Development Supplement 1, 1991, 47–54
Printed in Great Britain © The Company of Biologists Limited 1991

Dominant gain-of-function mutations that lead to misregulation of the *C. elegans* heterochronic gene *lin-14*, and the evolutionary implications of dominant mutations in pattern-formation genes

GARY RUVKUN, BRUCE WIGHTMAN, THOMAS BÜRGLIN and PREMA ARASU

Department of Molecular Biology, Massachusetts General Hospital, Department of Genetics, Harvard Medical School, Boston MA 02114, USA

Summary

The heterochronic gene *lin-14* controls the temporal sequence of developmental events in the *C. elegans* postembryonic cell lineage. It encodes a nuclear protein that is normally present in most somatic cells of late embryos and L1 larvae but not in later larval stages or adults. Two *lin-14* gain-of-function mutations cause an inappropriately high level of the *lin-14* nuclear protein late in development. These mutations delete 3′ untranslated sequences from the *lin-14* mRNAs and identify a negative regulatory element that controls the formation of the *lin-14* protein temporal gradient. The 21 kb *lin-14* gene contains 13 exons that are differentially spliced to generate two *lin-14* protein products with variable N-terminal regions and a constant C-terminal region. No protein sequence similarity to any proteins in various databases was found.

The temporal and cellular expression patterns of *lin-14* protein accumulation is altered by mutations in the heterochronic genes *lin-4* and *lin-28*. The *lin-4* gene is required to down-regulate *lin-14* protein levels during the mid-L1 stage. The *lin-4* gene product could be the trans-acting factor that binds to the negative regulatory element in the *lin-14* 3′ untranslated region. In contrast, the *lin-28* gene activity positively regulates *lin-14* protein levels during early L1. Thus, these genes act antagonistically to regulate the *lin-14* temporal switch.

The normal down-regulation of *lin-14* within 10 h of hatching is not determined by the passage of time *per se*, but rather is triggered when feeding induces postembryonic development. Loss of *lin-28* gene activity causes precocious down-regulation of *lin-14* protein levels before feeding, whereas loss of *lin-4* gene activity does not affect the level of *lin-14* protein before feeding. These data suggest that to trigger the *lin-14* temporal switch, the *lin-4* gene is up-regulated after feeding which in turn down-regulates *lin-14 via* its 3′ untranslated region.

We speculate on the evolutionary implications of dominant mutations in pattern-formation genes.

Key words: heterochronic genes, *C. elegans*, *lin-14* gene, cell lineage, dominant mutation, evolution.

Pattern-formation genes

During the ontogeny of multicellular animals, diverse cell types are generated from the zygote in an intricate series of cell divisions and differentiations. Over the last 20 years, genes that control this process have been discovered and isolated from *Drosophila* and *Caenorhabditis elegans* (Sternberg and Horvitz, 1984; Ingham, 1988; Scott *et al.* 1989). Molecular analyses of these pattern-formation and cell-specification genes have shown that in many cases these genes encode transcription or RNA processing factors (Costa *et al.* 1988; Finney *et al.* 1988; Scott *et al.* 1989; Bell *et al.* 1988) or genes involved in extracellular or cell–cell signal transduction pathways (Greenwald, 1985; Hafen *et al.* 1987). Genes homologous to these invertebrate pattern-formation genes have been isolated from mammals, and are also thought to control development (Graham *et al.* 1989; Balling *et al.* 1988).

In many cases, these pattern-formation genes are expressed or become active in spatial domains that presage the generation of differentiated cells and structures (Akam, 1987). Mutations in these genes that disrupt their normally asymmetric pattern of expression or activation have been shown to lead to homeotic changes (Ingham, 1988; Schneuwly *et al.* 1987; Costa *et al.* 1988; Ruiz i Altaba and Melton, 1989).

Just as pattern-formation genes define and interpret spatial information during development, in *C. elegans* the temporal sequence of developmental events has been shown to be explicitly controlled by heterochronic genes (Ambros and Horvitz, 1984, 1987; Ruvkun and Giusto, 1989). Heterochronic mutations cause many blast cells to undergo patterns of cell division and

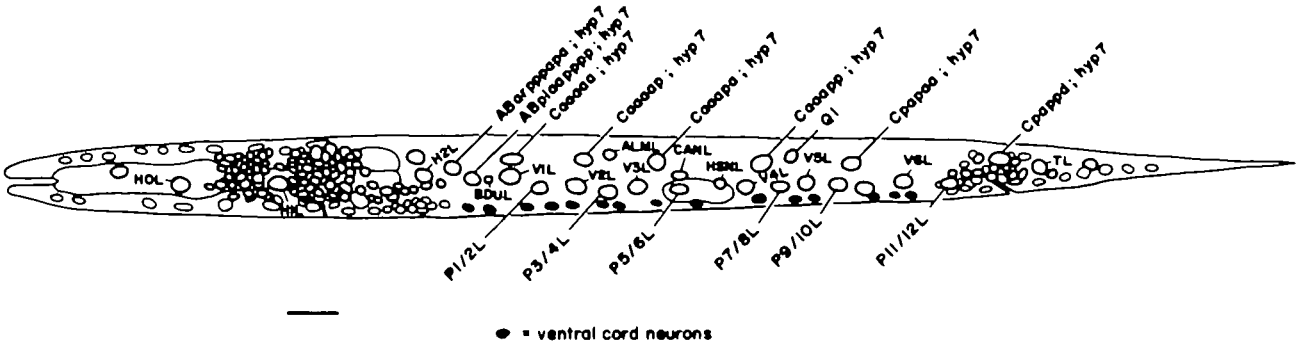

Fig. 1. Arrangement of neuronal and hypodermal nuclei on the left side of a newly hatched L1 (from Sulston *et al.* 1983). Cells that do not divide postembryonically are named by their lineage, for example, Cpaapp. Postembryonic blast cells are renamed H1, H2, etc. Nuclei that were difficult to identify, i.e. the nerve ring, are unnamed in this figure. Pattern on the right side is very similar. Bar, 10 μm.

differentiation normally observed at distinct developmental times, suggesting that these mutations perturb the definition or interpretation of developmental time (Ambros and Horvitz, 1984). Thus, both the time and space dimensions during development are defined by pattern-formation genes.

These discoveries about the molecular nature of pattern-formation genes allow us to ask very direct questions about the molecular mechanisms operating during development. We must explain how these genes or their products come to be activated or expressed non-uniformly, and we must explain how these gene activities that affect groups of cells interact to specify particular cell fates.

Here we present our analysis of how developmental time is specified by the heterochronic gene *lin-14* to form a temporal developmental switch, and how other heterochronic genes regulate the *lin-14* temporal switch.

Genetic and molecular studies of heterochronic genes in C. elegans

After 12 h of embryogenesis, the 550-cell L1 stage larvae of *C. elegans* hatch. About 80 of these cells will divide again over the next larval stages, to generate cell lineages that are in most cases distinct from those generated by blast cells at different times and locations. Many of these blast cells are shown in Fig. 1. Heterochronic mutations cause most of these blast cells (for example, the intestinal cells (E lineage), the mesoblast (M lineage) and the hypodermal cell lineage (H0, H1, H2, V1 to V6, and T lineages)) to adopt fates normally associated with cells at earlier or later stages of development.

The lin-14 *gene*

The *lin-14* heterochronic gene plays a central role in this temporal regulation of the cell lineage. Loss-of-function *lin-14* alleles cause the precocious execution of cell lineages normally observed in descendent cells one or two larval stages later. Gain-of-function *lin-14* alleles affect the same cell lineages but cause the opposite transformations in cell fate: early cell lineages are normal, but later cells reiterate the early cell lineages, normally associated with their ancestor cells. For example, in the development of the lateral hypodermis, in *lin-14* loss-of-function mutants, the blast cell T skips its characteristic larval stage 1 (L1) sublineage and instead expresses a sublineage normally associated with its granddaughter blast cell T.ap (Fig. 2). In *lin-14* gain-of-function mutants, the blast cell T expresses its normal L1-specific sublineage, but its granddaughter T.ap reiterates this L1-specific sublineage normally associated with blast cell T (Fig. 2) (Ambros and Horvitz, 1984). The *lin-14* gain-of-function mutations have been shown by genetic criteria to cause excess *lin-14* gene activity, while the *lin-14* loss-of-function mutations are due to loss of *lin-14* gene activity (Ambros and Horvitz, 1987). These data suggest that during normal development, a relatively high *lin-14* gene activity during early larval stages, for example in cell T, is reduced later in development, for example in cell T.ap, to form a temporal developmental switch. In an analogous way the *lin-14* gene coordinately controls the postembryonic fates of cells in a number of lineages: the E, M H0, H1, V1 to V6, and T lineages.

A regulatory hierarchy of heterochronic genes

Other genes have been shown to act in the same pathway as *lin-14* to control the temporal fates of these same cells or a subset of them, and their epistatic interactions have been determined (Ambros and Horvitz, 1984; Ambros, 1989). A recessive loss-of-function mutation in the gene *lin-4* causes the same retarded phenotype as *lin-14* gain-of-function mutations (Fig. 2). In addition, the *lin-4* phenotype depends on a functional *lin-14* gene: *lin-14* loss-of-function mutations are epistatic to *lin-4* (Ambros, 1989). These data show that the *lin-4* gene negatively regulates the *lin-14* gene, though they do not suggest whether this is direct or at what level the interaction occurs.

lin-28 mutations are not as pleiotropic as *lin-14* or *lin-4* mutations: these mutations cause precocious expression of later fates in P, H0, H1, V1 to V6, and T hypodermal lineages but not in the M lineages or E lineages (muscle or intestine) (Fig. 2). Mutations in *lin-*

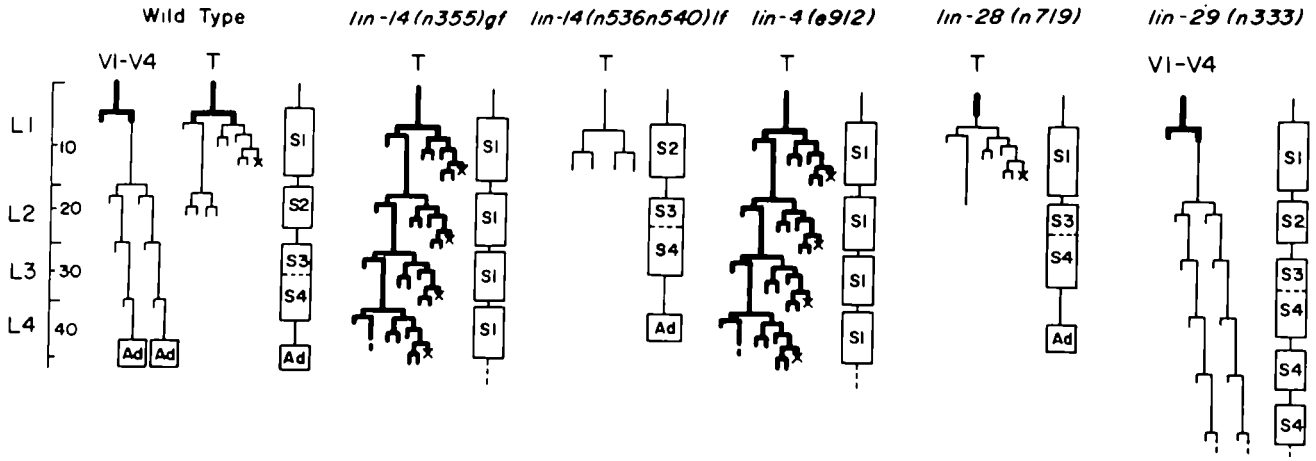

Fig. 2. Shown are two examples of cell lineages from the hermaphrodite lateral hypodermis that are affected by heterochronic mutations. The wild-type lineages of blast cells V6 and T are shown in the left panel. Shown are the lineage changes induced by mutations and superimposed in dark line are those cells that also express the *lin-14* protein. Wild type expresses *lin-14* protein only during the L1 stage. *lin-14(n355)gf* and *lin-4(e912)* express the *lin-14* protein at all developmental stages and therefore reiterate L1-specific cell lineages. *lin-14(n536n540)* shows no detectable *lin-14* protein so that precocious L2-specific lineages occur during the L1 stage. *lin-28(n719)* shows *lin-14* protein before the L1 cell division, but not after it so that a normal L1-specific lineage occurs, but L3-and later specific lineages occur precociously during the L2 stage. *lin-29(n333)* expresses *lin-14* normally during the L1 stage so that L1-specific lineages and L2-specific lineages are normally expressed. This mutant fails to express adult specific traits regardless of whether *lin-14* is mutant or not.

28 are epistatic to *lin-4* or *lin-14* gain-of-function mutations in these hypodermal lineages (Ambros, 1989). Thus the *lin-28* gene is necessary for the *lin-14* gene activity in the hypodermal lineages.

Loss-of-function mutations in the *lin-29* gene only affect larval stage 4 cell fates in the hypodermal lineages: these cells continue to express L4 fates during adult stages in *lin-29* mutants. *lin-29* mutants cause this phenotype in double mutant combinations with any of the other heterochronic genes, suggesting that it is farthest downstream in the pathway and directly regulates an L4/adult switch.

The lin-14 *gene encodes a nuclear protein that forms a temporal molecular gradient*

We cloned the *lin-14* gene by genetically mapping in parallel many linked DNA polymorphisms (RFLPs) associated with the transposon Tc1 to find the two closest to the gene (Ruvkun *et al.* 1989). We isolated *lin-14* cDNA clones and fused these to the *E. coli lacZ* gene to produce anti-*lin-14* antibodies so that the temporal, spatial, cellular and subcellular distribution of *lin-14* protein could be followed. These anti-*lin-14* antibodies detect specific somatic nuclei in wild-type preparations but not in strains bearing *lin-14* null alleles, showing that the antibody is specific for *lin-14* protein (Ruvkun and Giusto, 1989).

During wild-type development *lin-14* protein was first observed in embryos about half-way through embryogenesis and staining was most intense in nuclei of late embryos just before hatching, and in newly hatched L1 animals. *lin-14* protein was present in the nuclei of all postembryonic blast cells affected by *lin-14* mutations: the hypodermal blast cells H1, H2, V1 to V6, and T, the intestinal (E) cells, both neuroblasts QR and QL, the mesoblast M cell, and in the P cells. No staining was observed in the somatic or germ line gonadal blast cell nuclei Z1 to Z4, which are not affected by *lin-14* mutations.

The blast cells containing *lin-14* protein execute L1-specific lineages and their daughter cell nuclei initially contain *lin-14* protein but the level falls before the next cell division (Fig. 2). Late in the L1 stage, the *lin-14* protein staining in all nuclei rapidly falls. By L2 and in subsequent larval stages, only barely detectable *lin-14* protein staining remains in some neuronal nuclei (Arasu *et al.* 1991). Western blotting has shown that the level of the *lin-14* protein decreases by a factor of more than 25 from L1 to L2.

Gain-of-function *lin-14* mutations dramatically alter the temporal regulation of *lin-14* protein levels

Gain-of-function *lin-14* mutations cause reiterations of early larval cell lineages at late larval and adult stages in a number of cell lineages and tissues (Ambros and Horvitz, 1984). Embryonic and L1 stage staining with the anti-*lin-14* antibodies in these mutants was equivalent to wild type in amounts, nuclear localization and cellular distribution of the protein. However, unlike wild type, these mutants showed *lin-14* protein at high levels in many nuclei during larval stages 2, 3, 4 and in adults. All of the postembryonic hypodermal blast, intestinal blast, neuroblast and mesoblast cells known to be affected by gain-of-function *lin-14* mutations inappropriately accumulate the *lin-14* protein at these late stages. For example, in the case of the T lineage,

during normal development, the *lin-14* nuclear protein is present at high levels in the T cell but is not observable in cell T.ap (Fig. 2). In *lin-14* gain-of-function mutants, this temporal gradient is disrupted: the *lin-14* protein is now observed at all stages of development (Ruvkun and Giusto, 1989), for example in both cells T and T.ap (as well as T.apap, etc.), and these cells reiterate L1-specific cell lineages (Fig. 2).

Thus the normally sharp decrease in *lin-14* protein levels during the L1 stage causes cells to switch from L1-specific cell lineages to L2-specific cell lineages, and in *lin-14* gain-of-function mutants, the inappropriate presence of the *lin-14* nuclear protein late in development prevents this temporal switch in cell fate.

The lin-14 *gain-of-function mutations are located in the 3′ untranslated region of the* lin-14 *mRNA*

The *n355* mutation is an insertion or inversion of at least 10 kb of unknown DNA sequences 256 bases 3′ to the termination codon of the *lin-14* protein coding region common to both *lin-14* transcripts (Fig. 3). The other *lin-14* gain-of-function mutation, *n536*, is a 607 bp deletion, 300 bases downstream from the 3′ end of the *lin-14* open-reading frame (Fig. 3) and overlaps the region which is rearranged in the *n355* mutation.

The location of both *lin-14* gain-of-function mutations in the 3′ untranslated region suggests that they do not affect the *lin-14* protein. Immunoblot analysis of the *lin-14* proteins from wild-type and both the gain-of-function mutants confirmed this prediction; no difference in the size of the *lin-14* proteins was observed in these mutants (Wightman *et al.* 1991).

Computer analysis of the *lin-14* mRNA sequence identified a stem–loop structure with 21 out of 22 matching pairs in the main stem and side stems of 9 and 10 base pairs located about 60 bases downstream from the end of the *lin-14* protein coding region and about 80 bases upstream of the *n355* breakpoint. This structure would be expected to have a free energy of −38.3 kcal mole^{-1}, about 3× that of a randomized sequence of the same base composition. The location of this structure suggests that it would not be directly affected by the *lin-14* gain-of-function mutations.

Thus the 3′ region of the *lin-14* transcripts contains an element that encodes the down-regulation of *lin-14* protein levels after larval stage 1. Posttranscriptional regulation of translation or transcript stability has been demonstrated in the 3′ UTR of other eukaryotic genes (Casey *et al.* 1988) and proteins have been identified that bind to those elements (Leibold and Munro, 1988). In HIV, the *rev* transactivator protein has been shown to interact with an RNA structure in the 3′ UTR of the viral transcript (Zapp and Green, 1989) and to mediate export of the mRNA from the nucleus (Emerman *et al.* 1988). Thus, sequences in the 3′ UTR can regulate export of a transcript from the nucleus, the half-life of the transcript, or translation of that transcript.

Temporal down-regulation of lin-14 *expression is initiated by a developmental cue*

When newly hatched *C. elegans* L1 animals are starved, they do not begin postembryonic development, but instead can suspend development for up to 5 days, when upon feeding, they will reinitiate postembryonic development. Because the level of the *lin-14* protein normally falls within 12 h of hatching, we investigated whether the passage of time in these suspended L1s causes the *lin-14* protein levels to fall. We found that both the *lin-14* protein levels and the *lin-14* transcript levels are maintained at the high level of normal early L1 animals in these starved L1s (Arasu *et al.* 1991). Thus the *lin-14* down regulation that is necessary for normal developmental timing does not respond to clock time, but instead must first respond to developmental cues.

The high level of *lin-14* protein expression in these starved L1s could reflect a very stable *lin-14* protein or mRNA that is only destablized upon feeding, perhaps *via* induction of a protease or nuclease. We addressed this by inhibiting translation with cycloheximide and

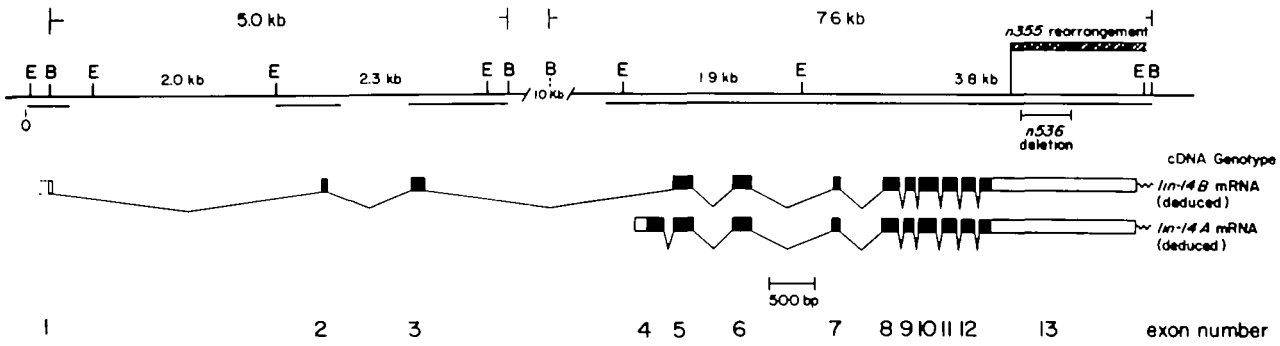

Fig. 3. A molecular map of the *lin-14* region showing the organization of the gene with the two *lin-14* transcripts inferred from cDNAs isolated from both wild-type (N2) and *lin-14* mutant strains. The cDNAs were isolated using either the 3.8 kb *Eco*RI fragment, the 7.6 kb *Bgl*II fragment, or the 5.0 kb *Bgl*II fragment as probe (B, *Bgl*II, E, *Eco*RI). Exons containing open reading frames are shown in black, whereas those containing untranslated regions are shown in white. Exons are numbered below. Those genomic regions that have been sequenced are shown under the genomic restriction map and generally flank the exons detected in the cDNA clones. This analysis could have missed exons that are incorporated rarely into *lin-14* transcripts. The location of the two *lin-14* gain of function mutations in the 3′ untranslated region are shown above and below the genomic restriction map.

found that the level of the *lin-14* protein quickly fell to zero, even though myosin levels remained constant. The level of the *lin-14* mRNA also remains at normal levels in starved L1s and cycloheximide-treated starved L1s. These data suggest that *lin-14* translation continues in these starved L1s.

How other heterochronic genes interact with lin-14 *to generate the temporal gradient or interpret it*

Other heterochronic genes could control developmental timing of the *C. elegans* cell lineage by participating in the generation or reception of the *lin-14* temporal gradient. The *lin-4* gene is necessary for the down regulation of *lin-14* protein levels: inappropriate *lin-14* protein staining at late stages is observed in this mutant (Fig. 2) (Arasu *et al.* 1991). Thus, directly or indirectly, *lin-4* negatively regulates *lin-14*. It is possible that the *lin-4* protein directly interacts with the *lin-14* 3′ regulatory sequences.

Mutations in *lin-28* lead to precocious expression of L3-specific hypodermal cell lineages during the L2 stage, and so are similar to *lin-14* mutations in these hypodermal cell lineages only (Ambros and Horvitz, 1984). The level of *lin-14* protein is decreased in *lin-28* mutants initially in the hypodermal cell lineages but in all cell lineages of starved L1 animals, suggesting that the gene acts upstream of *lin-14* (Arasu *et al.* 1991). We have not yet determined at what level the *lin-28*/*lin-14* gene interaction is taking place. *lin-28* could activate *lin-14* transcription or translation, or stabilize the *lin-14* protein by, for example, a heterodimer interaction.

The lin-14 *gene encodes two protein products*

The *lin-14* DNA sequence revealed 13 exons in the *lin-14* gene, with introns ranging in size from about 12 kb to 44 bp. Differential splicing of exons 1,2, and 3 (the *lin-14*B transcript) or exon 4 (the *lin-14*A transcript) to the common exons 5 to 13 was observed in *lin-14* cDNAs (Fig. 3). RNAase protection experiments using probes to exons 1,2, and 3 and exons 4 and 5 confirmed this differential splicing and showed that the two transcripts are expressed in wild type at similar levels and show the same temporal regulation of transcript levels.

The longest open reading frame in the *lin-14*B transcript yields a protein of 539 amino acids. The *lin-14*A transcript encodes a 537 amino acid *lin-14*A protein. This protein has a 63 amino acid N-terminal domain that is distinct from the 65 amino acid N-terminal domain of the *lin-14*B protein. No similarity between the two amino-termini of the *lin-14* proteins could be detected.

Databank searches using the amino acid sequences of both *lin-14* proteins revealed no significant sequence similarity to any proteins. A 20 amino acid region (positions 417 to 436), common to both *lin-14* proteins, could potentially form an amphipathic α-helix with two basic regions separated by an acidic region on the hydrophilic face (Fig. 4) (Wightman *et al.* 1991). This motif could encode a DNA- or RNA-binding domain.

The *lin-14* proteins also contain a high proportion of prolines, serines, threonines and glutamates, or PEST sequences, as has been observed in various unstable proteins (Rogers *et al.* 1986; Nash *et al.* 1988). Given that the protein appears to be quickly degraded late in larval stage 1 to control an L1/L2 switch in cell fates (Ruvkun and Giusto, 1989), these sequences could mediate that instability.

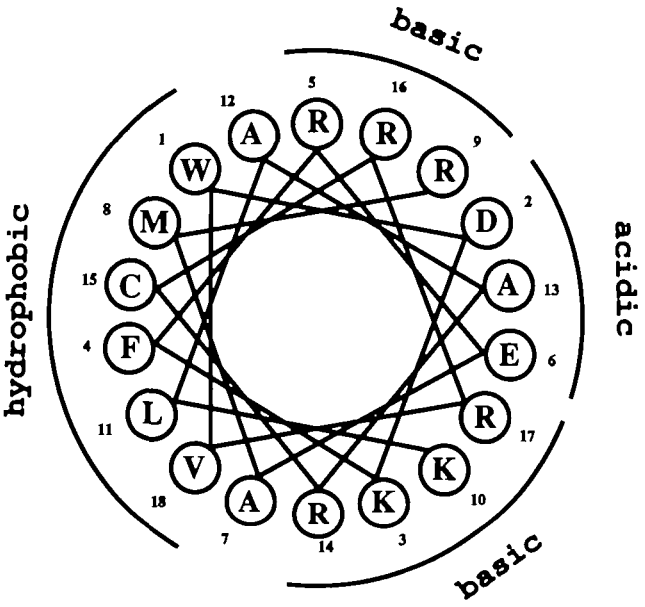

Fig. 4. A helical wheel of *lin-14* residues 417 to 434, showing a hydrophobic face, two basic patches, and an acidic patch.

Discussion

Generation of the lin-14 *temporal gradient*

The DNA sequence of the *lin-14* gene from wild-type and three *lin-14* mutants, and its expression pattern in wild type and various mutants, has revealed aspects of how the *lin-14* protein gradient is generated (Fig. 5). First, during normal development the down regulation of *lin-14* protein levels begins only after a feeding signal initiates postembryonic development. Both the *lin-14* and the *lin-28* gene activities are necessary to maintain the level of *lin-14* protein high before this food signal. After initiation of postembryonic development, down-regulation of *lin-14* is triggered. The *lin-4* gene activity is necessary for this down-regulation, and may in fact be up-regulated by feeding or postembryonic development. This leads to a 25-fold decrease in *lin-14* protein level from L1 to L2 that causes cells to switch to L2-specific fates. *lin-14* gain-of-function mutations abrogate the negative regulation of *lin-14* protein levels. This failure to reduce markedly the *lin-14* protein levels prevents or delays the normal L1 to L2 switch in cell fates in these mutants.

The stability of the *lin-14* protein is also relevant to the formation of the *lin-14* temporal gradient. The half-life of the previously synthesized *lin-14* protein must be less than 1 h to account for this observed rate of disappearance. The *lin-14* protein levels decrease at the same rate in both dividing cells and non-dividing cells. This

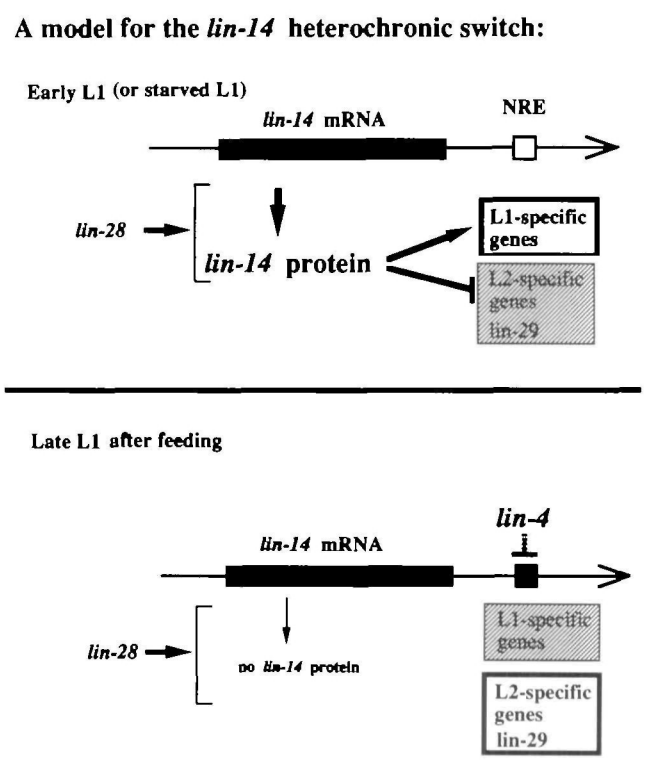

Fig. 5. A model for the generation of the *lin-14* protein gradient. In early L1 or starved L1 wild-type animals, *lin-14* protein is actively translated and specifies L1-specific cell lineages, perhaps by activating L1-specific genes in each blast cell and by repressing L2-specific and later gene activities including *lin-29*. During normal development, *lin-14* protein levels decrease at least 25-fold from the L1 to the L2 stage. *lin-28* is required either to activate positively *lin-14* expression or to stabilize the *lin-14* protein during the L1 stage of wild-type development and during the L2 and later stages of development in *lin-4* and *lin-14*gf mutants. A decrease in *lin-14* protein levels during the mid-L1 stage is triggered by feeding and/or by postembryonic developmental events. This triggering event may activate *lin-4* expression or activity to negatively regulate *lin-14* at a posttranscriptional level. This negative regulation of *lin-14* does not operate by down-regulating *lin-28* because the *lin-28* gene is still active at the L2 stage in as shown by the fact that *lin-28* mutations are epistatic to *lin-14*gf mutations. Thus *lin-4* negatively regulates *lin-14* independently of *lin-28*. The *lin-4* gene may directly or indirectly negatively regulate *lin-14* expression *via* the negative regulatory element (NRE) in the *lin-14* 3′ UTR (depicted as an open box in the L1 stage, when it is not acting and as a black box in the L2 stage, when it is necessary for posttranscriptional regulation). This negative regulatory element is inactivated in *lin-14* gain-of-function mutants allowing inappropriate expression of *lin-14* protein at L2 and later stages. Hence, this negative regulatory element functions at the same time as *lin-4*. The *lin-4* protein could bind to the *lin-14* NRE or another protein could act at the *lin-14* NRE while *lin-4* acts to regulate, for example, *lin-14* translation or protein stability. This disappearance of the *lin-14* protein prevents the expression of L1-specific cell lineage genes and allows the expression of L2-specific and subsequent cell lineage genes, including *lin-29*, that specifies the switch from larval to adult-stage cuticle formation (Ambros, 1989). Arrow denote positive regulation; bars denote negative regulation.

suggests that, at least in the non-dividing cells, breakdown of the nuclear membrane is not necessary for the degradation of the *lin-14* protein and that nuclearly localized proteases must control this process (Ruvkun and Giusto, 1989). The presence of PEST sequences in the *lin-14* protein supports the notion that the rate of *lin-14* protein degradation is relevant to the formation of the *lin-14* protein gradient.

While the DNA sequence of the *lin-14* gene did not reveal any homology that would suggest its molecular mechanism, the nuclear localization of the *lin-14* proteins suggests that they may regulate the pattern of gene expression of the cells that accumulate them. The observation that the *lin-14* protein is normally present only in embryos and larval stage 1 animals suggests that it either activates early genes or represses late genes so that the disappearance of *lin-14* after larval stage 1 causes a transition from the expression of early cell lineage genes to late cell lineage genes.

The particular early or late cell fate specified by the level of *lin-14* gene activity is distinct for many of the postembryonic cell lineages affected by *lin-14* mutations, although the fates inappropriately executed in these mutants are always fates normally executed by a closely related descendent or ancestor cell (Ambros and Horvitz, 1984). The *lin-14* gene product may function to convey general temporal information to these cell lineages. The specific response made by each cell must be caused by unique properties of that cell, defined by other developmental control genes, that either modify the *lin-14* signal or cause cells to interpret it differently. It is possible that this difference between cells is reflected in the distinct spectra of previously specified or partitioned nuclear factors which interact with the *lin-14* nuclear protein to cause a distinct response to this temporal signal in nearly every cell. The best current model for this type of signal integration in cell fate specification is from yeast: the three yeast cell types are specified in a combinatoric manner by the MATa1 and MATα2 nuclear proteins (Goutte and Johnson, 1986).

Pattern-formation genes and evolution

The spatial and temporal asymmetries in the patterns of developmental control gene activity during ontogeny has been shown to cause cells or groups of cells to become different from each other (Ingham, 1988). Mutations that change the expression or activation pattern of pattern-formation genes in genetically studied animals such as *Drosophila* and *C. elegans* lead to major changes in the morphology of the organisms, in many cases deleting structures and/or duplicating groups of cells and structures (Sternberg and Horvitz, 1986; Akam, 1987). For example, the *Drosophila Antennapedia* mutant adds a nearly complete leg where an antenna is normally located (Dennell, 1973), and particular *bithorax* mutants can add another pair of wings to a thorax segment that normally has no wing (Lewis, 1982). The *lin-14* gain-of-function heterochronic mutant duplicates entire sets of muscle, hypoderm, endoderm and neuronal cells, and is missing other such cells.

These are the same sorts of homeotic and heterochronic variations that have been observed in phylogeny and point to mutations in pattern-formation genes as a major cause of the variation necessary for evolutionary change

(Raff and Kauffman, 1983). For example, the *C. elegans* heterochronic mutations are analogous to the heterochronic variation between species noted in phylogenetic studies (Gould, 1977). This heterochronic variation in evolution could be due to mutation in one or a few heterochronic genes like *lin-14*, rather than many mutations that incidentally change developmental timing. More generally, mutations that change the spatial, temporal, or cellular asymmetries in pattern-formation gene activities may be the underlying cause of the many forms of metazoans and may be a significant force in evolutionary change.

While the most frequent class of mutations are those that lead to a decrease in gene activity, many mutations that cause an increase or inappropriate activity of pattern-formation genes have also been isolated. In the case of the *lin-14* heterochronic mutations, deletion of the *lin-14* 3′ untranslated regulatory region causes inappropriate expression of the *lin-14* protein at late developmental stages, leading to reiteration of early cell lineages. The *Drosophila Antennapedia* mutation has been shown to arise by a chromosomal inversion that causes more promiscuous expression of the gene due to fusion of the protein-coding regions to a novel regulatory region (Schneuwly *et al.* 1987). This causes the *Antp* protein to be uniformly expressed rather than expressed only in the thorax region.

The misregulated expression of pattern-formation genes does not completely disrupt all cell identity specifications during development. Rather, only a limited number of structures or cells are affected. For example, uniform expression of the *Antp* protein transforms only the antenna to a leg, but does not affect other segments (Schneuwly *et al.* 1987). Similarly, uniform expression of one of the *bithorax* complex proteins causes segmental transformations of only particular cells within particular segments (Mann and Hogness, 1990). Uniform expression of the *int-1* oncogene in *Xenopus* causes production of two organizers, leading to a duplication of the embryonic axis (McMahon and Moon, 1989). Thus only some cells are competent to respond to the presence of these pattern-formation gene products so that uniform expression causes a non-uniform response. This is most likely due to the combinatorial nature of pattern-formation genes: a spectrum of proteins is necessary to specify particular cell fates so that only cells already containing the interacting proteins will be responsive to the inappropriate presence of one such gene product. These data suggest that discrete developmental variation can result from mutations that lead to uniform expression of a particular pattern-formation gene.

Gain-of-function mutations that lead to misregulated gene expression or activity can also arise in any gene or gene product that contains a negative regulatory domain. For example, gain-of-function mutations in the *C. elegans* sex determination genes *tra-2* and *fem-3* that transform hermaphrodites into pseudomales or females, are located in a negative regulatory region of the 3′ end of the transcript, downstream of the protein coding regions (Rosenquist and Kimble, 1988; Okkema and Kimble, personal communication). Similarly, 3′ mutations in the vertebrate oncogene *c-fos* lead to activation of this protooncogene, possibly by stabilizing the mRNA (Curran *et al.* 1985). A gain-of-function mutation in the *Drosophila* sex determination gene *Sxl* has been shown to disrupt a region that negatively regulates one of the spliced mRNAs, leading to expression of the normally XX(female)-specific *Sxl* protein in XO males (Bell *et al.* 1988). Gain-of-function mutations in the *C. elegans lin-12* (Greenwald and Seydoux, 1990) and *Drosophila Notch* (Kelley *et al.* 1987) genes that control particular cell–cell signaling during development have been shown to be point mutations in protein coding regions, thus identifying a domain which negatively regulates the activity of these proteins. Negative regulation is a common form of feedback control, and those regions of genes, mRNAs and proteins that mediate down regulation would be the targets for such mutations. Dominant mutations have also been shown to result in mutant gene activities that interfere with the wild-type gene activity (Muller, 1932; Park and Horvitz, 1986), or that result in entirely new gene activities (Muller, 1932), or that reduce or eliminate gene activity (Muller, 1932).

Such gain-of-function mutations in pattern-formation genes lead to genetically dominant morphological transformations. The genetically dominant nature of such pattern-formation gene mutants has enormous evolutionary implications. Unlike recessive mutations, dominant mutations need not be homozygous to affect the phenotype. The dominant mutation will be segregated to and cause a phenotype in 1/2 of the progeny of a heterozygous carrier, independent of effective breeding populations. In this way, evolutionary change can arise and a selectively advantageous allele can become more frequent in a large non-isolated population.

The evolutionary implications of gain-of-function mutations seem to have been overlooked by population biologists and evolutionary theorists. While subtle recessive mutations are sufficient to explain the observed gradual variation between races and related species, sudden morphological changes have been noted in phylogeny and have been a major problem for evolutionary theories that suppose gradual change *via* recessive mutations. One camp of paleontologists resolves this conflict by stressing that the sudden changes are artefacts of the incomplete geological record and that evolutionary change indeed proceeds by incremental steps (Grant, 1985). On the other hand, advocates of punctuated evolution have stressed that bursts of major structural changes in the paleontological record are not due to gaps in the record but in fact accurately present phylogenetic history (Eldredge and Gould, 1972). For these rapid evolutionary changes, they argued that small isolated populations would be able to generate morphological variants due to a relatively high frequency of production of homozygous recessive mutants, and that the theory of allopatric speciation could explain the sudden appearance of such a new form in the paleontological record: the variant would arise and define a new species in a geographically isolated region and then suddenly (in geological time) appear in larger areas.

In contrast, we propose that because genetically dominant mutations in pattern-formation genes have been shown to be relatively common and to cause significant morphological changes, sufficient variation for

sudden evolutionary change could take place within large breeding populations. Thus, due to mutations in pattern-formation genes, even a large outbred population would produce a constellation of morphological variants, most of which are not viable. Rarely, perhaps due to recombination of two such mutations or due to a change in environment, viable and fertile mutants would be selected and either displace or speciate from the original form. Thus there is a molecular genetic mechanism to generate the much maligned 'hopeful monsters' of Goldschmidt (Raff and Kaufmann, 1983). These major variants could be the vanguard of major evolutionary events, like the flowering of metazoans during the Cambrian Period or the mammalian radiation.

References

Akam, M. (1987). The molecular basis for metameric pattern in *Drosophila* embryo. *Development* **101**, 1–22.

Ambros, V. (1989). A hierarchy of regulatory genes controls a larva-to-adult developmental switch in *C. elegans* *Cell* **57**, 49–59.

Ambros, V and Horvitz, H. R (1984). Heterochronic mutants of the nematode *Caenorhabditis elegans* *Science* **226**, 409–416.

Ambros, V. and Horvitz, H. R. (1987). The *lin-14* locus of *Caenorhabditis elegans* controls the time of expression of specific postembryonic developmental events. *Genes Dev.* **1**, 398–414.

Arasu, P., Wightman, B and Ruvkun, G. (1991). The *lin-14* heterochronic switch of *C. elegans* is triggered by feeding and is temporally regulated by the antagonistic action of *lin-4* and *lin-28*. manuscript submitted.

Aziz, N. and Munro, H. N (1986). Both subunits of rat liver ferritin are regulated at a translational level by iron induction. *Nucleic Acids Res.* **14**, 915–927.

Balling, R., Deutch, U. and Gruss, P. (1988). *undulated*, a mutation affecting the development of the mouse skeleton, has a point mutation in the paired box of Pax 1 *Cell* **55**, 531–535.

Beachy, P. A., Helfand, S. L. and Hogness, D. S. (1985). Segmental distribution of bithorax complex proteins during *Drosophila* development. *Nature* **313**, 545–551.

Bell, L. R., Maine, E. M., Schedl, P. and Cline, T. W. (1988). Sex-lethal, a *Drosophila* sex-determination switch gene, exhibits sex-specific RNA splicing and sequence similarity to RNA binding proteins. *Cell* **55**, 1037–1046.

Braun, R. E. *et al.* (1989). Protamine 3'-untranslated sequences regulate temporal translational control and subcellular localization of growth hormone in spermatids of transgenic mice. *Genes Dev.* **3**, 793–802

Casey, J. L. *et al.* (1988). Iron-responsive elements: Regulatory RNA sequences that control mRNA levels and translation. *Science* **240**, 924–928.

Chalfie, M , Horvitz, H. R and Sulston, J. E. (1981). Mutations that lead to reiterations on the cell lineages of *C elegans*. *Cell* **24**, 59–69.

Costa, M , Weir, M., Coulson, A., Sulston, J. E. and Kenyon, C. (1988). Posterior pattern formation in *C. elegans* involves position-specific expression of a gene containing a homeobox. *Cell* **55**, 747–756

Curran, T. C , Van Beveren, N., Ling and Verma, I M. (1985) Viral and cellular fos proteins are complexed with a 39,000-Dalton cellular protein. *Molec. cell. Biol.* **5**, 167–172.

Dennell, R E. (1973) Homeosis in *Drosophila* I Complementation studies with revertants of Nasabemia. *Genetics* **75**, 279–297.

Eldredge, N and Gould, S. J. (1972) *Punctuated Equilibria: An Alternative to Phyletic Gradualism in Models in Paleobiology*. Freeman, Cooper and Co , San Francisco, 82–115.

Emerman, M., Vazeux, R. and Peden, K. (1988). The rev gene product of the Human Immunodeficiency Virus affects envelope-specific RNA localization. *Cell* **57**, 1155–1165.

Finney, M., Ruvkun, G. and Horvitz, H. R. (1988). The C. elegans cell lineage and differentiation gene unc-86 encodes a protein with a homeodomain and extended similarity to transcription factors. *Cell* **55**, 757–769

Gould, S. (1977). *Ontogeny and Phylogeny*, Harvard University Press, Cambridge, MA.

Goutte, C. and Johnson, A. D. (1988). a1 protein alters the DNA binding specificity of a2 repressor. *Cell* **52**, 875–882.

Graham, A., Papalopulu, N. and Krumlauf, R. (1989). The *Murine* and *Drosophila* homeobox gene complexes have common features of organization and expression. *Cell* **57**, 367–378.

Grant, V. (1985). *The Evolutionary Process*. Columbia University Press.

Greenwald, I. (1985). *lin-12*, a nematode homeotic gene, is homologous to a set of mammalian proteins that includes epidermal growth factor. *Cell* **43**, 583–590.

Greenwald, I and Seydoux, G. (1990). Analysis of gain-of-function mutations of the *lin-12* gene of *Caenorhabditis elegans*. *Nature* **346**, 197–199.

Hafen, E., Basler, K., Edstrom, J. E. and Rubin, G. M. (1987). *Sevenless*, a cell-specific homeotic gene of *Drosophila*, encodes a putative transmembrane receptor with a tyrosine kinase domain *Science* **236**, 55–63.

Ingham, P. (1988). The molecular genetics of embryonic pattern formation in *Drosophila*. *Nature* **335**, 25–34

Kelley, M. R., Kidd, S., Deutsch, W A. and Young, M. W. (1987). Mutations altering the structure of epidermal growth factor-like coding sequences at the *Drosophila* notch locus. *Cell* **51**, 539–548.

Leibold, E. A. and Munro, H. N. (1988). Cytoplasmic protein binds *in vitro* to a highly conserved sequence in the 5' untranslated region of ferritin heavy- and light-subunit mRNAs. *Proc. natn Acad. Sci. U.S.A* **85**, 2171–2175.

Lewis, E. B. (1982). A gene complex controlling segmentation in *Drosophila*. *Nature* **276**, 565–570.

Mann, R and Hogness, D. S. (1990). Functional dissection of ultrabithorax proteins in *D. melanogaster*. *Cell* **60**, 597–610

Muller, H. J. (1932). *Proc. 6th Int Cong Genet* **1**, 213–255.

Nash, R., Tokiwa, G., Anand, S , Erickson, K. and Futcher, A. B (1988). The WH11+ gene Saccharomyces cerevisiae tethers cell division to cell size and is a cyclin homolog. *EMBO J.* **7**, 4335–4346.

Park, E.-C. and Horvitz, H R. (1986) Mutations of dominant effects on the behavior and morphology of the nematode. *Genetics* **113**, 821–852.

Raff, R. A. and Kaufman, T C. (1983). *Embryos, Genes and Evolution* Macmillan Publishing Co., New York

Rogers, S., Wells, R and Rechsteiner, M. (1986). Amino acid sequences common to rapidly degraded proteins: The PEST hypothesis. *Science* **234**, 364–368.

Rosenquist, T. and Kimble, J. (1988). Molecular cloning and transcript analysis for fem-3, a sex-determination gene *Caenorhabditis elegans*. *Genes Dev.* **2**, 606–616

Ruiz i Altaba, A. and Melton, D. A. (1989). Involvement of the *Xenopus* homeobox gene Xhox3 in pattern formation along the anterior–posterior axis. *Cell* **57**, 317–326.

Ruvkun, G., Ambros, V., Coulson, A., Waterston, R., Sulston, J. and Horvitz, H. R. (1989). Molecular genetics of the *Caenorhabditis elegans* heterochronic gene in lin-14. *Genetics* **121**, 501–516

Ruvkun, G and Giusto, J. (1989). The *Caenorhabditis elegans* heterochronic gene *lin-14* encodes a nuclear protein that forms a temporal developmental switch. *Nature* **338**, 313–319

Schneuwly, S., Kuroiwa, A. and Gehring, W. J. (1987) Molecular analysis of the dominant homeotic antennapedia phenotype. *EMBO J* **6**, 201–206.

Scott, M. P , Tamkun, J. W. and Hartzell, G. W. (1989). The structure and function of the homeodomain in *BBA Reviews on Cancer* **989**, 25–48.

Sternberg, P. S. and Horvitz, H. R. (1986). Genetic control of cell lineage during nematode development *Ann. Rev. Genet.* **18**, 489–524.

Sulston, J. E., Schierenberg, E., White, J. G. and Thomson, J. N. (1983). The embryonic cell lineage of the nematode Caenorhabditis elegans *Devl Biol.* **100**, 64–119.

Wightman, B C , Bürglin, T. R., Gatto, J., Arasu, P. and Ruvkun, G. (1991). Posttranscriptional regulation of the *C elegans* heterochronic gene *lin-14* generates a temporal developmental switch. Manuscript submitted

Wood, W. (ed.) (1988). *The Nematode* Caenorhabditis elegans Cold Spring Harbor Press.

Zapp, M. L. and Green, M. R. (1989) Sequence-specific RNA binding by the HIV-1 rev protein. *Nature* **342**, 714–716.