Supplementary Information for "The Emergence of Geometric Order in Proliferating Metazoan Epithelia"

Matthew C. Gibson, Ankit B. Patel, Radhika Nagpal & Norbert Perrimon

This supplementary document presents two extensions of the mathematical model, as well as a more detailed description of the Drosophila experimental data. In addition, we present some clarifications on the graph model and a detailed methods section.

Table of Contents:

1. Graph Model: Derivation in the Presence of Boundary Conditions

In the paper we *approximate* s_t, the average number of sides per cell at generation t as:

$$s_t \approx 2e_t / f_t = 2(e_{t-1} + 3f_{t-1}) / 2f_{t-1} = (s_{t-1} / 2) + 3 \qquad (1)$$

We solve this recurrence to get:

$$s_t \approx 6 + 2^{-t}(s_0 - 6) \qquad (2)$$

However, this equation for s_t is only an *upper bound approximation* because it double counts edges on the boundary of the epithelium. If we account for boundary cells, then the exact equation for s_t is:

$$s_t = (2e_t - e_t^{bd}) / f_t = 6 + 2^{-t}(s_0 - 6) - e_t^{bd} / f_t \qquad (3)$$

where e_t^{bd} is the number of boundary edges and f_t is the number of cells at generation t. This basically is the same as equation (1) but corrects for the over counted edges.

In a 2D epithelium comprised of similarly-sized cells, the number of boundary edges approximately measures the perimeter and the number of cells approximately measures area. Hence $e_t^{bd} \sim O(r)$ and $f_t \sim O(r^2)$, where r is the radius of the epithelium (in cells). The third term in equation (3) is therefore $O(1/r)$ and this approaches zero as t gets large and the epithelium gets large. This implies that *the average number of neighbours per cell will approach six* from below. In fact, the convergence rate is exponential since $f_t \sim$

$O(2^t)$ and $e_t^{bd} \sim O(\sqrt{f_t})$. [Note that this would not be true for a strip of cells that proliferate in a single direction since then $e_t^{bd} \sim O(f_t)$.]

Thus, the original approximate formula for s_t in (2) is still asymptotically correct. For a roughly circular region of 200 cells, the difference between equations (2) and (3) is less than 4% and for 30,000 cells it is less than 0.4%. Therefore equation (2) is a good approximation for the sizes of epithelial regions we sample, and a very good approximation for the imaginal disc itself, which eventually has well over 30,000 cells.

An important aspect of this analysis is that it is independent of cleavage plane orientation (condition 6), and is instead a deeper implication of the formation of mostly tricellular junctions (vertices of degree 3). Notice however that this result by itself is not sufficient to explain the predominance of hexagons or even their existence– for example, a planar graph with equal numbers of octagons and quadrilaterals has only tricellular junctions and an average of six neighbours, but no hexagons.

2. Extended Markov Model: Error in Interface Formation

One of the discrepancies between the mathematically predicted distribution and the experimentally observed distribution is the absence of 4-sided cells. The Markov model allows the creation of 4-sided cells through division but assumes that at the end of a round of division each cell has gained exactly one edge. As a result all 4-sided cells become 5-sided in each round, after gaining edges from neighbors, and thus the model predicts no 4-sided cells. In our experimental data, a small fraction of 4-sided cells are observed (2-3%). There are several potential factors that can explain the difference between the prediction and observation:

1. The experimental data does not represent the end of a round of division, and therefore not all cells will have gained a neighbor in the current round. Thus one should expect to see some 4-sided cells that have been created through splitting but not gained an edge from a neighbor yet.
2. In the Markov model, we assume that each cell gains exactly one edge. However in reality a cell may gain more or less. More precisely, the number of sides added via neighbor division is a random variable obeying a particular distribution. We prove that this distribution has a mean of exactly 1. By using the shift matrix S we are making the mean-field approximation, i.e. that this mean is "good enough" to calculate the equilibrium distribution. Our experiments with aged and mitotic cells suggest that the mean-field approximation is reasonable. This approximation likely adds a small amount of error to the prediction for all cell types, but is most obvious for 4-sided cells.
3. Our model assumes that all divisions result in the formation of a new interface between the daughters (condition 3). However, it is possible that this fails with some small rate as shown in our data on the occurrence of non-interface forming clones (Text Fig. 1). While we cannot easily measure the exact rate of failure, we

can model the effect of different error rates in interface formation, as shown below

Experimental data on interface formation

In the text we define three types of cell division, Type I, II and III (Text Fig. 1). Type I is a division that results in the formation of a normal interface between daughters. Type II division results in two daughters but no new interface is formed so they are connected only at a vertex. Type III similarly has two daughters that fail to form a new interface but also get separated.

In our experimental data on two-cell clones, 94% of the divisions are unambiguously Type I. However this is a conservative estimate, and the remaining cases are not unambiguously Type II/III. This is primarily because it is difficult to distinguish Type II from cytokinesis that has not yet resolved into an interface. More importantly, it is also difficult to distinguish a Type III division from two independent clone events that occur within a 1-cell radius (which was our scoring criteria for Type III). For these experiments, clones were tested within a 10-hour period, during which they are likely to divide once. However, some cells may not divide and some cell clones may occur close to each other. Therefore there are a small number of single cells and larger clones. If we consider only two-cell clones, then the fraction of non-Type I events is < 3%, whereas using all clones results in 6% non-Type I events. Including the larger clones increases the likelihood that independent events are mistaken for Type III divisions. In summary, the failure rate is difficult to measure unambiguously, however we expect it to be less than 6% and most likely around 2%.

Modeling error in interface formation

While our Markov model is parameterless, it is possible to introduce parameters to investigate failures in the conditions underlying the model. We introduce a parameter μ equal to the probability that a cell divides without forming a new interface (i.e. Type II or III). Hence with probability (1-μ) the cell divides normally, forming an interface between daughter cells (i.e. Type I). Our experimental data suggests that μ is between 2-6%.

We can model this process as a Markov Chain with the following division matrix:

$$P_\mu = (1-\mu)P_I + \mu P_{II/III}$$

where P_I is the normalized Pascal Matrix defined in the text for Type I divisions (Text Box 1) and $P_{II/III}$ is a new matrix that accounts for Type II/III divisions wherein a new interface is *not* formed. If the two daughters fail to form a new interface then they end up with one less side each. Thus the number of sides of a daughter in a non-interface forming division is just one less than an interface-forming one.

$$P_{II/III}(i,j) = P_I(i,j+1) \text{ for } i >= 4.$$

Thus $P_{II/III}$ looks like P_I, but shifted one column to the left:

$$\text{(unnormalized) } P_{II/III} = \begin{matrix} 1 & & & & & \\ 1 & & & & & \\ 1 & 1 & & & & \\ 1 & 2 & 1 & & & \\ 1 & 3 & 3 & 1 & & \\ 1 & 4 & 6 & 4 & 1 & \\ 1 & 5 & 10 & 10 & 5 & 1 \end{matrix}$$

Notice that the new Markov Chain is defined on the state space {3,4,5,6,…} because the division of a 4-sided cell can result in the formation of 3-sided daughters when interface formation fails. The division transition matrices P_I and $P_{II/III}$ have the extra provision that a 3-sided cell divides to form two 3-sided cells, to avoid the degenerate case of a 2-sided daughter cell.

***Modeling only Type II division effects*:** In Type II divisions, the two neighboring cells that are cut by the cleavage plane will gain one side each, regardless of whether a new interface is formed or not. Thus the shift matrix does not change, and $S_\mu = S_I$ where S_I is simply the standard shift matrix from the text.

The overall transition matrix is $T_\mu = P_\mu S_\mu$ and by the Perron-Frobenius theorem the overall process $p^{(t+1)} = p^{(t)} T_\mu$ converges to an equilibrium distribution p^* where p^* is the principal eigenvector of T_μ. Using MATLAB, we calculate $p^*(\mu)$ for several values of μ (Supplemental Figure 1). For small μ, as μ increases, p_4^* and p_5^* increase while p_6^*, p_7^*, p_8^*, p_9^* decrease. This leads to a decrease in average sidedness i.e. $s^*(\mu) < 6$ for $\mu > 0$. This is expected since a higher μ implies less new sides are formed and thus an overall decrease in sidedness. At the extreme when $\mu=1$, i.e. an epithelium with purely non-interface forming divisions, all cells eventually become 4-sided.

This analysis shows that, for small values of μ that are consistent with experimental estimates, the equilibrium polygon distribution *E* is robust (Supplementary Fig. 1a).

***Modeling Type III division effects*:** In the case of Type III divisions, neighbouring cells gain two edges each, rather than a single edge. Thus we must modify the shift matrix to allow +2 sides to the neighboring cells for such divisions. Let p_T = probability that a cell has division type T, where T is one of {I,II,III}. A fraction ($p_I + p_{II}$) of cells contribute +1 side to each of their neighbors that are cut by the cleavage-plane in Type I/II divisions. A fraction (p_{III}) of cells contribute +2 to each such neighbor (Text Fig. 1). Hence, the new shift matrix is:

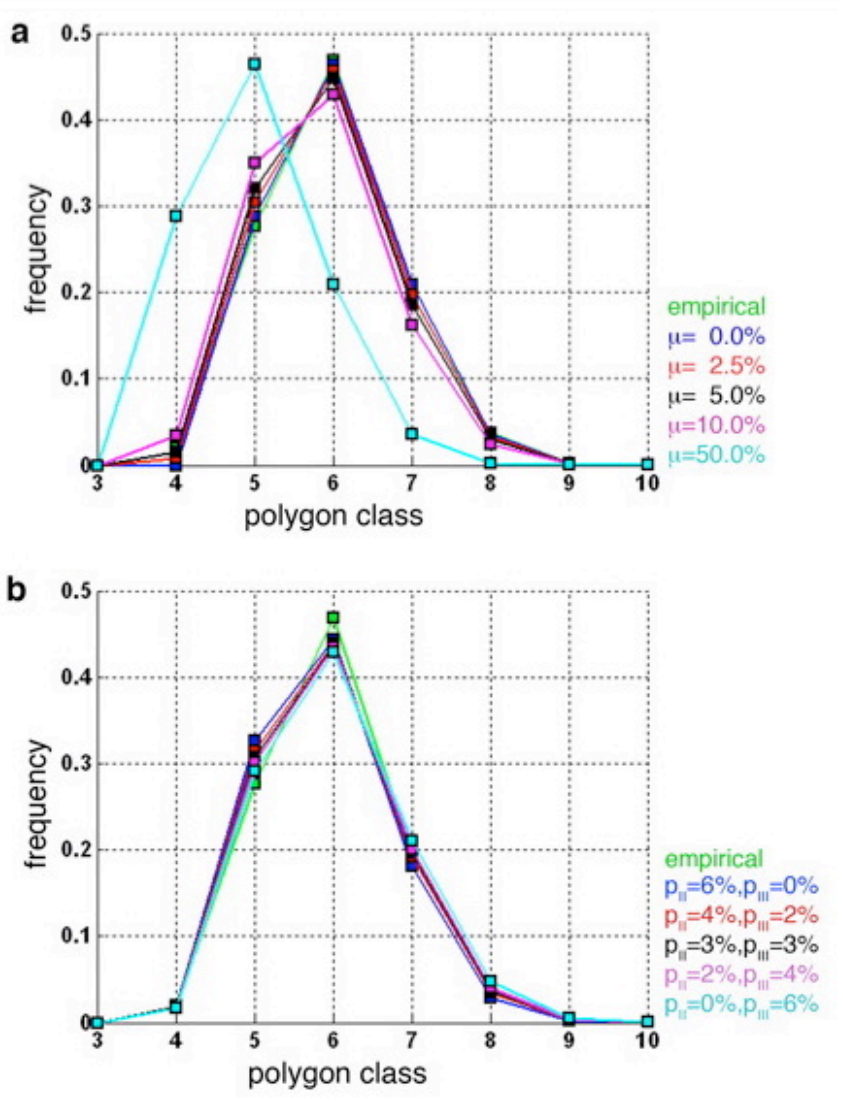
$$S(p_I, p_{II}, p_{III}) = (p_I + p_{II})S_I + p_{III}S_I^2$$

where S_I is the original shift matrix from the paper [Text Box 1].

The division transition matrix $P(p_I, p_{II}, p_{III})$ is just P_μ with $\mu = (1-p_I)$. Using $P(p_I, p_{II}, p_{III})$ and $S(p_I, p_{II}, p_{III})$, we can calculate p^* for several values of p_I, p_{II}, and p_{III} where $\mu = p_{II} + p_{III} = 6\%$. Once again, we find that the equilibrium distribution E remains robust to small errors, whether Type II or Type III (Supplementary Fig. 1b).

In summary,

1. For an error rate of $\mu = 2\text{-}6\%$, the equilibrium distribution remains largely unaffected. This is true for both Type II and III errors. Thus the system is robust to small errors in interface formation.
2. Cases where $\mu = 6\%$ predict that a small percentage of 4-sided cells (~2%) will be formed, which is consistent with observed percentages of 4-sided cells (~3%).
3. Lastly, this analysis suggests how one can create parameterized versions of the model to investigate small deviations from the idealized conditions.



Supplemental Figure 1.
(a) Equilibrium polygon distribution (*E*) as a function of the rate of non-interface-forming divisions, μ. As μ increases, the frequency of 4- and 5-sided cells increases while 6-, and higher-sided cells decrease implying that the overall mean decreases. For small rates of face-formation error, E is robust.
(b) Equilibrium polygon distribution (*E*) as a function of the rate of Type II and Type III non-face-forming divisions (p_{II} and p_{III}), for a fixed total error rate μ=6%. *E* is robust to small rates of Type II/III face formation errors

3. Extended Markov Model: Alternate Cleavage Plane Models

Here we examine alternate models of cleavage plane choice that illustrate how the transition probabilities in the division matrix P affect the final equilibrium distribution p^*. In particular, we show that changes in P lead to noticeable changes in p^*.

Uniformly at random sides: Consider a division model where each cell "chooses" two non-adjacent sides uniformly at random. More specifically (1) choose side s_1 at random from all n sides of a cell (2) choose side s_2 uniformly at random from the $n-3$ remaining non-adjacent sides (3) Create a cleavage plane between s_1 and s_2. This model has the following division transition matrix:

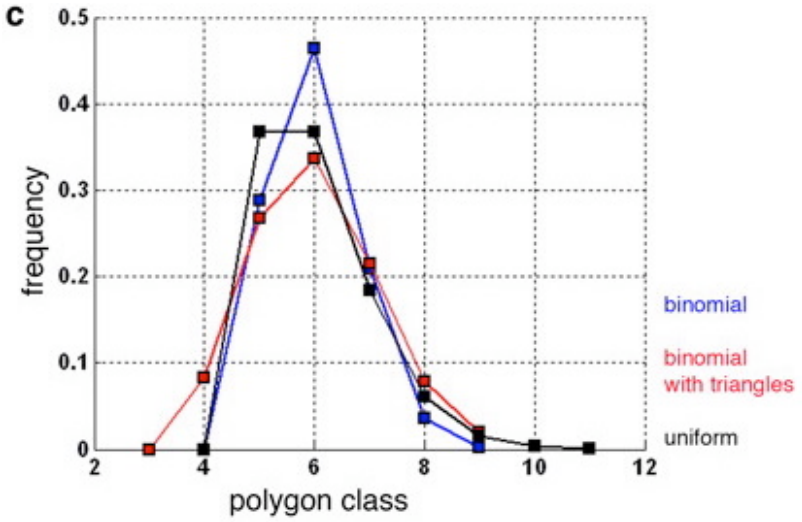
$$\text{(unnormalized) } P_U = \begin{matrix} 1 & & & & \\ 1 & 1 & & & \\ 1 & 1 & 1 & & \\ 1 & 1 & 1 & 1 & \\ 1 & 1 & 1 & 1 & 1 \end{matrix}$$

Note that P_U only differs from the Pascal matrix P_0 in rows 3 and higher. The shift matrix S_U remains the same as the standard shift matrix S_0. Using MATLAB, we calculate the equilibrium distribution for the overall Markov process with transition matrix $T_U = P_U S_U$ and get p* = [0.3679 0.3679 0.1839 0.0613 0.0153 0.0031 0.0006] (Supplemental Fig. 2; *uniform*).

Randomly distributed junctions, allowing 3-sided cells. In this case we modify the original division model (Text Box 1) to allow the case where a 3-sided daughter can be formed. We still disallow 2-sided cells (which could occur if no junction falls on one side of the cleavage plane). In this case the transition probability is defined as $P_T(i,j) = Comb(i-2, j-3)$ where $Comb(n,k) = n!/k!(n-k)!$, the number of k-subsets of a set of n elements. We can calculate the equilibrium distribution for the overall Markov process with transition matrix $T_T = P_T S_T$ and get p^* = [0.0000 0.0835 0.2682 0.3360 0.2144 0.0782 0.0197] (Supplementary Fig. 2; *binomial with triangles*). This shows that the condition disallowing 3-sided cells is important, even though we do not yet know the specific mechanisms that prevent it.

Equal junction split. Another simple model of division is equal junction split where a cell always divides its tricellular junctions as equally as possible between daughters. For even-sided cells $P_E(i,i/2) = 1$ and for odd-sided cells $P_E(i,(i-1)/2) = P_E(i,(i+1)/2) = 1/2$. For example, a 6-sided cell gives 3 tricellular junctions to each daughter while a 7-sided cell gives one daughter 3 tricellular junctions and the other daughter 4 tricellular junctions. This system is predicted to converge to a population of all hexagons. In reality, neighbors can gain more or less than one side, and therefore it is unlikely that one will observe a perfect hexagonal lattice. However the system is driven to a much tighter distribution than the previous models.

From Supplementary Fig. 2, it is clear that for each different model, the equilibrium distributions deviates significantly from p_0^* and thus the Markov model can discriminate between these cases. Notice that all the systems must have a mean of 6. However, which cell type is predominant and what percentage of predominant cell type is observed can vary considerably. Importantly, while we cannot prove that no other division mechanism will generate p_0^*, the distributions may indicate which division mechanisms are insufficient to generate p_0^*. Thus in cases where we cannot observe the division process directly, we may be able to gain some insight into potential cell-level mechanisms.



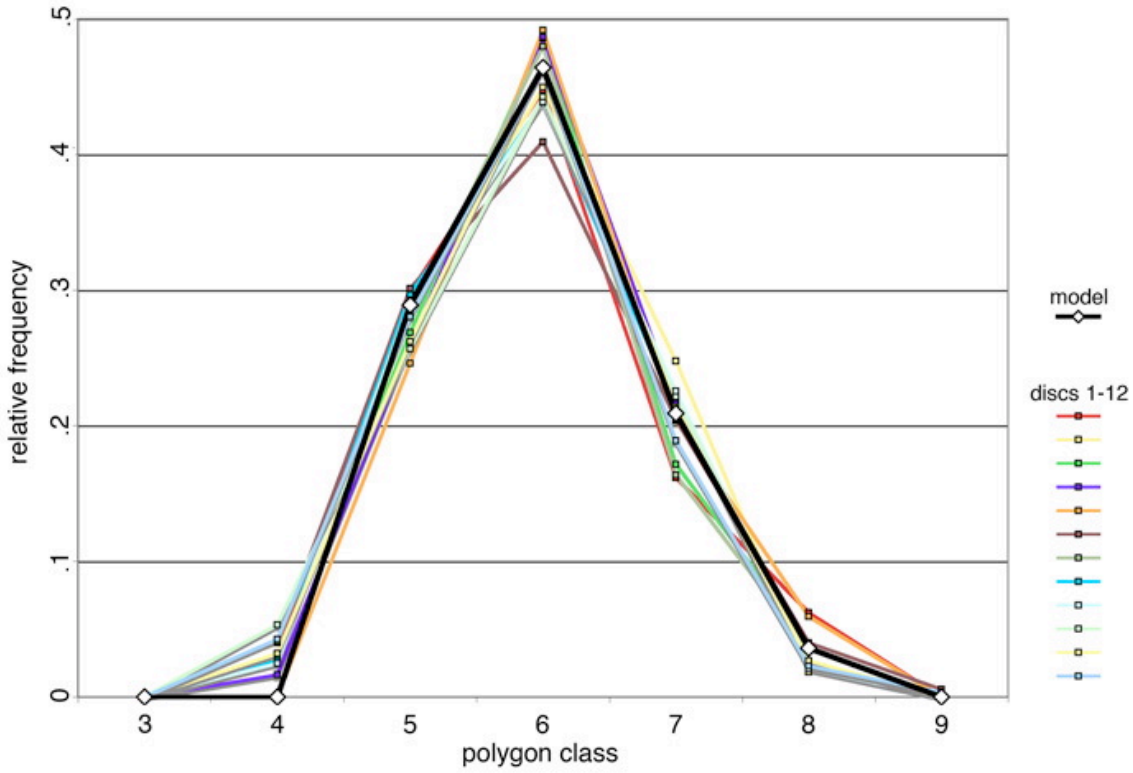
Supplemental Figure 2. Equilibrium polygon distribution (*E*) as a function of cleavage plane choice. Allowing 3-sided cells increases the variance and decreases the skew of *E*. Choosing cleavage planes uniformly at random, as opposed to binomially, increases the variance by increasing the frequency of 5-sided cells while decreasing the frequency of 6-sided cells. Thus *E* is sensitive to changes in cleavage plane choice.

4. Imaginal Disc Polygonal Cell Counts

In the paper we present average and standard deviation results from counting the polygonal cell types in fixed *Drosophila* imaginal discs. Here we present the data from the individual imaginal discs. The data is from twelve imaginal discs at the third instar larval stage, collected on two separate days. For each disc we examined an arbitrary region of ~100-300 cells in the presumptive blade region (excluding the wing margin territory).

The scoring of late-third instar imaginal disc cells was performed by hand on high-resolution digital images (raw data in Supplementary Table 1). While scoring polygon class for most cells was unambiguous, some errors in counting are likely. Images were ambiguous for a number of reasons: 1) Focal plane effects due to which the septate junction was not clearly resolved; 2) Variation in the length of cell sides; and 3) Cells adjoining mitotic cells were distorted, making it difficult to discern cell contacts.

Despite these caveats, the scored images showed a remarkably consistent distribution with a small deviation from the average (Supplementary Fig. 3). Importantly, each disc exhibited similar properties: no cells with less than four sides or more than nine sides were observed, the median cell sidedness was six, and the percentage of seven-sided cells was always lower than that of five-sided cells.



Supplemental Figure 3. Empirically observed polygon distributions for 12 imaginal discs in the late third instar stage (100-200 cells per disc). Note that disc-to-disc variations in the distribution are small and do not affect the overall characteristic shape.

Polygon distribution/ disc	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***total cells***
1	0	3.11	29.81	44.72	16.15	6.21	0	161
2	0	0	28.44	44.95	24.77	1.83	0	109
3	0	2.86	26.86	48.57	17.14	4	0.57	175
4	0	1.62	25.95	48.65	21.62	2.16	0	185
5	0	0	24.58	49.15	20.34	5.93	0	118
6	0	3.98	30.11	40.91	20.45	3.98	0.57	176
7	0	4.09	28.07	47.95	16.37	3.51	0	171
8	0	2.65	29.65	43.81	21.24	2.65	0	226
9	0	2.46	29.1	43.85	22.54	2.05	0	244
10	0	5.31	25.66	44.25	22.12	2.65	0	113
11	0	3.21	26.2	46.53	21.39	2.67	0	187
12	0	4.23	28.01	46.25	18.89	2.28	0.33	307
average	0	2.79	27.7	45.8	20.25	3.33	0.12	**2172**
std. dev.	0	1.55	1.75	2.39	2.53	1.4	0.22	
max.	0	5.31	30.11	49.15	24.77	6.21	0.57	
min.	0	0	24.58	40.91	16.15	1.83	0	
Markov model	0	0	28.9	46.4	20.9	3.6	0	

Supplemental Table 1. Frequency counts of data shown in Supplemental Figure 3, presented as polygon distributions per disc for twelve imaginal discs. No 3-sided or 10-sided cells are observed. The median is 6 and the frequency of 5-sided cells is greater than 7-sided cells for all discs. Summary statistics per polygon class and overall are shown.

5. Detailed Methods

Confocal Timelapse

Imaginal discs were dissected in droplets of Ringer's solution on siliconized slides and immediately transferred into culture media on standard glass slides. Strips of double-stick tape were employed as spacers to create a shallow chamber between the coverslip and the specimen. For short duration *ex vivo* culture of imaginal discs, it was possible to track mitotic cell junction dynamics in a variety of media. We used standard Ringer's Solution (130 mM NaCl, 5mM KCl, 1.5 mM $MgCl_2$) as well as Shields and Sang M3 Insect Media (Sigma-Aldrich; modified with 10% Fetal Bovine Serum, 10 mU/L insulin, 10U/mL penicillin, 10 ug/mL streptomycin) to obtain movies over culture periods of 1.5-2 hours at maximum. Movies were collected and processed using the Leica TCS SP2 AOBS Confocal Microscope system.

Animal Husbandry and Imaging

Drosophila stocks were maintained using standard methods at 25^{o}C. GFP-expressing mutant clones were induced with a 15-minute heat shock at 37^{o}C followed by a 10-hour recovery period. For immunocytochemistry and Phalloidin staining, imaginal discs were fixed in 4% paraformaldehyde in PBS at room temperature for 30-40 minutes. All subsequent steps were performed according to standard protocols. Phalloidin-546 (Molecular Probes) was used at a dilution of 1:250, Mouse anti-Discs Large (Developmental Studies Hybridoma Bank) was used at 1:500. Secondary antibodies were Goat anti-Mouse Alexa 647 or Goat anti-Rabbit Alexa 647 (1:500; Molecular Probes). Polygon counts were taken from nine late third instar larval imaginal discs. *Xenopus* tadpoles, approximately stage 44-50[26], were obtained live (Carolina Biological Supply) and immediately fixed in cold 4% paraformaldehyde in PBS. Tails were removed and stained overnight in 1:200 Phalloidin-546 at 4^{o}C in PBT and then rinsed 3X in PBT at room temperature for 2 hours. For examination, the epidermis was dissected away from the muscle prior to mounting. Counts were taken from five samples. Wild-caught *Hydra* populations (Carolina Biological Supply) were maintained in Poland Spring Water in plastic dishes on a diet of freshly-hatched brine shrimp nauplii (Carolina Biological Supply). For fixation, animals were permitted to engorge themselves with shrimp larvae and then fixed whole with 4% paraformaldehyde in Poland Spring Water. Staining was overnight at 4^{o}C in a 1:200 dilution of Phalloidin-546 in PBT, followed by 3 rinses in PBT over 2 hours at room temperature. Polygon counts were taken from five samples. All specimens were mounted in 70% glycerol/PBS using spacers as appropriate. Images were collected on a Leica TCS SP2 AOBS Confocal Microscope system and processed using Adobe Photoshop 7.0 software.

Clonal analysis and polygon distributions

Clones were induced in flies of the genotype: *yw hs-flp^{122}; Actin5c > >Gal4, UAS-GFP/+*. Polygon distributions were determined by eye in confocal micrographs; error was estimated as the average standard deviation between counts from different images.

Empirically, it was not possible to systematically account for certain rare but inevitable irregularities in real epithelia, such as occasional 4-way point junctions and dying or grossly misshapen cells. Indeed, such irregularities were most common in *Xenopus* epidermis, which deviated the most significantly from expectations. There are clearly regulative mechanisms and spatial irregularities in epithelia not accounted for by the present study, and overall polygon counts must therefore be taken as approximations that illustrate general design principles of default-state epithelial monolayers. The raw counts for cells of different sidedness are as follows: *Drosophila* disc columnar epithelium (*4*- 64; *5*- 606; *6*- 993; *7*- 437; *8*- 69; *9*- 3). *Hydra* (*4*- 16; *5*- 159; *6*- 278; *7*- 125; *8*- 23; *9*- 1). *Xenopus* (*3*- 2; *4*- 40; *5*- 305; *6*- 451; *7*- 191; *8*- 52; *9*- 8; *10*- 2). *Drosophila* peripodial controls (*4*- 11; *5*- 106; *6*-198; *7*-86; *8*- 10; *9*- 0). *Drosophila* peripodial *string* clones (only cells on the clone periphery were scored: *3*- 3; *4*- 27; *5*- 134; *6*-105; *7*-21; *8*- 5; *9*- 0).

Markov Chain convergence and equilibrium distribution calculation

The Perron-Frobenius Theorem in Markov Chain theory guarantees the convergence of $\boldsymbol{p}^{(t)}$ to a unique stable equilibrium distribution $\boldsymbol{p}^{*}$ independent of the initial distribution $\boldsymbol{p}^{(0)}$, provided that the chain is irreducible and aperiodic. Note that in our case all probability mass leaves state 4 of our Markov chain in 1 iteration (Text Fig. 2). Hence the Markov chain is technically defined only on states 5 and higher, with state 4 as a transient state. Restricted to this smaller state set, it is both irreducible and aperiodic and thus the theorem guarantees the existence of a unique stable equilibrium[23]. Furthermore the equilibrium distribution is the first eigenvector of the transition matrix T, and the second eigenvalue determines the rate of convergence. We truncated the infinite transition matrix *T* down to 20 rows and 20 columns and then used Matlab to calculate $\boldsymbol{p}^{*}$ as the principal eigenvector of *T*. We also used Matlab to compute λ_2, the second largest eigenvalue of *T* as 0.5, which implies that the topology approaches the predicted equilibrium distribution exponentially fast.