Data-Driven Discovery of Active Nematic Hydrodynamics

Chaitanya Joshi[1,2,*] Sattvic Ray[3] Linnea M. Lemma[3,1] Minu Varghese[4,1] Graham Sharp[3]
Zvonimir Dogic[3,†] Aparna Baskaran[1,‡] and Michael F. Hagan[1,§]

[1]*Department of Physics, Brandeis University, Waltham, Massachusetts 02453, USA*
[2]*Department of Physics and Astronomy, Tufts University, Medford, Massachusetts 02155, USA*
[3]*Department of Physics, University of California at Santa Barbara, Santa Barbara, California 93106, USA*
[4]*Department of Physics, University of Michigan, Ann Arbor, Michigan 48109 USA*

Active nematics can be modeled using phenomenological continuum theories that account for the dynamics of the nematic director and fluid velocity through partial differential equations (PDEs). While these models provide a statistical description of the experiments, the relevant terms in the PDEs and their parameters are usually identified indirectly. We adapt a recently developed method to automatically identify optimal continuum models for active nematics directly from spatiotemporal data, via sparse regression of the coarse-grained fields onto generic low order PDEs. After extensive benchmarking, we apply the method to experiments with microtubule-based active nematics, finding a surprisingly minimal description of the system. Our approach can be generalized to gain insights into active gels, microswimmers, and diverse other experimental active matter systems.

Active nematics demonstrate how energy-consuming motile constituents can self-organize into diverse non-equilibrium dynamical states [1–3]. They offer a versatile platform to advance our fundamental understanding of nonequilibrium physics and develop materials with properties that are thermodynamically forbidden in equilibrium. These twin goals require theoretical models that reveal the mechanisms underlying the emergent dynamics, and guide rational design to elicit desired spatiotemporal dynamics. Here, we combine data-driven model discovery with experiments and computational modeling to identify the most parsimonious model for an experimental realization of active nematics. Using the discovered model, we identify the relationship between key theoretical parameters, such as the magnitude of activity, and experimental control variables. The described methods can be applied to diverse active nematics ranging from shaken rods to motile cells [4–9], and other forms of active matter.

Our target is a quantitative description of microtubule-based active nematics. Being reconstituted from tunable and well-characterized components, they afford a unique opportunity to develop continuum theory models and connect these to the microscopic dynamics [10–12]. Hydrodynamic theories, built on purely symmetry considerations, have provided insight into dynamics of active nematics in general, and the microtubule-based system specifically. For example, such models have been used to describe defect dynamics [13–18], induced flows in the suspending fluid [19–21], and how confinement in planar [22–24] and curved geometries [25–27] controls defect proliferation and dynamics. These efforts employed a range of hydrodynamic models that assumed different symmetry-allowed terms, and the parameters of the model were largely undetermined. Thus, the field lacks a quantitative model and understanding of magnitudes and sources of error in existing approximations.

Data-driven approaches and machine learning have been successfully applied to study active matter [28]. However, previous studies for active nematics were limited to parameter optimization with a preassumed model [16,29], or machine learning forecasting [30,31] which, while successful, does not provide an analytical equation for the learned dynamics. To overcome these limitations, we build on the Sparse Identification of Nonlinear Dynamics (SINDy) framework [32,33] that was recently applied to particle-based simulations of active matter [34] and computational and experimental data of overdamped polar particles [35]. This method filters out the best parsimonious fit to the data from a highly generalized class of potential models. We adapt key improvements of this method [36–39] to the microtubule-based active nematics system. We then employ extensive birefringence and fluorescence measurements of microtubule alignment and PIV (particle image velocimetry) measurements of velocities, to identify equations governing both the orientational dynamics and the activity-driven flows. This enables direct inference of the underlying model that is rigorously supported by experimental data. In contrast to the hard-to-interpret deep neural nets generated by machine learning approaches, our method yields an optimal analytical model and estimated parameter values.

With the available alignment and velocity measurements, we seek models describing the active nematic as a single

2D fluid with nematic symmetry [12,40]. Hence, there are two fields: the symmetric-traceless nematic tensor order parameter $\mathbf{Q} = s[\mathbf{n} \otimes \mathbf{n} - (1/2)\mathbf{I}]$ and a flow field $\mathbf{u}$, with $\mathbf{n}$ as the local orientation unit vector and s the scalar order parameter. We assume constant density and an incompressible fluid: the former is justified since the scalar order parameter captures density variations near the defects (see below); the latter is validated by numerical measurements of the divergence of the velocity field [41]. Our model then consists of four independent scalar fields: Q_{xx}, Q_{xy}, u_x, u_y, and a latent variable P (pressure).

We begin by postulating the generalized form of the model. The Q-tensor dynamics takes the form common to all continuum theories of active nematics:

$$\partial_t Q_{ij} = \sum_k a_{ij}^k F_k(\mathbf{Q}, \mathbf{u}, \nabla\mathbf{Q}, \nabla\mathbf{u}, \ldots) \qquad (1)$$

where the F_k's are combinations (potentially nonlinear) of $\mathbf{Q}$, $\mathbf{u}$, and their spatial derivatives up to a maximum order, and the a_{ij}^k's are the corresponding phenomenological coefficients. For instance, in 2D, a well-known model for the Q equation is [12]:

$$\partial_t \mathbf{Q} + \mathbf{u} \cdot \nabla\mathbf{Q} - \mathbf{S} = D_r \mathbf{H} \qquad (2)$$

where $\mathbf{S} = -(\mathbf{\Omega} \cdot \mathbf{Q} - \mathbf{Q} \cdot \mathbf{\Omega}) + \lambda\mathbf{E} - 2\lambda\mathbf{Q}(\mathbf{Q}:\nabla\mathbf{u})$ is the corotation term and $\mathbf{H} = a_2\mathbf{Q} + a_4\mathrm{Tr}(\mathbf{Q}^2)\mathbf{Q} + K\nabla^2\mathbf{Q}$ is the negative gradient of the liquid crystal free energy. Here, $E_{ij} = (\partial_i u_j + \partial_j u_i)/2$ and $\Omega_{ij} = (\partial_i u_j - \partial_j u_i)/2$ are the strain rate and vorticity tensors, respectively, λ is the flow alignment parameter, D_r is the rotational diffusion coefficient, K is the elastic constant, and $a_2 > 0$, $a_4 < 0$ are phenomenological coefficients corresponding to the isotropic-nematic transition. (See Supplemental Material [42], which includes Refs. [30,43–47], for further discussion.) We build a library of the terms F_k ($n = 246$ terms) that can capture models well beyond Eq. (2). Further, we make no physics-based simplifying assumptions, e.g., translational, rotational, and Galilean invariance, for the alignment equation [Eq. (1)]. Hence, discovery of a model which satisfies these conditions is a test of the algorithm (see Supplemental Material [42]).

For the flow equation, the usual form assumed for model discovery is Navier-Stokes-like, with the time derivative on the left side and rest of the terms on the right side [33,37–39]. However, because the active nematic is in the low Reynolds number regime [15,20], the significance of the time-derivative term itself needs investigation. Indeed, active nematic flows have been modeled using pure Stokes [30,48,49], unsteady Stokes [20,50], and full Navier-Stokes [12–14,19,22,51–55] formulations. While these approaches have been compared numerically [56], there has yet to be a definitive indication of the contributions of the inertial terms for this system. Since the viscous

forcing is guaranteed to exist in this regime, we assume a form

$$\nabla^2\mathbf{u} = c_0\partial_t\mathbf{u} + \sum_i c_i\mathbf{H}_i(\mathbf{Q}, \mathbf{u}, \nabla\mathbf{Q}, \nabla\mathbf{u}, \ldots) \qquad (3)$$

with $\nabla \cdot \mathbf{u} = 0$, and the time derivative on the *right-hand side* so that its contribution can be evaluated. For instance, the lowest order symmetry-allowed "active stress" in the flow equation is the well-known $-\alpha\mathbf{Q}$, with $\alpha > 0$ being the extensile "activity" [12,57]. In our model form, this gives a general flow equation:

$$\nabla^2\mathbf{u} = c_0\partial_t\mathbf{u} + c_1\mathbf{u} \cdot \nabla\mathbf{u} + c_2\nabla P + c_3\nabla \cdot \mathbf{Q} + \ldots$$

with the coefficient c_3 as the ratio of the activity to the viscosity, α/η.

We perform model discovery from the data as follows [58]. Setting N_x, N_y, and N_t as the number of measurements in the two spatial dimensions and time, respectively, we randomly select m of the total $N_x N_y N_t$ space-time points. At each selected space-time point, we evaluate a linear system, e.g., for the Q_{xx} equation, $(\partial_t Q_{xx})_{m \times 1} = F_{m \times n} \cdot \vec{a}_{n \times 1}$. The derivatives are computed numerically, which amplifies noise in the data. To mitigate noise, we use two different approaches. In the *integral formulation*, for each of the m selected space-time points and n terms, we compute a local average in space and time in a small window (e.g., $5 \times 5 \times 5$ pixels) [36]. This approach is effective for model discovery, but leads to inaccurate parameter estimates for the flow equation—since pressure is not an observable in the experiments, we must perform the operation $\hat{z} \cdot \nabla\times$ on the flow equation [33,37], which adds one more order of derivatives, amplifying the noise. To obtain more powerful noise mitigation at the cost of additional analytical effort, we adapt a *weak formulation* of the PDE regression problem [38,39]. Briefly, we fit the data to the weak form of Eq. (3):

$$\int_{\Omega_k} \mathbf{w} \cdot [\nabla^2\mathbf{u} = c_0\partial_t\mathbf{u} + c_1\mathbf{u} \cdot \nabla\mathbf{u} + \ldots]. \qquad (4)$$

By choosing an appropriate test function $\mathbf{w}$ (such that the boundary terms vanish after integration-by-parts), we can move the derivatives from the noisy experimental data to the exact test functions, and also integrate out latent variables using integration-by-parts (in this case by making $\mathbf{w}$ divergence-free, see Supplemental Material [42]). The terms included in the library are in Table S1.

Next, we seek optimal fits to these equations with the minimum number of nonzero terms, thus yielding an interpretable model that accurately describes the data but avoids overfitting. To this end, we perform Ridge regression (least-squares gives similar results), starting with all the terms in the library, and then eliminating the least

important terms one by one to obtain a hierarchy of models [36]. Obtaining the R^2 value at each step, we plot the *optimality curve* as the logarithm of $(1 - R^2)$ as a function of the number of nonzero terms left in the model. We define the optimal number of terms n^* as the n value at which the second derivative of the curve is highest, indicating the largest drop in $\log(1 - R^2)$.

To demonstrate the validity of our approach, we first benchmark it against data generated by numerical simulations (Fig. S1, Table S2, which includes Refs. [59,60]). We consider two qualitatively different models for flow: one is purely Stokesian with substrate friction, and the other is unsteady Stokes flow [20,50]. After adding synthetic noise to the simulation data, we apply the integral formulation to the alignment equation and the weak formulation to the flow equation. The framework returns the correct equations with very small errors in the identified coefficients (Figs. S1, S2, and S3 [42]). Thus, we estimate important phenomenological parameters directly from the data, including the activity level α, bending modulus K, flow alignment coupling λ, and bulk free energy coefficients a_2 and a_4. Further benchmarking against varying window sizes and noise levels (see Supplemental Material [42]) indicates that the integral formulation benefits from high resolution, low noise data whereas the weak formulation benefits from a large amount (in space and time) of data.

Next, we perform model discovery on experimental microtubule-based active nematics [Fig. 1(a), Supplemental Material [42] which includes Refs. [61–66]]. Coarse-graining the director, we obtain a Q-tensor field that contains the spatially varying scalar order parameter and orientation [Fig. 1(b)]. The low-fluorescence-intensity regions, corresponding to low microtubule density near the defect cores, are correlated with the low-scalar-order-parameter regions, thus capturing the density variation near the defects (Movie S1 [42]). This justifies the constant density assumption. The velocity is obtained from PIV analysis [Fig. 1(c)]. We varied the ATP concentration, which determines the motor stepping speed and thus determines the structure and dynamics of active nematics. We collected the data on a field of view several vortex diameters wide [Fig. 1(c)] for long times (> 20 velocity autocorrelation times, defined below). In addition, we acquired one more dataset with higher resolution but a smaller field of view, denoted as the "HR-SF" data [67] (see Fig. S4 [42]). Optimality curves for the alignment and flow equations, respectively, [Figs. 1(d) and 1(e)] lead to the following optimal model:

$$\partial_t \mathbf{Q} = -\mathbf{u} \cdot \nabla \mathbf{Q} - (\mathbf{\Omega} \cdot \mathbf{Q} - \mathbf{Q} \cdot \mathbf{\Omega})$$
$$+ \mathbf{E} - 2(\mathbf{Q} : \nabla \mathbf{u})\mathbf{Q} + K\nabla^2 \mathbf{Q}$$
$$\eta\nabla^2 \mathbf{u} = +\alpha\nabla \cdot \mathbf{Q} + \nabla P. \qquad (5)$$
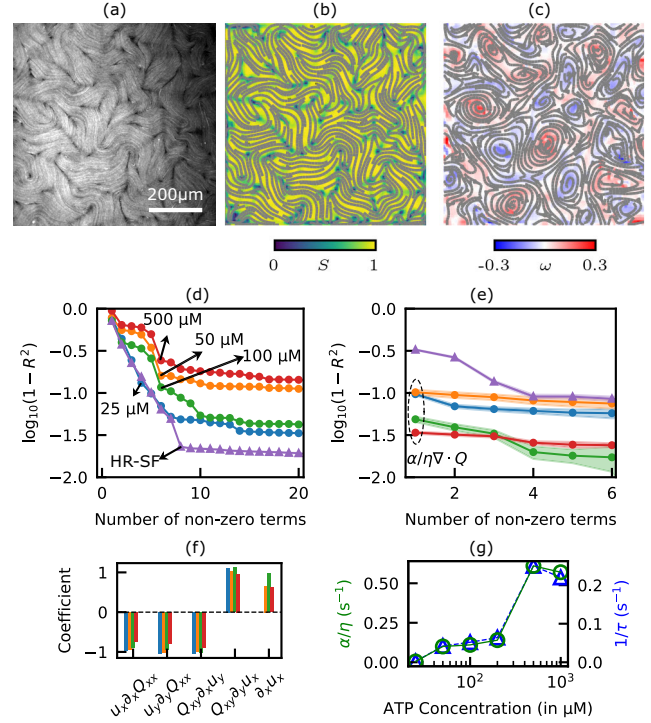

FIG. 1. Discovering active nematic hydrodynamics from experimental data. (a) A representative fluorescence image of the microtubule-kinesin active nematic at an ATP concentration of 100 μM (scale bar is 200 μm). (b) The computed director field and scalar order parameter S, and (c) the flow field and vorticity ω for the data in (a). (d) Optimality curves, $\log(1 - R^2)$ vs number of nonzero terms, for the Q_{xx} equation from the indicated datasets. The beginning of each arrow corresponds to the threshold corresponding to the highest-order term included in the optimal model. (e) Optimality curves for the weak-form flow equation. In the cases highlighted with the dashed oval, the optimal model contains only the activity term, $\nabla \cdot \mathbf{Q}$, consistent with Stokesian dynamics. In (d),(e) the purple triangles correspond to the high-resolution, small field-of-view (HR-SF) dataset, while the blue, orange, green, and red circles correspond to 25 μM, 50 μM, 100 μM, and 500 μM ATP, respectively. (f) Values of the coefficients of key flow-coupling terms appearing in the optimal models for various ATP concentrations. The colors are the same as in (d),(e). (g) The fit coefficient of the activity term, α/η, as a function of the ATP concentration (green circles). This quantity closely matches the inverse of the velocity correlation time (blue triangles), suggesting that α/η corresponds to a relevant timescale in the system.

Note that we added the term $K\nabla^2 \mathbf{Q}$ because this or an analogous term with higher order derivatives must be present for stability, discussed further below.

We arrived at this model as follows. For the alignment equation, the HR-SF dataset [purple triangles in Fig. 1(d)] has a low error ($R^2 = 0.97$) and an abrupt shoulder that clearly defines a threshold for the optimal model. In comparison, the lower resolution data sets have larger error (see Table S3 [42]) and less distinct thresholds.
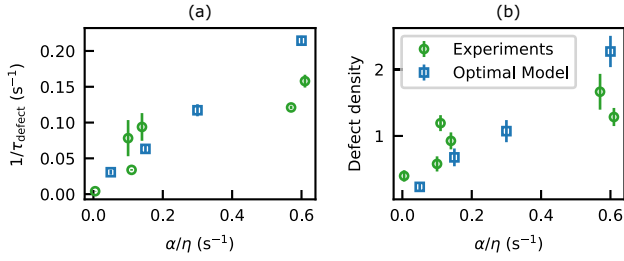
FIG. 2. Comparisons of results from simulations using the discovered optimal model against the experimental data. (a) Inverse lifetime of $+1/2$ defects plotted as a function of α/η for experiments at different ATP concentrations (green circles) and simulations using the optimal model performed with different values of α/η (blue squares). For the experiments, the value of α/η is obtained from the discovered optimal model at each ATP concentration. The height of the error bars is twice the standard error of mean. (b) Defect density plotted against α/η from experiments (green circles) and simulations using the optimal model (blue squares). The density in the simulations is scaled by a constant. The height of the error bars is twice the standard deviation.

Consistent with the benchmarking of numerical data described above and in Supplemental Material [42], these results show that high resolution is more important than a large field-of-view for determining the alignment equation. The threshold chosen for each dataset is indicated by the tail of the corresponding arrow in Fig. 1(d), and the resulting model for each data set is given in Table S3 [42]. Table S4 [42] gives the lowest-order terms beyond the threshold. For all datasets, the optimal model is dominated by flow-coupling terms, such as the convective and rotational derivatives and flow alignment. Equation (5) corresponds to the optimal model for the HR-SF dataset, and with the exception of the higher-order flow-alignment term $2(\mathbf{Q}:\nabla\mathbf{u})\mathbf{Q}$, two of the low-resolution datasets. For other datasets, there is some variability in the terms near the threshold (Tables S3 and S4), but the terms in Eq. (5) are all present near the threshold, and other near-threshold terms can be eliminated because they violate known symmetry or conservation criteria for the system. We include the higher-order flow-coupling term because the HR-SF dataset has the highest statistical accuracy and because it is expected theoretically for stability of the nematic order parameter. We attribute the variability in the near-threshold terms for the low-resolution datasets to statistical inaccuracies arising from the limited experimental data and the small contributions of these terms, rather than different physics being present at different ATP concentrations. These results highlight the importance of the amount and resolution of data for accurately determining the alignment equation.

The alignment equation recovers Galilean invariance from the data: the convective and corotational derivatives have coefficients of ~ 1 [Fig. 1(f)]. Furthermore, the flow alignment parameter, $\lambda \sim 1$ [Fig. 1(f)], is consistent with the theoretical result for the high aspect ratio $a \gg 1$ of the microtubules, $\lambda = (a^2 - 1)/(a^2 + 1) \to 1$ [68]. Importantly, the bulk liquid crystal free energy terms that stabilize nematic order [with coefficients $a_2 > 0$ and $a_4 < 0$, see Eq. (2)] are not present in the discovered model for any dataset [69]. This finding supports a previous model [54] which argued that active flow alignment acts as an effective free energy that drives nematic order. These results indicate that the alignment dynamics are dominated by flow coupling. In comparison, contributions from the free energy dissipation to the dynamics are negligible. Elastic distortion energy terms [70,71] only appear above the threshold (see Table S4 [42]). However, a term of the form $K\nabla^2\mathbf{Q}$, which contains the elastic terms in the single constant approximation, is required for numerical stability. Moreover, the elastic terms play a key role in determining the structure of a nematic in the vicinity of defects. To understand this apparent contradiction, we compare the contributions of the distortion energy with flow-coupling terms as a function of space (Movie S2 [42]). This shows that the elastic terms are small everywhere except near defects. When combined with the fact that the majority of the experimental data is far from defects due to their small core size and finite density, this is the likely reason for the absence of elastic terms in the discovered model (Fig. S5).

The optimality curves for the flow equation are almost flat [Fig. 1(e)], showing that the active force $\alpha/\eta\nabla\cdot\mathbf{Q}$ alone balances the viscous force. Noting that this is a fit to the *weak* form of the equation, we test the strong form of the discovered equation by comparing the spatial dependence of $\nabla\times\nabla^2\mathbf{u}$ with $\alpha/\eta\nabla\times\nabla\cdot\mathbf{Q}$ and find good agreement (Movies S3 and S4 [42]). The inertial terms are absent (not appearing until $n \sim 5$), indicating that the Stokes flow approximation accurately describes the experimental active nematic. Finally, the absence of the substrate friction term $\Gamma\mathbf{u}$ indicates that the screening length $\sqrt{\eta/\Gamma}$ is larger than the typical vortex size of the flows. This result likely depends on the active nematic system and experimental conditions; for example, changing the substrate depth affects the effective friction coefficient [72]. However, the framework presented here can be applied directly to other conditions or materials.

The discovered flow equation provides a direct estimate of the scaled activity parameter α/η, an intrinsic "active timescale" [73], as a function of the ATP concentration [Fig. 1(g)] [74]. Determining the relationship between activity and experimental control parameters has been a significant challenge [21]. The molecular motors that generate activity also act as passive cross-linkers between steps [76], and in a dense active nematic, forces generated by different motors are largely noncooperative [77]. To test the estimate of α against an independent observable, we compare the active timescale to the velocity

autocorrelation time τ, defined as $\bar{C}_v(\tau) = 1/e$, with the autocorrelation function $\bar{C}_v(t) = \langle\langle \mathbf{u}(\mathbf{r}, t'+t) \cdot \mathbf{u}(\mathbf{r}, t')\rangle_{t'} / \langle \mathbf{u}(\mathbf{r}, t') \cdot \mathbf{u}(\mathbf{r}, t')\rangle_{t'}\rangle_{\mathbf{r}}$. These observables closely agree at all ATP concentrations [Fig. 1(g)].

Finally, we test the optimal model by performing simulations of Eq. (5). For numerical stability, we include the $K\nabla^2\mathbf{Q}$ term in the $\mathbf{Q}$ equation with $K = 1$ in dimensionless simulation units. We compare the mean defect lifetime and defect density [78,79] as a function of α/η (Fig. 2). Remarkably, the defect lifetimes for experiments and simulations align well without any fit parameters [Fig. 2(a)]. The defect densities from experiment and simulation also match, up to a constant scaling factor [80] [Fig. 2(b)]. The latter cannot be specified—because the terms in the discovered alignment equation all have dimensionless coefficients, we cannot directly estimate a length scale [81].

In summary, we have applied a data-driven method to identify equations governing both the orientational dynamics and the activity-driven flows of microtubule-based active liquid crystals. The optimal model is surprisingly minimal. It demonstrates that (1) flow coupling dominates the orientational dynamics, and (2) the lowest-order active stress, proportional to the local orientational order, together with the vanishing Reynolds number limit describe the flow. This model is not only consistent with previous theoretical arguments [54], but is also less complex than most models considered in the literature. Our results also show that statistical uncertainty arising from limited experimental data impedes unequivocal identification of near-threshold terms, but suggest strategies to mitigate these effects. For example, acquiring a combination of high-resolution small field-of-view and low-resolution large field-of-view datasets would enable more accurate discovery of the alignment and flow equations, respectively. Further, acquiring more data in the vicinity of defects and/or additional analysis that preferentially weights data in the vicinity of defects may identify elastic energy terms.

The identified equations enable mapping between key model parameters and experimental control variables, including the elusive relationship between the magnitude of activity and ATP concentration. Thus, our results are the first to assess the quantitative variation of phenomenological theory parameters as a function of experimental control knobs in active nematics, while *also* providing evidence for the validity of the underlying model. Through comparison of several noise reduction approaches and extensive benchmarking, we have identified an approach to model discovery which is highly robust against experimental noise. This approach can be extended to study recently developed 3D active nematic materials [49,82,83], complementing existing theoretical and numerical efforts [48,52,84–88]. It can be applied to a wide variety of active matter systems, or more broadly, to any system for which observations of dynamics can be projected onto continuous fields. This process can shed light on relationships between physical quantities or even identify new physical mechanisms.

Note added.—Recently, we learned of a complementary, concurrent work that uses symbolic regression, whose findings are generally consistent with those of our work [89].

---

[*]chaitanya@brandeis.edu
[†]zdogic@physics.ucsb.edu
[‡]aparna@brandeis.edu
[§]hagan@brandeis.edu

[1] M. C. Marchetti, J. F. Joanny, S. Ramaswamy, T. B. Liverpool, J. Prost, M. Rao, and R. A. Simha, Rev. Mod. Phys. **85**, 1143 (2013).

[2] S. Ramaswamy, Annu. Rev. Condens. Matter Phys. **1**, 323 (2010).

[3] J. Toner, Y. Tu, and S. Ramaswamy, Ann. Phys. (Amsterdam) **318**, 170 (2005).

[4] V. Narayan, S. Ramaswamy, and N. Menon, Science **317**, 105 (2007).

[5] H. H. Wensink, J. Dunkel, S. Heidenreich, K. Drescher, R. E. Goldstein, H. Lowen, and J. M. Yeomans, Proc. Natl. Acad. Sci. U.S.A. **109**, 14308 (2012).

[6] S. Zhou, A. Sokolov, O. D. Lavrentovich, and I. S. Aranson, Proc. Natl. Acad. Sci. U.S.A. **111**, 1265 (2014).

[7] G. Duclos, C. Erlenkämper, J. F. Joanny, and P. Silberzan, Nat. Phys. **13**, 58 (2017).

[8] K. Kawaguchi, R. Kageyama, and M. Sano, Nature (London) **545**, 327 (2017).

[9] N. Kumar, R. Zhang, J. J. de Pablo, and M. L. Gardel, Sci. Adv. **4**, eaat7779 (2018).

[10] T. Sanchez, D. T. N. Chen, S. J. DeCamp, M. Heymann, and Z. Dogic, Nature (London) **491**, 431 (2012).

[11] S. J. DeCamp, G. S. Redner, A. Baskaran, M. F. Hagan, and Z. Dogic, Nat. Mater. **14**, 1110 (2015).

[12] A. Doostmohammadi, J. Ignés-Mullol, J. M. Yeomans, and F. Sagués, Nat. Commun. **9**, 3246 (2018).

[13] L. Giomi, M. J. Bowick, X. Ma, and M. C. Marchetti, Phys. Rev. Lett. **110**, 228101 (2013).

[14] L. Giomi and A. DeSimone, Phys. Rev. Lett. **112**, 147802 (2014).

[15] A. Doostmohammadi, T. N. Shendruk, K. Thijssen, and J. M. Yeomans, Nat. Commun. **8**, 15326 (2017).

[16] A. U. Oza and J. Dunkel, New J. Phys. **18**, 093006 (2016).

[17] D. Cortese, J. Eggers, and T. B. Liverpool, Phys. Rev. E **97**, 022704 (2018).

[18] T. N. Shendruk, K. Thijssen, J. M. Yeomans, and A. Doostmohammadi, Phys. Rev. E **98**, 010601(R) (2018).

[19] S. P. Thampi, R. Golestanian, and J. M. Yeomans, Phys. Rev. E **90**, 062307 (2014).

[20] L. Giomi, Phys. Rev. X **5**, 031003 (2015).

[21] L. M. Lemma, S. J. DeCamp, Z. You, L. Giomi, and Z. Dogic, Soft Matter **15**, 3264 (2019).

[22] T. N. Shendruk, A. Doostmohammadi, K. Thijssen, and J. M. Yeomans, Soft Matter **13**, 3853 (2017).

[23] M. M. Norton, A. Baskaran, A. Opathalage, B. Langeslay, S. Fraden, A. Baskaran, and M. F. Hagan, Phys. Rev. E **97**, 012702 (2018).

[24] T. Gao, M. D. Betterton, A.-S. Jhang, and M. J. Shelley, Phys. Rev. Fluids **2**, 093302 (2017).

[25] R. Zhang, Y. Zhou, M. Rahimi, and J. J. de Pablo, Nat. Commun. **7**, 13483 (2016).

[26] P. W. Ellis, D. J. G. Pearce, Y.-W. Chang, G. Goldsztein, L. Giomi, and A. Fernandez-Nieves, Nat. Phys. **14**, 85 (2018).

[27] F. Alaimo, C. Köhler, and A. Voigt, Sci. Rep. **7**, 5211 (2017).

[28] F. Cichos, K. Gustavsson, B. Mehlig, and G. Volpe, Nat. Mach. Intell. **2**, 94 (2020).

[29] H. Li, X.-q. Shi, M. Huang, X. Chen, M. Xiao, C. Liu, H. Chaté, and H. P. Zhang, Proc. Natl. Acad. Sci. U.S.A. **116**, 777 (2019).

[30] Z. Zhou, C. Joshi, R. Liu, M. M. Norton, L. Lemma, Z. Dogic, M. F. Hagan, S. Fraden, and P. Hong, Soft Matter **17**, 738 (2021).

[31] J. Colen, M. Han, R. Zhang, S. A. Redford, L. M. Lemma, L. Morgan, P. V. Ruijgrok, R. Adkins, Z. Bryant, Z. Dogic, M. L. Gardel, J. J. De Pablo, and V. Vitelli, Proc. Natl. Acad. Sci. U.S.A. **118**, e2016708118 (2021).

[32] S. L. Brunton, J. L. Proctor, J. N. Kutz, and W. Bialek, Proc. Natl. Acad. Sci. U.S.A. **113**, 3932 (2016).

[33] S. H. Rudy, S. L. Brunton, J. L. Proctor, and J. N. Kutz, Sci. Adv. **3**, e1602614 (2017).

[34] S. Maddu, Q. Vagne, and I. F. Sbalzarini, arXiv:2201.08623.

[35] R. Supekar, B. Song, A. Hastewell, A. Mietke, and J. Dunkel, arXiv:2101.06568.

[36] E. P. Alves and F. Fiuza, Phys. Rev. Res. **4**, 033192 (2022).

[37] P. A. K. Reinbold and R. O. Grigoriev, Phys. Rev. E **100**, 022219 (2019).

[38] P. A. K. Reinbold, D. R. Gurevich, and R. O. Grigoriev, Phys. Rev. E **101**, 010203(R) (2020).

[39] P. A. K. Reinbold, L. M. Kageorge, M. F. Schatz, and R. O. Grigoriev, Nat. Commun. **12**, 3219 (2021).

[40] A. N. Beris and B. J. Edwards, *Thermodynamics of Flowing Systems with Internal Microstructure* (Oxford University Press, New York, 1994).

[41] L. M. Lemma, M. M. Norton, A. M. Tayar, S. J. DeCamp, S. A. Aghvami, S. Fraden, M. F. Hagan, and Z. Dogic, Phys. Rev. Lett. **127**, 148001 (2021).

[42] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.129.258001 for details of the theory, model discovery, analysis, and experimental methods.

[43] G. P. G. de and J. Prost, *The Physics of Liquid Crystals* (Clarendon Press, Oxford, 1993).

[44] L. Giomi, M. J. Bowick, P. Mishra, R. Sknepnek, and M. Cristina Marchetti, Phil. Trans. R. Soc. A **372**, 20130365 (2014).

[45] T. Gao, R. Blackwell, M. A. Glaser, M. D. Betterton, and M. J. Shelley, Phys. Rev. Lett. **114**, 048101 (2015).

[46] E. J. Hemingway, P. Mishra, M. C. Marchetti, and S. M. Fielding, Soft Matter **12**, 7943 (2016).

[47] S. Ngo, A. Peshkov, I. S. Aranson, E. Bertin, F. Ginelli, and H. Chaté, Phys. Rev. Lett. **113**, 038302 (2014).

[48] M. Varghese, A. Baskaran, M. F. Hagan, and A. Baskaran, Phys. Rev. Lett. **125**, 268003 (2020).

[49] G. Duclos, R. Adkins, D. Banerjee, M. S. E. Peterson, M. Varghese, I. Kolvin, A. Baskaran, R. A. Pelcovits, T. R. Powers, A. Baskaran, F. Toschi, M. F. Hagan, S. J. Streichan, V. Vitelli, D. A. Beller, and Z. Dogic, Science **367**, 1120 (2020).

[50] L. Giomi, L. Mahadevan, B. Chakraborty, and M. F. Hagan, Nonlinearity **25**, 2245 (2012).

[51] S. Chandragiri, A. Doostmohammadi, J. M. Yeomans, and S. P. Thampi, Soft Matter **15**, 1597 (2019).

[52] S. Chandragiri, A. Doostmohammadi, J. M. Yeomans, and S. P. Thampi, Phys. Rev. Lett. **125**, 148002 (2020).

[53] S. P. Thampi, R. Golestanian, and J. M. Yeomans, Phys. Rev. Lett. **111**, 118101 (2013).

[54] S. P. Thampi, A. Doostmohammadi, R. Golestanian, and J. M. Yeomans, Europhys. Lett. **112**, 28004 (2015).

[55] S. P. Thampi, A. Doostmohammadi, T. N. Shendruk, R. Golestanian, and J. M. Yeomans, Sci. Adv. **2**, e1501854 (2016).

[56] C.-M. Koch and M. Wilczek, Phys. Rev. Lett. **127**, 268005 (2021).

[57] R. A. Simha and S. Ramaswamy, Phys. Rev. Lett. **89**, 058101 (2002).

[58] The codes used for this Letter are available under, https://github.com/joshichaitanya3/actnempy.

[59] J. Zhao and Q. Wang, J. Sci. Comput. **68**, 1241 (2016).

[60] S. P. Vanka, J. Comput. Phys. **65**, 138 (1986).

[61] M. Castoldi and A. V. Popov, Protein Expression Purif. **32**, 83 (2003).

[62] A. Hyman, D. Drechsel, D. Kellogg, S. Salser, K. Sawin, P. Steffen, L. Wordeman, and T. Mitchison, in *Methods in Enzymology*, Molecular Motors and the Cytoskeleton Vol. 196 (Academic Press, New York, 1991), pp. 478–485, 10.1016/0076-6879(91)96041-O.

[63] R. Subramanian and J. Gelles, J. Gen. Physiol. **130**, 445 (2007).

[64] A. W. C. Lau, A. Prasad, and Z. Dogic, Europhys. Lett. **87**, 48006 (2009).

[65] R. Oldenbourg, *Live cell imaging: A laboratory manual*, edited by R. D. Goldman and D. L. Spector (Cold Spring Harbor Laboratory Press Cold Spring Harbor, NY, 2005), pp. 205–237.

[66] D. Garcia, Exp. Fluids **50**, 1247 (2011).

[67] The higher resolution data has the PIV velocity measured on a 79×66 pixel grid, which is comparable to the rest of the data sets (71×71), but its field of view contains only one or two vortices (Fig. S4 [42]), thus resolving the flow field much more accurately.

[68] A. Maitra, P. Srivastava, M. Cristina Marchetti, J. S. Lintuvuori, S. Ramaswamy, and M. Lenz, Proc. Natl. Acad. Sci. U.S.A. **115**, 6934 (2018).

[69] Three data sets give (near threshold) $a_2 < 0$ (isotropic regime), with one of them containing an unstable component with $a_4 > 0$.

[70] E. Putzig, G. S. Redner, A. Baskaran, and A. Baskaran, Soft Matter **12**, 3854 (2016).

[71] N. J. Mottram and C. J. P. Newton, arXiv:1409.3542.

[72] K. Thijssen, D. A. Khaladj, S. A. Aghvami, M. A. Gharbi, S. Fraden, J. M. Yeomans, L. S. Hirst, and T. N. Shendruk, Proc. Natl. Acad. Sci. U.S.A. **118**, e2106038118 (2021).

[73] L. Giomi, L. Mahadevan, B. Chakraborty, and M. F. Hagan, Phys. Rev. Lett. **106**, 218101 (2011).

[74] Independent measurements of the viscosity such as in [75] can be then used to estimate α, the strength of the active force.

[75] P. Guillamat, J. Ignés-Mullol, S. Shankar, M. C. Marchetti, and F. Sagués, Phys. Rev. E **94**, 060602(R) (2016).

[76] D. A. Gagnon, C. Dessi, J. P. Berezney, R. Boros, D. T.-N. Chen, Z. Dogic, and D. L. Blair, Phys. Rev. Lett. **125**, 178003 (2020).

[77] A. M. Tayar, M. F. Hagan, and Z. Dogic, Proc. Natl. Acad. Sci. U.S.A. **118**, e2102873118 (2021).

[78] R. D. Kamien, Rev. Mod. Phys. **74**, 953 (2002).

[79] D. B. Allan, T. Caswell, N. C. Keim, C. M. van der Wel, and R. W. Verweij, Soft-matter/trackpy: Trackpy v0.5.0, Zenodo (2021), 10.5281/zenodo.4682814.

[80] We re-scale the defect density so that the simulation value for $\alpha/\eta = 0.3$ falls on the linear fit of the experimental values.

[81] We can force the framework to estimate dimensional parameters by constraining the regression procedure to include specific terms, while performing sparse regression on the remaining terms. For example, by forcing a term $K\nabla^2 Q_{xx}$ [see Eq. (2)], we obtain a value for the elastic modulus of $K \sim 1 \ \mu m^2 \ s^{-1}$. However, because this term has a negligible contribution to the dynamics of $\mathbf{Q}$, the quantitative accuracy of this estimate may be limited.

[82] P. Chandrakar, M. Varghese, S. A. Aghvami, A. Baskaran, Z. Dogic, and G. Duclos, Phys. Rev. Lett. **125**, 257801 (2020).

[83] B. Najma, M. Varghese, L. Tsidilkovski, L. Lemma, A. Baskaran, and G. Duclos, Nat. Comm. **13**, 6465 (2022).

[84] D. M. Sussman and D. A. Beller, Front. Phys. **7**, 204 (2019).

[85] S. Mandal and M. G. Mazza, Phys. Rev. E **99**, 063319 (2019).

[86] T. Kozhukhov and T. N. Shendruk, Sci. Adv. **8**, eabo5788 (2022).

[87] L. J. Ruske and J. M. Yeomans, Phys. Rev. X **11**, 021001 (2021).

[88] S. Čopar, J. Aplinc, Ž. Kos, S. Žumer, and M. Ravnik, Phys. Rev. X **9**, 031051 (2019).

[89] M. Golden, R. Grigoriev, J. Nambisan, and A. Fernandez-Nieves, arXiv:2202.12853.