**Int. J. Dev. Biol. 61:** 537-543 (2017)
doi: 10.1387/ijdb.170194oo

THE INTERNATIONAL JOURNAL OF DEVELOPMENTAL BIOLOGY
www.intjdevbiol.com

# An empirical model of Onecut binding activity at the sea urchin *SM50* C-element gene regulatory region

OCHAN OTIM*

*Division of Biology, California Institute of Technology, Pasadena, CA, USA*

**ABSTRACT Studying the formation of endoskeleton in many species is complex and difficult. The sea urchin embryo offers an unparalleled platform for understanding this process because of the ease with which its skeletogenic mesenchyme cells can be manipulated. In this study, preliminary evidence from biochemical studies towards understanding the role of the Onecut transcription factor during sea urchin skeletogenic mesenchyme cell specification is presented. Based on the evidence, an empirical model is proposed showing how Onecut, together with associated co-factors, may be using the C-element of the *SM50* gene regulatory region in advance of the sea urchin *Strongylocentrotus purpuratus* spicule development. In the model, Onecut recognizes and binds the DNA sequence CATCGATCTC in the C-element without temporal restriction. Onecut then utilizes different sets of co-factors to switch from its unknown function early in development (four cell stage to the mesenchyme blastula stage), to its known role in the oral-aboral boundary thereafter. At the writing of this report, definitive evidence as to whether the "early" factors are expressed in all cells except the micromere lineages, or whether the "late" factors are expressed in micromere descendants or ectodermal precursors only are lacking. The former would suggest a possible Onecut repression function for the early co-factors outside the micromere lineages; the latter scenario would suggest a Onecut activation function for the late co-factors in the presumptive ciliary band.**

KEY WORDS: *C-element, HNF6, Onecut, sea urchin, SM50*

## Introduction

Studying the formation of endoskeleton in many species is a difficult process because of the complex nature of tissues involved. The purple sea urchin (*Strongylocentrotus purpuratus, S. purpuratus*), with its elaborate endoskeleton formation, has been playing a central role in advancing knowledge in this area because of the ease with which the primary mesenchyme cells (PMC) can be manipulated micro-surgically (Ettensohn and McClay 1988, and references therein), or separated and cultured *in vitro* until deposition of mineralized tissue (Okazaki 1975; Lee *et al.,* 1999). The PMC is a group of 32 cells that synthesize the sea urchin endoskeleton (see Lyon *et al.,* 2014, and Minokawa 2016 for recent descriptions of the role of PMC during sea urchin development). Central to embryonic skeletal development studies in sea urchin are several cloned genes (Peled-Kamar *et al.,* 2002; Sucov *et al.,* 1987; Wilt 1999), but whose gene regulatory networks are yet to be completely dissected. The regulation of one such gene, the *SM50*, is of interest to us. *SM50* encodes a differentiation product utilized exclusively in the skeletogenic lineages. That the expression of *SM50* is spatially restricted during development was demonstrated by Sucov *et al.,* (1988) who used a fusion construct *SM50.CAT* to show accurate spatial embryonic expression in the PMC. The *SM50.CAT* fusion construct included a genomic sequence from -440 to +120 with respect to the transcription start site. This genomic sequence was later shown to contain four adjacent regulatory elements, three of which are positively acting, and one, the C-element positioned downstream of the transcription start site and the subject of this study, is an indispensable spatial control element (Makabe *et al.,* 1995). Another *S. purpuratus* gene which has also been shown to contain the C-element in its putative regulatory region is the *SM37* gene (Lee *et al.,* 1999). Compared to the *SM50* C-element, the *SM37* C-element has four base mismatches and one deletion. The role of C-element in *SM37* regulation is not defined as yet. Within the broader Strongylocentrotidae family, the C-element is conserved among five of the eight closely related species (Walters *et al.,* 2008).

*Abbreviations used in this paper:* EMSA, electrophoretic mobility shift assay; PMC, primary mesenchyme cell.

***Address correspondence to:** Ochan Otim. Environmental Monitoring Division, City of Los Angeles, 12000 Vista Del Mar, Playa Del Rey, CA 90293, USA. Tel: +1 (310) 648-5835. Fax: +1 (310) 648-5828. E-mail: ochan.otim@lacity.org https://orcid.org/0000-0001-7272-4356

*Submitted:* 9 August 2017; *Accepted:* 19 September, 2017.

ISSN: Online 1696-3547, Print 0214-6282

One sea urchin gene product which we believed has a role in regulating *SM50* is Onecut (also known as HNF6), a factor of the cut-domain family of homeodomain proteins. In mammalian cells, Onecut is enriched in the liver, where it stimulates transcription of liver-expressed genes, and antagonizes glucocorticoid-stimulated gene transcription (Rausa *et al.,* 2003). Recently, Onecut protein was shown to recognize the -5/-6 region of the human *F9* promoter (5'-AACTAATCGACCTTACCA-3', Funnell *et al.,* 2013), a sequence that includes an ATCGA motif in the Sucov *et al.,* (1988) *SM50. CAT* fusion construct.

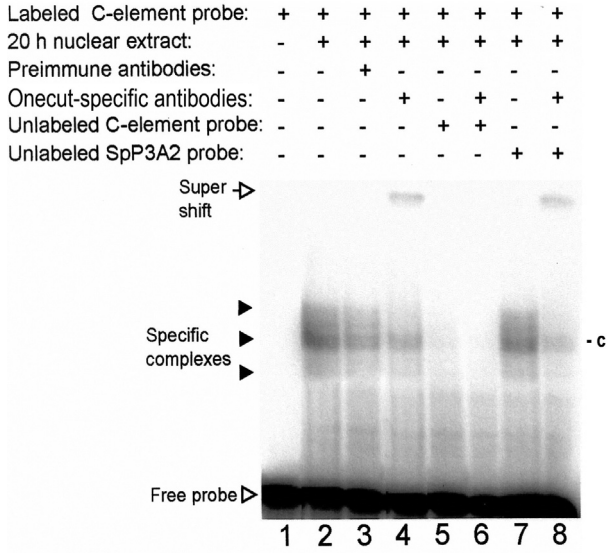
In a preliminary study using recombinant Onecut expressed in bacteria, we showed that the factor binds specifically to the C-element (Fig. 10 in Otim *et al.,* 2004) and, therefore, should be involved in embryonic endoskeleton formation. On that basis, we expected Onecut to regulate *SM50* specifically in the skeletogenic cells or the PMC. However, attempt to test this hypothesis was sidetracked by the discovery of the spectacular role this factor plays in the development of the ciliated band, the oral-aboral boundary marker. Out of curiosity, we also found by whole-mount *in situ* hybridization that Onecut plays a similar role in marking the oral-aboral boundary in a sea star *Asterina miniata* even though sea star and sea urchin diverged from a common ancestor around 500 million years ago (Otim *et al.,* 2005). Knockdown of Onecut in sea urchin led to multiple consequences on embryonic development. This included failure to form fully developed spicules, the abnormal growth of a perfectly spherical embryo, the uncharacteristic formation of a centralized archenteron which fails to bend and fuse with the oral ectoderm (presumably because the oral ectoderm is absent in the knockdown), and the proliferation of pigment cells symmetrically arranges around the archenteron (Fig. 7 in Otim *et al.,* 2004).

In this study (cited in its preliminary stage by Walters *et al.,* 2008), biochemical tools including extensive electrophoresis mobility shift assay (EMSA, Hellman and Fried 2007) are employed in an attempt to assess, sort, and define empirically Onecut binding activities at the C-element sequence of *SM50* regulatory region in *S. purpuratus*. Results reported should inform subsequent effort to understand the contribution of Onecut to the accurate expression of *SM50* in the PMC or the skeletogenic lineages. Such effort may involve identifying and functionally analyzing all the direct and indirect Onecut-associated C-element binding proteins identified in this work.

## Results

### *Onecut is present in nuclear extracts and binds C-element DNA sequence specifically*

Classic EMSA was used to definitively show that in a 20-h old embryonic nuclear extract, factors that bind to the C-element DNA sequence found in the regulatory region of *SM50* are present. The results confirm the presence of such proteins that can form specific complexes (indicated by three arrowheads in Fig. 1, lanes *2-4, 7-8*) with the wild-type C-element probe (Labeled C-element probe in Fig. 1). The complexes were lost with an excess amount of unlabeled C-element competitor (Fig. 1, lanes *5, 6*) and were supershifted with Onecut-specific antibodies (lanes *4, 8*), but not affected by the presence of preimmune antibodies (Fig. 1, lane *3*) or unlabeled SpP3A2 probe (Fig. 1, lanes *7-8*). The SpP3A2 target site is unrelated to the *SM50* C-element. The preliminary




**C-element probe: gatcTGGTAGTCGTGAATGCATCGATCTC**

**Fig. 1. Specificity test of Onecut-SM50 C-element binding by electrophoretic mobility shift assays (EMSA).** *EMSA of a 20-h* Strongylocentrotus purpuratus *embryonic nuclear proteins extract with the* SM50 *$^{32}$P-labeled C-element gatcTGGTAGTCGTGAATGCATCGATCTC oligonucleotide (lanes* 2-8*), rabbit pre-immune serum antibodies (lane* 3*), Onecut-specific rabbit antibodies (lanes* 4,6,8*), unlabeled C-element probe (lane* 5,6*) and unlabeled* S. purpuratus *P3A2 probe (lane* 7-8*). Each binding reaction contained nuclear extract (2 μL), poly(dI-dC)/poly(dI-dC) (200 ng) and $^{32}$P end-labeled probe (0.4 ng). The reactions were incubated on ice for 10 min prior to electrophoresis. The complex labeled "*c*" contains Onecut.*

conclusion from these experiments is that the observed DNA-protein complexes (labeled *c*) contain Onecut as the major protein bound to the C-element specifically. Note the reduction of the band intensities by adding antibodies to Onecut in Fig. 1 which occurs in all the bands referred to as "specific complexes" (compare lanes 4 and 8 with lanes 2 and 7, respectively).

### *Onecut binding on the C-element is limited to the sequence CATCGATCTC*

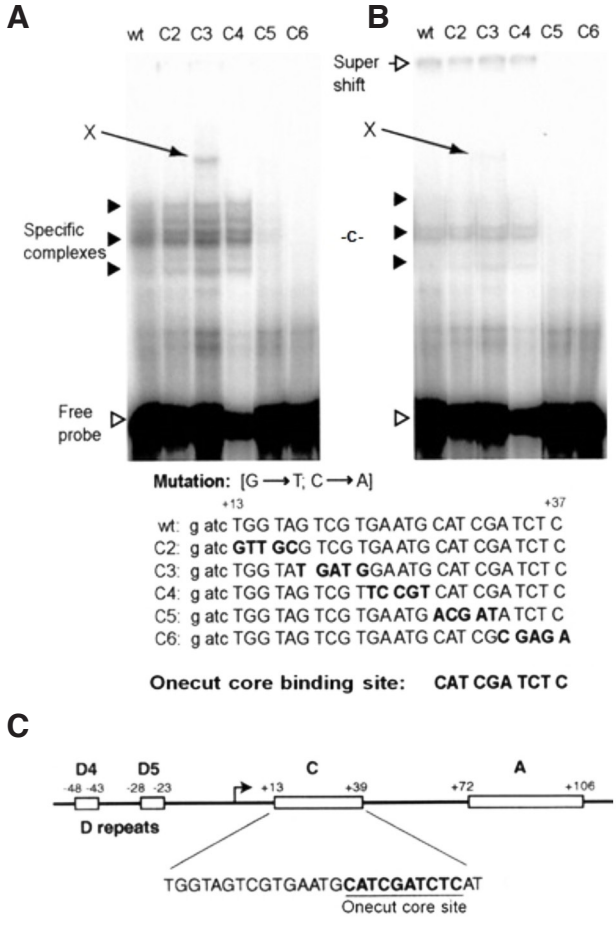
In traditional EMSA, the $^{32}$P-labeling of wild type probe is preferred. In this particular EMSA experiments however, $^{32}$P-labeled mutants C2-C6 competing with *limiting* amounts of unlabeled C-element wild type probe were preferred to accomplish two goals: to see if the mutations introduced specific binding sites, and to determine the exact target site on the C-element sequence of Onecut. EMSA was carried out with 20-h nuclear extract. The use of limiting amounts of the wild type probe unambiguously optimized signal differences between DNA binding site not damaged by mutation and when only the wild type probe contains the intact binding site. The results presented in Fig. 2 show the formation of specific DNA-protein complexes (indicated by three arrowheads) with three of five transversely mutated C-elements C2, C3 and C4 when compared to *wt*, the wild type (Fig. 2A, lanes *1-4*). The DNA-protein complexes were lost when $^{32}$P-labeled mutants C5 and C6 were used (Fig. 2A, lanes *5-6*) implying that the unlabeled wild type probe effectively outcompeted mutants C5 and C6 because these two mutant sequences are not recognized by the specific DNA binding proteins in the 20-h nuclear extract. (Indeed, it ap-

pears more bands than the major ones indicated by arrowhead are affected by C5 and C6 mutations.) These results indicate (i) the absolute requirement of the10 bases CATCGATCTC (+28 to +37 with respect to transcription start site) for Onecut binding activities at the C-element, and (ii) that more than one protein is involved in this binding. The indispensable sequence CATCGATCTC includes a 6-base palindromic sequence ATCGAT as expected for a DNA binding site used by regulatory protein (Wyrwicz *et al.,* 2007, Lis and Walther 2016), but with zero intervening length in the *S. purpuratus SM50* C-element.

All the DNA-protein complexes are supershifted by Onecut specific antibodies except for C5 and C6 mutants (Fig. 2B) confirming not only the specificity of this complexes to Onecut protein, but that Onecut (with co-factors in tow) binds the DNA sequence CATCGATCTC directly. The relative location of the indispensable 10 bases spanning C5 and C6 within the C-element is shown in Fig. 2C together with the three adjacent positively acting regulatory elements identified by Makabe *et al.,* (1995): two D-elements upstream and one A-element downstream of the transcription start site. Mutation represented by C3 introduces a non-*SM50* sequence recognized specifically by a factor labeled X. The activity of X would not have been visible had we used $^{32}$P-labeled wild type probe.

### *Onecut forms unique temporal DNA-protein complexes during development*

To acquire precursory information on contributors to the DNA-protein complexes formed with the participation of Onecut in sea urchin development, EMSA was carried out using nuclear extracts prepared at an interval of about 3 h over the first 26 h after fertilization (Fig. 3). The EMSA temporal profile reveals at least eight unique DNA-protein complexes (labeled *a-h*). Complexes *b, d-g* are clearly formed by early acting factors, and *h* by late factors. Complex *a* is observed transiently around 9 to 18 h. The complex labeled *c* contains Onecut (evidently supershifted by Onecut specific antibodies, Figs. 1-2) and is temporally unrestricted. The



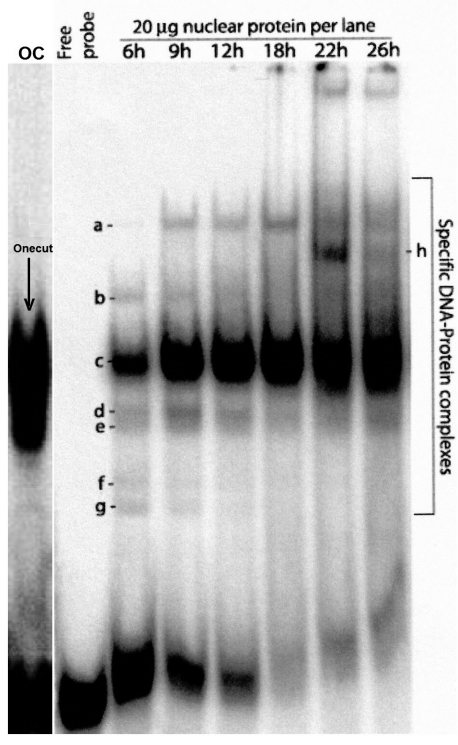




**Fig. 2 (left). Determining Onecut target within the C-element.** *The C-element contains an indispensable Onecut target site comprising 10 bases as revealed by EMSA using transversion mutations shown below in bold face* (G *to* T*, and* C *to* A). *In lane* 1*, the control experiment, the wild type probe (wt) is* $^{32}$*P-labeled. In lanes* 2-6*, the mutants C2-C6 were* $^{32}$*P-labeled instead of, and competing with, unlabeled wild type probe for binding protein in the 20-h nuclear extract. DNA-protein complexes (also labeled* c*) are indicated by three arrowheads.* **(A)** *No Onecut specific antibody added;* **(B)** *DNA-protein complexes supershifted by binding to Onecut specific antibodies. Note that specific and supershift complexes are not observed with C5 and C6 mutants because the unlabeled wild type effectively outcompeted the mutants C5 and C6. These mutants are not recognized by the DNA binding proteins in the 20-h nuclear extract.* **(C)** *Makabe* et al., *(1995)* cis*-regulatory control elements of the* SM50 *gene showing the indispensable 10 bases found in this study within the C-element to be required for Onecut binding.*

**Fig. 3 (right). Temporal profile of the C-element binding proteins over the first 26 h of embryonic development as revealed by EMSA.** *Embryonic nuclear proteins extracts were harvested from a vat of synchronously developing embryos at different stages of development (shown above) were incubated for 10 min with* SM50 $^{32}$*P-labeled C-element oligonucleotide in the presence of 200 ng poly(dI-dC)/poly(dI-dC) and analyzed by electrophoresis. Lane "OC" represents an experiment with* in vitro *Onecut (see Otim* et al.*, 2004 for details) and is included to identify complex containing Onecut,* c.

conclusion is that Onecut is constitutively present in all extracts and is associated collectively with at least eight factors within the developmental time frame of this study.

***Model of Onecut interaction at the C-element***

To determine the composition of the two distinct multimeric Onecut specific complexes detected early and late in development (Fig. 3), EMSA was carried out with nuclear extracts prepared from the four cells developmental stage embryos (6 h post-fertilization, Fig. 4A) and the blastula stage embryos (22 h post-fertilization, Fig. 4B) using limited amounts of (and this time) $^{32}$P-lebeled wild-type C-element competing with *excess* unlabeled mutants C2-C6. This reversed labeling *vis-a-vis* experiments in Fig. 2 was to allow contrast between these experiments and those presented in Fig. 2. Results show that in both the four cell stage and the blastula stage, the mutants C2, C3 and C4 outcompeted the wild-type for the Onecut major complex formation (indicated by *c* and the oval shape in Fig. 4, and identified by gel supershift assay in Figs. 1 and 2) confirming that the Onecut binding site is not affected by mutations C2, C3 and C4 of the C-element. The two adjacent mutations at C5 and C6 covering 10 bases render the unlabeled mutants uncompetitive for the Onecut associated complex. Although few in numbers in these experiments, the $^{32}$P-labeled wild type probe was able to bind effectively to the Onecut associated proteins in the nuclear extracts at the expense of the mutants. These results confirm the importance of the stretch of DNA sequence CATCGATCTC to Onecut binding.

Also shown in Fig. 4 as two 10-point star polygons (labeled "Additional not Onecut" and *b*) are factors that bind the 25 base pair study sequence in a way distinctively different from the pattern shown by Onecut. The signal strength of these DNA-protein complexes appear to suggest that these additional factors are either rare or that their binding activities are weak and not affected by mutations C2, C4, C5 and C6. One conclusion is clear from these results; that these additional factors are not competing with Onecut for the same site. Mutation C3 appears to reduce binding activities of these factors but enhances the presumptive Onecut binding activity (lane C3 in Fig. 4A). This may mean the *b* factors require sequence C3 to stay detached from Onecut; with the loss of C3, the *b* factors are attached to Onecut. We also believe that these factors are bound directly to DNA at C2 and C4 (possibly C3 as well) since C2, C3 and C4 mutants do not erase the $^{32}$P-labeled wild type probe competitive binding, but the Onecut associated DNA protein complex (oval, Fig. 4).

The relative signal strength of the early factors (Fig. 4A, pentagon, two bands *d-g* indicated by arrows) and the late factor (Fig. 4B, rectangle, single band *h* indicated by arrow) appears to mirror proportionally that of Onecut in that both sets of signals are diminished by mutations at C2, C3 and C4. This may suggests that the two factors (early and late) are *trans*-acting and are directly associated with Onecut.

In Fig. 4C is shown an integrated proposal summarizing all the interactions discussed at the two embryonic developmental stages (the four cell and the blastula stages, Figs. 4A and 4B, respectively). The constitutively present factors labeled collectively as *b* are bound at C2 and C4; Onecut with its two *trans*-acting switching factors (*d-g* and *h*) are bound to the DNA sequence stretch whose mutations are labeled C5 and C6.

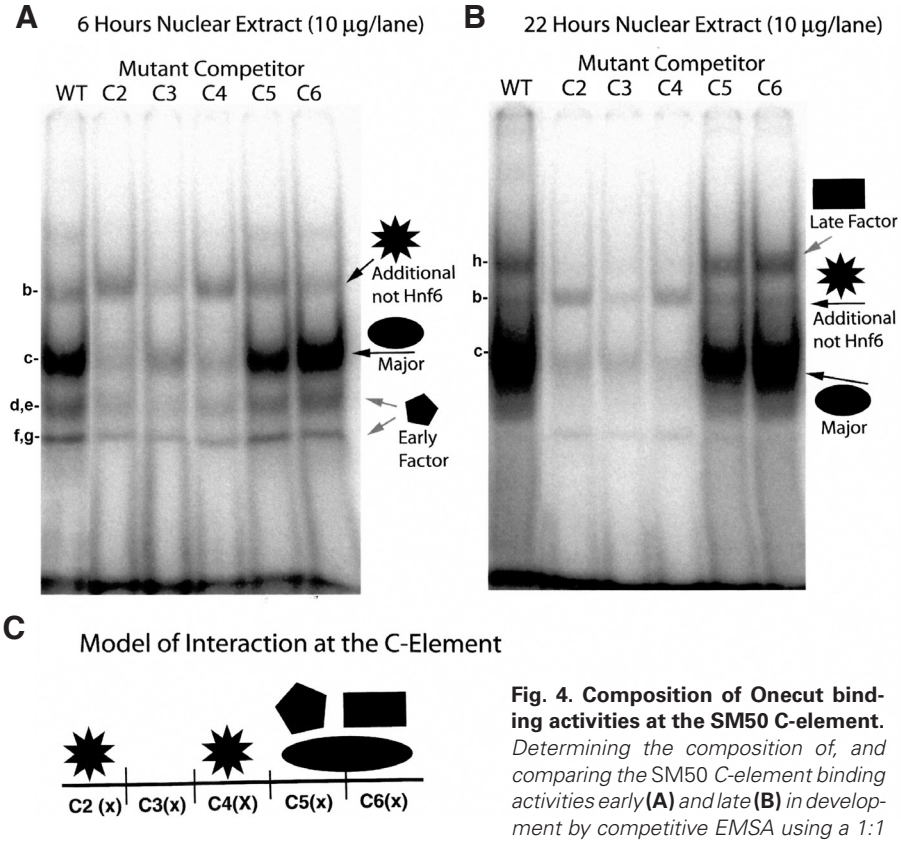



**Fig. 4. Composition of Onecut binding activities at the SM50 C-element.** *Determining the composition of, and comparing the* SM50 *C-element binding activities early* **(A)** *and late* **(B)** *in development by competitive EMSA using a 1:1 mixture of unlabeled mutants (Fig. 2) and $^{32}$P-labeled wild-type C-element in all lanes (unlike experiments in Fig.2 where the mutants are $^{32}$P-labeled in lanes* 2-6*). Other conditions are as described in Fig. 1.* **(C)** *An empirical model summarizing our interpretation of Onecut binding activity at the sea urchin SM50 C-element gene regulatory region as revealed by EMSA.*

***Onecut binding activity is ionic strength dependent***

To glean into the stoichiometry of binding at the C-element and, particularly, to determine whether Onecut binding depends on ionic strength, a 20-h nuclear extract was passed through a tandem of C-element oligonucleotide affinity columns, washed extensively and eluted in steps with solutions containing increasing amounts of KCl as shown on top of Fig. 5A. Ten fractions were collected and the specific DNA-protein binding activities were monitored in each fraction by EMSA on a 10% native polyacrylamide gel; other EMSA conditions are as described in Fig. 1 caption. The EMSA profile of the ten fractions were compared with that of an experiment performed using unpurified 20-h nuclear extract (lane C, Fig. 5A) to identify the specific complexes *c* and *h* present in the late embryonic stages (Fig. 3). The affinity purified wild-type Onecut protein was identified by Southwestern blot analysis (Fig. 5B) and its molecular mass was determined to be about 52 kDa as expected (the size of

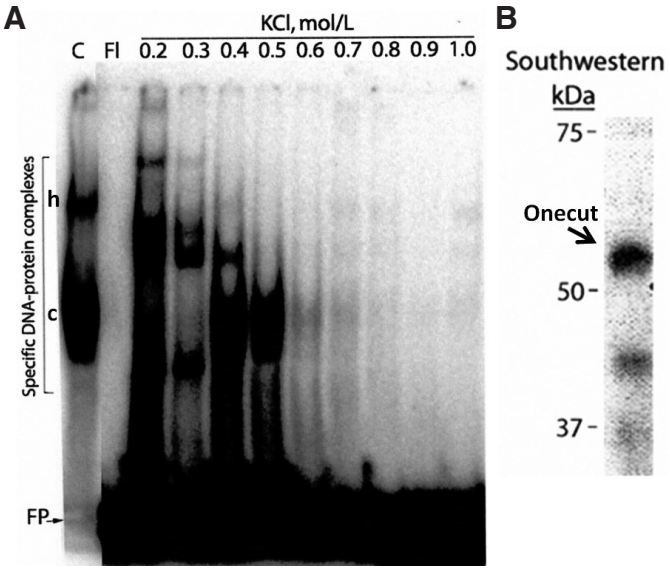



**Fig. 5. Purification of native Onecut proteins. (A)** *Partial purification of Onecut binding proteins (late acting factors) from a 20-h embryonic nuclear extract by affinity chromatography. The sequence of the concatenated oligonucleotide used for affinity chromatography was 5'gatcTGGTAGTCGTGAATGCATCGATCTC (probe) corresponding to +13 to +37 region from the transcription start site of Onecut (Fig. 2C without the last AT). C, Fl and FP refer to unpurified binding reaction used here as negative control of elution, flow-through column blank and free probe as the positive control, respectively. The nuclear extract was passed through a tandem of oligonucleotide affinity columns, washed extensively and eluted in steps with solutions containing increasing amounts of KCl as shown on top. Ten fractions were collected and the specific DNA-protein binding activities were monitored in each fraction by mobility shift assay on a 10% native polyacrylamide gel as described in Fig. 1 caption. At the minimum, C-element binding factors* c *and* h *(Fig. 3) are present in the 20 h nuclear extracts (no salt, lane C). Notice that the complex hardest to disrupt is the Onecut-DNA complex* c *(*c *is missing in 0.3 mol/L KCl eluent);* h *appears to be loosely bound and is completely eluted at 0.3 mol/L KCl.* **(B)** *Identifying affinity purified transcription factors by Southwestern blot analysis. Onecut is shown by the arrow (molecular mass ~ 52 kDa).*

full length Onecut synthesized from cDNA clone 195A17 (Otim *et al.*2004) is 52 kDa). It is clear from Fig. 5A that increasing KCl concentration disrupts rather than augments the binding of Onecut protein to DNA. This finding is outside recent discoveries that binding of some transcription factors to DNA displays virtually zero dependence on ionic strength under these conditions (Mikles *et al.,* 2014). The most stable DNA-protein complex to disruption is *c*, that which contains Onecut. Indeed, the DNA motif utilized by Onecut is known to form a stable DNA-protein complex (Nepveu 2001). The C-element binding factor *h* found in the 22-h nuclear extract shown in Fig. 3 is present as well in the 20-h nuclear extract used here (Fig. 5A); factor *h* appears to be loosely bound and is completely eluted at 0.3 mol/L KCl. Note that no band is present at the position corresponding to band *c* in the 0.3 mol/L KCl eluent; this is an anomaly. Note also that only *h* and *c* bands in the control experiment (no salt, lane C) can be labeled with certainty in Fig. 5A (and may be *c* elsewhere) because of ambiguity in predicting the effect of salt on this multiplex DNA-protein associations. Additionally, the range of ionic strengths used in these experiments is not found in living sea urchin cells and, therefore, these experimental results cannot be used to discuss the function of Onecut as a transcription factor.

## Discussion

### *Structure of Onecut binding site and co-factors involved*

Transcriptional regulation results from a complex organization of *cis*-acting sequences that serve as binding sites for a multitude of *trans*-acting factors which together determine the activity of a gene. In higher eukaryotes, these *cis*-acting sequences are typically clustered into discrete functional modules. We found such clustering to be present in the C-element regulatory region of the *SM50* gene. At least three short sequences labeled C2 (TGGTA), C3 (GTCGT), C4 (GAATG), and C5-C6 (CATCGATCTC) are clearly required for the regulatory control of this gene, the latter of which is specifically utilized by Onecut. Detected also were at least eight factors associated with the C-element; five of these interactions are preliminarily accounted for in Fig. 4C.

### *Early factor and late factor are potential determinants of Onecut function*

Based on data obtained, the function of "early factor" and "late factor" would be of interest to understanding the role of Onecut in the context of this work. Given that the nuclear extracts used in this study are a mixture of transcription factors expressed in different embryonic territories, the first question to ask would be whether these early and late factors are expressed in different portion of embryos (such as micromere descendants or ectodermal precursors). If so, these co-factors should be prime candidates for defining the role of the Onecut transcriptional complex. In one scenario, Onecut could be the portion responsible for DNA binding in the complex, while the co-factors define the specific function of the complex. The early factors could then function as transcription repressors of Onecut, and are expressed in all cells except the micromere lineages while the late factors functions as transcription activators, and are expressed in the presumptive ciliary band. In this arrangement, the early factors functionally compete with the late factors for Onecut binding which effectively means Onecut is neutral for transcriptional function without these co-factors. Such an explanation (that the DNA sequence CATCGATCTC within the C-element has no transcriptional activity on its own) would fit well with Makabe's hypothesis (Makabe *et al.,* 1995). A similar relationship is the well-known Tcf, β-catenin, and groucho transcription system whereby the transcription factor Tcf is responsible for DNA binding, the cofactor β-catenin then mediates transcription activation in embryonic territory where the co-factor groucho level is low, and groucho mediates transcription repression in territory where β-catenin level is low (Range *et al.,* 2005).

In an alternative scenario, Onecut may actually be using a very different regulatory element (not the C-element) to regulate the formation of the presumptive ciliary band, hence linking the oral-aboral axis formation to results describe here at this time would be premature.

The following conjecture could also be made about the function of Onecut even though the experiments described in this paper are inadequate. Of the four *SM50* elements (A, C, D4 and D5, Fig. 2), the C-element has been shown to have two very specific roles: one of directing *SM50* expression to the PMC, and the other whose mutation kills the ability of *SM50* to express at all (Makabe *et al.,*

1995). Furthermore, it is known that the palindromic ATCGAT DNA sequence (shown in this study to bind Onecut, C5-C6 in Fig. 2) enables factors with the cut homeodomain (such as Onecut), to repress transcription by forming a stable repressor DNA-protein complex (Nepveu, A. 2001). It then becomes impossible to rule out Onecut transcription complex from being part of a repression mechanism of *SM50* expression elsewhere in a developing embryo but the PMC or the skeletogenic lineages. This is supported by previous finding that (i) Onecut mRNA is present everywhere in the embryo until about 30 h after fertilization when expression is localized in the region where the ciliated band would arise (see Fig. 6 in Otim *et al.,* 2004; *hnf6* is *Onecut)* and that (ii) no correlation was observed between the expression pattern of *SM50* and Onecut 24 h after fertilization (see Fig. 11 in Otim *et al.,* 2004). This means the repressive role of Onecut complex in this context is not required everywhere except where it is needed to regulate the development of the ciliated band after the sea urchin's endoskeleton has been assigned to the right address.

## Concluding remarks

We have shown by EMSA that proteins in a 6-h and a 20-h sea urchin embryonic nuclear extract bind the C-element of the *SM50* gene specifically, and that one complex is the major (strongest $^{32}$P signal) and the most tightly bound to DNA (requiring high concentrations of KCl to disrupt). We have also demonstrated that the major DNA-protein complex is supershifted by antibodies raised against Onecut implying a protein in the major complex is Onecut. This was confirmed by both western and southwestern blots. Through mutational studies, the exact DNA region within the C-element recognized by Onecut protein was found to be this short sequence: CATCGATCTC. A five time-point EMSA "titration" after fertilization suggests that Onecut utilizes different co-factors as control switch at different regulatory stages of development. The detail of this complexity was investigated at the four cell stage and at the blastula stage of embryonic development by competitive EMSA using a 6-h and a 22-h sea urchin embryonic nuclear extract, respectively. Demonstratively, a distinct and low level collection of early factors, and of late factors appear to be but unaffected by mutations. A model of interactions at the C-element is proposed. Still needed though to firmly support the proposed model, a largely untested hypothesis, are experiments revealing the identity, function and expression of the putative auxiliary factors.

## Materials and Methods

***Expression and purification of Onecut bacterial protein***

The cDNA clone 224C2 (Fig. 1A, Otim *et al.,* 2004) was double digested with *EcoR*I (New England Biolabs, Ipswich, MA, USA) at an internal site 160 bp from the 5' end of the clone, and with *Hind*III (New England Biolabs, Ipswich, MA, USA) at a site that adds six amino acids (34 bp containing a stop codon) from a plasmid vector pSPORT (Invitrogen, Carlsbad, CA, USA) that was used to construct a 20-h cDNA library. The digest was cloned in frame into the *Hind*III and *EcoR*I sites of the pRSETC expression vector (Invitrogen, Carlsbad, CA, USA) and expressed to produce a 39 kD recombinant SpCut224 fusion protein. Antibodies were raised against this fusion protein, the specificity of which was tested by western blots using nuclear extracts prepare from 6-h, 12-h, 24-h, 48-h, and 72-h embryonic developmental stages (Otim *unpublished*), and by southwestern blot.

***Electrophoresis mobility shift assay***

The synthetic *SM50* C-element wild type (wt) sequence used as EMSA probes on the post-fertilization nuclear extracts was a double stranded DNA strand resulting from annealing a 1:1 mole ratio of sequences *SM50C+1*: 5' gatcTGGTAGTCGTGAATGCATCGATCTC and *SM50C-1*: 5' gatcGAGATCGATGCATTCACGACTACCA. The sequence spans the region from +13 to +37 nucleotides from the transcription start site. Nucleotides in lower case are terminal tags for end labeling. For experiments involving the C-element mutants, the region to be mutated was transversionally changed (i.e. G to T and C to A). Mutations were introduced in the form of synthetic oligonucleotides.

***Temporal scale definition***

In this report, early is arbitrarily defined as the first 6 h after fertilization; late as more than 20 h after fertilization. Hours between 6 and 20 are assumed to contain different shades of events at 6 h and 20 h.

*Acknowledgments*

*The author thanks Dr. Eric H. Davidson (EHD) of this institute for guidance and Dr. Chiou-Hwa Yuh, also of this institute, for unparalleled hands-on support. The author also thanks Dr. Takuya Minokawa of Tohoku University, Japan, and Dr. Gabriele Amore of the Istituto Regionale Vini E Oli, Italy, for critically reviewing this manuscript. This work was supported by NIH grant HD-37105 to EHD and by the Caltech Beckman Institute. The author has no industrial links or affiliations*

## References

COFFMAN J A and YUH C H (2004). Identification of Sequence-specific DNA Binding Proteins. *Methods Cell Biol* 74: 653-675.

ETTENSOHN C A and MCCLAY D R (1988). Cell lineage conversion in the sea urchin embryo. *Dev Biol* 125: 396-409.

FUNNELL A P, WILSON M D, BALLESTER B, MAK K S, BURDACH J, MAGAN N, PEARSON R C, LEMAIGRE F P, STOWELL K M, ODOM D T, FLICEK P and CROSSLEY M. (2013). A CpG Mutational hotspot in a ONECUT binding site accounts for the prevalent variant of Hemophilia B Leyden. *Am J Hum Genet* 92: 460-467.

HELLMAN L M and FRIED M G (2007). Electrophoretic mobility shift assay (EMSA) for detecting protein–nucleic acid interactions. *Nat Protoc* 2: 1849-1861.

LEE Y-H, BRITTEN R J and DAVIDSON E H (1999). *SM37*, a skeletogenic gene of the sea urchin embryo linked to the *SM50* gene. *Develop Growth & Differ* 41: 303-312.

LIS M and WALTHER D (2016). The orientation of transcription factor binding site motifs in gene promoter regions: does it matter? *BMC Genomics* 17: 1-21.

LYONS D C, MARTIK M L, SAUNDERS L R and MCCLAY D R (2014). Specification to biomineralization: following a single cell type as it constructs a skeleton. *Integr Comp Biol* 54: 723-733.

MAKABE K W, KIRCHHAMER C V, BRITTEN R J and DAVIDSON E H (1995). Cis-regulatory control of the *SM50* gene, an early marker of skeletogenic lineage specification in the sea urchin embryo. *Development* 121: 1957-1970.

MIKLES D C, BHAT V, SCHUCHARDT B J, MCDONALD C B and FAROOQ A (2014). Enthalpic factors override the polyelectrolyte effect in the binding of EGR1 transcription factor to DNA. *J Mol Recogni* 27: 82-91.

MINOKAWA T (2016). Comparative studies on the skeletogenic mesenchyme of echinoids. *Dev Biol* 427: 212-218.

NEPVEU A (2001). Role of the multifunctional CDP/Cut/Cux homeodomain transcription factor in regulating differentiation, cell growth and development. *Gene* 270: 1-15.

OKAZAKI K (1975). Spicule formation by isolated micromeres of the sea urchin embryo. *Amer Zool* 15: 567-581.

OTIM O, AMORE G, MINOKAWA T, MCCLAY D R and DAVIDSON E H (2004). SpHnf6, a transcription factor that executes multiple functions in sea urchin embryogenesis. *Dev Biol* 273: 226-243.

OTIM O, HINMAN V F and DAVIDSON E H (2005). Expression of AmHNF6, a sea star orthologue of a transcription factor with multiple distinct roles in sea urchin

development. *Gene Expr Patterns* 5: 381-386.

PELED-KAMAR M, HAMILTON P and WILT F H (2002). Spicule matrix protein LSM34 is essential for biomineralization of the sea urchin spicule. *Exp Cell Res* 272: 56-61.

RAFIQ K, CHEERS M S and ETTENSOHN C A (2012). The genomic regulatory control of skeletal morphogenesis in the sea urchin. *Development* 139: 579-590.

RANGE R C, VENUTI J M and MCCLAY D R (2005). LvGroucho and nuclear beta-catenin functionally compete for Tcf binding to influence activation of the endomesoderm gene regulatory network in the sea urchin embryo. *Dev Biol* 279: 252-267.

RAUSA F M, TAN Y and COSTA R H (2003). Association between hepatocyte nuclear factor 6 (HNF-6) and FoxA2 DNA binding domains stimulates FoxA2 transcriptional activity but inhibits HNF-6 DNA binding. *Mol Cell Biol* 23: 437-449.

SUCOV H M, BENSON S, ROBINSON J J, BRITTEN R J, WILT F and DAVIDSON E H (1987). A lineage-specific gene encoding a major matrix protein of the sea-urchin embryo spicule.2. Structure of the gene and derived sequence of the protein. *Dev Biol* 120: 507-519.

SUCOV H M, HOUGH-EVANS B R, FRANKS R R, BRITTEN R J and DAVIDSON E H (1988). A regulatory domain that directs lineage-specific expression of a skeletal matrix protein in the sea urchin embryo. *Genes Dev* 2: 1238-1250.

WALTERS J, BINKLEY E, HAYGOOD R, and ROMANO L A (2008). Evolutionary analysis of the *cis*-regulatory region of the spicule matrix gene SM50 in strongylocentrotid sea urchins. *Dev Biol* 315: 567-578.

WILT F H (1999). Matrix and mineral in the sea urchin larval skeleton. *J. Struct. Biol* 126: 216- 226.

WYRWICZ L S, GAJ P, HOFFMANN M, RYCHLEWSKI L and OSTROWSKI J (2007). A common *cis*-element in promoters of protein synthesis and cell cycle genes. *Acta Biochim Pol* 54: 89-98.

## Further Related Reading, published previously in the *Int. J. Dev. Biol.*

**Role of Mad2 expression during the early development of the sea urchin**
Odile Bronchain, Wael Jdey, Laetitia Caraty and Brigitte Ciapa
Int. J. Dev. Biol. (2017) 61: 451-457
https://doi.org/10.1387/ijdb.170053bc

**Genome-wide identification of enhancer elements**
Sarah Tulin, Julius C. Barsi, Carlo Bocconcelli and Joel Smith
Int. J. Dev. Biol. (2016) 60: 141-150
https://doi.org/10.1387/ijdb.160108jb

**Brachyury, Tbx2/3 and sall expression during embryogenesis of the indirectly developing polychaete Hydroides elegans**
Cesar Arenas-Mena
Int. J. Dev. Biol. (2013) 57: 73-83
https://doi.org/10.1387/ijdb.120056ca

**Genomic control of patterning**
Isabelle S. Peter and Eric H. Davidson
Int. J. Dev. Biol. (2009) 53: 707-716
https://doi.org/10.1387/ijdb.072495ip

**Developmental gene network analysis**
Roger Revilla-i-Domingo and Eric H Davidson
Int. J. Dev. Biol. (2003) 47: 695-703
http://www.intjdevbiol.com/web/paper/14756345


**5 yr ISI Impact Factor (2013) = 2.879**