Molecular recognition of the nucleosomal "supergroove"

Rajeswari S. Edayathumangalam*, Philipp Weyermann†, Joel M. Gottesfeld‡, Peter B. Dervan†§, and Karolin Luger*§

*Department of Biochemistry and Molecular Biology, Colorado State University, Fort Collins, CO 80523-1870; †Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125; and ‡Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037

Contributed by Peter B. Dervan, March 16, 2004

Chromatin is the physiological substrate in all processes involving eukaryotic DNA. By organizing 147 base pairs of DNA into two tight superhelical coils, the nucleosome generates an architecture where DNA regions that are 80 base pairs apart on linear DNA are brought into close proximity, resulting in the formation of DNA "supergrooves." Here, we report the design of a hairpin polyamide dimer that targets one such supergroove. The 2-Å crystal structure of the nucleosome–polyamide complex shows that the bivalent "clamp" effectively crosslinks the two gyres of the DNA superhelix, improves positioning of the DNA on the histone octamer, and stabilizes the nucleosome against dissociation. Our findings identify nucleosomal supergrooves as platforms for molecular recognition of condensed eukaryotic DNA. *In vivo*, supergrooves may foster synergistic protein–protein interactions by bringing two regulatory elements into juxtaposition. Because supergroove formation is independent of the translational position of the DNA on the histone octamer, accurate nucleosome positioning over regulatory elements is not required for supergroove participation in eukaryotic gene regulation.

Recent high-resolution structures of nucleosome core particles (NCPs) (1, 2) provide a wealth of details on the intricacies of protein–DNA interactions and provide a view of eukaryotic DNA in a physiologically relevant context (refs. 1–3; reviewed in ref. 4). Two features are of particular relevance. First, due to extensive interactions of the DNA with the core framework of the histone octamer, one face of the DNA is completely precluded from interactions with sequence-specific DNA binding factors over its entire 147-bp length. Although ≈70% of nucleosomal DNA is solvent accessible, transcription factors can access only six to eight consecutive base pairs before being sterically blocked by the histones and by extreme fluctuations in groove width. Thus, a single nucleosome limits access of transcription factors and other cellular factors to the underlying nucleosomal DNA. Second, the modulation of helical parameters within the left-handed DNA superhelix brings the minor (and major) grooves between the two superhelical gyres into alignment throughout almost the entire nucleosome (Fig. 1 *A* and *B*). This alignment creates seven minor and six major DNA "supergrooves" in which 14–16 bp of DNA would be accessible for site-specific interaction with a single DNA binding ligand. Each supergroove brings sequence elements that are 80 bp apart in linear DNA into close structural proximity (5). Supergrooves thus provide unique recognition platforms that occur only within the structural context of a nucleosome.

Pyrrole-imidazole polyamides (PAs) bind the DNA minor groove with high affinity and specificity and have been used to modulate gene expression (6). We previously characterized the binding of a small library of eight-ring hairpin PAs to several discrete 6-bp sites on nucleosomes containing 146-bp DNA (NCP146) (7), and determined the crystal structures of NCP146 in complex with three of these hairpin PAs (8). Here, we report the design of hairpin PA dimers that target two noncontiguous minor groove sites in the nucleosomal supergroove in a sequence-specific manner, thereby "clamping" the two gyres of nucleosomal DNA. Our aim was to generate a sequence-specific bivalent PA clamp that would expand the total size of DNA recognizable in a nucleosomal context, and that could stabilize the histone–DNA complex against dissociation. By using x-ray crystallography, we demonstrate that the clamp binds to NCP146 across the supergroove as designed, and with very high affinity and specificity. In addition, biochemical assays reveal that the clamp has a dramatic effect on the *in vitro* stability of nucleosomes against dilution-induced dissociation. We propose that nucleosomal supergrooves are molecular interaction platforms that can be exploited for synergistic protein–protein interactions, and that supergrooves add an additional level of complexity to eukaryotic gene regulation.

Materials and Methods

Cocrystallization of NCP–PA Complexes. Previously established protocols were used to reconstitute NCP146 from recombinant *Xenopus laevis* histones and a 146-bp DNA fragment derived from human α-satellite DNA (1, 9, 10). PAs were synthesized as described (11, 12). The purity and identity of the clamps were established by analytical HPLC and matrix-assisted laser desorption ionization-time-of-flight mass spectroscopy. PAs were purified by preparative HPLC, lyophilized, and stored at 4°C. Each PA was freshly dissolved in 20 mM potassium cacodylate (pH 6.0) and 1 mM EDTA, and its concentration was determined by measuring UV absorbance by using empirically determined extinction coefficients (ε_{316} = 68,800 for PA1; ε_{314} = 139,000 for PW12, PW13, and PW14). NCP146 (at 40–50 μM) was incubated with a 10-fold molar excess of clamp in solution at ambient temperature for 45 min. The integrity of the nucleosome preparation after incubation with the ligand was checked by electrophoretic mobility-shift assay (9). Crystals of NCP146–clamp complexes were grown 1–2 weeks in 40–45 mM $MnCl_2$, 35–38 mM KCl, and 20 mM potassium cacodylate (pH 6.0) containing ≈20 μM (≈4 mg·ml^{-1}) NCP146 at 19°C by using vapor diffusion. Crystals were harvested and flash cooled as described (2).

Data Collection, Structure Refinement, and Validation. X-ray data were collected at beamline 5.0.2 at the Advanced Light Source in Berkeley, CA. Data for each of the individual structures were collected on single crystals and processed with DENZO and SCALEPACK (13). Molecular replacement (using Protein Data Bank ID code 1AOI as the original search model) and refinement were done with CNS (14), and model building with O (15). Each model was checked by using simulated annealing omit maps during early stages of model building. PAs were clearly visible in the original difference density maps calculated after molecular replacement. An initial model for the PA was gener-

Abbreviations: NCP, nucleosome core particle; PA, polyamide.

Data deposition: The atomic coordinates and structure factors for NCP146–PW**12** have been deposited in the Protein Data Bank, www.pdb.org (PDB ID code 1S32).

§To whom correspondence may be addressed. E-mail: kluger@lamar.colostate.edu or dervan@caltech.edu.

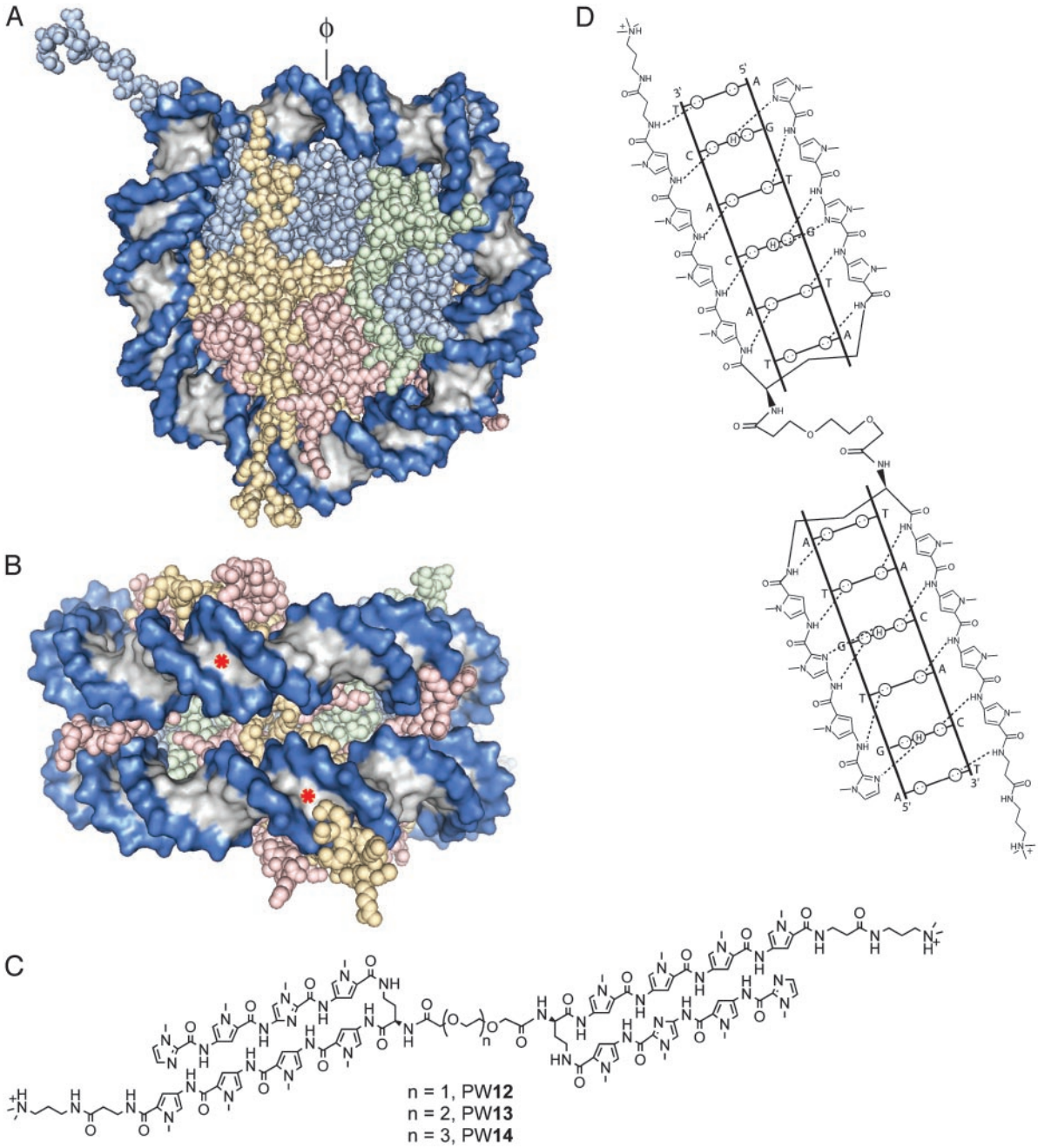


Fig. 1. Site-specific recognition of nucleosomal DNA by clamp PAs. (*A*) NCP146 structure (PDB ID code 1AOI, ref. 2) viewed with the superhelical axis along the *z* axis. The particle pseudo-two-fold dyad axis (ϕ) is shown for orientation. DNA (blue and white) and associated histone proteins (H2A, yellow; H2B, red; H3, blue; H4, green) are shown in sphere or surface representation. (*B*) Supergrooves in NCP146. Shown is a different view of NCP146 with the superhelical axis along the *y* axis. Color scheme is the same as in *A*. One of the DNA supergrooves is indicated by two asterisks. (*C*) Chemical structures of clamp PAs, PW**12** to -**14**. (*D*) Hydrogen bonding model of PW**12** to its target DNA site. Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines. Circles containing H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dotted lines.

ated by using (*i*) fragments of previously published PA structures (8), (*ii*) information from the Uppsala Software Factory HIC-Up server (16), and (*iii*) Engh and Huber stereochemical parameters (17). The PA was refined along with the rest of the model. The geometry of the final model is excellent (see Table 1), with 93.7% of the residues in the most-favored regions, 5.6% in additional allowed regions, 0.7% in the generously allowed regions, and no residues in the disallowed regions of the Ramachandran plot. The refined structures were compared with unliganded nucleosome structures by using the LSQMAN program from the Uppsala Software Factory (18). Several figures in the paper were made by using the molecular graphics program PYMOL [W. L. DeLano, PYMOL Molecular Graphics System (2002), www.pymol.org].

Nucleosome Dilution Assay. The nucleosome dilution experiments were performed as described (19). Briefly, NCP146 (reconstituted at a molar ratio of 0.6 histone octamer:DNA) was subjected to serial dilution from an initial concentration of 10 nM in steps of 3.3, 1.1, 0.34, 0.11, 0.035, 0.01, and 0.004 nM, in a

Table 1. Crystallographic statistics for data set NCP146–PW12*

Data collection	
Number of crystals merged	1
Space group	$P2_12_12_1$
Unit cell dimensions, Å	
a	105.4
b	109.7
c	181.8
Resolution range, Å	2.05–100
Unique reflections	124,523
Redundancy	4.75
Completeness (%)	94.4 (54.1)
$I/\sigma(I)$	22.99 (2.04)
R_{merge}†	0.076 (0.353)
Refinement statistics	
Reflections in test set	6050
R_{cryst}‡	0.2194
R_{free}§	0.2434
No. of amino acids	779¶
No. of DNA base pairs	292
No. of water molecules	890
No. of Mn^{2+} ions	14
No. of Cl^- ions	4
No. of polyamides	1
rmsd‖	
Bonds, Å	0.008
Angles, °	1.099
Average B-factors	
Protein	48.5
DNA	98.9
Polyamide	143.7
Solvent	65.1

Numerical values in parentheses are for highest resolution shell.
*Data collected at Advanced Light Source beamline 5.0.2.
†$R_{merge} = \Sigma|I_h - \langle I_h\rangle|/\Sigma I_h$, where $\langle I_h\rangle$ is the mean of the measurements for a single *hkl*.
‡$R_{cryst} = \Sigma|F_{obs} - F_{calc}|/\Sigma F_{obs}$.
§Calculated using 5% of the *hld* data chosen randomly and omitted from the start of refinement.
¶Protein structure also includes a second alternate side chain conformation for 16 histone residues.
‖rmsd, rms deviation from ideal geometry.

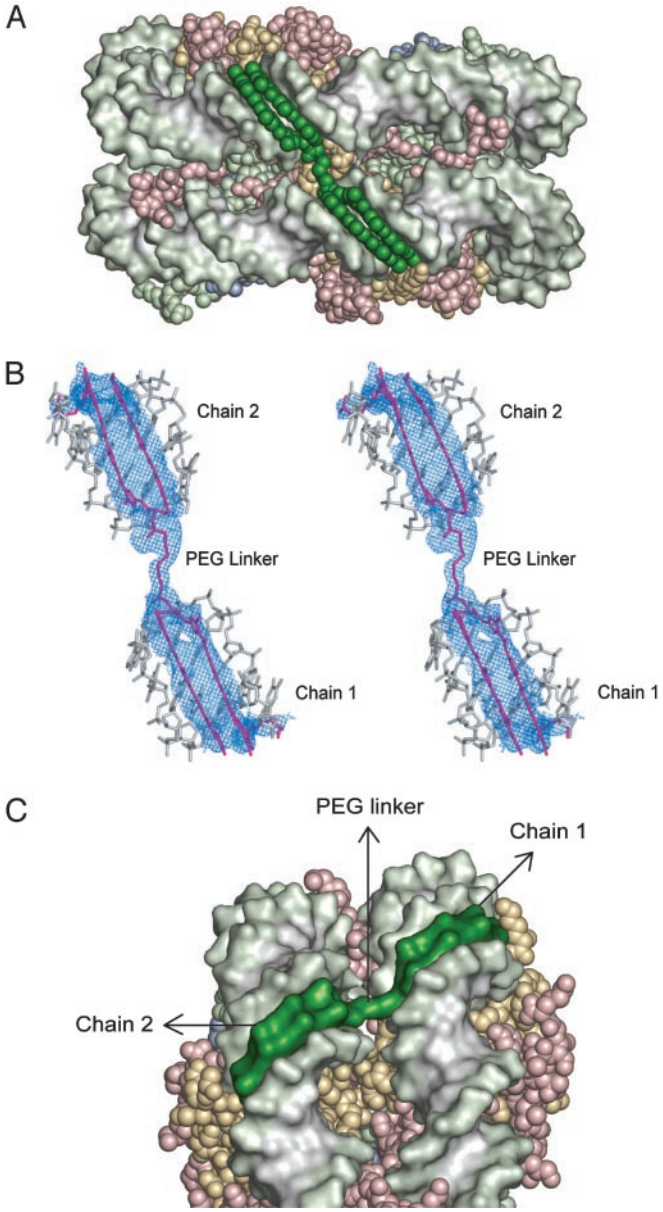


Fig. 2. Clamp binding in the NCP146–PW**12** complex. Color code for histone proteins is as in Fig. 1. DNA is shown in green and white, PW**12** in green or magenta. (*A*) Overview of NCP146–PW**12** structure, orientation same as in Fig. 1*B*. (B) Stereoview of PW**12** bound to its target DNA site. Omit density for the clamp is shown at 2 σ contour level. Chains 1 and 2 denote the two hairpin moieties. (*C*) The linker in the clamp is buried between the two gyres of superhelical DNA. A close-up surface representation of the NCP146–PW**12** structure is shown.

buffer containing 200 mM NaCl, 10 mM Tris·Cl (pH 7.5), 1 mM EDTA, 0.05% Nonidet P-40, and 10% (vol/vol) glycerol. After incubation at 37°C for 1 h, the samples were subjected to nondenaturing gel electrophoresis, followed by PhosphorImager (Molecular Dynamics) analysis of the gel.

Results and Discussion

Bivalent Clamp Binds to a Nucleosomal Supergroove. Three bivalent hairpin PA dimers (PW**12**, PW**13**, and PW**14**) were synthesized. A previously characterized eight-ring hairpin, whose binding sites were separated by 80 bp on the palindromic α-satellite DNA fragment used for previous crystallographic studies (PA**1**) (7, 8), was bridged in a head-to-head fashion with one, two, or three ethylene glycol units (resulting in distances of ≈11–18 Å between the two hairpins; Fig. 1*C*) to create the bivalent clamps. The predicted hydrogen-bonding pattern of one of the clamps, PW**12**, to its target DNA sequence is shown in Fig. 1*D*. The NCP146–clamp complexes have the same electrophoretic mobility as unliganded NCP146 and do not show signs of dissociation or aggregation upon binding the clamp (data not shown). Quantitative DNase I footprint titrations (20) (data not shown) demonstrate that the clamps bind NCP146 with affinities in the low nanomolar range. Nonspecific binding of the clamp to NCP146 was not observed.

To obtain unambiguous evidence for the mode of interaction of the clamp with nucleosomes, we crystallized NCP146 in complex with each of the three clamps (NCP146–PW**12**, NCP146–PW**13**, and NCP146–PW**14**). Because the biophysical and structural parameters were very similar for the three co-crystal structures, only NCP146–PW**12** will be further discussed. Clamp binding increased the size, order, and resolution of NCP146 crystals. High quality x-ray data to 2-Å resolution were obtained for NCP146–PW**12** from a single crystal, without detectable radiation damage (Table 1). In contrast, nucleosome

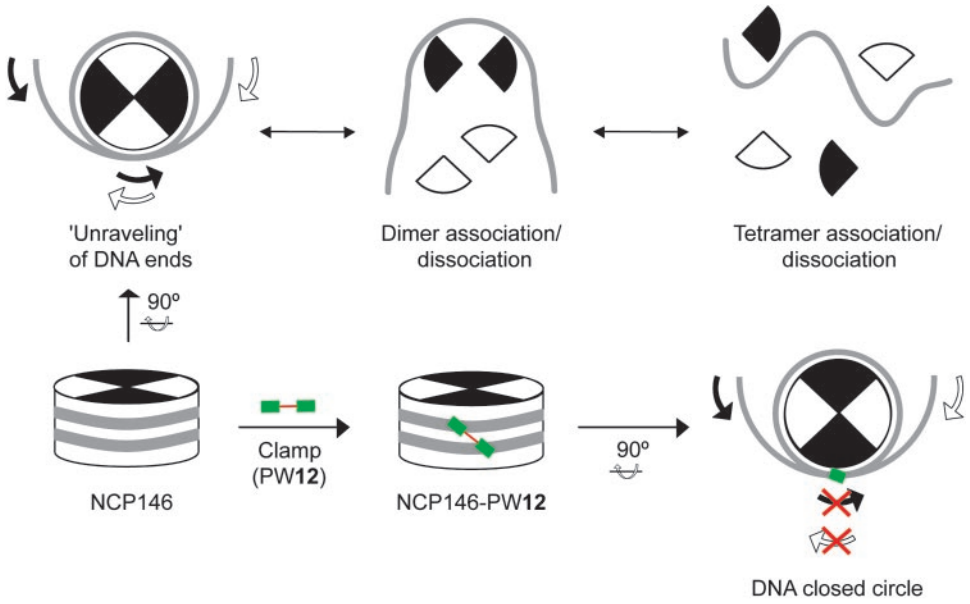


Fig. 3. Schematic illustration of the predicted effect of PA clamps on nucleosome dissociation. In the absence of ligand binding, nucleosome dissociation initiates with unraveling of the DNA ends, followed by dissociation of the (H2A-H2B) dimers, and finally by the dissociation of the $(H3\text{-}H4)_2$ tetramer. Binding clamp to the nucleosomes leads to the formation of a closed ≈80-bp DNA supercoil that prevents further disassembly of the nucleosomes.

crystals without bound ligand are highly sensitive to radiation damage, and the recently reported nucleosome structures at similarly high resolution required merging data from as many as 44 crystals (1). Because the diffraction qualities of nucleosome crystals depend on the positioning of the DNA with respect to the histone octamer with single bp accuracy (3, 21), our results strongly suggest that the clamp improves the accuracy of DNA positioning with respect to the histone octamer.

We refined the NCP146–PW**12** structure to a resolution of 2 Å (Table 1 and Fig. 2*A*). Electron difference density for PA moieties and linker was clearly visible after molecular replacement and during subsequent refinement cycles (Fig. 2*B*). The ethylene glycol linker is buried between the two gyres of the superhelix (Fig. 2*C*). Both subunits of the bivalent hairpin ligand bind to their target sites in accordance with previously established pairing rules (6) (Fig. 1*D*). Thus, the crystal structure reveals that the clamp indeed targets the supergroove, and that a designed ligand can exploit the special geometry of a DNA supergroove in the nucleosome for high-affinity and high-specificity binding. The structure of the nucleosome and of the hairpin PAs remains virtually unchanged compared with that of NCP146 in complex with the parent PA (PA**1**) (8), demonstrating that no additional structural distortions are imparted to the nucleosome upon clamp binding.

Clamp Binding to a Supergroove Prevents Nucleosome Dissociation. Nucleosome dissociation is likely to initiate by the unraveling of the DNA ends from the surface of the histone octamer, followed by the dissociation of one or both (H2A-H2B) dimers (Fig. 3). Furthermore, unwrapping of the ends of nucleosomal DNA by chromatin remodeling complexes such as SWI/SNF (22) and Swr1 (23) is thought to expose the DNA-binding surface of the (H2A-H2B) dimer and initiate dissociation and/or exchange of the (H2A-H2B) dimer. The PA clamp may counteract dissociation because it effectively cross-braces the two gyres of nucleosomal DNA and forms a closed DNA circle around the histone octamer (Fig. 3). Although partial unraveling of the DNA ends may occur in the presence of the clamp, complete dissociation of the DNA is likely to be precluded beyond the clamp binding sites (Fig. 3), which would in turn stabilize the nucleosome from dissociation.

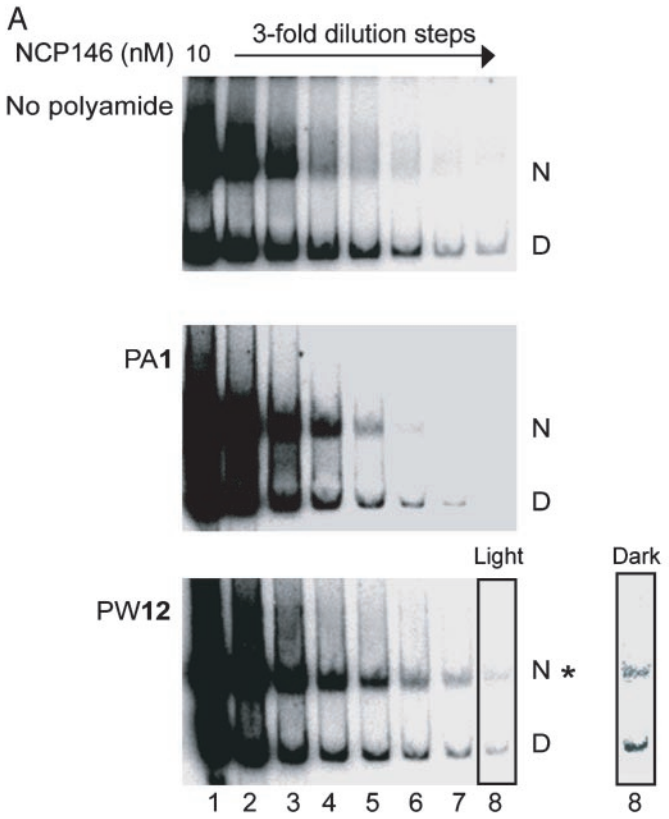
The effect of clamp binding on NCP146 stability was investigated by using a simple dilution assay (19, 24, 25). This method allows us to compare the kinetic stability of NCP146 in the presence and absence of the clamp. We find that a saturating concentration of PW**12** dramatically stabilizes NCP146 against dissociation into free DNA and histones (Fig. 4 *A* and *B*). In the absence of PA, we observed 50% dissociation at a NCP146 concentration of 0.33 nM. In the presence of the unlinked parent PA (PA**1**), 50% dissociation is observed at ≈9-fold lower concentration of NCP146 (0.038 nM). In contrast, dissociation is minimal for NCP146–PW**12** (compare lane 8 in the three panels in Fig. 4*A*, and see Fig. 4*B*). The off-rate for PW**12** on NCP146 is increased almost by an order of magnitude relative to that of PA**1** on NCP146 whereas the equilibrium affinities of both the clamp and PA**1** for NCP146 are similar (Table 2). Thus, it seems that the kinetic stability (k_{off}) of the clamp on NCP146 is significantly higher than that of PA**1** on NCP146. These results are also consistent with our observation that NCP146–clamp complexes form ordered and highly diffracting crystals. PA clamps could be potent reagents to determine whether chromatin remodelers transiently unravel DNA as a step en route to histone exchange. Furthermore, this class of molecules might be used to

Table 2. Dissociation rate constants for PA1 and PW12 with NCP146

Polyamide	PA**1**	PW**12**
K_a* (M^{-1})	1.1 (± 0.1) × 10^9	0.5 (± 0.05) × 10^9
Half-life, min†	15	180
k_{off}‡ (s^{-1})	7.7 × 10^{-4}	6.4 × 10^{-5}
k_{on}§ (M^{-1} s^{-1})	8.5 × 10^5	3.2 × 10^4

*Determined by quantitative DNase I footprinting (data not shown). Values in parentheses denote the SDs. Values reported here are averages and SDs of four independent determinations.
†Determined by competition footprinting using unlabeled 146-bp DNA as a competitor (data not shown).
‡Calculated according to $k_{off} = 0.693/t_{1/2}$.
§Calculated from the equation $k_{on} = (K_a)(k_{off})$.

BIOCHEMISTRY



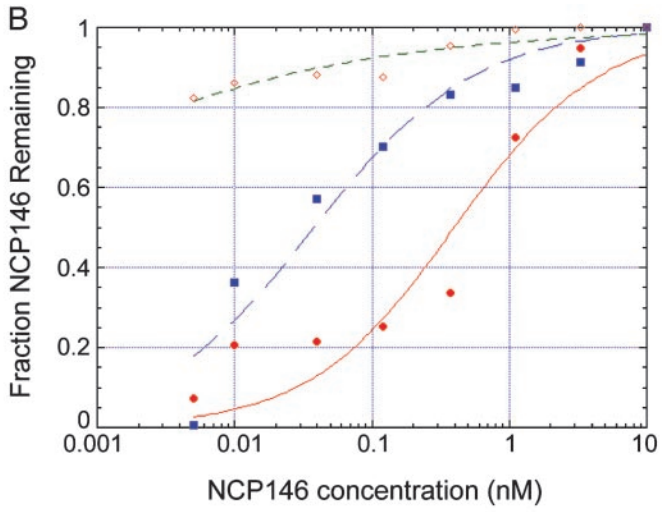


Fig. 4. Stabilization of NCP146 by the clamp. (*A*) Nondenaturing gel analysis. NCP146 was subjected to serial dilution in a dilution buffer that contained either no PA (*Top*), or 100 nM PA1 (*Middle*) or PW**12** (*Bottom*), as indicated. NCP146 samples were then subjected to electrophoresis on a 6% nondenaturing polyacrylamide gel, and a PhosphorImager of the gel is shown. N, NCP146; D, 146-bp DNA; A * denotes the stability of the NCP146 upon dilution in the presence of PW**12**. The side panel shows a darker contrast of lane 8 in the *Bottom*. (*B*) Graphical analysis of the dissociation data shown in *A*. The fraction of DNA remaining in NCP146, relative to the 10-nM sample, is plotted vs. NCP146 concentration. ●, no PA; ■, PA1; ◇, PW**12**.

explore and modulate many aspects of chromatin function: for example, to improve nucleosome positioning and stability on poorly positioned nucleosomal templates, to study the role of nucleosome dissociation and dynamics during transcription, replication, and chromatin remodeling, or to target reagents or recruit proteins specifically to particular nucleosomes.

Nucleosomal Supergrooves Are Platforms of Molecular Recognition. Here, we show *in vitro* that DNA architecture embedded within the nucleosome generates interaction platforms or supergrooves that can be targeted by sequence-specific bivalent DNA binding ligands to mediate short- and medium-range interactions, and to modulate nucleosome stability. *In vivo*, nucleosomal supergrooves may mediate medium- and long-range DNA interactions. Indeed, there is evidence that such interactions may regulate transcription of several genes. For example, on the mouse mammary tumor virus long terminal repeat, the precise spacing between two glucocorticoid receptor recognition elements brings about nucleosome-mediated synergistic transcriptional activation (ref. 26 and references therein). In several other genes, a positioned nucleosome brings two precisely spaced regulatory elements into juxtaposition; examples of this are the *X. laevis* vitellogenin B1 and TRβA genes, the *Drosophila* Adh, hsp26, and hsp27 genes, and the human U6 gene (refs. 26 and 27 and references therein). We found numerous examples of tandem or multiple repeat regulatory sequences that are arranged with a regular spacing of ≈80 bp. Such spacing may have evolved to promote crosstalk between factors bound to noncontiguous regulatory elements in the context of chromatin. Supergrooves occur all along the nucleosomal DNA (seven minor and six major supergrooves per nucleosome), and their occurrence requires only that the two DNA segments be separated by ≈80 bp and be within the context of a folded nucleosome. Because they do not require precise translational nucleosome positioning, the presence of nucleosomes over such regulatory elements, and their role in transcription regulation, may often go undetected, and synergistic interactions between regulatory factors may be much more frequent than previously assumed.

We thank Gerry McDermott and Keith Henderson at the Advanced Light Source (Berkeley, CA) for help with data collection; Stephanie Ebbesen and Pamela Dyer for help with reagents; and Laurie Stargell and Clara Kielkopf for helpful comments. This work was supported by a grant supplement from the National Institutes of Health (to K.L., J.M.G., and P.B.D.) and by a postdoctoral fellowship from the Swiss National Science Foundation (to P.W.).

1. Davey, C. A., Sargent, D. F., Luger, K., Maeder, A. W. & Richmond, T. J. (2002) *J. Mol. Biol.* **319,** 1097–1113.
2. Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. (1997) *Nature* **389,** 251–260.
3. Richmond, T. J. & Davey, C. A. (2003) *Nature* **423,** 145–150.
4. Luger, K. (2003) *Curr. Opin. Genet. Dev.* **13,** 127–135.
5. Srinivasan, A. R. & Olson, W. K. (1992) *Biophys. Chem.* **43,** 279–310.
6. Dervan, P. B. & Edelson, B. S. (2003) *Curr. Opin. Struct. Biol.* **13,** 284–299.
7. Gottesfeld, J. M., Melander, C., Suto, R. K., Raviol, H., Luger, K. & Dervan, P. B. (2001) *J. Mol. Biol.* **309,** 615–629.
8. Suto, R. K., Edayathumangalam, R. S., White, C. L., Melander, C., Gottesfeld, J. M., Dervan, P. B. & Luger, K. (2003) *J. Mol. Biol.* **326,** 371–380.
9. Luger, K., Rechsteiner, T. J. & Richmond, T. J. (1999) *Methods Enzymol.* **304,** 3–19.
10. Dyer, P. N., Edayathumangalam, R. S., White, C. L., Bao, Y., Chakravarthy, S., Muthurajan, U. M. & Luger, K. (2004) *Methods Enzymol.* **375,** 23–44.
11. Baird, E. E. & Dervan, P. B. (1996) *J. Am. Chem. Soc.* **118,** 6141–6146.
12. Weyermann, P. & Dervan, P. B. (2002) *J. Am. Chem. Soc.* **124,** 6872–6878.
13. Otwinowski, Z. & Minor, W. (1997) *Processing of X-Ray Diffraction Data Collected in Oscillation Mode* (Academic, New York).
14. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., *et al.* (1998) *Acta Crystallogr. D Biol. Crystallogr.* **54,** 905–921.

15. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjelgaard, M. (1991) *Acta Crystallogr. A* **47,** 110–119.
16. Kleywegt, G. J. & Jones, T. A. (1998) *Acta Crystallogr. D Biol. Crystallogr.* **54,** 1119–1131.
17. Engh, R. A. & Huber, R. (1991) *Acta Crystallogr. A* **47,** 392–400.
18. Kleywegt, G. J. (1996) *Acta Crystallogr. D* **52,** 842–857.
19. Gottesfeld, J. M. & Luger, K. (2001) *Biochemistry* **40,** 10927–10933.
20. Trauger, J. W. & Dervan, P. B. (2001) *Methods Enzymol.* **340,** 450–466.
21. Muthurajan, U. M., Park, Y. J., Edayathumangalam, R. S., Suto, R. K., Chakravarthy, S., Dyer, P. N. & Luger, K. (2003) *Biopolymers* **68,** 547–556.
22. Kassabov, S. R., Zhang, B., Persinger, J. & Bartholomew, B. (2003) *Mol. Cell* **11,** 391–403.
23. Mizuguchi, G., Shen, X., Landry, J., Wu, W. H., Sen, S. & Wu, C. (2004) *Science* **303,** 343–348.
24. Ausio, J., Seger, D. & Eisenberg, H. (1984) *J. Mol. Biol.* **176,** 77–104.
25. Cotton, R. W. & Hamkalo, B. A. (1981) *Nucleic Acids Res.* **9,** 445–457.
26. Collingwood, T. N., Urnov, F. D. & Wolffe, A. P. (1999) *J. Mol. Endocrinol.* **23,** 255–275.
27. Stunkel, W., Kober, I. & Seifart, K. H. (1997) *Mol. Cell. Biol.* **17,** 4397–4405.