Electronic Supplementary Information (ESI) for Molecular BioSystems

Structural characterization and biological properties of human gastrokine 1.

Luigi Michele Pavone[a,d], Pompea Del Vecchio*[b], Paolo Mallardo[a], Filomena Altieri[a], Valeria De Pasquale[a], Silviana Rea[a], Nicola M. Martucci[a,c], Chiara Stella Di Stadio[a], Pietro Pucci[b,d], Angela Flagiello[b,d], Mariorosario Masullo[a,c], Paolo Arcari*[a,d], Emilia Rippa[a]

[a]*Department of Biochemistry and Medical Biotechnologies, University of Naples Federico II, Via S. Pansini 5, 80131 Naples, Italy;*

[b]*Department of Chemical Sciences , University of Naples Federico II, Via Cintia, 80126 Naples, Italy;*

[c]*Department of DiSIST, University of Naples "Parthenope", Via Medina 40, 80133, Naples, Italy*

[d]*CEINGE Advanced Biotechnology scarl, Via Gaetano Salvatore 486, 80145 Naples, Italy*

Supplementary Experimental Procedure

Vector construction and *Pichia pastoris* cell transformation

pCDNA3.1 containing the GKN1 cDNA (pCDNA3.1/GKN1)[1] was used as a template for PCR amplification. The following primers were used: PIC-GKF1 (forward) 5'TACGTAATGAAGTTCACAATTGTCTT3' corresponding to M1KFTIVF of full length GKN1, PIC-GKF2 (forward) 5'TACGTAAACTATAATATCAACGTCAAT3' corresponding to the Y2NYNINVN sequence of mature GKN1 (truncated of the first 20 leader amino acids) and PIC-GKR1 (reverse) 5'GCGGCCGCTCAATGGTGATGGTGATG3' corresponding to the $(His)_6$ C-terminal region of the pCDNA3.1/GKN1 construct and containing SNAB1 and NOT1 restriction sites, respectively. The PCR (GKN1A and GKN1B) products were then cloned in pPIC9K digested with the same restriction enzymes, and the isolated positive clones were sequenced (pPIC9K/GKN1).

P. Pastoris (strain GS115) competent cells were prepared as it follows: cells were inoculated into 10 ml of Yeast Peptone Dextrose Medium (YPD) (Invitrogen), and grown overnight at 30 °C in a shaking bath. Subsequently, the inoculum was diluted in 100 ml of YPD to a starting cell concentration of 0.1 OD660nm. 100 ml of culture were grown at 30 °C to a concentration of 0.7 OD660nm. Cell culture was harvested by centrifugation at 4500 rpm for 20 min, and washed in 50 ml of solution A [1 M sorbitol (Fisher), 10 mM bicine (Sigma), pH 8.35, 3% (v: v) ethylene glycol (Merck)]. Cells were re-suspended in 4 ml of solution A, and divided into aliquots of 0.2 ml in sterile Eppendorf tubes of 1.5 ml. 1l ml of DMSO were added in each aliquot. Competent cells were frozen at -80 °C. *P. pastoris* competent cells were transformed according to the following procedure. 50 mg of pPIC9K/GKN1, linearized with SAL1, in a final volume of 20 ml were used. The DNA was added to competent cells still frozen together with salmon sperm (Roche) as DNA carrier. The sample was incubated at 37 °C in a water bath for 5 min under shaking. 1.5 ml of solution B [40% (w: v) polyethylene glycol 1000 (Sigma), bicine 0.2 M, pH 8.35] were added to the samples that were incubated for 1 h at 30 °C, centrifuged at 3000 rpm for 10 min, and re-suspended in 1.5 ml of solution C (0.15 m NaCl, 10 mM bicine, pH 8.35). After a final centrifugation, the pellet was re-suspended in 0.2 ml of solution C. The cells were plated on agarose plates containing selective His^- culture medium (Invitrogen), and grown for 2 days at 30 °C.

GKN1 expression and purification

Analytical expression was first carried out to identify the best GKN1-expressing clone. *P. pastoris* transformed colonies were inoculated into 10 ml of BMMY. The culture was incubated at 30 °C by vigorous shaking for 72 h. Cells were then harvested by centrifugation at 3500 rpm for 15 min at 4°C, and the supernatant of each clone was analyzed by SDS-PAGE electrophoresis and Western blotting using a mouse monoclonal anti-GKN1 antibody (2E5 clone from Abnova). GKN1A was not expressed. In contrast, the mature form of the protein, GKN1B, was expressed in all examined clones. The clone showing the highest GKN1 expression (clone 27) was grown, and stored at -80 °C. Preparative GKN1 expression was performed in 2 l of culture medium. Briefly, 100 ml of *P. pastoris* clone 27 were inoculated in 12 ml of BMGY, and grown overnight at 30 °C. 3 ml of this culture were then used to inoculate 500 ml of BMGY, and grown overnight up to a cell concentration of 5 OD 600nm. Cells were harvested by centrifugation, re-suspended in a volume of 2 l of BMMY, and grown for 72 h at 30 °C under vigorous shaking. After the first 24 h, methanol (0.5% final concentration) was added to the growing culture every day. Finally, cells were harvested by centrifugation at 8000 rpm for 40 min, and the resulting supernatant was cleaned using 0.22 mm filters, and stored at 4 °C.

Proteins from supernatant were precipitated with 70% ammonium sulfate (844 g/2 l) at 4°C, centrifuged at 13000 rpm for 50 min, and the resulting pellet was re-suspended in 50 ml of buffer A (50 mM TrisHCl, pH 8.5). The protein sample was then dialyzed against 2 l of buffer A, and applied on anion exchange DEAE Sepharose (1x60cm) (FF, Pharmacia Biotech). The column, equilibrated at 4°C in buffer A, was washed with 10 x volume column, and eluted with a linear gradient (0 - 500 mM NaCl) at a flow rate of 2 ml/min (total volume: 10x column volume). The elution profile was followed at 280 nm, and fractions (12 ml) were then analyzed by SDS-PAGE. Fractions containing GKN1 were pooled, dialyzed against buffer A, and applied on a Ni-NTA Agarose column (10 ml). After washing the column with buffer A (10x column volume), the column was washed with 5x column volume of buffer A containing 5 mM imidazole, and subsequently the protein was eluted with buffer A containing 100 mM imidazole. Fractions were analyzed by SDS-PAGE and Western blotting, and those fractions containing purified GKN1 were pooled, dialyzed against buffer A or 20 mM phosphate buffer, pH 7.0, and stored at –20°C.

Mass spectrometry

GKN1 was analyzed on a C8 column by High Performance Liquid Chromatography (HPLC) coupled with a LCMS EV 2010 Liquid Chromatograph Mass Spectrometer (Shimadzu). Sample (30 □l) was first eluted at 0.5 ml/min with a linear gradient from 5 to 40% acetonitrile (Romil) in water containing 0.1% trifluoroacetic acid (TFA) (Applied Biosystems) for 50 min, and then from 40 to 95% acetonitrile in water containing 0.1% TFA for 20 min.

Supplementary Results

Expression and purification of recombinant GKN1

The expressed recombinant GKN1 protein, secreted into culture medium, was identified after ammonium sulfate precipitation (70%) by mass spectrometry coupled to HPLC. As reported in Fig. S1, the HPLC chromatogram showed a peak of GKN1 accounting for about 2% of total proteins. The peak was collected, and the presence of GKN1 was confirmed by electrospray mass spectrometry analysis. After the ammonium sulfate precipitation, GKN1 was further purified by DEAE and Ni-NTA chromatography. Fig. S2A shows the DEAE chromatogram of GKN1 eluted by a linear salt gradient. Fractions containing the recombinant protein were analyzed by SDS-PAGE (Fig. S2B), and the appropriate pooled fractions were then separated on Ni-NTA agarose chromatography which was eluted by 100 mM imidazole. Finally, the purified protein was concentrated up to about 2-3 mg/ml, and stored in 20 mM phophate buffer, pH 7.4 at –20°C. Fig. S3 reports the SDS-PAGE electrophoretic profile of the purified GKN1.

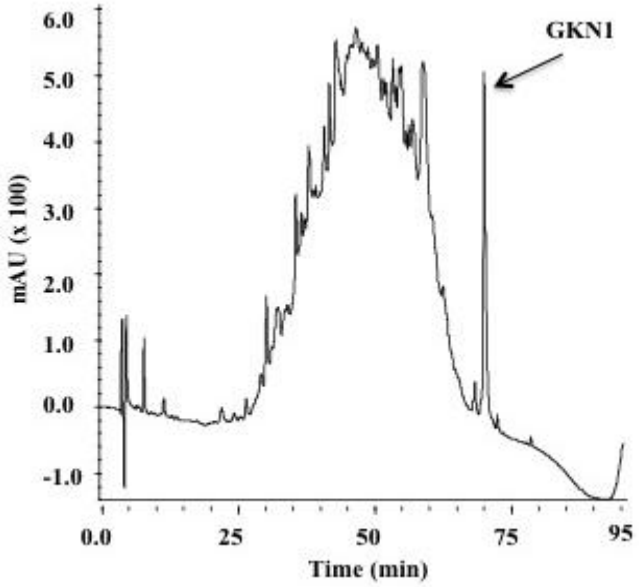

Fig. S1. Analysis by HPLC of *P. pastoris* culture medium.

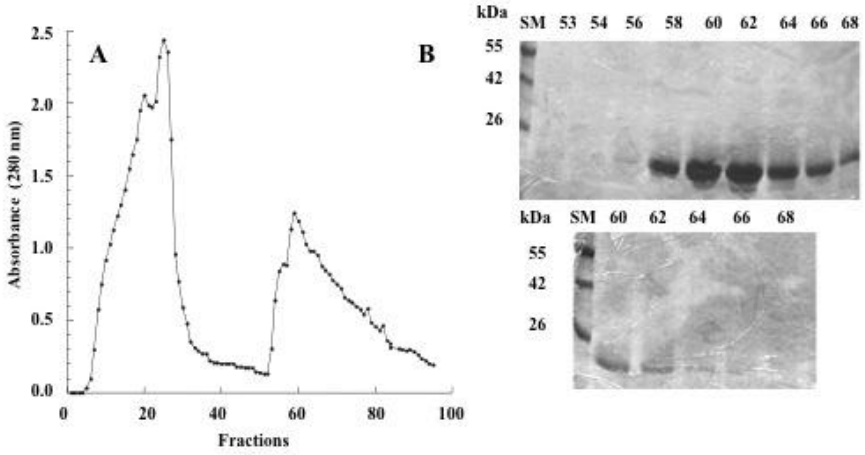

Fig. S2. Purification of GKN1. (A) DEAE chromatography after ammonium sulphate precipitation. (B) SDS-PAGE of DEAE fraction containing GKN1.

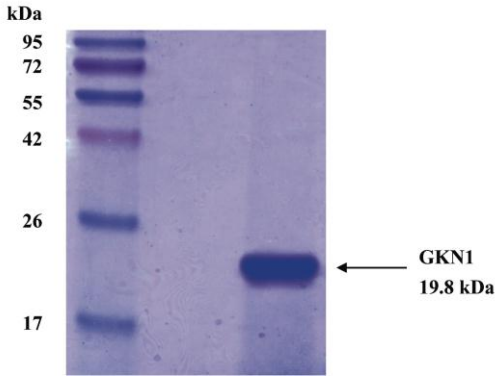

Fig. S3. SDS-PAGE of the purified recombinant GKN1.

Limited proteolysis of GKN1

Cleavage of GKN1 with trypsin led to the formation of two major fragments of about 17 and 11 kDa (Fig. S4A). The formation of proteolytic products was accompanied by the disappearance of intact recombinant GKN1 (Fig. S4B). Edman degradation of the largest protein fragment (17 kDa) showed the N-terminal sequence of the intact protein ($Y_{-2}VN_1YN$..), thus indicating the presence of a proteolytic cleavage site toward the C-terminal region; the N-terminal sequence of 11 kDa fragment (G_{90}PGGP..) showed the presence of a proteolytic site at the level of Lys89 (numbering mature GKN1).

Limited proteolysis of GKN1 with chimotrypsin (Fig. S4C) was accompanied by the formation of three major peptides of about 15, 10 and 5 kDa (Fig. S4D). The N-terminal sequence of the 15 kDa fragment (N_3INVN..) started at level of Tyr2, thus suggesting the occurrence of a further proteolytic cleavage site at the C-terminal region of the protein. The N-terminal sequence of the 10 kDa fragment (F_{140}YSGT..) indicated the presence of a cleavage site at level of Phe139. Also in this case, the molecular weight of this fragment suggests the presence of an additional cleavage site at the C-terminal. The third 5 kDa fragment gave two different N-terminal sequences (K_{65}MNKE.. and S_{102}VNPN..), indicating the presence of two proteolytic sites at level of His64 and Tyr101, respectively. Also for these two fragments, their size suggests the presence of an additional cleavage site at the C-terminal.

Following the digestion of GKN1 with thermolysin (Fig. S4E), four fragments of about 18, 16, 11 and 9 kDa were observed by SDS-PAGE (Fig. S4F). The first two fragments showed the same N-terminal sequences (I_4NVN..), indicating that both fragments were formed by cleavage of GKN1 at

position of Asn3 and at two different cleavage positions in the C-terminal region of the protein. The fragment of 11 kDa produced two N-terminal sequences (L_{86}QGKG.. and F_{140}YSGT..) indicating cuts at level of Lys85 and Phe139, respectively. On the basis of the sizes of these two fragments, the presence of an additional cleavage site at the C-terminal was supposed. From the 9 kDa band, three N-terminal sequences were found (L_{86}QGKG.., F_{140}YSGT.. and Y_{141}SGT..) corresponding to cleavage at level of Lys85, Phe139 and Phe140, respectively. Also in this case, additional cleavage sites are likely present at the C-terminal region of each fragment.

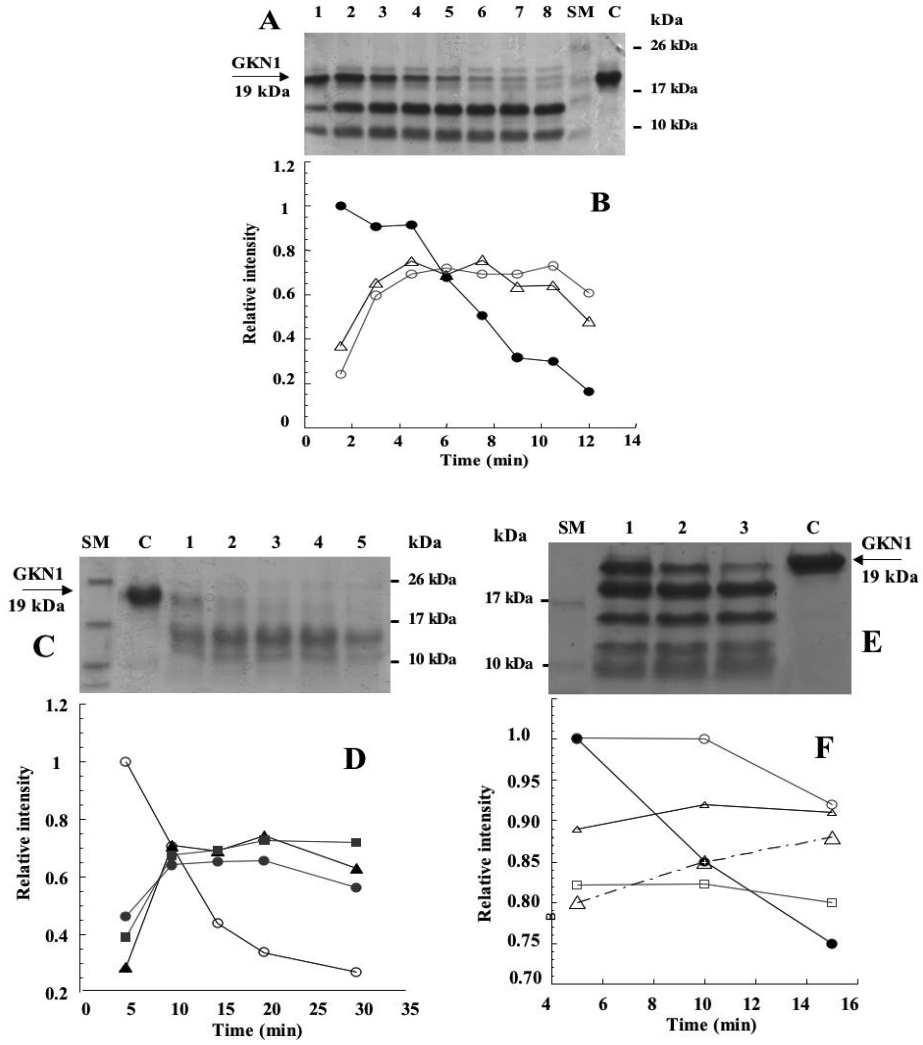

Fig. S4. Limited proteolysis of GKN1 with trypsin, chimotrypsin and thermolysin. GKN1 was incubated in the presence of trypsin (A), chimotrypsin (C) and thermolysin (E) as reported ion Materials and Methods and at the indicated times, aliquots were withdrawn from the reaction mixture and separated on SDS-PAGE. (A) Lanes: 1-8, samples after 1.5, 3.0, 4.5, 6.0, 7.5, 9.0, 10.5, 12.0 minutes incubation with trypsin; SM, size marker; C, intact GKN1 (●); 17 kDa fragment (○); 11 kDa fragment (△). (C) Lanes: SM, size marker; C, intact GKN1 (○); 1-5, samples after 5, 10, 15, 20 and 30 minutes incubation with chimotrypsin; 15 kDa fragment (■); 10 kDa fragment (▲); 5 kDa fragment (●). (E) Lanes: SM, size marker; C, intact GKN1 (●);1-3, samples after 5, 10 and 15 minutes incubation with thermolysin; 18 kDa fragment (○);16 kDa fragment (△); 11 kDa fragment (△). (B, D and F) Relative densitometric evaluation of the band intensity present in (A), (C) and (E), respectively.

Trypsin digestion of GKN1 in the absence or presence of DTT

The digestion products of GKN1 after trypsin digestion were then separated on a SDS-PAGE in non-reducing condition. The results reported in supporting Fig. S5 showed that in the presence of 1 mM DTT, GKN1 trypsin digestion gave rise to the two proteolytic products (17 and 11 kDa), whereas, in the absence of DTT, no fragments were produced.

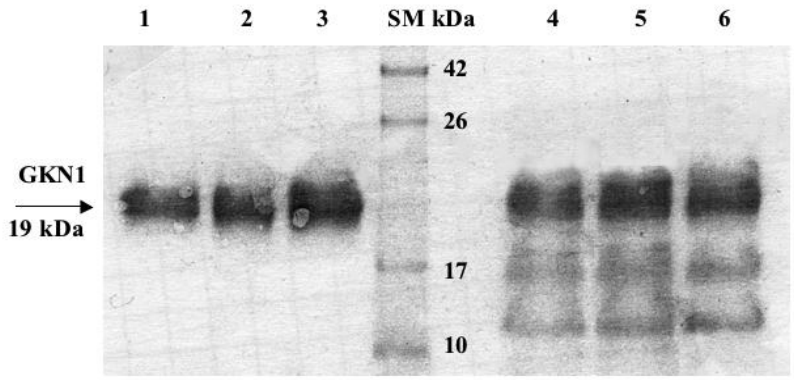

Fig. S5. Proteolysis of GKN1 in absence and presence of DTT. GKN1 was incubated in the presence of trypsin and at the indicated times, sample aliquots were withdrawn from the reaction mixture and separated on non-reducing SDS-PAGE. Lanes: 1-3, sample after 5, 10 and 15 minutes incubation in absence of DTT; 4 – 6, sample after 5, 10 and 15 minutes incubation in presence of 1 mM DTT.

GKN1 thermal stability

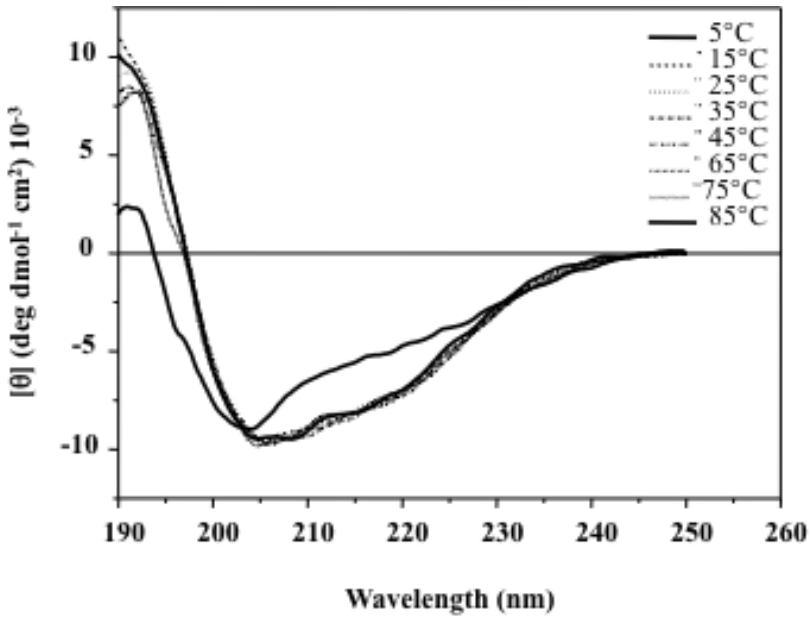

Fig. S6. Far-UV CD spectra of GKN1 recorded in the temperature interval 5°C – 85°C.

Supplementary GKN1 structure prediction

Prediction of GKN1 secondary structure

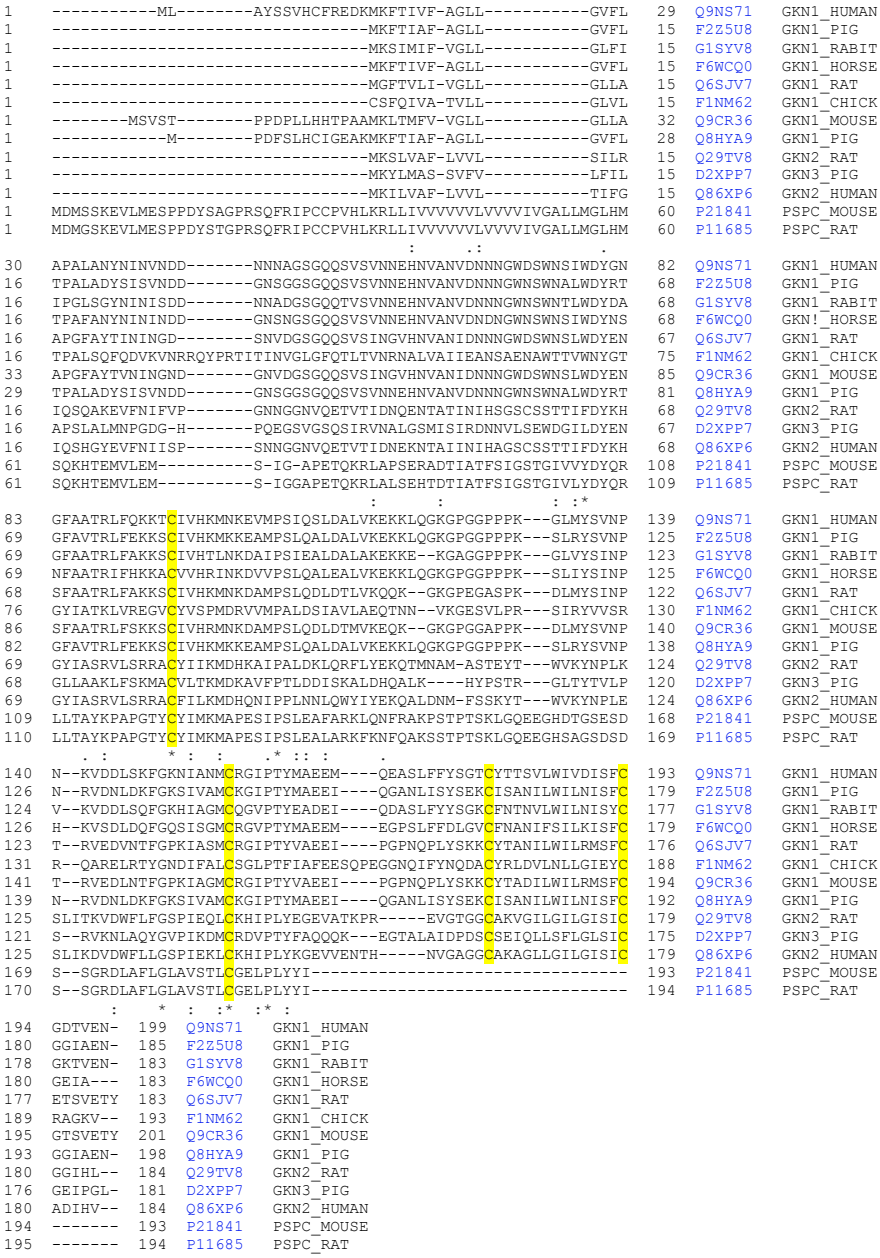

Fig. S7. Multiple alignment of GKN1 homologous proteins. GKN1 homologous, including pulmonary surfactant-associated protein C (PSPC) sequences were selected and aligned at the Uniprot server. Sequence legend**:** Homo sapiens (Human), Sus scrofa (Pig), Oryctolagus cuniculus (Rabbit,) Equus caballus (Horse), Rattus norvegicus (Rat), Gallus gallus (Chicken), Mus musculus (Mouse). In yellow are highlighet the conserved cysteine residues.

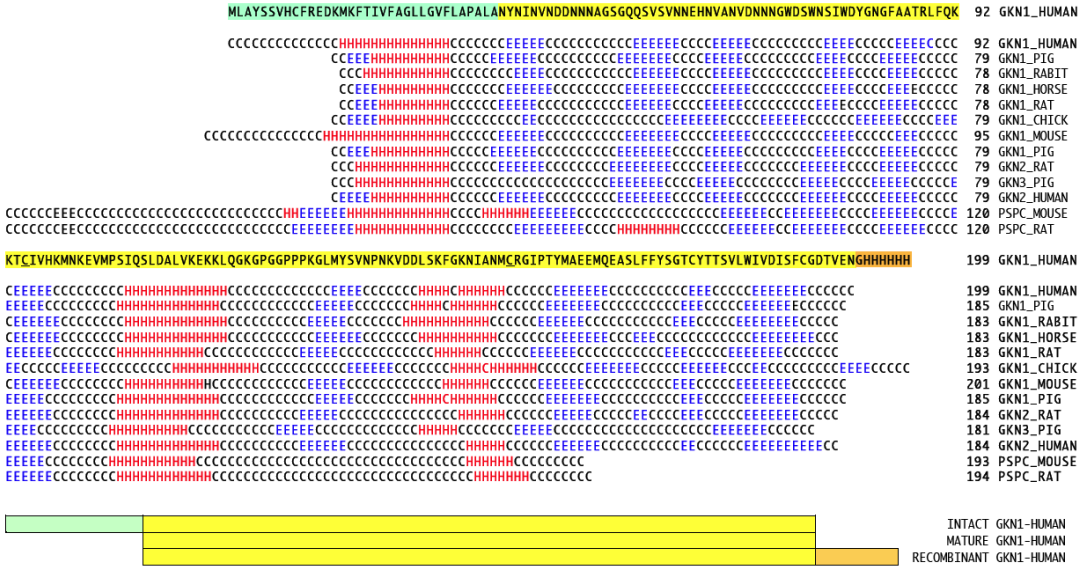

Fig. S8. Predicted secondary structure comparison between GKN1 homologous proteins. GKN1 and PSPC sequences were selected at the Uniprot database and their predicted secondary structure, obtained at PSIPRED Protein Structure Prediction Server, were compared on the basis of the sequence alignment reported in Fig. S7. Sequence legend**:** Homo sapiens (Human), Sus scrofa (Pig), Oryctolagus cuniculus (Rabbit,) Equus caballus (Horse), Rattus norvegicus (Rat), Gallus gallus (Chicken), Mus musculus (Mouse). H, α-helix; E, β-strand and C, coil. The upper line shows the amino acid sequence of human GKN1 (intact, mature or recombinant) according to the representation reported at the bottom.

Prediction of GKN1 tertiary structure

I-Tasser estimated accuracy of Model1-3: 0.29±0.09 (TM-score) 14.5±3.7Å (RMSD).

Table S1. Top 10 templates used by I-TASSER

Rank	*PDB Hit*	*Iden 1*	*Iden 2*	*Conv*	*Norm. Z-score*
1	2yadA	0.20	0.10	0.43	1.04
2	2yadA	0.19	0.10	0.43	1.50
3	2yadA	0.18	0.10	0.43	2.09
4	3v89A	0.10	0.21	0.96	0.67
5	1vjhA	0.07	0.08	0.49	0.54
6	2l69A	0.18	0.16	0.76	0.76
7	1vybA	0.07	0.19	0.97	0.86
8	2qs8A	0.08	0.17	0.97	0.84
9	1vjhA	0.10	0.08	0.50	0.69
10	2axpA	0.11	0.21	0.92	0.49

1. BRICHOS domain of Surfactant protein C precursor protein; 2. BRICHOS domain of Surfactant protein C precursor protein; 3. BRICHOS domain of Surfactant protein C precursor protein; 4. The crystal structure of transferrin binding protein A (TbpA) from Neisseria meningitides serogroup B in complex with the C-lobe of human transferrin; 5. Crystal structure of gene product of At1g24000 from Arabidopsis thaliana; 6. Solution NMR Structure of de novo designed protein, P-loop NTPase fold, Northeast Structural Genomics Consortium Target OR28; 7. Endonuclease domain of human line1 orf2p; 8. Crystal structure of a Xaa-Pro dipeptidase with bound methionine in the active site; 9. Crystal structure of gene product of At1g24000 from Arabidopsis thaliana; 10. X-Ray Crystal Structure of Protein BSU20280 from Bacillus subtilis. Northeast Structural Genomics Consortium Target SR256

Rank of templates represents the top ten threading templates used by I-TASSER.[2,3] Ident1 is the percentage sequence identity of the templates in the threading aligned region with the query sequence. Ident2 is the percentage sequence identity of the whole template chains with query sequence. Cov. represents the coverage of the threading alignment and is equal to the number of aligned residues divided by the length of query protein. Norm. Z-score is the normalized Z-score of the threading alignments. Alignment with a Normalized Z-score >1 mean a good alignment and vice versa.

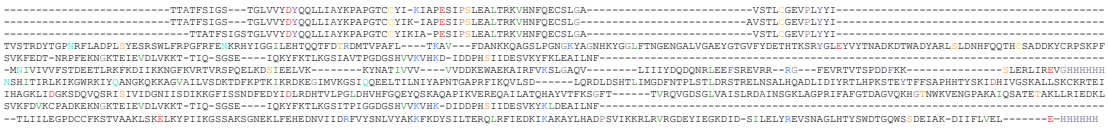

Fig. S9. The top 10 alignments reported above (in order of their ranking) are from the following threading programs.
1: SP3 2: PPA-I 3: SAM T99 4: MUSTER 5: HHSEARCH 6: SP3 7: PROSPECT2 8: PPA-I 9: HHSEARCH I 10: SPARKS.
All the residues are colored in black; however, those residues in template which are identical to the residue in the query sequence are highlighted in color. Coloring scheme is based on the property of amino acids, where polar are brightly coloured while non-polar residues are colored in dark shade.

Molprobity analysis

Both pdb files obtained by I-Tasser[2,3] and Swiss Model[4] (model 1-3.pdb and model 1.pdb, respectively) were uploaded to the Molprobity server[5,6] and after addition to these models of hydrogens and allowing Asn/Gln/His flips, new improved structures were generated (model 1-3FH.pdb and model 1FH.pdb, respectively). The analysis of the results evaluated in terms of poor rotamers, Ramachandran outliers, Ramachandran favored, Cβ deviations, MolProbity score, residues with bad bonds, residues with bad angles are reported in Table S2.

Table S2. Molprobity analysis output: all-atom contacts and geometry for GKN1 models

		Model 1-3FH.pdb		**Model 1FH.pdb**	
All-Atom Contacts	Clashscore, all atoms:	148.66	0th percentile* (N=1784, all resolutions)	105.56	0th percentile* (N=1784, all resolutions)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.				
Protein Geometry	Poor rotamers	2.68%	Goal: <1%	3.45%	Goal: <1%
	Ramachandran outliers	8.82%	Goal: <0.2%	2.00%	Goal: <0.2%
	Ramachandran favored	78.82%	Goal: >98%	90.00%	Goal: >98%
	Cβ deviations >0.25Å	13	Goal: 0	5	Goal: 0
	MolProbity score^	3.71	10th percentile* (N=27675, 0Å - 99Å)	3.44	10th percentile* (N=27675, 0Å - 99Å)
	Residues with bad bonds:	0.00%	Goal: 0%	2.94%	Goal: 0%
	Residues with bad angles:	8.72%	Goal: <0.1%	8.82%	Goal: <0.1%
		By adding H to this model and allowing Asn/Gln/His flips, 4 bad rotamers have been fixed and clashscore improved by 3.06 points.		By adding H to this model and allowing Asn/Gln/His flips, clashscore was improved by 0.63 points.	

*100th percentile is the best among structures of comparable resolution; 0th percentile is the worst.
^MolProbity score is defined as the following: 0.42574*log(1+clashscore)+ 0.32996*log(1+max(0,pctRotOut-1)) + 0.24979*log(1+max(0,100-pctRamaFavored-2)) + 0.5.

Fig. S8 shows the improved structures obtained from Molprobity. Model 1-3FH included also the BRICHOS domain of surfactant protein C precursor (proSP-C) (pdb 2yadA).[7] The model with highest C-score (-3.97) showed a content of β-sheet (29%) and α-helix (17%) similar to that predicted by CD spectroscopy. The model possessed structural features resembling those present in the 3D structure of proSP-C BRICHOS domain. In fact, the N-terminal region of GKN1 showed the presence of five consecutive strands forming an anti-parallel β–sheet structure (β1-β5, Fig. S8A) followed by two helices almost opposite to each side of the β-sheet structure (α1, α2; Fig. S8A). The β-sheet could denote two

faces “face A” that packs against helix 1, and “face B” for the face packing against helix 2. Moreover, some aromatic amino acids (W and Y) appear to be exposed to the solvent in agreement to the fluorescence emission spectrum (Fig. 6). Model 1FH instead was confined to the region corresponding to the BRICHOS domain of GKN1. The model possessed structural characteristic closer to that of its template *i.e.* four consecutive β-strand (β1-β4, Fig. 7B) facing the two α-helixs segments (α1, α2; Fig 7B).

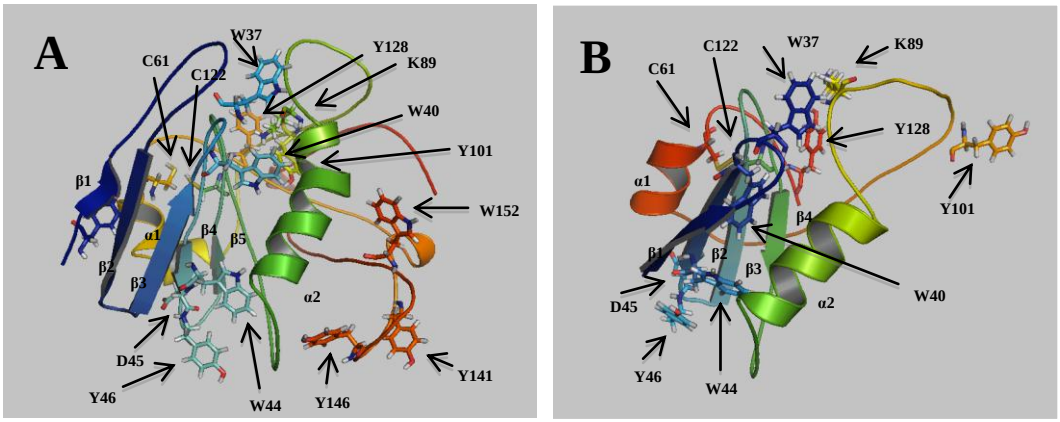

Fig. S10. Imitation of 3D structure of GKN1. Both model 1-3FH (A) and model 1FH (B) were visualized using PyMol. Arrows indicate the position of conserved cysteines and aromatic amino acids.

Supplementary references

1 E. Rippa, G. La Monica, R. Allocca, M.F. Romano, M. De Palma and P. Arcari, Over-expression of gastrokine 1 in gastric cancer cells induces fas-mediated apoptosis. *J. Cell. Physiol.*, 2011, 226(10), 2571-2578.

2 R. Ambrish, K. Alper and Z. Yang, I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols*, 2010, 5(4), 725-738.

3 Z. Yang, I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, 2008, 9, 40.

4 K. Arnold, L. Bordoli, J. Kopp and T. Schwede, The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. *Bioinformatics*, 2006, 22(2), 195-201.

5 V.B. Chen, W.B. Arendall, J.J. Headd, D.A. Keedy, R.M. Immormino, G.J. Kapral, L.W. Murray, J.S. Richardson and D.C. Richardson, MolProbity: all-atom structure validation for macromolecular crystallography, *Acta Crystallographica*, 2010, D66: 12-21.

6 I.W. Davis, A. Leaver-Fay, V.B. Chen, J.N. Block, G.J. Kapral, X. Wang, L.W. Murray, W.B. Arendall, J. Snoeyink, J.S. Richardson and D.C. Richardson, MolProbity: all-atom contacts and structure validation for proteins and nucleic acids, *Nucleic Acids Research*, 2007, 35(2), W375-W383.

7 H. Willander, G. Askarieh, M. Landreh, P. Westermark, K. Nordling, H. Keränen, E. Hermansson, A. Hamvas, L.M. Nogee, T. Bergman, A. Saenz, C. Casals, J. Åqvist, H. Jörnvall, H. Berglund, J. Presto, S.D. Knight and J. Johansson, High-resolution structure of a BRICHOS domain and its implications for anti-amyloid chaperone activity on lung surfactant protein C. *Proc. Natl. Acad. Sci. USA*, 2012, 109, 2325-2329.