Development 134, 4347-4355 (2007) doi:10.1242/dev.008409

collier transcription in a single *Drosophila* muscle lineage: the combinatorial control of muscle identity

Laurence Dubois, Jonathan Enriquez, Virginie Daburon, Fabien Crozet, Gaelle Lebreton, Michèle Crozatier and Alain Vincent*

Specification of muscle identity in *Drosophila* is a multistep process: early positional information defines competence groups termed promuscular clusters, from which muscle progenitors are selected, followed by asymmetric division of progenitors into muscle founder cells (FCs). Each FC seeds the formation of an individual muscle with morphological and functional properties that have been proposed to reflect the combination of transcription factors expressed by its founder. However, it is still unclear how early patterning and muscle-specific differentiation are linked. We addressed this question, using Collier (Col; also known as Knot) expression as both a determinant and read-out of DA3 muscle identity. Characterization of the *col* upstream region driving DA3 muscle specific expression revealed the existence of three separate phases of cis-regulation, correlating with conserved binding sites for different mesodermal transcription factors. Examination of *col* transcription in *col* and *nautilus* (*nau*) loss-of-function and gain-of-function conditions showed that both factors are required for *col* activation in the 'naïve' myoblasts that fuse with the DA3 FC, thereby ensuring that all DA3 myofibre nuclei express the same identity programme. Together, these results indicate that separate sets of cis-regulatory elements control the expression of identity factors in muscle progenitors and myofibre nuclei and directly support the concept of combinatorial control of muscle identity.

KEY WORDS: Cis-regulatory modules, *collier* (*knot*), *nautilus* (*MyoD*), *Drosophila*, Muscle identity

INTRODUCTION

Drosophila Collier (Col; also known as Knot) belongs to the COE (Col/Olf1/EBF) family of transcription factors, which contains a single member in metazoans, except for vertebrates, in which four genes have been identified (Dubois and Vincent, 2001; Liberg et al., 2002; Pang et al., 2004). *col* was initially characterised for its expression and function in a specific region of the embryonic head that corresponds to both a mitotic domain (MD2) and a gnathal parasegment (PS0) (Crozatier et al., 1996). *col* is also expressed in, and required for the formation of, a single somatic muscle, the embryonic Dorsal/Acute 3 (DA3) muscle (Crozatier and Vincent, 1999), thereby providing a unique entry site for studying the transcriptional control of muscle identity.

The embryonic musculature of *Drosophila melanogaster* is highly stereotyped, with a standard arrangement of around 30 somatic muscles in each trunk hemisegment. Each muscle fibre is an individual syncitium that can be distinguished by its position, shape, epidermal attachment sites and innervation (Bate, 1993; Baylies et al., 1998). Muscle fibres are seeded by founder cells (FCs), which are themselves generated from progenitor cells singled out from promuscular clusters by Notch-mediated lateral inhibition (Carmena et al., 1995; Ruiz Gomez and Bate, 1997). FCs undergo multiple rounds of fusion with fusion competent myoblasts (FCMs) to form a myofibre. The current view is that 'muscle identity' transcription factors (TFs) endow FCs with the capacity to execute the fusion and differentiation programme specific to each muscle fibre (Baylies and Michelson, 2001; Frasch and Leptin, 2000). The 'identity TF code', at least in part, reflects the initial position of the promuscular cluster and derived progenitor cell. Pioneering work on the control of expression of the homeodomain transcription factor Even-Skipped (Eve) in dorsal muscle progenitors showed that it involved the combinatorial activity of TFs functioning downstream of Wingless (Wg), Decapentaplegic (Dpp) and receptor tyrosine kinase (RTK) signalling, [dTCF (Pan – FlyBase), Mothers against Dpp (Mad) and Pointed (Pnt), respectively]. Integration of this positional information and tissue-specific (mesodermal) information at the level of the *eve* promoter was responsible for activating Eve-expression in promuscular clusters (equivalence groups) from which Eve progenitors were selected by Notch (N) signalling (Carmena et al., 2002; Carmena et al., 1998; Halfon et al., 2000; Halfon et al., 2002). Large-scale analyses of gene expression in conditions of perturbation of components of Eve regulation suggested that related transcriptional codes could be responsible for different patterns of progenitor gene expression (Estrada et al., 2006; Philippakis et al., 2006; Sandmann et al., 2006). The *eve* enhancer reproducing Eve expression in muscle progenitors was not active, however, in recruited FCM nuclei (Halfon et al., 2000), indicating that different cis-regulatory elements (and TFs) could be required for specifying promuscular clusters and maintaining a TF identity code.

Here we used Col expression as both a determinant and read-out of DA3 muscle identity to ask how positional information that defines promuscular clusters is relayed into the FC identity and extended to fused FCM nuclei. We first identified the cis-regulatory regions controlling *col* transcription at several steps during formation of the DA3 muscle and defined a DA3-muscle-specific cis-regulatory module (CRM). Detailed analysis of this CRM revealed the existence of three separate steps: Col activation in promuscular clusters, upregulation in the selected progenitor and DA3 FC and activation in the nuclei of FCM incorporated in the growing DA3 myofibre during the muscle fusion process. Comparison of the DA3 muscle CRM between several *Drosophila* species identified a set of conserved sequence motifs with functional significance supported by the expression patterns of reporter genes containing the *D. virilis* (*D. vir*)

Centre de Biologie du Développement, UMR 5547 CNRS/UPS, IFR 109, Institut d'Exploration Fonctionnelle des Génomes, 118 route de Narbonne, 31062 Toulouse cedex 9, France.

*Author for correspondence (e-mail: vincent@cict.fr)

Accepted 14 September 2007

DNA. Conserved binding sites for the mesodermal TFs Twist (Twi), Nautilus (Nau, the *Drosophila* orthologue of MyoD) and Mef2 (Andres et al., 1995; Huang et al., 1996; Ip et al., 1992; Kophengnavong et al., 2000) and a putative Col-binding site necessary for positive autoregulation were present in different subdomains of the DA3 muscle CRM, correlating with the separate phases of *col* regulation. We show that *col* auto-regulation is crucial for a reiterative, two-step activation of *col* transcription in each 'naïve' FCM incorporated into the DA3 muscle. Nau, which was previously reported to be required for DA3 muscle formation (Keller et al., 1998), is also required for *col* transcription in the DA3 muscle, beyond the FC stage. Pan-FC expression of either Col, Nau or both proteins resulted in ectopic *col* transcription in different sets of muscles. Together, our results show that separate sets of cis-regulatory elements ensure *col* activation in the DA3/D05 promuscular cluster, progenitor and DA3 myofibre. Nau and Col act together in ensuring that all nuclei within the DA3 myofibre activate Col and express the same differentiation programme, thereby directly supporting the concept of combinatorial control of muscle identity.

MATERIALS AND METHODS

Drosophila strains

The following strains were used: w^{1118} as a wild-type (wt) reference and for P element transformation using standard procedures (Rubin and Spradling, 1982); rp298-Gal4 (Menon and Chia, 2001); col^{1} (Crozatier et al., 1999) and nau^{188} (Balagopalan et al., 2001) EMS-induced loss-of-function alleles; $vg^{83b27\text{-}R}$, a γ-ray induced amorphic allele; UAS-*col* (Vervoort et al., 1999); *hs-col* (Crozatier and Vincent, 1999); UAS-*nau* (Keller et al., 1997); UAS-*lacZ* (Bloomington Stock Center, Indiana, USA). UAS-*mcd8::GFP* (Grueber et al., 2003).

Plasmid constructions and transgenic lines

The *P5cl* construct was described in Crozatier and Vincent (Crozatier and Vincent, 1999). Other *Pcl* constructs were generated by cloning different fragments of *col* upstream DNA (for the restriction sites used, see Fig. S2 in the supplementary material) into pCaSpeRβ-gal or pPTGal4 (Sharma et al., 2002). Mutagenesis of the putative Nau- and Col-binding sites in *P2.6cl* was done by PCR. The *D. vir* constructs were generated by restriction digestion of genomic DNA isolated from a λ phage library (J. Tamkun, University of California, Santa Cruz, CA).

Immunohistochemical staining and in situ hybridisation

Embryos were fixed and processed for antibody staining and/or in situ hybridisation as described (Crozatier et al., 1996). The *nau* intronic probe covers all three *nau* introns and the two corresponding exons. The following primary antibodies were used: rabbit anti-Col (1/400); mouse anti-Col (1/100); rabbit anti-MHC (1/500; D. Kiehart, Duke University, Durham, NC); mouse anti-β-gal (1/1000, Promega); rabbit anti-GFP (1/1000 Torrey Pines Biolabs); Secondary antibodies were Alexa Fluor 488 and Alexa Fluor 647 conjugated goat anti-rabbit, Alexa Fluor 647 conjugated goat anti-mouse (Molecular Probes 1/300); Rhodamin RedX conjugated donkey anti-mouse (Jackson Laboratory 1/300); biotinylated goat anti-mouse (Vector Laboratory, 1/1000). For double fluorescent in situ hybridisation/immunostaining, we used biotinylated *col* and digoxygenin-labelled *nau* intronic probes and the ABC kit from Vector Laboratory, followed by fluorescent tyramide staining (Alexa fluor 555 or 488 conjugated tyramide from Molecular Probes) and Fast Red. Primary antibodies against Col, GFP or MHC were used at five times the usual concentration. Monoclonal Col antibodies were generated in collaboration with Jeannine Boyes and Georges Delsol, U 563 INSERM, Toulouse Purpan.

Sequence alignments and transcription factor binding sites

Pairwise sequence alignments of *col* upstream sequences from various *Drosophila* species (http://flybase.bio.indiana.edu/static_pages/news/articles/2007_03/genomes_papers3.html) were done using NCBI-BLAST (bl2seq), Genome Browser (UC Santa Cruz) and Evoprinter (NINDS, NIH, Bethesda) and manually edited following eye-inspection. Search for individual binding sites for transcription factors made use of Genomatix Matinspector, Possum (http://zlab.bu.edu/~mfrith/possum/), cis-analyst (http://rana.lbl.gov/cis-analyst/cgi/viewer.php) and FlyEnhancer (http://genomeenhancer.org/fly; M. Markstein) and manual inspection based on the literature. Access to the Mef2 and Twi in vivo binding sites (Sandmann et al., 2007; Sandmann et al., 2006) was via the E. Furlong's lab site (http://furlonglab.embl.de/data/).

RESULTS

Modular organisation of the *col* cis-regulatory region

col belongs to the class of *Drosophila* regulatory genes with numerous introns, large amounts of flanking sequence and multiple expression sites (Crozatier and Vincent, 1999; Nelson et al., 2004; Philippakis et al., 2006). During embryogenesis, *col* is expressed in the MD2/PS0 head region, the somatic DA3 muscle, precursor cells of the lymph gland, a small set of multidendritic (md) neurons of the peripheral nervous system and specific neurons of the central nervous system (CNS) (Baumgardt et al., 2007; Crozatier et al., 2004; Crozatier et al., 1999; Crozatier and Vincent, 1999; Orgogozo and Schweisguth, 2004). We previously generated a *lacZ* reporter transgene (*P{5col::lacZ}*, abbreviated *P5cl*, Fig. 1A) containing 5 kb of *col* upstream DNA, which faithfully reproduced *col* transcription both in the MD2/PS0 and the DA3 muscle, starting at the progenitor stage and not in promuscular cluster(s) (Crozatier and Vincent, 1999). To identify the missing cis-regulatory information, we tested a longer construct containing the entire 9 kb region separating *col* from *CG10200,* the next predicted upstream gene (http://flybase.bio.indiana.edu/; *P9cl*, Fig. 1A). In addition to the head and DA3 muscle, *P9cl* expression reproduced *col* expression in md neurons and a subset of neurons in the CNS. A DNA fragment located further upstream, between CG10200 and the next predicted gene CG10202, was independently shown to drive *col* expression in the anteroposterior organiser of the wing imaginal disc (Hersh and Carroll, 2005). However, neither this construct nor P9cl reproduced Col expression in promuscular clusters (Fig. 1D). The *col* transcription unit is immediately flanked at its 3′ end by another gene, *BEAF32* (Fig. 1A), making rather unlikely the presence of cis-regulatory elements within this region. However, it contains ten different introns, of total length around 30 kb, the cis-regulatory content of which remains to be assessed (see Discussion).

To delineate more precisely the CRM driving *col* expression in the DA3 muscle, we tested a series of constructs containing 2.6, 2.3, 1.6 and 0.9 kb of DNA upstream of the *col* transcription start site, respectively (Fig. 1A). *P2.6cl* retained the information necessary for *col* expression in MD2/PS0 and the DA3 progenitor and muscle (Fig. 1C), although we noted that *P2.6cl* expression in muscle progenitors was less robust than *P9cl*. *P2.3cl* was also activated in MD2/PS0 at stage 6 and the DA3 muscle. However, unlike *P9cl* or *P2.6cl*, *P2.3cl* was not activated in the DA3/DO5 progenitor but only at the FC stage (Fig. 1C; ectopic *lacZ* expression was observed in clusters of neuroectodermal cells at embryonic stage 11). This difference indicated that cis-regulatory elements required for *col* expression in the DA3/DO5 progenitor reside between positions –2.6 and –2.3 and act separately from those required for expression in the DA3 FC and muscle. *P1.6cl* was only active in MD2/PS0, whereas no expression at all could be detected with *P0.9cl* (data not shown). Together, expression data from this series of reporter constructs allowed the mapping of the CRM required for *col*-specific expression in the DA3/DO5 muscle progenitor and DA3 FC/myofibre to a DNA fragment between positions –2.6 and –1.6 upstream of the *col* transcription start (Fig. 1E).

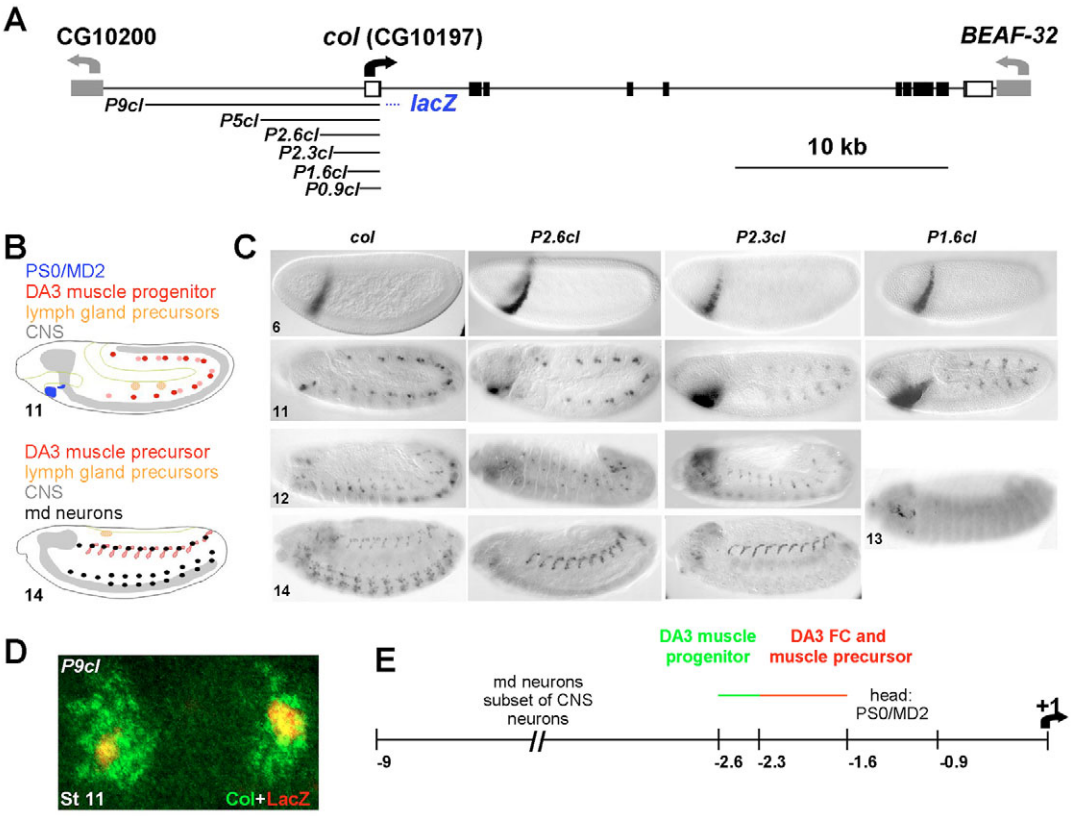

Fig. 1. Mapping of the *col* DA3 muscle CRM in *Drosophila*. (**A**). Schematic representation of the *col* genomic region. Coding exons and the 5′ and 3′ untranslated regions are indicated by black and white boxes, respectively. The positions of the immediately upstream and downstream predicted genes (http://flybase.bio.indiana.edu/), *CG10200* and *BEAF-32*, are indicated by grey boxes and their direction of transcription by arrows. The extent of *col* upstream region present in each *lacZ* reporter gene, *P9cl* to *P0.9cl* is indicated by a black line. (**B**) Diagrammatic, colour-coded representation of the different *col* expression sites in stage 11 and 14 embryos. (**C**) In situ hybridisation showing expression of *P2.6cl*, *P2.3cl* and *P1.6cl*, compared to endogenous *col*, at embryonic stages 6, 11, 12 and 14. (**D**) Close-up view of the DA3 promuscular cluster and progenitor in the T2 and T3 segments of a *P9cl* embryo at stage 11, stained for Col (green) and β-gal (red). Unlike Col, *lacZ* expression is restricted to the progenitor cell. (**E**) Schematic representation of the modular organisation of the *col* cis-regulatory region, underlining the position of the DA3 muscle CRM.

Conserved motifs and TF-binding sites in the *col* upstream region

We took advantage of the recently available genome sequences of several *Drosophila* species to search for conserved motifs in the *col* upstream DNA, as it has often proven to be effective to identify functionally important cis-regulatory elements (Wasserman et al., 2000; Yuh et al., 2002). Among these species, *D. virilis* (*D. vir*) is the most distant from *D. melanogaster* (*D. mel*) (Tamura et al., 2004). We first verified that Col expression in *D. vir* was similar to that in *D. mel* embryos (Fig. 2A and see Fig. S1 in the supplementary material) and could infer from this that the regulatory information controlling *col* transcription in the DA3 muscle lineage has been conserved. Sequence comparison of 9 kb of the *col* upstream region between *D. mel*, *D. vir* and four other *Drosophila* species, *D. yakuba*, *D. ananassae*, *D. pseudoobscura* and *D. mojavensis* revealed numerous stretches of high sequence conservation, of sizes up to 100 bp (see Fig. S1 in the supplementary material). Ten conserved motifs of size >20 bp, numbered 1 to 10 from 5′ to 3′, were found in the same order and at the same relative position between position –2.6 and the start of transcription in all six *Drosophila* species (Fig. 2B and see Fig. S2 in the supplementary material). To test the relevance of this conservation, we generated *lacZ* reporter constructs containing either *D. vir* or *D. mel* DNA (Fig. 2B). *P.3.4cl*^{vir} corresponds to *D. mel P2.6cl*, whereas *P3.4-1.3cl*^{vir} and *P2.6-0.9cl* are truncated versions covering motifs 1 to 10. All four reporter genes showed expression in the DA3 muscle, starting at the progenitor stage, confirming the evolutionary conservation of a DA3-muscle-specific CRM (Fig. 2A). A Gal4 driver line containing only the –2.6 to –1.6 region (*P2.6-1.6cG*), harbouring only motifs 1 to 7 (Fig. 2B), was also specifically expressed in the DA3 muscle (Fig. 2A). This confirmed that the DA3 muscle CRM is located between positions –2.6 and – 1.6. We noticed, however, that expression of *P2.6-1.6cG* was weaker and more sporadic than *P2.6-0.9cl*, suggesting the existence of cis-regulatory element(s) between positions –1.6 and –0.9 contributing to robust DA3 muscle expression. We then searched within the conserved motifs 1 to 10 for consensus binding sites of known TFs that could account for *col* activation in the DA3 muscle. This identified a binding site for the mesodermal basic helix-loop-helix (bHLH) protein Twi (Ip et al., 1992; Kophengnavong et al., 2000) (within motif 2 Fig. 2B), correlating well with the position of the muscle progenitor cis-element (Fig. 1E) and a potential EBF/Col-binding site (Travis et al., 1993) within motif 7. Further visual inspection of the sequence alignments identified other conserved TF-binding sites, including one Mef2-binding site (Andres et al., 1995) within the –1.6 to –0.9 fragment and one consensus binding site for Nau (Huang et al., 1996; Kophengnavong et al., 2000). On the one hand, the position of the Mef2 site correlated well with the

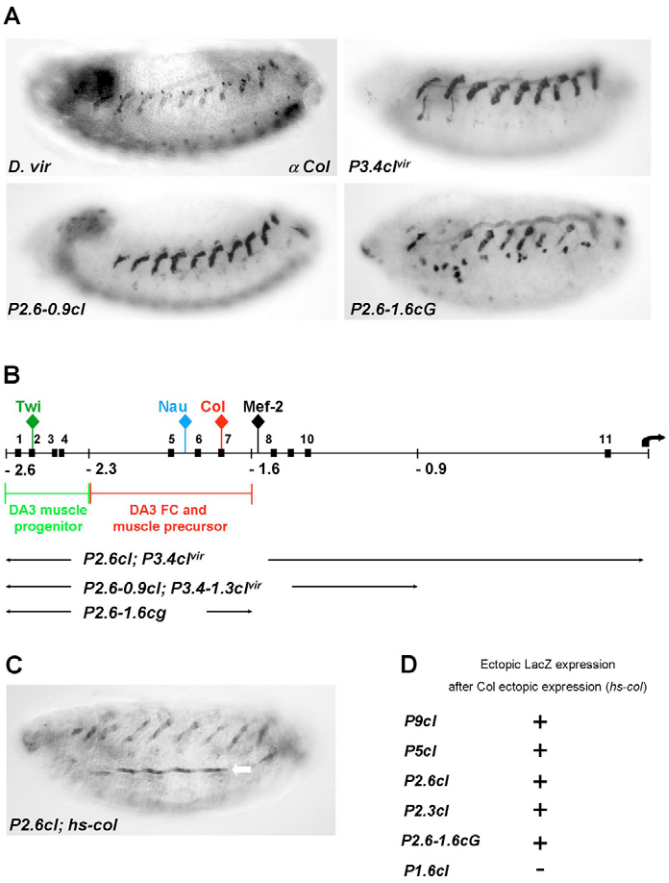

Fig. 2. Conserved cis-regulatory elements and TF-binding sites in the DA3 muscle CRM. (**A**) Col expression in a stage 14 *D. vir* embryo (top left) and in situ hybridisation to *lacZ* transcripts showing expression of different *D. vir* and *D. mel col* reporter genes, as indicated in each panel. Note that *P2.6-1.6cG* is a *Gal4/UAS-lacZ* line. (**B**) Diagrammatic representation of the relative positions of conserved motifs, numbered from 1 to 10 and potential binding sites for Twi, Nau, Col and Mef2 in the DA3 muscle CRM (for details, see Fig. S2 in the supplementary material). (**C**,**D**) Ubiquitous *hs-col* driven Col expression specifically activates *col-lacZ reporter* genes in the VL1 muscle (white arrow), as shown here for *P2.6cl* (C). This is mediated by conserved cis-regulatory elements in the DA3 muscle CRM (D).

requirement of the –1.6 to –0.9 fragment for robust DA3 muscle expression (Fig. 2A). On the other hand, the presence of a Nau-binding site was particularly intriguing as Nau is required for DA3 muscle formation (Keller et al., 1998). Potential binding sites for other TFs (Bergman et al., 2005; Vlieghe et al., 2006) could be found in the DA3 CRM, but we limited here our annotation to the conserved sites (see Fig. S2 in the supplementary material). The relative paucity of known TF-binding sites in the conserved sequence motifs found in the DA3 muscle CRM leaves largely open the question of the roles of these motifs in *col* regulation.

Ectopic activation of *col* transcription reveals a muscle TF code

Heat-shock-driven, ubiquitous expression of the Col protein activated endogenous *col* transcription in a few muscles other than DA3, mainly the VL1 and, more sporadically, the DA2 muscle (Crozatier and Vincent, 1999). By using different *col-lacZ* reporter genes, we mapped the cis-regulatory element(s) responsible for this muscle-specific activation to the DA3 muscle CRM (Fig. 2B-D and data not shown). As it is restricted to the DA3 and VL1 (and possibly DA2) muscles, we reasoned that *col* auto-activation was dependent upon a specific combination of TF expressed in these muscles. Of the known TFs expressed in somatic muscles, only Vestigial (Vg) and Nau are expressed in DA3 and VL1 (Bate et al., 1993; Dohrmann et al., 1990; Keller et al., 1998; Paterson et al., 1991). *nau* mutant embryos lack a subset of muscle fibres, with DA3 being the most severely affected (Balagopalan et al., 2001; Keller et al., 1998). By contrast, no muscle phenotype has yet been described for *vg* loss-of-function mutations. *vg* mutants show reduced wings and severe flight muscle defects but are viable and fertile, allowing the study of their maternal plus zygotic phenotype. We did not observe abnormal Col expression or abnormal DA3 muscle formation in *vg* mutant embryos, indicating that Vg is not required for formation of this muscle (data not shown).

col activation in nuclei of fused myoblasts: a reiterative process endowing all nuclei of the DA3 myofibre with the same transcriptional programme

In situ hybridisation with a *col* intronic probe that labels nascent transcripts revealed that *col* transcription is activated in the nuclei of those FCMs that are recruited to form the DA3 muscle (Crozatier and Vincent, 1999). To further investigate the mechanisms behind this observation, we compared the patterns of Col accumulation and *col* transcription during the process of DA3 muscle formation (Fig. 3A-C). We found that, throughout the FC/FCM fusion phase (stage 12-15), each DA3 muscle syncitium contains on average one or two nuclei, which stain positive for Col but do not transcribe *col* (see also Fig. 4). Close-up analysis of fusion events in stage 13 embryos further revealed that only nuclei containing high levels of Col protein activated *col* transcription (Fig. 3A). This strongly suggested that Col accumulation is a prerequisite for auto-activation in newly fused FCM nuclei. In support of this interpretation, all the DA3 muscle nuclei transcribe *col* after completion of the muscle fusion process (Fig. 3B), although this uniform expression phase is only transient, as *col* transcription declines abruptly during stage 16 to become undetectable (Fig. 3C). From these observations, we conclude that activation of *col* transcription occurs through a reiterative two-step mechanism, ensuring the same transcriptional programme to all nuclei of the DA3 myofibre. In a first step, nuclei from FCMs newly incorporated into the growing syncitium import some of the Col protein present in the muscle precursor (inset in Fig. 3A). In a second step, *col* transcription is turned on in these nuclei.

Col and Nau are required for *col* transcription during DA3 muscle fusion

First evidence for *col* autoregulation during DA3 muscle formation came from the observation that *col* transcription is not maintained in the DA3 FC in *col* mutant embryos (Crozatier and Vincent, 1999). In order to investigate this phenotype in more detail, we constructed a *P9col-Gal4* driver (*P9cG*), allowing us to express a membrane-bound form of GFP in the DA3 muscle progenitor and to specifically follow the fate of this progenitor in *col* mutant embryos (*UAS-mcd8GFP/P9cg*; Fig. 3D). In wt embryos, mCD8GFP remains expressed and is detected both intracellularly and at the plasma membrane of the DA3 myofibre. In *col* mutant embryos, mCD8GFP expression is lost early but stability of the protein at the plasma membrane allows the detection of the mutant DA3 fibres. This experimental set-up confirmed that fusion of FCMs with the DA3 FC is drastically impaired in *col* mutant embryos and that *col*

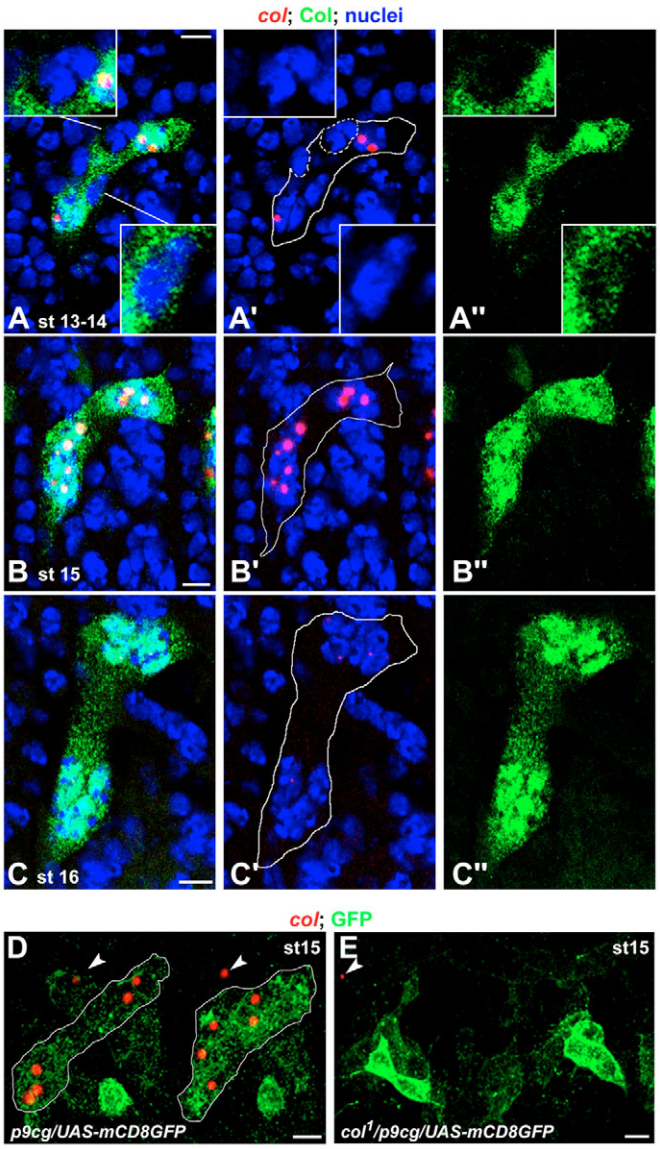

Fig. 3. *col* activation in FCM nuclei incorporated in the DA3 myofibre in *Drosophila*. (**A-C**) *col* transcription in wt DA3 muscle precursors, visualised by in situ hybridisation to *col* primary transcripts (red dots), immunostaining for Col (green) and nuclear staining (blue). (**A′-C′**) Blue and red channels; (**A″-C″**) green channel. (A) Stage 14 embryo. The DA3 muscle precursor contains several nuclei; the two distalmost have already accumulated a high level of Col protein and activated *col* transcription. One central nucleus starts accumulating Col protein (lower inset) but does not yet transcribe *col*. Two other FCM have probably fused but not started to import Col protein (surrounded by a dashed line in A′, upper inset). Another FCM has started engaging in the fusion process, (dashed notch in A′). (B) Stage 15 embryo. At this stage, each DA3 muscle nucleus contains high levels of Col protein and transcribes *col*. (C) Stage 16 embryo. All the DA3 muscle nuclei still contain high levels of Col protein but *col* transcription has almost completely ceased. (**D**,**E**) In situ hybridisation to *col* primary transcripts (red dots) in (D) wt and (E) col^{1} mutant embryos (two segments are shown). A membrane-targeted form of GFP expressed under control of the *col* promoter (*P9cg* construct) allows the visualisation of the DA3 muscle (green). Note the complete absence of *col* transcription in *col* mutant embryos (E). The white arrowhead points to a dorsal md neuron expressing Col. Scale bar: 5 μm.

transcription is neither maintained in the DA3 FC nor activated in the nuclei of FCM, which sometimes fuse to form an abortive DA3 muscle precursor (Fig. 3E). These data establish that *col* auto-regulation and the muscle DA3 identity programme are intimately connected.

Nau activity is also required for formation of the DA3 muscle, although the described DA3 *nau* mutant phenotype is not as severe as for *col* (Keller et al., 1998). The presence of a consensus Nau-binding site in the *col* DA3 muscle CRM raised the possibility that one Nau function could be to regulate *col* transcription. To address this possibility, we first compared *nau* and *col* transcription in wt DA3 muscles, using in situ hybridisation to primary transcripts and Col immunostaining. This revealed that *col* and *nau* are transcribed together in the DA3 progenitor, FC and muscle precursor up to early stage 13 (Fig. 4A-C). Subsequently, only *col* transcription persists in the DA3 muscle (Fig. 4D). We then looked at *col* transcription in embryos homozygous for the null allele nau^{188} (Balagopalan et al., 2001; Wei et al., 2007). Based on Col and Myosin heavy chain (MHC) antibody staining (Fig. 4E,F and data not shown) the DA3 muscle was completely absent in around 5% of segments, abnormal in orientation in 45% and rather normal-looking in about 50% of segments, consistent with previous reports (Keller et al., 1998). Low amounts of Col protein were observed in nuclei of the 'normal-looking' DA3 muscles (Fig. 4E,F), allowing us to look at *col* transcription in these nuclei. In wt embryos at stage 15, each DA3 muscle syncitium contains on average nine nuclei, which are all strongly stained with Col antibodies, and seven to eight are positive for *col* transcription (Fig. 4E; see also Fig. 3B). The DA3 fibres present in nau^{188} embryos contained only seven nuclei on average, with most showing a low level of Col protein. However, at most two or three of those transcribed *col* (Fig. 4F). This result indicated that Nau activity is required, in addition to Col, for activation of *col* transcription in the FCM nuclei that are recruited by the DA3 FC. The Col protein that is detected in *nau* mutant embryos probably derives from earlier, Nau-independent *col* transcription. Supporting this conclusion, one nucleus, probably the FC nucleus, shows high levels of *col* transcripts in *nau* mutant embryos at late stage 12, when DA3 muscle precursors contain two or three nuclei (Fig. 4G,H). In summary, *nau* and *col* are expressed in the DA3 FC and both Nau and Col are required for *col* activation in the nuclei of newly recruited FCM, thereby ensuring that all nuclei within the DA3 myofibre acquire the same identity.

The combinatorial activity of Nau and Col controls *col* expression

To further test the hypothesis of a combinatorial role of Nau and Col in conferring the DA3 muscle its identity, we examined the activation pattern of *P2.6cl* at stage 15 after either Nau alone, Col alone or Nau+Col were ectopically expressed in all muscle FCs (*rp298Gal4* driver) (Menon and Chia, 2001). *rp298Gal4*-driven Col expression resulted in ectopic *P2.6cl* expression in several muscles other than DA3, including DA2, DT1 and VL2, although this expression was most robust in VL1, as seen in *hs-col* experiments (Fig. 5A,B), without major phenotypic effects, at least at the level of muscle fibre morphology (data not shown). By contrast, *rp298Gal4*-driven Nau expression, while altering the pattern of muscle fibres, as previously documented with a heat-shock construct (Keller et al.,

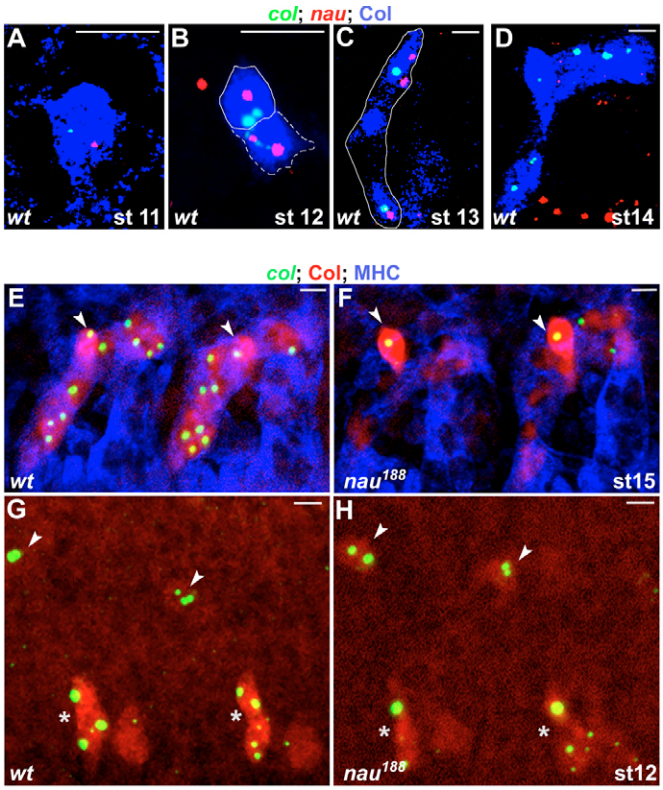

Fig. 4. *Nau*-dependent *col* transcription during the DA3 muscle fusion process in *Drosophila*. (**A-D**) Double in situ hybridisation with intronic probes for *col* (green) and *nau* (red) nascent transcripts and Col immunostaining (blue) show that *nau* and *col* are co-expressed in (A) the DA3/DO5 progenitor cell, (B) the DA3 FC (outlined by a plain line) and (C), the DA3 muscle precursor when it contains two to three nuclei (outlined). *nau* remains transcribed in the DO5 FC (dashed outline in B), whereas *col* transcription is rapidly turned down. (**E-H**) *col* transcription (green dots) in (E,G) wt and (F,H) nau^{188} mutant embryos (two segments are shown); the DA3 muscle is visualised by immunostaining for Col (red) and MHC (blue in E,F). In stage 15 nau^{188} mutant embryos (F), the DA3 muscle is reduced, compared to wt (E) and most nuclei do not transcribe *col*. At stage 12, *col* expression in the DA3 muscle precursor (asterisk) when it contains two to three nuclei is similar in (H) nau^{188} and (G) wt embryos, although only one nucleus, probably the FC nucleus, expresses high levels of *col* transcripts in nau^{188} embryos. Arrowheads indicate *col* transcription in a dorsal multidendritic neuron. Scale bars: 5 μm.

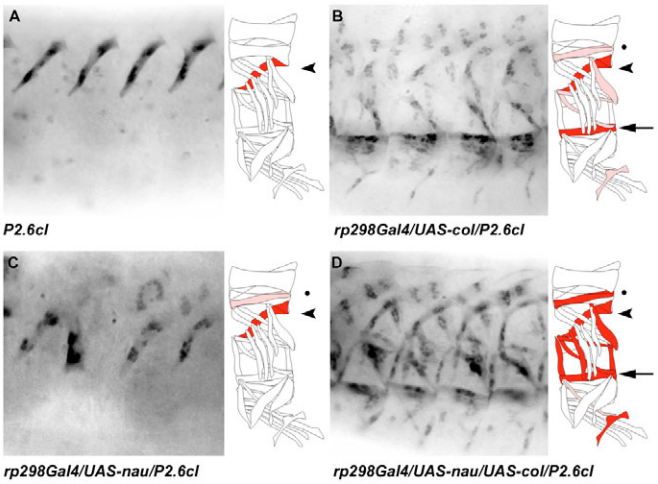

Fig. 5. Nau and Col separately and synergistically activate ectopic *col* transcription in specific subsets of muscles. (**A**) *P2.6cl* expression in the DA3 muscle in stage 15 wt embryos, visualised by β-gal antibody staining. (**B**) *rp298Gal4*-driven Col expression of in all FCs activates *P2.6cl* in a subset of somatic muscles cells, activation being most robust in the VL1 muscle. Nau expression (**C**) is unable to activate ectopic *P2.6cl* expression, except for, sporadically, the DA2 muscle. (**D**) Together, Col and Nau activate *P2.6cl* expression in a large number of somatic muscles in addition to VL1. A schematic representation of the abdominal muscle pattern is shown of the right side of each panel to indicate the *P2.6cl* expressing muscles. The DA3, DA2 and VL1 muscles are designated by an arrowhead, a dot and an arrow, respectively.

1997), provoked ectopic expression of *P2.6cl* only in a single muscle, the DA2 muscle (Fig. 5C). These data confirmed that, despite a more general role than Col in somatic myogenesis (Keller et al., 1997; Wei et al., 2007), Nau is generally unable by itself to ectopically activate *col* transcription. When Col and Nau were expressed together, *P2.6cl* was activated in the same muscles as with Col alone, but much more strongly (compare Fig. 5B with D), confirming that Nau potentiates the ability of Col to activate its own transcription. Interestingly, *P2.6cl* was activated by Nau+Col in a few muscles, including the SBM muscles, which did not respond to the presence of Col alone, indicating that Nau and Col may act synergistically. Still, many muscles remained refractory to this combination and did not express *P2.6cl*, suggesting that other competence factors are lacking or that negative regulation exerted by Notch and/or other factors may be dominant in these muscles.

The control of *col* transcription by Nau+Col is probably direct

The evolutionary conservation of a Nau-binding site and a potential EBF-binding site within the DA3 muscle CRM (Fig. 2B and see Fig. S2 in the supplementary material) suggested that regulation of *col* transcription by Nau and Col could be direct. We independently mutated the putative Nau- and EBF-binding sites within the *P2.6cl* construct, giving rise to $P2.6cl^{nau}$ and $P2.6cl^{col}$, respectively (Fig. 6F). $P2.6cl^{nau}$ expression was either lost from the DA3 muscle or much reduced compared with *P2.6cl* (Fig. 6A,C), suggesting that Nau directly regulates *col* transcription. Unlike the case with *P2.6cl*, however, ectopic $P2.6cl^{nau}$ expression was observed, indicating that the mutated E-box in the Nau site could also mediate binding of repressing factor(s) in absence of Nau. Col binds in vitro to the EBF consensus binding site (TTCT/CNNGGGAA) (Travis et al., 1993), consistent with sequence conservation of the COE DNA-binding domain (Dubois and Vincent, 2001) (V.D., unpublished). The closest match to the consensus EBF recognition site found within the DA3 CRM is the sequence ATGTCTGGGGAT, which is part of the conserved motif 7 (Fig. 6F and see Fig. S2 in the supplementary material). Gel-shift assays and immunoprecipitation of DNA-protein complexes formed by co-incubation of Col with synthetic oligonucleotides overlapping this predicted EBF-binding site failed to reveal strong binding in vitro (V.D., unpublished). Nevertheless, DA3-specific expression of $P2.6cl^{col}$ in vivo was almost undetectable when this site was mutated (Fig. 6B,F), suggesting that it mediates *col* auto-regulation. To provide a different test of this in vivo function, we looked at $P2.6cl^{col}$ activation in conditions of ectopic Col expression. Unlike *P2.6cl* (Fig. 6D), $P2.6cl^{col}$ expression was activated very weakly, if at all, in the VL1 (and DA2)

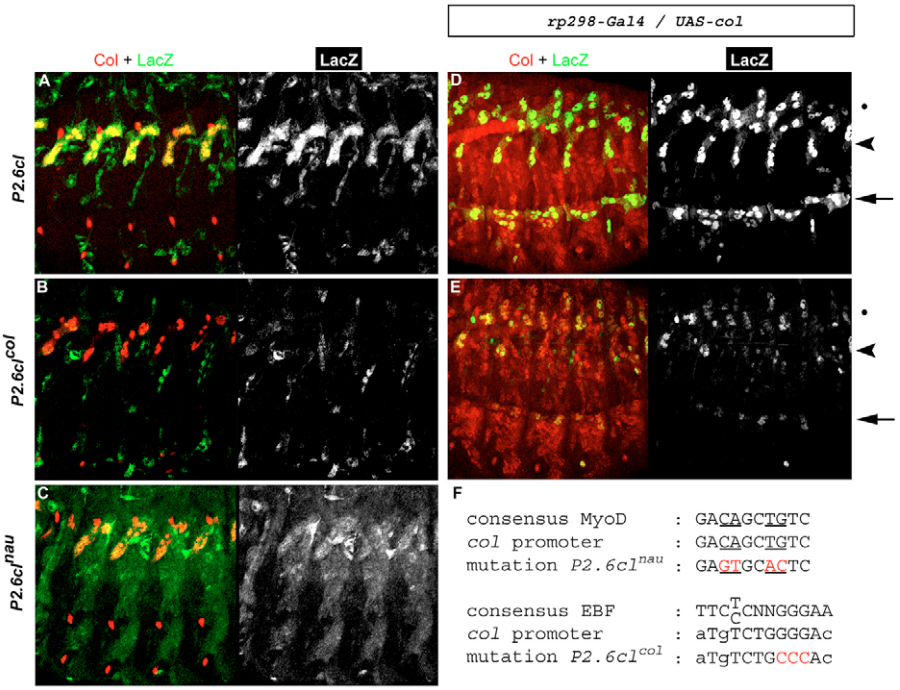

Fig. 6. *col* transcription in the DA3 muscle precursor depends upon a Nau and a potential Col-binding site in the DA3 muscle CRM. (**A**) Col and *P2.6cl* expression in wt *Drosophila* embryos at stage 15, visualised by Col (red) and β-gal (green in the right and white in the left panel) antibody staining. (**B**,**C**) *P2.6cl* expression is lost when either the putative EBF/Col- (B) or Nau- (C) binding site present in the DA3 muscle CRM is mutated. Red dots in A and C correspond to Col expression in md neurons. (**D**) Col expression in all FCs (*rp298-Gal4/ UAS-col*) induces ectopic *P2.6cl* expression in the VL1 (arrow) and DA2 (dot) muscles, as visualised by β-gal immunostaining; the arrowhead points to the DA3 muscle. (**E**) Ectopic expression of *P2.6cl* is not observed when the EBF/Col-binding site is mutated. (**F**) The consensus EBF- and MyoD-(Nau) binding sites (Huang et al., 1996; Travis et al., 1993) are represented above the predicted sites found within the DA3 muscle CRM. The mutated positions introduced in *P2.6cl*^{col} and *P2.6cl*^{nau} are shown in red.

muscles (Fig. 6E). These results reinforce the conclusion that the predicted EBF/Col-binding site present within the conserved motif 7 is required for positive *col* autoregulation.

DISCUSSION

The stereotyped pattern of *Drosophila* body wall muscles relies upon the specification of FCs that seed the formation of individual muscles at specific positions in the somatic mesoderm (Baylies et al., 1998; Rushton et al., 1995). The current view is that the properties specific to each muscle result from the selective expression, in each FC, of distinct combinations of 'muscle identity' TFs. However, experimental evidence for such a combinatorial code has remained sparse. We addressed here this question, using as a paradigm Col expression as both a determinant and read-out of DA3 muscle identity.

Three separate steps in the transcriptional control of muscle identity

Functional dissection of the DA3 muscle CRM present in the *col* upstream region showed that *col* expression in the DA3 FC can be separated from its expression in the DA3/D05 progenitor and the promuscular cluster. It thus revealed the existence of three steps in the transcriptional control of muscle identity (Fig. 7). That *col* expression in the DA3/D05 progenitor could be uncoupled from that in

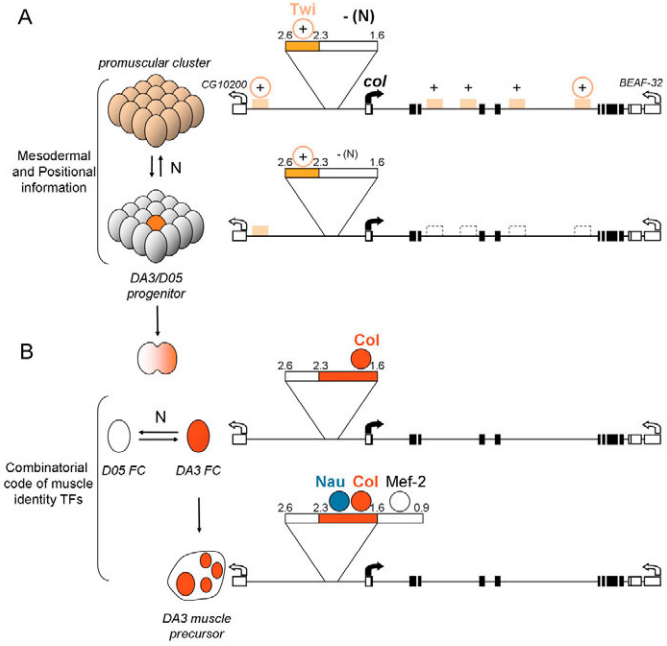

Fig. 7. A model for the combinatorial coding of DA3 muscle identity in *Drosophila*. (**A**) Col is activated in one (T1-T3 segments) and two (A1-A2 segments) promuscular clusters (Crozatier and Vincent, 1999), in response to positional and mesodermal cues. This first step is probably mediated by clusters of relevant TF-binding sites [light orange boxes (Philippakis et al., 2006)], including Twi-binding sites (+) (Sandmann et al., 2007) that are located within the *col* upstream region and introns. Col expression subsequently becomes restricted to the DA3/DO5 progenitor (orange cell) by lateral inhibition (Crozatier and Vincent, 1999). We postulate that positive inputs from TFs binding to the –2.6 to –2.3 fragment, including Twi, are sufficient to allow *P2.6cl* activation in the selected DA3/DO5 progenitor, upon relief of N repression. (**B**) Following division of the progenitor, restriction of Col expression to the DA3 FC involves positive auto-regulation in this FC and negative regulation by N in the sibling DO5 FC. From this stage, a combination of Nau and Col activity is required for *col* transcriptional activation in the FCM nuclei, which are recruited by the DA3 FC to form a myofibre, thereby ensuring that all nuclei in the DA3 muscle express the same identity programme.

promuscular clusters was in apparent contradiction with the previous conclusion from pioneering studies on Eve expression in dorsal muscle progenitors that this expression issued from Eve activation in promuscular clusters. Restriction of Eve expression to progenitors was considered a secondary step, mediated by N-signalling during progenitor selection by lateral inhibition (Carmena et al., 1998; Halfon et al., 2000). To reconcile our data and this model, we propose that the muscle DA3 CRM is only active in the DA3/D05 progenitor because it lacks some positively acting cis-elements necessary to counteract N-mediated repression of *col* transcription (Fig. 7A). We have indeed previously shown that *col* transcription is repressed by N during the progenitor selection process (Crozatier and Vincent, 1999). We also noted that a Twi-binding site is present in the 'progenitor' subdomain of the DA3 CRM (Fig. 2B and Fig. 7A). The functional importance of this site is supported by its in vivo occupancy in 4- to 6-hour-old embryos when selection of the DA3/DO5 progenitor takes place (Sandmann et al., 2007). Together, Twi in vivo binding and the *col/P2.6cl/P2.3cl* expression data suggest that Twi activity contributes to *col* expression in the DA3/DO5 progenitor but may not be sufficient to override N repression of *col* transcription before progenitor selection. Additional binding sites for Twi present in the *col* upstream region, between positions –8.7 and –8.3, are also bound by Twi in vivo (Sandmann et al., 2007) and probably contribute to the robustness of *P9cl* expression in progenitor cells, but the question of which cis-regulatory elements mediate *col* activation in promuscular clusters remains open. From their Eve expression studies, Michelson and colleagues developed a computational framework to identify other FC-specific genes (Estrada et al., 2006; Philippakis et al., 2006). This framework, named Codefinder, integrates transcriptome data and clustering of combinations of binding sites for five different TFs (Pnt, dTCF, Mad, Twi and Tin). *col/kn* was selected by Codefinder owing to the presence of five clusters of binding sites, four of which are located within introns (Philippakis et al., 2006). It remains to be determined which of these could be responsible for *col* activation in promuscular clusters, but it is interesting to note that another in vivo Twi-binding site in 4-6-hour-old embryos correlates with the 3′-most cluster (Sandmann et al., 2007). In addition to Twi, conserved binding sites for Nau and Mef2 are found within the DA3 CRM. The Mef2 binding site is located in a region required for robust DA3-muscle expression of a reporter gene (Fig. 2B, Fig. 7B; and see Fig. S2 in the supplementary material). A direct control of *col* transcription by Mef2 during the muscle fusion process is further supported by the recent finding that Mef2 binds in vivo to the *col* upstream region between 6 and 8 hours of embryonic development (Sandmann et al., 2006).

Propagation of transcriptional identity from the founder cell to fusion-competent myoblasts

Detailed analysis of *col* auto-activation revealed a reiterative, two-step process: import of pre-existing Col protein in the FCM nuclei that incorporate into the growing DA3 myofibre precedes activation of *col* transcription (Fig. 3). This process ensures that all incorporated FCM nuclei acquire the same identity. Nau is required for maintaining *col* transcription in the DA3 muscle precursor and this control is probably direct. The presence of a putative EBF-binding site in the DA3 muscle CRM also correlates with the Col requirement for maintaining its own transcription beyond the FC stage (Crozatier and Vincent, 1999). Thus, despite the failure of our assays to detect strong Col binding to this site in vitro, it appears to be essential for *col* auto-regulation in vivo. This suggests that in vivo binding is potentiated by one or more specific co-factor(s) present in the DA3 muscle. One co-factor is probably Nau, as the ability of Col to activate its own transcription in newly recruited FCM is dependent upon Nau activity (Fig. 7B). Nau is not sufficient, however, as many muscles containing both Nau and Col proteins do not activate *col* transcription (Fig. 5). Interestingly, mouse EBF (also known as Ebf1 and Olf1 – Mouse Genome Informatics) and E2A (Tcfe2a – Mouse Genome Informatics), a bHLH protein of the same subgroup as MyoD (Simionato et al., 2007), have been shown to act on the same target promoter and synergistically upregulate transcription of B-lymphocyte-specific genes, although no direct physical interaction between EBF and E2A could be found in vitro. This suggested that functional interaction of EBF and E2A, similar to Col and Nau, requires yet another factor (O'Riordan and Grosschedl, 1999). Taking into account the restricted pattern of ectopic *col* activation in *hs-col* conditions, we hypothesised that Vg could be another component of the DA3 combinatorial identity (Bate, 1993; Frasch, 1999). However, we found that Vg is not required for DA3 muscle specification, leaving open the question of which factor may bridge Col and Nau functions.

Temporal and combinatorial control of muscle identity

Unlike *col or P2.6cl*, *P2.3cl* is expressed in the DA3 FC and muscle precursor but not the DA3/DO5 progenitor, showing that *col* transcription in the progenitor and muscle precursor is under separate control. These two phases of *col* regulation are intimately linked, however, as Col is required for activating its own transcription in the nuclei of FCM recruited by the DA3 FC. This regulatory cascade may explain how pre-patterning of the somatic mesoderm and muscle identity are transcriptionally linked in the *Drosophila* embryo. As discussed above, the ability of Col to auto-regulate depends upon the presence of Nau, another muscle identity TF. Col and Nau act as obligatory co-factors for maintenance/activation of Col expression in all nuclei of the DA3 muscle, thus bringing to light a clear case of combinatorial coding of muscle identity.

We thank the Bloomington Stock Center, S. Menon and S. Abmayr for fly stocks, D. Kiehart for antibodies, A.M. Michelson for sharing unpublished results, M. Markstein for access to Fly Enhancer version 2, J. Boyes and G. Delsol for help with generating Col antibodies and S. Plaza and D. Cribbs for discussion. We acknowledge the help of the Toulouse RIO Imaging platform and B. Ronsin for confocal microscopy. This work was supported by CNRS, Ministère de la Recherche et de la Technologie, Université Paul Sabatier (grant to G. Delsol, Inserm U563 and A. Vincent) and Association Française contre les Myopathies. J.E. was supported by a fellowship from MRT.

Supplementary material

Supplementary material for this article is available at
http://dev.biologists.org/cgi/content/full/134/24/4347/DC1

References

Andres, V., Cervera, M. and Mahdavi, V. (1995). Determination of the consensus binding site for MEF2 expressed in muscle and brain reveals tissue-specific sequence constraints. *J. Biol. Chem.* **270**, 23246-23249.

Balagopalan, L., Keller, C. A. and Abmayr, S. M. (2001). Loss-of-function mutations reveal that the Drosophila nautilus gene is not essential for embryonic myogenesis or viability. *Dev. Biol.* **231**, 374-382.

Bate, M. (1993). The mesoderm and its derivatives. In *The Development of Drosophila melanogaster*. Vol. 2 (ed. M. Bate and A. Martinez-Arias), pp. 1013-1090. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Bate, M., Rushton, E. and Frasch, M. (1993). A dual requirement for neurogenic genes in Drosophila myogenesis. *Dev. Suppl.* **1993**, 149-161.

Baumgardt, M., Miguel-Aliaga, I., Karlsson, D., Ekman, H. and Thor, S. (2007). Specification of neuronal identities by feedforward combinatorial coding. *PLoS Biol.* **5**, e37.

Baylies, M. K. and Michelson, A. M. (2001). Invertebrate myogenesis: looking back to the future of muscle development. *Curr. Opin. Genet. Dev.* **11**, 431-439.

Baylies, M. K., Bate, M. and Ruiz Gomez, M. (1998). Myogenesis: a view from Drosophila. *Cell* **93**, 921-927.

Bergman, C. M., Carlson, J. W. and Celniker, S. E. (2005). Drosophila DNase I footprint database: a systematic genome annotation of transcription factor binding sites in the fruitfly, Drosophila melanogaster. *Bioinformatics* **21**, 1747-1749.
Carmena, A., Bate, M. and Jimenez, F. (1995). Lethal of scute, a proneural gene, participates in the specification of muscle progenitors during Drosophila embryogenesis. *Genes Dev.* **9**, 2373-2383.
Carmena, A., Gisselbrecht, S., Harrison, J., Jimenez, F. and Michelson, A. M. (1998). Combinatorial signaling codes for the progressive determination of cell fates in the Drosophila embryonic mesoderm. *Genes Dev.* **12**, 3910-3922.
Carmena, A., Buff, E., Halfon, M. S., Gisselbrecht, S., Jimenez, F., Baylies, M. K. and Michelson, A. M. (2002). Reciprocal regulatory interactions between the Notch and Ras signaling pathways in the Drosophila embryonic mesoderm. *Dev. Biol.* **244**, 226-242.
Crozatier, M. and Vincent, A. (1999). Requirement for the Drosophila COE transcription factor Collier in formation of an embryonic muscle: transcriptional response to notch signalling. *Development* **126**, 1495-1504.
Crozatier, M., Valle, D., Dubois, L., Ibnsouda, S. and Vincent, A. (1996). Collier, a novel regulator of Drosophila head development, is expressed in a single mitotic domain. *Curr. Biol.* **6**, 707-718.
Crozatier, M., Valle, D., Dubois, L., Ibnsouda, S. and Vincent, A. (1999). Head versus trunk patterning in the Drosophila embryo; collier requirement for formation of the intercalary segment. *Development* **126**, 4385-4394.
Crozatier, M., Ubeda, J. M., Vincent, A. and Meister, M. (2004). Cellular immune response to parasitization in Drosophila requires the EBF orthologue collier. *PLoS Biol.* **2**, E196.
Dohrmann, C., Azpiazu, N. and Frasch, M. (1990). A new Drosophila homeo box gene is expressed in mesodermal precursor cells of distinct muscles during embryogenesis. *Genes Dev.* **4**, 2098-2111.
Dubois, L. and Vincent, A. (2001). The COE – Collier/Olf1/EBF – transcription factors: structural conservation and diversity of developmental functions. *Mech. Dev.* **108**, 3-12.
Estrada, B., Choe, S. E., Gisselbrecht, S. S., Michaud, S., Raj, L., Busser, B. W., Halfon, M. S., Church, G. M. and Michelson, A. M. (2006). An integrated strategy for analyzing the unique developmental programs of different myoblast subtypes. *PLoS Genet.* **2**, e16.
Frasch, M. (1999). Controls in patterning and diversification of somatic muscles during Drosophila embryogenesis. *Curr. Opin. Genet. Dev.* **9**, 522-529.
Frasch, M. and Leptin, M. (2000). Mergers and acquisitions: unequal partnerships in Drosophila myoblast fusion. *Cell* **102**, 127-129.
Grueber, W. B., Ye, B., Moore, A. W., Jan, L. Y. and Jan, Y. N. (2003). Dendrites of distinct classes of Drosophila sensory neurons show different capacities for homotypic repulsion. *Curr. Biol.* **13**, 618-626.
Halfon, M. S., Carmena, A., Gisselbrecht, S., Sackerson, C. M., Jimenez, F., Baylies, M. K. and Michelson, A. M. (2000). Ras pathway specificity is determined by the integration of multiple signal-activated and tissue-restricted transcription factors. *Cell* **103**, 63-74.
Halfon, M. S., Grad, Y., Church, G. M. and Michelson, A. M. (2002). Computation-based discovery of related transcriptional regulatory modules and motifs using an experimentally validated combinatorial model. *Genome Res.* **12**, 1019-1028.
Hersh, B. M. and Carroll, S. B. (2005). Direct regulation of *knot* gene expression by Ultrabithorax and the evolution of cis-regulatory elements in *Drosophila*. *Development* **132**, 1567-1577.
Huang, J., Blackwell, T. K., Kedes, L. and Weintraub, H. (1996). Differences between MyoD DNA binding and activation site requirements revealed by functional random sequence selection. *Mol. Cell. Biol.* **16**, 3893-3900.
Ip, Y. T., Park, R. E., Kosman, D., Yazdanbakhsh, K. and Levine, M. (1992). dorsal-twist interactions establish snail expression in the presumptive mesoderm of the Drosophila embryo. *Genes Dev.* **6**, 1518-1530.
Keller, C. A., Erickson, M. S. and Abmayr, S. M. (1997). Misexpression of nautilus induces myogenesis in cardioblasts and alters the pattern of somatic muscle fibers. *Dev. Biol.* **181**, 197-212.
Keller, C. A., Grill, M. A. and Abmayr, S. M. (1998). A role for nautilus in the differentiation of muscle precursors. *Dev. Biol.* **202**, 157-171.
Kophengnavong, T., Michnowicz, J. E. and Blackwell, T. K. (2000). Establishment of distinct MyoD, E2A, and twist DNA binding specificities by different basic region-DNA conformations. *Mol. Cell. Biol.* **20**, 261-272.
Liberg, D., Sigvardsson, M. and Akerblad, P. (2002). The EBF/Olf/Collier family of transcription factors: regulators of differentiation in cells originating from all three embryonal germ layers. *Mol. Cell. Biol.* **22**, 8389-8397.
Menon, S. D. and Chia, W. (2001). Drosophila rolling pebbles: a multidomain protein required for myoblast fusion that recruits D-Titin in response to the myoblast attractant Dumbfounded. *Dev. Cell* **1**, 691-703.
Nelson, C. E., Hersh, B. M. and Carroll, S. B. (2004). The regulatory content of intergenic DNA shapes genome architecture. *Genome Biol.* **5**, R25.
Orgogozo, V. and Schweisguth, F. (2004). Evolution of the larval peripheral nervous system in Drosophila species has involved a change in sensory cell lineage. *Dev. Genes Evol.* **214**, 442-452.
O'Riordan, M. and Grosschedl, R. (1999). Coordinate regulation of B cell differentiation by the transcription factors EBF and E2A. *Immunity* **11**, 21-31.
Pang, K., Matus, D. Q. and Martindale, M. Q. (2004). The ancestral role of COE genes may have been in chemoreception: evidence from the development of the sea anemone, Nematostella vectensis (Phylum Cnidaria; Class Anthozoa). *Dev. Genes Evol.* **214**, 134-138.
Paterson, B. M., Walldorf, U., Eldridge, J., Dubendorfer, A., Frasch, M. and Gehring, W. J. (1991). The Drosophila homologue of vertebrate myogenic-determination genes encodes a transiently expressed nuclear protein marking primary myogenic cells. *Proc. Natl. Acad. Sci. USA* **88**, 3782-3786.
Philippakis, A. A., Busser, B. W., Gisselbrecht, S. S., He, F. S., Estrada, B., Michelson, A. M. and Bulyk, M. L. (2006). Expression-guided in silico evaluation of candidate cis regulatory codes for Drosophila muscle founder cells. *PLoS Comput. Biol.* **2**, e53.
Rubin, G. M. and Spradling, A. C. (1982). Genetic transformation of Drosophila with transposable element vectors. *Science* **218**, 348-353.
Ruiz Gomez, M. and Bate, M. (1997). Segregation of myogenic lineages in Drosophila requires numb. *Development* **124**, 4857-4866.
Rushton, E., Drysdale, R., Abmayr, S. M., Michelson, A. M. and Bate, M. (1995). Mutations in a novel gene, myoblast city, provide evidence in support of the founder cell hypothesis for Drosophila muscle development. *Development* **121**, 1979-1988.
Sandmann, T., Jensen, L. J., Jakobsen, J. S., Karzynski, M. M., Eichenlaub, M. P., Bork, P. and Furlong, E. E. (2006). A temporal map of transcription factor activity: mef2 directly regulates target genes at all stages of muscle development. *Dev. Cell* **10**, 797-807.
Sandmann, T., Girardot, C., Brehme, M., Tongprasit, W., Stolc, V. and Furlong, E. E. (2007). A core transcriptional network for early mesoderm development in Drosophila melanogaster. *Genes Dev.* **21**, 436-449.
Sharma, Y., Cheung, U., Larsen, E. W. and Eberl, D. F. (2002). PPTGAL, a convenient Gal4 P-element vector for testing expression of enhancer fragments in drosophila. *Genesis* **34**, 115-118.
Simionato, E., Ledent, V., Richards, G., Thomas-Chollier, M., Kerner, P., Coornaert, D., Degnan, B. M. and Vervoort, M. (2007). Origin and diversification of the basic helix-loop-helix gene family in metazoans: insights from comparative genomics. *BMC Evol. Biol.* **7**, 33.
Tamura, K., Subramanian, S. and Kumar, S. (2004). Temporal patterns of fruit fly (Drosophila) evolution revealed by mutation clocks. *Mol. Biol. Evol.* **21**, 36-44.
Travis, A., Hagman, J., Hwang, L. and Grosschedl, R. (1993). Purification of early-B-cell factor and characterization of its DNA-binding specificity. *Mol. Cell. Biol.* **13**, 3392-3400.
Vervoort, M., Crozatier, M., Valle, D. and Vincent, A. (1999). The COE transcription factor Collier is a mediator of short-range Hedgehog-induced patterning of the Drosophila wing. *Curr. Biol.* **9**, 632-639.
Vlieghe, D., Sandelin, A., De Bleser, P. J., Vleminckx, K., Wasserman, W. W., van Roy, F. and Lenhard, B. (2006). A new generation of JASPAR, the open-access repository for transcription factor binding site profiles. *Nucleic Acids Res.* **34**, D95-D97.
Wasserman, W. W., Palumbo, M., Thompson, W., Fickett, J. W. and Lawrence, C. E. (2000). Human-mouse genome comparisons to locate regulatory sites. *Nat. Genet.* **26**, 225-228.
Wei, Q., Rong, Y. and Paterson, B. M. (2007). Stereotypic founder cell patterning and embryonic muscle formation in Drosophila require nautilus (MyoD) gene function. *Proc. Natl. Acad. Sci. USA* **104**, 5461-5466.
Yuh, C. H., Brown, C. T., Livi, C. B., Rowen, L., Clarke, P. J. and Davidson, E. H. (2002). Patchy interspecific sequence similarities efficiently identify positive cis-regulatory elements in the sea urchin. *Dev. Biol.* **246**, 148-161.