Identification of hunchback cis-regulatory DNA conferring temporal expression in neuroblasts and neurons

Keiko Hirono, Jonathan S. Margolis, James W. Posakony, Chris Q. Doe

Abstract

The specification of temporal identity within single progenitor lineages is essential to generate functional neuronal diversity in Drosophila and mammals. In Drosophila, four transcription factors are sequentially expressed in neural progenitors (neuroblasts) and each regulates the temporal identity of the progeny produced during its expression window. The first temporal identity is established by the Ikaros-family zinc finger transcription factor Hunchback (Hb). Hb is detected in young (newly-formed) neuroblasts for about an hour and is maintained in the early-born neurons produced during this interval. Hb is necessary and sufficient to specify early-born neuronal or glial identity in multiple neuroblast lineages. The timing of hb expression in neuroblasts is regulated at the transcriptional level. Here we identify cis-regulatory elements that confer proper hb expression in "young" neuroblasts and early-born neurons. We show that the neuroblast element contains clusters of predicted binding sites for the Seven-up transcription factor, which is known to limit hb neuroblast expression. We identify highly conserved sequences in the neuronal element that are good candidates for maintaining Hb transcription in neurons. Our results provide the necessary foundation for identifying trans-acting factors that establish the Hb early temporal expression domain.

Keywords: Temporal identity, Hunchback, Neuroblast, Early-born neuron.
1. Results and discussion

1.1. Generation of \( h_b \) enhancer-reporter constructs

During previous studies of the \( h_b \) cis-regulatory region controlling early blastoderm expression (Margolis, 1992; Margolis et al., 1994; Margolis et al., 1995), we discovered an ~4 kb element that conferred expression in neuroblasts and GMCs but not neurons (HZ4 element; Fig. 1A). Here, we extend our analysis to subdivide this 4 kb element using a series of \( h_b \)-green fluorescent protein (\( h_b \)-GFP) constructs containing (from 5' to 3') \( h_b \) cis-regulatory DNA, the \( h_b \) P1 5' UTR including the first intron and the P2 promoter, a basal hsp70 promoter, GFP coding sequence, the \( h_b \) 3' UTR, the SV40 transcription termination sequence, and an attB sequence (Fig. 1B). We then used the PhiC recombinase system to insert all transgenese into the same genomic site on chromosome 2. This eliminates variability in transgene expression due to chromosomal position effects, which can be a problem when inserting transgenes randomly in the genome. We used EvoPrinterHD software (Odenwald et al., 2005; Yavatkar et al., 2008) to identify blocks of evolutionarily-conserved sequence within the largest construct (HG4-1; Fig. 2), and then designed smaller constructs centered around these blocks of conservation (HG4-2 through HG4-7) as shown in Figs. 1 and 2. We originally designed the constructs using a more restrictive EvoPrint showing sequences conserved among all twelve Drosophila species, but present the EvoPrint of sequences conserved among 11/12 species. This resulted in a small block of conservation between HG4-6 and HG4-7 not being included in these smaller constructs. This region of DNA could contain regulatory information that might help sum HG4-6 and HG4-7 to match the larger HG4-3 construct.

1.2. Identification of \( h_b \) cis-regulatory DNA conferring proper temporal expression in NBs

We performed RNA in situ hybridization with an antisense GFP probe to whole mount embryos expressing each of the \( h_b \)-GFP reporter transgenes (Fig. 3). The endogenous \( h_b \) gene (detected with a \( h_b \)-specific probe) shows maximal expression in NBs during stages 9 and 10 (Fig. 3, NB layer marked with black arrows). Importantly, expression in the neuroblast layer is lost by stage 11 (Fig. 3, black arrows), although there is clear expression in early-born GMCs at this stage (Fig. 3, white arrowheads). Using a GFP probe, we found that the HG4-1, HG4-3, and HG4-7 transgenes are all expressed in a pattern similar to endogenous \( h_b \); expression in NBs at stage 9 and stage 10, followed by early-born GMC expression at stage 11. Although these transgenes may have been expressed at lower levels than endogenous \( h_b \) (or the probes were less effective), it is important to note that all three transgenes showed the proper temporal pattern. The remaining HG transgenes (HG4-2, HG4-4, HG4-5, HG4-6) did not show NB expression (Fig. 3).

We do not show ventral views of the transgene RNA patterns in Fig. 3 because the expression of endogenous \( h_b \), and the transgenes, is transient in neuroblasts (at the time they form), and neuroblasts form at different times from the earliest at stage early 9 (e.g. 7–1) to the latest at late stage 11 (e.g. 7–3). Thus, the RNA in situ’s will never show the full complement of neuroblasts at a single time point. Instead, to determine if the three HG transgenes with proper temporal expression also had proper spatial expression within the NB population, we stained for GFP and the neuroblast marker Deadpan (Dpn) and imaged the field of NBs; we used the positional marker Engrailed to determine segment boundaries (Fig. 4). Because GFP has a relatively long half-life, this experiment allowed us to “sum up” the expression of the HG transgenes during stages 9 and 10 to determine if they were properly expressed in the majority of NBs in each segment. We found that the HG4-1, HG4-3, and HG4-7 transgenes each expressed in >90% of NBs per segment (Fig. 4), similar to endogenous Hb protein (Isshiki et al., 2001). We conclude that the nested DNA elements contained in HG4-1, HG4-3, and HG4-7 all possess the cis-regulatory DNA necessary for proper temporal expression in newly-formed NBs; this maps the element to the smallest transgene DNA, HG4-7, which contains 961 base pairs. We conclude that a 961 nucleotide cis-regulatory DNA element is sufficient to generate the \( h_b \) temporal expression pattern within embryonic NBs.

Interestingly, this region of DNA contains seven predicted Svp binding sites (AGGTCA or its complement, with one mismatch allowed; Fig. 2, blue highlighting), including one pair spaced 10 nucleotides apart that is predicted to allow Svp homo- or heterodimer binding (Zelhof et al., 1995); this site is not detected in any other conserved sequence blocks within the 4 kb HG4-1 region. In addition to the predicted Svp binding sites, which may
be important for limiting Hb expression, there are other blocks of conserved sequence within the HG4-7 DNA that may represent binding sites for direct or indirect Dan-mediated repression of hb transcription (Kohwi et al., 2011), or binding sites for currently unknown transcriptional activators of hb expression in NBs. Mutation of individual small blocks of conserved sequence would be necessary to test these hypotheses; sites found to be functionally important would be excellent candidates for biochemical purification and identification of trans-acting factors that regulate the timing of hb expression in neuroblasts.

1.3. Identification of hb cis-regulatory DNA conferring proper expression in early-born neurons

We next examined embryos at stage 16 to determine the expression pattern of each transgene in post-mitotic neurons. At this stage, early-born neurons are located in the deepest layer of the CNS, whereas later-born neurons occupy a more superficial position (Isshiki et al., 2001). The endogenous hb gene is weakly detected in deep layer neurons in lateral views (Fig. 5, top row). Using an antisense GFP probe, we found that the HG4-1, HG4-2, HG4-3, and HG4-6 transgenes are all expressed in deep layer neurons (Fig. 5); in addition, HG4-5 appears to be expressed in deep layer neurons and strongly expressed in superficial (late-born) neurons (Fig. 5). The transgenes HG4-4 and HG4-7 show little or no RNA expression in the stage 16 CNS (Fig. 5).

Fig. 2. Evoprint of the cis-regulatory DNA used in this study. Bold capital letters represent conservation in either all 12 or 11/12 Drosophila species for which genomic data exists. The Flybase coordinates for the last nucleotide of each row is shown to the right, and the approximate positions of the constructs used is shown to the left (see Supplemental data Table 1 for primers used in these constructs). Transcription factor predicted binding sites within the conserved blocks of DNA were determined using the TESS website (http://www.cbil.upenn.edu/cgi-bin/tess). Blue, Seven-up (Svp) binding sites; gray, Hb binding sites; yellow, Pou domain binding sites, including Nubbin/Pdm2; blue underlines, homeodomain core motifs (ATTA/TAAT) sites.
To determine the expression pattern of each transgene within early-born and late-born neurons of the CNS at higher resolution, we double stained each of the transgenic lines for GFP and Hb protein (a marker for deep-layer, early-born neurons). We found that HG4-1, HG4-2, HG4-3, and HG4-6 transgenes all showed GFP staining in deep layer neurons (Fig. 6A–G), although it was not a perfect 1-to-1 register with the Hb+ deep layer neurons. Some GFP+ neurons were in Hb-negative neurons, whereas some Hb+ neurons lacked GFP. For example, line HG4-6 had Hb+ neurons that were 19% strong GFP, 37% weak GFP, 44% no GFP (n = 328 neurons from two embryos); the lack of GFP in some Hb+ neurons is most likely due to lack of full cis-regulatory information in the transgene, but we can't rule out low levels of GFP being present. In addition, HG4-5 showed strong GFP staining in Hb-negative superficial neurons (Fig. 6K), whereas the other transgenes were weakly or not detected in superficial neurons (Fig. 6H–J, L–N); see next section. Unexpectedly, we noticed HG4-7 exhibited a clear GFP protein signal within the CNS (Fig. 6G and N) despite lack of detectable GFP RNA at this stage (Fig. 5). The GFP protein may be due to persistence of GFP from neuroblasts and GMCs into the neurons. We conclude that the smallest DNA element that drives early-born neuron expression is the 866 base pair HG4-6 element, and that all larger elements containing this fragment also show early-born neuron expression.

It is currently unknown how Hb expression is maintained in the GMCs and neurons born from Hb+ young neuroblasts. We can rule out simple positive autoregulation because misexpression of Hb in neurons does not activate hunchback transcription (Grosskortenhaus et al., 2005). The presence of several blocks of highly conserved sequence within the 866 base pair HG4-6 element would be excellent starting points for identifying the transcriptional activator(s) required to maintain Hb expression in early-born neurons.

1.4. Identification of a cis-regulatory DNA element conferring expression in late-born neurons

The HG4-5 showed strong reporter expression in Hb-negative superficial neurons (Fig. 6K), which was surprising because Hb is not detected in these neurons at any embryonic stage (Isshiki et al., 2001; Kambadur et al., 1998). None of the other transgenes were strongly expressed in these neurons (Fig. 6H–J, L–N), although we detected weak “stripy” GFP reporter expression in the superficial layer neurons in line HG4-7 (Fig. 6N). To test whether HG4-5 was preferentially expressed in late-born neurons, we assayed lateral confocal sections through the stage 16 CNS double stained for GFP and Hb protein. We observed a virtually complementary pattern of GFP and Hb expression (Fig. 6O–Q). This shows that the DNA contained in the HG4-5 transgene contains an enhancer that drives expression in late-born neurons. Interestingly, the HG4-2 construct encompasses the HG4-5 fragment but does not have similar strong expression in late-born neurons (compare Fig. 6I with 6K); this shows that HG4-2 must contain a late-born neuron silencer element. We do not know whether the late-born neuron enhancer found in HG4-5 is normally silenced by adjacent regulatory elements during wild type development, or if it drives expression of a flanking gene such as CG8112, whose promoter is closer to the HG4-5 element than the Hb promoter (CG8112 is transcribed in the opposite direction as hunchback, and its predicted promoter lies between the HG4-5 element and the hunchback promoters; i.e. the HG4-5 element is in the first intron of CG8112). Nevertheless, the serendipitous identification of a late-neuron enhancer will be useful as a tool (e.g. making a Gal4 line with late-born neuron expression);
future experiments will be necessary to determine its relevance to 
hb or CG6112 transcription.

HG4-5 is only 571 base pairs long, and contains just two blocks of 
conserved sequence. Interestingly, there are eight predicted Hb 
binding sites within these small blocks of conserved sequence, rais-
ing the possibility that Hb acts as a transcriptional auto-repressor 
to prevent the function of this enhancer fragment within early-born 
neurons (e.g. to keep the adjacent CG6112 gene off in early-born 
neurons). In this case, we might expect the remaining small region 
of conserved sequence to contain binding site(s) for a pan-neuronal 
transcriptional activator.

2. Experimental procedures

2.1. Enhancer-reporter transgenic lines and other stocks

The HZ4 construct was made as described previously (Margolis, 
1992; Margolis et al., 1994; Margolis et al., 1995); briefly, it con-
tains the DNA from Flybase position 4527893 (Xbal site) to 
4532199 (Not site) – coordinates shown in Fig. 1A – attached to 
a basal hsp70 promoter directly upstream of the lacZ coding se-
quences. Unlike the HG constructs described below, there is no 
hb 5’ or 3’ UTR sequences in the HZ4 construct; furthermore HZ4 
was randomly inserted into the genome. The HG4-1 construct con-
tains precisely the same upstream 4 kb of DNA as the HZ4 
construct, but driving a different reporter (GFP instead of lacZ). 
All HG constructs are summarized in Fig. 1A (different cis-regula-
tory fragments) and Fig. 1B (core construct that is common to all 
HG transgenes). To make the core HG construct, hb 5’UTR was 
amplified from a BAC clone, BACR01F13 using forward primer 
5’CAGTGTGCTCCGAGTCC and reverse primer 5’CTGGCCCTTGTC 
ACCATCTTTGGCGCCTTAGACCG, which contained GFP 5’ coding 
sequence. The hb 3’UTR was amplified from BACR01F13 using for-
ward primer 5’GCATGGAGAGCTTGTAACAGTTCATCCCATACC 
ATACCTTG, which contained GFP 3’ coding sequence and reverse 
primer 5’TATATTGAATAATGGATATTGTGATTGTC. We 
inserted the GFP coding sequence between the hb 5’ UTR and 3’ 
UTR by assembly PCR utilizing forward primer 5’CGGTCTAGAG 
CCGCGGATGTTGGAAGGGGGGAG, which contained hb 3’ 5’UTR 
sequence and reverse primer 5’CAAGGTGATGG TGATGGGAACCT-
TACTCTGTCAGCTCGTCATCG, which contained hb 5’ 3’UTR se-
quency. This core hb-gfp gene was cloned into SpeI-BamHI site of 
a modified pCaSpeR4 vector by inserting a basal hsp70 promoter 
(hsp70 TATA box) at NotI-SpeI site, SV40 termination signal at 
Xbal-Hpal site, and attB site at Xhol site to make the core HG vector 
shown in Fig. 1B.

Using EvoPrinterHD software (Odenwald et al., 2005; Yavatkar 
et al., 2008), we found 4 regions evolutionally conserved in Dro-
sophila melanogaster plus all 11 other Drosophila species, which 
was used to design fragments; a slightly more relaxed EvoPrint 
of D. melanogaster and 10/11 other species is shown (Fig. 2). En-

2.2. RNA in situ hybridization to embryos

Embryo whole-mount in situ RNA hybridization was carried out 
as previously published (Tautz and Pfeifle, 1989). The hb RNA 
probe was generated from BACR01F13 using forward primer 
5’GGTCCTACACTACATCCCTG and reverse primer 5’CTGGCATT 
CAGATTCGGCTT with a T7 promoter sequence. For the GFP RNA 
probe, we used forward primer 5’CTTCTCAAGTCCGATGC and reverse 
primer 5’AATCCGAGACGACCAGTG with a T7 promoter sequence. DIG 
Labeling Kit (Roche Diagnostics, Indianapolis, IN) was used for transcription and digoxigenin-labeling. Images were 
photographed with Zeiss AxioCam HRc camera on an Axioplan 
microscope DIC/Nomarski optics.

2.3. Immunofluorescence and DIC

Antibody staining was performed according to standard meth-
ods (Kohwi et al., 2011). Primary antibodies, dilutions and sources 
were: chicken anti-GFP 1:500 (Aves Labs, Inc., Tigard, OR, USA); rabbit 
anti-Hb 1:200 (Tran and Doe, 2008); mouse anti-Even-skipped 
(3C10-c) and mouse anti-Engrailed (4D9) (Developmental Studies 
Hybridoma Bank, University of Iowa, IA, USA); rat anti-Dnp (Doe lab). Donkey anti-chicken Dylight 488- and donkey anti-mouse Dyl-

2.4. Transcription factor binding site identification

Transcription Element Search System software (TESS; http:// 
www.cbil.upenn.edu/cgi-bin/tess) was used to identify predicted 
transcription factor binding sites within the conserved evoprinted 
regions shown in Fig. 2. The Svp sites were annotated as CF1/USP 
sites, but these are equivalent to Svp sites (Zelhof et al., 1995).
We only indicate high confidence sites; we excluded low information content sites under 5 base pairs. Where there were overlapping sites for a single factor, the site with the highest TESS score was used.

3. Conclusions

We conclude that the pattern of $hb$ expression in the CNS can be recapitulated in large part by the cis-regulatory DNA contained in the $\sim$4 kb HG4-1 construct. Within this stretch of DNA are contained smaller elements: a 961 bp element that confers proper $hb$ expression in newly-formed NBs (HG4-7), a 866 bp element that confers proper $hb$ expression in early-born neurons (HG4-6), and an unexpected 571 bp element that confers expression in $hb$-negative late-born neurons (HG4-5); summarized in Fig. 7. The NB element contains predicted binding sites for a known trans-acting factor that regulates $hb$ NB expression (Svp) and several other conserved regions that may represent transcriptional activator binding sites. The results of our study will provide a foundation for future functional analysis of this temporal cis-regulatory DNA, and the identification of trans-acting factors that control the precise timing of $hb$ expression in the CNS.

Acknowledgements

We thank Minoree Kohwi for comments on the manuscript. This work was funded by an NIH grant GM41100 to J.W.P. and NIH grant HD27056 to C.Q.D., who is an Investigator of the Howard Hughes Medical Institute.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.gep.2011.10.001.

References

K. Hirono et al. / Gene Expression Patterns 12 (2012) 11–17
