Cryo-EM structures reveal the activation and substrate recognition mechanism of human enteropeptidase

Xiaoli Yang1,3, Zhanyu Ding1,2,3, Lisi Peng1,3, Qiuyue Song1, Deyu Zhang1, Fang Cui1, Chuanchao Xia1, Keliang Li1, Hua Yin1, Shiyu Li1, Zhaoshen Li1, & Haojie Huang1

Enteropeptidase (EP) initiates intestinal digestion by proteolytically processing trypsinogen, generating catalytically active trypsin. EP dysfunction causes a series of pancreatic diseases including acute necrotizing pancreatitis. However, the molecular mechanisms of EP activation and substrate recognition remain elusive, due to the lack of structural information on the EP heavy chain. Here, we report cryo-EM structures of human EP in inactive, active, and substrate-bound states at resolutions from 2.7 to 4.9 Å. The EP heavy chain was observed to clamp the light chain with CUB2 domain for substrate recognition. The EP light chain N-terminus induced a rearrangement of surface-loops from inactive to active conformations, resulting in activated EP. The heavy chain then served as a hinge for light-chain conformational changes to recruit and subsequently cleave substrate. Our study provides structural insights into rearrangements of EP surface-loops and heavy chain dynamics in the EP catalytic cycle, advancing our understanding of EP-associated pancreatitis.

Human enteropeptidase (hEP), also known as enterokinase, is an essential enzyme in food digestion and is localized at the brush border of the duodenal and jejunal mucosa1–3. Activated hEP can stimulate the conversion of trypsinogen to trypsin via cleavage of a specific trypsinogen activation peptide, namely Asp-Asp-Asp-Asp-Lys (DDDDK), which is highly conserved in vertebrates4–6. Then trypsin initiates a cascade of activations of various pancreatic zymogens, for example, chymotrypsinogen, proelastase, procarboxypeptidase, and prolipase, to regulate nutrient absorption and metabolism7–10. hEP also plays a pivotal role in body homeostasis11–14. Duodenopancreatic reflux of activated hEP can trigger a pancreatic enzyme cascade leading to acute necrotizing pancreatitis15–18, whereas translocation of hEP into the pancreatic-biliary tract is associated with pancreatitis19, suggesting hEP as a potential drug target for pancreatitis treatment20–22.

hEP is a type II transmembrane serine protease that is synthesized as a catalytically inactive zymogen that can be activated by trypsin or related proteases23–25. However, the mechanism of hEP activation remains to be established26–28. hEP contains a multi-domain heavy chain and a catalytic light chain, linked by a disulfide bond29. The heavy chain contains a single-helix transmembrane (TM) domain and seven structural motifs: one copy of Sea urchin sperm protein, Enteropeptidase, and Agrin (SEA), meprin-like domain (MAM), and Scavenger Receptor Cysteine-rich Repeat (SRCR); and two copies of Low-Density Lipoprotein Receptor (LDLR) and Complement, Urchin embryonic growth factor, and Bone morphogenetic protein-1 (CUB)25,26 (Fig. 1a). The functions of these domains and motifs may be involved in protein anchoring, macromolecular substrate recognition and inhibitor specificity25,27,28. The light chain is homologous to trypsin-like serine proteinases with a typical Asp-His-Ser (D-H-S) catalytic triad responsible for peptidase activity29–31.

Structural exploration of full-length hEP is crucial for understanding how the heavy chain modulates the full function of hEP. Previous crystal structures of the EP light chain revealed that the light chain has a typical trypsin-like serine protease fold32–34. Structural
**Results**

**Cryo-EM structure of hEP in the inactive state**

We initially determined the cryo-EM structure of the hEP protein starting from the SEA domain (residues 48–1019), i.e., with the TM domain truncated (Fig. 1a). However, this cryo-EM reconstruction was only obtained at low resolution (Supplementary Fig. 2c). The poor resolution was attributed to the presence of the SEA domain as previous reports indicated that the SEA domain might be involved in protein autocleavage, leading to increased instability. Since the SEA domain has been shown to be dispensable for the enzymatic activity, a further truncated hEP including only residues 182–5019, i.e., also leaving out the SEA domain (Fig. 1a), was used for further high-resolution structural analysis. The resulting purified hEP was a heavily glycosylated single-chainzymogen with a molecular weight (>130 kDa) higher than that theoretically predicted of the amino acids alone (95 kDa). An enzymatic activity assay performed in vitro confirmed its inactive state with no detectable cleavage. The overall architecture of this inactive hEP was the same as that containing the SEA domain (residues 48–1019), showing a clamp shape adopted by the heavy chain with the light chain in the center (Fig. 1b and Supplementary Fig. 2c). The LDLR2, CUB2, and SRCR domains of the heavy chain and the peptidase domain of the light chain, remaining in lower local resolution likely due to high flexibility of the LCM relative to the core region (Fig. 1c and Supplementary Fig. 2c). In addition to the disulfide bond (Cys772-Cys896) between the light and heavy chains, the CUB2 and SRCR domains stabilized the peptidase domain by forming hydrogen bonds between Glu851 and Arg871, and between Ser771 and Gln893 (Supplementary Fig. 2e). The CUB2 and SRCR domains were interconnected by the LDLR2 domain, and clamped the peptidase domain with a total interaction surface of 1163.2 Å² (Supplementary Fig. 2e).

**Bone morphogenetic protein-1.** a Cryo-EM reconstruction of inactive hEP. The reconstruction is a composite map generated from the locally refined hEP-core and the low-resolution region of hEP-complete. The domain color scheme follows that in (a). Unless otherwise stated, the same domain color scheme is applied to all figures. c Structural model of the inactive hEP, with the map displayed as semi-transparent. The nine detected N-linked glycans are shown as stick atomic models, and the amino acid residues to which they are covalently attached are labeled. LCM: LDLR, CUB, and MAM domains.

---

**Fig. 1 | Cryo-EM structure of hEP in the inactive state.** a Domains of hEP. Numbers indicate the N- or C-terminus positions of the domains in the amino acid sequence. The unexpressed domains (TM and SEA), as well as the flexible domains with a lower resolution (LDLR, CUB, and MAM), were colored in grey, CUB2 domain in pink, LDLR2 domain in light green, SRCR domain in orange, and the peptidase domain in dodger blue. This domain color scheme was used throughout this study. SEA: Sea urchin sperm protein, Enteropeptidase, and Agrin; MAM: meprin-like domain; SRCR: Scavenger Receptor Cysteine-rich Repeat; LDLR: Low-Density Lipoprotein Receptor; CUB: Complement, Urchin embryonic growth factor, and LDLR2 domain in light green, SRCR domain in orange, and the peptidase domain in dodger blue. This domain color scheme was used throughout this study. SEA: Sea urchin sperm protein, Enteropeptidase, and Agrin; MAM: meprin-like domain; SRCR: Scavenger Receptor Cysteine-rich Repeat; LDLR: Low-Density Lipoprotein Receptor; CUB: Complement, Urchin embryonic growth factor, and LDLR2 domain in light green, SRCR domain in orange, and the peptidase domain in dodger blue. This domain color scheme was used throughout this study. SEA: Sea urchin sperm protein, Enteropeptidase, and Agrin; MAM: meprin-like domain; SRCR: Scavenger Receptor Cysteine-rich Repeat; LDLR: Low-Density Lipoprotein Receptor; CUB: Complement, Urchin embryonic growth factor, and LDLR2 domain in light green, SRCR domain in orange, and the peptidase domain in dodger blue. This domain color scheme was used throughout this study.
**Mechanism of the activation of hEP**

Upon cleavage of the hEP into heavy and light chains by trypsin, the enzyme became fully active as shown by the protease activity assay (Supplementary Fig. 1a, b). The protruding LCM region was not resolved in the active state, corresponding to the surface loops L1, L2 and LD not resolved (Supplementary Fig. 2a, b), suggesting a near-physiological state of our inactive hEP structure.

**Inhibition of hEP activity**

Inhibition of serine proteases has emerged as a novel therapy for many diseases, including Alzheimer’s disease, autoimmune disease, and COVID-19. The development of specific inhibitors of hEP has aided clinical intervention in pancreatitis, with nafamostat and camostat as the most promising of these inhibitors (Fig. 3a, b). Nafamostat, whose structure is shown in Fig. 3c, is a broad-spectrum synthetic serine protease inhibitor, with an IC50 value of 1.5 μM for the inhibition of hEP activity, and hence more potent than camostat in inhibiting other serine proteases; moreover, nafamostat has been reported to be clinically useful in the treatment of acute pancreatitis.

We measured the binding affinity of nafamostat to active hEP to be 95.1 μM (Supplementary Fig. 5a). And active hEP incubated with nafamostat was subjected to cryo-EM structure determination.
After 3D classification, ~58% of the particles showed no density for the LCM domain, resulting in a 3.1-Å-resolution map of the inhibited hEP-core (Fig. 3d, Supplementary Fig. 5e–g, and Supplementary Table 1). Although this inhibited hEP-core map was almost identical to the active map, with an r.m.s. deviation of 0.605 Å (Supplementary Fig. 6a, b), extra density was found in the catalytic pocket, which could be modelled as the reaction product of nafamostat with hEP. The model of this reaction product was found to covalently bind to the catalytic residue Ser971 and form an electrostatic interaction with the conserved Asp965 (Fig. 3e and Supplementary Fig. 6c), similar to the binding of camostat to the hEP light chain10.

Cryo-EM reconstruction of another ~29% of the particles of the active hEP-nafamostat complex showed an intact hEP with the LCM domain rotated ~120 degrees from the position of the inactive hEP around the loop between MAM and CUB2 domains, with this inhibited hEP-LCM in close proximity to the LDLR2 domain (Fig. 3f, g, Supplementary Figs. 5e, 6d, e). This considerable flexibility was thought to likely facilitate substrate recruitment and expose the catalytic site for subsequent cleavage. The light chain was observed to remain firmly clamped by the CUB2, LDLR2, and SRCR domains of the heavy chain within the core region (Fig. 3f), allowing for the intimate interaction between the heavy and light chains after the hEP-activating cleavage. The inhibitor nafamostat was thus found to covalently bind to Ser971 and induce an apparent change in the position of the hEP core region relative to the LCM domain.
Substrate-engaged hEP

Given the critical role of pancreatic enzymes in the pathogenesis of acute pancreatitis, knowledge of the normal activation process of humanzymogens initiated by hEP-cleaved trypsinogen to trypsin is apparently of great interest. In a previous study, the heavy chain of hEP was found to be necessary for efficient substrate recognition. In order to capture the substrate-bound state of hEP, we incubated active hEPmut and hEPwild with the physiological substrate trypsinogen. As expected, active hEPmut cleaved the substrate, while the active hEPwild failed to do so (Supplementary Fig. 7a). The association of trypsinogen with active hEPmut was validated by a distinct migration pattern in native gel electrophoresis (Supplementary Fig. 7b). Thus, we incubated the active hEPmut with trypsinogen and purified the resulting complex using size exclusion chromatography, and did so in the presence of the cross-linker glutaraldehyde to maintain the complex for cryo-EM analysis (Supplementary Fig. 7c, d).

3D reconstruction of trypsinogen-bound hEPmut yielded a cryo-EM map at a resolution of 4.9 Å, with this relatively low resolution due to the particles having adopted a preferred orientation, resulting in a conformation of active hEPmut with fewer high-resolution features (Fig. 4a, Supplementary Fig. 8, and Supplementary Table 1). However, the active hEPmut alone can generate a normal hEP map, showing an overall shape equivalent to that of the active hEPmut with a well-organized catalytic pocket (Supplementary Fig. 4d–h). Thus, the engagement of substrate changed the performance of the active hEPmut under the cryoEM condition.

The overall architecture of the substrate-bound hEPmut was observed to be similar to that of the inactive hEPwt, with extra density observed to be attached to CUB2; this extra density could be attributed to trypsinogen (Fig. 4a, b and Supplementary Fig. 9a). A full-length model of trypsinogen including its uncleaved N-terminal peptide was constructed based on a prediction using AlphaFold2 and the model was docked as a rigid body into the extra density (Supplementary Fig. 9b). This apparent binding of trypsinogen to CUB2 validated the previous reports suggesting CUB2 to be essential for the recognition of natural substrates, and explaining the slow cleavage requires trypsin or other related proteases for activation. Upon activation of the zymogen, the flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LD loops form a flexible L1, L2 and LDLR2 located in the catalytic triad of hEP. The catalytic triad on hEP is colored red. The cleavage site, DDDDK, is also labeled.

Discussion

Acute pancreatitis is a complex inflammatory disease of the pancreas, and its exact pathogenesis remains unclear. hEP was discovered over 100 years ago as the physiological activator of trypsinogen. However, the structural basis of full-length hEP processing, substrate binding, and small-molecule inhibition remains elusive. Inappropriate activation of trypsinogen, such as overactivation of trypsinogen by duodenopancreatic reflux of hEP, and premature or reduced degradation of trypsinogen caused by mutations in trypsinogen are thought to contribute to acute pancreatitis, and acquired hEP deficiency may be the origin of indigestion and malabsorption caused by functional pancreatic insufficiency. Our structural study of hEP and its complexes has marked an important step in deciphering the pathogenesis of pancreatitis, and may open up a direction for the treatment of this disease.

Cryo-EM structures of hEP in inactive, active, inhibited and substrate-bound states revealed conformational changes of the hEP core region relative to the LCM domain, stabilization of the L1, L2, and LDL loops around the catalytic site in the active structure, a mechanism of inhibition of hEP activity by nafamostat involving covalent binding, and a snapshot of substrate recognition. In general, hEP zymogen is anchored on the brush border of the duodenal and jejunal mucosa with a single transmembrane helix at the N-terminus, with this domain followed by the SEA, LCM, CUB2, LDLR2, SCR and light-chain peptidase domains (Fig. 1a). In the catalytic process, the zymogen first requires trypsin or other related proteases for activation (Fig. 5). Upon activation of the zymogen, the flexible L1, L2 and LDL loops form a complete and rigid catalytic pocket ready for performing catalysis. The binding of nafamostat or other covalent inhibitors to the catalytic site apparently results in a change of the conformation of the hEP core available in the published PDB structures, for they were all in the active state. Active hEP stimulates the conversion of trypsinogen to trypsin by cleaving this N-terminal tail.

![Fig. 4](https://doi.org/10.1038/s41467-022-34364-9)
substrate binding and nafamostat inhibition, and has thus provided an inhibition mechanism that may shed light on pancreatitis.

CUB2 with the substrate N-terminal tail placed into the catalytic site for cleavage. The complex marked by the dashed line displays one of the core becomes dynamic to facilitate the exposure of the catalytic site and recruit-EP to its dynamic state to recruit and catalyze the cleavage of more trypsinogen. Substrates such as trypsinogen bind to CUB2 with the N-terminal tail placed into the catalytic site for cleavage (Fig. 5).

Our study has revealed the structural basis of hEP activation, substrate binding and nafamostat inhibition, and has thus provided an improved understanding of the hEP catalysis process and covalent inhibition mechanism that may shed light on hEP-associated pancreatitis.

Methods

Protein expression and purification

Two forms of wild type hEP (corresponding to residues 48–1019 and residues 182–1019) and a mutant hEP (H825A, D876A and S971A, residues 182–1019) were cloned using homologous recombination into, respectively, a pcDNA3 vector with an N-terminal signal peptide and 10 × His tags (HieffCloneTM One Step Cloning Kit, Yeasen). For each protein, HEK293F cells (A14635, ThermoFisher Scientific Inc.) with a ratio of transfection reagent: DNA is 5 μg/μl were transfected by plasmid and Sinofection reagent (STF02, Sino Biological Inc.), with a ratio of transfection reagent: DNA is 5 μg/μl. After 7 days of cell expansion, the cell supernatant was harvested, protected care) in PBS buffer. Then the column was washed with a wash buffer containing 95 μl of trypsinogen (1 mg/ml in buffer A (20 mM Tris-HCl, pH 7.6, 150 mM NaCl, and 5 mM CaCl2)) was incubated with purified 50 ng/μl hEP in buffer A (20 mM Tris-HCl, pH 7.6, 150 mM NaCl, and 5 mM CaCl2) at a final mass ratio of I: 100 for 2 h at 37 °C.

The hEP activity was determined using an enteropeptidase activity assay kit (K758-100, Biovision). Briefly, after a volume of 5 μl of hEP alone or that incubated with trypsin was combined with 95 μl of enteropeptidase assay buffer containing 2 μl of enteropeptidase substrate. The relative fluorescence units (RFU) (excitation/emission wavelengths = 380/500 nm) per well at 5 minutes and 0 minutes were subsequently measured using a fluorescence microplate reader (SpectraMax i3, Molecular Devices). Then the enteropeptidase activity levels of the samples were calculated from the fluorescence readings according to the manufacturer’s instructions. Experiments were performed at least three times. GraphPad Prism software v8.3.0 was used to construct the statistic graphs. The proteins were also assessed using 8% SDS-PAGE and 8% native-PAGE.

The ability of hEP cleaves trypsinogen was determined by the trypsin activity colorimetric assay kit (MAK290, Sigma-Aldrich). Briefly, add 0, 2, 4, 6, 8, and 10 μl of p-NA Standard into a series of wells, respectively. Adjust volume to 50 μl/well with trypsin assay buffer to generate 0, 8, 12, 16, and 20 nmol/well of the p-NA standard. A volume of 5 μl of hEP, active-hEP, hEPmut (50 ng/μl) and active hEPmut (50 ng/μl) in buffer A were combined, respectively, with 95 μl of trypsin assay buffer containing 2 μl of trypsin substrate and 5 μl trypsinogen (100 ng/μl). The trypsin activity levels of the samples were calculated from the optical density readings according to the manufacturer’s instructions. Experiments were performed at least three times. GraphPad Prism software v8.3.0 was used to construct the statistic graphs.

Preparation of active hEP by using trypsin

Trypsin at 10 ng/μl (T1426, Sigma-Aldrich) (20 mM Tris-HCl, pH 4.0, 150 mM NaCl, and 5 mM CaCl2) was incubated with purified 50 ng/μl hEP in buffer A (20 mM Tris-HCl, pH 7.6, 150 mM NaCl, and 5 mM CaCl2) at a final mass ratio of I: 100 for 2 h at 37 °C.

The hEP activity was determined using an enteropeptidase activity assay kit (K758-100, Biovision). Briefly, after a volume of 5 μl of hEP alone or that incubated with trypsin was combined with 95 μl of enteropeptidase assay buffer containing 2 μl of enteropeptidase substrate. The relative fluorescence units (RFU) (excitation/emission wavelengths = 380/500 nm) per well at 5 minutes and 0 minutes were subsequently measured using a fluorescence microplate reader (SpectraMax i3, Molecular Devices). Then the enteropeptidase activity levels of the samples were calculated from the fluorescence readings according to the manufacturer’s instructions. Experiments were performed at least three times. GraphPad Prism software v8.3.0 was used to construct the statistic graphs.

Preparation of inhibited hEP

Various compounds, including nafamostat, camostat, gabexate, PMSF, and ulinastatin (Topscience), were first tested to find the best inhibitor of hEP. PMSF was diluted in DMSO and the other above inhibitors were diluted in buffer A, in each case to a final concentration of 10 mg/ml for screening. Each inhibitor solution was then mixed with 5 μl of the solution of active hEP (50 ng/μl) at a molar ratio of 3000:1 at 37 °C for 30 minutes. Then the hEP enzymatic activity of each mixture was determined using the above-mentioned enteropeptidase activity assay kit (K758-100, Biovision). GraphPad Prism software v8.3.0 was used to construct the statistic graphs.

As stated in Results, nafamostat was more potent than camostat in inhibiting other serine proteases41,42, and reported to be clinically useful in the treatment of acute pancreatitis14,15, hence it was chosen to find the best inhibit efficacy. Samples of the nafamostat solution (10 mg/ml) were mixed with samples of 5 μl of active hEP (50 ng/μl) at molar ratios of 50:1, 100:1, 500:1, and 1000:1, respectively. The hEP activity of each mixture was then determined as mentioned above. GraphPad Prism software v8.3.0 was used to construct the statistic graphs.

Preparation of substrate-bound hEP

A mixture containing a 1:5 molar ratio of hEP (1 mg/ml in buffer A) to trypsinogen (1 mg/ml in 20 mM Tris-HCl, pH 4.0, 150 mM NaCl, and 5 mM CaCl2) was incubated in 1× PBS at 37 °C for 30 minutes, protected from light; then the products of this incubation were detected using native-PAGE to evaluate the association of trypsinogen with hEP. In addition, two experiments involving each an incubation of a mixture of
the hEPmut (1 mg/ml in buffer A) with trypsinogen (1 mg/ml) at a 1:5 molar ratio were also conducted for 30 minutes at room temperature in the dark with one experiment also including 0.05% glutaraldehyde to crosslink the hEPmut and trypsinogen in order to stabilize their complex, and the other experiment without crosslinker added; in the former experiment, the cross-linking was stopped after the 30 minutes of incubation by adding Tris, pH 8.0, to a final concentration of 1 mM. To obtain the highest purity of substrate-bound hEPmut, gel filtration was performed on each incubated sample by passing the sample through a Superdex 200 (24 ml) column (GE Healthcare) with buffer A. The gel fractions containing trypsinogen-hEPmut were identified by using 8% SDS-PAGE, and proteins were concentrated for cryo-EM data collection.

Surface plasmon resonance analysis

Commercial hEP (7136-50, BioVision) was immobilized onto a CM5 sensor chip surface by using the NHS/EDC method with a Biacore Sk (Cytiva) and 1× PBS-T running buffer (1× PBS with 0.05% Tween 20). Then nafamostat solutions at concentrations of 1.95 μM, 3.9 μM, 7.8 μM, 15.625 μM, and 31.25 μM in the 1× PBS-T buffer were injected to flow over a different chip at a rate of 30 ml/min. The binding is monitored by subsequent changes in the refractive index of the medium close to the sensor surface upon injection, then the quantitative binding parameters were obtained. The resulting surface plasmon resonance data were analyzed using Affinity implemented in Biacore Insight Evaluation Software v.3.0.

Cryo-EM sample preparation and data collection

Samples with a volume each of 3 μl and concentrations of 1–1.5 mg/ml were placed onto glow-discharged holey amorphous nickel-titanium alloy film supported by 400-mesh gold grids2, then blotted by deploying a Vitrobot Mark IV (FEI/Thermo Fisher Scientific), and flash frozen in liquid ethane.

Images were taken by using a Titan Krios transmission electron microscope (Thermo Fisher Scientific) operated at 300 kV and equipped with a Bio Quantum post-column energy filter with a zero-loss energy selection slit set to 20 eV. Images were collected by using a K2 Summit direct electron detector (Gatan) (K3 for the substrate-bound dataset) in super-resolution counting mode, corresponding to a pixel size of 0.523 Å (0.425 Å for the substrate-bound dataset) at the specimen level. Except for that in the substrate-bound dataset, each movie was dose-fractioned into 40 frames with a dose rate of 8 e per pixel per second on the detector. The exposure time was 7.2 s with 0.2 s for each frame, generating a total dose of ~52 e−/Å2. Defocus values varied from ~0.7 to ~3.5 μm. All of the images were collected by using the SerialEM 3.9.0 automated data collection software package8. For the substrate-bound dataset, each movie was dose-fractioned into 40 frames with a dose rate of 16 e per pixel per second on the detector. The exposure time was 2.18 seconds, generating a total dose of ~50 e−/Å2. Defocus values varied from ~1.4 to ~2.4 μm. All of the images were collected by using EPU 2.8.1 (Thermo Fisher Scientific).

Cryo-EM data processing

Unless otherwise specified, single-particle analysis was mainly executed in cryoSPARC 2.15, including patch motion correction and patch CTF estimation. In the dataset collected from inactive hEP, the particles were auto-picked by using the blob picker in cryoSPARC9, generating a dataset of 5,550,766 particles. After 2 rounds of 2D classification (particles binned 4 × 4 during extraction), the good class averages appeared white on a black background, showing internal secondary structural elements. The class averages showed relatively few features and a noisy background were discarded. The remaining 1,527,467 particles were subjected to ab initio reconstruction, followed by one round of heterogeneous refinement. After visual inspection of the resulting 4 classes using UCSF Chimera 1.1410, the map in class one showed clear features with continuous density and could be made to match the 2D class averages. Thus, 474,076 particles within class one were selected to perform another round of 2D classification in order to generate good class averages as the template for picking particles in the remaining datasets. Four rounds of heterogeneous refinement were performed to classify the particles from the template picker, resulting in one class with good features. The selected good classes from Blob and template pickers were combined together and further classified by carrying out two rounds of heterogeneous refinement. The particles from the good class were re-extracted with the original size (pixel size = 1.046 Å), and were subjected to further classification. The bad class was removed to obtain a cleaner particle stack, and the remaining three classes were combined to be reconstructed into a complete map, while the core map was locally refined by applying a hEP-core mask. The locally refined hEP-core and the low-resolution region in the whole map were merged in UCSF Chimera 1.14 and used for subsequent model building and analysis.

Model building and validation

The model of inactive hEP predicted by AlphaFold2,46 was initially fitted into the density maps using UCSF Chimera 1.1411 and manually adjusted using COOT 0.8.9.11. All the high-resolution models were subjected to multiple rounds of real-space refinement against the corresponding maps by using PHENIX 1.17.165. In the inhibited hEP-complete structure, the EP-core from inhibited hEP-core structure and the LCM domain from inactive hEP-complete structure were fitted into the density map as rigid bodies. In the substrate-engaged hEP map, the EP model from inactive hEP-complete structure and a full-length model of trypsinogen predicted using AlphaFold2,46 were fitted into the density map as rigid bodies, except that the N-terminus of the substrate was manually adjusted. Thus, the related overall structure models were not further flexible refined.

The nafamostat and glycans were added using COOT 0.8.9.11. To identify the glycans, we referred to the information in Uniprot database (https://www.uniprot.org/uniprot/PT98073#tm_processing). Uniprot showed 18 potential N-linked glycosylation sites, displaying GlcNAc-branched N-glycans. Nine N-linked GlcNAc were clearly detected in our cryo-EM maps of the hEP-core structure.

Model validation was performed following the phenix.molprobity protocol2,46. UCSF Chimera 1.1411 and ChimeraX 1.117 were used for map segmentation and figure generation.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The data that support this study are available from the corresponding authors upon reasonable request. EM maps have been deposited in the Electron Microscopy Data Bank (EMDB) under accession codes of EMD-32715 (Inactive-EP), EMD-32714 (Active-EP), EMD-32716 (Active-mutEP), EMD-32717 (Inhibited-EP), EMD-32828 (Inhibited EP-complete), and EMD-32829 (Substrate bound EP). Atomic coordinates have been deposited in the Protein Data Bank (PDB) under accession numbers of 7WQX (Inactive-EP), 7WQW (Active-EP), 7WQZ...
(Active-mutEP), and 7W87 (Inhibited-EP). For the two low resolution maps, inhibited hEP-complete and substrate-bound hEP, we deposited structural models only modeled as poly-Ala in the PDB under accession numbers of 8H3U (inhibitor-bound EP, polyA model) and 8H3S (substrate-bound EP, polyA model). Other structural model used in this study is available in the PDB with entry code of 4DGJ (X-ray model of hEP light chain variant). The source data underlying Figs. 2F, 3a, b, Supplementary Figs. 1a, b, 7 are provided as a Source Data file. Source data are provided with this paper.

References
1. Kunitz, M. Formation of trypsin from crystalline trypsinogen by means of enterokinase. J. Gen. Physiol. 22, 429–446 (1939).
2. Zheng X. L., Kitamoto Y. Fau-Sadler J. E., Sadler J. E. Enteropeptidase, a type II transmembrane serine protease. Front Biosci (Elite Ed) 2, 242–249 (2009).
3. Light, A. & Janska, H. Enterokinase (enteropeptidase): comparative aspects. Trends Biochem Sci. 14, 110–112 (1989).
4. Light, A. & Fonseca, P. The preparation and properties of the catalytic subunit of bovine enterokinase. J. Biol. Chem. 259, 13195–13198 (1984).
5. Savithri, H. S. & Light, A. Retention of enzymatic activity of bovine enterokinase after a limited reduction of disulfide bonds. Biochim Biophys. Res. Commun. 94, 360–365 (1980).
6. Zamolodchikova, T. S., Sokolova, E. A., Lu, D. & Sadler, J. E. Activation of recombinant proenteropeptidase by duodenase. FEBS Lett. 466, 295–299 (2000).
7. Eggermont, E., Molla, A. M., Tytgat, G. & Rutgeerts, L. Distribution of enterokinase activity in the human intestine. Acta Gastroenterol. Belg. 34, 655–662 (1971).
8. Maroux, S., Baratti, J. & Desnuelle, P. Purification and specificity of porcine enterokinase. J. Biol. Chem. 246, 5031–5039 (1971).
9. Rutgeerts, L. & Eggermont, E. Human enterokinase. Tijdschr. Gastroenterol. 19, 231–246 (1976).
10. Sun, W. et al. Targeting enteropeptidase with reversible covalent inhibitors to achieve metabolic benefits. J. Pharm. Exp. Ther. 375, 510–521 (2020).
11. Wang, L. et al. Novel compound heterozygous TMPRSS15 gene variants cause enterokinase deficiency. Front Genet. 11, 538778 (2020).
12. Austin, A. E. Enterokinase in infancy. J. Med. Res. 20, 137–147 (1909).
13. Hadorn, B., Tarlow, M. J., Lloyd, J. K. & Wolff, O. H. Intestinal enterokinase deficiency. Lancet 1, 812–813 (1969).
14. Holzinger, A. et al. Mutations in the proenteropeptidase gene are associated with congenital enteropeptidase deficiency. Am. J. Hum. Genet. 70, 20–25 (2002).
15. Szabo, R. & Buggle, T. H. Type II transmembrane serine proteases in development and disease. Int J. Biochem Cell Biol. 40, 1297–1316 (2008).
16. Rinderknecht, H. Activation of pancreatic zymogens. Normal activation, premature intrapancreatic activation, protective mechanisms against inappropriate activation. Dig. Dis. Sci. 31, 314–321 (1986).
17. Mann, N. S. & Mann, S. K. Enterokinase. Proc. Soc. Exp. Biol. Med. 206, 114–118 (1994).
18. Grant, D. Acute necrotising pancreatitis - a role for enterokinase. Int J. Pancreatol. 1, 167–183 (1986).
19. Rashid, M. Y. et al. Role of SCO-792, a novel enteropeptidase inhibitor, in the prevention of post-endoscopic retrograde cholangio-pancreatography pancreatitis. Cureus 13, e13724 (2021).
20. Yashiro, H. et al. SCO-792, an enteropeptidase inhibitor, improves disease status of diabetes and obesity in mice. Diabetes Obes. Metab. 21, 2228–2239 (2019).
21. Sugama J., Morito Y., Yashiro H., Tsuchimori K., Watanabe M. Enteropeptidase inhibition improves obesity by modulating gut microbiota composition and enterobacterial metabolites in diet-induced obese mice. Pharmacol. Res. 163, https://doi.org/10.1016/j.phrs.2020.105337 (2021).
22. Sasaki, M. et al. Discovery and characterization of a small-molecule enteropeptidase inhibitor, SCO-792. Pharm. Res. Perspect. 7, e00517 (2019).
23. Lu, D., Yuan, X., Zheng, X. & Sadler, J. E. Bovine proenteropeptidase is activated by trypsin, and the specificity of enteropeptidase depends on the heavy chain. J. Biol. Chem. 272, 31293–31300 (1997).
24. Zamolodchikova, T. S. et al. Subcellular localization, substrate specificity and crystallization of duodenase, a potential activator of enteropeptidase. Eur. J. Biochem. 249, 612–621 (1997).
25. Kitamoto, Y., Yuan, X., Wu, Q., McCourt, D. W. & Sadler, J. E. Enterokinase, the initiator of intestinal digestion, is a mosaic protease composed of a distinctive assortment of domains. Proc. Natl Acad. Sci. USA 91, 7588–7592 (1994).
26. Wu, Q. Type II transmembrane serine proteases. Curr. Top. Dev. Biol. 54, 167–206 (2003).
27. Wang, Z., Guo, C., Liu, L. & Huang, H. Effects of N-glycosylation on the biochemical properties of recombinant bEKL expressed in Pichia pastoris. Enzym. Micro. Technol. 14, 40–47 (2010).
28. Collins-Racie, L. A. et al. Production of recombinant bovine enterokinase catalytic subunit in Escherichia coli using the novel secretory fusion partner DsBA. Biotechnol. (N. Y) 13, 982–987 (1995).
29. Lu, D. et al. Crystal structure of enteropeptidase light chain complexed with an analog of the trypsinogen activation peptide. J. Mol. Biol. 292, 361–373 (1999).
30. Simeonov, P., Zahn, M., Sträger, N. & Zuchner, T. Crystal structure of a supercharged variant of the human enteropeptidase light chain. Proteins 80, 1907–1910 (2012).
31. Xu, J. et al. Structure basis for the unique specificity of medaka enteropeptidase light chain. Protein Cell 5, 178–181 (2014).
32. Levitin, F. et al. The MUC1 SEA module is a self-cleaving domain. J. Biol. Chem. 280, 33374–33386 (2005).
33. Johansson, D. G., Macao, B., Sandberg, A. & Härd, T. SEA domain autoproteolysis accelerated by conformational strain: mechanistic aspects. J. Mol. Biol. 377, 1130–1143 (2008).
34. Deeb, S. J., Cox, J., Schmidt-Supprian, M. & Mann, M. N-linked glycosylation enrichment for in-depth cell surface proteomics of diffuse large B-cell lymphoma subtypes. Mol. Cell. Proteom. 13, 240–251 (2014).
35. Perona, J. J. & Craik, C. S. Evolutionary divergence of substrate specificity within the chymotrypsin-like serine protease fold. J. Biol. Chem. 272, 29987–29990 (1997).
36. Light, A. & Janska, H. The amino-terminal sequence of the catalytic subunit of bovine enterokinase. J. Protein Chem. 10, 475–480 (1991).
37. Brown, A. J. H. et al. From structure to clinic: Design of a muscarinic M1 receptor agonist with potential to treatment of Alzheimer’s disease. Cell 184, 5886–5901 (2021).
38. Wagner, A. et al. Metabolic modeling of single Th17 cells reveals regulators of autoimmunity. Cell 184, 4168–4185.e4121 (2021).
39. Kreutzberger, A. J. B. et al. Synergistic Block of SARS-CoV-2 Infection by Combined Drug Inhibition of the Host Entry Factors Pikfyve Kinase and TMPRSS2 Protease. J. Virol. 95, e0097521 (2021).
40. Fraser, B. J. et al. Structure and activity of human TMPRSS2 protease implicated in SARS-CoV-2 activation. Nat. Chem. Biol. 18, 963–971 (2022).
41. Yamamoto, M. et al. The anti-infective nafamostat potently inhibits SARS-CoV-2 S protein-mediated fusion in a cell fusion assay system and viral infection in vitro in a cell-type-dependent manner. Viruses 12, 629 (2020).
42. Zhu, H. et al. Spontaneous binding of potential COVID-19 drugs (Carnosat and Nafamostat) to human serine protease TMPRSS2. Comput Struct. Biotechnol. J. 19, 467–476 (2021).
43. Iwaki, M. et al. Pharmacological studies of FUT-175, nafamostat mesilate. V. Effects on the pancreatic enzymes and experimental acute pancreatitis in rats. Jpn J. Pharm. 41, 155–162 (1986).

44. Choi, C. W. et al. Nafamostat mesilate in the prevention of post-ERCP pancreatitis and risk factors for post-ERCP pancreatitis. Gastrointest. Endosc. 69, e11–e18 (2009).

45. Jumpers, J. et al. Highly accurate protein structure prediction with AlphaFold. Nature 586, 583–589 (2021).

46. Tunyasuvunakool, K. et al. Highly accurate protein structure prediction for the human proteome. Nature 596, 590–596 (2021).

47. Bork, P. & Beckmann, G. The CUB domain. A widespread module in developmentally regulated proteins. J. Mol. Biol. 231, 539–545 (1993).

48. Thielens, N. M. et al. The N-terminal CUB-epidermal growth factor module pair of human complement protease C1r binds Ca2+-dependent interaction with C1s. J. Biol. Chem. 274, 9149–9159 (1999).

49. Gardner, T. B. Acute Pancreatitis. Ann. Intern Med. 174, Itc17–Itc32 (2021).

50. Mederos, M. A., Reber, H. A. & Girgis, M. D. Acute Pancreatitis: A Review. Jama 325, 382–390 (2020).

51. Hamill, J. M. On the identity of trypsinogen and enterokinase respectively, invertebrates. J. Physiol. 33, 476–478 (1906).

52. Mellany, J. & Woolley, V. J. The ferments of the pancreas Part I. The generation of trypsin from trypsinogen by enterokinase. J. Physiol. 45, 370–388 (1912).

53. Huang, H. et al. Transgenic expression of PRSS1(R122H) sensitizes mice to pancreatitis. Gastroenterology 158, 1072–1082.e1077 (2020).

54. Yuan, X. et al. Structure of murine enterokinase (enteropeptidase) and expression in small intestine during development. Am. J. Physiol. 274, G342–G349 (1998).

55. Huang, X. et al. Amorphous nickel-titanium alloy film: A new choice for cryo electron microscopy sample preparation. Prog. biophysics Mol. Biol. 156, 3–13 (2020).

56. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. J. Struct. Biol. 152, 36–51 (2005).

57. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSMTM: A new choice for cryo electron cryomicroscopy. J. Struct. Biol. 152, 36–51 (2005).

58. Pettersen, E. F. et al. UCSF Chimera-a visualization system for exploratory research and analysis. J. computational Chem. 25, 1605–1612 (2004).

59. Wang, N. et al. Structural basis of human monocarboxylate transporter 1 inhibition by anti-cancer drug candidates. Cell 184, 370–383.e313 (2021).

60. Dang, S. et al. Cryo-EM structures of the TMEM16A calcium-activated chloride channel. Nature 552, 426–429 (2017).

61. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. J. Mol. Biol. 333, 721–745 (2003).

62. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. Nat. methods 11, 63–65 (2014).

63. Sanchez-Garcia, R. et al. DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. Commun. Biol. 4, 874 (2021).

64. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. Acta Crystallogr. Sect. D., Biol. Crystallogr. 60, 2126–2132 (2004).

65. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta. Crystallogr. Sect. D., Biol. Crystallogr. 66, 213–221 (2010).

66. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. Acta. Crystallogr. Sect. D., Biol. Crystallogr. 66, 12–21 (2010).

67. Goddard, T. D. et al. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. Protein Sci. 27, 14–25 (2018).

Acknowledgements
We thank members of the cryo-EM facility of Southern University of Science and Technology for providing facility support. We also thank Hong Wu and Zhangyun Jiang from Shanghai YueXin Life-Science Information Technology Co., Ltd for help with computation. This work was supported by National Natural Science Foundation of China Grants 8202108005 (to Z.S.L.) and 82022008 (to H.J.H.), and a grant from the technology support plan of China 2015BAI13808 (to Z.S.L.).

Author contributions
Conceived and designed the experiments: X.L.Y., Z.Y.D., L.S.P., and H.J.H. Designed constructs for structural studies: X.L.Y., L.S.P. and Q.Y.S. Protein purification: L.S.P., Q.Y.S., F.C., and D.Y.Z. Performed functional analysis: X.L.Y., Z.Y.D., C.C.X., K.L.L., H.Y., S.Y.L., Z.S.L., and H.J.H. Performed EM data collection and analysis: X.L.Y., and Z.Y.D. Structure reconstruction: X.L.Y., and Z.Y.D. Model building: Z.Y.D. SPR analysis: X.L.Y. and L.S.P. Wrote the paper: X.L.Y., Z.Y.D., L.S.P., and H.J.H.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-022-34364-9.

Correspondence and requests for materials should be addressed to Zhaoshen Li or Haojie Huang.

Peer review information Nature Communications thanks Bernhard Loll and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022