Supplementary information, Data S1

Experimental procedures

Protein expression and purification

The plasmids containing the pGEX-6p-1 (TALEs, including AvrBs4, dTale1 (designer TALE protein 1), dTale2) were constructed through standard PCR (polymerase chain reaction) and verified by sequencing. The dTale1 and dTale2 are engineered TALE proteins based on the sequence of AvrBs4. The TALE proteins were expressed in Escherichia coli BL21 (DE3) cells by induction with 1mM IPTG for 12 hours at 15 ℃.

Cells were harvested by centrifugation at 4000 rpm for 20 min and were thawed and resuspended in buffer A (25 mM Tris–HCl pH 8.0, 500 mM NaCl, 3 mM DTT) containing protease-inhibitor PMSF. Then the suspension was subjected to lysis by sonication and cell debris was removed by centrifugation at 18000 rpm for 50 min at 4 ℃. The supernatant was subsequently loaded onto a 5 ml Glutathione Sepharose 4B (GS4B) beads (Qiagen). Bound proteins were washed and cleaved by precision protease overnight at 4 ℃.

The proteins were further purified by anion-exchange chromatography using Heparin column (GE Healthcare) with buffer B (25 mM Tris–HCl pH 8.0, 3 mM DTT) and buffer C (25 mM Tris–HCl pH 8.0, 2 M NaCl, 3 mM DTT). The proteins in the peak fraction were concentrated and then purified by size exclusion chromatography using Superdex 200 gel-filtration (GE Healthcare) with buffer D (10 mM Tris–HCl

pH 8.0, 100 mM NaCl, 3 mM DTT, 5 mM $MgCl_2$, 5% v/v glycerol). The purity of the dTale2 proteins was monitored at all stages of the purification process using SDS–PAGE (polyacrylamide gel electrophoresis). The similar method was used to purify the different constructs and various mutants of dTale2 protein.

Limited proteolysis

The limited protease digestion reactions were set up with 5 µg of the dTale2 protein and gradient concentrations of trypsin (0, 0.1, 0.05, 0.025, 0.0125, 0.00625, 0.003125, 0.0015625, 0.00078125, 0.000390625 µg) in a 10 µl volume in buffer D (10 mM Tris–HCl pH 8.0, 100 mM NaCl, 3 mM DTT, 5 mM $MgCl_2$, 5% v/v glycerol) at 4 ℃ for 30 min. The digestion reactions were stopped with the addition of SDS loading buffer and the products were separated on 15% SDS - PAGE and stained with Coomassie brilliant blue (CBB).

Crystallization, data collection and structure determination

The TALE proteins were directly used for crystallization trials with Hampton Research Crystallization Kits. Hanging-drop vapor-diffusion crystallization method was carried out by mixing the samples in a 1:1 ratio with reservoir solution over 0.4 ml mother liquor and incubating at 18 ℃. The dTale2 (148-610) alone gave rise to high quality crystals in three days in the reservoir solution (0.05 M $MgSO_4$, 0.05 M HEPES pH 7.0, 1.6 M Li_2SO_4).

Prior to data collection, the crystals were transferred into cryo-protectant buffer (the mother buffer containing 20% (w/v) glycerol) and flash-cooled in liquid nitrogen. Diffraction data were collected at 100 K using the synchrotron-radiation sources beamline at BL17U at Shanghai Synchrotron Radiation Facility (SSRF, China). Indexing, scaling and merging of the data were performed using HKL2000 [1].

The structure of dTale2 (148-610) was determined by Molecular Replacement method using the program PHASER [2] in the CCP4 suit [3] with the dHax3 as the searching models (3V6P and 3V6T in PDB). The initial atomic model built by Phenix [4] was extended and completed manually in Coot [5]. Refinement was carried out using Phenix and REFMAC [6], together with appropriate manual model rebuilding in Coot. The structure was refined to final R factor and R_{free} values of 20.60 % and 24.28 %, respectively.

DNA annealing and gel shift assay

The oligonucleotides containing dTale2 binding site (TATCAATCGT) were purchased from BGI (The Beijing Genomics Institute, China) (supplementary information, Table S3). The dsDNA were formed in the buffer D (10 mM Tris–HCl pH 8.0, 100 mM NaCl, 3 mM DTT, 5 mM $MgCl_2$, 5% v/v glycerol) via annealing procedure as follows: 95 ℃, 5 min; 50 ℃, 60 min. The gel shift assay was performed in the buffer D with the ratio of protein and dsDNA at 1:1.2. The reaction was incubated for 30 min at 4 ℃ and the samples were then loaded onto an 8% native denaturing polyacrylamide gel in 1× TG buffer.

Electrophoresis was carried out at 150 V for 45 min at 4 ℃. The native gel was stained with ethidium bromide (EB) and then coomassie brilliant blue (CBB).

Isothermal titration calorimetry (ITC)

DNA oligonucleotides for ITC were dissolved in the HEPES buffer containing 25 mM HEPES pH7.5, 100 mM NaCl. Equal amounts of sense and antisense strands were then annealed by heating to 95 °C for 5min and slowly cooling to 16 ℃ over several hours, followed by incubation on ice. The dTale2 proteins were purified as usual except that the buffer for Superdex 200 was replaced by the HEPES buffer containing 25mM HEPES pH7.5, 100mM NaCl. The concentrations of the proteins and dsDNA were 0.015-0.03 mM, 0.4-0.6 mM, respectively. ITC experiments were carried out on a MicroCal iTC_{200} calorimeter (GE Healthcare) with the following settings: total injections, 20-25; cell temperature, 25 ℃; reference power, 5 μcal/sec; Initial delay, 60sec.; stiring speed, 800rpm; injection volume, 1.5 μl; spacing time, 150s. The titration data were processed and fitted using Origin 7.0 software.

References

1 Otwinowski Z., Minor W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 1997; **276**:307-326.

2 McCoy AJ. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr D Biol Crystallogr* 2007; **63**:32-41.

3 Winn MD, Ballard CC, Cowtan KD, *et al*. Overview of the CCP4 suite and current

developments. *Acta Crystallogr D Biol Crystallogr* 2011; **67**:235-242.

4 Adams PD, Afonine PV, Bunkoczi G, *et al*. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 2010; **66**:213-221.

5 Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 2010; **66**:486-501.

6 Winn MD, Murshudov GN, Papiz MZ. Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol* 2003; **374**:300-321.