Supplementary Information

Interaction Energetics and Druggability of the Protein-Protein Interaction between Kelch-like ECH-associated Protein 1 (KEAP1) and Nuclear Factor, Erythroid 2 Like 2 (Nrf2)

Mengqi Zhong[1], Andrew Lynch[1], Stefan Jehle[1,2], Lingqi Luo[2,3], David R. Hall[4], Reina Iwase[1], James P. Carolan[1], Kristina Streu[1], Christine M. Harvey[1], Dima Kozakov[5,*], Sandor Vajda[1,2,3,6,*], Karen N. Allen[1,3,6,*] and Adrian Whitty[1,6,*]

Supplementary Figures

Supplementary Figure S1: Characterization data for the KEAP1 protein constructs used in this study. *(SDS-PAGE, CD, thermal melting).*

Supplementary Figure S2: Characterization data for the Nrf2 protein constructs used in this study. *(SDS-PAGE, CD, thermal melting).*

Supplementary Figure S3: Inhibition data for $Nrf2_{1-100}$ and $Nrf2_{34-100}$, measured using the fluorescence anisotropy competition binding assay.

Supplementary Figure S4: Intramolecular interactions involving D77 and T80.

Supplementary Figure S5: New crystal from of KEAP1 β-propeller domain suitable for ligand soaking, and comparison of soaked structures of wild-type (LDEETGEFL) and T80A (LDEEAGEFL) 9-mer peptides.

Supplementary Figure S6: Sequence alignment of Nrf2 paralogs, and sequence conservation of key residues in KEAP1.

Supplementary Figure S7: Minor conformational change of Tyr525 upon ligand binding.

Supplementary Figure S8: Close-up view showing how R415 shifts upon ligand binding and the resulting effects on the relative strengths and locations of FTMap hot spots.

Supplementary Figure S9: Illustration showing how hot spot A can be approached to achieve higher binding affinity, using as an example Biogen compound 16.

Supplementary Tables

Supplementary Table 1: Extent of surface area that is buried by each residue in the DxETGE motif, upon binding to KEAP1, and the degree of overlap of each side-chain with the FTMap Consensus Clusters.

Supplementary Table 2: Fragment hits, with chemical structures and results observed in ThermoFluor and FA screening assays.

Supplementary Table 3: Locations and strengths of Consensus Clusters Apo and peptide-bound KEAP1 structures, and correlation with hot spots identified by alanine scanning.

Supplementary Table 4: List of Arg415 poses in 5 our structures, 17 human KEAP1 structures, and 3 co-crystal structures of Astex fragments bound mouse KEAP1.

Supplementary Table 5: Categorization of alanine scanning results into True Positive, False Positive, True Negative, False Negative, and Null results, plus correspondence with FTMap hot spots.

Supplementary Table 6. Crystallographic data collection and refinement statistics.

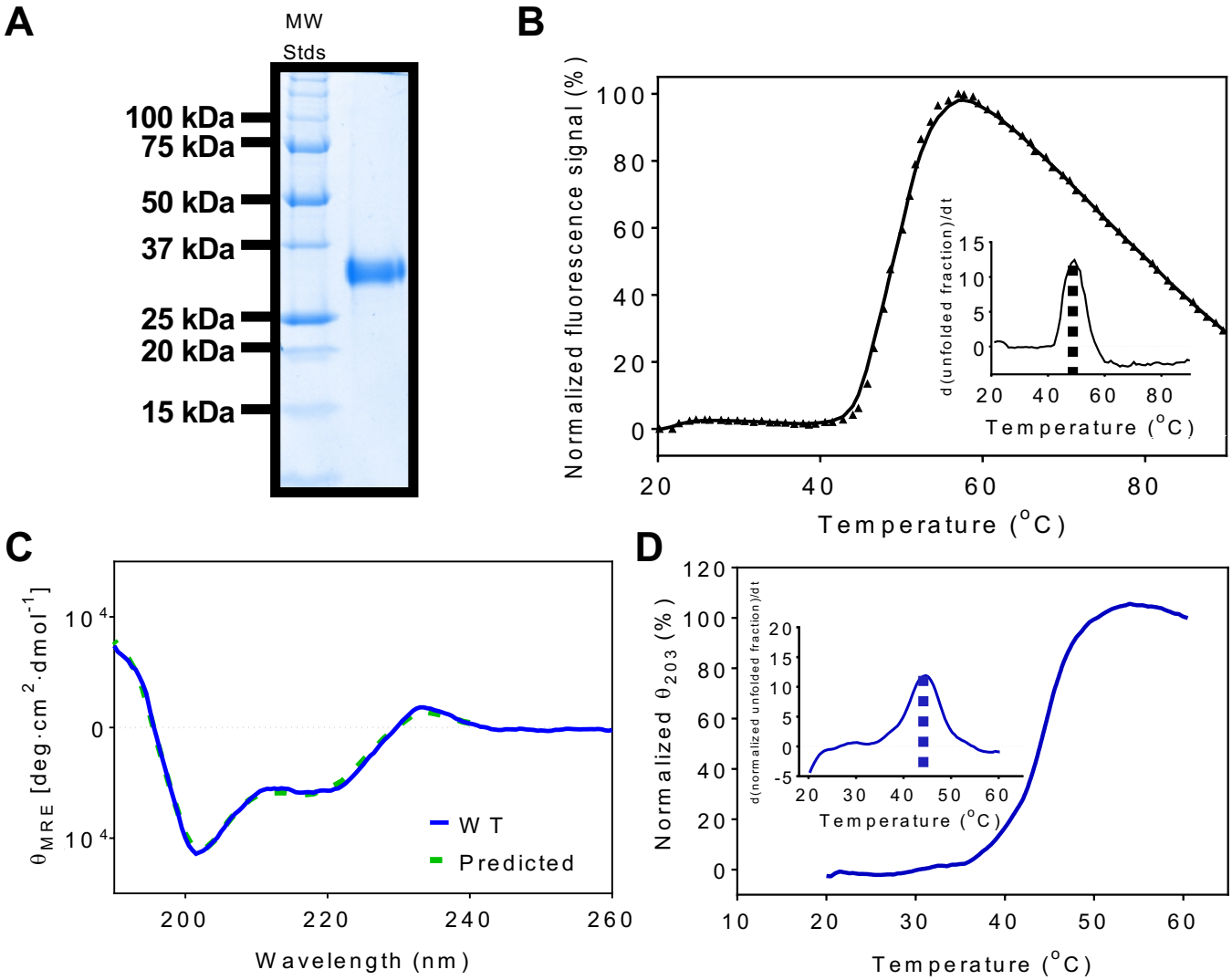


Figure S1. Characterization of KEAP1$_{312\text{-}624}$ construct. (A) SDS-PAGE analysis of the KEAP1$_{312\text{-}624}$ construct used in the study, under reducing and denaturing conditions. The intact mass of the protein was confirmed by ESI-qTOF mass spectrometry (not shown). (B) Thermal denaturation curve of KEAP1$_{312\text{-}624}$ using the ThermoFluor method, monitored by the change of fluorescence signal at 480 nm of SYPRO Orange dye. Data are representative of three independent experiments. Inset shows the first derivative of the melting curve. The T_m value for KEAP1$_{312\text{-}624}$ is 48.4℃. (C) The Circular Dichroism spectrum of KEAP1$_{312\text{-}624}$ in 25 mM Tris, pH 8.0 revealed a β-sheet secondary structure content. The program CDSSTR provided by DichroWeb was used to deconvolute the CD spectrum (dashed line). (D) Thermal denaturation data of KEAP1$_{312\text{-}624}$ using CD, monitored by the change in θ_{203}. Data are representative of two independent experiments. Inset shows the first derivative of the melting curve. The T_m value for KEAP1$_{312\text{-}624}$ construct determined by CD is 44.2℃. The thermal denaturation process is irreversible.

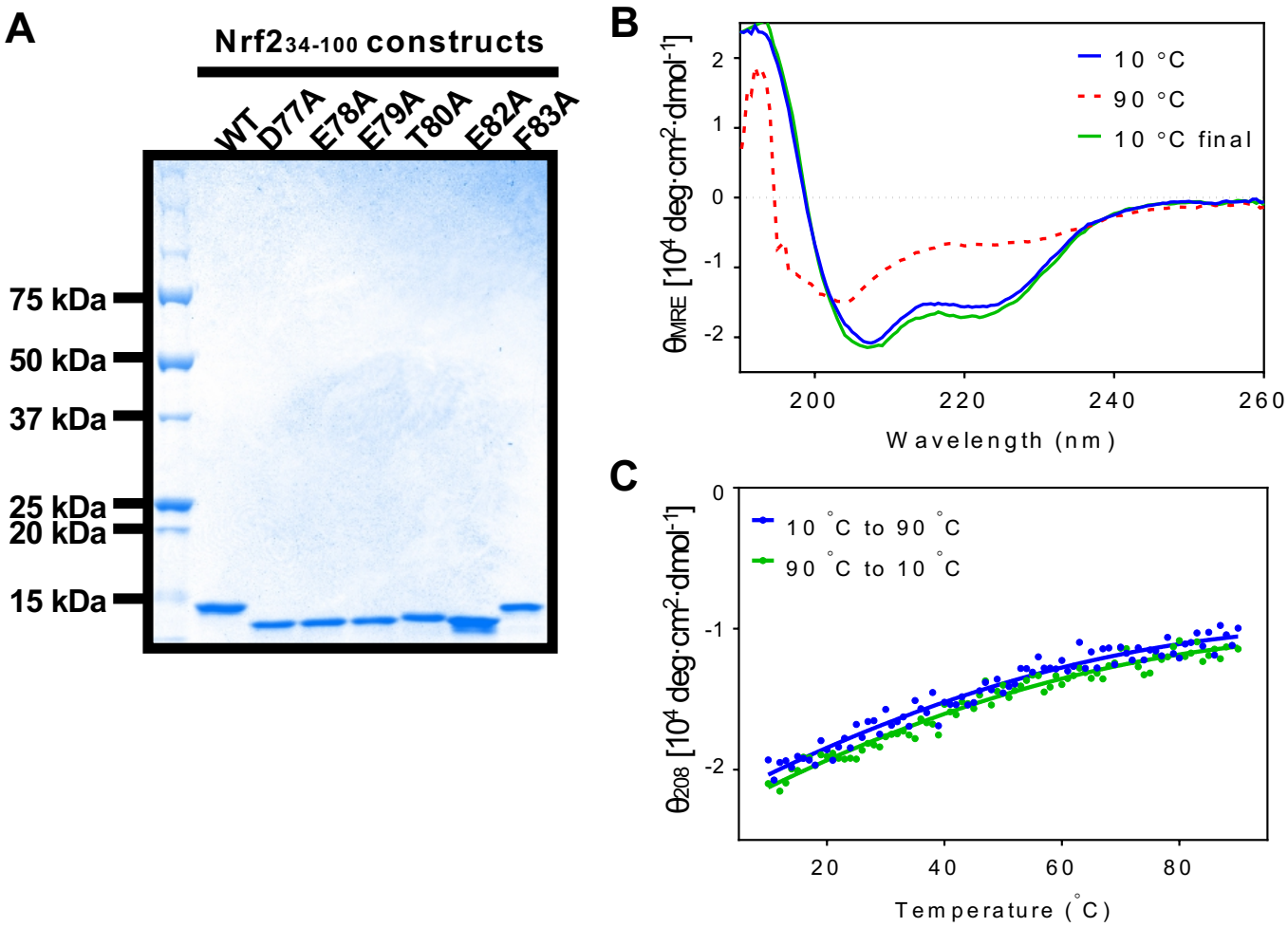


Figure S2. Characterization of Nrf2$_{34\text{-}100}$ constructs. (A) SDS-PAGE analysis of the Nrf2$_{34\text{-}100}$ constructs used in the study, under reducing and denaturing conditions. The intact mass of each protein was confirmed by ESI-qTOF mass spectrometry (not shown), showing that the small differences in apparent MW seen for some of the Nrf2 variants is a gel artifact. (B) Circular Dichroism spectrum of Nrf2$_{34\text{-}100}$ in 20 mM phosphate, 200 mM NaCl, pH 7.4 at 10℃ (blue line), after being heated to 90℃ at 1℃/min (red dashed line), and after being incubated at 90℃ for 30 min and then cooled back to 10℃ at 1℃/min (green line). (C) Thermal denaturation of Nrf2$_{34\text{-}100}$ using CD, monitored by the change in θ_{208}. Data are representative of two independent experiments. No cooperative unfolding transition was observed.

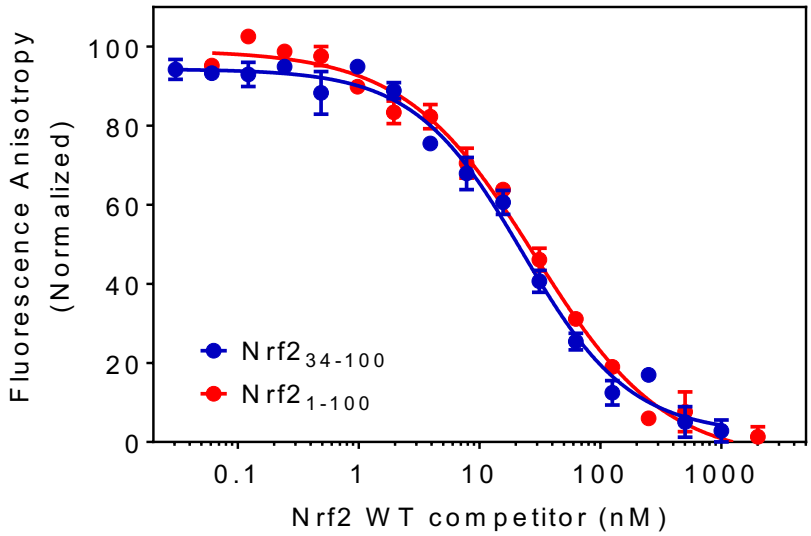


Figure S3. Comparison of the KEAP1 binding properties of $Nrf2_{1-100}$ versus $Nrf2_{34-100}$. $Nrf2_{1-100}$, which contains both the DLG and the DxETGE motifs (red), was evaluated in the FA competition assay in parallel with $Nrf2_{34-100}$ (blue), which contains DxETGE but not DLG. Anisotropy values were normalized to reflect the fractional changes between maximum and minimum anisotropy signal. Error bars show the range of duplicate experiments. Data were fitted to a competitive equilibrium binding model using DYNAFIT 4 software, as described in Materials and Methods (main text). The fits returned binding affinities of $K_D = 10 \pm 2$ nM for $Nrf2_{34-100}$ and 10 ± 1 nM for $Nrf2_{1-100}$. Results shown are representative of at least three independent experiments.

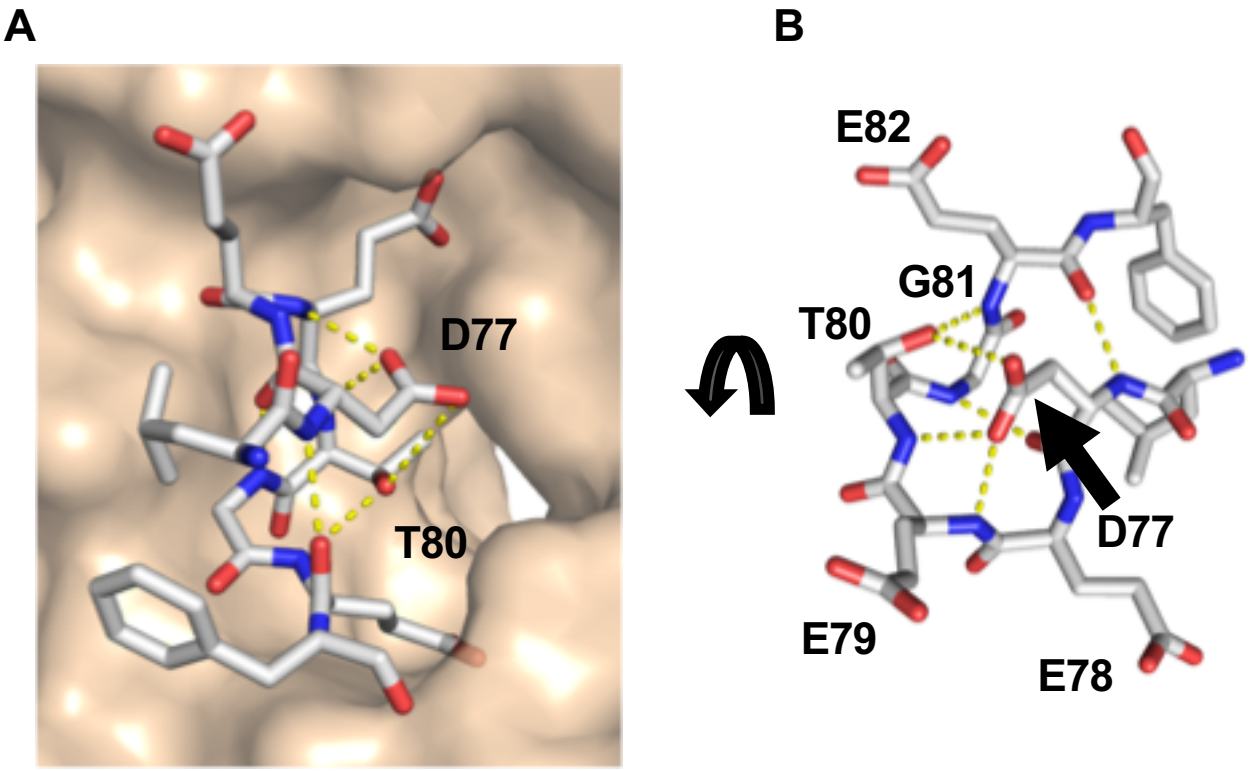


Figure S4. Intramolecular interactions involving D77 and T80 (dashed yellow lines indicate likely polar interactions). **A.** View of the bound Nrf2 peptide (white sticks), from a perspective looking down into the KEAP1 binding site (wheat surface). **B.** Peptide alone, viewed as looking from the bottom of the KEAP1 binding site (approximately a 180° vertical rotation from the view in A). Figures were created using the KEAP1/Nrf2 peptide structure from PDB 4IFL, with portions of the bound peptide lying outside the segment 76-LDEETGEF-83 omitted for clarity.

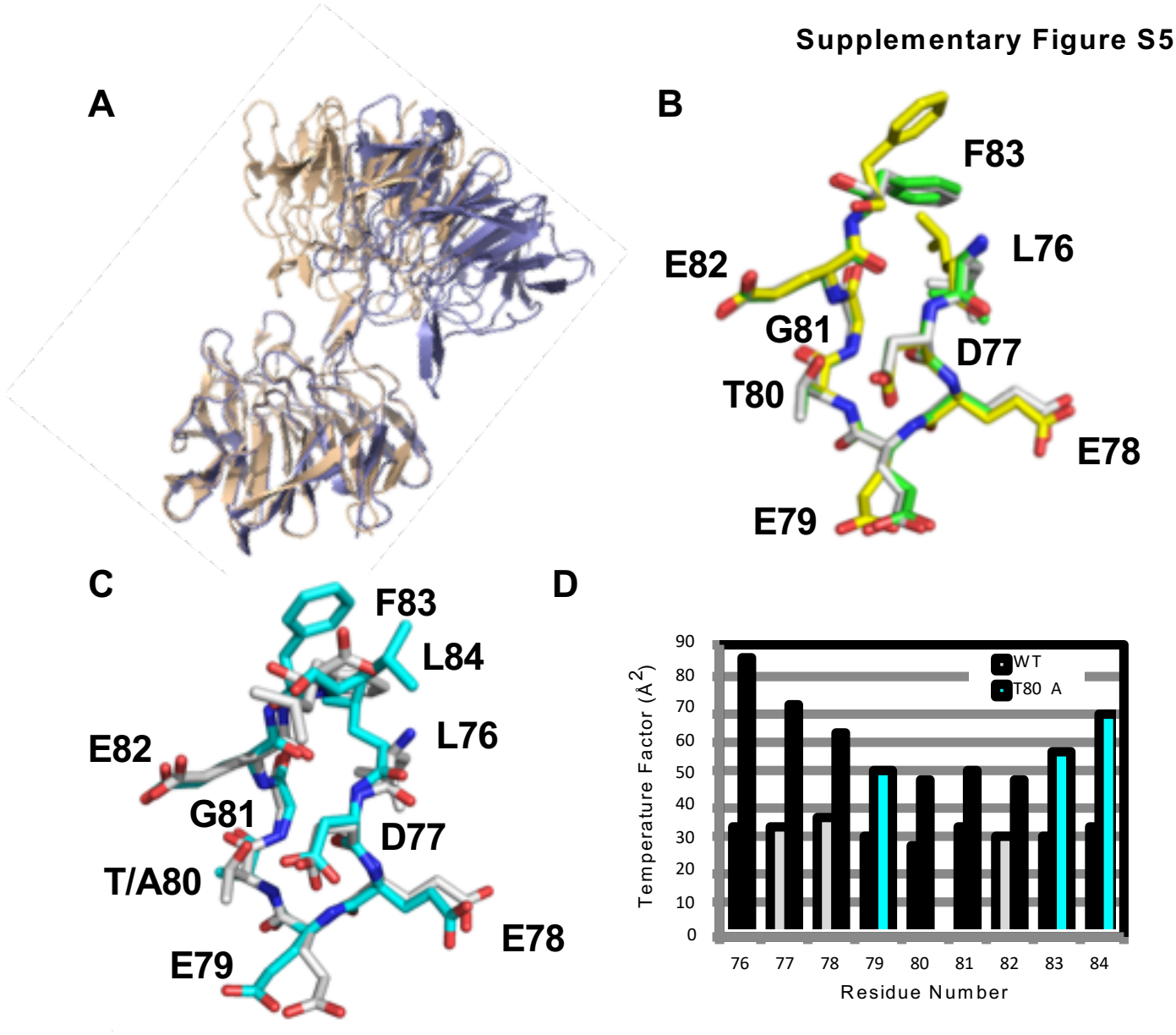


Figure S5. New crystal from of KEAP1 β-propeller domain suitable for ligand soaking. **A.** Crystallographic dimer seen in our new crystal form (purple), compared with the previously reported structure 3ZGD (wheat). In our structure, the blade 2 BC-loop of chain B interacts with the DA loop between blades 1 and 2 of chain A. This difference in interaction, compared to previous KEAP1 structures, results in a shifting of chain B away from the chain A ligand binding site, fully opening the chain A ligand binding site to the solvent channel. **B.** X-ray crystal structure obtained by soaking KEAP1 crystals with the 8-mer LDEETGEF from PDB 4IFL (white) and 2FLU (green), peptide (yellow) superimposed with the corresponding regions of published co-crystal structure of Nrf2 bound to KEAP1, showing that the Nrf2-derived peptide maintains the same binding pose and thus that the crystal form used in these studies is suitable for fragment-soaking experiments. **C.** X-ray crystal structure of an Nrf2-derived 9-mer peptide that contains a T80A substitution (LDEEAGEFL), soaked into our new crystal form of KEAP1, superimposed on the soaked wild-type structure (white sticks), showing that the binding mode of the core ETGE motif residues is largely unchanged by the T80A mutation. **D**. Histogram showing how crystallographic B-factors for each residue of the bound T80A mutant peptide from (C), averaged over all residue atoms, increase towards the termini, whereas they remain constant for the wild-type peptide. A similar trend was seen for main-chain atoms only.

			D	x	E	T	G	E			
Homo sapiens	76	L	D	E	E	T	G	E	F	L	84
Pongo abelii	60	L	D	E	E	T	G	E	F	L	68
Bos taurus	76	L	D	E	E	T	G	E	F	L	84
Macaca mulatta	76	L	D	E	E	T	G	E	F	L	84
Rattus norvegicus	76	L	D	E	E	T	G	E	F	L	84
Myotis davidii	166	L	D	E	E	T	G	E	F	L	174
Myotis brandtii	82	L	D	E	E	T	G	E	F	L	90
Fukomys damarensis	60	L	D	E	E	T	G	E	F	L	68
Heterocephalus glaber	245	L	D	E	E	T	G	E	F	L	253
Cricetulus griseus	60	L	D	E	E	T	G	E	F	L	68
Pteropus alecto	60	L	D	E	E	T	G	E	F	L	68
Capra hircus	60	L	D	E	E	T	G	E	F	L	68
Mauremys reevesii	71	L	D	E	E	T	G	E	F	I	79
Pelodiscus sinensis	71	L	D	E	E	T	G	E	F	V	79
Amazona aestiva	71	L	D	E	E	T	G	E	F	V	79
Alligator mississippiensis	68	L	D	E	E	T	G	E	F	V	76
Xenopus tropicalis	75	L	D	E	E	T	G	E	F	I	83
Anas platyrhynchos	88	L	D	E	E	T	G	E	F	V	96
Mus musculus	76	L	D	E	E	T	G	E	F	L	84
Coturnix coturnix	60	L	D	E	E	T	G	E	F	V	68
Gallus gallus	60	L	D	E	E	T	G	E	F	V	68
Oncorhynchus kisutch	75	L	D	E	E	T	G	E	F	V	83
Ctenopharyngodon idella	75	L	D	E	E	T	G	E	F	V	83
Larimichthys crocea	114	L	D	E	E	T	G	E	Y	I	122
Salmo salar	75	L	D	E	E	T	G	E	F	V	83
Xenopus laevis	75	L	D	E	E	T	G	E	F	I	83
Danio rerio	75	L	D	E	E	T	G	E	F	L	83

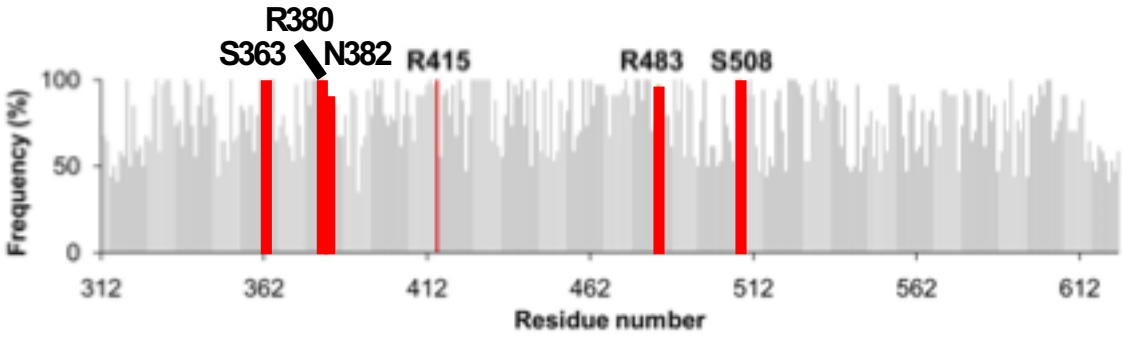


Figure S6. A. Conservation of the DxETGE motif and surrounding residues among 27 Nrf2 paralogs. **B.** Frequency of human amino acid at each position among 34 KEAP1 paralogs, showing that the six KEAP1 residues that interact with atoms on Nrf2 involved in energetically important contacts are highly conserved.

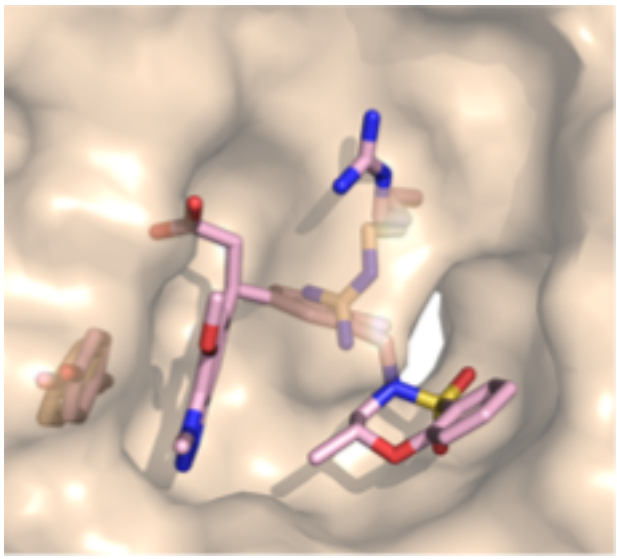


Figure S7. KEAP1 residue Tyr525 shows only minor conformattional adjustment to ligand binding. The side chains of Tyr525 and, for comparison, the highly mobile residue Arg415 in unbound KEAP1 (wheat), and in a published structure of KEAP1 bound by a small molecule (light pink, PDB 5FNU). Among the 17 reported crystal structures of human KEAP1 and 16 of mouse KEAP1, including unbound and liganded, only minor conformational shifts in Tyr525, such as shown above, are seen.

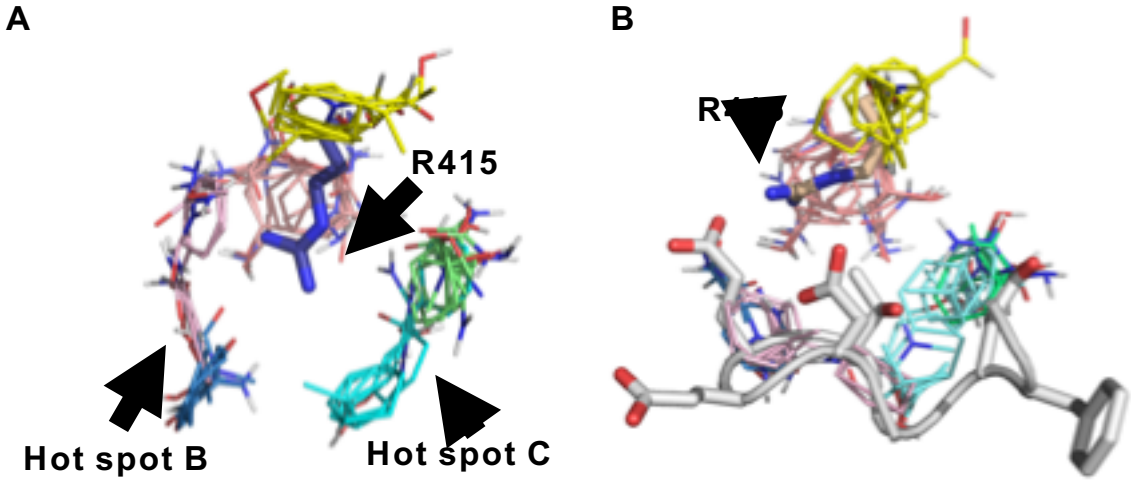


Figure S8. The repositioning of KEAP1 residue R415 upon Nrf2 binding indicates an induced fit component to the KEAP1/Nrf2 interaction. R415 conformations (sticks), relative position and strength of hot spot B and C (lines) in (**A**) unbound KEAP1 and (**B**) Nrf2 bound KEAP1 crystal structures. Upon ligand binding, the R415 side-chain moves into a position enabling the formation of a salt bridge with E79, shifting the locations of key hot spots B and C to achieve substantially greater overlap with the side-chain methyl group of T80.

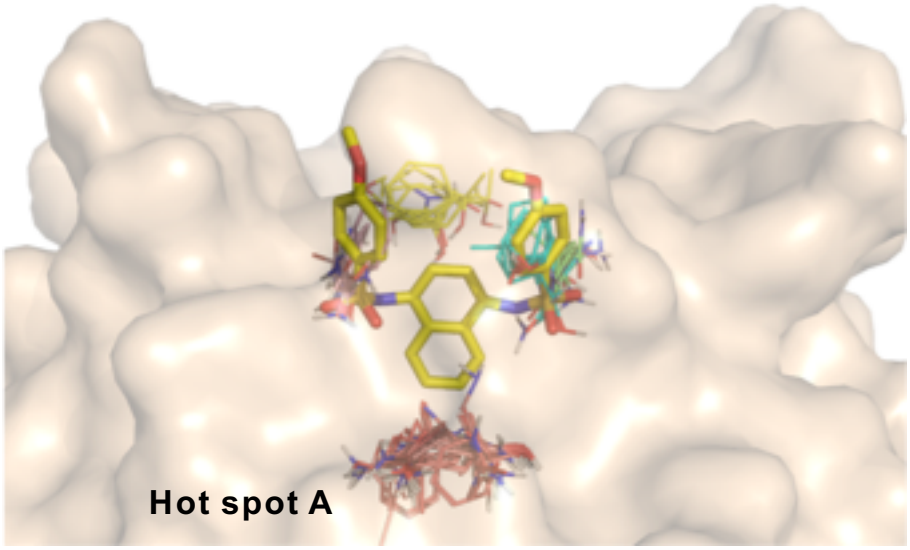


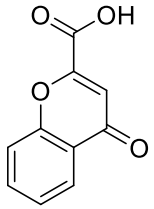
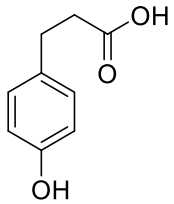
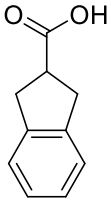
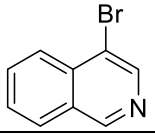
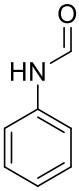
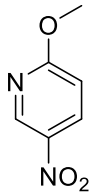
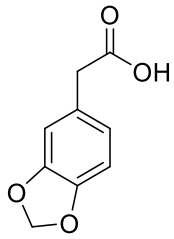
Figure S9. Example of reported small molecule inhibitor that approaches hot spot A to position an aromatic moiety close to this strong hot spot. KEAP1 shown in wheat surface, compound shown in sticks (yellow), FTMap hot spots shown in lines.

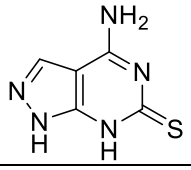
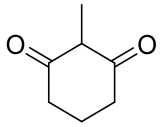
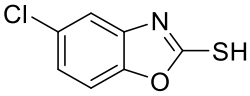
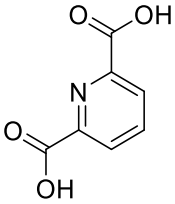
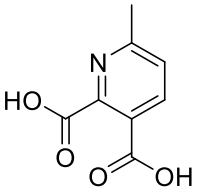
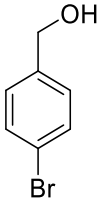
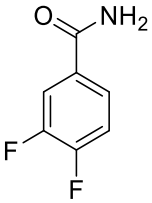
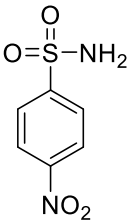
Supplementary Table 1: Extent of contact with KEAP1 and overlap with FTMap hot spots, for each residue in the Nrf2 binding motif

Nrf2 Residue	Contact Area (ΔASA, Å^2)[1]		Overlap with FTMap Clusters (Probe atoms within 2 Å of specified side-chain atom)[1]		
	All atoms[3]	Side-chain after Cβ[4]	Atom 1[5]	Atom 2[5]	Total[6]
L76	14.9	14.9	Cγ, 0	-	0
D77	18.5	16.2	Oδ1, 92	Oδ2, 74	166
E78	83.7	34.4	Oγ1, 0	Oγ2, 0	0
E79	172.3	126.7	Oγ1, 6514	Oγ2, 3212	9726
T80	82.0	37.9	Oγ1, 1453	Cδ1, 1207	2660
G81	41.7	n/a	n/a	n/a	n/a
E82	117.2	104.6	Oγ1, 1272	Oγ2, 6536	7808
F83	32.8	3.4	Cγ, 0	Cζ, 0	0

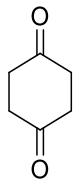
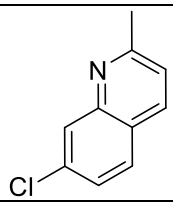
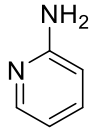
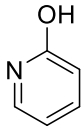
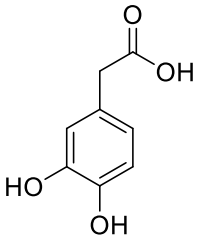
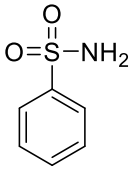
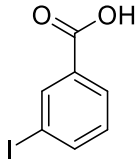
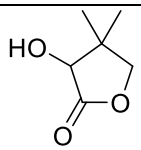
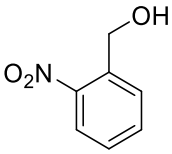
[1]Burial of solvent-accessible surface area in the KEAP1/Nrf2 complex (4IFL) attributable to the Nrf2 residue in question, calculated using PyMol as described in Materials and Methods. [2]Overlap numbers represent the average of values calculated from FTMap analysis of unbound KEAP1 (3ZDG, chain A) and bound KEAP1 (4IFL, chain x) after removal of the atoms of Nrf2. [3]Buried of solvent-accessible surface area attributable to all atoms of the residue in question. [4]Burial of solvent-accessible surface area attributable to the side-chain atoms beyond Cβ; i.e. those atoms eliminated when the residue is mutated to alanine. [5]Specific side-chain atoms that were used to calculate overlap with FTMap hot spots. [6]Overlap with FTMap hot spots, summed for all side-chain atoms that were evaluated.

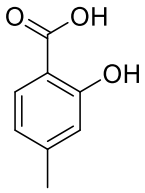
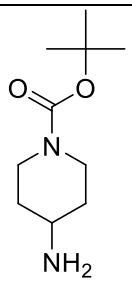
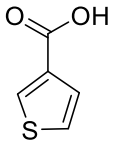
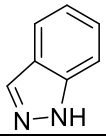
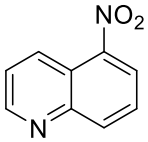
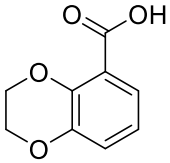
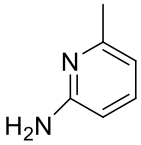
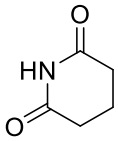
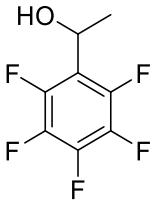
Supplementary Table 2: Table showing fragment hits, with chemical structures and results observed in ThermoFluor and FA assays.

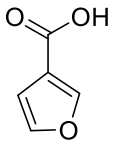
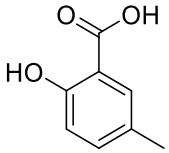
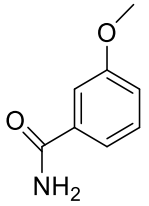
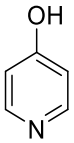
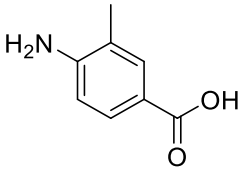
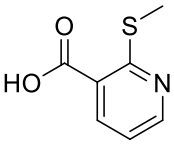
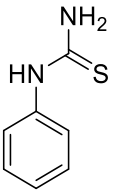
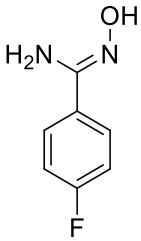
Compound number	Structure	% inhibition in FA assay	ΔT_m in ThermoFluor Assay (℃)
ZT0256		85.5	+ 4.6
ZT0633		50.5	+ 2.8
ZT0802		75.7	+ 1.3
ZT0204		90.5	+ 0.5
ZT0391		95.2	+ 0.3
ZT0418		89.0	+ 0.2
ZT0589		62.0	+ 0.2

ZT0017		Not testable poor solubility	+ 4.8
ZT0372		Not testable Fluorescence interference	+4.2
ZT0010		Not testable poor solubility	+ 3.9
ZT0707*		25.8	+3.8
ZT0676		Not testable Fluorescence interference	+3.1
ZT0563		94.5	/
ZT0602		90.3	/
ZT0740		84.6	/

* ZT0707 was excluded from the soaking attempts because it contains two unmodified carboxylates, which is structurally unlikely to be a good hit.

ZT0356		83.6	/
ZT0199		81.6	/
ZT0039		77.8	/
ZT1664		77.2	/
ZT0588		73.1	/
ZT0274		71.5	/
ZT0642		70.3	/
ZT0478		70.0	/
ZT0747		67.0	/

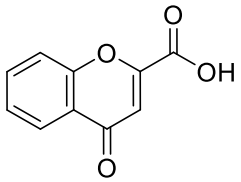
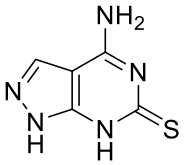
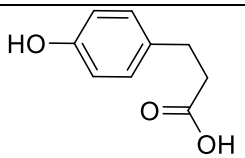
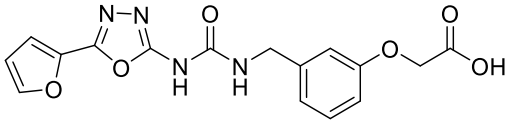
ZT0732		65.9	/
ZT0861		63.7	/
ZT0852		62.5	/
ZT0074		60.7	/
ZT0450		58.4	/
ZT1680		58.2	/
ZT0045		57.8	/
ZT0373		56.5	/
ZT0709		55.4	/

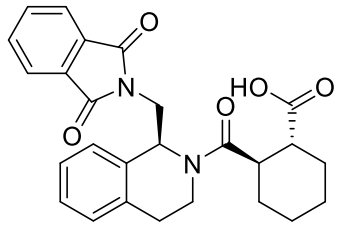
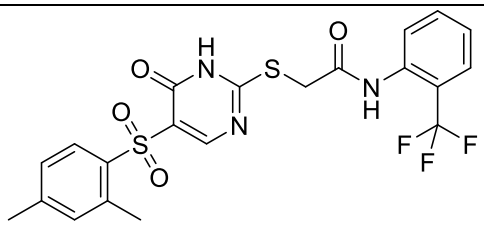
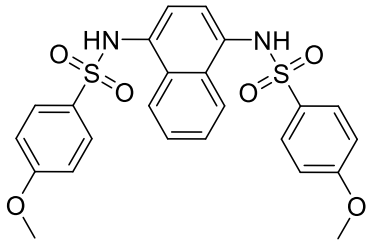
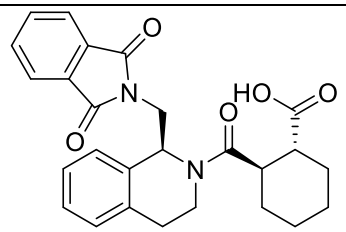
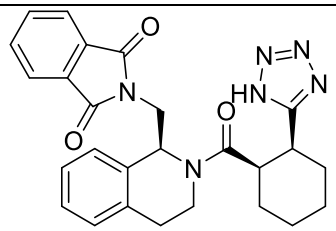
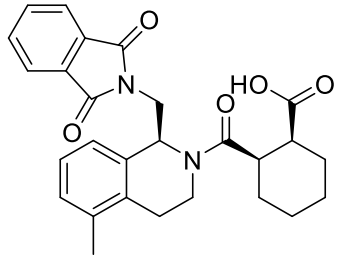
ZT0173		55.1	/
ZT0730		55.0	/
ZT0546		54.4	/
ZT0040		52.4	/
ZT0687		52.1	/
ZT0724		52.1	/
ZT0772		51.5	/
ZT0600		50.5	/

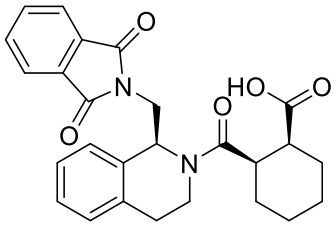
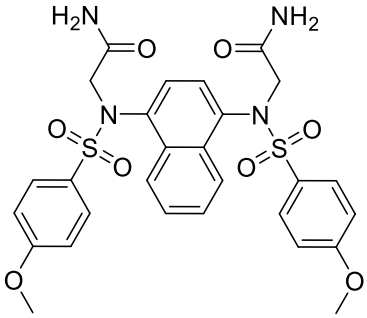
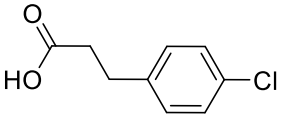
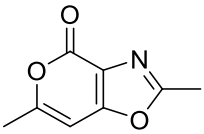
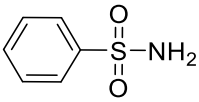
Supplementary Table 3. Consensus clusters analysis on both Apo and peptide-bound KEAP1 structure

Apo KEAP1 (No. of clusters) *Compare to fragment binding*	Peptide-bound KEAP1 *Compare to ASM*	Location (Apo/Bound)	Color in figure
CC1 (28)	CC1 (30)	Hot spot A	Red
CC2 (16)	CC2 (16)	Hot spot B	Pink
CC4 (14)	CC3 (16)	Hot spot C	Cyan
CC5 (12)	CC4 (14)	Hot spot D	Yellow
CC3 (15)	CC5 (7)	Hot spot C	Green
CC6 (9)	CC6 (4)	Hot spot B	Blue

Supplementary Table 4. Variable Arginine 415 poses in different reported crystal structures of KEAP1

PDB ID	Ligand	R415 pose	References
Unbound KEAP1	No	Down	/
Nrf2 9mer peptide bound KEAP1	LDEETGEFL	Side	/
ZT0256-bound KEAP1		Side	/
ZT0017-bound KEAP1		Side	/
ZT0633-bound KEAP1		Down (two side chain poses)	/
1U6D	No	Down	
1ZGK	No	Down	(1)
2FLU	AFFAQLQLDEETGEFL	Down	(2)
3VNG *(co-crystalized)* 3VNH *(soaking)*		Down	(3)
3ZGC	Cyclic GDEETGE	Side	(4)
3ZGD	No	Down	(4)
4IFJ	No	Down	

4IFL	AFFAQLQLDEETGEFL	Side	
4IFN		Up	
4IN4		Up	(5)
4IQK		Up	(5)
4L7B		Up	(6)
4L7C		Up	(6)
4L7D		Up	(6)

4N1B		Up	(6)
4XMB		Up	(7)
Astex Fragments Bound Mouse KEAP1			
5FNQ	 *(Fragment 1)*	Up	(8)
5FZJ	 *(Fragment 2)*	Side	(8)
5FZN	 *(Fragment 3)*	Side	(8)

Supplementary Table 5. Categorization of alanine scanning mutagenesis results on Nrf2

	ASM results	FTMap results
True Positives	Glu79, Thr80* and Glu82	Hot spot B and C
False Positives	Asp 77	No probes overlap
True Negatives	Glu78 and Phe83	No probes overlap
False Negatives	/	/
Null results	No ligand residue makes contact with the protein receptor at the site in question	Hot spot A and D

*T80 is a true positive in a qualitative but not in a quantitative sense, in that the alanine scanning result correctly identified a hot spot, but the magnitude of the loss of binding energy observed upon mutating T80 was amplified by indirect effects (see main text).

Supplementary Table 6. Crystallographic data collection and refinement statistics.

	Unliganded	**ETGE**	**EAGE**	**ZT0256**	**ZT0017**	**ZT0633**
PDB Accession Code	**5WFL**	**5WFV**	**5WG1**	**5WHL**	**5WIY**	**5WHO**
Wavelength (Å)	0.97954	0.979100	0.97946	1.075	1.075	0.9792
Resolution range (Å)	34.37 - 1.93 (1.999 - 1.93)	52.14 - 1.91 (1.978 - 1.91)	34.32 - 2.021 (2.094 - 2.021)	51.85 - 2.23 (2.31 - 2.23)	51.93 - 2.23 (2.31 - 2.23)	71.22 - 2.5 (2.589 - 2.5)
Space group	*C* 1 2 1	*C* 1 2 1	*C* 1 2 1	*I* 1 2 1	*C* 1 2 1	*C* 1 2 1
Unit cell	a = 162.314, b = 68.73, c = 77.167; β = 117.729º	a = 162.919, b = 69.253, c = 79.228; β = 118.498º	a = 161.809, b = 68.649, c = 77.126; β = 117.54º	a = 78.46, b = 69.09, c = 143.206; β = 91.1118º	a = 162.04, b = 69.18, c = 78.6; β = 117.83º	a = 160.849, b = 68.306, c = 77.406; β = 117.684º
Total reflections	191776 (18279)	205616 (20860)	117920 (11396)	73711 (7131)	328656 (30639)	97584 (9639)
Unique reflections	55204 (5414)	58943 (5843)	46561 (4570)	37484 (3718)	37651 (3724)	25927 (2545)
Multiplicity	3.5 (3.4)	3.5 (3.6)	2.5 (2.5)	2.0 (1.9)	8.7 (8.2)	3.8 (3.8)
Completeness (%)	97.32 (96.02)	92.77 (89.22)	94.62 (94.03)	96.55 (89.38)	99.41 (98.37)	99.56 (99.69)
Mean I/sigma(I)	10.26 (3.05)	15.18 (3.74)	6.27 (2.03)	12.62 (2.85)	28.50 (2.46)	11.60 (1.27)
Wilson B-factor	31.55	28.16	35.87	33.19	35.95	41.93
R-merge	0.09152 (0.2129)	0.06899 (0.3304)	0.1538 (0.3269)	0.039 (0.279)	0.8581 (2.115)	0.1519 (1.05)
R-meas	0.1083 (0.2523)	0.08143 (0.389)	0.1895 (0.4091)	0.05516 (0.3946)	0.9134 (2.263)	0.1772 (1.226)
R-pim	0.05721 (0.1338)	0.04283 (0.2037)	0.1089 (0.2428)	0.039 (0.279)	0.3104 (0.7927)	0.09077 (0.6288)
CC1/2	0.99 (0.965)	0.994 (0.886)	0.971 (0.861)	0.998 (0.835)	0.705 (0.225)	0.986 (0.482)
CC*	0.997 (0.991)	0.999 (0.969)	0.993 (0.962)	0.999 (0.954)	0.909 (0.606)	0.996 (0.806)
Reflections used in refinement	55164 (5408)	55913 (5372)	46547 (4570)	36243 (3359)	37445 (3682)	25815 (2545)
Reflections used for R-free	1650 (160)	1352 (133)	1994 (195)	1936 (178)	1990 (191)	2015 (200)
R-work	0.1995 (0.2352)	0.2006 (0.2557)	0.2320 (0.2402)	0.2091 (0.2507)	0.1967 (0.3131)	0.2164 (0.3295)
R-free	0.2276 (0.2761)	0.2266 (0.2759)	0.2676 (0.2413)	0.2436 (0.2841)	0.2387 (0.3298)	0.2700 (0.3776)
CC(work)	0.950 (0.934)	0.954 (0.847)	0.929 (0.891)	0.946 (0.863)	0.830 (0.501)	0.942 (0.695)
CC(free)	0.951 (0.837)	0.949 (0.812)	0.926 (0.859)	0.921 (0.835)	0.762 (0.337)	0.894 (0.606)

Number of non-hydrogen atoms	4578	4606	4502	4546	4463	4420
Protein	4400	4412	4365	4357	4396	4362
Ligand/Bound Ions	25	74	71	14	11	12
Solvent	153	115	56	155	41	31
Protein residues	573	573	569	569	573	569
RMS (bonds)	0.014	0.007	0.004	0.002	0.007	0.008
RMS (angles)	1.36	0.90	0.70	0.51	0.94	1.01
Ramachandran favored (%)	96.31	96.88	95.80	94.51	96.84	95.75
Ramachandran allowed (%)	3.51	3.12	4.02	5.49	3.16	4.25
Ramachandran outliers (%)	0.18	0.00	0.17	0.00	0.00	0.00
Rotamer outliers (%)	0.00	0.00	0.00	0.00	0.00	0.65
Clashscore	8.36	6.84	9.23	17.34	10.34	20.12
Average B-factor	53.73	37.12	58.47	43.97	49.47	59.73
Protein	54.00	37.30	58.57	44.16	49.51	59.77
Ligand	N/A	34.11	64.34	60.16	65.83	89.85
Solvent	41.48	31.18	37.05	36.39	32.55	35.28

Statistics for the highest-resolution shell are shown in parentheses.

* Friedel mates were averaged when calculating reflection statistics.

References

1. Beamer LJ, Li XC, Bottoms CA, & Hannink M (2005) Conserved solvent and side-chain interactions in the 1.35 angstrom structure of the Kelch domain of Keap1. *Acta Crystallogr D* 61:1335-1342.
2. Lo SC, Li XC, Henzl MT, Beamer LJ, & Hannink M (2006) Structure of the Keap1 : Nrf2 interface provides mechanistic insight into Nrf2 signaling. *Embo J* 25(15):3605-3617.
3. Satoh M*, et al.* (2015) Multiple binding modes of a small molecule to human Keap1 revealed by X-ray crystallography and molecular dynamics simulation. *Febs Open Bio* 5:557-570.
4. Horer S, Reinert D, Ostmann K, Hoevels Y, & Nar H (2013) Crystal-contact engineering to obtain a crystal form of the Kelch domain of human Keap1 suitable for ligand-soaking experiments. *Acta Crystallogr F* 69:592-596.
5. Marcotte D*, et al.* (2013) Small molecules inhibit the interaction of Nrf2 and the Keap1 Kelch domain through a non-covalent mechanism. *Bioorgan Med Chem* 21(14):4011-4019.
6. Jnoff E*, et al.* (2014) Binding Mode and Structure-Activity Relationships around Direct Inhibitors of the Nrf2-Keap1 Complex. *Chemmedchem* 9(4):699-705.
7. Jain AD*, et al.* (2015) Probing the structural requirements of non-electrophilic naphthalene-based Nrf2 activators. *Eur J Med Chem* 103:252-268.
8. Davies TG*, et al.* (2016) Monoacidic Inhibitors of the Kelch-like ECH-Associated Protein 1: Nuclear Factor Erythroid 2-Related Factor 2 (KEAP1:NRF2) Protein Protein Interaction with High Cell Potency Identified by Fragment-Based Discovery. *J Med Chem* 59(8):3991-4006.