel therapeutic strategies aimed at disease intervention. However, the information available to identify the substrate and cleavage sites of ADAM10 is still insufficient. Therefore, it is essential to identify and classify the features of substrates and to elucidate cleavage sites through experimental approaches. However, these studies across numerous proteins present significant challenges.

To address the promise of these investigations, we developed a model that leverages protein structure prediction to decipher substrate features, classify substrates, and predict cleavage sites. Through the analysis of predicted protein complexes between ADAM10 and its substrates using PDB files, we evaluated protein-protein interaction (PPI) data, the structural information of cleavage sites, and the spatial relationships between the cleavage sites and metal ions.

Finally, we present a computational model that effectively classifies substrates and accurately predicts cleavage sites in this study. Our study demonstrates the potential for application not only to ADAM10 but also to other members of the ADAM family and, more broadly, to additional metalloproteases. By leveraging computed protein structural information, our approach offers a novel perspective for substrate classification.

Keyword: ADAM10, Metalloprotease, Cleavage, Protein Structure Prediction, Protein-

Protein-Interaction (PPI)

Introduction

A metalloprotease is a type of protease enzyme whose catalytic mechanism depends on the involvement of a metal ion. Among these, ADAM (A Disintegrin And Metalloprotease) is known to play a crucial role in regulating immune functions, particularly during inflammatory processes, as reported in several studies **[1-2]**. Furthermore, ADAMs have been implicated in a wide range of diseases, including cancer metastasis, inflammatory diseases, and neurological disorders **[3-6]**. The ADAM family is classified based on their proteolytic activity into proteolytic ADAMs, non-proteolytic ADAMs, and pseudogenes. Proteolytic ADAMs include ADAM8, 9, 10, 12, 15, 17, 19, 20, 21, 28, 30, and 33, while non-proteolytic ADAMs include ADAM2, 7, 11, 18, 22, 23, 29, and 32 **[7]**.

Proteinases from the ADAM family are known to regulate the functions of various cell surface proteins. Among them, ADAM10 is reported to cleave substrates such as Notch, amyloid precursor protein (APP), N-cadherin, E-cadherin, and TACSTD2, which play critical roles in cancer cell proliferation, migration, and invasion **[8-14]**. These characteristics suggest that ADAM10 can serve as a target antigen for cancer therapy through two distinct approaches. The first approach involves using ADAM10 itself as an antigen, which can be utilized to inhibit cancer cell growth and enhance the efficacy of existing treatments in combination therapies **[15]**. The second approach focuses on using substrates of ADAM10 as antigens, particularly by targeting the cleavage sites generated during proteolytic processes associated with cancer cell growth, thereby suppressing tumor progression **[9-14]**.

These antigens have significant potential to be utilized in antibody-drug conjugates (ADCs), which have recently gained attention. ADCs are anticancer agents designed to deliver cytotoxic payloads specifically to tumor sites by linking the payload to antibodies targeting

tumor-associated antigens, showing reduced systemic toxicity compared to conventional chemotherapies **[16]**. However, antigen selection remains a major challenge in ADC development **[17]**. Due to the inherent properties of ADCs that employ highly potent cytotoxic agents, inappropriate antigen selection can result in unintended off-target effects and increased vulnerability to drug resistance, leading to the discontinuation of many ADC development efforts **[17-18]**. For these reasons, antigen selection requires a multifaceted approach distinct from traditional strategies **[19]**.

In this study, we focused on a multifaceted approach that uses the cleavage sites of ADAM10 substrates, such as TACSTD2, which encompass critical processes for cancer cell growth, as antigens. Through this approach, we aim to provide antigens that can be utilized for more effective and precise therapeutic options. While ADAM10, a proteolytic enzyme, holds great potential as a novel cancer therapeutic target, the number of candidate substrates cleaved by ADAM10 remains limited **[8]**. Several studies have employed mass spectrometry to identify cleavage motifs in proteins that are substrates of ADAM10. However, precise predictions of cleaved proteins have not been achieved **[20]**. Additionally, it is highly challenging to separate and test millions of proteins under experimental conditions to verify specific cleavage events due to the associated costs and time constraints.

In this study, we aim to leverage the medical applications that have emerged after recent advancements in artificial intelligence-based protein structure prediction technologies such as AlphaFold and RosettaFold **[21-22]**. These software tools have been yielding increasingly promising outcomes **[23]**. Based on its primary application to protein binding, we saw potential for this technology to be utilized in analyzing cleavage processes involving the binding between two proteins, a protease and its substrate. To address these challenges, we developed a prediction algorithm based on the characteristics of proteins cleaved by

ADAM10 using these protein structure predictions. The algorithm will not only provide insights into various potential substrate candidates but also offer information about positions where antibodies can recognize substrates without interference.

Our objective is to predict protein structures by leveraging the structural characteristics of ADAM10, classify candidate substrates based on cleavage-related features, and develop an algorithm to support this process.

Method

Data collection

To obtain structural and sequence information for ADAM10, we utilized data from UniProt (https://www.uniprot.org/) using the ADAM10 entry (O14672). Additionally, we identified the sequence and positional information of ADAM10's domains, which include the Signal peptide (S, 1-19), Pro-domain (Pro, 20-213), Metalloprotease domain (MP, 214-456), Disintegrin domain (Dis, 457-551), Cysteine-rich domain (Cys, 552-672), Transmembrane domain (M, 673-693), and Cytoplasmic domain (Cyto, 694-748) **(Figure 1A) [24]**. Similarly, we identified the sequence and positional information of the Nitrogen epsilon 2 (NE2) atoms of HIS383, HIS 387, and HIS393, which are amino acid residues involved in binding the metal ion associated with cleavage **[25]**.

We obtained the sequence and cleavage site information of proteins cleaved by ADAM10 not only for TACSTD2 but also using data provided by Biochem J et al., 2009 **[26]**. These substrates include Amyloid-beta precursor protein (APP), Probetacellulin (BTC), Cadherin-1 (CDH1), Cadherin-2 (CDH2), Delta-like protein 1 (DLL1), Pro-epidermal growth factor (EGF), Tumor necrosis factor ligand superfamily member 6 (FASLG), Low affinity immunoglobulin epsilon Fc receptor (FCER2), Proheparin-binding EGF-like growth factor (HBEGF), Interleukin-6 receptor subunit alpha (IL6R), Neurogenic locus notch homolog protein 1 (NOTCH1), and Melanocyte protein PMEL (PMEL). Additionally, 38 substrates were included through literature review. The exact cleavage sites of these added substrates are not known **(Table 1)**. The substrates were also identified for their sequences and positional information through UniProt, similar to ADAM10.

Protein structure prediction

To discover the interactions between ADAM10 and its substrates after cleavage, we examined both the precursor form, where the Pro-domain sequence remains intact, and the mature form, in which the Pro-domain is removed **(Figure 1A)**. The sequence of the Pro-domain is as follows:

“N-
QYGNPLNKYIRHYEGLSYNVDSLHQKHQRAKRAVSHEDQFLRLDFHAHGRHFNLR
MKRDTSLFSDEFKVETSNKVLDYDTSHIYTGHIYGEEGSFSHGSVIDGRFEGFIQTRG
GTFYVEPAERYIKDRTLPFHSVIYHEDDINYPHKYGPQGGCADHSVFERMRKYQMTG
VEEVTQIPQEEHAANGPELLRKKR-C”.

To analyze the structure of proteins with and without the Pro-domain, we utilized a system running Ubuntu 22.04.4 LTS with Linux kernel version 6.5.0-44-generic and used AlphaFold2 (AF2) release 2.3.2 for the structural predictions. During the prediction of multimeric structures between ADAM10 and candidate substrates, we set the "model_preset=multimer" flag. Next, we employed random seeds to improve prediction quality. High-confidence regions are predicted stably regardless of the random seed, but low-confidence regions are not. Therefore, by using random seeds to set different initial conditions, we enabled the exploration of various structural possibilities. This allowed us to obtain Protein Data Bank (PDB) files consisting of five ranked models, and to obtain one PDB file per rank, we set the "num_multimer_predictions_per_model=1" flag. Finally, to enhance the stability and quality of the predicted structures, we applied the AMBER (Assisted Model Building with Energy Refinement) relaxation process to all PDB models by setting the "models_to_relax=all" flag **(Supplementary Figure 1)**.

For verification of the protein structures, we used Chimera X (version 1.6.1) to explore the

spatial structures obtained from the PDB files generated by AF2.

Analysis of multimeric protein (ADAM10, Substrate) PDB files

The predicted PDB files, based on the identified sequence information, contain essential data such as chains distinguishing the two proteins, atom names, and atomic coordinates. These files were analyzed to extract additional insights using various computational algorithms. First, hydrogen bond interactions between ADAM10 and its substrate, as well as intra-substrate hydrogen bonds, were identified using Biopython. Subsequently, the secondary structures of the proteins (α-helix, β-sheet) and Solvent-Accessible Surface Area (SASA) were analyzed using the DSSP algorithm **[27]**. Solvent-Accessible Surface Area (SASA) serves as a robust tool for quantifying and describing protein-protein interactions (PPI). A SASA value approaching zero indicates that the corresponding protein region is involved in interactions **[28]**.

To evaluate the predicted structures, the identifiers of each substrate provided by UniProt were used to retrieve experimentally validated structures from the Protein Data Bank (www.rcsb.org). The Root Mean Square Deviation (RMSD) was calculated to compare the experimental and predicted structures. RMSD is a widely used metric for assessing the similarity between two protein structures, where lower values indicate higher similarity. An RMSD value below 2 Å suggests a high degree of structural similarity, making it an appropriate measure for assessing the accuracy of predicted protein structures **[29-30]**.

All analyses were performed using Python (version 3.11.9) with Biopython (version 1.83) and mkDSSP (version 4.0.4).

Results

Identification of ADAM10 composition, substrate structure, and common binding regions of ADAM10

We investigated the domain structure of ADAM10, and the structural changes induced by cleavage. In its protease-inactive form, ADAM10 is comprised of seven domains: S, Pro, MP, Dis, Cys, M, and Cyto. Upon cleavage of the Pro-domain, the protease-active form of ADAM10 consists of MP, Dis, Cys, M, and Cyto **(Figure 1A)**. TACSTD2, one of the substrates of ADAM10, is composed of an extracellular region (Extra), a transmembrane region (M^{Sub}), and a cytoplasmic region ($Cyto^{Sub}$), which is the case for other substrates as well **(Figure 1B)**.

Based on this, we used AlphaFold-Multimer to predict and observe the complex structures of ADAM10 and its 13 substrates **[12, 26]** to compare the interactions between them depending on protease activity. In the presence of the Pro-domain, interactions between the GLY329-SER340 region of ADAM10 and TACSTD2 were not observed due to steric hindrance caused by the Pro-domain. However, upon removal of the Pro-domain, hydrogen bond interactions were observed between ALA331, VAL333, SER339, and SER340 of ADAM10 and ALA2, GLY4, and PRO5 of TACSTD2 **(Figure 1C and 1D)**.

Upon examining the remaining substrates, consistent interactions were observed in the GLY329–SER340 region of ADAM10 for the 11 substrates, excluding CDH2, in the absence of the Pro-domain **(Supplementary Table 1)**.

To further confirm these findings quantitatively, we measured the SASA. The results indicated that, in the presence of the Pro-domain, the SASA values for the Pro-domain region (designated as region ①) were VAL178 (11 Å²), MET182 (14 Å²), and GLN186 (8 Å²). For

the GLY329-SER340 region of ADAM10 (designated as region ②), the SASA values were ALA331 (0 Å²), VAL333 (34 Å²), SER339 (22 Å²), and SER340 (10 Å²) **(Figure 1E)**. In contrast, in the absence of the Pro-domain, the SASA values for the GLY329-SER340 region (designated as region ②') were ALA331 (0 Å²), VAL333 (12 Å²), SER339 (31 Å²), and SER340 (4 Å²), while the SASA values for TACSTD2 (designated as region ③) were ALA2 (1 Å²), GLY4 (4 Å²), and PRO5 (25 Å²) **(Figure 1F)**.

Similarly, SASA measurements for the remaining 11 substrates, excluding CDH2, showed comparable patterns in their interactions with ADAM10 **(Supplementary Tables 2-5)**. These results quantitatively demonstrate that the substrates share the characteristic of binding to the GLY329-SER340 region of protease-active ADAM10.

Next, we assessed the feasibility of actual binding between the GLY329-SER340 region and various substrates. This region resides within the 1-672 region **(Figure 1A)**, which is external to the membrane, indicating that for substrate binding to occur, the binding site must also be located in the extracellular region.

In the case of TACSTD2, the extracellular region spans residues 1-274 region **(Figure 1B)**, with binding sites at ALA2, GLY4, and PRO5, all located within the extracellular region. Thus, we confirmed that binding between the GLY329-SER340 region and TACSTD2 is feasible. Similarly, among the 11 substrates that bind to the GLY329-SER340 region, 9 substrates, excluding NOTCH1 and DLL1, were confirmed to bind within the extracellular region, supporting the feasibility of these interactions **(Supplementary Table 6 and 7)**.

Structural Features of Cleavage Sites in ADAM10 Substrates

We examined the structure of Cleavage Sites to predict and characterize features of the substrates. The ARG87-THR88 site of TACSTD2 is cleaved by ADAM10 **(Figure 2A) [12]**. Based on this, we conducted a structural investigation of TACSTD2. The SER81-VAL90 region containing the cleavage site in TACSTD2 was observed to be in a linear structure **(Figure 2B)**.

Similarly, we verified the Cleavage Sites of the remaining 12 substrates, finding that BTC, EGF, FASLG, and FCER2 each possessed two Cleavage Sites **(Table 2) [26]**. In total, we examined the structures of 17 Cleavage Sites identified across all 13 substrates, including TACSTD2 **(Figure 2C)**. 12 cleavage sites, including TACSTD2 (ARG87-THR88), FASLG (SER126-LEU127), and FASLG (LYS129-GLN130), were observed to have a linear structure **(Figure 2B, D)**. 3 cleavage sites, including FCER2 (ARG80-GLN81) and FCER2 (ARG101-LEU102), were observed in an α-helix structure **(Figure 2E)**. Finally, 2 cleavage sites, including PMEL (GLN583-LEU584), were observed in a β-sheet structure **(Figure 2F) (Supplementary Figure 2)**. We confirmed that over 70% of the Cleavage Sites exhibit a linear structure **(Figure 2C)**.

Next, the predicted structures were evaluated using RMSD **(Supplementary Table 8)**. Among the 13 substrates, CDH2 was excluded from the analysis due to the absence of experimentally validated structural data.

For TACSTD2, the RMSD between the experimentally determined structure, the crystal structure of the extracellular part of human TACSTD2 (PDB ID: 7PEE), and the predicted structure was found to be 0.774 Å. At a value lower than 2 Å, this indicates a high degree of similarity. Similarly, the remaining 11 substrates also exhibited RMSD values below 2 Å for at least one experimentally determined structure, demonstrating a strong structural alignment

with the predicted models **[29, 31]**.

We examined whether the experimentally determined cleavage site structures matched the cleavage site regions predicted in our study. 7PEE revealed that the actual cleavage site exhibited a linear structure, consistent with our predictions **[31]**.

We also aimed to perform similar validations for other substrates. Notably, only APP and NOTCH1 had experimental data available for their cleavage sites. APP exhibited both the predicted linear structure and a secondary structure at the cleavage site, while NOTCH1 showed a β-sheet structure consistent with our predictions. For the remaining substrates, either no experimental data were available for the cleavage sites, or the data pertained to findings unrelated to the cleavage sites **(Supplementary Table 8)**.

Subsequently, we investigated whether regions with fewer hydrogen bonds among neighboring amino acids were more likely to be Cleavage Sites by observing the number of hydrogen bonds in and around each Cleavage Site. For the Cleavage Site in TACSTD2, ARG87 had one hydrogen bond while THR88 had zero. Furthermore, an examination of hydrogen bonds in the adjacent ALA82-THR96 and LEU89-VAL90 sites showed that ALA82, LEU89, and VAL90 each had zero hydrogen bonds; PRO83, LYS84, and ASN85 each had one; and THR86 had two. For the remaining 12 Cleavage Sites with linear structures (excluding the five Cleavage Sites with secondary structures like α-helix or β-sheet that generally require multiple interactions with surrounding residues), we also counted hydrogen bonds in the surrounding amino acids **(Supplementary Table 9)**. Consequently, we found that Cleavage Sites in a linear structure generally had two or fewer hydrogen bonds in their surrounding amino acids.

Spatial Characteristics between Cleavage Sites and the Zinc ion

The extracellular domain of ADAM10 consists of the MP, Dis, and Cys domains, excluding the PRO, M, and Cyto domains from the full structure. Within the MP domain, key histidine residues (HIS383, HIS387, and HIS393) are involved in binding a zinc ion, which plays a critical role in proteolytic cleavage **(Figure 3A) [32]**.

Our aim was to investigate the spatial characteristics of the cleavage sites in substrates, focusing on the zinc ion associated with ADAM10's proteolytic activity. We first needed to predict the location of the zinc ion within the PDB file of ADAM10. We utilized the PDB structure of ADAM10 Extracellular Domain (PDB ID: 6BE6) from the Protein Data Bank Site **(Figure 3B)**. The 6BE6 structure consists of four ADAM10 extracellular domains, each with a zinc ion. We measured the distances between the zinc ion and the NE2 atoms of the three histidines involved in its coordination: HIS383, HIS387, and HIS393. The distances between the zinc ion and HIS383 NE2 were 2.134 Å, 2.189 Å, 1.998 Å, and 2.106 Å, while for HIS387 NE2, the distances were 2.058 Å, 2.082 Å, 2.082 Å, and 2.062 Å. The distances for HIS393 NE2 were 2.058 Å, 2.100 Å, 2.341 Å, and 2.290 Å.

Based on these measurements, we were able to predict the location of the zinc ion in the previously predicted PDB structures of substrates **(Figure 3B)**. To explore the spatial characteristics of the zinc ion relative to the cleavage sites, we visualized the coordinates of the zinc ion along with those of 85 cleavage sites (derived from 17 cleavage sites across five ranked models). We aligned the coordinates such that the zinc ion was positioned at the origin (0,0,0) **(Figure 3C)**.

The cleavage sites of TACSTD2 were located at (0.998, 2.692, 35.137), (23.290, 5.868, 21.558), (-1.262, 8.469, 19.099), (6.444, 10.229, 27.228), and (-5.523, 10.754, 32.696). The

cleavage sites of other substrates were also positioned at similar coordinates **(Supplementary Table 10)**.

The results show that, with the exception of 1 site in APP, 3 in CDH1, 4 in FCER2, 1 in IL6R, and 2 in PMEL, a total of 74 cleavage sites fell within the coordinate range of (-87.3 to 101.912, 1.853 to 135.016, -66.860 to 80.174). This indicates that over 87% of the cleavage sites are located in Octants 1, 4, 5, and 8 relative to the zinc ion at the origin **(Supplementary Figure 3)**.

Classification of Substrates and Algorithm Development Based on Identified Features

Using the aforementioned characteristics, we classified the substrates into 13 substrates with known cleavage sites (in black) and 38 substrates with unknown cleavage sites (in red) **(Figure 4A)**.

First, we used AF2 to predict the complex structures of ADAM10 and its substrates, obtaining PDB files. Next, we extracted hydrogen bonds between the two proteins from the PDB files to determine whether substrates were bound to the common binding region of ADAM10. If no binding was observed, the substrates were classified into Groups 3, 4, and 6, and if binding was observed, they were classified into Groups 1, 2, 4, 5, 7, 8, and 9. For Groups with binding, we further examined whether the substrates were bound to the feasible region, the extracellular region (Extra).

If binding occurred in the extracellular region, the substrates were classified into Groups 1, 5, and 8; if not, they were classified into Groups 2, 7, and 9. Next, we examined whether the substrates contained predicted cleavage sites with the structural features of known cleavage sites, based on hydrogen bonding and structural information within the substrates. If no predicted cleavage site was identified, the substrates were finalized into Groups 4, 7, and 8. To determine the spatial orientation of the predicted cleavage sites relative to the zinc ion serving as the origin, we examined whether the cleavage sites of Group 1, 2, 3, 5, 6, and 9 are positioned within Octants 1, 4, 5, and 8. If this condition was met, the substrates were finalized into Groups 1, 2, and 3; otherwise, they were classified into Groups 5, 6, and 9 **(Figure 4B)**.

Among the 51 substrates, hydrogen bond formation with the common binding region of ADAM10 was observed in 45 substrates, excluding 6 **(Supplementary Table 11)**. Furthermore, of these 45 substrates, 30 were confirmed to bind within the feasible

extracellular region (Extra) **(Supplementary Table 12 and 13)**. Of the 51 substrates, 49 were found to contain predicted cleavage sites, excluding 2 **(Supplementary Table 14)**. Additionally, among the 49 substrates with predicted cleavage sites, 46 had their cleavage sites located within Octants 1, 4, 5, and 8, relative to the zinc ion serving as the origin **(Supplementary Table 15)**.

Consequently, 28 out of the 51 substrates were classified into Group 1, which satisfied all identified features. Although the binding region of these substrates in Group 2 was predicted to be outside the feasible binding region in ADAM10's common binding site, 15 substrates were classified into Group 2 as they met all other features. The remaining substrates were also classified into respective Groups based on their satisfaction of the various features **(Figure 4A) (Supplementary Table 16)**.

Discussion

In our study, we predicted potential antigen cleavage candidates by ADAM10, a member of the ADAM family, which belongs to the metalloprotease group. While the role of ADAM10 in tumors has been well-known with its cleavage effects reported, information regarding antigens cleaved by ADAM10 has been limited **[33]**. Comprehensive cleavage information for membrane proteins localized on the membrane remains unavailable, and experimental limitations make it challenging to validate. Therefore, we proposed a novel substrate classification approach in this study.

This study utilized 13 known membrane proteins as a foundation to perform protein structure prediction and identify potential cleavage sites based on structural features. As a result, about 70% of the proteins were classified into the G1 Group. We investigated CDH1, CDH2, DLL1, and NOTCH1, which were classified into Groups other than G1, and confirmed that all four proteins were associated with adhesion. Building upon this, we examined all 51 substrates, including the previously mentioned 13 substrates, and identified 18 out of the 51 substrates as adhesion-related proteins. Among these, 11 substrates were not classified into G1, indicating that nearly half (47.8%) of the 23 substrates not classified into G1 are associated with adhesion **(Supplementary Table 17)**. We confirmed that 70% of the substrates were classified into the G1 Group, excluding adhesion-related substrates that were not classified into G1.

Membrane-associated antigens in tumors play critical roles. For instance, ERBB2 (HER2) is expressed not only in breast cancer but also in gastric cancer and serves as an important biomarker **[34]**. Other examples include CD79b **[35]**. Recent studies have focused on developing ADCs to target the aforementioned antigens. For instance, Trastuzumab deruxtecan, an ADC developed for ERBB2 (HER2), and Polatuzumab vedotin-piiq, an ADC

targeting CD79b, have been approved by the FDA. TACSTD2 also has an FDA-approved ADC and is actively being investigated in clinical trials **[36]**. TACSTD2, a protein cleaved by ADAM10, has been identified in its cleaved form in tumors **[12]**. ADCs targeting these cleavage sites are currently in clinical trials, showing promising results **[37]**.

To utilize such approaches, a precise understanding of antigen-antibody recognition (epitopes) is essential. Therefore, we initiated this study based on the hypothesis that many membrane proteins cleaved by metalloproteases like ADAM10 in tumors could serve as potential antigen candidates.

To identify antigens cleaved by metalloproteases, we employed the following methods. First, we utilized protein structure prediction to extract protein-protein interactions and secondary structural features (linear, α-helix, β-sheet) **[20]**. Second, we analyzed the three-dimensional spatial distribution of cleavage sites based on physical distances in 3D space **[38]**.

In other studies, the existence of cleavage events has been demonstrated solely through experimental methods **[12, 26]**. Although potential motifs for cleavage are provided, no consensus motifs have been identified. Therefore, our approach offers a novel perspective on this topic. Leveraging 3D protein structure prediction applications using AlphaFold software, which has advanced to AlphaFold 3, we predict protein-protein interactions with increasing accuracy. Furthermore, mathematical and engineering advancements have enabled precise calculations of positions and distances within protein structures **[39]**. Thus, there is no issue regarding the feasibility of our method.

Nevertheless, a few limitations remain. First, limitations arising from data insufficiency were identified. The limited number of predicted protease-substrate complex structures and known substrates may lead to potential biases in the results. The exact cleavage sites were not

explicitly defined in the experimental data, preventing the identification of consensus sequences composed of amino acids at the cleavage sites **(Supplementary Figure 4)**. Consequently, cleavage site predictions were provided as regions rather than specific positions. The RMSD analysis of the predicted structures showed that similarities with NMR and EM structures were relatively lower compared to X-ray structures. This outcome highlights the limitation stemming that 92% of experimentally validated structural data in the PDB are derived from X-ray crystallography, resulting in a lack of sufficient NMR and EM data for AlphaFold 2 training **[40]**.

Second, limitations arising from biological complexity were noted. We examined the structural and functional differences between the Precursor Form and the Mature Form of ADAM10 based on the presence or absence of the Pro-domain sequence. However, unlike cell-level microscopy, this approach does not account for the actual cellular environment, which may lead to insufficient reflection of biological complexity under certain conditions.

Nevertheless, providing targets for antigens cleaved by Metalloproteases in tumors using AI-based approaches will play a crucial role in drug development.

References

1. Khokha, R., A. Murthy, and A. Weiss, *Metalloproteinases and their natural inhibitors in inflammation and immunity.* Nature Reviews Immunology, 2013. **13**(9): p. 649-665.
2. Lambrecht, B.N., M. Vanderkerken, and H. Hammad, *The emerging role of ADAM metalloproteinases in immunity.* Nature Reviews Immunology, 2018. **18**(12): p. 745-758.
3. Duffy, M.J., et al., *Role of ADAMs in cancer formation and progression.* Clinical Cancer Research, 2009. **15**(4): p. 1140-1144.
4. Sikora-Skrabaka, M., et al., *ADAM10 and ADAM17 as biomarkers linked to inflammation, metabolic disorders and colorectal cancer.* Current Issues in Molecular Biology, 2022. **44**(10): p. 4517-4527.
5. Zhou, M., et al., *Association of ADAM10 gene variants with sporadic Parkinson's disease in Chinese Han population.* The Journal of Gene Medicine, 2021. **23**(3): p. e3319.
6. Epis, R., et al., *Blocking ADAM10 synaptic trafficking generates a model of sporadic Alzheimer's disease.* Brain, 2010. **133**(11): p. 3323-3335.
7. Edwards, D.R., M.M. Handsley, and C.J. Pennington, *The ADAM metalloproteinases.* Molecular aspects of medicine, 2008. **29**(5): p. 258-289.
8. Reiss, K., et al., *Scramblases as regulators of proteolytic ADAM function.* Membranes, 2022. **12**(2): p. 185.
9. Sakach, E., R. Sacks, and K. Kalinsky, *Trop-2 as a therapeutic target in breast cancer.* Cancers, 2022. **14**(23): p. 5936.
10. Tomiyama, E., et al., *Trop-2 in upper tract urothelial carcinoma.* Current Oncology, 2022. **29**(6): p. 3911-3921.
11. Belluomini, L., et al., *Antibody-drug conjugates (ADCs) targeting TROP-2 in lung cancer.* Expert Opinion on Biological Therapy, 2023. **23**(11): p. 1077-1087.
12. Trerotola, M., et al., *Trop□2 cleavage by ADAM10 is an activator switch for cancer growth and metastasis.* Neoplasia, 2021. **23**(4): p. 415-428.
13. Smith Jr, T.M., A. Tharakan, and R.K. Martin, *Targeting ADAM10 in cancer and autoimmunity.* Frontiers in immunology, 2020. **11**: p. 499.
14. Woods, N., et al., *Fendiline inhibits proliferation and invasion of pancreatic cancer cells by interfering with ADAM10 activation and β-catenin signaling.* Oncotarget, 2015. **6**(34): p. 35931.
15. Pece, R., et al., *Inhibitors of ADAM10 reduce Hodgkin lymphoma cell growth in 3D microenvironments and enhance brentuximab-vedotin effect.* Haematologica, 2021. **107**(4): p. 909.
16. Chau, C.H., P.S. Steeg, and W.D. Figg, *Antibody–drug conjugates for cancer.* The Lancet, 2019. **394**(10200): p. 793-804.
17. Loganzo, F., M. Sung, and H.-P. Gerber, *Mechanisms of resistance to antibody–drug conjugates.* Molecular cancer therapeutics, 2016. **15**(12): p. 2825-2834.
18. Tsuchikama, K., et al., *Exploring the next generation of antibody–drug conjugates.* Nature Reviews Clinical Oncology, 2024. **21**(3): p. 203-223.
19. Nelson, B.E. and F. Meric-Bernstam, *Leveraging TROP2 antibody-drug conjugates in solid tumors.* Annual Review of Medicine, 2024. **75**(1): p. 31-48.
20. Tucher, J., et al., *LC–MS based cleavage site profiling of the proteases ADAM10 and ADAM17 using proteome-derived peptide libraries.* Journal of proteome research, 2014. **13**(4): p. 2205-2214.
21. Jumper, J., et al., *Highly accurate protein structure prediction with AlphaFold.* nature,

2021. **596**(7873): p. 583-589.
22. Baek, M., et al., *Accurate prediction of protein structures and interactions using a three-track neural network.* Science, 2021. **373**(6557): p. 871-876.
23. Callaway, E., *Chemistry Nobel goes to developers of AlphaFold AI that predicts protein structures*. Nature, 2024. **634**(8034): p. 525-526.
24. Ebsen, H., et al., *Identification of SH3 domain proteins interacting with the cytoplasmic tail of the a disintegrin and metalloprotease 10 (ADAM10).* PloS one, 2014. **9**(7): p. e102899.
25. Seegar, T.C., et al., *Structural basis for regulated proteolysis by the α-secretase ADAM10.* Cell, 2017. **171**(7): p. 1638-1648. e7.
26. Caescu, C.I., G.R. Jeschke, and B.E. Turk, *Active-site determinants of substrate recognition by the metalloproteinases TACE and ADAM10.* Biochemical Journal, 2009. **424**(1): p. 79-88.
27. Nagy, G. and C. Oostenbrink, *Dihedral-based segment identification and classification of biopolymers I: proteins.* Journal of chemical information and modeling, 2014. **54**(1): p. 266-277.
28. Konstantinidis, K., et al., *On the estimation of the molecular inaccessible volume and the molecular accessible surface of a ligand in protein–ligand systems.* Molecular Systems Design & Engineering, 2021. **6**(11): p. 946-963.
29. Reva, B.A., A.V. Finkelstein, and J. Skolnick, *What is the probability of a chance prediction of a protein structure with an rmsd of 6 å?* Folding and Design, 1998. **3**(2): p. 141-147.
30. Castro-Alvarez, A., A.M. Costa, and J. Vilarrasa, *The performance of several docking programs at reproducing protein–macrolide-like crystal structures.* Molecules, 2017. **22**(1): p. 136.
31. Pavšič, M., *Trop2 forms a stable dimer with significant structural differences within the membrane-distal region as compared to EpCAM.* International journal of molecular sciences, 2021. **22**(19): p. 10640.
32. Gomis-Rüth, F.X., *Structural aspects of the metzincin clan of metalloendopeptidases.* Molecular biotechnology, 2003. **24**: p. 157-202.
33. Mochizuki, S. and Y. Okada, *ADAMs in cancer cell proliferation and progression.* Cancer science, 2007. **98**(5): p. 621-628.
34. Cao, D., et al., *Molecular characteristics of gastric cancer with ERBB2 amplification.* Heliyon, 2023. **9**(8).
35. Schlette, E., et al., *CD79b expression in chronic lymphocytic leukemia: association with trisomy 12 and atypical immunophenotype.* Archives of pathology & laboratory medicine, 2003. **127**(5): p. 561-566.
36. Liu, X., et al., *Trop2-targeted therapies in solid tumors: advances and future directions.* Theranostics, 2024. 14(9): p. 3674.
37. Kim, H., et al., *LCB84, a TROP2-targeted ADC, for treatment of solid tumors that express TROP-2 using the hu2G10 tumor-selective anti-TROP2 monoclonal antibody, a proprietary site-directed conjugation technology and plasma-stable tumor-selective linker chemistry.* Cancer Research, 2022. 82(12_Supplement): p. 328-328.
38. Gheisi, A. and G. Naser, *Multi-aspect performance analysis of water distribution systems under pipe failure.* Procedia Engineering, 2015. 119: p. 158-167.
39. Abramson, J., et al., *Accurate structure prediction of biomolecular interactions with AlphaFold 3.* Nature, 2024: p. 1-3.
40. Laurents, D.V., *AlphaFold 2 and NMR spectroscopy: partners to understand protein structure, dynamics and function.* Frontiers in molecular biosciences, 2022. **9**: p.

906437.

Figure.1 . Identification of ADAM10 composition, substrate structure, and common binding regions of ADAM10

A: Domain composition of ADAM10. S: Signal sequence(Light orange), Pro: Pro-domain(Gray), MP: Metalloprotease domain(Red), Dis: Disintegrin domain(Sky blue), Cys: Cystein-rich domain(Light green), M: Transmembrane region(Navy blue), Cyto: Cytoplasmic region(Purple)

ADAM10(Pro cleavage) represents the protease-active state of ADAM10, where the Pro-domain, which inhibits the protease function, has been removed. The G329-S340 region in the MP domain of ADAM10(Pro cut) is the sequence that commonly binds to substrates exclusively in the Pro cleavage state.

B: Domain composition of the representative ADAM10 substrate (TACSTD2). Extra: Extracellular region, MSub: Transmembrane region of the substrate(Navy blue), CytoSub: Cytoplasmic region of the substrate(Purple)

C: Predicted structure of the ADAM10+Substrate (TACSTD2) complex. The enlarged section shows that, due to the presence of the Pro-domain (), no interaction occurs between the G329-S340 region () of ADAM10 and the substrate (TACSTD2).

D: Predicted structure of the ADAM10(Pro cleavage)+Substrate (TACSTD2) complex. The enlarged section shows that the G329-S340 region (') of ADAM10(Pro cleavage) interacts with the substrate (TACSTD2) (), forming hydrogen bonds (H-bonds).

E: SASA graph of the enlarged region of the ADAM10+Substrate (TACSTD2) complex. The vertical axis represents SASA values, and the horizontal axis represents amino acid residue numbers. The mutually low SASA values of the Pro-domain () and the G329-S340 region () of ADAM10 indicate interactions between these regions, which, through the high SASA values of the substrate (TACSTD2), confirm that interaction with region is prevented.

F: SASA graph of the enlarged region of the ADAM10(Pro cleavage)+Substrate (TACSTD2) complex. The low SASA values of the G329-S340 region (') of ADAM10(Pro cleavage) and the substrate (TACSTD2) () confirm their interaction.

Figure.2 . Structural Features of Cleavage Sites in ADAM10 Substrates

A: Domain composition of the representative ADAM10 substrate (TACSTD2), including its cleavage site.

B: Predicted structure of the cleavage sites R87 and T88 of the substrate (TACSTD2) cleaved by ADAM10. The cleavage site of the substrate (TACSTD2) was predicted to adopt a linear structure.

C: Pie chart of the structural characteristics of substrate cleavage sites. Among the 17 cleavage sites analyzed, 12 were predicted to adopt a linear structure, 3 were α-helix structures, and 2 were β-sheet structures.

D: Predicted structure of the cleavage sites S126 and L127, and K129 and Q130 of the substrate (FASLG). The cleavage sites of the substrate (FASLG) were predicted to adopt a linear structure.

E: Predicted structure of the cleavage sites A80 and Q81, and R101 and L102 of the substrate (FCER2). The cleavage sites of the substrate (FCER2) were predicted to adopt an α-helix structure.

F: Predicted structure of the cleavage site Q583 and L584 of the substrate (PMEL). The cleavage site of the substrate (PMEL) was predicted to adopt a β-sheet structure.

Figure.3 . Spatial Characteristics between Cleavage Sites and the Zinc ion

A: Metalloprotease domain (MP) in the extracellular domain of ADAM10. The MP contains three histidines (H383, H387, H393) that serve as binding sites for the zinc ion, a metal ion directly involved in proteolytic cleavage.

B: The PDB structure of ADAM10 Extracellular Domain (PDB ID: 6BE6), showing the spatial relationship between ADAM10 and the zinc ion. The enlarged section highlights the zinc ion and the Nitrogen epsilon 2 (NE2) atoms of histidines H383, H387, and H393 that coordinate its binding.

C: Coordinate distribution of 85 cleavage sites (17 cleavage sites × 5 ranked) aligned with the zinc ion as the origin (0, 0, 0). Among these, 74 cleavage sites were found to be located within Octants 1, 4, 5, and 8, excluding 1 site from APP, 3 from CDH1, 1 from FCER2, 1 from IL6R, and 1 from PMEL.

Figure.4 . Classification of Substrates and Algorithm Development Based on Identified Features

A: Decision tree showing the classification of ADAM10 substrates based on specific conditions.
The classification was performed on a total of 51 substrates, including 13 substrates with known cleavage sites (black) and 38 substrates with unknown cleavage sites (red).

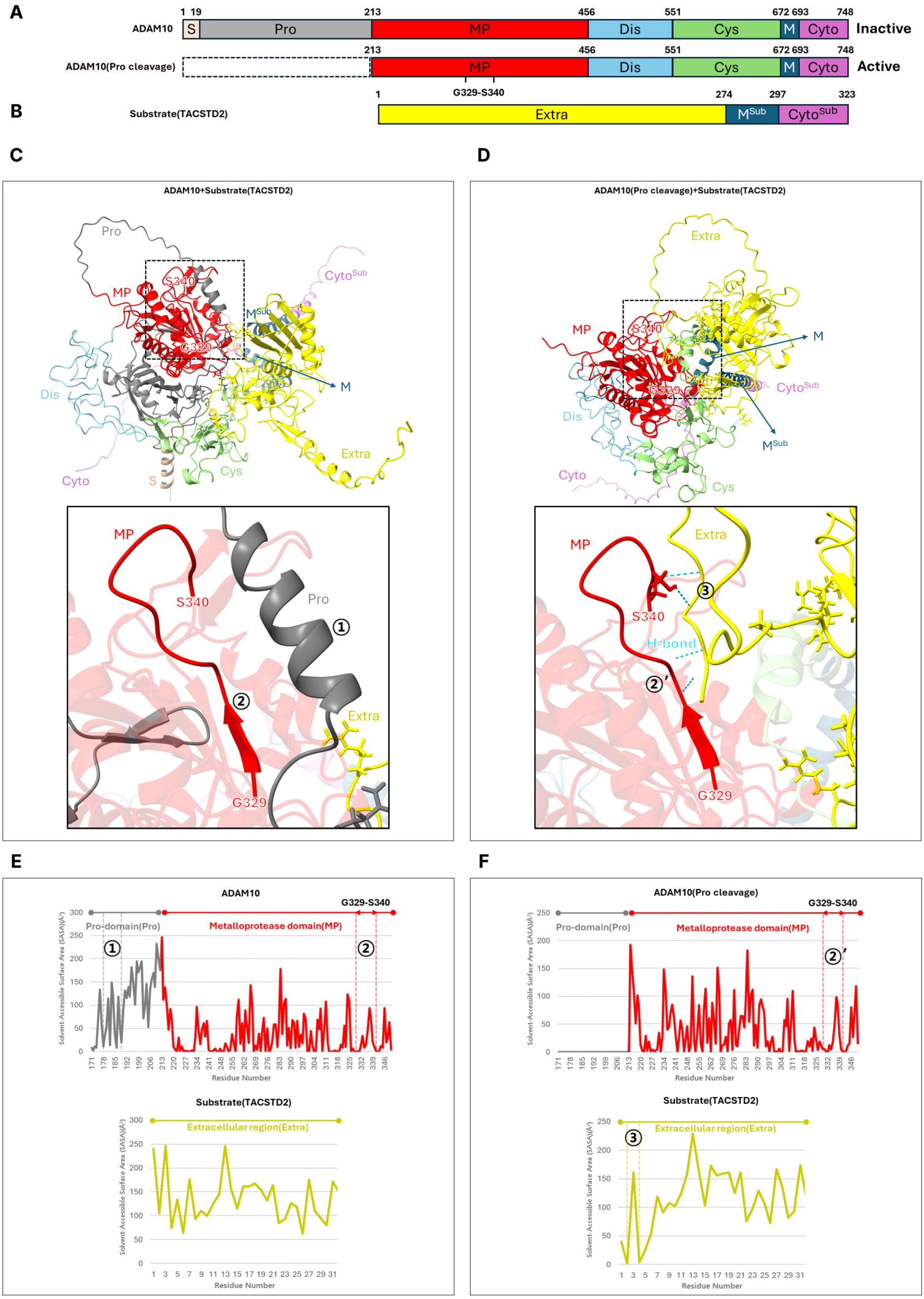
B: Flowchart depicting the classification process of the substrates.

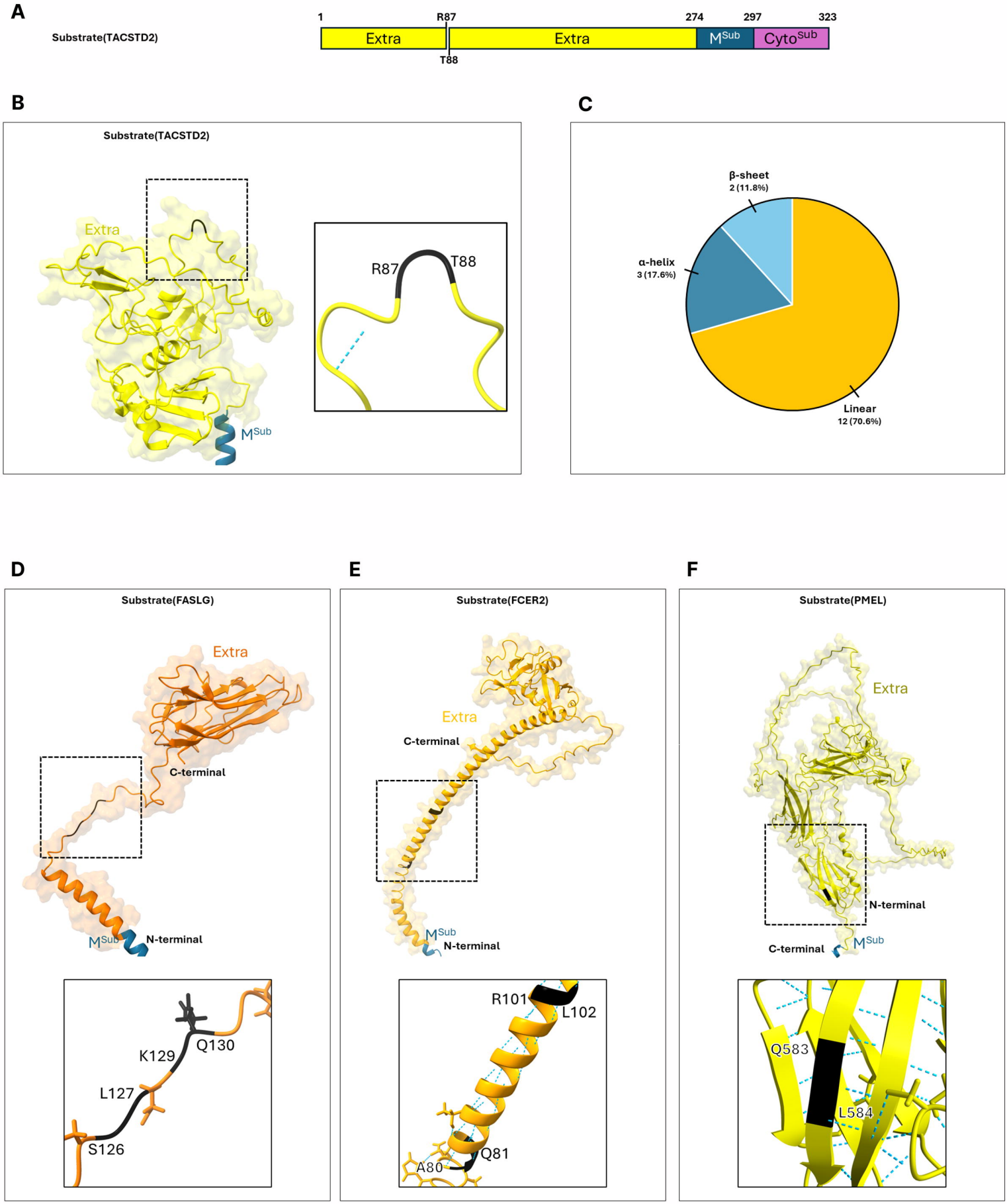
Table.1 . Information on ADAM10 substrates

UniProt ID	Gene name	Amino acid length	Protein name
O14672	ADAM10	748	Disintegrin and metalloproteinase domain-containing protein 10
P05067	APP	770	Amyloid-beta precursor protein
P35070	BTC	178	Probetacellulin
P12830	CDH1	882	Cadherin-1
P19022	CDH2	906	Cadherin-2
O00548	DLL1	723	Delta-like protein 1
P01133	EGF	1207	Pro-epidermal growth factor
P48023	FASLG	281	Tumor necrosis factor ligand superfamily member 6
P06734	FCER2	321	Low affinity immunoglobulin epsilon Fc receptor
Q99075	HBEGF	208	Proheparin-binding EGF-like growth factor
P08887	IL6R	468	Interleukin-6 receptor subunit alpha
P46531	NOTCH1	2555	Neurogenic locus notch homolog protein 1
P40967	PMEL	661	Melanocyte protein PMEL
P09758	TACSTD2	323	Tumor-associated calcium signal transducer 2
P50895	BCAM	628	Basal cell adhesion molecule
Q9BY67	CADM1	442	Cell adhesion molecule 1
Q04900	CD164	197	Sialomucin core protein 24
Q9NZQ7	CD274	290	Programmed cell death 1 ligand 1
P29965	CD40LG	261	CD40 ligand
P16070	CD44	742	CD44 antigen
P15529	CD46	392	Membrane cofactor protein
P04233	CD74	296	HLA class II histocompatibility antigen gamma chain
Q9Y6N8	CDH10	788	Cadherin-10
P33151	CDH5	784	Cadherin-5
P55285	CDH6	790	Cadherin-6
P07333	CSF1R	972	Macrophage colony-stimulating factor 1 receptor
P78423	CX3CL1	397	Fractalkine
Q9H2A7	CXCL16	254	C-X-C motif chemokine 16
P43146	DCC	1447	Netrin receptor DCC
P16422	EPCAM	314	Epithelial cell adhesion molecule
Q15303	ERBB4	1308	Receptor tyrosine-protein kinase erbB-4
P21802	FGFR2	821	Fibroblast growth factor receptor 2
P22455	FGFR4	802	Fibroblast growth factor receptor 4
Q9HCN6	GP6	339	Platelet glycoprotein VI
Q8TDQ0	HAVCR2	301	Hepatitis A virus cellular receptor 2
O75144	ICOSLG	302	ICOS ligand
Q5VWK5	IL23R	629	Interleukin-23 receptor
P78504	JAG1	1218	Protein jagged-1
Q9Y219	JAG2	1238	Protein jagged-2
Q9UEF7	KL	1012	Klotho
P18627	LAG3	525	Lymphocyte activation gene 3 protein
O75581	LRP6	1613	Low-density lipoprotein receptor-related protein 6
Q16820	MEP1B	701	Meprin A subunit beta
Q96NY8	NECTIN4	510	Nectin-4
Q8N2Q7	NLGN1	863	Neuroligin-1
Q04721	NOTCH2	2471	Neurogenic locus notch homolog protein 2
Q9UM47	NOTCH3	2321	Neurogenic locus notch homolog protein 3
Q99466	NOTCH4	2003	Neurogenic locus notch homolog protein 4
P01135	TGFA	160	Protransforming growth factor alpha
O60603	TLR2	784	Toll-like receptor 2
O00206	TLR4	839	Toll-like receptor 4
Q07011	TNFRSF9	255	Tumor necrosis factor receptor superfamily member 9

Table.2 . Cleavage sites targeted by ADAM10

Gene	Number(P1)	Number(P1')	P5	P4	P3	P2	P1		P1'	P2'	P3'	P4'	P5'
APP	687	688	V	H	H	Q	K	-	L	V	F	F	A
BTC	42	44	S	P	E	T	N	(G)	L	L	C	G	D
	111	112	V	D	L	F	Y	-	L	R	G	D	R
CDH1	700	701	R	K	A	Q	P	-	V	E	A	G	L
CDH2	714	715	T	D	V	D	R	-	I	V	G	A	G
DLL1	535	536	D	L	T	E	K	-	L	E	G	Q	G
EGF	970	971	H	Y	S	V	R	-	N	S	D	S	E
	1023	1024	W	W	E	L	R	-	H	A	G	H	G
FASLG	126	127	H	T	A	S	S	-	L	E	K	Q	I
	129	130	S	S	L	E	K	-	Q	I	G	H	P
FCER2	80	81	G	D	Q	M	A	-	Q	K	S	Q	S
	101	102	A	E	Q	Q	R	-	L	K	S	Q	D
HBEGF	149	150	G	L	S	L	P	-	V	E	N	R	L
IL6R	357	358	S	L	P	V	Q	-	D	S	S	S	V
NOTCH1	1720	1721	Y	K	I	E	A	-	V	Q	S	E	T
PMEL	583	584	V	V	S	T	Q	-	L	I	M	P	G
TACSTD2	87	88	P	K	N	A	R	-	T	L	V	R	P

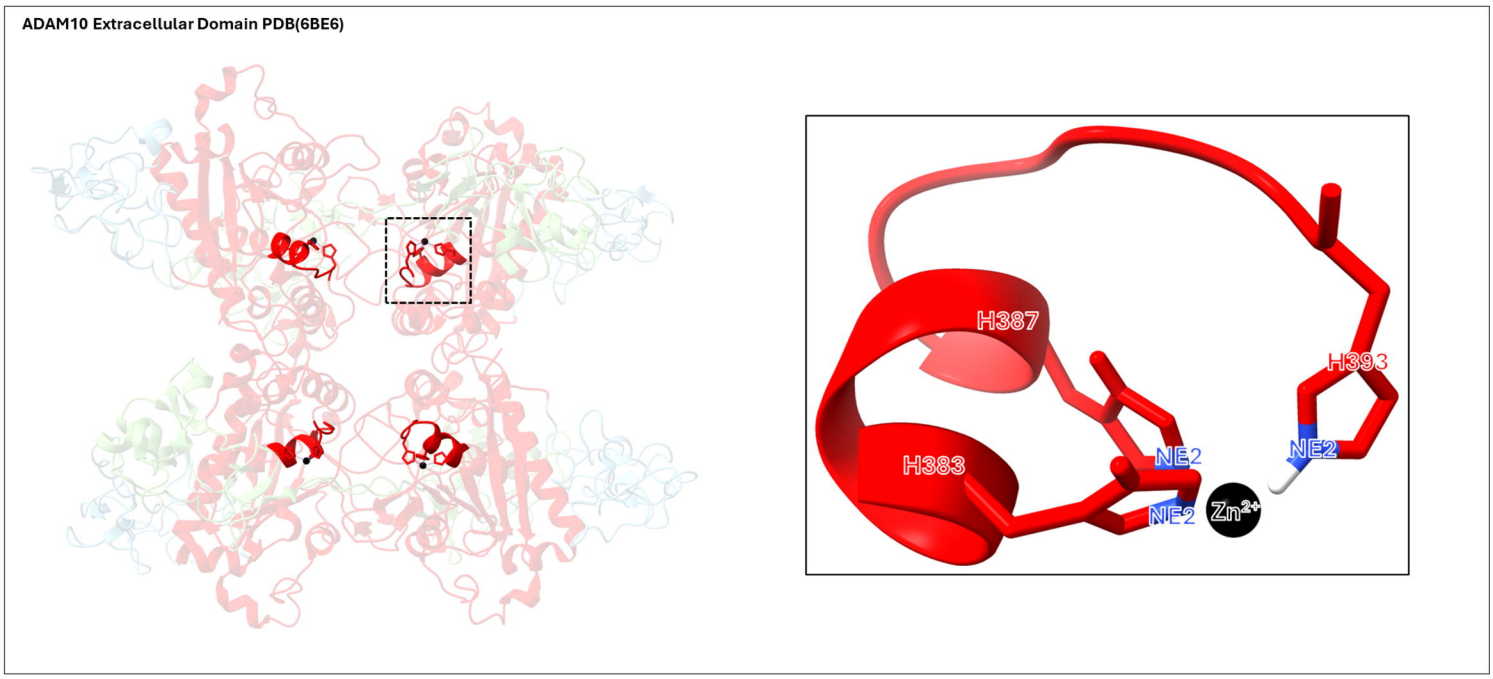

A
ADAM10
1 19 213 456 551 672 693 748
S Pro MP Dis Cys M Cyto
Inactive
ADAM10(Pro cleavage)
213 456 551 672 693 748
MP Dis Cys M Cyto
Active
G329-S340
B
Substrate(TACSTD2)
1 274 297 323
Extra M^Sub Cyto^Sub
C
ADAM10+Substrate(TACSTD2)
Pro
MP
S340
G329
Cyto^Sub
M^Sub
M
Dis
Extra
Cyto
S
Cys
MP
S340
Pro
①
②
Extra
G329
D
ADAM10(Pro cleavage)+Substrate(TACSTD2)
Extra
MP
S340
G329
M
Cyto^Sub
Dis
M^Sub
Cyto
Cys
MP
Extra
③
S340
H-bond
②′
G329
E
ADAM10
G329-S340
Pro-domain(Pro)
Metalloprotease domain(MP)
①
②
Solvent-Accessible Surface Area (SASA)(Å²)
Residue Number
Substrate(TACSTD2)
Extracellular region(Extra)
Solvent-Accessible Surface Area (SASA)(Å²)
Residue Number
F
ADAM10(Pro cleavage)
G329-S340
Pro-domain(Pro)
Metalloprotease domain(MP)
②′
Solvent-Accessible Surface Area (SASA)(Å²)
Residue Number
Substrate(TACSTD2)
Extracellular region(Extra)
③
Solvent-Accessible Surface Area (SASA)(Å²)
Residue Number


A
Substrate(TACSTD2)
1
R87
274
297
323
Extra
Extra
M Sub
Cyto Sub
T88
B
Substrate(TACSTD2)
Extra
R87
T88
M Sub
C
β-sheet
2 (11.8%)
α-helix
3 (17.6%)
Linear
12 (70.6%)
D
Substrate(FASLG)
Extra
C-terminal
M Sub
N-terminal
K129
Q130
L127
S126
E
Substrate(FCER2)
Extra
C-terminal
M Sub
N-terminal
R101
L102
Q81
A80
F
Substrate(PMEL)
Extra
N-terminal
C-terminal
M Sub
Q583
L584

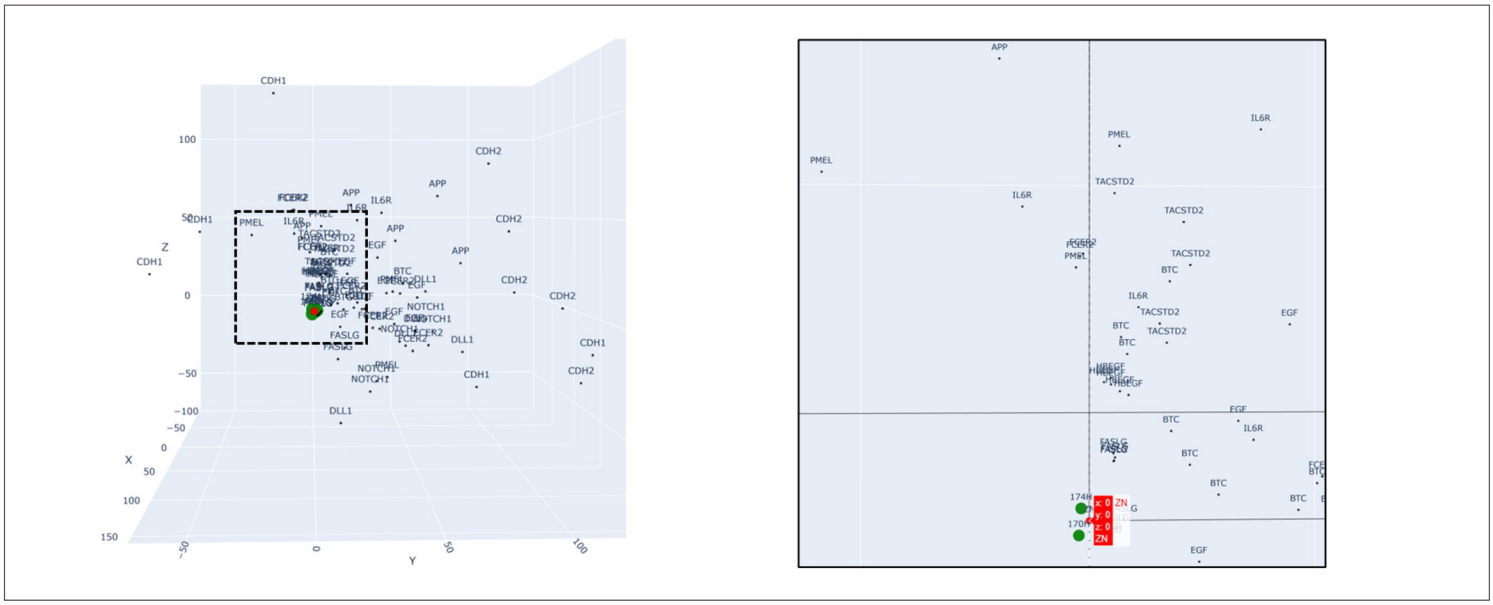
A

ADAM10(Extracellular Domain)
213
456
551
672
MP
Dis
Cys
H383, H387, H393

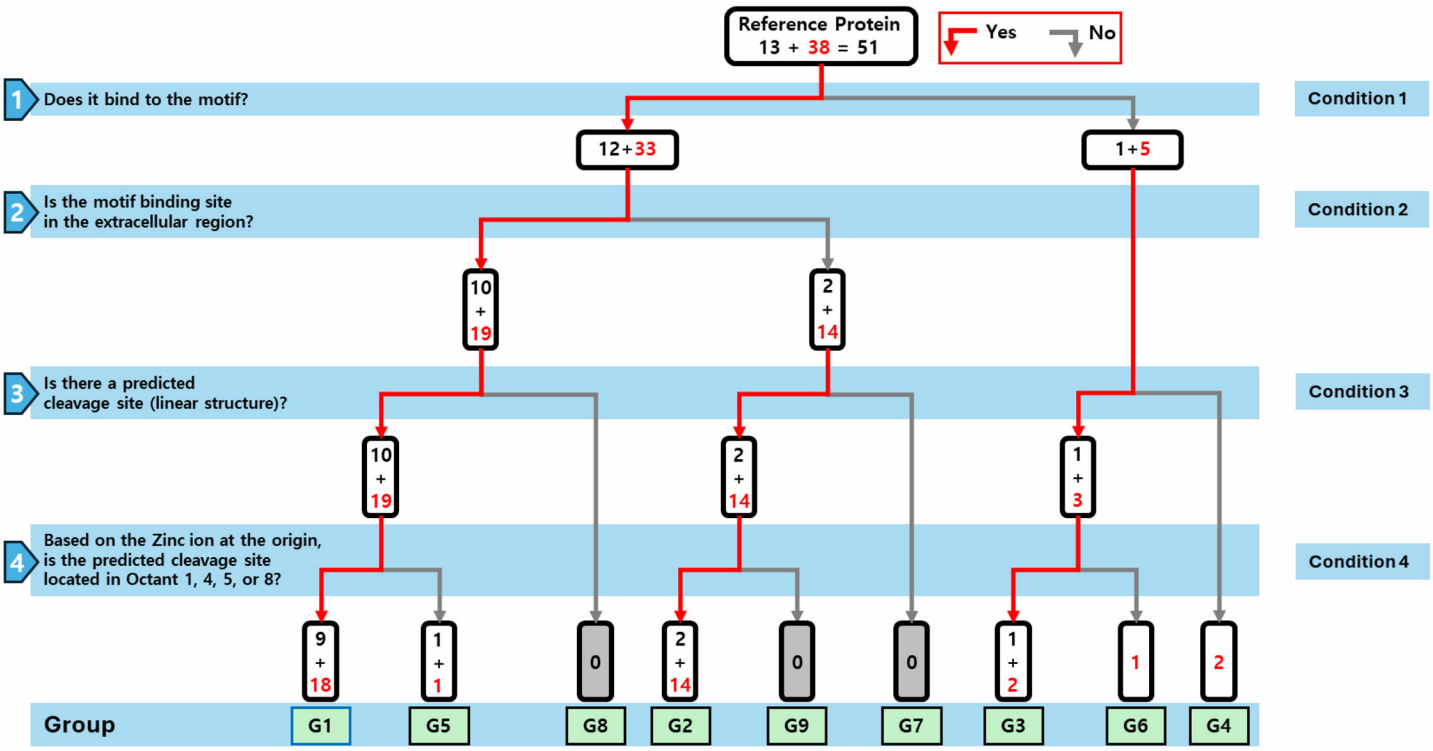
B


ADAM10 Extracellular Domain PDB(6BE6)
H387
H393
H383
NE2
NE2
NE2
Zn2+

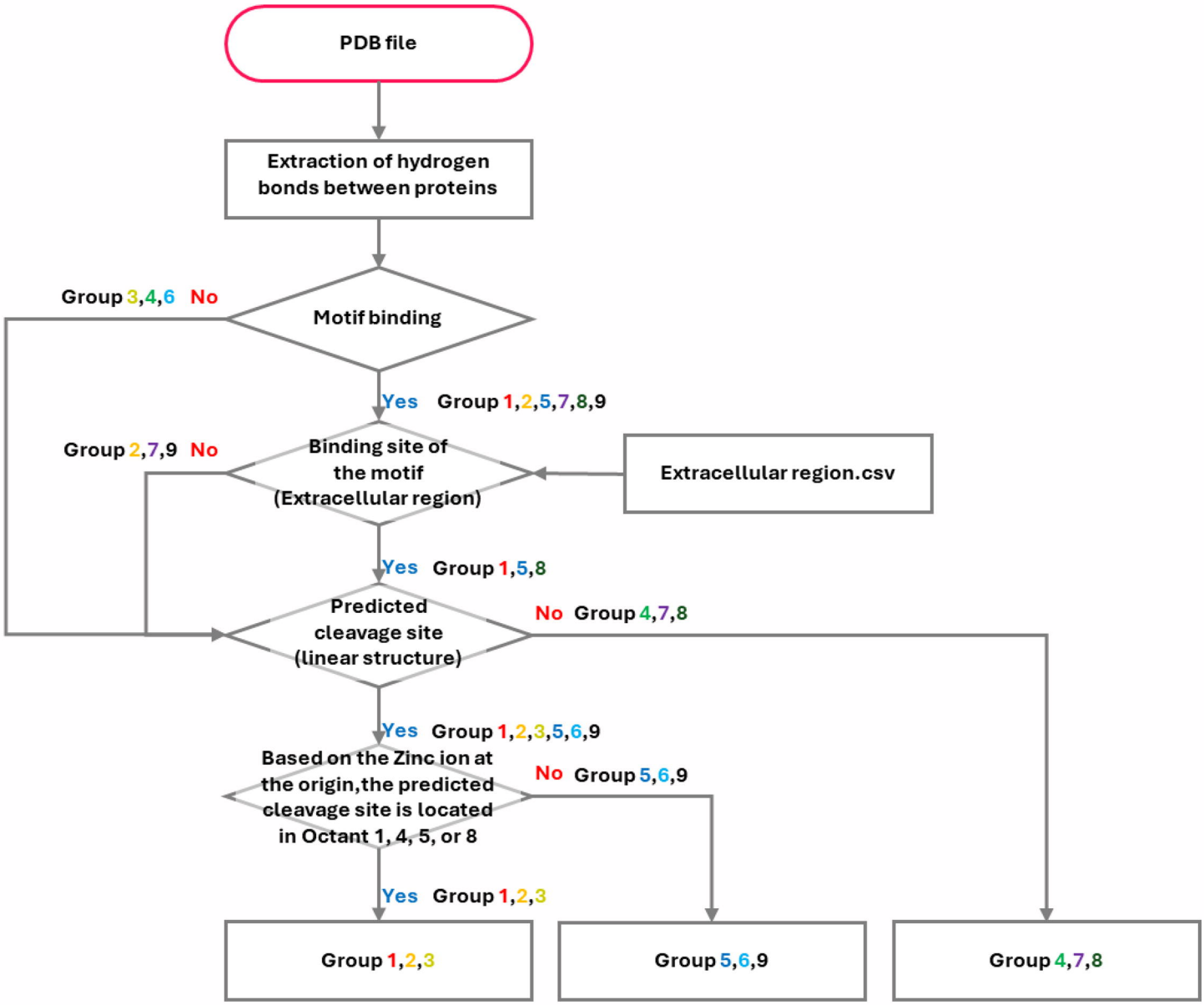
C


CDH1
CDH2
APP
IL6R
PMEL
TACSTD2
EGF
BTC
NOTCH1
DLL1
FASLG
X
Y
Z

A


Reference Protein
13 + 38 = 51
Yes
No
1 Does it bind to the motif?
Condition 1
12+33
1+5
2 Is the motif binding site in the extracellular region?
Condition 2
10 + 19
2 + 14
3 Is there a predicted cleavage site (linear structure)?
Condition 3
10 + 19
2 + 14
1 + 3
4 Based on the Zinc ion at the origin, is the predicted cleavage site located in Octant 1, 4, 5, or 8?
Condition 4
9 + 18
1 + 1
0
2 + 14
0
0
1 + 2
1
2
Group
G1
G5
G8
G2
G9
G7
G3
G6
G4

B


PDB file
Extraction of hydrogen bonds between proteins
Motif binding
Group 3,4,6 No
Yes Group 1,2,5,7,8,9
Binding site of the motif (Extracellular region)
Group 2,7,9 No
Extracellular region.csv
Yes Group 1,5,8
Predicted cleavage site (linear structure)
No Group 4,7,8
Yes Group 1,2,3,5,6,9
Based on the Zinc ion at the origin, the predicted cleavage site is located in Octant 1, 4, 5, or 8
No Group 5,6,9
Yes Group 1,2,3
Group 1,2,3
Group 5,6,9
Group 4,7,8