Crystal structure of the BEACH domain reveals an unusual fold and extensive association with a novel PH domain

Gerwald Jögl, Yang Shen, Damara Gebauer, Jiang Li, Katja Wiegmann¹, Hamid Kashkar¹, Martin Krönke¹ and Liang Tong²

Department of Biological Sciences, Columbia University, New York, NY 10027, USA and ¹Institute of Medical Microbiology, Immunology and Hygiene, University of Cologne, D-50935 Köln, Germany

²Corresponding author
e-mail: tong@como.bio.columbia.edu

The BEACH domain is highly conserved in a large family of eukaryotic proteins, and is crucial for their functions in vesicle trafficking, membrane dynamics and receptor signaling. However, it does not share any sequence homology with other proteins. Here we report the crystal structure at 2.9 Å resolution of the BEACH domain of human neurobeachin. It shows that the BEACH domain has a new and unusual polypeptide backbone fold, as the peptide segments in its core do not assume regular secondary structures. Unexpectedly, the structure also reveals that the BEACH domain is in extensive association with a novel, weakly conserved pleckstrin-homology (PH) domain. Consistent with the structural analysis, biochemical studies show that the PH and BEACH domains have strong interactions, suggesting they may function as a single unit. Functional studies in intact cells demonstrate the requirement of both the PH and the BEACH domains for activity. A prominent groove at the interface between the two domains may be used to recruit their binding partners.

Keywords: Chediak–Higashi syndrome/protein structure/TNF signaling/vesicle trafficking

Introduction

Chediak–Higashi syndrome (CHS) is a rare, autosomal recessive disorder that can cause severe immunodeficiency, albinism and other diseases in humans and other mammals (Spritz, 1998; Introne *et al.*, 1999; Ward *et al.*, 2000). Unless treated by bone marrow transplantation, CHS patients generally die in childhood (Spritz, 1998; Introne *et al.*, 1999; Certain *et al.*, 2000). At the cellular level, a hallmark of the CHS disease is the presence of giant, perinuclear vesicles (lysosomes, melanosomes and others) in many cells of the patients (Introne *et al.*, 1999; Dell'Angelica *et al.*, 2000). Therefore, it has been proposed that the CHS protein, with 3801 amino acid residues (Figure 1A) (Barbosa *et al.*, 1996; Nagle *et al.*, 1996; Perou *et al.*, 1996; Ward *et al.*, 2000), may have a crucial function in the fusion, fission or trafficking of these vesicles (Introne *et al.*, 1999; Ward *et al.*, 2000).

A domain of ~300 amino acid residues near the C-terminus of CHS is highly conserved in a large family of eukaryotic proteins (Figure 1B). This domain is known as the BEACH (for beige and CHS) domain (beige is the name for the CHS disease in mice) (Figure 1A) (Nagle *et al.*, 1996). Many of the proteins that contain this domain are very large (with >2000 residues) (Figure 1A) and have putative functions in vesicular transport or membrane dynamics, such as CHS, neurobeachin (Nbea) (Wang *et al.*, 2000), LBA (also known as BGL, CDC4L) (Feuchter *et al.*, 1992; Wang *et al.*, 2001) and LvsA (Kwak *et al.*, 1999; Cornillon *et al.*, 2002). Nbea has been implicated in membrane traffic in neuronal cells. It also contains a motif that binds to the regulatory subunit of protein kinase A (PKA; Wang *et al.*, 2000), and therefore can be classified as an A kinase anchoring protein (AKAP) (Colledge and Scott, 1999) (Figure 1A). LBA may have a function in polarized vesicle trafficking, and is localized to vesicles after stimulation by lipopolysaccharide (LPS) (Wang *et al.*, 2001). Another function for BEACH-containing proteins is illustrated by the protein FAN, which is involved in signaling by the p55 tumor necrosis factor receptor (TNFR1) (Figure 1A) (Adam-Klages *et al.*, 1996; Kreder *et al.*, 1999).

The BEACH domain is crucial for the normal function of these proteins. Studies with FAN showed that its BEACH domain is required for downstream signaling by TNFR1 (Adam-Klages *et al.*, 1996). Genetic analyses of the naturally occurring mutations associated with CHS also reveal the functional importance of the BEACH domain. These are all non-sense or frame-shift mutations, leading to premature termination of the protein and the loss of the BEACH domain (Figure 1A) (Introne *et al.*, 1999; Certain *et al.*, 2000). In particular, one of these mutations occurs within the BEACH domain itself (Figure 1A) (Karim *et al.*, 1997). Moreover, the severity of the disease is not correlated with the length of the remaining CHS protein, suggesting the full-length protein is required for the normal function (Certain *et al.*, 2000).

The exact molecular function of the BEACH domain is currently unknown. Moreover, the domain does not share any recognizable sequence homology to other proteins in the database. To help obtain a complete understanding of this highly conserved domain, we have determined the crystal structure of the BEACH domain of human Nbea at 2.9 Å resolution. The crystal structure reveals that the BEACH domain has a new and unusual polypeptide backbone fold. In addition, our structural studies reveal the presence of a novel, weakly conserved pleckstrin-homology (PH) domain just before the BEACH domain in the primary sequence (Figure 1A). The structural analysis suggests strong interactions between the PH and the BEACH domains, which we have confirmed using protein binding assays. Functional studies with the protein FAN

Fig. 1. Primary structures of the BEACH domain. (A) Schematic drawing of the primary structures of neurobeachin (Nbea), CHS and FAN. The BEACH and WD40 domains are shown in purple and cyan, respectively. The PH domains, identified from the current study, are shown in green. The bars in CHS indicate naturally occurring frame-shift (black) or non-sense (red) mutations in CHS patients. The blue bar in Nbea represents the A-kinase anchoring (AKAP) motif. (B) Alignment of the PH and BEACH domain sequences. The start of the PH and BEACH domains is indicated with the green and purple arrows, respectively. The secondary structure elements (S.S.) are shown and labeled. The residue numbers shown are for Nbea. Residues in the core of the structure are colored green in the Nbea sequence. The symbol '=' represents a residue that is strictly conserved among 30 BEACH domain sequences, and 17 out of 18 identified PH domain sequences. '-' indicates a residue that is identical to that in Nbea, and '.' represents a deletion.

demonstrate that both the PH and the BEACH domains are required to transduce the signal from TNFRI. Our studies therefore suggest that these two domains may function as a single unit. A prominent groove at the interface between these two domains may be important for binding the partner of these domains. The structural information provides a foundation for studying and understanding the role of these BEACH-containing proteins in vesicle trafficking, membrane dynamics and receptor signaling.

Results

Structure determination

To obtain crystals for structural analysis, we screened many BEACH domains from several organisms for their bacterial expression, solution properties and crystallization behavior. An expression construct based on human

Nbea produced crystals suitable for X-ray structure determination. It covers residues 2137–2553 of human Nbea, and contains ~130 additional residues N-terminal to the conserved BEACH domain (Figures 1B and 3A, see below).

The crystal structure of this BEACH domain was determined at 2.9 Å resolution. Initial phases for the reflections were obtained from the selenomethionyl (Se-Met) single- and multi-wavelength anomalous diffraction method (Hendrickson, 1991). After solvent-flattening, several α -helices could be recognized in the electron density map, but the overall quality of the map was still rather poor. The parameters of a non-crystallographic symmetry (NCS) axis were obtained based on the Se positions, and refined in reciprocal space using the solvent-flattened phases (Tong *et al.*, 1992; Tong, 1993). The electron density map after 2-fold NCS averaging was of

Fig. 2. Electron density and crystal packing of the BEACH domain. (A) The electron density after 2-fold NCS averaging at 2.9 Å resolution for strands $\beta 5$, $\beta 6$ and $\beta 7$ in the PH domain. The contour level is at 1σ . (B) The crystal packing. The crystal unit cell is viewed along the c -axis. The PH domains are shown in red and blue, and the BEACH domains are shown in yellow and green. The locations of the 6_1 , 3_1 and 2_1 screw axes are shown. The channel along the c -axis of the unit cell extends through the entire crystal. (A) was produced with Setor (Evans, 1993) and (B) was produced with Molscript and Raster3D (Kraulis, 1991; Merritt and Bacon, 1997).

excellent quality (Figure 2A), and allowed the tracing of the entire protein. The current R -factor for the structure is 23.0% for all observed reflections to 2.9 Å resolution. The statistics for structure determination and refinement are summarized in Table I.

There are two copies of the BEACH domain molecule in the asymmetric unit of the crystal, which belongs to space group $P6_1$. This gives rise to rather high solvent content (75%) and V_m value (4.5 Å³/Da) for the crystal. The molecules are clustered around the 3_1 -screw axes in the unit cell, forming long fibers along the direction of the c -axis of the unit cell (Figure 2B). The contacts between neighboring fibers in the crystal are mediated by crystallographic 2_1 -screw axes, also along the c -axis. Such a packing arrangement produces large channels (~140 Å in diameter) along the 6_1 -screw axes that extend throughout the entire crystal (Figure 2B). The high solvent content of these crystals may be correlated with their poor X-ray diffraction quality.

Light scattering studies showed that the protein exists as monomers in solution (data not shown). In the crystal, the

two molecules in the asymmetric unit are related by an improper NCS, with an angle of rotation of 163° and a translation element of 32 Å. Each molecule has ~700 Å² of buried surface area in this improper dimer, half of which is due to the docking of a loop (residues 2438–2444) from one monomer into a groove on the surface of the other monomer (see below). The equivalent residues in the other monomer (2440'–2452', with ' indicating residues in the second monomer) are disordered.

The BEACH domain has a new and unusual fold

The structural analysis reveals that the BEACH domain has a new polypeptide backbone fold, consistent with its unique amino acid sequence. Moreover, the fold of the BEACH domain is quite unusual in that it contains several segments that are either completely buried in the hydrophobic core or help to enclose it, but these segments can not be classified as β -strands as they are not fully extended (Figure 3A and B). In addition, only a few main-chain hydrogen bonds are made among these segments. Many of the main-chain amides and carbonyls of these

Fig. 3. Structures of the BEACH and PH domains. (A) Schematic stereoview of the structure of the PH-BEACH domains of human Nbea. The PH domain is shown in green, and the linker between the two domains in orange. (B) Schematic stereoview of the structure of the BEACH domain of human Nbea. The extended segments are shown in cyan, the α -helices in yellow, and the loops in purple. The side chains of residues W2279, R2299, R2331, E2423 and R2553 are shown as stick models. (C) Schematic drawing of the structure of the PH domain of human Nbea. The β -strands are shown as arrowed ribbons. Also shown is the linker segment that connects to the BEACH domain (in orange). (D) Superposition of the structure of the PH domains of Nbea (in green) and phospholipase C (PLC, in purple) (Ferguson *et al.*, 1995). The phospholipid ligand in the latter structure is shown as a stick model. The αC - $\epsilon 3$ segment from the BEACH domain (residues 2297-2307) is located in the same position as the $\beta 5$ - $\beta 6$ insertion in the PLC PH domain, which is also the place where several other PH domains bind peptide ligands. Produced with Ribbons (Carson, 1987).

segments are hydrogen-bonded to the side-chains of conserved amino acids in the domain instead.

For ease of discussion, these partially extended segments that help to form the hydrophobic core of the domain are named ε1 to ε7, starting from the N-terminus of the domain (Figure 3B). Of the seven segments, ε1, ε4 and ε7 contain regions that are completely buried in the core of the domain (Figure 3B). These residues are highly conserved among the BEACH domains (Figure 1B), including a buried ion pair between Arg2331 and Glu2423 (Figure 3B). In addition to these segments, the structure also contains 11 α-helices (αA through αK). They are arranged on the periphery of the structure, but help to enclose the core of the domain (Figure 3B).

Based on the structural analysis, the BEACH domain of human Nbea covers residues 2264–2553 and contains 290

residues (Figure 1B). However, the first conserved residue of the domain is Trp2279, and the sequence conservation for the 15 residues N-terminal to this Trp residue (2264–2278) is much weaker than the rest of the BEACH domain (Figure 1B). This exemplifies the difficulty of domain parsing based solely on sequence comparisons. Remarkably, the side chain of Trp2279 shows amino-aromatic interactions with the side chain of Arg2553 (Figure 3B), the last highly conserved residue in the BEACH domain (Figure 1B).

The structure reveals a novel PH domain just prior to the BEACH domain

While the BEACH domain is highly conserved, the amino acid sequences outside this region show much greater variation (Figure 1B). From a detailed sequence analysis, we found that a 130-residue segment just N-terminal to the BEACH domain in the primary sequence is weakly conserved among many of these proteins, with ~20% sequence identity (Figure 1B). Therefore, we engineered expression constructs that contained this additional 130-residue segment at the N-terminus. The crystal structure revealed that this N-terminal segment forms a complete domain in itself (Figure 3A and C). Unexpectedly, the backbone fold of this domain is identical to that of PH domains (Blomberg *et al.*, 1999). Extensive sequence searches failed to classify this region in Nbea as a PH domain, identifying this as a PH domain with a novel sequence.

Like the other PH domains, the PH domain in Nbea contains a seven-stranded β-sandwich (β1 to β7), with an α-helix (α3) near the C-terminus that closes off one of the open ends of the sandwich (Figure 3C). The structure of the β-sandwich core of this PH domain is highly homologous to those of other PH domains, with a root mean squares (r.m.s.) distance of ~2.5 Å for ~80 equivalent Cα atoms (Figure 3D). The homology to the other PH domains at the amino acid sequence level, however, is much lower, in the 6–12% range for structurally equiva-

Table I. Summary of crystallographic information

Data processing statistics			
Wavelength	λ2	λ3	λ4
Maximum resolution (Å)	2.9	2.9	2.9
No. of observations	316 221	355 843	391 076
R _{merge} (%) ^a	8.3	7.9	8.5
No. of reflections	80 562	80 025	79 148
Completeness (%)	99.4	99.2	98.7
Structure refinement statistics			
Resolution range for refinement	30–2.9 Å		
Completeness (%)	99.2		
R-factor ^b (%)	23.0		
Free R-factor (%)	26.4		
R.m.s. deviation in bond lengths (Å)	0.008		
R.m.s. deviation in bond angles (°)	1.3		

^a $R_{\text{merge}} = \frac{\sum_h \sum_i |I_{hi} - \langle I_h \rangle|}{\sum_h \sum_i I_{hi}}$

^b $R = \frac{\sum_h |F_h^o - F_h^c|}{\sum_h F_h^o}$

Fig. 4. The interface between the PH and BEACH domains. (A) Molecular surface of the PH and BEACH domains. Residues in the PH–BEACH interface are shown in yellow for hydrophobic residues, green for polar residues, red for acidic residues, and blue for basic residues. (B) Schematic drawing of part of the interface between the PH and BEACH domains. The exposed residues of the back sheet (β5, β6 and β7) of the PH domain, shown in green, interact with the αC–ε1 linker of the BEACH domain (in purple). (A) was produced with Grasp (Nicholls *et al.*, 1991) and (B) was produced with Ribbons (Carson, 1987).

Fig. 5. Biochemical evidence for interactions between the PH and BEACH domains. (A) Protein binding assays with wild-type PH (as GST fusion) and BEACH domains of human FAN. To show the dose-responsive nature of the interaction, the amount of GST-PH was increased, while that of BEACH was held constant. The specificity of the interaction was indicated with the GST control. (B) Binding isotherm of GST-PH and BEACH. The experimental observations from the GST pull-down experiment are shown in the inset and plotted as the blue dots. The red line represents a hyperbolic fit to the experimental data. (C) Mutations in the PH-BEACH interface disrupts their interactions. Mutations in the PH and BEACH domains were selected based on the structural information. The two control mutations are indicated with red asterisks.

lent residues. In addition, the Nbea PH domain contains an insertion of two helices ($\alpha 1$ and $\alpha 2$) between strands $\beta 3$ and $\beta 4$ (Figure 3C). The low sequence homology, together with the unique pattern of insertions/deletions, may be the reason why sequence analyses failed to identify this region in Nbea as a PH domain.

PH domains are known to have a variety of functions, including phospholipid binding, phosphotyrosine binding [for the related phosphotyrosine binding (PTB) domains] and protein-protein interactions (Blomberg *et al.*, 1999). The phospholipids generally bind on either side of the loop connecting $\beta 1$ and $\beta 2$ (Figure 3D). In the PH domain of Nbea, however, one of these binding sites is occupied by the $\alpha 2$ helix, while the other is blocked by residues 2302–2307 from the BEACH domain (Figure 3D). The structural analyses therefore suggest that the PH domain of Nbea is probably not involved in phospholipid binding. Instead, this domain may be involved in protein-protein interactions (see below).

An extensive, conserved interface between the PH and BEACH domains

The structural analysis revealed an extensive interface between the PH and BEACH domains in Nbea (Figure 3A). About 1100 Å² of surface area in each domain is buried at

the interface, which is mostly hydrophobic or polar in nature (Figure 4A). A major portion of this interface consists of residues 2295–2306 from the BEACH domain (the αC - $\epsilon 1$ linker) packing against the ‘back sheet’ (consisting of strands $\beta 5$, $\beta 6$, $\beta 7$ and $\beta 1$) of the β -sandwich of the PH domain (Figure 4B). Interestingly, the equivalent surface areas in several other PH domains are also used for protein-protein interactions (Blomberg *et al.*, 1999).

Many of the residues in the PH-BEACH interface are conserved among these proteins. For example, four of the five highly conserved residues in the PH domains (Figure 1B) are located in this interface, a remarkable distribution considering the low degree of sequence conservation of the PH domains. These conserved residues include Arg2208 (in strand $\beta 5$) and Glu2218 ($\beta 6$), which form an ion pair on the surface of the PH domain and are buried in this interface (Figure 4B). In addition, Arg2208 is located near the C-terminus of helix αC in the BEACH domain, and may therefore have favorable interactions with the dipole of this helix (Figure 4B). This sequence conservation suggests that similar PH-BEACH interfaces may be present in other BEACH-containing proteins.

The linker between the PH and BEACH domains covers residues 2246–2263 in Nbea (Figure 3A). Neither its sequence nor its length are conserved among the different proteins. Nbea appears to have an insertion of >13 residues as compared with CHS or FAN in this region (Figure 1B), while the yeast YCR032W protein has an insertion of >40 residues. The structure shows, however, that the C α atoms of 2248 and 2262 in Nbea are within 5 Å of each other (Figure 3C), suggesting that insertions (such as in YCR032W) or deletions (such as in CHS and FAN) in this linker may be accommodated with minimal disturbances to the PH-BEACH interface.

Biochemical evidence for interactions between the PH and BEACH domains

The extensive interface between the PH and BEACH domains suggests there may be strong interactions between the two domains. To obtain biochemical evidence for such interactions, we performed protein binding assays using purified PH domain fused with glutathione S-transferase (GST-PH) and the purified His-tagged BEACH domain of the protein FAN (Figure 1A) (Adam-Klages *et al.*, 1996). The experiments clearly demonstrate the strong and dose-responsive interactions between the PH and the BEACH domains of FAN (Figure 5A). This confirms the structural information, and also supports the conserved nature of the PH-BEACH interface as observed in our Nbea structure.

To obtain an estimate for the K_d of the interactions between these two domains, we immobilized a constant amount of the GST-PH domain on glutathione-agarose beads and introduced increasing amounts of the BEACH domain. The resulting binding data can be readily fitted to a hyperbolic curve, suggesting a 1:1 ratio in the interaction between these two domains (Figure 5B). The K_d value obtained from this binding isotherm is $\sim 1 \mu M$, consistent with the structural information. However, in the native protein, the two domains are linked covalently. This should produce an even stronger interaction between the two domains.

Fig. 6. Functional studies of the PH and BEACH domains of FAN. (A) The activity of neutral sphingomyelinase after TNF activation was determined for mouse FAN^{-/-} fibroblasts transfected with vector control, full-length FAN, BEACH-WD40 ($\Delta 1-275$) and PH-BEACH-WD40 ($\Delta 1-178$) deletion mutants. (B) The effects of FAN mutants in the activation of neutral sphingomyelinase in mouse FAN^{-/-} fibroblasts. The error bars represent observations from three independent experiments.

To characterize further the interactions at the PH-BEACH interface, we introduced mutations in this interface in the protein FAN based on the structural information, and determined their effects on the interactions by the protein binding assay. The mutants created include R246E (R246 in FAN is equivalent to R2208 in Nbea) and E256R (2218) in the PH domain, and N328A (2302), Q332A (2306), Y409D (2388) and Q562A (2540) in the BEACH domain. (R2208, E2218, N2302 and Q2306 can be seen in Figure 4B, and residues 2388 and 2540 can be seen in Figure 4A.) Two additional mutants, Q234K (2196) in the PH domain and D495Q (2473) in the BEACH domain, were generated as controls, as these residues are poorly conserved and located outside the PH-BEACH interface based on the structure (Figure 1B).

Our results showed that mutations at the interface in either the PH or the BEACH domain can significantly affect the interactions between the two domains, whereas mutations outside this interface had little effect (Figure 5C). In particular, the R246E and E256R mutations in the PH domain reduced the interactions such that they were barely detectable by Coomassie Blue staining, consistent with the important roles of these residues in the

interface from the structural information. Similarly, the effects of mutations in the BEACH domain are generally in agreement with the structural observation (Figure 4B). For example, the N328A mutation has a greater effect on the interaction than the Q332A mutation, and N328 has a larger surface area burial ($120 \text{ \AA}^2$) than Q332 ($41 \text{ \AA}^2$) in the interface. Overall, the results from the protein binding experiments confirm our observations from the crystal structure. Therefore, the strong association between the PH and BEACH domains may be a conserved feature, and the two domains may function as a single unit.

Both the PH and the BEACH domains are required for signaling by FAN

Next, we characterized the effects of the PH and BEACH domains on the biological functions of the protein FAN. We selected this protein for assessing the structural information as it is one of the few BEACH-containing proteins that have been examined in some detail (Adam-Klages *et al.*, 1996; Kreder *et al.*, 1999). Prior studies demonstrated that the WD40 domain in this protein, located at the extreme C-terminus (Figure 1A), is necessary and sufficient for binding to a peptide segment in the intracellular domain of TNFRI (Adam-Klages *et al.*, 1996). Transfection experiments showed that this domain by itself has dominant-negative effects on FAN signaling, suggesting that the other parts of the molecule (BEACH and N-terminal domains; Figure 1A) may have crucial functions in downstream events in the pathway.

To assess the functional roles of the PH and BEACH domains in the signal transduction by FAN, we created the BEACH-WD40 (missing the N-terminal 275 residues, $\Delta 1-275$) and the PH-BEACH-WD40 ($\Delta 1-178$) deletion mutants. The ability of these mutants to rescue signaling in FAN^{-/-} mouse fibroblasts (Kreder *et al.*, 1999) were then determined. While the BEACH-WD40 mutant could not rescue the knock-out phenotype, the PH-BEACH-WD40 mutant restored wild-type level FAN activity to these cells (Figure 6A). Moreover, the BEACH-WD40 mutant produced a dominant-negative effect (95% inhibition) when transfected into human embryonic kidney (HEK) 293 cells that express endogenous wild-type FAN. In contrast, transfection of the wild-type or the PH-BEACH-WD40 mutant has little effect on the function of the endogenous FAN. This shows that the BEACH domain cannot function independently of the PH domain in FAN signaling, and is therefore consistent with our structural and biochemical observations that the PH and BEACH domains function as a single unit.

Moreover, single-site mutations in the PH-BEACH interface can also reduce the signaling by the protein FAN (Figure 6B), supporting the functional importance of this interface. The expression levels of the different mutants were checked using anti-FLAG western analyses. Although there are some inconsistencies in individual experiments, we obtained functional data on every single-site mutant that is expressed at significant levels. A clear reduction in the signaling activity of FAN was observed for those mutants in the PH-BEACH interface (R246A, F258A and N328A) that have comparable expression levels. The effects of mutations at the Arg246 and Asn328 positions have also been assessed by the GST pull-down experiments (Figure 5C). The effects of the mutations in

Fig. 7. A groove at the interface between PH and BEACH domains. (A) Molecular surface of the PH-BEACH domains, with the PH domain in green and BEACH domain in purple. Strictly conserved residues are colored in yellow, and mostly conserved residues in cyan. (B) Docking of residues in the $\epsilon 4$ - $\epsilon 5$ loop of one BEACH molecule into the groove of the other in the crystal. The molecular surface of the second PH-BEACH molecule (residue numbers 2137'-2553') is colored according to electrostatic potential. Produced with Grasp (Nicholls *et al.*, 1991).

the signaling assay (Figure 6B) are smaller than those in the biochemical assay (Figure 5C), which may be for two reasons. First, the biochemical assays are based on Coomassie Blue staining. Weaker interactions cannot be detected with this assay, but such interactions might still be able to function in FAN signaling. Secondly, and perhaps more importantly, the biochemical assays detect the interactions of PH and BEACH domains as separate entities, while the two domains are covalently linked in the natural FAN protein. This should increase the affinity between the two domains and dampen the effect of mutations in their interface. We have also checked the effects of mutations outside this interface. The Asn459 and Leu463 residues are in the $\epsilon 4$ - $\epsilon 5$ loop, one of the longest loops in the BEACH domain structure (Figure 3B). Their mutations have little impact on the function of FAN

(Figure 6B). Glu448 (equivalent to Glu2427 in Nbea) is highly conserved and partly exposed on the surface (Figure 7B). Our mutation result shows that this residue is not required for the function of FAN (Figure 6B).

A possible binding site at the interface between the PH and BEACH domains

To identify possible biological functions for the BEACH domain based on the crystal structure, we first examined the locations of the conserved residues in this domain. This analysis showed that most of these residues are located in the core of the domain (Figure 1B). Few of the strictly conserved residues that could have catalytic activities are exposed on the surface of the BEACH domain (Figure 7A), and we have mutated some of these residues in the functional assay (Figure 6). Therefore, based on the currently available structural and mutagenesis data, it is unlikely that the BEACH domain is an enzyme.

Our structural and biochemical evidence showed that the BEACH domain may function together with the PH domain. There is a prominent groove at the interface between the two domains in the current structure, and there is a higher concentration of exposed, mostly conserved residues in this region (Figure 7A). Moreover, in the improper dimer in the crystal, the loop between the $\epsilon 4$ and $\epsilon 5$ segments of one molecule is docked into part of this groove of the other molecule (Figure 7B). The strongest interaction in this area is mediated by residues Tyr2441-Asn-Leu2443 (Figure 7B). Preliminary binding assays based on fluorescence perturbation measurements suggest that the affinity between a 12-residue peptide (2436-VNSNGYNLGVRE-2447) and Nbea is in the micromolar range, whereas little binding was observed for the equivalent peptide carrying the Y2441A mutation (data not shown). While it is probably unlikely that this mode of interaction between the two BEACH domain molecules is biologically relevant, this observation does offer the tantalizing suggestion that this groove between the PH and BEACH domains may be involved in binding a partner (possibly a peptide segment) in the natural function of the PH-BEACH domains.

Discussion

The BEACH domain has so far only been found in eukaryotes, and the sequences of this domain are highly conserved among these proteins. For example, the BEACH domains of human Nbea (Wang *et al.*, 2000) and yeast YCR032W share 46% amino acid sequence identity. A total of 31 BEACH domain sequences are currently known, which show that many genomes carry more than one BEACH-containing protein. The human genome may have eight such proteins, including CHS, Nbea, FAN, LBA and KIAA1607 (Nagase *et al.*, 2000).

Nbea and LBA belong to a subfamily of BEACH-containing proteins that have orthologs in many other organisms, such as AKAP550 in *Drosophila melanogaster* (Han *et al.*, 1997) and the uncharacterized open reading frame F10F2.1 in the *Caenorhabditis elegans* genome. In addition, many of the proteins in this subfamily can also be classified as AKAPs, as they contain the motif that binds to the regulatory subunit of PKA (Colledge and Scott, 1999). Therefore, one function of these proteins may be to direct

PKA to proper locations in the cell. However, the AKAP motif covers only ~20 amino acid residues, and therefore accounts for only a tiny fraction of the residues in these proteins (Figure 1A).

In all the BEACH-containing proteins, the BEACH domain is located just prior to a WD40 domain in the primary sequence, which is located at the extreme C-terminus of most of these proteins (Figure 1A). The WD40 domain is found in a large family of proteins, and is believed to mediate protein–protein interactions (Neer *et al.*, 1994). This domain in the protein FAN is necessary and sufficient for the interactions with TNFRI (Adam-Klages *et al.*, 1996), thereby mediating the recruitment of FAN to the receptor. The function of the WD40 domains in the other BEACH-containing proteins is currently unknown, although it is likely that these domains are also involved in recruiting the proteins to the proper locations in the cell.

The BEACH domain is the only domain that is highly conserved among this family of proteins, many of which contain >2000 amino acid residues (Figure 1A). This strong conservation gives the impression that this domain might function as an independent module (a cassette) in these proteins. Surprisingly, our studies showed that the BEACH domain has intimate contacts with a novel PH domain just before it in the primary sequence. The biochemical and biological data confirm that the two domains may function as a unit, even though the PH domain is conserved at a much lower level. Therefore these two domains do not behave like ‘beads on a string’, as might be implied from sequence comparisons. It will be interesting to study whether the functions of these proteins require close associations between the PH–BEACH unit and other domains within them (e.g. the WD40 domain).

In most protein structures, the main-chain segments that are buried in the core of the structure assume regular secondary structures (α -helix or β -sheet). The structure of the BEACH domain is rather unusual in that none of the segments in the core of the structure assume regular α - or β -conformation (Figure 3B), and the BEACH domain lacks extensive main-chain hydrogen-bonding interactions among the segments (ϵ 1– ϵ 7) in its core. Therefore, interactions involving side chains are likely to be extremely important for the stability of this domain. This is consistent with our observation that most of the conserved residues are located in the core of the domain, and the high degree of sequence conservation of this domain may be required to stabilize its fold.

In contrast, our identification of a PH domain with novel amino acid sequences demonstrates the strong stability of the PH fold, which is probably provided by the many main-chain hydrogen-bonding interactions among its seven β -strands (Figure 3C). This would in turn allow large variation in the side chains, and hence little conservation of the amino acid sequences of the domain. The resulting differences in the surface decorations on this fold define the unique biochemical functions of the various PH domains, such as phospholipid binding, phospho-tyrosine binding and protein–protein interactions (Blomberg *et al.*, 1999).

The large sizes of the BEACH-containing proteins generally make it difficult to study their functions. Moreover, with the exception of the WD40 domain, the

amino acid sequences of these proteins appear to be unique. As illustrated by the BEACH domain, they are not homologous to other protein sequences in the database, which makes it nearly impossible to infer the biological functions of these proteins based solely on sequence analysis. Therefore, elucidating the three-dimensional structures of these proteins is a crucial component in understanding their functions. Our structural, biochemical and functional analyses of the PH and BEACH domains of human Nbea represent the first step in this process, and suggest that there may be many more surprises in the studies of this important family of proteins.

Materials and methods

Protein expression and purification

Residues 229–645 of the KIAA1544 protein (Nagase *et al.*, 2000), a putative human ortholog of murine Nbea (sharing 99% amino acid sequence identity) (Gilbert *et al.*, 1999; Wang *et al.*, 2000), was subcloned into the pET28a vector (Novagen) and overexpressed in *Escherichia coli* at 20°C. The expression construct contains an N-terminal His₆ tag, and covers the conserved BEACH domain together with an additional 130 residues at the N-terminal end. The soluble protein was bound to nickel–agarose affinity resin (Qiagen), and eluted with a buffer containing 20 mM Tris pH 8.5, 250 mM NaCl and 150 mM imidazole. The protein was purified further by anion exchange chromatography at pH 8.5, using a linear gradient of 10–500 mM NaCl concentration. Finally, the protein sample was purified by gel filtration chromatography, in a running buffer containing 20 mM Tris pH 8.5, 200 mM NaCl and 10 mM dithiothreitol (DTT). The protein fractions from this column were pooled and concentrated to 30 mg/ml. Glycerol was added to 5% (v/v) concentration, and the protein sample was flash-frozen in liquid nitrogen and stored at –80°C. The N-terminal His tag was not removed for crystallization.

For the production of selenomethionyl proteins, the expression construct was transformed into DL41(DE3) cells. Bacterial growth was carried out in defined LeMaster media (Hendrickson *et al.*, 1990), and the protein was purified using the same protocol as for the wild-type protein. The successful incorporation of selenomethionyl residues was confirmed by MALDI-TOF mass spectrometry.

Protein crystallization

Crystals of the BEACH domain were obtained at 4°C by the sitting-drop vapor diffusion method. The reservoir solution contained 100 mM Tris pH 7.6, 4% (w/v) PEG8000 and 2 mM DTT. Crystals in the shape of hexagonal prisms generally took 2 weeks to grow to full size ($0.3 \times 0.3 \times 1$ mm³). For cryo protection, the crystals were transferred in a few steps to an artificial mother liquor containing 20 mM Tris pH 8.3, 5% (w/v) PEG8000 and 25% (v/v) PEG200. The largest crystals generally did not survive this treatment, and X-ray diffraction analyses and data collection were performed with crystals of size $0.1 \times 0.1 \times 0.6$ mm³. Most of the crystals were highly mosaic and had very weak X-ray diffraction.

Data collection

X-ray diffraction data to 2.9 Å resolution were collected on a Mar CCD at the 32-ID beamline (ComCAT) of the Advanced Photon Source (APS) (Table I). Four wavelengths were used for collecting the selenomethionyl MAD data: λ 1 (12 500 eV, 0.9764 Å, low-energy remote), λ 2 (12 660 eV, 0.9793 Å, edge), λ 3 (12 663 eV, 0.9791 Å, peak) and λ 4 (12 800 eV, 0.9686 Å, high-energy remote). The crystal-to-detector distance was 230 mm, and inverse beam geometry was used to collect the anomalous data. The diffraction images were processed and scaled with the HKL package (Otwinowski and Minor, 1997). The crystal belongs to the space group $P6_1$, with cell dimensions of $a = b = 179.9$ Å, and $c = 98.5$ Å. Even though the crystal was kept at 100 K throughout the data collection, significant decay in the X-ray diffraction of the crystal was observed. The data set at λ 1, which was collected last, was not used in phasing due to this decay.

Structure determination and refinement

The locations of six selenium atoms were determined from the anomalous difference Patterson maps with the program Patsol (Tong and Rossmann,

1993), using the data collected at the peak wavelength (λ_3). Subsequent least-squares refinement against these anomalous differences, with the program MADSYS (Hendrickson, 1991), revealed the positions of 12 additional, weaker Se sites. Initial phase information was calculated using diffraction data at the peak wavelength only (SAD phasing) (Wang, 1985), as well as using three wavelengths (MAD phasing) with the program Mphare (CCP4, 1994). After solvent flattening, several helices could be recognized in the electron density map.

We expected several copies of the protein molecule in the asymmetric unit of the crystal. However, self-rotation functions could not define the orientation of the non-crystallographic symmetry (NCS) axes in the crystal (Tong and Rossmann, 1990). The NCS axis was found based on the positions of the Se atoms. Each BEACH domain molecule studied here contains eight Met residues (excluding the two Met residues at the N-terminus for the introduction of the His tag). The presence of 16 Se sites suggests that there are only two molecules in the asymmetric unit, giving a V_m of 4.5 Å³/Da and a solvent content of ~75%. The two molecules are related by improper NCS. The parameters of the NCS axis were refined based on the solvent-flattened phase information with the program GLRF (Tong *et al.*, 1992; Tong, 1993). The phases were then improved by 2-fold NCS averaging and solvent flattening, using a locally written program (L.Tong, unpublished data). The resulting electron density map was of excellent quality and could be easily interpreted based on the sequence of the BEACH domain (Figure 2). The atomic model was built into the electron density with the program O (Jones *et al.*, 1991).

The structure refinement was carried out with the program CNS (Brunger *et al.*, 1998), using all the observed reflections between 30 and 2.9 Å resolution in the data set collected at λ_3 . The statistics on the structure refinement are summarized in Table I.

Protein binding assays

The PH domain of FAN (residues 183–298) was expressed and purified as a GST fusion protein, immobilized with glutathione-agarose, and incubated with the BEACH domain of FAN (residues 295–579) that has been expressed and purified as a His-tagged protein. The incubation buffer contained 20 mM Tris pH 8.5, 300 mM NaCl and 1 mM EDTA. After washing, the bound proteins were eluted, separated by SDS-PAGE, and stained with Coomassie Blue. Mutations at the PH-BEACH interface were designed based on the structural information. The mutants were made with the QuikChange kit (Stratagene) and sequenced for confirmation. They were purified and assayed for protein interaction under the same condition as the wild-type protein.

Functional studies with FAN mutants

The deletion mutants were made using PCR and the site-specific mutants were made using the QuikChange kit (Stratagene). The expression vectors were transfected into mouse FAN^{-/-} fibroblasts (Kreder *et al.*, 1999). After 2 days, the cells were stimulated with TNF, and the activity of neutral sphingomyelinase was determined following protocols reported earlier (Adam-Klages *et al.*, 1996).

Atomic coordinates

The atomic coordinates have been deposited in the Protein Data Bank (accession code 1MI1).

Acknowledgements

We thank Kevin D'Amico and Steve Wasserman for setting up the beamline at APS, Randy Abramowitz and Craig Ogata for setting up the beamline at the National Synchrotron Light Source, and Kazusa DNA Research Institute for providing the KIAA1544 cDNA. We thank Reza Khayat and Zhiru Yang for help with data collection at the synchrotron sources, and Hao Wu for helpful discussions. This research is supported in part by a grant (GM066753 to L.T.) from the National Institutes of Health and Center for Molecular Medicine, Cologne (to M.K.).

References

Adam-Klages, S., Adam, D., Wiegmann, K., Struve, S., Kolanus, W., Schneider-Mergener, J. and Kronke, M. (1996) FAN, a novel WD-repeat protein, couples the p55 TNF-receptor to neutral sphingomyelinase. *Cell*, **86**, 937–947.
 Barbosa, M.D.F.S. *et al.* (1996) Identification of the homologous beige and Chediak-Higashi syndrome genes. *Nature*, **382**, 262–265.
 Blomberg, N., Baraldi, E., Nilges, M. and Saraste, M. (1999) The PH

superfold: a structural scaffold for multiple functions. *Trends Biochem. Sci.*, **24**, 441–445.
 Brunger, A.T. *et al.* (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D*, **54**, 905–921.
 Carson, M. (1987) Ribbon models of macromolecules. *J. Mol. Graph.*, **5**, 103–106.
 CCP4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D*, **50**, 760–763.
 Certain, S. *et al.* (2000) Protein truncation test of LYST reveals heterogeneous mutations in patients with Chediak-Higashi syndrome. *Blood*, **95**, 979–983.
 Colledge, M. and Scott, J.D. (1999) AKAPs: from structure to function. *Trends Cell Biol.*, **9**, 216–221.
 Cornillon, S., Dubois, A., Bruckert, F., Lefkir, Y., Marchetti, A., Benghezal, M., de Lozanne, A., Letourneur, F. and Cosson, P. (2002) Two members of the beige/CHS (BEACH) family are involved at different stages in the organization of the endocytic pathway in *Dictyostelium*. *J. Cell Sci.*, **115**, 737–744.
 Dell'Angelica, E.C., Mullins, C., Caplan, S. and Bonifacio, J.S. (2000) Lysosome-related organelles. *FASEB J.*, **14**, 1265–1278.
 Evans, S.V. (1993) SETOR: hardware lighted three-dimensional solid model representations of macromolecules. *J. Mol. Graph.*, **11**, 134–138.
 Ferguson, K.M., Lemmon, M.A., Schlessinger, J. and Sigler, P.B. (1995) Structure of the high affinity complex of inositol triphosphate with a phospholipase C pleckstrin homology domain. *Cell*, **83**, 1037–1046.
 Feuchter, A.E., Freeman, J.D. and Mager, D.L. (1992) Strategy for detecting cellular transcripts promoted by human endogenous long terminal repeats: identification of a novel gene (CDC4L) with homology to yeast CDC4. *Genomics*, **13**, 1237–1246.
 Gilbert, D.J., Engel, H., Wang, X., Grzeschik, K.-H., Copeland, N.G., Jenkins, N.A. and Kilimann, M.W. (1999) The neurobeachin gene (Nbea) identifies a new region of homology between mouse central chromosome 3 and human chromosome 13q13. *Mamm. Genome*, **10**, 1030–1031.
 Han, J.-D., Baker, N.E. and Rubin, C.S. (1997) Molecular characterization of a novel A kinase anchor protein from *Drosophila melanogaster*. *J. Biol. Chem.*, **272**, 26611–26619.
 Hendrickson, W.A. (1991) Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science*, **254**, 51–58.
 Hendrickson, W.A., Horton, J.R. and LeMaster, D.M. (1990) Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J.*, **9**, 1665–1672.
 Introne, W., Boissy, R.E. and Gahl, W.A. (1999) Clinical, molecular and cell biological aspects of Chediak-Higashi syndrome. *Mol. Gen. Metab.*, **68**, 283–303.
 Jones, T.A., Zou, J.Y., Cowan, S.W. and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A*, **47**, 110–119.
 Karim, M.A., Nagle, D.L., Kandil, H.H., Burger, J., Moore, K.J. and Spritz, R.A. (1997) Mutations in the Chediak-Higashi syndrome gene (CHS1) indicate requirement for the complete 3801 amino acid CHS protein. *Hum. Mol. Gen.*, **6**, 1087–1089.
 Kraulis, P.J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.*, **24**, 946–950.
 Kreder, D. *et al.* (1999) Impaired neutral sphingomyelinase activation and cutaneous barrier repair in FAN-deficient mice. *EMBO J.*, **18**, 2472–2479.
 Kwak, E., Gerald, N., Laroche, D.A., Vithalani, K.K., Niswonger, M.L., Maready, M. and de Lozanne, A. (1999) LvsA, a protein related to the mouse beige protein, is required for cytokinesis in *Dictyostelium*. *Mol. Biol. Cell*, **10**, 4429–4439.
 Merritt, E.A. and Bacon, D.J. (1997) Raster3D—photorealistic molecular graphics. *Methods Enzymol.*, **277**, 505–524.
 Nagase, T., Kikuno, R., Nakayama, M., Hirosawa, M. and Ohara, O. (2000) Prediction of the coding sequences of unidentified human genes. XVIII. The complete sequences of 100 new cDNA clones from brain which code for large proteins *in vitro*. *DNA Res.*, **7**, 273–281.
 Nagle, D.L. *et al.* (1996) Identification and mutation analysis of the complete gene for Chediak-Higashi syndrome. *Nat. Genet.*, **14**, 307–311.
 Neer, E.J., Schmidt, C.J., Narabudripad, R. and Smith, T.F. (1994) The ancient regulatory-protein family of WD-repeat proteins. *Nature*, **371**, 297–300.

- Nicholls,A., Sharp,K.A. and Honig,B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*, **11**, 281–296.
- Otwinowski,Z. and Minor,W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.
- Perou,C.M. *et al.* (1996) Identification of the murine beige gene by YAC complementation and positional cloning. *Nat. Genet.*, **13**, 303–308.
- Spritz,R.A. (1998) Genetic defects in Chediak–Higashi syndrome and the beige mouse. *J. Clin. Immunol.*, **18**, 97–105.
- Tong,L. (1993) Replace: a suite of computer programs for molecular-replacement calculations. *J. Appl. Crystallogr.*, **26**, 748–751.
- Tong,L. and Rossmann,M.G. (1990) The locked rotation function. *Acta Crystallogr. A*, **46**, 783–792.
- Tong,L. and Rossmann,M.G. (1993) Patterson-map interpretation with noncrystallographic symmetry. *J. Appl. Crystallogr.*, **26**, 15–21.
- Tong,L., Choi,H.-K., Minor,W. and Rossmann,M.G. (1992) The structure determination of Sindbis virus core protein using isomorphous replacement and molecular replacement averaging between two crystal forms. *Acta Crystallogr. A*, **48**, 430–442.
- Wang,B.-C. (1985) Resolution of phase ambiguity in macromolecular crystallography. *Methods Enzymol.*, **115**, 90–112.
- Wang,J.-W., Howson,J., Haller,E. and Ker,W.G. (2001) Identification of a novel lipopolysaccharide-inducible gene with key features of both A kinase anchor proteins and chs1/beige proteins. *J. Immunol.*, **166**, 4586–4595.
- Wang,X., Herberg,F.W., Laue,M.M., Wullner,C., Hu,B., Petrasch-Parwez,E. and Kilimann,M.W. (2000) Neurobeachin: a protein kinase A-anchoring, beige/Chediak–Higashi protein homolog implicated in neuronal membrane traffic. *J. Neurosci.*, **20**, 8551–8565.
- Ward,D.M., Griffiths,G.M., Stinchcombe,J.C. and Kaplan,J. (2000) Analysis of the lysosomal storage disease Chediak–Higashi syndrome. *Traffic*, **1**, 816–822.

*Received May 17, 2002; revised July 23, 2002;
accepted July 31, 2002*