Kinesin subfamily UNC104 contains a FHA domain: boundaries and physicochemical characterization

Ann Westerholm-Parvinen[a,b,*], Isabelle Vernos[a], Luis Serrano[b]

[a] *Cell Biology and Cell Biophysics Program, European Molecular Biology Laboratory, D-69117 Heidelberg, Germany*
[b] *Structural and Computational Biology Program, European Molecular Biology Laboratory, D-69117 Heidelberg, Germany*

Received 31 October 2000; revised 17 November 2000; accepted 18 November 2000

First published online 30 November 2000

Edited by Gunnar von Heijne

Abstract By sequence analysis we show that the U104 domain found in the UNC104 subfamily of kinesins is a forkhead homology-associated domain (FHA). A combination of limited proteolysis, mass spectroscopy, and physicochemical analysis define this domain as a genuine autonomously folding domain. Our data show that the FHA domain is shorter than previously reported since the C-terminal α-helix is not part of its minimum core. Key amino acids postulated to recognize phosphorylated residues are conserved. These data suggest that the kinesin FHA domains are functional domains involved in protein–protein interactions regulated by phosphorylation.

Key words: Forkhead homology-associated domain; *Xenopus* kinesin-like protein 4; Kinesin-like protein; UNC104 subfamily

1. Introduction

Kinesin-like proteins (KLPs) are motor proteins that use the energy derived from ATP hydrolysis to produce directional movements along microtubules. KLPs have a conserved motor domain containing the ATP and microtubule binding sites, linked to diverse tail regions. Based on phylogenetic sequence analysis of the motor domains, KLPs can be grouped into several subfamilies. The variable tail sequences are thought to define the specificity for cargo binding and/or the targeting to different cellular locations.

Very few known protein domains have been identified so far in the non-motor regions of KLPs. The only clearly defined one is a pleckstrin homology domain present in the tail of *Caenorhabditis* UNC104 [1] and its homologues [2–4]. Multiple sequence alignments have also revealed the presence of a new putative domain, called U104, in two KLPs from the UNC104 subfamily and three non-kinesin proteins [5].

Xklp4 (*Xenopus* kinesin-like protein 4) is a member of the UNC104 kinesin subfamily [6]. Members from this subfamily are functionally very diverse and participate in the transport of a variety of cellular cargoes [2–4,7–11]. In addition, some of them are monomers [2,9,12] whereas others are homodimers [4,10]. As a first step to characterize Xklp4, we have obtained new cDNA clones that extend the previously known sequence. We find that Xklp4 has a putative U104 domain at the same position as the members of the UNC104 subfamily. To investigate this domain, we have performed sequence analysis searches and found regions of homology between U104 and forkhead homology-associated (FHA) domains. We have determined the exact boundaries of the Xklp4 U104/FHA homology region and using physicochemical analysis we show that it is a true autonomously folding domain. Our data show that this FHA domain is shorter than reported previously and does not include a C-terminal α-helix. In addition, they confirm that the U104 putative domain found in some KLPs is a true protein domain that may be involved in phosphorylation-dependent protein–protein interactions.

2. Materials and methods

2.1. Isolation of partial Xklp4 cDNA

The PCR fragment of Xklp4 described by Vernos et al. [6] (accession no.: PIR:D48835) was used for screening an oocyte cDNA library kindly provided by D. Melton. Two clones were isolated and contained overlapping sequences lacking the 5′ and the 3′ ends of the Xklp4 open reading frame (ORF). To obtain the 5′ end, a SMART RACE protocol (Clonetech) was used. The partial Xklp4 cDNA sequence can be found in the EMBL database under accession number AJ297516.

2.2. Cloning, expression and purification

DNA fragments encoding amino acids (aa) 463–589 for FHA129, 472–580 for FHA110 and 472–623 for FHA153 were obtained by PCR on the Xklp4 cDNA clone and cloned into *Escherichia coli* expression vector pBat4 [13]. A methionine codon present in the *Nco*I site was used as translation initiator. An aspartate residue was introduced after the first methionine residue when necessary to accommodate the *Nco*I restriction site in the linker DNA.

Protein expression was performed in BL21(DE3) at 37°C for 4 h after induction with IPTG. Soluble proteins were first purified on a Q-Sepharose FF column (Pharmacia) run with 10 mM sodium phosphate, 1 mM EDTA and 1 mM dithiothreitol (DTT) and protein was eluted with a linear salt gradient at ~290 mM NaCl. Gel filtration in a Superdex 75 26/60 preparation grade column (Pharmacia) was performed with 50 mM sodium phosphate buffer, 150 mM NaCl, 1 mM EDTA and 1 mM DTT pH 7.0. Purified protein was concentrated to 10–20 mg/ml.

The molecular weights of purified proteins were verified by the Mass Spectroscopy Service at EMBL. For FHA153 it was shown that the N-terminal methionine was formylated and the C-terminus was cleaved at position Arg583.

Protein concentration was determined spectrophotometrically in 6 M guanidium–HCl using a calculated molar absorption coefficient of 9035/M at 278 nm [14].

2.3. Limited proteolysis of FHA129

Limited proteolysis was carried out using 0.2 mg/ml of FHA129 in 500 mM ammonium bicarbonate pH 8.5 at 37°C using elastase

*Corresponding author. Fax: (49)-6221-387306.
E-mail: westerho@embl-heidelberg.de

Abbreviations: FHA, forkhead homology-associated domain; Xklp4, *Xenopus* kinesin-like protein 4; KLP, kinesin-like protein

0014-5793/00/$20.00
PII: S0014-5793(00)02310-3

(Boehringer), trypsin (Boehringer) or α-chymotrypsin (Sigma) at 0.4 μg/ml. Samples were removed from the reaction at different time points and proteolysis stopped by directly boiling half of the sample in Laemmli sample buffer [15] for analysis and freezing the other half in liquid nitrogen. The samples in Laemmli's sample buffer were analyzed by SDS–PAGE and visualized by Coomassie brilliant blue R250 staining. The frozen samples were used for molecular weight determinations by mass spectroscopy. The mass spectroscopy analyses were performed immediately on thawing by the Mass Spectroscopy Service at EMBL.

2.4. Circular dichroism (CD) analysis

The far UV CD spectra of a 10 μM protein sample in 10 mM sodium phosphate buffer pH 7.0 were recorded on a Jasco-710 spectropolarimeter at 298 K in a cuvette with a 0.2 mm path. Measurements were averaged for 20 scans recorded at 50 nm/min.

The near UV spectra were recorded at 298 K in a cuvette with a 1 cm path at a protein concentration of 10 μM in 10 mM sodium phosphate buffer pH 7.0.

2.5. Chemical denaturation experiment

Urea solutions were prepared and accurate concentrations determined with a refractometer. For each data point denaturant was mixed with protein in sodium phosphate buffer pH 7.0 to a final buffer concentration of 10 mM and protein concentration of 4 μM. Fluorescence was measured in an Aminco Bowman Series 2 luminescence spectrometer. Excitation was at 290 nm with a 2 nm slit and fluorescence was detected through a 4 nm slit at 298 K. The free energy of unfolding, ΔG, was calculated using the equations described in [16].

2.6. Thermal denaturation analysis

Thermally induced unfolding was monitored by following the change in signal at 215 nm with a CD spectropolarimeter over a temperature range of 278–368 K with a 50°C/h heating rate. The protein concentration was 10 μM and the measurement was carried out in 10 mM sodium phosphate buffer pH 7.0 in a cuvette with a 2 mm path. Reversibility was checked by the comparison of the original spectrum with the one obtained after recooling the sample. The curve was fitted with a two-state model using the equations described in [17]. ΔCp was approximated to be −251+0.19·ΔASA and ΔASA is derived from −907+93·(number of residues) [18].

3. Results and discussion

3.1. Protein database search

To define and characterize the U104 domain from Xklp4 we made a sequence profile based on sequence alignments of the U104 domains found in the UNC104 subfamily members and three other proteins reported by Ponting [5] and searched the protein database. We found three more proteins containing a similar domain, Cer144 from *Caenorhabditis*, Dun1p, a protein kinase from *Saccharomyces cerevisiae* and afadin from rat.

We then run Xklp4 sequence on the SMART program (http://smart.embl-heidelberg.de/) designed to recognize domains based on sequence profiles, and found that the U104 domain is a FHA domain [19]. The KLP FHA domains give E-values varying from 0 to 10^{-10}, in the same range as the values for other FHA domains.

The FHA domain is found in more than 170 proteins in a non-redundant compilation of SwissProt and TREMBL databases from both prokaryotes and eukaryotes. Based on sequence comparisons, the FHA domain was initially predicted to be 55–75 aa long with three conserved regions separated with spacers of varying length. Only two aa residues are conserved (the Gly and His marked with stars in Fig. 1). The

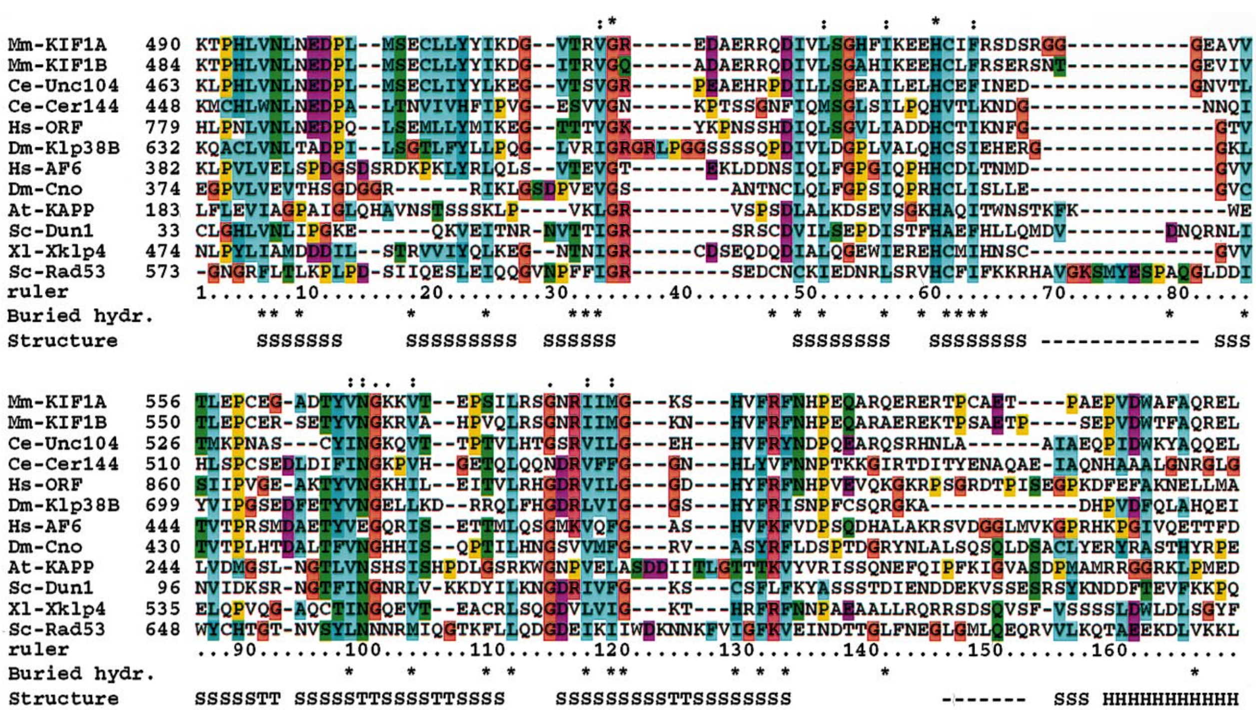

Fig. 1. Multiple sequence alignment of selected FHA domains. The alignment was constructed with Clustal X [27] and manually refined. In the first column the protein names are preceded with species names: At, *Arabidopsis thaliana*; Ce, *Caenorhabditis elegans*; Dm, *Drosophila melanogaster*; Hs, *Homo sapiens*; Mm, *Mus musculus*; Sc, *S. cerevisiae*; Xl, *Xenopus laevis*. The second column gives the positions of the first aligned residue in each of the sequences. At the bottom of the alignment the structure of the Rad53 FHA2 domain is shown; S for β-sheets, T for turn, H for α-helix and - for unstructured loops. The hydrophobic residues in Rad53 FHA2 that are less than 10% accessible are indicated with stars. The accession numbers from GenBank/EMBL/DDBJ are the following: KIF1A, P33173; KIF1B, Q60575; UNC104, Q18778; Xklp4, AJ297516; Cer144, U23515; ORF-KIAA0042, Q15058; Klp38B, Y15247; AF-6, P55196; Cno, Q24279; Dun1p, P39009; Rad53, P22216; KAPP, A55174.

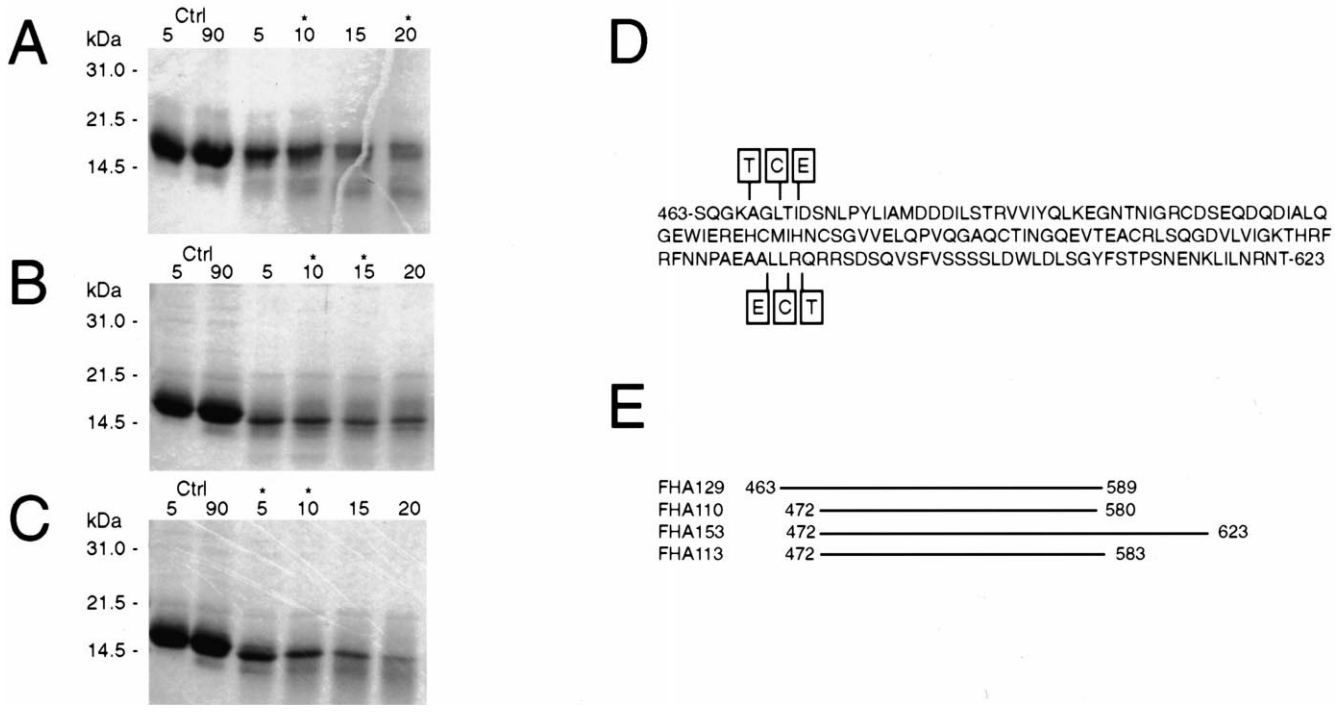

Fig. 2. Identification of protease-resistant fragments of FHA129. FHA129 was incubated with (A) elastase, (B) trypsin or (C) α-chymotrypsin for the times (min) indicated at the top. The molecular mass of marker proteins is indicated on the left. The samples marked with a star on top were analyzed by mass spectroscopy. The first two lanes in each gel show undigested protein as a control at 5 and 90 min. (D) Positions of the cleavage sites from the limited proteolysis; C, α-chymotrypsin; E, elastase; T, trypsin. (E) Constructs of Xklp4 FHA domain used in this paper.

solution structure of FHA2, the C-terminal FHA domain from Rad53, a checkpoint protein involved in DNA damage response in yeast, showed that it consists of a β-sandwich containing two antiparallel β-sheets and a short C-terminal α-helix [20,21]. The structure also shows that the FHA domain is larger than previously thought, with flanking sequences required to form a stable β-sandwich structure. This result is supported by functional data indicating that around 30 aa flanking the FHA domain on each side are needed for interaction between the Rad53 FHA2 domain and phosphorylated Rad9 [22], and between the *Arabidopsis* kinase-associated protein phosphatase (KAPP) FHA domain and RLK5 [23]. An experimental approach to determine FHA domain boundaries of the N-terminal FHA (FHA1) domain from Rad53 and the FHA domain from another yeast checkpoint protein, Dun1p, by limited proteolysis also indicates that the FHA domain is indeed remarkably larger than first predicted [24].

The sequence alignment between the FHA domains found in the UNC104 subfamily kinesins and some representative FHA sequences as well as their predicted secondary structure [20] are shown in Fig. 1. The sequence homology is low but spans almost the entire length of the Rad53 FHA2 domain except at the last C-terminal 20–30 residues. All the residues that are thought to be important for interaction with the

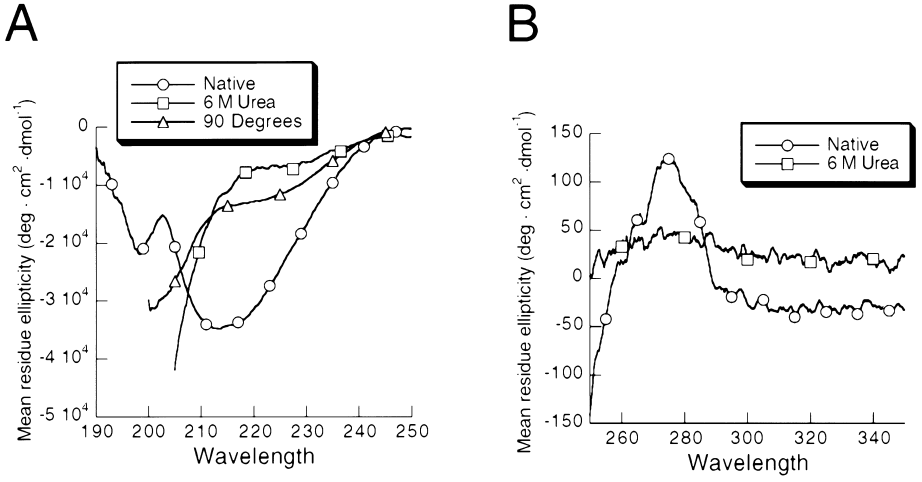

Fig. 3. CD far (A) and near (B) UV spectra of native and denatured FHA113. CD spectra were recorded at 298 K for native FHA113 in 10 mM phosphate, pH 7.0 (○), and denatured in 6 M urea (□). Denaturation of FHA113 by heat was recorded at 363 K (△).

phosphorylated aa [20,22] are conserved in the UNC104 subfamily of kinesins. According to a recent classification of FHA domains into two families [25], all the FHA domains shown in Fig. 1 belong to the Rad53-like family.

3.2. Determination of the Xklp4 FHA domain boundaries

Although the low sequence homology did not allow an accurate prediction of the domain boundaries, we generated construct FHA129 corresponding to aa 463–589 of Xklp4 (Fig. 2E) based on the sequence alignment shown in Fig. 1. To define precisely the domain boundaries we then performed limited proteolysis on the purified protein under mild conditions to cleave the unstructured tails and preserve the folded core. Three different proteases were used: elastase (Fig. 2A), trypsin (Fig. 2B) and α-chymotrypsin (Fig. 2C). Mass spectroscopy analysis of the higher molecular weight bands showed that cleavages had occurred at residues 466–471 and 580–583 (Fig. 2D). Based on these results we made construct FHA110 corresponding to aa 472–580 of Xklp4 (Fig. 2E).

Based on the first solution structure of the Rad53 FHA2 domain (Fig. 5) and its sequence alignment with Xklp4 (Fig. 1) it seemed that FHA110 might be lacking some C-terminal aa. We thus prepared FHA153, a longer construct corresponding to aa 472–623 of Xklp4 (Fig. 2E). Purification of this protein resulted consistently in a single well-defined protein of lower molecular weight than predicted. Mass spectroscopy analysis and N-terminal protein sequencing showed that the N-terminal methionine was present but that the C-terminus was cleaved off at position Arg583. The purified FHA153 protein was thus only three aa longer at the C-terminus than FHA110. We called this truncated version of FHA153, FHA113. Both FHA110 and FHA113 were very similar from a spectroscopic point of view (data not shown). However, we found that FHA113 produced slightly more cooperative chemical denaturation curves (data not shown). Therefore, we performed all further physicochemical characterizations on FHA113.

3.3. CD and fluorescence spectroscopy analysis

Far UV CD spectroscopy showed that FHA113 consisted of mainly β-sheet structure with a deep minimum at around 217 nm (Fig. 3A), as expected from secondary sequence predictions and the solved Rad53 FHA2 structure. The secondary structure was lost in a reversible manner upon protein denaturation by heat or chemical denaturant (Fig. 3A). After either cooling the heated sample, or diluting the sample from 6 M urea to native conditions, the far UV spectrum was identical to the original native one (data not shown). The signal from the aromatic residues in the near UV CD was also lost upon denaturation with urea, indicating that the domain has tertiary structure (Fig. 3B).

Fluorescence emission spectra of the tryptophan in

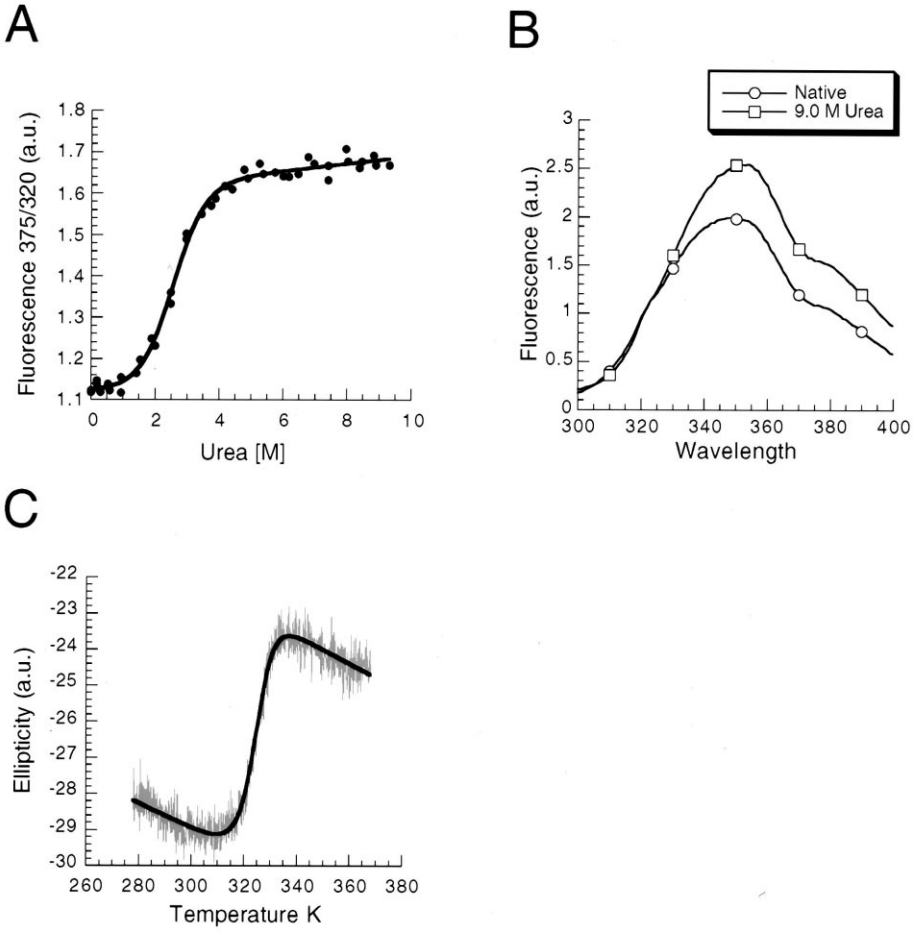

Fig. 4. Denaturation of FHA113. (A) Chemical denaturation of FHA113 followed by fluorescence. (B) Fluorescence emission spectra of native (○) and denatured (□) FHA113. (C) Thermal denaturation of FHA113 followed by CD. The unfolding was followed by the changes in the CD signal at 215 nm.

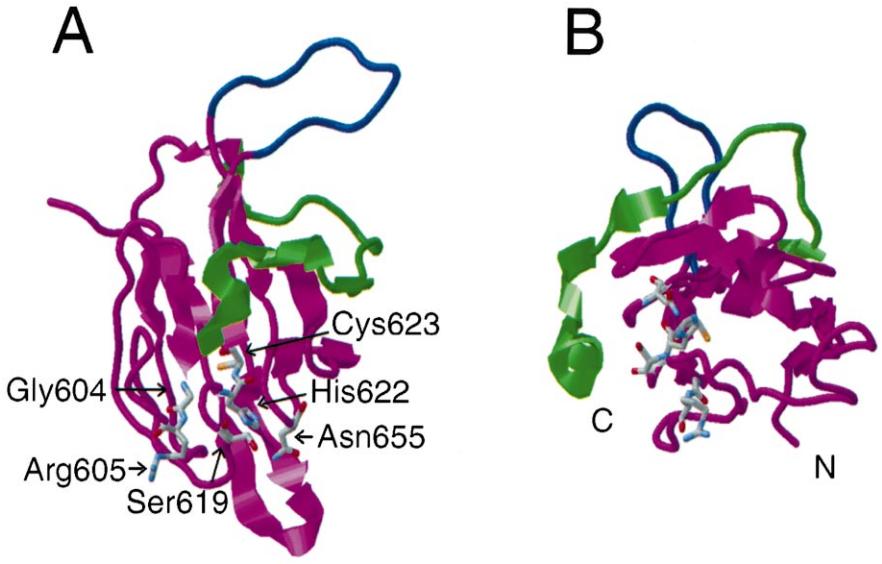

Fig. 5. Structure of the Rad53 FHA2 domain. The unstructured long loop between residues 630 and 644 is in blue and the C-terminal part of the domain in green. (A) The conserved residues postulated to be important for phosphoprotein binding are shown: Gly604, Arg605, Ser619, His622, Cys623 and Asp655. (B) The C-terminal α-helix and the preceding connecting loop are not conserved in FHA domains from the kinesin subfamily UNC104.

FHA113 also showed that the domain was folded: the intrinsic fluorescence increased and the maximum was slightly shifted upon denaturation with urea (Fig. 4B).

3.4. Chemical and thermal denaturation

A hallmark of folded proteins is that they exhibit a cooperative sigmoidal curve when denatured by chemical agents or temperature. Fig. 4A shows the urea denaturation of FHA113, followed by fluorescence. Denaturation of this domain was cooperative and could be fitted to a two-state denaturation model. Sequence homology and the three-dimensional structure of Rad53 FHA2 suggested that the tryptophan in FHA113 was relatively solvent accessible. This results in a small change in fluorescence upon unfolding and in a high sloping of the folded baseline due to the decrease in hydrophilicity of the solvent upon increasing urea concentration. To eliminate the sloping and facilitate the fitting we used the ratio between the fluorescence values at 375 and 320 nm (Fig. 4A; using a single wavelength gave the same results but with a larger error; data not shown). From the average of three different experiments we determined that the free energy of unfolding, ΔG, was 3.0 ± 0.2 kcal/mol. The m value, related to the difference between the solvent accessibility of the folded and unfolded states, was 1.17 ± 0.09 kcal/mol·M, which is reasonable for a protein of this size [18].

Fig. 4C shows that the thermal denaturation curve of FHA113 was sigmoidal as expected for a well-folded protein. The curve was fitted with a two-state model using a fixed ΔCp value. The ΔCp value was obtained from the correlation analysis between size and ΔCp, done by Pace and co-workers [18]. We found a midpoint temperature (T_m) of 324.48 ± 0.13 K and a ΔH_m of 255.22 ± 8.08 kJ/mol. Using these values and the approximated ΔCp value of 6.59 kJ/K·mol, the free energy of unfolding was calculated to be $\sim$3.3 kcal/mol. The ΔG values obtained from the two denaturation experiments were equal within the experimental error. Therefore, these results confirmed that the FHA113 domain is a folded and stable domain.

3.5. Comparison between Rad53 FHA2 domain and Xklp4 FHA113

The solution structure of the FHA2 domain of Rad53 shows that it consists of a β-sandwich containing two antiparallel β-sheets and a short C-terminal α-helix [20]. The residues important for phosphoprotein binding [20,22] are conserved in the FHA domains of the KLPs shown in Fig. 5A, suggesting a role for these FHA domains in phosphorylation-dependent protein–protein interactions.

The conservation of the buried hydrophobic residues in Rad53 FHA2 and the FHA domains of KLPs, together with the CD spectra and biophysical characterization of Xklp4 FHA113 indicate that the three-dimensional core structure of these domains must be very similar. Outside the core structure, we find that the long loop between residues 630 and 644 in Rad53 FHA2 (blue in Fig. 5) is absent in the FHA domains of KLPs. Interestingly, the C-terminal α-helix of Rad53 FHA2 and its connecting loop (green in Fig. 5) are not conserved at the sequence level in the KLPs. In fact, the hydrophobic residues from the β-sheets that pack against the C-terminal α-helix in Rad53 FHA2 are replaced by polar residues in Xklp4 (Phe602 to Asn501; Val621 to Glu523; Phe624 to Met526; Phe626 to His528; Cys650 to Gln537). This observation together with our experimental data presented above lead us to propose that the minimum folded core of the FHA domain is shorter than that found for Rad53 FHA2 and does not require the C-terminal α-helix. This has also been seen by limited proteolysis of Dun1p [24].

3.6. Conclusions

Our results show that the members of the UNC104 kinesin subfamily contain a FHA domain that constitutes an autonomously folding unit. Most of the proteins containing FHA domains have been shown to be nuclear or at least to have nuclear localization signals. Whether or not the KLP FHA domains have nuclear function remains to be determined. So far the only member of this KLP subfamily, chromokinesin Klp38B from *Drosophila* [11], has been shown to localize to

the nucleus. Recent studies indicate that the FHA domain might be involved in phosphorylation-dependent protein–protein interaction [22,23]. It has also been shown that different FHA domains bind peptides in a phosphorylation-dependent manner [26]. Although it is still unclear whether the FHA domain binds exclusively to phosphorylated tyrosine [21] or threonine residues, convincing data show that some FHA domains bind to phosphorylated threonine peptides [26]. Determining the partners of the kinesins FHA domain may turn out to be very important to understand their cellular function.

Acknowledgements: We thank Anna Shevchenko and Keith Ashman at the EMBL Mass Spectroscopy Service. We also thank Jörg Schultz for information about FHA and SMART and David Stephens for critical reading of the manuscript.

References

[1] Otsuka, A.J., Jeyaprakash, A., Garcia-Anoveros, J., Tang, L.Z., Fisk, G., Hartshorne, T., Franco, R. and Born, T. (1991) Neuron 6, 113–122.
[2] Okada, Y., Yamazaki, H., Sekine-Aizawa, Y. and Hirokawa, N. (1995) Cell 81, 769–780.
[3] Furlong, R.A., Zhou, C.Y., Ferguson-Smith, M.A. and Affara, N.A. (1996) Genomics 33, 421–429.
[4] Pollock, N., de Hostos, E.L., Turck, C.W. and Vale, R.D. (1999) J. Cell Biol. 147, 493–506.
[5] Ponting, C.P. (1995) Trends Biochem. Sci. 20, 265–266.
[6] Vernos, I., Heasman, J. and Wylie, C. (1993) Dev. Biol. 157, 232–239.
[7] Hall, D.H. and Hedgecock, E.M. (1991) Cell 65, 837–847.
[8] Yonekawa, Y., Harada, A., Okada, Y., Funakoshi, T., Kanai, Y., Takei, Y., Terada, S., Noda, T. and Hirokawa, N. (1998) J. Cell Biol. 141, 431–441.
[9] Nangaku, M., Sato-Yoshitake, R., Okada, Y., Noda, Y., Takemura, R., Yamazaki, H. and Hirokawa, N. (1994) Cell 79, 1209–1220.
[10] Dorner, C., Ullrich, A., Haring, H.U. and Lammers, R. (1999) J. Biol. Chem. 274, 33654–33660.
[11] Molina, I., Baars, S., Brill, J.A., Hales, K.G., Fuller, M.T. and Ripoll, P. (1997) J. Cell Biol. 139, 1361–1371.
[12] Okada, Y. and Hirokawa, N. (1999) Science 283, 1152–1157.
[13] Peränen, J., Rikkonen, M., Hyvönen, M. and Kääriäinen, L. (1996) Anal. Biochem. 236, 371–373.
[14] Gill, S.C. and von Hippel, P.H. (1989) Anal. Biochem. 182, 319–326.
[15] Laemmli, U.K. (1970) Nature 227, 680–685.
[16] Villegas, V., Azuaga, A., Catasus, L., Reverter, D., Mateo, P.L., Aviles, F.X. and Serrano, L. (1995) Biochemistry 34, 15105–15110.
[17] Viguera, A.R., Martinez, J.C., Filimonov, V.V., Mateo, P.L. and Serrano, L. (1994) Biochemistry 33, 2142–2150.
[18] Myers, J.K., Pace, C.N. and Scholtz, J.M. (1995) Protein Sci. 4, 2138–2148.
[19] Hofmann, K. and Bucher, P. (1995) Trends Biochem. Sci. 20, 347–349.
[20] Liao, H., Byeon, I.J. and Tsai, M.D. (1999) J. Mol. Biol. 294, 1041–1049.
[21] Wang, P., Byeon, I.J., Liao, H., Beebe, K.D., Yongkiettrakul, S., Pei, D. and Tsai, M.D. (2000) J. Mol. Biol. 302, 927–940.
[22] Sun, Z., Hsiao, J., Fay, D.S. and Stern, D.F. (1998) Science 281, 272–274.
[23] Li, J., Smith, G.P. and Walker, J.C. (1999) Proc. Natl. Acad. Sci. USA 96, 7821–7826.
[24] Hammet, A., Pike, B.L., Mitchelhill, K.I., Teh, T., Kobe, B., House, C.M., Kemp, B.E. and Heierhorst, J. (2000) FEBS Lett. 471, 141–146.
[25] Boudrez, A., Beullens, M., Groenen, P., Van Eynde, A., Vulsteke, V., Jagiello, I., Murray, M., Kranier, A.R., Stalmans, W. and Bollen, M. (2000) J. Biol. Chem. 275, 25411–25417.
[26] Durocher, D., Henckel, J., Fersht, A.R. and Jackson, S.P. (1999) Mol. Cell 4, 387–394.
[27] Jeanmougin, F., Thompson, J.D., Gouy, M., Higgins, D.G. and Gibson, T.J. (1998) Trends Biochem. Sci. 23, 403–405.