structural communications

Crystal-contact engineering to obtain a crystal form of the Kelch domain of human Keap1 suitable for ligand-soaking experiments

**Stefan Hörer,* Dirk Reinert,
Katja Ostmann, Yvette Hoevels
and Herbert Nar**

Department of Lead Identification and
Optimization Support, Boehringer
Ingelheim Pharma GmbH and Co. KG,
Birkendorferstrasse 65, 88400 Biberach,
Germany

Correspondence e-mail:
stefan.hoerer@boehringer-ingelheim.com

Keap1 is a substrate adaptor protein for a Cul3-dependent ubiquitin ligase complex and plays an important role in the cellular response to oxidative stress. It binds Nrf2 with its Kelch domain and thus triggers the ubiquitinylation and degradation of Nrf2. Oxidative stress prevents the degradation of Nrf2 and leads to the activation of cytoprotective genes. Therefore, Keap1 is an attractive drug target in inflammatory diseases. The support of a medicinal chemistry effort by structural research requires a robust crystallization system in which the crystals are preferably suited for performing soaking experiments. This facilitates the generation of protein–ligand complexes in a routine and high-throughput manner. The structure of human Keap1 has been described previously. In this crystal form, however, the binding site for Nrf2 was blocked by a crystal contact. This interaction was analysed and mutations were introduced to disrupt this crystal contact. One double mutation (E540A/E542A) crystallized in a new crystal form in which the binding site for Nrf2 was not blocked and was accessible to small-molecule ligands. The crystal structures of the apo form of the mutated Keap1 Kelch domain (1.98 Å resolution) and of the complex with an Nrf2-derived peptide obtained by soaking (2.20 Å resolution) are reported.

1. Introduction

Cullin–RING ubiquitin ligase complexes are the largest family of multi-subunit E3 ligases in eukaryotes (Zimmerman *et al.*, 2010). Specificity is achieved by a large set of adaptor and substrate-binding proteins. Keap1, one of the substrate-binding proteins, consists of several domains: an N-terminal BTB domain which interacts with Cul3, an intervening region (IVR) which contains the oxidation-sensitive cysteine residues and the C-terminal Kelch domain which binds the substrate Nrf2. In the response to oxidative stress, modification of the cysteines in the IVR leads to a conformational change which prevents the ubiquitinylation of Nrf2. Nrf2 translocates to the cell nucleus and induces cytoprotective genes. A model suggested by Tong *et al.* (2007) proposes that two Keap1 proteins dimerize *via* the BTB domain and that the two Kelch domains in this complex recognize two distinct sites at the N-terminus of Nrf2 and thereby fix the orientation of Nrf2 within the ubiquitin ligase complex. These two binding sites on Nrf2 show a 200-fold difference in binding affinity for Keap1. The structure of the human Keap1 Kelch domain has been described in its apo form (Beamer *et al.*, 2005; Li *et al.*, 2004; PDB entries 1zgk and 1u6d) and in complex with a peptide comprising the high-affinity binding site of Nrf2 (Lo *et al.*, 2006; PDB entry 2flu). The low-affinity binding site has been structurally characterized using Keap1 from mouse (Tong *et al.*, 2007; PDB entry 2dyh). None of the crystal forms identified for human Keap1 can be soaked with small-molecule ligands, as the binding site is either occupied by a crystal contact (in the apo form) or by the Nrf2-derived peptide, which binds with high affinity and is furthermore itself involved in crystal contacts. We analyzed the crystal contact at the Nrf2-binding site and designed several mutations to disrupt this crystal contact. One of these constructs (E540A/E542A) crystallized under different conditions to the wild-type protein and structure determination revealed a different crystal form in which one of the two crystallographically independent protomers in the asymmetric unit has an accessible Nrf2-binding site.
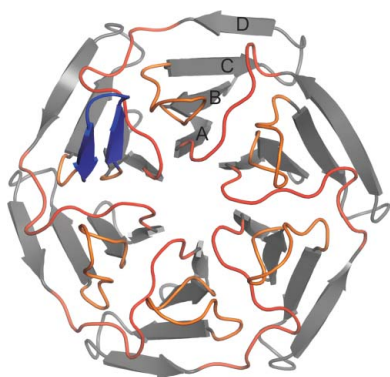
2. Methods

2.1. Cloning, mutagenesis, expression and purification

The DNA coding for the Keap1 Kelch domain (residues 321–609) was cloned into pET-15b using the *Nde*I and *Bam*HI restriction sites to obtain a construct with an N-terminal six-histidine tag. Mutagenesis was performed with a QuikChange Multi Site-Directed Mutagenesis Kit (Stratagene) using the sequence 5′-GCGCTACGA-TGTGGCAACAGCGACGTGGACTTTCG-3′ as a primer to obtain the two mutations E540A and E542A. The protein was expressed in *Escherichia coli* BL21(DE3)pLysS cells at 303 K. The cells were harvested by centrifugation, resuspended in 20 m*M* Tris–HCl, 500 m*M* NaCl, 5 m*M* imidazole pH 7.9 and disrupted by sonification. The cleared lysate was loaded onto an Ni–NTA column (Qiagen) with a 5 ml bed volume. The column was washed five times with 5 ml 20 m*M* Tris–HCl, 500 m*M* NaCl, 20 m*M* imidazole pH 7.9 and was eluted with 20 m*M* Tris–HCl, 500 m*M* NaCl, 250 m*M* imidazole pH 7.9. After dialysis into 20 m*M* Tris, 2 m*M* EDTA, 5 m*M* DTT pH 7.9, the protein was loaded onto a 5 ml HiTrap Q HP column (GE Healthcare) and eluted with a linear gradient from 0 to 0.5 *M* NaCl over five column volumes. The protein eluted at 250 m*M* NaCl. Fractions containing Keap1 were pooled and desalted prior to crystallization. Typical yields were 5 mg per litre of expression medium.

2.2. Crystallization and data collection

Initial crystal hits were found using the SaltRx screen (Hampton Research) in a condition consisting of 4.0 *M* ammonium acetate, 0.1 *M* sodium acetate trihydrate pH 4.6. This condition was refined using Additive Screen (Hampton Research). The best crystals grew within 10 d at a protein concentration of 9.5 mg ml^{-1} using either 0.01 *M* taurine, 0.01 *M* nickel chloride, 4% γ-butyrolactone or 0.025% dichloromethane as additives. The Nrf2-derived peptide used

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the outer resolution shell.

	Apo Keap1	Keap1–Nrf2(77–82)
Data collection		
Wavelength (Å)	1.002	0.979
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters (Å)	$a = 75.72$, $b = 75.77$, $c = 202.04$	$a = 76.10$, $b = 76.06$, $c = 207.54$
No. of molecules per asymmetric unit	2	2
Resolution (Å)	38.00–1.98 (2.08–1.98)	76.88–2.20 (2.27–2.20)
Mosaicity (*XDS*) (°)	0.2	0.13
No. of observations	534488 (77292)	249285 (22520)
No. of unique reflections	82414 (11878)	61950 (5482)
Multiplicity	6.5 (6.5)	4.0 (4.1)
R_{merge} (%)	5.2 (48.4)	10.1 (49.5)
Mean $I/\sigma(I)$	22.9 (3.4)	14.1 (2.9)
Completeness (%)	100 (100)	99.8 (99.9)
Refinement statistics (*autoBUSTER*)		
Resolution range (Å)	37.86–1.98	40.00–2.20
R_{cryst}† (%)	16.4	17.2
R_{free}‡ (%)	18.1	18.8
R.m.s.d., bond distances (Å)	0.008	0.008
R.m.s.d., bond angles (°)	0.98	1.02
Total No. of non-H atoms in asymmetric unit	5100	5012
No. of solvent molecules	679	560
B value from Wilson plot (Å^2)	42.4	44.6
Average B value, protein (Å^2)	44.8	40.6
Average B value, solvent (Å^2)	55.3	50.8

† $R_{\mathrm{cryst}} = \sum_{hkl} ||F_{\mathrm{obs}}| - |F_{\mathrm{calc}}||/\sum_{hkl} |F_{\mathrm{obs}}|$, where F_{obs} and F_{calc} are observed and calculated structure factors, respectively. ‡ R_{free} is the R factor calculated for 5% of the reflections which were not included in refinement.

in this study consisted of residues 77–82 of human Nrf2 cyclized by an additional glycine residue (cycloGD$_{77}$EETGE$_{82}$; Thermo Fisher). A crystal was transferred into a solution containing the crystallization reservoir buffer and the peptide at a concentration of 1 mg ml^{-1} for 2 h. The crystals were then flash-cooled in liquid nitrogen. The
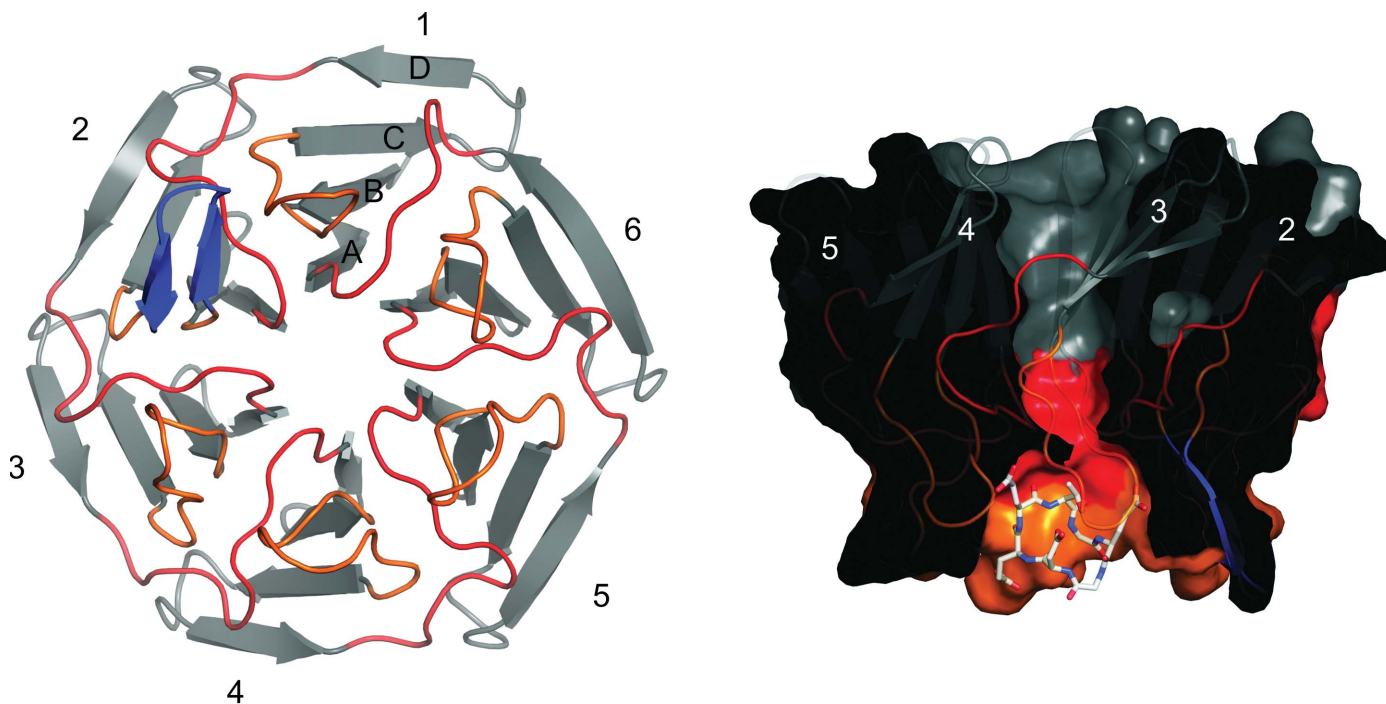

Figure 1
Architecture of the Keap1 Kelch domain. The six individual blades of the highly symmetric β-propeller (grey ribbons) are connected by the interblade A–D loops (red); the B–C loops are coloured orange. The Nrf2-binding site is formed by the interblade A–D loops and the B–C loops. The small β-sheet inserted into the B–C loop of the second blade is coloured blue. All figures were created with *PyMOL* (Schrödinger).

crystallographic experiments were performed on the X06SA beam-line at the Swiss Light Source, Paul Scherrer Institut, Villigen, Switzerland. Data were processed with *XDS* (Kabsch, 2010) and *SCALA* (Evans, 2006) within *APRV* (Kroemer *et al.*, 2004).

2.3. Structure solution and refinement

The structure of the apo form of the Keap1 Kelch domain was determined by molecular replacement with *Phaser* (McCoy *et al.*, 2007) using PDB entry 1u6d (Li *et al.*, 2004) as a starting model and was refined with *autoBUSTER* (Bricogne *et al.*, 2011). Model building was performed with *Coot* (Emsley *et al.*, 2010). The final model consisted of two protomers with 570 residues and 668 water molecules. In addition, ten acetate molecules present in the crystallization buffer and one sodium ion were visible in the electron density and were incorporated into the final model. The amino-acid numbering in this manuscript is according to the complete Keap1 protein. The first 25 residues of the protein comprising the histidine tag, the thrombin-cleavage site and residues 321–324 of the Kelch domain were not visible in the electron-density map. Four side chains were not well defined and their occupancy was reduced and refined. Two side chains were modelled in two conformations. The structure of the Keap1–Nrf2 peptide complex was determined using the difference Fourier method. Manual adjustments of the model were performed with *Coot* (Emsley *et al.*, 2010). The quality of the final models was assessed with *MolProbity* (Chen *et al.*, 2010). 98% of all residues were in favoured regions of the Ramachandran plot and there were no outliers. Data-collection and refinement statistics are given in Table 1.

3. Results

3.1. Overall structure

The Keap1 Kelch domain forms a six-bladed β-propeller with four β-strands in each blade. Following the notation of Beamer *et al.* (2005), the interaction with Nrf2 is at the 'bottom' side of the structure. This binding site is formed by two concentric rings of loops. The inner ring, which is formed by the interblade D–A loops, forms
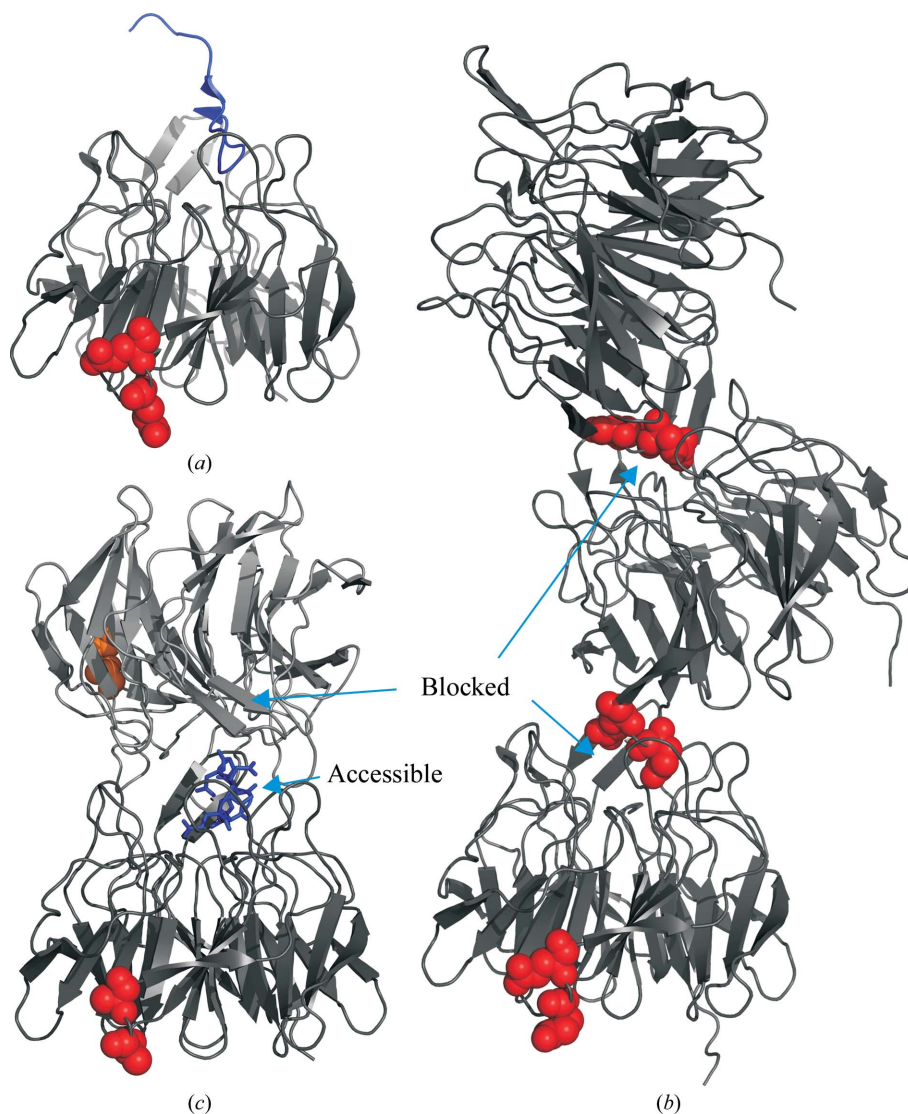

Figure 2
Comparison of the crystal packing of (*a*) the wild-type Keap1 Kelch domain in complex with an Nrf2-derived peptide (PDB entry 2flu; Lo *et al.*, 2006), (*b*) unliganded Keap1 (PDB entry 1zgk; Beamer *et al.*, 2005) and (*c*) the E540A/E542A double mutant. Residues 540 and 542 in monomers *A* and *B* are shown as red and orange spheres, respectively. The access to the Nrf2-binding site is blocked in wild-type Keap1 (*b*) and free in one of the two protomers of the E540A/E542A double mutant (*c*). This binding site is occupied by the cyclic peptide.
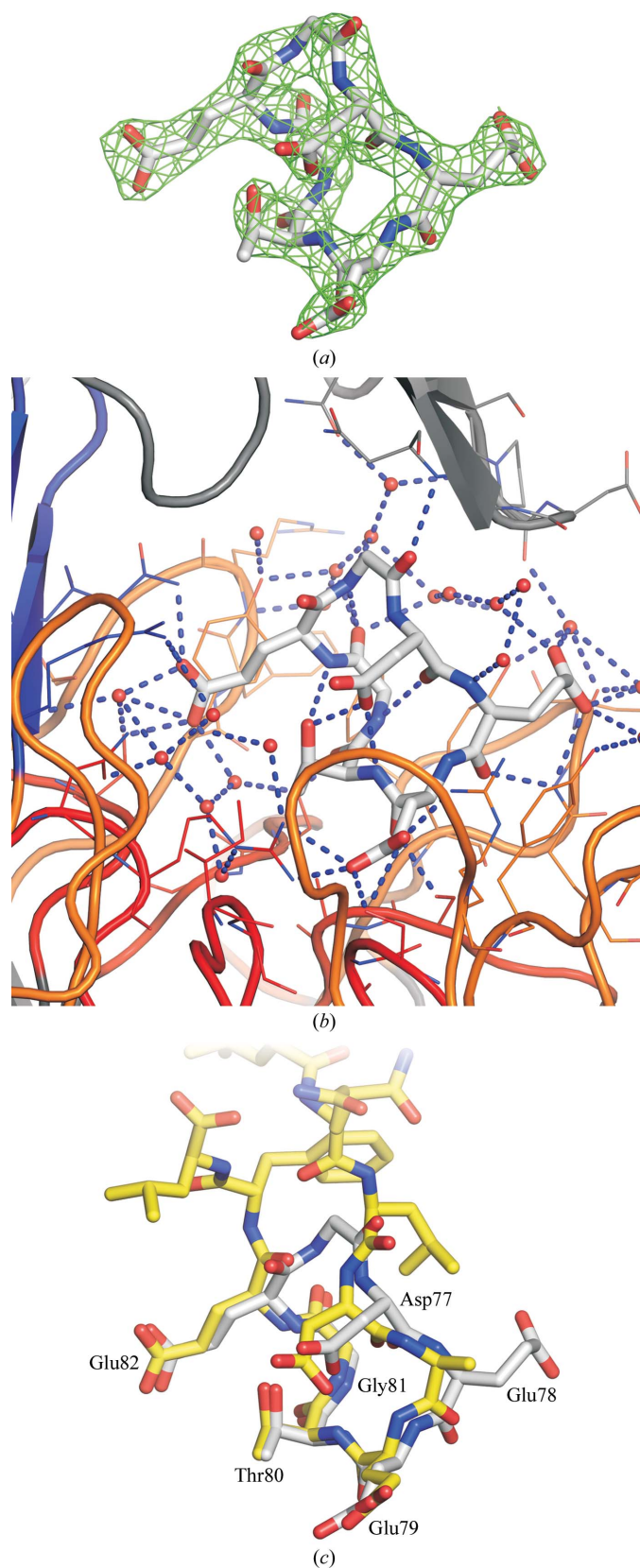
(a)

(b)

(c)

Figure 3
Structure of the Keap1 Kelch domain in complex with the Nrf2-derived peptide cycloGD$_{77}$EETGE$_{82}$. (a) shows the initial $F_o - F_c$ density at a contour level of 5.0σ. In (b) the polar contacts involved in binding the cyclic peptide are shown. (c) depicts a superpositioning of the cyclic peptide used in this study (grey) and the linear peptide from PDB entry 2flu (yellow).

the bottom of the binding site, whereas the outer ring, which consists of the B–C loops, forms the rim of the binding site. The B–C loop of the second blade (residues 378–394) is four amino acids longer than the other B–C loops and forms an additional antiparallel β-sheet (Fig. 1).

3.1.1. Comparison of the structures of the E540A/E542A double mutant and the wild-type protein. The structure of the E540A/E542A double mutant superimposes well with the high-resolution structure of the wild-type protein (Beamer *et al.*, 2005; PDB entry 1zgk). The r.m.s.d. for 1739 aligned atoms is 0.26 Å. The only notable change is in the tip of the B–C loop within the second blade (residues 385). This loop is involved in different crystal contacts and therefore shifts by 2.6 Å.

3.1.2. Differences in the crystal packing of the E540A/E542A double mutant and the wild-type protein. In the wild-type form, the Keap1 Kelch domain crystallizes with one molecule per asymmetric unit. Crystal contacts are formed between the C–D loop of the fifth blade of the β-propeller and residues within the Nrf2-binding pocket of the adjacent molecule. The two molecules interact in a 'top-to-bottom' fashion in which both Nrf2-binding pockets of the crystallographic dimer are blocked (Fig. 2b). Glu540 forms a salt bridge to Arg415 and hydrogen bonds to four water molecules. One of these water molecules bridges Glu540 and Ser602. Glu542 forms a bidentate salt bridge to Arg380 and Arg415 and hydrogen bonds to six water molecules. The hydrogen-bonding network formed by these two residues involves almost all of the water molecules within the Nrf2-binding pocket.

In contrast to the wild-type crystal form, two protomers of the E540A/E542A double mutant occupy the asymmetric unit of the crystals. The two molecules (annotated A and B) interact in a 'bottom-to-bottom' fashion employing three distinct interaction sites. The first consists of the long B–C loop of the second β-propeller blade of molecule A, which occupies the Nrf2-binding pocket of the neighbouring molecule B. Asp385 interacts with Gln530 and Ser555, while Arg415, the carbonyl of Asn387 and the side chain of Thr388 interact with Arg483. Again, additional contacts are mediated by water molecules within the Nrf2-binding pocket. The second contact is an intermolecular disulfide bridge between Cys434 within the B–C loops of the third blade of both molecules. The third contact is made by the B–C loop of the sixth blade of molecule A with the B–C loop of the second β-propeller blade of molecule B. His575 forms a hydrogen bond to Asp385. The resulting noncrystallographic dimer displays no proper local symmetry, resulting in the Nrf2-binding pocket of molecule A being unoccupied and accessible to small molecules (Fig. 2c).

3.1.3. Comparison of the apo structure and the Keap1–Nrf2 peptide complex. The overall structure of the Keap1 apo protein is very similar to that of the Keap1–Nrf2 peptide complex (Fig. 3). The r.m.s.d. on C$^\alpha$ atoms is 0.20 Å for chain A and 0.15 Å for chain B. The major change is a rigid-body movement of chain B relative to chain A, which can be described as a rotation of 6° around the region where the disulfide bond connects chains A and B. This leads to an elongation of the asymmetric unit content by approximately 2 Å and thus to a change in the c axis from 202.0 to 207.5 Å. Further changes occur at the Nrf2-binding site, in which side chains move as a consequence of the peptide interaction. Arg415 moves to avoid a clash with Thr80 from Nrf2 and forms a salt bridge with Glu79. The side chain of Arg380 moves to form a salt bridge to Glu82. Tyr572 shifts outwards by 1.6 Å from the centre of the cavity to achieve the optimal distance to form a water-mediated hydrogen bond to Glu78. The same phenomenon is observed for Glu530, which forms a hydrogen bond to the carbonyl O atom of Glu78. All side chains of the Nrf2 peptide

are well ordered and visible in the electron density. The binding mode of the peptide used in this study is consistent with the binding mode reported by Lo and coworkers (PDB entry 2flu; Lo *et al.*, 2006). Differences can be observed in the side-chain conformation of Glu79. The greatest C^α deviation can be seen for Asp77 and Glu78 (1.5 Å). This is probably owing to the cyclic nature of our peptide, which does not possess the same degree of conformational freedom as the linear peptide in PDB entry 2flu (A_{69}FFAQLQLDEETGEFL$_{84}$). In the structure of Keap1 in complex with the cyclic peptide described here, the side chain of Glu78 is well defined in the electron density, which is in contrast to the structure of Keap1 in complex with the linear peptide, despite its higher resolution (1.5 Å compared to 2.2 Å). This might arise from additional water-mediated crystal contacts.

4. Discussion

Two glutamate residues in the wild-type Keap1 Kelch domain have been mutated to alanines in order to disrupt a crystal contact which blocks the binding site for Nrf2. The mutated protein no longer crystallizes in the wild-type crystal form but adopts a new crystal form in which two molecules occupy the asymmetric unit of the crystals. In one of these protomers the Nrf2-binding site is empty and is accessible to small organic compounds. This E540A/E542A double mutant is therefore suitable for soaking experiments. This greatly facilitates the generation of protein–ligand complex structures. The applicability of this crystal form has been demonstrated by solving the structure of the Keap1 Kelch domain in complex with a cyclic peptide derived from Nrf2. Soaking of this peptide into the apo crystals led to the straightforward determination of the structure of the complex without major changes in the quality of the crystal.

References

Beamer, L. J., Li, X., Bottoms, C. A. & Hannink, M. (2005). *Acta Cryst.* D**61**, 1335–1342.

Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, W., Roversi, P., Sharff, A., Smart, O. S., Vonrhein, C. & Womack, T. O. (2011). *BUSTER* v.2.11.2. Cambridge: Global Phasing Ltd.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.

Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.

Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.

Kroemer, M., Dreyer, M. K. & Wendt, K. U. (2004). *Acta Cryst.* D**60**, 1679–1682.

Li, X., Zhang, D., Hannink, M. & Beamer, L. J. (2004). *J. Biol. Chem.* **279**, 54750–54758.

Lo, S.-C., Li, X., Henzl, M. T., Beamer, L. J. & Hannink, M. (2006). *EMBO J.* **25**, 3605–3617.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.

Tong, K. I., Padmanabhan, B., Kobayashi, A., Shang, C., Hirotsu, Y., Yokoyama, S. & Yamamoto, M. (2007). *Mol. Cell. Biol.* **27**, 7511–7521.

Zimmerman, E. S., Schulman, B. A. & Zheng, N. (2010). *Curr. Opin. Struct. Biol.* **20**, 714–721.