Mutations in the $\alpha 2(\text{IV})$ basement membrane collagen gene of *Caenorhabditis elegans* produce phenotypes of differing severities

Marion H.Sibley¹, Patricia L.Graham¹,
Nicola von Mende² and James M.Kramer^{1,3}

¹Department of Cell, Molecular and Structural Biology, Northwestern University Medical School, Chicago, IL 60611, USA and ²Rothamsted Experimental Station, Harpenden, Herts, AL5 2JQ, UK

³Corresponding author

Communicated by R.Timpl

Type IV collagen forms a network that provides the major structural support of basement membranes. We have determined the nucleotide alterations and phenotypes of 17 mutant alleles of the *Caenorhabditis elegans* $\alpha 2(\text{IV})$ collagen gene *let-2*. All 17 mutations are within the triple helical (Gly-X-Y) repeat domain of the molecule. Fifteen of the mutations are replacements of Gly-X-Y repeat glycines with aspartate, glutamate or arginine, and they cause a wide range of phenotypes. The mildest alleles are nearly wild-type at 15 and 20°C but embryonic lethal at 25°C, while the most severe allele is embryonic lethal at all three temperatures. Mutations resulting in severe phenotypes are generally located in areas of lower calculated thermal stability of the type IV collagen molecule. An alanine to threonine substitution at position X of a Gly-X-Y triplet immediately following an interruption results in a severe phenotype. This mutation is unusual because substitutions at positions X or Y have not generally been found to cause strong phenotypes in *C.elegans* or human collagens. An intron splice acceptor mutation causes a strict embryonic lethal phenotype, but does not completely abolish gene function. Pairs of independent mutations affect each of three glycines, indicating a non-random distribution of mutations in the molecule. It is suggested that this clustering results because many glycine substitutions may cause dominant lethal or sterile phenotypes.

Key words: basement membrane/collagen/embryonic lethal/mutations/temperature-sensitive

Introduction

Type IV collagen is a major structural component of basement membranes, assembling into a complex three-dimensional network of highly branched filaments (Yurchenco and Ruben, 1987; Yurchenco and Schittny, 1990). The type IV collagen molecule is most commonly a heterotrimer of two $\alpha 1(\text{IV})$ chains and one $\alpha 2(\text{IV})$ chain. Four additional type IV collagen chains, $\alpha 3$ – $\alpha 6$, have been identified in mammals (Saus *et al.*, 1988; Gunwar *et al.*, 1990; Hostikka *et al.*, 1990; Zhou *et al.*, 1993). The $\alpha 1$ and $\alpha 2$ type IV collagen chains are found in all basement membranes, while the other chains have rela-

tively restricted tissue distributions (Desjardins and Bendayan, 1991; Paulsson, 1993).

Complete amino acid sequences have been determined for the $\alpha 1(\text{IV})$ and $\alpha 2(\text{IV})$ collagen chains of human (Brazel *et al.*, 1987; Hostikka and Tryggvason, 1988), mouse (Muthukumaran *et al.*, 1989; Saus *et al.*, 1989), and the nematode *Caenorhabditis elegans* (Guo and Kramer, 1989; Guo *et al.*, 1991; Sibley *et al.*, 1993). In addition, the complete sequences of human $\alpha 5$ (Zhou *et al.*, 1992), *Drosophila* $\alpha 1$ (Blumberg *et al.*, 1988), sea urchin $\alpha 1$ (Exposito *et al.*, 1993) and *Ascaris suum* $\alpha 2$ (Pettitt and Kingston, 1991) type IV collagen chains have been determined. All type IV collagen sequences have conserved structural features. The (Gly-X-Y)_n triple helical domain makes up about three-quarters of the chain and is punctuated with numerous interruptions. Each type IV chain has two non-collagenous domains, one at the N-terminus that forms part of the 7S domain, and the highly conserved C-terminal NC1 domain. The type IV collagen network in basement membranes is formed by tetramerization of 7S domains, dimerization of NC1 domains, and lateral interactions along the length of the triple helical domain (Yurchenco and Ruben, 1987). In addition to these self-interactions, type IV collagen binds cell surface receptors and other basement membrane components, such as laminin, heparan sulfate proteoglycans and nidogen (Timpl, 1989; Charonis and Tsilibary, 1990).

In *C.elegans*, basement membranes surround the pharynx, intestine and gonad, and cover the pseudo-coelomic face of the hypodermis (Albertson and Thomson, 1976; White *et al.*, 1976, 1986). *C.elegans* has two type IV collagen genes that encode proteins very similar to the mammalian $\alpha 1(\text{IV})$ and $\alpha 2(\text{IV})$ collagen chains (Guo and Kramer, 1989; Guo *et al.*, 1991; Sibley *et al.*, 1993; Kramer, 1994). The $\alpha 1(\text{IV})$ chain is encoded by *emb-9* located on chromosome III, and the $\alpha 2(\text{IV})$ chain is encoded by *let-2* on the X chromosome. The diversity of type IV collagen molecules is increased in *C.elegans* by alternative splicing of the $\alpha 2(\text{IV})$ transcript, resulting in the production of two transcripts having either of two mutually exclusive exons. These two exons encode very similar Gly-X-Y sequences and differ primarily in the sequence of a nine or 10 amino acid interruption. The relative levels of the two alternatively spliced $\alpha 2(\text{IV})$ collagen transcripts are developmentally regulated, such that one transcript predominates during embryogenesis, while the other transcript predominates during larval and adult stages (Sibley *et al.*, 1993).

Mutations in the *C.elegans* $\alpha 1$ and $\alpha 2(\text{IV})$ collagen genes, *emb-9* and *let-2*, can cause embryonic lethality (Guo *et al.*, 1991; Sibley *et al.*, 1993), indicating that defects in basement membranes disrupt normal embryonic development. Sequence analysis of three *C.elegans* $\alpha 1(\text{IV})$ collagen mutations showed that they were all substitutions

Table I. Phenotypes of *let-2* homozygotes^a

Genotype ^b	15°C			20°C			25°C		
	Emb	Acc	Ad	Emb	Acc	Ad	Emb	Acc	Ad
N2	2	1	97	1	3	96	2	4	94
<i>unc-3</i>	2	0	98	0	2	98	3	7	90
<i>g25</i>	3	0	97	2	16	82	96	4	0
<i>b246</i>	6	2	92	9	13	78	97	3	0
<i>mn147</i>	13	56	31	7	93	0	100	0	0
<i>mn139</i>	4	0	96	20	80	0	95	5	0
<i>mn109</i>	10	54	36	37	63	0	100	0	0
<i>mn152</i>	4	50	46	50	50	0	100	0	0
<i>g30</i>	2	5	93	90	10	0	100	0	0
<i>e1470</i>	6	4	90	94	6	0	100	0	0
<i>mn151</i>	13	8	79	96	4	0	100	0	0
<i>mn103</i>	35	24	41	98	2	0	100	0	0
<i>mn126</i>	7	75	18	100	0	0	100	0	0
<i>mn129</i>	6	85	9	100	0	0	100	0	0
<i>mn143^c</i>	100	0	0	100	0	0	100	0	0
<i>mn153^c</i>	100	0	0	100	0	0	100	0	0
<i>mn114^e</i>	n.d. ^d	n.d.	27	n.d.	n.d.	14	n.d.	n.d.	0
<i>mn101^c</i>	n.d.	n.d.	8	n.d.	n.d.	0	n.d.	n.d.	0

^aHomozygous mutant animals from homozygous mothers were examined to determine the percentage that die without hatching (Emb), arrest as larvae (Acc), or become adults (Ad).

^bN2 and *unc-3* are control strains that are wild-type for *let-2*. Strains carrying *let-2* mutations are listed by allele designation. All *let-2* mutations, except *g30*, *g25* and *b246*, were marked with an *unc-3* mutation. *g37* has the same nucleotide alteration as *b246*; its phenotype was not quantitatively determined.

^c*mn143* and *mn153* could not be grown as homozygotes, but all homozygous mutant embryos fail to hatch.

^dn.d., not determined.

^eBecause these alleles are essentially sterile as homozygotes, it was not possible to determine embryonic and larval lethality. An estimate of the percentage of homozygous mutant animals that reach the adult stage was obtained by analyzing the offspring of mothers carrying a duplication; *unc-3 let-2(m); Dp[let-2(+)]/+* (see Materials and methods).

for glycines in the (Gly-X-Y) repeat domain of the molecule. A variety of mutations in the human $\alpha 5(\text{IV})$ collagen gene have been shown to cause the X-linked form of Alport syndrome, a progressive hereditary glomerulonephritis associated with hearing loss and ocular lesions (Tryggvason *et al.*, 1993). Recently, deletions that inactivate both the human $\alpha 5$ and $\alpha 6(\text{IV})$ collagen genes have been shown to cause benign smooth muscle tumors in conjunction with Alport syndrome (Zhou *et al.*, 1993). The localized defects that result from mutations in the $\alpha 5$ and $\alpha 6$ chains presumably reflect their relatively restricted tissue distribution compared with the ubiquitous $\alpha 1$ and $\alpha 2$ chains.

In this paper, the phenotypic effects of 17 mutations in the *C.elegans* $\alpha 2(\text{IV})$ collagen gene *let-2* are described. The mutations are molecularly identified and the relationship between the molecular nature and phenotypic consequences of the mutations is evaluated.

Results

Phenotypes of *let-2* mutant animals

Nineteen alleles of the *C.elegans* $\alpha 2(\text{IV})$ collagen gene *let-2* have been described (Meneely and Herman, 1979, 1981; Wood *et al.*, 1980; Cassada *et al.*, 1981). The mildest alleles are essentially wild-type at 15 and 20°C, but embryonic lethal at 25°C. The most severe alleles are embryonic lethal at all temperatures. At the non-permissive temperature, *let-2* embryos arrest development at approximately the twofold stage of embryogenesis and die before hatching. At lower temperatures many *let-2* animals com-

plete embryogenesis but arrest during larval development, never reaching the adult stage. A detailed analysis of the phenotypes caused by different *let-2* mutations was performed by scoring embryonic and larval lethality at 15, 20 and 25°C (Table I). In contrast to the wild-type N2 strain of *C.elegans*, which is viable and fertile at 15, 20 and 25°C, *let-2* mutants die as embryos at one or more of these temperatures. In Table I, the *let-2* alleles whose mutations have been molecularly identified are listed in order of increasing phenotypic severity at 20°C. The mildest alleles, *g25* and *b246*, are essentially wild-type at 15°C, exhibit moderate levels of larval lethality at 20°C, and are embryonic lethal at 25°C. Two alleles, *mn143* and *mn153*, are strict embryonic lethal at all temperatures. Most alleles show intermediate effects, with increasing severity as the temperature increases. In general, these results are consistent with different mutations causing different degrees of disruption of $\alpha 2(\text{IV})$ collagen function.

Some alleles show different degrees of temperature sensitivity. For example, *mn147* is the third most severe allele at 20°C, but ranks tenth at 15°C, indicating that its phenotype is less affected by increasing temperature than are most alleles. *g30* and *e1470* have very mild phenotypes at 15°C (at least 90% viable adults) but are very severe at 20°C (at least 90% embryonic lethality), indicating that they are more strongly affected by increasing temperature than most alleles. These differences in response to temperature suggest that the function of certain mutant $\alpha 2(\text{IV})$ chains is more severely disrupted by increases in temperature than are others.

mn101 and *mn114* could not be classified by the

Fig. 1. Locations of mutations in *C.elegans* $\alpha 2(\text{IV})$ collagen. The structure of the $\alpha 2(\text{IV})$ collagen chain is shown, with the N-terminal and C-terminal non-Gly-X-Y sequences heavily shaded. Interruptions in the Gly-X-Y repeat sequence are shown as vertical bars, with wider bars indicating longer interruptions. The alternatively spliced region of the $\alpha 2(\text{IV})$ gene (Sibley *et al.*, 1993) is also indicated, with light shading. Mutations are shown in boxes above and below the sequence. In each box, the allele name is in the top section, the amino acid change is in the center section, and the percentage embryonic lethality at 20°C in the bottom section.

procedure used for the other alleles since they have such low fertility, even at 15°C, that sufficient numbers of embryos could not be obtained from homozygous mothers. Analysis of progeny from hermaphrodites carrying a duplication that covers *let-2*, however, provides an estimate of the percentage of homozygous mutant animals that develop to the adult stage (see Materials and Methods). Approximately 8% of *mn101* animals reach the adult stage at 15°C, while no adults are seen at higher temperatures (Table I). This percentage of adults is similar to that seen with *mn129*; however, *mn129* adults are moderately fertile. Approximately 14% of *mn114* animals develop to the adult stage at 20°C, making it the third mildest allele by this criterion. Yet, *mn114* adults are virtually sterile at all temperatures, in contrast to the other mild alleles which produce highly fertile adults. These results indicate that certain mutations may affect the function of particular tissues more severely than others. The sterility of *mn114* strongly suggests that this mutation may disproportionately affect the function of the gonad.

Molecular identification of *let-2* mutations

We have identified point mutations in the *C.elegans* $\alpha 2(\text{IV})$ collagen chain in 17 *let-2* alleles by comparison with the wild-type $\alpha 2(\text{IV})$ collagen gene sequence (Sibley *et al.*, 1993). Figure 1 shows the location of the *let-2* mutations on the $\alpha 2(\text{IV})$ collagen chain, and Table II lists the exact location of the mutations and their phenotypes. All of the mutations are G to A nucleotide changes, the most common result of the EMS mutagenesis used to produce the alleles. All 17 mutations are within the triple helical (Gly-X-Y) domain of the $\alpha 2(\text{IV})$ chain, and all but two of the mutations cause glycine substitutions.

Glycine substitutions

Glycine substitutions probably inhibit folding of the protein into a triple helical conformation, as glycine is the only amino acid that can be easily accommodated in the center of the helix (Engel and Prockop, 1991; Kuivaniemi *et al.*, 1991). The 15 glycine substitution alleles cause

Table II. Molecular identification of *let-2* mutations

<i>let-2</i> allele	% Emb at 20°C ^a	Site of G to A substitution ^b	Amino acid substitution ^c
<i>mn114</i>	73 ^d	1440	Gly48Glu
<i>mn153</i>	100	1860	splice acceptor
<i>mn126</i>	100	4019	Ala366Thr
<i>mn109</i>	37	4679	Gly570Glu
<i>mn103</i>	98	4732	Gly588Arg
<i>mn151</i>	96	4732	Gly588Arg
<i>mn152</i>	50	4759	Gly597Arg
<i>mn101</i>	100 ^d	5038	Gly690Arg
<i>mn129</i>	100	5039	Gly690Glu
<i>mn143</i>	100	5232	Gly737Glu
<i>g30</i>	90	5651	Gly877Arg
<i>e1470</i>	94	5732	Gly904Arg
<i>mn139</i>	20	6307	Gly1003Glu
<i>g25</i>	2	6673	Gly1125Asp
<i>mn147</i>	7	6754	Gly1152Asp
<i>g37</i>	n.d.	7201	Gly1286Asp
<i>b246</i>	9	7201	Gly1286Asp

^aPercentage of animals that arrest during embryogenesis at 20°C, taken from Table I.

^bNumbering in nucleotides, with start of transcription designated +1 (Sibley *et al.*, 1993).

^cThe wild-type amino acid and its number (Sibley *et al.*, 1993) are followed by the mutant amino acid.

^dThese numbers are estimates of total embryonic plus larval lethality derived from *unc-3 let-2*; *mnDp1/+* hermaphrodites.

glycine to glutamate, arginine or aspartate changes, and produce a wide range of phenotypes. They include the mildest allele, *g25*, as well as one of the most severe alleles, *mn143* (Table I). The location of a glycine substitution may be one important factor in determining its effect on type IV collagen function. The *g25*, *b246*, *mn147* and *mn139* alleles have the mildest phenotypes, and are all located in the C-terminal one-third of the Gly-X-Y repeat domain (Figure 1). Alleles with more severe phenotypes are located in the central one-third of the Gly-X-Y repeat domain. The only glycine replacement allele in the N-terminal

one-third, *mn114*, has a relatively mild, but unusual, phenotype as noted above. The effects of different glycine substitutions in close proximity can be quite different, indicating that factors other than a simple C- to N-terminal severity gradient must affect phenotypes. For example, *mn129* and *mn143* are glycine to glutamate substitutions only 47 amino acids apart, but cause 6 and 100% embryonic lethality, respectively, at 15°C.

The *mn114* allele is unusual because it has a relatively mild effect on embryonic and larval development, but a very severe effect on fertility. The *mn114* mutation causes a glycine to glutamate substitution in the third Gly-X-Y triplet of the $\alpha 2(\text{IV})$ collagen chain. This defect at the extreme amino end of the molecule may interfere more severely with $\alpha 2(\text{IV})$ collagen function in the gonad than in other tissues. The other mutant that is essentially sterile, *mn101*, causes a degree of lethality similar to that caused by *mn129*, which is fertile. Both alleles cause substitutions for the same glycine, but it is replaced with arginine in *mn101* and with glutamate in *mn129* (Table II; Figure 1). Substitution with arginine or glutamate results in similar effects on embryonic and larval development, but different effects on fertility.

g30 and *e1470* were noted above because they are more severely affected by increasing temperature than most alleles. Both mutations cause glycine to arginine substitutions 27 amino acids apart from one another. These mutations may define a region of the molecule that is particularly sensitive to temperature.

We identified identical nucleotide alterations in the *mn151* and *mn103* alleles (Table II). However, the strains carrying these alleles show significantly different degrees of embryonic and larval lethality at 15°C (Table I). We repeated the molecular analyses of these alleles and obtained the same results. It is likely that one of these strains also contains a mutation in another gene, or a second mutation within *let-2* that we did not detect.

A mutation at position X

The *mn126* mutation results in an alanine to threonine substitution at position X of the first triplet following a four amino acid interruption (Thr-Phe-Asn-Asp-Gly-Ala-Pro). This change results in a more severe phenotype than most of the glycine substitution mutations, as only 18% of embryos survive to adult even at 15°C. In contrast to the glycine substitutions, it is not likely that the *mn126* mutation would have a major effect on triple helix assembly, since the side chains at positions X and Y project away from the triple helix. This alanine residue may be involved in interactions between type IV collagen molecules or with other basement membrane constituents. Thus, the severe phenotype of *mn126* may result from altered intermolecular interactions in the basement membrane.

A splicing mutation

The *mn153* allele has a G to A change in the terminal G of the 3' splice acceptor of intron 5 (TTGCAG), located near the N-terminus of the Gly-X-Y sequence (Figure 1; Table II). This terminal G is found at the 3' end of all eukaryotic introns (Senepathy *et al.*, 1990) and is normally required for proper splicing. Failure to splice this intron would result in production of a severely truncated $\alpha 2(\text{IV})$

chain, since the intron contains several in-frame stop codons. *mn153* causes strict, non-temperature-sensitive embryonic lethality. It is possible that *mn153* is a null allele, producing no functional $\alpha 2(\text{IV})$ collagen. However, it has recently been shown that the dinucleotide AG found at the 3' end of virtually all eukaryotic introns is not essential for splicing in *C.elegans* (Aroian *et al.*, 1993). In *C.elegans* *dpy-10* and *let-23* mutants, which have the same 3' splice acceptor G to A change as *mn153*, a significant amount of splicing was observed at or near the mutated splice site. In light of these results, it is possible that *mn153* produces aberrantly spliced gene products that could interfere with normal type IV collagen function.

All of the amino acid replacement alleles of *let-2* show dominant effects, such that some *let-2(m)/+* animals arrest or do not develop normally (P.Graham, M.Sibley and J.Kramer, unpublished results). If *mn153* is a null allele we would not expect it to show dominant effects. To investigate this possibility, we compared the embryonic lethality of progeny from *mn153/+* and *mnDf41/+* animals. *mnDf41* is a deficiency that removes the entire *let-2* gene. The expected survival to adulthood of progeny from *mn153/+* and *mnDf41/+* animals, excluding any dominant effects, is 75%. While 76% of the progeny of *mnDf41/+* animals survived to the adult stage ($n = 111$), only 55% of the progeny of *mn153/+* animals survived to adulthood. The survival of *mn153/+* progeny to adulthood (55%) is significantly lower than the expected value (75%; $P < 0.001$, $n = 83$), indicating that some *mn153/+* animals arrest during embryonic or larval development. Since animals with a single copy of *let-2* (*mnDf41/+*) develop normally, the *mn153* allele must produce some abnormal $\alpha 2(\text{IV})$ collagen that interferes with development of *mn153/+* animals. Clearly, *mn153* does not represent a true null allele.

Discussion

The mutations in the $\alpha 2(\text{IV})$ collagen gene, *let-2*, described in this study cause a wide range of phenotypes. At 20°C, for example, phenotypes range from nearly wild-type to 100% embryonic lethality, with many mutations showing intermediate levels of embryonic lethality. The phenotypic diversity of *let-2* alleles permits an analysis of the relationship between the nature of a particular $\alpha 2(\text{IV})$ collagen mutation and the characteristics of the resulting phenotype. It should be noted that none of these mutations completely abolish production of $\alpha 2(\text{IV})$ collagen, so their phenotypes may result from some combination of the lack of wild-type product and the presence of interfering mutant product. We will discuss several attributes of the *C.elegans* $\alpha 2(\text{IV})$ chain and type IV molecule that may account for the phenotypic differences observed in these *let-2* mutants.

Glycine substitution mutations

Most of the mutations identified in the *C.elegans* $\alpha 2(\text{IV})$ collagen gene result in glycine substitutions in the Gly-X-Y repeats of the molecule, creating an additional interruption in the mutant $\alpha 2(\text{IV})$ chain. Non-fibrillar collagens, which include type IV, normally contain interruptions of their Gly-X-Y repeats. However, analyses of mutations in human $\alpha 5(\text{IV})$ collagen (Tryggvason *et al.*, 1993), *C.elegans* $\alpha 1$ (Guo *et al.*, 1991) and $\alpha 2(\text{IV})$ collagen, and

Fig. 2. Calculated relative stability of *C.elegans* type IV collagen ($\alpha_1\alpha_2$ stoichiometry). Thermal stability of successive tripeptides is plotted along the triple helix proceeding from the N-terminus to the C-terminus of the molecule, calculated using a window of 11 triplets (Bachinger and Davis, 1991). The dashed line indicates the average relative stability of the ($\alpha_1\alpha_2$) molecule. The locations of mutations in α_2 (IV) collagen chain are indicated.

C.elegans cuticle collagens (Johnstone *et al.*, 1992; Kramer and Johnson, 1993; Levy *et al.*, 1993) demonstrate that glycine substitutions can severely disrupt the function of non-fibrillar collagens. More than 60 different glycine substitutions in the human type I collagen genes have been identified in patients suffering from osteogenesis imperfecta (Kuivaniemi *et al.*, 1991; Byers and Steiner, 1992). Glycine substitutions prevent the zipper-like folding of the triple helix because only glycine, the smallest amino acid, can easily be accommodated in the center of the helix (Engel and Prockop, 1991). Glycine substitutions in type I collagen result in increased post-translational modification N-terminal to the site of the mutation, decreased thermal stability of molecules containing abnormal chains, and a general reduction in the efficiency of secretion (Kuivaniemi *et al.*, 1991). It is likely that glycine substitutions in type IV collagen would have similar effects on post-translational processing and secretion, but as yet there are no studies of mutant type IV proteins.

Correlation of thermal stability and phenotype

Since collagen assembly is a temperature-sensitive process, the thermal stability of the region of the triple helix that contains a glycine substitution may influence the phenotypic consequences of that mutation. Based on studies using synthetic peptides, the thermal stabilities of different Gly-X-Y tripeptide units have been classified (Dolz and Heidemann, 1986; Bachinger and Davis, 1991; Bachinger *et al.*, 1993). In Figure 2, the relative stability of *C.elegans* type IV collagen is plotted and the location of the α_2 (IV) chain mutations are indicated. Most of the low points of thermal stability are in the N-terminal half of the triple helical domain, primarily because most (22 of 35) interruptions of the Gly-X-Y repeats are located in

this portion of the molecule. Mammalian type IV collagen also has more Gly-X-Y repeat interruptions in the N-terminal half of the molecule (Yamada and Kuhn, 1993). In addition, the C-terminal half of the *C.elegans* type IV molecule has fewer low-stability triplets than the N-terminal half. Most notably, there is a stretch of 97 Gly-X-Y repeats in the C-terminal half of the α_1 (IV) chain that includes no low-stability triplets (triplets 318–415).

let-2 mutations located in low-stability regions of the type IV collagen molecule tend to have more severe phenotypes than those in high-stability regions (Figure 2). *mn143*, *mn126*, *mn129* and *mn101* are located in regions of relatively low stability, and all cause severe phenotypes. Alleles that cause phenotypes of intermediate severity tend to be in regions that are close to the mean thermal stability value. The mildest alleles, *mn147*, *g25* and *b246*, are located in areas of relatively high stability. These results suggest that α_2 (IV) collagen mutations in regions of low calculated thermal stability are not as well tolerated as mutations in high-stability regions. Mutations in these low-stability regions could result in molecules that are too unstable to assemble at normal *C.elegans* growth temperatures, resulting in embryonic lethality even at 15°C. Differences in the thermal stability values do not account for the phenotypic severity of all *let-2* mutants, however. *mn152* is located in a region of very low stability and *mn109* is in a high-stability region, but both cause moderately severe phenotypes. Some other factors must influence the phenotypes produced by these mutations.

Nature of glycine substitution

The nature of the amino acid substituted for glycine may also be a factor in determining phenotype. *mn129* and *mn101* alter the same glycine, to glutamate and arginine,

respectively. Both have similarly severe effects on development to the adult stage, but *mn101* has a much more severe effect on adult fertility. Arginine has a larger side chain than glutamate and may cause greater interference with triple helix assembly. It is also possible that substitution of a positively charged residue is more destabilizing than substitution of a negatively charged residue. Different effects due to substituting different amino acids for a particular glycine have also been seen in human $\alpha 1(\text{I})$ collagen; for example, Gly883Asp causes lethal osteogenesis imperfecta while Gly883Ser causes mild osteogenesis imperfecta (Cohn *et al.*, 1990; P.Byers, personal communication).

Clustering of type IV collagen mutations

let-2 mutations are not randomly distributed along the length of the molecule. Of the 17 mutations characterized, two pairs are identical nucleotide substitutions and a third pair affects adjacent nucleotides in the same codon. Clustering of mutations was also noted in the *C.elegans* $\alpha 1(\text{IV})$ collagen gene *emb-9*, where two of the three mutations characterized were the same nucleotide change and the third was located just 18 nucleotides away (Guo *et al.*, 1991). No mutations were identified in the non-collagenous NC1 domain of either gene, although a random distribution would give 2.6 (13%) expected NC1 mutations. This situation contrasts strongly with analyses of mutations in the human $\alpha 1$ and $\alpha 2(\text{I})$ collagen chains in which multiple independent mutations affecting the same glycine are relatively uncommon (Kuivaniemi *et al.*, 1991; Byers and Steiner, 1992; P.Byers, personal communication).

There are two simple explanations for the non-random distribution of mutations in the *C.elegans* type IV collagen genes. The glycine substitution mutations that we have analyzed cause a wide range of phenotypes and it is likely that other glycine substitutions can have milder or stronger effects than those characterized. Mutations that did not noticeably affect development or fertility would not be identified in genetic screens, and could contribute to the non-randomness. In addition, all of the mutations in the *C.elegans* $\alpha 1$ and $\alpha 2(\text{IV})$ collagen genes cause some degree of dominant embryonic and/or larval lethality (P.Graham, M.Sibley, M.Gupta and J.Kramer, unpublished results). Indeed, some of these mutations have temperature-sensitive dominant lethal phenotypes. It is possible that some glycine substitutions could cause strict dominant lethal or sterile phenotypes and would, therefore, be lost in genetic screens. The mutations that we have analyzed were selected to have phenotypes that were not too mild or too severe, and substitutions for only a small number of glycine residues may produce phenotypes within this range.

Alteration of the amino acid at position X of a Gly-X-Y repeat

The relatively severe phenotype of *mn126*, which causes an alanine to threonine change at position X, is surprising because alterations at positions X and Y have not generally been identified in mutant collagens. Amongst the hundreds of mutant human collagens of all types that have been analyzed, just one arginine to glutamine change at position Y in $\alpha 2(\text{I})$ (Phillips *et al.*, 1990) and three arginine to

Fig. 3. Interallelic complementation between $\alpha 2(\text{IV})$ collagen mutations as a function of the distance (in amino acid residues) between the mutations in each *trans* heterozygote. Complementation scores for these mutations were taken from Meneely and Herman (1981), and were assigned as follows: 0 = embryonic lethal at 20°C; 1 = sterile at 20 and 25°C; 2 = slightly fertile at 20°C, sterile at 25°C; 3 = fertile at 20°C, sterile at 25°C; 4 = fertile at 20 and 25°C (wild-type). N + N refers to heterozygotes with two mutations in the N-terminal half of the chain, N + C refers to heterozygotes with one mutation in the N-terminal half and one mutation in the C-terminal half, and C + C refers to heterozygotes with two mutations in the C-terminal half of the chain.

cysteine alterations at position Y in $\alpha 1(\text{II})$ (Ala-Kokko *et al.*, 1990; Chan *et al.*, 1993; Williams *et al.*, 1993) have been reported. These mutations result in relatively mild phenotypes. The type II mutations that insert cysteine may affect function due to formation of ectopic disulfide bonds, rather than loss of the particular amino acid at position Y. Presumably most substitutions at positions X and Y have little or no effect on the phenotype and therefore are not detected. We can think of two reasons why the *mn126* alanine to threonine substitution at position X may cause a strong change in phenotype. It is in the first Gly-X-Y triplet on the carboxyl side of a four amino acid interruption and changes a triplet of highest predicted stability (Gly-Ala-Pro) to one of lower stability (Gly-Thr-Pro) (Bachinger and Davis, 1991). The combination of an interruption and an adjacent lower-stability triplet may interfere with triple helix assembly. Such effects would not occur in the fibrillar collagens, which lack interruptions, possibly explaining the paucity of X and Y mutations in human types I, II and III collagen. Alternatively, *mn126* may alter a region of the molecule that is involved in some critical intermolecular interaction. Such an interaction could be between type IV collagen molecules, or between type IV and some other matrix or cell surface component.

Interallelic complementation

An unusual feature of the $\alpha 2(\text{IV})$ collagen gene *let-2* is its complex pattern of interallelic complementation (Meneely and Herman, 1981). Every *let-2* allele complements at least one other allele to some extent, such that *trans* heterozygote animals have milder phenotypes than either homozygote. If each *C.elegans* type IV molecule contains one $\alpha 2$ chain, as in vertebrate type IV, then complementation between different $\alpha 2$ chain alleles can only result from some type of intermolecular interaction. This interaction may be between type IV collagen molecules, or between type IV collagen and other basement

membrane components. When the interallelic complementation scores for pairs of *let-2* alleles are plotted against the distance between the mutations (Figure 3), a weak positive correlation is seen. No two alleles show full complementation to a wild-type phenotype unless they are at least 134 amino acids apart. All alleles of >800 amino acids apart show full complementation, while alleles located <47 amino acids apart show no complementation. In general, animals carrying two mutations in the C-terminal half of the chain show greater complementation than animals carrying two mutations in the N-terminal half. This greater complementation may be due to the milder homozygous phenotypes that result from mutations in the C-terminal half.

The relationship between interallelic complementation and distance may reflect lateral interactions between *C.elegans* type IV collagen molecules. The tendency for mutations that are more distant from one another to show higher complementation could result from a simple side-by-side association of type IV collagen molecules with little stagger between adjacent molecules. In animals homozygous for a single $\alpha 2(\text{IV})$ mutation, side-by-side packing of type IV molecules may be inhibited due to the proximity of the triple helix defects. In *trans* heterozygotes for two different mutations, side-by-side packing may be more normal when the triple helix defects in adjacent molecules are separated by a sufficient distance. Lateral interactions between type IV molecules in mammalian basement membranes are complex, forming a network of superhelically twisted, irregularly branched strands of variable diameter (Yurchenco and Ruben, 1987). The scatter in the plot of complementation versus distance may indicate that there are several types of interactions between type IV collagen molecules in *C.elegans* basement membranes.

Materials and methods

Nematode strains

Maintenance and handling of *C.elegans* strains were as described by Brenner (1974). The *let-2* alleles analyzed were originally isolated by Meneely and Herman (1979, 1981), Wood *et al.* (1980) and Cassada *et al.* (1981) following ethylmethanesulfate mutagenesis, and selection for lethal or sterile phenotypes. The *let-2* alleles *e1470*, *mn101*, *mn103*, *mn109*, *mn111*, *mn114*, *mn126*, *mn129*, *mn131*, *mn139*, *mn143*, *mn147*, *mn151*, *mn152* and *mn153* were maintained at 20°C in strains of genotype *mnDp1/+*; *unc-3 let-2*. The *mnDp1* duplication carries *unc-3+* and *let-2+*, and is attached to chromosome V (Meneely and Herman, 1979). The *unc-3(e151)* mutation confers an uncoordinated phenotype, and is used as a visible marker for the absence of *mnDp1*. Strains carrying the *let-2* alleles *b246*, *g25*, *g30*, and *g37* were maintained as homozygotes, without *mnDp1*, at 15°C.

Scoring *let-2* lethal phenotypes

let-2 mutant animals were scored for viability at 15, 20 and 25°C. To examine lethality, strains of genotype *unc-3 let-2* were maintained at 15°C. To obtain embryos grown at non-permissive temperatures, L4 or young adult hermaphrodites were picked to temperature-equilibrated plates. Embryos that had begun development at the permissive temperature were eliminated by transferring adult hermaphrodites to a fresh plate after 3 h and discarding the first plate. Embryos laid within a 24 h period were picked to a fresh temperature-equilibrated plate and counted. One day later they were examined to determine the number that had failed to hatch (Emb). They were scored again after 3–4 days (25°C), 5–6 days (20°C) or 7–8 days (15°C) to determine the number that had developed to adults (Ad). Embryos not accounted for as unhatched or as adults were considered to have arrested during larval development (Acc). The number of Acc animals could be reliably

inferred, since the N2 and *unc-3* controls have very low percentages of Acc larvae at all temperatures, and the *let-2* animals are slow-moving (Unc), having a limited ability to escape from the plate. At least 100 embryos were scored for each strain at each temperature. Strains carrying *mn101*, *mn114*, *mn143* or *mn153* could not be examined by this procedure. In contrast to previous observations (Meneely and Herman, 1981), we find *mn139* to be fertile at 15°C.

To test for fertile homozygotes of *mn101*, *mn114*, *mn143* or *mn153*, we looked for fertile Unc progeny from *mnDp1/+*; *unc-3 let-2(m)* animals grown at 15°C. More than 500 progeny were examined for each strain. No larval or adult Unc progeny were seen with strains carrying *mn143* or *mn153*, indicating that they are embryonic lethal. Unc adults homozygous for *mn101* or *mn114* were identified, but they produced no or very few embryos.

Estimating *mn101* and *mn114* development to adult

Embryos produced by *mnDp1/+*; *unc-3 let-2(mn101 or mn114)* hermaphrodites at 15, 20 and 25°C were transferred and counted as described above. Approximately 200 embryos were analyzed for each strain at each temperature. The estimate of the percentage of *+/+*; *unc-3 let-2(m)* animals that developed to the adult stage was calculated as: (Unc adults/total embryos) $\times$ 400.

Identification of nucleotide alterations in *let-2* alleles

Genomic DNAs were prepared from *let-2* mutant strains as previously described (Kramer *et al.*, 1988). The *b246*-, *g25*-, *g30*- and *e1470*-containing strains were grown as homozygotes, while all other alleles were maintained in strains carrying *mnDp1* (see above). Nucleotide alterations in *let-2* alleles were located using hydroxylamine modification and piperidine cleavage, as described in Montandon *et al.* (1989). PCR fragments of the $\alpha 2(\text{IV})$ collagen gene were generated so as to overlap one another by 200–300 bp. Seven 1–2 kb fragments were used to span the *let-2* gene, including 170 bp 5' of the start codon and 330 bp 3' of the stop codon. Point mutations were identified by direct sequencing of PCR fragments in the areas of mismatch. For mutations maintained with *mnDp1*, the mutant and wild-type nucleotides were detected in a ratio of ~2:1, reflecting the stoichiometry of mutant and wild-type chromosomes [*mnDp1/+*; *let-2(m)*]. The Sequenase dideoxy sequencing kit was used according to the procedure recommended by the manufacturer (United States Biochemical Corp.), but with a modified annealing procedure (Kusukawa *et al.*, 1990). We were not able to identify nucleotide alterations in the *mn111* or *mn131* alleles.

Strains carrying *mn151* and *mn103* were found to have somewhat different phenotypes, but sequence analysis indicated that they have the identical nucleotide change. We reconfirmed the presence of the same nucleotide alteration in both of these mutations by reisolating genomic DNA from homozygous mutant animals, and sequencing the PCR fragment from the region previously identified as containing the mismatch with wild-type DNA.

Analysis of *mn153* dominant effects

Survival to the adult stage was scored for the offspring of *unc-3 let-2(mn153)/unc-7* and *mnDf41/unc-7* hermaphrodites at 20°C. To construct *unc-3(e151) let-2(mn153)/unc-7(e5)*, N2 males were crossed to *mnDp1/+*; *unc-3 let-2(mn153)* hermaphrodites. Non-Unc male cross progeny of genotype *mnDp1/+*; *unc-3 let-2(mn153)/0* were crossed to *unc-7* hermaphrodites. Individual non-Unc L4 hermaphrodite cross progeny were picked to separate plates and allowed to lay eggs for 1 day. Each hermaphrodite was then transferred to a new plate and crossed with N2 males. Self-progeny plates were examined after 1 day for embryos that failed to hatch, indicating that the hermaphrodite carried *let-2(mn153)*. Cross plates were examined after 3 days for wild-type males, indicating that the hermaphrodite carried *mnDp1*. Hermaphrodites with ~1/4 unhatched self-progeny and no wild-type male cross progeny were considered to be of genotype *unc-3 let-2(mn153)/unc-7*. A strain of genotype *mnDf41/unc-7* was constructed in a similar manner.

Calculation of sequence specific relative stability of *C.elegans* type IV collagen

The sequence specific relative stability of the collagen triple helix was calculated using a program kindly provided by Dr Hans Peter Bachinger, Oregon Health Sciences University, Portland, OR. This program used the empirical relative stabilities of Bachinger and Davis (1991), with the exception of interruptions in the Gly-X-Y, which were assigned a value of -1.5. In this program, the typical collagenous tripeptides Gly-Pro-Ala and Gly-Pro-Hyp are assigned the highest stability value (2.0), and tripeptides containing charged residues, such as Gly-Ala-Lys and

Gly-Asp-Arg, are assigned a lower stability value (0.0). An $[\alpha 1(\text{IV})]_2 \alpha 2(\text{IV})$ stoichiometry was designated, and the first tripeptide unit of the helix was used as the starting point for consecutive numbering.

Acknowledgements

We thank Hans Peter Bachinger for assistance in generating helix stability plots and Peter Byers for illuminating discussions. We thank Robert Herman and Iva Greenwald for providing some of the *C.elegans* strains used in these studies. Some strains were provided by the *Caenorhabditis* Genetics Center, which is funded by the National Institutes of Health National Center for Research Resources. This work was supported by a grant from the National Institutes of Health (HD27211) to J.M.K.

References

- Ala-Kokko, L., Baldwin, C.T., Moskowitz, R.W. and Prockop, D.J. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 6565–6568.
- Albertson, D.G. and Thomson, J.N. (1976) *Phil. Trans. R. Soc. Lond. B*, **275**, 299–325.
- Aroian, R.V., Levy, A.D., Koga, M., Ohshima, Y., Kramer, J.M. and Sternberg, P.W. (1993) *Mol. Cell. Biol.*, **13**, 626–637.
- Bachinger, H.P. and Davis, J.M. (1991) *Int. J. Biol. Macromol.*, **13**, 152–156.
- Bachinger, H.P., Morris, N.P. and Davis, J.M. (1993) *Am. J. Med. Genet.*, **45**, 152–162.
- Blumberg, B., MacKrell, A.J. and Fessler, J.H. (1988) *J. Biol. Chem.*, **263**, 18328–18337.
- Brazel, D., Oberbaumer, I., Dieringer, H., Bable, W., Glanville, R.W., Duetzmann, R. and Kuhn, K. (1987) *Eur. J. Biochem.*, **168**, 529–536.
- Brenner, S. (1974) *Genetics*, **77**, 71–94.
- Byers, P.H. and Steiner, R.D. (1992) *Annu. Rev. Med.*, **43**, 269–282.
- Cassada, R., Isnenghi, E., Culotti, M. and von Ehrenstein, G. (1981) *Dev. Biol.*, **84**, 193–205.
- Chan, D., Taylor, T.K.F. and Cole, W.G. (1993) *J. Biol. Chem.*, **268**, 15238–15245.
- Charonis, A.S. and Tsilibary, E.C. (1990) In Adair, W.S. and Mecham, R.P. (eds), *Organization and Assembly of Plant and Animal Extracellular Matrix*. Academic Press, New York, pp. 85–117.
- Cohn, D.H., Starman, B.J., Blumberg, B. and Byers, P.H. (1990) *Am. J. Hum. Genet.*, **46**, 591–601.
- Desjardins, M. and Bendayan, M. (1991) *J. Cell Biol.*, **113**, 689–700.
- Dolz, R. and Heidemann, E. (1986) *Biopolymers*, **25**, 1069–1080.
- Engel, J. and Prockop, D.J. (1991) *Annu. Rev. Biophys. Biophys. Chem.*, **20**, 137–152.
- Exposito, J.-Y., D'Alessio, M., DiLiberto, M. and Ramirez, F. (1993) *J. Biol. Chem.*, **268**, 5249–5254.
- Gunwar, S., Saus, J., Noelken, M.E. and Hudson, B.G. (1990) *J. Biol. Chem.*, **265**, 5466–5469.
- Guo, X. and Kramer, J.M. (1989) *J. Biol. Chem.*, **264**, 17574–17582.
- Guo, X., Johnson, J.J. and Kramer, J.M. (1991) *Nature*, **349**, 707–709.
- Hostikka, S.L. and Tryggvason, K. (1988) *J. Biol. Chem.*, **263**, 19488–19493.
- Hostikka, S.L., Eddy, R.L., Byers, M.G., Hoyhyta, M., Shows, T.B. and Tryggvason, K. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 1606–1610.
- Johnstone, I.L., Shafi, Y. and Barry, J.D. (1992) *EMBO J.*, **11**, 3857–3863.
- Kramer, J.M. (1994) *FASEB J.*, **8**, 329–336.
- Kramer, J.M. and Johnson, J.J. (1993) *Genetics*, **135**, 1035–1045.
- Kramer, J.M., Johnson, J.J., Edgar, R.S., Basch, C. and Roberts, S. (1988) *Cell*, **55**, 555–565.
- Kuivaniemi, H., Tromp, G. and Prockop, D. (1991) *FASEB J.*, **5**, 2052–2060.
- Kusukawa, N., Uemori, T., Asada, K. and Kato, I. (1990) *BioTechniques*, **9**, 66–72.
- Levy, A.D., Yang, J. and Kramer, J.M. (1993) *Mol. Biol. Cell*, **4**, 803–817.
- Meneely, P.M. and Herman, R.K. (1979) *Genetics*, **92**, 99–115.
- Meneely, P.M. and Herman, R.K. (1981) *Genetics*, **97**, 65–84.
- Montandon, A.J., Green, P.M. and Bentley, D.R. (1989) *Nucleic Acids Res.*, **17**, 3347–3358.
- Muthukumar, G., Blumberg, B. and Kurkinen, M. (1989) *J. Biol. Chem.*, **264**, 6310–6317.
- Paulsson, M. (1993) In Rohrbach, D.H. and Timpl, R. (eds), *Molecular and Cellular Aspects of Basement Membranes*. Academic Press, New York, pp. 177–187.
- Pettitt, J. and Kingston, I.B. (1991) *J. Biol. Chem.*, **266**, 16149–16156.
- Phillips, C.L., Shrago-Howe, A.W., Pinnell, S. and Wenstrup, R.J. (1990) *J. Clin. Invest.*, **86**, 1723–1728.
- Saus, J., Wieslander, J., Langeveld, J.P.M., Quinones, S. and Hudson, B.G. (1988) *J. Biol. Chem.*, **263**, 13374–13380.
- Saus, J., Quinones, S., MacKrell, A., Blumberg, B., Muthukumar, G., Pihlajaniemi, T. and Kurkinen, M. (1989) *J. Biol. Chem.*, **264**, 6318–6324.
- Senepathy, P., Shapiro, M.B. and Harris, N.L. (1990) *Methods Enzymol.*, **183**, 252–278.
- Sibley, M.H., Johnson, J.J., Mello, C.C. and Kramer, J.M. (1993) *J. Cell Biol.*, **123**, 255–264.
- Timpl, R. (1989) *Eur. J. Biochem.*, **180**, 487–502.
- Tryggvason, K., Zhou, J., Hostikka, S.L. and Shows, T.B. (1993) *Kidney Int.*, **43**, 38–44.
- White, J.G., Southgate, E., Thomson, J.N. and Brenner, S. (1976) *Phil. Trans. R. Soc. Lond. B*, **275**, 327–348.
- White, J.G., Southgate, E., Thomson, J.N. and Brenner, S. (1986) *Phil. Trans. R. Soc. Lond. B*, **314**, 1–340.
- Williams, C.J., Considine, E.L., Knowlton, R.G., Reginato, A., Neumann, G., Harrison, D., Buxton, P., Jimenez, S. and Prockop, D.J. (1993) *Hum. Genet.*, **92**, 499–505.
- Wood, W.B., Hecht, R., Carr, S., Vanderslice, R., Wolf, N. and Hirsh, D. (1980) *Dev. Biol.*, **74**, 446–469.
- Yamada, Y. and Kuhn, K. (1993) In Rohrbach, D.H. and Timpl, R. (eds), *Molecular and Cellular Aspects of Basement Membranes*. Academic Press, New York, pp. 121–146.
- Yurchenco, P.D. and Ruben, G.C. (1987) *J. Cell Biol.*, **105**, 2559–2568.
- Yurchenco, P.D. and Schittny, J.C. (1990) *FASEB J.*, **4**, 1577–1590.
- Zhou, J., Hertz, J.M., Leinonen, A. and Tryggvason, K. (1992) *J. Biol. Chem.*, **267**, 12475–12481.
- Zhou, J., Mochizuki, T., Smeets, H., Antignac, C., Laurila, P., de Paepe, A., Tryggvason, K. and Reeders, S.T. (1993) *Science*, **261**, 1167–1169.

Received on April 22, 1994