Supplementary information for

Binding of transcription factor GabR to DNA requires recognition of DNA shape at a location distinct from its cognate binding site

*Walid A Al-Zyoud[1], *Robert MG Hynson[2], Lorraine Ganuelas[2], Adelle CF Coster[3], Anthony P Duff[4], Matthew Baker[2], Alastair G Stewart[2], Eleni Giannoulatou[2], Joshua WK Ho[2], Katharina Gaus[1,5], Dali Liu[6], ^Lawrence K Lee[1,2,5], ^Till Böcking[1,5]

[1]School of Medical Sciences, The University of New South Wales, Australia
[2]The Victor Chang Cardiac Research Institute, 405 Liverpool St. Darlinghurst, Australia
[3]School of Mathematics and Statistics, University of New South Wales, Australia
[4]Australian Nuclear Science and Technology Organisation, Lucas Heights, Australia
[5]EMBL Australia Node for Single Molecule Science, The University of New South Wales, Australia
[6]Department of Chemistry and Biochemistry, Loyola University, Chicago, USA

*These authors contributed equally

^Joint Corresponding Authors:

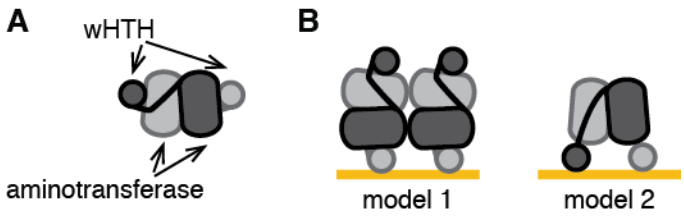
Lawrence K. Lee
School of Medical Sciences
EMBL Australia Node for Single Molecule Science
The University of New South Wales
Corner Botany and High Street,
Kensington Campus, Australia 2052
Tel. +612 9385 8252

Honorary Faculty, The Victor Chang Cardiac Research Institute
Lowy-Packer Building
405 Liverpool Street,
Darlinghurst, Australia 2010
Tel: +612 9295 8651 Fax: +612 9295 8601

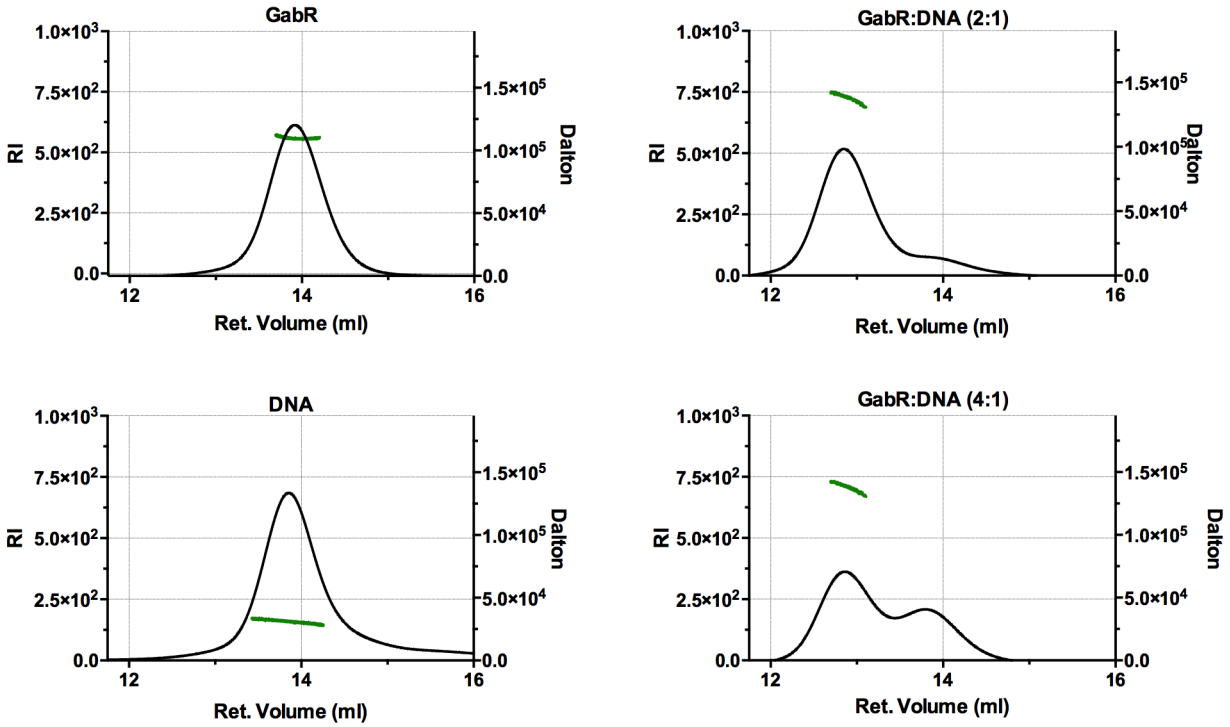
Email: lawrence.lee@unsw.edu.au

Till Böcking
School of Medical Sciences
EMBL Australia Node for Single Molecule Science
The University of New South Wales
Corner Botany and High Street,
Kensington Campus, Australia 2052

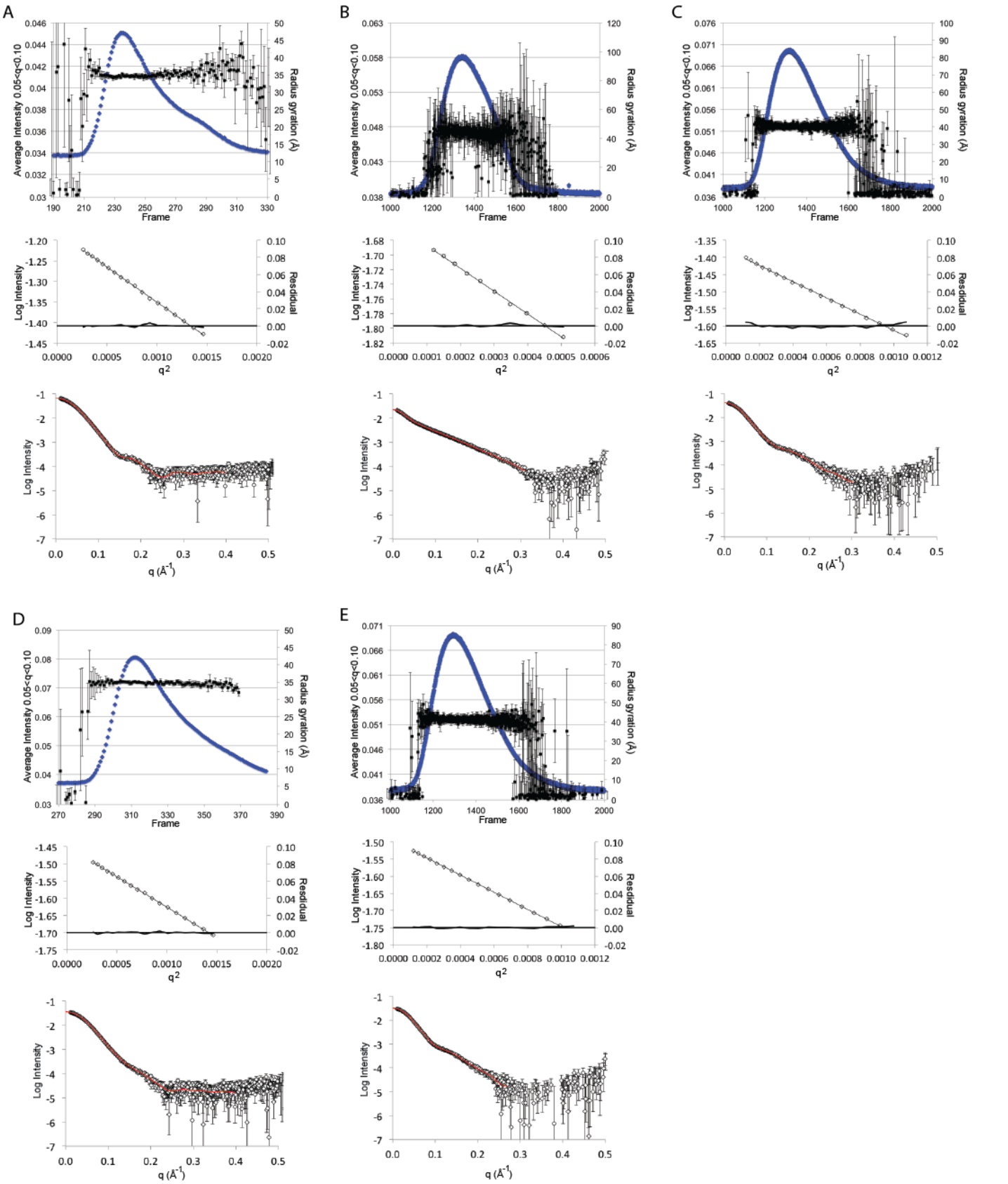
Email: t.boecking@unsw.edu.au



Supplementary Figure S1. Models for the GabR-DNA complex.
A. Schematic representation of the GabR dimer structure with arrows indicating the positions of the N-terminal winged helix-turn-helix (wHTH) domains and the C-terminal aminotransferase domain. **B.** Models for the binding of GabR to the regulatory DNA region containing two direct repeat sequences (ATACCA). In model 1, two dimers bind to the DNA with each dimer binding via one of its wHTH domains to one ATACCA motif. In model 2, a single dimer binds to the regulatory DNA region. A conformational change is required such that the wHTH domains can each contact one of the two ATACCA motifs.

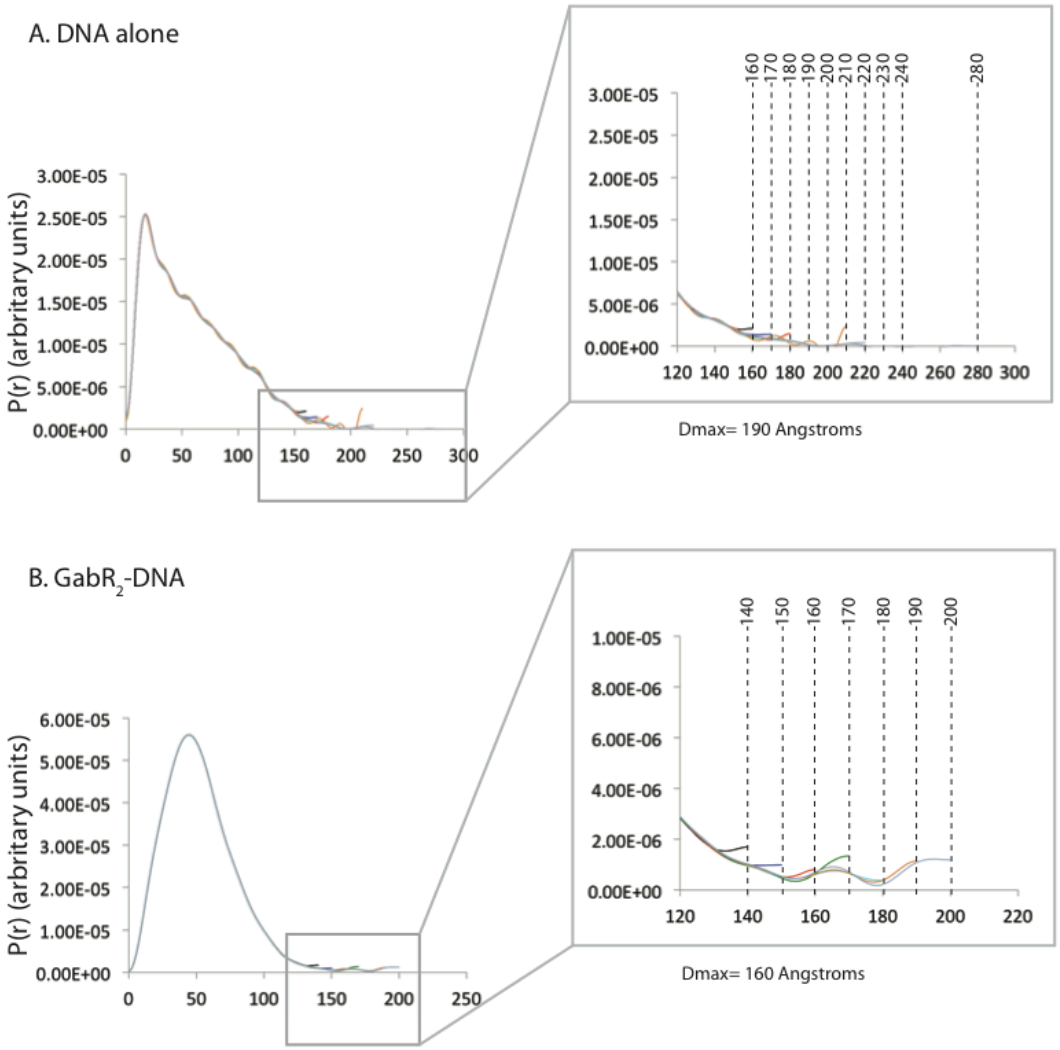


Supplementary Figure S2. Multi-angle laser light scattering of GabR, DNA and GabR-DNA complex.
Size exclusion chromatography elution profiles monitored by refractive index of samples containing GabR, DNA and GabR + DNA pre-incubated at molar ratios of 2:1 and 4:1. The molecular mass determined from light scattering for the major peak is shown in green.



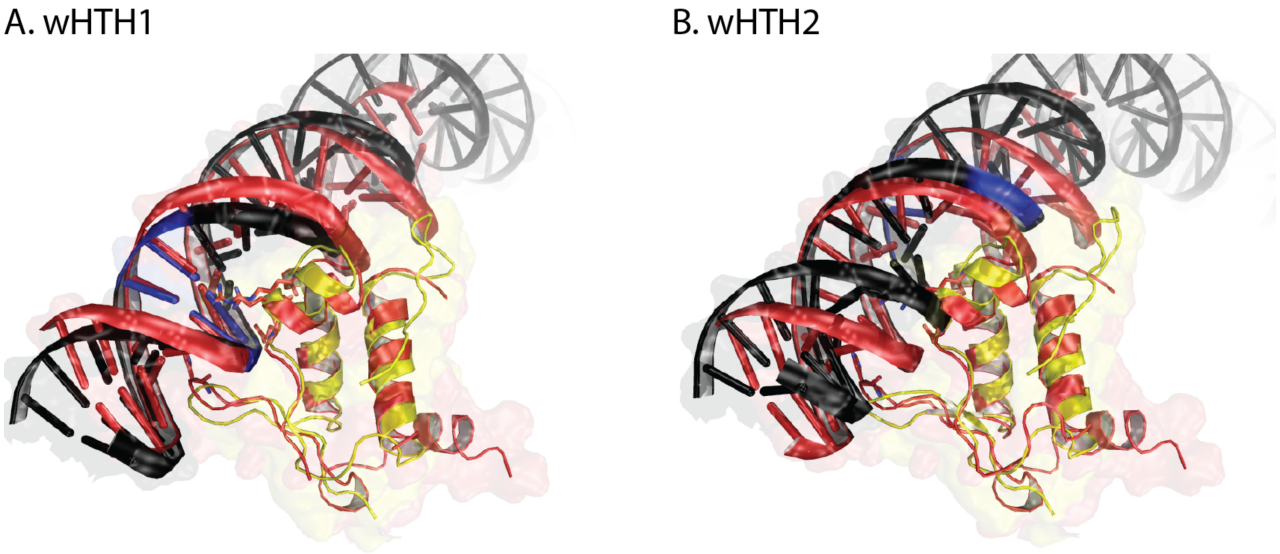
Supplementary Figure S3. SAXS analysis.

Plots of average low q intensity ($0.01 < q < 0.5$) and calculated radius gyration from background-subtracted data vs frame (top panel) where the frame indices that were scaled and averaged for further analysis are indicated at the top of the plot in text, subtracted scattering overlaid with fit used to generate P(r) vs r plots (middle panel) and linear fits to Guinier plots with residuals (bottom panel) are displayed for all datasets presented in this work: (A) GabR, (B) DNA, (C) GabR in complex with DNA, (D) GabR in the presence of γ-aminobutyric acid, (E) GabR in complex with DNA in the presence of γ-aminobutyric acid.



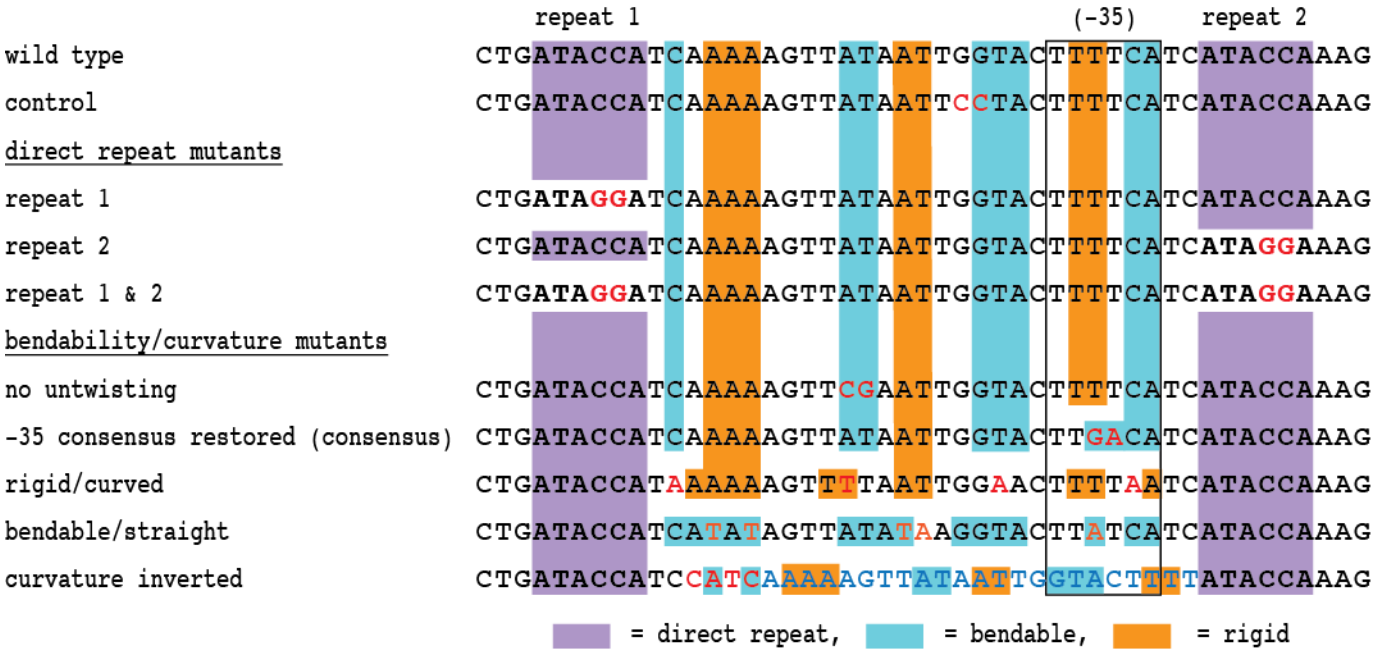
Supplementary Figure S4. Determining maximum dimensions from SAXS.
Maximum dimensions were determined by calculating unrestrained interatomic distance distribution (P(r)) plots with a range of specified maximal dimensions (Dmax) sampled every 10 Å. These P(r) plots are shown for the (A) 53 base pair DNA strand used in this study alone and (B) the $GabR_2$-DNA complex. The inset shows a enlarged section of the P(r) plots around the maximum dimensions tested, which are marked by a dashed line. The Dmax was chosen as the shortest interatomic distance above which these P(r) plots did not systematically increase.

Supplementary Figure S5. 3D models of the wild type gabRTD regulatory region.
3D structures of the DNA sequence shown in Fig 1A generated using the DNA curvature analysis tool (C. Gohlke, www.lfd.uci.edu/~gohlke/dnacurve/) with the following models: straight DNA (white), Calladine & Drew (yellow), nucleosome positioning (orange), DNAse I consensus (green), Cacchione & De Santis (magenta), AA wedge (cyan), Bolshoi and Trifonov (blue).

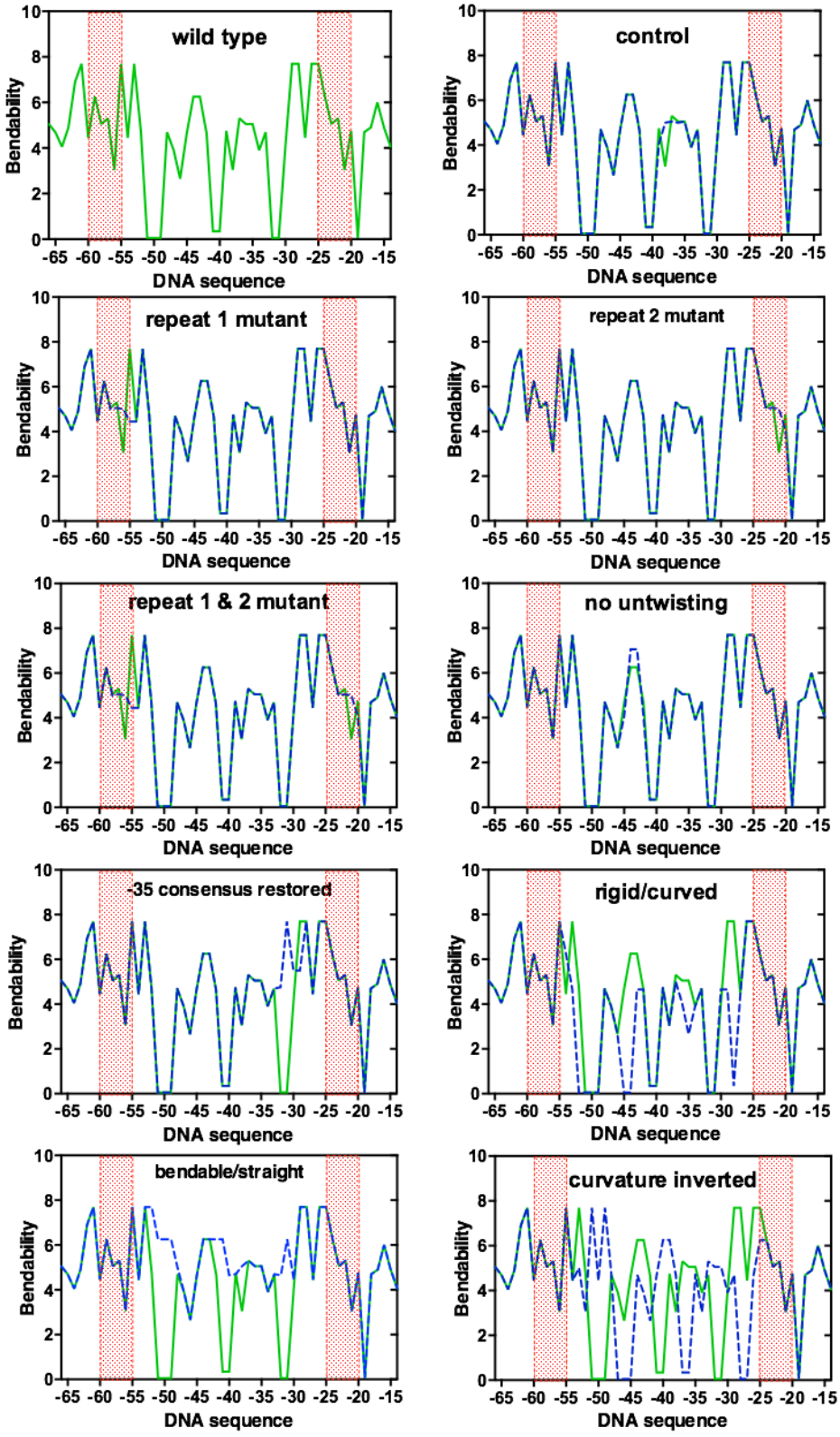


Supplementary Figure S6.
An enlarged view of the GabR wHTH binding domains in the model shown in Figure 4C. The co-crystal structure of the HTH domain of FadR in complex with a short strand of DNA (PDBID:1H9T) is rendered in red and structurally superimposed with the crystal structure of the GabR wHTH domains (yellow). The model of the DNA fragment targeted by GabR is rendered in black. The tandem cytosine bases in the repeated ATACCA cognate sites that are involved in binding to the GabR wHTH domains are in blue. At the first wHTH site (A), these tandem cytosines are superimposed with the equivalent cytosine repeat in the structure of the DNA fragment in the FadR-DNA complex. After bending the GabR target DNA strand around the GabR dimer, the proxmity of the tandem cytosine in the second ATACCA repeat at the second wHTH domain is shown in B.

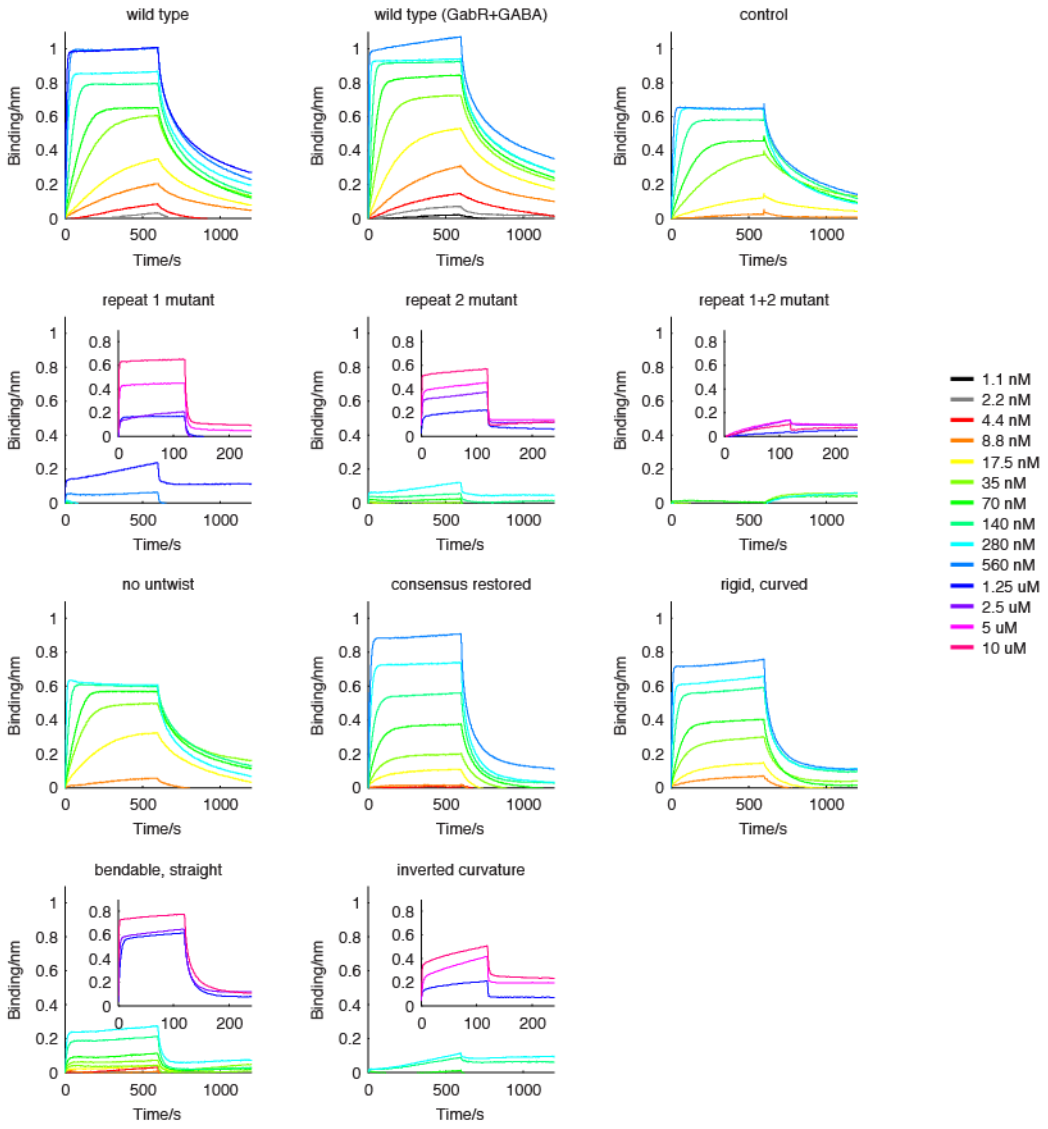
Supplementary Figure S7. Design of DNA oligonucleotides for biolayer interferometry.
The biotinylated DNA duplex was formed by hybridisation of two short oligonucleotides (oligo 1 and biotinylated oligo 2) with a long complementary oligonucleotide (oligo 3) and captured onto a streptavidin-modified biosensor.



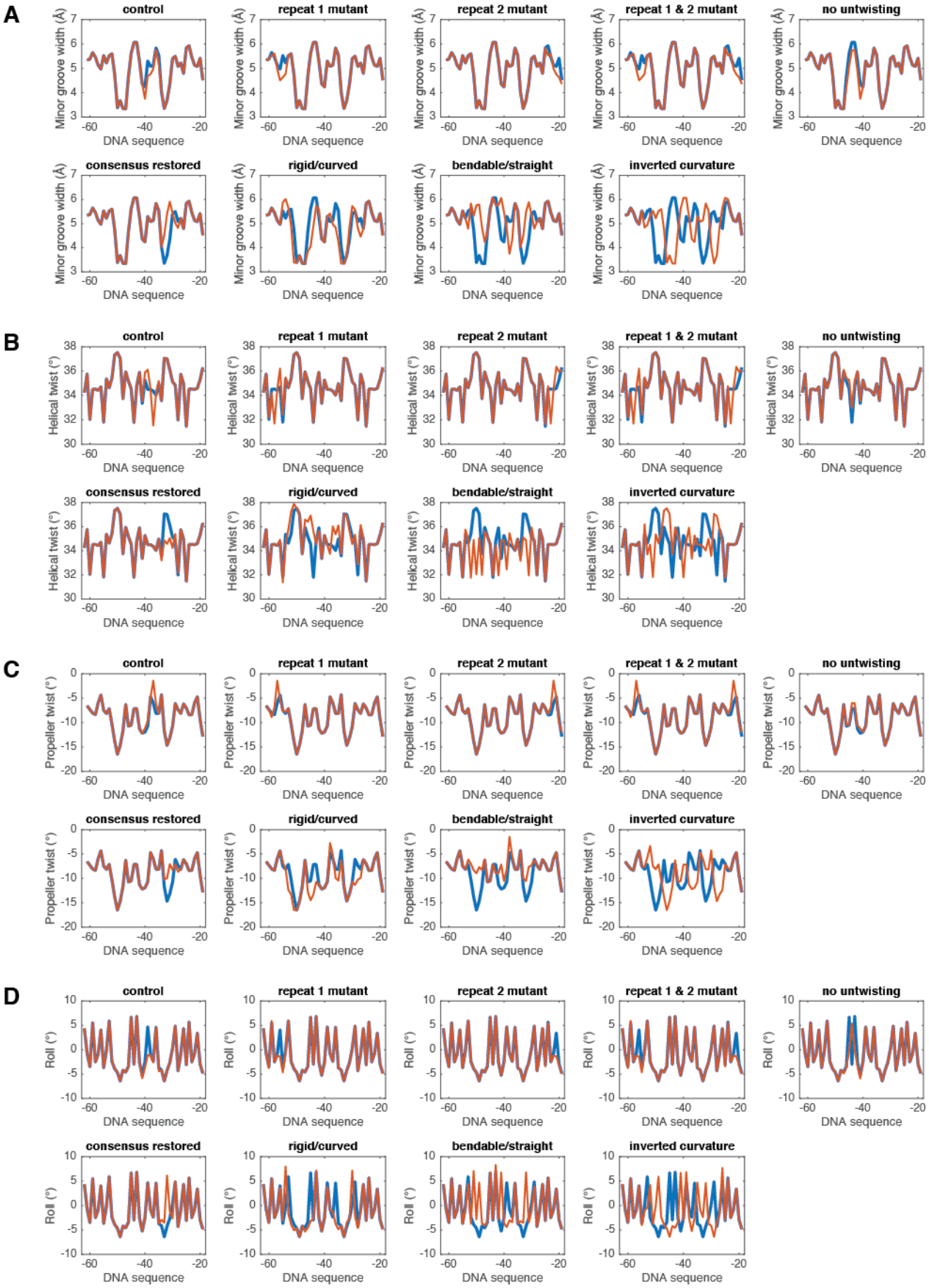
Supplementary Figure S8. Sequences of mutants analysed by biolayer interferometry.
Mutations used in this study to disrupt binding sites or alter the bendability of the bridging region between the repeat sequences. Mutations are shown in red. The DNA sequence shifted by 4 nucleotides in the "curvature inverted" mutant is shown in blue. The direct repeat sequences (purple) and regions of high (cyan) and low (orange) bendability are highlighted.



Supplementary Figure S9. Bendability plots of DNA sequences.
Bendability of wild type and mutant DNA sequences calculated using a trinucleotide model based on DNAse I digestion and nucleosome binding data (consensus scale).



Supplementary Figure S10.
Biolayer interferometry traces of GabR binding to and dissociation from surface-immobilised DNA duplexes (wild type and mutants).



Supplementary Figure S11.
Predicted minor groove width (A), helical twist (B), propeller twist (C) and roll (D) of wild type (blue) and mutant sequences (orange) calculated using DNAshape [31].

Supplementary Table S1. Summary of structural parameters derived from SAXS data

	GabR	DNA	GabR + DNA	GabR + GABA	GabR + GABA + DNA
Data collection parameters					
Wavelength (Å)	1.127	1.127	1.127	1.127	1.127
Q-range	0.0110 - 0.544	0.0110 – 0.503	0.0110 – 0.503	0.0110 – 0.544	0.0110 – 0.544
Exposure time (Sec)	5	1	5	1	1
Temperature (K)	293	293	293	293	293
Processing parameters					
P(r) Q-range ($Å^{-1}$)	0.0162 – 0.397	0.110 – 0.310	0.0122 – 0.304	0.110 – 0.398	0.0110 – 0.265
Guinier Q×Rg range	0.557 – 1.32	0.506 – 1.04	0.441 – 1.32	0.431 – 1.25	0.454 – 1.31
Structural parameters					
I(0) (cm^{-1}) [from P(r)]	$0.0661 \pm 8.04\times10^{-5}$	$0.0224 \pm 7.30\times10^{-5}$	$0.0421 \pm 9.23\times10^{-5}$	$0.0356 \pm 3.60\times10^{-5}$	$0.0320 \pm 6.26\times10^{-5}$
Rg (Å) [from P(r)]	34.6 ± 0.0770	49.7 ± 0.341	41.2 ± 0.172	35.4 ± 0.0661	42.6 ± 0.158
I(0) (cm^{-1}) [from Guinier]	0.66	0.036	0.022	0.042	0.032
Rg (Å) [from Guinier]	34.4 ± 0.0670	46.30 ± 0.372	40.4 ± 0.107	35.0 ± 0.0750	41.6 ± 0.103
Maximum dimension (Å)	125	190	160	130	160
Porod volume estimate ($Å^3$)	191739	46633	200056	212270	215342
Shape reconstructions					
Number of reconstructions	20	Not calculated	Not calculated	24	Not calculated
Average normalized spatial distribution (NSD)	0.517 ± 0.00428	Not calculated	Not calculated	0.507 ± 0.00325	Not calculated
DAM volume ($Å^3$)	207165 ± 262	Not calculated	Not calculated	214363 ± 156	Not calculated
Molecular-mass estimates					
Calculated monomeric MW from sequence (g/mol)	55166	31800	Not calculated	55166	Not calculated
Calculated MW from Porod volume (g/mol)	112788	27431	117680	124864	126672
Calculated MW from DAMMIN (g/mol)	103583	31796	117675	107182	125498

Supplementary Table S2. Summary of kinetic parameters (with 95% confidence interval) derived from biolayer interferometry data.

DNA	$k_{on}/\mu M^{-1} s^{-1}$	k_{off}/s^{-1}	KD/µM (#)	KD/µM (*)
wild type	0.139 [0.103, 0.175]	0.007 [0.003, 0.01]	0.048 [0.022, 0.083]	0.027 [0.019, 0.036]
wild type (+GABA)	0.343 [0.309, 0.376]	0.006 [0.003, 0.008]	0.017 [0.009, 0.024]	0.013 [0.012, 0.014]
control	0.150 [0.141, 0.16]	0.006 [0.003, 0.01]	0.042 [0.015, 0.068]	0.044 [0.023, 0.065]
repeat 1 mutant	0.062 [0.034, 0.091]	0.287 [0.17, 0.403]	4.604 [2.001, 11.765]	6.356 [5.482, 7.23]
repeat 2 mutant	0.064 [0.037, 0.091]	0.891 [0.491, 1.29]	13.915 [5.812, 32.078]	7.099 [5.226, 8.971]
no untwisting	0.305 [0.12, 0.491]	0.005 [0.004, 0.006]	0.017 [0.009, 0.066]	0.025 [0.013, 0.037]
consensus restored	0.125 [0.09, 0.161]	0.012 [0.01, 0.015]	0.099 [0.067, 0.154]	0.115 [0.092, 0.137]
rigid/curved	0.205 [0.131, 0.279]	0.012 [0.01, 0.014]	0.061 [0.04, 0.108]	0.101 [0.085, 0.117]
bendable/straight	0.076 [0.058, 0.094]	0.076 [0.054, 0.097]	0.993 [0.641, 1.485]	1.156 [0.702, 1.611]
curvature inverted	0.027 [-0.012, 0.066]	0.459 [-0.049, 0.968]	16.753 [-9.256, -9.256]	11.727 [5.794, 17.661]

(#) Calculated from the dissociation and association rate constants.
(*) Derived using equilibrium analysis.