The formation of triple helical DNA has been evoked in several cellular processes including transcription, replication, and recombination. Using conventional and affinity chromatography, we purified from Saccharomyces cerevisiae whole-cell extract a 35-kDa protein that avidly and specifically bound a purine motif triplex (with a K$_D$ of 61 pM) but not a pyrimidine motif triplex or duplex DNA. Peptide microsequencing identified this protein as the product of the STM1 gene. Confirmation that Stm1p is a purine motif triplex-binding protein was obtained by electrophoretic mobility shift assays using either bacterially expressed, recombinant Stm1p or whole-cell extracts from stm1Δ yeast. Stm1p has previously been identified as G4p2, a G-quartet nucleic acid-binding protein. This suggests that some proteins actually recognize features shared by G4 DNA and purine motif triplexes, e.g. Hoogsteen hydrogen-bonded guanines. Genetically, the STM1 gene has been identified as a multicopy suppressor of mutations in several genes involved in mitosis (e.g. TOM1, MPT5, and POP2). A possible role for multiplex DNA and its binding proteins in mitosis is discussed.

It has long been recognized that, under the proper conditions, certain DNA sequences preferentially adopt a structure composed of three nucleic acid strands (1). Triple helical or triplex DNA is a thermodynamically favored structure characterized by a third pyrimidine-rich (Py triplex)$^\dagger$ or purine-rich (Pu triplex) DNA strand located within the major groove of a homopurine/homopyrimidine stretch of duplex DNA (reviewed in Ref. 2). Both intermolecular triplexes, where the third stand originates from a separate DNA molecule, and intramolecular triplexes (H-DNA), where the third stand originates from a proximal site on the same DNA molecule as its duplex acceptor, have been described. In intermolecular and intramolecular triplexes, stable interaction of the third strand is achieved through either specific Hoogsteen (Py triplex) or reverse Hoogsteen (Pu triplex) hydrogen bonding to the homopurine strand of the duplex, with the third strand adopting either a parallel (Py triplex) or antiparallel (Pu triplex) orientation relative to the homopurine acceptor. Base triplets in the pyrimidine motif include T$^+$AT and C$^+$GC, whereas those in the purine motif include G$^+$GC, A$^+$AT, and T$^+$AT. Because cytosine protonation requires acidic pH (3) and the G$^+$GC base triplet is the most stable in the purine motif (4), T-rich Py motif or G-rich Pu motif triplexes would be expected to predominate under physiological conditions.

Do triplexes occur in vivo? Although direct proof is lacking, long oligopurine tracts with triplex-forming potential are quite common in eukaryotic genomes, ranging from yeast to human (5). These tracts are distributed nonrandomly and are typically located near gene promoters, recombination hot spots, and matrix attachment regions (6, 7). Additionally, multiple lines of evidence have implicated intramolecular triplexes in several cellular processes, including transcription, replication, and recombination (8–13). Finally, monoclonal antibodies generated against triplex DNA were found to interact nonuniformly with metaphase chromosomes and interphase nuclei, preferentially staining centromeric regions (14, 15). Taken together, these data support the existence of triplex DNA at some point during the life cycle of a eukaryotic cell and suggest an important role for these structures in DNA-dependent biological processes.

If triplexes form in vivo, whether as required intermediates or as undesired side products of a necessary process, then cellular proteins might also exist that specifically recognize this particular DNA form. To date, four examples of triplex-binding proteins (3BPs) have been described in the literature. These include two reports of similar 55-kDa human proteins that exhibit binding specificity for Py triplex DNAAs (16, 17), our findings of several human proteins that specifically recognize a Pu triplex (18), and evidence that the Drosophila GAGA factor can bind to Py triplexes (but not Pu triplexes) containing a (GA(TC)$_{20}$ sequence (19). Using electrophoretic mobility shift assays (EMSA), we have also found evidence for Pu motif 3BPs in extracts from organisms ranging from bacteria to human.$^\ddagger$ These data suggest that 3BPs are present in all eukaryotes and that they play important cellular roles.

To better understand the biological roles of 3BPs, we sought to identify their corresponding genes. We chose the yeast Saccharomyces cerevisiae as our model system, given its completely sequenced genome and the wealth of biochemical and genetic information presently available for this organism (20, 21). Here we describe the purification and characterization of the major S. cerevisiae 3BP, y3BP1, and its identification as the product of the STM1 gene.

**EXPERIMENTAL PROCEDURES**

**DNAAs—Sequences and structures of duplex, triplex, and quadruplex DNA probes and competitor DNAs used in this study are shown in Fig. 1. Paosolated oligonucleotides, indicated by a “P” prefix in their name or by a “P~” appended to their sequence, contained a 4′-hydroxy-
methyl-4,5,8-trimethylpsoralen-hexyl (Glen Research) moiety attached to their 5' terminus. All oligonucleotides were purified by n-butanol precipitation (22). Those used in constructing duplex and triplex probes were further purified by denaturing PAGE.

Duplex probes and competitor DNAs were made by annealing equimolar amounts of the two oligonucleotides at room temperature. In the case of labeled probes, annealed duplexes were 3'-end-filled using the Klenow fragment of DNA polymerase and deoxyribonucleotides, including [α-32P]dATP. Duplex DNAs used in this study included the G/C-rich Pu duplex and the A/T-rich Py duplex (Fig. 1). Pu triplexes used in this study contained either the noncovalently associated tetrameric parallel oligonucleosome of the tetrameric parallel structure GL-tetramer, the intermolecular quadruplex tetramer ODN 1 dimer, and the intramolecular quadruplex ODN 1 pentamer. Pu triplex probes were further purified by denaturing PAGE.

For pyrimidine motif triplex formation, the Py duplex, TFO PODN 1, and a buffer composed of 25 mM Tris-HCl (pH 6.0), 20 mM MgCl₂, and 70 mM NaCl were used instead. Psoralenated TFOs were covalently attached to both strands of the duplex DNA following triplex formation by irradiation at 365 nm for 10 min at 0 °C with a 6-watt hand-held UV lamp. Under these conditions, greater than 90% of the probe is converted to photo-cross-linked triplex (24). G-quartet (G4)-containing competitors used in this study included the tetratomic parallel structure GL-tetramer, the intermolecular quadruplex tetramer ODN 1 dimer, and the intramolecular quadruplex ODN 1 pentamer. Each G4 DNA was formed following published procedures (25–27). All DNA probes and competitor DNAs were analyzed by PAGE prior to use.

**EMSA**—Protein mixtures containing 3BP1s were incubated for 30 min at 24 °C in a 10-μl volume containing buffer A (25 mM HEPES- Na⁺, pH 7.9, 50 mM KCl, 10% glycerol, 0.5 mM dithiothreitol), 1 mM MgCl₂, and 0.1 μM probe DNA, together with 2 μg of poly(dI-dC) carrier DNA or additional competitor nucleic acids as indicated. Resulting protein-probe complexes were resolved by nondenaturing PAGE at 7 V/cm for 90 min through a 5% acrylamide, 0.13% bisacrylamide gel containing 22 mM Tris borate, 0.5 mM EDTA. Probe-containing species were visualized by autoradiography and quantitated by a Storm 840 PhosphorImager (Molecular Dynamics).

**Stm1, a Purine Motif Triplex DNA-binding Protein**

**Yeast Extract Preparation**—Haploid yeast (S. cerevisiae, strain FY5S-α) was cultured in eight 2-liter flasks containing 500 ml of YPD (1% yeast extract, 2% peptone, 1% dextrose), harvested at midlog phase (L₆₀₈₉ = 0.8), and lysed by vortexing with glass beads according to published protocols (28). Proteins were extracted from 18.6 g of lysed cells in 46 ml of buffer B (50 mM NaH₂PO₄, pH 7.4, 1 mM EDTA, 1 mM MgCl₂, and 0.01% Nonidet P-40) (20). For pyrimidine motif triplex formation, the Py duplex, TFO PODN 1, and a buffer composed of 25 mM Tris-HCl (pH 6.0), 20 mM MgCl₂, and 70 mM NaCl were used instead. Psoralenated TFOs were covalently attached to both strands of the duplex DNA following triplex formation by irradiation at 365 nm for 10 min at 0 °C with a 6-watt hand-held UV lamp. Under these conditions, greater than 90% of the probe is converted to photo-cross-linked triplex (24). G-quartet (G4)-containing competitors used in this study included the tetratomic parallel structure GL-tetramer, the intermolecular quadruplex tetramer ODN 1 dimer, and the intramolecular quadruplex ODN 1 pentamer. Pu triplex probes were further purified by denaturing PAGE.

**Peptide Sequencing and Data Base Searches**—Purified y3BP1 (8 μg) obtained from the Mono S column (fraction 8) was fractionated on a 10% SDS-PAGE gel and visualized by brief Coomassie blue staining, and a 100-μm² piece of acrylamide containing the major protein band was excised from the gel. This was sent to the Harvard Microchemistry Facility, which determined the sequences of two tryptic peptides, VNQQGWGDG and DVSNLPSLA, by tandem mass spectrometry on a Finnigan LCQ Quadrupole Ion Trap Mass Spectrometer. These peptide sequences were used in a search of the Yeast Protein Data base (21). Both peptides mapped with complete identity to sequences within the protein encoded by the S. cerevisiae STM1 gene.

**RESULTS**

**Identification of Yeast Pu Triplex-binding Proteins**—To determine whether yeast have proteins that specifically recognize Pu triplex DNA, an EMSA of yeast whole-cell extract using a covalently bound Pu triplex probe (Fig. 1B) was performed. This triplex is based on the well characterized 19-mer triplex-forming oligonucleotide ODN 1 (5’-TGGTTGGTGGTGTT-GGGTT-3’), which demonstrates strong, sequence-specific binding with an antiparallel orientation to a corresponding double-stranded DNA target (4). Using 5’-psoralenated TFO (PODN 1) as a triplex target containing a presumptive 5´-TA-3´ sequence allows the formation of a photo-cross-linked triplex (X-Pu triplex) that remains intact even under conditions that normally promote triplex dissociation (22, 31). Incubation of 1 μM radiolabeled X-Pu triplex probe, 2 μg of poly(dI-dC) carrier DNA, and additional competitor nucleic acids as indicated with 4.5-μg proteins from a yeast whole-cell extract for 20 min at 80 °C.
most likely not the result of psoralen photoadduct recognition. Competition with equivalent molar excesses of Pu duplex probe had no effect on the quantities of C2 and C1 species observed (lanes 7 and 8). Likewise, competition with 1000-fold molar excesses of the individual oligonucleotides that comprise the Pu triplex, i.e. the G/A-rich or C/T-rich strands of the Pu duplex (lanes 9 and 10), or 1 μM concentrations of the G/A-rich or C/T-rich strands of the Pu duplex, had no appreciable effect on C2 or C1 complex formation. Taken together, these data demonstrated that yeast have proteins that specifically recognize a Pu motif triplex.

Purification of the Major Yeast 3BP—To further characterize proteins that bind Pu triplexes, we purified the protein(s) responsible for the major protein-triplex EMSA complex, C1. A combination of conventional, affinity, and high performance liquid chromatography was employed. Table I outlines this purification, whereas SDS-PAGE and EMSA analyses of the relevant protein fractions are shown in Fig. 3, A and B, respectively. Initially, whole-cell yeast extract was loaded onto a phosphocellulose column. The proteins responsible for C1 complex formation were eluted in the range of 450–550 mM KCl, as determined by an ultraviolet spectrophotometer using an X-Pu triplex covalently attached to a Sepharose 4B matrix. The y3BPs present in C1 eluted in the 1000 mM KCl step fraction. Note that the overall complexity of this fraction was not greatly changed through this chromatographic step, though the major 50-kDa contaminant, believed to be yeast cytoplasmic elongation factor 1, was effectively removed (Fig. 3A, compare lanes 4 and 5). This limited purification achieved with the X-Pu triplex-Sepharose column may be due to there being a mixture of Pu duplex and X-Pu triplex.

Fig. 1. Schematics of the probe and competitor DNAs used for characterizing yeast 3BPs. A, purine motif triplex (Pu triplex) composed of the Pu duplex and the TFO ODN 1. The box indicates Klenow end-filling bases. B, cross-linked Pu triplex (X-Pu triplex). P–, a psoralen-hexyl moiety covalently attached to the 5′-end of the oligonucleotide. C, the site of psoralen cross-linking. D, cross-linked pyrimidine motif triplex composed of the Py duplex and the TFO PODN 3. E, parallel-oriented tetraplex composed of four GL oligonucleotides. Guanines involved in G-quartet structures are boxed. F, intramolecular quadruplex composed of a single ODN 1 oligonucleotide.

room temperature allowed formation of protein-triplex complexes that could be resolved by nondenaturing PAGE and visualized by autoradiography (Fig. 2). Under these conditions, nearly complete shifing of the triplex probe into two slower mobility species (C1 and C2) could be observed (compare lanes 1 and 2). The major species, C1, had a relative mobility (Rf) compared with the free probe of 0.43 and comprised 87% of the total triplex probe, whereas the minor species, C2, had a Rf of 0.53 and comprised 11% of the total triplex probe. The specificity of these protein-triplex interactions was demonstrated by competition binding with other nucleic acids. As shown here, the C1 species was reduced to 64% and less than 3% of its normal amount when 100- and 200-fold molar excess unlabeled, noncovalent Pu triplex DNA was present in the binding reaction, respectively (lanes 4 and 5). Formation of complex C2 was inhibited to a similar but lesser extent under these conditions. Note that the competitor triplex did not contain a psoralen photo-cross-link, indicating that this competition was
TABLE I

| Step               | Volume (ml) | Mass (mg) | Activity (units × 10⁶) | SA (units/mg × 10³) | Yield (%) | Factor |
|--------------------|-------------|-----------|------------------------|---------------------|-----------|--------|
| Yeast extract      | 46.0        | 751       | 6.4                    | 8.5                 | 100       | 1      |
| Phosphocellulose   | 5.7         | 14.3      | 2.1                    | 147                 | 33        | 17     |
| Triplex-Sepharose  | 3.6         | 1.7       | 1.1                    | 647                 | 17        | 76     |
| Mono Q/Mono S      | 0.7         | 0.097     | 0.64                   | 6600                | 10        | 780    |

*Mass was measured by a Bradford dye binding assay.

*a One unit is equivalent to the amount of binding activity necessary to drive 50% of a labeled X-Pu triplex probe into complex C1 under our standard EMSA reaction conditions.

Fig. 3. Purification of y3BP1. A. Coomassie Blue-stained SDS-PAGE of y3BP1-containing protein fractions at various stages of purification. Fractions analyzed included 15 μg of yeast extract (Y, lane 3), 17 μg of phosphocellulose gradient pool (P, lane 4), 4.9 μg of X-Pu triplex-Sepharose pool (3S, lane 5), and 15-μl aliquots of Mono S fractions 6–10 (lanes 6–10). The band corresponding to y3BP1 is indicated at the right by a star. Molecular masses were determined by comparison with prestained (M1, lane 1) or standard (M2, lane 2) protein molecular weight markers and are indicated at the left. B. EMSA analysis of protein fractions through the course of y3BP1 purification. Fractions analyzed included 3.3 or 16 μg of yeast extract (Y, lanes 2 and 3), 0.01 or 0.1 μg of phosphocellulose gradient pool (P, lanes 4 and 5), 0.021 or 0.2 μg of Pu triplex-agarose pool (3S, lanes 6 and 7), or 1 μl of a 1:200 dilution of Mono S fractions 6–10 (lanes 8–12). C1 corresponds to the y3BP1-triplex complex.

Stm1, a Purine Motif Triplex DNA-binding Protein

sites on this column (estimated 35:65, respectively), the local high molar concentration of these sites present, or the absence of competing sites normally provided by poly(dI-dC) carrier DNA. Final purification was achieved using high performance liquid chromatography. X-Pu triplex affinity-purified C1 35Ps were loaded onto tandem Mono Q and Mono S FPLC columns. These were separated, and the proteins bound to the downstream Mono S column were eluted with a 100–1000 mM linear gradient of KCl. Triplex binding activity was concentrated in a single 350 mM KCl fraction (Fig. 3B), which contained a prominent (>80% total protein) 35-kDa polypeptide (Fig. 3A). Using proteins eluted from SDS-PAGE gel slices, we determined that a 35–40-kDa protein was responsible for the C1 shifted species (data not shown). Based on these data, we concluded that the 35-kDa polypeptide present in Mono S fraction 8, referred to as y3BP1, was the sole protein responsible for the major protein-DNA species observed with a cross-linked Pu triplex probe and yeast whole-cell extracts.

Characterization of Purified y3BP1—Whereas our previous experiments suggested that y3BP1 recognized an intact Pu triplex DNA species, this need not be the case with our covalent X-Pu triplex probe. It is possible that the psoralen photoadduct itself, or a change it induces in the duplex DNA structure, is actually being recognized, as might be expected for a protein involved in DNA repair (33). Alternatively, because noncovalent Pu triplexes are inherently unstable under our standard gel electrophoresis conditions (absence of Mg²⁺ in the gel buffer) (18), our probe might be expected to adopt a structure composed of a single-stranded DNA attached at its 5'-end to a duplex DNA, which is somewhat reminiscent of structures found in DNA replication (34). To verify that an intact Pu triplex was actually being recognized by y3BP1, EMSAs were performed with different labeled probes, including the covalent X-Pu triplex, the noncovalent Pu triplex, and the Pu duplex. Note that in this experiment, electrophoresis was performed at 4°C and at a lower voltage (4 V/cm) to maintain maximal stability of the y3BP1-DNA complex. As shown in Fig. 4, the X-Pu triplex remained stable throughout electrophoresis, as did the y3BP1-triplex complex C1 (lanes 1 and 2). Also, as expected, the noncovalent Pu triplex probe dissociated under these conditions, quantitatively yielding Pu duplex (lane 3).

However, a significant fraction of the labeled probe (15%, as compared with 60% with the X-Pu triplex probe) remained intact as part of complex C1, when purified y3BP1 was present in the binding reaction (lane 4). This was likely not the result of binding to the Pu duplex part of the dissociated Pu triplex probe, because no interactions between y3BP1 and the Pu duplex were observed under these conditions (lane 6). Taken together, these and the competition data from Fig. 2 indicate that y3BP1 recognizes an intact purine motif triplex.

The y3BP1 protein may bind a Triplex specifically, but does it do so with high affinity? To measure its dissociation binding constant, a constant concentration of y3BP1 was incubated in a near physiological buffer (25 mM HEPES-Na+, pH 7.9, 50 mM KCl, 10% glycerol, 0.5 mM dithiothreitol, and 1 mM MgCl₂) with a fixed concentration of labeled, covalent X-Pu triplex probe and increasing concentrations of unlabeled, X-Pu triplex DNA. Protein-DNA complexes resulting from these reactions were analyzed by EMSA. Plotting the ratio of bound to free triplex DNA as a function of C1 concentration, we found that y3BP1 exhibited an apparent Kd of 61 μM for Pu triplexes (Fig. 5). We also determined that our Mono S fraction 8 contained 1.1 μM active y3BP1, which is comparable to the concentration of a 35-kDa protein estimated from the Coomassie Blue-stained SDS-PAGE gel (3.2 μM). This binding constant is similar to those of many duplex DNA-binding proteins (e.g.
valently attached Pu triplexes. EMSAs were performed on 10- 

binding reactions (10-

binding of y3BP1 to X-Pu triplex DNA. reaction mixtures containing 0.1 nM labeled probe, either the covalently 

concentration of active y3BP1 in each reaction was determined to be 0.27

[C1]), to the concentration of complex C1. From this analysis the con- 

were carried out with a 0.1 nM-labeled covalent X-Pu triplex probe, 0.69

lanes 3 and 4), or the corresponding Pu duplex alone (lanes 5 and 6). Minus and plus refer to the absence or presence of 0.69 ng of y3BP1 (Mono S fraction 8) in the reaction mixtures, respectively.

Though yeast 3BP1 may have a high affinity for Pu motif

triplexes, it is possible this structure is not the true target of this protein in vivo. To better understand the binding specificity of y3BP1, competition experiments were undertaken with a variety of different DNA structures. These included several quadruplex DNAs (both parallel tetraplexes and antiparallel hairpin dimers and intramolecular quadruplexes), triplex DNAs (both Pu motif and Py motif), duplex DNAs (both A/T- and G/C-rich), and the single-stranded TFO ODN 1. Their sequences and structures are shown in Fig. 1. Binding reactions were modified in some cases to maintain the integrity of these DNA structures. For example, competition experiments with single-stranded ODN 1 were performed in buffer A containing HEPES-LiCl instead of HEPES-NaCl and KCl, respectively, to minimize formation of G4-containing species with this G-rich oligonucleotide (26, 27). Quantitation from a series of competition experiments is shown graphically in Fig. 6, with apparent competitor concentrations necessary to observe a 50% decrease in complex C1 formation (EC50) provided in Table II. As shown here, the covalent Pu triplex had the highest affinity for y3BP1 followed by the parallel G4 tetraplex GL-tetramer, which had a 5.6-fold lower affinity. Surprisingly, y3BP1 exhibited significant, albeit lower (33-fold less than the Pu triplex) binding affinity to the Pu duplex. However, this binding might be explained by the fortuitous formation of a small amount of a purine motif triplex DNA, containing two G/A-rich strands and one C/T-rich strand from the Pu duplex, under our incubation conditions. Even lower binding affinities were observed with the Py triplex and the G4 DNAs ODN 1 dimer and ODN 1 monomer, indicating that not all DNA structures possessing high negative charge density are avidly bound by y3BP1. Single-stranded ODN 1 bound y3BP1 with very low

[20], a sequence of as few as five amino acid residues can be
determined to the gene responsible for y3BP1, we obtained amino

acid sequence information from the purified protein. Given that the complete nucleotide sequence of S. cerevisiae is available (20), a sequence of as few as five amino acid residues can be sufficient to define a single gene. Tryptic fragments from SDS-

PAGE-purified y3BP1 were sequenced by tandem mass spectrometry, and two amino acid sequences, VNQGWGDDK and DVSNLPSLA, were identified. A search of the Yeast Proteome
probe. This demonstrates that the product of the complex. Neither extract contained appreciable amounts of similar relative mobility as native y3BP1 (Fig. 7). Recombinant Stm1p also demonstrated a Pu triplex binding specificity identical to that of the native protein, as shown by the concentration of competitor DNA that reduces C1 complex formation 50% when a 0.1 nM-labeled X-Pu triplex probe and 0.27 nM purified y3BP1 were present in an EMSA binding reaction.

Data base (21) showed that these sequences mapped exactly to amino acids 106–114 and 265–273, respectively, of the protein encoded by the STM1 gene (30). As known as MPT4 (35) and STO1, this gene codes for a 30-kDa, basic (pI = 9.8) protein. This protein was originally found to have a specific affinity for some quadruplex nucleic acids and was referred to as G4p2 (36).

Two approaches were used to verify that the STM1 gene product was responsible for the primary yeast protein-Pu triplex complex, C1. In the first approach, the entire STM1 open reading frame was cloned into the bacterial expression vector pV2a (29). This allowed the production of a recombinant Stm1 protein as an N-terminal oligohistidine fusion. Bacterially expressed (His)6-Stm1p was purified to homogeneity by immobilized metal affinity chromatography and Mono S FPLC. When assayed by EMSA using a covalent X-Pu triplex probe, (His)6-Stm1p demonstrated a single protein-DNA complex with an identical relative mobility as native y3BP1 (Fig. 7). Recombinant Stm1p also demonstrated a Pu triplex binding specificity similar to that of the native protein, as shown by the concentrations of Pu triplex and Pu duplex competitor required to demonstrate a 50% decrease in complex C1 formation (220 pM and 19 nM, respectively). That a greater concentration of (His)6-Stm1p than native Stm1p was required most likely reflects the limited specific activity of our recombinant protein preparations. These data directly demonstrate that Stm1p is a bona fide triplex-binding protein and that bacterially expressed Stm1p retains its DNA binding properties. This also suggests that post-translational modifications absent in bacteria (e.g., proper phosphorylation, acetylation, and methylation) are not essential for the specific recognition of triplex DNA by Stm1p.

As a second approach to verifying the involvement of Stm1p in the major y3BP-Pu triplex complex, we obtained S. cerevisiae strain A1454 (35), which is deficient in Stm1p. Disruption of the STM1 gene is nonlethal in haploid yeast grown in rich medium, with the only observed phenotypic change for stm1Δ yeast being a slightly increased doubling time at elevated temperatures in media containing nonfermentable carbon sources (30). As shown in Fig. 8, proteins were present in wild-type but not stm1Δ yeast extracts that generate the C1 protein-triplex complex. Neither extract contained appreciable amounts of proteins that specifically bound the corresponding Pu duplex probe. This demonstrates that the product of the STM1 gene is responsible for activity of the major Pu triplex-binding protein in yeast extracts and is consistent with Stm1p being a triplex-binding protein. Note that other, minor protein-triplex complexes (C0 and C2) appeared in both extracts and that the amounts of these minor complexes approximately doubled in reactions containing the stm1Δ yeast extract. The former observation would indicate that Stm1p is not present in these other complexes, while the latter is suggestive of a possible functional compensation by other yeast proteins that bind triplex DNA.

**DISCUSSION**

Using a well defined purine motif triplex and EMSA, we identified at least two different proteins in S. cerevisiae whole cell extracts that specifically recognized triple helical DNA. Using a combination of conventional and affinity chromatography, we purified a 35-kDa protein, y3BP1, that was responsible for the major protein-triplex complex C1. Microsequencing of this protein and a comparison with the Yeast Proteome Data base indicated that y3BP1 was encoded by the STM1 gene. That Stm1 protein is a bona fide triplex-binding protein and responsible for the C1 complex was verified by bacterially expressed recombinant Stm1p and whole-cell extracts from a stm1Δ mutant yeast strain, respectively.

**TABLE II**

| DNA                  | EC50 (nM) |
|----------------------|-----------|
| X-Pu triplex         | 4.3 x 10^-10 |
| GL tetramer          | 2.4 x 10^-9  |
| Pu duplex            | 1.4 x 10^-8  |
| Py triplex           | 2.5 x 10^-8  |
| ODN 1 dimer          | 9.5 x 10^-6  |
| ODN 1 monomer        | 2.4 x 10^-7  |
| Single-stranded ODN 1| 3.2 x 10^-6  |
| Py duplex            | >1 x 10^-5   |

*a EC50 refers to the concentration of competitor DNA that reduces C1 complex formation 50% when a 0.1 nM-labeled X-Pu triplex probe and 0.27 nM purified y3BP1 were present in an EMSA binding reaction.

As shown in Fig. 8, proteins were present in wild-type but not stm1Δ yeast extracts that generate the C1 protein-triplex complex. Neither extract contained appreciable amounts of proteins that specifically bound the corresponding Pu duplex probe. This demonstrates that the product of the STM1 gene is responsible for activity of the major Pu triplex-binding protein in yeast extracts and is consistent with Stm1p being a triplex-binding protein. Note that other, minor protein-triplex complexes (C0 and C2) appeared in both extracts and that the amounts of these minor complexes approximately doubled in reactions containing the stm1Δ yeast extract. The former observation would indicate that Stm1p is not present in these other complexes, while the latter is suggestive of a possible functional compensation by other yeast proteins that bind triplex DNA.
Previous studies have described 3BPs in human cell extracts, but their conclusive identification with particular gene products had not been achieved (16–18). Similarly, a portion of the *Drosophila* GAGA factor has been shown to bind triplex DNA with less affinity than its duplex target (19), though it is debatable whether crude preparations of the native protein would preferentially recognize triplex DNA under physiological conditions. Thus, the work presented in this paper is the first reported purification of a Pu motif triplex-binding activity from crude cell extracts and the conclusive identification of its gene.

**STM1**, also referred to as **MPT4** and **STO1**, was identified through genetic screens as a multicopy suppressor of temperature-sensitive *tom1*, *htr1*, and *pop2* mutants (30, 35). **TOM1** encodes a ubiquitin ligase that regulates activation of the ADA histone acetyltransferase A complex and is required for G2/M progression (37). **POP2** encodes a subunit of the CCR4 general transcriptional complex, which regulates the expression of a number of genes during the late mitotic part of the cell cycle (38). **HTR1** (also known as **MPT5**) encodes a protein involved in mating pheromone-induced G1 arrest and for progression through G2/M (39). **MPT5** has been found to affect the lifespan of yeast cells, presumably by affecting the strength of transcriptional silencing at telomeres compared with the ribosomal DNA locus (40). Interestingly, **MPT5** is also a multicopy suppressor of **POP2** (35). Because high copy numbers of **STM1** were required to produce even partial phenotype suppression for *tom1*, *htr1*, and *pop2* mutations, it is likely Stm1p either acts far downstream of these proteins (e.g. for Tom1p) or is not a fully functional substitute for them (e.g. for Mpt5p and Pop2p). This and the viability of *STM1*Δ yeast (30) make it difficult to genetically determine an exact function for Stm1p. However, common themes among these proteins, e.g. transcriptional regulation and mitosis, strongly suggest that Stm1p has a role in these processes.

Evidence for a protein’s biological role can sometimes be found through an analysis of its pattern of expression under different conditions. For a baseline, **STM1** is moderately well expressed in yeast, averaging 51 copies/cell during log phase growth as determined by serial analysis of gene expression (41). However, as determined by microarray analysis, the level of **STM1** expression does not change significantly during the cell cycle (42), as a function of different cell mating type (43), or in response to stimuli such as heat shock (43) or treatment with DNA alkylators (44). The largest changes reported were observed upon a shift from fermentation to respiration (4.5-fold lower expression) and during sporulation (4.2-fold lower), though values of this magnitude were relatively common among all yeast genes investigated (45, 46). Likewise, an analysis of the **STM1** promoter region (47) provided few clues as to its possible regulation, because sites for relatively common transcription factors (e.g. GCN4 and GCR1) were primarily identified. More striking was the observation of three consensus **PHR1** upstream activator sequence (UASPHR) sites within 250 base pairs of the **STM1** translation start site, given that UASPHR sites are a hallmark of genes involved in nucleotide excision repair and recombination (e.g. RAD1, RAD4, RAD23, and RAD50) (48). However, it should be noted that large changes in expression are not necessarily a hallmark of an important DNA-binding protein. For example, the Cbf1 protein, which binds to an element in the centromere and is involved in mitosis (49), does not significantly change its expression in any of the aforementioned circumstances (42–46) nor has noteworthy transcription factor binding sites within its proposed promoter region been found (47).

Stm1 protein was first biochemically identified as G4p2, a yeast protein that exhibits specific affinity for quadruplex nucleic acids (36). Quadruplex nucleic acids are four-stranded, right-handed helical structures composed of stacked pairs (or greater) of G-quartets, square planar arrays of guanines Hoogsteen hydrogen bonded to one another (reviewed in Ref. 50). Quadruplexes can be composed of four parallel-oriented nucleic acid strands (see Fig. 1D for an example), two folded nucleic acid strands (Fig. 1E), or with certain G-rich sequences, a unmolecular species composed of only intramolecular G-quartets (Fig. 1F). Quadruplexes form with low to moderate kinetics under physiological conditions, and sequences exist within the yeast genome that can form quadruplexes *in vitro*. Examples include the 3′-G-rich single-stranded extensions present in the telomeres on chromosome ends (51) and transcripts of a G-rich 26 S rRNA gene (52). It is not conclusively known whether quadruplexes exist *in vivo*, nor whether the recognition of such structures is the biological function of proteins like Stm1p. In our studies, we found that Stm1p bound a Pu motif DNA triplex better than a DNA tetraplex and considerably better than a dimeric or a monomeric quadruplex DNA. Thus our quadruplex data are consistent with those described by Frantz and Gilbert (36). However, the importance of higher triplex binding affinity should not be exaggerated, given that the identity of the putative nucleic acid recognized by Stm1p *in vivo*, let alone its exact structure, is not known. Efforts to determine its identity, through cross-linking and immunoprecipitation experiments, are presently underway. It is interesting to speculate that structural elements common to both quadruplexes and Pu motif triplexes, e.g. a large negative charge density or reverse Hoogsteen hydrogen bonded guanines, might be recognized by 3BPs like Stm1p. However, none of the human Pu motif 3BPs so far reported bind G4 DNAs with high affinity (18). Likewise, not all quadruplex-binding proteins recognize Pu motif triplexes. For example, extracts made from an *arc1Δ* yeast strain (53), which does not produce the quadruplex-binding protein and cofactor for methionyl- and glutamyl-tRNA synthetases G4p1 (54), did not demonstrate the loss of any protein-Pu triplex complex observable by EMSA. Thus the relationship between quadruplex- and triplex-binding proteins, like the existence of their supposed binding sites *in vivo*, remains an open question.

Following its predicted N-terminal acetylation (55), mature Stm1p should be composed of 272 amino acids, with a calculated molecular mass of 29,903 Da (21). From an analysis of its amino acid composition, Stm1p would be predicted to be a soluble protein with a pI of 9.8. These characteristics stem from the unusual abundance of basic residues (19% of the total amino acids) and the significant paucity of the major hydrophobic residues leucine, valine, isoleucine, phenylalanine, and methionine (15%) in Stm1p compared with other yeast proteins (56). This high percentage of basic residues would not be unexpected for a triplex- and/or quadruplex-binding protein, given the necessity to complement the high negative charge density present on these multistranded nucleic acids. Interestingly, basic residues in Stm1p are distributed throughout the protein in patches of only slightly greater density than would be expected randomly (Fig. 9). This contrasts with studies done with basic oligopeptides, which indicated that high positive charge densities were best for stabilizing intermolecular and intramolecular triplexes (57, 58). However, without knowing the three-dimensional structure of Stm1p, it is difficult to predict the exact charge density on any part of its surface. Stm1p should be a nuclear protein, given the presence of two overlapping nuclear localization domains, one a pal7 motif and the other a bipartite nuclear localization domain, located between

---

4 L. D. Nelson, unpublished observations.
Stm1p

FIG. 9. Schematic representation of S. cerevisiae Stm1p. Shaded boxes indicate lysine- and arginine-rich domains. Dashed lines indicate other functional domains.

amo acids 33 and 50 (59). Otherwise, the only other noteworthy feature of Stm1p is a region from amino acids 119–141 that is very rich in alanines (48%) and acidic residues (39%) and is postulated to adopt a coiled-coil α-helical conformation characteristic of many protein-protein interaction domains (60).

Taken together, these data would suggest that multiple regions of Stm1p might be involved in making contacts with triple helical DNA and that Stm1p might form a complex with other proteins at some point in its life cycle.

Stm1p has been estimated to be present at about 35,000 copies/yeast cell (36). We found that purified Stm1p has a measured affinity for Pu triplex DNA of 61 pM under our standard reaction conditions. This value is quite reasonable for a moderately abundant DNA-binding protein. However, the conditions we employed for our binding reactions were those that demonstrated maximal binding affinity in vitro and are not those expected in vivo. We have investigated purified Stm1p binding under more physiological conditions (140 mM KCl, 12 mM MgCl₂, 1 mM spermine, pH 7.5) and found that it retained substantial binding activity (~50%) under these conditions. This would suggest that Stm1p could bind Pu triplexes in vivo, should they occur.

Though we have observed 3BPs in species ranging from S. cerevisiae to human, are there any direct analogs of Stm1p in other organisms? A BLASTP (61) search of the nonredundant protein data base identified two homologous proteins in other organisms. To human, are there any direct analogs of Stm1p in other organisms? A BLASTP (61) search of the nonredundant protein data base identified two homologous proteins in other organisms. At present, we can only speculate about the possible biological roles of triplexes and 3BPs. Two 3BPs have been identified so far, the Drosophila GAGA factor (19) and S. cerevisiae Stm1p. As mentioned above, genetic studies have shown STM1 to be a multipurpose suppressor of several genes that have roles in transcriptional controls and progression through mitosis. Interestingly, GAGA, a well known DNA-binding transcription factor involved in the regulation of many Drosophila genes (63), has also been found to be a constituent of centromeric heterochromatin in mitotic chromosomes (64), whose absence has been shown to result in mitotic defects including failures in chromosome condensation and segregation (65). We have recently identified a second S. cerevisiae gene, CDPI, which also encodes a 3BP. Genetically, CDPI is a complement of CBF1, a centromere-binding protein involved in mitosis (49, 66). A yeast strain harboring a deletion in CDPI exhibited defects in chromosomal segregation and a temperature-sensitive arrest in G2/M (66).

Biochemically, Cdp1p is part of a general transcription factor complex containing Cdc73p and Pat1p that interacts directly with RNA polymerase II (67, 68). Together these findings are consistent with 3BPs having a role in cell cycle progression, specifically mitosis. Interestingly, triple helical DNA has also been implicated in mitosis. Agazie et al. (69) found that when anti-triplex monoclonal antibodies were introduced into synchronized myeloma cells, they had their greatest effect on cell growth when introduced at the end of S phase and during G2 (69). From these studies they proposed a model in which transmolecular triplex formation and dissociation were involved in the processes of chromosome condensation and decondensation, respectively. Transmolecular triplexes are structures that involve triplex-forming single-stranded DNA and duplex acceptor from either different DNA molecules or, unlike intramolecular triplexes, from distal sites on the same DNA. From their model, these antibodies would bind triplexes in condensed DNA but would lack the ability to be readily reversible, thereby retarding chromosome decondensation and cell cycle progression. In extension, a possible role for 3BPs in this process might be to provide a reversible means of alternatively protecting and deprotecting triplex DNA from the actions of other proteins (e.g. naturases and helicases) responsible for triplex formation and dissociation. Note that it need not be necessary to limit this model to only triplexes, given that association between different DNAs can also be achieved through G4 (e.g. hairpin dimer) formation. Thus transmolecular G4 formation may be the actual initiator of chromosome condensation, and G4 DNA may be the true target of 3BPs in vivo.

Acknowledgments.—We thank Y. Kikuchi for the STM1-containing plasmid pFU151, A. Sakai for the yeast strains A1454 (stm1Δ) and A1455 (STM1 wt), and G. Simos for the yeast strain arc1Δ. We also wish to thank Sharon Roth and Michele Sawadogo for critically reading the manuscript.

REFERENCES

1. Felsenfeld, G., Davies, D. R., and Rich, A. (1957) J. Am. Chem. Soc. 79, 2023–2024
2. Frank-Kamenetskii, M. D., and Mirkin, S. M. (1995) Annu. Rev. Biochem. 64, 1–25
3. Singleton, S. F., and Dervan, P. B. (1992) Biochemistry 31, 10995–11003
4. Beal, P. A., and Dervan, P. B. (1992) Nucleic Acids Res. 20, 2773–2776
5. Behe, M. (1995) Nucleic Acids Res. 23, 689–695
6. Wells, R. D., Collier, D. A., Hanvey, J. C., Shimizu, M., and Wohlrab, F. (1988) FASEB J. 2, 2939–2949
7. Boulikas, T. (1996) J. Cell. Biochem. 60, 287–296
8. Kohwi, Y., and Kohwi-Sigamatsu, T. (1991) Genes Dev. 5, 2545–2554
9. Chen, S., Supakar, P. C., Vellanoweth, R. L., Song, C. S., Chatterjee, B., and Roy, A. K. (1997) Mol. Endocrinol. 11, 3–15
10. Bianchi, A., Wells, R. D., Heintz, N. H., and Caddle, M. S. (1990) J. Biol. Chem. 265, 21769–21796
11. Baran, N., Lapidot, A., and Manor, H. (1991) Proc. Natl. Acad. Sci. U. S. A. 88, 507–511
12. Kohwi, Y., and Panchenko, Y. (1993) Genes Dev. 7, 1766–1778
13. Rooney, S., and Moore, P. D. (1995) Proc. Natl. Acad. Sci. U. S. A. 92, 2141–2144
14. Burkholder, G. D., Latimer, L. J. P., and Lee, J. S. (1988) Chromosoma (Berl.) 97, 185–192
15. Burkholder, G. D., Latimer, L. J. P., and Lee, J. S. (1991) Chromosoma (Berl.) 101, 1–18
16. Kiyama, R., and Camerini-Otero, R. D. (1991) Proc. Natl. Acad. Sci. U. S. A. 88, 10450–10454
17. Gu¨geys, A.-L., Prasseuth, D., and Hel¨ene, C. (1997) J. Mol. Biol. 267, 289–298
18. Musso, M., Nelson, L. D., and Van Dyke, M. W. (1998) Biochemistry 37, 3086–3095
19. Jimenez-Garcia, E., Vaquero, A., Eginas, M. L., Soliva, R., Orozco, M., Bernue¨s, J., and Azorite, F. (1998) J. Biol. Chem. 273, 25440–25448
20. Goffeau, A., Barrett, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Gilbert, F., Hoheisel, J. D., Jag, C., Johnston, M., Lesina, E. J., Mewes, H. W., Moriyama, Y., Philipp, S., Teitelman, H., and Oliver, S. G. (1996) Science 274, 563–567
21. Hodges, R. E., McKee, A. H., Davis, B. P., Payne, W. E., and Garrels, J. I. (1999) Nucleic Acids Res. 27, 69–73
22. Sawadogo, M., and Van Dyke, M. W. (1991) Nucleic Acids Res. 19, 674
23. Musso, M., and Van Dyke, M. W. (1995) Nucleic Acids Res. 23, 2320–2327

5 M. Musso, L. D. Nelson, G. Bianchi-Scarrà, and M. W. Van Dyke, manuscript in preparation.

6 M. Musso, unpublished observations.
The Yeast *STM1* Gene Encodes a Purine Motif Triple Helical DNA-binding Protein
Laura D. Nelson, Marco Musso and Michael W. Van Dyke

*J. Biol. Chem.* 2000, 275:5573-5581.
doi: 10.1074/jbc.275.8.5573

Access the most updated version of this article at [http://www.jbc.org/content/275/8/5573](http://www.jbc.org/content/275/8/5573)

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 68 references, 23 of which can be accessed free at [http://www.jbc.org/content/275/8/5573.full.html#ref-list-1](http://www.jbc.org/content/275/8/5573.full.html#ref-list-1)