Molecular genetic analysis of the glycosyltransferase Fringe in *Drosophila*

Trudy Correia*, Venizelos Papayannopoulos*†, Vladislav Panin*‡, Pamela Woronoff*, Jin Jiang§¶, Thomas F. Vogt‖ and Kenneth D. Irvine*††**

*Howard Hughes Medical Institute, Waksman Institute and Department of Molecular Biology and Biochemistry, Rutgers University, Piscataway, NJ 08854; §Howard Hughes Medical Institute, Department of Genetics and Development, Columbia University College of Physicians and Surgeons, New York, NY 10032; and ‖Department of Molecular Biology, Princeton University, Princeton NJ 08544

Edited by Michael S. Levine, University of California, Berkeley, CA, and approved March 25, 2003 (received for review February 11, 2003)

Fringe proteins are β1,3-*N*-acetylglucosaminyltransferases that modulate signaling through Notch receptors by modifying O-linked fucose on epidermal growth factor domains. Fringe is highly conserved, and comparison among 18 different Fringe proteins from 11 different species identifies a core set of 84 amino acids that are identical among all Fringes. Fringe is only distantly related to other glycosyltransferases, but analysis of the predicted *Drosophila* proteome identifies a set of four sequence motifs shared among Fringe and other putative β1,3-glycosyltransferases. To gain functional insight into these conserved sequences, we genetically and molecularly characterized 14 point mutations in *Drosophila fringe*. Most nonsense mutations act as recessive antimorphs, raising the possibility that Fringe may function as a dimer. Missense mutations identify two distinct motifs that are conserved among β1,3-glycosyltransferases, and that can be modeled onto key motifs in the crystallographic structures of bovine β1,4-galactosyltransferase 1 and human glucuronyltransferase I. Other missense mutations map to amino acids that are conserved among Fringe proteins, but not among other glycosyltransferases, and thus may identify structural motifs that are required for unique aspects of Fringe activity.

Fringe (FNG) proteins modulate signaling through Notch receptors (reviewed in refs. 1 and 2). Notch receptors are transmembrane proteins that are activated by the binding of DSL ligands to epidermal growth factor repeats in the Notch extracellular domain (reviewed in ref. 3). Two ligands are known in *Drosophila*, Delta and Serrate, whereas mammals possess three Delta-related ligands and two Serrate-related ligands. Notch signaling plays important roles during the development of most tissues in the body, and human genetic diseases have been associated with mutations in genes in the Notch pathway, including cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL), leukemia, spondylocostal dystosis, and Alagille (reviewed in ref. 4). Although the Notch pathway mediates a wide range of cell fate decisions, the influence of *fringe* (*fng*) genes on Notch signaling has been best studied during the development of the *Drosophila* wing. In the developing wing imaginal disk, FNG inhibits the ability of Serrate to activate Notch, and potentiates the ability of Delta to activate Notch. These effects of FNG position a stripe of Notch activation along the border between dorsal and ventral cells, which is essential for the subsequent patterning, growth, and compartmentalization of the wing (reviewed in refs. 5 and 6).

Drosophila and vertebrate FNG proteins are highly conserved and possess an unusual β1,3-*N*-acetylglucosaminyltransferase activity that elongates O-linked fucose within EGF domains (7, 8). Although FNG is not closely related to other glycosyltransferases, it does share a few short sequence motifs with certain other glycosyltransferases (9), and we report here that it is most closely related to members of a β1,3-glycosyltransferase (β3GT) superfamily. These enzymes use a variety of sugar donors and acceptors, but all transfer the sugar from a UDP-sugar donor, creating a β linkage between the 1 carbon of the donor and the 3 carbon of the acceptor. With the exception of site-directed mutations in a DDD motif of FNG (7, 8, 10, 38), the functional requirements for conserved sequence motifs in FNG or other β3GTs have not been assessed. Moreover, the only crystallographic structure for a β3GT determined to date is that of human glucuronyltransferase I (GlcAT-I), which does not share significant primary sequence similarity with FNG.

Here, we employ two complementary approaches to identify functionally important amino acids within FNG: sequence conservation, and genetic and molecular characterization of mutations in *Drosophila fng*. Because of the great interest in the Notch pathway and its modulation by FNG, sequence information is available for FNG from a wider variety of organisms than for any other eukaryotic glycosyltransferase, and we report here the sequence of an additional insect *fng* gene. Genetically, *fng* is the best characterized glycosyltransferase in *Drosophila*, which is, in turn, one of the best eukaryotic systems for genetic analysis. The distinct requirements for *fng* in *Drosophila* have facilitated the isolation of a large number of randomly induced point mutations, which provides for an unbiased means of identifying key amino acid residues. By combining molecular analysis of these mutations with global comparisons to other putative β3GTs, and by modeling three conserved motifs onto the crystallographic structures of β1,4-galactosyltransferase 1 (β4GalT1) and GlcAT-I, we identify key motifs within FNG that are essential for function and are conserved among β3GTs, as well other motifs that are essential for function and are unique to FNG.

Methods

***Drosophila* Stocks.** $fng^{35UZ\text{-}1}$, fng^{52}, fng^{80}, fng^{129}, fng^{2}, fng^{7}, fng^{8}, fng^{13}, and fng^{14} have been described (11). fng^{1}, fng^{3}, fng^{5}, and fng^{11} were isolated in the same screen, although they were not identified by name in that publication. fng^{L19}, fng^{L73}, fng^{L81}, fng^{L83}, and fng^{M69} were isolated in a screen for mutations that affect imaginal development in homozygous clones (12). All alleles were tested for temperature sensitivity by test crosses at 18, 25, and 29°C. Except in the case of fng^{M69}, all additional crosses were then conducted at 25°C.

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: FNG, Fringe protein; EMS, ethyl methanesulfonate; β3GT, β1,3-glycosyltransferase; β4GalT, β1,4-galactosyltransferase; GlcAT, glucuronyltransferase.

Data deposition: The sequence reported in this paper has been deposited in the GenBank database [accession no. AY228759 (*P. coenia fng*)].

†Present address: Department of Biochemistry and Molecular Biophysics, University of California, San Francisco, CA 94143.

‡Present address: Department of Biochemistry and Biophysics, Texas A & M University, College Station, TX 77843.

¶Present address: Center for Developmental Biology and Department of Pharmacology, University of Texas Southwestern Medical Center, Dallas, TX 75390.

**Present address: Department of Molecular Pharmacology, Merck Research Laboratories, West Point, PA 19486.

††To whom correspondence should be addressed. E-mail: irvine@waksman.rutgers.edu.

www.pnas.org/cgi/doi/10.1073/pnas.1131007100

Fig. 1. Sequence identity among FNG proteins. The predicted amino acid sequences of four insect FNG proteins [*Drosophila melanogaster* (D.m.), *P. coenia* (P.c.), *Schistocerca gregaria* (S.g.), and *Anopheles gambiae* (A.g.)], aligned by the PILE UP program of GCG 10. Nonconserved amino- and carboxyl-terminal regions are not shown. This sequence corresponds to amino acids 149–410 of *Drosophila* FNG. Black boxes identify amino acids identical among all cloned FNG proteins, including 14 vertebrate FNG proteins. An alignment including these vertebrate genes is presented in Fig. 5. Gray boxes identify amino acids identical among all insect FNGs, but not all vertebrate FNGs. Lines above the sequence letters demarcate the conserved sequence motifs identified in Fig. 2. Letters above the sequence letters identify the locations of missense mutations in *Drosophila* FNG, and asterisks (*) above the sequence letters identify the locations of nonsense mutations (see Table 1).

Cloning of *P. coenia fng.* A cDNA library from *Precis coenia* (gift of S. Carroll, Howard Hughes Medical Institute, University of Wisconsin, Madison) was screened by low stringency hybridization using ^{32}P-labeled *Drosophila fng* cDNA 103 (11) as a probe. Labeled plaques were isolated and characterized by Southern blotting and DNA sequencing.

Cloning and Sequencing of *Drosophila fng* Mutations. Homozygous mutant animals were identified by using the GFP-marked chromosome *TM3*, *Kr-Gal4 UAS-GFP* (13). 50 Non-GFP-expressing animals were ground in 100 mM Tris·HCl, pH 7.5/100 mM EDTA/100 mM NaCl/0.5% SDS, incubated at 65°C for 30 min, and protein and RNA were precipitated by addition of 400 μl of 1.4 M KOAc/4.3 M LiCl, followed by incubation on ice for 10 min and centrifugation for 15 min. Supernatant (500 μl) was transferred to a new tube, and DNA was precipitated by the addition of 300 μl of isopropanol and centrifugation for 15 min, washed in 70% ethanol, and resuspended in 75 μl of TE buffer.

The coding portions and flanking DNA of the seven exons of *fng* were amplified by PCR using Herculase polymerase (Stratagene). Both strands were completely sequenced for each exon and any ambiguities or mutations were resequenced. Primer sequences are available in *Supporting Methods*, which is published as supporting information on the PNAS web site, www.pnas.org.

Sequence Comparisons and Database Analysis. FNG proteins and glycosyltransferase families B and F were aligned by using the PILE UP algorithm of GCG 10. *fng*-related genes were identified using the National Center for Biotechnology Information PSI-BLAST search engine (www.ncbi.nlm.nih.gov/blast). All of the genes identified in Fig. 2 *A–H* were identified after five iterations, and no additional genes were identified through additional iterations. Each related gene was then used as a probe in standard BLASTP against the predicted *Drosophila* proteome (14), using the *Drosophila* genome database at the Berkeley *Drosophila* Genome Project (www.fruitfly.org/blast). PSI-BLAST E values and alignments, multiple sequence alignments for families b and f, and BLASTP E values are available (see Fig. 4, which is published as supporting information on the PNAS web site).

Statistics. PRISM software (GraphPad, San Diego) was used to compare phenotypic scores by two-tailed, unpaired *t* test with Welch's correction for unequal variances.

Results and Discussion

Sequence Similarity Among *fng* Genes. The conservation of amino acids between distantly related species provides an indication that these amino acids make an important contribution to normal protein function. In work that preceded the initial identification of vertebrate *fng*-related genes, a homologue of *Drosophila fng* was cloned from the butterfly species *P. coenia*, which diverged from *Drosophila* ≈300 million years ago. Although the sequences of 14 different vertebrate *fng* genes have now been reported, only two additional insect *fng* genes have been reported to date, and both are incomplete at the amino terminus. In Fig. 1, the predicted amino acid sequence of *P. coenia fng* is aligned together with the three other insect *fng* genes. Although the four insect *fng* genes are more similar to each other than they are to any of the vertebrate *fng* genes, the majority of the amino acids conserved among predicted insect FNG proteins are also conserved among vertebrate FNG proteins. Indeed, within the 261-aa domain that is conserved among FNG proteins (corresponding to amino acids 149–410 of *Drosophila* FNG), 32% of the amino acids are identical among all 18 sequenced *fng* genes (Fig. 1, and Fig. 5, which is published as supporting information on the PNAS web site). Multiple alignment provides a more stringent comparison than simple pairwise alignments, and these 84 amino acids constitute a core of structural elements that are presumably essential for normal FNG function. The high degree of sequence similarity among *fng* genes is consistent with the conserved enzymatic activity of *Drosophila* and mammalian FNG (8), and with the biological activity of mouse *fng* genes expressed in *Drosophila* (15, 16).

Sequence Similarities Among β3GTs. FNG is not closely related to any other proteins encoded by the *Drosophila* genome (ref. 14; Fig. 4). However, by using the PSI-BLAST program (17), a set of additional *Drosophila* proteins was identified with weak similarity to FNG (Fig. 2 *B–H* and Fig. 4). Notably, most of these proteins are related to mammalian proteins that have been identified as having a β3GT activity (Fig. 2 and Fig. 4). We therefore propose that all of these proteins comprise a superfamily of *Drosophila* β3GTs. To more clearly define the relationships among these putative β3GTs, each was used as the query in a BLASTP search of the predicted *Drosophila* proteome. This analysis, together with consideration of shared sequence motifs, allowed their division into two families including several proteins more closely related to each other than to the other β3GTs (Fig. 2 *B* and *F* and Fig. 4), and six families with only a single member.

All but three of these putative β3GTs include the same four highly conserved sequence motifs (Fig. 2, motifs i–iv), in the same order. The remaining genes (CG3038, CG2983, and CG4351) lack good matches to all four motifs, but do appear to

BIOCHEMISTRY

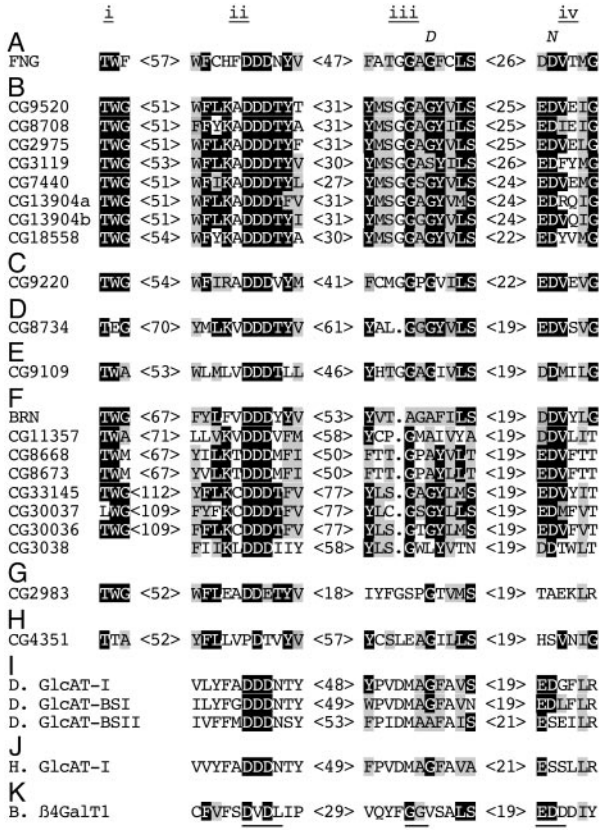

Fig. 2. Sequence motifs conserved among known and putative *Drosophila* β3GTs. The small roman numerals above the sequence letters indicate the designations for the motifs described in the text. The letters above the FNG sequence indicate the location of missense mutations in these motifs (Table 1). The genes identified in *A–H* were aligned by considering both PSI-BLAST and PILE UP alignments (see Fig. 5), and those in *I–K* were aligned by manually searching for the conserved motifs. Black boxes denote amino acids identical among 13 or more β3GT domains, and gray boxes denote amino acids that are similar among 13 or more β3GT domains. Numbers in brackets indicate the number of amino acids between motifs. Known or presumed sugar donors/acceptors for each glycosyltransferase (*A*) GlcNAc/Fuc, known (7, 8). (*B*) Gal/GalNAc, presumed by similarity to rat core 1 galactosyltransferase (32). (*C*) GlcUA/GalNAc, presumed by similarity to mammalian chondroitin synthase (33). (*D*) Gal/Gal, presumed by similarity to mammalian galactosyltransferase II (34). (*E*) Unknown donor/acceptor. (*F*) GlcNAc/Man, for BRN, known (35, 36), for the remainder, presumed to transfer either Gal to GlcNac, or GlcNac to Gal, based on similarity to mammalian enzymes (18, 20, 35, 37). (*G* and *H*) Donor/acceptor unknown. (*I*) GlcUA/Gal, known (22). (*J*) Human GlcAT-I. (*K*) Bovine β4GalT1. The underlined amino acids face the substrate binding pocket (25).

include two to three of them. These motifs overlap sequences identified as being shared within certain families of glycosyltransferases (9, 18–21), but, with the exception of the DXD sequence within motif ii (19), have not been identified as common to all β3GTs. Only one known class of β3GTs in *Drosophila* consisting of three GlcATs (22), was not identified by PSI-BLAST or BLASTP as being related to our proposed β3GT superfamily. Nonetheless, on manual inspection, three of the conserved motifs could be identified in these GlcATs and aligned with those of the other β3GTs (Fig. 2*I*).

Sequence comparison among different families of glycosyltransferases led to identification of a highly conserved DXD motif in a wide range of glycosyltransferases (19), which likely corresponds to the DDD sequence within motif ii of β3GTs. Intriguingly, almost all of the β3GTs we identified also contain an acidic motif (DD or ED, within motif iv) carboxyl-terminal to the DDD sequence. Acidic motifs have been identified at a similar location in β4GalTs (EDDD) and polypeptide *N*-acetylgalactosaminyltransferases (EXXE; reviewed in ref. 23). Moreover, although β4GalTs have a highly conserved motif, including GGV, in between the two acidic motifs, most members of the β3GT family (Fig. 2) have GAG or similar sequences within motif iii. Based on the order, spacing, and structural similarities among these motifs, we propose that they serve identical functions. Implicit in this suggestion is the notion that one can model these motifs in the primary sequences of β3GTs onto the tertiary structure of β4GalT.

The central feature of the crystallographic structure of bovine β4GalT1 is a large open pocket (24, 25). The amino acids DVD, GG, and EDD within motifs ii, iii, and iv are all solvent-exposed and are located near the bottom of the pocket (Fig. 2, and Fig. 6, which is published as supporting information on the PNAS web site). The structure of β4GalT1 has been solved in complex with its cofactor, Mn^{++}, and its donor substrate, UDP-galactose (24, 25). Amino acids within all three of these motifs make contacts with UDP-galactose. In addition, DVD (motif ii) participates in coordination of Mn^{++}, the small amino acids (GG) within motif iii allow the donor sugar to fit deep into the pocket, and an acidic amino acid within motif iv is proposed to make contact with the sugar acceptor and to participate in nucleophilic attack on it. Similarly, in the crystallographic structures of human GlcAT-I with its substrates (26, 27), DDD (motif ii) interacts with UDP and Mn^{++}, and the E within motif iv interacts with the acceptor substrate and is proposed to participate directly in catalysis. Indeed, even though bovine β4GalT1 and human GlcAT-I exhibit no significant primary sequence similarity, when the locations of conserved motifs ii, iii, and iv are mapped onto their crystallographic structures, it is evident that the spatial relationships of these motifs to each other and to the substrate binding pocket are similar (Fig. 6). We suggest that analogous spatial relationships and functions exist among these motifs in members of the β3GT superfamily.

Genetic Analysis of *Drosophila fng* Mutations. Although sequence conservation provides insights into important motifs, it does not indicate whether a specific amino acid is absolutely required for function, or whether substitutions can be made that only partially impair function. Conversely, it is possible that critical amino acids are not absolutely conserved, because substitution of similar amino acids can be tolerated, or because compensatory changes can occur elsewhere in the protein. Genetic analysis provides a means of identifying amino acids that are essential for function without regard to their conservation. Nine ethyl methanesulfonate (EMS)-induced alleles of *fng* were recovered in an F1 screen for mutations that failed to complement a weak, viable allele of *fng*, fng^{35UZ-1} (11). Although this screen was reported in the initial characterization of *fng*, several of these alleles have not been described previously. We also characterized five EMS-induced alleles isolated in a screen for mutations that influence the patterning of adult tissues in homozygous mutant clones (12), none of which have been described previously.

Of the 14 different EMS-induced alleles examined, 12 die during larval stages when they are homozygous, hemizygous, or transheterozygous with the strong alleles fng^{80} or fng^{13} (Table 1). The hemizygous condition was created by crossing *fng* alleles to a chromosomal deficiency that uncovers *fng*, *DfriXT1*. Only one of the alleles, fng^{2}, is fully viable (Table 1). A second allele, fng^{M69}, is semiviable. To check for temperature sensitivity, crosses with all alleles were conducted at three different temperatures (18, 25, and 29°C); fng^{M69} was the only temperature-sensitive allele identified. Even at 18°C, the fraction of surviving adult flies is <5% of the expected frequency, and the surviving flies have very strong wing phenotypes. A second semiviable allele, fng^{52}, was previously isolated by partial excision of a P element (11). Although fng^{52} is similar in strength to fng^{M69}, they

Table 1. Mutant alleles of *Drosophila fng*

Allele	Wing phenotype of transheterozygotes with DfriXT1	80	13	M69 at 18°	52	2	Mutagen	Mutation	Genetic background
1	L	L	L	5.0 (16)	3.5 (25)	1.8 (121)	EMS	TCG to ACG S(350) to T	FRT80B P[w^+]70C
2	2.3 (186)	1.1 (98)	2.4 (133)	2.7 (113)	1.3 (168)	1.0 (79)	EMS	CTG to CAG L(398) to Q	*st e*
3	L	L	L	4.9 (14)	3.2 (58)	1.9 (122)	EMS	GAA to AAA E(220) to K	FRT80B mwh jv
5	L	L	L	5.0 (10)	**4.0** (14)	**2.7** (122)	EMS	TCG to TTG S(350) to *	FRT80B P[w^+]70C
7	L	L	L	ND	**3.9** (29)	2.4 (159)	EMS	TTG to TAG L(221) to *	*st e*
8	L	L	1	5.0 (6)	**3.9** (14)	2.4 (125)	EMS	AAG to TAG K(196) to *	FRT80B P[w^+]70C
11	L	L	L	ND	**4.2** (16)	**2.6** (84)	EMS	GT to AT K(196) to N E *	FRT80B P[w^+]70C
13	L	L	L	5.0 (18)	**3.7** (31)	2.4 (133)	EMS	TGG to TAG W(288) to *	FRT80B *mwh jv*
14	L	L	L	ND	**4.3** (8)	**2.7** (134)	EMS	GAT to AAT D(327) to N	FRT80B *mwh jv*
L19	L	L	L	5.0 (12)	**3.6** (18)	**2.5** (181)	EMS	TGG to TGA W(181) to *	FRT80B P[w^+]70C
L73	L	L	L	5.0 (30)	**4.1** (23)	**2.8** (127)	EMS	TGG to TAG W(181) to *	FRT80B P[w^+]70C
L81	L	L	L	5.0 (4)	3.6 (10)	**2.5** (181)	EMS	GGC to GAC G(295) to D	FRT80B P[w^+]70C
L83	L	L	L	5.0 (14)	3.4 (54)	2.1 (204)	EMS	CTT to TTT L(213) to F	FRT80B P[w^+]70C
M69 at 18°	5.0 (8)	3.9 (24)	5.0 (18)	L	**4.2** (29)	**2.7** (113)	EMS	ACG to ATG T(184) to M	FRT80B P[w^+]70C
M69 at 25°	L	4.6 (32)	L	L	2.8 (20)	2.3 (164)			
M69 at 29°	L	L	L	L	3.3 (12)	**2.7** (142)			
35UZ-1	ND	ND	0.9 (89)	ND	ND	0.4 (67)	*P* element	Insertion	w^{1118}
52	3.3 (34)	4.6 (10)	3.7 (31)	4.2 (29)	3.0 (41)	1.3 (168)	Transposase	ND	35UZ-1
129	L	L	L	5.0 (4)	3.2 (28)	2.1 (116)	Transposase	Deletion	35UZ-1
DfriXT1	L	L	L	5.0 (8)	3.3 (34)	2.3 (186)	X-ray	Deletion	*ru st e ca*

Columns 2–7 show the average phenotypes of combinations of the indicated alleles, with the number of wings scored in parentheses. Phenotypic scores for combinations with fng^2 or fng^{52} that are significantly greater with EMS-induced alleles than with DfriXT1 are indicated in bold, and scores that are significantly less are underlined. L, Lethal; ND, not determined. For point mutations, both the nucleotide changed and amino acid affected are indicated. *, stop codon. All EMS-induced mutations except fng^2 and fng^7 were originally induced on a *P*[w^+] FRT80B chromosome. In addition to these mutations, the two following polymorphisms were identified: in fng^2 and fng^7, at mRNA nt 690/aa 193, GAG(E) to GAA(E); in fng^{52} at nt 1605 (3′ UTR), C to A.

do not fall into a simple allelic series because fng^{M69} affects the development of the wing more strongly than the development of the head, whereas the opposite is true for fng^{52} (Table 1 and data not shown; refs. 11 and 28). We think this result is because fng^{52} is a regulatory rather than a structural mutation (see below).

To characterize the relative *fng* function provided by different mutant alleles, we crossed all of the EMS-induced alleles to each of the three weakest alleles and scored the phenotypes of the transheterozygous progeny. According to well established genetic criteria, a chromosomal deficiency or a deletion allele can be used as a baseline for comparison in transheterozygous combinations (29). If a mutant allele behaves identically to a deficiency, it is an amorph (genetically null allele). If an allele results in less severe phenotypes than a deficiency, it is a hypomorph (partial loss-of-function allele). If an allele results in more severe phenotypes than a deficiency, it is an antimorph. Antimorphic behavior implies that an allele produces a defective product that antagonizes the activity of the product of the other allele.

Although *fng* is required for the normal development of several different tissues, the wing is the most sensitive to reduced fng activity (11). The degree to which tissue is lost from the wing varies among viable *fng* alleles, allowing them to be ordered into a phenotypic series, with $fng^{M69} > fng^{52} > fng^2$. Similarly, differences in the wing phenotypes of transheterozygous combinations of different EMS induced *fng* alleles were observed in combination with these viable alleles. All allelic combinations give some range of wing phenotypes; to quantify them we created a five-point scale in which a wild-type wing phenotype was scored as 0 (Fig. 3*A*), a wing with only vein phenotypes was scored as

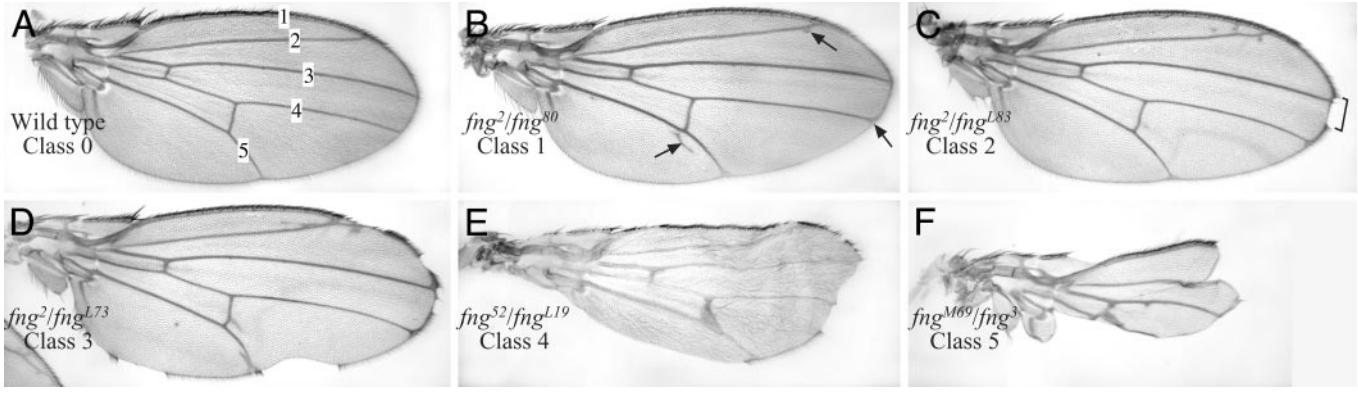

Fig. 3. Sample *fng* mutant wing phenotypes. (*A*) Wild-type wing with class 0 phenotype. The wing veins are numbered. (*B*) fng^2/fng^{80} wing with class 1 phenotype. Arrows point to extra vein material. (*C*) fng^2/fng^{L83} wing with class 2 phenotype. The bracket marks a gap in the wing margin. (*D*) fng^2/fng^{L73} wing with class 3 phenotype. (*E*) fng^{52}/fng^{L19} wing with class 4 phenotype. (*F*) fng^{M69}/fng^3 wing (at 18°C) with class 5 phenotype.

1 (Fig. 3*B*), a wing with very mild wing notching (less than the distance between veins 3 and 4, Fig. 3*C*) was scored as 2, a wing with moderate notching (less than one-third of the wing circumference, Fig. 3*D*) was scored as 3, and a wing with strong wing notching (less than two-thirds of the wing circumference, Fig. 3*E*) was scored as 4, and a wing with severe wing notching (more than two-thirds of the wing circumference, Fig. 3*F*) was scored as 5.

Surprisingly, many of the EMS-induced *fng* alleles appear to behave as recessive antimorphs. fng^{2} over *DfriXT1* has mild wing notching (average phenotype of 2.3, Table 1). However, nine of the EMS-induced lethal alleles gave a more severe phenotype than *DfriXT1* in combination with fng^{2} (Table 1). Although the average difference is slight, *t*-test analysis indicates that it is significant at the 99% confidence interval for six of the nine transheterozygous combinations (Table 1). fng^{52} has a stronger wing phenotype over *DfriXT1* than does fng^{2}, with an average score of 3.3. Although the phenotypic scores are based on smaller numbers because of the relatively low viability of fng^{52} (<10% of expected flies survive in most cases), we again found that the strongest EMS-induced alleles gave more severe phenotypes in combination with fng^{52} than did *DfriXT1*, and the difference is significant for eight combinations (Table 1). Although some of the differences could be because of modifiers in the genetic background, this possibility does not suffice as a general explanation because the EMS-induced alleles tested are in different backgrounds, as are fng^{2} and fng^{52}. Additionally, to exclude the possibility that a modifier exists in the *DfriXT1* stock, we also examined phenotypes in combination with an allele, fng^{129}, which deletes the first two exons of *fng* and makes no detectable mRNA, and hence should be a true null (11). This mutation gives phenotypes in combination with fng^{2} and fng^{52} that are similar to those of *DfriXT1*.

Two of the lethal alleles, fng^{1} and fng^{3}, gave phenotypes slightly weaker than *DfriXT1* or fng^{129} in combination with fng^{2}, and hence are characterized as hypomorphs. Although the wing phenotypes of transheterozygous combinations of these alleles with fng^{52} or fng^{M69} were not significantly weaker than for *DfriXT1* (Table 1), fng^{3} is fully viable over fng^{52}, which contrasts with the reduced viability of fng^{52} over *DfriXT1* or any of the stronger alleles.

Molecular Analysis of *fng* Mutations. The existence of a large collection of EMS-induced *fng* mutations presents an opportunity to identify essential structural features of FNG without biases based on similarity to other β3GTs or sequence conservation among *fng* genes. To identify the molecular lesions responsible for these *fng* mutations, we sequenced genomic DNA from all 14 EMS-induced alleles, as well as fng^{52}. Because homozygous mutant animals are lethal, they were identified at late embryonic or early larval stages by the absence of fluorescence generated by a GFP transgene on a balancer chromosome. The exons and portions of the introns of *fng* were then amplified in fragments, and the sequences of both strands of the entire coding region and splice junctions were determined. The results of these analyses are summarized in Table 1 and Fig. 1.

Single nucleotide changes that alter the predicted amino acid sequence were identified in all 14 of the EMS-induced alleles. DNA sequencing also revealed a polymorphism between fng^{2} and fng^{7}, which were induced on a *st e* chromosome, and the remaining alleles, which were induced on a *P*[*w*+] *FRT80B* chromosome (Table 1). The absence of other polymorphisms, together with the detection of unique changes within each mutant line, argues that the detected alterations are responsible for the mutation of *fng*. No alteration of the predicted coding region was detected in fng^{52}. fng^{52} retains some P element sequences, and the lack of alteration in the coding region implies that this is a regulatory, rather than a structural, mutation.

Six of the EMS-induced alleles are associated with point mutations that introduce a stop codon. One additional allele, fng^{11}, lacks mutations in coding sequences, but has a GT to AT mutation in the splice-donor site at the beginning of intron 3. This mutation is predicted to prevent splicing of this intron, resulting in the addition of two intron-encoded amino acids and then a stop codon. Altogether, these seven alleles encode truncated proteins that contain as few as 180 and as many as 349 amino acids of wild-type FNG (Table 1, Fig. 1). Notably, all seven mutations behave as amorphs or recessive antimorphs in genetic complementation tests. Thus, the truncation of as few as 63 amino acids from the C terminus completely eliminates normal FNG function. Five of these seven mutations delete all three of the motifs shared among putative β3GTs, and one deletes two of the three motifs. We have proposed that these motifs are related to motifs in β4GalT1 and GlcAT-I that make key interactions with substrates, and the absence of *fng* function in these mutants is consistent with this hypothesis. Indeed, given the substantial truncation of FNG, the mutant proteins would likely be unable to interact with substrates. However, some glycosyltransferases are known to form dimers (30, 31), and in the case of β4GalT1 the amino-terminal transmembrane domain has been implicated in dimerization (30). Thus, one explanation for the antimorphic behavior of truncated alleles is that FNG too may function as a dimer, and that the truncated proteins retain some ability to dimerize. Truncated proteins could then trap some of the protein produced by the other allele in nonfunctional complexes. Some variation in the strength of different truncated alleles was observed in our genetic analysis, which, according to this hypothesis, could be caused by differences in the stability of truncated products.

Seven of the EMS-induced alleles are associated with missense mutations, and six of these occur in amino acids that are identical in all FNG proteins. Only fng^{2}, the weakest *fng* allele, is associated with a mutation in an amino acid that is not identical in all FNGs, and even in this case it makes a nonconservative substitution at a site where only conservative substitutions are found among wild-type FNG.

Notably, two of the missense mutations affect sequence motifs that are conserved among β3GTs. fng^{14} is a mutation of D to N within conserved motif iv, and fng^{L81} is a mutation of G to D within conserved motif iii. Importantly, these mutations act as null or slightly antimorphic alleles and are genetically indistinguishable from mutations associated with stop codons that truncate FNG amino-terminal to motif ii. The recovery of these point mutations thus lends further support to the inference that these conserved motifs identify amino acids critical for the function of FNG, as well as that of other members of the β3GT superfamily. Moreover, the nature of these mutations is consistent with homology modeling to the β4GalT1 structure (Fig. 6). The proposed homology implies that DD will be in a position to participate in a nucleophilic attack on the sugar acceptor. Although N is similar is size to D, it cannot act as a nucleophile. The GAG motif is proposed to allow donor sugars to fit into the central pocket; the disruption of this motif caused by mutation to D in fng^{L81} could preclude donor sugar binding. The GAG and DD motifs overlap motifs that have been recognized as being conserved among certain glycosyltransferases (9), but these *fng* mutations constitute a demonstration that they are in fact essential for β3GT function. Finally, we also note that the recovery of two independent point mutations in motifs that are conserved among β3GTs provides further support for the conclusion that the glycosyltransferase activity of FNG is essential to its ability to modulate Notch signaling (7, 8, 10).

Four missense mutations occur in amino acids that are identical among all FNG proteins, but are not obviously conserved among other β3GTs. They may thus be required for unique aspects of FNG function, such as recognition of its EGF-*O*-

fucose substrate. All four of these mutations are strong alleles, but are nonetheless slightly weaker than nonsense mutations or the missense mutations in β3GT motifs. Interestingly, two of these mutations, fng^{L83} and fng^{3}, occur seven amino acids apart, along a predicted amphipathic helix amino-terminal to motif ii. fng^{M69} has a T to M change, and the conditional nature of this mutation suggests that it may affect the stability of the correctly folded protein. fng^{1} is intriguing because it is a strong *fng* mutation, yet it is associated with the conservative substitution of a S with a T. This S occurs in the middle of a six amino acid motif that is identical in all FNG proteins (Fig. 1), which constitutes the most extensive block of sequence identity outside of the β3GT motifs. As four of the six amino acids are polar or charged, the S is most likely solvent exposed. The fact that its replacement by T severely compromises FNG function suggests that the positioning of the hydroxyl group of the S is absolutely critical for normal FNG activity.

We thank S. Carroll for the *P. coenia* cDNA library; S. Johnston and J. Ratchford for assistance in cloning and sequencing *P. coenia* cDNAs; the Bloomington stock center for *Drosophila* stocks; U. Haecker for pointing out the similarity of CG8673 and CG18558 to other family members; R. Haltiwanger for comments on the manuscript; and G. Struhl, in whose laboratory fng^{L19}, fng^{L73}, fng^{L81}, fng^{L83}, and fng^{M69} were isolated, for his support. Research in K.D.I.'s laboratory is supported by the Howard Hughes Medical Institute and by National Institutes of Health Grant R01-GM54594.

1. Panin, V. M. & Irvine, K. D. (1998) *Semin. Cell Dev. Biol.* **9,** 609–617.
2. Haltiwanger, R. S. & Stanley, P. S. (2003) *Biochim. Biophys. Acta* **1573,** 328–335.
3. Artavanis-Tsakonas, S., Rand, M. D. & Lake, R. J. (1999) *Science* **284,** 770–776.
4. Joutel, A. & Tournier-Lasserve, E. (1998) *Semin. Cell Dev. Biol.* **9,** 619–625.
5. Irvine, K. D. & Vogt, T. F. (1997) *Curr. Opin. Cell Biol.* **9,** 867–876.
6. Irvine, K. D. & Rauskolb, C. (2001) *Annu. Rev. Cell Dev. Biol.* **17,** 189–214.
7. Bruckner, K., Perez, L., Clausen, H. & Cohen, S. (2000) *Nature* **406,** 411–415.
8. Moloney, D. J., Panin, V. M., Johnston, S. H., Chen, J., Shao, L., Wilson, R., Wang, Y., Stanley, P., Irvine, K. D., Haltiwanger, R. S., *et al.* (2000) *Nature* **406,** 369–375.
9. Yuan, Y. P., Schultz, J., Mlodzik, M. & Bork, P. (1997) *Cell* **88,** 9–11.
10. Munro, S. & Freeman, M. (2000) *Curr. Biol.* **10,** 813–820.
11. Irvine, K. D. & Wieschaus, E. (1994) *Cell* **79,** 595–606.
12. Jiang, J. & Struhl, G. (1995) *Cell* **80,** 563–572.
13. Casso, D., Ramirez-Weber, F. & Kornberg, T. B. (2000) *Mech. Dev.* **91,** 451–454.
14. Misra, S., Crosby, M. A., Mungall, C. J., Matthews, B. B., Campbell, K. S., Hradecky, P., Huang, Y., Kaminker, J. S., Millburn, G. H., Prochnik, S. E., *et al.* (2002) *Genome Biol.* **3,** RESEARCH0083-3.
15. Cohen, B., Bashirullah, A., Dagnino, L., Campbell, C., Fisher, W. W., Leow, C. C., Whiting, E., Ryan, D., Zinyk, D., Boulianne, G., *et al.* (1997) *Nat. Genet.* **16,** 283–288.
16. Johnston, S. H., Rauskolb, C., Wilson, R., Prabhakaran, B., Irvine, K. D. & Vogt, T. F. (1997) *Development (Cambridge, U.K.)* **124,** 2245–2254.
17. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997) *Nucleic Acids Res.* **25,** 3389–3402.
18. Zhou, D., Dinter, A., Gutierrez Gallego, R., Kamerling, J. P., Vliegenthart, J. F., Berger, E. G. & Hennet, T. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 406–411.
19. Wiggins, C. A. & Munro, S. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 7945–7950.
20. Amado, M., Almeida, R., Carneiro, F., Levery, S. B., Holmes, E. H., Nomoto, M., Hollingsworth, M. A., Hassan, H., Schwientek, T., Nielsen, P. A., *et al.* (1998) *J. Biol. Chem.* **273,** 12770–12778.
21. Isshiki, S., Togayachi, A., Kudo, T., Nishihara, S., Watanabe, M., Kubota, T., Kitajima, M., Shiraishi, N., Sasaki, K., Andoh, T., *et al.* (1999) *J. Biol. Chem.* **274,** 12499–12507.
22. Kim, B.-T., Tsuchida, K., Lincecum, J., Kitagawa, H., Bernfield, M. & Sugahara, K. (2003) *J. Biol. Chem.* **278,** 9116–9124.
23. Breton, C. & Imberty, A. (1999) *Curr. Opin. Struct. Biol.* **9,** 563–571.
24. Ramakrishnan, B., Balaji, P. V. & Qasba, P. K. (2002) *J. Mol. Biol.* **318,** 491–502.
25. Gastinel, L. N., Cambillau, C. & Bourne, Y. (1999) *EMBO J.* **18,** 3546–3557.
26. Pedersen, L. C., Darden, T. A. & Negishi, M. (2002) *J. Biol. Chem.* **277,** 21869–21873.
27. Pedersen, L. C., Tsuchida, K., Kitagawa, H., Sugahara, K., Darden, T. A. & Negishi, M. (2000) *J. Biol. Chem.* **275,** 34580–34585.
28. Papayannopoulos, V., Tomlinson, A., Panin, V. M., Rauskolb, C. & Irvine, K. D. (1998) *Science* **281,** 2031–2034.
29. Muller, H. J. (1932) in *Proceedings of the Sixth International Congress of Genetics*, ed. Jones, D. F. (Brooklyn Botanical Gardens, Menasha, WI), pp. 213–255.
30. Yamaguchi, N. & Fukuda, M. N. (1995) *J. Biol. Chem.* **270,** 12170–12176.
31. Ju, T., Cummings, R. D. & Canfield, W. M. (2002) *J. Biol. Chem.* **277,** 169–177.
32. Ju, T., Brewer, K., D'Souza, A., Cummings, R. D. & Canfield, W. M. (2002) *J. Biol. Chem.* **277,** 178–186.
33. Kitagawa, H., Uyama, T. & Sugahara, K. (2001) *J. Biol. Chem.* **276,** 38721–38726.
34. Bai, X., Zhou, D., Brown, J. R., Crawford, B. E., Hennet, T. & Esko, J. D. (2001) *J. Biol. Chem.* **276,** 48189–48195.
35. Schwientek, T., Keck, B., Levery, S. B., Jensen, M. A., Pedersen, J. W., Wandall, H. H., Stroud, M., Cohen, S. M., Amado, M. & Clausen, H. (2002) *J. Biol. Chem.* **277,** 32421–32429.
36. Muller, R., Altmann, F., Zhou, D. & Hennet, T. (2002) *J. Biol. Chem.* **277,** 32417–32420.
37. Shiraishi, N., Natsume, A., Togayachi, A., Endo, T., Akashima, T., Yamada, Y., Imai, N., Nakagawa, S., Koizumi, S., Sekine, S., *et al.* (2001) *J. Biol. Chem.* **276,** 3498–3507.
38. Chen, J., Moloney, D. J. & Stanley, P. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 13716–13721.