Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Conserved solvent and side-chain interactions in the 1.35 Å structure of the Kelch domain of Keap1

Lesa J. Beamer,[a]* Xuchu Li,[a,b] Christopher A. Bottoms[c] and Mark Hannink[a,b]

[a]Department of Biochemistry, University of Missouri-Columbia, Columbia, MO 65211, USA, [b]Life Sciences Center, University of Missouri-Columbia, Columbia, MO 65211, USA, and [c]Department of Chemistry, University of Missouri-Columbia, Columbia, MO 65211, USA

Correspondence e-mail: beamerl@missouri.edu

Received 6 May 2005
Accepted 14 July 2005

PDB Reference: Kelch domain of Keap1, 1zgk, r1zgksf.

The Kelch repeat is a common sequence motif in eukaryotic genomes and is approximately 50 amino acids in length. The structure of the Kelch domain of the human Keap1 protein has previously been determined at 1.85 Å, showing that each Kelch repeat forms one blade of a six-bladed β-propeller. Here, use of 1.35 Å SAD data for *de novo* structure determination of the Kelch domain and for refinement at atomic resolution is described. The high quality and resolution of the diffraction data and phase information allows a detailed analysis of the role of solvent in the structure of the Kelch repeat. Ten structurally conserved water molecules are identified in each blade of the Kelch β-propeller. These appear to play distinct structural roles that include lining the central channel of the propeller, interacting with residues in loops between strands of the blade and making contacts with conserved residues in the Kelch repeat. Furthermore, we identify a conserved C—H$\cdots\pi$ hydrogen bond between two key residues in the consensus Kelch repeat. This analysis extends our understanding of the structural roles of conserved residues in the Kelch repeat and highlights the potential role of solvent in maintaining the fold of this common eukaryotic structural motif.

1. Introduction

The Keap1 protein participates in a key signal transduction pathway in mammalian cells, in which it functions to sequester the Nrf2 transcription factor in the cytosol, thereby targeting it for degradation by the 26S proteosome. Keap1 is a multi-domain protein containing an N-terminal BTB domain, a central conserved linker region and a C-terminal Kelch domain. BTB-Kelch proteins are a large and evolutionary conserved family, with ~60 representatives in the human genome. Kelch domains typically contain six repeated sequence motifs, generally 45–55 residues in length, which were first identified in the *Drosophila melanogaster* Kelch protein (Xue & Cooley, 1993). Each Kelch repeat contains several highly conserved residues and is predicted to form one blade of a β-propeller structure.

Previously, we have determined the crystal structure of the Kelch domain of Keap1 at 1.85 Å resolution, providing the first view of a mammalian Kelch domain and the first high-quality template for modeling other members of the large BTB-Kelch family (Li *et al.*, 2004*b*). This structure shows that the Kelch domain of Keap1 is a highly symmetric six-bladed β-propeller, in which a strand from the C-terminus of the protein closes the ring of the propeller (Fig. 1*a*). Key residues in the consensus Kelch motif, including two adjacent glycine residues and a Tyr/Trp pair, form extended hydrogen-bond networks. The highly conserved Tyr/Trp pair packs in the

protein core and participates in amino-acid contacts involved in inter-blade interactions, while the glycine doublet participates in a conserved hydrogen-bond network within each blade.

In this report, we describe the structure determination of the Kelch domain of human Keap1 to 1.35 Å *via* single-wavelength anomalous diffraction (SAD). Owing to the high quality of the SAD data, phases were available to nearly the resolution limit of the crystals, providing unbiased experimental electron-density maps. Using these phases, the model was refined to 1.35 Å, revealing new insights into the structure of the Kelch motif, including ten structurally conserved solvent molecules in each blade and a C—H$\cdots\pi$ hydrogen bond between key residues in the motif. This work extends the understanding of the structural features of the Kelch repeat and has implications for understanding the fold and sequence preference of members of the extended BTB-Kelch protein family.

2. Methods

2.1. Crystallization and data collection

Purification and crystallization of SeMet Kelch was carried out as previously described (Li *et al.*, 2004*a*). A single SeMet Kelch crystal was used to collect a highly redundant SAD data set at 93 K on beamline SBC 19-ID at the Advanced Photon Source, Argonne National Laboratory. The data set consisted of 360 frames with an oscillation angle of 0.5° per frame, a detector distance of 125 mm, a detector θ of −5° and an exposure time of 3 s per frame. All data were indexed and

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the outer resolution shell (1.4–1.35 Å).

SAD data collection	
Wavelength (Å)	0.979
Space group	$P6_522$
Unit-cell parameters (Å)	$a = b = 85.72$, $c = 148.66$
No. of molecules per AU	1
Resolution (Å)	50–1.35
Mosaicity (°)	0.17
No. of observations	1244831
No. of unique reflections	71295
Redundancy	17.5 (10.0)
R_{merge} (%)	7.9 (45.2)
Mean $I/\sigma(I)$	86.8 (3.1)
Completeness (%)	99.9 (100.0)
Refinement statistics	
Resolution range (Å)	28.1–1.34
R_{cryst}†	12.8
R_{free}‡	14.2
R.m.s.d. bond distance (Å)	0.010
R.m.s.d. bond angle (°)	1.26
Total No. of non-H atoms in AU	2527
No. of solvent molecules	377
B value from Wilson plot (Å^2)	17.1
Average protein B value (Å^2)	9.9
Average solvent B value (Å^2)	22.2

† $R_{cryst} = \sum|F_o - F_c|/\sum|F_c|$, where F_o and F_c are observed and calculated structure factors, respectively. ‡ R_{free} is the R factor calculated from 5% of the reflections not included in refinement. No σ-cutoff of the data was used.

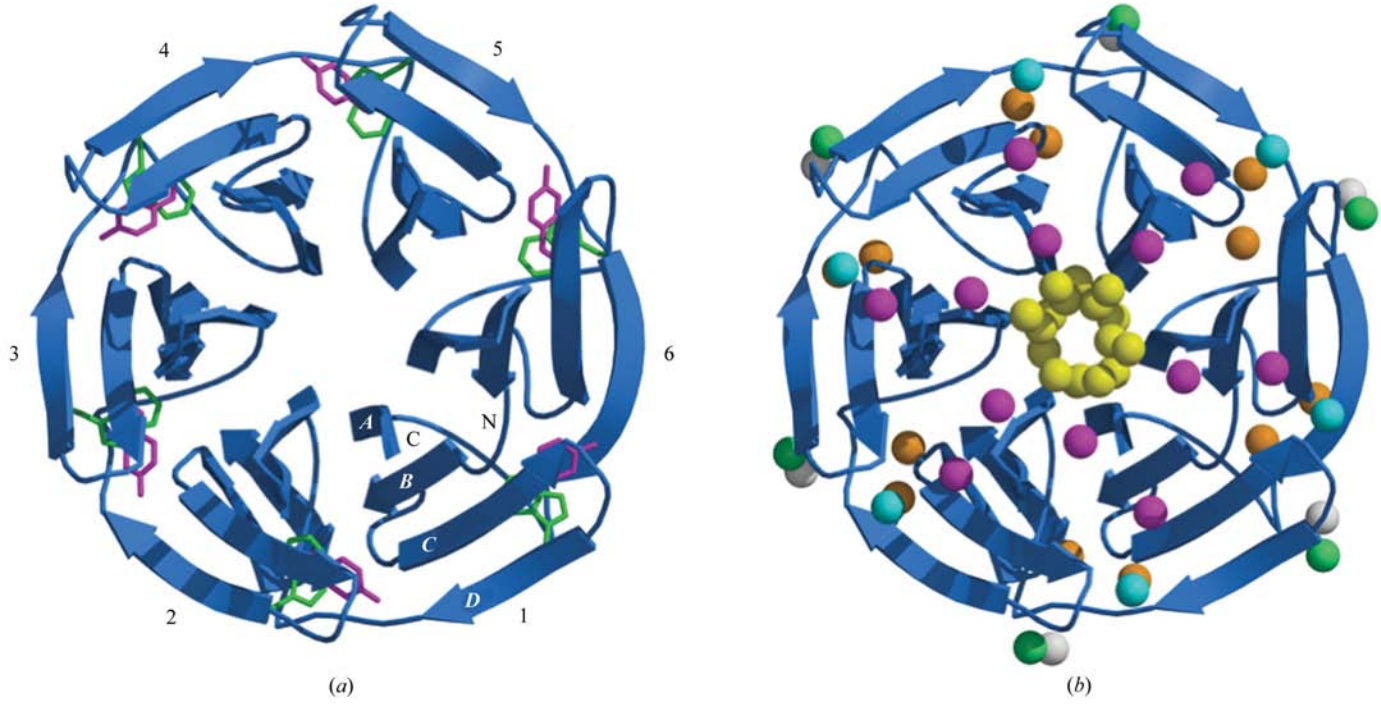

Figure 1
(*a*) A ribbon diagram of the Kelch domain of human Keap1 (residues 322–609) showing the six-bladed β-propeller structure. Side chains of the conserved Tyr/Trp pair are shown as sticks in magenta and green. Consecutive blades are numbered 1–6 and the strands of blade 1 are labeled *A*–*D*. Both the N- and C-termini of the domain are located in blade 1 and are labeled N and C, respectively. (*b*) View of the Kelch domain corresponding to that in (*a*), with the 60 conserved water molecules (ten per blade) shown as spheres. Waters are colored according to their structural roles: yellow waters occupy the central channel, magenta waters interact with the loop between strands *A* and *B*, orange waters interact with the loop between strands *C* and *D*, and green and white waters line the solvent-exposed edge of strand *D* on the outside of the β-propeller.

integrated with *HKL*2000 (Otwinowski & Minor, 1997). See Table 1 for the statistics.

2.2. Structure solution and refinement

Although the structure of the Kelch domain had already been determined by Se-SIRAS at 1.85 Å, we chose to redetermine the structure at 1.35 Å *via* SAD in order to take advantage of unbiased phases to this resolution. All seven of the possible SeMet sites were located with *SOLVE* (Terwilliger & Berendzen, 1999; Z score = 50.2 and figure of merit = 0.40 to 1.39 Å resolution). High-quality electron-density maps were further improved through solvent flattening *via RESOLVE* (figure of merit = 0.69; Terwilliger, 2000). The automatic chain-tracing feature of *RESOLVE* built 243 of the 288 residues (not including 22 disordered residues at the N-terminus, most of which are from the histidine-affinity tag). Comparison with the structure refined at 1.85 Å revealed only minor differences in side-chain conformations. These were removed from the model along with solvent molecules and the resulting model was used for refinement.

Refinement was performed with *REFMAC*5.0 (Murshudov *et al.*, 1999), initially using individual isotropic restrained B factors. Progress was monitored by the use of $R_{\rm free}$ and 5% of the data were set aside for cross-validation before refinement. After several rounds of refinement, the use of individual anisotropic B factors was introduced, resulting in a decrease in $R_{\rm free}$, so these parameters were included in the final model. Water molecules were placed automatically using *WATPEAK* (Collaborative Computational Project, Number 4, 1994) in peaks greater than 3.0σ in $F_{\rm o} - F_{\rm c}$ maps and within hydrogen-bonding distance to N or O atoms of the protein or other solvent atoms. Substantial peaks of negative electron density on the Se atom of the SeMet side chains were apparent after initial refinement; changing the occupancy of these atoms from 1.0 to 0.7 eliminated these peaks. This apparent reduced occupancy may reflect incomplete SeMet substitution in the protein or may be a consequence of radiation damage during data collection (Burmeister, 2000). TLS refinement (Winn *et al.*, 2001) was performed with the protein as a single rigid body. Owing to the high resolution of the data, the B-factor restraints for side-chain atoms were relaxed relative to the default values in

Table 2
Water clusters.

		Interacting residue						
Site	Blade	Backbone	Distance (Å)	Side chain	Distance (Å)	Water No.	B factor (Å^2)	Structural role
W1	1	Val608 N	2.95			74	19.9	Channel
	2	Val369 N	2.93			60	18.2	
	3	Val420 N	2.89			13	15.3	
	4	Val467 N	2.97			22	14.1	
	5	Val514 N	2.96			112	29.2	
	6	Val561 O	2.81			247	33.6	
W2	1	Val606 N	2.95			2	9.6	Channel
	2	Gly367 N	2.88			2	9.6	
	3	Val418 N	2.98			26	15.0	
	4	Val465 N	2.88			155	12.6	
	5	Val512 N	2.98			3	10.4	
	6	Ile559 N	3.04			20	14.5	
W3	1	Val604 O	2.76			161	16.3	Channel
	2	Leu365 O	2.67			16	15.2	
	3	Ile416 O	2.75			5	13.7	
	4	Val463 O	3.02			156	13.4	
	5	Ala510 O	2.63			141	21.8	
	6	Leu557 O	2.73			54	16.1	
W4	1			Thr609 OG	2.66	108	19.3	*A*–*B* loop
	2	Gly371 N	2.72			193	27.1	
	3	Asp422 N	2.89			92	18.3	
	4	Asn469 N	2.86			79	17.7	
	5	His516 N	2.93			148	21.3	
	6	Gln563 N	2.99			233	24.2	
W5	1	Thr609 O	2.70	Tyr329 OH	2.59	41	14.8	*A*–*B* loop/strand *B*
	2	Val370 O	2.76	Tyr375 OH	2.74	18	11.2	
	3	Ile421 O	3.00	Tyr426 OH	2.69	140	17.5	
	4	Leu468 O	2.82	Tyr473 OH	2.78	27	11.6	
	5	Leu515 O	2.89	Tyr520 OH	2.60	66	17.3	
	6			Tyr567 OH	2.51	192	20.3	
W6	1	Pro347 O	2.87	Tyr345 OH	2.69	182	24.5	*C*–*D* loop/strand *C*
	2	Pro398 O	2.76	Tyr396 OH	2.77	377	11.6	
	3	Pro445 O	2.78	Tyr443 OH	2.64	275	22.1	
	4			Tyr490 OH	2.93	143	18.1	
	5	Val539 O	2.76	Tyr537 OH	2.78	32	12.2	
	6	Pro586 O	2.76	Tyr584 OH	2.75	187	21.0	
W7	1	Ser338 O	2.84			25	11.6	*B*–*C* loop
	2	Asp389 N	2.84			174	21.6	
	3	His436 N	2.92			119	22.1	
	4	Arg483 N	2.90			215	29.3	
	5	Gln530 N	2.88			210	27.9	
	6	Phe577 N	2.92			212	20.9	
W8	1	Ser340 N	2.89	Ser340 OG	3.18	84	20.3	*B*–*C* loop
	2	Ser391 N	2.94			144	21.3	
	3	Asn438 N	2.97	Asn438 OD1	2.98	44	13.6	
	4	Asn485 N	3.06	Asn485 OD1	2.86	168	17.8	
	5	Asn532 N	3.03	Asn532 OD1	2.92	131	16.9	
	6	Asp579 N	3.19	Asp579 OD1	2.89	124	22.8	
W9	1	Asp357 O	2.69			316	32.7	*D*–*A* loop
	2	Pro408 O	2.69			197	21.6	
	3	Pro455 O	2.66			37	16.3	
	4	Ala502 O	2.62			242	25.2	
	5	Pro549 O	2.69			374	29.0	
	6	Arg596 O	2.73			200	29.0	
W10	1	Trp352 N	2.96			246	29.8	Strand *D*
	2	Trp403 N	2.91			115	19.1	
	3	Trp450 N	2.93			50	15.7	
	4	Trp497 N	2.94			11	11.1	
	5	Trp544 N	2.90			72	16.6	
	6	Trp591 N	2.86			132	22.8	

REFMAC. Model building was performed interactively using *O* (Jones *et al.*, 1991) and *COOT* (Emsley & Cowtan, 2004)

The final model of the wild-type protein consists of 288 residues and 377 water molecules (Table 1). (Residues in this manuscript are numbered according to the sequence of the intact Keap1 protein, which has 624 residues in total; the model of the Kelch domain begins at residue 322 and ends at residue 609.) Density for the side chains of ten residues was not well defined and they have been truncated; 25 residues have been modeled in two conformations, including one with two different conformations for backbone atoms. This is an increase from the ten residues with alternate conformations in the 1.85 Å model. Three waters were modeled in strong peaks at 50% occupancy owing to inter-residue distances that were too short to permit full occupancy.

The model was evaluated by *SFCHECK* (Vaguine *et al.*, 1999) and *WHAT_CHECK* (Hooft *et al.*, 1996). Figures were prepared with *MOLSCRIPT* (Kraulis, 1991), *PyMOL* (DeLano, 2002) and *RASTER3D* (Merritt & Bacon, 1997). The coordinates have been deposited in the PDB with code 1zgk. The Kelch model has good geometry, with 90.4% of its residues lying in the most favored regions of the Ramachandran plot and 0.0% in the disallowed region (Laskowski *et al.*, 1993). Calculation of H-atom positions for the C—H$\cdots\pi$ interaction analysis was performed with *HGEN* (Collaborative Computational Project, Number 4, 1994).

2.3. Identification of structurally conserved water molecules

A shell of water molecules within 5 Å of the Kelch-domain monomer was generated using *XPAND* (Kleywegt *et al.*, 2001). The six blades of the β-propeller and waters were superimposed with *TOP3D* (Lu, 2000). Solvent molecules with analogous structural roles in the blades of the Kelch β-propeller were identified following the method of Bottoms *et al.* (2005). Briefly, clusters of waters are defined as being within 2 Å of a central point after superposition; waters in clusters with at least one conserved noncovalent interaction with a polar atom of the protein are considered to be structurally equivalent. A cutoff of 3.2 Å was used for the water–protein interactions. Ten conserved clusters (W1–W10) with analogous contacts in all six blades were identified by the above method and then visually inspected. Many of the waters in these clusters had additional contacts from other residues; these are not shown in Table 2 unless the second contact occurs in more than one blade (*e.g.* W5). We included two cases where alternative atoms within aligned residues were considered to make analogous contacts with a water molecule. This occurs once in cluster W1, where all contacts are made by the backbone amide except in blade 6, where the carbonyl coordinates the water, and a second time in cluster W4, where blades 2–5 contact the conserved water *via* their backbone amide, while in blade 1 (in the strand where the β-propeller closes) this contact is made by the side chain of a threonine residue.

3. Results

3.1. Overall structure

Using synchrotron radiation, a data set from selenomethionine (SeMet) crystals of the Kelch domain was collected to 1.35 Å resolution and the structure refined using SAD phases (Table 1). As originally seen at 1.85 Å, the Kelch domain of Keap1 is a highly symmetric six-bladed β-propeller (Fig. 1*a*), in which each blade is a twisted β-sheet composed of four antiparallel β-strands (*A*–*D*). The 'top' of the propeller is formed by the loops between strands *A* and *B* and between strands *C* and *D*, while the 'bottom' contains residues from the *B*–*C* loop and the interblade *D*–*A* loops. The ring of the propeller is closed by a strand from the C-terminus, which completes a β-sheet with three strands from blade 1 at the N-terminus of the protein.

The structural differences between the 1.85 Å Kelch structure and the 1.35 Å structure are minor, with a C^{α} r.m.s.d. for 289 residues of 0.1 Å^2. The high quality of the phases is apparent in Fig. 2, which shows a typical region of the electron-density map. The model of the Kelch domain at 1.35 Å has 377 water molecules (~1.3 waters per residue of the protein), an increase of 64 waters compared with the 1.85 Å structure. Of these, about two-thirds were apparent in the SAD experimental map. All waters exhibited $2F_o - F_c$ density at 1.0σ or higher at the end of refinement; *B* factors range from 7.7 to 48.2 Å^2.

3.2. Structurally conserved solvent molecules in the Kelch repeat

Our initial structural characterization of the Kelch domain revealed key structural roles for the highly conserved residues in the Kelch repeat, such as the glycine doublet and Tyr/Trp pair mentioned previously. Refinement of the Kelch domain to 1.35 Å allows us to extend analysis of structurally conserved features in the six blades of the β-propeller to include solvent

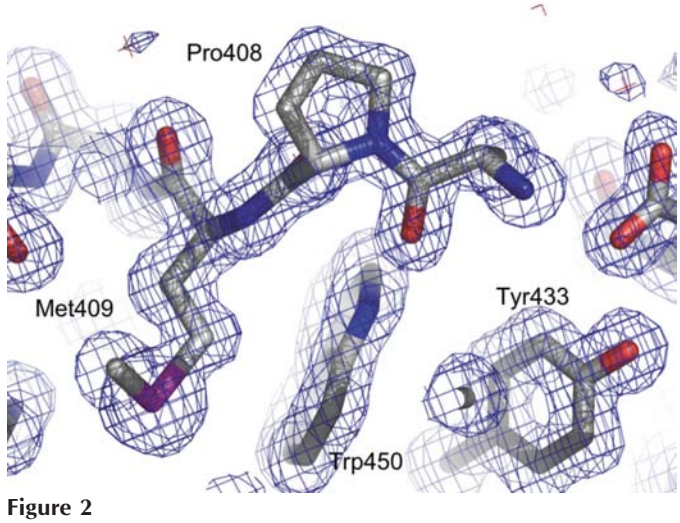

Figure 2
The SAD experimental electron-density map at 1.0σ after solvent flattening by *RESOLVE* (Terwilliger, 2000). The high quality of the 1.35 Å phases is apparent. The map shows the vicinity of the conserved Tyr/Trp pair in blade 3 of the Kelch β-propeller.

molecules. To approach this systematically, we used the method of Bottoms *et al.* (2005), originally developed to find structurally conserved waters in groups of homologous proteins. Waters with conserved structural interactions were identified as described in §2.

Our analysis identified ten clusters of waters (W1–W10) which occupy highly conserved positions in all six blades of the β-propeller (Table 2). The location of these waters in the protein are illustrated in Fig. 1(*b*). An examination of their structural context shows that they are found throughout the Kelch β-propeller in a variety of structural roles. One major category lines the central channel of the propeller (clusters W1–W3). Many of the other clusters (W4–W9) interact with residues in loops connecting the four strands of the twisted β-sheets that form the blades of the Kelch propeller. Several clusters (W5 and W6) interact with the side chains of conserved residues in the strands. Cluster W10, in contrast, interacts with the backbone atom of a highly conserved residue in strand *D* of each blade. Occasionally these categories overlap, *e.g.* clusters W5 and W6 interact with both conserved side chains and residues in loops. In one case, a water–water bridging interaction is found between two of the conserved clusters (W9 and W10). Eight of the conserved waters from six different clusters were not previously visible in the 1.85 Å maps, demonstrating the importance of the high-resolution data and phases for this analysis.

A detailed description of several of the representative clusters is found below. In this discussion, the term 'conserved' is used both to refer to contacts made by the water clusters (*e.g.* structural conservation) and also to refer to sequence conservation between the six blades of the Kelch β-propeller, as determined by their structural superposition (Li *et al.*, 2004*b*). It is also true that most of the residues that are conserved between the blades of the Kelch domain are conserved when comparing proteins within the large BTB-Kelch family (Prag & Adams, 2003).

3.3. Waters in the central channel

Three clusters of structurally conserved water molecules (W1–W3) are found in the central channel of the Kelch

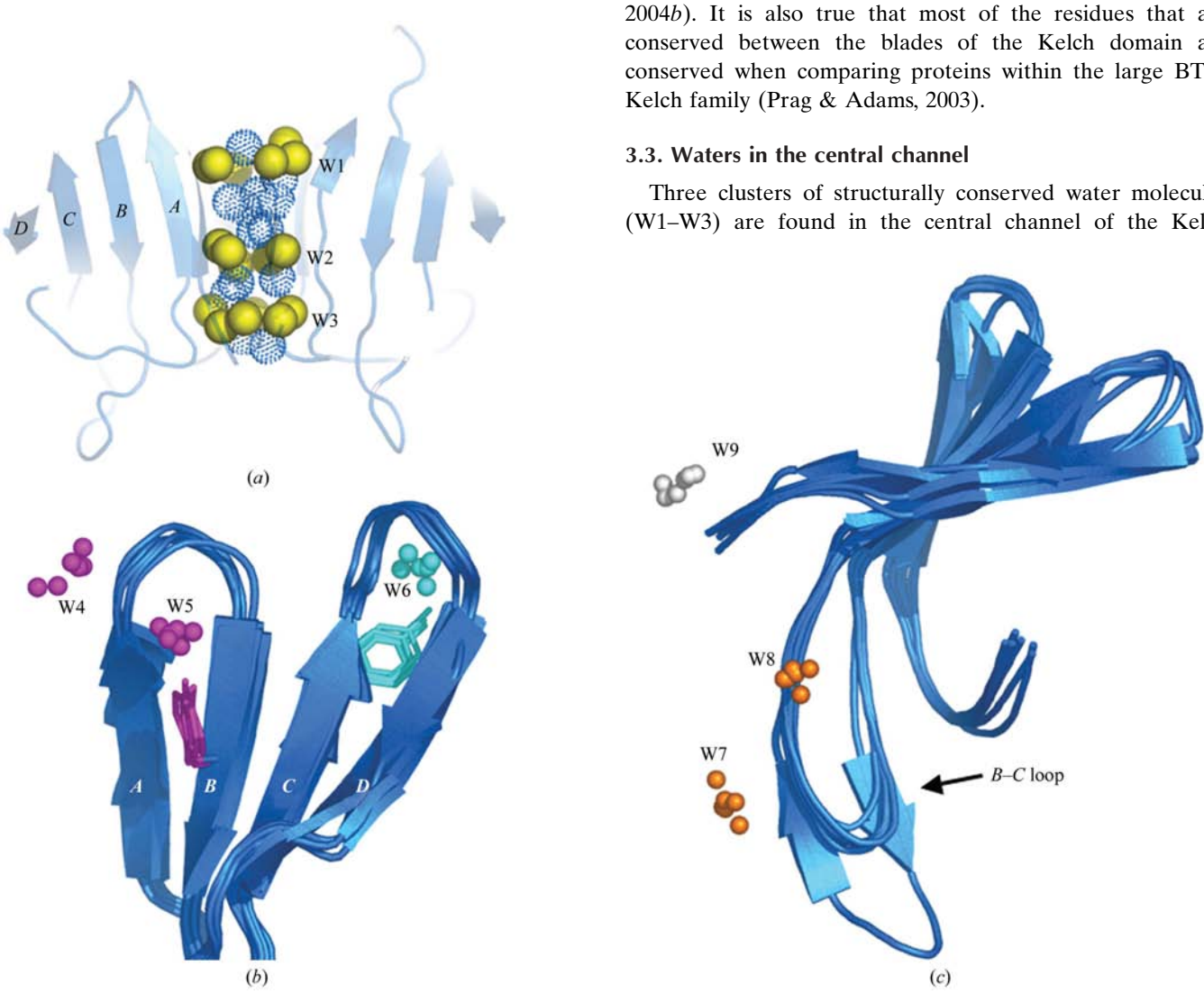

Figure 3
(*a*) A slice through a side view of the Kelch β-propeller (semi-transparent blue) showing the three clusters of conserved water molecules that line the central channel as solid spheres in yellow; non-conserved waters in the channel are shown as spheres of blue dots. (*b*) A superposition of the six blades of the Kelch domain, illustrating the conserved water molecules that interact with residues in the *A*–*B* and *C*–*D* loops of the Kelch repeat. Water clusters W4 and W5 are shown in magenta and W6 in cyan (the color scheme matches that in Fig. 1*b*). The side chains of the conserved tyrosine residues that interact with W5 and W6 are shown in matching colors. (*c*) An illustration of water clusters W7–W9 relative to a backbone superposition of the six blades of the Kelch β-propeller. Clusters W7 and W8 (orange) interact with the *B*–*C* loop of the blades, while W9 (white) interacts with the intra-blade *D*–*A* loop.

domain. This is the largest group of conserved waters and they form hydrogen bonds exclusively with the backbone amide or carbonyl groups of residues in strand *A* of each blade. The three clusters are found in distinct layers within the central channel, as can be seen in a side view of the propeller (Fig. 3*a*). In one case, a single water molecule is shared by strand *A* of two different blades (cluster W2, blades 1 and 2). These three clusters of waters have generally the lowest *B* factors of those identified in our analysis (Table 2). The three clusters represent slightly more than half of the solvent atoms that occupy the channel.

Owing to the constraints of the β-propeller fold, it is often not possible for other regions of the protein to interact with the unfulfilled hydrogen-bonding donors or acceptors on the innermost β-strand of the blades. The important role of solvent molecules in lining the central channel is emphasized by their conserved interactions with the protein and by the observation that many other β-propellers contain well ordered solvent molecules in the central channel (Wimmerova *et al.*, 2003; Takagi *et al.*, 2003; Sprague *et al.*, 2000; Cheng *et al.*, 2004).

3.4. Waters bridging loops and β-strands

Many of the waters in clusters W5 and W6 play dual roles, contacting both highly conserved residues in strands of the β-propeller blades and also residues in loops (Table 2 and Fig. 3*b*). In all six blades, waters in these two clusters contact the side chain of two completely conserved tyrosines in the Kelch repeat, while most of the waters make additional interactions with backbone atoms in a loop. All of the W5 waters hydrogen bond with the side-chain hydroxyl of a tyrosine residue in strand *B*; five of the six waters in W5 also interact with the backbone carbonyl atom of the first residue following strand *A*. The residues between these strands form a β-turn (most commonly type I′). The W6 waters interact with a conserved tyrosine in strand *C*: this tyrosine is part of the highly conserved Tyr/Trp interaction characteristic of the Kelch repeat. Five of the six waters in the W6 cluster also interact with a backbone carbonyl of a residue, usually proline, found two positions later in the polypeptide chain in the *C–D* loop. This proline is the generally the first residue in a type I β-turn. Thus, both the W5 and W6 water clusters make similar 'bridging' interactions between a conserved side chain of a residue in a strand and a backbone atom of a residue in a β-turn.

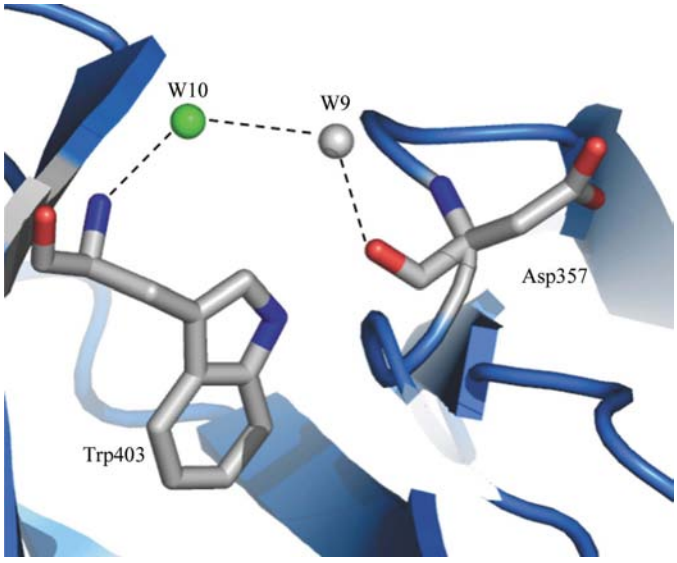

Figure 4
The water–water bridge between blades of the Kelch β-propeller, as formed by two representative waters in clusters W9 (white) and W10 (green). The bridging interaction between residues in blades 1 (Asp357) and 2 (Trp403) is illustrated. Dashed lines indicate hydrogen bonds.

3.5. Waters bridging blades of the β-propeller

The waters in clusters W9 and W10 are involved in a water–water bridge which spans adjacent blades of the Kelch β-propeller. Waters in cluster W9 are involved in contacts with residues in the extended interblade *D–A* loop, which links the individual blades of the propeller. The W9 waters interact with the backbone carbonyl of a residue in this loop, often a proline, located about halfway up the side of the propeller. Waters in cluster W10 interact with backbone atoms of a tryptophan residue found in strand *D* of the Kelch propeller. This contact is made to the same tryptophan that is involved in the conserved Tyr/Trp interaction found in every blade of the β-propeller. The interaction of this water cluster is not sequence-specific, however, since the contact is made with the backbone amide of the tryptophan, not a side-chain atom.

Visual inspection of these two clusters revealed contacts between pairs of waters in W9 and W10, creating a water–water bridge between blades. This interaction can perhaps be best visualized by following the contacts from one blade to the next (Fig. 4). For example, in blade 1, the backbone carbonyl of Asp357 contacts a water molecule in the W9 cluster. This water in turn interacts with a solvent atom in the W10 cluster, which forms a second hydrogen bond to the backbone amide of Trp403. This series of protein–water–water–protein contacts is found at the interface of each blade, including the blade 6–1 interface where the propeller closes. In Fig. 1(*b*), the water–water bridges are illustrated by the pairs of green and white spheres that encircle the outer edge of the Kelch propeller. Although this interaction is water-mediated, similar inter-blade interactions also occur between protein residues in the protein (Li *et al.*, 2004*b*); presumably, these interactions work together to help stabilize the assembly of blades into a closed β-propeller.

3.6. C—H$\cdots\pi$ hydrogen bond between the conserved Tyr/Trp pair of the Kelch repeat

The high-resolution refinement of the Kelch domain led us to re-examine the conserved sequence features of the Kelch motif. Two of these key residues, a conserved Tyr/Trp pair found in strands *C* and *D*, respectively, adopt highly similar conformations in all six blades and are in van der Waals contact with each other (Fig. 1*a*; Li *et al.*, 2004*b*). Furthermore, they participate in an extended hydrogen-bond network that

connects adjacent blades of the propeller and helps form the hydrophobic core of the protein.

A close inspection of the conserved Tyr/Trp interaction in the 1.35 Å Kelch structure revealed a previously unobserved feature: the presence of a C—H$\cdots\pi$ interaction between these two residues. In all six blades of the propeller, the C^{β} atom of the tyrosine is located approximately 3 Å from the center of the six-membered ring of the tryptophan side chain and the calculated position for one of the tyrosine C^{β} H atoms places it in the appropriate orientation for a C—H$\cdots\pi$ interaction. As defined by Brandl *et al.* (2001), these include: (i) a distance of <4.5 Å between the C atom (in this case the tyrosine C^{β}) and the center of mass of the π-system (hereafter referred to as *X*) and (ii) an angle of >120° between the carbon (C^{β}), hydrogen and *X*. In the case of Tyr443 and Trp450 depicted in Fig. 5(*a*), these values are 3.4 Å and 175°; similar values are found for each Tyr/Trp pair in the structure (Fig. 5*b*).

In the C—H$\cdots\pi$ interactions observed in the Kelch domain, the C^{β} hydrogen of the tyrosine acts as the donor and the six-membered ring of the tryptophan acts as the π-acceptor. This category of C—H$\cdots\pi$ interactions, where the donor is an aliphatic C—H group and the acceptor is an aromatic π-system, is the most frequently observed in protein structures (Brandl *et al.*, 2001). While not commonly included in hydrogen-bonding analyses of proteins, it has been proposed that C—H$\cdots\pi$ interactions may be an underappreciated component of protein structure and stability. Although estimates vary, typical values for O—H$\cdots$O or N—H$\cdots$O hydrogen bonds in proteins range from $\sim$5 to 10 kJ mol^{-1}, while C—H$\cdots\pi$ hydrogen bonds are likely to have values closer to 0.5 to 1.0 kJ mol^{-1} (Weiss *et al.*, 2001). However, C—H$\cdots\pi$ interactions are only one category of 'non-classical' hydrogen bonds, which also include C—H$\cdots$O and C—H$\cdots$N interactions; the cumulative effect of these weak but numerous interactions may have a significant effect on protein structure and stability (Weiss *et al.*, 2001).

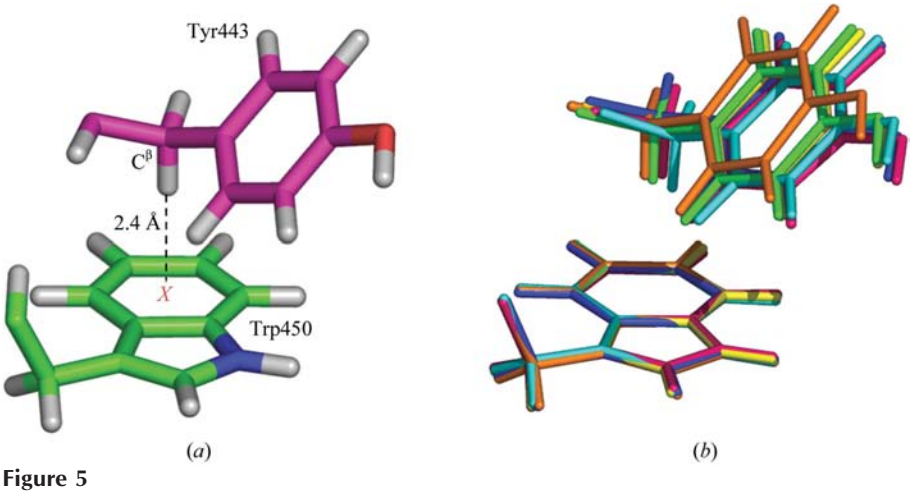

Figure 5
Close-up view of the proposed C—H$\cdots\pi$ interaction between the conserved Tyr/Trp pair found in all six blades of the Kelch domain. (*a*) Details of the interaction shown with Tyr443 and Trp450. The dashed line indicates the distance between the C^{β} hydrogen of the tyrosine and the center of mass (*X*) of the six-membered aromatic ring of the tryptophan side chain. Side chains are shown as stick models, with H atoms (gray) in their calculated positions. (*b*) An overlay of the Tyr/Trp pairs from all six blades of the Kelch domain, showing the orientation of the C^{β} hydrogen relative to the tryptophan ring. Orientation is the same as in (*a*); superposition is based on atoms in the tryptophan side chain. The Tyr/Trp pair from each blade is shown in a different color.

4. Discussion

Our analysis of the Kelch domain of Keap1 refined at 1.35 Å has revealed ten clusters of water molecules that make conserved structural interactions with the protein. These waters are found throughout the structure of the β-propeller, with the largest group occupying the central channel. In addition, they also form bridges connecting highly conserved residues in the strands of the blades to residues in the loops between strands. These contacts help explain the sequence conservation of the residues in the Kelch motif that are involved in these interactions. In addition, the waters also function as hydrogen-bond donors or acceptors for backbone atoms in residues of the polypeptide chain without regular secondary structure (*e.g.* in loops and turns). It is possible that these structurally conserved waters also serve to help stabilize the fold (or folding process) of the blades and/or the Kelch β-propeller by linking β-strands to loops and blades to blades.

Structurally conserved waters have been identified in a variety of protein families and proposed to play various important roles, including interacting with ligands and co-factors (Loris *et al.*, 1994; Ogata & Wodak, 2002; Bottoms *et al.*, 2002). With one exception, where a single conserved water molecule was identified in the Asp-box β-hairpin motif of sialidase-like proteins (Copley *et al.*, 2001), such an analysis has not yet been conducted for any members of the large β-propeller fold family. These proteins vary in function, sequence and structure and consist of propellers that contain from four to eight blades (Jawad & Paoli, 2002; Fülöp & Jones, 1999). Despite their common topology, the significant sequence/structure differences between the various subtypes of β-propellers (*e.g.* WD repeat, Asp-box *etc.*) make it implausible that other members of this family will share the same waters identified in the Kelch propeller. However, novel conserved waters are likely to be observed in the high-resolution structures of other β-propeller subtypes. As in the case of the Kelch β-propeller, the possible roles of conserved waters in the structure and folding of these distinct family members remains to be elucidated.

In the Kelch β-propeller of Keap1, a high degree of structural symmetry is apparent and therefore it may not seem surprising that waters with similar structural interactions are found in each blade. Our previous analysis showed that the blades have an r.m.s.d. of $\sim$0.5–1.0 Å^2 for pairwise superpositions of C^{α} atoms, demonstrating their three-dimensional similarity (Li *et al.*, 2004*a*). However, structure-based sequence

alignments of the blades show only moderate identities, from ~30 to 50%. Since this range of sequence identity is similar to that seen between members of the BTB-Kelch family (Li *et al.*, 2004*a*), it is possible that many of these proteins will share the waters identified in this analysis. In particular, since most of the water clusters identified in the Kelch domain interact with either backbone atoms of the protein (and are therefore not sequence-specific) or with highly conserved residues in the Kelch repeat, we expect that at least some of these waters will be found in the structures of other BTB-Kelch proteins. Consideration of highly conserved waters has been shown to improve the predicted structure of homology models, in at least one protein family (Henriques *et al.*, 1997).

Our analysis of the 1.35 Å structure of the Kelch domain also revealed a novel interaction between the Tyr/Trp pair characteristic of the Kelch repeat. The occurrence of a C—H$\cdots\pi$ interaction between these two highly conserved residues in the Kelch repeat is striking and may add to the propensity for sequence conservation at these positions in both the individual blades of the propeller and also in members of the BTB-Kelch family. While many side chains, *e.g.* leucine, could donate a C^{β} hydrogen for this interaction, the side chains of the conserved Tyr/Trp pair in the Kelch repeat provide the opportunity for additional favorable interactions, such as the ability to participate in the previously described interblade hydrogen-bond network (Li *et al.*, 2004*b*). However, we note that the most frequent substitution in BTB-Kelch proteins for the conserved tyrosine is phenylalanine, which could be expected to make a C—H$\cdots\pi$ interaction and van der Waals interactions with the tryptophan, but would not have a side-chain hydroxyl to participate in hydrogen bonds. As the structures of other proteins containing Kelch repeats are determined, it will be interesting to see whether the C—H$\cdots\pi$ interaction is retained when different residues are found in these positions.

In summary, we have refined the structure of the Kelch domain of Keap1 to 1.35 Å using high-quality SAD phases. Analysis of this high-resolution structure shows ten clusters of conserved water molecules which play a variety of structural roles in the protein. Furthermore, a C—H$\cdots\pi$ interaction is found between two key residues in the Kelch repeat. Both of these results emphasize the correlation between conserved amino acids of the Kelch repeat and structural interactions within the protein and lend further insight into the fold of this common eukaryotic structural motif.

We thank Manfred Weiss for pointing out the possibility of C—H$\cdots\pi$ interactions. We acknowledge the staff of the Structural Biology Center at beamline 19-ID of the Advanced Photon Source of Argonne National Laboratory for synchrotron time and technical assistance. This work was supported by the University of Missouri Molecular Biology Program, the University of Missouri Food for the 21st Century program, research grants from NIH (GM59213 and P50 CA103130) and by a grant from the University of Missouri Research Board.

References

Bottoms, C. A., Smith, P. E. & Tanner, J. J. (2002). *Protein Sci.* **11**, 2125–2137.
Bottoms, C. A., White, T. A. & Tanner, J. J. (2005). Submitted.
Brandl, M., Weiss, M. S., Jabs, A., Suhnel, J. & Hilgenfeld, R. (2001). *J. Mol. Biol.* **307**, 357–377.
Burmeister, W. P. (2000). *Acta Cryst.* D**56**, 328–341.
Cheng, Z., Liu, Y., Wang, C., Parker, R. & Song, H. (2004). *Protein Sci.* **13**, 2673–2684.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760.
Copley, R. R., Russell, R. B. & Ponting, C. P. (2001). *Protein Sci.* **10**, 285–292.
DeLano, W. L. (2002). *The PyMOL Molecular Graphics System.* http://www.pymol.org.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Fülöp, V. & Jones, D. T. (1999). *Curr. Opin. Struct. Biol.* **9**, 715–721.
Henriques, E. F., Ramos, M. J. & Reynolds, C. A. (1997). *J. Comput. Aided Mol. Des.* **11**, 547–556.
Hooft, R. W., Vriend, G., Sander, C. & Abola, E. E. (1996). *Nature (London)*, **381**, 272.
Jawad, Z. & Paoli, M. (2002). *Structure*, **10**, 447–454.
Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kleywegt, G. J., Zou, J. Y., Kjeldgaard, M. & Jones, T. A. (2001). *International Tables for Crystallography*, Vol. *F*, edited by M. G. Rossmann & E. Arnold, pp. 353–356. Dordrecht: Kluwer Academic Publishers.
Kraulis, P. J. (1991). *Science*, **254**, 581–582.
Laskowski, R. A., Moss, D. S. & Thornton, J. M. (1993). *J. Mol. Biol.* **231**, 1049–1067.
Li, X., Zhang, D., Hannink, M. & Beamer, L. J. (2004*a*). *Acta Cryst.* D**60**, 2346–2348.
Li, X., Zhang, D., Hannink, M. & Beamer, L. J. (2004*b*). *J. Biol. Chem.* **279**, 54750–54758.
Loris, R., Stas, P. P. & Wyns, L. (1994). *J. Biol. Chem.* **269**, 26722–26733.
Lu, G. (2000). *J. Appl. Cryst.* **33**, 176–183.
Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.
Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. (1999). *Acta Cryst.* D**55**, 247–255.
Ogata, K. & Wodak, S. J. (2002). *Protein Eng.* **15**, 697–705.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Prag, S. & Adams, J. C. (2003). *BMC Bioinformatics*, **4**, 42.
Sprague, E. R., Redd, M. J., Johnson, A. D. & Wolberger, C. (2000). *EMBO J.* **19**, 3016–3027.
Takagi, J., Yang, Y., Liu, J. H., Wang, J. H. & Springer, T. A. (2003). *Nature (London)*, **424**, 969–974.
Terwilliger, T. C. (2000). *Acta Cryst.* D**56**, 965–972.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.
Vaguine, A. A., Richelle, J. & Wodak, S. J. (1999). *Acta Cryst.* D**55**, 191–205.
Weiss, M. S., Brandl, M., Suhnel, J., Pal, D. & Hilgenfeld, R. (2001). *Trends Biochem. Sci.* **26**, 521–523.
Wimmerova, M., Mitchell, E., Sanchez, J. F., Gautier, C. & Imberty, A. (2003). *J. Biol. Chem.* **278**, 27059–27067.
Winn, M. D., Isupov, M. N. & Murshudov, G. N. (2001). *Acta Cryst.* D**57**, 122–133.
Xue, F. & Cooley, L. (1993). *Cell*, **72**, 681–693.