# Supporting Information

## Ibrahim et al. 10.1073/pnas.0904309106

### *SI Text*

**Protein Expression and Purification.** The ORFs of tTFIIB (residues 1–345) and tTFIIB$_C$ (residues 87–345) were amplified from *T. brucei* Lister 427 genomic DNA, inserted into pDONR221 (Invitrogen) and then recombined into the pDEST-HisMBP destination vector (Addgene) to create pDEST-HisMBP-tTFIIB and pDEST-HisMBP-tTFIIB$_C$ expression vectors. The tTFIIB mutant variant expression vectors were generated by the Change-IT Mutagenesis Kit (USB) using the pDEST-HisMBP-tTFIIB vector as template. Expression vectors were confirmed by DNA sequencing and transformed into Rosetta2 *E. coli* cells (EMD Biosciences). Cultures (1 L) harboring each construct were grown in Terrific Broth with 50 $\mu$g/mL ampicillin and 35 $\mu$g/mL chloramphenicol at 37 °C to an $OD_{600}$ of 1.8, adjusted to 25 °C, induced with 0.5 mM isopropyl-1-thio-beta-D-galactopyranoside, and incubated for 6 h. Cells were harvested by centrifugation, resuspended in 50 mM sodium phosphate, pH 7.5, 500 mM NaCl, 10 mM imidazole, 1 mM DTT, and 5% glycerol, at 0.5 g wet cell weight/mL, lysed by sonication, and centrifuged. The supernatant was loaded onto Ni Sepharose Fast Flow His-Trap columns, and the fusion protein was eluted in an imidazole gradient. The $His_6$-MBP tag was removed by TEV protease cleavage overnight at 4 °C, leaving an additional glycine at the N terminus. tTFIIB and variants were separated from TEV protease and the $His_6$-MBP tag by gel-filtration on a Sephacryl S-100 column in 50 mM sodium phosphate, pH 7.5, 150 mM NaCl, 1 mM DTT, and 5% glycerol. After TEV cleavage, the tTFIIB$_C$ preparation was desalted into 25 mM sodium phosphate, pH 7.5, 50 mM NaCl, 1 mM DTT, and 5% glycerol, and separated from TEV protease and the $His_6$-MBP tag by anion exchange on an SP Sepharose column in a 50–500 mM NaCl gradient. Protein samples were concentrated to ~10 mg/mL (Vivaspin) and stored at −80 °C. SeMet incorporation was 4–5 of the 6 methionines, as judged by mass spectrometry. Concentrations of tTFIIB monomers ($\varepsilon_{280}$ = 12,950 $M^{-1}$ $cm^{-1}$) and tTFIIB$_C$ monomers ($\varepsilon_{280}$ = 9,970 $M^{-1}$ $cm^{-1}$) were determined by UV absorbance.

**Structure Determination of tTFIIB$_C$.** Crystals were cryo-protected in crystallization buffer plus 5% ethylene glycol for 1 min and flash-cooled in a nitrogen gas stream at −160 °C. X-ray diffraction data to 2.3 Å were collected from a native crystal on a 007-HF (Rigaku Americas) copper K-alpha source with an Raxis-IV++ imaging plate detector at the PHRI X-ray Crystallography Core Facility. A 3-wavelength MAD data set was collected at the National Synchrotron Light Source beamline X29 ( Brookhaven National Laboratory) from SeMet- tTFIIB$_C$ crystals, which initially diffracted to 3.0 Å but suffered radiation damage and diffracted to 3.2 Å by collection of data at the third wavelength. Data were reduced with HKL2000 (1) and CCP4 (2). Phases (3.0 Å) were calculated from the MAD data set in PHENIX (3) using 5 SeMet positions with 1 molecule in the asymmetric unit. Electron density maps from MAD phasing were continuous, allowing manual building of residues 94–261 and 275–313 in COOT (4). Using the native data set, this model was then subjected to refinement in REFMAC5 (5) and additional automated building in ARP/wARP (6), which was unable to add residues to the model but assisted in correct placement of side-chain positions. Water molecules and an ethylene glycol molecule were added using ARP/wARP (6) in the later stages of refinement and manual rebuilding. Residues 87–93, 262–274, and 314–345 could not be built owing to poor electron density. The final TLS-refined model has Ramachandran statistics of 100% in preferred or allowed regions (7). Secondary structure assignments were made by DSSP (8), and alignments were performed in DALI or by SSM in COOT (9, 10). Figs. 2*A*, 2*B*, 3, and S4 were generated with PyMOL (11). Data collection and refinement statistics are in Table S1 and Table S2.


1. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 276:307–326.
2. Potterton E, Briggs P, Turkenburg M, Dodson E (2003) A graphical user interface to the CCP4 program suite. *Acta Crystallogr D Biol Crystallogr* 59:1131–1137.
3. Adams PD, *et al.* (2002) PHENIX: Building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* 58:1948–1954.
4. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60:2126–2132.
5. Winn MD, Murshudov GN, Papiz MZ (2003) Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol* 374:300–321.
6. Langer G, Cohen SX, Lamzin VS, Perrakis A (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat Protoc* 3:1171–1179.
7. Murshudov GN, Vagin AA, Lebedev A, Wilson KS, Dodson EJ (1999) Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallogr D Biol Crystallogr* 55:247–255.
8. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22:2577–2637.
9. Holm L, Kaariainen S, Rosenstrom P, Schenkel A (2008) Searching protein structure databases with DaliLite v. 3. *Bioinformatics* 24:2780–2781.
10. Krissinel E, Henrick K (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* 60:2256–2268.
11. DeLano WL (2002) *The PyMOL User's Manual* (DeLano Scientific, San Carlos, CA).
12. Malik S, Lee DK, Roeder RG (1993) Potential RNA polymerase II-induced interactions of transcription factor TFIIB. *Mol Cell Biol* 13:6253–6259.
13. Pace CN, Scholtz JM (1997) in *Protein Structure: A Practical Approach*, ed. Creighton TE (IRL Press, New York, NY), vol 316, pp 299–322.

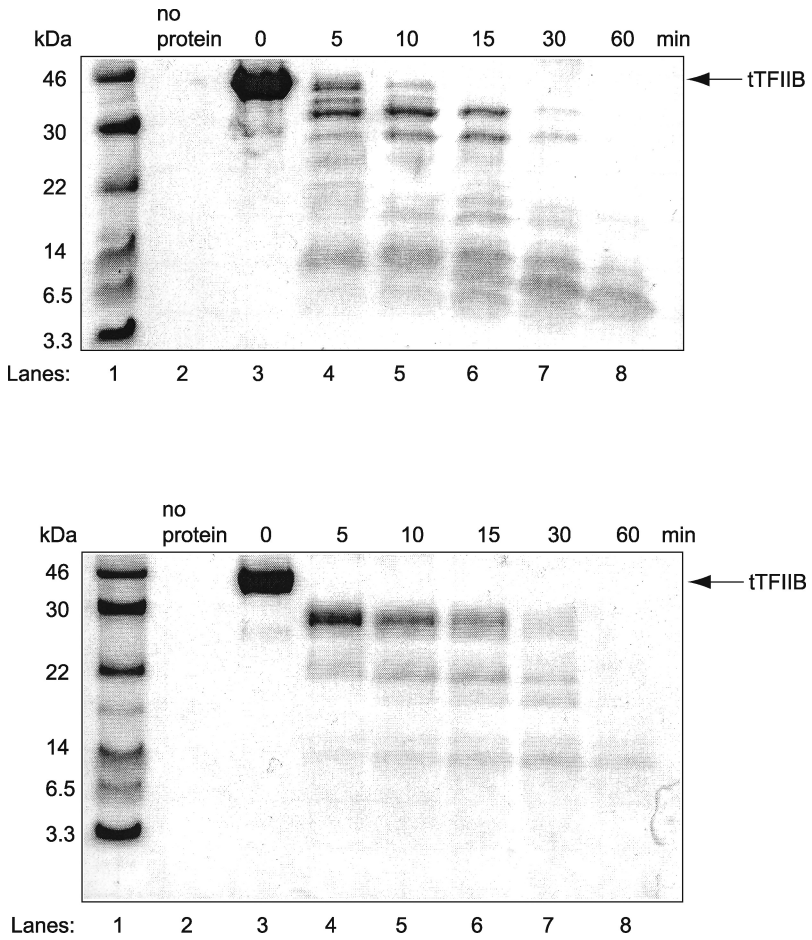


**Fig. S1.** Limited proteolysis of tTFIIB. tTFIIB was unstable in the presence of trypsin (*Top*) or subtilisin (*Bottom*) as detected by SDS/PAGE over 60 min (lanes 3–8, both panels). tTFIIB (20 $\mu$M) was completely degraded at 25 °C (enzyme:substrate molar ratio of 1:400) within 60 min by trypsin (*Top*, lane 8) and by subtilisin (*Bottom*, lane 8). In contrast, the C-terminal domain of human TFIIB was resistant to digestion by trypsin (enzyme:substrate molar ratio of 1:7.5) after 45 min (12). Molecular weight marker is in lane 1, and no protein was added in lane 2.

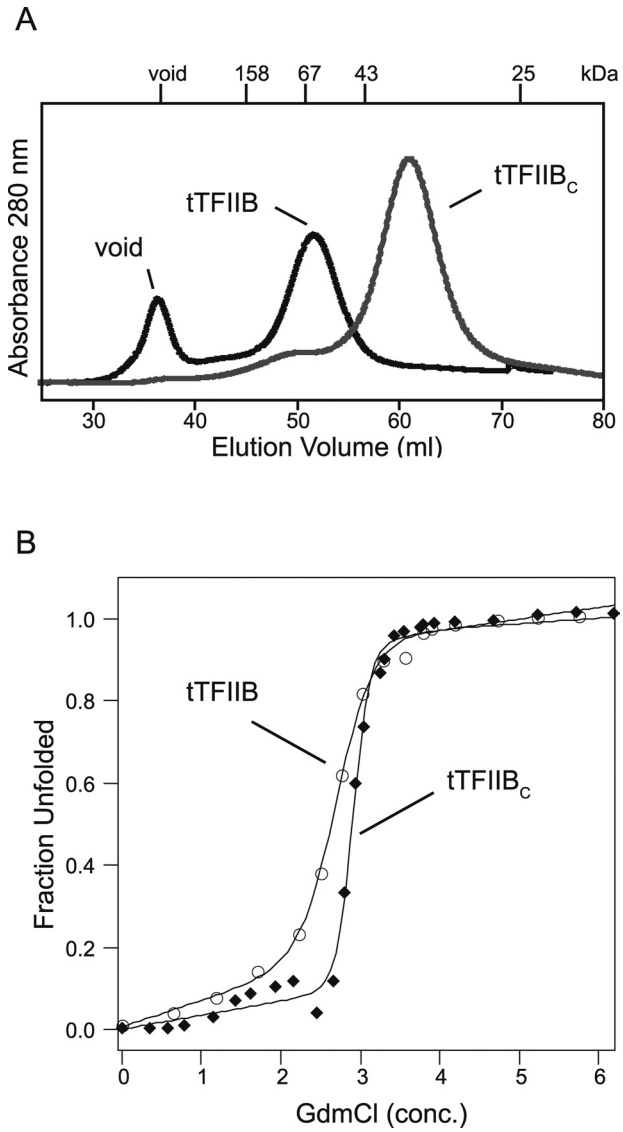


**Fig. S2.** Biophysical properties of tTFIIB and tTFIIB$_C$. (*A*) Sephacryl S-200 gel filtration of tTFIIB and tTFIIB$_C$ at 4 °C. tTFIIB eluted at a position corresponding to an apparent $M_r$ = 61.7 kDa and tTFIIB$_C$ eluted at apparent $M_r$ = 38.2 kDa. tTFIIB and tTFIIB$_C$ have monomer molecular masses of 37.7 and 28.9 kDa, respectively, as determined by MALDI-TOF mass spectrometry. (*B*) Guanidinium chloride stability (25 °C) of tTFIIB (open circles) and tTFIIB$_C$ (filled diamonds) at 3 $\mu$M assayed by fluorescence. The curves represent the least-squares fit of an equation to the data. This equation assumes a 2-state folded <-> unfolded equilibrium and linear pre- and posttransition baselines (13). This fit yields values of $\Delta G$ = 6.9 kcal/mol, $m$ = 2.5 kcal/mol·M for tTFIIB and $\Delta G$ = 16.4 kcal/mol, $m$ = 5.6 kcal/mol·M for tTFIIB$_C$.

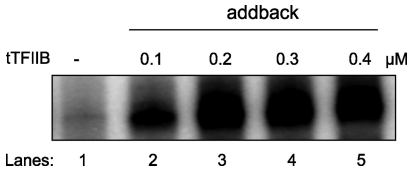


**Fig. S3.** Activity of tTFIIB in in vitro SL RNA transcription assays. Transcription activity in tTFIIB-depleted extracts (lane 1) was restored upon addback of between 0.1 to 0.4 $\mu$M (lanes 2–5). Maximal restoration occurred at 0.2 $\mu$M tTFIIB (lane 3).

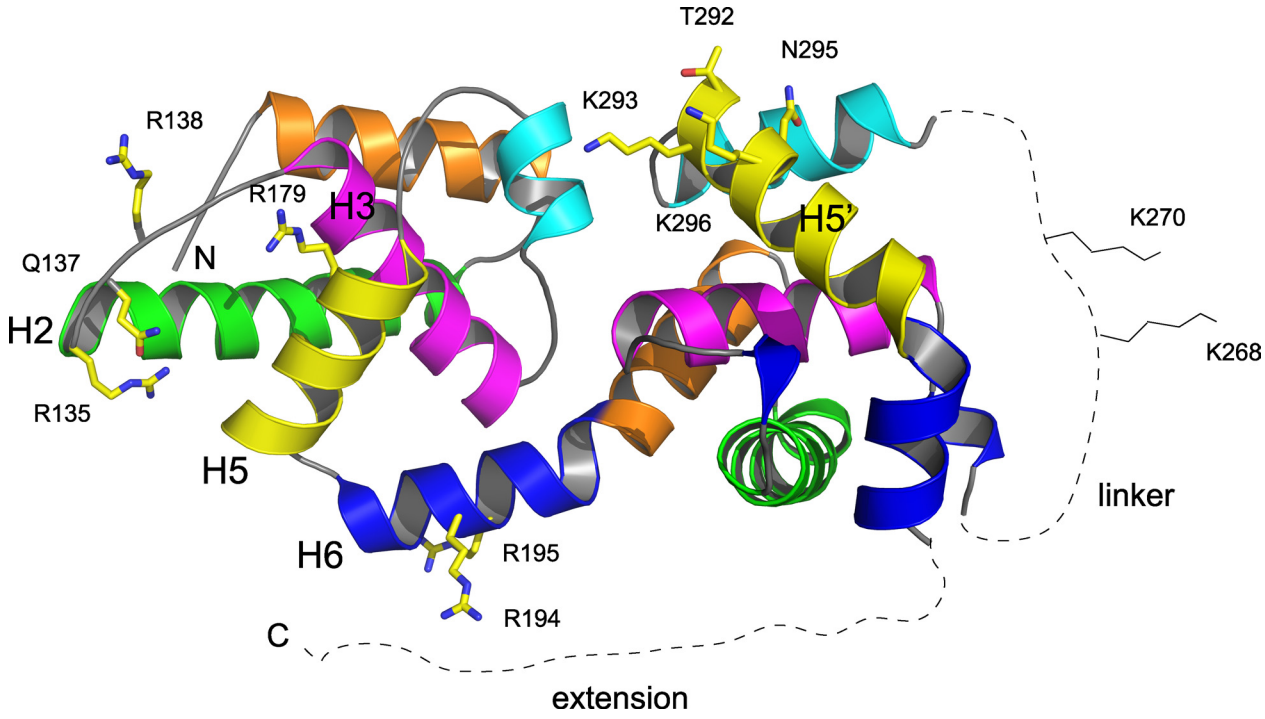


**Fig. S4.** Location of amino acids mutated in this study. Arg-135, Gln-137, and Arg-138 are in the loop between helices H2 and H3. Arg-179 is helix H5 and Arg-194 and Arg-195 are in helix H6. Lys-268 and Lys-270 are in the linker between H3′A and H4′. The residues Thr-292, Lys-293, Asn-295, and Arg-296 reside in helix H5′.

**Table S1. Data collection statistics**

| | | SeMet | | |
|---|---|---|---|---|
| | Native | λ1-peak | λ2-inflection | λ3-remote |
| Source | Cu$\alpha$ | NSLS X29 | NSLS X29 | NSLS X29 |
| Wavelength, Å | 1.5418 | 0.9791 | 0.9793 | 0.9700 |
| Resolution, Å | 2.3 (2.38–2.30) | 3.0 (3.11–3.0) | 3.0 (3.11–3.0) | 3.2 (3.31–3.20) |
| Unit cell $a = b$, $c$ | 109.23, 51.12 | 109.42, 51.34 | 109.45, 51.34 | 109.62, 51.41 |
| No. of reflections | 60,094 | 43,733 | 43,827 | 34,930 |
| Unique reflections | 13,573 | 6,621 | 6,633 | 5,490 |
| Completeness, % | 94.8 (94.3) | 99.6 (98.3) | 99.7 (98.6) | 99.3 (95.6) |
| Mean $I/\sigma I$ | 22.0 (4.0) | 14.5 (4.6) | 15.4 (4.4) | 13.6 (2.8) |
| $R_{merge}$,* % | 6.3 (41.2) | 11.2 (39.0) | 11.1 (38.8) | 12.1 (49.6) |

Values in parentheses correspond to the highest-resolution shell.
*$R_{merge} = \Sigma|I_{obs} - \langle I\rangle|/\Sigma\langle I\rangle$, calculated for all data.

**Table S2. Refinement statistics**

| | |
|---|---|
| Resolution range, Å | 7.99–2.30 (2.36–2.30) |
| No. of reflections | 11,861 |
| Completeness, % | 95.2 |
| Data cutoff, $F/\sigma F$ | none |
| $R_{work}$* | 0.204 (0.240) |
| $R_{free}$† | 0.249 (0.294) |
| rmsd, bond lengths, Å | 0.008 |
| rmsd, bond angles, ° | 1.07 |

Values in parentheses correspond to the highest-resolution shell.

*$R_{work} = \Sigma|F_{obs} - F_{calc}|/\Sigma|F_{obs}|$, where $F_{obs}$ and $F_{calc}$ are the observed and calculated structure factor amplitudes, respectively.

†$R_{free}$ is the same as the $R_{work}$ but is calculated from 10% of the data excluded from the refinement.