SUPPLEMENTARY INFORMATION

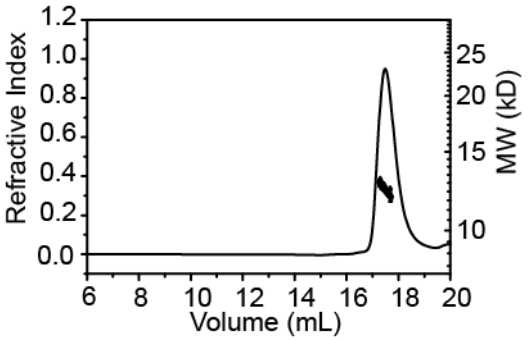

Figure S1. GATA1$_{NC}$ is monomeric in solution. SEC/MALLS analysis of GATA1$_{NC}$ (loading concentration 200 μM) showing protein concentration in refractive index units (solid line) and the calculated weight average molecular weight (solid symbols).

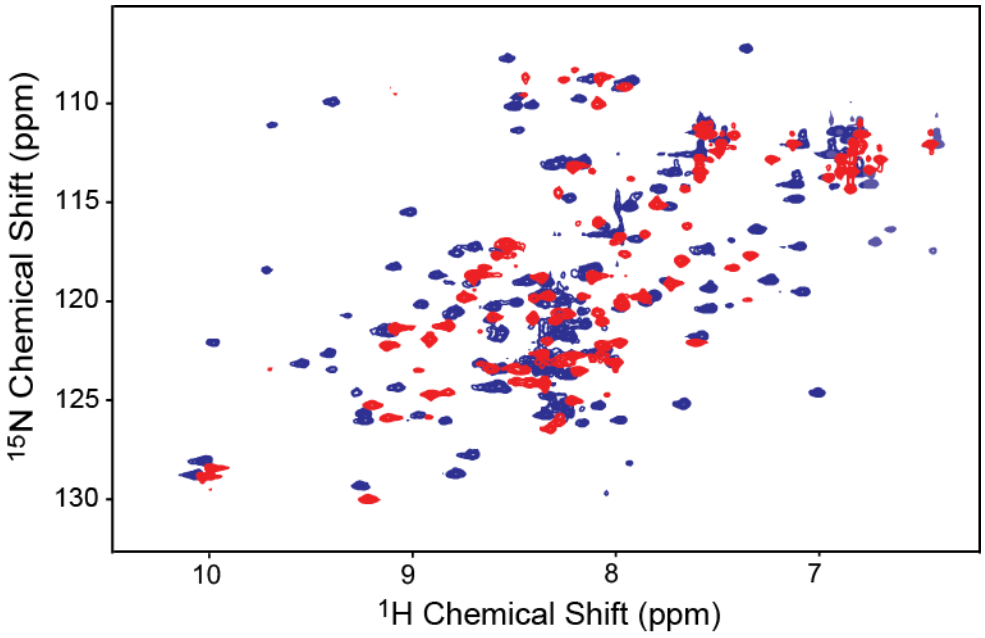

Figure S2: GATA1$_{NC}$ in complex with mPal DNA. Overlay of the 2D ^{1}H-^{15}N HSQC spectra of ^{15}N-GATA1$_{NC}$ alone (220 μM, red peaks) and in complex with 2.2 molar equivalents of mPal oligonucleotide (Blue peaks).

PROTEIN CRYSTALLOGRAPHY AND STRUCTURE COMPARISONS

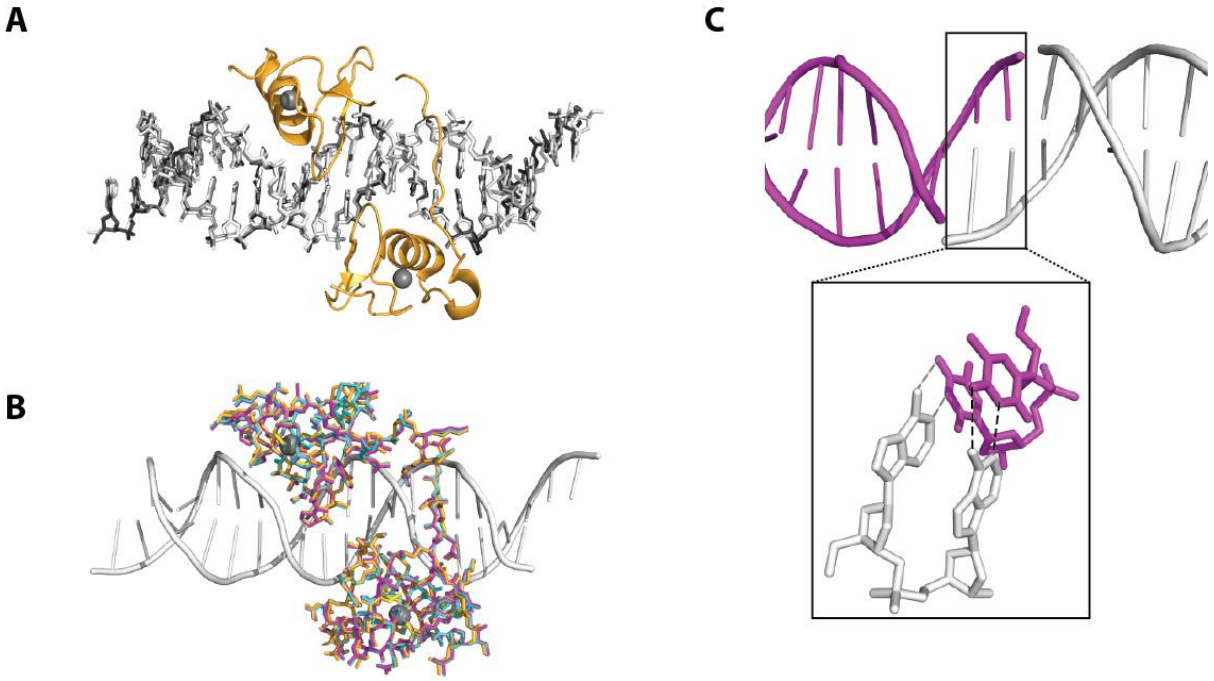

Figure S3. Features and comparisons of GATA1$_{NC}$-mPal crystal structures. A) Complexes from the P_1 crystal form (**PDB ID: 3vek**) with DNA in grey and dark grey and the $P2_1$ crystal form (**PDB ID: 3vd6**) with DNA in white and protein in orange. **B)** Complexes from the P_1 crystal form (**PDB ID: 3vek**) with protein in cyan and magenta and the $P2_1$ crystal form (**PDB ID: 3vd6**) with DNA in white and protein in orange. Images show the same alignment and orientation in both panels, but focus on the DNA and protein components, respectively, for clarity. **C)** Base-pairing of AA (white) and TT (magenta) overhangs from (PDB ID: 3vd6.

Note that compared to the P_1 crystal form (**PDB ID: 3vek**), the $P2_1$ structure (**PDB ID: 3vd6**) shows six additional crystal contacts between the DNA and amino acids of C- and N-terminal finger (Ade17 and Gln257, Gua25 and Arg219, Thy32 and Thr220) and six contacts between amino acids within the N-terminal finger (Ala201 and His232, Glu203 and Gly236, Tyr223 and Asn235). In addition to the crystal contacts between the AA and TT overhangs, DNA and amino acids (T6 and Asp300, C40 and 292 Asn), contacts between amino acids within the N- and C- terminal finger (Glu203 and His232, Val205 and Arg219, His 222 and Asn235, Leu288 and Thr 296, Asp300 and Tyr285), crystal contacts are established by additional zinc atoms where zinc coordinates two amino acids of different chains

(Zn1/Glu203/His232 and Zn2/Glu203/His232) or one DNA base and amino acid of one chain with a amino acid of a different chain (Zn5/T6/His289/Asp300 and Zn8/ T6/His289/Asp300).

Structures previously determined for GATA1 were generated using NMR methods and include the mouse NF in isolation (**PDB ID: 1gnf**)(1) , or in complex with FOG (**PDB ID: 1y0j**)(2), and the chicken $GATA1_{CF}$ (88% identity in the CF domain) bound to DNA containing an AGATAA motif (**PDB IDs: 1gat, 2gat** (3); and, with refinement with residual dipolar couplings, **PDB IDs: 3gat and 4gat** (4). The coordinates of 4gat are used for comparisons). GATA proteins share a highly conserved sequence in the ZF domains. Crystal structures of $GATA3_{CF}$ bound to two double-site oligonucleotides (**PDB IDs: 3dfv** and **3dfx)(5)**, and $GATA3_{NC}$ bound to mPal DNA (**PDB ID: 4hca**) and other double-site oligonucleotides have been determined (PDB Ids: **4hc7 and 4hc9)(6)**. Solution structures of the single GATA zinc-finger from an Aspergillus nidulans protein, AREA, bound to a GCGATAG site (**PDB ID: 5gat and 6gat**) have also been determined(7).

Apart from the orientation of binding sidechains, the conformations of NF in the structures are highly conserved (**Fig. S4A**). The structures of $GATA_{CF}$ structures are also essentially identical, apart from some differences in the CF-tail (**Fig. S4B**). Although most structures exhibit minor groove-binding by the $GATA3_{CF\text{-}tail}$, in one $GATA3_{CF}$ structure (**pdb ID: 3dfx**) there is an alternative minor conformation of the CF-tail into the adjacent major groove with the sidechain of GATA3-R364 binding a phosphate moiety in the DNA backbone and GATA3-R366 making no direct contacts with the DNA (Supp (5). A general difference between NMR and X-ray structures, is that residues equivalent to R305 and R307 in the mouse protein, make contacts with the phosphate backbone rather than binding into the minor groove as in our crystal structures, which may arise from flexibility and/or alternate conformations of the CF-tail region and differences in the ability of alternate methods to highlight alternate conformations, such as crystal packing and or detection of NOEs.

The structures of $GATA1_{NC}$ and $GATA3_{NC}$ bound to pseudo-palindromic DNA are structurally conserved (**Fig S4C**; r.m.s.d. over C_α of proteins = 0.55 Å). The few sequence differences between these two proteins (**Fig SD**; many of which are homologous) are restricted to the ends of structured regions of the domains or to surface exposed residues that do not have an obvious role in binding to DNA. Although not present in the deposited

coordinates of **4hca**, the authors reported the presence of sufficient electron density from the $GATA3_{NF}$ tail region that allowed the backbone, but not sidechains, of that region to be traced indicating that the $GATA3_{NF}$ tail had low occupancy in the minor groove in a conformation equivalent to that of the CF tail, but made apparently little contribution to DNA binding (6). In our structure we can see only low levels of electron density data for one or two residues at each end of the $GATA1_{NF}$ tail that are consistent with the same wrapping model. In each of the **4hc7 and 4hc9** structures, the $GATA3_{NF}$ tail is structured, (in both cases allowing the N- and C-fingers to bind different strands of dsDNA), but assumes different conformations (6), while other structures of N-terminally extended constructs of $GATA3_{CF}$ bound to DNA that include some of those residues which differ in sequence between GATA1 and GATA3 (SAARRAG in **3dfx** and RRAG in **3dfv** in GATA3) adopt other conformations. None of these structures suggest that the $GATA3_{NF}$ tail region makes significant contacts with DNA, but do indicate that region is likely to be flexible or plastic (6).

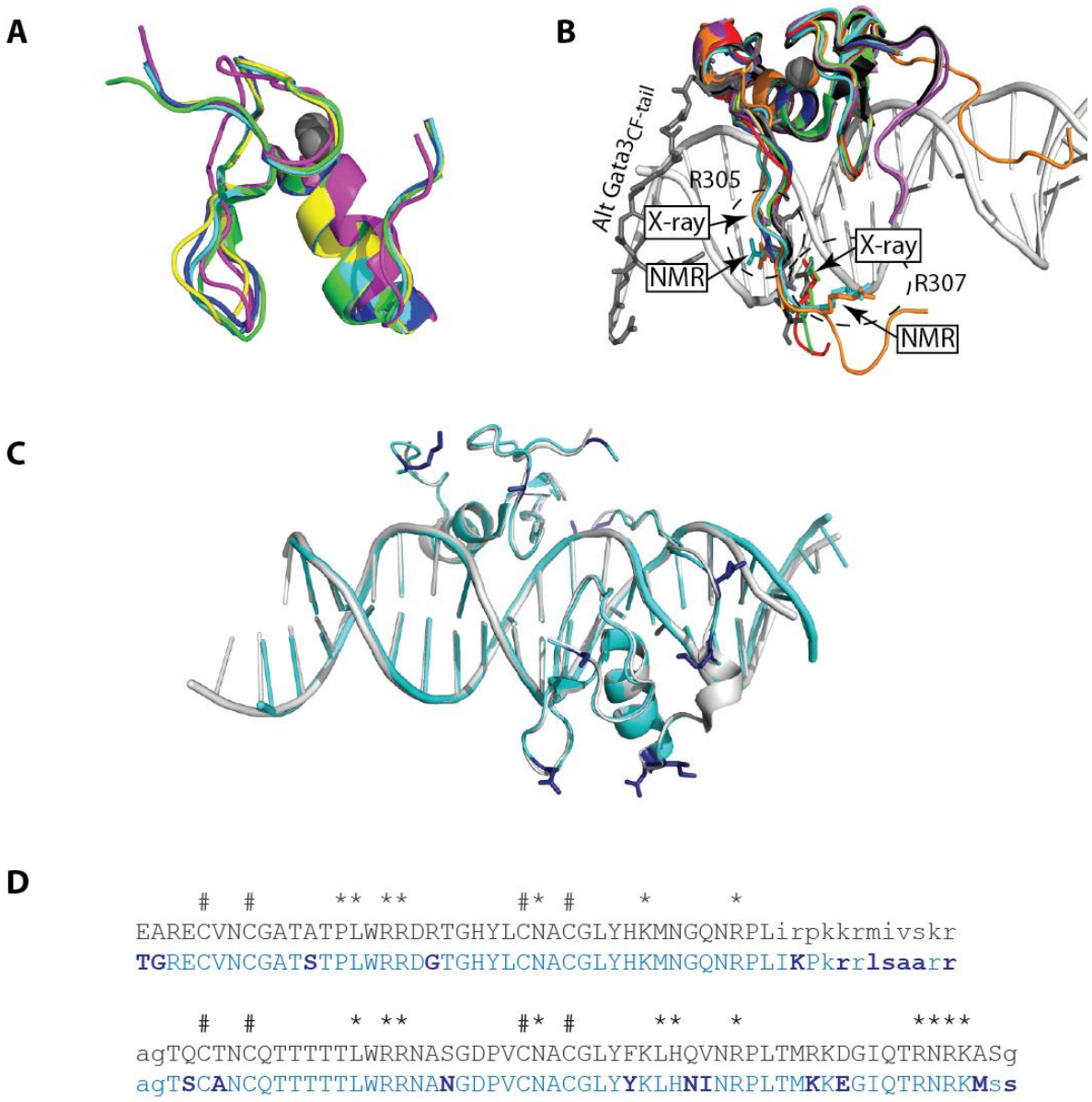

Figure S4 Comparison of GATA$_{NC}$ structures A) Structures of GATA$_{NF}$s from GATA1 (**PBD ID: 1gnf**, magenta, NMR structure), and GATA3 (**4hac**, green; **4hc7**, blue; and **4hc9**, yellow; X-ray structures) are aligned with GATA1$_{NF}$ from **3vd6** (cyan). **B)** Structures of GATA$_{CF}$s from chicken GATA1 (**4gat**, cyan, NMR structure), and GATA3 (**3dfv** -Chain C, black; **3dfv**-Chain D, grey; **3dfx**-Chain A, lilac; **3dfx**-Chain B, purple; **4hac**, green; **4hc7**, blue; and **4hc9**, yellow; Xray structures), and the GATA finger from *Aspergillus nidulans* Area (**5gat**, orange, NMR structure) are aligned with GATA1$_{NF}$ from **3vd6** (cyan) bound to DNA (white). Note the alternate conformation for **3dfv**-Chain D (stick representation). Sidechains of residues equivalent to GATA1-R305 and R307 are shown in stick

representation. **C)** Comparison of $GATA_{NC}$-mPal structures for GATA1 (white) and GATA3 (cyan). Residues from GATA3 that differ from GATA1 are shown as blue sticks. In all relevant panels zinc ions are shown as spheres. **D**) Sequence comparison of murine $GATA1_{NC}$ (black) as used in this study and human $GATA3_{NC}$ (blue) as used for the structure of **4hca.** Sequences that differ in GATA3 are shown in dark blue. Residues present and absent in the models of deposited structures are in upper and lower case, respectively. DNA binding (*) and zinc-coordinating (#) residues are indicated.

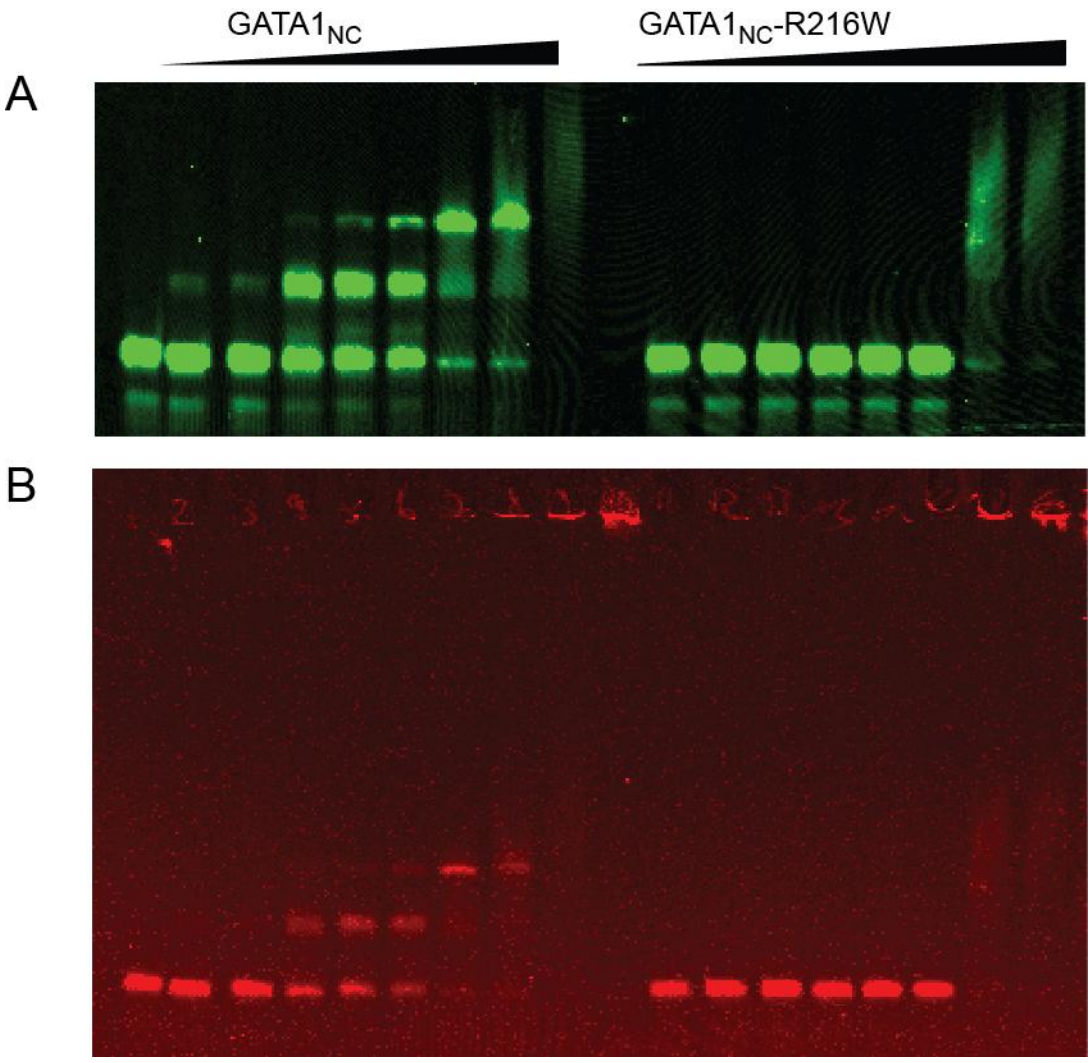

Figure S5: $GATA1_{NC}$ is unable to loop two binding sites on a single piece of DNA. EMSA analysis of $GATA1_{NC}$ (Lanes 2–9) and $GATA1_{NC}$-R216W (Lanes 11–18) at concentrations of 50, 100, 200, 400 and 600 nM under physiological salt (150 mM) binding to 30-mer DNA sequences containing a GATA site and a Cy5 tag or a GATG site and a Cy3 tag linked by a single stranded poly-dT 20mer. Images show the fluorescence emission of the Cy3 donor (green) following excitation of the Cy3 (A) or that of the Cy5-donor (red) following excitation of the Cy3 (B).

The binding of $GATA1_{NC}$-R216W is apparently weaker (signs of binding only evident at higher concentrations of protein) and is smeared rather than forming discrete gel-shift bands, similar to EMSA data for $GATA1_{NC}$ binding to GG_M1 DNA in Figure 4C.

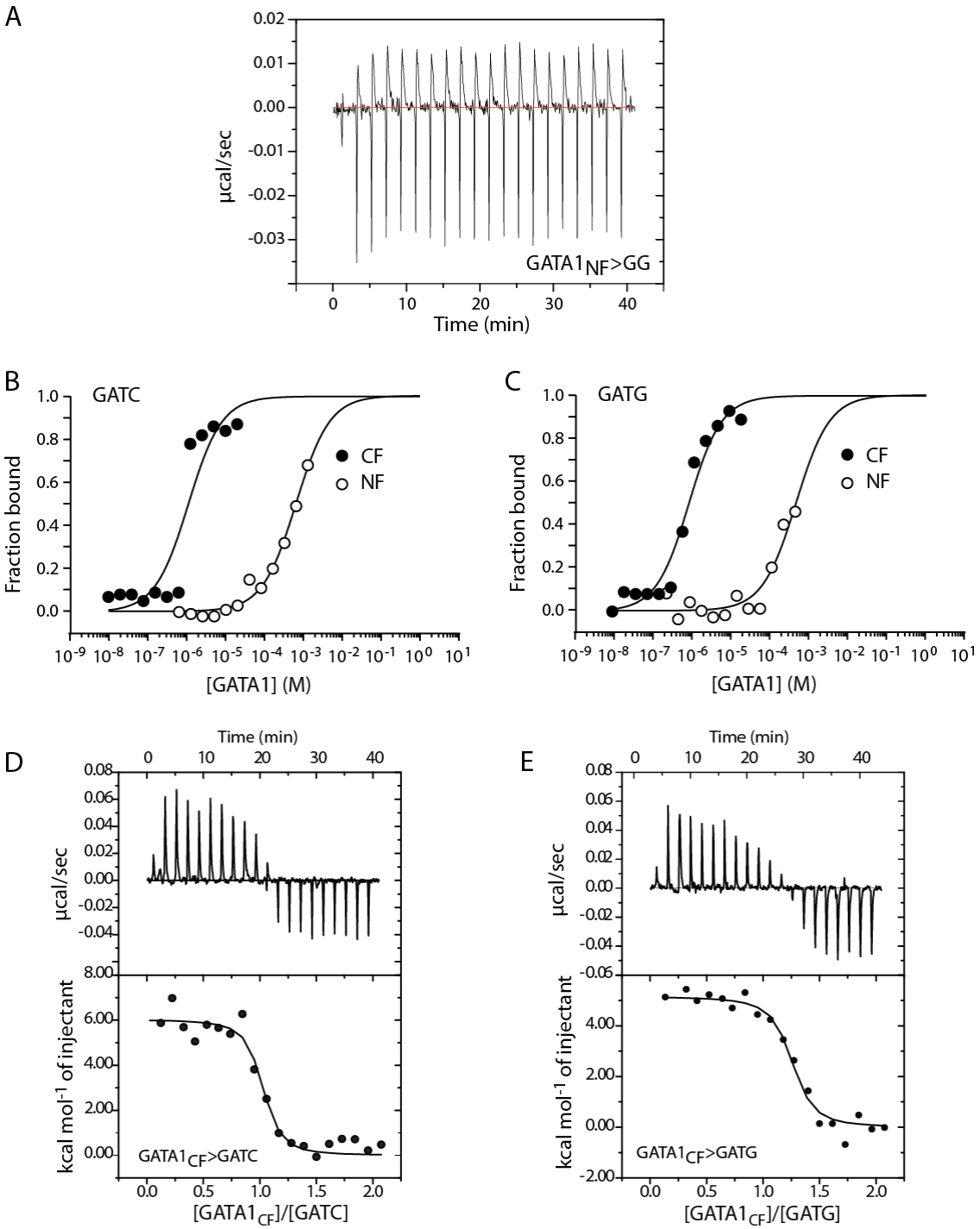

Figure S6: $GATA1_{NF}$ shows negligible levels of binding to DNA under physiological salt concentrations. A) ITC data showing the titration of $GATA1_{NF}$ (200 µM) into the GG oligonucleotide (20 µM) in 20 mM Tris, 150 mM NaCl, 1 mM TCEP pH 7.4 at 4 °C. **B)** MST data showing binding of $GATA1_{NF}$ and $GATA1_{CF}$ to GATC and **C**) GATG DNA at 25 °C **D)** ITC data for of $GATA1_{CF}$ (~200 µM) into the GATC (20 µM) at 15 °C and **E)** GATG at 25 °C. The buffer used for panels B-E was 20 mM Tris, 1 mM DTT pH 7.5 and 150 mM NaCl.

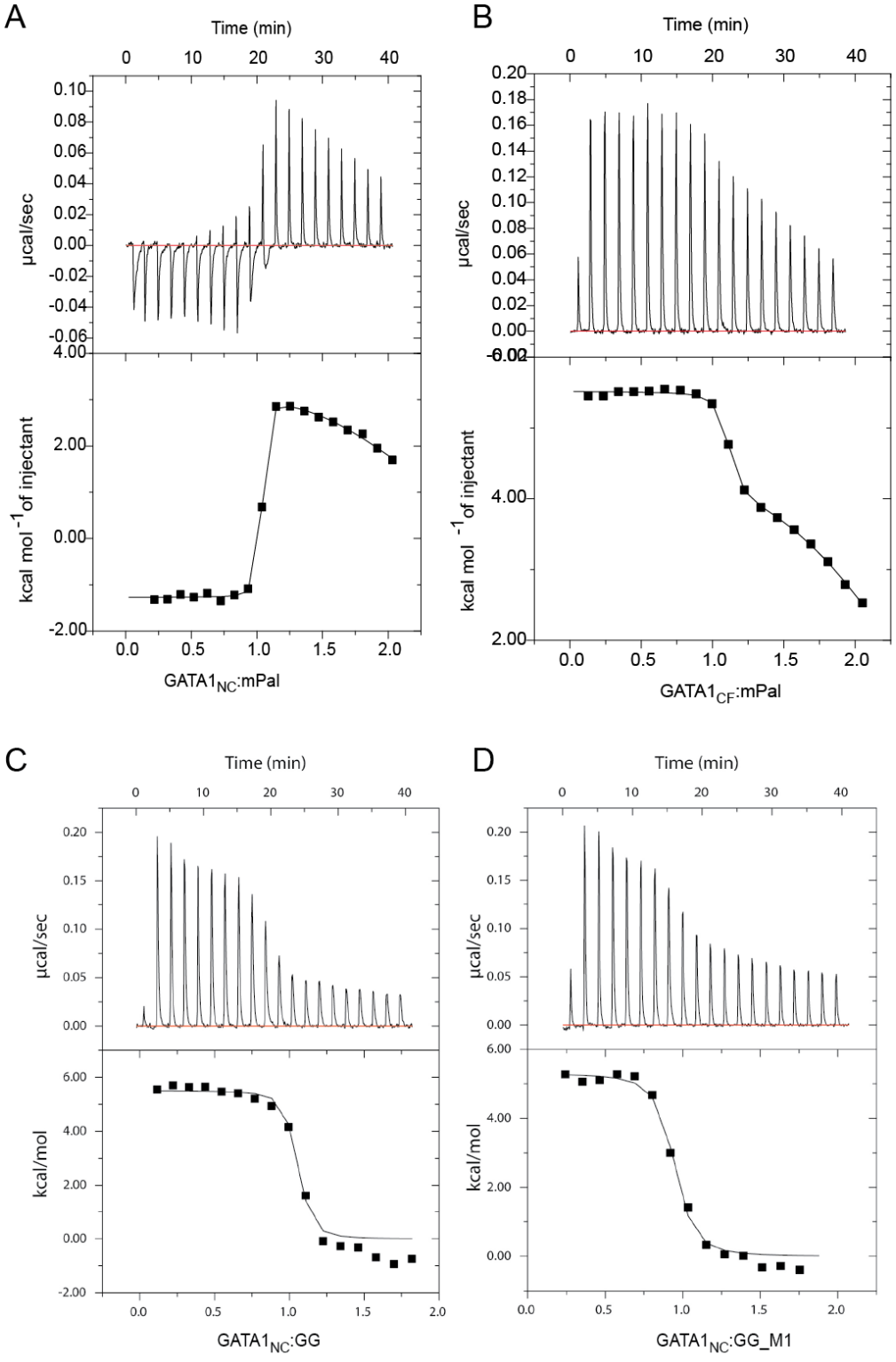

Figure S7: $GATA1_{NC}$ **or** $GATA1_{CF}$ **shows biphasic binding to GG and mPal DNA**. ITC data showing the titration of $GATA1_{NC}$ or $GATA1_{CF}$ (200 µM) into either the mPal, mPalM1, GG or GG_M1 oligonucleotide (20 µM) in 20 mM Tris, 150 mM NaCl, 1 mM TCEP pH 7.4 at 10 °C (except $GATA1_{NC}$ into mPal, which was collected at 20 °C).

A

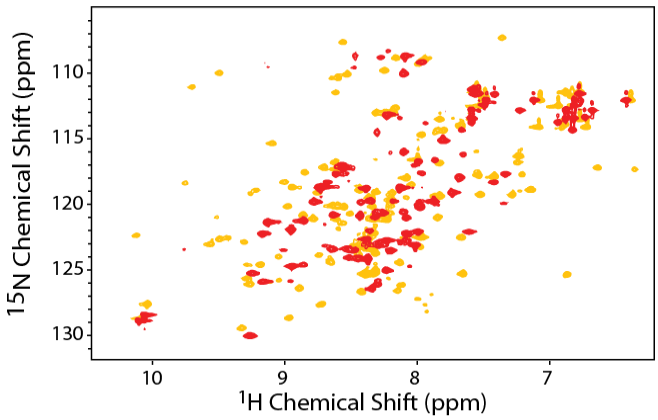

Figure S8: GATA1$_{NC}$ in complex with GG DNA. A) Overlay of the 2D ^{1}H-^{15}N HSQC spectra of ^{15}N-GATA1$_{NC}$ alone (220 µM, red peaks) and in complex with 2.2 molar equivalents of GG oligonucleotide (orange peaks). Figure 5. Model of GATA1$_{NC}$ in complex with GG DNA.

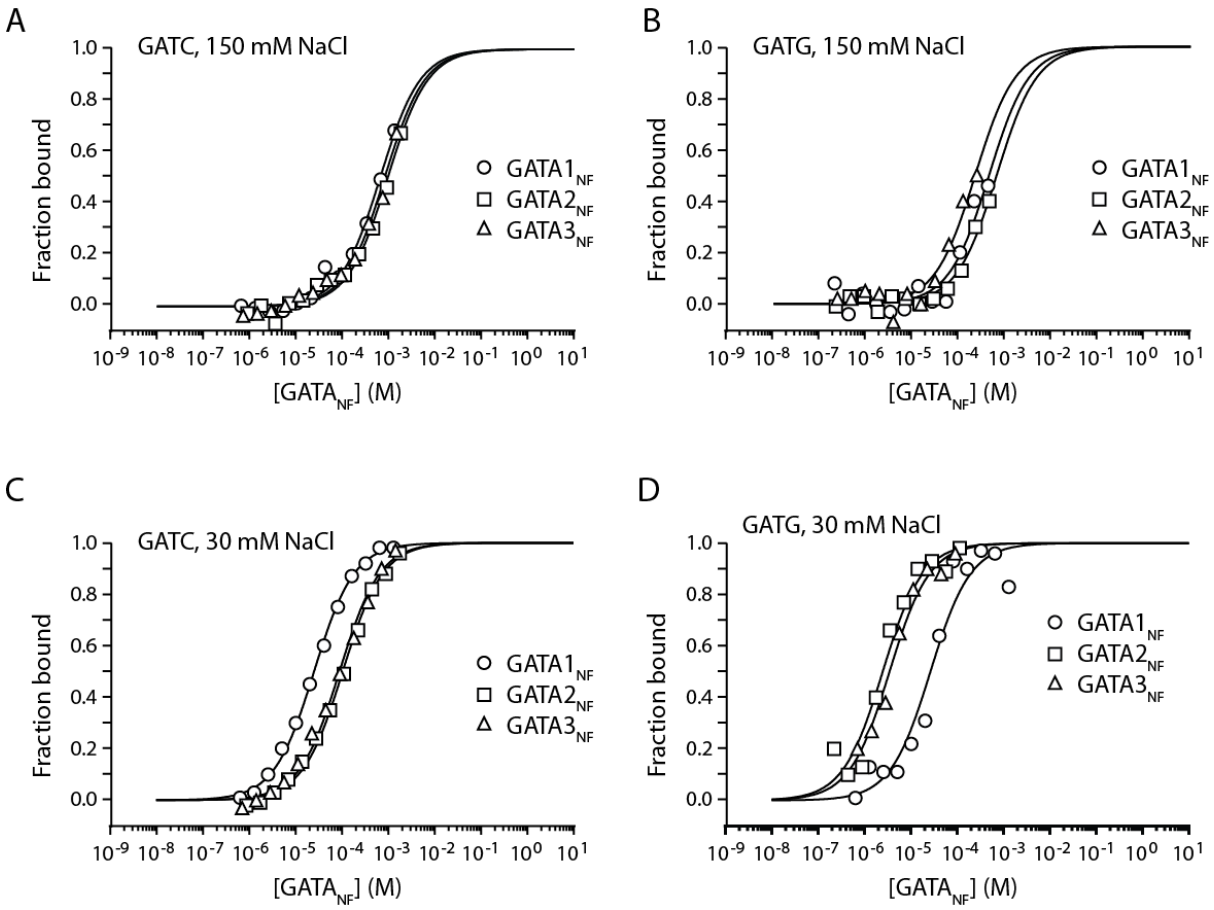

Figure S9. MST data shows negligible binding of GATA2–3$_{NF}$ binding to GATC-containing and GATG-containing DNA at physiological salt concentrations, with moderate binding of GATA1–3$_{NF}$ at low salt concentrations. Buffers were 20 mM Tris, 1 mM DTT pH 7.5 and 30 mM or 150 mM NaCl as indicated. Experiments were carried out at 25 °C. Data at higher concentrations (>100 μM) of protein appear to involve a non-specific binding component.

SUPPLEMENTARY METHODS

Oligonucleotide Sequences used in this paper

Sequences used for crystallography

Name	Sequences
mPal-Xtal top	AAGAGTCCATCTGATAAGAC
mPal-Xtal reverse	CTCAGGTAGACTATTCTGTT
GG-Xtal top	TTCGGGATAAAGATCTTAAG
GG-Xtal reverse	GCCCTATTTCTAGAATTCAA

Sequences of oligonucleotides used in radiolabelled EMSA, ITC and MST experiments

Name	Sequence of top strand (5'to 3')
mPal	AGTCCATCTGATAAGACTTCAGTGCTGCCC
mPal_M1	AGTCCCAGTGATAAGACTTCAGTGCTGCCC
GG	TTCAGCTTCGGGATAAAGATCTTAAATTC
GG_M1	TTCAGCTTCGGGATAAACAGCTTAAATTC
GATA1 (29-mer single GATA-site used in Figure 4C)	GATCTCCGGCAACTGATAAGGATTCCCTG
GATC (leading stand carries a 3' FAM tag for MST)	GCAACTGATCTGGACT
GATG (leading stand carries a 3' FAM tag for MST)	AGTCCATCTGTTAAGACTTA

Sequences used for In Gel FRET and two-colour fluorescence EMSA.

#	Oligonucleotide (5'-to-3')
1	GAT CCG GGC AGC ACT GAA GTC TTA ACA GAT GGA CT-CY3
2	AAT TC AGT CCA ACT GAT AAG ACT TCA GTG CTG CCC
3	AGT CCA TCT GTT AAG ACT TCA GTG CTG CCC G
4	CY5-GGG CAG CAC TGA AGT CTT ATC AGT TGG ACT G
5	CY5- AGA GAC TCT AGA TTC TGA AGT CAC GAC TTT
6	CCC GTC GTG ACT TCA GAA TAG TCT ACC TGA-CY3

For two-colour fluorescent EMSA, the following complementary pairs of oligonucleotides were used: #1/#3; #2/#4.

For in-gel FRET experiments oligos #5 and #6 were annealed to a ssDNA 80-mer containing complementary sequence to the two 30-mers, separated by a region of 20-nt poly(dT).

Supplementary Data References

1. Kowalski, K., Czolij, R., King, G.F., Crossley, M. and Mackay, J.P. (1999) The solution structure of the N-terminal zinc finger of GATA-1 reveals a specific binding face for the transcriptional co-factor FOG. *J Biomol NMR*, **13**, 249-262.
2. Liew, C.K., Simpson, R.J., Kwan, A.H., Crofts, L.A., Loughlin, F.E., Matthews, J.M., Crossley, M. and Mackay, J.P. (2005) Zinc fingers as protein recognition motifs: structural basis for the GATA-1/friend of GATA interaction. *Proc Natl Acad Sci U S A*, **102**, 583-588.
3. Omichinski, J.G., Clore, G.M., Schaad, O., Felsenfeld, G., Trainor, C., Appella, E., Stahl, S.J. and Gronenborn, A.M. (1993) NMR structure of a specific DNA complex of Zn-containing DNA binding domain of GATA-1. *Science*, **261**, 438-446.
4. Tjandra, N., Omichinski, J.G., Gronenborn, A.M., Clore, G.M. and Bax, A. (1997) Use of dipolar 1H-15N and 1H-13C couplings in the structure determination of magnetically oriented macromolecules in solution. *Nat Struct Biol*, **4**, 732-738.
5. Bates, D.L., Chen, Y., Kim, G., Guo, L. and Chen, L. (2008) Crystal Structures of Multiple GATA Zinc Fingers Bound to DNA Reveal New Insights into DNA Recognition and Self-Association by GATA. *J Mol Biol*, **381**, 1292-1306.
6. Chen, Y., Bates, Darren L., Dey, R., Chen, P.-H., Machado, Ana Carolina D., Laird-Offringa, Ite A., Rohs, R. and Chen, L. (2012) DNA Binding by GATA Transcription Factor Suggests Mechanisms of DNA Looping and Long-Range Gene Regulation. *Cell Rep*, **2**, 1197-1206.
7. Starich, M.R., Wikström, M., Arst Jr, H.N., Clore, G.M. and Gronenborn, A.M. (1998) The solution structure of a fungal AREA protein-DNA complex: an alternative binding mode for the basic carboxyl tail of GATA factors. *J Mol Biol*, **277**, 605-620.