







ARTICLE

<https://doi.org/10.1038/s41467-019-13577-5>

OPEN

# Solution structure of human myeloid-derived growth factor suggests a conserved function in the endoplasmic reticulum

Valeriu Bortnov <sup>1</sup>, Marco Tonelli<sup>2,3</sup>, Woonghee Lee <sup>2,3</sup>, Ziqing Lin <sup>4,5</sup>, Douglas S. Annis <sup>1</sup>, Omar N. Demerdash<sup>6</sup>, Alex Bateman<sup>7</sup>, Julie C. Mitchell <sup>6</sup>, Ying Ge <sup>4,5</sup>, John L. Markley<sup>2,3</sup> & Deane F. Mosher<sup>1,8\*</sup>

Human myeloid-derived growth factor (hMYDGF) is a 142-residue protein with a C-terminal endoplasmic reticulum (ER) retention sequence (ERS). Extracellular MYDGF mediates cardiac repair in mice after anoxic injury. Although homologs of hMYDGF are found in eukaryotes as distant as protozoans, its structure and function are unknown. Here we present the NMR solution structure of hMYDGF, which consists of a short  $\alpha$ -helix and ten  $\beta$ -strands distributed in three  $\beta$ -sheets. Conserved residues map to the unstructured ERS, loops on the face opposite the ERS, and the surface of a cavity underneath the conserved loops. The only protein or portion of a protein known to have a similar fold is the base domain of VNN1. We suggest, in analogy to the tethering of the VNN1 nitrilase domain to the plasma membrane via its base domain, that MYDGF complexed to the KDEL receptor binds cargo via its conserved residues for transport to the ER.

<sup>1</sup>Department of Biomolecular Chemistry, University of Wisconsin-Madison, Madison, WI 53706, USA. <sup>2</sup>Department of Biochemistry, University of Wisconsin-Madison, Madison, WI 53706, USA. <sup>3</sup>National Magnetic Resonance Facility at Madison, University of Wisconsin-Madison, Madison, WI 53706, USA. <sup>4</sup>Departments of Cell and Regenerative Biology and Chemistry, University of Wisconsin-Madison, Madison, WI 53706, USA. <sup>5</sup>Human Proteomics Program, University of Wisconsin-Madison, Madison, WI 53706, USA. <sup>6</sup>Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA. <sup>7</sup>European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton CB10 1SD, UK. <sup>8</sup>Department of Medicine, University of Wisconsin-Madison, Madison, WI 53706, USA. \*email: [dfmosher@wisc.edu](mailto:dfmosher@wisc.edu)

**H**uman myeloid-derived growth factor (hMYDGF) is a member of the widely distributed MYDGF family of proteins found in organisms as distant as protozoans<sup>1</sup>. In humans, MYDGF is abundant in nearly 200 different tissues, fluids, and cell lines as compiled by Proteomics DB, a repository of quantitative proteomics data<sup>2</sup>. hMYDGF comprises a 31-residue signal sequence followed by a 142-residue mature protein ending in C-terminal residues RTEL. Experiments appending candidate endoplasmic reticulum (ER) retention sequences (ERS) to the C-terminus of a model secreted protein for expression in HeLa cells demonstrated that RTEL binds to human KDEL receptors KDELRL1, KDELRL2, and (to a lesser extent) KDELRL3 (ref.<sup>3</sup>). KDELRLs retain target proteins in the ER by engaging a C-terminal ERS at the slightly acidic pH of the Golgi (6.0–6.7) and dissociating under the more neutral pH conditions of the ER (7.2) after retrograde transport<sup>4–6</sup>. Removal of the C-terminal Glu-Leu residues from hMYDGF exogenously expressed in HEK293 cells demonstrated that absence or presence of an intact ERS determines, in a nearly all-or-none fashion, whether hMYDGF is retained in ER or secreted<sup>1</sup>.

The name “myeloid-derived growth factor” was assigned when MYDGF was identified as a protein secreted from monocytes/macrophages that promotes tissue repair in a murine model of myocardial infarction<sup>7</sup>. MYDGF-knockout mice, which lacked a developmental phenotype, developed larger infarct scars and impaired angiogenesis compared to control mice after cardiac ischemia followed by reperfusion<sup>7</sup>. Delivery of recombinant mouse MYDGF also ameliorated effects of cardiac ischemia and improved survival of mice in which MYDGF was not knocked out<sup>7</sup>. Human plasma has a median hMYDGF concentration of 3.3 ng/mL (0.2 nM) as analyzed by multiple reaction monitoring-mass spectrometry, with increased levels in patients with acute myocardial infarctions<sup>8</sup>. Endogenous MYDGF, other resident ER proteins, and a reporter construct containing the last 7 residues of hMYDGF have been shown to be released from SH-SY5Y human neuroblastoma cells in response to ER calcium depletion by thapsigargin<sup>9</sup>. The same result was obtained with cells exposed to oxygen-glucose deprivation<sup>9</sup>, an *in vitro* model of ischemia tied to depleted intracellular calcium stores<sup>10</sup>. Thus, the literature indicates that although MYDGF is predominantly retained in the ER as a result of engaging KDELRLs through its ERS, it is secreted upon cellular stress to act as a paracrine/autocrine survival factor with therapeutic potential.

MYDGF homologs are grouped into the Uncharacterized Protein Family (UPF) 0556 in the Pfam database (Pfam: PF10572 (<http://pfam.xfam.org/family/PF10572>)), which is annotated on the basis of sequence analysis as having unknown structure and no similarity to any other protein family. To fill the gap in knowledge regarding the structure of a protein expressed widely in nature and guide future functional studies of MYDGF, we now report the solution structure of hMYDGF determined at pH 6 by protein NMR. We demonstrate that the face of hMYDGF predicted to interact with KDELRLs is perturbed by changes in pH, whereas residues conserved in MYDGF homologs are clustered on the opposite face. A search for structural folds similar to hMYDGF identified the base domain of the vanin family of pantetheinases, which tethers the vanin nitrilase domain to the outer plasma membrane. We suggest, in analogy to vanins, that MYDGF complexed to KDELRLs binds cargo via its conserved residues for transport to the ER.

## Results

**Production and characterization of recombinant hMYDGF.** A gene coding for hMYDGF lacking the signal peptide residues (mature hMYDGF: V32-L173; UniProt Q969H8 ([https://www.](https://www.uniprot.org/uniprot/Q969H8)

[uniprot.org/uniprot/Q969H8](https://www.uniprot.org/uniprot/Q969H8))) with a thrombin-cleavable N-terminal 6xHis-tag was expressed in *Escherichia coli* cells grown in medium containing [<sup>13</sup>C]-glucose and <sup>15</sup>N-ammonia as, respectively, the sole sources of carbon and nitrogen. The protein was extracted in 8 M urea after cell lysis, purified by immobilized metal affinity chromatography (IMAC), renatured by dialysis against dilute acetic acid (pH 3.7) and then against Tris-buffered saline (TBS; pH 8.5), treated with thrombin to remove the N-terminal 6xHis-tag, and purified from the cleavage mix by size exclusion chromatography. The final product comprised the mature hMYDGF with the sequence GSKGT introduced at the N-terminus as a cloning artifact.

To evaluate whether hMYDGF folded and purified by this method has the same structure as hMYDGF processed through the ER, we compared the bacterially expressed protein to hMYDGF secreted from High Five insect cells. Cells were infected with recombinant baculovirus encoding the mature hMYDGF sequence flanked by an N-terminal gp67 signal sequence that targets the protein to the ER and a C-terminal 6xHis-tag that masks the ERS, enabling the construct to be secreted and subsequently isolated and purified from culture medium by IMAC. We performed top-down mass spectrometry (MS)<sup>11</sup> to confirm the sequences of the proteins and to determine the oxidation state of the two conserved cysteines (C63, C92). The experimental molecular masses of the tagged insect cell-derived and untagged bacteria-derived hMYDGF matched the masses calculated for the proteins with oxidized cysteines (Supplementary Fig. 1a, c). Tandem MS (MS/MS) of the insect cell-derived construct by collision-induced dissociation (CID) and the bacteria-derived construct by CID or electron-transfer dissociation (ETD), which produce backbone cleavages in the protein, did not yield fragmentation between C63 and C92 (Supplementary Fig. 1b, d), confirming that hMYDGF forms a disulfide bond in both expression systems. As a second comparison, C-terminus tagged insect cell-derived hMYDGF and N-terminus tagged bacteria-derived hMYDGF were examined by far UV circular dichroism (CD) spectroscopy (Supplementary Fig. 1e, Supplementary Table 1). The two spectra overlaid closely, with both exhibiting small negative peaks centered around 231 nm and 218 nm and a large positive peak near 202 nm. The secondary structure content predicted by the Beta Structure Selection (BeStSel) web server<sup>12</sup> on the basis of the CD spectra of the two constructs also closely matched: both with ~40% antiparallel  $\beta$ -strand, 2%  $\alpha$ -helix, and ~58% irregular (Supplementary Table 1). Finally, we measured the intrinsic fluorescence of the two hMYDGF tryptophan residues (W77, W95) (Supplementary Fig. 1f). Excitation at 295 nm produced single emission peaks for the tagged insect cell-derived protein and untagged bacteria-derived protein at 338.5 nm and 338.0 nm, respectively, and nearly identical peak contours. We conclude that the refolding protocol for bacterially expressed hMYDGF yielded a product experimentally equivalent to hMYDGF secreted from a eukaryotic cell. These experiments, therefore, justify use of the recombinant protein isotopically labeled by *E. coli* to determine the native structure of hMYDGF.

**Solution structure of hMYDGF solved by protein NMR.** The structure of [<sup>13</sup>C, <sup>15</sup>N]-hMYDGF in solution at pH 6 was determined by protein nuclear magnetic resonance (NMR) spectroscopy. This pH is at the lower end of the Golgi pH range at which KDELRLs efficiently engage ERS-containing proteins<sup>4,13</sup>. In brief, two-dimensional (2D) and three-dimensional (3D) NMR spectra were used to assign the <sup>1</sup>H, <sup>13</sup>C, and <sup>15</sup>N resonances from the backbone and sidechains of recombinant hMYDGF; the chemical shift assignment completeness was 91% for all atoms.

**Table 1 NMR and refinement statistics for the solution structure of hMYDGF.**

NMR constraints and structure statistics	hMYDGF
Distance constraints	
Total NOE	3703
Short-range ( $ i - j  \leq 1$ )	2932
Medium-range ( $1 <  i - j  \leq 5$ )	151
Long-range ( $ i - j  > 5$ )	620
Hydrogen bonds	28
Dihedral angle restraints	
Total	239
$\phi$	117
$\psi$	122
Structure statistics <sup>a</sup>	
Violations	
Distance constraints ( $>0.5 \text{ \AA}$ )	0
Dihedral angle constraints ( $>5^\circ$ )	0
Van der Waals ( $>0.2 \text{ \AA}$ )	0
Average pairwise RMSD <sup>b</sup> ( $\text{\AA}$ )	
Heavy	$1.30 \pm 0.09$
Backbone	$0.72 \pm 0.07$
Xplor-NIH pseudopotential energy ( $\text{kJ mol}^{-1}$ )	5481
MOLPROBITY mean score/clash score	2.10/11.21
MOLPROBITY Ramachandran plot summary <sup>b</sup> (%)	
Favored regions	95.7
Allowed regions	4.3
Disallowed regions	0.0

<sup>a</sup>Structure statistics were calculated using the 20 lowest pseudo-potential energy conformers out of the 100 total calculated conformers. The average pairwise RMSD was calculated against the lowest-energy conformer. <sup>b</sup>RMSD and Ramachandran statistics were obtained using ordered hMYDGF residues P35-A126 and D133-A168 as defined by CYRANGE

The resonance assignments and 3D NOESY spectra were utilized for Xplor-NIH-based structure calculation through the PONDEROSA-C/S suite, in an iterative process of structure calculation and constraint validation. This method calculated the top 100 most energetically stable hMYDGF conformers representing the 3D structure, from which the NMR statistics of the top 20 are summarized in Table 1. The final structure calculation was based on 3703 distance (620 long-range, 151 medium-range, 2932 short-range), 239 dihedral angle, and 28 hydrogen bond constraints with no constraint violations among the top 20 conformers. Backbone  $\phi/\psi$  angles in these structures were in favored regions of the Ramachandran plot for 96% of all ordered residues as assessed by MOLPROBITY, with none falling in disallowed regions. NMR data were deposited in the BioMagResBank<sup>14</sup> (BMRB 30584 ([http://www.bmrwisc.edu/data\\_library/summary/?bmrId=30584](http://www.bmrwisc.edu/data_library/summary/?bmrId=30584))), and the structural coordinates and restraints were deposited in the Protein Data Bank<sup>15,16</sup> (PDB 6O6W (<https://www.rcsb.org/structure/6O6W>)).

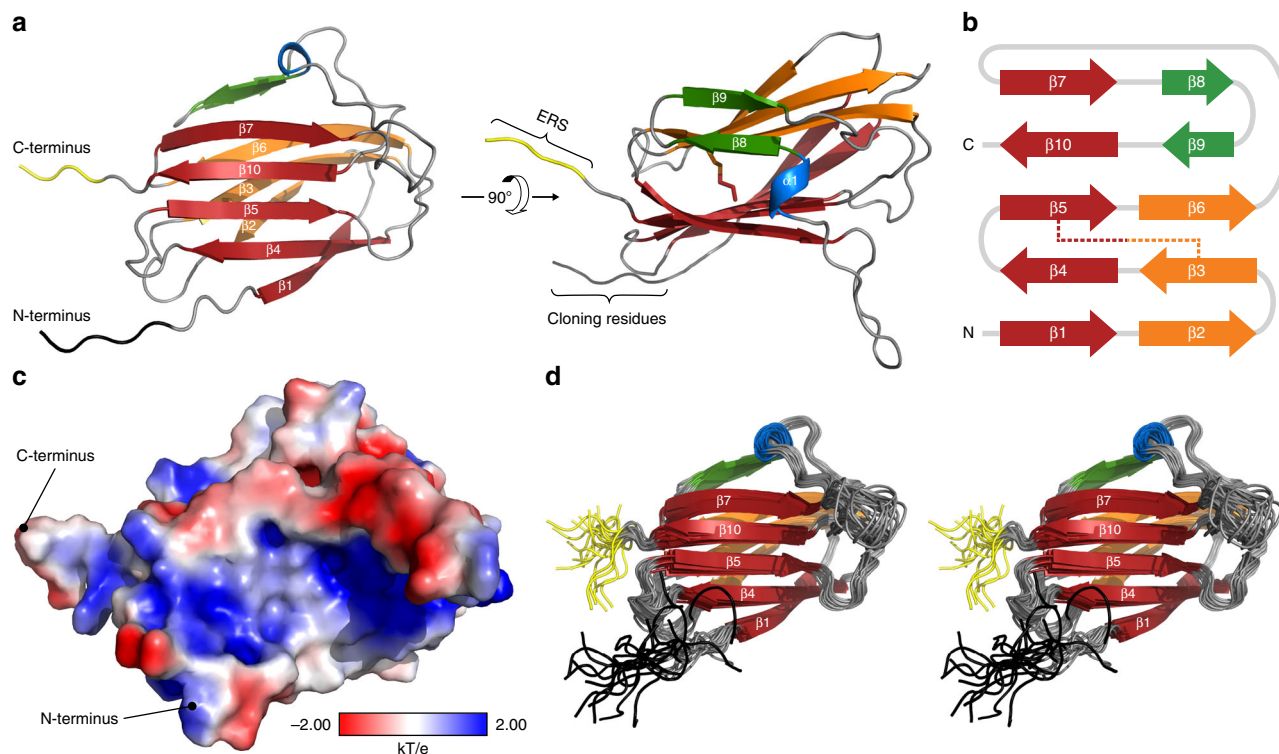
As shown for the most energetically stable conformer (Fig. 1a), the global fold of hMYDGF comprises three antiparallel  $\beta$ -sheets (ten  $\beta$ -strands:  $\beta 1$ – $\beta 10$ ) and a single  $\alpha$ -helical turn ( $\alpha 1$ ). The largest  $\beta$ -sheet ( $\beta 1$ ,  $\beta 4$ ,  $\beta 5$ ,  $\beta 10$ , and  $\beta 7$ ; red) is linked to a smaller  $\beta$ -sheet ( $\beta 2$ ,  $\beta 3$ , and  $\beta 6$ ; orange) by the disulfide bridge (stick representation between  $\beta 3$  and  $\beta 5$ ) to form a  $\beta$ -sandwich. The  $\beta$ -sandwich is capped at the  $\beta 6/\beta 7$  edge by the  $\alpha$ -helical turn (blue) and the third  $\beta$ -sheet ( $\beta 8$  and  $\beta 9$ ; green). The two terminal ends of hMYDGF are on the same face of the protein, with the five N-terminal residues having been introduced as a result of cloning (black) and the four C-terminal residues (RTEL) comprising the ERS (yellow). A schematic of  $\beta$ -strand connectivity is depicted in Fig. 1b, color-coded by  $\beta$ -sheet with the disulfide bridge represented by a dotted line. Positively charged (blue), negatively

charged (red), and uncharged (white) patches are scattered over the protein surface (Fig. 1c).

The 20 most energetically stable hMYDGF conformers (Fig. 1d) align with a root-mean-square deviation (RMSD) of 1.3  $\text{\AA}$  for all heavy atoms and 0.7  $\text{\AA}$  for backbone heavy atoms of ordered residues (Table 1). The regions with the highest degree of flexibility are located at the N-terminus, C-terminus, and the elongated loop between  $\beta 7$  and  $\alpha 1$ , which projects outward from the globule. These and other regions of hMYDGF lacking secondary structure align well with <sup>15</sup>N relaxation data, which showed longer  $T_2$  and smaller heteronuclear nuclear Overhauser effect (NOE) values when compared to the more structured sections of the backbone (Supplementary Fig. 2). The average relaxation times calculated from amides in secondary structure, 763.6 ms for  $T_1$  and 78.6 ms for  $T_2$ , were used to estimate a rotational correlation time ( $\tau_c$ ) of 9.4 ns for hMYDGF (Equation 2 in Rossi et al.<sup>17</sup>). This value corresponds to a molecular mass of 15.5 kDa<sup>17</sup>, which is close to the recombinant hMYDGF mass of 16.25 kDa and confirms the protein is monomeric in solution under the tested conditions.

The secondary structure composition of the hMYDGF NMR structure is 48% antiparallel  $\beta$ -strand, 2%  $\alpha$ -helix, and 50% irregular. This aligns well with the secondary structure predicted by BeStSel<sup>12</sup> of 45% antiparallel  $\beta$ -strand, 0%  $\alpha$ -helix, and 54% irregular calculated from CD spectra of hMYDGF in buffers ranging from pH 4.0 to 7.5 (Fig. 2a, Supplementary Table 1). The two tryptophan residues, while in close proximity, are on opposite sides of the largest  $\beta$ -sheet: W77 oriented toward the core on  $\beta 4$  and W95 toward the solvent on  $\beta 5$  (Supplementary Fig. 3a). However, only a single tryptophan fluorescence emission peak at 338 nm was present after excitation at 295 nm (Supplementary Fig. 3b). The peak was not perturbed by addition of 10 mM dithiothreitol (DTT), whereas addition of 6 M guanidine without or with DTT resulted in a single peak red-shifted to  $\sim 357$  nm and increased fluorescence intensity (Supplementary Fig. 3b). These results indicate that the tryptophan residues share a hydrophobic environment and transition to a hydrophilic environment upon denaturation. Consistent with this interpretation, we did not observe backbone or sidechain amide peaks for either tryptophan in clean solvent exposed amide (SEA) heteronuclear single quantum correlation (HSQC) spectra recorded with mixing times ranging from 10 to 140 ms to allow for hydrogen exchange with bulk water (Supplementary Fig. 3c, d; 140 ms spectrum presented). This result indicates that solvent exchange is relatively slow for both tryptophan residues.

**pH and calcium titration of hMYDGF.** Because hMYDGF may reside in multiple cellular microenvironments including calcium-rich ER (pH 7.2), Golgi (pH 6.0–6.7), and potentially other intracellular compartments of the secretory and endocytic pathways (pH down to 5.5<sup>4</sup>), we determined whether the structure of hMYDGF varies as a function of pH or calcium concentration. The overall fold of hMYDGF as assessed by CD spectroscopy was not sensitive to pH (Fig. 2a, Supplementary Table 1). <sup>1</sup>H, <sup>15</sup>N HSQC peaks from individual residues of hMYDGF, however, exhibit pH-dependent changes. Figure 2b displays an overlay of <sup>1</sup>H, <sup>15</sup>N HSQC spectra of hMYDGF at pH conditions ranging pH 5.5–8.0 colored-coded maroon to blue. Inasmuch as all backbone amides except the first two cloning residues have been assigned for hMYDGF at pH 6, we were able to efficiently assign the HSQC spectra recorded at the range of pH values. Peak perturbations for backbone amides varied depending on the residue. For example, the backbone amide peak for E34 remained at the same ppm values for all pH conditions, whereas H49 had the second highest pH shift perturbation (Fig. 2b, expanded view). We



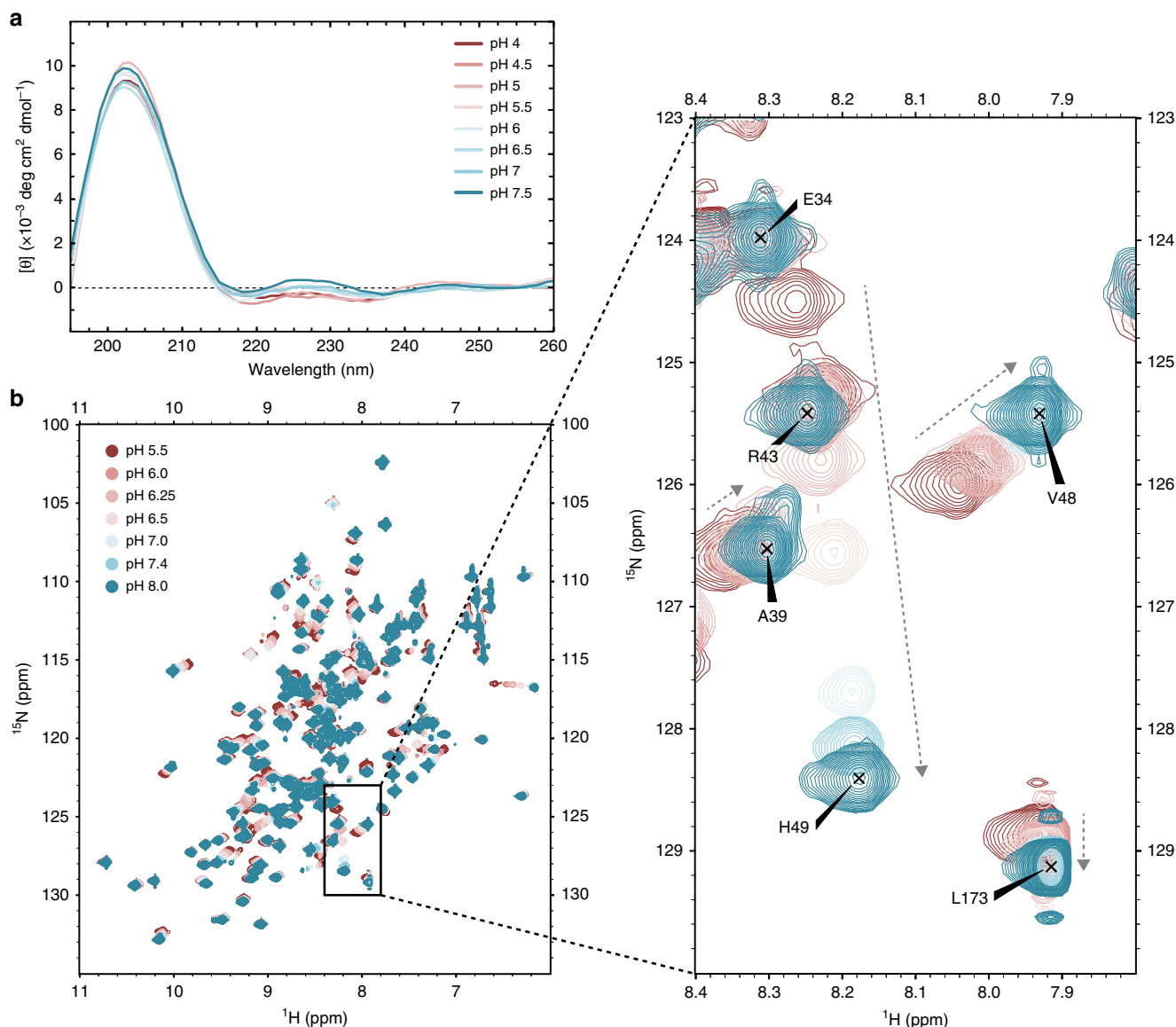
**Fig. 1** Solution structure of hMYDGF solved by protein NMR at pH 6. **a** The lowest pseudo-potential energy NMR conformer of hMYDGF. The structure is made up of ten  $\beta$ -strands ( $\beta 1$ – $\beta 10$ ) that constitute three  $\beta$ -sheets (red, orange, green), a short  $\alpha$ -helix (blue), and a disulfide bridge between strands  $\beta 3$  and  $\beta 5$  (right image, stick representation). The first five N-terminal residues were introduced as a result of cloning (black) and the last four C-terminal residues (RTEL) comprise the ERS (yellow). **b** Diagram of the  $\beta$ -strand connectivity colored by  $\beta$ -sheet with the disulfide linkage displayed as a dotted line. **c** Surface charge distribution of hMYDGF (lacking cloning residues) calculated at pH 6 depicting positively charged (blue), negatively charged (red), and uncharged (white) regions scattered across the protein. **d** Overlay of the 20 most energetically stable hMYDGF conformers presented in stereo view. Ordered hMYDGF residues align with an RMSD of 0.72 Å for backbone heavy atoms (Table 1), with the highest degree of flexibility at the N-terminus (black), the C-terminus (yellow), and the loop between  $\beta 7$  and the  $\alpha$ -helix (gray).

calculated the peak perturbations ( $\Delta\delta_{\text{NH}}$ ) from pH 5.5 to pH 8.0 for each assigned backbone amide and binned them from lowest (gray) to highest (green) ppm difference (Fig. 3a). The residues in the hMYDGF NMR structure were then color-coded to match the  $\Delta\delta_{\text{NH}}$  bins (Fig. 3b). Since this pH titration crosses the typical  $pK_a$  of a histidine imidazole sidechain ( $\sim 6.0$ ), it is reasonable that areas of the highest peak perturbations surrounded the hMYDGF histidine residues H49, H53, H87, and H89 (Fig. 3a, b; green). The backbone amides of the fifth hMYDGF histidine (H150) and the surrounding residues, however, were not perturbed by variations in pH. The edge of the  $\beta$ -sandwich that includes H53, H87, and H89, and from which the ERS extends, was the most pH-sensitive region (Fig. 3b). Plotting the  $\Delta\delta_{\text{NH}}$  from each spectrum relative to the pH 5.5 spectrum for the five histidine residues (W95 was included as a pH-independent amide peak) revealed noticeable differences among the titration curves (Fig. 3c). The curve for H150 resembled W95 in that both residue amides were minimally perturbed by pH. The titration curves for the remaining four histidine residues had slightly varying inflection points (most clearly seen when plotted as a percent maximum of  $\Delta\delta_{\text{NH}}$ ; Fig. 3c inset), with  $pK_a$  values of 6.4 for H49, 6.0 for H87, and 6.8 for H89 calculated from their fitted curves. The  $pK_a$  for H53 was estimated to be 5.4 based on an incomplete titration curve. We explored the effect of protonation state on surface charge distribution using the PDB2PQR server<sup>18</sup> and APBS plugin in PyMOL with  $pK_a$  values of ionizable groups calculated by PROPKA3.1 (refs. 19,20) (Fig. 3d). The surface of mature hMYDGF adjacent to the ERS, which includes H53, H87, and H89, is predicted to be predominantly positive at pH 6 and

neutral/negative at pH 7.4 (outlined in Fig. 3d) whereas remaining surface charge distribution is largely unchanged.

We used the HADDOCK protein docking web server<sup>21–23</sup> to dock the most energetically stable hMYDGF NMR conformer onto the crystal structure of chicken KDELR2 (cKDELR2) bound to an ERS-containing peptide (PDB 6I6H (<https://www.rcsb.org/structure/6I6H>)<sup>6</sup>; peptide was removed for docking calculations) (Supplementary Table 2, Supplementary Fig. 4). The crystal structure, like the NMR structure, was solved at pH 6, and cKDELR2 is 96% identical to human KDELR2 with sidechains of the nine variant residues all facing away from the binding cavity. The 400 complexes calculated by HADDOCK were grouped evenly into two clusters, with hMYDGF less centered along the cKDELR2 cavity and rotated  $\sim 130^\circ$  along the ERS in cluster 2 relative to cluster 1. In the lowest-energy models of both complexes, the C-terminal ERS of hMYDGF (residues RTEL) was bound inside the cKDELR2 cavity and up to 12 additional hMYDGF residues (including H89) were within 5 Å of cKDELR2 that formed interfaces stabilized by multiple polar contacts. Of the close-proximity hMYDGF residues, 12 were the same in cluster 1 (16 total) and cluster 2 (14 total). However, none of the polar contacts with cKDELR2 residues were the same. As highlighted in Fig. 3e (only cluster 1 shown), these interfaces coincide with the location of the pH-sensitive region of hMYDGF depicted in Fig. 3b and d, suggesting that these MYDGF residues have the potential to promote binding to KDELR2 in the Golgi and/or destabilize the complex for MYDGF release in the ER.

Because the ER is the dominant calcium store in the cell with concentrations reaching millimolar<sup>24</sup> and depletion of ER



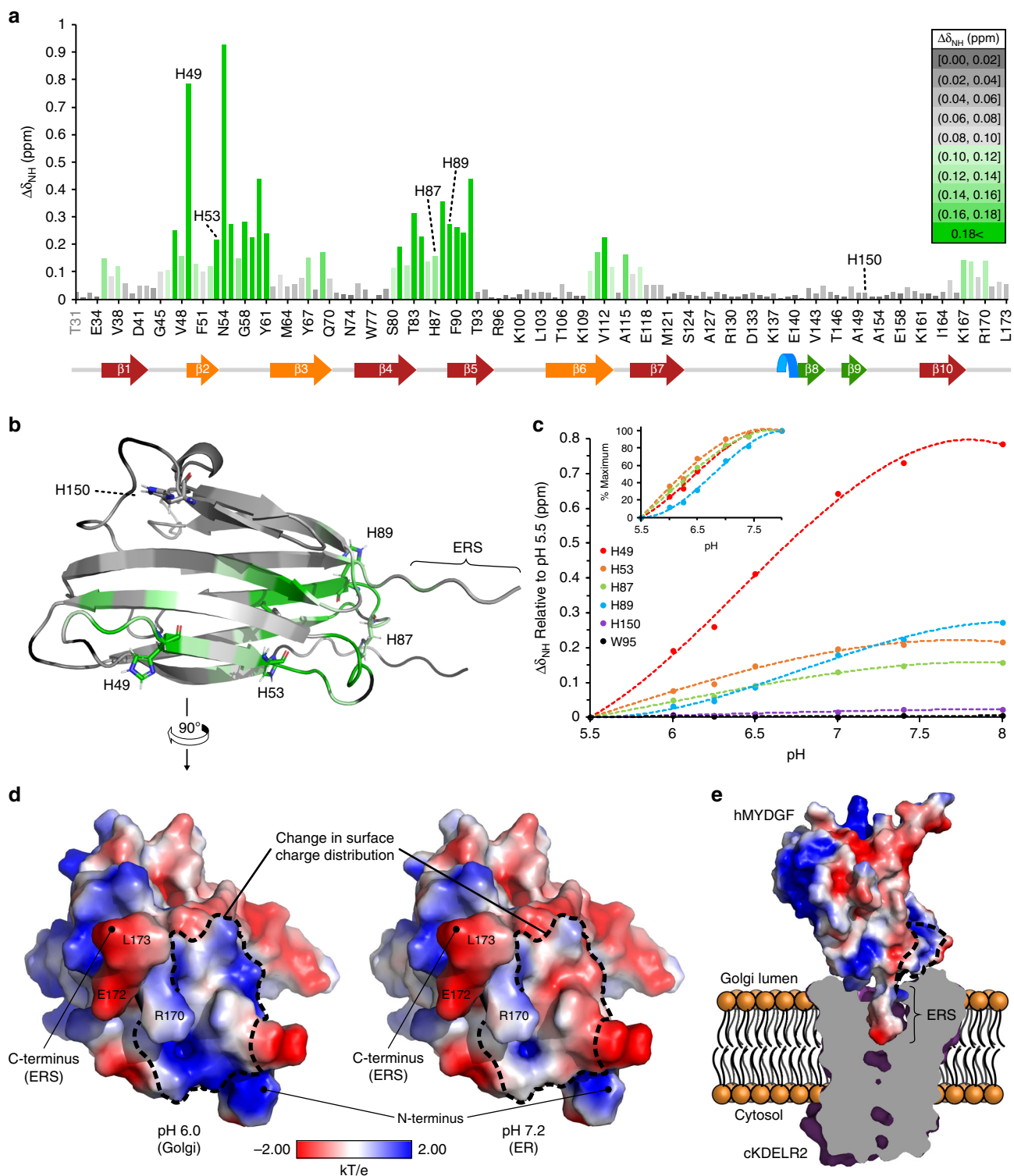
**Fig. 2** Effects of pH on hMYDGF. **a** Smoothed CD spectra of hMYDGF in buffer ranging from pH 4.0 (maroon) to pH 7.5 (blue). Spectra were baseline subtracted from their corresponding buffer spectra. Source data for each spectra are provided as a Source Data file. **b**  $^1\text{H}$ ,  $^{15}\text{N}$  HSQC spectra of hMYDGF in buffers ranging from pH 5.5 (maroon) to pH 8.0 (blue). Full spectra (left) and a representative expanded view (right) display a range of peak shifts from no movement (e.g., E34), to highly pH-sensitive peaks (e.g., H49). Arrows represent direction of peak perturbation as a result of increasing pH. Assignments shown are for peaks at pH 8.0.

calcium results in secretion of ERS-containing proteins including hMYDGF<sup>9</sup>, we determined whether the presence of calcium impacts hMYDGF structure. In contrast to the differences found as a function of pH, the  $^1\text{H}$ ,  $^{15}\text{N}$  HSQC spectrum of hMYDGF in the presence of up to 1 mM calcium (fourfold molar excess) was unperturbed relative to hMYDGF in the absence of calcium (Supplementary Fig. 5).

**MYDGF sequence similarity mapped onto the hMYDGF structure.** We used the ConSurf web server<sup>25</sup> to map conserved residues on the structure of hMYDGF. With mature hMYDGF as the input protein sequence, the HMMER homolog search algorithm in ConSurf identified 87 non-redundant MYDGF homologous sequences from the UniRef90 protein database. A phylogenetic tree of these 87 sequences based on MAFFT multiple sequence alignment is presented in Fig. 4a, with UniProtKB/UniParc protein identifiers listed at each branch. The tree is

color-coded based on the MYDGF homolog's phylum and class (Fig. 4a legend), which spans across the animal kingdom (hMYDGF is starred) and also includes slime molds and protozoans. A multiple sequence alignment of one MYDGF homolog from each class (hMYDGF representing Mammalia) is shown in Fig. 4b with the consensus logo and hMYDGF secondary structure underneath the sequences. The 25 positions marked with an asterisk have  $\geq 85\%$  sequence identity across the 236 UniRef90 MYDGF sequences found by Basic Local Alignment Search Tool (BLAST), with the 12 asterisks in black having  $\geq 90\%$  identity. Only ten of the 25 most conserved residues map to regions with secondary structure. The remaining 15 residues map to the ERS and loops that link  $\beta 1/\beta 2$ ,  $\beta 3/\beta 4$ ,  $\beta 5/\beta 6$ , and  $\beta 9/\beta 10$  strands.

Each residue of the 87 identified homologous sequences was assigned to a conservation level bin by ConSurf; these are color-coded from most variable (orange) to most conserved (blue) and mapped onto the most energetically stable NMR structure of



hMYDGF (Fig. 5a; residues from the cloning artifact are not shown). Among the most conserved residues are the two cysteines (Fig. 5a, center of the  $\beta$ -sandwich) and the two C-terminal residues that comprise the ERS. The ERS extends from an edge of the  $\beta$ -sandwich that contains predominantly variable residues (Fig. 5b). In contrast, the protein surface opposite the ERS, consisting of  $\beta$ 1/ $\beta$ 2,  $\beta$ 3/ $\beta$ 4,  $\beta$ 5/ $\beta$ 6, and  $\beta$ 9/ $\beta$ 10 loops, contains highly conserved residues (Fig. 5c). Underneath the surface of this conserved face between the  $\beta$ -sheets of the  $\beta$ -sandwich is an apparent cavity for which NOE interactions

between adjacent residue pairs lining the cavity were identified and validated, but no NOE interactions could be identified between residue pairs across it despite directed searches. When we analyzed the 20 most energetically stable conformers by Computed Atlas of Surface Topography of proteins (CASTp)<sup>26</sup> using the default probe radius of 1.4 Å (comparable to that of a water molecule), the solvent-accessible volume in this cavity ranged from 6 Å<sup>3</sup> to 67 Å<sup>3</sup>, with an average of 37 Å<sup>3</sup> and standard deviation of 16 Å<sup>3</sup> (Fig. 5d; cavity in red for the most energetically stable structure). The side-chains identified by CASTp to line the

**Fig. 3 pH-dependent alterations of hMYDGF backbone  $^1\text{H}$ ,  $^{15}\text{N}$  NMR peaks and surface charge distribution.** **a** Absolute value magnitude of the change in the combined  $^1\text{H}$ ,  $^{15}\text{N}$  chemical shift of each residue between pH 5.5 and 8.0. The  $\Delta\delta_{\text{NH}}$  values were binned and color-coded from lowest (gray) to highest (green) perturbation, with four of the five hMYDGF histidine residues falling into the highest bin. **b** Color-coded  $\Delta\delta_{\text{NH}}$  bins from **a** mapped onto the structure of hMYDGF. Aside from H150, the hMYDGF histidines (stick representations) mark regions of highest pH-sensitive peak perturbation between pH 5.5 to pH 8.0. Residues that could not be assigned in at least one of the  $^1\text{H}$ ,  $^{15}\text{N}$  HSQC spectra are colored black. **c** pH titration curves for the five hMYDGF histidines and W95 (no change) with  $\Delta\delta_{\text{NH}}$  plotted for each pH condition relative to the pH 5.5  $^1\text{H}$ ,  $^{15}\text{N}$  HSQC spectrum. Curves for H49, H53, H87, and H89 are also graphed as a percent of maximum change (inset). Source data used for chemical shift perturbation calculations are provided as a Source Data file. **d** Surface charge distribution (blue, positive; red, negative; white, uncharged) of hMYDGF lacking the five N-terminal cloning residues calculated at pH 6 (lower end of the Golgi pH range) and pH 7.2 (pH in the ER) using PDB2PQR<sup>18</sup> and APBS based on  $\text{pK}_a$  values predicted by PROPKA3.1 (refs. 19,20). The most drastic changes were in a region enclosed by dotted lines in which the positive surface charge became more neutral/negative with increasing pH. This region was adjacent to the protruding ERS (three labeled residues). **e** hMYDGF depicting surface charge distribution at pH 6 docked onto cKDEL2 (PDB 616H (<https://www.rcsb.org/structure/616H>)<sup>6</sup>; purple) from HADDOCK cluster 1 (see Supplementary Fig. 4). cKDEL2 was clipped (interior cross-section in gray) to show how the hMYDGF ERS is situated into the cKDEL2 binding pocket. The region of greatest change in surface charge highlighted in **d** is largely at the interface of hMYDGF and cKDEL2 (enclosed by dotted lines) and may play a role in stabilizing the interaction in the Golgi and destabilizing for hMYDGF release in the ER.

cavity were well-conserved and exclusively hydrophobic (Fig. 5d; color-coded based on ConSurf conservation bin).

**MYDGF and vanin base domain protein homology.** The MYDGF family of proteins is categorized under UPF0556 (<http://pfam.xfam.org/family/PF10572>) in Pfam based on sequence homology. The hMYDGF NMR structure allowed us to search for structural similarity to other proteins using the Dali server, which matches and ranks the query structure to the structures of all the proteins in the PDB<sup>27</sup>. The server identified the base domain from the crystal structure of human vanin-1 (VNN1; pantetheinase; PDB 4CYF (<https://www.rcsb.org/structure/4CYF>)<sup>28</sup>) as the top match to the structure of hMYDGF, with a Dali z-score of 7.6 and 110 ordered residues of hMYDGF superimposing with a Ca RMSD of 4.0 Å (Fig. 6a, left; VNN1 in cyan). Vanins are a family of enzymes (comprising VNN1, VNN2, and VNN3 in humans) that hydrolyze a carboamide linkage in D-pantetheine, thus releasing cystamine and recycling pantothenic acid (vitamin B5)<sup>29,30</sup>. The base domain is linked to the N-terminal enzymatic, nitrilase domain<sup>28</sup>. VNN1 and VNN2 are ectoenzymes, which contain a base domain with a C-terminal glycosylphosphatidylinositol (GPI)-anchored cleavage site that is modified to tether the enzymes to the outer leaflet of the plasma membrane<sup>29–31</sup>.

The human VNN1 base domain has the same  $\beta$ -strand connectivity as hMYDGF shown in Fig. 1b. The major structural differences are two additional disulfide bonds that are conserved in the vanin protein family (one between  $\beta_2$  and  $\beta_3$  strands and a second in the loop connecting  $\beta_5$  and  $\beta_6$  strands), hydrogen bonding between  $\beta_6$  and  $\beta_9$  to form a larger  $\beta$ -sheet ( $\beta_2$ ,  $\beta_3$ ,  $\beta_6$ ,  $\beta_9$ , and  $\beta_8$ ) rather than the two separate sheets in hMYDGF, and a smaller pocket (10 Å<sup>3</sup>) identified by CASTp in the vicinity of where the hMYDGF central cavity is located. A multiple sequence alignment of hMYDGF, human VNN1 base domain (UniProtKB: O95497 (<https://www.uniprot.org/uniprot/O95497>), residues V325–G491), and the Pfam seed sequences for MYDGF and vanin base domain (Pfam: PF10572 (<http://pfam.xfam.org/family/PF10572>) and Pfam: PF19018 (<http://pfam.xfam.org/family/PF19018>), respectively), revealed a number of conserved residues among the two protein families (Fig. 6b, consensus logo corresponds to the alignment from 82 sequences). Human MYDGF and VNN1 base domain have a sequence identity of 15%, with eight residues (asterisks) having  $\geq 85\%$  sequence identity among all sequences (black asterisks mark seven residues with  $\geq 90\%$  sequence identity). The eight residues are also conserved in over 75% of MYDGF homologs identified from the UniRef90 database (Fig. 5b). These residues appear to be structurally conserved as well, inasmuch as they are in close

proximity in the superposition of the two structures (Fig. 6a, right; VNN1 residues in cyan). The eight well-conserved residues include the disulfide connecting  $\beta_3$  to  $\beta_5$ , five core residues that line the hMYDGF cavity (Fig. 5d), and a glycine located on the same face as the C-terminus.

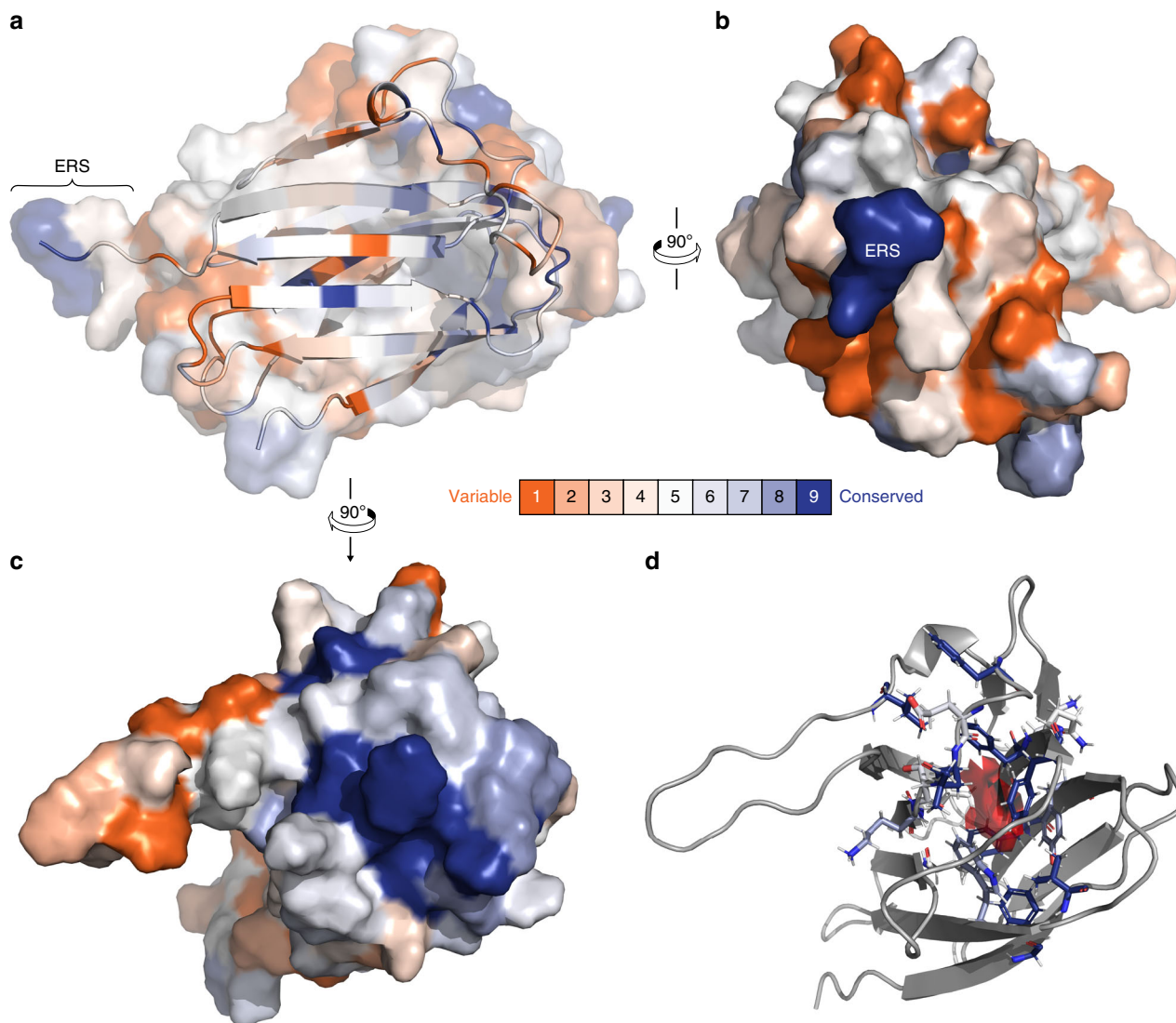
## Discussion

Our goal in initiating these studies was to determine the solution structure of hMYDGF as a model for other members of the widespread MYDGF/UPF0556 family of proteins and to provide insights into the functions of MYDGF as a well-conserved resident ER protein<sup>1</sup> and as a paracrine/autocrine survival factor with therapeutic potential after myocardial infarction<sup>7,32</sup>. We found that hMYDGF consists of a  $\beta$ -sandwich occluded at one edge by an  $\alpha$ -helical turn and small  $\beta$ -sheet (Fig. 1a, b). The protein is well-ordered except for the N-terminus, C-terminus (ERS), and elongated loop between  $\beta_7$  and the  $\alpha$ -helix. A disulfide bond bridges the sheets of the  $\beta$ -sandwich, which, remarkably, encloses a cavity surrounded by hydrophobic residues (Fig. 5d). The structure allowed us to discern that the same fold is present in the “base domain” of the crystal structure of VNN1. Based on this discovery, a new Pfam entry was created for the vanin base domain called “Vanin\_C” and given the accession PF19018 (<http://pfam.xfam.org/family/PF19018>). The similar fold and the presence of invariant residues are strong evidence that MYDGF and the vanin base are homologous protein domains.

The base domain of vanins tethers the nitrilase domain to the plasma membrane through a lipid anchor (Fig. 6c)<sup>29</sup>. As depicted in the same figure, ERS-bearing proteins such as MYDGF interact with multi-pass transmembrane KDELRs in the lumen of the Golgi<sup>6</sup>. The HADDOCK models of hMYDGF bound to cKDEL2 reveal a topology in which the ERS protrudes from the base of hMYDGF and the conserved loop residues are on the opposing face, available to interact with potential cargo. The nitrilase domain, in contrast, is at one side of the base domain in the crystal structure of VNN1 with a linker between the two domains wrapping around the base domain. KDELRs undergo a conformational change as a function of pH that favors binding of the ERS at pH 6 as found in the Golgi and release of the ERS at the higher pH found in the ER<sup>4,6,13</sup>. Although the global fold of hMYDGF is stable across this pH range, groups of residues, predominantly at the base of the  $\beta$ -sandwich from which the ERS protrudes and including three of the five hMYDGF histidines and surrounding residues, are sensitive to environmental pH as assessed by  $^1\text{H}$ ,  $^{15}\text{N}$  HSQC perturbation analysis and surface charge distribution calculations. Docking hMYDGF onto cKDEL2 revealed that a number of these residues are at the







**Fig. 5** Amino acid conservation mapped onto the structure of hMYDGF. **a** Using the ConSurf server, each residue of hMYDGF was binned from 1 (most variable, orange) to 9 (most conserved, blue) based on the alignment of 87 unique MYDGF protein homolog sequences (see Fig. 4) and mapped onto the structure of hMYDGF. In addition to the two cysteines that form the disulfide in the core of  $\beta$ -sandwich, the most conserved residues include, as highlighted in **b**, **c**, the C-terminus Glu-Leu sequence of the ERS and those in loops opposite the ERS. **d** A cavity (red) lined with conserved, hydrophobic residues (sticks; color-coded by conservation) beneath the surface shown in **c** was identified by CASTp and recreated in PyMOL as displayed here. The cavity has an average volume of  $37 \text{ \AA}^3$  among the 20 most energetically stable conformers that represent the structure of hMYDGF.

binding interface between the two proteins, raising the possibility that pH-dependent changes in both MYDGF and KDELRs modulate MYDGF–KDELR interactions. The significance of the two predicted modes of MYDGF binding to KDELR2 at pH 6 is not known.

MYDGF homologs are present in organisms throughout and beyond the animal kingdom, including species as distant to humans as slime molds and protozoans. Conservation analysis using ConSurf identified the C-terminal Glu-Leu residues that are part of the ERS among the most conserved residues (Fig. 5; bin 9, dark blue), indicating that the ERS is a critical component of nearly all MYDGF-family proteins. Of the most-conserved MYDGF residues in addition to those in the ERS, 12 do not overlap with conserved residues shared with vanin base family members (Fig. 6b, asterisks) and are highlighted as dark blue sticks in Fig. 6d. Eight of these residues reside in the loops on the opposite face from the ERS and the remaining four are between the  $\beta$ -sheets of the  $\beta$ -sandwich and surround the cavity (Fig. 6d,

red). While none of the 20 conformers individually has a clear inlet to the cavity, an overlay of these conformers reveals a potential entrance path from underneath the  $\beta 9$ – $\beta 10$  loop. We hypothesize that these conserved hMYDGF residues constitute a dynamic external (loops) and/or internal (cavity) binding interface for conserved and as-of-yet unknown interactor(s). Binding to this conserved interface of KDELR-tethered MYDGF is a potential means for the cell to return metabolic intermediates or non-ERS-containing proteins to the ER.

A possible alternative function of MYDGF is as part of the “tool kit” of ER components that assist in protein folding and movement through the ER. MYDGF itself likely does not require interaction with protein-folding machinery. The protein produced in *E. coli* renatured readily after exposure to 8 M urea into a structure with all peptidyl-prolyl peptide bonds in the trans configuration and a disulfide that is buried and apparently inert. There is no site for N-linked glycosylation in MYDGF, and we detected no post-translational modifications by MS. How



**Fig. 6 Comparison of hMYDGF and human VNN1 base domain.** **a** Left: superposition of ordered residues (P35-A126, D133-A168) from the NMR structure of hMYDGF upon the crystal structure of the human VNN1 base domain (PDB 4CYF (<https://www.rcsb.org/structure/4CYF>)<sup>28</sup>; cyan). 110 residues aligned with a  $\alpha$  RMSD of 4.0 Å. The VNN1 base domain is flanked by a nitrilase domain at the N-terminus and a GPI anchor at the C-terminus (cyan arrows highlight the strands of the base domain that lead to these features). Right: stick representations of highly conserved residues (circled) shared by human VNN1 (cyan) and hMYDGF superimposed on the hMYDGF structure. **b** Multiple sequence alignment and consensus logo of hMYDGF, human VNN1 base domain, and Pfam seed sequences for MYDGF (Pfam: PF10572 (<http://pfam.xfam.org/family/PF10572>)) and vanin base domain (Pfam: PF19018 (<http://pfam.xfam.org/family/PF19018>)) families (82 sequences total). hMYDGF and the human VNN1 base domain are displayed as representative sequences, with hMYDGF sharing 15% sequence identity with the VNN1 base domain. Asterisks represent eight highly conserved residues among all sequences (black,  $\geq 90\%$  identity; gray,  $\geq 85\%$  identity), which align well in space between hMYDGF and the human VNN1 base domain as shown in **a**. **c** The VNN1 base domain (cyan) is connected to the nitrilase domain (blue) via a linker strand (pink) at its N-terminus and tethered to the plasma membrane by a GPI anchor at the C-terminus. An analogous model of hMYDGF (colored by residue conservation as in Fig. 5: blue, most conserved; orange, most variable) is presented in a similar orientation as the VNN1 base domain and engaging kDEL2 (PDB 6L6H (<https://www.rcsb.org/structure/6L6H>)<sup>6</sup>; purple) from HADDOCK cluster 1 (Supplementary Fig. 4) in the Golgi via its C-terminal ERS with the remaining highly conserved residues on the opposite face. **d** The 14 most-conserved MYDGF residues (Fig. 5, bin 9) that do not overlap with conserved residues shared with vanin base family members (asterisks in **b**) are highlighted as dark blue sticks. Note that aside from the C-terminal Glu-Leu residues of the ERS, these MYDGF-specific conserved residues reside in the loops on the face opposite to the ERS and surface of the cavity (red).

human protein (V32-L173; UniProt Q969H8 (<https://www.uniprot.org/uniprot/Q969H8>)) was subsequently amplified by polymerase chain reaction and cloned into the pAcGP67.coco plasmid<sup>39</sup> using 5' primer CCG GCG GAT CCG GTG TCC GAG CCC ACG ACG (BamHI restriction site) and 3' primer GC TGC TTC TAG AAG CTC AGT GCG CGA TGC CTT GG (XbaI restriction site), and into the pET.ELMER plasmid<sup>40</sup> using 5' primer CCG GCG GGT ACC GTG TCC GAG CCC ACG ACG GTG (KpnI restriction site) and 3' primer CAG GGC GCT AGC TCA CAG CTC AGT GCG CGA TGC C (NheI restriction site).

Sf9 insect cells (11496015, Thermo Fisher Scientific) were co-transfected with the pAcGP67.coco-MYDGF plasmid and Sapphire Baculovirus DNA (ABP-BVD-10001, Allele Biotechnology), from which high titer stocks of recombinant baculovirus containing MYDGF were obtained by plaque purification and three rounds of amplification<sup>39</sup>. High Five insect cells (B85502, Thermo Fisher Scientific) were then infected with the recombinant baculovirus for hMYDGF production<sup>39</sup>. In this system, hMYDGF was expressed with an N-terminal gp67 signal sequence and a C-terminal 6xHis-tag that masks the ERS, ultimately resulting in the protein's secretion. The recombinant protein was purified from the medium by IMAC using nickel-nitrilotriacetic acid resin (30230, Qiagen). The purified insect cell-derived hMYDGF construct contained N-terminal residues ADP and C-terminal residues LELVPRGSAAGHHHHHH as a result of the cloning process.

hMYDGF was also expressed in and purified from either BL21(DE3) (69450, MilliporeSigma) or Rosetta 2(DE3) competent cells (71400, MilliporeSigma) transformed with the pET.ELMER-MYDGF plasmid. The only observable difference between the competent cells was that use of Rosetta 2(DE3) resulted in slightly higher protein yield. Transformed cells were grown at 37 °C in LB medium containing 30 µg/mL kanamycin and 25 µg/mL chloramphenicol until the OD<sub>600</sub> was between 1 to 1.5, upon which protein expression was induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 4 h. Cells were harvested by centrifugation then lysed by freeze-thaw and incubation in a cell lysis solution (100 mM sodium phosphate, 10 mM Tris Base, 8 M urea, 5 mM imidazole, 1 mM β-mercaptoethanol, pH 8). The recombinant protein, which was expressed with N-terminal residues MGS SHHHHHH GSLVPRGSKGT preceding the mature hMYDGF sequence, was purified by IMAC, dialyzed against dilute acetic acid (100 mM, then 1 mM; pH 3.7) to remove components of the cell lysis solution, dialyzed against phosphate-buffered saline (PBS; pH 6) or TBS (pH 8.5), and purified by size-exclusion chromatography on a HiLoad 16/600 Superdex 75 prep grade column (28-9893-33, GE Healthcare) equilibrated in and eluted with the same buffer. In some batches, the N-terminal 6xHis-tag was cleaved off leaving only cloning residues GSKGT at the N-terminus (labeled as G27-T31 in figures). This was performed prior to size-exclusion chromatography by incubation for 4 h in a room temperature reaction in TBS using 0.01 units of thrombin (HT 1002a, Enzyme Research Laboratories) per 1 µg of recombinant hMYDGF. Bacteria-derived hMYDGF was concentrated after size-exclusion chromatography using Amicon Ultra-4 centrifugal filters (UFC801024, MilliporeSigma) and dialyzed against desired buffer depending on the experiment (specified for each section; all within  $\pm 0.02$  pH units from reported pH). Protein concentrations were determined using Beer's law, with absorbance measurements at 280 nm obtained using a SpectraMax M5 Microplate Reader equipped with SoftMax Pro v6.3 software (Molecular Devices) and extinction coefficients predicted by the Expasy ProtParam tool.

**Top-down MS.** Insect cell-derived hMYDGF (containing the C-terminal 6xHis-tag) and bacteria-derived hMYDGF (lacking the N-terminal 6xHis-tag) were dialyzed against 1 mM acetic acid (pH 3.7) and diluted with equal volume of acetonitrile to 12.5 µg/ml for top-down MS and MS/MS analysis. Both MS and MS/MS were performed on a Bruker maXis II ETD quadrupole time-of-flight mass spectrometer (Bruker Daltonics, Bremen, Germany) by direct infusion at a flow rate of

6 µL/min. Mass spectra of insect cell-derived hMYDGF and bacteria-derived hMYDGF were acquired at a scan rate of 1 Hz over 200–2000 *m/z* and 500–3000 *m/z* range, respectively. Targeted MS/MS of CID was performed with the selected charge states of insect cell-derived MYDGF (21+) at 1 Hz over 200–2000 *m/z*. The isolation window was set to 2 *m/z*, and the collision direct current bias was fixed to 23 V. Targeted MS/MS of CID and ETD was performed with the selected charge states of bacteria-derived MYDGF (19+) at 1 Hz over 500–3000 *m/z* range. The isolation window was set to 2 *m/z*. The collision direct current bias was fixed to 18 V for CID. The precursor ion accumulation was set to 800 ms with a reagent (3,4-hexanedione) injection duration of 8 ms and an additional 0 ms reaction for ETD. The MS spectra were deconvoluted by the Maximum Entropy algorithm with a resolving power of 80,000 using Bruker DataAnalysis 4.3 (Supplementary Fig. 1). All fragment ions from both CID and ETD were manually validated using MASH Suite Pro<sup>41</sup>. Fragments of *b*, *y*, *c*, *c* − 1, *z*, and *z* + 1 ions were assigned with a mass tolerance of 15 ppm after the validation to generate the ion fragment maps (Supplementary Fig. 1). All masses are reported as the monoisotopic masses.

**CD spectroscopy.** Far UV CD spectroscopy data were collected on an AVIV Model 420 Circular Dichroism Spectrometer and analyzed by Igor Pro v6.3 software. Scans were collected at 25 °C using the following parameters: 0.1 cm cuvette pathlength, 1 cm bandwidth, 1 nm steps, and 10 s averaging time. All scans were subtracted from corresponding buffer scans and converted to units of molar ellipticity. For comparison between 6xHis-tagged insect cell- and bacteria-derived MYDGF (Supplementary Fig. 1e), the insect cell-derived protein (5.9 µM) was dialyzed against 10 mM sodium phosphate, 10 mM NaCl (pH 7.5) and bacteria-derived protein (4.8 µM) was dialyzed against 10 mM sodium phosphate (pH 6). For the pH titration scans (Fig. 2a), bacteria-derived MYDGF lacking the 6xHis-tag (concentrations ranging 4.4–4.7 µM) was dialyzed in 10 mM sodium acetate for pH 4.0–5.5 conditions, and 10 mM sodium phosphate for pH 6.0–7.5 conditions. The BeStSel server<sup>42</sup> was used for secondary structure prediction based on the CD spectra (Supplementary Table 1) and the CD Analysis and Plotting Tool (CAPITO)<sup>42</sup> was utilized to smooth the CD spectra by applying a Savitzky-Golay filter for clearer visual comparison among the scans.

**Intrinsic tryptophan fluorescence.** Intrinsic tryptophan fluorescence of insect cell-derived hMYDGF (containing an N-terminal 6xHis-tag) and bacteria-derived hMYDGF (lacking the N-terminal 6xHis-tag) in TBS (pH 7.4) was measured using a Fluoromax-3 Spectrofluorometer in conjunction with Datamax v2.2 spectroscopy software (HORIBA Jobin Yvon). When comparing insect cell- and bacteria-derived hMYDGF (Supplementary Fig. 1f), scans were collected in 2 mm × 10 mm quartz cuvettes at 25 °C using the following parameters: excitation λ: 295 nm, emission λ: 305–400 nm, excitation bandwidth: 2 nm, emission bandwidth: 4 nm, 0.5 nm steps, and 0.5 s integration time. The same conditions/parameters were used for analysis of native versus denatured and/or reduced bacteria-derived hMYDGF (Supplementary Fig. 3b), with the exception of the following: excitation bandwidth: 2 nm, emission bandwidth: 4 nm, and 1 s integration time. All samples contained 1 µM recombinant hMYDGF. Resulting spectra were an average of three scans that were baseline subtracted from their corresponding, buffer scans. Fluorescence peak maximums were determined from sixth-order polynomial trendlines, which were fit to the spectra using Microsoft Excel.

**Production of isotopically labeled MYDGF.** The method for producing single (<sup>15</sup>N) and double (<sup>13</sup>C, <sup>15</sup>N) labeled hMYDGF was adapted from multiple protocols<sup>43–45</sup>. A 1-mL glycerol stock of Rosetta 2(DE3) cells transformed with pET.ELMER-MYDGF in MDG medium<sup>43</sup> was used to inoculate 50 mL of sterile MDG

medium containing 60 µg/mL kanamycin and 50 µg/mL chloramphenicol. Culture was grown overnight at 25 °C in a shaking incubator. The following day, 1 L of sterile-filtered, heavy isotope-enriched minimal growth medium was prepared composed of the following: 1× M9 salts<sup>44</sup>, 0.1× metal mix<sup>43</sup>, 2 mM MgSO<sub>4</sub>, 0.4% w/v glucose (<sup>13</sup>C-glucose for double-labeled sample), 0.2% w/v <sup>15</sup>NHCl<sub>4</sub>, 0.1 mM CaCl<sub>2</sub>, 1× vitamin solution<sup>45</sup>, 1× vitamin B<sub>12</sub> solution<sup>45</sup>, 30 µg/mL thiamine, 60 µg/mL kanamycin, and 50 µg/mL chloramphenicol. One liter of minimal growth medium was inoculated with 20 mL of overnight culture and incubated at 37 °C with shaking until OD<sub>600</sub> ≈ 1. Protein expression was induced with 1 mM IPTG for 25 h at 25 °C with shaking. The protocol described above was performed to extract, renature, remove the 6xHis-tag, purify from the cleavage mix, dialyze against desired buffer, and determine concentration for the isotopically labeled proteins. NMR samples were supplemented to achieve final concentrations of the following: 10% v/v D<sub>2</sub>O, 15 µM 4,4-dimethyl-4-silapentane-1-sulfonic acid (internal chemical shift reference), and 0.02% w/v NaN<sub>3</sub> (bacteriostat).

**NMR data collection and processing.** NMR spectra were acquired on Bruker AVANCE III or Varian VNMR5 spectrometers ranging from 600 MHz to 900 MHz using TopSpin 3.5 and VNMRJ, respectively, and equipped with cryogenic triple-resonance probes. The temperature of the sample was regulated at 298 K for all recorded experiments. Data processing was conducted using the NMRPipe package<sup>46</sup>. Non-uniform sampling (NUS) was employed for the triple resonance experiments with sampling rates ranging from 30 to 40%. A Poisson-gap sampling<sup>47</sup> schedule was used for the NUS data acquisition and the SMILE plug-in<sup>48</sup> in NMRPipe was used for spectral reconstruction.

**NMR structure determination of MYDGF.** Backbone resonances were manually assigned in NMRFAM-SPARKY<sup>49</sup> using 2D <sup>1</sup>H, <sup>15</sup>N HSQC, 3D CBCA(CO)NH, 3D HNCACB, and 3D HNCO spectra with an additional 3D <sup>1</sup>H, <sup>15</sup>N HSQC - <sup>1</sup>H, <sup>13</sup>C HSQC NOESY experiment used for assigning the A157 backbone amide (4.6 ppm in the <sup>1</sup>H-dimension). All backbone amides were assigned aside from the first two cloning residues of the recombinant protein (labeled as G27–G28 in figures). Side chain resonances were also manually assigned using 2D <sup>1</sup>H, <sup>13</sup>C HSQC (aliphatic), 2D <sup>1</sup>H, <sup>13</sup>C HSQC (aromatic), 3D C(CO)NH, 3D HBHA(CO)NH, 3D H(CCO)NH, 3D HCCH-TOCSY (aliphatic), 3D HCCH-TOCSY (aromatic), 2D (HB)CB(CGCD)HD (aromatic), and 2D (HB)CB(CGDCDCE)HDHE (aromatic) spectra. Manually assigned resonances were cross-validated using the I-PINE web server<sup>50</sup> after submitting a job with the PINE-SPARKY.2 plugin<sup>51</sup>, and the structure-independent probabilistic validation algorithm ARECA<sup>52</sup>. The <sup>13</sup>C chemical shifts for the Ca and Cβ of the two hMYDGF cysteines (C63, C92) also matched reported resonances of cysteines that form disulfides in β-sheets of other proteins<sup>53</sup>.

The solution structure of <sup>13</sup>C, <sup>15</sup>N hMYDGF (508 µM) in PBS (pH 6) was solved using an Integrative NMR approach<sup>54</sup> with three additional NMR spectra: 3D <sup>1</sup>H, <sup>15</sup>N HSQC NOESY, 3D <sup>1</sup>H, <sup>13</sup>C HSQC NOESY (aliphatic), and 3D <sup>1</sup>H, <sup>13</sup>C HSQC NOESY (aromatic). Initial folding was calculated with the PONDEROSA refinement option, which utilizes CYANA<sup>55</sup> for obtaining inter-residual proximity information used in the Ponderosa Analyzer white-list/black-list manager for efficient NOE assignment in consequent steps. Xplor-NIH-based calculations (AUDANA<sup>56</sup>) were used for the remaining majority of submissions in the PONDEROSA-C/S package<sup>57</sup>. Distance and angle constraints were generated and used for structure calculation by running PONDEROSA-X refinement, which utilized the AUDANA algorithm and TALOS-N<sup>58</sup> optimization from the resonance assignment list and NOESY spectra inputs. Constraints were carefully validated using Ponderosa Analyzer interfaced with the Ponderosa Connector plug-in (two-letter-code “up”) in NMRFAM-SPARKY and PyMOL 2.0 programs. After several iterative calculations using the Constraints Only-X option, we finalized constraint refinement by running Final Step with the explicit water refinement option. This provided the top 20 out of 100 lowest pseudo-potential energy conformers and conducted energy minimization in the water box. The final structures were validated using the wwPDB Validation Service (<https://validate-rcsb-2.wwpdb.org/>)<sup>15,16</sup> and MolProbity<sup>59</sup> online servers. Structural NMR statistics for the 20 most energetically stable hMYDGF conformers are summarized in Table 1.

**MYDGF relaxation times and hydrogen exchange with solvent.** <sup>15</sup>N relaxation experiments (*T*<sub>1</sub>, *T*<sub>2</sub>, and heteronuclear NOE; Supplementary Fig. 2), and amide hydrogen exchange spectra (Supplementary Fig. 3c, d) were recorded on Varian VNMR5 spectrometers operating at 600 and 800 MHz, and equipped with a cryogenic triple-resonance probe. The same sample that was used for structure determination was also used for collecting these NMR experiments. All spectra were recorded with the temperature of the sample regulated at 298 K.

For measuring <sup>15</sup>N *T*<sub>2</sub> values, multiple 2D <sup>1</sup>H, <sup>15</sup>N spectra were recorded in an interleaved fashion using relaxation delays of 10, 30, 50, 70, 90, 110, 130, 150, 170, 190, and 210 ms. Similarly, <sup>15</sup>N *T*<sub>1</sub> values were measured using relaxation delays of 80, 160, 240, 320, 400, 560, 720, 960, 1200, 1520, and 2000 ms. <sup>15</sup>N heteronuclear NOE experiments were recorded in duplicate using a relaxation delay of 5 s with and without saturation of the amide protons. All the spectra were processed in NMRPipe and analyzed in NMRFAM-SPARKY. <sup>15</sup>N *T*<sub>1</sub> and *T*<sub>2</sub>

relaxation times were calculated by fitting the decaying signals to a single exponential function using the “rh” extension in NMRFAM-SPARKY. <sup>15</sup>N heteronuclear NOE values were calculated from the ratio of corresponding intensities of the spectra recorded with and without <sup>1</sup>H saturation using the “np” extension.

To map SEA groups, amide hydrogen exchange experiments were acquired by using a clean SEA HSQC experiment<sup>60</sup>. In short, magnetization from all protein hydrogens is first eliminated by a double <sup>15</sup>N/<sup>13</sup>C filter and then allowed to recover through exchange with bulk water during a mixing period (ranging from 10 to 140 ms), such that only amides that are exposed to the solvent are observed. All spectra were processed with NMRPipe and analyzed in NMRFAM-SPARKY.

**Calcium and pH titration.** 2D <sup>1</sup>H, <sup>15</sup>N HSQC spectra of <sup>15</sup>N hMYDGF (257 µM) in TBS (pH 7.4) were collected in the presence of 0.5-, 1-, 2-, and 4-fold molar excess of CaCl<sub>2</sub>. The spectrum of the sample with the highest calcium concentration is compared to the spectrum of hMYDGF in the absence of calcium in Supplementary Fig. 5.

For pH titration, 2D <sup>1</sup>H, <sup>15</sup>N HSQC experiments were recorded at various pH values (Figs. 2b and 3a–c). <sup>15</sup>N hMYDGF (concentrations ranging 215–356 µM) was dialyzed against 10 mM acetic acid, 10 mM sodium phosphate, 150 mM NaCl at pH 4.0, 5.5, 6.0, 6.25, 6.5, 7.0, and 8.0. hMYDGF at pH 4.0 precipitated out of solution and a suitable <sup>1</sup>H, <sup>15</sup>N HSQC spectrum could not be obtained. The same spectrum from the calcium titration studies (before addition of CaCl<sub>2</sub>) was used as the pH 7.4 condition. All spectra were aligned to the W77 backbone amide peak which displayed one of the smallest chemical shift changes as a function of pH. The chemical shift perturbation ( $\Delta\delta_{\text{NH}}$  in ppm) for each assigned peak was calculated using following equation:

$$\Delta\delta_{\text{NH}} = \sqrt{(\Delta\delta_{\text{H}})^2 + (\Delta\delta_{\text{N}}/5)^2}, \quad (1)$$

where  $\Delta\delta_{\text{H}}$  and  $\Delta\delta_{\text{N}}$  are the peak displacements (ppm) for the hydrogen and nitrogen dimensions, respectively. Peaks with the lowest displacement were categorized in the first bin of Fig. 3a (darkest gray). The upper bound of the first bin was set to the third quartile +1.5 × (interquartile range) of  $\Delta\delta_{\text{NH}}$  values between the pH 6 <sup>1</sup>H, <sup>15</sup>N HSQC spectrum for this set of experiments and the pH 6 <sup>1</sup>H, <sup>15</sup>N HSQC spectrum of hMYDGF in PBS used for backbone resonance assignments, which had minimal overall peak perturbations. Consequent bins were set and color-coded based on multiples of the first, with the tenth bin (darkest green) containing the peaks with the highest  $\Delta\delta_{\text{NH}}$  values. Trendlines fit to the  $\Delta\delta_{\text{NH}}$  values of the five hMYDGF histidines and W95 (Fig. 3c) were third-order polynomials, with pK<sub>a</sub> values calculated from their second derivatives set to 0.

The surface charge distribution for the most energetically favorable NMR conformer of hMYDGF (lacking the five N-terminal cloning residues) was calculated at pH 6 and pH 7.2 (Fig. 3d, e) through the PDB2PQR server<sup>18</sup> using PROPKA3.1 (refs. 19,20) to assign protonation states at the desired pH values. The APBS Tools 2.1 Plugin in PyMOL was then utilized to visualize the surface charge distribution.

**MYDGF docking onto KDELR2.** Predicted complexes between the most energetically stable NMR conformer of hMYDGF (omitting the five N-terminal residues introduced from the cloning process) and the crystal structure of cKDELR2 (PDB 616H (<https://www.rcsb.org/structure/616H>)<sup>6</sup>) were generated using the webserver implementation of the computational docking program HADDOCK<sup>21–23</sup>. This program is predicated on the fact that knowledge of the contacting interfacial residues in experimentally determined complexes involving homologous proteins can be used as soft restraints when predicting a given unknown complex. The experimentally solved complex between cKDELR2 and the ERS-containing peptide TAEKDEL at pH 6 provided information regarding the cKDELR2 residues (specifically R5, R47, Y48, E117, R159, Y162, N165, W166, and R169) that are implicated in binding ERS sequences in the Golgi by forming a network of hydrogen bonds<sup>6</sup>. The soft restraints that the HADDOCK webserver uses in its docking calculation were specified by these cKDELR2 residues in addition to the hMYDGF ERS sequence (residues RTEL). In brief, the algorithm consists of three steps: (1) randomization of the orientation of one of the proteins and rigid body energy minimization, (2) simulated annealing and energy minimization in torsion angle space, and (3) refinement with energy minimization and molecular dynamics in explicit water, retaining soft pairwise restraints between the constituent atoms of the user-specified residues at all three stages (for additional algorithmic details, refer to cited publications<sup>21–23</sup>). HADDOCK grouped 400 predicted output complexes into two relatively equal-sized clusters (statistics summarized in Supplementary Table 2). These clusters had average energetic scores (unitless HADDOCK scores) of  $-94.9 \pm 3.6$  and  $-73.2 \pm 0.6$ , which are calculated as weighted sums of various energies and buried surface areas (the cluster with the lowest HADDOCK score is ranked first). The lowest-energy complexes from each of these two clusters were examined for further analysis (Supplementary Fig. 4).

**Analysis of MYDGF sequence similarity.** MYDGF sequence similarity was mapped onto the structure of hMYDGF using the ConSurf web server<sup>25</sup>. A PDB file of hMYDGF (lowest pseudo-potential energy structure lacking the first five cloning residues) was submitted to ConSurf and a multiple sequence alignment was

generated with parameters set to use the HMMER homolog search algorithm (1 iteration, 0.0001 *E*-value cutoff) against the UniRef90 protein database. Homolog selection parameters were set to 500 sequences that sample the list of homologs to the reference sequence, with cutoffs of 100% for maximum identity between sequences and 20% for minimum identity for homologs. MAFFT(L-INS-i) was used to build the multiple sequence alignment of the 87 identified, unique sequences and the Bayesian calculation method was used to determine the rate of evolution at each site in the alignment. Resulting ConSurf output files were used as the basis for creating a phylogenetic tree using the Interactive Tree Of Life (iTOL v4) web server<sup>61</sup> and mapping sequence homology onto the structure of hMYDGF in PyMOL (Figs. 4a and 5).

Jalview v2.1 was used to generate, visualize, and create images of protein sequence alignments. A multiple sequence alignment that samples the different classes of MYDGF homologs was generated by MAFFT(L-INS-i) using protein sequences derived from UniProt (Fig. 4b; UniProtKB/UniParc IDs listed next to class). Signal peptides, either annotated in UniProt or predicted using the SignalP 5.0 server<sup>62</sup>, were omitted from this alignment. If a signal peptide could not be identified, the C-terminal truncated protein sequence identified by ConSurf was used instead. A MAFFT(L-INS-i) alignment of 236 representative UniRef90 sequences found through BLAST (mature hMYDGF input sequence, 0.0001 *E*-value cutoff) identified 25 residues with  $\geq 85\%$  sequence identity across all sequences (Fig. 4b, asterisks; black asterisks correspond to 12 residues with  $\geq 90\%$  homology).

**MYDGF and vanin base sequence and structural homology.** The Dali server<sup>27</sup> was used to identify proteins structurally similar to the hMYDGF structure. The most energetically stable conformer lacking the first five cloning residues (mature hMYDGF) was used as the query protein structure against the full PDB. The crystal structure of the human VNN1 base domain (PDB 4CYF (<https://www.rcsb.org/structure/4CYF>), residues V314–G491<sup>28</sup>) was identified as the top hit with a Dali Z-score of 7.6. All other hits had a Dali Z-score of  $< 5$  (scores  $< 2$  are considered invalid hits) and none appeared to have a complete hMYDGF fold. A total of 110 hMYDGF ordered residues (P35–A126, D133–A168) superimposed with the human VNN1 base domain with a Ca RMSD of 4.0 Å using the default settings for the “super” command in PyMOL.

The MUSCLE alignment of 82 sequences for MYDGF- and vanin-family proteins (Fig. 6b) was generated through Jalview using their corresponding seed sequences from Pfam (Pfam: PF10572 (<http://pfam.xfam.org/family/PF10572>) and Pfam: PF19018, (<http://pfam.xfam.org/family/PF19018>), respectively) in addition to the mature hMYDGF sequence (UniProtKB: Q969H8 (<https://www.uniprot.org/uniprot/Q969H8>), residues V32–L173) and the human VNN1 base domain (UniProtKB: O95497 (<https://www.uniprot.org/uniprot/O95497>), residues V325–G491). This alignment was the basis for defining the human VNN1 base domain as residues V325–G491. Inasmuch as UniProtKB annotating the enzymatic domain as residues A39–S306, we define residues H307–E324 as a linker that connects the two domains of VNN1 (Fig. 6c).

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

## Data availability

The solution structure of hMYDGF along with structural restraints was deposited in the PDB under accession code 6O6W (<https://www.rcsb.org/structure/6O6W>). NMR data were deposited in BMRB under accession number 30584 ([http://www.bmrwisc.edu/data\\_library/summary/?bmrblid=30584](http://www.bmrwisc.edu/data_library/summary/?bmrblid=30584)). MYDGF and vanin base protein families can be viewed in Pfam under accession codes PF10572 (<http://pfam.xfam.org/family/PF10572>) and PF19018 (<http://pfam.xfam.org/family/PF19018>), respectively. The lowest-energy hMYDGF/cKDEL2R complexes from HADDOCK clusters 1 and 2 were deposited in PDB-Dev under accession code PDBDEV\_00000036 (<https://pdb-dev.wwpdb.org/>). The source data underlying Figs. 2a, 3a, c, Supplementary Table 1, and Supplementary Figs. 1e, f, and 3b are provided as a Source Data file. The data that support the findings of this study are available from the corresponding author upon reasonable request.

Received: 31 August 2019; Accepted: 13 November 2019;

Published online: 09 December 2019

## References

- Bortnov, V. et al. Myeloid-derived growth factor is a resident endoplasmic reticulum protein. *J. Biol. Chem.* **293**, 13166–13175 (2018).
- Schmidt, T. et al. ProteomicsDB. *Nucleic Acids Res.* **46**, D1271–D1281 (2018).
- Raykhel, I. et al. A molecular specificity code for the three mammalian KDEL receptors. *J. Cell. Biol.* **179**, 1193–1204 (2007).
- Paroutis, P., Touret, N. & Grinstein, S. The pH of the secretory pathway: measurement, determinants, and regulation. *Physiology* **19**, 207–215 (2004).
- Capitani, M. & Saltese, M. The KDEL receptor: new functions for an old protein. *FEBS Lett.* **583**, 3863–3871 (2009).
- Bräuer, P. et al. Structural basis for pH-dependent retrieval of ER proteins from the Golgi by the KDEL receptor. *Science* **363**, 1103–1107 (2019).
- Korf-Klingebiel, M. et al. Myeloid-derived growth factor (C19orf10) mediates cardiac repair following myocardial infarction. *Nat. Med.* **21**, 140–149 (2015).
- Polten, F. et al. Plasma concentrations of myeloid-derived growth factor in healthy individuals and patients with acute myocardial infarction as assessed by multiple reaction monitoring-mass spectrometry. *Anal. Chem.* **91**, 1302–1308 (2019).
- Trychta, K. A., Bäck, S., Henderson, M. J. & Harvey, B. K. KDEL receptors are differentially regulated to maintain the ER proteome under calcium deficiency. *Cell Rep.* **25**, 1829–1840.e6 (2018).
- Wang, C., Nguyen, H. N., Maguire, J. L. & Perry, D. C. Role of intracellular calcium stores in cell death from oxygen-glucose deprivation in a neuronal cell line. *J. Cereb. Blood. Flow. Metab.* **22**, 206–214 (2002).
- Chen, B., Brown, K. A., Lin, Z. & Ge, Y. Top-down proteomics: ready for prime time? *Anal. Chem.* **90**, 110–127 (2018).
- Micsonai, A. et al. BeStSel: a web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra. *Nucleic Acids Res.* **46**, W315–W322 (2018).
- Wilson, D. W., Lewis, M. J. & Pelham, H. R. B. pH-dependent binding of KDEL to its receptor in vitro. *J. Biol. Chem.* **268**, 7465–7468 (1993).
- Ulrich, E. L. et al. BioMagResBank. *Nucleic Acids Res.* **36**, D402–D408 (2008).
- Berman, H., Henrick, K. & Nakamura, H. Announcing the worldwide Protein Data Bank. *Nat. Struct. Mol. Biol.* **10**, 980–980 (2003).
- wwPDB Consortium. Protein Data Bank: the single global archive for 3D macromolecular structure data. *Nucleic Acids Res.* **47**, D520–D528 (2019).
- Rossi, P. et al. A microscale protein NMR sample screening pipeline. *J. Biomol. NMR* **46**, 11–22 (2010).
- Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. & Baker, N. A. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* **32**, W665–W667 (2004).
- Søndergaard, C. R., Olsson, M. H. M., Rostkowski, M. & Jensen, J. H. Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pKa values. *J. Chem. Theory Comput.* **7**, 2284–2295 (2011).
- Olsson, M. H. M., Chresten, R. S. øndergaard, Rostkowski, M. & Jensen, J. H. PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. *J. Chem. Theory Comput.* **7**, 525–537 (2011).
- van Zundert, G. C. P. et al. The HADDOCK2.2 web server: user-friendly integrative modeling of biomolecular complexes. *J. Mol. Biol.* **428**, 720–725 (2016).
- de Vries, S. J. et al. HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets. *Proteins Struct. Funct. Bioinform.* **69**, 726–733 (2007).
- Dominguez, C., Boelens, R. & Bonvin, A. M. J. J. HADDOCK: a protein–protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **125**, 1731–1737 (2003).
- Raffaello, A., Mammucari, C., Gherardi, G. & Rizzuto, R. Calcium at the center of cell signaling: interplay between endoplasmic reticulum, mitochondria, and lysosomes. *Trends Biochem. Sci.* **41**, 1035–1049 (2016).
- Ashkenazy, H. et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44**, W344–W350 (2016).
- Tian, W., Chen, C., Lei, X., Zhao, J. & Liang, J. CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Res.* **46**, W363–W367 (2018).
- Holm, L. Benchmarking fold detection by DaliLite v.5. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/btz536> (2019).
- Boersma, Y. L. et al. The structure of vanin 1: a key enzyme linking metabolic disease and inflammation. *Acta Crystallogr. D Biol. Crystallogr.* **70**, 3320–3329 (2014).
- Maras, B., Barra, D., Duprè, S. & Pitari, G. Is pantetheinase the actual identity of mouse and human vanin-1 proteins? *FEBS Lett.* **461**, 149–152 (1999).
- Bartucci, R., Salvati, A., Olinga, P. & Boersma, Y. L. Vanin 1: its physiological function and role in diseases. *Int. J. Mol. Sci.* **20**, 3891 (2019).
- Mariani, F. & Roncucci, L. Role of the vanins-myeloperoxidase axis in colorectal carcinogenesis. *Int. J. Mol. Sci.* **18**, 918 (2017).
- Yuan, Z. et al. Injectable citrate-based hydrogel as an angiogenic biomaterial improves cardiac repair after myocardial infarction. *ACS Appl. Mater. Interfaces* **11**, 38429–38439 (2019).
- Schaeffer, R. D., Liao, Y. & Grishin, N. V. Searching ECOD for homologous domains by sequence and structure. *Curr. Protoc. Bioinformatics* **61**, e45 (2018).
- Ni, M., Zhang, Y. & Lee, A. S. Beyond the endoplasmic reticulum: atypical GRP78 in cell viability, signalling and therapeutic targeting. *Biochem. J.* **434**, 181–188 (2011).
- Marzec, M., Eletto, D. & Argon, Y. GRP94: an HSP90-like protein specialized for protein folding and quality control in the endoplasmic reticulum. *Biochim. Biophys. Acta Mol. Cell Res.* **1823**, 774–787 (2012).

36. Glembofski, C. C. et al. Mesencephalic astrocyte-derived neurotrophic factor protects the heart from ischemic damage and is selectively secreted upon sarco/endoplasmic reticulum calcium depletion. *J. Biol. Chem.* **287**, 25893–25904 (2012).
37. Henderson, M. J., Richie, C. T., Airavaara, M., Wang, Y. & Harvey, B. K. Mesencephalic astrocyte-derived neurotrophic factor (MANF) secretion and cell surface binding are modulated by KDELR receptors. *J. Biol. Chem.* **288**, 4209–4225 (2013).
38. Park, S.-J. et al. Discovery of endoplasmic reticulum calcium stabilizers to rescue ER-stressed podocytes in nephrotic syndrome. *Proc. Natl Acad. Sci. USA* **116**, 14154–14163 (2019).
39. Mosher, D. F., Huwiler, K. G., Misenheimer, T. M. & Annis, D. S. Expression of recombinant matrix components using baculoviruses. *Methods Cell Biol.* **69**, 69–81 (2002).
40. Maurer, L. M. et al. Extended binding site on fibronectin for the functional upstream domain of protein F1 of *Streptococcus pyogenes*. *J. Biol. Chem.* **285**, 41087–41099 (2010).
41. Cai, W. et al. MASH suite pro: a comprehensive software tool for top-down proteomics. *Mol. Cell. Proteom.* **15**, 703–714 (2016).
42. Wiedemann, C., Bellstedt, P. & Görlach, M. CAPITO—A web server-based analysis and plotting tool for circular dichroism data. *Bioinformatics* **29**, 1750–1757 (2013).
43. Studier, F. W. Protein production by auto-induction in high-density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).
44. Marley, J., Lu, M. & Bracken, C. A method for efficient isotopic labeling of recombinant proteins. *J. Biomol. NMR* **20**, 71–75 (2001).
45. Fox, B. G. & Blommel, P. G. Autoinduction of protein expression. *Curr. Protoc. Protein Sci.* **56**, 5.23.1–5.23.18 (2009).
46. Delaglio, F. et al. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
47. Hyberts, S. G., Takeuchi, K. & Wagner, G. Poisson-gap sampling and forward maximum entropy reconstruction for enhancing the resolution and sensitivity of protein NMR data. *J. Am. Chem. Soc.* **132**, 2145–2147 (2010).
48. Ying, J., Delaglio, F., Torchia, D. A. & Bax, A. Sparse multidimensional iterative lineshape-enhanced (SMILE) reconstruction of both non-uniformly sampled and conventional NMR data. *J. Biomol. NMR* **68**, 101–118 (2017).
49. Lee, W., Tonelli, M. & Markley, J. L. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325–1327 (2015).
50. Lee, W. et al. I-PINE web server: an integrative probabilistic NMR assignment system for proteins. *J. Biomol. NMR* **73**, 213–222 (2019).
51. Lee, W. & Markley, J. L. PINE-SPARKY.2 for automated NMR-based protein structure research. *Bioinformatics* **34**, 1586–1588 (2018).
52. Dashti, H. et al. Probabilistic validation of protein NMR chemical shift assignments. *J. Biomol. NMR* **64**, 17–25 (2016).
53. Sharma, D. & Rajarathnam, K. <sup>13</sup>C NMR chemical shifts can predict disulfide bond formation. *J. Biomol. NMR* **18**, 165–171 (2000).
54. Lee, W. et al. Integrative NMR for biomolecular research. *J. Biomol. NMR* **64**, 307–332 (2016).
55. Güntert, P. Automated NMR structure calculation with CYANA. in *Protein NMR Techniques* 353–378 (Humana Press, 2004).
56. Lee, W., Petit, C. M., Cornilescu, G., Stark, J. L. & Markley, J. L. The AUDANA algorithm for automated protein 3D structure determination from NMR NOE data. *J. Biomol. NMR* **65**, 51–57 (2016).
57. Lee, W., Stark, J. L. & Markley, J. L. PONDEROSA-C/S: client–server based software package for automated protein 3D structure determination. *J. Biomol. NMR* **60**, 73 (2014).
58. Shen, Y. & Bax, A. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J. Biomol. NMR* **56**, 227–241 (2013).
59. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).
60. Lin, D., Hung Sze, K., Cui, Y. & Zhu, G. Clean SEA-HSQC: a method to map solvent exposed amides in large non-deuterated proteins with gradient-enhanced HSQC. *J. Biomol. NMR* **23**, 317–322 (2002).
61. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259 (2019).
62. Almagro Armenteros, J. J. et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* **37**, 420–423 (2019).

## Acknowledgements

We thank and acknowledge the support of William Milo Westler at NMRFAM for his input on the hMYDGF NMR structure; Ronnie Frederick for offering his protocols, expertise, and reagents for producing isotopically labeled protein; and Darrell McCaslin for offering advice on conducting CD experiments and interpreting results. V.B., D.S.A., and D.F.M. were supported by NIH grants P01 HL088594 and R01 AI125390. M.T., W.L., and J.L.M. were supported by NIH grant P41 GM103399. This study made use of the National Magnetic Resonance Facility at Madison, which is supported by NIH grant P41 GM103399, formerly P41 RR002301. Equipment was purchased with funds from the University of Wisconsin-Madison, the NIH (P41 GM103399, S10 RR002781, S10 RR008438, S10 RR023438, S10 RR025062, S10 RR029220), and the NSF (DMB-8415048, OIA-9977486, BIR-9214394). Z.L. and Y.G. acknowledge top-down proteomics software grant R01 GM125085 and high-end instrument grant S10 OD018475.

## Author contributions

D.S.A. cloned the human MYDGF gene into expression vectors and expressed and purified insect cell-derived hMYDGF. V.B. expressed and purified non-labeled and isotopically labeled, bacteria-derived hMYDGF, conducted C.D. and intrinsic tryptophan experiments, quantified resonance perturbation results, and generated phylogenetic trees. A.B. carried out sequence analysis and updated Pfam based on finding regarding MYDGF and vanin base homology. V.B. and A.B. generated multiple sequence alignments. M.T. collected and processed all NMR data, and performed relaxation experiment calculations; V.B. assigned and validated NMR resonances and analyzed N.O.E. constraints; V.B. and W.L. conducted structure calculation and validation. M.T., W.L., V.B., and J.L.M. assessed and validated NMR results. Z.L. collected and processed mass spectrometry experiments; V.B., Z.L., and Y.G. analyzed and validated mass spectrometry results. J.C.M. and O.N.D. carried out computational docking of hMYDGF onto cKDELRL2. V.B. and D.F.M. planned the experiments and wrote the manuscript. All authors reviewed and edited the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41467-019-13577-5>.

Correspondence and requests for materials should be addressed to D.F.M.

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

Reprints and permission 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) 2019