Journal of Biomolecular NMR (2006) 35: 73–78
DOI 10.1007/s10858-006-9013-1

NMR structure note

Second Kunitz-type protease inhibitor domain of the human WFIKKN1 protein

Edvards Liepinsh[a,b], Alinda Nagy[c], Mária Trexler[c], László Patthy[c] & Gottfried Otting[d,*]
[a]*Department of Medical Biochemistry and Biophysics, Karolinska Institute, S-17177, Stockholm, Sweden;* [b]*Institute of Organic Synthesis, LV-1006, Riga, Latvia;* [c]*Institute of Enzymology, Biological Research Center, Hungarian Academy of Sciences, H-1113, Budapest, Hungary;* [d]*Research School of Chemistry, Australian National University, Canberra, ACT, 0200, Australia*

Received 2 February 2006; Accepted 22 March 2006

Key words: human WFIKKN1, Kunitz domain, NMR structure

Biological context

Two closely related multi-domain proteins, WFIKKN and WFIKKNRP (recently renamed to WFIKKN1 and WFIKKN2), have been described that contain a WAP-domain, a Follistatin/Kazal domain, an Immunoglobulin domain, two Kunitz-domains and an NTR domain (Trexler et al., 2001, 2002). WAP-, Kazal-, Kunitz- and NTR-modules are protease inhibition domains, suggesting that these proteins act as multivalent protease inhibitors. The WFIKKN1 and WFIKKN2 proteins have been shown to be differentially expressed in different tissues (Trexler et al., 2001, 2002). Inhibition of trypsin, but not of chymotrypsin, elastase, plasmin, pancreatic kallikrein, lung tryptase, plasma kallikrein, thrombin, urokinase or tissue plasminogen activator has been demonstrated for the second Kunitz-type module (KU2) of the human WFIKKN1 protein (Nagy et al., 2003). Yet, the inhibition constant of the human WFIKKN1 KU2 domain for bovine trypsin was about five orders of magnitudes lower than that of bovine pancreatic trypsin inhibitor (BPTI) for trypsin (Krowarsch et al., 1999; Nagy et al., 2003). The present structure determination of the KU2 domain was conducted to explore the molecular basis of its protease-inhibitory specificity. To the best of our knowledge, this is the first three-dimensional structure of a Kunitz domain with a tryptophan residue at the P'_2 site of the protease-recognition loop and the only such Kunitz domain for which trypsin-inhibitory activity has been demonstrated.

*To whom correspondence should be addressed:
E-mail: gottfried.otting@anu.edu.au

Methods and results

The second Kunitz domain of the human WFIKKN1 protein (KU2) was expressed in *Pichia pastoris*. The cDNA encoding the KU2 domain was cloned after the yeast α-mating factor signal sequence allowing the secretion of the recombinant protein into the medium. Secreted protein was purified by nickel-chromatography, desalted on a Sephadex G-25 column and lyophilised from 0.1 M ammonium bicarbonate. A ^{15}N-labelled sample was prepared by expression in a minimal medium containing 0.34% Yeast nitrogen base, 100 mM KH_2PO_4 pH 6.0, 0.2% $(^{15}NH_4)_2SO_4$, 4×10^{-3}% histidine and 4×10^{-5}% biotin. Protein secretion was induced by the daily addition of 1% methanol. About 4.5 mg of KU2 domain were obtained from 500 ml medium after 96 h expression. The construct comprised the amino-acid sequence shown in Figure 1 extended by the peptides EAEAEFT and VDHHHHHH at the N- and C-termini, respectively. The protein was natively folded and unglycosylated. N-terminal sequencing showed that part of the protein was missing the N-terminal EA dipeptide, reflecting the presence of different cleavage sites for removal of the α-mating factor signal peptide. This heterogeneity did not interfere

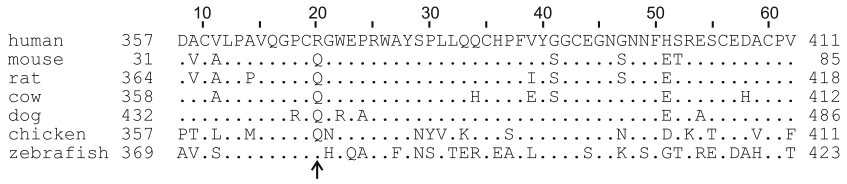

Figure 1. Amino-acid sequence alignment of the KU2 domain from human WFIKKN1 with homologous domains from WFIKKN1 proteins from mouse, rat, cow, dog, chicken and zebrafish. Residues identical to those in the human KU2 domain are replaced by dots. The residue numbering used in the structure determination of the human KU2 domain is indicated at the top. In addition, the sequence positions in the wild-type sequences are shown for the first and last amino acids. An arrow identifies the putative P_1 residue in the KU2 domain.

with the subsequent NMR analysis, since the N-terminal seven residues were highly mobile. No medium- or long-range NOEs were observed for the first seven and last seven residues.

NMR spectra were recorded at pH 4.5, 25 °C, using 1.1 mM solutions of KU2 domain in 90% H_2O/10% D_2O on a Varian Inova 800 MHz NMR spectrometer. NOE distance restraints were collected from a NOESY spectrum recorded in 144 h with 40 ms mixing time, t_{1max} = 100 ms, t_{2max} = 225 ms. Scalar coupling constants were derived from a DQF-COSY spectrum recorded in D_2O solution. Additional NOE distance restraints were obtained from a NOESY spectrum with 60 ms mixing time recorded in D_2O solution on a Bruker DMX-600 NMR spectrometer equipped with a cryoprobe. In addition, a 3D NOESY-^{15}N-HSQC spectrum was recorded at 600 MHz, using 60 ms mixing time, 68 h total recording time and $t_{1max}(^{15}N)$ = 18 ms, $t_{2max}(^{1}H)$ = 28.5 ms, $t_{3max}(^{1}H)$ = 73 ms.

The NMR structure was calculated using the program DYANA (Güntert et al., 1997) starting from 50 random conformers. All residues present in the construct were included in the structure calculations. The 20 conformers with the lowest residual restraint violations were energy minimised in water using the program OPAL (Luginbühl et al., 1996) with standard parameters. The Ramachandran plot was analysed using PROCHECK-NMR (Laskowski et al., 1996). Table 1 shows an overview of the restraints used and structural statistics. Root mean square deviation (r.m.s.d.) values were calculated using the program MOLMOL (Koradi et al., 1996) which was also used to create Figure 3.

The NMR spectra of the KU2 domain were complicated by peak doubling for most of the residues due to *cis–trans* isomerization of Pro61 (Figure 2). Both forms were about equally populated. Peak doubling was pronounced for the C-terminal residues following Phe50 and for residues near Cys10 (due to the disulfide bond between Cys10 and Cys60), but peak doubling was barely noticeable in the putative protease-binding loop (residues 16–24). Stereo-specific resonance assignments were obtained for 40 $C^{\beta}H_2$, 11 $C^{\gamma}H_2$, 7 $C^{\delta}H_2$, 6 $C^{\alpha}H_2$, 4 $C^{\gamma}H_3$, 3 $C^{\delta}H_3$ and 4 NH_2

Table 1. Structural characteristics for the NMR conformers of the KU2 domain from human WFIKKN1 with *cis*- and *trans*-Pro61, respectively[a]

	cis	trans
Number of assigned NOE cross peaks	970	970
Number of non-redundant NOE upper-distance limits	644	644
Number of scalar coupling constants[b]	71	71
Number of dihedral-angle restraints	173	173
Intra-protein AMBER energy (kcal/mol)	−3048 ± 60	−2973 ± 99
Maximum NOE-restraint violations (Å)	0.10 ± 0.00	0.10 ± 0.01
Maximum dihedral-angle restraint violations (°)	2.3 ± 0.2	2.4 ± 0.1
R.m.s.d. to the mean for N,C^{α} and C′ (Å)[c]	0.54 ± 0.13	0.61 ± 0.14
R.m.s.d. to the mean for all heavy atoms (Å)[c]	0.80 ± 0.16	0.87 ± 0.15
Ramachandran plot appearance[c]		
Most favoured regions (%)	80.4	78.4
Additionally allowed regions (%)	19.6	21.4
Generously allowed regions (%)	0.0	0.2
Disallowed regions (%)	0.0	0.0

[a]Data for 70 residues, including the N-terminal cloning artefact and C-terminal His-tag.
[b]71 $^{3}J(H^{\alpha} - H^{\beta})$.
[c]For residues 7–62.

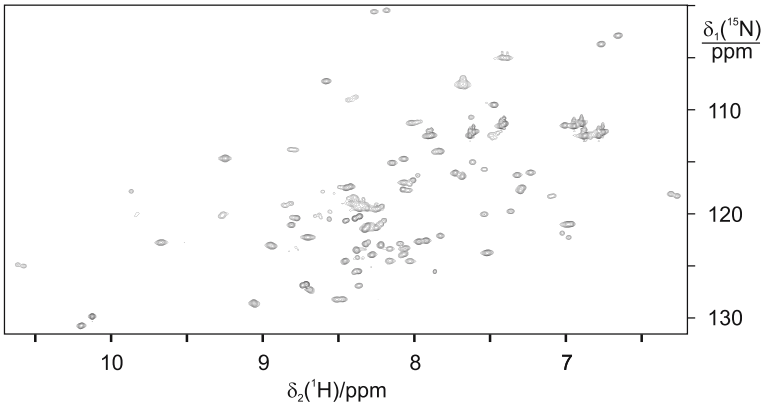

Figure 2. ^{15}N-HSQC spectrum of a 1 mM solution of the KU2 domain from human WFIKKN1. The spectrum was recorded at pH 3.6 and 25 °C. It contained more cross-peaks than expected due to *cis–trans* isomerization of Pro61.

groups. The NMR chemical shifts have been deposited at the BMRB (accession code 6963).

Two sets of coordinates were calculated with *cis* and *trans* peptide bonds of Pro61, respectively. Most of the restraints were identical for both molecules and the resulting structures were closely similar in quality. The structural statistics for both forms are reported in Table 1. The conformation of the protease-binding loop (top of Figure 3a), the buried amino-acid side chains and of the disulfide bridges was well defined, but the backbone following Pro61 (bottom of Figure 3a) was less well determined. The atomic coordinates have been deposited at the PDB (accession codes 2DDI and 2DDJ for the *trans*- and *cis*-Pro61 conformers, respectively).

The structure of the second Kunitz domain from human WFIKKN1 is very similar to that of BPTI, with a backbone rmsd of 0.9 Å between the mean structure of WFIKKN1 KU2 (residues 8–62, Figure 1) and residues 3–57 of BPTI (PDB code 5PTI). In particular, the protease-binding loops of the two proteins are closely similar. A model of the complex between WFIKKN1 KU2 and trypsin can thus be obtained by superposition of the KU2 domain with BPTI in the crystal structure of the trypsin–BPTI complex (Helland et al., 1999).

Figure 3b shows a superposition of the KU2 domain with the complexes between BPTI and bovine β-trypsin (Helland et al., 1999) and between the domain II of tissue factor pathway inhibitor (TFPI) and porcine trypsin (Burgering et al., 1997). The figure illustrates the small magnitude of structural differences between these two highly homologous trypsin–inhibitor complexes. The backbone of the protease-binding loop of the KU2 domain closely traces the backbone conformations of the other two inhibitors. Also the side chains superimpose remarkably closely. In particular, the P_1 residue of the KU2 domain points in the same direction as the corresponding residues of BPTI and TFPI. A notable difference, however, is presented by the side-chain orientation of Trp22 at the P'_2 position which is incompatible with a trypsin complex. Trp22 would have to assume the χ_1 angle of the P'_2-residue Tyr17 in domain II of TFPI (Figure 3b) to allow complex formation. The population of alternative side-chain conformations of Trp22 could not be assessed from our data. Yet, a conformation analogous to that of Tyr17 in TFPI would have resulted in strong NOEs between the methyl groups of Val39 and the indol H^N resonance of Trp22. These were not observed in the NOESY spectrum recorded with 40 ms mixing time; instead, the side chain of Val39 showed NOEs with protons on the opposite side of the ring ($H^{\varepsilon 3}$ and $H^{\zeta 3}$). Unfavourable side-chain conformations of Trp22 may hence reduce the effective on-rate of the KU2 domain in its complex with trypsin and thus contribute to its much lower trypsin affinity compared with BPTI.

Discussion and conclusions

There are a number of indications that trypsin may not be the prime target of the KU2 domain. (i) The inhibition constant of the human WFIKKN1

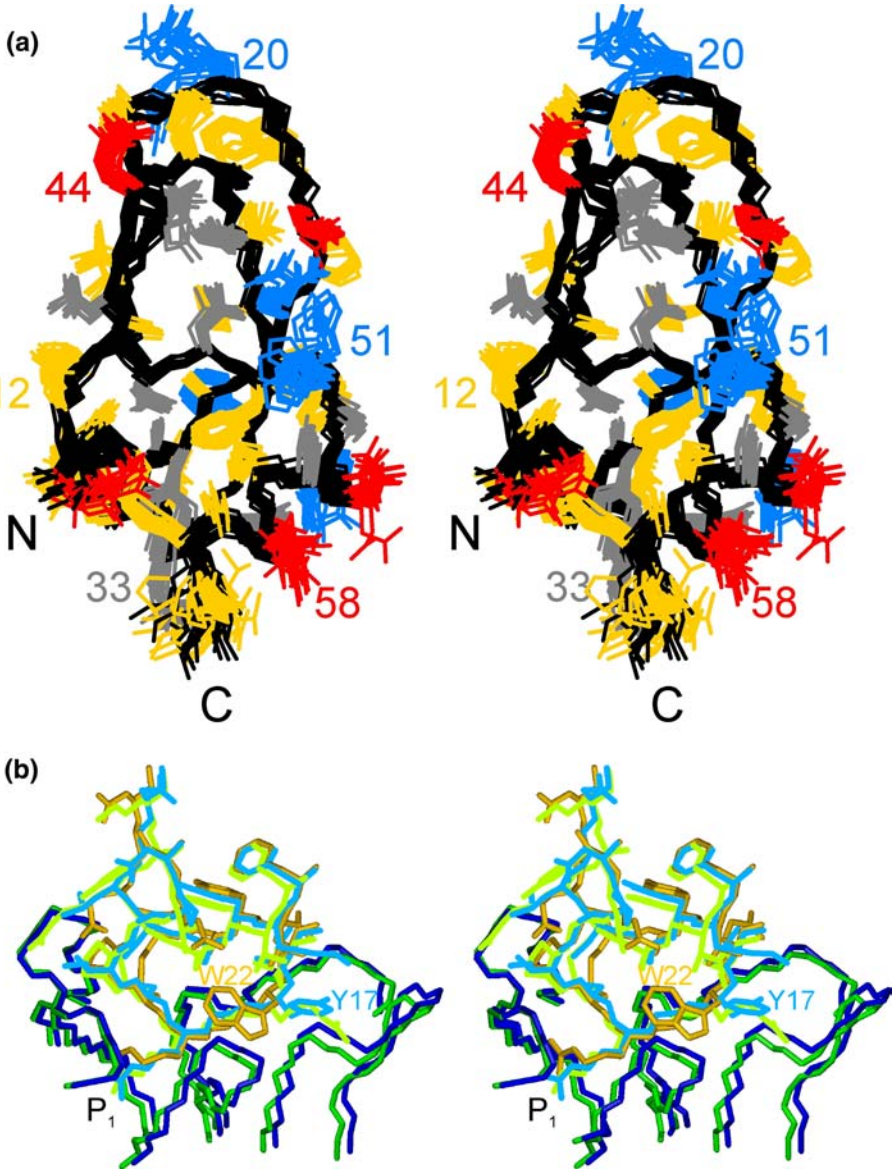

Figure 3. NMR structure of the KU2 domain from human WFIKKN1 and putative binding mode to trypsin. (a) Superposition of the 20 best *trans*-Pro61 conformers. The following colours were used: black (backbone), yellow (Ala, Cys, Leu, Phe, Pro, Trp, Val), red (Glu, Asp), blue (Arg, Lys, His); grey (Asn, Gln, Ser, Tyr). The N- and C-termini are labelled. (b) Comparison of the interface between BPTI and bovine β-trypsin (3BTK; Helland et al., 1999; plotted in green) and the interface between domain II of tissue factor pathway inhibitor (TFPI) and porcine trypsin (1TRX; Burgering et al., 1997; plotted in blue). In addition, the protease-binding loop of the KU2 domain (yellow) is superimposed onto the Kunitz domains. The trypsin molecules (bottom) are plotted with darker colours than the Kunitz domains (at the top). Side chains are shown only for the Kunitz domains and only residues near the interface are plotted. The P_1 side chains are identified, as well as Trp22 of KU2 and Tyr17 of domain II of TFPI.

KU2 domain with trypsin (K_i = 9.6 nM; Nagy et al., 2003) is several orders of magnitude weaker than that of BPTI (Krowarsch et al., 1999). (ii) Typical trypsin inhibitors of the Kunitz family feature a Tyr residue or other less bulky residues at the P'_2 site. Among 1000 Kunitz domains identified by a BLAST search, only about 3% feature a Trp residue at the P'_2 site and about 80% of those do not simultaneously display the Arg or Lys residue at the P_1 site characteristic of trypsin inhibitors.

(iii) The main contacts between Trp22 of KU2 and trypsin are expected to occur with Tyr151 and Gly193 of trypsin (using the sequence numbering of bovine trypsin) and possibly also with Gly142, Asn143, Cys191, Gln192. All these residues are strictly conserved in trypsins of mouse, human, dog, chicken, rat, zebrafish, bovine. Remarkably, however, the P_1 residue of the WFIKKN1 KU2 domains of these organisms is most frequently a Gln (Figure 1). Gln instead of Lys at the P_1 site in BPTI reduces the association with trypsin by almost five orders of magnitude (Krowarsch et al., 1999), indicating that the Gln residue in the KU2 domain would largely destroy trypsin inhibition.

Among all Kunitz-type inhibitors for which 3D structures have been determined, only β2 bungarotoxin has a Gln at the P_1 site, but the protein is not a protease inhibitor; it binds to ion-channels rather than proteases, its structure shows a very different conformation of the peptide loop corresponding to the protease binding loop in typical Kunitz-type protease inhibitors such as BPTI and ion-channel binding is via a surface remote from the protease-binding loop (Kwong et al., 1995). As shown by the present work, the structure of the KU2 domain is closely related to that of the canonical protease inhibitors. If the WFIKKN1 KU2 domain had a conserved protease-inhibitory function between the organisms shown in Figure 1, the matching protease would be expected to accommodate arginine as the P_1 residue in humans and zebrafish, and glutamine in the other organisms. We could not identify such a protease. Possibly, the presence of the KU1 domain prevents proteases from access to the KU2 domain in analogy to the situation reported for the second Kunitz domain of bikunin (Xu et al., 1998). Like in bikunins, the length of the linker peptide between KU1 and KU2 domains is remarkably conserved between different organisms, although it is longer by two residues compared to the bikunin linker and the bikunin interface residues are not conserved in WFIKKN proteins.

The WFIKKN-related protein (WFIKKNRP or WFIKKN2) has the same domain organization as the WFIKKN1 protein but different expression patterns (Trexler et al., 2002). It exists in all the species of Figure 1. Recent studies have identified WFIKKN2 as a myostatin-binding protein and implicated it in the regulation of muscle development (Hill et al., 2003). Myostatin, a member of the TGFβ superfamily is produced from precursor protein by proteolytic processing. After cleavage of a single peptide bond by a furin-type protease the N-terminal myostatin propeptide and the C-terminal mature myostatin remain associated, forming an inactive latent complex. Active mature myostatin must be liberated from the latent complex by the proteolytic degradation of the propeptide, but the identity of the latter proteases is unknown (Zimmers et al., 2002).

It is noteworthy in this respect that the KU2 domain of WFIKKN2 from all these species contains an absolutely conserved GPCKAY sequence at the $P_4 - P_2'$ site which is closely related to the GPCRAF sequence in the Kunitz domain of trypstatin. Trypstatin inhibits factor Xa and tryptase with inhibition constants of $1–4 \times 10^{-10}$ M (Kido et al., 1988), raising the possibility that the KU2 domains of WFIKKN2 proteins target proteases with specificities that are conserved between different organisms.

Acknowledgements

We thank Dr Aaron Oakley for an MD simulation probing the time scale of conformational exchange of Trp22. G. O. thanks the Australian Research Council for a Federation Fellowship. This work was supported by the Swedish and Australian Research Councils, and by grants NKFP1/044/2001 from the Hungarian National Research and Development Program, and grant OTKA/TS049890 from the Hungarian Research Fund, Budapest.

References

Burgering, M.J.M., Orbons, L.P.M., Vanderdoelen, A., Mulders, J., Theunissen, H.J.M., Grootenhuis, P.D.J., Bode, W., Huber, R. and Stubbs, M.T. (1997) *J. Mol. Biol.*, **269**, 395–407.

Güntert, P., Mumenthaler, C. and Wüthrich, K. (1997) *J. Mol. Biol.*, **273**, 283–298.

Helland, R., Otlewski, J., Sundheim, O., Dadlez, M. and Smalås, A.O. (1999) *J. Mol. Biol.*, **287**, 923–942.

Hill, J.J., Qiu, Y., Hewick, R.M. and Wolfman, N.M. (2003) *Mol. Endocrinol.*, **17**, 1144–1154.

Kido, H., Yokogoshi, Y. and Katunuma, N. (1988) *J. Biol. Chem.*, **263**, 18104–18107.

Koradi, R., Billeter, M. and Wüthrich, K. (1996) *J. Mol. Graph.*, **14**, 51–55.

Krowarsch, D., Dadlez, M., Buczek, O., Krokoszynska, I., Smalås, A.O. and Otlewski, J. (1999) *J. Mol. Biol.*, **289**, 175–186.

Kwong, P.D., McDonald, N.Q., Sigler, P.B. and Hendrickson, W.A. (1995) *Structure*, **3**, 1109–1119.
Laskowski, R.A., Rullmann, J.A.C., MacArthur, M.W., Kaptein, R. and Thornton, J.M. (1996) *J. Biomol. NMR*, **8**, 477–486.
Luginbühl, P., Güntert, P., Billeter, M. and Wüthrich, K. (1996) *J. Biomol. NMR*, **8**, 136–146.
Nagy, A., Trexler, M. and Patthy, L. (2003) *Eur. J. Biochem.*, **270**, 2101–2107.
Trexler, M., Bányai, L. and Patthy, L. (2001) *Proc. Natl. Acad. Sci.*, **98**, 3705–3709.
Trexler, M., Bányai, L. and Patthy, L. (2002) *Biol. Chem.*, **383**, 223–228.
Xu, Y., Carr, P.D., Guss, J.M. and Ollis, D.L. (1998) *J. Mol. Biol.*, **276**, 955–966.
Zimmers, T.A., Davies, M.V., Koniaris, L.G., Haynes, P., Esquela, A.F., Tomkinson, K.N., McPherron, A.C., Wolfman, N.M. and Lee, S.J. (2002) *Science*, **296**, 1486–1488.