Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Exact ensemble bifurcation analysis reveals hidden bistability and makes signalling dynamics predictable from kinetic parameters

Guturu L Harika
IIIT-Delhi, Okhla Phase III
K. Sriram
sriramk@iiitd.ac.in

IIIT-Delhi, Okhla Phase III

Article

Keywords:

Posted Date: November 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11522901/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Exact ensemble bifurcation analysis reveals hidden bistability and makes signalling dynamics predictable from kinetic parameters

Guturu L Harika[1] and K. Sriram[2,*]

[1]Department of Computational Biology, IIIT-Delhi, Okhla Phase III, New Delhi 110020, India
[2]Department of Computational Biology, Center for Computational Biology, and Centre for Artificial Intelligence, IIIT-Delhi, Okhla Phase III, New Delhi 110020, India
[*]Corresponding author: sriramk@iiitd.ac.in

Abstract

Whether a signalling motif acts as a graded sensor, a switch or a memory is set by kinetic parameters that are rarely known, yet bifurcation analysis examines one parameter set at a time. We introduce an inverse method that parametrises every steady state of the dual phosphorylation–dephosphorylation cycle by the free kinase-to-phosphatase ratio, giving the complete bifurcation diagram of a parameter set exactly, without initial conditions, in ms. Applied to 40,000 Latin hypercube parameter sets, it yields an ensemble of exact bifurcation diagrams labelled into ten bifurcation classes. Bistability is common (65% of sets) but lies mostly on detached branches that single-start continuation misses (19–27% against 72–88% of asymmetric sets); a closed-form criterion predicts every detached branch. Gradient-boosted classifiers trained on these labels predict the class from kinetic parameters alone (93% accuracy) and rediscover the analytical conditions, identifying measurable rate ratios for designing synthetic switches.

Introduction

Cells make decisions with molecular circuits. Mathematical models based on ordinary differential equations (ODEs) are a foundational way to understand how such circuits turn molecular interactions into cellular behaviour [1, 2, 3]. They reveal the nonlinear phenomena on which decisions rest, such as multistability, oscillations and ultrasensitive responses [4, 5, 6], and they are increasingly used not only to explain natural circuits but to design synthetic ones, from transcriptional toggle switches [7] to engineered protein-phosphorylation networks [8].

Whether a circuit acts as a graded sensor, a reversible switch or an irreversible memory is decided not by its wiring alone but by its kinetic parameters. Bifurcation analysis identifies the parameter values at which the number or stability of steady states changes, and therefore where the behaviour changes qualitatively [6, 9]. In practice, however, it varies one parameter while all others are held fixed, and so it probes a single slice of a high-dimensional parameter space. Kinetic parameters are rarely known: they are uncertain and span several orders of magnitude across biochemical contexts [10, 11]. This has motivated approaches that treat parameters as ensembles rather than fixed values [12], and the observation that distinct parameter combinations often produce nearly indistinguishable behaviour suggests that qualitative dynamics are governed by a few combinations of parameters rather than by individual values [13, 14]. Yet how the bifurcation structure of a network is organised across its whole kinetic parameter space, which combinations of parameters control the transitions, and how often each behaviour occurs, remain largely unknown.

Multisite phosphorylation is an ideal system in which to ask these questions. Phosphorylation–dephosphorylation (PdP) cycles are canonical models of nonlinear signalling, exhibiting ultrasensitivity, bistability, symmetry breaking and biphasic responses [15, 16, 17, 18]. The dual PdP cycle lies at the core of the MAPK cascade [19], and, depending on parameters and mechanism, it can show saddle-node, transcritical and pitchfork bifurcations and their combinations [20, 21]. Distributive mechanisms in particular generate bistability and switch-like behaviour, whereas processive mechanisms tend to suppress it [22]. For the ordered distributive mechanism with a common kinase and phosphatase, the number of positive steady states is at most $2n - 1$ for n sites [23], and this bound is attained [24]. The motif is thus small enough to be analysed exactly, rich enough to display every codimension-1 bifurcation of equilibria, and relevant both to natural signalling and to the design of synthetic switches [8].

Existing approaches answer parts of the question. Chemical Reaction Network Theory determines from the reaction topology alone whether a network has the capacity for multiple steady states [25, 26], and is implemented in software such as the Chemical Reaction Network Toolbox [27]. Algebraic methods partition parameter space into regions with one or several steady states, and have been applied to two-site phosphorylation [28, 29]; automated structural screening has been applied to tens of thousands of small networks [30]. Structural bifurcation analysis predicts from the topology alone which species bifurcate and, for power-law kinetics, whether a bifurcation is a saddle-node, transcritical or pitchfork [31, 32]; being independent of the kinetics, it cannot say for which parameter values these bifurcations occur. Sampling-based methods address the parameter dependence directly: Latin hypercube sampling with numerical equilibrium search, guided by logistic-regression or neural-network surrogates and summarised by rule induction, approximates the regions of multistationarity and multistability of several networks, including the sequential dual phosphorylation cycle [33], and equilibrium-informed neural networks learn the inverse map from equilibrium states to parameters to locate saddle-node thresholds [34]. These methods establish *whether* and *where* multistationarity is possible, and which bifurcations a topology admits, but not the complete type and arrangement of the bifurcations that a given parameter set exhibits along a control parameter. Numerical continuation provides exactly this information, and it can be scripted: MATCONT has a command-line version for batch use [35, 36, 37]. In EMBED [38] we combined MATCONT with Latin hypercube sampling to perform ensemble bifurcation analysis of the dual PdP cycle, and concluded that bifurcation-driven behaviour occupies narrow regions of parameter space whereas graded, ultrasensitive and biphasic responses dominate.

Two problems limit what any such ensemble can reveal. The first is completeness. Continuation follows the branch of steady states on which it is started; a branch that is disconnected from it is never found, whatever the continuation settings. An ensemble labelled by continuation from a single steady state, as in EMBED and most ensemble studies, may therefore under-count multistability systematically, and its labels need a measured accuracy before they can support conclusions about prevalence or be used to train predictive models. The second problem concerns design. Synthetic biology engineers circuits through Design–Build–Test–Learn (DBTL) cycles [39, 40], but designs that work *in silico* often fail when built and tested. EMBED attributed part of this gap to the narrowness of desired regimes in parameter space [38]. A further source is suggested by the analogy between computation and experiment: the standard test of a switch, a dose–response measured with increasing input from an unactivated state, follows a single branch of steady states exactly as single-start continuation does, and would miss the same switches. If so, the place of simulation in the design cycle needs to be reconsidered.

Here we address six problems. The first concerns global organisation (**P1**): how is the bifurcation structure of the dual PdP cycle organised across its full kinetic parameter space, rather than along single parameter slices? The second concerns completeness and prevalence (**P2**): does continuation

from a single steady state miss multistability, and if so, is bistability in this motif actually rare, as single-start ensembles suggest, or typical? The third concerns verifiable labels (**P3**): can thousands of bifurcation diagrams be labelled completely, in a fixed window, with an independently measured accuracy? The fourth concerns symmetry and mechanism (**P4**): how does kinase–phosphatase symmetry, and its controlled breaking, determine the type of bifurcation (pitchfork, transcritical, fold on the main or on a detached branch), and can the boundaries between classes be written in closed form? The fifth concerns prediction and interpretation (**P5**): can the bifurcation class be predicted from the kinetic parameters by machine learning, and does the classifier learn the mechanism or merely the sampling design? The sixth concerns transitions (**P6**): which changes of the kinetic parameters move a parameter set from one class to a neighbouring one, and which of these changes are controlled by a closed-form condition?

To answer them, we build an ensemble of 40,000 parameter sets of the ordered, distributive dual PdP cycle in its full mass-action form, generated under a controlled symmetry-breaking design: full kinase–phosphatase symmetry, symmetry broken only in the first binding step, and kinase- or phosphatase-dominant asymmetry (P1). We label every set with one of ten bifurcation classes in the total substrate concentration A_{TOT}, using an explicit parametrisation of all steady states by the ratio of free kinase to free phosphatase, which finds every branch, including detached ones, without initial conditions or root finding, and we validate the labels against MATCONT seeded on every branch and by normal-form coefficients at all singularities (P3). Within the sampled ranges the map shows that bistability is common, but mostly hidden on detached branches that single-start continuation cannot reach (P2), whereas bifurcations at branch points require kinetic symmetry; all class boundaries but one are closed-form conditions on a few ratios of enzyme capacities and Michaelis constants (P4). Gradient-boosted classifiers predict the class with 93% accuracy and, within each sampling stratum, rediscover these conditions without being given them (P5); per-class contributions show which conditions each class is learnt from. Finally, we map the transitions between neighbouring classes: single changes of a catalytic ratio move a set across the exact boundaries, whereas the fold on the main branch, the one boundary without a closed form, responds to several parameters together (P6). What this implies for the design cycle of synthetic biology is discussed as future work. To our knowledge, this is the first ensemble of exact and complete bifurcation diagrams for a signalling motif, and the first classifier trained on such labels to predict the bifurcation class of a circuit from its kinetic parameters.

Results

A symmetry-designed ensemble of the dual phosphorylation cycle

We consider a substrate A that is phosphorylated twice by a kinase K ($A \to AP \to APP$) and dephosphorylated twice by a phosphatase P ($APP \to AP \to A$), each step proceeding through its own enzyme–substrate complex (reactions R1–R8; Fig 1 and Methods). This ordered, distributive dual phosphorylation–dephosphorylation (PdP) cycle is described by six ODEs and three conservation laws (Eqs (1)–(2)), with 14 fixed quantities (K_{TOT}, P_{TOT} and twelve rate constants). The total substrate concentration A_{TOT} is the bifurcation parameter, and the output of the switch is the doubly phosphorylated substrate $x_2 = [APP]$.

We built an ensemble of 40,000 parameter sets in four sampling iterations of 10,000 sets each, drawn by Latin hypercube sampling uniformly in log space (Table 1). The iterations are a designed experiment on the kinase–phosphatase symmetry of the network. Iteration 3 imposes full swap symmetry ($K_{\mathrm{TOT}} = P_{\mathrm{TOT}}$, $k_1 = k_5$, $k_{11} = k_{55}$, $k_2 = k_6$, $k_3 = k_7$, $k_{33} = k_{77}$, $k_4 = k_8$); Iteration 4 breaks it only in the first binding step (k_1, k_{11}, k_5, k_{55} free); Iterations 1 and 2 break it completely, in

Table 1. Parameter bounds, symmetry constraints and bifurcation classes for the four sampling iterations ($N = 10{,}000$ per iteration; sampling uniform in log space). Baseline bounds: 10–500 nM for K_{TOT} and P_{TOT}, 10^{-3}–10^{1} for all rate constants ($\mathrm{nM}^{-1}\,\mathrm{s}^{-1}$ for the binding constants k_1, k_3, k_5, k_7; s^{-1} for the others); boldface bounds deviate from the baseline. Classes are the exact labels over $A_{\mathrm{TOT}} \in [0, 10^6]$. Code: `Ensemble_LHS_v3.m`, `Manuscript_numbers_v3.m` (folder A).

		Iter. 1: kinase-dominant			Iter. 2: phosphatase-dominant			Iter. 3: full symmetry			Iter. 4: first-binding asym.		
Par.	Description	Lower	Upper	Constr.	Lower	Upper	Constr.	Lower	Upper	Constr.	Lower	Upper	Constr.
K_{TOT}	Total kinase	**100**	500	free	10	**100**	free	10	500	$= P_{\mathrm{TOT}}$	10	500	$= P_{\mathrm{TOT}}$
P_{TOT}	Total phosphatase	10	**100**	free	**100**	500	free	10	500	$= K_{\mathrm{TOT}}$	10	500	$= K_{\mathrm{TOT}}$
k_1	A + K binding	10^{-3}	$\mathbf{10^{-2}}$	free	$\mathbf{10^{-1}}$	10^{1}	free	10^{-3}	10^{1}	$= k_5$	10^{-3}	10^{1}	free
k_{11}	AK unbinding	10^{-3}	10^{1}	free	10^{-3}	$\mathbf{10^{-1}}$	free	10^{-3}	10^{1}	$= k_{55}$	10^{-3}	10^{1}	free
k_2	AK → AP + K	10^{-3}	$\mathbf{10^{-1}}$	free	10^{-3}	$\mathbf{10^{-1}}$	free	10^{-3}	10^{1}	$= k_6$	10^{-3}	10^{1}	$= k_6$
k_3	AP + K binding	10^{-3}	$\mathbf{10^{0}}$	free	$\mathbf{10^{-1}}$	10^{1}	free	10^{-3}	10^{1}	$= k_7$	10^{-3}	10^{1}	$= k_7$
k_{33}	APK unbinding	10^{-3}	10^{1}	free	10^{-3}	$\mathbf{10^{-1}}$	free	10^{-3}	10^{1}	$= k_{77}$	10^{-3}	10^{1}	$= k_{77}$
k_4	APK → APP + K	$\mathbf{10^{-1}}$	10^{1}	free	$\mathbf{10^{-1}}$	10^{1}	free	10^{-3}	10^{1}	$= k_8$	10^{-3}	10^{1}	$= k_8$
k_5	APP + P binding	$\mathbf{10^{-1}}$	10^{1}	free	10^{-3}	$\mathbf{10^{-2}}$	free	10^{-3}	10^{1}	$= k_1$	10^{-3}	10^{1}	free
k_{55}	APPC unbinding	10^{-3}	$\mathbf{10^{-1}}$	free	10^{-3}	10^{1}	free	10^{-3}	10^{1}	$= k_{11}$	10^{-3}	10^{1}	free
k_6	APPC → AP + P	10^{-3}	$\mathbf{10^{-1}}$	free	10^{-3}	$\mathbf{10^{-1}}$	free	10^{-3}	10^{1}	$= k_2$	10^{-3}	10^{1}	$= k_2$
k_7	AP + P binding	$\mathbf{10^{-1}}$	10^{1}	free	10^{-3}	10^{1}	free	10^{-3}	10^{1}	$= k_3$	10^{-3}	10^{1}	$= k_3$
k_{77}	APC unbinding	10^{-3}	$\mathbf{10^{-1}}$	free	10^{-3}	10^{1}	free	10^{-3}	10^{1}	$= k_{33}$	10^{-3}	10^{1}	$= k_{33}$
k_8	APC → A + P	10^{-3}	10^{1}	free	$\mathbf{10^{-1}}$	10^{1}	free	10^{-3}	10^{1}	$= k_4$	10^{-3}	10^{1}	$= k_4$
Bifurcation class (exact labels)													
1	LP only, S-shaped	2,032 (20.32%)			0			0			0		
2	S-shaped LP + detached LP	1,333 (13.33%)			0			0			0		
3	LP only, Z-shaped	0			478 (4.78%)			0			0		
4	Z-shaped LP + detached LP	0			1,906 (19.06%)			0			0		
5	LP on a detached branch only	3,845 (38.45%)			6,399 (63.99%)			0			0		
6	BP only	0			0			4,971 (49.71%)			0		
7	LP and BP, $x_2^{LP} > x_2^{BP}$	0			0			15* (0.15%)			2,563* (25.63%)		
8	LP and BP, $x_2^{LP} < x_2^{BP}$	0			0			16* (0.16%)			2,439 (24.39%)		
9	No singularity, biphasic	0			1,217 (12.17%)			0			0		
10	No singularity, monotone	2,790 (27.90%)			0			4,998 (49.98%)			4,998 (49.98%)		
Multistationary (Classes 1–8)		72.1%			87.8%			50.0%			50.0%		

*Includes flagged sets whose fold coincides with the branch point (relative difference in $A_{\mathrm{TOT}} < 10^{-9}$; the turning point of the crossing branch at the pitchfork, with vanishing fold coefficient): all 31 sets of Classes 7 and 8 in Iteration 3 and one Class-7 set in Iteration 4. The 32 flagged sets are reported here and in all figures, and excluded from machine learning. Total: 40,000 parameter sets (39,968 used for machine learning).

opposite directions of enzyme dominance (kinase-dominant, high K_{TOT}; phosphatase-dominant, high P_{TOT}). Every parameter set is labelled with one of ten bifurcation classes in $A_{\mathrm{TOT}} \in [0, 10^6]$ (Fig 1; Table 1): (1) LP only on the main branch, S-shaped hysteresis; (2) S-shaped hysteresis together with a fold on a detached branch; (3) LP only, Z-shaped; (4) Z-shaped hysteresis together with a detached fold; (5) a fold on a detached branch only; (6) BP only; (7) LP and BP with x_2 at the LP above its value at the BP; (8) LP and BP with x_2 at the LP below it; (9) no singularity, biphasic; (10) no singularity, monotone. The labels are exact, obtained from an explicit parametrisation of all steady states (the inverse method, Fig 3; Methods), and validated against MATCONT seeded on every branch.

Iteration 1: kinase-dominant. With high total kinase and low total phosphatase ($K_{\mathrm{TOT}} \in [100, 500]$, $P_{\mathrm{TOT}} \in [10, 100]$) and phosphorylation-favouring rate constants, every multistationary parameter set has at least one fold, and 72.1% of the sets are multistationary (Fig 2A). Main-branch hysteresis has the S orientation: LP only, S-shaped, 20.3% ($n = 2{,}032$), and S-shaped hysteresis together with an additional fold on a detached branch, 13.3% ($n = 1{,}333$). The largest class is a fold on a detached branch only, 38.5% ($n = 3{,}845$): the main branch is monotone, but a second, disconnected branch of steady states appears at a fold and makes the system bistable for all larger A_{TOT}. The remaining 27.9% ($n = 2{,}790$) have no singularity and a monotone response. No branch point or biphasic response occurs.

Iteration 2: phosphatase-dominant. Reversing the dominance (low $K_{\mathrm{TOT}} \in [10, 100]$, high $P_{\mathrm{TOT}} \in$

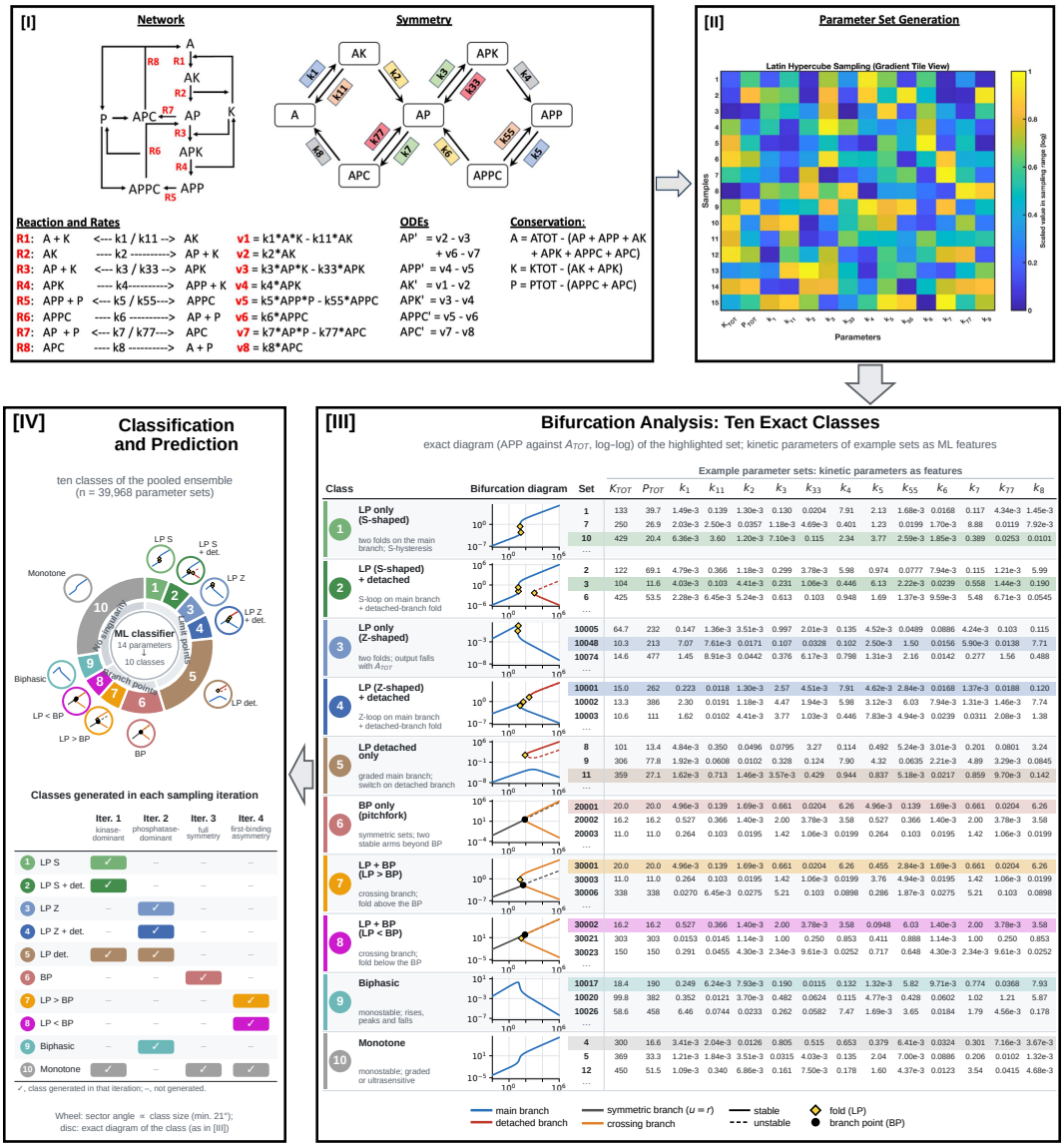


Fig 1. Workflow for high-throughput bifurcation analysis and classification of the dual phosphorylation–dephosphorylation (PdP) network. K, kinase; P, phosphatase; A, AP, APP, un-, singly and doubly phosphorylated substrate; AK, APK, APC, $APPC$, enzyme–substrate complexes. **[I] Network.** Reactions R1–R8, rate laws v_1–v_8, the six ODEs and the conservation laws for A_{TOT}, K_{TOT} and P_{TOT}. Mirrored kinase and phosphatase steps share a colour (e.g. k_1 and k_5); these pairs define the symmetry imposed or broken across iterations. **[II] Parameter set generation.** Latin hypercube sampling of the 14 parameters, uniform in log space; tiles show the scaled position of each value in its range for 15 sets. An iteration is one sampling run over a defined set of parameter ranges (Table 1); each iteration generates a different subset of classes. **[III] Bifurcation analysis.** Each set is labelled from its exact diagram of APP against A_{TOT} (inverse method, Fig 3). For each of the ten classes: example parameter sets with their 14 parameters (set n is row $n - 10000(i-1)$ of Iteration i) and the diagram of the highlighted set. **[IV] Classification and prediction.** Wheel: the ten classes of the pooled ensemble (39,968 sets), grouped into limit points, branch points and no singularity; sector angle proportional to class size (minimum 21°); discs as in [III]. Grid: the classes generated in Iterations 1–4 (kinase-dominant, phosphatase-dominant, full symmetry, first-binding asymmetry); ✓, generated. These labels are the targets of the machine-learning models. Code: `Ensemble_LHS_tiles_v3.m` ([II]), `Ensemble_class_table_v3.m` ([III]), `Ensemble_class_wheel_v3.m` ([IV]); labels from `Ensemble_LHS_v3.m` (folder A).

$[100, 500]$, dephosphorylation-favouring rate constants) reverses the orientation of main-branch hysteresis and makes multistationarity even more frequent: 87.8% of the sets are multistationary (Fig 2B). LP only, Z-shaped, accounts for 4.8% ($n = 478$) and Z-shaped hysteresis together with a fold on a detached branch for 19.1% ($n = 1{,}906$); a fold on a detached branch only is the dominant class, 64.0% ($n = 6{,}399$). The remaining 12.2% ($n = 1{,}217$) have no singularity but a biphasic response, in which APP first rises and then falls with A_{TOT}; a monotone response without singularity does not occur in this iteration.

Iteration 3: full kinase–phosphatase symmetry. Imposing the full swap symmetry removes folds and detached branches altogether and replaces them by branch points (Fig 2C). Half of the sets have a branch point and no fold, 49.7% ($n = 4{,}971$), and the other half no singularity with a monotone response, 50.0% ($n = 4{,}998$); the split is exact up to sampling noise because a branch point exists if and only if $k_4 > k_2$ (below). The remaining 31 sets (0.3%) are labelled as coexisting LP and BP (15 with $x_2^{LP} > x_2^{BP}$, 16 with $x_2^{LP} < x_2^{BP}$), but in all of them the reported fold coincides with the branch point to within 10^{-9} in relative A_{TOT}: it is the turning point of the crossing branch at the pitchfork, not a separate fold. These 31 sets are reported in all tables and figures, flagged, and excluded from machine learning.

Iteration 4: first-binding asymmetry. Freeing only the first binding step unfolds every pitchfork into a transcritical crossing accompanied by a fold on one arm of the crossing branch (Fig 2D). Coexisting LP and BP therefore become the generic outcome whenever a branch point exists: with the fold above the branch point in x_2, 25.6% ($n = 2{,}563$), and below it, 24.4% ($n = 2{,}439$); which of the two occurs is decided by the sign of $K_{M5} - K_{M1}$ (below). The other 50.0% ($n = 4{,}998$) have no singularity and a monotone response. One set of the class with $x_2^{LP} > x_2^{BP}$ is flagged as in Iteration 3.

S- and Z-shaped hysteresis are kept as separate classes throughout. Although both arise from folds on the main branch of the same output, they have opposite orientations, arise under kinase- and phosphatase-favoured conditions respectively, and occupy different regions of the plane of catalytic dominance (Fig 7); kinase–phosphatase imbalances are frequently associated with disease, and both enzyme classes are therapeutic targets [41, 42, 43, 44]. Within each orientation, sets in which main-branch hysteresis coexists with a fold on a detached branch (Classes 2 and 4) are separated from those with main-branch hysteresis only (Classes 1 and 3), because the two routes to bistability have different mechanisms and different parameter conditions (criterion (10); Fig 4C).

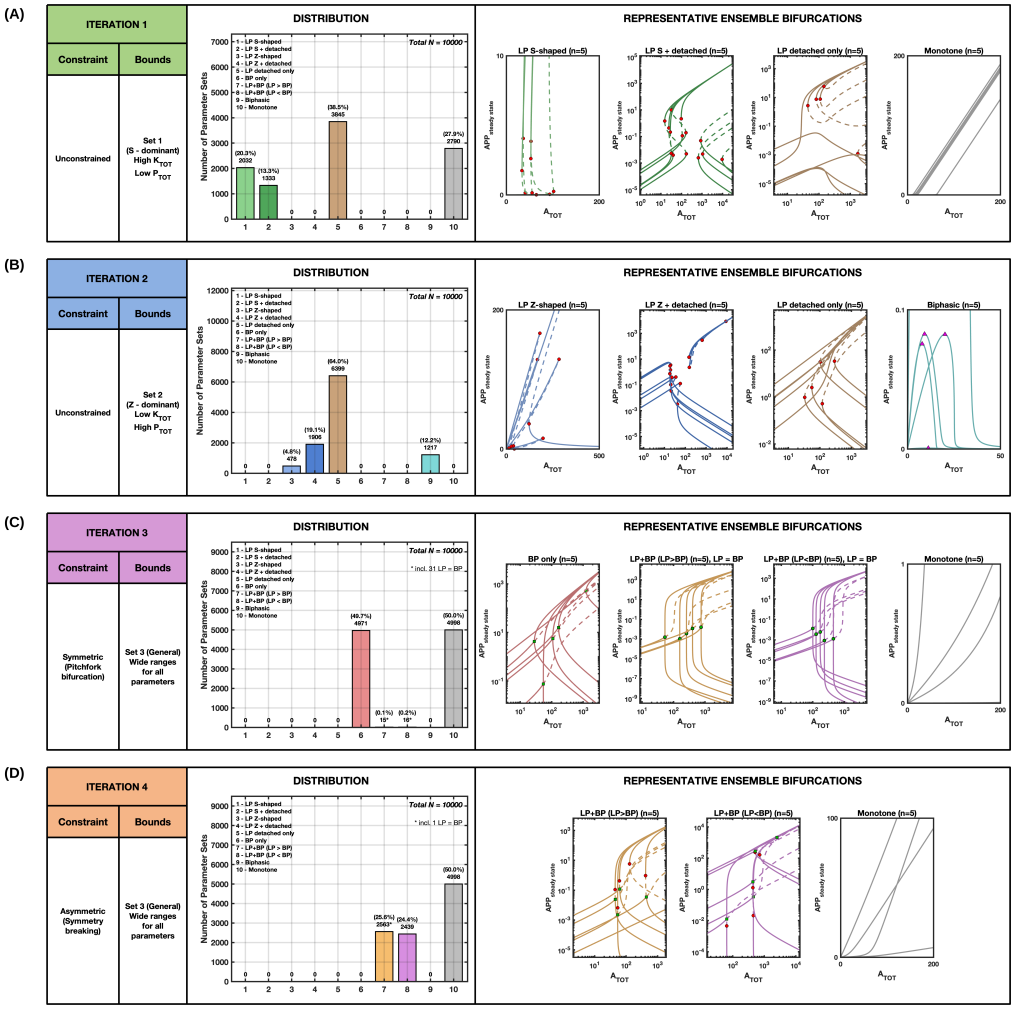


Fig 2. The four sampling iterations: class distribution and representative bifurcation diagrams. Each row: left, sampling constraint and parameter bounds (Table 1); middle, number of parameter sets per class ($N = 10{,}000$ per iteration); right, exact bifurcation diagrams (APP against A_{TOT}) of five randomly chosen parameter sets of each class, overlaid (solid, stable; dashed, unstable, from the eigenvalues of the analytic Jacobian; red circles, limit points; green squares, branch points; magenta triangles, maximum of a biphasic response). Classes with a detached branch are shown on log–log axes zoomed around the folds. (A) Iteration 1 (kinase-dominant, unconstrained): Classes 1 (2,032), 2 (1,333), 5 (3,845) and 10 (2,790); main-branch hysteresis is S-shaped. (B) Iteration 2 (phosphatase-dominant, unconstrained): Classes 3 (478), 4 (1,906), 5 (6,399) and 9 (1,217); hysteresis is Z-shaped (the fold at smaller A_{TOT} is the lower one), and biphasic responses rise to a single maximum and decay within the window. (C) Iteration 3 (full kinase–phosphatase symmetry: $K_{\mathrm{TOT}} = P_{\mathrm{TOT}}$ and every phosphorylation rate constant equal to its dephosphorylation counterpart): Classes 6 (4,971) and 10 (4,998), and 31 flagged sets of Classes 7 and 8 (asterisk) whose fold coincides with the branch point; the symmetric branch $u = K/P = 1$ loses stability at a pitchfork. (D) Iteration 4 (first-binding asymmetry, $k_1 \neq k_5$, $k_{11} \neq k_{55}$): Classes 7 (2,563; one flagged set), 8 (2,439) and 10 (4,998); the crossing is transcritical, and a fold on the kinase-rich arm above the branch point (Class 7) or on the phosphatase-rich arm below it (Class 8) starts the bistable range. Code: `Ensemble_iteration_panels_v3.m` on the output of `Ensemble_LHS_v3.m` (folder A).

Single-start continuation misses detached branches

The first labels of this ensemble were obtained, as in our earlier EMBED framework [38], by integrating the ODEs to a steady state at $A_{\mathrm{TOT}} = 0.1$ and continuing that state in A_{TOT} with MATCONT [35, 36], with branch switching at every branch point ("MATCONT without the inverse method"; Methods). EMBED applied this workflow to the reduced Michaelis–Menten model; here it is applied to the full mass-action model, and the two are compared in Supplementary Table 23. Such a continuation can only report singularities on the branch on which it starts and on branches that cross it (Fig 3A). To label the ensemble completely we parametrised all steady states by the ratio $u = K/P$ of free kinase to free phosphatase (Fig 3B; Eqs (7)–(9), Methods). For each u the free substrate follows in closed form, all concentrations follow from u and A, and the substrate conservation law, used in reverse, gives the A_{TOT} to which the steady state belongs; this map is a one-to-one correspondence with the set of all positive steady states over all A_{TOT}.

A closed-form criterion for detached branches. $A(u)$ can change sign only at $u = r$ and at $u^* = -\beta/\alpha$, so the set of admissible u consists of at most two intervals, each a branch of steady states. A second, detached branch exists if and only if criterion (10) holds, $k_8 P_{\mathrm{TOT}} > k_2 K_{\mathrm{TOT}}$ and $k_4 K_{\mathrm{TOT}} > k_6 P_{\mathrm{TOT}}$: the phosphatase has the larger maximal rate at the first site and the kinase the larger maximal rate at the second. Fig 4A places every parameter set of Iterations 1 and 2 on the two corresponding log-ratios, coloured by the outcome of the full steady-state computation. All 14,254 parameter sets of the ensemble with a detached branch (at any A_{TOT}) lie in the quadrant where the phosphatase dominates the first site and the kinase the second, and every set in that quadrant has one. With this crossed dominance the substrate can rest either mostly unphosphorylated, because the first phosphate is removed faster than it is added, or mostly doubly phosphorylated, because once the first phosphate is on, the second is added faster than it is removed. In 13,483 of these sets the fold of the detached branch lies inside the window $A_{\mathrm{TOT}} \leq 10^6$: 10,244 are bistable only through the detached branch (Class 5), and 3,239 also have S- or Z-shaped hysteresis on the main branch (Classes 2 and 4; 1,333 and 1,906 sets; Fig 4B,C). In the remaining 771 the detached fold lies above the window, and they are labelled by their main branch (open circles in Fig 4B). When the kinase dominates both sites (4,657 sets, all from Iteration 1) the main branch is S-shaped (Class 1; 1,985) or monotone (Class 10; 2,672); when the phosphatase dominates both (1,089 sets, Iteration 2) it is Z-shaped (Class 3; 407) or biphasic (Class 9; 682) (Fig 4C). The boundary between "detached branch" and "no detached branch" is therefore not a diffuse region but a pair of planes in log-parameter space, $k_8 P_{\mathrm{TOT}} = k_2 K_{\mathrm{TOT}}$ and $k_4 K_{\mathrm{TOT}} = k_6 P_{\mathrm{TOT}}$. Inequalities on the catalytic constants have previously been shown to guarantee multistationarity in dual phosphorylation [28, 29], and the kinetic parameters that permit multistationarity have been characterised completely [45]. Criterion (10) is complementary: it is the exact condition, including the enzyme totals, for a detached branch, and it states on which branch the second stable state lies. With the totals left free it can be met if and only if $k_4 k_8 > k_2 k_6$ (some $K_{\mathrm{TOT}}/P_{\mathrm{TOT}}$ then lies between k_6/k_4 and k_8/k_2), a condition on the catalytic constants alone, in line with their role identified in these studies.

Why single-start continuation under-counts multistationarity. Single-start continuation was run on the same 40,000 parameter sets; for 362 of them (119, 241, 2 and 0 in Iterations 1–4) the run failed or ended on negative concentrations (Supplementary Note 5). Comparing its labels with the exact structure of the same sets identifies two mechanisms by which coexisting states were missed; single-start continuation never reports a singularity where the exact labels have none. First, at $A_{\mathrm{TOT}} = 0.1$ only the main branch exists, so every run started on it; by construction a detached branch cannot be reached from there (Fig 3A). Among the single-start "biphasic" and "no bifurcation" sets of Iterations 1 and 2, 10,665 have a detached branch with its fold in the window, and for 6,979 of them the fold lies inside the A_{TOT} range that the continuation covered: the continuation

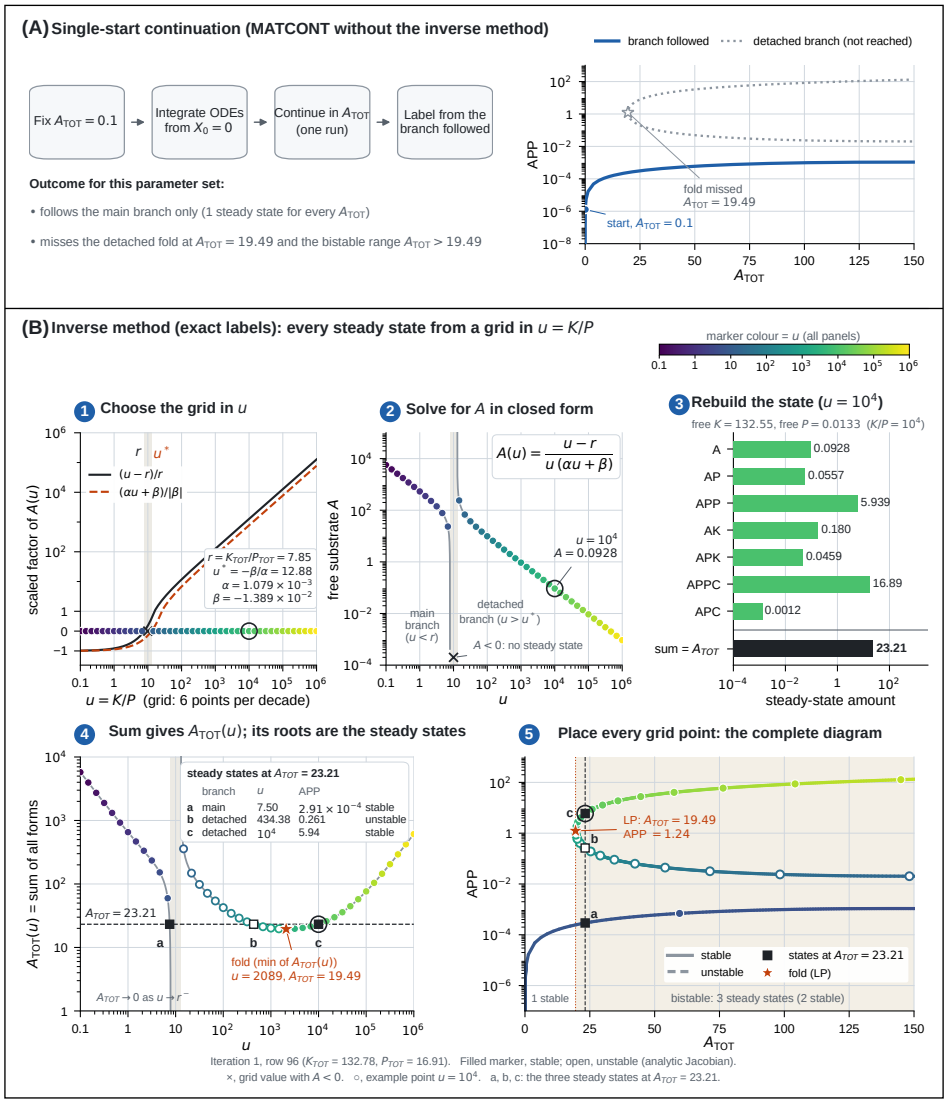


Fig 3. The inverse (u-parametrisation) method. (A) Single-start continuation (MATCONT without the inverse method) starts at $A_{\mathrm{TOT}} = 0.1$ from $X_0 = 0$ and follows only the main branch. It misses the detached branch (dotted), its fold at $A_{\mathrm{TOT}} = 19.49$ and the bistable range beyond it. (B) The inverse method parametrises every steady state by $u = K/P$. (1) A grid in u (6 points per decade; colour encodes u throughout). The numerator and denominator of $A(u)$ (Eq. (8)) vanish at $r = K_{\mathrm{TOT}}/P_{\mathrm{TOT}} = 7.85$ and $u^* = -\beta/\alpha = 12.88$. Here $\alpha = 1.079 \times 10^{-3}$ and $\beta = -1.389 \times 10^{-2}$ (Eq. (8)) are set by the second and the first phosphorylation site, respectively. (2) $A(u)$ in closed form (Eq. (8)). In the shaded band $r < u < u^*$, $A < 0$, so these grid values (×) are not steady states. (3) K, P and all substrate forms follow from u and A (Eq. (7)). Their sum gives A_{TOT} (Eq. (9)); shown for the circled point $u = 10^4$ ($A_{\mathrm{TOT}} = 23.21$). (4) The steady states at a given A_{TOT} are the roots of $A_{\mathrm{TOT}}(u)$. At $A_{\mathrm{TOT}} = 23.21$ there are three (a, b, c). The detached fold (star) is the minimum of $A_{\mathrm{TOT}}(u)$ ($u = 2089$, $A_{\mathrm{TOT}} = 19.49$). (5) Placing each grid point at its A_{TOT} gives the complete bifurcation diagram: one steady state below the fold, and three (two stable) above it. Filled markers and solid lines, stable; open markers and dashed lines, unstable (analytic Jacobian). No initial condition, time integration or grid in A_{TOT} is needed. Parameter set: Iteration 1, row 96 (S-dominant, Class 5). Code: `support/ss_structure.m` (folder A).

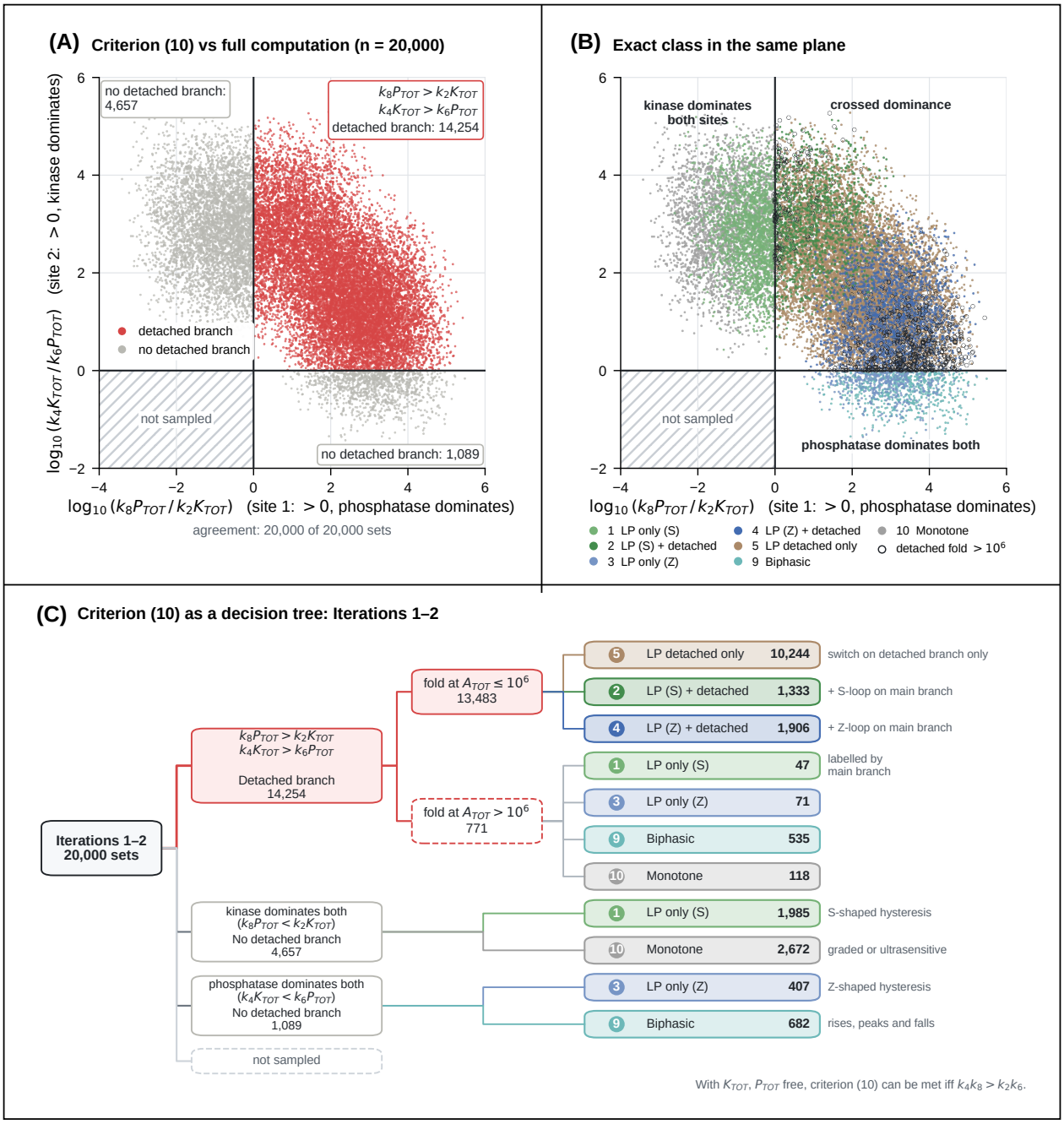


Fig 4. A closed-form criterion predicts every detached branch. All 20,000 parameter sets of Iterations 1 and 2, placed on the two catalytic-dominance ratios of criterion (10): $\log_{10}(k_8P_{\mathrm{TOT}}/k_2K_{\mathrm{TOT}})$ (positive, the phosphatase dominates site 1) and $\log_{10}(k_4K_{\mathrm{TOT}}/k_6P_{\mathrm{TOT}})$ (positive, the kinase dominates site 2). (A) Coloured by the full steady-state computation: red, detached branch at any A_{TOT}; grey, none. Criterion (10) predicts a detached branch if and only if both ratios exceed 1 (upper-right quadrant); it agrees with the computation for all 20,000 sets. (B) Coloured by the exact class; open circles, detached fold above the window ($A_{\mathrm{TOT}} > 10^6$). (C) The criterion as a decision tree. Crossed dominance gives a detached branch (14,254 sets): with the fold inside the window (13,483) the set is bistable through it (Class 5) or also has S- or Z-shaped hysteresis on the main branch (Classes 2 and 4); with the fold above the window (771) it is labelled by its main branch. Kinase dominance at both sites gives Classes 1 and 10; phosphatase dominance at both gives Classes 3 and 9. The lower left quadrant is not sampled. Code: `Fig_criterion_v3.m` (folder A).

passed through a bistable region without detecting it (3,962 of the 5,101 single-start "no bifurcation" sets with a detached branch; Supplementary Fig. 4). The initial condition itself is not the cause: integrating from $X_0 = 0$ at twice the A_{TOT} of the fold reaches the detached stable state in 41 of 100 test cases. Second, the single-start runs were limited by a fixed number of continuation points, so the A_{TOT} range they covered differed between parameter sets (median upper end 1,244; 5–95% range 257–1,745). All 1,116 branch points missed in Iterations 3 and 4 lie beyond that range (median $A_{\mathrm{TOT}} \approx 3{,}100$). Wherever single-start continuation did reach a singularity, it located it accurately (all 8,890 branch points to a relative difference below 10^{-5}, median 1.2×10^{-9}; the 4,115 folds of coexisting LP and BP to below 10^{-4} in 99.95% of cases; Supplementary Fig. 3): its errors were omissions. Supplementary Table 24 summarises the three labelling procedures; in its "Bistable" rows both fractions refer to all 10,000 sets of each iteration, a failed single-start run counting as no singularity found.

Class distribution under exact labelling. Table 2 gives the ten-class distribution of the exactly labelled ensemble. Fig 5 compares, for all 10,000 parameter sets of each iteration, the single-start labels with the exact labels of the same sets (the exact counts are those of Fig 2); failed single-start runs are shown as a separate bar. Because single-start continuation cannot distinguish a main-branch fold with and without a detached branch, main-branch hysteresis is shown there as one class per orientation (Classes 1+2 and 3+4). Classes defined by singularities that continuation did find are confirmed almost completely; the differences are concentrated in the "biphasic" and "no bifurcation" classes. In Iteration 1, 3,845 parameter sets have a fold on a detached branch only (Class 5); they come from the single-start biphasic class (2,519 → 0) and from "no bifurcation" (4,627 single-start sets against 2,790 exact monotone sets), and S-shaped hysteresis on the main branch also becomes more frequent (2,734 → 3,365) because the exact labels include folds beyond the single-start continuation range. In Iteration 2 the effect is strongest: 6,399 sets have a fold on a detached branch only, "no bifurcation" disappears (4,083 → 0) and biphasic responses fall from 3,763 to 1,217. In Iteration 3, BP-only sets increase (4,437 → 4,971) as pitchforks beyond the single-start continuation range are found; the 43 sets labelled LP+BP by single-start continuation have a fold that coincides with the branch point (41 of them are exact BP only), and the 31 exact LP+BP sets of this iteration are the flagged sets of Table 2. In Iteration 4 both LP+BP classes grow (2,126 → 2,563 and 2,006 → 2,439), at the expense of BP only (278 → 0), where the fold on the crossing branch had been missed, and of "no bifurcation" (5,590 → 4,998). In the ten-class ensemble itself, 1,333 of the 3,365 S-shaped sets of Iteration 1 (40%) and 1,906 of the 2,384 Z-shaped sets of Iteration 2 (80%) also carry a fold on a detached branch, so that a detached fold is present in 51.8% and 83.1% of all sets of these iterations.

Multistationarity is common, not rare. In the exactly labelled ensemble (per-iteration details in Supplementary Note 2 and Supplementary Table 2) 72.1%, 87.8%, 50.0% and 50.0% of the parameter sets of Iterations 1–4 (Classes 1–8; 65.0% of all 40,000 sets) are bistable for some A_{TOT} in the window, against 27.4%, 19.1%, 44.8% and 44.1% with single-start continuation on the same parameter sets. In the asymmetric iterations most of the increase comes from detached branches, which carry a fold in the window in 51.8% and 83.1% of the sets (Classes 2, 4 and 5); in the symmetric iterations a branch point exists if and only if $k_4 > k_2$, so the bistable fraction equals the fraction of sampled sets with $k_4 > k_2$. Because the labels refer to a finite window and detached folds span a wide range of A_{TOT} (5–95% range 37 to 1.1×10^6), Supplementary Table 22 reports how the multistationary fraction depends on the window; the conclusion that multistationarity is the typical behaviour of this motif whenever criterion (10) holds is independent of the choice. The pooled fraction (65.0%) is an equal-weight average over four strata of our design (Table 1); it describes this ensemble, not how often bistability occurs in cells, which would require experimentally informed priors on the parameters (Discussion). The statement that does not depend on the design is the conditional one:

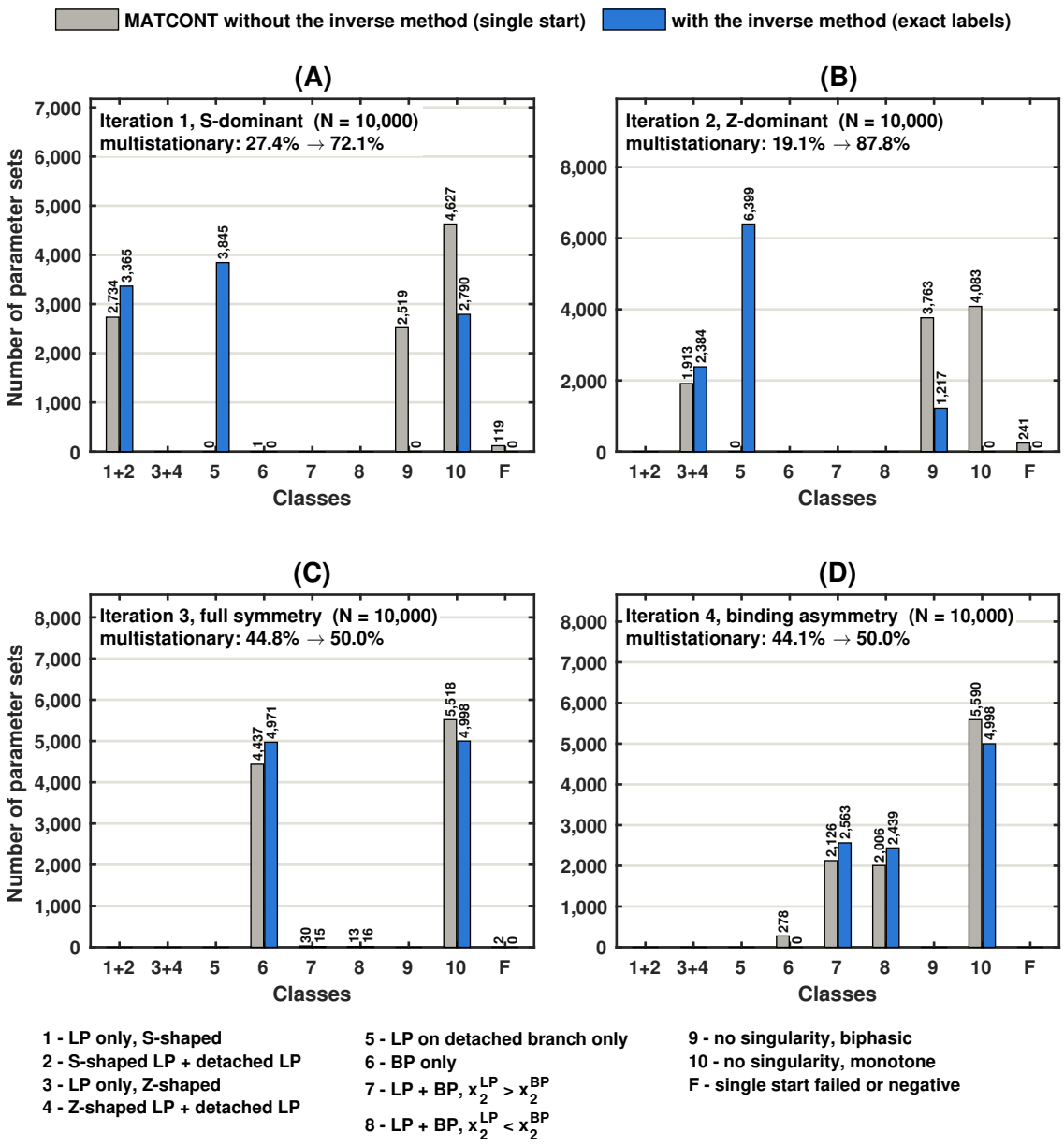


Fig 5. Single-start continuation under-counts multistationarity. Class counts per sampling iteration obtained with MATCONT without the inverse method (grey; single start) and with the inverse method (blue; exact labels) for all 10,000 parameter sets of each iteration, (A)–(D) Iterations 1–4, $A_{\mathrm{TOT}} \leq 10^6$. Both sets of bars add up to 10,000 per iteration; the blue bars are the counts of Fig 2 and Table 2. Classes as in Table 1; main-branch hysteresis is shown as one class per orientation (1+2, S-shaped; 3+4, Z-shaped), because single-start continuation cannot tell whether a detached branch is also present; the single-start "no bifurcation" label is counted as Class 10. F: single-start runs that failed or ended on negative concentrations (119, 241, 2 and 0). Class 5 (LP on a detached branch only) cannot be detected by single-start continuation. The text in each panel gives the fraction of multistationary parameter sets (Classes 1–8) found without and with the inverse method. Code: `make_comparison_figures.m` (folder B).

Table 2. Class distribution of the exactly labelled ensemble. Number of parameter sets per class and sampling iteration ($A_{\mathrm{TOT}} \leq W = 10^6$). *Includes flagged sets whose fold coincides with the branch point (31 in Iteration 3, 1 in Iteration 4); these are excluded from machine learning. Code: `Ensemble_LHS_v3.m`, `Manuscript_numbers_v3.m` (folder A).

Class	Label	Iter. 1	Iter. 2	Iter. 3	Iter. 4	Total
1	LP only, S-shaped	2,032	–	–	–	2,032
2	S-shaped LP + detached LP	1,333	–	–	–	1,333
3	LP only, Z-shaped	–	478	–	–	478
4	Z-shaped LP + detached LP	–	1,906	–	–	1,906
5	LP on detached branch only	3,845	6,399	–	–	10,244
6	BP only	–	–	4,971	–	4,971
7	LP + BP, $x_2^{LP} > x_2^{BP}$	–	–	15*	2,563*	2,578
8	LP + BP, $x_2^{LP} < x_2^{BP}$	–	–	16*	2,439	2,455
9	no singularity, biphasic	–	1,217	–	–	1,217
10	no singularity, monotone	2,790	–	4,998	4,998	12,786
	Multistationary (Classes 1–8)	7,210	8,783	5,002	5,002	25,997
	Total	10,000	10,000	10,000	10,000	40,000
	of which flagged	0	0	31	1	32

every set that satisfies criterion (10) has a detached branch with a fold (Fig 4A), and therefore two stable states over a range of A_{TOT} (verification (iii), Methods).

Branch points lie on the enzyme-balanced branch, and their orientation defines Classes 7 and 8

Why every branch point satisfies $K = P$. In the symmetric and binding-asymmetric iterations, every branch point found satisfies $K = P$, together with $x_3 = x_5$ and $x_4 = x_6$, to machine precision, although the locus $K = P$ is not invariant under the flow for general parameters. The parametrisation explains why. In Eq. (8) the sign of $A(u)$ can change only at $u = r = K_{\mathrm{TOT}}/P_{\mathrm{TOT}}$ and at $u^* = -\beta/\alpha$, and substituting α and β gives (derived in Mathematica [46]; Supplementary File 6, notebook `Analytical_Derivations_StepByStep_final.nb`, Section 7)

$$u^* = \frac{K_{M3}\, k_8\, (1 - r k_2/k_8)}{K_{M7}\, k_2\, (r k_4/k_6 - 1)}, \tag{11}$$

in which K_{M1} cancels and K_{M5} does not appear: u^* is independent of k_1, k_{11}, k_5 and k_{55}. Under the constraints shared by Iterations 3 and 4 ($K_{\mathrm{TOT}} = P_{\mathrm{TOT}}$, $k_6 = k_2$, $k_7 = k_3$, $k_{77} = k_{33}$, $k_8 = k_4$) one has $K_{M3} = K_{M7}$ and therefore $u^* = r = 1$, whatever the values of the first-binding constants. The steady-state condition then factorises as $(u - 1)(\alpha u A - 1) = 0$, and the steady states form two branches: the line $u = 1$, on which $K = P$ for every value of A and hence of A_{TOT}; and the curve $A = 1/(\alpha u)$, which is positive only when $\alpha > 0$. On the first branch, $K_{\mathrm{TOT}} = P_{\mathrm{TOT}}$ and $K = P$ imply $x_3 + x_4 = x_5 + x_6$, and $K_{M3} = K_{M7}$ gives $x_4 = x_6$ and hence $x_3 = x_5$. The two branches cross at $u = 1$, $A = 1/\alpha$: this is the branch point, and its location $A_{\mathrm{TOT}}^{BP} = A_{\mathrm{TOT}}(u = 1, A = 1/\alpha)$ follows from Eq. (9). Since $\alpha > 0$ if and only if $k_4 > k_2$ under these constraints, a branch point exists exactly when $k_4 > k_2$: 5,003 sets per iteration have $k_4 > k_2$, and in all but one of them per iteration the branch point lies inside the window (5,002 sets of Classes 6–8 in each iteration). Breaking the symmetry only in the first binding step (Iteration 4) cannot remove the branch point or move it off $K = P$, because $u^* = r$ does not involve the constants that are perturbed (Fig 6).

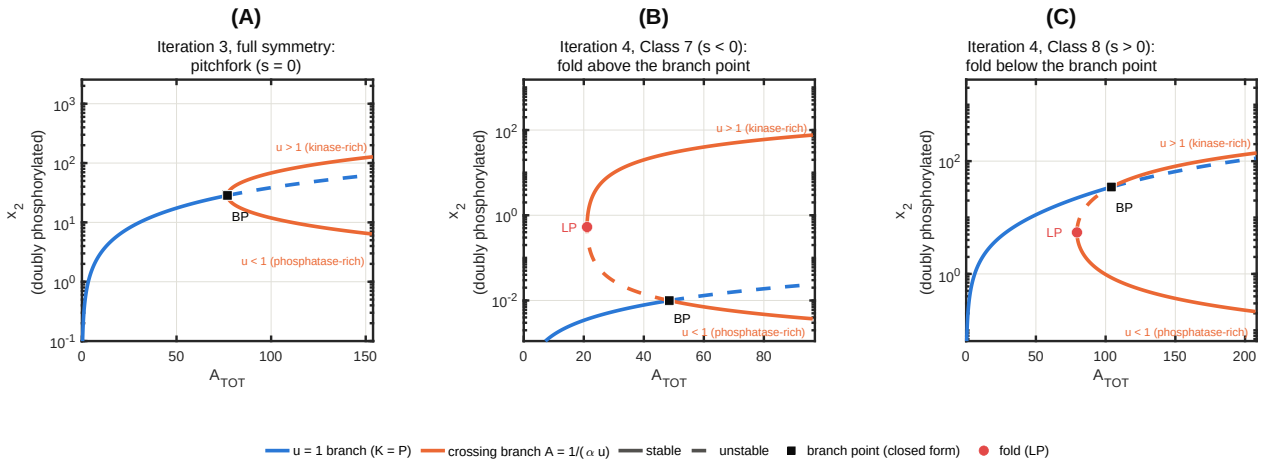


Fig 6. Branch points arise where two branches of steady states cross on the line $K = P$. Exact steady states of three parameter sets, plotted as the output x_2 against A_{TOT}. Blue: the branch $u = K/P = 1$, which exists for all A_{TOT} whenever $u^* = r$ (Iterations 3 and 4). Orange: the crossing branch $A = 1/(\alpha u)$, with its kinase-rich arm ($u > 1$, high x_2) and phosphatase-rich arm ($u < 1$, low x_2). Black square: the branch point at $u = 1$, $A = 1/\alpha$ (closed form); red circle: fold. Solid, stable; dashed, unstable. (A) Iteration 3: both arms of the crossing branch leave the branch point towards larger A_{TOT} and the $u = 1$ branch loses stability there (pitchfork, $s = 0$). (B,C) Iteration 4: with asymmetric first binding the crossing is tilted (transcritical, $s \neq 0$) and one arm turns back at a fold before the branch point, so that bistability begins at the fold. (B) $s < 0$: the fold lies on the kinase-rich arm, above the branch point in x_2 (Class 7). (C) $s > 0$: the fold lies on the phosphatase-rich arm, below the branch point (Class 8). Code: `make_comparison_figures.m` (folder B).

Pitchfork or transcritical: the slope of the crossing branch. The type of a branch point is determined by the slope of the crossing branch where it meets the line $u = 1$,

$$s = \left.\frac{\mathrm{d}\ln A_{\mathrm{TOT}}}{\mathrm{d}\ln u}\right|_{u=1,\ A=1/(\alpha u)}. \tag{12}$$

If $s = 0$ the two arms of the crossing branch ($u > 1$ and $u < 1$) leave the branch point on the same side in A_{TOT} and the bifurcation is a pitchfork; if $s \neq 0$ they leave on opposite sides and it is transcritical [9, 47]. Evaluated for every parameter set (Supplementary Table 25), s is zero to within 10^{-10} for all 5,002 branch points of Iteration 3 in the window, which are therefore pitchforks, and differs from zero for all 5,002 branch points of Iteration 4 ($|s|$ from 10^{-6} to 1), which are transcritical. Breaking the symmetry of the first binding step preserves the crossing but unfolds every pitchfork into a transcritical bifurcation, as expected when a symmetry that produces a pitchfork is broken [47].

Classes 7 and 8 are the two orientations of the transcritical crossing. In Iteration 4 the arm of the crossing branch that extends towards lower A_{TOT} turns back at a fold. The sign of s determines which arm this is: for $s < 0$ it is the kinase-rich arm ($u > 1$, higher x_2), so the fold lies above the branch point in output concentration (Class 7); for $s > 0$ it is the phosphatase-rich arm ($u < 1$), and the fold lies below it (Class 8). This holds without exception: $s < 0$ for all 2,563 Class-7 sets and $s > 0$ for all 2,439 Class-8 sets, and the fold lies on the $u > 1$ arm in every Class-7 set and on the $u < 1$ arm in every Class-8 set (Supplementary Fig. 24B). The branch-point normal form gives the same split: the two branch slopes obtained from the branching equation have opposite signs in all 2,563 Class-7 sets and the same sign in all 2,439 Class-8 sets. In the plane of the two first-binding

asymmetries, $\log_{10}(k_1/k_5)$ and $\log_{10}(k_{11}/k_{55})$, the boundary between the two orientations is close to a straight line, which is the boundary the classifier recovers (Fig 7; Supplementary Fig. 16C). In Iteration 3 the 31 sets labelled 7 or 8 are the degenerate limit of this picture: under full symmetry $s = 0$, the turning point of the crossing branch coincides with the pitchfork, and their fold coefficient vanishes to round-off ($|a_{LP}| \leq 9.4 \times 10^{-13}$); they are flagged rather than counted as folds.

A closed-form criterion for Classes 7 and 8. Along the crossing branch the parametrisation also gives the slope s in closed form. Under the constraints of Iterations 3 and 4, substituting $A = 1/(\alpha u)$ into Eqs (7)–(9) and differentiating at $u = 1$ gives (Methods; derived and simplified in Mathematica, Supplementary File 6, `Analytical_Derivations_StepByStep_final.nb`, Section 8)

$$s = \frac{k_4^2\,(K_{M5} - K_{M1})}{K_{\mathrm{TOT}}\,(k_4^2 - k_2^2) + k_4^2\,(K_{M1} + K_{M5}) + k_2 k_4\,K_{M3}}, \tag{13}$$

with $K_{M1} = (k_{11} + k_2)/k_1$, $K_{M3} = (k_{33} + k_4)/k_3$ and $K_{M5} = (k_{55} + k_2)/k_5$ (recall $k_6 = k_2$). The denominator is positive whenever a branch point exists ($k_4 > k_2$). Under full symmetry $K_{M5} = K_{M1}$, so $s = 0$ and every branch point is a pitchfork; with first-binding asymmetry $s \neq 0$, and

$$\text{Class 7} \iff K_{M5} < K_{M1}, \qquad \text{Class 8} \iff K_{M5} > K_{M1}. \tag{14}$$

The class is thus decided by which enzyme binds its substrate more tightly at the outer steps of the cycle: the fold lies on the kinase-rich arm (Class 7) when the phosphatase binds the doubly phosphorylated substrate more tightly than the kinase binds the unphosphorylated substrate, and on the phosphatase-rich arm (Class 8) in the opposite case. Criterion (14) reproduces the class of all 5,002 Iteration-4 parameter sets of Classes 7 and 8 without exception; the slope formula is also evaluated in Mathematica for the symmetric Class-6 example and for Class-7 and Class-8 examples taken from the ensemble (Supplementary File 6, `Excel_Parameter_Examples_Condition_Checks_final.nb`, Section 4).

Normal-form coefficients confirm the type of every singularity. Close to a bifurcation point the dynamics of a model of any size can be reduced to a short polynomial, the normal form, whose coefficients state in a model-independent way which bifurcation occurs and whether it is degenerate [9]. At a fold, a non-zero quadratic coefficient a_{LP} identifies a generic saddle-node; at a branch point, the quadratic coefficient a of the branching equation is zero for a pitchfork and non-zero for a transcritical bifurcation. The coefficients, computed in MATCONT's conventions from the analytic derivatives and checked against MATCONT (Methods), are summarised in Supplementary Table 25 and Supplementary Fig. 24 for all 40,018 singularities inside the window. All 10,004 branch points are simple crossings of two branches (positive discriminant of the branching equation). The coefficient a separates the two iterations completely: it vanishes to round-off at every branch point of Iteration 3 ($|a| \leq 3 \times 10^{-14}$; pitchforks) and is non-zero at every branch point of Iteration 4 ($|a|$ from 4.3×10^{-13} to 0.12; transcritical), in agreement with the slope s. All 24,981 folds on the main and detached branches of Iterations 1 and 2 have $|a_{LP}| \geq 3.9 \times 10^{-9}$ and are non-degenerate saddle-nodes (per-iteration values: Supplementary Note 4 and Supplementary Table 4). On the crossing branch of Iteration 4 the fold coefficient is non-zero as well, but becomes small (below 10^{-10} in 12 of 5,002 sets) when the fold approaches the branch point as the first-binding symmetry is restored.

A ready reckoner of the ten classes

The exact analysis identifies two ratios that control the branch structure: $x = \log_{10}(k_2 K_{\mathrm{TOT}}/k_8 P_{\mathrm{TOT}})$, which compares the maximal phosphorylation and dephosphorylation rates at the first site (the negative of the horizontal axis of Fig 4), and $y = \log_{10}(k_4 K_{\mathrm{TOT}}/k_6 P_{\mathrm{TOT}})$, which compares them at

the second site. Fig 7 places every class in this plane against the whole ensemble and shows the exact bifurcation diagram of a representative parameter set, so that the region a class occupies, the condition that defines it and its diagram can be read together. The classes occupy distinct, mechanistically interpretable regions. Folds on a detached branch (Classes 2, 4 and 5) occur only where the phosphatase dominates the first site and the kinase the second (upper-left quadrant, $x < 0$ and $y > 0$, criterion (10); upper-right quadrant of Fig 4B): Class 5 fills this quadrant, and Classes 2 and 4 are the part of it in which the main branch also has S- or Z-shaped hysteresis. S-shaped hysteresis alone (Class 1) and the monotone sets of Iteration 1 occur where the kinase dominates both sites (upper-right quadrant), and Z-shaped hysteresis alone (Class 3) and biphasic responses (Class 9) where the phosphatase dominates both (lower-left quadrant).

Open circles: detached branches beyond the window. Some sets of Classes 1, 3, 9 and 10 lie inside the upper-left quadrant (open circles in Fig 7A and Fig 4B; 771 sets in total: 2.3% of Class 1, 14.9% of Class 3, 44.0% of Class 9 and 0.9% of Class 10). They are not exceptions to criterion (10): every one of them has a detached branch, but its fold lies above the labelling window, at A_{TOT} between $10^{6.0}$ and $10^{10.7}$, so within the window only the main branch is seen and the set keeps its main-branch class. Fig 7B shows the diagram of the starred set of each of these classes extended beyond $A_{\mathrm{TOT}} = 10^6$: the detached branch appears outside the window, that is, outside the range of total substrate concentrations that can be reached experimentally, and with the window widened to include its fold the set would be Class 2, 4 or 5 (Supplementary Note 19 and Supplementary Fig. 22, two such sets of each class). These sets lie close to the quadrant edges (median distance $\min(-x, y)$ of 0.45 decades, against 1.21 decades for Classes 2, 4 and 5), where the detached fold moves to large A_{TOT}.

How well the regions separate the classes. Each class is largely confined to its region of the plane: 97.7% of Class 1 lies in the upper-right quadrant, 85.1% of Class 3 in the lower-left, 99.1% of Class 10 in the upper-right quadrant or on the lower-right half of the anti-diagonal, and all sets of Classes 2, 4, 5 and 6 in their regions; the sets outside are the open circles, together with two Class-10 sets of Iterations 3 and 4 whose branch point lies above the window. Class 9 occupies two quadrants, the lower-left (682 sets, 56.0%) and the upper-left (535 sets, 44.0%, all with the detached fold above the window). Because pairs of classes share a region (1 and 10; 2 and 5; 4 and 5; 3 and 9), the fraction of the sets in a region that belong to one class is lower (21–72% for Classes 1–5 and 9, 86.5% for Class 10, 94–100% for Classes 6–8). The members of each pair differ only in whether the main branch folds, which has no closed-form condition and no region in this plane; the best single-feature thresholds separate them at 70% ($\log_{10}(k_3/k_7) < -1.6$, Class 1 against 10) and 86% ($\log_{10}(k_3/k_7) > 1.5$, Class 3 against 9), and not above the majority baseline for Classes 2 and 4 against Class 5. Classes 7 and 8 share the upper-left half of the anti-diagonal and are separated by the first-binding asymmetries (see below); their exact boundary $K_{M5} = K_{M1}$, written in the features as $(k_1/k_5)(1 + k_{55}/k_2) = 1 + k_{11}/k_2$, classifies all 5,001 sets.

Two further exact statements complete the picture. First, the side of $u = r$ on which the main branch lies follows from Eq. (7): near $u = r$ the numerator of $A(u)$ changes sign, so $A > 0$ requires $u > r$ if $\alpha r + \beta > 0$ and $u < r$ otherwise. All 3,365 sets of Classes 1 and 2 have their main branch on $u > r$, and all 3,601 sets of Classes 3, 4 and 9 on $u < r$: the orientation of the main-branch hysteresis, S or Z, is the side of the main branch. Second, on the side $u < r$ with $x < 0$ ($\beta < 0$) the main branch runs to $u \to 0$, where $x_2 \approx -a_1 a_2 r u/\beta \to 0$, so x_2 rises and falls with A_{TOT}; on every other main branch $x_2 \to \infty$. Among the sets without a singularity, Class 9 (biphasic) therefore occurs if and only if $x < 0$ and the main branch lies on $u < r$, which holds in all 1,217 Class-9 sets and in none of the 12,786 Class-10 sets. In the symmetric and binding-asymmetric iterations $x = \log_{10}(k_2/k_4) = -y$, so these sets lie on the anti-diagonal: those with a branch point (Classes 6–8) on its upper-left half, where $k_4 > k_2$, and the monotone sets on its lower-right half. Classes 6–8 coincide in this

plane and are separated by the first-binding asymmetries $\log_{10}(k_1/k_5)$ and $\log_{10}(k_{11}/k_{55})$: under full symmetry all Class-6 sets lie at $k_1 = k_5$, $k_{11} = k_{55}$; with first-binding asymmetry, Class 7 lies above and Class 8 below a straight line, $\log_{10}(k_{11}/k_{55}) \approx 1.5\ \log_{10}(k_1/k_5)$, which separates them at 95.2% (Supplementary Fig. 16C). The exact boundary is $K_{M1} = K_{M5}$ (Eq. (14)); it reduces to the diagonal $k_1/k_5 = k_{11}/k_{55}$ when the unbinding constants exceed the catalytic constant k_2, and to the vertical $k_1 = k_5$ when k_2 dominates, so the fitted line lies between the two limits and the misclassified sets lie close to it (Supplementary Fig. 16C). Of the class boundaries, only one, whether the main branch carries a fold pair (1 against 10, 2 against 5, 3 against 9 and 4 against 5), has no closed form and no region in this plane.

Class boundaries are sharp surfaces in specific parameter ratios. Because each boundary depends on a specific ratio of a few parameters, it is not aligned with Euclidean proximity in the space of all 14 log-parameters, and a neighbourhood-preserving embedding of that space does not reveal it. A two-dimensional t-SNE embedding [48] of the log-parameters recovers the sampling iteration of a parameter set at 82.7% (baseline 25.2%) but its class at only 64.4% (baseline 31.8%), compared with 89.1% for a supervised classifier trained on the same 14 log-parameters and 92.7% with all 105 features (Supplementary Note 12, Supplementary Table 12 and Supplementary Fig. 15). Centred log-ratio coordinates [49] change this only marginally (65.3–66.8%), and re-embedding each sampling regime separately does not close the gap: within the regimes $K_{\mathrm{TOT}} > P_{\mathrm{TOT}}$, $K_{\mathrm{TOT}} < P_{\mathrm{TOT}}$ and $K_{\mathrm{TOT}} = P_{\mathrm{TOT}}$ the embedding recovers the class at 62–74%, against 76–93% for a supervised classifier on the same representation, and in the regime $K_{\mathrm{TOT}} < P_{\mathrm{TOT}}$ it is no better than the majority baseline (64.1% against 64.3%). The class structure of the ensemble is visible only in coordinates that follow the mechanism, such as those of Fig 7.

Predicting the bifurcation class from kinetic parameters

We next asked how well the bifurcation class of a parameter set can be predicted from its kinetic parameters alone, and which parameter combinations carry that information (Fig 8). All analyses use the ten exact class labels for the 39,968 parameter sets that remain after removal of the 32 flagged sets, split into training (70%), validation (10%) and held-out test (20%) sets stratified by sampling iteration and class (Methods).

The sampling design defines four strata. The symmetry constraints of the iterations leave a signature in the features. Seven pairwise ratios ($K_{\mathrm{TOT}}/P_{\mathrm{TOT}}$, k_1/k_5, k_{11}/k_{55}, k_2/k_6, k_3/k_7, k_{33}/k_{77}, k_4/k_8) are exactly equal to one in every parameter set of Iteration 3, and five of them in every set of Iteration 4 (Supplementary Table 26). The pattern of these ratios identifies the sampling iteration of every parameter set without error, and LightGBM trained to predict the iteration reaches 100% accuracy (baseline 25.0%). Because each iteration produces its own subset of classes, part of the information that separates the classes in the pooled data is the sampling design itself: eight of the ten classes occur in a single iteration, Class 5 occurs in Iterations 1 and 2, and Class 10 in Iterations 1, 3 and 4. In particular, the S-shaped (Classes 1, 2) and Z-shaped (Classes 3, 4) classes are separated in the pooled data by the sign of $K_{\mathrm{TOT}} - P_{\mathrm{TOT}}$, i.e. by construction, and the design-constrained ratios $K_{\mathrm{TOT}}/P_{\mathrm{TOT}}$ and k_1/k_5 are among the top-ranked pooled features of every model (Supplementary Fig. 10). We therefore report pooled performance, but base all mechanistic interpretation on analyses within each iteration, where the design-constrained ratios have zero variance (Supplementary Note 7).

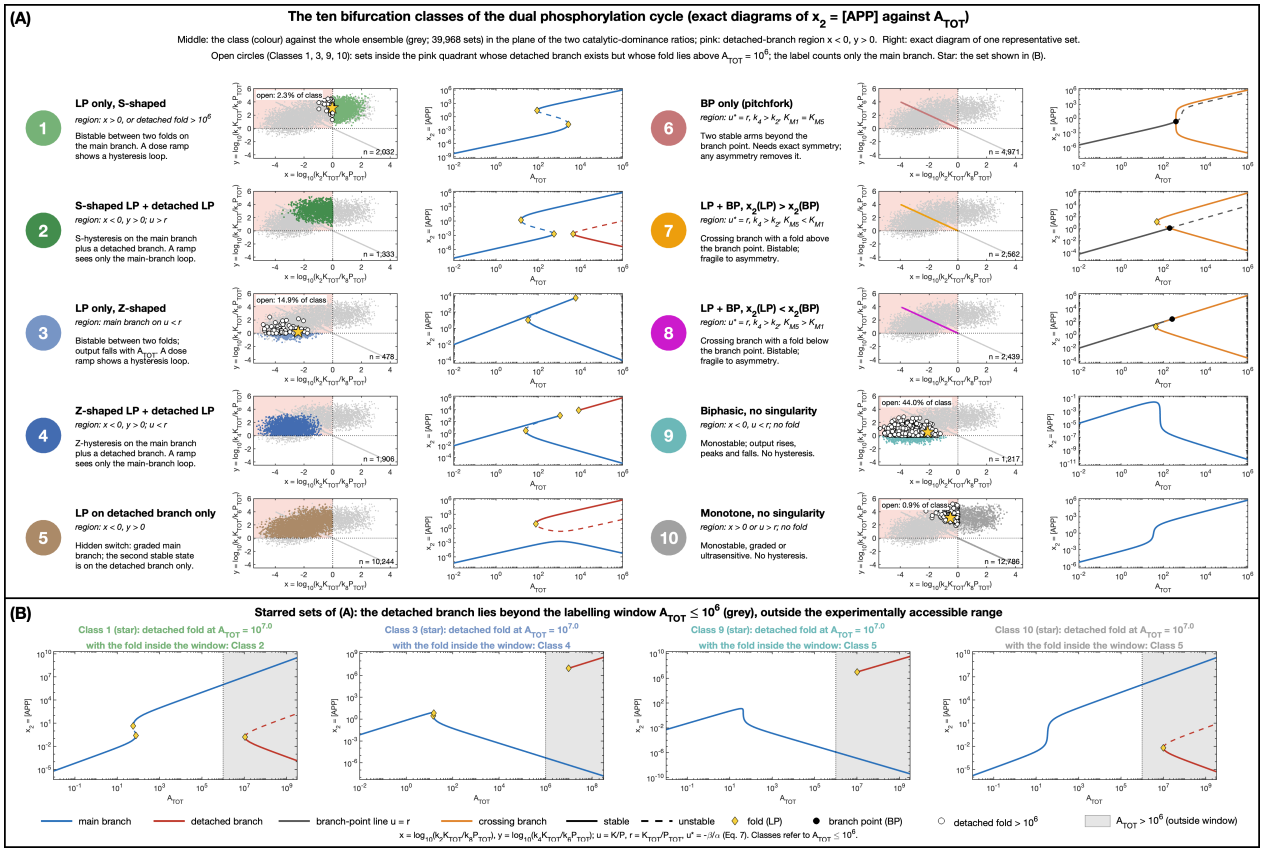


Fig 7. A ready reckoner of the ten bifurcation classes. (A) For every class: its definition and the exact region it occupies, a one-line reading of its experimental meaning, the class (colour) against the whole ensemble (grey; random subset of 9,000 of the 39,968 sets used for machine learning) in the plane of the two catalytic-dominance ratios $x = \log_{10}(k_2 K_{\mathrm{TOT}}/k_8 P_{\mathrm{TOT}})$ and $y = \log_{10}(k_4 K_{\mathrm{TOT}}/k_6 P_{\mathrm{TOT}})$, and the exact bifurcation diagram ($x_2 = [APP]$ against A_{TOT}) of one representative parameter set. Pink: the detached-branch region $x < 0$, $y > 0$ (criterion (10)). Open circles (Classes 1, 3, 9 and 10): sets inside the pink region whose detached branch exists but whose fold lies above $A_{\mathrm{TOT}} = 10^6$, so that the label counts only the main branch (2.3%, 14.9%, 44.0% and 0.9% of the four classes); star: the set shown in (B). Sets of Iterations 3 and 4 lie on the anti-diagonal $y = -x$, with branch points (Classes 6–8) on its upper-left half ($k_4 > k_2$). (B) Exact diagrams of the starred sets of (A), extended beyond the labelling window (grey, $A_{\mathrm{TOT}} > 10^6$): the detached branch and its fold lie outside the window, i.e. outside the experimentally accessible range of total substrate; with the fold inside the window the sets would be Classes 2, 4, 5 and 5. Line styles and symbols: legend at the bottom. Representative and starred sets: Methods (Class transitions). Classes in (A) refer to $A_{\mathrm{TOT}} \leq 10^6$. All open circles and two examples per class: Supplementary Fig. 22; the map in the plane of the first-binding asymmetries: Supplementary Fig. 16C. Code: `Class_transitions/TR_01_ready_reckoner.m` (folder A).



Fig 8. Machine-learning pipeline: from the labelled ensemble to the ten-class classifier. (1) Input vector. Each parameter set of the Latin hypercube ensemble (Iterations 1–4; Fig 1) is a vector of the 14 kinetic parameters, labelled with its exact class (39,968 sets after removal of the 32 flagged sets); four parameters of six example sets and the class distribution are shown. Descriptors derived from the labels are never used as inputs. (2) Feature engineering. All features are $\log_{10}$ values: the 14 log-parameters and the 91 pairwise log-ratios form set A (105 features); sets B–F are the variants compared in Supplementary Table 27. Zero-variance and duplicated columns are removed in each training split. (3) Feature selection (two-class). For every pair of classes the features are ranked by the Jensen–Shannon divergence between the two class-conditional histograms, retained up to the broken-stick cut-off, and pooled over all pairs (set D). Example, computed for Classes 1 and 2 of Iteration 1: the classes separate at $\log_{10}(k_2K_{\mathrm{TOT}}/k_8P_{\mathrm{TOT}}) = 0$, and k_2/k_8 ranks first of the 13 retained features. In the decisive two-class tests, two log-ratios with a linear boundary separate Classes 1/2, 3/4 and 7/8. (4) Multiclass classification. Stratified split into training (70%), validation (10%) and held-out test (20%) sets, class weights $n/(Kn_c)$, five classifiers, one class per set. Bars: pooled test accuracy with feature set A (mean ± SD, three splits; naive Bayes, range over feature sets); dashed line, majority-class baseline. Below: LightGBM within each iteration. Code: `ML_v3_10class_parallel.m` (classifiers, feature sets, accuracies) and `ML_v3_feature_engineering.m` (feature engineering and JSD selection, block 3) (folder A).

Table 3. Feature engineering and feature selection. Number of features at each step for the pooled data (in brackets: within Iterations 1–4) and the purpose of each step. Accuracies of every set are given in Supplementary Table 27; the JSD profiles of every class pair, the selected features and two-parameter views in Supplementary Note 9 (Supplementary Figs 6–9 and Supplementary Tables 6 and 7). Code: `ML_v3_feature_engineering.m`, `ML_v3_10class_parallel.m` (folder A).

Step	Operation	Features	Purpose
Input	K_{TOT}, P_{TOT} and the 12 rate constants	14	quantities derived from the labels are never used
Engineering	$\log_{10}$ of each parameter	14	products and ratios become sums and differences
	all pairwise log-ratios $\log_{10}(p_i/p_j)$	$+91 \rightarrow 105$ (set A)	a two-parameter boundary becomes one split
Preprocessing	zero-variance and duplicated columns removed (training rows)	105 (105, 105, 43, 59)	the symmetry constraints make ratios constant or identical
Selection	B: remove the 7 design-constrained ratios	98	does the classifier only recognise the iteration?
	C: remove the 91 log-ratios	14	is feature engineering needed?
	D: Jensen–Shannon divergence for every class pair, broken-stick cut-off, union	64.3 (18.0, 36.3, 2.0, 5.0)	filter selection of the original pipeline
Transformation	E: centred log-ratios of the 14 parameters; F: of the 12 rate constants, with $\log_{10}(K_{\mathrm{TOT}}/P_{\mathrm{TOT}})$	14; 13	scale-free representation

Feature engineering, feature selection and classification

Feature engineering. The classifiers see only the 14 sampled quantities, but the exact analysis shows that the class boundaries are conditions on products and ratios of a few of them (criterion (10), $k_4 > k_2$, the sign of $K_{M5} - K_{M1}$). Tree ensembles split on one feature at a time, so a boundary that is oblique to the parameter axes can only be approximated by a staircase of splits, whereas the log-ratio of the two parameters involved turns it into a single threshold. We therefore engineered the features in two steps, without telling the models which ratios matter: every parameter is log-transformed, and all $\binom{14}{2} = 91$ pairwise log-ratios are added, giving 105 features (feature set A; Fig 8, panel 2, and Table 3). Ratios of three or more parameters, such as those of criterion (10), were deliberately left out, so that a feature ranking pointing to them is a result rather than an input.

Feature selection. From set A, five further feature sets were derived by selection or transformation, each testing one step of the pipeline: B removes the seven design-constrained ratios (does the classifier only recognise the sampling iteration?), C keeps the 14 log-parameters (is feature engineering needed at all?), D applies the divergence-based selection of our original pipeline, and E and F are scale-free centred log-ratio representations. Table 3 lists the number of features left at each step.

The selection itself is informative (Supplementary Note 9). In the pooled data it is driven by the sampling design: for every pair of classes that never occur in the same iteration, the features kept are design-constrained ratios, often separating the two classes perfectly, and all seven of them are kept. Within each iteration, by contrast, it recovers the analytical conditions: k_2/k_8 (site 1 of criterion (10)) heads every Iteration-1 pair that differs in the detached branch, k_4/k_6 (site 2) the corresponding pairs of Iteration 2, k_2/k_4 alone separates the branch-point classes of Iterations 3 and 4, and k_1/k_5, k_1/k_{11} and k_5/k_{55} separate Classes 7 and 8. Because the divergence scores one

feature at a time, it misses boundaries that exist only in combination: $K_{\mathrm{TOT}}/P_{\mathrm{TOT}}$ alone hardly separates Classes 1 and 5 in Iteration 1, yet together with k_2/k_8 it places them on either side of the line $k_8 P_{\mathrm{TOT}} = k_2 K_{\mathrm{TOT}}$ (Supplementary Fig. 9). For this reason the main analyses use all 105 features and let the tree ensembles select.

Classification. Every feature set is used to train every model (LightGBM, XGBoost, CatBoost, random forest and gradient boosting, with naive Bayes and the majority class as baselines) on the training set (70%) of three stratified splits, with training samples weighted inversely to the frequency of their class and five-fold cross-validation inside the training set (Methods); each model returns one of the ten classes, with a class probability, for every held-out parameter set. The feature importances and TreeSHAP contributions of the trained models then show which features carry the class, and within each iteration they point to the ratios of the analytical criteria (Fig 10). The comparison of the six feature sets follows.

Classifier performance

The class is predictable by every tree ensemble, and does not depend on the design-constrained ratios. On the pooled ten-class data, LightGBM with feature set A (the 14 log-parameters and their 91 pairwise log-ratios) reaches $92.66 \pm 0.20\%$ held-out accuracy (balanced accuracy 86.87%, macro-F1 87.29%, Cohen's $\kappa = 0.908$), against a majority-class baseline of 32.00% (Fig 9A; Table 4). XGBoost is close behind ($91.87 \pm 0.44\%$), followed by the random forest ($89.25 \pm 0.38\%$) and CatBoost ($86.49 \pm 0.65\%$); naive Bayes, which assumes independent features, performs poorly (38.9–54.6%), as expected for features that are deterministic functions of one another. CatBoost has the lowest accuracy of the tree ensembles but the highest balanced accuracy (91.29%): with class weights and 300 rounds it favours the small classes at the expense of the large Class 5, and it has not yet converged. Removing the seven design-constrained ratios (set B) lowers LightGBM accuracy by only 1.2 percentage points, to $91.43 \pm 0.08\%$, and the 14 parameters alone (set C) still give $89.13 \pm 0.46\%$. Divergence-based feature selection (set D, on average 64 features) performs as well as the full set ($92.71 \pm 0.17\%$), and the two compact centred log-ratio representations reach 88.8–89.8% with 13–14 features. The ranking of the models is the same on every feature set.

The model comparison with cross-validation inside the training set gives the same picture (Fig 9B). With feature set A, five-fold cross-validated accuracy is $92.22 \pm 0.36\%$ for LightGBM, $91.48 \pm 0.50\%$ for XGBoost, $89.69 \pm 0.42\%$ for gradient boosting, $89.22 \pm 0.29\%$ for the random forest, $86.07 \pm 0.43\%$ for CatBoost and $44.78 \pm 0.31\%$ for naive Bayes, against 31.99% for the most-frequent-class predictor; the held-out accuracies agree with the cross-validated values to within 0.6 percentage points. These accuracies measure interpolation within the four sampled strata: training and test sets are random splits of the same design, so they do not show how a classifier generalises to other parameter ranges, other priors or other mechanisms. The analytical conditions, which the classifier rediscovers, carry no such restriction. Every tree ensemble separates the classes almost perfectly in the ranking sense: the macro-averaged one-vs-rest AUC on the test set is 0.9955 for LightGBM, 0.9952 for XGBoost, 0.9943 for CatBoost, 0.9937 for gradient boosting and 0.9925 for the random forest (naive Bayes 0.956; Supplementary Fig. 25A). The per-class AUC is at least 0.998 for Classes 1, 3 and 6–10 and lowest, 0.984–0.988 for LightGBM, for Classes 2, 4 and 5, the three classes with a fold on a detached branch. Cross-validated accuracies and the per-class results of every model, pooled and within each iteration, are given in Supplementary Note 10 (Supplementary Tables 8–10).

Within each iteration, where the constrained ratios carry no information, all tree ensembles remain far above the baseline (Fig 9B, Supplementary Table 27). With set A, which after removal of constant and duplicated columns contains 105, 105, 43 and 59 features in Iterations 1–4, LightGBM reaches $86.82 \pm 1.33\%$ in Iteration 1 (baseline 38.45%), $84.48 \pm 0.78\%$ in Iteration 2 (baseline 64.00%),

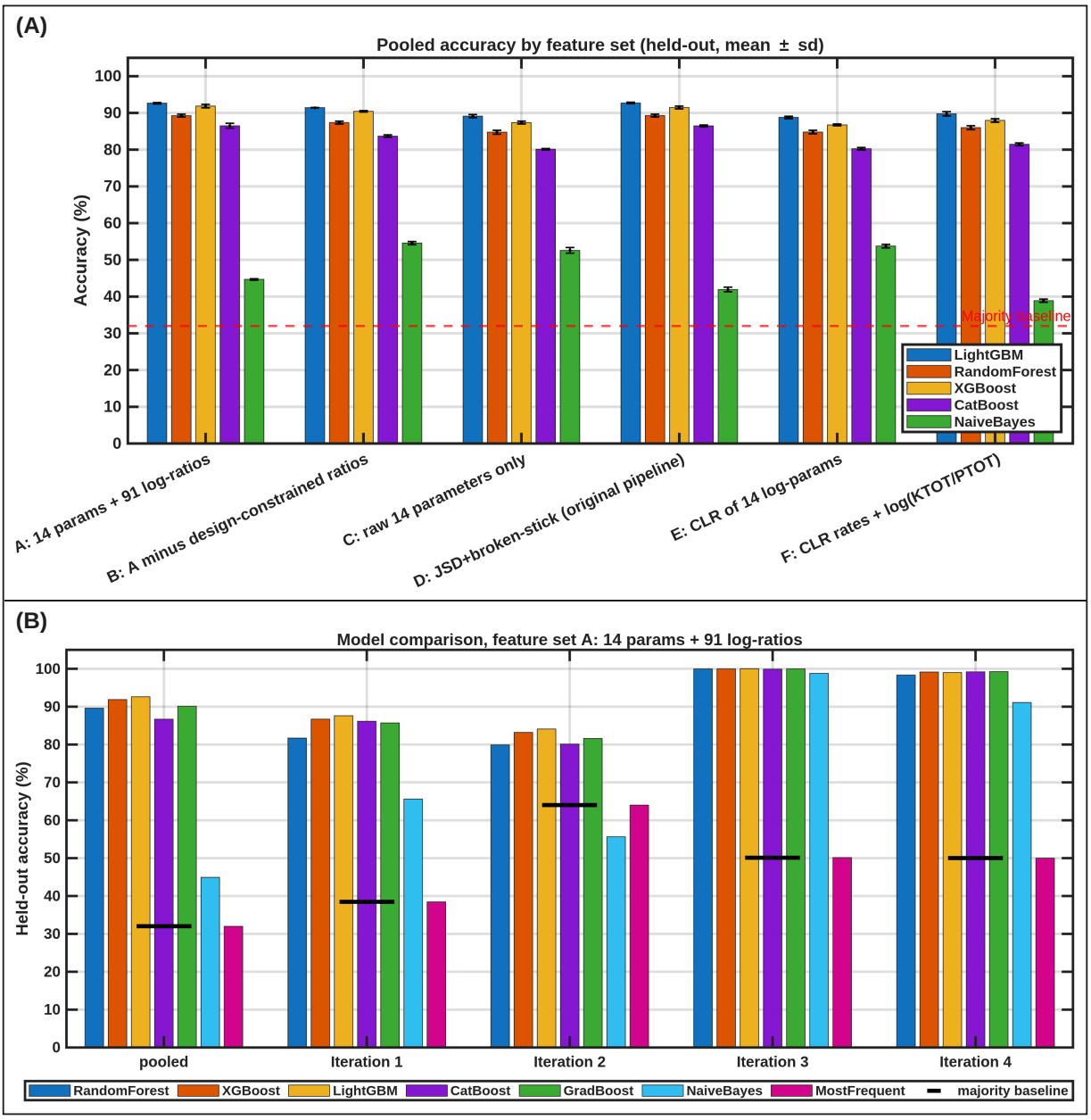


Fig 9. Classifier accuracy by feature set and by model. (A) Pooled held-out accuracy (mean ± SD over three stratified splits) of five classifiers on the ten-class data, for six feature sets: A, the 14 log-parameters and all 91 pairwise log-ratios (105 features); B, set A without the seven design-constrained ratios of Supplementary Table 26 (98); C, the 14 log-parameters alone; D, the ratios selected by Jensen–Shannon divergence and the broken-stick criterion (64 on average); E, centred log-ratio (CLR) of the 14 log-parameters; F, CLR of the 12 rate constants plus $\log(K_{\mathrm{TOT}}/P_{\mathrm{TOT}})$ (13). Red dashed line: majority-class baseline (32.0%). Within-iteration results are given in Supplementary Table 27. (B) Held-out test accuracy with feature set A of the random forest, XGBoost, LightGBM, CatBoost, gradient boosting, naive Bayes and the most-frequent-class predictor, on the pooled data and within each sampling iteration; black bars, majority-class baseline of the test set. All tree ensembles are far above the baseline in every scope; within Iterations 3 and 4 they are almost perfect. Code: `ML_v3_10class_parallel.m` (folder A).

Table 4. Classifier performance (pooled ten-class data, feature set A, held-out test set). Code: `ML_v3_10class_parallel.m` (folder A).

Model	Held-out accuracy (%)	5-fold CV accuracy (%)	Macro AUC
LightGBM	92.66 ± 0.20	92.22 ± 0.36	0.9955
XGBoost	91.87 ± 0.44	91.48 ± 0.50	0.9952
Gradient boosting	90.11	89.69 ± 0.42	0.9937
Random forest	89.25 ± 0.38	89.22 ± 0.29	0.9925
CatBoost	86.49 ± 0.65	86.07 ± 0.43	0.9943
Naive Bayes	38.9–54.6	44.78 ± 0.31	0.956
Majority class	32.00	31.99	–

$99.98 \pm 0.03\%$ in Iteration 3 (baseline 50.15%) and $99.03 \pm 0.28\%$ in Iteration 4 (baseline 50.00%); the within-iteration macro AUC is 0.978, 0.962, 1.000 and 0.9998. The class of a parameter set is therefore predictable from the kinetic parameters themselves and not only from the stratum in which it was sampled. Iterations 3 and 4 are classified almost perfectly by every tree ensemble and with any feature set, including the 14 parameters alone; Iterations 1 and 2 benefit from the explicit pairwise ratios (set A is 4.6–4.8 points above set C), indicating that in the asymmetric regimes the class depends on specific products and ratios of parameters that a tree ensemble cannot easily construct from the raw parameters. In Iteration 3 the divergence-based feature selection retains only two features and still classifies 99.98% of test sets correctly.

300 boosting rounds are sufficient, and the remaining errors are not a matter of training length. The convergence curves (Supplementary Fig. 25B) show the training and validation log-loss and error as functions of the number of boosting rounds. The random forest is stable after about 20 trees. XGBoost and gradient boosting still improve slowly at the end (relative change of the validation log-loss over the last 20% of rounds 2.0% and 2.4%), at a validation error that no longer changes. LightGBM reaches its smallest validation log-loss after 63 of 300 rounds; the log-loss then rises slightly (-6.1% over the last 20% of rounds) while the validation error stays at about 8%, i.e. the model becomes more confident without becoming less accurate. CatBoost alone is not converged at 300 rounds (5.8% change over the last 20%; validation error of Class 5 still 38%), which explains its lower accuracy. The per-class curves (Supplementary Fig. 13) show where the errors lie: Classes 6, 7, 8 and 10 quickly reach a validation error below 4%, whereas for the boosted models the errors of Classes 2 and 4 increase with the number of rounds while that of Class 5 decreases (LightGBM: Class 4 from 21% to 39%, Class 5 from 48% to 9%): the models trade the two small mixed classes against the large Class 5. Within Iterations 3 and 4 every model converges to a per-class validation error of at most 2.5% (Supplementary Fig. 14; Supplementary Note 11 and Supplementary Table 11).

Within each stratum, the classifier rediscovers the analytical conditions for bistability. The most informative features within each iteration (Fig 10A) are the parameters that enter the exact conditions derived from the steady-state parametrisation (Table 5, Fig 10B). In Iteration 1 (Classes 1, 2, 5 and 10) the top-ranked feature is k_2/k_8, followed by $K_{\mathrm{TOT}}/P_{\mathrm{TOT}}$. A detached branch, which distinguishes Classes 2 and 5 from Classes 1 and 10, exists if and only if $k_8 P_{\mathrm{TOT}} > k_2 K_{\mathrm{TOT}}$ and $k_4 K_{\mathrm{TOT}} > k_6 P_{\mathrm{TOT}}$ (criterion (10), Fig 4); in Iteration 1 the second inequality holds in every parameter set, so the first, which involves exactly k_2/k_8 and $K_{\mathrm{TOT}}/P_{\mathrm{TOT}}$, decides (Fig 10B(a)). In Iteration 2 (Classes 3, 4, 5 and 9) the roles are reversed: the first inequality always holds, the second decides, and the top-ranked features are k_4/k_6, k_3/k_7, $K_{\mathrm{TOT}}/P_{\mathrm{TOT}}$ and K_{TOT}/k_6 (Fig 10B(b)). Whether the main branch also has a fold, the second distinction in both iterations, has no closed-form condition; the next-ranked features (k_3/k_7, k_3/k_6, k_2/k_7, k_2/k_6) carry this information. In Iterations 3 and 4 the top-ranked feature is k_2/k_4, carrying all of the importance in Iteration 3:

under the constraints of these iterations a branch point exists if and only if $k_4 > k_2$ (Fig 10B(c,d)), so the classifier has found the single algebraic condition that separates Class 6 from Class 10. In Iteration 4 the next most important features are k_1/k_5 and k_{11}/k_{55}, followed by k_{11}/k_2 and k_2/k_{55}; together these are the quantities that make up $K_{M1} = (k_{11} + k_2)/k_1$ and $K_{M5} = (k_{55} + k_2)/k_5$, whose order decides between Classes 7 and 8 (Eq. (14)). Fig 10B shows these exact separations iteration by iteration, as histograms of the single quantity that decides; Fig 7A shows the same result for all ten classes at once, as regions of the (x, y) plane. The two figures carry the same information and are shown deliberately in both forms: Fig 7 emphasises where each class lies relative to the others, Fig 10B that the condition ranked highest by the classifier splits the sets of an iteration exactly at zero, with no set on the wrong side.

Per-class contributions show which condition each class is learnt from. Split gain is summed over all trees of a multiclass model, so it ranks features for the classifier as a whole and cannot say which class a feature informs. We therefore computed, within each iteration, the TreeSHAP contribution [50] of every feature to the score of every class on the held-out test rows (Methods), and ranked the features of each class by their mean absolute contribution (Table 5; Supplementary Note 17). The per-class rankings follow the exact conditions class by class. In Iteration 1, k_2/k_8 leads for all four classes, with opposite signs: a larger k_2/k_8 (and $K_{\mathrm{TOT}}/P_{\mathrm{TOT}}$) raises the score of Classes 1 and 10 and lowers that of Classes 2 and 5, as the site-1 condition $k_8 P_{\mathrm{TOT}} > k_2 K_{\mathrm{TOT}}$ requires. In Iteration 2 the site-2 features lead, k_4/k_6 for Classes 3 and 5 and K_{TOT}/k_6 for Class 9, again with the sign of $k_4 K_{\mathrm{TOT}} > k_6 P_{\mathrm{TOT}}$. In Iterations 3 and 4, k_2/k_4 carries 92–97% of the contribution to Classes 6 and 10, and Classes 7 and 8 are separated by k_1/k_5 and k_{11}/k_{55} with opposite signs, the two ratios that set $K_{M5} < K_{M1}$ (Eq. (14)). The features that do not belong to any closed-form condition, k_3/k_7, k_3/k_6 and k_2/k_7, appear exactly for the classes that differ from a neighbour only by the fold on the main branch: Class 2 (against 5), Class 4 (against 5, where they lead) and Classes 1 and 10 (against each other). They are therefore candidates for the open boundary, which we test directly below.

Decisive tests: two log-ratios suffice where the classes differ in one analytical condition. Three pairs of classes occur in the same iteration and differ in a single condition. *Classes 1 and 2* (Iteration 1) both have S-shaped hysteresis on the main branch and differ only in whether a fold on a detached branch is also present. LightGBM on all features separates them at $97.37 \pm 0.82\%$ (baseline 60.33%), with the importance concentrated in k_2/k_8 (normalised gain 0.790) and $K_{\mathrm{TOT}}/P_{\mathrm{TOT}}$ (0.140), exactly the quantities of the site-1 condition. The two mechanistic log-ratios $\log_{10}(k_2 K_{\mathrm{TOT}}/k_8 P_{\mathrm{TOT}})$ and $\log_{10}(k_4 K_{\mathrm{TOT}}/k_6 P_{\mathrm{TOT}})$ alone give $98.02 \pm 0.82\%$ with boosted trees and $97.77 \pm 0.26\%$ with a linear boundary, which lies at $\log_{10}(k_2 K_{\mathrm{TOT}}/k_8 P_{\mathrm{TOT}}) \approx 0$ (Supplementary Note 13, Supplementary Table 13 and Supplementary Fig. 16). *Classes 3 and 4* (Iteration 2), the same distinction for Z-shaped hysteresis, are separated at $95.88 \pm 0.12\%$ (baseline 79.87%), with importance concentrated in k_4/k_6 (0.595), K_{TOT}/k_6 (0.102) and P_{TOT}/k_4 (0.084); the two log-ratios give 95.18% (boosted trees) and 95.11% (linear), with the boundary at $\log_{10}(k_4 K_{\mathrm{TOT}}/k_6 P_{\mathrm{TOT}}) \approx 0$. In both cases the two mechanistic features are as good as all 105, and the residual error is close to the limit set by the window: 47 of the 3,365 sets of Classes 1 and 2 and 71 of the 2,384 sets of Classes 3 and 4 satisfy criterion (10) but have their detached fold above $A_{\mathrm{TOT}} = 10^6$ (Fig 4C), so that the exact criterion itself would classify 98.6% and 97.0% of the sets correctly. *Classes 7 and 8* (Iteration 4) are separated at $98.37 \pm 0.40\%$ (baseline 51.20%), with the importance concentrated in k_1/k_5 (0.616) and k_{11}/k_{55} (0.239), followed by k_{11}/k_2 (0.058) and k_2/k_{55} (0.050). The two first-binding log-ratios alone separate the classes with a linear boundary at $95.20 \pm 0.44\%$ (boosted trees on the same two features: $93.43 \pm 0.15\%$; Supplementary Fig. 16C). The exact boundary is the surface $K_{M1} = K_{M5}$ (Eq. (14)); it also involves the shared catalytic constant k_2, which is why a straight line in the two log-ratios separates the classes well but not perfectly, and why the next-ranked features are the

Table 5. Per-class contributions recover the analytical conditions. For every class in each sampling iteration, the three features (set A) with the largest mean absolute TreeSHAP contribution to that class's score (LightGBM, held-out test rows, mean over three stratified splits; the percentage is the share of the class total). Arrow: a larger value of the feature raises (↑) or lowers (↓) the class score (sign of the correlation between the feature and its contribution). Right: the exact condition that the leading features encode. Full rankings, contributions grouped by parameter and dependence plots: Supplementary Note 17, Supplementary Table 18 and Supplementary Figs 17 and 18. Code: `ML_v3_SHAP_per_class.m` (folder A).

Iter.	Class	Largest per-class contributions	Exact condition
1	1	$k_2/k_8\uparrow$ (37%), $K_{\mathrm{TOT}}/P_{\mathrm{TOT}}\uparrow$ (7%), $k_3/k_7\downarrow$ (6%)	no detached branch ($k_8P_{\mathrm{TOT}} < k_2K_{\mathrm{TOT}}$); S-shaped fold
	2	$k_2/k_8\downarrow$ (26%), $k_3/k_6\uparrow$ (8%), $k_3/k_7\uparrow$ (5%)	detached branch ($k_8P_{\mathrm{TOT}} > k_2K_{\mathrm{TOT}}$); S-shaped fold
	5	$k_2/k_8\downarrow$ (38%), $K_{\mathrm{TOT}}/P_{\mathrm{TOT}}\downarrow$ (13%), $K_{\mathrm{TOT}}/k_8\downarrow$ (5%)	detached branch; no main-branch fold
	10	$k_2/k_8\uparrow$ (29%), $K_{\mathrm{TOT}}/P_{\mathrm{TOT}}\uparrow$ (11%), $k_3/k_7\uparrow$ (6%)	no detached branch; no fold
2	3	$k_4/k_6\downarrow$ (24%), $K_{\mathrm{TOT}}/k_6\downarrow$ (10%), $K_{\mathrm{TOT}}/P_{\mathrm{TOT}}\downarrow$ (9%)	no detached branch ($k_4K_{\mathrm{TOT}} < k_6P_{\mathrm{TOT}}$); Z-shaped fold
	4	$k_3/k_7\uparrow$ (14%), $k_2/k_7\uparrow$ (10%), $K_{\mathrm{TOT}}/P_{\mathrm{TOT}}\downarrow$ (5%)	detached branch ($k_4K_{\mathrm{TOT}} > k_6P_{\mathrm{TOT}}$); Z-shaped fold
	5	$k_4/k_6\uparrow$ (9%), $K_{\mathrm{TOT}}/P_{\mathrm{TOT}}\uparrow$ (9%), $k_3/k_7\uparrow$ (7%)	detached branch; no main-branch fold
	9	$K_{\mathrm{TOT}}/k_6\downarrow$ (13%), $K_{\mathrm{TOT}}/P_{\mathrm{TOT}}\downarrow$ (11%), $k_4/k_6\downarrow$ (10%)	no detached branch; biphasic ($x < 0$, $u < r$)
3	6	$k_2/k_4\downarrow$ (92%), $k_2\downarrow$ (2%), $k_4/k_{55}\uparrow$ (1%)	branch point $\iff k_4 > k_2$
	10	$k_2/k_4\uparrow$ (92%), $k_2\uparrow$ (2%), $k_4/k_{55}\downarrow$ (1%)	no branch point ($k_4 < k_2$)
4	7	$k_2/k_4\downarrow$ (30%), $k_1/k_5\downarrow$ (25%), $k_{11}/k_{55}\uparrow$ (12%)	branch point; $K_{M5} < K_{M1}$
	8	$k_2/k_4\downarrow$ (30%), $k_1/k_5\uparrow$ (24%), $k_{11}/k_{55}\downarrow$ (11%)	branch point; $K_{M5} > K_{M1}$
	10	$k_2/k_4\uparrow$ (97%), $k_4/k_6\downarrow$ (2%), $k_2\uparrow$ (0%)	no branch point ($k_4 < k_2$)

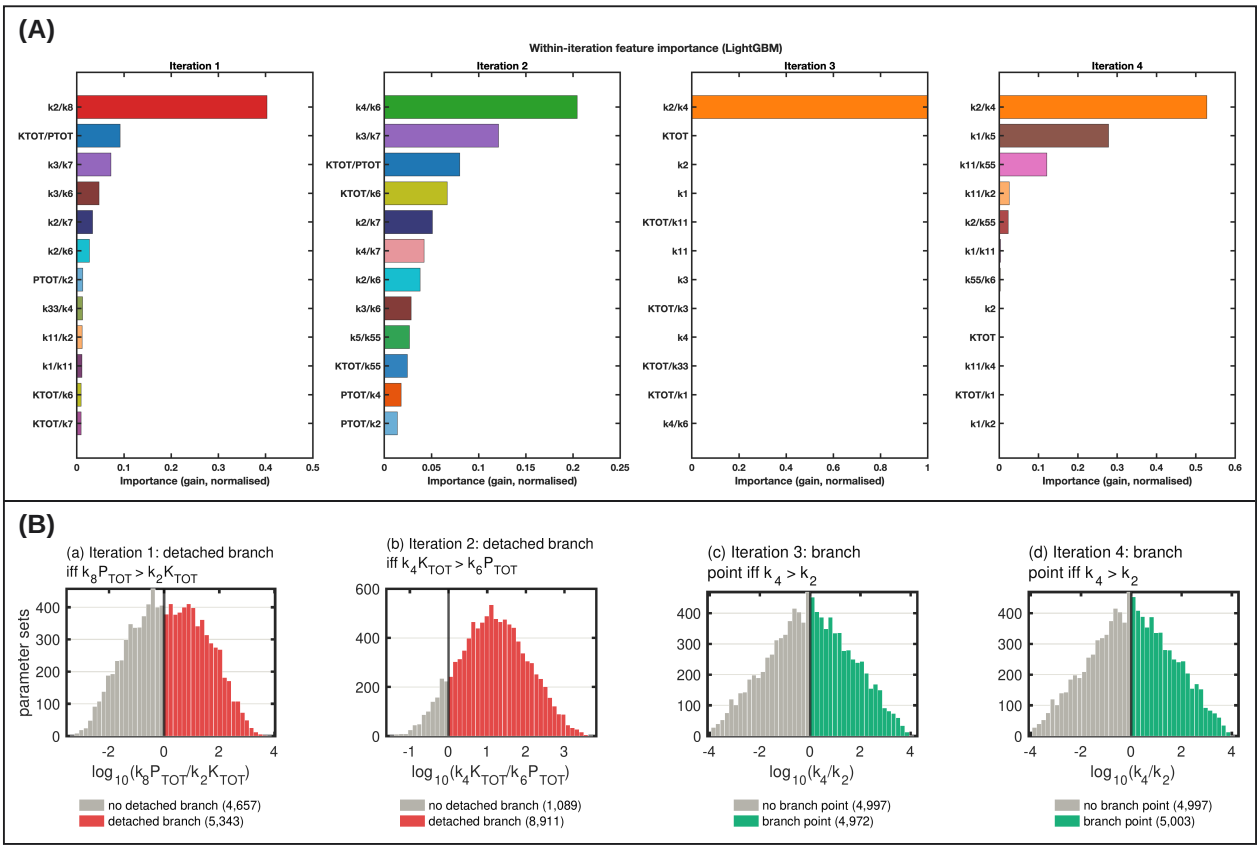


Fig 10. The features ranked highest by the classifier encode exact conditions for detached branches and branch points. (A) Feature importance within each sampling iteration: Normalised gain of the twelve most important LightGBM features (set A) in each iteration. Within an iteration the design-constrained ratios are constant and are excluded automatically, together with duplicated columns. Iteration 1: k_2/k_8 and $K_{\mathrm{TOT}}/P_{\mathrm{TOT}}$ (site-1 detached-branch condition); Iteration 2: k_4/k_6 (site-2 condition); Iteration 3: k_2/k_4 alone (branch-point condition $k_4 > k_2$); Iteration 4: k_2/k_4, then the first-binding asymmetries k_1/k_5 and k_{11}/k_{55}. Each feature has one fixed colour, the same in every figure (pooled importances of all models: Supplementary Fig. 10). (B) Histograms of the parameter combination named on each horizontal axis, coloured by the outcome of the exact steady-state computation (over all A_{TOT}). (a) Iteration 1: sets with a detached branch (red) are exactly those with $k_8 P_{\mathrm{TOT}} > k_2 K_{\mathrm{TOT}}$. (b) Iteration 2: sets with a detached branch are exactly those with $k_4 K_{\mathrm{TOT}} > k_6 P_{\mathrm{TOT}}$. (c,d) Iterations 3 and 4: sets with a branch point (green) are exactly those with $k_4 > k_2$. In every panel the separation at zero is exact: no parameter set lies on the wrong side. Panels (a), (b) and (d) show all 10,000 sets of each iteration; panel (c) shows the 9,969 sets of Iteration 3 used for machine learning (the 31 flagged sets removed), so its branch-point count (4,972) is 31 lower than the 5,003 sets with $k_4 > k_2$ quoted in the text. Panels (a) and (b) are the detached-branch quadrant of Fig 7A seen along one axis at a time, and (c) and (d) its anti-diagonal. Code: (A) `ML_v3_10class_parallel.m` (folder A); (B) `make_comparison_figures.m` (folder B).

ratios of k_2 to k_{11} and k_{55}. Class 7 occupies the region where $K_{M1} > K_{M5}$, Class 8 the opposite region.

Classification errors are concentrated where two routes to bistability coexist. The pooled confusion matrix (LightGBM panel of Supplementary Fig. 12) shows that the errors are not spread uniformly. Class 6 (BP only) is recovered without error, Classes 7 and 8 at 98.6% and 98.2%, Class 10 at 98.0%, Classes 1 and 5 at 91.6% and 91.4% and Class 9 at 88.1%. Most errors involve the detached-branch classes: 41.2% of the Class-4 and 34.1% of the Class-2 test sets are predicted as Class 5, and 4.1% and 3.2% of the Class-5 sets as Class 4 and Class 2. Classes 2, 4 and 5 all have a fold on a detached branch, which the classifier recognises reliably through criterion (10) (Fig 4); what it misses is whether the main branch carries an additional S- or Z-shaped fold, a distinction for which no closed-form condition exists. The orientation of main-branch hysteresis, by contrast, is almost never confused (Class 1 against 2: 7 and 11 sets; Class 3 against 4: 6 and 4 sets). The remaining errors are main-branch folds near the onset of hysteresis that are predicted as monotone (Class 1 to 10, 5.9%) or biphasic (Class 3 to 9, 12.5%). A second, smaller group of errors comes from the labelling window rather than from the classifier. Of the 591 test errors, 42 (7.1%, whereas such sets are 1.9% of the ensemble) are open circles of Fig 7A: sets in the detached-branch quadrant whose detached fold lies just above the window ($A_{\mathrm{TOT}} = 10^{6.0}$ to $10^{7.2}$). The classifier recognises the detached branch from the signs of x and y and predicts the class that the set would have with a wider window, whereas the label counts only the main branch: all 18 Class-9 sets predicted as Class 5, 7 of the 9 Class-10 sets predicted as Class 5, all 7 Class-1 sets predicted as Class 2, 5 of the 6 Class-3 sets predicted as Class 4, and 4 further sets (Supplementary Note 19, Supplementary Fig. 23 and Supplementary Table 21). These errors would disappear if the window were widened; the errors of Supplementary Fig. 26 would not. Apart from the main-branch onset errors and these window errors, all confused classes are bistable; the dominant error therefore confuses the route to bistability rather than its presence. The within-iteration confusion matrices (Supplementary Fig. 11) and those of all models (Supplementary Fig. 12) show the same pattern.

Misclassified parameter sets lie close to a class boundary. To see what the errors are, we refitted the pooled LightGBM model on the same training rows (the refit reproduces the test accuracy of 92.61%, i.e. the same 591 misclassified sets of the 7,994 test sets) and plotted the exact bifurcation diagrams of representative misclassified sets for the six most frequent confusions, which together account for 463 of the 591 errors (78%) (Supplementary Fig. 26). No error is a wrong label: every misclassified set has exactly the singularities of its true class, and the errors fall into two groups. The four largest confusions (Supplementary Fig. 26A–D; all confusions in Supplementary Note 14 and Supplementary Table 14) concern whether the main branch folds in addition to a detached branch. In panels A and B the main branch carries a clear Z- or S-shaped hysteresis loop (widths of 30–120% of A_{TOT} in most examples, but as narrow as 0.3% in one S-shaped set) next to the detached fold, yet the sets are predicted as Class 5. Conversely, the Class-5 sets predicted as Class 4 (panel C) have a main branch that does not fold but drops very steeply, close to the onset of a Z-shaped loop, and the Class-5 sets predicted as Class 2 (panel D) have a monotone main branch with a detached fold at high substrate levels ($A_{\mathrm{TOT}} \approx 10^4$–$10^5$). The second group (panels E and F) lies at the onset of S-shaped hysteresis in Iteration 1: monotone sets predicted as Class 1 have a threshold-linear (zero-order) response, and S-shaped sets predicted as monotone have a narrow loop (widths of 5–30%) whose lower fold lies at very small x_2. In both groups the misclassified sets lie near a class boundary, where a small change of the parameters creates or removes a fold.

Moving between neighbouring classes

The exact conditions and the per-class contributions say which quantities separate neighbouring classes; they do not say whether changing one of them moves a given parameter set across the boundary. We tested this directly in 1,560 parameter sets, 120 of every class in each iteration, by changing every parameter, one at a time, by a factor of $10^{\pm 0.5}$ and $10^{\pm 1}$ and labelling every result exactly (92,160 labels), and by sweeping single quantities through each boundary (Methods; Supplementary Note 18). The ten classes form a transition map whose edges fall into three kinds (Supplementary Fig. 19 and Supplementary Table 19).

Exact conditions behave as exact conditions. Setting $x = -0.3$ through k_8 gave a detached branch in 97.5% of Class-1 sets, and setting $y = +0.3$ in Classes 3 and 9 gave one whenever its fold fell inside the window: of the 720 sets moved across the $x = 0$ or $y = 0$ boundary, 77.5% had a detached fold below $A_{\mathrm{TOT}} = 10^6$ and all 162 others had it above. Setting $k_4 = k_8 = 2k_2$ turned every symmetric Class-10 set into Class 6, and setting $\log_{10}(K_{M1}/K_{M5}) = \pm 0.3$ through k_5 turned every Class-6 set into Class 7 or 8 (100%). Any single change that breaks $u^* = r$ removed the branch point (14,400 of 14,400 moves), and the (x, y) quadrant then predicted whether the new set has a detached branch in 99.2% of cases; all 110 exceptions had their detached fold above the window. A change of k_3 or k_7, which enters neither x nor y, turned 98% of the branch-point sets into Class 5: the upper half of the pitchfork becomes the detached branch (Supplementary Fig. 21). Changes of k_1, k_{11}, k_5 or k_{55} preserve the symmetry and kept the branch point in all 5,760 moves.

The fold on the main branch is controlled by k_3/k_7. A quantity that sets the detached-branch condition also reshapes the main branch, so a single parameter that crosses $x = 0$ often removes the main-branch fold at the same time: $x = -0.3$ through k_8 turned 54% of Class-1 sets into Class 5 rather than Class 2. When x was swept back and forth through zero, only the two classes of the edge occurred along the sweep in 21% of Class-1 sets and 60% of Class-10 sets, and y in 51% (Class 3) and 61% (Class 9); moving x through k_2, K_{TOT} or P_{TOT}, or y through k_6, K_{TOT} or P_{TOT}, never gave a clean two-class switch (Supplementary Table 20). The ratio k_3/k_7, which the per-class contributions single out for the fold classes (Table 5), behaves differently. Because k_3 and k_7 enter neither x nor y, sweeping k_3 leaves the detached-branch condition fixed and switches the main-branch fold alone: the sweep was two-class in 84% of Class-1 sets (1 and 10), 68% of Class-2 sets (2 and 5), 74% of Class-3 sets (3 and 9) and 71% of Class-4 sets (4 and 5) (Supplementary Fig. 20). The best single moves remove the fold readily ($K_{\mathrm{TOT}} \div 10$ turns 100% of Class-2 sets into Class 5 and $K_{\mathrm{TOT}} \times 10$ 99% of Class-4 sets into Class 5), whereas the reverse moves, which must create a fold, succeed in only 27–36% of sets. The fold is therefore the one property that no single quantity controls, consistent with its lack of a closed-form condition; k_3/k_7 is the best single handle on it and the natural starting point for deriving that condition from the inverse parametrisation.

The class of a construct can be assigned in milliseconds

The exact labelling, including the normal-form coefficients, takes 11 ms per parameter set on average (9–14 ms in the four iterations), i.e. 0.13 core-hours for all 40,000 sets; MATCONT seeded on every branch takes 8.0 s per set, about 700 times the cost per set of the exact labelling (Methods). The trained classifier assigns a class, with a class probability, without computing any steady state: LightGBM takes 0.11 ms per set when many sets are classified together and 0.45 ms for a single set, on one core and including the computation of the 105 features (Methods). Exact class assignment is therefore fast enough to be applied to every candidate design, to every construct as built, and to a local ensemble of parameter sets sampled within the measurement uncertainty of a construct.

Discussion

Bifurcation analysis is one of the sharpest tools of systems biology, but it answers one question at a time: what does this network do with these parameters? We have asked a different question: what does the network do with *any* parameters? For the dual phosphorylation–dephosphorylation cycle the answer is a map in which every point of the sampled 14-dimensional kinetic parameter space carries its complete bifurcation structure in $A_{\mathrm{TOT}} \leq 10^6$. Three features of this map stand out. Within the sampled ranges bistability is common, and wherever criterion (10) holds it is the rule, but most of it is hidden from the standard numerical approach. Bifurcations at branch points, in contrast, are fragile and require kinetic symmetry. And the boundaries between classes are not diffuse regions but sharp surfaces defined by a handful of parameter ratios, which a classifier trained within the sampled design rediscovers on its own.

The six problems posed in the Introduction are answered as follows. (P1) The bifurcation structure is organised as a map of ten classes over the sampled parameter space (Fig 1, Fig 2; Tables 1 and 2), which reduces to a ready reckoner on two catalytic-dominance ratios (Fig 7). (P2) Single-start continuation misses most of the bistability: it finds 19–27% bistable sets in the asymmetric iterations, against 72–88% with exact labels, because a detached branch cannot be reached from low A_{TOT} (Fig 3A, Fig 5; Supplementary Table 24); within the sampled design bistability is typical (65.0% of all sets; Supplementary Note 2 and Supplementary Table 2). (P3) Every set is labelled completely in the window $A_{\mathrm{TOT}} \leq 10^6$ by the inverse method (Fig 3B), and the accuracy of the labels is measured independently by MATCONT seeded on every branch (87.3% agreement on 4,000 sets, no spurious singularity; Supplementary Note 3, Supplementary Table 3, Supplementary Figs 1 and 2) and by the normal-form coefficients of all 40,018 singularities in the window (Supplementary Note 4, Supplementary Table 25, Supplementary Fig. 24). (P4) Symmetry decides the type of bifurcation: a branch point requires $u^* = r$ and, under the symmetry constraints, exists if and only if $k_4 > k_2$; full symmetry makes it a pitchfork, and breaking the first binding step makes it a transcritical crossing whose orientation is set by the sign of $K_{M5} - K_{M1}$ (Fig 6, criterion (14)); without symmetry, a detached branch exists if and only if criterion (10) holds (Fig 4). Every class boundary except the fold on the main branch is a closed-form condition. (P5) The class is predicted from the kinetic parameters with 92.7% accuracy (Table 4, Fig 9), through the feature pipeline of Fig 8 and Table 3; within single iterations, where the sampling design carries no information, accuracy remains at 84.5–99.98% and the most informative features are the ratios of the analytical conditions (Fig 10, Table 5), so the classifier learns the mechanism and not only the design. (P6) Single changes of a catalytic ratio move parameter sets across the closed-form boundaries as those conditions predict, and every single change that breaks $u^* = r$ removes the branch point (14,400 of 14,400 moves); the fold on the main branch, in contrast, responds to several parameters together, with k_3/k_7 as the best single handle (Supplementary Note 18, Supplementary Figs 19–21, Supplementary Tables 19 and 20).

Why has bistability in this motif appeared rare? Because what continuation sees depends on where it starts. Most bistable parameter sets carry a second branch of steady states that is detached from the branch at low substrate levels: it appears at a fold and persists for all larger A_{TOT}, but it is separated from the main branch by a band in which no positive steady state exists. A continuation started at low A_{TOT}, as in our earlier EMBED framework [38] and in most ensemble studies, cannot cross that band, whatever its settings. In the asymmetric iterations it found bistability in 19–27% of the parameter sets; exact labels give 72–88%. The detached branch exists exactly when the phosphatase dominates the first site and the kinase the second (criterion (10), Fig 4), a condition with a transparent biochemical meaning: the first phosphate is removed faster than it is added, so the substrate can rest unphosphorylated, but once it is on, the second is added faster than it is

removed, so the substrate can also rest doubly phosphorylated. Where single-start continuation did reach a singularity it located it to within 10^{-5}, and seeded on every branch MATCONT agrees with the exact labels for 87.3% of 4,000 sets and never reports a spurious singularity. The limitation was therefore never the continuation engine but the completeness of the steady-state set.

This result sharpens rather than contradicts the conclusion of EMBED that bifurcation-driven behaviour occupies narrow regions of parameter space [38]. It holds, exactly, for branch points. A branch point requires the second branch of steady states to meet the enzyme-balanced line $K = P$, which happens only when $u^* = r$, a condition satisfied on a set of measure zero in unconstrained parameter space; under the symmetry constraints it holds automatically, and a branch point then exists if and only if $k_4 > k_2$. Full symmetry makes it a pitchfork; breaking only the first binding step unfolds every pitchfork into a transcritical crossing with a fold on one arm, and the sign of $K_{M5} - K_{M1}$ selects the arm. EMBED observed this orientation rule numerically; criterion (14) states it exactly. The picture that emerges is simple: *bifurcations at branch points are fragile; bistability is not.*

The class boundaries are strikingly simple. Detached branches are bounded by the planes $k_8 P_{\mathrm{TOT}} = k_2 K_{\mathrm{TOT}}$ and $k_4 K_{\mathrm{TOT}} = k_6 P_{\mathrm{TOT}}$ in log-parameter space, branch points by the plane $k_4 = k_2$, and the two orientations of the transcritical crossing by the surface $K_{M1} = K_{M5}$. Each condition compares the two enzymes at a specific step, and each refines known conditions on catalytic constants for multistationarity in dual phosphorylation [28, 29]. Structural methods establish whether multistationarity is possible [25, 26, 27] and which bifurcations a topology admits [31, 32], and sampling with numerical equilibrium search approximates where multiple steady states occur [33]; the map adds, exactly and for every parameter set, through which bifurcation and on which branch. Because every boundary is a ratio of a few parameters, none is aligned with distance in the full parameter space, which is why a neighbourhood-preserving embedding blurs the classes that mechanistic coordinates separate cleanly (Fig 7A, Supplementary Fig. 15). One boundary remains open: whether the main branch itself folds, which separates S- or Z-shaped hysteresis from graded or biphasic responses.

Machine learning turns the map into a predictor and, at the same time, audits the analysis. Gradient-boosted trees predict the ten classes from the kinetic parameters alone with 92.7% accuracy (macro AUC 0.996; majority baseline 32.0%), and every tree ensemble performs within a few percentage points of the best. Once the information carried by the sampling design is removed, by training within each iteration, accuracy remains at 84.5–99.98%, and the most informative features are precisely the ratios of the analytical conditions, which the models were never given. Where two classes differ in a single condition, those two ratios alone separate them as well as all 105 features. The remaining errors are equally telling: they fall almost entirely at the one boundary without a closed form, where the classifier recognises a detached branch but misses an additional fold on the main branch. The classifier thus confuses the route to bistability, never its presence, and its errors point to exactly the boundary that theory has yet to resolve.

Machine learning has been used before to steer sampling towards multistationary regions [33] and to learn the map from equilibria to parameters [34]. Here its role is different: trained on exact and complete labels, it predicts the full bifurcation class rather than the presence of several steady states, and its feature ranking can be checked against closed-form boundaries. The step from steady states to parameters that such networks learn approximately is the one the inverse method takes exactly, which is why every branch is found and every label can be trusted.

These findings bear directly on synthetic biology, which we leave for future work. Circuits are engineered through Design–Build–Test–Learn (DBTL) cycles [39, 40], and phosphorylation-based switches can now be built *de novo* [8], alongside transcriptional switches such as the toggle switch [7]. Yet the standard Test, a dose–response measured with increasing substrate from an unphosphorylated

state, is the experimental twin of single-start continuation: it follows the main branch and will report a monostable response for a construct whose bistability lies on a detached branch. Simulation can close this gap if the as-built construct is placed on the class map before it is tested. The class is fixed by parameter ratios that are known only after Build, because expression levels and enzyme kinetics drift with genetic context: a twofold shortfall in phosphatase expression shifts $\log_{10}(k_8 P_{\mathrm{TOT}}/k_2 K_{\mathrm{TOT}})$ by 0.3, enough to erase a detached branch designed with a smaller margin. The transition map (Fig 7, Supplementary Note 18) states which such drifts change the class, and the class states which experiment can reveal the switch, such as a decreasing sweep or a measurement from a pre-phosphorylated state for a detached-branch switch [51, 52]. Simulation can serve this role only if it is complete, fast, of measured accuracy and expressed in measurable quantities; the exact method meets all four conditions (all branches, 11 ms per parameter set, validated against MATCONT, boundaries in catalytic rates and Michaelis constants; how each of these quantities can be tuned and measured is summarised in Supplementary Note 21). Using the class map in this way, in silico and experimentally, is the subject of a separate study.

The same logic applies to natural signalling. If bistability in dual phosphorylation is common but often detached, as the model predicts, a cell's state may depend on its history more often than dose–response experiments suggest, and increasing-dose protocols that start from the unphosphorylated state, and therefore follow the main branch, may systematically under-report switches in MAPK and related cascades [19]. Because the class boundaries are set by the balance of catalytic capacity between the two sites and by the relative binding affinities at the outer steps, they also indicate which perturbations, such as kinase or phosphatase inhibitors, expression changes or affinity mutants, are expected to create, remove or relocate a switch, a question of direct relevance to diseases driven by kinase–phosphatase imbalance [41, 42, 43, 44].

Several limitations define the next steps. The analysis concerns a single ordered, distributive mechanism in its full mass-action form, and the reduced model used in EMBED may place some boundaries differently. The iterations were designed to enrich specific regimes, so the class fractions hold for the sampled ranges; an unconstrained ensemble with experimentally informed priors [10, 11] would estimate how often each class occurs *in vivo*. Labels refer to the window $A_{\mathrm{TOT}} \leq 10^6$, although all singularities are stored for re-thresholding. The window matters near the edges of the detached-branch quadrant, where the detached fold moves to very large A_{TOT}: 1.9% of the sets satisfy criterion (10) but are labelled by their main branch (open circles in Fig 4B and Fig 7A), and they account for 7.1% of the classifier's errors. These errors reflect where the window is drawn, not a failure of the classifier, and a window matched to the experimentally accessible range of a given substrate would remove or shift them. The open boundary of main-branch hysteresis is a semi-algebraic condition on the rational function $A_{\mathrm{TOT}}(u)$ and should yield to discriminant or Sturm-sequence methods. Beyond these, the inverse method extends to multisite and mixed mechanisms with a common kinase and phosphatase [20, 22, 23, 24], and continuation or homotopy seeded on every branch extends the validation protocol to networks with feedback and oscillations. Coupling the labeller and the classifier to active learning would close such a design loop computationally, and placing binding and unbinding rates from molecular simulations on the map would link molecular structure directly to circuit behaviour.

For a new parameter set, the class can be obtained by continuation, by the exact inverse method or by the trained classifier, and the three serve different purposes. Seeded continuation (8.0 s per set) is the general-purpose reference but requires a starting equilibrium on every branch; single-start continuation, the common practice, misses detached branches and failed for 362 of the 40,000 sets. The exact method (11 ms per set) is the reference for this network but depends on a closed-form parametrisation that is not available for most networks. The classifier requires neither a steady state nor an initial condition, has no step sizes, tolerances or parameter window to tune, and assigns

a class in 0.1–0.5 ms per set, about 20,000 to 70,000 times faster than seeded continuation. It returns a class probability that identifies the sets worth confirming numerically, and its feature importances recover the analytical conditions (criterion (10), $k_4 > k_2$) that it was never given, whereas continuation returns a diagram for one set and says nothing about why. Machine-learning classification is therefore preferred for screening large design spaces, libraries of constructs and local ensembles sampled within the measurement uncertainty of a construct; inside optimisation, inverse-design or active-learning loops, where continuation at every step is impractical; and for networks with feedback, more phosphorylation sites or mixed mechanisms, where no exact parametrisation exists and seeded continuation becomes expensive and uncertain about completeness. Continuation or exact labelling remains the reference for confirming individual cases. For the network studied here the advantage of the classifier over the exact method is interpretability and transfer rather than speed or accuracy: it is only 25–100 times faster per set, it is correct for 92.7% of the sets, its errors concentrate in Classes 2, 4 and 5, where it detects the detached branch but misses an additional fold on the main branch, and its predictions hold only within the sampled ranges and designs; outside them they are extrapolations.

Ensemble bifurcation analysis can be exact, and when it is, it changes the picture it was meant to describe. In the model of the dual phosphorylation cycle, bistability is not a fine-tuned exception: it is common across the sampled ranges and the rule wherever criterion (10) holds, often on branches that single-start continuation cannot reach and that, in the model, an increasing-dose protocol from the unphosphorylated state would not reach either; symmetry-dependent bifurcations, by contrast, are genuinely fragile; and a few ratios of enzyme capacities and affinities decide which behaviour a circuit has. Completeness, not computational power, was the missing ingredient, and with it the map becomes a design tool: once the kinetic parameters of a construct have been estimated, it can be placed on the map and its class predicted before the switch is tested. The exact labels, singularities, normal-form coefficients and code are released as a benchmark for predicting dynamical behaviour from kinetic parameters.

Methods

Reaction network

We consider a mechanistic model of the dual phosphorylation and dephosphorylation of a substrate A through a sequence of reversible Michaelis–Menten steps mediated by a kinase K and a phosphatase P. The substrate exists in three phosphorylation states, unphosphorylated (A), singly phosphorylated (AP) and doubly phosphorylated (APP), and every elementary step proceeds through its own enzyme–substrate complex, so that the mechanism is ordered and distributive. Reactions R1–R4 describe the sequential conversion $A \to AP \to APP$ by the kinase, and R5–R8 the reverse conversion

$APP \to AP \to A$ by the phosphatase:

$$\begin{aligned}
&\text{R1: } A + K \underset{k_{11}}{\overset{k_1}{\rightleftharpoons}} AK, && v_1 = k_1[A][K] - k_{11}[AK],\\
&\text{R2: } AK \xrightarrow{k_2} AP + K, && v_2 = k_2[AK],\\
&\text{R3: } AP + K \underset{k_{33}}{\overset{k_3}{\rightleftharpoons}} APK, && v_3 = k_3[AP][K] - k_{33}[APK],\\
&\text{R4: } APK \xrightarrow{k_4} APP + K, && v_4 = k_4[APK],\\
&\text{R5: } APP + P \underset{k_{55}}{\overset{k_5}{\rightleftharpoons}} APPC, && v_5 = k_5[APP][P] - k_{55}[APPC],\\
&\text{R6: } APPC \xrightarrow{k_6} AP + P, && v_6 = k_6[APPC],\\
&\text{R7: } AP + P \underset{k_{77}}{\overset{k_7}{\rightleftharpoons}} APC, && v_7 = k_7[AP][P] - k_{77}[APC],\\
&\text{R8: } APC \xrightarrow{k_8} A + P, && v_8 = k_8[APC].
\end{aligned}$$

Comparison with the Markevich model. Both our model and that of Markevich, Hoek and Kholodenko [19] describe distributive dual phosphorylation–dephosphorylation, but they differ in complexity and symmetry. The Markevich model has ten reactions with multiple enzyme–substrate complexes, enzyme redistribution and intermediate recycling, particularly in dephosphorylation by MKP3, which account for enzyme sequestration and rebinding. Our model has eight reactions, each step proceeding through a single intermediate without feedback or recycling; it nevertheless retains enough nonlinearity and saturation to exhibit bistability and a rich repertoire of codimension-1 bifurcations. The Markevich model uses the total kinase ($\mathrm{MAPKK_{TOT}}$) as bifurcation parameter and generates bistability primarily through saddle-node bifurcations arising from enzyme sequestration; here A_{TOT} is the bifurcation parameter, which preserves the kinase–phosphatase symmetry of the network and permits symmetry breaking.

Dynamical equations and conservation laws

We write $x_1 = [AP]$, $x_2 = [APP]$, $x_3 = [AK]$, $x_4 = [APK]$, $x_5 = [APPC]$ and $x_6 = [APC]$. The free substrate, kinase and phosphatase follow from the three conservation laws

$$A = A_{\mathrm{TOT}} - \sum_{i=1}^{6} x_i, \qquad K = K_{\mathrm{TOT}} - x_3 - x_4, \qquad P = P_{\mathrm{TOT}} - x_5 - x_6, \tag{1}$$

and the dynamics are given by six ODEs,

$$\dot{x}_1 = v_2 - v_3 + v_6 - v_7, \quad \dot{x}_2 = v_4 - v_5, \quad \dot{x}_3 = v_1 - v_2, \quad \dot{x}_4 = v_3 - v_4, \quad \dot{x}_5 = v_5 - v_6, \quad \dot{x}_6 = v_7 - v_8, \tag{2}$$

implemented in the model file `MAPKfull.m`. Eqs (1) and (2) and the analytic Jacobian of Eq. (2) were derived symbolically in Mathematica [46] from the list of species, complexes and reactions (Supplementary File 6, `Analytical_Derivations_StepByStep_final.nb`, Sections 1 and 3). A_{TOT} is the bifurcation parameter; the other 14 quantities (K_{TOT}, P_{TOT} and twelve rate constants) are fixed for each parameter set. The output of the switch is $x_2 = [APP]$.

Structural network analysis

The structural properties of the network were analysed with the Chemical Reaction Network Toolbox, version 2.35 [27], which implements the deficiency theory of chemical reaction network theory together

with the advanced deficiency algorithms [25, 26, 53]. The network has nine species and ten complexes organised into two linkage classes. The stoichiometric matrix has rank $\rho = 6$, so there are three independent conservation laws,

$$P_{\mathrm{TOT}} = APC + APPC + P, \tag{3}$$

$$K_{\mathrm{TOT}} = AK + APK + K, \tag{4}$$

$$A_{\mathrm{TOT}} = A + AK + AP + APC + APK + APP + APPC, \tag{5}$$

identical to Eq. (1). The deficiency is

$$\delta = c - l - \rho = 10 - 2 - 6 = 2, \tag{6}$$

where c is the number of complexes and l the number of linkage classes. The conservation laws, Eqs (3)–(5), and the deficiency, Eq. (6), were verified independently in Mathematica: the three conserved totals annihilate every reaction vector, and $c - l - \rho = 2$ (Supplementary File 6, `Analytical_Derivations_StepByStep_final.nb`, Section 2). The network is not weakly reversible, so the Deficiency Zero and Deficiency One Theorems do not apply; it is conservative, and its reaction vectors are positively dependent, so positive steady states exist for some rate constants. Because the deficiency is two (each linkage class has deficiency zero), the network was analysed with higher deficiency theory. The Higher Deficiency Report shows that, with mass-action kinetics, the network has the capacity for multiple positive steady states, and for a positive steady state with a zero eigenvalue in the stoichiometric subspace (Supplementary File 1). The example it returns has $k_1 = 12.463$, $k_{11} = k_2 = 30.061$, $k_3 = 25.087$, $k_{33} = k_4 = 1.910$, $k_5 = 898.91$, $k_{55} = k_6 = 1$, $k_7 = 782.52$ and $k_{77} = k_8 = 12.438$, with two distinct positive steady states in the same stoichiometric compatibility class ($A_{\mathrm{TOT}} = 18.314$, $K_{\mathrm{TOT}} = 5.283$, $P_{\mathrm{TOT}} = 9.140$): an asymptotically stable state with $APP = 6.8 \times 10^{-3}$ and an unstable state with $APP = 1.007$. Applying the parametrisation of Eqs (7)–(9) to these rate constants recovers both states and a third, asymptotically stable state with $APP = 1.332$ at the same totals, so the example is bistable. This agrees with the bound of at most $2n - 1 = 3$ positive steady states for the ordered distributive cycle with a common kinase and phosphatase [23], which is attained [24]. The structural analysis establishes the capacity for multistationarity; which parameter regions realise it, and through which bifurcations along A_{TOT}, is determined by the exact steady-state analysis below (Supplementary Note 1 and Supplementary Table 1).

Codimension-1 bifurcations in A_{TOT}

All parameters are held fixed except A_{TOT}. Varying the total substrate with fixed kinase and phosphatase totals preserves the kinase–phosphatase symmetry of the network, and this permits saddle-node (fold, limit point, LP), transcritical and pitchfork bifurcations through variation of a single parameter. At a fold a stable and an unstable steady state are created or annihilated, producing an abrupt jump, bistability and hysteresis. At a transcritical bifurcation two branches cross and exchange stability. Pitchfork bifurcations arise in kinetically symmetric networks, where a symmetric steady state destabilises and gives rise to two asymmetric states [4]; they are structurally unstable, and breaking the symmetry unfolds them, generically into a transcritical crossing accompanied by a fold [47]. The steady states can also form a detached branch, disconnected from the branch at low A_{TOT}, which appears at a fold and persists for all larger A_{TOT}; and the output can be biphasic, first increasing and then decreasing with A_{TOT}.

Sampling of the ensemble

Kinetic parameters were sampled by Latin hypercube sampling [54, 55] with the MATLAB function `lhsdesign`, uniformly in logarithmic space: K_{TOT} and P_{TOT} over $[10, 500]$ nM and the twelve rate

constants over $[10^{-3}, 10^{1}]$ ($\mathrm{nM}^{-1}\,\mathrm{s}^{-1}$ for the binding constants k_1, k_3, k_5 and k_7, s^{-1} for the unbinding and catalytic constants), with the iteration-specific bounds and constraints of Table 1; concentrations, including A_{TOT}, are in nM throughout. The ranges span four orders of magnitude for every rate constant and a 50-fold range of enzyme totals. Because every class boundary is a condition on ratios of parameters (Eqs (10), (11), (14)), the ranges affect the results only through the distribution of these ratios: they set the class fractions of each iteration, but not the conditions that separate the classes. Each iteration contains 10,000 parameter sets. A fixed random seed (20260930) makes the ensembles reproducible. The pipeline does not depend on the choice of design; Sobol or Halton sequences can be used without change to the labelling step.

Exact steady-state structure: the inverse method

Parametrisation of all steady states. With $K_{M1} = (k_{11} + k_2)/k_1$, $K_{M3} = (k_{33} + k_4)/k_3$, $K_{M5} = (k_{55}+k_6)/k_5$ and $K_{M7} = (k_{77}+k_8)/k_7$, the four complex equations at steady state give $x_3 = AK/K_{M1}$, $x_4 = x_1K/K_{M3}$, $x_5 = x_2P/K_{M5}$ and $x_6 = x_1P/K_{M7}$. These are exact consequences of Eq. (2); no quasi-steady-state approximation is made. Flux balance at the two sites ($v_4 = v_6$, $v_2 = v_8$) then gives

$$x_1 = a_1 u A, \quad x_2 = a_1 a_2 u^2 A, \quad u \equiv \frac{K}{P}, \quad a_1 = \frac{k_2 K_{M7}}{k_8 K_{M1}}, \quad a_2 = \frac{k_4 K_{M5}}{k_6 K_{M3}}, \tag{7}$$

so the free enzymes enter only through their ratio u. Dividing the kinase by the phosphatase conservation law eliminates K and P and leaves an equation that is linear in A for fixed u:

$$A(u) = \frac{u - r}{u\,(\alpha u + \beta)}, \quad \alpha = \frac{a_1}{K_{M3}}\left(\frac{r k_4}{k_6} - 1\right), \quad \beta = \frac{1}{K_{M1}}\left(\frac{r k_2}{k_8} - 1\right), \quad r = \frac{K_{\mathrm{TOT}}}{P_{\mathrm{TOT}}}. \tag{8}$$

For any $u > 0$ with $A(u) > 0$, the free kinase is $K = K_{\mathrm{TOT}}/(1 + A/K_{M1} + a_1 u A/K_{M3})$, $P = K/u$, and $x_1, \ldots, x_6$ follow from the relations above. The substrate conservation law, used in reverse, gives the total substrate to which this steady state belongs:

$$A_{\mathrm{TOT}}(u) = A + x_1 + x_2 + x_3 + x_4 + x_5 + x_6. \tag{9}$$

The map $u \mapsto (A_{\mathrm{TOT}}(u), x_1(u), \ldots, x_6(u))$ on $U = \{u > 0 : A(u) > 0\}$ is a one-to-one correspondence with the set of all positive steady states over all $A_{\mathrm{TOT}} > 0$. The number of steady states at $A_{\mathrm{TOT}} = a$ is therefore the number of solutions of $A_{\mathrm{TOT}}(u) = a$, at most three, the bound $2n - 1$ for $n = 2$ sites [23, 24]. Parametrising steady states by the free-enzyme ratio follows Wang and Sontag [23]; rational parametrisations of this kind are a standard tool for multisite phosphorylation systems [22, 28]. We call this procedure the inverse method: instead of fixing A_{TOT} and searching for steady states, one chooses a steady state, through u, and computes the A_{TOT} to which it belongs. Eqs (7)–(9) were derived in Mathematica [46], and the two site-flux balances and the total-enzyme balance were checked symbolically to vanish identically on the parametrisation (Supplementary File 6, `Analytical_Derivations_StepByStep_final.nb`, Sections 5 and 6).

Branches, folds and branch points. $A(u)$ can change sign only at $u = r$ and at $u^* = -\beta/\alpha$, so U consists of at most two intervals, each a branch of steady states. The main branch has $u = r$ as an endpoint, where $A_{\mathrm{TOT}} \to 0$; it is the branch on which a continuation started at low A_{TOT} lies. A second, detached branch exists if and only if

$$k_8 P_{\mathrm{TOT}} > k_2 K_{\mathrm{TOT}} \quad \text{and} \quad k_4 K_{\mathrm{TOT}} > k_6 P_{\mathrm{TOT}} \qquad (u^* \neq r), \tag{10}$$

that is, when the phosphatase has the larger maximal rate at the first site and the kinase the larger maximal rate at the second (Fig 4). Criterion (10), the endpoint u^* of Eq. (11) and the branch-point

condition $u^* = r$ were derived in Mathematica from the factors of $A(u)$ (Supplementary File 6, `Analytical_Derivations_StepByStep_final.nb`, Section 7), and both inequalities of criterion (10) are evaluated there for one ensemble parameter set of every class (`Excel_Parameter_Examples_Condition_Checks_final.nb`, Section 3). $A_{\mathrm{TOT}} \to \infty$ at both ends of a detached branch, so the branch appears at a fold at its minimum A_{TOT} and exists for all larger A_{TOT}. It is separated from the main branch by values of u at which no positive steady state exists, and therefore cannot be reached by continuation from the main branch. When $u^* = r$, which holds under the symmetry constraints of Iterations 3 and 4, the two intervals touch and the branches cross at a branch point on the line $K = P$ (Eqs (11)–(14), Results). Folds are turning points of $A_{\mathrm{TOT}}(u)$, $\mathrm{d}A_{\mathrm{TOT}}/\mathrm{d}u = 0$, located on the u grid and refined; the stability of every steady state is given by the eigenvalues of the analytic 6×6 Jacobian of Eq. (2).

Slope of the crossing branch, Eq. (13). Under the constraints of Iterations 3 and 4 ($K_{\mathrm{TOT}} = P_{\mathrm{TOT}}$, $k_6 = k_2$, $k_8 = k_4$, $K_{M7} = K_{M3}$), Eqs (7) and (8) give $a_1 = k_2 K_{M3}/(k_4 K_{M1})$, $a_1 a_2 = K_{M5}/K_{M1}$ and $\alpha = (k_4 - k_2)/(k_4 K_{M1})$. On the crossing branch $A = 1/(\alpha u)$ the steady-state concentrations follow from the relations above Eq. (7):

$$x_1 = \frac{a_1}{\alpha}, \quad x_2 = \frac{a_1 a_2 u}{\alpha}, \quad x_3 = \frac{K}{\alpha K_{M1} u}, \quad x_4 = \frac{a_1 K}{\alpha K_{M3}}, \quad x_5 = \frac{a_1 a_2 K}{\alpha K_{M5}}, \quad x_6 = \frac{a_1 K}{\alpha K_{M3} u},$$

with $K(u) = K_{\mathrm{TOT}} \big/ \big(1 + 1/(\alpha K_{M1} u) + a_1/(\alpha K_{M3})\big)$, so that Eq. (9) becomes

$$A_{\mathrm{TOT}}(u) = \frac{1}{\alpha}\Big(\frac{1}{u} + a_1 + a_1 a_2 u\Big) + \frac{K(u)}{\alpha}\Big(\frac{1}{K_{M1} u} + \frac{a_1}{K_{M3}} + \frac{a_1 a_2}{K_{M5}} + \frac{a_1}{K_{M3} u}\Big).$$

Differentiating, $s = u\, A'_{\mathrm{TOT}}(u)/A_{\mathrm{TOT}}(u)$ at $u = 1$, and substituting a_1, $a_1 a_2$ and α gives Eq. (13) after simplification (derived and simplified in Mathematica [46], Supplementary File 6, `Analytical_Derivations_StepByStep_final.nb`, Section 8, and checked against a finite-difference derivative of $A_{\mathrm{TOT}}(u)$ for random parameter sets). Every term of the denominator is positive when $k_4 > k_2$, i.e. whenever a branch point exists, so the sign of s is the sign of $K_{M5} - K_{M1}$. Under full symmetry ($k_5 = k_1$, $k_{55} = k_{11}$, Iteration 3) $K_{M5} = K_{M1}$ and $s = 0$; with first-binding asymmetry (Iteration 4) $s \neq 0$. Which arm of the crossing branch turns back at a fold follows from the sign of s (Results), which gives criterion (14).

Labelling pipeline

For every parameter set, `Ensemble_LHS_v3.m` computes all steady states for A_{TOT} in a fixed window $[0, W]$, identical for all samples ($W = 10^6$), with 10,000 grid points in u per branch accumulating at the branch endpoints $u = r$ and $u = u^*$. The parametrisation, the branch structure and the criteria (10)–(14) are exact; the labels are computed numerically from them: folds are located as sign changes of $\mathrm{d}A_{\mathrm{TOT}}/\mathrm{d}u$ on this grid and refined by one-dimensional minimisation, stability from the eigenvalues of the analytic Jacobian, and the class refers to the window. The resolution and the window are the two limits of the labels; both are quantified by the verification tests below, by the dependence on the window (Supplementary Table 22) and, for sets whose detached fold lies just beyond the window, in Supplementary Note 19. Each parameter set is assigned to one of the ten classes of Table 1 from the singularities inside the window. Main-branch hysteresis is S-shaped when $K_{\mathrm{TOT}} > P_{\mathrm{TOT}}$ and Z-shaped when $K_{\mathrm{TOT}} < P_{\mathrm{TOT}}$, which coincides with the kinase- and phosphatase-dominant Iterations 1 and 2. A residual class for any other combination of singularities is provided but contains no parameter set. A parameter set whose fold coincides with its branch point (relative difference in A_{TOT} below 10^{-9}) is flagged: in such a set the reported fold is the turning point of the crossing branch at the branch point and its normal-form coefficient vanishes.

This applies to 32 sets (31 in Iteration 3, labelled 7 or 8, and one in Iteration 4); they are reported in all tables and figures and excluded from machine learning. Each set also records the locations of all folds and branch points, whether a detached branch is present and whether its fold lies in the window, the maximum number of coexisting steady states, the A_{TOT} ranges of bistability and the normal-form coefficients of every singularity, so that users can re-threshold the window. As an independent check, MATCONT was run on a random subsample of 1,000 parameter sets per iteration (4,000 in total), seeded on every branch (including detached branches), continued in both directions and trimmed to the same window; agreement between the MATCONT and exact labels is reported with Wilson 95% confidence intervals [56] (Supplementary Note 3, Supplementary Table 3 and Supplementary Figs 1 and 2). The complete procedure is summarised in Algorithm 1.

Algorithm 1. Exact ensemble labelling with independent MATCONT validation.

Require: ODE model (Eq. (2)), parameter bounds and constraints of iteration i, window $[0, W]$, $W = 10^6$.

Ensure: ten-class labels, flags, singularity locations, normal-form coefficients, validation statistics.

1. For each iteration $i \in \{1, 2, 3, 4\}$: draw $N = 10{,}000$ parameter sets by log-uniform LHS (fixed seed) and impose the equality constraints of iteration i.
2. For each parameter set p_k (in parallel):
   (a) compute $r = K_{\mathrm{TOT}}/P_{\mathrm{TOT}}$, α, β and $u^* = -\beta/\alpha$ (Eq. (8));
   (b) on each interval of $\{u > 0 : A(u) > 0\}$, place a grid in u refined towards the endpoints r and u^*;
   (c) for every grid point compute A, K, P, $x_1, \ldots, x_6$ and $A_{\mathrm{TOT}}(u)$ (Eqs (7)–(9)); keep $A_{\mathrm{TOT}} \le W$;
   (d) locate folds as roots of $\mathrm{d}A_{\mathrm{TOT}}/\mathrm{d}u = 0$ on every branch; locate the branch point at $u = 1$, $A = 1/\alpha$ if $u^* = r$ and $\alpha > 0$;
   (e) determine the stability of every steady state from the eigenvalues of the analytic Jacobian;
   (f) assign one of the ten classes from the singularities inside the window; flag the set if a fold coincides with the branch point;
   (g) compute the normal-form coefficients of every fold and branch point (analytic derivatives);
   (h) record all fold and branch-point locations, the maximum number of coexisting steady states and the bistable ranges of A_{TOT};
   (i) if p_k is in the validation subsample (1,000 per iteration): seed MATCONT on a stable steady state of every branch, continue in both directions, trim to $[0, W]$ and assign a MATCONT label.
3. Aggregate the four iterations; compute agreement, sensitivity and specificity of the MATCONT labels with 95% Wilson confidence intervals.
4. Export labels, singularities and validation statistics (Excel, CSV, MAT).

Verification. An independent re-implementation of the parametrisation reproduces the exact class of 39,968 of the 39,969 parameter sets compared (`make_comparison_figures.m`, Supplementary File 4). The remaining set is the flagged Iteration-4 set whose fold lies within a relative distance of 10^{-7} in A_{TOT} of its branch point; there the orientation of the crossing is decided at numerical resolution, and criterion (14) (K_{M5} exceeds K_{M1} by 0.17%) agrees with the label of the main pipeline. (i) At 1,802 steady states from 1,500 random parameter sets and random $A_{\mathrm{TOT}} \in [0.1, 10^3]$ (log-uniform), the scaled residual of Eq. (2) has a median of 3.6×10^{-14} (maximum 1.2×10^{-7}). (ii) Criterion (10) agrees with the full computation for all 20,000 parameter sets of Iterations 1 and 2 (14,254 with a detached branch, 5,746 without; Fig 4A). (iii) For 600 random parameter sets with a detached branch, evaluated at twice the A_{TOT} of the fold, all have exactly three steady states, two of them stable. (iv) All branch points and limit points reported by single-start MATCONT are reproduced, to median relative differences of 1.2×10^{-9} ($n = 8{,}890$ branch points) and 8.2×10^{-8}

($n = 4{,}115$ folds of coexisting LP and BP). (v) Seeded on all branches, MATCONT gives the same class as the exact labels for 87.3% of the 4,000 checked parameter sets of the ten-class ensemble (3,493 sets; 95% Wilson CI 86.3–88.3%; 82.0, 81.3, 99.4 and 86.6% in Iterations 1–4). It detects a fold or branch point in 93.3–99.8% of the multistationary sets of each iteration and never reports one in a set without (specificity 100% in every iteration). The disagreements are singularities that MATCONT misses or does not classify: in Iteration 4, the fold that lies close to the branch point when the first-binding asymmetry is small (133 sets labelled BP only by MATCONT); in Iterations 1 and 2, main-branch hysteresis that MATCONT assigns to its residual class or to "no singularity", and detached folds that it does not reach; in Iteration 3, three flagged LP+BP sets and three BP-only sets. (vi) At the 2,422 folds of the validation subsample at which MATCONT reports its fold coefficient, our coefficient a_{LP} agrees with MATCONT's to a median relative difference of 1.4×10^{-8} (95th percentile 1.7×10^{-7}).

Single-start MATCONT labels

The labels against which the exact method is compared (Fig 5; Supplementary Note 5) were obtained with the workflow of EMBED [38], applied here to the full mass-action model. For each parameter set, Eq. (2) was integrated with `ode15s` [57] (RelTol = 10^{-8}, default AbsTol) from the state with all substrate unphosphorylated and all enzyme free ($x = 0$) over $t \in [0, 5000]$ at $A_{\mathrm{TOT}} = 0.1$, and the final state was used, without a convergence test, to initialise equilibrium continuation in A_{TOT} with MATCONT [35, 36] (`init_EP_EP`, `cont`) with the settings of Supplementary Table 28. Continuation ran forward only and stopped after 1,500 points, so the range of A_{TOT} covered differed between parameter sets (median upper end 1,244; 5–95% range 257–1,745). At every detected branch point the branch was switched (`init_BP_EP`, amplitude 10^{-8}) and continued in both directions for up to 1,500 points, and limit points on the new branch were recorded. A parameter set without singularities was labelled biphasic when `findpeaks`, with default arguments, found exactly one maximum of x_2 along the computed branch. Two properties of this procedure limit what it can detect (Supplementary Fig. 4). First, the fixed number of continuation points ends each run at a different A_{TOT}, mostly below 2,000, whereas the branch points of Iterations 3 and 4 that it missed lie between 10^3 and 10^6 (median 3,056). Second, a run on the main branch cannot reach a detached branch even when the fold of that branch lies inside the range covered; this happened for 3,962 of the 5,101 single-start "no bifurcation" sets of Iterations 1–2 that have a detached branch. In addition, the integration used to start the continuation was not checked for convergence: repeating it with `ode15s` for 500 random parameter sets, 29 (5.8%) had not reached the steady state at $t = 5000$ (relative distance $> 10^{-3}$), and only 7 of these would have been flagged by a residual test with threshold $|\dot{x}|_\infty > 10^{-6}$.

Biphasic criterion

In the exact pipeline a parameter set is labelled biphasic (Class 9) when its main branch has no fold or branch point in the window and x_2 has an interior maximum along it. Specifically, the steady states of the main branch inside the window are sorted by A_{TOT} (at least five points are required); successive differences of x_2 smaller in magnitude than 10^{-9} times the largest value of x_2 are discarded as numerical noise, and the branch is biphasic when the remaining differences change sign from positive to negative exactly once, i.e. when x_2 has exactly one interior local maximum (function `is_biphasic` in `support/classify_branches.m`). To test whether the label depends on how small a maximum is accepted, we required the peak prominence (the height of the maximum above the higher of the two flanking minima) to exceed a fraction τ of the amplitude of x_2 along

the branch. For the 1,217 biphasic parameter sets, no label changes at $\tau = 0.1\%$, 1%, 5% or 10%; the minimum prominence is 99.99% of the amplitude. The maximum lies well inside the window (median $A_{\mathrm{TOT}} \approx 11$; 5–95% range 6–38). The biphasic class is therefore insensitive to the prominence threshold (Supplementary Note 6 and Supplementary Fig. 5).

Analytic derivatives and normal-form coefficients

The model file `MAPKfull.m` supplies, in closed form, the Jacobian of Eq. (2) with respect to the state and to the parameters, the corresponding second derivatives and the third derivatives with respect to the state. Singularity detection and normal-form coefficients in MATCONT therefore do not rely on finite differences, and the same Jacobian determines the stability of every steady state in the exact labelling. At every limit point and branch point inside the window the normal-form coefficients are computed in `Ensemble_LHS_v3.m` from these derivatives, in the conventions of MATCONT [35, 36]. At a fold, $a_{LP} = \frac{1}{2}\, p^{\mathsf{T}} B(q, q)$, where q and p are the right and left null vectors of the Jacobian normalised by $\|q\| = 1$ and $p^{\mathsf{T}} q = 1$; this is the coefficient MATCONT reports at limit points (its sign depends on the orientation of q, so $|a_{LP}|$ is reported). At a branch point MATCONT reports no normal-form coefficient, so the coefficients a, b, c of the algebraic branching equation, its discriminant and the slopes of the two crossing branches are computed in the same way. The symbolic definitions of these coefficients are given in Mathematica (Supplementary File 6, `Analytical_Derivations_StepByStep_final.nb`, Section 9). The fold coefficients were checked against MATCONT on the validation subsample (verification (vi) above).

Machine-learning classification

Data. The classifiers were trained on the ten-class exact labels, read directly from the output files of the labelling pipeline. The 32 flagged parameter sets were removed, leaving 39,968 parameter sets (10,000, 10,000, 9,969 and 9,999 in Iterations 1–4). Only the 14 sampled quantities are used as inputs; descriptors derived from the labels (bistability, multistationarity, presence of a detached branch, singularity locations) are never used as features.

Features. All features are computed from $\log_{10}$ of the parameters, so a pairwise ratio is a difference of log-parameters. Six feature sets were compared: A, the 14 log-parameters and all $\binom{14}{2} = 91$ pairwise log-ratios (105 features); B, set A without the seven ratios fixed by the sampling design ($K_{\mathrm{TOT}}/P_{\mathrm{TOT}}$, k_1/k_5, k_{11}/k_{55}, k_2/k_6, k_3/k_7, k_{33}/k_{77}, k_4/k_8); C, the 14 log-parameters alone (for tree-based models equivalent to the raw parameters [58]); D, the features selected by divergence-based selection, i.e. for every pair of classes the features are ranked by the Jensen–Shannon divergence [59] between the two class-conditional histograms (30 bins), the top features are retained up to the broken-stick cut-off [60], and the union over class pairs is kept, recomputed within each training split; E, the centred log-ratio transform [49] of the 14 log-parameters; and F, the centred log-ratio transform of the 12 rate constants together with $\log_{10}(K_{\mathrm{TOT}}/P_{\mathrm{TOT}})$. Before training, zero-variance columns are removed and exactly duplicated columns are collapsed to one (within Iterations 3 and 4 many ratios coincide because of the symmetry constraints), so that importance is not split between identical features (Supplementary Table 27).

Models. Five classifiers were compared: three gradient-boosted tree ensembles [61], LightGBM [62] (300 boosting rounds, 63 leaves, learning rate 0.1, minimum 3 samples per leaf, gain-based importance), XGBoost (300 rounds, maximum depth 6, learning rate 0.1, histogram method) and CatBoost (300 iterations, depth 6, learning rate 0.1); a random forest [63] (scikit-learn [64]; 300 trees, 30% of features considered at each split, minimum 2 samples per leaf); and Gaussian naive Bayes. The model comparison additionally includes scikit-learn gradient boosting (200 rounds, depth 5, learning

rate 0.1) and a most-frequent-class predictor as a baseline. All models were called from MATLAB R2026a through its Python interface (Python 3.9, NumPy 2.0.2 [65], LightGBM 4.6.0, XGBoost 2.1.4, CatBoost 1.2.10, scikit-learn 1.6.1); the experiments were distributed over 64 parallel workers, one thread per model. Training samples were weighted inversely to the frequency of their class (weight $n/(Kn_c)$ for a class with n_c of n training samples and K classes). LightGBM is the primary model for feature importance, confusion matrices and the figures.

Dataset splitting (Supplementary Note 8 and Supplementary Table 5). Every data set (the pooled data, or one sampling iteration) was split, stratified, into a training set (70%), a validation set (10%) and a held-out test set (20%). Pooled splits were stratified jointly by sampling iteration and class, within-iteration splits by class. Models are fitted on the training rows only; all reported accuracies refer to the held-out test rows, and the validation rows are used only for the convergence curves. Everything that depends on the data, i.e. the removal of constant and duplicated columns and the feature selection of set D, is computed on the training rows of each split.

Evaluation (full results in Supplementary Notes 9–14). (i) Feature-set ablation: each combination of feature set and model was evaluated on three independent splits with fixed seeds (43, 44 and 45 for the pooled data; $42 + 100a + r$ for repeat r within iteration a); we report the mean and standard deviation of the held-out accuracy, balanced accuracy (mean per-class recall) [66], macro-averaged F1 and Cohen's κ [67], together with the majority-class baseline of the same test set. (ii) Model comparison: for feature sets A and D, every model was evaluated by five-fold stratified cross-validation inside the training rows and on the held-out test rows, for the pooled data and within each iteration. (iii) ROC curves: one-vs-rest receiver operating characteristic curves from the predicted class probabilities on the held-out test rows (MATLAB `perfcurve`), with the area under the curve per class and its macro average. (iv) Convergence: the multiclass log-loss and the error rate on the training rows (a random subsample of up to 5,000) and on the validation rows were recorded at 30 checkpoints along the boosting rounds (number of trees for the random forest), together with the validation error of every class. (v) Feature importance: LightGBM split gain, normalised to sum to one, from the first split. (vi) Decisive tests: Classes 1 and 2 in Iteration 1, Classes 3 and 4 in Iteration 2 and Classes 7 and 8 in Iteration 4 were classified separately, with all non-constant features and with two mechanistic log-ratios only ($\log_{10}(k_2 K_{\mathrm{TOT}}/k_8 P_{\mathrm{TOT}})$ and $\log_{10}(k_4 K_{\mathrm{TOT}}/k_6 P_{\mathrm{TOT}})$ for the first two pairs; $\log_{10}(k_1/k_5)$ and $\log_{10}(k_{11}/k_{55})$ for the third), and a linear support-vector classifier [68] (MATLAB `fitcsvm`, linear kernel, box constraint 1) was fitted on the two log-ratios; three splits each. (vii) Design confound: LightGBM was trained to predict the sampling iteration instead of the class. (viii) Embedding audit: a two-dimensional t-SNE embedding [48] (MATLAB `tsne`, perplexity 30) was computed for 6,000 random parameter sets, and the accuracy with which a k-nearest-neighbour classifier [69] ($k = 15$, five-fold cross-validation) recovers the iteration and the class from the embedding was compared with that of a supervised classifier; the same was done separately for the regimes $K_{\mathrm{TOT}} > P_{\mathrm{TOT}}$, $K_{\mathrm{TOT}} < P_{\mathrm{TOT}}$ and $K_{\mathrm{TOT}} = P_{\mathrm{TOT}}$ (3,000 sets each). (ix) Per-class contributions: within each iteration, LightGBM was refitted on the training rows of the same three splits as in (i), and the TreeSHAP contribution [50] of every feature to the raw score of every class was computed on the held-out test rows (LightGBM `pred_contrib`). Each class was summarised by the mean absolute contribution of each feature over the test rows, averaged over the three splits, and by the direction of the effect, the sign of the correlation between the feature value and its contribution. Contributions were also grouped by parameter, by splitting the contribution of each log-ratio equally between its two parameters (SHAP values are additive).

Class transitions

Single-parameter scan. In each iteration, 120 parameter sets of every class were drawn at random from the ensemble (sets used for machine learning; 1,560 sets in total). Each of the 14 parameters, and in Iterations 3 and 4 the pairs (k_4, k_8) and (k_1, k_{11}), which preserve the symmetry $u^* = r$, was multiplied by 10^d with $d \in \{-1, -0.5, 0.5, 1\}$, and every result was labelled exactly with the labeller of the ensemble (92,160 labels). For every pair of classes, the move that converts the largest fraction of the source sets is reported. *Exact-condition moves.* The quantity of an exact condition was set to a margin by changing a single parameter: $x = -0.3$ through k_8 (Classes 1 and 10 of Iteration 1), $y = +0.3$ through k_6 or k_4 (Classes 3 and 9 of Iteration 2), $k_4 = k_8 = 2k_2$ (Class 10 of Iterations 3 and 4) and $\log_{10}(K_{M1}/K_{M5}) = \pm 0.3$ through k_5 (Class 6). *Two-class sweeps.* For each edge of the transition map, one quantity was swept through the boundary in 80 sets of the source class, at 41 points: x or y over $[-1, 1]$ through one parameter (k_8 or k_4; k_2, k_6, K_{TOT} and P_{TOT} for comparison), $\log_{10}(k_4/k_2)$ over $[-1, 1]$ with $k_4 = k_8$, $\log_{10}(K_{M1}/K_{M5})$ over $[-1, 1]$ through k_5, or k_3 (and, for comparison, k_7) multiplied by 10^{2v}, $v \in [-1, 1]$, which changes $\log_{10}(k_3/k_7)$ without changing x or y. A sweep is two-class when exactly the two classes of the edge, and no other, occur along it. *Representative sets* of Fig 7: among the 150 sets nearest to the class median, the set with all singularities in $10^{0.5} \le A_{\mathrm{TOT}} \le 10^5$ and the widest main-branch fold pair (Classes 1–4, with the detached fold at least 0.7 decades from the main folds for Classes 2 and 4) or the largest separation of fold and branch point (Classes 7 and 8); the nearest such set for the other classes (Class 10 from Iterations 1–2); the starred set of Classes 1, 3, 9 and 10 is the open circle whose detached fold is nearest to $A_{\mathrm{TOT}} = 10^7$. Bifurcation diagrams were drawn from the exact parametrisation, with stability from the eigenvalues of the analytic Jacobian. *Sets beyond the window* (open circles of Fig 7): all sets used for machine learning with $x < 0$, $y > 0$, $u^* \neq r$ and label 1, 3, 9 or 10; for each, the A_{TOT} of the detached fold was computed exactly from the parametrisation (anywhere, not only in the window), and the class with the window widened to include it (`TR_05_window_sets.m`). The misclassified held-out test sets of the pooled LightGBM model were checked in the same way (same script). Random choices use a portable generator, so that MATLAB and GNU Octave select the same sets.

Software, hardware and computational cost

Supplementary Table 29 lists the software required by each step, and Supplementary Note 15 (Supplementary Table 15) the cost of each iteration; the exact labelling requires only base MATLAB [70]; the analytical results (Eqs (1)–(14)) were derived with Wolfram Mathematica 15 [46]. All computations were run on a Linux server with an AMD EPYC 9334 processor (32 cores) and 251 GB RAM, with MATLAB R2026a. MATCONT can be downloaded from `https://sourceforge.net/projects/matcont/`. The labelling pass is embarrassingly parallel and was run with `parfor` on 64 workers. We report the cost in core-hours (serial CPU time, the sum of the per-set times). The exact labelling, including the normal-form coefficients, takes 11 ms per parameter set on average (9–14 ms in the four iterations), i.e. 0.13 core-hours for all 40,000 sets. The MATCONT validation runs seeded on every branch take 8.0 s per set (7.7–8.3 s), i.e. 8.9 core-hours for the 4,000 validated sets, about 700 times the cost per set of the exact labelling. Single-start continuation of all 40,000 sets takes 0.44 s per set on average (median 0.30 s), i.e. 4.9 core-hours and 4.8 min wall clock on 64 workers: about 40 times the cost of the exact labelling, which in addition finds the detached branches and the singularities at large A_{TOT} that it misses. The main loop of all four iterations, exact labelling and validation together, took 48 min wall-clock on 64 workers. The trained LightGBM classifier (feature set A) assigns a class in 0.11 ms per set when the 7,994 held-out sets are classified together and in 0.45 ms (median) for a single set, on one core and including the computation of the 105 features from the 14

parameters; training on 31,974 sets takes 27 s (`ML_v3_inference_time.m`).

Use of large language models

Claude (Anthropic) was used to assist with figure preparation and language editing. The authors reviewed all output and take full responsibility for the content of the manuscript.

Code availability

All code is publicly available on GitHub [72] (`https://github.com/HarikaGL/High_throughput_bifurcation`) and is archived with the data in Zenodo [71] (DOI: 10.5281/zenodo.23609482; licence MIT). It comprises the exact labelling pipeline (`Ensemble_LHS_v3.m`, the model file `MAPKfull.m` and the labeller in `support/`; MATLAB), the machine-learning analysis (`ML_v3_10class_parallel.m`, `ML_v3_SHAP_per_class.m`; models called through MATLAB's Python interface), the class-transition analysis (`Class_transitions/TR_00_run_all.m`), the comparison with single-start continuation (`make_comparison_figures.m`), every figure script, and the Mathematica notebooks of the analytical derivations (Supplementary Files 2–6). Supplementary Table 17 lists the code behind each figure and table, and `README` files give the order of the runs, their inputs, outputs and run times.

Data availability

The labelled ensemble and all results are published and publicly available in Zenodo [71] (DOI: 10.5281/zenodo.23609482; licence CC BY 4.0). For every one of the 40,000 parameter sets the deposit gives the 14 parameters, the sampling iteration, the ten-class label, the flag for a fold coinciding with the branch point, all fold and branch-point locations with their normal-form coefficients, the maximum number of coexisting steady states, the bistable ranges of A_{TOT} and the single-start labels for comparison (CSV and Excel files, one per iteration), together with the complete steady-state results (MATLAB MAT files), the machine-learning, transition and comparison results, and the Chemical Reaction Network Toolbox reports (Supplementary File 1). The Supplementary Files are described in Supplementary Note 16 and Supplementary Table 16.

Acknowledgements

The authors sincerely thank Dr. P. K. Vinod (Centre for Computational Natural Sciences and Bioinformatics, IIIT Hyderabad) and Dr. Jainendra Shukla (Department of Computer Science and Engineering, IIIT-Delhi) for their valuable contributions to the machine learning component of this work. Their discussions and insights helped shape the development of the machine learning framework presented in this study. We also thank Alka Singh for thoroughly reading the manuscript and providing thoughtful comments that significantly improved its clarity and presentation.

Funding

The author(s) received no specific funding for this work.

Author contributions

G.L.H.: methodology, software, formal analysis, visualisation, writing. K.S.: conceptualisation, supervision, writing. Both authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

References

[1] Kitano H. Systems biology: a brief overview. Science. 2002;295(5560):1662–4. doi:10.1126/science.1069492.

[2] Alon U. An introduction to systems biology: design principles of biological circuits. 2nd ed. Chapman and Hall/CRC; 2019.

[3] Tyson JJ, Novák B. A dynamical paradigm for molecular cell biology. Trends in Cell Biology. 2020;30(7):504–15. doi:10.1016/j.tcb.2020.04.002.

[4] Ferrell JE. Bistability, bifurcations, and Waddington's epigenetic landscape. Current Biology. 2012;22(11):R458–66. doi:10.1016/j.cub.2012.03.045.

[5] Tyson JJ, Csikász-Nagy A, Novák B. The dynamics of cell cycle regulation. BioEssays. 2002;24(12):1095–109. doi:10.1002/bies.10191.

[6] Strogatz SH. Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering. Chapman and Hall/CRC; 2024.

[7] Gardner TS, Cantor CR, Collins JJ. Construction of a genetic toggle switch in *Escherichia coli*. Nature. 2000;403(6767):339–42. doi:10.1038/35002131.

[8] Mishra D, Bepler T, Teague B, Berger B, Broach J, Weiss R. An engineered protein-phosphorylation toggle network with implications for endogenous network discovery. Science. 2021;373(6550):eaav0780. doi:10.1126/science.aav0780.

[9] Kuznetsov YA. Elements of Applied Bifurcation Theory. 3rd ed. Applied Mathematical Sciences, vol. 112. New York: Springer; 2004. doi:10.1007/978-1-4757-3978-7.

[10] Bar-Even A, Noor E, Savir Y, Liebermeister W, Davidi D, Tawfik DS, et al. The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. Biochemistry. 2011;50(21):4402–10. doi:10.1021/bi2002289.

[11] Schreiber G, Haran G, Zhou HX. Fundamental aspects of protein–protein association kinetics. Chemical Reviews. 2009;109(3):839–60. doi:10.1021/cr800373w.

[12] Szep G, Dalchau N, Csikász-Nagy A. Parameter inference with bifurcation diagrams. Advances in Neural Information Processing Systems. 2021;34:5620–30.

[13] Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, Sethna JP. Universally sloppy parameter sensitivities in systems biology models. PLoS Computational Biology. 2007;3(10):e189. doi:10.1371/journal.pcbi.0030189.

[14] Nienałtowski K, Włodarczyk M, Lipniacki T, Komorowski M. Clustering reveals limits of parameter identifiability in multi-parameter models of biochemical dynamics. BMC Systems Biology. 2015;9:65. doi:10.1186/s12918-015-0205-8.

[15] Goldbeter A, Koshland DE. An amplified sensitivity arising from covalent modification in biological systems. Proc Natl Acad Sci USA. 1981;78(11):6840–4. doi:10.1073/pnas.78.11.6840.

[16] Ferrell JE. Tripping the switch fantastic: how a protein kinase cascade can convert graded inputs into switch-like outputs. Trends Biochem Sci. 1996;21(12):460–6. doi:10.1016/s0968-0004(96)20026-x.

[17] Ramesh V, Krishnan J. Symmetry breaking meets multisite modification. eLife. 2021;10:e65358. doi:10.7554/eLife.65358.

[18] Ramesh V, Krishnan J. A unified approach to dissecting biphasic responses in cell signaling. eLife. 2023;12:e86520. doi:10.7554/eLife.86520.

[19] Markevich NI, Hoek JB, Kholodenko BN. Signaling switches and bistability arising from multisite phosphorylation in protein kinase cascades. Journal of Cell Biology. 2004;164(3):353–9. doi:10.1083/jcb.200308060.

[20] Suwanmajo T, Krishnan J. Mixed mechanisms of multi-site phosphorylation. Journal of The Royal Society Interface. 2015;12(107):20141405. doi:10.1098/rsif.2014.1405.

[21] Jolley CC, Ode KL, Ueda HR. A design principle for a posttranslational biochemical oscillator. Cell Reports. 2012;2(4):938–50. doi:10.1016/j.celrep.2012.09.006.

[22] Thomson M, Gunawardena J. Unlimited multistability in multisite phosphorylation systems. Nature. 2009;460(7252):274–7. doi:10.1038/nature08102.

[23] Wang L, Sontag ED. On the number of steady states in a multiple futile cycle. Journal of Mathematical Biology. 2008;57(1):29–52. doi:10.1007/s00285-007-0145-z.

[24] Flockerzi D, Holstein K, Conradi C. N-site phosphorylation systems with 2n-1 steady states. Bulletin of Mathematical Biology. 2014;76(7):1892–916. doi:10.1007/s11538-014-9984-0.

[25] Feinberg M. Chemical reaction network structure and the stability of complex isothermal reactors. Chemical Engineering Science. 1987;42(10):2229–68. doi:10.1016/0009-2509(87)80099-4.

[26] Feinberg M. Foundations of Chemical Reaction Network Theory. Springer; 2019. doi:10.1007/978-3-030-03858-8.

[27] Ellison P, Feinberg M, Ji H, Knight D. The Chemical Reaction Network Toolbox, Version 2.35. The Ohio State University; 2018. https://crnt.osu.edu/CRNTWin.

[28] Conradi C, Feliu E, Mincheva M, Wiuf C. Identifying parameter regions for multistationarity. PLOS Computational Biology. 2017;13(10):e1005751. doi:10.1371/journal.pcbi.1005751.

[29] Conradi C, Mincheva M. Catalytic constants enable the emergence of bistability in dual phosphorylation. Journal of the Royal Society Interface. 2014;11(95):20140158. doi:10.1098/rsif.2014.0158.

[30] Siegal-Gaskins D, Mejia-Guerra MK, Smith GD, Grotewold E. Emergence of switch-like behavior in a large family of simple biochemical networks. PLOS Computational Biology. 2011;7(5):e1002039. doi:10.1371/journal.pcbi.1002039.

[31] Huang YJ, Okada T, Mochizuki A. Uncovering bifurcation behaviors of biochemical reaction systems from network topology. Scientific Reports. 2025;15:27596. doi:10.1038/s41598-025-10688-6.

[32] Alamin AJLJ, Hernandez BS. Computational framework NetBiF predicts and classifies steady-state bifurcations in biochemical systems via reaction network structure. arXiv:2610.18087 [math.DS]; 2026.

[33] Talabis DASJ, Mendoza VMP, Hernandez BS. Approximating parameter regions for multistationarity and multistability in mass-action systems. arXiv:2610.05242 [q-bio.QM]; 2026.

[34] Pal S, Melnik R. Machine learning for complex systems dynamics: detecting bifurcations in dynamical systems with deep neural networks. arXiv:2603.04420; 2026.

[35] Dhooge A, Govaerts W, Kuznetsov YA. MATCONT: a MATLAB package for numerical bifurcation analysis of ODEs. ACM Transactions on Mathematical Software. 2003;29(2):141–64. doi:10.1145/779359.779362.

[36] Dhooge A, Govaerts W, Kuznetsov YA, Meijer HGE, Sautois B. New features of the software MatCont for bifurcation analysis of dynamical systems. Mathematical and Computer Modelling of Dynamical Systems. 2008;14(2):147–75. doi:10.1080/13873950701742754.

[37] MatCont development team. MatCont and CL MatCont: continuation toolboxes in MATLAB – manual. Ghent University and Utrecht University; 2019. https://sourceforge.net/projects/matcont/.

[38] Harika GL, Sriram K. EMBED: Ensemble MATCONT-Based Detection and Classification of Bifurcations in Dual Phosphorylation–Dephosphorylation Reaction Networks. ACS Omega. 2026;11(38):57307–31. doi:10.1021/acsomega.6c05467.

[39] Nielsen J, Keasling JD. Engineering cellular metabolism. Cell. 2016;164(6):1185–97. doi:10.1016/j.cell.2016.02.004.

[40] Carbonell P, Jervis AJ, Robinson CJ, Yan C, Dunstan M, Swainston N, et al. An automated Design-Build-Test-Learn pipeline for enhanced microbial production of fine chemicals. Communications Biology. 2018;1:66. doi:10.1038/s42003-018-0076-9.

[41] Cohen P. Protein kinases—the major drug targets of the twenty-first century? Nature Reviews Drug Discovery. 2002;1(4):309–15. doi:10.1038/nrd773.

[42] Blume-Jensen P, Hunter T. Oncogenic kinase signalling. Nature. 2001;411(6835):355–65. doi:10.1038/35077225.

[43] Julien SG, Dubé N, Hardy S, Tremblay ML. Inside the human cancer tyrosine phosphatome. Nature Reviews Cancer. 2011;11(1):35–49. doi:10.1038/nrc2980.

[44] Tonks NK. Protein tyrosine phosphatases: from genes, to function, to disease. Nature Reviews Molecular Cell Biology. 2006;7(11):833–46. doi:10.1038/nrm2039.

[45] Feliu E, Kaihnsa N, de Wolff T, Yürük O. The kinetic space of multistationarity in dual phosphorylation. J Dyn Differ Equ. 2022;34:825–852. doi:10.1007/s10884-020-09889-6.

[46] Wolfram Research, Inc. Mathematica, Version 15. Champaign, Illinois, United States: Wolfram Research, Inc.; 2026.

[47] Golubitsky M, Schaeffer DG. Singularities and Groups in Bifurcation Theory, Volume I. Applied Mathematical Sciences, vol. 51. New York: Springer; 1985. doi:10.1007/978-1-4612-5034-0.

[48] van der Maaten L, Hinton G. Visualizing data using t-SNE. Journal of Machine Learning Research. 2008;9:2579–605.

[49] Aitchison J. The Statistical Analysis of Compositional Data. London: Chapman and Hall; 1986.

[50] Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, Katz R, Himmelfarb J, Bansal N, Lee SI. From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence. 2020;2(1):56–67. doi:10.1038/s42256-019-0138-9.

[51] Xiong W, Ferrell JE Jr. A positive-feedback-based bistable 'memory module' that governs a cell fate decision. Nature. 2003;426(6965):460–465. doi:10.1038/nature02089.

[52] Ozbudak EM, Thattai M, Lim HN, Shraiman BI, van Oudenaarden A. Multistability in the lactose utilization network of *Escherichia coli*. Nature. 2004;427(6976):737–740. doi:10.1038/nature02298.

[53] Ellison P. The Advanced Deficiency Algorithm and Its Applications to Mechanism Discrimination. PhD thesis, University of Rochester; 1998.

[54] McKay MD, Beckman RJ, Conover WJ. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics. 1979;21(2):239–45. doi:10.2307/1268522.

[55] Stein M. Large sample properties of simulations using Latin hypercube sampling. Technometrics. 1987;29(2):143–51. doi:10.2307/1269769.

[56] Wilson EB. Probable inference, the law of succession, and statistical inference. Journal of the American Statistical Association. 1927;22(158):209–12. doi:10.1080/01621459.1927.10502953.

[57] Shampine LF, Reichelt MW. The MATLAB ODE Suite. SIAM Journal on Scientific Computing. 1997;18(1):1–22. doi:10.1137/S1064827594276424.

[58] Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed. New York: Springer; 2009. doi:10.1007/978-0-387-84858-7.

[59] Lin J. Divergence measures based on the Shannon entropy. IEEE Transactions on Information Theory. 1991;37(1):145–51. doi:10.1109/18.61115.

[60] Jackson DA. Stopping rules in principal components analysis: a comparison of heuristical and statistical approaches. Ecology. 1993;74(8):2204–14. doi:10.2307/1939574.

[61] Friedman JH. Greedy function approximation: a gradient boosting machine. The Annals of Statistics. 2001;29(5):1189–232. doi:10.1214/aos/1013203451.

[62] Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. LightGBM: a highly efficient gradient boosting decision tree. In: Advances in Neural Information Processing Systems 30 (NeurIPS 2017); 2017. p. 3146–54.

[63] Breiman L. Random forests. Machine Learning. 2001;45(1):5–32. doi:10.1023/A:1010933404324.

[64] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. Journal of Machine Learning Research. 2011;12:2825–30.

[65] Harris CR, Millman KJ, van der Walt SJ, et al. Array programming with NumPy. Nature. 2020;585(7825):357–62. doi:10.1038/s41586-020-2649-2.

[66] Brodersen KH, Ong CS, Stephan KE, Buhmann JM. The balanced accuracy and its posterior distribution. In: Proceedings of the 20th International Conference on Pattern Recognition (ICPR); 2010. p. 3121–4. doi:10.1109/ICPR.2010.764.

[67] Cohen J. A coefficient of agreement for nominal scales. Educational and Psychological Measurement. 1960;20(1):37–46. doi:10.1177/001316446002000104.

[68] Cortes C, Vapnik V. Support-vector networks. Machine Learning. 1995;20(3):273–97. doi:10.1007/BF00994018.

[69] Cover TM, Hart PE. Nearest neighbor pattern classification. IEEE Transactions on Information Theory. 1967;13(1):21–7. doi:10.1109/TIT.1967.1053964.

[70] The MathWorks, Inc. MATLAB and Statistics and Machine Learning Toolbox, Release R2026a. Natick, Massachusetts, United States: The MathWorks, Inc.; 2026.

[71] Harika GL, Sriram K. Exact ensemble bifurcation analysis reveals hidden bistability and makes signalling dynamics predictable from kinetic parameters: labelled ensemble, results and code [Dataset]. Version 1.0. Zenodo; 2026. `https://doi.org/10.5281/zenodo.23609482`.

[72] Harika GL, Sriram K. High_throughput_bifurcation: code for the exact bifurcation labelling of dual phosphorylation ensembles. Version 1.0.0. GitHub; 2026. `https://github.com/HarikaGL/High_throughput_bifurcation`.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- npjSYSBIOSUPP.pdf