

Preface

How embryos “self-assemble” has fascinated thinkers for millennia [2918, 3064, 3190]. Among the ancient Greeks, Aristotle (384–322 BCE) made copious observations and coined the term “morphogenesis,” which is still in use today [2989, 4305]. For the past century, the science of “developmental mechanics” has hammered at this problem relentlessly, but it is only in the last decade that the core mysteries have finally cracked [1487]. The deepest secrets have come from a fairylike fly named *Drosophila melanogaster*, probably the same species of “gnat” that Aristotle himself noticed hovering over vinegar slime [217, 3361, 4184]. Unfortunately, these insights can only be fully appreciated in the arcane language of fly genetics. Hence this book full of runes and rules.

This book concerns cuticular patterns, the cellular machinery that makes them, and the genetic circuitry that runs the machinery. Although it is mainly a survey, it is also a narrative that traces the roots of our knowledge. The story that it tells – albeit in condensed form – rivals the *Iliad* in scope (legions of researchers devoting decades to attacking thousands of genes) and the *Odyssey* in wonderment (monstrous mutants posing riddles that challenge even the most clever explorer-heroes). Indeed, truth is often stranger than a fairy tale in the realm of the fly. Believe it or not, there are even remote islands where giant drosophilids with dappled wings and feathery legs have been spied dancing and fighting in the misty forests [668, 669].

Ever since 1910 when T. H. Morgan’s first “fly paper” was published [2948], the field of fly genetics has brimmed with intriguing curiosities [820, 2951, 3673] and equally colorful human personalities [120, 327, 2283, 4183]. Added to these delights is a menagerie of recently discovered molecules

– e.g., the midget “Bearded” (81 a.a.) [2499] and the giant “Dumpy” (3680 a.a.) [4668]. Now that the fly genome project is ending [14], the world is peering into this circus. What newcomers may not realize is that this field offers many diversions beyond its databases.

Like other holometabolous insects, flies live two lives – first as a grub, then as a flying adult [82]. During metamorphosis, 19 “imaginal discs” erupt from inside the maggot and are quilted together to form most of the adult skin. The gold-colored cuticle secreted by that skin is exquisitely ornate. The head is embossed with hundreds of domes that focus light onto bundles of photoreceptors, the thorax is sculpted into dozens of jointed parts that form a contraption for walking and flying, and the abdominal wall (built from non-disc tissue [2648]) is pleated into an expansible chamber for digestion and reproduction. Nearly everywhere, the body surface sprouts bristles whose patterns can be as orderly as soldiers on parade.

Why do only some cells make bristles? That is a problem of differentiation. Why do bristles arise only at certain sites? That is a problem of pattern formation, and these questions can be asked for structures in general. Beneath both problems is a coding enigma: how does the fly’s 1-dimensional genome encode the 2-dimensional cuticular landscape? Once, it seemed that each body part might be governed by its own set of genes [4509, 4512], but this notion proved wrong [1094, 1114, 2410, 4643]. In fact, most patterns are built by the same ensembles of genes. These modules arose eons ago in the mythical common ancestor of insects and vertebrates [1439, 3840]. Since then, evolution has customized the circuitry by making new intra- and inter-modular links [968, 1440].

What is the nature of the circuitry, and how does it program cells to “compute” patterns? That is the subject of this book. Topics are arranged roughly in order of increasing complexity. Chapter 1 establishes one simple fact: in contrast to nematodes, flies rely primarily on intercellular signaling (vs. cell lineage) to assign cell fates. The rest of the book traces how signaling occurs. Chapter 2 delves into the 5-cell cluster that constructs a mechanosensory bristle. The bristle is an exception to the signaling rule: its cell fates are dictated almost entirely by lineage. Chapter 3 uses bristle *patterns* to show how cells communicate in populations larger than a bristle but smaller than a disc, and Chapter 4 sets the stage for a discussion of larger-scale patterning by reviewing how discs arise and grow. Chapters 5 to 7 explore how leg, wing, and eye discs use similar toolkits of genes in idiosyncratic ways. The other two major discs – haltere and genital – are excluded because their strategies so closely resemble wing [16, 51, 3875, 4683, 4694] and leg discs [679, 735, 1163, 2343, 2942, 3732], respectively. (Fly genitalia are evolutionarily modified appendages [1137, 1179, 1562].) Chapter 8 contemplates the phenomenon of homeosis in the context of evolution.

Overall, the book's quest is to understand cellular “epistemology” (what do cells know?) and “psychology” (how do they think?). Its approach involves de- and re-construction: to cut through the jargon, tease out the facts, and then try to make sense of the models by piecing the clues back together using *a priori* reasoning.

The bad news is that there are so many pieces in the puzzle that persistence will be needed. The good news is that their interactions are so limited that no fancy math is required to learn the rules of the game [3598, 3841]. A recurrent theme in the saga is how cellular riddles were solved by molecular genetics. The abiding moral is that there is much more experimental work to be done if we are to comprehend how the fly's ~14,000 genes [14, 1559, 3618, 3674] – or a large portion thereof [280, 615, 963, 4273] – are orchestrated during patterning [698, 2162, 2237, 2845, 4084]. In short, the fly still holds many secrets, and genomics will need genetics to ferret them out [465].

Thus, the book is a sampler of case studies and gedanken exercises, not an encyclopedia. That function is served by the Internet databases, and readers should consult two main websites: *FlyBase* (flybase.bio.indiana.edu) [124, 279] and *The Interactive Fly* (sdb.bio.purdue.edu) [484]. Fly lore is best savored by browsing the classics: the 1993 Cold Spring Harbor 2-volume compendium on development [238], its gargan-

antuan 12-volume predecessor *The Genetics and Biology of Drosophila* [122], Mike Ashburner's huge “handbook” [118], Lindsley and Zimm's dictionary of fly genes [2561], Bridges and Brehme's Barnumesque catalog of freakish mutants [470], and the Morgan team's *magnum opus* of 1925 [2951]. However, the fun of fly research is best portrayed in the charming *Fly* by Martin Brookes (2001, Harper-Collins, N.Y.).

Despite this disclaimer about breadth, a few topics are covered in depth in the appendices. Appendix 1 is a glossary of protein domains. Appendix 2 lists most of the ideas that have guided research in this field. Appendices 3 to 5 catalog the well-studied genes that affect bristles, sensilla, or bristle patterns, and Appendix 6 outlines three of the key signaling pathways in disc development (Hedgehog, Wingless, and Decapentaplegic). The other two pathways are discussed in Chapters 2 (Notch) and 6 (EGFR). Appendix 7 contains additional comments about the figures.

Historically, disc research has been reviewed intermittently. Disc histology was codified by Dietrich Bodenstern in 1950 [377]. Disc development and genetics were surveyed by Gehring and Nöthiger (1973) [1421], Postlethwait and Schneiderman (1973) [3448], Bryant (1978) [526], Shearn (1978) [3881], Poodry (1980) [3422], and Oberlander (1985) [3165]. The first blush of molecular-genetic data was evaluated by Stephen Cohen in 1993 [834], and the fundamentals of signaling were summarized by Seth Blair in 1999 [358]. Two books that nicely bracket the last 30 years of investigation are *The Biology of Imaginal Discs* (1972, H. Ursprung and R. Nöthiger, eds.) [4426] and *Developmental Genetics of Drosophila* (1998, A. Ghysen, ed.) [1452].

Conventional nomenclature is used. Locations of genes are stated in terms of the salivary gland chromosome map [2561]: the 3-part code (e.g., “92E12–14”) denotes the chromosome section (1–20 span the X, 21–60 the 2nd, 61–100 the 3rd, and 101–102 the tiny 4th chromosome), the lettered subdivision (A–F), and the band or range of bands. Genes are italicized, but gene complexes (e.g., Bar-C) are not. All proteins are in plain type. Mutations are superscripted (e.g., *numb*^{LOF}), whereas wild-type alleles are not (*numb*) or are labeled with “+” (*numb*⁺). Null alleles are designated by a “null” or “–” superscript. Most gene names record the dominant (capital) or recessive (lowercase) nature of early mutations (e.g., *Notch* vs. *numb*). Capital “D” (*Drosophila*) is used for paralogs within the species (e.g., *Dfz2* [310] in the *frizzled* series), whereas lowercase “d” refers to

orthologs of vertebrate genes (e.g., *dTcf* [692, 1517]). Proteins are always capitalized (e.g., Numb).

Given these rules, the normal symbols for *Hairless* (*H*) and *hairy* (*h*) are distinct for the genes but not for the proteins ("H" in both cases), so "H" will be used only for *Hairless*, while "Hairy" will be written out. Likewise, *Beadex* will be written out to avoid confusion with the protein encoded by *bithorax* (both would be "Bx"). Small capitals are employed for Boolean states (ON, OFF), conditions (IF, THEN, NOT), and conjunctions (AND, OR). Amino acid and nucleotide sequences are underlined. Boundaries are denoted by slash marks (e.g., "A/P") and axes by hyphens (e.g., "A-P"). Short gene names (≤ 5 letters) are not usually abbreviated.

Abbreviations include a.a. (amino acid), AEL (after egg laying), AP (after pupariation) a.k.a. (also known as), b.p. (base pair), h (hour), hs (heat shock), kb (kilobase), kD (kilodalton), MC (macrochaete), mC (microchaete), St. (stage of embryogenesis), t.s. (temperature-sensitive), pers. comm. (personal communication), and unpub. obs. (unpublished observations). Times (h AEL or h AP) refer to a culture temperature of 25°C, unless stated otherwise. Polypeptide lengths are for the unprocessed (nascent) precursor. Genes that are usually called "neurogenic" (based on mutant phenotype) [436] are here termed "antineural" (based on function) [4367] to contrast them with "proneural" (based on function) genes [2018]. "Eye disc" refers to both the eye and antennal parts, and "wing disc" denotes the entire dorsal mesothoracic disc (wing, notal, and pleural parts). By tradition (quirky though it may be), fate maps employ *left* legs (Ch. 5), *right* wings (Ch. 6), and *left* eyes (Ch. 7) [185, 320, 526, 531], although right eyes are used by some authors [2962].

Readers must be familiar with the basics of fly development [358, 2434, 3517] and the methods of modern genetics [354, 4671], including (1) induction of cell clones by *flp*-mediated recombination [1530–1532, 3952, 4781] and the *flp*-out trick [4159], (2) regional misexpression of genes via *Gal4-UAS* constructs [435, 3857], (3) temporal misexpression via heat-sensitive alleles [4214] or heat-shock promoters [2953], (4) enhancer trapping using *lacZ* reporter genes [278, 329, 1286, 4667], and (5) two-hybrid screening for protein interactions [222, 763, 1228, 1229, 1316].

Wherever possible, circuits are formulated in terms of Boolean logic [399] because this format shows syntax better than the "spaghetti diagrams" of genetics, electronics, or neural networks [2870]. The temptation to compare fly circuits with vertebrate or nematode circuits

is generally resisted here for the sake of conciseness. Such comparisons can be found in Eric Davidson's book *Genomic Regulatory Systems* [968] and at Tom Brody's website *The Interactive Fly*.

The term "link" is used in the sense of "causal linkage." Links are symbolized as " $\rightarrow$ " (activation) or " $\dashv$ " (inhibition). When a gene is the object (e.g., "*Dpp* $\rightarrow$ *omb*"), the effect is always at the transcriptional level, but pathways may be distilled in terms of either genes (*en* $\dashv$ *ci* $\rightarrow$ *ptc*) or proteins (*En* $\dashv$ *Ci* $\rightarrow$ *Ptc*), and any attendant ambiguities will be clarified by context. Epistatic links need not be direct. Thus, "*a* $\rightarrow$ *c*" could reflect a longer chain such as "*a* $\rightarrow$ *b* $\rightarrow$ *c*" or "*a* $\dashv$ *b* $\dashv$ *c*." The reason for listing so many links in this book is to facilitate Aristotle's goal of delineating the entire chain of causes from the fertilized egg to the adult [1993, 2919, 4305]. Only by concatenating all the known fragments can we see the gaps that remain to be filled.

The terms "LOF" (Loss of Function) and "GOF" (Gain of Function) typically denote decreases or increases in levels of gene activity (i.e., under- or overexpression) [1117, 1455], but in the broader sense that will be used here, GOF also includes ectopic misexpression where the "gain" is regional (cf. Fig. 6.13). For example, clones of cells that express a wild-type allele of *engrailed* (*en*⁺) outside the territory where *en*⁺ is normally transcribed will be called "*en*^{GOF}" [4848]. Cases do arise where overexpressing a wild-type allele has effects that differ from expressing a constitutively active construct [3545], and these will be so indicated. Mutations that are neither LOF nor GOF (e.g., neomorphs and antimorphs) are rarer, and allele-specific superscripts will be retained for them (e.g., *ci*^D [3818] and *en*^I [1636]).

LOF and GOF tests are used to assess the necessity (LOF) and sufficiency (GOF) of a specific gene for a particular process [173, 3643, 4333, 4671], and they are valuable tools. However, neither is foolproof. For example, if we delete gene "*a*" and see no effect on bristles (a negative LOF result), then *a* is clearly dispensable for bristle formation, but we cannot conclude that *a* is irrelevant because it might be acting redundantly with gene "*b*" [2845, 4584]: "*a* OR *b* $\rightarrow$ bristle." GOF data can also be misleading [6, 682, 1329]. For instance, if we drive the expression of gene "*a*" in a region where it is not normally transcribed and find that it induces bristles (a positive GOF result), then *a* is clearly sufficient for evoking bristles [1458, 1854, 2019, 3267], but this does not mean that *a* promotes bristle formation in wild-type flies because GOF perturbations can saturate limiting components (e.g., bHLH

partners [438, 918, 1854] or external ligands [421]) or provoke interactions with other pathways (e.g., converging RTK cascades [326, 1117, 2623] or branched Frizzled chains [3912, 4365, 4867]), resulting in all sorts of artifacts [6]. Researchers beware!

It is... unsafe to deduce normal gene function [when] the product is forced into inappropriate cells, perhaps in the absence of proteins with which it normally interacts and the presence of others that it does not normally encounter. [1304]

Results derived from mutant analyses or from utilizing ectopic expression of a gene product reveal the potential of a particular interaction to occur, not whether the interaction actually occurs during wild-type development. [3248]

Artifacts can be minimized by combining LOF and GOF tests [147, 3462]. Indeed, that is the only way to distinguish factors that are “instructive” for cell fates from those that are merely “permissive.” Instructive agents have both LOF and GOF effects, whereas permissive agents have a LOF but no GOF effect [449, 1455]. Even this 2-pronged approach may not be able to resolve epistatic relations, however, where (1) interactions are cooperative as in multiprotein complexes, (2) pathways are nonlinear, (3) feedback obscures causality, or (4) the “upstream” vs. “downstream” ranking of genes contradicts the order of cellular actions in time. An example of the last difficulty involves *scute* and *Notch*. In general, *scute* is epistatic to *Notch* (i.e., *scute*^{LOF} *Notch*^{LOF} flies show the *scute*^{LOF} missing-bristle trait instead of the *Notch*^{LOF} extra-bristle trait) [918, 1797, 1802, 3270, 3983], so *scute* should be acting downstream of *Notch*, but in fact *scute* must endow cells with “proneural competence” before *Notch* can enforce any “lateral inhibition” (cf. Ch. 3). The situation is even more complex at certain sites where *Notch* also acts before *scute* during a “prepattern” (pre-proneural) stage [461, 886].

Not all the fly’s circuitry is as inscrutable as the *Notch-scute-Notch* cascade, but our view of every subsystem is distorted by the imperfect lens of genetic dissection [2917, 3981, 4085, 4671]. Conclusions must therefore be qualified by layers of caveats about this or that alternative interpretation. The problem with such equivocation, of course, is that it can put readers to sleep.

How much of this blather can readers tolerate? Why not just present “best guess” models and avoid all the dithering? Good advice on this issue comes from a delightful little essay entitled, “Wingless signaling: The inconvenient complexities of life.” Therein, Rachel Cox and Mark Peifer argue that cartoon-like abstractions are essential but must be tempered by critiques that convey

the subtleties. Around every “gospel truth” there is a Talmudic aura of uncertainty. The author’s goal should be to make the material as accessible as possible without hiding any ambiguities. This book will attempt to do just that.

Nature is a home handywoman. Constrained by evolution, she does the job with the tools at hand, using a screwdriver for a hammer if necessary.... This machinery is neither elegant nor simple, but consists rather of a complex set of interacting proteins that were cobbled together by evolution.... Models help to organize our thoughts and offer testable hypotheses. Of course, in constructing a model, some data may need to be hammered into place, and the inconvenient data that cannot be coaxed into place have to be left out. The models that are frequently illustrated in minireviews... thus cannot be viewed as the “truth,” or they would narrow thought processes and squelch novel lines of research. We must be thoughtful iconoclasts, remembering that ultimately all models are wrong, fundamentally flawed or lacking the full complexity of systems shaped by evolution rather than intelligent design. We will thus use this forum to critique rather than prop up our model. It is increasingly clear that life is more complicated than portrayed there. [894]

Only by venturing into the ocean of literature can novices experience the richer Fly World beyond the Internet harbors. Alas, it is all too easy to get lost in those rougher seas. For that reason, an effort is made to supply the equivalents of charts and buoys. To wit, all key mysteries that have taunted investigators are set in boldface when introduced. So are the models and metaphors that have been contrived to explain the mysteries, plus the epiphanies encountered whenever great mysteries were slain. All these concepts are inventoried in Appendix 2. Some of the coined names for the concepts are whimsical, but no more so than the silly names of many fly genes. Indeed, working in this field has been so much fun *because* of its playful irreverence – a legacy of the neophyte pioneers in Morgan’s team [119]. Even “the Boss” himself loved to clash ideas [2947] and smash idols [2946]. Ideas are contrasted here wherever possible, and the style is decidedly iconoclastic.

All statements are source-referenced, and cross-references that are not direct attributions are listed as “cf. such-and-such” – a style that is common in the humanities but rare in the sciences [1630]. The cf.’s mean to compare, confer, or just “see also.” Due to space limitations, some citation strings had to be truncated. Those cases are flagged with a “Δ” superscript to alert readers who want to trace earlier sources thereby. An unabridged bibliography is posted at *The Interactive Fly*.

Esoterica are banished to tables, figures, and appendices wherever possible, and supportive evidence is crammed into indented blocks of text so that readers can skip them if they want. Even so, readers may find some sections of the text unnavigable without looking up the cited papers and tracing their lines of reasoning. Subheadings are worded as sentences so that the Table of Contents reads like a summary for each chapter. Gene abbreviations are defined wherever they are used in the text. Overall, the layout is designed to avoid boring the expert without confusing the novice. I still remember how hard it was to make my way into this field as an apprehensive apprentice.

This field has seen paradigm clashes of Promethean proportions, and those wars must be recounted to do the subject justice. For that reason, the modern facts have been woven into a historical tapestry, with a few homilies stitched in for good measure. Admitting past mistakes can help in spotting future pitfalls... even in the Olympian realm of molecular genetics, which surprisingly has more than a fair share of mortal foibles [1873, 2414, 3909, 4669, 4673]. The potential pitfalls include not only (1) the aforementioned LOF and GOF artifacts, but also (2) reporter anomalies (e.g., perdurance of β -gal [3764, 4188]), (3) antibody limitations (e.g., misleading epitopes on proteins that are cleaved [155, 3271] or reshaped [1990]), (4) confocal illusions [3293, 4760], and (5) *in vitro* infidelities relative to *in vivo* conditions [655, 871]. For the next generation of researchers, some of the parables may sound quaint, but for those of us who toiled through this period, they are a chronicle worth preserving.

Readers accustomed to color photos may bemoan the book's reliance on black-and-white diagrams. I am sorry for any disappointment. The latter style just seemed more fitting for an abstract analysis. All the figures were drawn in ADOBE *Illustrator* by me (a hopeless attempt to compete with my truly artistic siblings). They evolved from cartoons into montages. When many grew too big to fit the standard 6 $\times$ 9-in. size of this series, I tried breaking them into pieces but found that the surgery was lethal. The montages had acquired a life of their own. They tell whole stories (some of which spill over into App. 7). I thank Cambridge for approving a

larger trim size and for letting me set my own deadline. The cusp of the millennium seemed an apt time to step back and take a wide-angle "snapshot" of this blossoming field. The last batch of citations came from the annual *Drosophila* Research Conference (in Washington, DC) entitled, "2001: A Fly Odyssey."

This project began in 1992 when Robin Smith (then Life Sciences Editor at Cambridge) asked me to write a book for this series at the behest of Paul Green (a series editor). The topic took shape gradually, and the contract was signed in 1996. By 1997, my other professional pursuits had to be sidelined as the writing became all-consuming. I thank Peter Barlow (another series editor) for calming my fears and Ellen Carlin (Assistant Life Sciences Editor) for trusting my judgment.

Encouragement was provided by my dear parents (Maj. Gen. Lewis I. Held and Minnie Cansino Held), siblings (Lloyd, a.k.a. Grey, and Linda), other relatives and sundry friends – most of whom remain skeptical that any sane adult can adore flies. Maybe this book will change their minds? Probably not!

Critical comments on portions of the manuscript were kindly furnished by Seth Blair, Tom Brody, Ian Duncan, Matt Gibson, Robert Holmgren, Teresa Orenic, Grace Panganiban, Amy Ralston, Allen Shearn, David Sutherland, and Tanya Wolff. The idea about Notch and Argos in the Skeptic-Theorist debate (Ch. 6) was Seth's. I regret any overlooked errors.

As one foot soldier in the global army of fly pushers, I have met many "generals" over the years who figure prominently in this saga. By far the greatest – and humblest – was Curt Stern. His musings on the mysteries of patterning were the siren songs that lured me to this lovely fly. Those of us who heeded his call have long dreamt of finding insights one day. Little did any of us suspect, though, that the bounty of revelations in the last decade would go so far beyond merely sating our curiosity. As we sift the treasure, the sparkle of so many answers is fostering – even in the saltiest among us – a profound sense of awe.

Lewis I. Held, Jr.
Lubbock, Texas
April 2001