J. Biochem. 125, 809-817 (1999)

Recognition of Multiple Patterns of DNA Sites by *Drosophila* Homeodomain Protein Bicoid[1]

Dong Yuan,[2] Xiuguang Ma,[3] and Jun Ma[4]

Division of Developmental Biology, Children's Hospital Research Foundation, and Graduate Program in Developmental Biology, University of Cincinnati College of Medicine, 3333 Burnet Avenue, Cincinnati, Ohio 45229-3039, USA

Received October 19, 1998; accepted January 7, 1999

Our previous studies demonstrated that the *Drosophila* homeodomain protein, Bicoid (Bcd), binds DNA cooperatively. In this study, we determined the patterns of adjacent DNA sites required for cooperative recognition by Bcd. Our *in vitro* selection and biochemical experiments demonstrated that Bcd binds preferentially to both head-to-head and tail-to-tail symmetric sites that are separated by short spacing. An increase in the spacing reduces the strict requirement of symmetric patterns of adjacent sites, permitting Bcd to recognize tandem repeat sites cooperatively. Our further experiments *in vivo* showed that the only pair of optimally spaced symmetric Bcd sites in a *hunchback* (*hb*) enhancer element contributes the most to transcriptional activation by Bcd, demonstrating the biological importance of the binding site patterns revealed by our *in vitro* selection studies.

Key words: Bicoid, DNA binding, *Drosophila*, gene activation, *in vitro* selection.

Control of gene expression requires specific interactions between regulatory protein molecules and their DNA binding sites (*1-3*). How individual sites are arranged with respect to each other can affect both the specificity and efficiency of these interactions, leading to dramatically different biological consequences. This is best illustrated by the nuclear receptor superfamily of transcription factors (*4, 5*). The core half-site elements that are in tandem or symmetric arrangements with different spacing are recognized by different members of the superfamily in response to different physiological stimuli. Such specific protein-DNA interactions are facilitated by interactions between these transcription factors, both homodimeric and heterodimeric, involving multiple interaction surfaces.

Homeodomain proteins play an important role in controlling the expression of genes required for many different biological processes, such as cell type specification in yeast and embryonic pattern formation in animals (*6-8*). A homeodomain contains a helix-turn-helix motif which directs specific DNA recognition (*9-11*). Recent studies showed that DNA recognition by homeodomain proteins can also be influenced by the spatial arrangements of their DNA sites. For example, DNA recognition by the homeodomain of the *Drosophila* paired protein (Prd) requires symmetric sites separated by a specific spacing (*12*). In addition, cooperative DNA binding by the homeodomain proteins, labial (Lab) and extradenticle (Exd), requires a specific arrangement of their DNA sites (*13, 14*). In yeast, different arrangements of DNA sites allow homeodomain protein $\alpha 2$ to interact with different protein partners to specify different cell types (*15*).

The *Drosophila* homeodomain protein, Bicoid (Bcd), is required for specifying the anterior structures during early embryonic development (*16, 17*). Embryos from females lacking the *bicoid* (*bcd*) gene fail to develop a head of thorax (*18, 19*). Several zygotic genes, including *hunchback* (*hb*), *Kruppel* (*Kr*), *knirps* (*kni*), and *even-skipped* (*eve*), have been shown to be directly activated by Bcd (*20-23*). It has been proposed that cooperative DNA recognition by Bcd multiple sites found in the enhancer elements of these target genes is an important mechanism for proper target gene regulation (*19-26*). However, no systematic analysis has been conducted to determine how individual Bcd sites should be aligned with each other for cooperative Bcd recognition.

In vitro selection (*27*) is a powerful approach for determining the patterns of DNA sites recognized by regulatory proteins (*12, 28*). It was reported previously that when the homeodomain of Bcd was used in such an *in vitro* selection experiment, only single sites were isolated, generating a consensus sequence of TAATCC (*29*). One possibility for the failure to select multiple sites in those experiments may have been that the homeodomain of Bcd is unable to bind DNA cooperatively (*30*). Here, we use an intact recombinant Bcd protein and an *in vitro* selection approach to identify preferred patterns of Bcd binding sites. Both symmetric tail-to-tail and head-to-head patterns of DNA sites separated by short spacing have been isolated from random sequences, suggesting preferential recognition of these

[1] This work was supported in part by a grant from NIH to J.M. X.M. was a Ryan Fellow of the University of Cincinnati, and J.M. was a recipient of an American Cancer Society Junior Faculty Research Award.

Present addresses: [2] Department of Internal Medicine, Danbury Hospital, 24 Hospital Avenue, Danbury, CT 06810, USA; [3] Department of Internal Medicine, New Hanover Regional Medical Center, 2131 South 17th Street, Wilmington, NC 28402, USA.

[4] To whom correspondence should be addressed. Tel: +1-513-636-7977, Fax: +1-513-636-4317, E-mail: jun.ma@chmcc.org

sites by Bcd. Additional experiments *in vivo* show that the only pair of optimally spaced head-to-head Bcd sites in a *hb* enhancer element plays a most critical role in supporting gene activation by Bcd, demonstrating the biological importance of our *in vitro* selected patterns of binding sites.

MATERIALS AND METHODS

In Vitro Selection of Bcd Binding Sites—Oligonucleotides comprising 48 and 13 random nucleotide sequences were used for the selection of symmetric tail-to-tail and head-to-head sites, respectively. Both random sequences are flanked by two 20 nucleotide defined sequences for PCR and subsequent cloning. The double-stranded DNA was generated by annealing the primer, 5′AGTCGTGCGTCGTCTAGAG3′ (for the selection of tail-to-tail sites) or 5′GGCTGAGTCTGAACGGATCC3′ (for the selection of head-to-head sites), with the corresponding templates, followed by extension with Taq or Klenow DNA polymerase. The double-stranded DNA was purified by gel electrophoresis and incubated with a recombinant Bcd protein expressed in Sf-9 cells (*26*). The experimental procedure of immunoprecipitation for separating protein-DNA complexes is similar to that of co-immunoprecipitation described previously (*26, 30*). Briefly, protein-DNA complexes were separated with Bcd antibodies (*31*). After incubation on ice for 1 h, pre-washed and equilibrated *Staphylococcus aureus* cells (Calbiochem) were added, followed by further incubation on ice for 1 h. The reaction mixtures were then diluted to a final volume of 100 μl in Buffer B (*26*), and precipitated by spinning for 3 min at 4°C, and then the pellet was gently washed three times with 100 μl Buffer B. The samples were then heated for 10 min at 85°C, followed by spinning for 3 min at 4°C. DNA in the supernatant was amplified by PCR using primers 5′AAGTCGTGCGTCGTCTAGAG3′ and 5′ATGCCCTTGACAGTCTCGAG3′ for the selection of head-to-head sites, or primers 5′CGCGACGCTCAGTAAGCTTG3′ and 5′GGCTGAGTCTGAACGGATCC3′ for the selection of tail-to-tail sites. PCR products were purified by gel electrophoresis, and one-fifth of each product was used for subsequent rounds of binding, immunoprecipitation and PCR amplification. For the selection of tail-to-tail sites, the products of the 14th cycle were digested with *Xho*I and *Xba*I, and then cloned into the *Xho*I/*Xba*I sites of pBluescript KS(−). For the selection of head-to-head sites, the PCR products of the 14 cycle were digested with *Bam*HI and *Hin*dIII, and then cloned into the *Bam*HI/*Hin*dIII sites of pBluescript KS(−). All oligonucleotides were synthesized at the University of Cincinnati DNA Core Facility.

Plasmids Construction—pDY82 contains tail-to-tail double sites separated by 8 base pairs, while pDY83 contains the same sequence as pDY82 except for a change of 4 base pairs at one of the Bcd sites. These two plasmids were constructed by inserting double-stranded oligonucleotides at the *Pst*I/*Sal*I sites of pBluescript KS(−). pDY66 and 67 contain head-to-tail and head-to-head doubles sites separated by 3 base pairs, respectively. pDY69 contains the same sequence as pDY67 except for a change of 2 base pairs at one Bcd site. All three plasmids were constructed by inserting double-stranded oligonucleotides at the *Bam*HI/*Hin*dIII sites of pBluescript KS(−). The head-to-head sequence used in the DNase I footprint experiment shown in Fig. 2D was carried on pDY40-1, which was generated by removing the *Hin*dIII-*Sal*I fragment from pDY67 and then circularization after Klenow treatment. The lower site in pDY40-1 was changed from TAAGCT to GAAGCT; this site and the upper site appeared to be protected by Bcd at the same protein concentration. pDY72, which is based on pBluescript KS(−) and contains a single site, TAATCC, was used in the experiment shown in Fig. 2D. pMAX7 was described previously (*26*). pMAX12 and pMAX13 containing site A1 in different orientations relative to X1 were constructed by inserting a double-stranded oligonucleotide composed of 5′TCGAACGTAATCCCC3′ and 5′TCGAGGGGATTACGT3′ at the *Xho*I site of pMAX7. In both constructs, A1 and X1 are separated by 36 base pairs (end-to-end).

Reporter plasmids pDY88 and pDY89 contain mutations at sites X3s and X3t in a *hb* enhancer element (−298 to −160), respectively. X3s was mutated from "GGATCATCC" to "GGGTCGACC." X3t was mutated from "CCAAATCCA" to "CCGAATTCA." These mutated sites do not bind Bcd, as revealed by gel retardation assays (data not shown). pDY88 and 89 were generated by cloning the *Xho*I-*Xba*I fragments (with the *Xba*I site filled-in with Klenow) of the bacterial vectors pDY86 and pDY87, respectively, into the *Xho*I/*Sma*I sites of the yeast integrating plasmid, LR1Δ1Δ2μ (*32*). pMAX108 (*33*) was used as the template in a PCR-based site-directed mutagenesis procedure to generate the mutations at X3s and X3t. To generate pDY86 and pDY87, the *Xho*I-*Sac*II fragments obtained on PCR mediated site-directed mutagenesis were cloned into the *Xho*I/*Sac*II sites of the pBluescript KS(−) vector. Mutations at sites X3s and X3t were generated using primer R, primer F and the corresponding primer containing mutations at site X3s or X3t. Effector plasmids expressing Bcd or Bcd-VP16 and reporter plasmids pMAX104 and pMAX116, which contain the reporter genes, hb-Δ5 and mut-X3, respectively, have already been described (*33*).

Gel Retardation and DNase I Footprint Assays—To generate radioactively labeled DNA probes for gel retardation and DNase I footprint assays, *Kpn*I-*Xba*I fragments containing Bcd binding sites and part of the polylinker were first isolated from the respective plasmids, and then filled-in with reverse transcriptase or Klenow in the presence of [α-^{32}P]dCTP. The experimental procedures and conditions for gel retardation and DNase I footprint assays were described previously (*26, 30*). The Molecular Dynamics PhosphorImager ImageQuant program was used for quantitative analysis.

Yeast Strains, Yeast Transformation, and β-Galactosidase Activity Assays—The reporter plasmids were integrated at the *URA3* locus of the yeast strain, GGY1 (*34*), as described previously (*33*). Only strains containing single copy reporter genes were used for subsequent studies. To assay the activities of Bcd and Bcd-VP16 from different reporter genes, the corresponding effector plasmids were transformed into appropriate yeast strains. Five colonies from each transformation were cultured and assayed for β-galactosidase activity according to the procedure described previously (*26*).

RESULTS

Bcd Preferentially Binds to Symmetric Tail-to-Tail Double Sites—To identify the optimal arrangements of Bcd sites for cooperative Bcd binding, we performed an *in vitro* selection experiment using a pool of 48 base pair random oligonucleotides. The DNA sequences bound by Bcd were selected by immunoprecipitation with antibodies against Bcd and then amplified by the Polymerase Chain Reaction (PCR). A similar selection approach involving immunoprecipitation was described previously (*35, 36*). Because our previous studies demonstrated that the interaction between Bcd molecules is weak (*26*), we conducted fourteen rounds of selection and amplification to maximize our chance of isolating DNA sequences that are preferentially recognized by multiple Bcd molecules. The PCR products were then cloned and 10 randomly chosen individual clones were sequenced (see "MATERIALS AND METHODS" for details). We used the consensus sequence, TAATCC, which was generated in a previous *in vitro* selection experiment using the homeodomain of Bcd (*12*), to analyze our selected sequences. All the sites indicated in Fig. 1A have at least 5 positions matching this consensus sequence, and the majority of them (76%) exhibit a perfect match, suggesting that the consensus site for the intact protein is also TAATCC. As shown in Fig. 1A, all the selected DNA sequences contain tail-to-tail double sites that are separated by 7-15 base pairs (end-to-end). Most of the selected sites (80%) are separated by 7-10 base pairs, only two sequences having 13 and 15 base pair spacing (#2 and #9, respectively). A symmetric double site consensus sequence deduced from our sequence data is shown in Fig. 1B. Only one of the selected sequences (#5) shown in Fig. 1A (also see legend) contains a third site in addition to the tail-to-tail double sites; this third site and its neighboring site are arranged in a head-to-head fashion.

Since the sequences shown in Fig. 1A were selected from a pool of random sequences, they represent sites that are preferentially recognized by Bcd, suggesting that Bcd can bind to these selected tail-to-tail sites cooperatively. To further demonstrate cooperative Bcd binding to these sites, we performed gel retardation assays using two different probes (Fig. 1C). While one probe contains a selected sequence with two sites, the other probe has a 4 base pair mutation to eliminate one of the Bcd-binding sites but is otherwise identical to the first probe. Figure 1C shows that, at low concentrations of Bcd, the majority of the complexes formed on the double site probe was Complex 1, which contains one Bcd molecule (lane 2; also see legend to Fig. 1). An increase in the Bcd concentration led to the formation of more Complex 2 containing two Bcd molecules (lanes 3-5) and, eventually, almost all the probe was shifted to Complex 2 (lane 5). As expected, only Complex 1 was formed on the single site probe (lanes 7-10).

The following observations suggest that Bcd binds to the selected double sites cooperatively. First, at the same concentration of Bcd, the total amount of complex formed on the double site probe (lanes 2-5) was always more than twice the amount of the complex formed on the single site probe (lanes 7-10). However, as noted above, the majority of the complexes formed on the double site probe was Complex 1, rather than Complex 2 (except lane 5). Such unusual behavior has been reported previously for other proteins that, like Bcd molecules (*26*), interact with each other weakly. For example, one protein can help another protein to bind DNA but is absent in the final complex, presumably because it is dissociated from the complex during electrophoresis (*37*). Second, at the highest concentration of Bcd used in our assay, virtually all the double site probe was shifted to Complex 2 (lane 5), while a significant amount of the single site probe remained unbound (lane 10). Had there been no cooperativity on the double site probe, all the single site probe would have been similarly shifted at the same protein concentration. Third, we conducted quantitative analysis to estimate the Hill coefficient by measuring the amounts of both bound and unbound probes at different Bcd concentrations (see Fig. 1D legend for further details). Figure 1D shows that the Hill coefficient (the slope) for the double site probe was 1.56, while the value for the single site probe was 0.70, further indicating that Bcd binds the selected double sites cooperatively. Although the results of DNase I footprint analysis of the selected double site probe were less compelling than desired because there were very few respectable bands at the Bcd sites in the absence of the protein, the limited information was nevertheless consistent with cooperative Bcd recognition (not shown).

Head-to-Head Sites Are Also Isolated in a Different Selection Experiment—Analysis of adjacent Bcd sites in the natural enhancer elements of several Bcd target genes revealed three different arrangements: head-to-tail tandem repeats, and symmetric head-to-head and tail-to-tail alignments (data not shown; also see Refs. *16* and *21-23*). However, all the sequences isolated in the *in vitro* selection experiment shown in Fig. 1A contain tail-to-tail sites with only one other arrangement (fragment #5). In a different, "forced" selection experiment, a pre-existing site that can be recognized by Bcd was placed next to a 13 base pair random sequence (Fig. 2A; also see its legend). We note that this 13 base pair random sequence was too short for the tail-to-tail sites, which require at least 19 base pairs (Fig. 1B), thus making it possible to isolate other patterns of adjacent sites, *i.e.*, head-to-head or tandem repeat sites.

After fourteen rounds of selection and amplification using the random pool of oligonucleotides shown in Fig. 2A, the PCR products were cloned and 10 randomly chosen individual clones were sequenced. Figure 2B shows that symmetric head-to-head double sites were isolated in this forced selection experiment (also see legend). No sequences with tandem repeat sites were isolated. It is interesting to note that, unlike the selected tail-to-tail double sites that are separated by different spacing (Fig. 1, A and B), all the selected head-to-head sites are separated by 3 base pairs. Figure 2C shows the deduced head-to-head double site consensus sequence with one of the sites from the pre-existing sequence. The fact that one of the two sites was selected from a pool of random sequences indicates preferential recognition of Bcd to the head-to-head double sites over single sites. In addition, the quantitative analysis shown in Fig. 1D revealed a higher Hill coefficient for the selected head-to-head sites (1.13) than for a single site (0.70), further supporting the notion that Bcd binds cooperatively to the selected head-to-head sites. Moreover, a DNase I footprint experiment showed that Bcd binds to a head-to-head double site probe better than to a single site

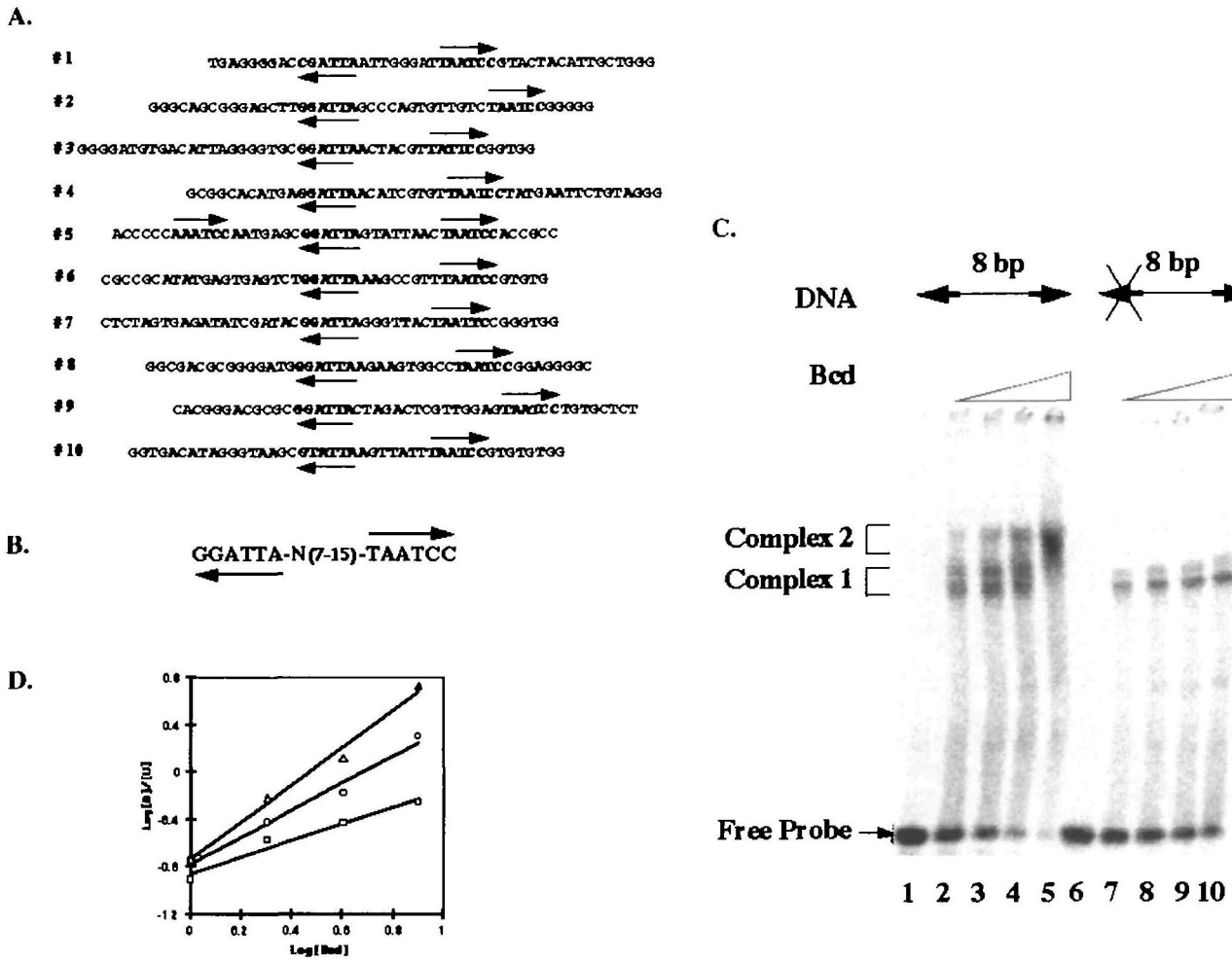

Fig. 1. **Bcd preferentially binds to symmetric tail-to-tail double sites.** (A) Compilation of sequences selected from a pool of 48 base pair random oligonucleotides. Only the random portion of the oligonucleotide sequence is shown. The flanking sequences upstream and downstream of the random sequences are: 5′ATGCCCTTGACAG-TCTCGAG3′ and 5′AAGTCGTGCGTCGTCTAGAG3′, respectively. Another sequence in fragment #7, GGGTTA, which is located between the two marked sites, is not indicated in this figure because Bcd cannot tolerate a G residue at position 3 according to our unpublished results. Similarly, another sequence in fragment #3 (ACATTA) is not indicated because it lacks the critical G residue at position 2. The sequences flanking the Bcd sites appear to be GC-rich, suggesting that they may contribute, either directly or indirectly, to the stability of Bcd/DNA complexes during our selection procedure. The flanking sequences shown in Fig. 2 are even more strikingly GC-rich. We note that no tandem repeat sites separated by long spacing were isolated in the experiment shown in this figure, suggesting that long-spaced tandem repeat sites may be less optimal than the selected tail-to-tail sites for Bcd recognition. (B) The deduced consensus sequence of the symmetric tail-to-tail pattern of double sites. (C) Cooperative binding of Bcd to the selected tail-to-tail double sites in gel retardation assays. The probe shown in the left panel is #6 listed in A and contains one pair of selected Bcd double sites separated by 8 base pairs. The probe shown on the right panel contains a 4 base pair mutation to eliminate one Bcd site but is otherwise identical to the other probe. The estimated amounts of active Bcd were: 0, 0.12, 0.23, 0.46, and 0.92 μM for lanes 1 through 5, and 6 through 10, respectively. The presence of two bands for Complex 1 is due to two different forms of Bcd, one of which is a partially degraded product presumably missing one of the extreme termini of the protein. Although we currently do not have experimental evidence indicating that these two forms of Bcd behave identically in cooperative DNA binding, any potential defect of the truncated form should not invalidate our experiments because it would only lead to underestimation, rather than overestimation, of Bcd cooperativity. It is noted that only one broad band representing Complex 2 is observable, presumably because the three complexes arising from the two forms of Bcd are not resolved in this gel. (D) A plot to estimate the Hill coefficient for a single site (squares) and tail-to-tail sites (triangles), as well as head-to-head sites (circles; see below). The amounts of free probe, and Complexes 1 and 2 at different Bcd concentrations for the experiments shown in C (and unpublished results) were determined using a Molecular Dynamics PhosphorImager system. The bound (B) and unbound (U) fractions of the Bcd sites were calculated according to the following formulase: [B] = [Complex 2] + 1/2[Complex 1], and [U] = [Free probe] + 1/2[Complex 1]. In this figure, the slope of each line, calculated by linear regression using Microsoft Excel, represents the Hill coefficient. The concentrations of Bcd shown in this figure are expressed in arbitrary units of 1, 2, 4, and 8, representing estimated active Bcd concentrations of 0.12, 0.23, 0.46, and 0.92 μM, respectively; since the experiments were carried out under low probe conditions, the free Bcd concentration should be approximately the same as the total Bcd concentration. We note that the Hill coefficient calculated for the single site probe was less than the theoretical number of 1.0, suggesting systematic underestimation of the Hill coefficient presumably due to the gel electrophoresis procedure; nevertheless, the results shown in this figure clearly demonstrate cooperative Bcd binding to the double site probes.

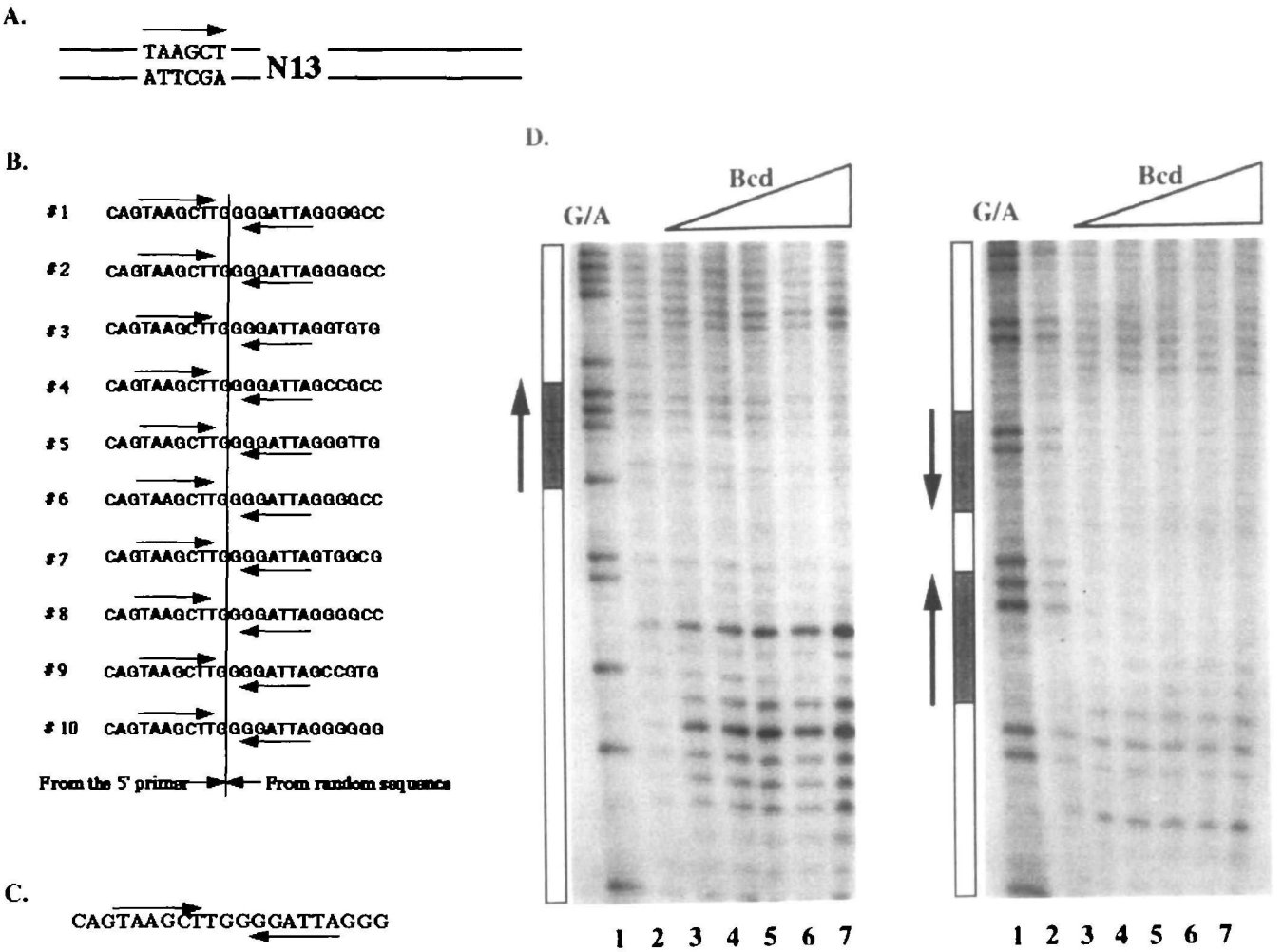

Fig. 2. **Head-to-head sites are isolated in a different selection experiment.** (A) A schematic diagram showing the sequence used in a "forced" selection experiment. In this sequence, a pre-existing Bcd site (TAAGCT) is located two base pairs away from the 13 base pair random sequence. (B) Compilation of sequences selected from the pool of random oligonucleotides shown in A. In addition to the random sequence, only part of the upstream sequence containing the pre-existing Bcd site is shown. Flanking sequences not shown in this figure are: 5′CGCGACGCT3′ (upstream) and 5′GGCTGAGTCTGAACGGATCC3′ (downstream). We note that four of the selected sequences (#1, #2, #6, and #8) are identical. Although we do not know whether these sequences represent independent clones from the original pool of random oligonucleotides or clones resulting from PCR amplification, they appear to be preferentially enriched in our selection experiment. We also note that the head-to-head double sites could only be isolated from the pool of oligonucleotides containing a pre-existing site, suggesting that they may be less optimal than the tail-to-tail sites for Bcd recognition. This idea is consistent with the observation that the Hill coefficient for the selected tail-to-tail sites was higher than that for the selected head-to-head sites (1.56 *vs.* 1.13) on quantitative analysis (Fig. 1D and unpublished results). (C) The head-to-head double site consensus sequence deduced from the selection results shown in B. The pre-existing Bcd site in the forced selection experiment matches the consensus sequence poorly but is identical to sites representing X1 and X2 in the *hb* enhancer element. In a similar forced selection experiment using the consensus Bcd site, TAATCC, as the pre-existing site, we failed to select any sequence resembling a Bcd site (data not shown). We currently cannot explain this result, but it may be related to how Bcd may interact differently with the consensus site and X sites that match the consensus sequence poorly. (D) DNase I footprint analysis of a single site probe (left panel) and a head-to-head double site probe (right panel). These two experiments were performed side-by-side. For the experiment shown in the right panel, protection of the marked sites (particularly the lower site) was evident in lane 3, in addition to several hypersensitive sites at or immediately near the marked sites. In contrast, for the experiment shown in the left panel, no protection of the marked site was observed at the same or even higher concentrations of Bcd. We do not understand why several bands in the lower part of the left panel increased in intensity, but this was presumably a non-specific effect because these bands are outside the Bcd site. Lane 1 shows a G+A DNA sequencing ladder. The estimated amounts of active Bcd were: 0, 0.06, 0.12, 0.23, 0.46, and 0.92 μM for lanes 2 through 7, respectively. See "MATERIALS AND METHODS" for further details.

probe (Fig. 2D).

The Effect of Long Spacing on Cooperative Bcd Recognition—It was shown previously that Bcd can bind cooperatively to the long-spaced tandem repeat sites, A1 and X1, from the *hb* enhancer element (*26*). Since adjacent tandem repeat Bcd sites separated by different spacing are found in natural enhancer elements (*16, 21-23*), we were interested in determining whether or not Bcd could also recognize short-spaced tandem repeat sites cooperatively. Gel retardation experiments were performed to directly compare the binding of Bcd to sites A1 and X1, which are separated by natural 36 base pairs, and to two sites (X1 and A1) that

are separated by 3 base pairs. The two sites in both of these probes are aligned in a tandem repeat fashion. Figure 3A shows that Bcd binds to the long-spaced sites more efficiently than to the short-spaced ones. In fact, no Complex 2 was detected on the short-spaced sites even at the highest concentration of Bcd in our assay (lane 5), suggesting that Bcd binding to these two short-spaced tandem repeat sites is mutually exclusive. These experiments suggest that short-spaced tandem repeat sites are unfavorable sites for cooperative Bcd binding and an increase in spacing enables Bcd to recognize them cooperatively. We currently do not know the minimal spacing required for cooperative Bcd binding to tandem repeat sites. However, our selection experiments shown in Fig. 2 suggest, but do not prove, that more than 7 base pairs may be required: had 7 base pairs been sufficient, a Bcd site of 6 base pairs arranged in tandem with the existing site might have been isolated.

The following two sets of experiments further demonstrate that no strict arrangements of long-spaced sites are required for cooperative recognition by Bcd. First, we modified the probe containing the natural A1-X1 sites from the *hb* enhancer element by site-directed mutagenesis. We inverted A1 so that the two sites became aligned in a symmetric tail-to-tail fashion, rather than the natural tandem repeats. DNase I footprint assays involving these two otherwise identical probes suggested a similar affinity to Bcd (Fig. 3B). As demonstrated previously (*26*), the affinity of Bcd to the natural A1-X1 tandem repeat sites is almost one magnitude higher than that to the individual sites. Second, we introduced 4 additional base pairs between A1 and X1 to change the relative positions of these sites on the surface of the DNA double helix. DNase I footprint experiments suggested that Bcd binds similarly to these two probes (data not shown).

Critical Contribution to Gene Activation by a Pair of Symmetric Sites—Our recent experiments designed to determine the intrinsic contribution of individual Bcd sites to transcriptional activation in the context of a *hb* enhancer element revealed the greatest contribution by sites X3 and X2 (*33*). In those experiments, a minimal enhancer element that was sufficient for mediating transcriptional activation by Bcd in yeast was systematically analyzed by mutating each of the Bcd sites individually (A1, X1, X2, X3, and A2). It was shown that mutations at X3 or X2 resulted in a more severe loss in transcriptional activation by Bcd than mutations at any other sites (*33*). The importance of X3 and X2 was even more evident when a Bcd fusion protein (Bcd-VP16) containing the strong activation domain, VP16 (*38*), was used to trigger transcription from these *hb* enhancer derivatives. In this case, while mutations at X3 and X2 decreased the levels of reporter gene expression by over 10-fold and 5-fold, respectively, mutations at other sites had virtually no effect (less than 20%).

Sequence analysis, using the consensus Bcd site, TAATCC, revealed that the region near X3 contains two previously unrecognized sites. Figure 4A illustrates the Bcd sites in

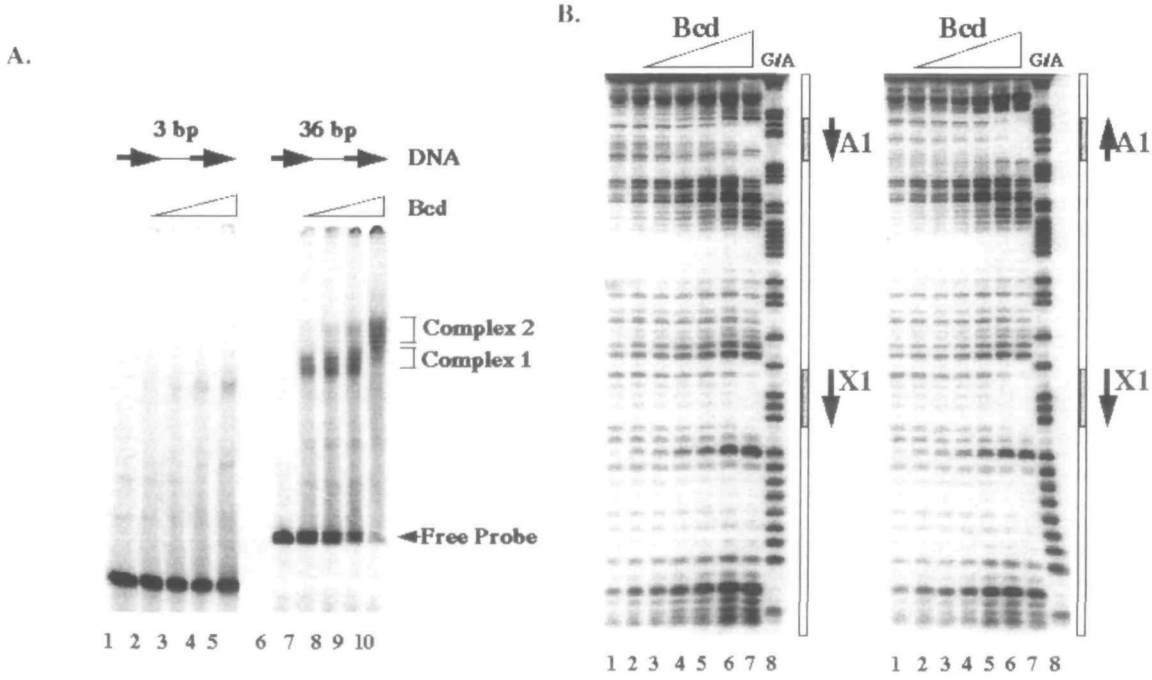

Fig. 3. **The effect of long spacing between adjacent DNA sites on Bcd binding.** (A) Gel retardation assays involving probes containing either short-spaced (left panel) or long-spaced (right panel) head-to-tail double sites. The long-spaced sites, A1 and X1, are from the *hb* enhancer element with the natural spacing of 36 base pairs (*26*). The short-spaced tandem repeat sites, which are identical to the sequences representing X1 (TAAGCT) and A1 (TAATCC), respectively, are separated by 3 base pairs. The estimated amounts of active Bcd added to the reactions were: 0, 0.12, 0.23, 0.46, and 0.92 μM for lanes 1 through 5, and 6 through 10, respectively. See legend to Fig. 1 for additional details. (B) DNase I footprint assays on long-spaced double sites that are aligned differently. The DNA probe shown in the left panel contains the naturally aligned head-to-tail sites, A1 and X1, from the *hb* enhancer element. The DNA probe shown in the right panel is identical to that shown in the left panel, except that the orientation of A1 was inverted by site-directed mutagenesis. The estimated amounts of active Bcd for both experiments were: 0, 0.03, 0.06, 0.12, 0.23, 0.46, and 0.92 μM for lanes 1 through 7, respectively. Lane 8 represents a G+A DNA sequencing ladder. The two experiments shown in this figure were performed side-by-side.

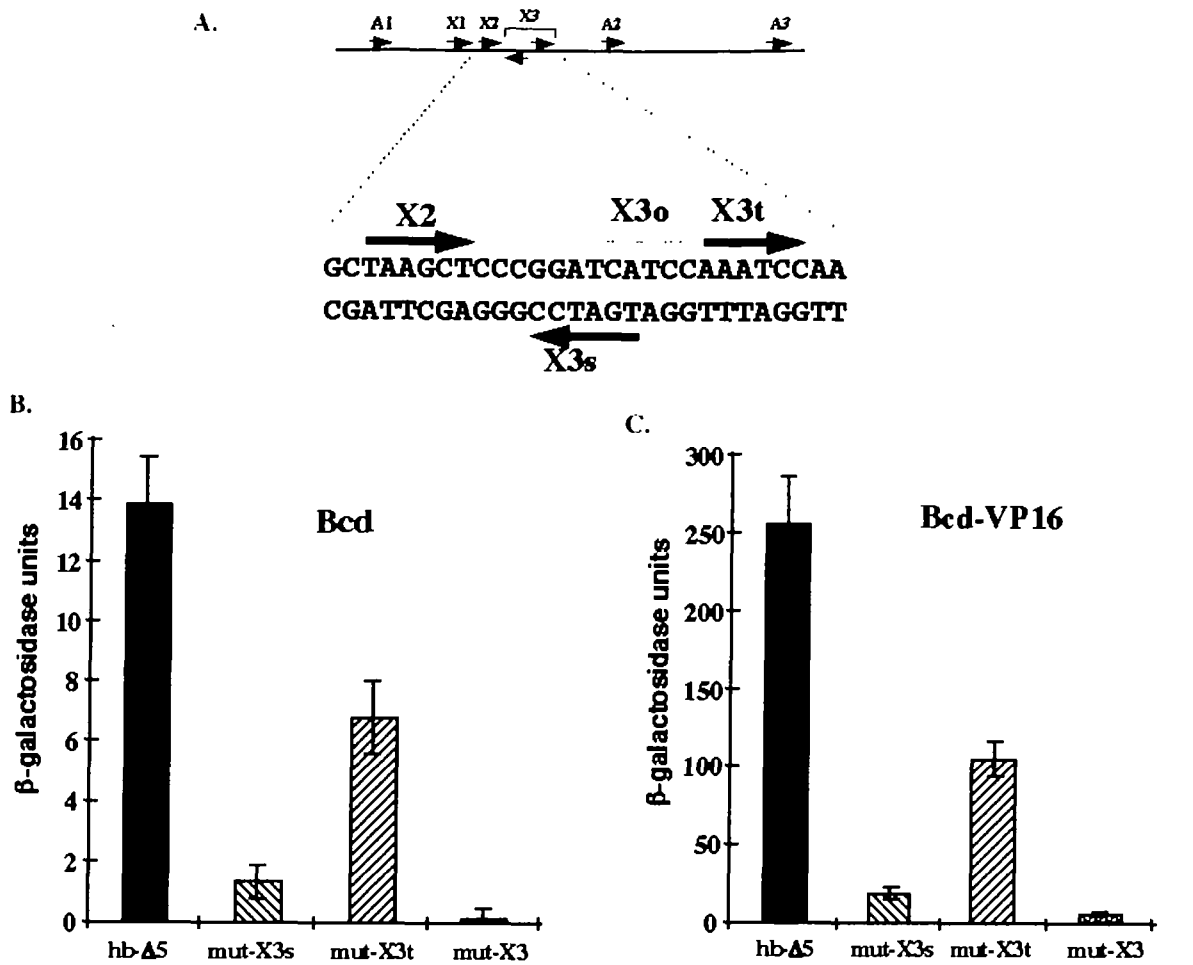

Fig. 4. **Critical contribution to gene activation by a pair of symmetric sites in a *hb* enhancer element.** (A) Arrangements of Bcd sites in the *hb* enhancer element (top), and detailed sequence information in the X2-X3 region (bottom). The two previously unrecognized sites, X3s and X3t, each have 5 positions matching the consensus site, TAATCC. The 6 base pair site, X3o, representing the originally described site (*16*), is not shown in the top diagram, but is indicated by a broken line in the bottom diagram to reflect our finding that it failed to be recognized by Bcd as an isolated individual site (data not shown; also see text). It should be noted that we currently do not know whether or not X3o can be recognized by Bcd in the context of the entire *hb* enhancer element. The *hb* enhancer element used for the mutagenesis studies in Ma *et al.* (*33*) and in this study (hb-⊿5; see below) contains the following sites: A1-X1-X2-X3-A2, where X3 collectively represents X3s and X3t (see text for further details). (B and C) Transcriptional activation by Bcd (B) and Bcd-VP16 (C) from the following *GAL1-lacZ* reporter genes in yeast cells: hb-⊿5, mut-X3s, mut-X3t, and mut-X3. hb-⊿5 contains the minimal *hb* enhancer element sufficient to support gene activation by Bcd in yeast (*33*). mut-X3s and mut-X3t carry mutations at X3s and X3t, respectively, and mut-X3 contains mutations destroying both X3s and X3t (*33*).

the *hb* enhancer element, depicting in detail the X2-X3 region. For simplicity, we term the two newly recognized sites, X3s and X3t (reflecting their symmetric and tandem alignments with X2, respectively), and the originally recognized site, X3o. To determine whether or not the three 6 base pair sites representing X3s, X3t, and X3o can be recognized by Bcd individually, we performed gel retardation assays. We found that Bcd can recognize both X3s and X3t, but not X3o, as isolated individual sites (data not shown; also see Fig. 4A legend). It should be noted that we were aware of the existence, although not the functional significance, of both X3s and X3t when we generated mutations in the study by Ma *et al.* (*33*). The X3 mutation described in that report and in Fig. 4 here was designed to destroy both X3s and X3t. Therefore, throughout both reports, X3 refers to X3s and X3t collectively.

In the *hb* enhancer element, there are only two pairs of symmetric adjacent sites, X2-X3s and X3s-X3t, all the other pairs of adjacent sites being aligned in tandem repeats (Fig. 4A). While the 3 base pair spacing for the head-to-head pair of X2-X3s is optimal according to our *in vitro* selection studies, the spacing for the tail-to-tail pair of X3s-X3t is not. In fact, our gel retardation experiments indicated that the entire X3 region can only accommodate one Bcd molecule, suggesting mutually exclusive binding of Bcd to X3s and X3t *in vitro* (data not shown). Taken together, these results suggest that the only pair of optimally spaced adjacent symmetric sites in the *hb* enhancer element is X2 and X3s.

We hypothesized that sites X2 and X3 contributed more to gene activation than the other sites due to the pair of optimally spaced head-to-head sites, X2 and X3s. To directly test this hypothesis, we mutated X3s and X3t individually, anticipating that the X3s mutation would

result in a more severe loss in transcriptional activation than the X3t mutation. *GAL1-lacZ* reporter genes bearing the mutated enhancer elements were integrated into the yeast genome as single copies to assay transcriptional activation by Bcd. As shown in Fig. 4B, while the X3t mutation decreased the reporter gene activity by less than 2-fold, the X3s mutation reduced the reporter gene activity by 10-fold. We also assayed these reporter genes for transcriptional activation by the Bcd-VP16 fusion protein, which contains nearly full-length Bcd (residues 1–479 of the 489 amino acid protein) fused to the activation domain of VP16 (*38*). Figure 4C shows that, although higher expression levels were observed with Bcd-VP16 for all the reporter genes as expected (note the scale difference), Bcd-VP16 responded to the Bcd site mutations similarly to Bcd. Taken together, these experiments demonstrate that the sites, X2 and X3s, which are preferentially aligned according to our *in vitro* selection findings, are also most active in supporting transcriptional activation by Bcd in an *in vivo* test.

DISCUSSION

Cooperative DNA binding by the morphogenetic protein, Bcd, has been proposed to be an important mechanism facilitating the establishment of restricted patterns of target gene expression in the embryo (*26*). In this study, we used an *in vitro* selection approach to identify preferred arrangements of adjacent sites for cooperative recognition by Bcd. Sequences with Bcd sites arranged in both tail-to-tail and head-to-head patterns have been isolated (Figs. 1 and 2), demonstrating preferential recognition of these sites by Bcd. Our studies also revealed a remarkable difference in Bcd binding to short-spaced sites and long-spaced sites. Specific alignments between short-spaced sites (symmetric tail-to-tail or head-to-head patterns) are required for cooperative Bcd binding. In contrast, Bcd can bind cooperatively to long-spaced sites with a high degree of flexibility (Fig. 3B), permitting Bcd to bind cooperatively to tandem repeat sites (Fig. 3A). Finally, we show that a pair of preferentially aligned Bcd sites in a *hb* enhancer element contributes the most to transcriptional activation *in vivo* (Fig. 4), demonstrating the biological importance of the binding site patterns revealed by our *in vitro* selection studies.

How does Bcd recognize different patterns of adjacent sites cooperatively? One possibility is that Bcd uses a single interaction interface to facilitate cooperative recognition to multiple patterns of adjacent sites due to a high degree of protein and/or DNA flexibility. Alternatively, Bcd may use different surfaces to facilitate the cooperative binding to differently aligned sites. Several lines of evidence are consistent with this latter idea. First, Bcd recognizes two symmetric patterns of double sites differently (Fig. 1D), suggesting different Bcd-DNA and Bcd-Bcd interaction profiles on these differently aligned sites (see Fig. 2B legend for further discussion). Second, multiple regions of Bcd have been identified that are involved in protein-protein interactions between Bcd molecules (*30*; C. Zhao and J.M., unpublished results). Third, complexes containing multiple Bcd molecules appear to be more stable than dimeric complexes on DNA, suggesting that Bcd can form multimers with multiple interaction surfaces (*26*). It remains to be determined how specific surfaces of Bcd can facilitate the cooperative recognition to various types of double sites.

Our current and previous studies revealed multiple modes of the interaction between Bcd and its DNA sites. First, without requiring stable dimers or oligomers as the obligatory form for DNA binding, Bcd can bind DNA as a monomer (Fig. 1; also see Ref. *26*). Second, Bcd can bind cooperatively to two sites that are aligned properly: short-spaced symmetric double sites and long-spaced tandem or symmetric sites. We propose that cooperative binding to these properly aligned double sites by Bcd is facilitated by specific interfaces of Bcd. We note that the selected head-to-head sites are separated by shorter spacing (3 base pairs) than the selected tail-to-tail sites (7–15 base pairs). This finding may help explain why some inappropriately aligned sites, *e.g.*, tandem repeat or tail-to-tail sites that are separated by 3 base pairs, can only accommodate one Bcd molecule (Fig. 3A and data not shown). Third, Bcd can bind to an increasing number of sites in a concerted manner with an increasingly higher degree of cooperativity (*26*). We propose that multiple interaction surfaces of Bcd enable one molecule to interact with at least two other molecules simultaneously on DNA, resulting in a complex containing multiple Bcd molecules.

Our gene activation studies on yeast demonstrated that sites X2 and X3s, which represent the only pair of preferentially aligned adjacent sites in the *hb* enhancer element, play a most critical role in supporting transcriptional activation by Bcd (Fig. 4B and Ref. *33*). It is interesting to note that sites X2 and X3s actually match the consensus Bcd site poorly relative to A1 and A2. Although specific sequences of Bcd sites and their alignments are both likely to be important for mediating DNA binding and transcriptional activation *in vivo*, the alignment appears to be more important in some cases. For example, it has been shown that the six Bcd sites found in the *kni* enhancer element, which are preferentially aligned according to our *in vitro* selection results, although matching the consensus sequence poorly, can drive the expression of a reporter gene almost throughout the entire embryo (*22*).

We propose that the preferentially aligned pair of adjacent sites, X2 and X3s, in the *hb* enhancer element may play a "nucleation" role in initiating the event of overall cooperative recognition of all the sites by Bcd molecules. Although we do not know the exact network of all interactions between Bcd molecules bound at the individual sites in the *hb* enhancer element, such a proposed nucleation role of X2 and X3s may help explain how other sites contribute to transcriptional activation. In particular, among the other three sites that were systematically examined in our experiments, X1, which is the closest to the X2-X3s pair, contributed the most to transcriptional activation, whereas A1, which is the farthest from this pair, contributed the least (*33*). It is interesting to note that in our previous DNase I footprint experiments on the *hb* enhancer element (*26*), A3, which is the farthest from the X2-X3s pair (*16*), showed the weakest occupancy for Bcd among all the sites in the context of the *hb* enhancer element, although it matches the TAATCC consensus perfectly. Our recent studies (*33*) demonstrated that the strong activation domain, VP16, can compensate for the loss of certain Bcd sites. The failure of VP16 to do so for sites X2 and X3s further suggests a qualitatively different role for these two

sites in the mediation of Bcd binding and transcriptional activation (Fig. 4C). We note that, although X3s accounts for most of the X3 region's contribution to transcriptional activation by Bcd, X3t does make some minor contributions (Fig. 4). It is possible that X3t may be recognized by Bcd *in vivo* in the context of the entire enhancer element even though the recognition of X3s and X3t by Bcd *in vitro* appears to be mutually exclusive.

We wish to thank I. Cartwright, S. Potter, T. Bugge, B. Aronow, M. Sussman, and members of this laboratory for discussions and/or comments on the manuscript, and W. Driever for kindly providing the Bcd antibody.

REFERENCES

1. Triezenberg, S.J. (1995) Structure and function of transcriptional activation domains. *Curr. Opin. Genet. Dev.* **5**, 190-196
2. Ptashne, M. and Gann, A. (1997) Transcriptional activation by recruitment. *Nature* **386**, 569-577
3. Tjian, R. and Maniatis, T. (1994) Transcriptional activation: a complex puzzle with few easy pieces. *Cell* **77**, 5-8
4. Mangelsdorf, D.J., Thummel, C., Beato, M., Herrlich, P., Schutz, G., Umesono, K., Blumberg, B., Kastner, P., Mark, M., Chambon, P., and Evans, R.E. (1995) The nuclear receptor superfamily: the second decade. *Cell* **83**, 835-839
5. Mangelsdorf, D.J. and Evans, R.M. (1995) The RXR heterodimers and orphan receptors. *Cell* **83**, 841-850
6. Scott, M.P., Tamkun, J.W., and Hartzell, G.W.D. (1989) The structure and function of the homeodomain [Review]. *Biochim. Biophys. Acta* **989**, 25-48
7. Kaufman, T.C., Seeger, M.A., and Olsen, G. (1990) Molecular and genetic organization of the antennapedia gene complex of *Drosophila melanogaster*. *Adv. Genet.* **27**, 309-362
8. Kappen, C., Schughart, K., and Ruddle, F.H. (1993) Early evolutionary origin of major homeodomain sequence classes. *Genomics* **18**, 54-70
9. Kissinger, C.R., Liu, B.S., Martin-Blanco, E., Kornberg, T.B., and Pabo, C.O. (1990) Crystal structure of an engrailed homeodomain-DNA complex at 2.8 Å resolution: a framework for understanding homeodomain-DNA interactions. *Cell* **63**, 579-590
10. Qian, Y.Q., Billeter, M., Otting, G., Muller, M., Gehring, W.J., and Wuthrich, K. (1989) The structure of the *Antennapedia* homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors. *Cell* **59**, 573-580
11. Wolberger, C., Vershon, A.K., Liu, B., Johnson, A.D., and Pabo, C.O. (1991) Crystal structure of a MAT alpha 2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions. *Cell* **67**, 517-528
12. Wilson, D., Sheng, G., Lecuit, T., Dostatni, N., and Desplan, C. (1993) Cooperative dimerization of paired class homeodomains on DNA. *Genes Dev.* **7**, 2120-2134
13. Chan, S.K. and Mann, R.S. (1996) A structural model for a homeotic protein-extradenticle-DNA complex accounts for the choice of HOX protein in the heterodimer. *Proc. Natl. Acad. Sci. USA* **93**, 5223-5228
14. Chan, S.K., Jaffe, L., Capovilla, M., Botas, J., and Mann, R. (1994) The DNA binding specificity of Ultrabithorax is modulated by cooperative interactions with extradenticle, another homeoprotein. *Cell* **78**, 603-615
15. Goutte, C. and Johnson, A.D. (1994) Recognition of a DNA operator by a dimer composed of two different homeodomain proteins. *EMBO J.* **13**, 1434-1442
16. Driever, M. and Nusslein-Volhard, C. (1989) The bicoid protein is a positive regulator of *hunchback* transcription in the early *Drosophila* embryo. *Nature* **337**, 138-143
17. Struhl, G., Struhl, K., and Macdonald, P.M. (1989) The gradient morphogen *bicoid* is a concentration-dependent transcriptional activator. *Cell* **57**, 1259-1273
18. Nusslein-Volhard, C., Frohnhofer, H.G., and Lehmann, R. (1987) Determination of anteroposterior polarity in *Drosophila*. *Science* **238**, 1675-1681
19. Driever, W. (1992) The Bicoid morphogen: concentration dependent transcriptional activation of zygotic target genes during early *Drosophila* development, in *Transcriptional Regulation*, pp. 1221-1250, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
20. Tautz, D., Lehmann, R., Schnurch, H., Schuh, R., Seifert, E., Kienlin, A., Jones, K., and Jackle, H. (1987) Finger protein of novel structure encoded by *hunchback*, a second member of the gap class of *Drosophila* segmentation genes. *Nature* **327**, 383-389
21. Hoch, M., Schroder, C., Seifert, E., and Jackle, H. (1990) cis-acting control elements for Kruppel expression in the *Drosophila* embryo. *EMBO J.* **9**, 2587-2595
22. Rivera-Pomar, R., Lu, X., Perrimon, N., Taubert, H., and Jackle, H. (1995) Activation of posterior gap gene expression in the *Drosophila* blastoderm. *Nature* **376**, 253-256
23. Small, S., Kraut, R., Hoey, T., Warrior, R., and Levine, M. (1991) Transcriptional regulation of a pair-rule stripe in *Drosophila*. *Genes Dev.* **5**, 827-839
24. Lawrence, P.A. and Struhl, G. (1996) Morphogens, compartments, and pattern: lessons from *Drosophila*? *Cell* **85**, 951-961
25. Arnosti, D.N., Barolo, S., Levine, M., and Small, S. (1996) The *eve* stripe 2 enhancer employs multiple modes of transcriptional synergy. *Development* **122**, 205-214
26. Ma, X., Yuan, D., Diepold, K., Scarborough, T., and Ma, J. (1996) The *Drosophila* morphogenetic protein Bicoid binds DNA cooperatively. *Development* **122**, 1195-1206
27. Szostak, J.W. (1992) In vitro genetics. *TIBS* **17**, 89-93
28. Blackwell, K.T. and Weintraub, H. (1990) Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. *Science* **250**, 1104-1110
29. Wilson, D.S., Sheng, G., Jun, S., and Desplan, C. (1996) Conservation and diversification in homeodomain-DNA interactions: A comparative genetic analysis. *Proc. Natl. Acad. Sci. USA* **93**, 6886-6891
30. Yuan, D., Ma, X., and Ma, J. (1996) Sequences outside the homeodomain of bicoid are required for protein-protein interaction. *J. Biol. Chem.* **271**, 21660-21665
31. Driever, W. and Nusslein-Volhard, C. (1988) A gradient of bicoid protein in *Drosophila* embryos. *Cell* **54**, 83-93
32. West, R.W.J., Yocum, R.R., and Ptashne, M. (1984) *Saccharomyces cerevisiae GAL4-GAL10* divergent promoter region: location and function of the upstream activating sequences UAS_G. *Mol. Cell. Biol.* **4**, 2467-2478
33. Ma, X., Yuan, D., Scarborough, T., and Ma, J. (1999) Contributions to gene activation by multiple functions of Bicoid. *Biochem. J.* **338**, 447-455
34. Gill, G. and Ptashne, M. (1987) Mutants of GAL4 protein altered in an activation function. *Cell* **51**, 121-126
35. Wright, W.E., Binder, M., and Funk, W. (1991) Cyclic amplification and selection of targets (CASTing) for the myogenin consensus binding site. *Mol. Cell. Biol.* **11**, 4101-4110
36. Pollock, R. and Treisman, R. (1990) A sensitive method for the determination of protein-DNA binding specificities. *Nucleic Acids Res.* **18**, 6197-6204
37. Grueneberg, D.A., Natesan, S., Alexandre, C., and Gilman, M.Z. (1992) Human and *Drosophila* homeodomain proteins that enhance the DNA-binding activity of serum response factor. *Science* **257**, 1089-1095
38. Sadowski, I., Ma, J., Triezenberg, S., and Ptashne, M. (1988) GAL4-VP16 is an unusually potent transcriptional activator. *Nature* **335**, 563-564