DNA binding properties of the *Saccharomyces cerevisiae* DAT1 gene product

Brian J. Reardon, Donna Gordon⁺, Michelle J. Ballard[§] and Edward Winter^{*}

Department of Biochemistry and Molecular Biology, Thomas Jefferson University, Philadelphia, PA 19107, USA

Received July 31, 1995; Revised and Accepted October 27, 1995

ABSTRACT

The *DAT1* gene of *Saccharomyces cerevisiae* encodes a DNA binding protein (Dat1p) that specifically recognizes the minor groove of non-alternating oligo(A)·oligo(T) tracts. Sequence-specific recognition requires arginine residues found within three perfectly repeated pentads (G-R-K-P-G) of the Dat1p DNA binding domain [Reardon, B. J., Winters, R. S., Gordon, D., and Winter, E. (1993) *Proc. Natl. Acad. Sci. USA* 90, 11327–11331]. This report describes a rapid and simple method for purifying the Dat1p DNA binding domain and the biochemical characterization of its interaction with oligo(A)·oligo(T) tracts. Oligonucleotide binding experiments and the characterization of yeast genomic Dat1p binding sites show that Dat1p specifically binds to any 11 base sequence in which 10 bases conform to an oligo(A)·oligo(T) tract. Binding studies of different sized Dat1p derivatives show that the Dat1p DNA binding domain can function as a monomer. Competition DNA binding assays using poly(I)·poly(C), demonstrate that the minor groove oligo(A)·oligo(T) constituents are not sufficient for high specificity DNA binding. These data constrain the possible models for Dat1p/oligo(A)·oligo(T) complexes, suggest that the DNA binding domain is in an extended structure when complexed to its cognate DNA, and show that Dat1p binding sites are more prevalent than previously thought.

INTRODUCTION

Sequence-specific recognition of DNA by proteins is often achieved through conserved structural domains (i.e. helix–turn–helix, zinc finger, leucine zipper and β -ribbon) that allow for hydrogen bonding between amino acid side chains and functional groups exposed in the major groove of the DNA helix (2–8). Sequence-specific recognition of DNA via the minor groove appears to be less common. Of the proteins known to recognize minor groove features in DNA, most have a preference for A+T-rich DNA, but do not recognize a highly specific sequence of bases (9–13). Many of these proteins appear to interact with the minor groove of A+T-rich DNA through relatively short (5–10 residue) peptides that are rich in glycine, arginine, lysine and

proline (GRKP motifs). GRKP motifs were first described in the HMG-I/Y chromosomal protein (also known as α -protein), a relatively abundant mammalian protein that recognizes any sequence of six A·T base pairs in duplex DNA (13). Synthetic peptides containing GRKP motifs have been shown to preferentially bind to A+T-rich DNA with equilibrium dissociation constants of 10^{-6} – 10^{-7} M (11,12). It has been proposed that GRKP A+T binding peptides in HMG-I bear structural similarities to the A+T DNA binding drugs netropsin and distamycin A and have therefore been referred to as A·T-hooks (11). Minor groove binding GRKP motifs are found in many sequence-specific DNA binding proteins that also contain major groove binding structural motifs. For example, the homeodomain proteins (including *antennapedia*, *engrailed*, as well as the yeast MAT α 2) contain GRKP motifs in their N-termini. X-ray crystallographic studies have shown that these motifs interact with A+T base pairs in the minor groove (14–19). A large number of other sequence-specific DNA binding proteins contain GRKP motifs in addition to major groove binding domains. It is likely that GRKP motifs act in concert with a wide variety of major groove binding domains to generate high affinity and specificity protein–DNA interactions.

The *DAT1* gene of *Saccharomyces cerevisiae* encodes a protein, Dat1p, that was identified based on its ability to bind non-alternating oligo(A)·oligo(T) tracts (hereafter referred to as A·T DNA) in a highly sequence-specific manner (20). Previous work has shown that the N-terminal 36 residues of Dat1p bind to A·T tracts with an equilibrium dissociation constant of 4×10^{-10} M and are thus sufficient for high affinity DNA recognition (1). Within this sequence is a pentad motif, GRKPG, that is perfectly repeated three times. Mutations that change a single K or P pentad residue to A had no effect on DNA binding; however, mutations that result in single R→A or R→K substitutions within the pentad greatly reduced binding (1). Any pairwise combination of pentad R substitutions completely eliminated binding activity of the doubly mutant Dat1p (1). Based on these results as well as chemical interference assays (1) and competition studies using minor groove binding drugs (netropsin and distamycin A), it was proposed that the Dat1p pentad arginines interact in a cooperative manner with the minor groove of the A·T helix to achieve high affinity sequence-specific binding.

Non-alternating A·T tracts are common intergenic elements in many organisms. In yeast uninterrupted A·T tracts 10 base pairs (bp) or longer are found on average once every 5 kb, and are found in promoter regions of 25% of yeast genes (defined for statistical

^{*} To whom correspondence should be addressed

Present addresses: ⁺University of Pennsylvania, Department of Physiology, Philadelphia, PA 19104, USA and [§]DuPont Merck Pharmaceutical Co., Glenolden Laboratory, Glenolden, PA 19036, USA

purposes as 500 bp upstream of the initiator ATG) (21). A·T tracts adopt a unique DNA structure. X-ray diffraction studies of oligonucleotide duplexes containing A·T tracts shows that this structure is characterized by a short helical repeat, a narrow minor groove, a high degree of propeller twist, and a system of bifurcated hydrogen bonds (22–24). This unique structure results in a rigid duplex with a number of unusual physical properties: runs of ~5 bp of the homopolymer phased every helical repeat cause an overall bend in the DNA helix (22), A·T tracts have a melting temperature that is 10°C higher than alternating (AT)·(AT) tracts (23), and long tracts (>20 bp) disfavor nucleosome formation (25).

In yeast, A·T tracts have been shown to promote transcription (26,27). A number of differences between the transcriptional activating properties of A·T tracts and other *cis*-acting transcriptional elements indicate that A·T tracts exert their effects through their unique physical properties and not through A·T-specific transcription factors. Iyer and Struhl have shown that A·T tract transcriptional potency is length dependent in a linear non-threshold dependent fashion, and that A·T tracts increase occupancy of nearby transcription factor binding sites (27). These and other observations have led to the proposal that A·T tracts promote transcription through a unique DNA structure and possibly through a mechanism that involves destabilization of nucleosomes and concomitant increase in the accessibility of promoters to transcription factors. Promoters containing A·T tracts are expressed at 1.5–2.5-fold higher levels in *dat1* strains than in coisogenic wild-type strains, and 2 μ -high copy *DAT1* plasmids repress the expression of promoters containing A·T tracts 2–3-fold over wild type (20,27, our unpublished results). These results show that Dat1p can repress A·T promoter activity. One possibility is that Dat1p binding may repress A·T specific promoter activity by altering the chromatin structure of A·T tracts.

In this report, we describe a rapid purification scheme for the DNA binding domain of Dat1p based primarily on its solubility properties. This purified material was used to show that the Dat1p DNA binding domain can bind to A·T DNA as a monomer and to precisely define the DNA sequences that interact with Dat1p. These results show that the recognition site for Dat1p is any 11 bp sequence in which 10 bases conform to an uninterrupted

oligo(A)·oligo(T) tract. Additionally, we show that the presence of A·T minor groove functional groups are not sufficient for specific interaction with Dat1p. These observations are discussed with respect to a possible model for the Dat1p–A·T DNA interaction.

MATERIALS AND METHODS

Expression and purification of D-90

Escherichia coli (*E.coli*) strain JM101 harboring the plasmid pD90 (1) that encodes the first 90 residues of Dat1p (D-90) under the control of the isopropyl β -D-thiogalactoside (IPTG) inducible pKK233 promoter (20) was grown to an OD₆₀₀ of 0.5, IPTG added to 1 mM, and the cells incubated for an additional 1.5 h before being pelleted by centrifugation. Cell lysates were prepared using the Triton/lysozyme method as previously described (20). D-90 was purified from the crude lysate using two different methods. Method 1 is identical to that previously described for the purification of Dat1p from yeast (1,20). In method 2, cell lysates were heated in a boiling water bath for 10 min and subsequently cooled in an ice water bath. The heat-treated lysate was centrifuged to remove the flocculent precipitate, and the supernatant decanted (fraction II, Table 1). All centrifugation steps were at 10 000 g for 10 min at 4°C. Solid ammonium sulfate was slowly added to the collected supernatant with stirring at 4°C to a final concentration of 45% saturation, incubated from 1 h to overnight at 4°C, and the supernatant recovered after centrifugation. To this supernatant, solid ammonium sulfate was added to a final concentration of 75% saturation, incubated from 1 h to overnight at 4°C and the D-90 containing precipitate (fraction IV) recovered by centrifugation. The precipitate was dissolved in 20 mM Tris–HCl, pH 8.0, and incubated with a 0.1 vol DEAE resin equilibrated in the same buffer in order to remove RNA. The unbound fraction was collected by pouring the slurry through a glass fritted funnel and washed with 1 vol column buffer. The material was loaded onto a Pharmacia FPLC Mono-S column, and developed using a 0–1 M NaCl linear 47 ml gradient (1 ml fractions were collected). D-90 eluted between 0.45 and 0.55 M NaCl. In our hands, fraction IV is relatively stable and suitable for most binding studies, behaving indistinguishably from the highly purified preparations.

Table 1. Purification of D90

Fraction ^a	Total protein (mg)	Relative binding activity ^b (%)	Fold purification ^c
I. Crude lysate	75	100	1
II. Boiled lysate	3.0	93	23
III. 45% (NH ₄) ₂ SO ₄ supernatant	1.6	83	39
IV. 75% (NH ₄) ₂ SO ₄ pellet	1.2	80	50
V. DEAE flow-through	0.59	70	89
VI. Mono S fraction	0.11	35	239

^aDetails of fractionation steps are described in Materials and Methods.

^bRelative binding activity determined by analyzing radiolabeled DNA probe-1 binding activity in equivalent portions of each sample, quantitating the fraction of bound and unbound probe using a PhosphorImager (Molecular Dynamics), expressed relative to binding activity of fraction I (see Materials and Methods).

^cFold purification of each fraction is the relative binding activity/mg protein normalized to fraction I.

DNA binding assay

A·T DNA binding activity was assayed using the band-shift procedure as previously described (20). DNA probe-1, a 138 bp *AluI* restriction endonuclease fragment of pLP-119 (28) containing a 19 bp A·T tract, was radiolabeled with ^{32}P to roughly 2×10^8 dpm/ μg using T4 kinase. Oligonucleotide duplexes for Dat1p binding studies consisted of the sequences; 5'-CGC[A]₉₋₁₃GCG-3' and 5'-CGC[T]₉₋₁₃GCG-3'. Double-stranded oligonucleotide DNA probes were prepared by radiolabeling each of the complementary oligonucleotides using T4 kinase to $\sim 1 \times 10^{11}$ dpm/ μg before mixing equimolar concentrations in 10 mM Tris-HCl, pH 7.0, 100 mM NaCl, and 1 mM EDTA, heating to 90°C and slow cooling the mixture to room temperature.

Dat1p mixing/denaturation/renaturation experiments

Mono S-purified derivatives of Dat1p were mixed, subjected to heat or chemical denaturation/renaturation, and assayed for DNA binding as described above. Heat denaturation/renaturation consisted of mixing Dat1p derivatives in 20 μl of binding buffer (10 mM Tris-HCl, pH 7.5, 70 mM NaCl, 1 mM EDTA, 2% sucrose, 0.1% Triton X-100, and 10 mM 2-mercaptoethanol), boiling 2 min in a water bath followed by cooling to room temperature. Chemical denaturation/renaturation consisted of combining the Dat1p derivatives in 6 M guanidine HCl for 30 min, diluting 1:14 with water, and diluting a second time 1:10 in binding buffer before being assayed for DNA binding as described above.

Purification of yeast chromosomal Dat1p binding sites

Yeast genomic DNA digested to completion with *Sau3AI* was incubated at 5 mg/ml with 1 mg/ml D-90, purified by method 2, for 30 min in 1 ml of buffer A (10 mM Tris-HCl, pH 7.5, 80 mM NaCl, and 1 mM EDTA). The binding reactions were supplemented with 1 ng/ml of ^{32}P radiolabeled DNA probe-1 to monitor the recovery of Dat1p binding sites. Total rabbit antisera raised against a synthetic peptide corresponding to Dat1p residues 42–61 (1) was then added to a final dilution of 1:100 and incubated for 30 min before 0.1 vol protein A-sepharose was added and incubated for an additional 30 min. All incubations were at 4°C. The protein A-sepharose beads were pelleted by a brief centrifugation, the supernatant was removed, and the beads were washed three times in buffer A. Bound DNA was eluted with 1 M NaCl in buffer A, and precipitated by the addition of 2 vol ethanol using yeast tRNA as a carrier. The overall recovery of ^{32}P radiolabeled DNA probe-1 was 18–25% as determined by Cherenkov counting. Typically, three cycles of the above binding reactions were performed before the precipitated DNA was ligated to *Bam*HI digested, calf intestinal phosphatase treated, pUC19 plasmid DNA for transformation into *E. coli* strain DH5 α F'.

Plasmids harboring yeast Dat1p binding sites were identified by their ability to act as specific competitors in a band-shift assay. Unlabeled, linearized plasmid DNA (10–100 ng, purified from single colonies) was added to 25 μl binding reactions containing 1×10^{-8} M D-90, ~ 1000 dpm (10–100 pg) ^{32}P -radiolabeled DNA probe-1, and 200 ng *E. coli* competitor DNA in binding buffer (10 mM Tris-HCl, pH 7.5, 70 mM NaCl, 1 mM EDTA, 2% sucrose, 0.1% Triton X-100, and 10 mM 2-mercaptoethanol). Those plasmids whose inclusion reduced more than 90% of binding activity to DNA probe-1 were sequenced using the dideoxynucleo-

tide chain-termination method (United States Biochemical, Sequenase Kit V2.0).

RESULTS

Production of the Dat1p DNA binding domain in *E. coli*

The plasmid pD90 which directs the expression of the 90 N-terminal residues (D-90) was used to produce the Dat1p DNA binding domain in *E. coli*. Expression of D-90 was chosen for the following reasons: (i) it specifically bound to A·T tracts with high affinity, $K_d = 3 \times 10^{-10}$ M (1), (ii) it was more resistant to degradation by bacterial proteases than all of the longer and the shorter Dat1p derivatives tested, (iii) its DNA binding activity was unaffected by heat (boiling) and chemical (guanidine HCl) treatment (see below), and (iv) unlike the longer derivatives (including full length Dat1p) it was highly soluble in physiologic salt and pH conditions.

Two purification schemes were used for the purification of D-90. Method 1 was identical to that developed for the purification of Dat1p from yeast and utilizes phosphocellulose, an A·T affinity matrix, and Mono S FPLC ion exchange chromatography (1,20). While method 1 yielded D-90 of high purity, it was labor intensive, particularly for preparation of relatively large amounts of material, and required a specialized affinity matrix. Method 2, described in this report, took advantage of D-90's resistance to boiling, its solubility properties in ammonium sulfate, and a single ion exchange resin to achieve protein whose purity and specific binding activity was indistinguishable from D-90 purified by method 1. The purification of D-90 by method 2 is shown in Figure 1. This method routinely yielded D-90 that was at least 90% pure as judged by densitometric tracings of Coomassie stained electrophoretograms (Fig. 1B).

In addition to the purified 90 residue peptide a series of smaller Dat1p degradation products were resolved in the final Mono-S purification step (Fig. 1A, lanes 11–13). These chromatographically resolved proteolytic products were also shown to bind A·T DNA specifically. Dat1p proteolytic fragment/DNA probe complexes migrated at distinct positions with smaller degradation products forming complexes that migrated more rapidly than those formed by larger peptides (Fig. 1 and data not shown). The ability of small Dat1p derived peptides to specifically bind to A·T tracts is consistent with our previous demonstration that the N-terminal 36 Dat1p residues are sufficient for high affinity sequence-specific DNA binding (1).

The minimal A·T tract recognized by Dat1p is 10 bp

The minimal A·T tract required for Dat1p binding was measured by testing the binding of D-90 (from *E. coli* lysates) to ^{32}P radiolabeled duplex oligonucleotides with 9, 10, 11, 12 and 13 nonalternating A·T bp flanked by 3 bp G/C-rich clamps (referred to as duplexes A₉–A₁₃ respectively, see Fig. 2). D-90 bound with high affinity to the A₁₁, A₁₂ and A₁₃ duplexes and less tightly to the A₁₀ duplex. Quantitation of the amount of radiolabeled duplex probe bound by D-90 showed that 5-fold less A₁₀ duplex was bound compared to the A₁₁–A₁₃ duplexes (see Table 2). Binding of D-90 to the A₉ duplex was undetectable even at 10-fold overexposure of the autoradiogram shown in Figure 2 or when the concentration of D-90 was increased 10-fold. Binding of the A₁₁, A₁₂ and A₁₃ duplexes were indistinguishable. Thus, D-90 required at least 10 bp of A·T DNA for high affinity recognition, and the

a.

b.

Figure 1. Purification of D-90 from *E. coli*. (a) Fractions were assayed for binding to DNA probe-1, a 119 bp fragment containing a 19 bp oligo(A)·oligo(T) tract using the band shift assay, as described in Materials and Methods. Lane 1, Free DNA probe; lane 2, fraction I (crude lysate); lane 3, fraction II (boiled lysate); lane 4, 45% ammonium sulfate pellet; lane 5, fraction IV (DEAE flow-through; Mono-S loaded material); lane 6, Mono-S flow-through. Lanes 7–15 correspond to the peak of Mono-S purified D-90. For comparative purposes equivalent portions of fractions I, II and III were analyzed and equivalent portions of the Mono-S column load, and resolved fractions were analyzed. Dashed line indicates NaCl concentrations corresponding to fractions 7–15. (b) Protein from fraction 10 (10 µg) was analyzed by SDS–polyacrylamide gel electrophoresis and stained with Coomassie blue. The stained gel and its densitometric tracing are shown as are the molecular masses of protein standards.

strength of the interaction was not dependent on the length of the A·T tract in binding sites longer than 11 bp.

Dat1p binding sites in the yeast genome

In order to isolate representative yeast genomic DNA sequences that interact with Dat1p we made use of antisera raised against a synthetic peptide corresponding to Dat1p residues 42–61 (sera 159). Sera 159 specifically recognized Dat1p expressed in both yeast and *E. coli* (1). In addition, sera 159 specifically interacted with the D-90/A·T DNA complex since its inclusion in DNA binding assays retarded the electrophoretic migration of the peptide–DNA complex (Fig. 3). This interaction was exploited in order to purify genomic Dat1p binding sites. D-90 was incubated with total genomic DNA which had been digested with *Sau3A*I. Sera 159 was added, followed by the addition of protein A–sepharose. The complexes were pelleted, bound DNA was eluted with 1 M NaCl, and subjected to two more cycles of

Figure 2. Measurement of the minimal A·T tract required for Dat1p binding. Radiolabeled oligonucleotide duplex alone (lanes 1, 4, 7, 10 and 13), D-90 containing reactions (lanes 2, 5, 8, 11 and 14) and D-90 containing reactions supplemented with 50 ng of poly (A)·poly (T) (lanes 3, 6, 9, 12 and 15) were assayed by the band shift method as described in Materials and Methods. Oligonucleotide duplexes consisted of the sequence shown flanked by three G/C base pairs on either end.

Figure 3. Retardation of the D-90–DNA complex by anti-Dat1p antisera. The interaction of sera 159 with the D-90/DNA probe-1 complex was assayed by the band shift method. The migration position of the free (unbound) probe (F), the D-90–DNA complex (B), and the antibody-bound D-90–DNA complex (SC) are indicated.

D-90/antisera selection. The purified DNA was subsequently cloned into a plasmid vector (see Materials and Methods). ³²P radiolabeled DNA probe-1 was included to monitor specific DNA–protein interactions and in control experiments was used to determine the appropriate concentrations of D-90 and genomic DNA in order to achieve a high degree of specific interaction.

Table 2. Duplex oligonucleotide interactions with Dat1p

Oligonucleotide ^a	Relative binding activity ^b
(A) ₈	0
(A) ₉	0
(A) ₁₀	0.2
(A) ₁₁	1.0
(A) ₁₂	1.0
(A) ₁₃	1.0
(A) ₅ T(A) ₅	1.0
(A) ₅ C(A) ₅	1.0
(A) ₅ G(A) ₅	1.0
(A) ₄ T(A) ₅	0.02
(A) ₂ T(A) ₇	0.02

^aOligonucleotides were complementary duplexes in which the A-rich strand consisted of the sequence: 5'-CGC[X]GCG-3' where 'X' is the sequence indicated (see Materials and Methods).

^bRelative binding activities show the amount of duplex oligonucleotide that was specifically bound compared to the [A]₁₁ oligonucleotide which was defined as 1.0. Specificity for the interactions was confirmed using 50 ng of poly (A) · poly (T) DNA or 50 ng of *E.coli* DNA as specific and non-specific competitors, respectively.

The cloned DNA fragments were subsequently tested for their ability to specifically compete for binding of D-90 to DNA probe-1. In control experiments, 1 of 25 randomly cloned *Sau*3AI fragments were able to specifically compete for binding. However, 20 of 52 DNA fragments isolated and cloned through the above strategy were able to specifically compete for binding of Dat1p to the DNA probe. Thus, this method represents a relatively rapid and efficacious method for generating libraries of Dat1p binding sites.

The complete nucleotide sequences of 11 yeast chromosomal fragments that specifically interacted with Dat1p were determined. Seven of the 11 sequences contained A · T tracts between 10 and 27 bp in length, conforming to the requirements for Dat1p binding described above. The other four sequences contained A · T tracts longer than 10 bp that harbored single base A or T interruptions of the homopolymeric stretch (Table 3). Clone 6, which contained the sequence A₅TA₅TA₅ bound to Dat1p with an affinity that was indistinguishable from that of DNA probe-1. Additionally, hydroxyl-radical footprinting experiments showed that Dat1p bound directly to this sequence (data not shown).

A series of A · T-tract oligonucleotide duplexes containing single base pair interruptions were prepared and tested for binding to D-90 in order to confirm and extend the characterization of interrupted Dat1p binding sites (see Experimental Procedures). Interrupting the 11 bp A · T tract in duplex-A₁₁ (Table 2) by switching the central A · T base pair to a T · A, a G · C or a C · G (to yield duplexes A₅TA₅, A₅GA₅, or A₅CA₅) had no detectable effect on the affinity towards D-90 or on its relative specificity for binding as measured by the amount of nonspecific *E.coli* DNA required to compete for binding (see Experimental Procedures). Interrupting the 10 bp A · T tract in duplex-A₁₀ by switching either the third or the fifth A · T bp to T · A yielding duplexes A₂TA₇, or A₄TA₅ respectively, reduced binding affinity towards D-90 by ~10-fold. This reduction in apparent binding affinity was accompanied by an approximate ~10-fold reduction in specificity since

Figure 4. Mixing experiments of Dat1p derivatives. Derivatives of Dat1p were subjected to denaturation/renaturation treatments as indicated followed by band shift assays using the oligo(A) · oligo(T) containing DNA probe-1. The larger peptide fragment (D-90) corresponds to the N-terminal 90 Dat1p residues while the smaller peptide fragment represents a proteolytic fragment purified by Mono S FPLC (lane 13, Fig. 1). Control experiments show that the purified small fragment corresponds approximately to the 60 N-terminal Dat1p residues (data not shown). The position of the large peptide fragment–DNA complex (A), the small peptide fragment–DNA complex (B), and the free (unbound) DNA (F) are indicated.

roughly ~10-fold less nonspecific *E.coli* DNA was required to compete 50% of binding compared to the uninterrupted A₁₀ duplex. Thus, the affinity of D-90 towards the interrupted A₁₀ duplex was ~50-fold lower than its affinity towards the interrupted A₁₁ duplex (see Table 2).

The Dat1p DNA binding domain can function as a monomer

Mixtures of D-90 and a smaller FPLC purified A · T DNA binding peptide (Fig. 1A, lane 13) were denatured using heat or guanidine HCl, renatured by cooling or dilution of the denaturant, and subsequently challenged for binding to DNA probe-1. This mixture generated two complexes with DNA whose electrophoretic mobilities were indistinguishable from those observed when the peptides were analyzed separately (Fig. 4). Complexes of intermediate mobility were not detected, indicating that heteromultimers were not formed. This data suggests that the Dat1p DNA binding domain can interact with A · T DNA as a monomer.

Minor groove functional groups of A · T DNA are not sufficient for high affinity binding

We have previously shown that Dat1p recognizes A · T tracts through the minor groove (1). To determine if the minor groove of A · T DNA is sufficient for Dat1p recognition, we assessed whether the interaction of Dat1p with DNA probe-1 could be competed for by oligo(I) · oligo(C) DNA, which has the same minor groove constituents as A · T DNA. Binding reactions were carried out in the presence of increasing amounts of either alternating poly(AT) · (AT), homopolymeric (A) · (T), or homopolymeric (I) · (C) DNA (Fig. 5). As expected, the alternating polymer was a poor competitor for D-90 binding with binding activity being unaffected at >100-fold molar excess over the DNA probe. In contrast, the non-alternating homopolymeric A · T

Figure 5. A·T DNA minor groove functional groups are not sufficient for high affinity binding. Binding of purified D-90 to DNA probe-1 using the indicated unlabeled competitor DNA at 5 ng (lanes 3, 6 and 9), 50 ng (lanes 4, 7 and 10) and 500 ng (lanes 5, 8 and 11). Lanes 1 and 2 represent free (unbound) probe and D-90 complex in the absence of specific competitor DNA, respectively.

DNA strongly competed for D-90 binding at even the lowest concentration tested. Poly(I)·(C) was the poorest competitor, with D-90 binding activity unaffected even at >10 000-fold molar excess over the DNA probe-1. Therefore, the minor groove functional groups in A·T tracts are not sufficient for high affinity Dat1p recognition.

Table 3. Genomic Dat1p DNA binding sites

Clone ^a	Sequence
1	(T) ₁₂
2	(T) ₁₆
3	(A) ₃ T(A) ₆ T(A) ₈ T(A) ₃
4	(T) ₁₃
5	TA(T) ₂ A(T) ₉ A(T) ₈ A(T) ₂
6	(T) ₅ A(T) ₅ A(T) ₅
7	(T) ₇ A(T) ₅
8	(T) ₂₇
9	(T) ₁₁
10	(T) ₁₀ TA(T) ₁₀ A(T) ₂
11	(T) ₁₁

^aClones 1, 7, 9 and 10 match DNA sequences U11581, D42127, J01477 and U0060, respectively [Wisconsin Sequence Analysis Package (GCG) Version 8.0; GenBank, Release 87.0].

DISCUSSION

Many different sequence-specific DNA binding proteins use GRKP motifs to recognize A·T base pairs. Dat1p appears to be a specialized example in which GRKP motifs are exclusively used to recognize a relatively long and specific DNA sequence. An understanding of the requirements for Dat1p/A·T interaction

will provide information concerning the structure/function of GRKP motifs. In addition, these studies will shed light on the role of Dat1p in cellular physiology. This report provides evidence that Dat1p can interact with A·T DNA as a monomer, that the minimal A·T tract length recognized is 10 bp (single substitutions within longer A·T tracts are tolerated), and that minor groove constituents of A·T DNA are not sufficient for high affinity binding.

The ability of the DNA binding domain of Dat1p to function as a monomer limits the possible structural models that can account for its interaction with A·T tracts. Previous studies have shown that the N-terminal 36 Dat1p residues are sufficient for high affinity binding to A·T tracts and that this binding requires the arginine residues in the three GRKPG pentads of the Dat1p DNA binding domain that interact with the A·T minor groove (1). The helical repeat of A·T DNA is 10.0 bp (the minimal Dat1p recognition site) with a pitch that has been proposed to be 32.4 Å (22). The surface distance along the minor groove through the minimal binding site is ~70 Å assuming a helical diameter of 20 Å. The maximum distance that could separate the most N-terminal pentad arginine from the most C-terminal pentad arginine, if the intervening 18 peptide bonds are in a completely extended beta conformation, is ~65 Å. If it is assumed that pentad arginines contact the minor groove adjacent to the first and last base pair of the minimal binding site and that the DNA binding domain functions as a monomer, then it follows that the DNA binding domain must form an extended structure. The relative dimensions of the DNA and peptide suggest that it is possible that the Dat1p peptide backbone adopts a conformation which allows it to lie in juxtaposition to the minor groove path. It is also possible that the Dat1p DNA binding domain is extended when not complexed with DNA. It follows then that the Dat1p DNA binding domain may be permissive for sequence-specific binding in that it allows the arginine residues to interact specifically with the A·T minor groove as opposed to rigidly positioning the arginine side chains for the interaction. A prediction of this induced-fit model is that a subset of possible conformations should predominate when it is complexed to its cognate DNA site. Dat1p's inability to bind poly(I)·poly(C) DNA suggests that it cannot accommodate the minor groove dictated by this homopolymer and that the characteristic three dimensional structure of A·T DNA is critical for high affinity binding.

The oligonucleotide binding experiments and the characterization of genomic Dat1p binding sites allowed us to define the DNA sequence requirements for Dat1p binding. We propose that Dat1p specifically interacts with any 11 bp sequence which contains at least 10 bp that define an A·T tract. Single G·C, C·G or T·A interruptions in 11 bp A·T tracts have no discernible effect on Dat1p binding, while the same interruptions in 10 bp A·T tracts substantially reduce binding. We suggest that there are 11 different minimal binding sites which can be represented by the following series; XA₁₀, AXA₉, A₂XA₈, ..., A₁₀X (where X represents any base). Dat1p's tolerance of different single base interruptions within its DNA binding site (see Table 2 and 3) suggest that there may be flexibility in the allowable pentad arginine/minor groove contacts that result in high affinity interaction. These data also show that the unique structure of A·T DNA recognized by Dat1p can occur in non-contiguous A·T stretches. The three dimensional crystallographic structure of short A·T stretches has been reported for A·T tracts 3, 5 and 6 bp in length (23,24 and 22, respectively). These studies are in

close agreement, and show that short A·T tracts adopt a non-B-DNA structure characterized by a high degree of propeller twist and a distinctly narrow minor groove. Dat1p's accommodation of interruptions in its binding site suggest that the distinct features recognized by Dat1p are not the result of a structure that requires long (>9 bp) uninterrupted A·T tracts, making it likely that the distinct A·T minor groove recognition features are present in the model of A·T tracts generated from the existing crystallographic studies. These observations are consistent with the previously proposed modular nature of Dat1p/A·T DNA recognition.

Analysis of the nucleotide sequence of the yeast chromosome III shows that there are a total of 145 Dat1p binding sites: 63 uninterrupted and 82 interrupted A·T tracts which conform to the series described above. The sites are distributed throughout chromosome III, occurring roughly every 2.2 kb. A majority (60%) of the predicted protein coding genes contain Dat1p binding sites in their 5' intergenic regions. That a significant fraction of promoters do not contain Dat1p binding sites suggests that these sequences are not required for basal promoter function *in vivo*. The prevalence and location of Dat1p binding sites are consistent with a global role for Dat1p in gene expression.

ACKNOWLEDGEMENTS

We thank Perry Hall and Marisa Wagner for critical reading of this manuscript, and Michael Pierce for technical assistance. This work was supported by a grant from the NIH (GM45772).

REFERENCES

- Reardon, B. J., Winters, R. S., Gordon, D., and Winter, E. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 11327–11331.
- Phillips, S. E. V. (1994) *Annu. Rev. Biophys. Biomol. Struct.*, **23**, 671–701.
- Pabo, C. O., and Sauer, R. T. (1992) *Annu. Rev. Biochem.*, **61**, 1053–1095.
- Freemont, P. S., Lane, A. N., and Sanderson, M. R. (1991) *Biochem. J.*, **278**, 1–23.
- Churchill, M. E. A., and Travers, A. A. (1991) *Trends Biochem. Sci.*, **16**, 92–97.
- Latchman, D. S. (1990) *Biochem. J.*, **270**, 281–289.
- Struhl, K. (1989) *Trends Biochem. Sci.*, **14**, 137–141.
- Schleif, R. (1988) *Science*, **241**, 1182–1187.
- Lang, D., and Stamminger, T. (1994) *Nucleic Acids Res.*, **22**, 3331–3338.
- Starr, D. B., and Hawley, D. K. (1991) *Cell*, **67**, 1231–1240.
- Reeves, R., and Nissen, M. S. (1990) *J. Biol. Chem.*, **263**, 8573–8582.
- Churchill, M. E. A., and Suzuki, M. (1989) *EMBO J.*, **8**, 4189–4195.
- Solomon, M. J., Strauss, F., and Varshavsky, A. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 1276–1280.
- Gehring, W. J., Qian, Y. Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A. F., Resendez-Perez, D., Affolter, M., Otting, G., and Wuthrich, K. (1994) *Cell*, **78**, 211–223.
- Botfield, M. C., Jancso, A., and Weiss, M. A. (1994) *Biochemistry*, **33**, 8113–8121.
- Gehring, W. J., Affolter, M., and Burglin, T. (1994) *Annu. Rev. Biochem.*, **63**, 487–526.
- Laughon, A. (1991) *Biochemistry*, **30**, 11357–11367.
- Wolberger, C., Vershon, A. K., Liu, B., Johnson, A. D., and Pabo, C. O. (1991) *Cell*, **67**, 517–528.
- Kissinger, C. R., Liu, B., Martin-Blanco, E., Kornberg, T. B., and Pabo, C. O. (1990) *Cell*, **63**, 579–590.
- Winter, E., and Varshavsky, A. (1989) *EMBO J.*, **8**, 1867–1877.
- Karlin, S., Blaisdell, B. E., Sapolsky, R. J., Cardon, L. and Burge, C. (1993) *Nucleic Acids Res.*, **21**, 703–711.
- Nelson, H. C. M., Finch, J. T., Luisi, B. F., and Klug, A. (1987) *Nature*, **330**, 221–226.
- Coll, M., Frederick, C. A., Wang, F. H. -J., and Rich, A. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 8385–8389.
- DiGabriele, A. D., Sanderson, M. R., and Steitz, T. A. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 1816–1820.
- Kunkel, G. R., and Martinson, H. G. (1981) *Nucleic Acids Res.*, **9**, 6869–6888.
- Struhl, K. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 8419–8423.
- Iyer, V., and Struhl, K. (1995) *EMBO J.*, **14**, 2570–2579.
- Peck, L. J., and Wang, J. C. (1981) *Nature*, **292**, 375–378.