# Supporting Information

## Hartmann et al. 10.1073/pnas.0907256106

### SI Text

**Bioinformatics.** The SadA protein used in this study is ORF STM3691 of *Salmonella enterica* subsp. *enterica* serovar *typhimurium* (formerly *S. typhimurium* LT2), GenBank accession NP_462591. The domains of this protein were annotated with the help of the daTAA platform (1) and their structural properties were predicted in the MPI Bioinformatics Toolkit (http://toolkit.tuebingen.mpg.de/; ref. 2). Results have partly been reported elsewhere (3). The detailed annotation is shown in Fig. S1.

For the survey of known structures, coiled-coil assignments were derived from protein structures available from the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB) before February 1, 2009, using the SOCKET algorithm (4). The resulting coiled-coil-positive structures were collated in the CC+ database (http://coiledcoils.chm.bris.ac.uk/ccplus; ref. 5). Searches of CC+ were configured to return only N@d motifs found within parallel, homooligomeric coiled-coil assignments longer than 14 aa. The returned motifs were further refined, discarding redundant examples of structures with more than 50% sequence identity and with complex coiled-coil architectures, as detailed in Table S1. The local coiled-coil and $\alpha$-helical parameters shown in Fig. 4 and Fig. S3*D* were determined using TWISTER (6). The values used for the periodicity plot were averaged over a window of 3 residues to mask local fluctuations. The ideal Crick angle values for each heptad position, required for the calculation of the shift in the Crick angle, are average values obtained from a structure of a trimeric coiled coil GCN4-II (1GCM). Plots were prepared with PyX, a Python graphics package (http://pyx.sourceforge.net/).

**Protein Sample Preparation.** The SadAK3 construct contains residues 479–519 of SadA, fused at both the N and C termini in the correct heptad register to GCN4 leucine zippers in the expression vector pIBA-GCN4tri (3). SadAK3b is a similar construct containing residues 483–523, out of register by 4 residues to the leucine zippers, the heptad register at the junctions being *..fga fga..* and *..bcd abc...* A sixfold His-tag was fused to the C-terminal GCN4 adaptor to simplify purification.

Cloning, expression, and purification of SadAK3 have been described previously (3). For construction of SadAK3b the primers GACCATGGTCTCCGATTGAGAAAGTGGATCAGAACACCGCTGATATCACCACCAATACCAACAGCATCAATCAGAACACCACTGATATTGC and GACCATGGTCTCCTCATCAGGGTGGTAATGGAATCGCTCAGGTTGTTGATATTGGTGGTGTTGGTGGCAATATCAGTGGTGTTCTG were used. Cloning, expression, and purification of SadAK3b were performed as described for SadAK3His (3). Pure SadAK3b was refolded by dialysis against 20 mM Tris/HCl pH 7,5; 40 mM NaCl; 5% (vol/vol) glycerin and concentrated to 3.7 mg/mL. Peptide concentrations were determined by measuring tyrosine absorbance at 280 nm in water.

Peptides were synthesized by the core facility of the Max Planck Institute of Biochemistry, or purchased from EMC Microcollections. They have been tested for purity by MALDI MS. For crystallization all peptides were dissolved in 20 mM Mops, pH 7.7, 50 mM NaCl + 0.05% azide at a concentration of 10 mg/mL.

**CD Spectroscopy.** Circular dichroism (CD) measurements were performed using a JASCO J-810 spectropolarimeter (Jasco) equipped with a thermoelectric temperature controller. For CD studies, peptide and protein concentrations and buffers are denoted in Fig. S3 and Fig. S4. Spectra were recorded using a 0.1-cm path length cuvette at a 1-nm band width with response of 2 s. For far-UV CD spectra, a scanning speed of 100 nm/min and a data pitch of 1 nm were used. Thermal denaturation was monitored at 222 nm with a temperature ramp of 1 °C per min and a data pitch of 0.5 °C. For subtracting the buffer baseline, smoothing, and calculation of molecular ellipticities, the software provided by Jasco was used.

**Protein Digestion and Mass Spectrometry.** Proteinase K digestion of SadAK3b was performed by incubation of 40 $\mu$g protein with 12 ng proteinase K (900 units/mL, Fermentas) in 20 mM Tris/HCl pH 7.5; 40 mM NaCl; 5% (vol/vol) glycerin for 15 min at indicated temperatures. The digestion was stopped by addition of PMSF with a final concentration of 1 mM.

Proteins were separated over a 18% urea SDS/PAGE gel and stained with coomassie, bands of interest excised and digested in gel with trypsin (Promega) (7). The extracted peptides were separated by reversed-phase HPLC (nanoLC2D, Eksigent) using a 90-min gradient. Mass spectrometric analysis was performed on an ion trap (HCTultra, Bruker Daltonics) equipped with a nanoESI source from Proxeon. Mascot generic data files were searched against a modified SwissProt protein database including an entry for the recombinant GCN4pII-SadAK using Mascot Server (V2.2). The following settings were used: No enzyme, carbamylation of cystein as fixed modification, oxidation of methionin as variable, 0.5 Da peptide mass accuracy and 0.3 Da for fragmentation masses.

**Crystallization, Data Collection, and Structure Solution.** All crystallization screens were performed with ≈1,200 conditions by mixing 400 nL of reservoir with 400 nL of protein solution on 96-well Corning 3550 plates, using a Honey bee 961 crystallization robot (Genomic Solutions). The reservoir volume was 75 $\mu$L. The conditions for all crystals used in the diffraction experiments and, when applicable, soaking conditions are listed in Table S2. In particular, crystals of M3-IxxNT-chloride were soaked in their mother liquor supplemented with either NaI or $KNO_3$ to substitute chloride by iodide or nitrate. All crystals were loop mounted, flash frozen in liquid nitrogen, and all data collected at beamline PXII at the SLS (Paul Scherrer Institute, Villigen, Switzerland) under cryo conditions at 100 K using a mar225 CCD detector. Diffraction images were processed and scaled using the *XDS* program suite (8). The structures of SadAK3, SadAK3b-V1, SadAK3b-V2, M1-VxxNx, and M2-VxxNx were solved using the program MOLREP (9) and the GCN4-structure 1GCM as a search model. For SAD phasing of M3-IxxNT-bromide, anomalous data were collected at a wavelength of 0.9184 Å, slightly above the Br K edge. Initial phases were calculated to 1.3 Å using the 3 Br sites identified with SHELXD/SHELXE (10) and used for automatic chain tracing with ARP/WARP (11). M3-IxxNT-chloride, M3-IxxNT-iodide, and M3-IxxNT-nitrate were later solved on the basis of the M3-IxxNT-bromide coordinates. Where applicable, ARP/WARP was used for automated rebuilding. All structures were completed in cyclic manual modeling with Coot (12) and refinement with REFMAC5 (13). Analysis with Procheck (14) showed good geometries for all structures, without any Ramachandran outliers. Data collection and refinement statistics are summarized in Table S3 together with PDB accession codes. All

molecular depictions were prepared using MolScript (15) and Raster3D (16).


1. Szczesny P, Lupas A (2008) Domain annotation of trimeric autotransporter adhesins–daTAA. *Bioinformatics* 24(10):1251–1256.
2. Biegert A, Mayer C, Remmert M, Soding J, Lupas AN (2006) The MPI Bioinformatics Toolkit for protein sequence analysis. *Nucleic Acids Res* 34(Web Server issue):W335–W339.
3. Hernandez Alvarez B, et al. (2008) A new expression system for protein crystallization using trimeric coiled-coil adaptors. *Protein Eng Des Sel* 21(1):11–18.
4. Walshaw J, Woolfson DN (2001) Socket: A program for identifying and analysing coiled-coil motifs within protein structures. *J Mol Biol* 307(5):1427–1450.
5. Testa OD, Moutevelis E, Woolfson DN (2009) CC+: A relational database of coiled-coil structures. *Nucleic Acids Res* 37(database issue):D315–D322.
6. Strelkov SV, Burkhard P (2002) Analysis of alpha-helical coiled coils with the program TWISTER reveals a structural mechanism for stutter compensation. *J Struct Biol* 137(1–2):54–64.
7. Shevchenko A, Wilm M, Vorm O, Mann M (1996) Mass spectrometric sequencing of proteins from silver stained polyacrylamide gels. *Anal Chem* 68(5):850–858.
8. Kabsch W (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J Appl Crystallogr* 26:795–800.
9. Vagin A, Teplyakov A (2000) An approach to multi-copy search in molecular replacement. *Acta Crystallogr D Biol Crystallogr* 56(Pt 12):1622–1624.
10. Sheldrick GM (2008) A short history of SHELX. *Acta Crystallogr A* 64(Pt 1):112–122.
11. Perrakis A, Morris R, Lamzin VS (1999) Automated protein model building combined with iterative structure refinement. *Nat Struct Biol* 6(5):458–463.
12. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2126–2132.
13. Murshudov GN, Vagin AA, Lebedev A, Wilson KS, Dodson EJ (1999) Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallogr D Biol Crystallogr* 55(Pt 1):247–255.
14. Laskowski RA, Macarthur MW, Moss DS, Thornton JM (1993) Procheck: A program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 26:283–291.
15. Kraulis PJ (1991) Molscript: A program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr* 24:946–950.
16. Merritt EA, Bacon DJ (1997) Raster3D: Photorealistic molecular graphics. *Methods Enzymol* 277:505–524.

**The domains of the sequence are marked as follows:**
signal sequence; the predicted cleavage site is marked by /
YadA-like head repeats (left-handed beta roll)
neck sequence
lefthanded (heptad) coiled coil; hydrophobic core residues, polar core residues
righthanded (pentadecad) coiled coil; YTD motif
FGG motif
DALL2 motif
HANS motif
DALL1 motif
anchor domain
region used for the SadAK3 and SadAK3b constructs

**In repetitive regions, individual sequence repeats are written underneath each other.**

>gi|16766976|ref|NP_462591.1| putative inner membrane protein [Salmonella typhimurium LT2]
MNRIFKVLWNAATGTFVVTSETAKSRGKKNGRRKLAVSALIGLSSIMVSADALA/NAGNDTGDGVTPTGTQTGGKGWIAIGTDATANTYTNVDGA
SAAMGYKASAMGKW
STAIGSYSQSTGDS
SLALGVKSVSAGDR
AIAMGASSSASGSY
SMAMGVYANSSGAK
SVALGYKSVASGAT
SSALGYQATASGDD
SAAFGNGAKAIGTN
SVALGSGSVAQEDN
SVAVGNSTTQRQ
ITYVAKGDINSTSTDAVTGAQIYSLSQSVADRLGGGASVNSDGTVNAPLYEVGTGIYNNVGSALSALNTS----------ITNTEASVAGLAEDALLWDESISAFSASHTGNAS-K
ITNLAAGTLAADSTDAVNGSQLFDTNEKVD----------KNTADIATNTGSINQNTADITANTDSINQNTTDIAANTTSINQNTTDIATNTTNINSLSDSVTTLTDDALLWDAASGAFSAKHNGSDS-K
ITNLAAGTLAADSTDAVNGSQLFDTNEKVD----------QNTADITTNTNSINQNTTDIATNTTNINNLSDSITTLTDDALLWDAASGAFSANHNGSAS-K
ITNLAAGTLAADSTDAVNGSQLFATNENVS----------QNTADITTNTNSINQNTTDIATNTTSINNLSDSITTLTDDALLWDAASGTFSASRSGSAS-K
ITNLAAGTLAADSTDAVNGSQLYETNQKVD----------QNTS----------AIADINTSITNLSSDNLSWNETTSSFSASHGSSTTNK
ITNVAAGELSEESTDAVNGSQLFETNEKVD----------QNTTDIAANTTNITQNSTAIENLNTSVSDINTSITGLTDNALLWDEDTGAFSANHGGSTS-K
ITNVAAGALSEDSTDAVNGSQLYETNQKVD----------QNTS----------AIADINTSITNLGTDALSWDDEEGAFSASHGTSGTNK
ITNVAAGEIASDSTDAVNGSQLYETNMLISQYNESISQLAGDTSETYITENGTGVKYIRTNDNGLEGQDAYATGNG
ATAVGYDAVASGAG
SLALGQNSSSSIEG
SIALGSGSTSNR
AITTGIRETSATSD
GVVIGYNTTDRELLG
ALSLGTDGESYRQ
ITNVADGSEAQDAVTVRQLQNAIGAVTTTPTKYYHANSTEEDSLAVGTD
SLAMGAKTIVNADA
GIGIGLNTLVMADAIN
GIAIGSNARANHAN
SIAMGNGSQTTRGAQTDYTAYNMDTPQNSVG
EFSVGSEDGQRQ
ITNVAAGSADTDAVNVGQLKVTDAQVSRNTQSITNLNTQVSNLDTRVTNIENGIGDIVTTGSTKYFKTNTDGADANAQGAD
SVAIGSGSIAAAEN
SVALGTNSVADEAN
TVSVGSSTQQRR
ITNVAAGVNNTDAVNVAQLKASEAGSVRYETNADGSVNYSVLNLGDGSGGTTR
IGNVSAAVNDTDAVNYAQLKRSVEEANTYTDQKMGEMNSKIKGVENKMSGGIASAMAMAGLPQAYAPGANMTSIAGGTFNGESAVAIGVSMVSESGGWVYKLQGTSNSQGDYSAAIGAGFQW

**Fig. S1.** Annotation of the SadA sequence.

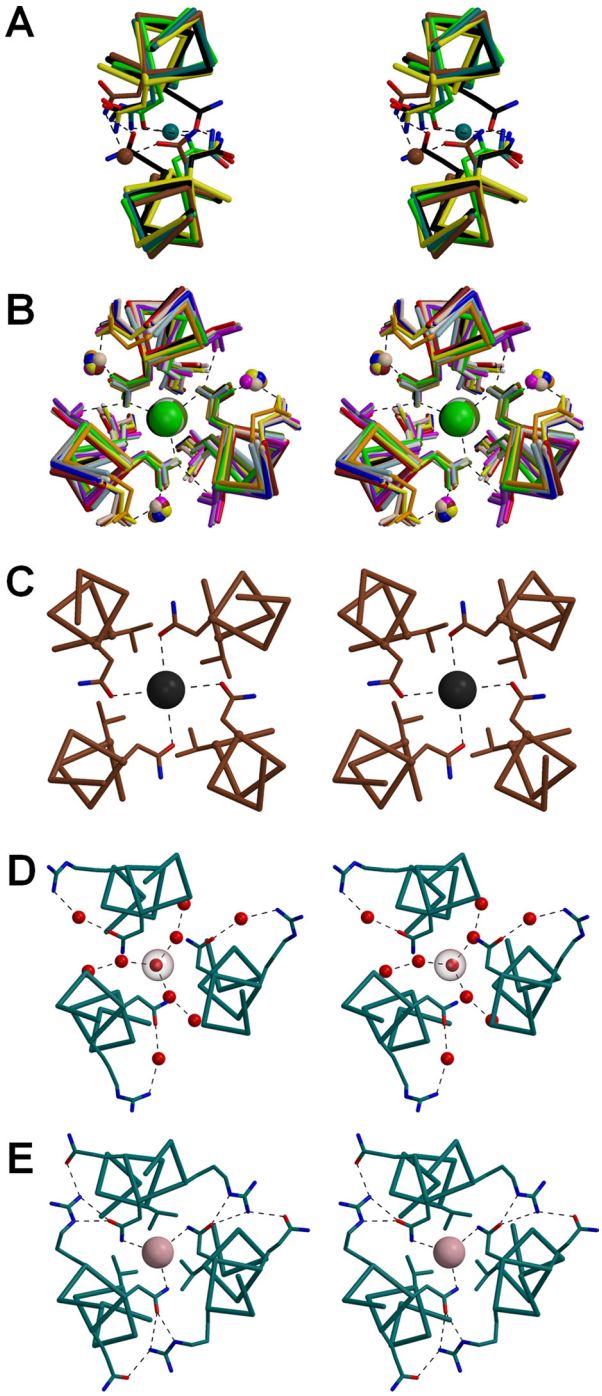


**Fig. S2.** N@*d* layers from proteins of known structure. (*A*) Superposition of 5 dimeric exemplars as listed in Table S1, some of them binding water. (*B*) Superposition of 16 trimeric exemplars, including 3 instances from this study as listed in Table S1, all coordinating chloride. Six of the structures include a T@*e*, 7 include an N@*g* or D@*g*. (*C*) The single tetrameric example (1EZJ) coordinating calcium. (*D*) Special case of a trimer (3EFG) with glycine in position *a* and an alanine in the preceeding *d* position, resulting in a water network penetrating the core upsteam the N@*d* layer. The coordination of chloride (transparent) is not shown for clarity. (*E*) The elaborate network of interactions of Q@*c*, N@*d*, and R@*e* in 1Y4M. Structures 3EFG, 2POH, 2FYZ, and 1Y4M needed to be re-refined on the basis of the experimental data deposited in the PDB.

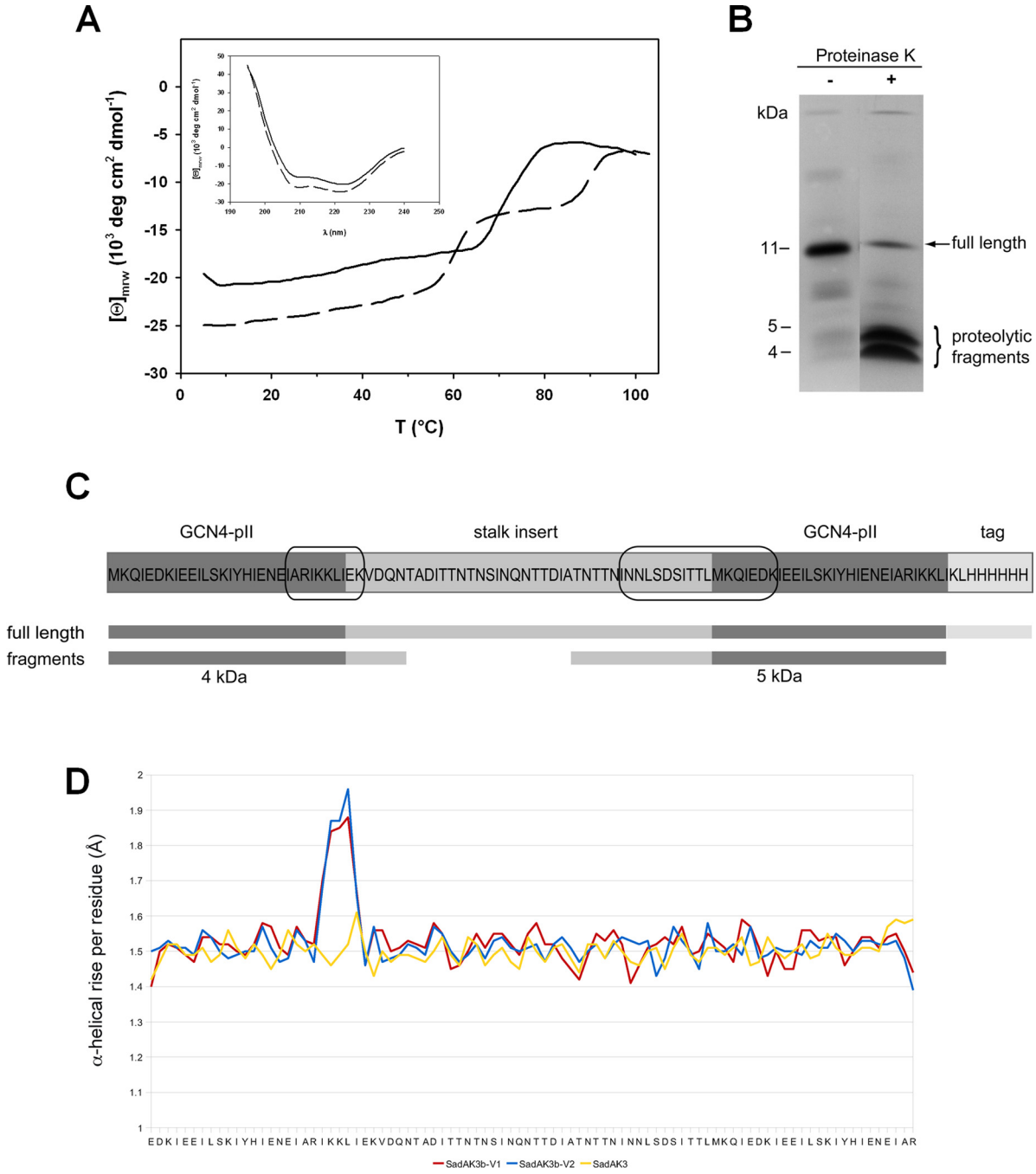


**Fig. S3.** Proteolytic stability and structural properties of SadAK3 and SadAK3b. (*A*) Thermal unfolding and CD spectra (*Inset*) of SadA constructs. SadAK3 and SadAK3b are shown as solid and broken lines, respectively. The measurements were performed at a concentration of 15 $\mu$M protein in 20 mM Tris/HCl pH 7.5, 40 mM NaCl, 5% glycerol. For the single CD spectra $[\Box]_{mrw}$ were determined at 10 °C. (*B*) SadAK3b was incubated with and without proteinase K at 60 °C as described. Samples were separated using 18% urea SDS PAGE. Protein bands corresponding to the undigested, full-length protein and the marked proteolytic fragments with a size of 4 and 5 kDa were excised from the gel and after tryptic in-gel digestion analyzed by LC-MS/MS. (*C*) Schematic picture showing the proteolytic fragments identified by mass spectroscopy aligned to the SadAK3b sequence. The GCN4-pII adaptors are highlighted with dark gray, the insert and the His-tag with light gray. Local discontinuities of the heptad repeat at the interface of the insert and the adjacent GCN4-pII adaptors are enframed. (*D*) $\alpha$-Helical rise per residue as calculated by TWISTER (6) for SadAK3, SadAK3b-V1, and -V2. For $\alpha$-helices this parameter is expected to be constant at 1.5 Å, in $3_{10}$ helices it is increased to 2 Å. In contrast to SadAK3, which is all $\alpha$-helical, both SadAK3b-V1 and -V2 contain a short $3_{10}$ helix within the stammer (compare Fig. 4). The sequence below the graph corresponds to SadAK3b and is aligned to panel *C*.

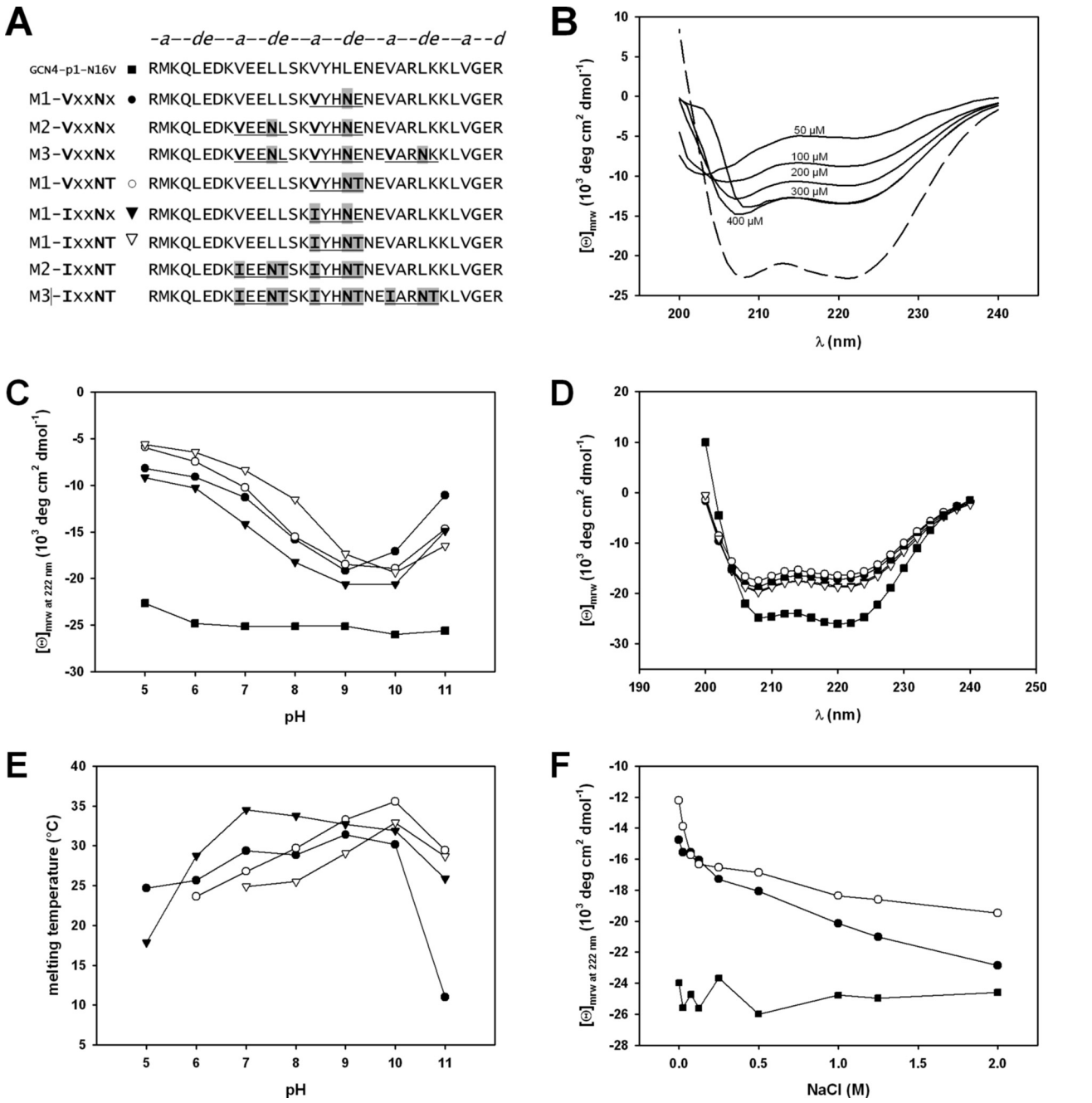


**Fig. S4.** N@d containing peptides are destabilized in their secondary structure. (*A*) Sequence of synthetic peptides used in this work. The sequence of GCN4-p1 N16V was modified by mutating specific amino acid residues (highlighted in gray) to create V/IxxNxx and V/IxxNTxx heptad motifs. (*B*) CD spectra of peptide M1-VxxNT in the concentration range from 50 to 400 $\mu$M (solid lines) in comparison to 100 $\mu$M GCN4-p1 N16V (broken line). Measurements were performed at 20 °C in 25 mM sodium phosphate buffer, pH 7, 25 mM NaCl. (*C–F*) Folding of N@d containing peptides is dependent on pH and ionic strength. In all panels the peptides are depicted by the following symbols: GCN4-p1 N16V (■), M1-VxxN (●), M1-VxxNT (○), M1-IxxN (▼), and M1-IxxNT (▽). (*C*) Folding of N@d-containing peptides is pH dependent. Molar ellipticities at 222 nm were determined as a function of pH at 10 °C in 25 mM buffer (sodium acetate pH 5, sodium phosphate pH 6 to 8, CHES pH 9 and 10, and CAPS pH 11) containing 25 mM NaCl. The peptide concentration was 100 $\mu$M. (*D*) Single CD spectra of GCN4-p1 N16V, M1-VxxN, M1-VxxNT, M1-IxxN, and M1-IxxNT at pH 9. Measurements were performed in 25 mM CHES pH 9, 25 mM NaCl at 10 °C, and a peptide concentration of 100 $\mu$M. (*E*) The thermal stability of N@d-containing peptides is pH dependent. Melting temperatures were determined by measuring $[\Theta]_{222\ nm}$ of M1-VxxN, M1-VxxNT, M1-IxxN, and M1-IxxNT at different pH. Thermal denaturation was followed at a peptide concentration of 100 $\mu$M in 25 mM buffer (sodium acetate pH 5, sodium phosphate pH 6–8, CHES pH 9–10, and CAPS pH 11) containing 25 mM NaCl. No $T_m$-values could be evaluated for M1-VxxNT at pH 5 and M1-IxxNT at pH 5 and 6, respectively. (*F*) The stability of N@*d*-containing peptides is dependent on ionic strength. Molar ellipticities at 222 nm of 100 $\mu$M M1-VxxN, M1-VxxNT, and GCN4-p1 N16V were determined as a function of ionic strength at 10 °C in 25 mM CHES, pH 9 containing 0–2 M NaCl.

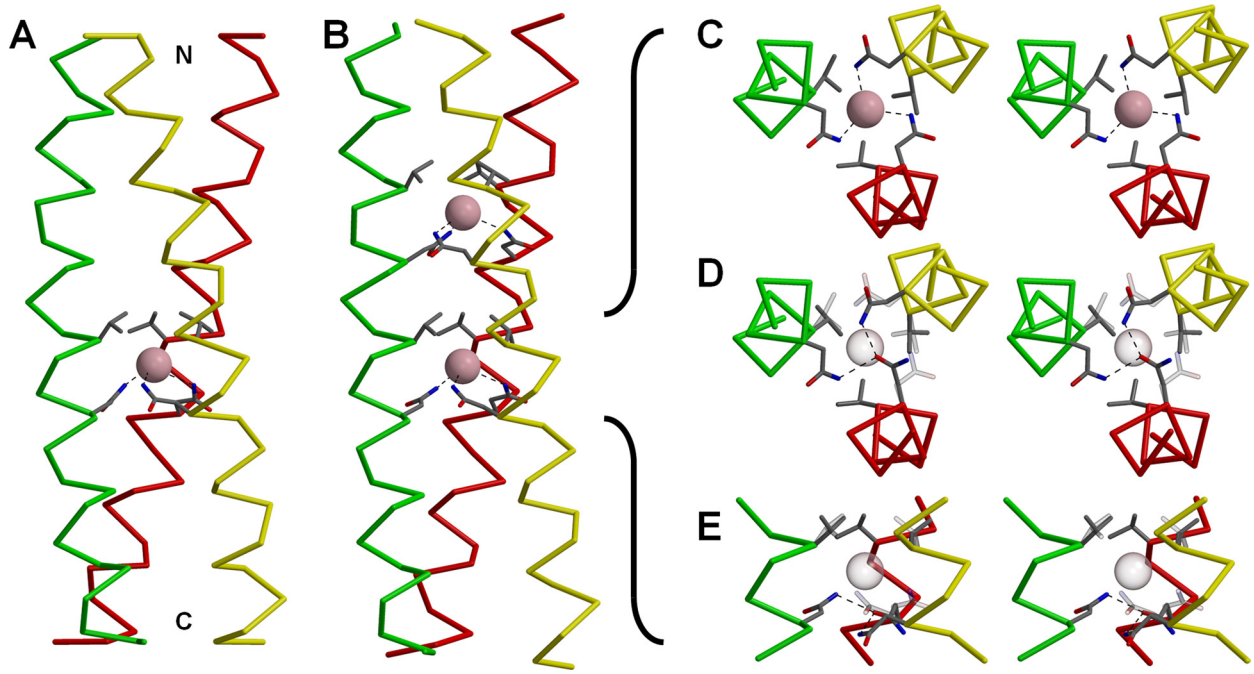


**Fig. S5.** Structure of the constructs M1-VxxNxxx (*A*) containing 1, and M2-VxxNxxx (*B*) containing 2 chloride-coordinating N@*d* layers. (*C*) Stereoview upstream the helical bundle, showing the chloride-bound state of V16 and N19 in M2-VxxNxxx. (*D*) Same view as *C*, showing the conformation in absence of chloride. The chloride-bound state is superposed transparently. (*E*) Side view of *D*.

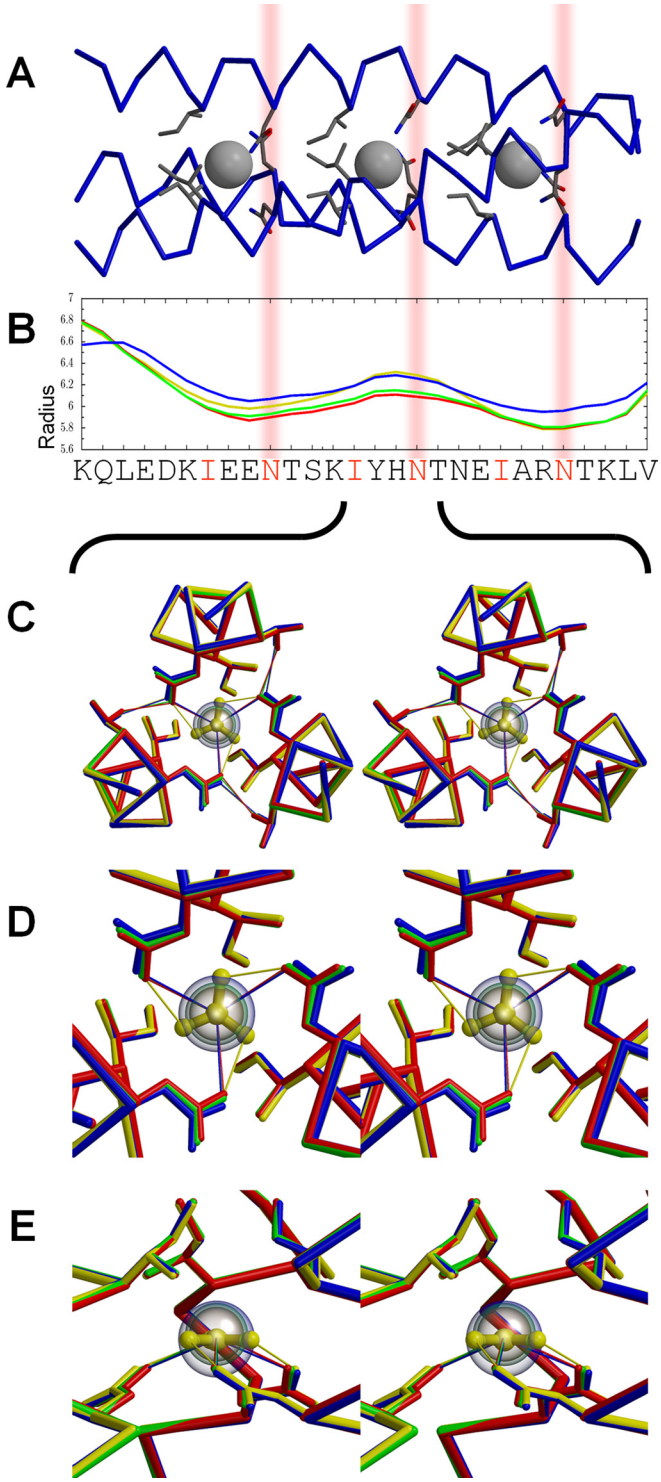


**Fig. S6.** Structures of the M3-IxxNTxx peptides coordinating chloride (red), bromide (green), iodide (blue), or nitrate (yellow). Exemplarily, the M3-IxxNTxx-iodide structure (*A*) is shown together with a diagram showing the coiled-coil radius of the 4 structures in the respective colors (*B*). For the 3 halogens, the radius of the structures is increasing systematically in all 3 N@*d* layers with an increasing radius of the anion. For M3-IxxNT-nitrate, the radius at the central N@*d* layer is even higher than for iodide. (Note that nitrate is only bound in the central layer; see text.) However, the change in radius is hardly visible in the structural superposition in *C*. As can be seen in the close-up in (*D*, *Top* view) and (*E*, *Side* view), the increasing anion radius is also compensated for by retreating N@*d* side chains and a shift of the anion further downstream, away from the isoleucines in positions *a*. The halogens are depicted transparently with their radii in scale to each other. Their coordination distances are ≈3.2 Å for chloride, ≈3.4 Å for bromide, and ≈3.6 Å for iodide. The coordination distances of the single nitrate are ≈3 Å. Despite these variations, the amino groups of the core asparagines form hydrogen bonds to the hydroxyl groups of threonines in position *e* of neighboring chains in all structures, with the same geometry as described for SadAK3.

**Table S1. N@*d* layers in coiled coils of known structure identified using CC+**

| | Sequence | Identical to | Brief PDB info | Not depicted because. . . | Figure |
|---|---|---|---|---|---|
| **DIMERS** | | | | | |
| **PDB** | | | | | |
| 1deb-26–34 | ELED N SNHL | | Human APC protein @ 2.4 Å | | S2*A*; brown |
| 1ik9–133–141 | TIAE N QAKN | | Human DNA repair protein XRCC4 @ 2.3 Å | | S2*A*; cyan |
| 1pl5–1305–1314 | EVMR N EIRI | 1nyh | *Saccharomyces cerevisiae* SIR4P C-terminal coiled coil @ 2.5 Å | | S2*A*; green |
| 1tpz-36–44 | QEIL N LIEL | 1tq2, 1tqd | Interferon-inducible GTPase @ 2 Å | | S2*A*; yellow |
| 2d4u-64–72 | RNTL N RAGI | | Bacterial serine chemoreceptor TSR @ 1.95 Å | | S2*A*; black |
| 1jyo-67–75 | IMVI N GELA | | *Salmonella* virulence effector SPTP @ 1.9 Å | Asn in divergent part of the coiled coil | |
| 2jk1–115–123 | SSAR N AARM | 2vuh, 2vui | *Rhodobacter capsulatus* HUPR1 @ 2.1 Å | Asn in divergent part of the coiled coil | |
| 2 hy6–13–21 | LASA N YHLA | | Seven-helix GCN4 mutant @ 1.25 Å | Staggered heptameric GCN4 mutant | |
| 2ipz-13–21 | LLSK N YHLV | | General control protein GCN4 @ 1.35 Å | Staggered tetrameric GCN4 mutant | |
| 1r5k-515–523 | RHMS N KGME | 3ert | Human estrogen receptor @ 2.7 Å | Right-handed cc (pentadecad), N in position *h* | |
| **TRIMERS** | | | | | |
| **This study** | | | | | |
| 3IxxNT-CL-15–23 | KIYH N TNEI | | Engineered GCN4-p1 mutant | | 3, S2*B*; red |
| SadAK3–484–492 | KVDQ N TADI | | *Salmonella* SadA | | 3, S2*B*; magenta |
| SadAK3–505–513 | DIAT N TTNI | | | | 3, S2*B*; bisque |
| **PDB** | | | | | |
| 2qih-541–549 | GIAE N KKDA | | *Moraxella catarrhalis* USPA1 protein @ 1.9 Å | | 3, S2*B*; orange |
| 2qih-555–563 | QANE N KDGI | | | | 3, S2*B*; white |
| 2qih-562–570 | GIAK N QADI | | | | 3, S2*B*; yellow |
| 2qih-590–598 | AVGN N TQGV | | | | 3, S2*B*; light blue |
| 2qih-597–605 | GAVT N KADI | | | | 3, S2*B*; pink |
| 2qih-604–612 | DIAK N QADI | | | Redundant with 2qih-562–570 | |
| 2qih-611–619 | DIAN N IKNI | | | | 3, S2*B*; brown |
| 2qih-639–647 | VSAA N TDRI | | | | 3, S2*B*; purple |
| 2qih-646–654 | RIAK N KAEA | | | N located in the *de* layer of a hendecad | |
| 3efg-43–51 | IGAR N AELI | | SLYX protein from *Xanthomonas campestris* @ 2 Å | | S2*D*; re-refined |
| 1aa0–421–429 | SIKA N ETNI | | Bacteriophage T4 fibritin deletion mutant E @ 2.2 Å | | 3, S2*B*; blue |
| 2poh-90–98 | RITA N TKAI | | Phage P22 tail needle GP26 @ 2.1 Å | | 3, S2*B*; dark gray, re-refined |
| 1 svf-129–137 | KANE N AAAI | | Simian virus 5 strain W3 fusion glycoprotein @ 1.4 Å | | 3, S2*B*; green |
| 2fyz-129–137 | QAQT N ARAI | | Mumps virus fusion protein core @ 2.20 Å | | 3, S2*B*; olive, re-refined |
| 2vrs-120–128 | TVDG N STAI | 2jjl | Avian reovirus Sigma C @ 1.75 Å | | 3, S2*B*; light gray |
| 2vrs-134–142 | DISS N GLAI | 2jjl | | | 3, S2*B*; black |
| 1y4m-67–75 | VVLQ N RRGL | 1mof | HERV-FRD envelope protein syncitin-2 @ 1.6 Å | | S2*E*; re-refined |
| 3duz-315–323 | LLKM N IELM | | Baculovirus fusion protein GP64 @ 2.95 Å | N located in the *de* layer of a hendecad | |

| | Sequence | Identical to | Brief PDB info | Not depicted because. . . | Figure |
|---|---|---|---|---|---|
| 1wdg-1001–1009 | VVNA N AEAL | 1wdf | Murine hepatitis virus spike protein @ 2.06 Å | N located in the *de* layer of a hendecad | |
| 2bez-933–941 | VVNQ N AQAL | 1wnc,1wyy, 1zv8,2beq | SARS core fusion protein @ 1.60 Å | N located in the *de* layer of a hendecad | |
| 2bez-947–955 | QLSS N FGAI | 1wyy | | N located in the *de* layer of a hendecad | |
| 1mg1–394–402 | AIVK N HKNL | | HTLV-1 GP21 ectodomain/maltose-binding protein @ 2.50 Å | Poor side-chain geometry, X-ray data not deposited | |
| 1mg1–408–416 | YAAQ N RRGL | | | Poor side-chain geometry, X-ray data not deposited | |
| 1ebo-81–89 | FSIL N RKAI | 2ebo, 3csy | Ebola virus envelope protein chimera @ 3 Å | Low resolution | |
| 2ebo-582–590 | FSIL N RKAI | 1ebo, 3csy | Core structure of GP2 from Ebola virus @ 1.90 Å | Poor side-chain geometry, X-ray data not deposited | |
| 1 g5 g-207–215 | GVEL N LYLT | | Newcastle disease virus fusion protein @ 3.30 Å | Low resolution | |
| **TETRAMERS PDB** | | | | | |
| 1ezj-67–75 | KVDE N KQLL | | Phosphoprotein from Sendai virus @ 1.9 Å | | S2*C* |

After inspection, some matches were excluded as listed. The remaining structures are shown in Fig. 3 and Fig. S2 in the indicated color. PDB entries 3efg, 2poh, 2fyz, and 1y4m were re-refined using the deposited experimental data before inclusion in the figure.

**Table S2. Crystallization conditions and crystal treatment**

| Structure | Protein solution | Reservoir solution (RS) | Soaking |
|---|---|---|---|
| SadAK3 | 5 mg/mL protein, 150 mM NaCL, 20 mM MOPS pH 7.2 | 20% (wt/vol) PEG 3350, 0.2 M $KNO_3$ | - |
| SadAK3b-V1 and SadAK3b-V2 | 3.7 mg/mL protein, 5% (v/v) glycerol, 40 mM NaCl, 20 mM Tris/HCl pH 7.5 | 10% (wt/vol) PEG 4000, 10% (v/v) isopopanol, 3% (wt/vol) 1,5-diaminopentane dihydrochloride, 100 mM Na-citrate pH 5.6 | - |
| M3-IxxNT-bromide | 10 mg/mL peptide, 0.05% azide, 50 mM NaCl, 20 mM MOPS pH 7.7 | 20% (wt/vol) PEG 3350, 200 mM NaBr, 100 mM Bis-Tris Propane pH 8.5 | - |
| M3-IxxNT-chloride | ″ | 2.4 M $(NH_4)_2HPO_4$, 100 mM Tris pH 8.5 | - |
| M3-IxxNT-iodide | ″ | ″ | 45 min in RS + 50 mM NaI + 5 mM NaCl |
| M3-IxxNT-nitrate | ″ | 25.5% (wt/vol) PEG 4000, 15% (v/v) glycerol, 170 mM Na-acetate, 90 mM Tris pH 8.5 | 180 min in RS + 50 mM $KNO_3$ +5 mM NaCl |
| M1-VxxNx | ″ | 3.2 M NaCl, 100 mM Na-acetate, pH 4.6 | - |
| M2-VxxNx | ″ | 30% PEG 4000, 200 mM Na-acetate, 0.1 M Tris, pH 8.5 | - |

**Table S3. Data collection and refinement statistics**

| Structure | SadAK3 | SadAK3b-V1 | SadAK3b-V2 | M1-VxxNx | M2-VxxNx | M3-IxxNT-chloride | M3-IxxNT-bromide | M3-IxxNT-iodide | M3-IxxNT-nitrate |
|---|---|---|---|---|---|---|---|---|---|
| PDB code | 2WPQ | 2WPR | 2WPS | 2WPY | 2WPZ | 2WQ0 | 2WQ1 | 2WQ2 | 2WQ3 |
| Monomers/AU | 3 | 3 | 3 | 1 | 3 | 1 | 1 | 1 | 1 |
| Space group | $P2_1$ | $P2_12_12_1$ | $P2_12_12_1$ | P321 | P1 | $I2_13$ | $I2_13$ | $I2_13$ | $I2_13$ |
| a (Å) | 26.0 | 36.1 | 36.7 | 33.3 | 25.4 | 56.4 | 56.6 | 56.3 | 56.7 |
| b (Å) | 37.0 | 62.5 | 55.9 | 33.3 | 25.4 | 56.4 | 56.6 | 56.3 | 56.7 |
| c (Å) | 178.4 | 172.6 | 169.0 | 50.8 | 33.7 | 56.4 | 56.6 | 56.3 | 56.7 |
| $\alpha$ (°) | 90 | 90 | 90 | 90 | 84.0 | 90 | 90 | 90 | 90 |
| b (°) | 92.7 | 90 | 90 | 90 | 84.9 | 90 | 90 | 90 | 90 |
| $\gamma$ (°) | 90 | 90 | 90 | 120 | 78.5 | 90 | 90 | 90 | 90 |
| Resolution (Å) | 1.85 | 2.65 | 2.60 | 1.75 | 1.25 | 1.12 | 1.08 | 1.35 | 1.22 |
| Completeness (%) | 98.3 | 87.8 | 98.7 | 99.5 | 91.1 | 99.3 | 99.5 | 99.4 | 99.4 |
| Redundancy | 3.97 | 3.29 | 3.47 | 10.0 | 2.00 | 4.74 | 4.11 | 2.83 | 3.57 |
| $I/\sigma$ (I) | 10.1 | 8.0 | 7.61 | 19.2 | 10.8 | 18.8 | 10.02 | 16.3 | 17.6 |
| $R_{merge}$ (%) | 7.7 | 12.8 | 12.7 | 6.3 | 3.6 | 3.7 | 7.9 | 3.3 | 3.8 |
| $R_{cryst}$ (%) | 22.1 | 26.7 | 23.8 | 21.4 | 17.8 | 14.8 | 12.7 | 14.2 | 14.1 |
| $R_{free}$ (%) | 28.8 | 32.7 | 32.4 | 25.2 | 22.3 | 19.7 | 16.7 | 18.4 | 20.0 |