

D.melanogaster Ama + 3kb upstream - Genomic Relaxed EvoPrint

Black capital letters represent bases conserved in all species and colored bases represent sequences present in all species except *D.simulans* or *D.pseudoobscura*. Yellow labeled letters represent coding sequence of Amalgam.

Hox binding sites described in Affolter *et al.*, 1990, Ebner *et al.*, 2005, Joshi *et al.*, 2007 and Joshi *et al.*, 2010.

xxx binding site for **Exd/Dfd** GATTAATGG (TG) (TA)

xxx binding site for **Exd/Lab**
 TGATGGATGG
 TGATGCATG
 TGATCAATTA

xxx binding site for **Scr** ATTAATG (similar to Dfd)

xxx binding site for **Antp**
 CATTAGAC
 CAATTACA

```

CGGAAGAGATGTAGAGATGgCACAAATGGCGGCCcTcCCTCGATGTtCaTTCTTTTGAGTAATTTaAAAAATCAAT 75
AAgCGTCTFAcaATAcATTtAGAcAAATatTTggacaTTTTATTTtAAATGTTtGCCTCTaATTTAAAAGTTtt 150
TAaTCGAACAGTACGGCAaaGCTGCAaaCcGatTgAGTgCAcAGcGTtTTCaCCAGgGGGCGtgaagAgAcAGCA 225
gTagaGAGCCCCCTGTGCGGcTGtATTgAaAGGTTGTGCTCaCTGCACaAAAGGTGAAtGTTCTAAaacGTCCGgA 300
GTTTATTTATTTtGtAGAgATTGcATTGAGAAaTAATAATAAAAaaAAAGCTTgAATATTgTAACAAaTAAGTAGA 375
aCCAATTAACcTTaaactCTTTGctCTGtTtGGAACTTCATAATAtTaAcAttTAAttAAcAAAAGGCcCCCA 450
AATtTAGGAAT AATT AAGTgTCGAA TgAATTTTTgTTAATATtATAAT TGTGCTcC 525
AGTCcagtccagCCTGAATTGAACAAAATGTGCACCTTTGTAGTTTATTcTGTTACATACTTcCGCTATTaataac 600
tacattattcaatccttaGTATCGCgagatatataaggacttGGGACAAAGAATAAGGATgTGGAGaTCgAGTTT 675
GTTTTCCAAatgATCTCaAACGgCTAcgTACGAAAGAAAATATCATATAATCGTTtGTAAgAcAGCaTATAGT 750
gcttgctatatgTACCTGttaGGCAGTGcATTATAAAATgTATtagAAgAATAaAGGTAAcCAGTGTGTCGAGTT 825
AaTTAgCTgTAGCGTTCgATTCTTTGttaaTTTTtttaagttggttactcatttcTGATgTTAAGCTAGGAT 900
TGGCaATTATTAttTcAGGCTCcATATTTGGAAAATTCGATATGTTTAGCGTtgAAGAATATTGTgtaaatc 975
tACAAgCATTAAGTGTGTCAGTGCATTGCCcAAGCTTTAAcATTTTAAAACACACCCAACGAGCCAGTTGACCT 1050
TTAAGCAAGTGTcCAAGGCCtTAAGTGAATATTTTTTAATGACATTTgCCaATAATgTcGTGTTTGTGTTTT 1125
GGGATcTGGCAGGAGCTTGTGCTGTTGCGGGcTTTGTCTGCTTCTGCGCTGCCAACTGGCAACAAAAGTCAA 1200
TTCAcAGAcACTCCAGTCGACGGAGTCGAGTGCAGCCAAAAGGCAAAACAATAACGtTCCATAGGCCAAAGCCC 1275
AAGGATACACACAAAGGTACAGTGGTgCTCTGACTACAAATACAAATTAACAGGgccagagctttaataccaa 1350
Ttcgataccattgcaacaacagatcttcaaatgctactttttatcatcagataaagataggtagaaaaggctata 1425
Tttacglttaaaggcaagctgatatttttagaaatttattcaaaaaaaaataattactttgacacataatattgt 1500
attagctcatctgactagaaactagaaatctctgaaatgtaacgaagttctgtgcatctctacAAATaTTAcAAAA 1575
AGTATCtgctagaTACCAGTGTGCGGAACtGGGCCACAAATTCaACCGAGTGTGTGACAAAGGACTTCcCCCTT 1650
CCCCGCTCGGCGACCCtGAATGCTGAATGCCCCCGTGTcTGCCGCTGCTGAAGCGAAACACCCAAAGAGTCACTG 1725
ACGAaAACAGGATTcGGGTAATAAATTACAAATTTGACATTTCTGTTTGGATCCGCTTGCcCTGAATCGCA 1800
CGGGCAGCCCTCTGTCGCACGCCCTcTAGCCCTcTTCTGTTCCGGTTTCCGGTGTGTGCTCCAGTTCCGTCGGG 1875
CCAAGTCGAGCGTTGCATGAAGCAGCCGACCTGcCAGGACCTGGACcAGGACCTGAACCTGCATCCAAgAGGCA 1950
TGCCCGCAATGGTCATAATGACCTGCCGGATGCAGTGAAGCaAAGGAGAGCCGtCaGTGGgcAGCgaGAAAG 2025
GGAGGAATGAAGCGCCCTATGcGGTCCGttGATAGTTAACTGAAATTTATTTTACCAATTTTACAAATTA 2100
TCTGATTCGAATGTGCATCAGcGACGAGGATTCGGCCAGATGGGcAGTACGAGTGAAGTCCGGGggTCCaAGt 2175
TCGGACGACAGATGCAGTGTcCTCCCATGGGAACGGCCcATTTCAGCAGAAGTCACCgAAACGATTTAAGATG 2250
TGATTTAATTTCCAGAAACTGTGGCCATAGTTTCTGTCCTCTTCCGGGGAATTTCTTCGCCAGACAGACAGCGT 2325
CGTCATCAATCATCTcGCAGAGTAAAGCCGGTTAACCTCGATGGGTTTGGGGGAaACTTGGCCCGGGGCTGA 2400
TCAGGAaTCaGGCTGGAAATGAAAATTAATGCGTTTGGGCaGCAGAAAAGTTTTCGAATGGTTCGGTGTCTAT 2475
ACAACAAGTGAAaTTTTgCAcATcAAaCGgcccTTTGACCACaTTTAATCtTTAAATCAcCTGCCATAGgAACTAA 2550
TTGATCGAGATCCAGACAcATtAAAcaCATcAATATAtATacacaAAATAATCAAtATAgACTATGTTATAT 2625
TTcgacgttttccctatcctcgtatgctgtGAGTACACATATTAACCtTCAGCCGTgTAAAGAAATTTCTAcGAACA 2700
GCCAAGTTAAAGGTacCCAATTTCCCGAATaCTGCTCTTTAAATATCCtTCAGAAATTTCCAGCATGGCTGAAC 2775
CAGGACTCGATTCAAGTcCACCCCAATGCCTTGTaCACCACCCgTTGTTTTGcCTTTTCGCGCATGCGCCCTC 2850
GCTCCGGCGCAAAAGAGACAGCACTAAAGAGACGGCGCTGGATCGCTGGCATTTCAGAGCGGctAGGATATTC 2925
CAGGCAAGCTCGTaaTCGTTGCTGTTGTTGTTAGCCGCCAGCGGAAAGTAGtGCGCCATGCCTAAATCAcACA 3000
AATGGGTAGAGTTCCACTTTCCACCCcGCCGctGCTCTTTTaTTGCTTCCAGAACCCCAACGTTTAT 3075
GGCCAAATGGCGGCAGaATTTGAATTTGAATTTGCGcACGCGGCTCTcAGTTCGCTGCGGCCGCTTTTCGCTGG 3150
CGCTTTGTTTTGCACTGACGCTTATCTTGGGGCGGGGCTGTCAAGTGGGcgGGGCAGaGGGCCaAGGCAc 3225
GAGGCAaTAACaACAGCAGCGAGcAGCAAAAACAATCAATTAAGACCAAGAAAGTGCATTGATAACTGAGC 3300
CCGAAGTCAAACTCAGCACAAGTAAACAATTTAAATACGAaCTaAAATGCAAAACAGATACAAAATACAGATA 3375
GAGATACTGAACAAACACAGATACAAATACTGTGAAGGCGACAAAGTACATAACCGAGGGGTGATAAGATCCAAA 3450
ACGTCTAGACTTCTTCACTCGGCAATACACTGAAAAAAGCTGtTACTaATTATTAaAAATATG-AAATAATA 3525
AAAGGTAaaCTATTTTTtttactaacttttaactatttactaactgcccATTAAATCTTTTTTCACATC 3600
TTAACTACAaAAAAATtAAtTTTAAATCAATTTGATtCGTTTTACCTGgCTATTCAtTaaaAAAAAAAAtAAGTA 3675
tTGATATATCgAAAAcTAAAGTTATCTccgTTTTTaccAGTGTATcCATTaGAAGCAACGTTGTCTAGACACAG 3750
CCGCGTAAATCAATCAATAGgTGTCTATCTGCTcTcATTCCTCGGGGTACAAAATTTCCGGACTTTCCTTGAGAA 3825
ACCGAGCCaTCGACGTGCGCGCTCAGGAAGAATACaAAAAtAAGAACcCGATTCTCAGaCCTCAGCAGAGATA 3900
    
```

CGGATTTTGTAGCCGAGCTCGGTCGCCTCTTCAATTCAGTCGAGTCCCGAACGCGAAGTGAAGCGCATTGAACC 3975
CACAttgaaactgaGAGATAGTCGAGGAAAAAGAGTCAAAAATCAACAGCAACATCCTGTAAATAAACAGAAAAG 4050
TATGAATTATATTTAAAAAAGACGATAAAgCTGaaAAGTGTGTaCaAGCtTTATAAATGTTTCAAAACGAGTGTG 4125
TGTGCAGAAATTCCTTTCcttcttatttacttttgaagataactttTAAAGATATACTTGTGAcCTTtAAATTATTA 4200
TTCTATAaTTGAACcTAGTTTTAGTTTTATATATGAATTCGAAGATGGGTAAAaatgGCAGTTCTCATTACA 4275
TACGAAATATTTAGAATGGCTAAcGCCATCAAAGACTCAACATAAAGTGATACACTaaTAAATCTCAAATCTAT 4350
TTTGATTCTTACAGACCAGACATGGCGCGTCTGCGGCTTtAATCGGTCTTATAtTCTGCCTTGCGATCAGCtT 4425
GGAtTCGGTGCCTtGTGCCCAAGTATCAGCCAATCAgCAAGGACGTGGTGGCCAGTGTGGCGACTCTGTGGA 4500
GTCAACTGCACcGTGGAGGAGTGGGAAGCCTCTCGGTGAGcTGGGCCAAGCGTCCCAGCGAGTCGGACACCAA 4575
CTCGGTGTGTCTTCCATGCGGAATATcCTCAGTCTGCCGACcAGCGGTACAACGTAAcGGTCACCGAAGGACC 4650
CAAGACCgGAAGTGCCATtTACACCTTCCGCATTcAGAACATCGAAGTCAgTGACATGGGTCCGTACGAGTtCA 4725
GGTGCTtGTTTCGGCCACGGAGAGGTgACCAAGAAGCTGAGcCTGCAGATCAAGACCCCGCCGTGATTGCGGA 4800
GAACACCcCCAAGAGcACCCTGGTgACgAGGGaCAGAATCTGGAGCTGACCTGCCATGCCAATGGCTTTCCCAA 4875
GCCtACGATCTCCTGGGCGCGGAcATAATGCaGTGATGCCtGCCGGtGGCCATTTGCTGGCgGAACCAACTCT 4950
CCGGATTCGtTCCGTGCACCGcATGGATCGCGGcGGCTACTACTGCATCGCGCAGAACGGAGAGGGCCAGCCGGA 5025
CAAGAGGTtGATCCGGGTGGAaGTGGAGTCCGTCCGCAGATTGCCGTcCAGAGcCCAAGATCGCCCAATGgT 5100
GTCCCAcTCCCGGAGCTGGAGTGTCTCCGTGCAGGGCTACCCGcCTCCGACTGTAGTTTGGCACAAgAAATGGCGT 5175
GCCCTGCAGTCTAGCCGGcCACcACGAGGTGGCCAACCCGCCCTCCTCGGGAACCACAACCTCTGTCTGCGG 5250
CATCGAaAGCGTgCGCAGGAGGACTtGGGACTACTACTGCAATGCCACCAATaAGCTGGGAcACGCCGATGC 5325
GCGACTGCATCTTCCAGACCCTCATcCCAGTGCCTCCTtTCGTAaTAAGCCCTCAACGTCTGATTCCATT 5400
CCCACGCAACAAAGACTGGAACCCtCACTTTcCTCTCGtCTCGCTTTTTTCGGCAATTCGCCAGTCCGGATGGG 5475
tCTGTGTTTATATTtTTAATGcATTTCATtCAATTAGCTCCATTAGACTTCATTATTTTCGGACACACTTGAGAT 5550
CGAGCCACCcCTTGTGCCTTATGAATTTTGTCTGCATCCGTCCGCTtGTGCAAAcACATTAATATGCTTAAG 5625
TAATAAAcGTTCTCTGATAAAcTTAATCGGTCTTTTCTATGTGGTTGGGTTGGCTTGATACGGtGTAAAGTTA 5700
aATACCAcAGAGaATAAAcACCTCTTgaAGCaACgaatTAAAGGGTTTGTcATTCAATAAAACTAAGCcAGGAG 5775
AcAAGcCATTGtCAAAAGT 5850

Affolter, M., Percival-Smith, A., Müller, M., Leupin, W. and Gehring, W. J. (1990).
DNA binding properties of the purified Antennapedia homeodomain. *Proc Natl Acad Sci U S A* **87**, 4093-4097.

Ebner, A., Cabernard, C., Affolter, M. and Merabet, S. (2005). Recognition of distinct target sites by a unique Labial/Extradenticle/Homothorax complex. *Development* **132**, 1591-1600.

Joshi, R., Passner, J. M., Rohs, R., Jain, R., Sosinsky, A., Crickmore, M. A., Jacob, V., Aggarwal, A. K., Honig, B. and Mann, R. S. (2007). Functional specificity of a Hox protein mediated by the recognition of minor groove structure. *Cell* **131**, 530-543.

Joshi, R., Sun, L. and Mann, R. (2010). Dissecting the functional specificities of two Hox proteins. *Genes Dev* **24**, 1533-1545.