Supporting Information for

Highly disordered histone H1–DNA model complexes and their condensates

Abigail L. Turner[a,1], Matthew Watson[a,1], Oscar G. Wilkins[a,1], Laura Cato[a,2], Andrew Travers[a,b], Jean O. Thomas[a] & Katherine Stott[a,3]

[a]Department of Biochemistry, University of Cambridge, 80 Tennis Court Road, Cambridge CB2 1GA, UK ; and [b]Medical Research Council Laboratory of Molecular Biology, Cambridge Biomedical Campus, Cambridge CB2 0QH, UK

[1]These authors contributed equally to this work

[2]Present address: Molecular Oncology, Sanofi Oncology, Cambridge, MA 02139, USA

[3]To whom correspondence may be addressed. Email: ks123@cam.ac.uk, ORCID: 0000-0002-4014-1188

www.pnas.org/cgi/doi/10.1073/pnas.1805943115

FIGURE S1

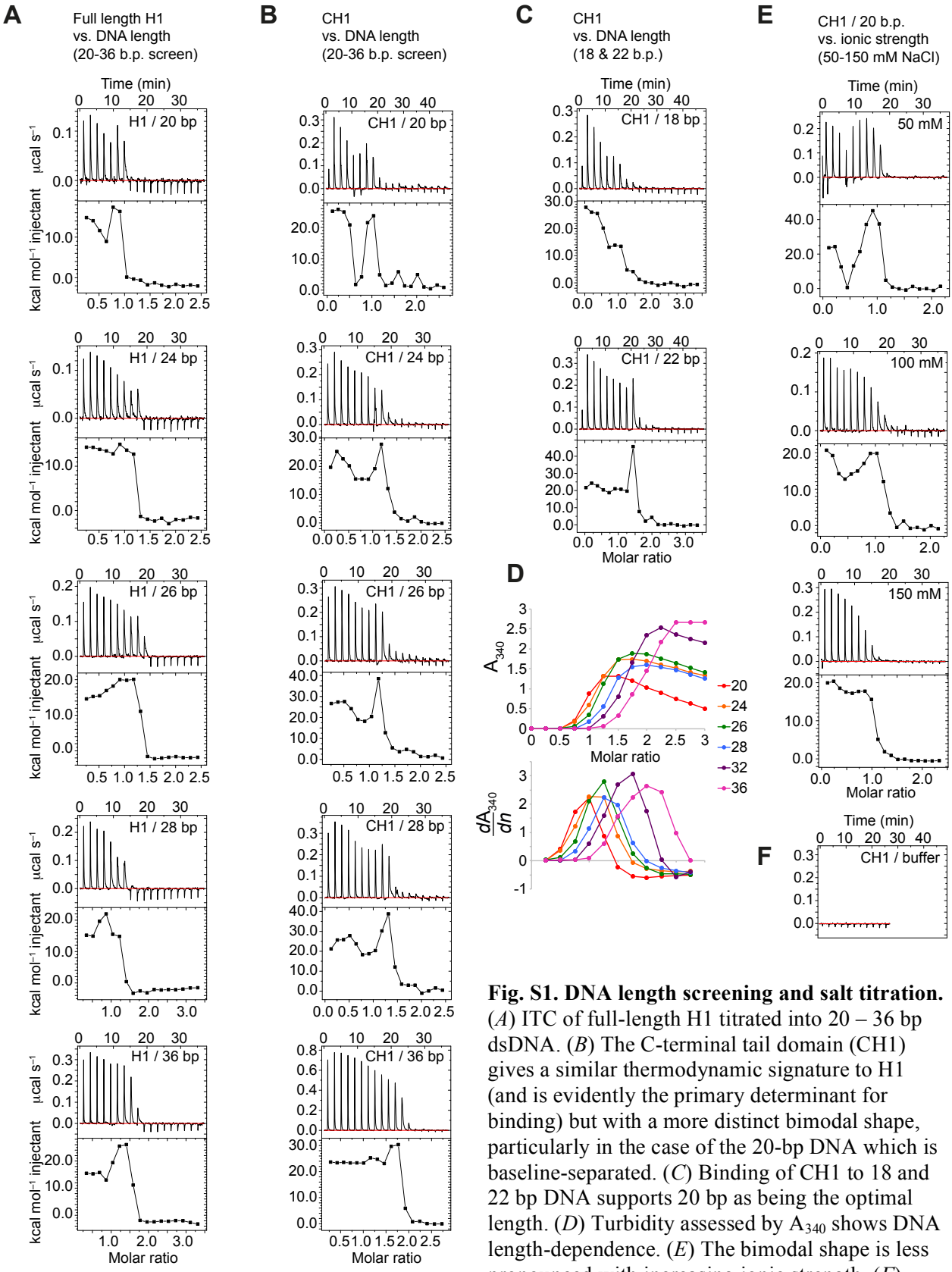

Fig. S1. DNA length screening and salt titration. (*A*) ITC of full-length H1 titrated into 20 – 36 bp dsDNA. (*B*) The C-terminal tail domain (CH1) gives a similar thermodynamic signature to H1 (and is evidently the primary determinant for binding) but with a more distinct bimodal shape, particularly in the case of the 20-bp DNA which is baseline-separated. (*C*) Binding of CH1 to 18 and 22 bp DNA supports 20 bp as being the optimal length. (*D*) Turbidity assessed by A_{340} shows DNA length-dependence. (*E*) The bimodal shape is less pronounced with increasing ionic strength. (*F*) Control heats were negligible. (*A – D* & *F* were performed at low ionic strength.)

MATERIALS AND METHODS

Expression and purification of proteins

Recombinant chicken histone H1 subtype H1.11L (here referred to as H1) was expressed and purified as described previously (1). The DNA sequence for the C-terminal domain of H1.11L, CH1 (residues ^{115}KPGEV … AAKKK^{225}), was cloned into pET13a using the NdeI and BamH1 sites to give pET13a CH1. CH1 was expressed in *E. coli* BL21(DE3) cells grown in 2YT (or M9 minimal medium, supplemented with 0.4 g/L $^{15}NH_4Cl$, or 0.4 g/L $^{15}NH_4Cl$ and 2 g/L $^{13}C_6$ D-glucose as the sole nitrogen and carbon sources for isotopic labelling), supplemented with 50 μg/ml kanamycin. At an OD_{600} of 0.7, cells were induced with 0.1 mM isopropyl β-D-1-thiogalactopyranoside and grown for a further 4 h at 30 ºC (2YT) or overnight at 23 ºC (M9 minimal medium) before harvesting. CH1 was purified as for H1. Protein purity was determined by SDS/18%-PAGE with Coomassie Blue staining (2).

Protein phosphorylation

Recombinant CDK2/cyclin A was expressed from plasmids pGEX3C CDK2 and pET21d cyclin A3 as described (3). ^{15}N-labelled CH1 (500 μl of ca. 1 mM) was phosphorylated with CDK2/cyclin A (in 10 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, with a 100-fold molar excess of ATP over protein) for 8 h at 30 ºC. The kinase reaction was stopped with 10 mM (final) EDTA pH 8, and ATP was removed by dialysis into 10 mM sodium phosphate pH 6.0, 1 M NaCl; the sample was further dialysed into salt-free buffer, after which phosphorylation was assessed by ^{15}N-HSQC by monitoring the appearance of peaks with a pronounced downfield ^{1}H shift corresponding to pSer, and disappearance of the original peaks (4, 5).

Preparation of dsDNAs

Oligonucleotides were purchased from Sigma. Sequences were designed by removal of 5' and 3' nucleotides from the 36-mer oligonucleotides (6) shown, in a symmetrical manner (20-mer sequences underlined):

5′–ATCAAGCTACGCCTGAAGAGTCTGGTGAGCAAGGGT–3′

3′–TAGTTCGATGCGGACTTCTCAGACCACTCGTTCCCA–5′

Oligonucleotides were dissolved in 10 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM EDTA. The two strands were combined in equimolar amounts and annealed by heating to 94 ºC for 10 min before cooling to 4 ºC over 6 h. Annealing was assessed by native 20% PAGE in 0.5 x Tris/borate/EDTA (TBE) buffer.

Concentration measurements

Protein concentration was determined from A_{205} for CH1 and A_{280} for full-length H1. DNA concentration was determined from A_{260}. Due to the absence of aromatic residues, the extinction coefficient for CH1 was checked by amino acid analysis in duplicate and, in the course of AUC experiments, by Rayleigh interferometry (described below). Turbidity was measured from A_{340} in 50 μl, 1 cm path length cuvettes using a NanoDrop One^C instrument (ThermoFisher).

Isothermal titration calorimetry

ITC experiments were performed on a Malvern iTC200 instrument at 25 ºC on samples that had been dialysed extensively into 10 mM sodium phosphate pH 6 containing 0-150 mM NaCl as indicated. The protein (at 40-85 μM) was injected into the DNA (at 5 μM); between 18 and 38 injections of 2 or 1 μl protein were performed, respectively, at intervals of 120 or 150 s with stirring at 750 rpm. Baseline correction and integration were performed in Origin. Isotherms that were single sigmoidal curves were fitted to the one-site model in Origin. Isotherms with distinct bimodal features were first re-acquired at high resolution and in triplicate, before fitting using the iterative least-squares approach in Matlab (release 2017b), to a two-step sequential process based on the Multiple Non-Interacting Sites (MNIS) model. This model has been developed (7, 8) to fit two-site processes where $K_{d,2} < K_{d,1}$, because the 'Two Sets of Sites' model provided in the manufacturer's software is not suitable since it assumes the stronger-binding site 1 titrates before the weaker site 2, i.e. $K_{d,2} > K_{d,1}$. Due to the two processes, ΔH is the sum of ΔH_1 and ΔH_2, for the first and second processes. In the modified MNIS model, to account for the sequential nature of the processes an extra sigmoidal function $\alpha(x)$ is introduced (7) such that $\Delta H(x) = \Delta H_1(x, n_1, r_1) + \alpha(x)\, \Delta H_2(x, n_2, r_2)$

where x = protein/DNA molar ratio, n = stoichiometry, $r = K_d$/[cell], $\Delta H\,(x, n, r) = (1/2)\,\Delta H\,[1 + (n–x–r)/\{(n+x+r)^2 – 4xn\}^{1/2}]$ and $\alpha(x)$ is the fraction of coacervate, $\alpha(x) = [1 + \exp\{–(x–n_2)/0.1\}]^{–1}$, i.e. a rising sigmoid of height 1, centred at n_2 and with lateral extension 0.1. The molar ratio, x, was fixed; all other parameters (ΔH_1, ΔH_2, n_1, n_2, r_1, and r_2) were allowed to float. The x-extension of the $\alpha(x)$-function was set to 0.1 as it corresponds approximately to the range where the differential power exhibits long relaxations around $x = n_2$ due to the slow kinetics of the coacervate phase formation (8); for our experiments this was usually over 1 - 3 injections.

NMR spectroscopy

NMR measurements were made on ^{15}N-labelled proteins (~ 0.5 mM) in 10 mM sodium phosphate pH 6.0 containing 10% ^2H_2O and either 0 or 150 mM NaCl, as indicated. Experiments were recorded at 5 ºC or 25 ºC on Bruker DRX500 or 600 spectrometers. Data were processed using AZARA (v. 2.7, © 1993-2018; Wayne Boucher and Department of Biochemistry, University of Cambridge). Assignments were made using CcpNmr Analysis v. 2.4 (9). Chemical shifts were referenced to 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS). Backbone ^{15}N, $^1H^N$ assignments were primarily derived from HNN and HN(C)N experiments (10), which connect three sequential residues *via* (H^N_i, N_i, $N_{i–1}$), (H^N_i, N_i, N_i) and (H^N_i, N_i, N_{i+1}) cross-peaks in 3D. (There are two main benefits to such an approach for disordered proteins, particularly those with low sequence complexity. One is that by recording two ^{15}N dimensions, the superior spectral dispersion of the ^{15}N nucleus relative to C^α and C^β can be fully exploited. The other is that by making sequential connections using triplets of three heteronuclear frequencies, much of the ambiguity that would otherwise be present with the usual two is removed.) Once the backbone H^N and N nuclei were assigned, the remaining C^α, C^β, C' and H^α assignments could be obtained using a conventional triple-resonance approach (HNCA, HN(CO)CA, HNCO, HNCACB, HN(CO)CACB) alongside TOCSY-^{15}N-HSQC (11). Chemical-shift differences were calculated using $\Delta\delta = [(\Delta\delta^H)^2 + (0.15 \times \Delta\delta^N)^2]^{1/2}$ (12). Heteronuclear NOE values were obtained at 600 MHz with either 4 s of ^{1}H saturation using a 120º pulse train or a 4 s delay prior to the first ^{15}N pulse (13). Diffusion was measured using

the 'BPP-LED' stimulated echo sequence (14); the diffusion coefficient (D) was obtained from the slope of an xy plot: $-\ln(I_2/I_1) = (\Delta - \delta/3 - \tau/2)\,\gamma^2\,\delta^2\,(G_2^2 - G_1^2)\,D$, where Δ = diffusion delay, δ = gradient length, τ = bipolar gradient separation, γ = gyromagnetic ratio and G_1 & G_2 = gradient amplitudes in two separate experiments leading to signal intensities I_1 & I_2. For the experiments described here, Δ = 500 ms, δ = 2 ms, τ = 200 μs, G_1 = 65 G cm^{-1} and G_2 was chosen to give a ca. 50% drop in intensity.

Far-UV CD spectroscopy

Spectra of CH1 and CH1-P and their complexes with 20-bp DNA were acquired at ca. 0.2 mg/ml over a 190-350 nm range at 25 ºC in 10 mM sodium phosphate pH 6.0 with 0 or 150 mM NaF (Cl^- absorbs strongly < 195 nm) as stated. For preparation of the complexes, 70 μM CH1 or CH1-P was titrated into 15 μM DNA to reach the required ratio. Spectra were acquired using an AVIV 410 spectrometer, in 1 mm path-length cuvettes and in 1 nm wavelength steps, averaged over three accumulations and baseline-corrected using buffer before smoothing, using the manufacturer's software. Millidegree units were converted to mean residue (amino acid or nucleotide) ellipticities (MRE) with units 10^{-3} deg cm^2 dmol^{-1} res^{-1} using MRE = millideg. / {(no. residues – 1) × c × l × 10}, where c = molar concentration and l = path length in cm.

Analytical ultracentrifugation

Sedimentation velocity (SV) was measured in an Optima XL-I (Beckman Coulter) centrifuge using an An60 Ti eight-hole rotor. Standard 12 mm double-sector Epon centrepieces with sapphire windows contained 400 μl of DNA at 3 or 12 μM, alone and with CH1 or CH1-P at a protein:DNA molar ratio of 0.5:1 or 1:1, at low I. Absorbance (λ = 260 nm) and interference data were acquired at 50 krpm, at 20 ºC, with a scan interval of ca. 540 s. The density and viscosity of the buffer (ρ = 0.99914 g/ml; η = 0.0010048 P) and the partial specific volume of the protein ($\bar{\nu}$ = 0.78 ml/g) were calculated using Sednterp (15). The partial specific volume of the DNA ($\bar{\nu}$ = 0.58 ml/g) was taken from tabulated values (16). Multi-component sedimentation coefficient $c(s)$ distributions were obtained from the first 50 scans (collected

over 7.5 h) by direct boundary modelling of the Lamm equation using Sedfit v.14.1 (17). The composition of the complex could be calculated by comparison of the UV and interferometry $c(s)$ distribution data: the DNA content of the peaks was first estimated by integration of the absorbance signals in Sedfit (the protein has no aromatic residues, hence no absorbance at 260 nm). The DNA contribution to the interferometry signals was calculated using the refractive index increment for DNA (dn/dc = 0.168 ml/g; (18)). The refractive index increment for CH1 (dn/dc = 0.173 ml/g) was calculated from the sequence using Sedfit (17); dn/dc is therefore approximately the same for both protein and DNA. The number of interference fringes per mg/ml was calculated as 3.0 using $Y = c \times dn/dc \times l/\lambda$, where Y = fringe displacement in units of fringes, c = concentration, dn/dc = refractive increment, l = path length (12 mm) and λ = laser wavelength (675 nm) (19). In the case of the complex, subtracting the contribution of the DNA from the total revealed the contribution from the protein. For the purposes of mass estimations in Sedfit, the partial specific volume of the complex was assumed to be the appropriately mass-weighted average.

Light microscopy

Samples were prepared by gradually adding 70 μM CH1 or CH1-P into 70 μM DNA solution to a final molar ratio of 1:1 over the course of an hour at room temperature. Bright field phase-contrast images were taken with a Nikon Eclipse Ti system at 20X magnification, and post-processed with Nikon NIS-Elements Advanced Research software suite.

REFERENCES

1. Cato L, Stott K, Watson M, Thomas JO (2008) The interaction of HMGB1 and linker histones occurs through their acidic and basic tails. *J Mol Biol* 384(5):1262–72.
2. Thomas JO, Kornberg R (1978) The study of histone-histone associations by chemical cross-linking. *Methods Cell Biol* 18:429–40.
3. Welburn J, Endicott J (2005) Methods for preparation of proteins and protein complexes that regulate the eukaryotic cell cycle for structural studies. *Methods Mol*

Biol 296:219–235.

4. Du JT, et al. (2005) Low-barrier hydrogen bond between phosphate and the amide group in phosphopeptide. *J Am Chem Soc* 127(47):16350–16351.
5. Theillet FX, et al. (2012) Cell signaling, post-translational protein modifications and NMR spectroscopy. *J Biomol NMR* 54(3):217–236.
6. Machha VR, et al. (2014) Exploring the energetics of histone H1.1 and H1.4 duplex DNA interactions. *Biophys Chem* 185:32–38.
7. Kim W, Yamasaki Y, Kataoka K (2006) Development of a fitting model suitable for the isothermal titration calorimetric curve of DNA with cationic ligands. *J Phys Chem B* 110(22):10919–10925.
8. Vitorazi L, et al. (2014) Evidence of a two-step process and pathway dependency in the thermodynamics of poly(diallyldimethylammonium chloride)/poly(sodium acrylate) complexation. *Soft Matter* 10:9496–9505.
9. Vranken WF, et al. (2005) The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins Struct Funct Genet* 59(4):687–696.
10. Panchal SC, Bhavesh NS, Hosur R V (2001) Improved 3D triple resonance experiments, HNN and HN(C)N, for H^N and ^{15}N sequential correlations in (^{13}C, ^{15}N) labeled proteins : Application to unfolded proteins. *J Biomol NMR* 20(C):135–147.
11. Cavanagh J, Skelton NJ, Fairbrother WJ, Rance M, Palmer AGI (2006) *Protein NMR Spectroscopy Principles and Practice* (Academic Press). 2nd Ed.
12. Zuiderweg ERP (2002) Mapping protein-protein interactions in solution by NMR spectroscopy. *Biochemistry* 41(1):1–7.
13. Farrow NA, et al. (1994) Backbone dynamics of a free and a phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. *Biochemistry* 33(19):5984–6003.
14. Wu DH, Chen A, Johnson CS (1995) An improved diffusion-ordered spectroscopy experiment incorporating bipolar-gradient pulses. *J Magn Reson Ser A* 115(2):260–264.
15. Laue TM, Shah BD, Ridgeway TM, Pelletier SL (1992) Computer-aided interpretation of analytical sedimentation data for proteins. *Anal ultracentrifugation Biochem Polym*

Sci:90–125.

16. Ralston G (1993) *Introduction to analytical ultracentrifugation* (Beckman Coulter Life Sciences).

17. Schuck P (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys J* 78(3):1606–1619.

18. Thiesen A, Johann C, Deacon M, Harding SE (2000) *Refractive increment data-book for polymer and biomolecular scientists* (Nottingham University Press, Nottingham, UK).

19. Laue TM (1996) Choosing which optical system of the Optima^{TM} XL-I analytical ultracentrifuge to use. *Beckman Coulter Tech Rep* A-1821-A:1–18.