Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 11327–11331, December 1993
Biochemistry

A peptide motif that recognizes A·T tracts in DNA

(yeast/*DAT1*/DNA binding motif)

BRIAN J. REARDON, R. SCOTT WINTERS, DONNA GORDON, AND EDWARD WINTER†

Department of Biochemistry and Molecular Biology and The Jefferson Institute of Molecular Medicine, Thomas Jefferson University, Philadelphia, PA 19107

Communicated by Darwin J. Prockop, August 30, 1993 (received for review July 2, 1993)

ABSTRACT **The *DAT1* gene of *Saccharomyces cerevisiae* encodes a DNA binding protein that specifically interacts with nonalternating oligo(A)·oligo(T) tracts (A·T tracts). Deletion analysis of *DAT1* coding information showed that the amino-terminal 36 residues are sufficient for specific DNA binding activity. Furthermore, a 35-residue synthetic peptide corresponding to amino acids 2–36 bound to A·T tracts with an equilibrium dissociation constant of 4×10^{-10} M. Within this region the pentad Gly-Arg-Lys-Pro-Gly is repeated three times. Mutational analysis revealed that the Arg side chains are required for high-affinity binding, whereas the other pentad side chains are dispensable. Chemical interference experiments showed that the DAT1 protein interacts with the minor groove of the double helix. The data suggest that the pentad arginines interact in a cooperative manner with a repeated minor groove feature of A·T tract DNA to achieve high-affinity recognition. Amino acid similarities with other DNA binding proteins suggest that the DAT1 protein pentad represents a specialized example of a widespread motif used by proteins to recognize A·T base pairs.**

In the yeast *Saccharomyces cerevisiae*, as in many other organisms, nonalternating oligo(A)·oligo(T) tracts (hereafter termed A·T tracts) are common intergenic elements. In yeast chromosome III, uninterrupted A·T tracts >10 bp occur on average once every 5 kbp and are found in the promoter regions of roughly 25% of yeast genes (defined for statistical purposes as 500 bp upstream of the initiator ATG) (1). A·T tracts have a number of unusual physical properties: they are involved in DNA bending; they have a melting temperature 10°C higher than alternating (A-T)·(A-T) tracts; they can preferentially exclude nucleosomes (for review, see ref. 2). Single-crystal x-ray analyses of oligonucleotide duplexes have shown that A·T tracts adopt a unique DNA structure characterized by a series of bifurcated hydrogen bonds, a narrow minor groove, and a high degree of propellor twist (2–4).

A·T tracts have been shown to promote transcription in yeast (5, 6). It is unclear whether the transcriptional activation results from the interaction of A·T tracts with specific transcription factors. Several differences in the transcriptional-activating properties between A·T tracts and other cis-acting transcriptional elements indicate that A·T tracts promote transcription through a distinct mechanism (6, 7). It has been proposed that A·T tracts promote transcription by causing local perturbations in chromatin structure that influence assembly of transcription complexes (5, 7).

One approach toward understanding the function of A·T tracts is to study the proteins that interact with them. We have previously identified a yeast protein that interacts with A·T tract DNA. This protein (previously referred to as "datin") has been purified and its gene (*DAT1*) has been cloned (8). *dat1* null mutants grow in rich and defined media, mate, and sporulate indistinguishably from congenic wild-type strains. We have been unable to detect any A·T tract-specific DNA binding activity in *dat1* deletion strains (E.W., unpublished results). Interestingly, the expression of many promoters that contain A·T tracts is higher in *dat1* mutants than in wild-type strains in stationary phase.

The amino acid sequence of the DAT1 protein (referred to hereafter as Dat1p) does not contain any obvious DNA binding motifs. To define and characterize its interaction with A·T tract DNA, a series of mutant Dat1p peptides were generated, and their DNA binding properties were examined. The data suggest a model in which high-affinity recognition is achieved through cooperative interactions between arginine residues in the Gly-Arg-Lys-Pro-Gly motifs found three times in the amino-terminal region of Dat1p and a repeated minor groove feature of A·T tract DNA.

MATERIALS AND METHODS

Plasmids. The starting plasmid used to construct the *DAT1* truncation plasmids was pDAT101, which contains the complete *DAT1* coding region under the control of the isopropyl β-D-thiogalactoside-inducible pKK233 promoter (8). The D-90 expressing plasmid (pD90) was generated from pDAT101 by digestion between *DAT1* codons 90 and 91 with restriction endonuclease *Sph* I, making the DNA ends flush with T4 DNA polymerase, and ligation of a duplex oligonucleotide linker that contained an *Xba* I restriction endonuclease site and stop translation codons in all reading frames (sequence, 5′-CTAGTCTAGACTAG-3′). The D-60- and D-22-expressing plasmids were generated by linearizing pD90 with *Xba* I and *Aat* II to create 5′ and 3′ protruding ends, respectively, followed by the unidirectional deletion from codon 90 toward the amino-terminal coding in formation with exonuclease III (9). The D-48- and D-36-expressing plasmids were generated from pD90 by inserting translational stop signals immediately after codons 48 and 36, respectively, by using an oligonucleotide-directed mutagenesis kit (Amersham) according to the manufacturer's instructions. Nucleotide sequences of the plasmids were determined from the initiator ATG to the terminator codons by the dideoxynucleotide chain-termination method (9). As a result of the deletion strategy, the D-60 and D-22 plasmid constructs encoded additional carboxyl-terminal extensions of 8 and 23 amino acid residues, respectively, that are not encoded by *DAT1*. Missense mutants in pD90 were generated using oligonucleotide-directed mutagenesis kits obtained from either Amersham or Clontech and mutagenized regions were sequenced using the dideoxynucleotide chain-termination method. pUC-19-A_{11} was constructed for use in the interference assays by cloning a duplex oligonucleotide [sequence, 5′-GCG$(A)_{11}$CGC-3′] into the *Sma* I restriction endonuclease site of pUC-19.

Band-Shift Assay. *Escherichia coli* harboring the indicated plasmids were grown to midlogarithmic phase and induced

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: Dat1p, DAT1 protein.

†To whom reprint requests should be addressed.

with isopropyl β-D-thiogalactoside, and extracts were prepared by the Triton/lysis method (9). Total lysate (200 ng) was analyzed by the band-shift assay as described (8) using 5–20 pg of the 138-bp ^{32}P-labeled Alu-A·T tract DNA probe that contains a 19-bp A·T tract. Nonspecific *E. coli* DNA (1 μg) and specific poly(A)·poly(T) competitor DNA (100 ng) were added to the reaction mixtures prior to the addition of protein, where indicated. To compare the binding activities of mutated to wild-type D-90 peptides, the ratios of bound to unbound probe were quantitated with a PhosphorImager (Molecular Dynamics). Differences in D-90 levels between extracts were normalized by measuring the D-90 immunoreactivity of immunoblots (see below) by laser densitometry (LKB Ultroscan XL). The standard deviation from standard curves was ±9%. The detection limit of the band-shift assay was <0.1% of the wild-type D-90 activity, as estimated from multiple autoradiographic exposures of control samples.

Major and Minor Groove Interference Assays. The plasmid pUC-A_{11} (see above) was digested with *Hin*dIII restriction endonuclease, radiolabeled with ^{32}P on the $(A)_{11}$ strand with the large fragment of DNA polymerase I, and digested with *Eco*RI to release the singly end-labeled 72-bp probe, which was isolated by polyacrylamide gel electrophoresis. The purified DNA probe was methylated with dimethyl sulfate (9) or carbethoxylated with diethyl pyrocarbonate (10). Scaled-up band-shift assays were performed under conditions of D-90 excess (20–40 nM D-90); the bound and unbound DNA were localized by autoradiography, excised, and eluted overnight in 0.5 M $NH_4C_2H_3O_2$/1 mM EDTA. After precipitation, DNA was cleaved at the site of modification with 1 M piperidine, before being electrophoresed through a denaturing 10% polyacrylamide gel and exposed to x-ray film (9).

Miscellaneous Assays and Procedures. For immunoblot analysis, the total protein lysate was electrophoretically resolved on a 15% polyacrylamide/SDS gel for 1400 V·H using an acrylamide/*N,N'*-methylenebisacrylamide weight ratio of 30:0.8. Protein was transferred to an Immobilon-P membrane (Millipore) and incubated with a rabbit polyclonal sera (sera 159) raised against a synthetic peptide corresponding to Dat1p residues 42–61 [Dat1p-(42–61)] and immunopurified using Dat1p-(42–61)-Affi-10 (Bio-Rad). Immunoblot incubations were performed as described (11) except that 20% (vol/vol) fetal calf serum was used as a nonspecific blocking agent. D-35 peptide (corresponding to residues 2–35) was synthesized with *t*-butoxy carbonyl chemistry using an Applied Biosystems model 430 peptide synthesizer and purified using C_{18} reverse-phase HPLC. D-90 was purified from extracts prepared from induced *E. coli* harboring pD90 by using phosphocellulose, poly(dA)·poly(dT)-agarose, and a Pharmacia FPLC Mono S column as described (8). Purified D-90 was >90% pure, as shown by laser densitometry (LKB Ultrascan XL) of Coomassie blue-stained polyacrylamide gels. The chromatographic properties of D-90 produced in *E. coli* were indistinguishable from the binding activity purified from yeast (8).

RESULTS

The 36 Amino-Terminal Dat1p Residues Are Sufficient for Specific DNA Binding. The DNA binding domain of the 248-residue Dat1p resides within its amino-terminal half since a proteolytic fragment from this region of the protein purified from yeast bound with high affinity and specificity to A·T tract DNA (8). To further delineate the DNA binding domain, translation termination signals were inserted after the first 90, 60, 48, 36, or 22 codons, the truncated peptides (referred to as D-90, -60, -48, -36, or -22, respectively) were expressed in *E. coli*, and total lysates were prepared and assayed for their ability to bind to A·T tract DNA. The D-90, -60, -48, and -36 peptides bound to A·T tract DNA specifically. Interaction of

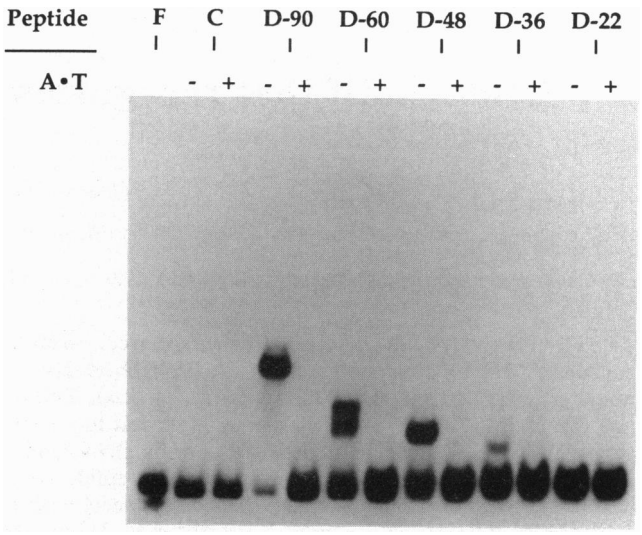

FIG. 1. Binding of carboxyl-terminal truncated Dat1p peptides to a DNA probe containing an A·T tract. Extracts were prepared from *E. coli* expressing D-90, D-60, D-48, D-36, or D-22 as indicated and equivalent amounts of total lysate were assayed for A·T tract binding activity using the band-shift assay. F, free probe; C, negative control extract prepared from *E. coli* harboring the parental vector pKK233. Binding reactions were carried out in the presence (+) or absence (−) of specific poly(A)·poly(T) competitor DNA as indicated.

D-22 with the DNA probe was not detected (Fig. 1). These data show that the amino-terminal 36 Dat1p residues are sufficient for high-affinity sequence-specific binding. To examine the DNA binding properties of this region in more detail, a 35-residue peptide corresponding to Dat1p amino acids 2–36 was chemically synthesized. This peptide (referred to as D-35) bound specifically to A·T tract DNA with an equilibrium dissociation constant of 4×10^{-10} M (Fig. 2). Thus, D-35 bound more tightly than did D-90 (D-90 $K_d = 3 \times 10^{-9}$ M; B.J.R., unpublished data). Residues 2–36 are, therefore, sufficient for the interaction with A·T tract DNA. The data suggest that residues 22–36 are required for high-affinity binding since D-22 failed to bind the DNA probe. Experiments described below show that missense mutations in this interval can severely reduce binding activity.

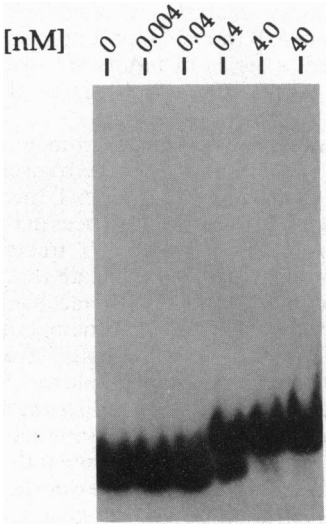

FIG. 2. Binding of D-35 to A·T tract DNA. The indicated final concentrations of the chemically synthesized D-35 were incubated with 4 pM labeled DNA probe. The relative amount of bound DNA in each sample was used to calculate the equilibrium dissociation constant (K_d) of 4×10^{-10} M.

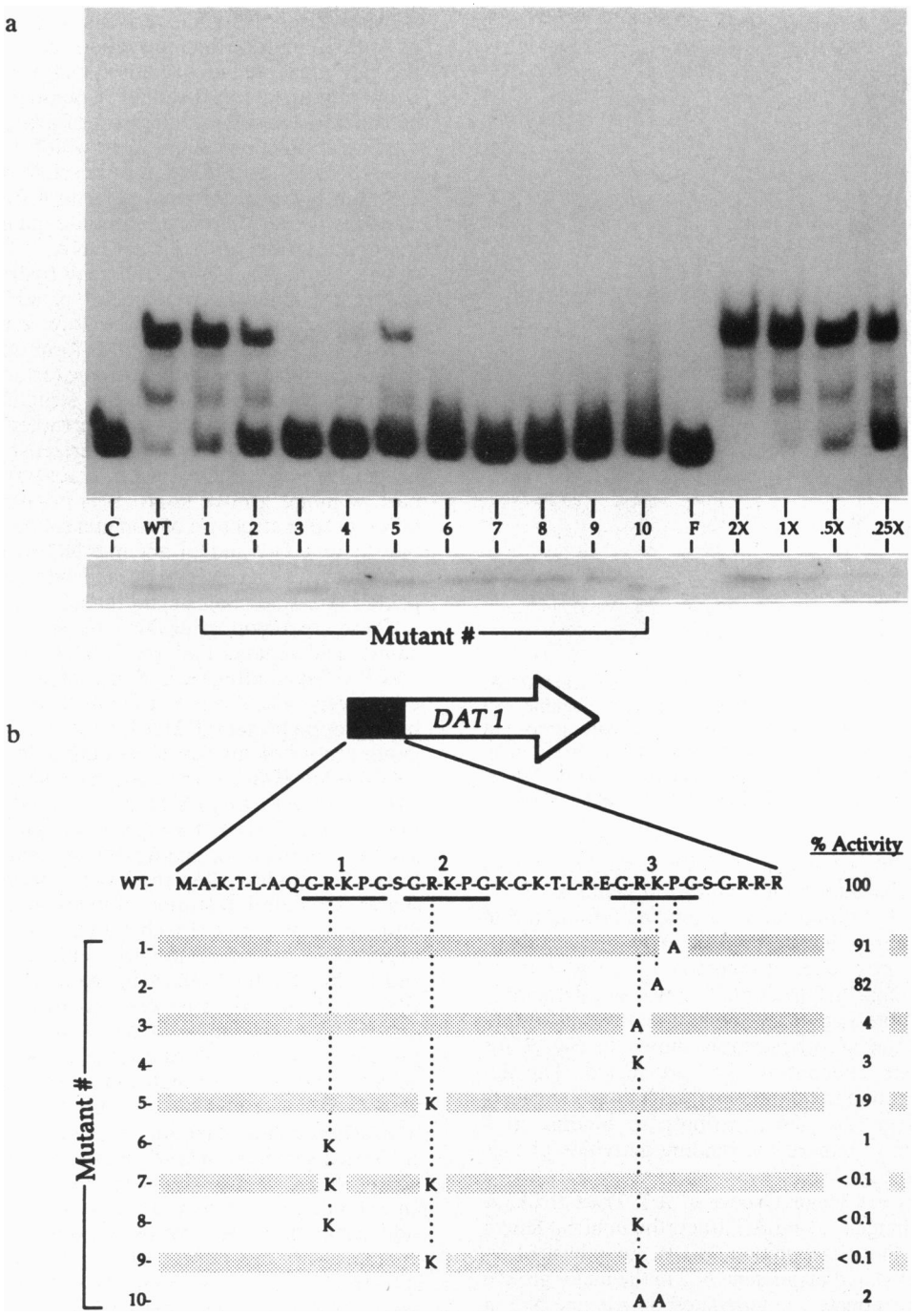

FIG. 3. Binding of D-90 pentad mutants to A·T tract DNA. (*a*) Band-shift assay of extracts prepared from *E. coli* expressing D-90 mutants and corresponding immunoblot used to determine relative D-90 levels. Extracts were prepared from *E. coli* harboring the following: C, the negative control pKK233 parent vector; WT, wild-type pD-90; mutants 1–10, pentad substitution mutants diagrammed in *b*. F, free probe; 2×, 1×, 0.5×, and 0.25×, 1:2 serial dilutions of the wild-type D-90 lysate. (*b*) Schematic representation of the *DAT1* gene showing the 36 amino-terminal residues and pentads 1, 2, and 3. Mutant designations show the substituted residue(s) and correspond to the lane designations in *a*. Relative binding activities were calculated by dividing the fraction of bound DNA (*a Upper*) by the relative amount of D-90 immunoreactivity (*a Lower*) expressed relative to wild type.

Arginine Residues in the Dat1p Pentads Are Required for DNA Binding Activity. The sequence Gly-Arg-Lys-Pro-Gly is repeated three times within the Dat1p DNA binding domain starting at residues 8, 14, and 26 (referred to below as pentads 1, 2, and 3, respectively, see Fig. 3). To determine whether the pentad residues are involved in the sequence-specific recognition of A·T tract DNA, they were changed using oligonucleotide-directed mutagenesis in pD90, and the binding activities of the mutated peptides were compared to wild-type D-90. Measurements of the D-90–DNA complex formation were corrected for variations in D-90 protein levels by quantitating immunoreactivity in the extracts using a peptide antisera directed against residues 42–61 (Fig. 3).

The Lys, Pro, or Arg codons in pentad 3 were separately changed to encode Ala. The Lys or Pro → Ala mutations had only modest effects on A·T tract binding, whereas the Arg → Ala mutation in pentad 3 reduced binding 25-fold. A conservative Arg → Lys substitution in pentad 3 and the Arg → Ala substitution reduced binding to a similar extent. These results suggest that the Arg residue in pentad 3 plays an important

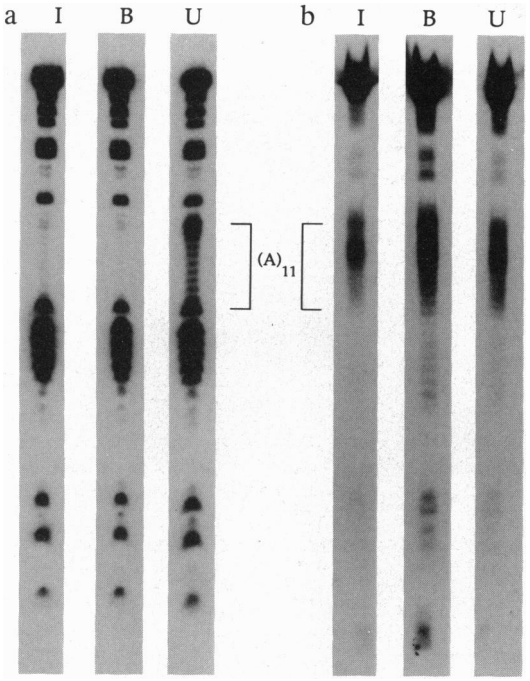

FIG. 4. Interference assays of D-90 binding to A·T tract DNA. The A·T tract was methylated in the minor groove at adenine N-3 using dimethyl sulfate (*a*) or carbethoxylated in the major groove at adenine N-7 using diethyl pyrocarbonate (*b*), prior to being challenged for binding by purified D-90. Lanes: I, input DNA; B, bound DNA; U, unbound DNA. The bracket indicates the position of the 11-bp A·T tract.

role in the recognition of A·T tract DNA. The Arg → Lys substitution was subsequently introduced in pentads 1 or 2. In pentad 2, this substitution reduced binding 5-fold, and in pentad 1 binding was reduced 100-fold. Arg → Lys substitutions were then introduced in pentads 1, 2, and 3 in all pairwise combinations. A·T tract binding was not detectable with any of the double Arg → Lys pentad mutants in 10-fold overexposures of the autoradiograph shown in Fig. 3, or when higher protein concentrations were tested. The decrease in binding seen in the double pentad mutants suggests that the pentad Arg residues contribute to binding in a cooperative fashion (compare the binding activities of mutants 4, 5, and 9 in Fig. 3).

Dat1p Recognizes the Minor Groove of A·T Tract DNA. A DNA probe containing an 11-bp A·T tract (the minimal length required for high-affinity recognition; B.J.R., unpublished results) was carbethoxylated on adenine N-7 in the major groove using diethyl pyrocarbonate or methylated on adenine N-3 in the minor groove using dimethyl sulfate. DNA was then challenged for binding to excess purified D-90, free and protein-bound probes were resolved from one another, and the DNA was cleaved at the site of modification and electrophoresed on a denaturing gel. As seen in Fig. 4, major-groove carbethoxyl modifications had no detectable effect on binding. In contrast, methylation of any of the 11 adenine N-3 minor groove groups strongly interfered with binding. These data indicate that D-90 interacts with the minor groove of A·T tract DNA. This conclusion is supported by the observation that the drugs netropsin and distamycin, which interact with the minor groove of A+T-rich DNA, can specifically compete for the binding of A·T tract DNA by D-90 (B.J.R., unpublished results).

DISCUSSION

The data suggest that Arg side chains in each of the Dat1p pentads interact with a repeated minor-groove feature of A·T tract DNA. We propose that single pentad Arg residues achieve a relatively modest binding affinity and that the overall strength of the interaction represents the product of the individual pentad-Arg affinities for A·T tract DNA and a cooperativity factor. It should be pointed out that the pentads can also be viewed as being part of a larger duplicated motif starting at positions 3 and 21 in which 11 out of 13 residues are repeated (see Fig. 3). It is possible that this duplication represents a superimposed structural framework that positions the pentads in the appropriate manner to interact with the minor groove of A·T tract DNA.

The pentad Arg side chains may hydrogen bond with the O-2 of thymine and/or the N-3 of adenine. These atoms occupy similar positions on the floor of the minor groove in both A·T and T·A base pairs in B-form DNA (12), suggesting that Dat1p might not distinguish the former base pair from the latter if it was constrained in this structure. X-ray crystallographic analysis shows that uninterrupted A·T tracts adopt an unusual DNA structure characterized by a high degree of propeller twist, a series of bifurcated hydrogen bonds, and narrow minor groove (2–4). It is possible that the unusual minor groove structure of nonalternating A·T DNA positions the thymine O-2 and/or adenine N-3 hydrogen-bond acceptors in a way that promotes their interaction with the three pentad Arg side chains simultaneously.

There are a number of DNA binding proteins that contain amino acid regions that are similar to the Dat1p Gly-Arg-Lys-Pro-Gly motifs (13). Some of these proteins interact exclusively with A+T-rich DNA. The mammalian high mobility group I protein (HMG-I, or α-protein) interacts with the minor groove of any 6 or more consecutive A or T base pairs (14). Within HMG-I, the consensus sequence Lys-Arg-Pro-Arg-Gly-Arg-Pro-Lys-Lys is repeated three times (15). Reeves and Nissen (16) have shown that an 11-residue synthetic peptide containing this sequence binds specifically to HMG-I binding sites and have proposed that the peptide shares structural features with the minor groove binding drugs netropsin and Hoechst 33258. The *Drosophila melanogaster* D1 protein preferentially binds A+T-rich DNA (17), and within D1 the consensus sequence Lys/Arg-Lys-Arg-Gly-Arg-Pro-Lys-Lys is repeated 10 times (18). The three-dimensional structures of helix–turn–helix proteins complexed with DNA show that sequence-specific binding can be achieved not only through the helix–turn–helix interaction with the major groove but also through motifs that are similar to Dat1p pentads that interact with the minor groove of A+T-rich subsites. In homeodomains, for example, the consensus Arg-Lys-Arg-Gly-Arg* in the amino terminus forms an extended region in which the practically invariant Arg* side chain (indicated by the asterisk) has been shown to interact with the minor groove of the TAAT consensus subsite (19–21). These contacts contribute substantially to sequence-specific DNA binding. Such contacts may also be important in contributing to the binding of HMG-box proteins that contain the consensus Pro-Lys-Arg-Pro in their extended amino-terminal region and bind to a strongly conserved AAA core sequence (22, 23). Sequences related to Dat1p pentad motifs are also used by prokaryotic DNA binding proteins. In the sequence Lys-Arg-Pro-Arg* of phage 434 repressor, the Arg* side chain interacts with the sequence ATA through its minor groove (24). In addition, Hin recombinase of *Salmonella typhimurium* and the other closely related recombinases Gin, Pin, and Cin have in their carboxyl-terminal domain a conserved Gly-Arg*-Arg/Pro-Pro-Lys motif in which the Arg* is reported to be important in recognition of the minor groove of an AAA sequence (25). These residues are amino-terminal to a helix–turn–helix motif that is necessary but not sufficient for sequence-specific DNA binding.

Why are minor-groove-binding Arg residues found in the context of Gly, Lys, and Pro residues? The mutagenesis studies described here show that the Lys and Pro side chains

in Dat1p pentads are not required for binding. Perhaps the role of these amino acids is to present the Arg side chain in an extended and/or nonconstrained conformation such that it is sterically unhindered and, therefore, able to interact in the minor groove. If so, then the intervening amino acids that dictate the molecular distances between minor-groove-binding Arg residues may play an important role in defining DNA binding site specificity.

We thank Perry Hall, Irene Weber, Mark Hochstrasser, and Ya Ming Hou for helpful discussions and comments on the manuscript. This work was supported by Grant R01GM45772 from the National Institutes of Health.

1. Karlin, S., Blaisdell, B. E., Sapolsky, R. J., Cardon, L. & Burge, C. (1993) *Nucleic Acids Res.* **21,** 703–711.
2. Nelson, H. C. M., Finch, J. T., Luisi, B. F. & Klug, A. (1987) *Nature (London)* **330,** 221–226.
3. Coll, M., Frederick, C. A., Wang, A. H.-J. & Rich, A. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 8385–8389.
4. DiGabriele, A. D., Sanderson, M. R. & Steitz, T. A. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 1816–1820.
5. Struhl, K. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 8419–8423.
6. Lue, N. F., Buchman, A. R. & Kornberg, R. D. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 486–490.
7. Chen, W., Tabor, S. & Struhl, K. (1987) *Cell* **50,** 1047–1055.
8. Winter, E. & Varshavsky, A. (1989) *EMBO J.* **8,** 1867–1877.
9. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K., eds. (1991) *Current Protocols in Molecular Biology* (Greene/Wiley-Interscience, New York).
10. Strum, R., Baumruker, T., Franza, B. R., Jr., & Herr, W. (1987) *Genes Dev.* **1,** 1147–1160.
11. Ballard, M. J., Tyndall, W. A., Shingle, J. M., Hall, D. J. & Winter, E. (1991) *EMBO J.* **10,** 3753–3758.
12. Seeman, N. C., Rosenberg, J. M. & Rich, A. (1976) *Proc. Natl. Acad. Sci. USA* **73,** 804–808.
13. Churchill, M. E. A. & Travers, A. A. (1991) *Trends Biochem. Sci.* **16,** 92–97.
14. Solomon, M. J., Strauss, F. & Varshavsky, A. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 1276–1280.
15. Johnson, K. R., Lehn, D. A., Elton, T. S., Barr, P. J. & Reeves, R. (1988) *J. Biol. Chem.* **263,** 18338–18342.
16. Reeves, R. & Nissen, M. S. (1990) *J. Biol. Chem.* **265,** 8573–8582.
17. Levinger, L. & Varshavsky, A. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 7152–7156.
18. Ashley, C. T., Pendleton, C. G., Jennings, W. W., Saxena, A. & Glover, C. V. G. (1989) *J. Biol. Chem.* **264,** 8394–8401.
19. Laughon, A. (1991) *Biochemistry* **30,** 11357–11367.
20. Kissinger, C. R., Liu, B., Martin-Blanco, E., Kornberg, T. B. & Pabo, C. O. (1990) *Cell* **63,** 579–590.
21. Otting, G., Qian, Y. Q., Billeter, M., Muller, M., Affolter, M., Gehring, W. J. & Wuthrich, K. (1990) *EMBO J.* **9,** 3085–3092.
22. Ner, S. S. (1992) *Curr. Biol.* **2,** 208–210.
23. Weir, H. M., Kraulis, P. J., Hill, C. S., Raine, A. R. C., Laue, E. D. & Thomas, J. O. (1993) *EMBO J.* **12,** 1311–1319.
24. Aggarwal, A. K., Rodgers, D. W., Drottar, M., Ptashne, M. & Harrison, S. C. (1988) *Science* **242,** 899–907.
25. Sluka, J. P., Horvath, S. J., Glasgow, A. C., Simon, M. I. & Dervan, P. B. (1990) *Biochemistry* **29,** 6551–6561.