SUPPLEMENTAL DATA

Experimental procedures

Cloning, protein production and purification. The DNA sequences corresponding to residues R10-S250 (DBD) and E258-T351 (IPT/TIG) in human EBF1 (gi:31415878), and E251-H386 (TIG-HLH) and S250- V406 (TIG-HLHLH) in human EBF3 (gi:53828926) were sub-cloned into plasmid pNIC28-Bsa4 (SGC Oxford, GI:124015065) and expressed in E. coli BL21(DE3) R3 pRARE cells, where R3 denotes a derivative of BL21(DE3) resistant to a strain of T1 bacteriophage (SGC Oxford) and the pRARE plasmid originates from the Rosetta strain (Novagen).

Bacterial cells were cultured at 37 °C in Terrific Broth (TB) supplemented with 8 g/l glycerol, 50 µg/ml kanamycin, 34 µg/ml chloramphenicol, and approximately 5-10 drops of Antifoam 204 (Sigma-Aldrich). The EBF1 DNA binding domain was also produced labelled with seleno-methionine using a minimal medium to which an amino acid mix including L(+)-selenomethionine was added (1). In all cases, expression was induced by addition of 0.5 mM IPTG and left overnight at 18°C. Cells were harvested by centrifugation and resuspended in lysis buffer (100 mM HEPES, 500 mM NaCl, 10% glycerol, 10 mM imidazole, 0.5 mM TCEP, pH 8.0) supplemented with one tablet of Complete EDTA-free protease inhibitor (Roche Applied Science) and 2000 U Benzonase (Merck). Lysis buffer was used in the ration of 1.5 ml buffer per 1 gram of wet cell pellet. Re-suspended pellets were stored at -80 °C.

Cells were disrupted by sonication. Purification of the His-tagged proteins was performed in two steps using Ni-charged HiTrap Chelating HP and HiLoad 16/60 Superdex 75 Prep Grade columns (GE Healthcare) pre-equilibrated with IMAC buffer (20 mM HEPES, 500 mM NaCl, 10% glycerol, 10 mM imidazole, 0.5 mM TCEP, pH 7.5) and gel filtration buffer (20 mM HEPES, 300 mM NaCl, 10% glycerol, 0.5 mM TCEP, pH 7.5), respectively. The filtered lysates were loaded onto the HiTrap Chelating columns and washed with IMAC buffer containing 25 mM imidazole. Bound protein was eluted with IMAC buffer containing 500 mM imidazole and automatically loaded onto the gel filtration column. In the purification of the EBF1 DNA binding domain the NaCl content of the gel filtration buffer was raised to 1 M in order to release bound E. coli nucleic acids. Fractions containing the target proteins were pooled and concentrated using a VIVASPIN 2 (Sartorius Stedim) centrifugal filter device with a cutoff size of 10 kDa.

Sodium chloride concentrations of DBD samples were were decreased to 300 mM to remove N-terminal His-tags. The tags were proteolytically removed by incubation of the proteins with His-tagged TEV protease at a molar ratio of 50:1 over night. The target proteins were subsequently purified by passing the reaction mixtures over a HisTrap HP column (GE Healthcare). The cleaved proteins were eluted with IMAC buffer. The proteins were concentrated and the buffer was changed to gel filtration buffer with 2 mM TCEP. Protein identities were confirmed by mass spectrometry and samples were stored at -80 °C.

The DBD was complexed with the DNA duplex: 5' GAGAGAGAGACTCAAGGGAATTGTGGCC and 5'GGCCACAATTCCCTTGAGTCTCTCTCTC. Protein and double stranded DNA

oligomer were mixed at a ratio of 2.1: 1 and subjected to size exclusion chromatography using a HiLoad 16/60 Superdex 200 Prep Grade column (GE Healthcare) in gel filtration buffer containing 2 mM TCEP. The peak at a retention volume corresponding to a 2:1 complex was collected and the complex was concentrated to an approximate concentration of 0.5 mM.

Crystallization, data collection, structure determination and validation. Crystals of the DBD, IPT/TIG and TIG-HLH were obtained by the sitting drop vapour diffusion method. The crystallization conditions were; DBD: 0.2 µl complex (0.5 mM) was mixed with 0.1 µl of well solution consisting of 0.1 M MES pH 5.8, 2.1 M ammonium sulphate, IPT/TIG: 0.2 µl protein solution (6.9 mg/ml) was mixed with 0.1 µl of well solution consisting of 0.1 M Tris pH 9, 0.3 M trimethylamine n-oxide and 23% PEG monomethyl ether 2000, TIG-HLH: 0.2 µl protein solution (10.2 mg/ml) was mixed with 0.1 µl of well solution consisting of 0.1 M Bis-Tris propane pH 6.9, 0.2 M sodium formate and 21% PEG 3350. Crystals of TIG-HLHLH were obtained by hanging drop vapour diffusion by mixing 2 µl of protein solution (10.7 mg/ml) with 1 µl of well solution consisting of 0.1 M Bis-Tris propane pH 7.5, and 2.9 M NaCl. All crystallisation plates were incubated at 4 ºC. IPT/TIG crystals were soaked 48 hours in ethyl mercury thiosalicylate (EMTS) before freezing. All crystals were flash from in liquid nitrogen after quick transfer to cryogenic solution containing; DBD: well solution and 20% glycerol, IPT/TIG: well solution, 1 mM EMTS and 20% glycerol, TIG-HLH: 20% Ethylene glycol, TIG-HLHLH: 20% ethylene glycol.

EBF1 DBD data was collected for single wavelength SAD to 2.5 Å at DIAMOND (I04), Oxfordshire, England. Data was integrated with XDS (2) and scaled with SCALA, from the CCP4 suite(3). Phases were calculated with SHELX (4). Buccaneer was used to build the initial model, which was then improved several rounds of refinement using Refmac5 and manual editing in COOT. The asymmetric unit consisted of a single monomer. IPT/TIG data was collected for single wavelength SAD to 2.3 Å at BESSY (BL14-1), Berlin, Germany. Data was integrated with XDS and scaled with XSCALE. Phases were calculated with SOLVE (5). RESOLVE (5) was used to build the initial model, which was then improved by manual editing in COOT and refinement in Refmac5. The asymmetric unit consisted of three molecules. Data for both TIG-HLH and TIG-HLHLH were collected at a single wavelength to 1.9 Å and 3.1 Å, respectively, at BESSY (BL14-1), Berlin, Germany. Data were integrated with XDS and scaled with XSCALE. Both structures were solved using IPT/TIG domain as a model for molecular replacement using PHASER (6). Models were improved by successive rounds of manual model building in COOT and refinement using PHENIX(7). All data collection and refinement statistics are presented in Table 1. All figures were generated using ccp4mg (8).

Docking. Docking was performed using the program HADDOCK (9-10) with default settings of the program using the DBD structure 3LYR and a DNA dublex with sequence 5'-AATTCCCAAGGAATT-3'. To enforce dimerization of the protein C2 and non-crystallographic symmetries were applied with default force constants. Initial rigid body docking was performed from 1000 starting configurations with the two proteins randomly

oriented. The 200 docking solution with the lowest overall energy (score) were promoted for further development, including semi-flexible refinement and fully flexible refinement with added water molecules. Cluster analysis normally performed in this stage of studies did not result in reasonably defined clusters, most likely due to the lack of sufficient number of structures. Therefore we decided to concentrate on the best solution according to the HADDOCK score.

Table S1: Data collection and refinement statistics for EBF1 (DBD and IPT/TIG) and EBF3 (TIG + HLH and TIG + HLHLH). High resolution shell values are indicated in brackets

Protein	EBF1 DBD	EBF1 TIG	EBF3 TIG+HLH	EBF3 TIG+HLHLH
PDB entry	3LYR	3MQI	3MUJ	3N50
Beamline	DIAMOND I04	BESSY BL14.1	BESSY BL14.1	BESSY BL14.1
Wavelength (Å)	0.9789	1.00764	0.91841	0.91841
Space group	P 65 2 2	C 1 2 1	P 65	I23
Cell dimensions				
a, *b*, *c* (Å)	134.22, 134.22, 72.39	86.29, 57.07, 69.14	134.68,134.68, 41.15	229.64, 229.64, 229.64
α, *β*, *γ* (°)	90, 90, 120	90, 93.17, 90	90, 90, 120	90, 90, 90
Resolution (Å)	58.12-2.51 (2.64-2.51)	69.03-2.3 (2.36-2.3)	50.0-1.92 (1.96-1.92)	20.0-3.1 (3.18-3.1)
R_{merge} §	0.064 (0.317)	0.1 (0.55)	0.044 (0.37)	0.09 (0.66)
I / (σI)	39.3 (12.8)	16.78 (2.72)	30.82 (4.64)	20.1 (2.55)
Completeness (%)	99.5 (99.4)	97.5 (96.9)	97.0 (76.6)	99. (99.9)
Redundancy	27.8 (28.7)	7.6 (7.8)	10.9 (7.5)	5.0 (5.0)
Refinement				
Resolution (Å)	58.12-2.51 (2.57-2.51)	69.03-2.3 (2.36-2.3)	38.88-1.91 (1.96-1.91)	19.84-3.1 (3.18.3.1)
No. reflections	13359 (966)	13970 (1023)	14254(1887)	13731(2695)
R_{work} †/ R_{free} ‡	23.3/27.4 (35.4/39.7)	23.5/27.3 (30.3/31.1)	16.9/19.3 (20.57/25.93)	18.8/22.7 (28.1/28.3)
No. atoms				
Protein	1681	2055	2130	6892
Ligands	7	23	8	0
Water	25	76	340	0
B-factors (Å^2)				
Protein	46.01	28.77	26.16	94.55
Ligands	58.4	77.17	32.62	
Water	42.32	30.56	39.53	
R.m.s deviations				
Bond lengths (Å)	0.017	0.009	0.015	0.005
Bond angles (°)	1.473	1.151	1.476	0.837
Ramachandran plot (%)				
Favored regions	96.14	98.1	98.47	94.38
Additionally allowed regions	3.86	1.9	1.53	5.62

References:

1. Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L., and Clardy, J. (1993) *J Mol Biol* **229**, 105-124
2. Kabsch, W. (2010) *Acta Crystallogr D Biol Crystallogr* **66**, 125-132
3. COLLABORATIVE COMPUTATIONAL PROJECT NUMBER 4. (1994) *Acta Crystallogr D Biol Crystallogr* **D50**, 760-763
4. Schneider, T. R., and Sheldrick, G. M. (2002) *Acta Crystallogr D Biol Crystallogr* **58**, 1772-1779
5. Terwilliger, T. (2004) *J Synchrotron Radiat* **11**, 49-52
6. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) *J Appl Crystallogr* **40**, 658-674
7. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) *Acta Crystallogr D Biol Crystallogr* **66**, 213-221
8. Potterton, L., McNicholas, S., Krissinel, E., Gruber, J., Cowtan, K., Emsley, P., Murshudov, G. N., Cohen, S., Perrakis, A., and Noble, M. (2004) *Acta Crystallogr D Biol Crystallogr* **60**, 2288-2294
9. Dominguez, C., Boelens, R., and Bonvin, A. M. (2003) *J Am Chem Soc* **125**, 1731-1737
10. de Vries, S. J., van Dijk, A. D., Krzeminski, M., van Dijk, M., Thureau, A., Hsu, V., Wassenaar, T., and Bonvin, A. M. (2007) *Proteins* **69**, 726-733