Core cysteine residues in the Plasminogen-Apple-Nematode (PAN) domain are critical for HGF/c-MET signaling

Debjani Pal1,3, Kuntal De1,3, Carly M. Shanks1, Kai Feng1, Timothy B. Yates1,2, Jennifer Morrell-Falvey1, Russell B. Davidson1, Jerry M. Parks1 & Wellington Muchero1,2✉

The Plasminogen-Apple-Nematode (PAN) domain, with a core of four to six cysteine residues, is found in > 28,000 proteins across 959 genera. Still, its role in protein function is not fully understood. The PAN domain was initially characterized in numerous proteins, including HGF. Dysregulation of HGF-mediated signaling results in multiple deadly cancers. The binding of HGF to its cell surface receptor, c-MET, triggers all biological impacts. Here, we show that mutating four core cysteine residues in the HGF PAN domain reduces c-MET interaction, subsequent c-MET autophosphorylation, and phosphorylation of its downstream targets, perinuclear localization, cellular internalization of HGF, and its receptor, c-MET, and c-MET ubiquitination. Furthermore, transcriptional activation of HGF/c-MET signaling-related genes involved in cancer progression, invasion, metastasis, and cell survival were impaired. Thus, targeting the PAN domain of HGF may represent a mechanism for selectively regulating the binding and activation of the c-MET pathway.
D

Despite its prevalence in a wide variety of organisms, the role of the Plasminogen-Apple-Nematode (PAN) domain in protein function has largely remained elusive. A key contributing factor is that this domain is found in many phylogenetically unrelated proteins across divergent organisms falling into categories including Alveolata, Archea, Amoebozoa, Bacteria, Cryptophyta, Euglenozoa, Haptophyceae, Opisthokhonta, Rhizaria, Rhodophyta, Stramenopiles, Viridiplantae, and Viruses. The PAN domain was first identified by Tordai et al.1 in which they noted that the domain was shared by the plasminogen/hepatocyte growth factor (HGF) protein family, the prekallikrein/coagulation factor XI protein family, and nematode proteins. The domain possesses the characteristic 4-6-cysteine residues in its core that are strictly conserved. These cysteine residues have been proposed to engage in two or three disulfide bridges to form a hairpin loop structure1–2. The PAN domain has been suggested to mediate protein-protein or carbohydrate-protein interactions and facilitate receptor dimerization3–6. HGF is secreted by mesenchymal cells as a single-chain, biologically inert precursor and is converted into its biologically active form when extracellular proteases cleave the bond between Arg494 and Val4957. The mature form of HGF consists of an α- and β-chain, which are held together by a disulfide bond8–10. The α-subunit of HGF includes N-terminal hairpin loop structure and four kringle domains (K1–K4) whereas the β-subunit consists of serine protease homology (SPH) domain11.

To provide evidence for its role in cellular mesenchymal-epidermal transition (c-MET) interaction and signal transduction, we characterized the functional role of the PAN domain in the heparin binding glycoprotein HGF, which functions as a ligand for the high-affinity receptor, c-MET. HGF/c-MET signaling has been shown to mediate cellular processes including angiogenesis, anti-apoptosis, mitogenesis, morphogenesis, mitogenesis and neurite extension12. Dysregulation of the HGF/c-MET signaling cascade can lead to tumorigenesis by transforming normal cells into tumor cells and c-MET hyperactivation has been reported in cancers including lung cancer, colorectal cancer, glioblastoma, and acute myeloid lymphoma among others13,14.

A crucial step in HGF/c-MET cascade activation is that the binding of HGF to c-MET brings the dimerization of c-MET that enables its intracellular kinase domains to undergo autophosphorylation14. The phosphorylated kinase domain recruits downstream cytosolic effector proteins, leading to the activation of downstream signaling pathways.

The HGF/c-MET pathway has emerged as a prime target for cancer pathways and tumorigenesis15. However, despite the clinical therapeutic significance of this pathway, the mechanism by which HGF activates c-MET is not well understood. HGF is a bivalent ligand with high-affinity binding pockets in the α chain and relatively low-affinity binding pockets in the β chain16. Studies demonstrated that the N- domain and the first kringle domain are enough for c-MET binding; however, for activation of c-MET, the β chain is essential16. Although the affinity of the α chain of HGF and the SEMA domain of MET is structurally well-characterized and functionally validated, it is not clear what regions of the α chain bind with c-MET.

Results

PAN domain-carrying proteins are enriched in cell recognition and cellular signaling processes. PAN domain-carrying proteins do not share a clear evolutionary or phylogenetic trajectory suggesting that this domain may have been co-opted independently by organisms to serve protein functions that are yet to be revealed. To provide evidence of its putative function, we scanned the InterPro (https://www.ebi.ac.uk/interpro/) and UniProt (https://G-LecRK.uniprot.org) databases and found 28,300 proteins across 2496 organisms falling into 959 genera (Supplementary Data 1). Gene Ontology (GO) enrichment analyses using the 28,300 proteins revealed 39 unique GO-terms that were enriched at p < 0.05 (Supplementary Data 2). These included terms such as cell recognition (p-value = 1E−30), cell communication (p-value = 1E−30), proteolysis (p-value = 1E−30), pollen-pistil interaction (p-value = 1E−30), reproduction (p-value = 1E−30), response to stimulus (p-value = 1E−30) and response to stress (p-value = 1E−30). Interestingly, the PAN domain resembles striking similarities with cysteine-rich peptides (CRPs) that have been widely implicated in the regulation of immune responses17–23. Based on a GO enrichment analysis and predicted cellular localization involving 28,000 PAN domain-containing proteins, we found that these proteins were highly enriched in protein proteolysis and immune signaling processes. Almost all of them were associated with the extracellular matrix (ECM) as cell surface receptors or as ligands for cell surface receptors and typically associated with cellular signal transduction, albeit in different pathways for divergent organisms. We therefore hypothesized that the HGF PAN domain would be essential for HGF and c-MET interaction and the downstream signaling. To test this possibility, we designed experiments to mutate core cysteine residues and assess implications on downstream signaling cascades.

Core PAN domain cysteine residues modulate HGF overall-abundance. Alignment of the PAN domain of proteins from 14 model organisms revealed four strictly conserved cysteine residues occurring at amino acid positions 70, 74, 84, and 96 of the HGF protein (Supplementary Figs. 1a, b, 2a–c). To evaluate the functional significance of these residues, we sequentially mutated single cysteines and simultaneously mutated all four cysteine residues and examined HGF stability in 293T cells. Mutant HGFs, in which cysteines 70 (C70A), 74 (C74A), 84 (C84A), and 96 (C96A) were substituted with alanine, led to a marked increase in the protein expression of exogenously expressed HGFs in cells compared to their wild-type (WT) counterpart (Fig. 1a, b). Furthermore, mutating all four core cysteines in the PAN domain dramatically increased the abundance of exogenously expressed HGF-4Cys-4Ala in the cells (Fig. 1c). Moreover, the c-MET receptor exhibited a similar increase in abundance in response to stimulation by the HGF 4Cys-4Ala mutant compared to WT HGF. These results suggest that mutant HGF and its c-MET receptor did not undergo the expected downregulation that was previously reported via ligand-induced internalization following ubiquitination and lysosomal degradation24.

The kringle domains of HGF have been reported to be crucial for protein–protein interactions. In particular, the first and second kringle domains are especially important for the proper biological function of the protein11. Kringle domains are usually disulfide crosslinked domain and previous studies have established that kringle domains in the α-subunit and the SPH domain in the β-subunit provide c-MET binding sites on HGF25,26. Recently, Uchikawa et al., determined structures of c-MET/HGF complex mimicking their active state at 4.8 Å resolution using cryo-electron microscopy27. The study identified multiple distinctive c-MET binding sites on the HGF protein including the N-terminal and kringle domains but did not implicate the PAN domain in the HGF/c-MET interaction26. Here, we exclusively investigated the role of the PAN domain of HGF on the c-MET binding and downstream signaling cascade.

As we established a connection between the core cysteines in the PAN domain of HGF with HGF abundance, we introduced four more mutations of additional cysteines located in the K1, K2, and SPH domains of HGF (C128A, C149A, C271A, and C535A).
None of these mutant HGFs showed increased protein expression when expressed exogenously in cells as did the HGF 4Cys-4Ala (Fig. 1d, e). Thus, cysteine residues in the kringle and SPH domains are apparently not involved in HGF abundance. Rather, induced expression of HGF is specifically defined by the four strictly conserved cysteine residues in PAN domain.

We hypothesized that the enhanced HGF expression of the 4Cys-4Ala mutant may result from an intra-PAN domain structural change that prevents HGF degradation. To test this hypothesis, we performed molecular dynamics (MD) simulations of the WT and 4Cys-4Ala PAN domains using models generated with AlphaFold227 and compared the resulting structural

---

*COMMUNICATIONS BIOLOGY | https://doi.org/10.1038/s42003-022-03582-8 ARTICLE*

---

COMMUNICATIONS BIOLOGY | (2022) 5:646 | https://doi.org/10.1038/s42003-022-03582-8 | www.nature.com/commsbio
Fig. 1 PAN domain modulates HGF stability. a, b Core cysteines in PAN domain are crucial for HGF stability. a Immunoblot analysis of whole cell lysates derived from 293T cells, transfected with Flag-HGF WT and different single cysteine mutants of Flag-HGF constructs as indicated. 30 h post-transfection, whole-cell lysates were prepared for immunoblot analysis. Representative image of n = 3 biological replicates. b Quantification of the band intensities in (a). The intensities of Flag-HGF (WT and mutants) bands were normalized to actin and then normalized to Flag-HGF WT. Data are represented as mean ± SEM, as indicated. 30 h post-transfection, whole-cell lysates were prepared for immunoblot analysis. Representative image of n = 3 biological replicates. c Mutation in all four core cysteine residues in PAN domain remarkably alters HGF stability. Immunoblot analysis of whole cell lysates derived from 293T cells, transfected with Flag-HGF WT and Flag-HGF 4Cys-4Ala constructs as indicated. 30 h post-transfection, whole-cell lysates were prepared for immunoblot analysis. d Cysteines in kringle domains and SPH domain of HGF have no impact on the protein abundance in cells. Immunoblot analysis of whole cell lysates derived from 293T cells, transfected with Flag-HGF WT, Flag-HGF 4Cys-4Ala, and different single cysteine mutants of Flag-HGF constructs as indicated. 30 h post-transfection, whole-cell lysates were prepared for immunoblot analysis. Representative image of n = 3 biological replicates. e Quantification of the band intensities in (d). The intensities of Flag-HGF (WT and mutants) bands were normalized to actin and then normalized to Flag-HGF WT. Data are represented as mean ± SEM, n = 3, and \( p < 0.05, ** p < 0.005, *** p < 0.0005 \) (student’s t-test). f Localization of Flag-HGF WT and different Flag-HGF mutants’ expression by confocal immunofluorescence microscopy in HeLa cells. The cells were transiently transfected with Flag-HGF WT and different mutants of Flag-HGFs as indicated. 30 h post-transfection cells were fixed, mounted and protein expression patterns were visualized using a Zeiss LSM 710 confocal microscope outfitted with a 63x objective. Scale bars represent 20 μm. The images shown are representative from three independent biological experiments (average 100 cells were observed per experimental condition per replicate). g Percentage of transfected HeLa cells showing perinuclear/nuclear staining for Flag-HGF WT and Flag-HGF 4Cys-4Ala were quantified. Data are represented as mean ± SD, n = 3 (average 100 cells were observed for each condition per experiment), and \( p < 0.05, ** p < 0.005, *** p < 0.0005 \) (Student’s t test).

enssembles. Five models were generated for each system and used as starting structures for the simulations. A 200 ns simulation was then performed for each model. Thus, the cumulative simulation time was 1 μs for each system. Analysis of the time evolution of the root-mean-square deviation (RMSD) of each system revealed only minor structural deviations in both sets of simulations. Using the top WT model as a reference structure, the average RMSDs and standard errors of the means were 1.61 ± 0.02 Å for the WT and 1.71 ± 0.01 Å for the 4Cys-4Ala system, indicating highly similar overall structures despite the four cysteine substitutions in the mutant (Supplementary Fig. 3a, b). The structural analysis indicates that no large-scale structural changes occurred in the 4Cys-4Ala mutant PAN domain compared to the WT on this time scale (Supplementary Fig. 4).

Upon binding with its high-affinity protooncogenic receptor, c-MET, HGF activates a wide range of cellular signaling pathways. Autophosphorylation in the carboxy-terminal tail of c-MET creates a binding sites for more downstream adaptors and effectors that lead to the activation of the Ras pathway (RAS)/extracellular signal-regulated kinase (ERK) and phosphatidylinositol 3-kinase (PI3K)/protein kinase B (AKT) axis via phosphorylation. To determine the influence of core cysteines on c-MET phosphorylation, 293T, glioblastoma U-87 MG, and HeLa cells were stimulated with purified wild-type HGF and HGF 4Cys-4Ala proteins. Western blot analyses revealed the absence of phosphorylated c-MET, AKT, and ERK in unstimulated controls and in cells stimulated with the HGF 4Cys-4Ala protein in both cell types (Fig. 2a and Supplementary Fig. 5a, b). Additionally, purified proteins of HGF C70A, HGF C74A, HGF C84A, and HGF C96A mutants failed to promote downstream c-MET phosphorylation following stimulation as evident in Supplementary Fig. 6. This result suggests that each cysteine residues in the PAN HGF domain are critical for HGF/c-MET signaling in a wide variety of cells and thus indicates the potential for designing novel HGF specific inhibitor targeting dysregulated c-MET signaling cascade in pathological conditions.

To characterize the incompetence of PAN mutant HGF in turning on the phosphorylation cascade of c-MET, we performed a direct interaction assay between endogenous c-MET and in vitro translated HGF proteins and showed that HGF 4Cys-4Ala was unable to interact with c-MET (Fig. 2b, c). We further got interested in whether there was a complete lack of interaction between c-MET and PAN mutant HGF or whether it could not maintain stable interaction. We employed both in vitro and in vivo cross-linking approaches to demonstrate the interaction stability. Purified HGF WT protein and HGF 4Cys-4Ala protein were activated by incubation with biotin-containing trifunctional cross-linking reagent Sulfo-SBED (Thermo). This reagent has amine group-specific reactivity, a nonspecific photo reactivity, and a biotin as a reactive handle. The cross-linking can be reversed by thiol cleavage. Upon reduction of the disulfide bond, the biotin label is transferred to the interacting protein. Non-reacted crosslinker was removed by overnight dialysis. Cross-linking was induced by UV irradiation following incubation with purified c-MET. Biotin transfer from PAN mutant HGF to c-MET was reduced significantly as compared to HGF WT (Supplementary Fig. 7a, b). The result suggests that PAN mutant HGF makes direct contact with c-MET but is unable to maintain stable interaction which attenuates downstream signaling.

PAN domain cysteine residues are essential for c-MET, AKT, and ERK phosphorylation. Following binding of HGF at the MET semaphoring homology (SEMA) domain, c-MET homodimerizes and autophosphorylates at two tyrosine residues (Y1234 and Y1235), followed by subsequent phosphorylation of two additional tyrosines in the carboxy-terminal tail (Y1349 and Y1356). This series of events creates a multifunctional docking site for downstream adaptors and effectors that lead to the activation of the Rat sarcoma (RAS)/extracellular signal-regulated kinase (ERK) and phosphatidylinositol 3-kinase (PI3K)/protein kinase B (AKT) axis via phosphorylation. To determine the influence of core cysteines on c-MET phosphorylation, 293T, glioblastoma U-87 MG, and HeLa cells were stimulated with purified wild-type HGF and HGF 4Cys-4Ala proteins. Western blot analyses revealed the absence of phosphorylated c-MET, AKT, and ERK in unstimulated controls and in cells stimulated with the HGF 4Cys-4Ala protein in both cell types (Fig. 2a and Supplementary Fig. 5a, b). Additionally, purified proteins of HGF C70A, HGF C74A, HGF C84A, and HGF C96A mutants failed to promote downstream c-MET phosphorylation following stimulation as evident in Supplementary Fig. 6. This result suggests that each cysteine residues in the PAN HGF domain are critical for HGF/c-MET signaling in a wide variety of cells and thus indicates the potential for designing novel HGF specific inhibitor targeting dysregulated c-MET signaling cascade in pathological conditions.

To characterize the incompetence of PAN mutant HGF in turning on the phosphorylation cascade of c-MET, we performed a direct interaction assay between endogenous c-MET and in vitro translated HGF proteins and showed that HGF 4Cys-4Ala was unable to interact with c-MET (Fig. 2b, c). We further got interested in whether there was a complete lack of interaction between c-MET and PAN mutant HGF or whether it could not maintain stable interaction. We employed both in vitro and in vivo cross-linking approaches to demonstrate the interaction stability. Purified HGF WT protein and HGF 4Cys-4Ala protein were activated by incubation with biotin-containing trifunctional cross-linking reagent Sulfo-SBED (Thermo). This reagent has amine group-specific reactivity, a nonspecific photo reactivity, and a biotin as a reactive handle. The cross-linking can be reversed by thiol cleavage. Upon reduction of the disulfide bond, the biotin label is transferred to the interacting protein. Non-reacted crosslinker was removed by overnight dialysis. Cross-linking was induced by UV irradiation following incubation with purified c-MET. Biotin transfer from PAN mutant HGF to c-MET was reduced significantly as compared to HGF WT (Supplementary Fig. 7a, b). The result suggests that PAN mutant HGF makes direct contact with c-MET but is unable to maintain stable interaction which attenuates downstream signaling.
As c-MET is an integral membrane protein, we used a whole-cell lysate crosslinking assay to demonstrate the significance of the PAN domain on c-MET interaction. We transfected HEK293T cells transfected with a vector expressing GFP-c-MET as target cells. The activation of HGF WT protein and HGF 4Cys-4Ala protein was done similarly as described above.

Activated purified biotinylated HGF WT protein and HGF 4Cys-4Ala protein were incubated with GFP-c-MET expressing cell lysate in the dark for 45 min, and crosslinking was induced by UV irradiation. After crosslinking, biotin-containing complexes were immunoprecipitated on streptavidin beads and analyzed by Western blotting for the presence of GFP-c-MET. As shown in...
Supplementary Fig. 7c, PAN mutant HGF show substantially reduced interaction with c-MET, which bolstered our conclusion that the PAN domain of HGF plays a crucial role in maintaining stable interaction.

Our results thus far indicate a central role for the PAN domain in its initial recognition by c-MET. In addition, HGF 4Cy4-Ala, like the unstimulated controls, failed to promote c-MET perinuclear translocation compared to wild-type HGF based on stimulation assays using HeLa cells (Fig. 2d).

As our cross-linking experiment suggested a significantly reduced interaction of c-MET and HGF PAN mutant compared to its wild-type counterpart, we were interested in checking if that reduced interaction was sufficient to turn ON the ubiquitination status of c-MET. We performed an in-vivo ubiquitination assay with c-MET. Ubiquitination was inhibited with the alanine mutants of HGF compared to wild type (Fig. 2e). Based on these observations, we propose that due to the lack of activation, c-MET receptors are unable to become internalized by endocytosis, suggesting that the mutated cysteine residues impose an overall retardation of c-MET activity and its endocytic trafficking.

Mutating core cysteine residues suppresses STAT3 phosphorylation and nuclear translocation. Signal transducer and activator of transcription 3 (STAT3) is a transcription factor that is present in the cytoplasm, forms dimers upon activation, and functions as a downstream effector molecule of the HGF/c-MET signaling pathway. STAT3 is reported to be constitutively active in several cancers, which leads to malignant transformation by playing a critical role in stimulating cell proliferation and arresting apoptosis. We tested the impact of the core cysteines in the HGF PAN domain on STAT3 activation by following its phosphorylation in glioblastoma U-87 MG and HeLa cells. STAT3 phosphorylation was notably reduced in cells post-stimulation with the HGF 4Cy4-Ala mutant compared to the wild type (Fig. 3a and Supplementary Fig. 8). Delayed time points were chosen for this assay because HGF induces delayed STAT3 phosphorylation. Confocal imaging confirmed that HGF 4Cy4-Ala, like the unstimulated controls, was unable to promote STAT3 nuclear translocation, while wild-type HGF promoted normal STAT3 nuclear localization (Fig. 3b). These results are consistent with the above observations, suggesting that mutating the cysteine residues in the PAN domain leads to disrupted HGF/c-MET signaling.

PAN mutations downregulate HGF/c-MET-dependent cell proliferation and alters the expression of genes essential for diverse cellular responses. Dysregulated expression of HGF/c-MET acts as a catalyst in many cancers, with overexpression of HGF often leading to aberrant cell proliferation and extracellular matrix invasion. Direct evidence has been established connecting a primary role for HGF with increased expression of matrix metalloproteinase-9 (MMP9), which is crucial for angiogenesis. HGF facilitates MMP9 expression via the PI3K/AKT and p38 mitogen-activated protein kinases (MAPK) axis. Given the reported role of HGF in modulating expression of MMP9, we evaluated the impact of mutated cysteine residues on its transcriptional response. To evaluate the role of PAN domain mutation on the pattern of the signature gene expression as well as cell proliferation, both 293T and U-87 MG cells were stimulated with wild-type HGF or 4Cy4-Ala HGF for 24 h following serum starvation. Cell viabilities, as assessed by the MTT assay, were notably reduced by treatment with HGF 4Cy4-Ala, whereas wild-type HGF could overcome serum starvation-mediated growth inhibition in both cell types (Supplementary Fig. 9a). 293T cells transiently transfected with Flag-WT HGF showed a significant increase in cell proliferation over time compared to cells transfected with Flag-HGF 4Cy4-Ala (Supplementary Fig. 10a). Quantitative real-time PCR (qRT-PCR) was performed to determine whether MMP9 expression was similarly impacted. mRNA levels of MMP9 were markedly decreased in HGF 4Cy4-Ala-stimulated 293T and U-87 MG cells compared to the wild type in both cell types (left panel, Supplementary Fig. 9b). Similarly, we observed a significant decline in c-MET mRNA expression in both cell types stimulated with HGF 4Cy4-Ala (Right panel, Supplementary Fig. 9b). Increased c-MET expression might be a crucial determinant for the overall balance of the HGF/c-MET cascade in cells and could be a trigger for the malignant transformation of normal cells. These data suggest that induction of the MMP9 receptor is correlated with overall c-MET expression in an HGF-dependent manner and could be regulated by minimally mutating the four core cysteines in the PAN domain of HGF.

Based on the apparent PAN domain-dependent transcriptional modulation of MMP9 expression by HGF, we characterized the transcriptional responses of additional downstream genes in the HGF/c-MET signaling pathway. Total RNA was extracted from 293T cells following a 24 h post treatment with wild-type HGF or 4Cy4-Ala HGF. Based on this analysis, we identified significant differences in the expression of genes previously implicated in c-MET signaling, cell cycle regulation, and cancer-related processes (Fig. 4; Supplementary Fig. 10b). Specifically, we observed a striking reduction in the expression levels of the downstream targets of the c-MET signaling cascade, ETS translocation variants 1, 4, and 5 (ETV1, ETV4, ETV5), in cells stimulated with HGF 4Cy4-Ala compared to wild type HGF.

ARTICLE COMMUNICATIONS BIOLOGY | https://doi.org/10.1038/s42003-022-03582-8
These transcription factors are members of the polyoma enhancer activator 3 (PEA3) subgroup of the E-twenty six (ETS) family and confer resistance to early growth factor response (EGFR) targeted therapy in lung cancer. Some known target genes of PEA3 are matrix metalloproteinase-2 (MMP2), matrix metalloproteinase-7 (MMP7) and MMP9, which are well recognized for their role in the invasiveness of cancer. Mutating the core cysteine residues in the PAN domain led to reduced expression of early growth response 1 (EGR1), as well as the metalloproteinasas, A disintegrin and metalloprotease domain-containing protein 9 and 10 (ADAM9 and ADAM10), all of which were previously linked to increased adhesion and cancer progression including hepatocellular carcinoma, and triple negative breast cancer via the AKT/NF-κB axis. Hyperactivation of focal adhesion kinase (FAK) and PI3K/receptor for activated C kinase 1 (RAC1) are responsible for metastasis progression and mutations of the core cysteines in the HGF PAN domain also resulted in their reduced expression. Apart from these examples, relative expression of proteins involved in DNA damage repair (double-strand-break repair protein or RAD21), chromosomal integrity and cohesion (structural maintenance of chromosomes protein 1 A or SMC1A), cell cycle progression (kinesin family member 2 A or KIF2A), posttranslational modification, and RNA splicing (protein arginine N-methyltransferase 5 or PRTM5) were also reduced. (Fig.4)

Furthermore, this analysis revealed well-known cancer biomarkers whose expression levels were affected by mutating the core cysteine residues. These biomarkers included calponin-3 or CNN3 (marker for colorectal cancer), dolichyl-diphosphooligosaccharide-protein glycosyltransferase subunit 2 or RPN2 (which inhibits autophagy and upregulates MMP9 expression), branched chain amino acid transaminase 1 or BCAT1 (which promotes hepatocellular carcinoma and chemoresistance), ski interacting protein or SKIP, epithelial cell transforming sequence 2 oncogene or ECT2 (overexpressed in different cancers including non-small cell lung cancer), and elastin microfibril interfacer 2 or EMILIN2 (which promotes angiogenesis and inflammation).
Discussion

c-MET receptor, which is a receptor tyrosine kinase (RTK) family, plays essential roles(s) in number of cellular processes which includes cell proliferation, survival, morphogenesis, and motility. Abnormal activation of c-MET leads to the progression of tumor growth and metastatic cancer cells, which makes it an important for drug target for cancer treatments. We report here experimental evidence that strongly supports the role of the PAN domain of HGF in providing the functional catalytic core and plays a crucial role in c-MET binding. Although our MD simulation data suggested that there is no change in overall structure of HGF because of PAN mutation, stable cross-linking of c-MET and HGF occurs only in the presence of intact PAN domain in HGF. Additionally, we demonstrated that mutating core cysteine residues in the HGF PAN domain results in a cascade of negative regulation of HGF/c-MET signaling, which starts with attenuated HGF degradation, impaired c-MET interaction, disrupted perinuclear localization, and is followed by a subsequent lack of

Fig. 4 Transcriptome analysis post HGF stimulation in 293T cells. Differential expression analysis by RNA seq in 293T cells following HGF WT and HGF 4Cys-4Ala treatment as indicated confirms that core cysteines in HGF PAN domain are necessary for the expression of a wide range of genes. Responsive genes were normalized to FPKM value for non-treated cells and then normalized to HGF WT treated cells. Data represents the average of three independent biological replicates and *p < 0.05, **p < 0.005 and ***p < 0.0005 were calculated with a student’s t-test.

phosphorylation for c-MET and its downstream targets AKT, ERK and STAT3 and c-MET ubiquitination.

This disruption of upstream events was confirmed by the lack of transcriptional activation of markers genes associated with cell migration and invasion including ADAM9/10, EGR1, ETV1, ETV4, ETV5, and MMP9 (Fig. S10B). Targeting the PAN domain in HGF could act as a multi-pathway target as it entirely shuts down the c-MET signaling cascade and downstream expression of relevant proteins known for their role in cancer prognosis and other diseases (Fig. S5). Taken together, the PAN domain and its core cysteine residues are essential for HGF/c-MET signaling in human cells. The observed enhanced stability of HGF with mutated cysteine residues is due to the lack of its stable interaction with c-MET. Considering the results described in this study, it would be essential to determine the molecular mechanism(s) by which the PAN domain regulates the c-MET interaction, which requires additional studies. This report identifies the HGF PAN domain as a potential site for c-MET binding and MET signaling cascade activation. Targeting the PAN domain can block cell proliferation by inhibiting MAPK phosphorylation and altering the expression of different cancer biomarkers. These observations point to the exciting prospect of treating HGF-overexpressing tumors by selectively targeting the PAN domain.

Materials and methods

PAN domain sequence alignment. Proteins with PAN domains were identified from Uniprot and were selected from 14 model organisms according to Bateman et al. The PAN domain coordinates from Uniprot were used to extract the corresponding sequences from the full-length proteins, which were then aligned with MAFFT linsi. The alignment was visualized with Geneious and a phylogenetic tree accompanying the alignment was constructed via the neighbor-joining method with Geneious (Supplementary Data 3).

GO enrichment of 14 PAN domain categories. GO terms of each category were extracted from InterPro2GO database from InterPro. GO enrichment analysis for each category against all PAN domain genes was performed by Fisher’s exact test via the TopGO package. Only the biological process GO category was used for GO enrichment.

Mammalian cell culture, transfection, and drug treatment. HeLa, HEK 293T, and Glioblastoma U87 cells were obtained from ATCC and maintained in a humidified atmosphere at 5% CO2 in Dulbecco’s Modified Eagle’s (DMEM) complete medium (Corning) supplemented with 10% fetal bovine serum (FBS; Seradigm) in 37 °C. Plasmid transfections were done with TransIT-LT1 (Mirus Bio) per the manufacturer’s instructions.

Plasmids and recombinant proteins. Flag-HGF (GeneBankTM accession number NM_000601.5) and various mutants of Flag-HGF, cloned into pcDNA 3.1 were obtained from GenScript. Purified Flag tagged HGF wild type, HGF C70A, HGF C74A, HGF C84A, HGF C96A, and HGF 4Cys-4Ala proteins were customized and synthesized using HD CHO-S cell expression (GenScript) (Supplementary Fig. 11). Briefly, Target DNA sequence was designed, optimized and synthesized, and subcloned into pcDNA3.4 vector and transfected into CHO-S cells. Following expression, Cell culture broth was centrifuged and followed by filtration. Filtered cell culture supematant was loaded onto an affinity purification column (Anti-DYKDDDK G1 Affinity Resin 3 ml) at an appropriate flowrate (1 ml/min). After washing and elution with appropriate buffers (Washing buffer: PBS, pH7.2; Elution buffer: 0.1 M Gly-HCl, pH 3.5), the eluted fractions were pooled, and buffer exchanged to the final formulation buffer (PBS, pH7.2). The purified protein was analyzed by SDS-PAGE, Western-blot, HPLC analysis to determine the molecular weight and purity. c-MET-C-GFPSpark® Clone was obtained from
For c-MET and STAT3 localization assays, cells were serum-starved for 24 h before the
primary antibodies in 5% BSA in 1X-PBS with 0.5% Triton X-100
were incubated with appropriate
were incubated with primary antibodies in 5% BSA in PBST for 30 min at room temperature. DNA was
incubated in 4% paraformaldehyde. Next, the cells were permeabilized with 0.5% Triton X-100 in PBS, and washed three times with the lysis buffer and eluted with SDS-gel loading buffer (with reducing agent added). Proteins were resolved by SDS-PAGE and immuno
tohibled with the indicated antibodies.

Cell proliferation assay. 3-(4, 5-dimethylthiazol-2-yl)-2, 5-diphenyl tetrazolium
bromide (MTT) assay was used to determine cell viability. HEK293T and U-87
MG cells were plated in 96-well plates with 500 cells per well in triplicate in serum-
free medium for 24 h prior to HGF stimulation. 100 ng mL
HGF WT, HGF 4Cys-4Ala, or both were added to the cells and incubated for 24 h prior to the
to the manufacturer’s instruction. The assays were performed in tripli
cate experiments were performed three times. Data were expressed as the
mean ± SD. Statistical analyses were performed. *P < 0.05 was considered to indi
cate a statistically significant difference.

Quantitative real-time PCR and RNAseq analysis. Total RNA was extracted from
gliblastoma U-87 MG and HEK293T cell line 24 h post HGF stimulation using TRIzol reagent (Invitrogen), and reverse transcription was performed using the
to the manufacturer’s instructions. The c-MET and MMP9 mRNA
expression levels were determined using 2−ΔΔCT. All the primer sequences are listed in
thermoe
erase, Recombinant (Invitrogen, no. 10342-020). Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was used as the internal control. The specific primers for c-MET, MMP9 and GAPDH were designed with Primer Premier software. The primers were used: c-MET, forward: 5′-TTAAAGGAGCTTACCATGTAATC-
and reverse: 5′-CTTGTGGAGTAAATTACGAGCACTTTGCA-3′; MMP9, forward: 5′-
and reverse: 5′-GGCCACCTTGATTGGTTGAA-3′. The PCR reaction was performed
according to the manufacturer’s instructions. The PCR conditions were as follows: amplification protocol was performed for 35 cycles consisting of 30 s at
94°C (denaturation), annealing 30 s at 45°C and extension 30 s at 72°C.

For RNA-seq analysis, Total RNA was extracted from the HEK293T cell line 24 h post HGF stimulation using TRIzol reagent (Invitrogen) according to
the manufacturer’s instructions. Quantification and quality control of isolated RNA
was performed by measuring absorbance at 260 and 280 nm on a NANO DROP ONEC spectrophotometer (Thermo Scientific, USA). The RNA-seq run was performed with four biological replicates. Library prep and sequencing was performed by BGI using the DNaseq-G400 platform which generated 100 bp
paired-end reads. The raw RNA-seq reads have been deposited at NCBI under BioProject ID PRJNA18997. RNA-seq data were aligned to
human reference genome GRCh38. Reads were mapped with bowtie2 v2.2.5. Expression levels for RNAs were calculated using fragments per kilobase per million reads (FPKM) values with RSEM v1.2.860. Differential expression analysis was performed with DESeq2 and genes with an adjusted
p-value less than 0.05 were considered differentially expressed.

Molecular dynamics simulations. MD simulations were initiated from the top five
generated with ColbertFold27,28 of the WT and 4Cys-4Ala mutant
were performed using AmberTools20. At least a 12 Å solvent
buffer between the protein and the periodic images. Sodium and chloride ions were added to neutralize charge and maintain a 0.15 M concentration. The simulations were performed with OpenMM version 7.5.139 on the CUDA platform (version 11.0.3) using Python 3.8.0. ParmEd was used to incorporate the force field parameters into the OpenMM platform57. The Langevin integrator and Monte Carlo barostat were used to maintain the systems at 300 K and 1 bar, respectively. Direct non-bonded interactions were calculated up to a 12 Å distance cutoff. All bonds involving hydrogen atoms were constrained to their equilibrium values. The particle mesh Ewald method was used to compute long-range Coulombic interactions58. A 2 fs integration time step was used with energies and positions written every 2 ps.

Simulation analysis. Analyses of MD trajectories were performed using Python
and the MDAnalysis version 1.0.19670. Matplotlib was used to plot the data.
1. Tordai, H., Bánai, L. & Pathy, L. The PAN module: The N-terminal domains of plasminogen and hepatocyte growth factor are homologous with the apple domains of the prekallikrein family and with a novel domain found in numerous nematode proteins. FEBS Lett. 461, 63–67 (1999).

2. Zhou, H. et al. The solution structure of the N-terminal domain of hepatocyte growth factor reveals a potential heparin-binding site. Structure 6, 109–116 (1998).

3. Herwald, H. et al. Mapping of the discontinuous kinogen binding site of prekallikrein: A distal binding segment is located in the heavy chain domain AA1. J Biol Chem. 271, 13061–13067 (1996).

4. Ho, D. H., Badellino, K., Baglia, F. A. & Walsh, P. N. A binding site for heparin in the apple 3 domain of factor XI. J Biol Chem. 273, 16382–16390 (1998).

5. Baglia, F. A., Badellino, K. O., Ho, D. H., Dasari, V. R. & Walsh, P. N. A binding site for heparin in the apple 2 domain of prothrombin in the apple 1 domain of factor XI. J Biol Chem. 275, 31954–31962 (2000).

6. Naithani, S., Chookajorn, T., Ripoll, D. R. & Nasrallah, J. B. Structural and functional analysis of the S-locus receptor kinase extracellular domains for receptor dimerization in the S-locus receptor kinase extracellular domain. Proc Natl Acad Sci USA 104, 12211–12216 (2007).

7. Organ, S. L. & Tsao, M. S. An overview of the c-MET signaling pathway. Ther Adv Med Oncol. 3, S7–S19 (2011).

8. Gohda, E. et al. Purification and partial characterization of hepatocyte growth factor from plasma of a patient with fulminant hepatic failure. J Clin Invest. 81, 414–419 (1988).

9. Nakamura, T. et al. Molecular cloning and expression of human hepatocyte growth factor. Nature 342, 440–443 (1989).

10. Zarnegar, R. & Michalopoulos, G. Purification and biological characterization of human hepatopoietin A, a polypeptide growth factor for hepatocytes. Cancer Res. 49, 3314–3320 (1989).

11. Mizuno, K. et al. Hairpin loop and second kringle domain are essential sites for heparin binding and biological activity of hepatocyte growth factor. J Biol Chem. 269, 1131–1136 (1994).

12. Funakoshi, H. & Nakamura, T. Hepatocyte growth factor: From diagnosis to clinical applications. Clin Chim Acta 327, 1–23 (2003).

13. Jeffers, M. et al. Activating mutations for the Met tyrosine kinase receptor in gastric tumors. Mol Cell 13, 791–805 (2001).

14. Lee, B.-S., Park, M., Cha, H.-Y. & Lee, J.-H. Hepatocyte growth factor induces delayed STAT3 phosphorylation through interleukin-6 expression. Cell Signal. 21, 419–427 (2009).

15. Koh, S. & Lee, K. H. HGF mediated upregulation of lipocalin 2 regulates MMp9 through nuclear factor-kB activation. Oncol Rep. 34, 2179–2187 (2015).

16. Qi, T. et al. Function and regulation of the PEA3 subfamily of ETS transcription factors in cancer. Am J Cancer Res. 10, 3083 (2020).

17. Kherrouche, Z. et al. PEA3 transcription factors are downstream effectors of Met signaling involved in migration and invasiveness of Met-addicted tumor cells. Mol Oncol. 9, 1852–1867 (2015).

18. Fischer, O., Hart, S., Gehwind, A. & Ulrich, A. EGFR signal transactivation in cancer cells. Biochem. Soc. Trans. 31, 1203–1208 (2003).

19. Duffy, M. J. et al. The ADAMs family of proteases: New biomarkers and therapeutic targets for cancer. Clin Proteom. 10, 213 (2011).

20. Zhou, R. et al. ADAM9 mediates triple-negative breast cancer progression via AKT/NF-kB pathway. Front Med. 7, 214 (2020).

21. Beauchene, N. A. et al. Rad21 is required for centromere integrity in human cells independently of its role in chromosome cohesion. Cell Cycle 9, 1774–1780 (2010).

22. Pan, X. W. et al. SMG1 promotes growth and migration of prostate cancer in vitro and in vivo. Int J Oncol. 49, 1963–1972 (2016).
46. Miyamoto, T. et al. The microtubule-depolymerizing activity of a mitotic kinesin protein KIF2A drives primary cilia disassembly coupled with cell proliferation. Cell Rep. 10, 664–673 (2015).

47. Andreu-Pérez, P. et al. Protein arginine methyltransferase 5 regulates ERK1/2 signal transduction amplitude and cell fate through CRAF. Sci. Signal. 4, ra38 (2011).

48. Nakarai, C. et al. Expression of AKR1C3 and CNN3 as markers for detection of lymph node metastases in colorectal cancer. Clin. Exp. Med. 15, 333–341 (2015).

49. Huang, L. et al. RPN2 promotes metastasis of hepatocellular carcinoma cell and inhibits autophagy via STAT3 and NF-κB pathways. Aging. 11, 6674–6690 (2019).

50. Zheng, Y. H. et al. BCAT 1, a key prognostic predictor of hepatocellular carcinoma, promotes cell proliferation and induces chemotherapy resistance to cisplatin. Liver Int. 36, 1836–1847 (2016).

51. Justilien, V. et al. Ect2-dependent rRNA synthesis is required for KRAS-TRP53-driven lung adenocarcinoma. Cancer Cell 31, 256–269 (2017).

52. Liu, Y. Q. et al. Skp1 in lung cancer: Clinical significance and therapeutic efficacy of its small molecule inhibitors. Oncotarget. 6, 34953–34967 (2015).

53. Andreatuzzi, E. et al. Deregulated expression of Elastin Microfibril Interface 2 (EMILIN2) in gastric cancer affects tumor growth and angiogenesis. Matrix Biol. 6, 100029 (2020).

54. Mo, H.-N. & Liu, P. Targeting MET in cancer therapy. J. Mol. Biol. Evol. 30, 772–780 (2013).

55. Mitchell, A. L. et al. InterPro in 2019: Improving coverage, classification, and accuracy of protein sequence annotations. Nucleic Acids Res. 47, D351–D360 (2019).

56. Alexa, A. & Rahnenfuehrer, J. toppGO: enrichment analysis for gene ontology. R Package Version. 2, 2010 (2010).

57. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357 (2012).

58. Li, B. & Dewey, C. N. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinform. 12, 323 (2011).

59. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and the false discovery rate. Genome Biol. 15, 550 (2014).

60. Mirdita, M. et al. ColabFold: making protein folding accessible to all. Nat Methods 19, 679–682 (2022).

61. Case, D. A. et al. AMBER 2020. (University of California, San Francisco, 2020).

62. Maier, J. A. et al. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. J. Chem. Theory Comput. 11, 3696–3713 (2015).

63. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impy, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 79, 926–935 (1983).

64. Eastman, P. et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. PLoS Comput. Biol. 13, e1005659 (2017).

65. Tian, C. et al. ff19SB: Amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. J. Chem. Theory Comput. 16, 528–532 (2020).

66. Cheatham, T. E. III, Miller, J. I., Fox, T., Darden, T. A. & Kollman, P. A. Molecular dynamics simulations on solvated biomolecular systems: The particle mesh Ewald method leads to stable trajectories of DNA, RNA, and proteins. J. Am. Chem. Soc. 117, 4193–4194 (1995).

67. Michaud-Agrawal, N., Deming, E. J., Wootf, T. B. & Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. J. Comput. Chem. 32, 2319–2327 (2011).

68. Nakarai, C. et al. Expression of AKR1C3 and CNN3 as markers for detection of lymph node metastases in colorectal cancer. Clin. Exp. Med. 15, 333–341 (2015).

69. Huang, L. et al. RPN2 promotes metastasis of hepatocellular carcinoma cell and inhibits autophagy via STAT3 and NF-κB pathways. Aging. 11, 6674–6690 (2019).

70. Gowers, R. J. et al. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. 98–105 (United States, N. p., 2019).