Supplementary Material for

Capicua is involved in Dorsal-mediated repression of *zerknüllt* expression in *Drosophila* embryo

Dong-Hyeon Shin and Joung-Woo Hong

Graduate School of East-West Medical Science, Kyung Hee University, Yongin, 446-702, South Korea
Supplementary Discussion

This current study was motivated by the discovery of a nucleotide sequence homology between the AT-rich sites in the VRE and Cic-binding sites retrieved from the bacterial 1-hybrid (B1H) library. It is debatable whether or not the result of the B1H analysis is reliable enough to reflect the in vivo DNA-binding specificities of Cic. There exist a few features of the B1H system that are superior to other methods for predicting the DNA binding specificity of a transcription factor (Nat. Biotechnol., 2005, 23, 988-994). In contrast to other in vitro methods, such as footprinting analysis and electrophoretic mobility shift assay (EMSA), the B1H system does not require purification of the transcription factors being investigated, thereby leading to an analysis of the binding specificity of a particular transcription factor in a similar manner to in vivo conditions. Additionally, the B1H system ensures more efficient transformation, compared to the yeast one-hybrid (Y1H) system. These features facilitate precise prediction of DNA binding specificity of a transcription factor and ensure easiness to employ the system. Indeed, a recent study provides biochemical support for the B1H system as an accurate predictor of DNA-binding specificity of a transcription factor (Development, 2011, 138, 4291-4299). New knirps (kni) binding sites were identified by extensive EMSA with a series of oligonucleotides corresponding to the full length of an even skipped (eve) 3+7 enhancer. The consensus sequence of the new Kni-binding sites was remarkably similar to the Kni consensus sequence derived from the B1H system. Taken together, these results strengthen the notion that the DNA binding motif library produced by the B1H system can faithfully recapitulate in vivo binding specificities of various transcription factors.
Table S1. DNA oligonucleotide sequences

| Name          | Sequence (5’ to 3’ direction)\(^1\)                  | Method | Note\(^2\) |
|---------------|------------------------------------------------------|--------|-------------|
| *cic* probe-F | CATCCTACACCTCACCTGTT                                 | PCR    | Fig. 2A, B  |
| *cic* probe-R | CAGGTCAATCTCTTTTACGG                                  | PCR    | Fig. 2A, B  |
| *zen* probe-F | CACGATGTTAACCACATC                                   | PCR    | Fig. 2C, D  |
| *zen* probe-R | AAAATGGTTAAACTTCACGATAA                               | PCR    | Fig. 2C, D  |
| *cic* HMG-F   | CGGAGGAGTCAATCGTT                                    | PCR    | Fig. 3A     |
| *cic* HMG-R   | TGTCGAGGGATGACTTT                                    | PCR    | Fig. 3A     |
| EMSA AT1-F    | CCTATTTTTCTTGTATATAGTTTTTGGGAAATCCAGAAGTC            | EMSA   | Fig. 4A, D  |
| EMSA AT1-R    | GACTTCTGAAATCCCAAAACTTATCAAAGAAAAATAGG              | EMSA   | Fig. 4A, D  |
| EMSA AT2-F    | GCCTATATGAAGTAATTTGATGGGTTCTCCTCCAGTT               | EMSA   | Fig. 4B, D  |
| EMSA AT2-R    | AACTGGGAGGAAAACCAATCATATCTGTTCATATAGGC              | EMSA   | Fig. 4B, D  |
| EMSA AT3-F    | TTGATTTGCTTCTCAGCTATAGGTGGTTTTATGATCTGGGG           | EMSA   | Fig. 4C, D  |
| EMSA AT3-R    | CCCAAGATCAATAAAACACTCTATAACTGGGAGAACCACATCAA        | EMSA   | Fig. 4C, D  |

\(^1\) All sequences are presented in the 5’ to 3’ direction relative to the physiological orientation of *zen* transcription.

\(^2\) Indicates section(s) in the manuscript where DNA oligonucleotides were used.
Figure S1. DNA sequences of the *zen* VRE

>zen_VRE_645bp
CCTTCTAGAATGAAACGAAAACAGTATCTGGTTTTCCCGAAAATCTTTTCTGAAATTAAAAATGCCACTTTATTGCACAT
ACTCACACATGCCCTCCCATATAATAATATGGATTCGCGATTTTTTCCCGGAACACCCCGGAGATCATAAACATTTTGACCAG
CTGCCCTCTGTATATCTCACCCCTGGAACCCATACCTATATCGCTGATCTCCGCGGGTCGCACTATTAGGTAGACACTGTAAGGCAACACTCGGGCGGCGCCAGAGCAACTTTGTGCCCTATTTTCTCTGTGAGATTGGTTCCCAACGTTTATAGAGTTTTATTGATCTTGCGCGGTTTATTAGGATGATCGGCCGCTTTGATGATTATTCTGGGTCCCATTTATCAAGCATTACTTACTCTACGCAAGAC
TTGTTTCTGTTGCCCCATCTCACACCTAGTTAACTCTAAATTAAACTACTTTTTTATCCGCCATTATGAGATAATTTTTATAGCATCTATGGCATCTTTTAAATATCGAGATATCGAAGATATTCATAAATTAGTTTTTTATTGTAAGCCTTTTTATTTTGCTCAGG

Sequences are presented in the 5’ to 3’ direction relative to the physiological orientation of *zerknüllt* (*zen*). DNA binding sites for Dorsal (Dl), Capicua (Cic) and Zelda (Zld) are identified in blue, magenta and green, respectively.
Figure S2. A position frequency matrix (PFM) derived from the five AT-rich sequences in the *zen* VRE

|   | A  | C  | G  | T  |
|---|----|----|----|----|
| A | 0  | 0  | 0  | 5  |
| C | 0  | 3  | 0  | 0  |
| G | 0  | 2  | 0  | 3  |
| T | 5  | 5  | 2  | 0  |
Figure S3. Position Frequency Matrices (PFMs) from the B1H database
(http://pgfe.umassmed.edu/ffs/) used to search clusters of Dl, Zld, and Cic across the ~40 kb zen locus

74225883_d1_NBT_FBgn0000462 PFM:
A: 0.007 0.007 0.068 0.558 0.558 0.248 0.068 0.007 0.007 0.218
C: 0.007 0.007 0.007 0.007 0.007 0.007 0.248 0.978 0.917 0.618
G: 0.978 0.978 0.738 0.007 0.007 0.007 0.007 0.007 0.007 0.158
T: 0.007 0.007 0.188 0.427 0.427 0.738 0.677 0.007 0.068 0.007

74225883_cic_FBgn0028386 PFM:
A: 0.059 0.000 0.000 1.000 0.000 0.000 0.000 1.000
C: 0.588 0.588 1.000 0.000 0.000 0.000 0.588 0.000
G: 0.000 0.000 0.000 0.000 0.000 0.000 0.412 0.000
T: 0.353 0.412 0.000 0.000 1.000 1.000 0.000 0.000

74225883_zld_SOLEXA_5_FBgn0259789 PFM:
A: 0.070 0.010 0.978 0.007 0.007 0.007 0.978 0.102
C: 0.010 0.970 0.007 0.007 0.007 0.007 0.007 0.102
G: 0.620 0.010 0.007 0.948 0.968 0.007 0.007 0.733
T: 0.300 0.010 0.007 0.037 0.018 0.978 0.007 0.063
Figure S4. Expression and purification of the recombinant Cic

(A) Two major isoforms of cic protein expressed in Drosophila are depicted. The short and long forms of Cic (Cic-S, NCBI protein database accession number AAF55751; Cic-L, AFH06521) are composed of 1403 and 2141 amino acids, respectively. A DNA binding domain, called the high mobility group (HMG) box, is shared by both forms. C1 and C2 motifs have been known to be essential for transcription repression and MAP kinase recruitment, respectively. 115 amino acids (1194~1309, relative to the start codon of Cic-L as a +1) encompassing the HMG box region were fused to glutathione-S transferase. (B) GST-Cic fusion protein is expressed in E. coli. 1mM of IPTG (Isopropyl β-D-1-thiogalactopyranoside) was used to induce expression of GST and the GST-Cic fusion protein. E.coli on the top of lanes 1 and 2 means that the BL21 strain contains no expression plasmid. (-) No IPTG, (+) 1mM of IPTG added. Expected molecular weights of GST and GST-Cic are 27.6 kDa and 40.4 kDa, respectively. (C) GST and GST-Cic were purified with glutathione-Sepharose and the purified proteins were separated by 8 % SDS-PAGE (left panel). Fusion proteins smaller than ~40 kDa were thought to be degraded forms of the GST-Cic. The SDS-PAGE shows multiple GST-Cic bands, but Western blot analysis shows only three major bands. These results suggest that random degradation occurred at both termini of the Cic fusion protein. The purified proteins were analyzed by Western blot with anti-GST antibody (SC-138; Santa Cruz Biotechnology).
MATERIALS AND METHODS

Fly stocks

*Drosophila melanogaster* strain (cic\(^{1}\)/TM6B Antp\(^{Hu}\)) was used for producing the cic homozygous mutant embryos and for *in situ* hybridization. The cic\(^{1}\) mutation was initially isolated in a P-element screen for female-sterile mutations affecting anteroposterior polarity of embryos, and embryos homozygous for the cic\(^{1}\) allele are fully viable but 100 % sterile (1). The cic mutant strain was kindly provided by Jordi Casanova (Barcelona, Spain).

Bioinformatics

To examine if AT-rich sequences are similar to DNA binding motifs in the FlyFactorSurvey database (http://pgfe.umassmed.edu/ffs/), the TOMTOM motif comparison tool (http://meme.nbcr.net/meme/cgi-bin/tomtom.cgi), fed with a position frequency matrix (PFM) (Fig. S2) of the five AT-rich sequences in the VRE, was used to search the FlyFactorSurvey database. The ClusterDraw algorithm (http://line.bioinfolab.net/webgate/submit.cgi) was fed with a ~40 kb genomic sequence of the zen locus (3R, 2,559,251~2,599,250) and PFMs (Fig. S2 and S3) of Dl, Cic, Zld, and AT-rich sites. The genomic sequence and the PFMs were obtained from the Flybase GBrowse (http://flybase.org/cgi-bin/gbrowse2/dmel/BDGP genome assembly 5 and *D. melanogaster* annotation 5.56) and the FlyFactorSurvey database (http://pgfe.umassmed.edu/ffs/), respectively.

Whole-mount *in situ* hybridization

Whole-mount *in situ* hybridization was performed as described in a previous study (2). Briefly, embryos were collected 0-4 hours after egg deposition (AED), dechorionated, fixed, and
hybridized with digoxigenin (DIG) UTP-labeled antisense RNA probes. To examine the pattern of endogenous cic and zen expression in wild-type and homozygous cic mutant embryos, antisense cic and zen RNA probes (approximately 1.97 kb and 1.1 kb, respectively) were produced by polymerase chain reaction (PCR) (Table S1 in Supplementary Material). Exon 2 and exon 3 in the cic locus were amplified by PCR using the full length of cic cDNA as a template (BDGP Gold cDNA, LD05430, 5175bp). An approximately 1.1 kb region of the 3’ largest exon of zen was amplified by genomic PCR. The amplified DNA fragments were introduced into the pGEM®-T Easy vector and used as a template for in vitro transcription. We used Campos-Ortega and Hartenstein’s definition for developmental stages during embryogenesis (3).

Overexpression and purification of GST-Cic fusion protein

A protein expression construct for GST-Cic was generated by cloning the PCR fragment of the cic HMG-box (amino acids 1194 ~ 1309 of Cic-L, GeneBank Accession number AFH06521) into the pGEX 6P-1 vector (GE Healthcare). The amplified cic HGM-box region was cloned between the BamHI and XbaI sites. Cic recombinant protein was expressed as a GST fusion in Escherichia coli BL21(DE3)/pLysS using the GST expression vector described above. Bacteria were grown to an optical density of 0.5~0.8 at 600 nm and 37°C and then induced with 1 mM of IPTG for three hours. GST-Cic was purified with glutathione-Sepharose 4B resin (GE Healthcare) under native conditions. The expressed and purified proteins were analyzed in 8 % SDS-PAGE and Western Blotting with anti-GST antibody (SC-138; Santa Cruz Biotechnology).

Electrophoretic mobility shift assay (EMSA)
The GST-Cic used for the EMSA was purified from bacterial extract. EMSAs were performed with a biotin-labeled DNA fragment containing an AT1, AT2, or AT3 site (See Table S1 for DNA nucleotide sequences of the probes). The biotin-labeled DNA probes were prepared by hybridization of single strand oligonucleotides conjugated with biotin molecules. Binding reactions and non-isotopic detections were done with the LightShift Chemiluminescent EMSA Kit (Pierce Biotechnology). Briefly, binding reactions were performed in 20 μl of a buffer containing 25 nM of a biotin-labeled DNA probe and 1 μg of GST or 0.2~0.4μg of GST-Cic protein. For competition experiments, 100-fold molar excess of unlabeled probes were added to the binding reaction. The binding mixtures were separated in 6 % non-denaturing polyacrylamide gels, electronically transferred to a nylon membrane (15 V, 400 mA), and then UV-cross linked (120mJ/cm²). Shifted and non-shifted probes on the nylon membrane were visualized by adding stabilized streptavidin conjugated with Horseradish Peroxidase (HRP) and HRP substrates, and then they were scanned with the EZ Capture II Chemiluminescent imaging system (ATTO Technology).
REFERENCES

1. Jimenez, G., Guichet, A., Ephrussi, A. and Casanova, J. (2000) Relief of gene repression by torso RTK signaling: role of capicua in Drosophila terminal and dorsoventral patterning. *Genes Dev.* **14**, 224-231.

2. Jiang, J. and Levine M. (1993) Binding affinities and cooperative interaction with bHLH activators delimit threshold responses to the dorsal gradient morphogen. *Cell* **72**, 741-752.

3. Campos-Ortega, J.A. and Hartenstein, V. (1985) The embryonic development of *Drosophila melanogaster*. *Springer-Verlag, Berlin; Heidelberg.*