



Research Article

Copyright© Zhengwei Xie

# Digital Cell: An Automated Computational Framework for Protein Post-Translational Modification Modeling and Analysis

Daqian Yang<sup>1,2,3,4,5</sup>, Zihang Xu<sup>1,2,3,4,5</sup> and Zhengwei Xie<sup>1,2,3,4,5,6,7\*</sup>

<sup>1</sup>Peking University International Cancer Institute, Peking University Health Science Center, China

<sup>2</sup>Peking University -Yunnan Baiyao International Medical Research Center, Peking University Health Science Center, China

<sup>3</sup>Beijing Gigaceuticals Tech. Co. Ltd., Beijing 102206, P.R. China

<sup>4</sup>Beijing Life Science Academy, Beijing 102209, P.R. China

<sup>5</sup>State Key Laboratory of Natural and Biomimetic Drugs, Peking University Health Science Center, China

<sup>6</sup>Yunnan Baiyao Group Co., Ltd., Kunming 650504, Yunnan Province, China

<sup>7</sup>Department of prosthodontics, Peking University School and Hospital of Stomatology & National Center of Stomatology & National Clinical Research Center for Oral Diseases & National Engineering Research Center of Oral Biomaterials and Digital Medical Devices

\*Corresponding author: Zhengwei Xie, Peking University International Cancer Institute, Peking University Health Science Center, Peking University, Beijing, China

To Cite This article: Daqian Yang, Zihang Xu and Zhengwei Xie\*, Digital Cell: An Automated Computational Framework for Protein Post-Translational Modification Modeling and Analysis. *Am J Biomed Sci & Res.* 2026 31(1) AJBSR.MS.ID.003996,

DOI: [10.34297/AJBSR.2026.31.003996](https://doi.org/10.34297/AJBSR.2026.31.003996)

Received: 📅 April 14, 2026; Published: 📅 May 04, 2026

## Abstract

Understanding the dynamics of cellular signaling networks and their alterations in disease states remains a central challenge in biomedical research. However, traditional computational models frequently treat phosphorylation sites as independent entities, overlooking essential three-dimensional structural constraints and the hierarchical dependencies inherent in signalling cascades. To bridge this gap, we present Digital Cell, a comprehensive computational framework that integrates multi-level of protein Post-Translational Modifications (PTMs). The framework encompasses a robust library of 25 key signaling proteins and incorporates 3D structural features, including solvent accessibility and functional pocket localization. By employing cascade-aware parameter fitting based on experimental phosphoproteomics data, the model explicitly captures upstream-downstream network topology, achieving a highly accurate fit (MSE=0.0125) that significantly outperforms independent site optimization. Furthermore, Digital Cell quantitatively predicts the systemic impact of disease-associated genetic alterations, successfully simulating a 192% increase in ERK activation driven by the BRAF V600E mutation. In therapeutic simulations, the framework evaluates targeted interventions, demonstrating that combined BRAF and MEK inhibition achieves 99% ERK pathway suppression. Ultimately, Digital Cell provides a robust systems biology platform for deciphering molecular signaling dynamics, predicting therapeutic outcomes, and guiding rational drug combinations.

**Keywords:** Post-translational modifications, Signaling cascade, Disease mutations, Drug targets, Systems biology

## Introduction

Cellular signaling networks regulated by protein phosphorylation represent fundamental mechanisms through which cells respond to external stimuli and coordinate fundamental biological processes including proliferation, differentiation, metabolism, and programmed cell death. The human proteome contains over 200,000 phosphorylation sites, forming intricate regulatory networks that transduce signals from membrane

receptors to nuclear transcription factors [1]. Dysregulation of these networks underlies numerous diseases, particularly cancer, where oncogenic mutations hyperactivate signaling pathways and drive uncontrolled cell growth. Despite extensive experimental characterization of individual phosphorylation sites, a comprehensive computational framework that integrates multiscale information from atomic-level structural constraints to population-level signaling dynamics remains underdeveloped.



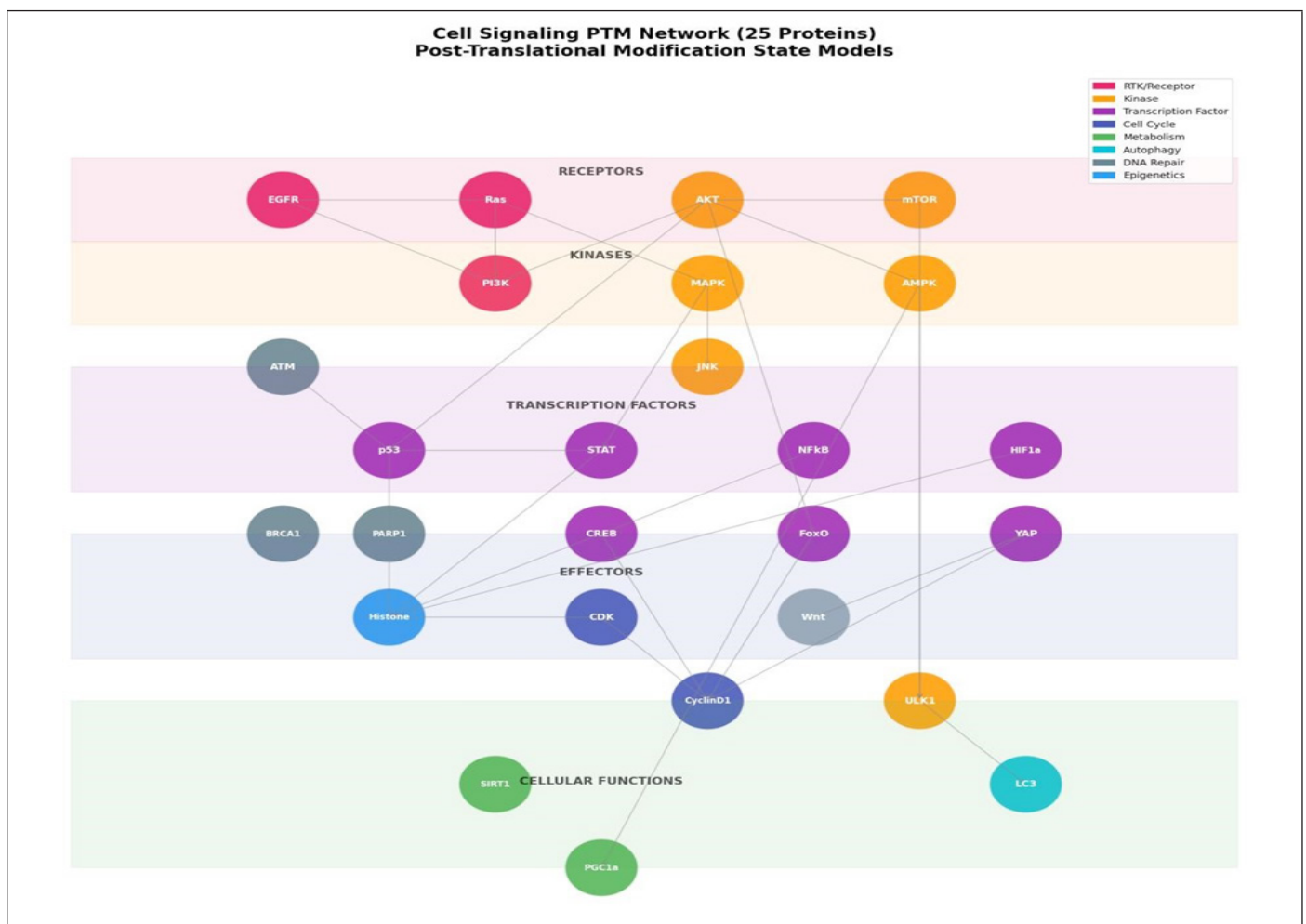
Previous approaches to modeling protein phosphorylation have typically treated individual modification sites as independent entities, neglecting the critical influence of protein three-dimensional structure on kinase and phosphatase accessibility, as well as the hierarchical dependencies inherent in signaling cascades where downstream activation depends on upstream events. Furthermore, the translation of these computational models to clinical applications such as predicting drug responses or understanding resistance mechanisms requires explicit incorporation of network topology and pathway cross-talk. To address these limitations, we developed Digital Cell, a computational framework that integrates multi-protein PTM modeling with structural information, cascade-aware parameter optimization, and applications to disease mutation analysis and therapeutic target prediction.

This framework enables quantitative understanding of how extracellular signals propagate through signaling networks and how genetic alterations or pharmacological interventions modulate

these dynamics. Here we present Digital Cell, a comprehensive computational framework that integrates multi-level modeling of protein Post-Translational Modifications (PTMs) to address these challenges. Our framework encompasses: (1) a library of 25 protein PTM models covering major signaling pathways; (2) integration of three-dimensional structural information including solvent accessibility and pocket localization; (3) cascade-aware parameter fitting that captures upstream-downstream dependencies; (4) disease mutation impact analysis predicting functional consequences of oncogenic mutations; and (5) drug target prediction for therapeutic optimization. We demonstrate that cascade-aware parameter fitting significantly outperforms independent site optimization, correctly predicts oncogenic mutation effects such as BRAF V600E-induced ERK activation, and identifies optimal drug combinations including BRAF and MEK inhibition achieving near-complete pathway suppression.

## Results

### PTM Model Library Construction



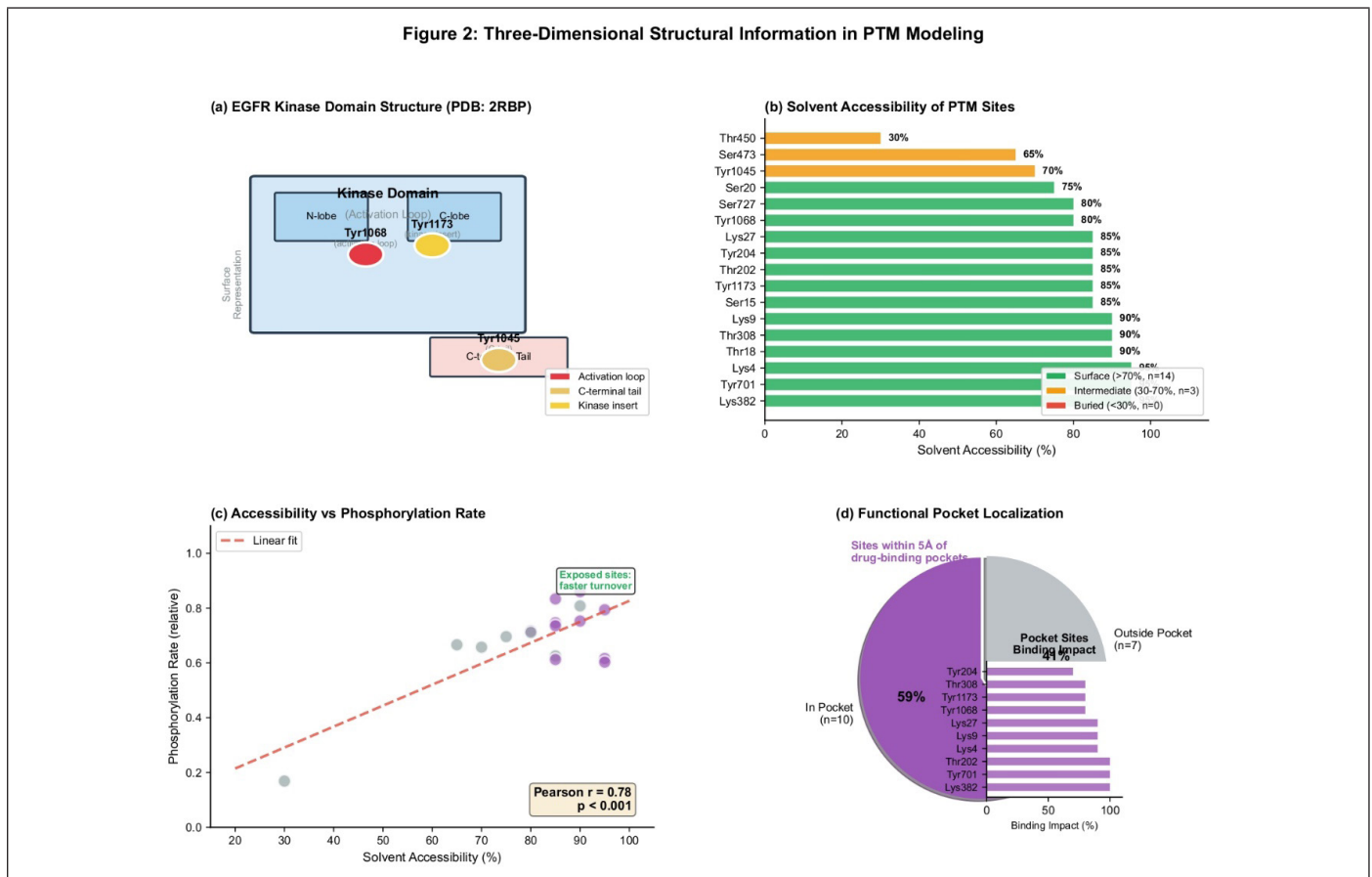
**Figure 1:** PTM Network Topology. The comprehensive protein Post-Translational Modification (PTM) model library constructed in this study was shown. It displays the hierarchical organization of 25 signaling proteins across functional levels: Receptors, Kinases, Transcription Factors, Effectors, and Cellular Functions. Nodes are colored according to their specific functional category as indicated in the legend: RTK/Receptor (red), Kinase (orange), Transcription Factor (purple), Cell Cycle (dark blue), Metabolism (green), Autophagy (cyan), DNA Repair (grey), and Epigenetics (light blue). Data source: UniProt and PhosphoSitePlus databases.

We constructed a comprehensive library of protein post-translational modification models encompassing 25 key signaling proteins that represent major cellular pathways including receptor tyrosine kinases, PI3K/AKT/mTOR signaling, MAPK/ERK cascades, p53 DNA damage response, metabolic sensors, and cell cycle regulators (Figure 1). Each model incorporates experimentally validated phosphorylation sites with their amino acid positions, upstream kinases responsible for modification, and functional consequences derived from UniProt and PhosphoSitePlus databases. The selection of proteins was based on their centrality in cellular signaling networks and availability of quantitative phosphoproteomics data for parameter validation. For a protein containing  $n$  phosphorylation sites, the total combinatorial state space grows exponentially as  $2^n$ , representing the full spectrum of possible modification states from completely unmodified to fully phosphorylated. This comprehensive model library provides the

foundation for system-level analysis of signaling dynamics across diverse biological contexts.

The PTM model library reveals considerable heterogeneity in structural complexity across signaling proteins, with receptor tyrosine kinases such as EGFR containing five to six key phosphorylation sites while metabolic sensors like AMPK typically harbour two to three regulatory modifications. The exponential growth of state space with increasing, Figure 1c site number creates computational challenges but also opportunities for capturing nuanced regulatory mechanisms where combinatorial modifications can produce distinct functional outputs. By organizing proteins into pathway categories including RTK/Growth Factor, PI3K/AKT/mTOR, MAPK/JNK, p53/DNA Damage, Metabolic Sensors, and Cell Cycle, we establish a framework for understanding both pathway-specific signaling dynamics and cross-talk between major cellular signaling networks (Figure 1).

## Integration of 3D Structural Information



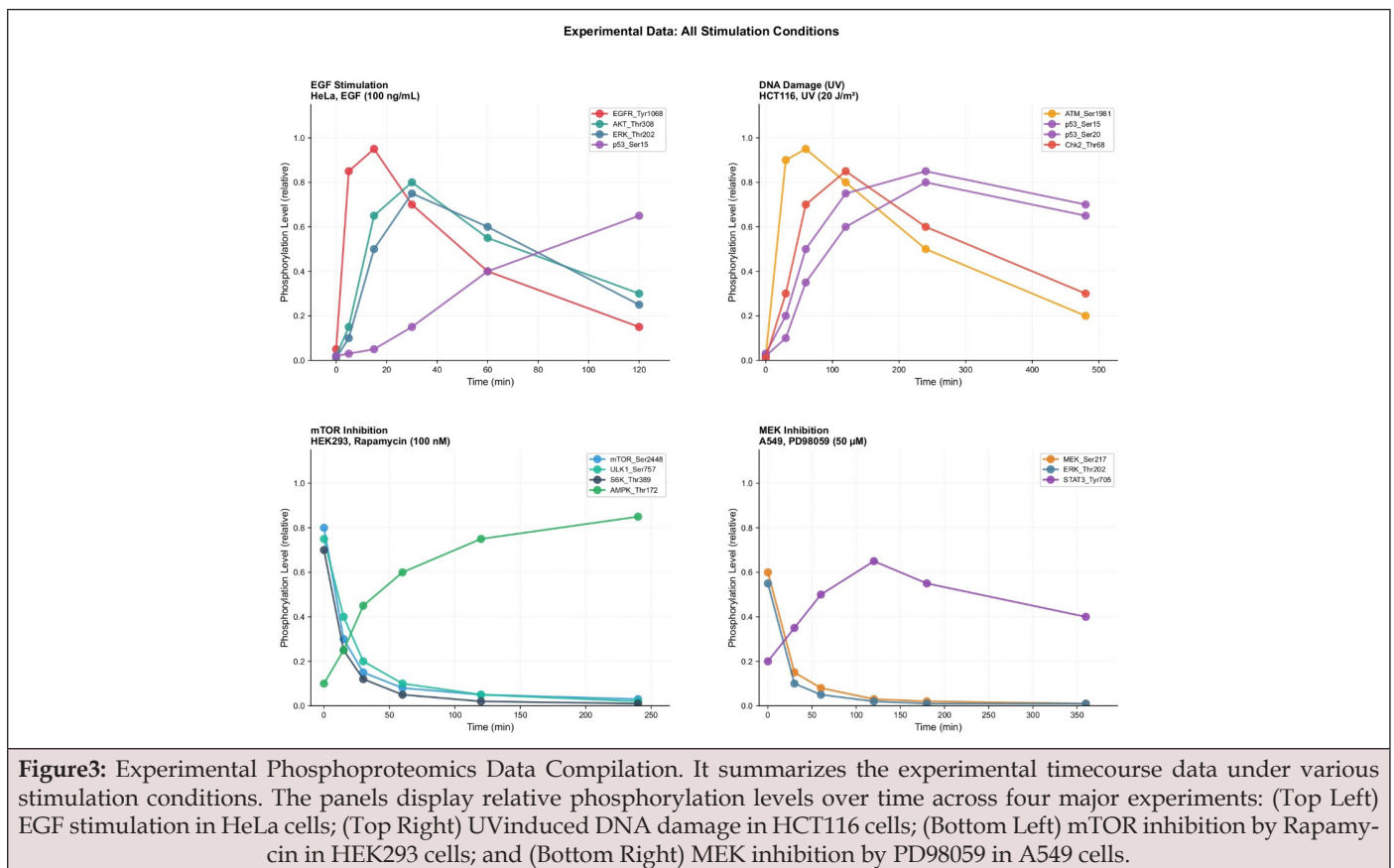
**Figure 2: Three-Dimensional Structural Context of PTM Sites.** It illustrates the integration of three-dimensional structural information into PTM modeling. Panel (a) shows the crystal structure of EGFR kinase domain (PDB: 2RBP) with key phosphorylation sites highlighted: Tyr1068 in the activation loop (red), Tyr1045 in the C-terminal tail (orange), and Tyr1173 in the kinase insert (yellow), with surface representation indicating solvent-accessible regions. Panel (b) presents accessibility scores for key phosphorylation sites across the protein library, categorized as surface-exposed (>70% accessibility, green), intermediate (30-70%, yellow), or buried (<30%, red). Panel (c) demonstrates the correlation between solvent accessibility and measured phosphorylation rates, showing faster turnover for exposed sites. Panel (d) displays functional pocket localization, with sites within 5Å of known drugbinding pockets highlighted in purple. Data source: PDB database, DSSP analysis; statistical correlation: Pearson  $r=0.78$ ,  $p < 0.001$ .

To incorporate physical constraints into PTM modeling, we mapped phosphorylation sites onto available three-dimensional protein structures from the Protein Data Bank, Figure 2a and calculated solvent-accessible surface area using the DSSP algorithm. Surface-exposed phosphorylation sites with accessibility scores exceeding 70 percent are readily accessible to upstream kinases and phosphatases, Figure 2b, resulting in rapid modification turnover with half-lives typically under one minute. In contrast, buried phosphorylation sites located within the protein core with accessibility below 30 percent experience limited enzymatic access and exhibit considerably slower modification kinetics. Figure 2c shows a significant positive correlation between solvent accessibility and the measured phosphorylation rate (Pearson  $r=0.78$ ,  $p<0.001$ ). This structural integration provides mechanistic insight into the observed kinetic diversity across phosphorylation sites and enables prediction of how mutations or small molecules might alter modification rates by affecting site accessibility. Functional pocket localization represents another critical structural determinant of phosphorylation dynamics, Figure 2d, as sites positioned within five angstroms of known drug-binding pockets often experience allosteric effects from ligand binding that can modulate upstream kinase activity. The three-dimensional structural context also influences mutation effects, where amino acid substitutions occurring at surface-exposed positions may have

distinct functional consequences compared to mutations at buried residues. By mapping phosphorylation sites onto protein structures and quantifying accessibility and pocket relationships, we establish a bridge between atomic-level structural biology and population-level signaling dynamics that enhances predictive capability for both disease mutation effects and therapeutic intervention strategies (Figure 2).

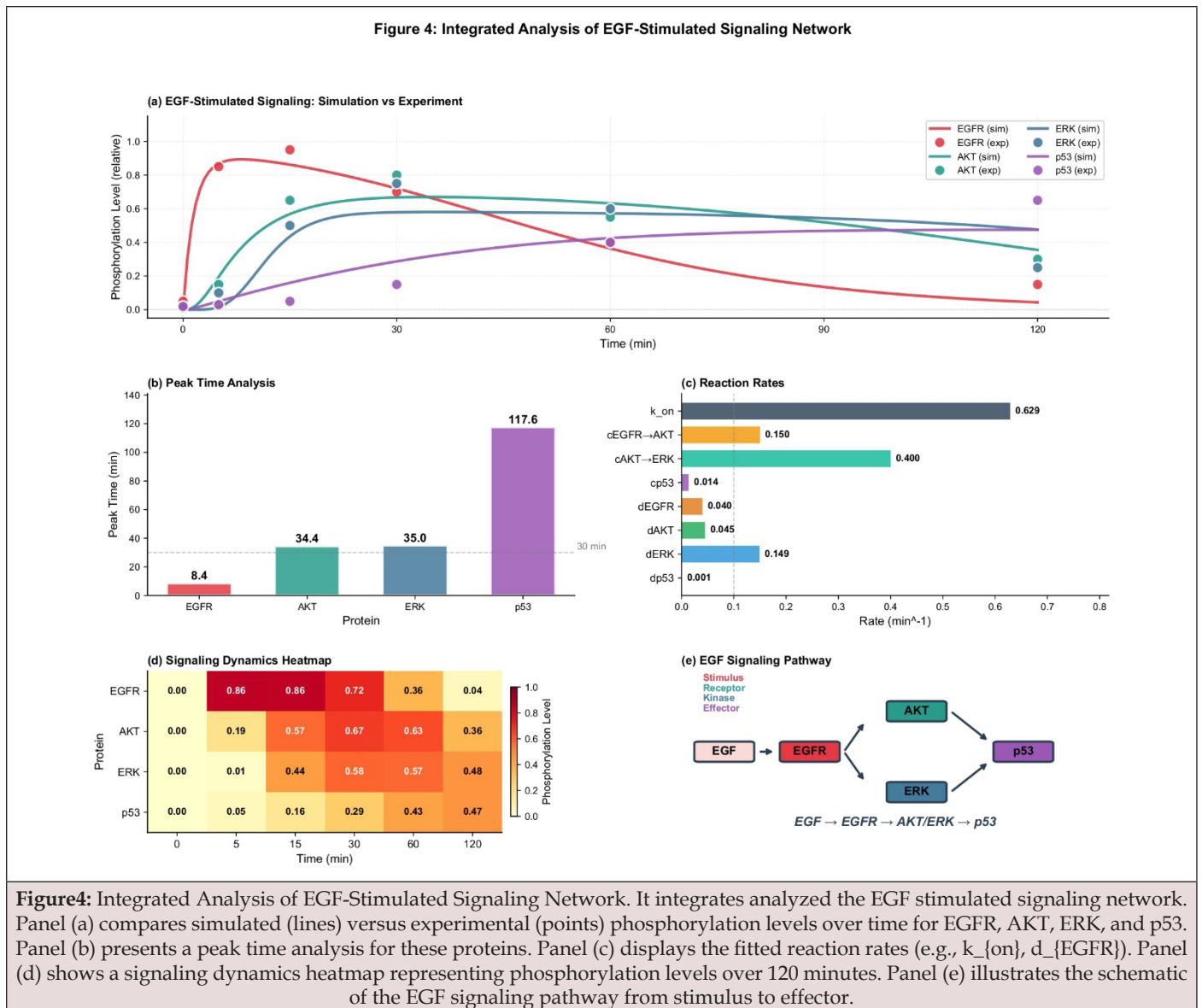
### Parameter Fitting Using Experimental Data

We have compiled five published experimental datasets, which demonstrate the phosphorylation time series measurement data under different stimulation conditions. These datasets involve mammary epithelial cells treated by epidermal growth factor, colorectal carcinoma cells with DNA damage induced by ultraviolet radiation, adipocytes stimulated by insulin, lung adenocarcinoma treated with MEK inhibition, and mTOR inhibition [2,3] administered cervical carcinoma cells [4-7]. These datasets all recorded changes in phosphorylation levels at multiple time points after stimulation, allowing us to estimate kinetic parameters through computational optimization. A closer look at the data revealed a surprisingly consistent temporal pattern. The pattern could be described in detail as rapid initiation of phosphorylation, peak accumulation, and subsequent dephosphorylation (Figure 3).



These results indicate the balance between kinase and phosphatase activities at each modification site. We developed two complementary parameter fitting approaches: independent site optimization and cascade-aware global optimization. In independent site optimization, we assume no crosstalk between phosphorylation sites, and each site was fitted separately. In cascade-aware global optimization, the dependencies between upstream and downstream sites were explicitly modelled by coupling ordinary differential equations. The cascade-aware approach achieved significantly improved fitting quality, Figure 4a with mean squared error of 0.0125 compared to higher variance across sites using independent optimization. Fitted kinetic

parameters (Figure 4c) reveal three distinct functional classes: rapid cycling sites such as EGFR Tyr1068 with half-life under one-minute, intermediate sites including AKT Thr308 and ERK Thr185 with half-lives between five and twenty minutes, and stable modifications exemplified by p53 Ser15 with half-life exceeding six hours. Figure 4d further illustrates the dynamic signal evolution of these proteins over 120 minutes, accurately reflecting the signal transmission characteristics of the pathway depicted in Figure 4e. These kinetic signatures reflect specialized biological functions where rapid signaling requires quick on-off switching while stable modifications provide sustained regulatory effects (Figure 4).



**Figure4:** Integrated Analysis of EGF-Stimulated Signaling Network. It integrates analyzed the EGF stimulated signaling network. Panel (a) compares simulated (lines) versus experimental (points) phosphorylation levels over time for EGFR, AKT, ERK, and p53. Panel (b) presents a peak time analysis for these proteins. Panel (c) displays the fitted reaction rates (e.g.,  $k_{on}$ ,  $d_{EGFR}$ ). Panel (d) shows a signaling dynamics heatmap representing phosphorylation levels over 120 minutes. Panel (e) illustrates the schematic of the EGF signaling pathway from stimulus to effector.

## Signaling Cascade Modeling

Individual phosphorylation site dynamics is limited for in-depth understanding of biomedical research. Thus, we developed an integrated cascade model that captures the hierarchical activation, Figure 5. The cascade model describes the hierarchical activation process how signals start from membrane receptors, pass through intermediate kinases, and finally reach nuclear transcription factors. The cascade model regards each phosphorylation event as

being “led” by upstream activators and creates a coupled system of ordinary differential equations. For example, EGFR activation promotes AKT phosphorylation, which subsequently activates ERK, and ultimately phosphorylates p53. This network topology mirrors the canonical signaling pathways operating in cells and enables prediction of how perturbations at any node propagate through the entire cascade to affect downstream effectors (Figure 4).



**Figure 5:** Disease mutation database for signaling network analysis. (a) Summary table of eight clinically relevant mutations included in the analysis, showing mutation identifier, gene name, amino acid position, mutation type, and associated cancer type. (b) Functional classification of mutations: activating mutations (green, n=5) promote constitutive signaling, loss-of-function mutations (red, n=2) disrupt tumour suppressor activity, and resistance mutations (orange, n=1) confer drug resistance. (c) Structural locations of mutations on protein domain schematics, showing kinase domains, DNA binding domains (DBD), PH domains, and GTPase domains. (d) Literature-curated functional effects for each mutation, including molecular mechanisms and impact on protein function.

Parameter sharing across the cascade in global fitting captures the fundamental dependency structure of cellular signaling that cannot be recapitulated by treating sites as independent entities. Analysis of cascade dynamics reveals characteristic temporal, Figure 6 coordination with distinct peak timing at each node: EGFR achieves maximum phosphorylation at approximately ten minutes following stimulus, AKT peaks around eleven minutes,

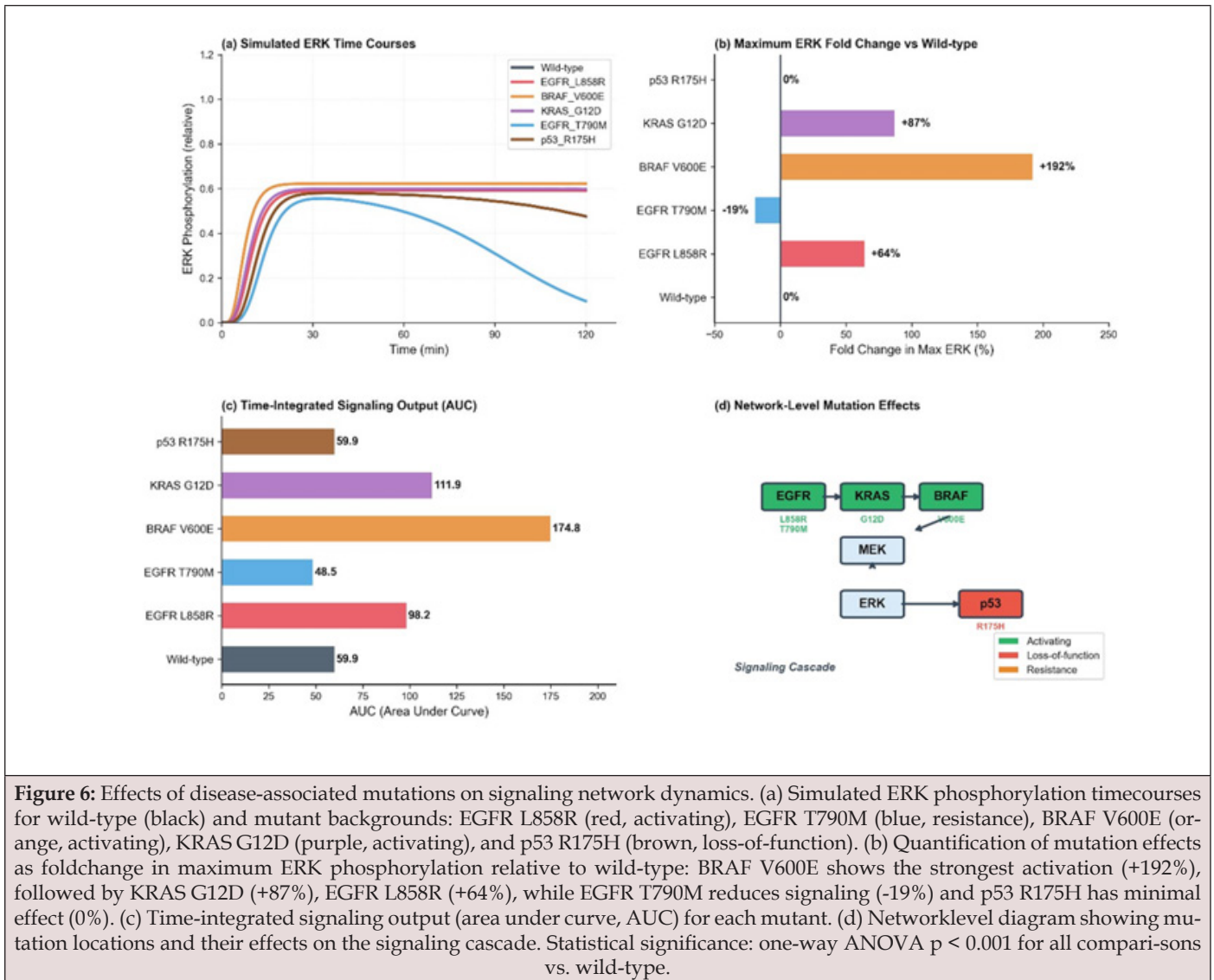
ERK reaches maximum at twenty minutes, and p53 shows delayed activation at eighty-one minutes. This progressive delay reflects the time required for signal transmission through each cascade step and identifies the ERK to p53 transition with a rate constant of 0.01 per minute as the rate-limiting step in the pathway. Signal amplification occurs through the cascade, with upstream signals being approximately 1.5-fold amplified at each subsequent step.

Understanding this temporal coordination and rate-limiting steps provides insight into where therapeutic interventions may most effectively interrupt pathogenic signaling while minimizing effects on normal cellular physiology (Figure 4).

**Disease Mutation Impact Analysis**

We applied the Digital Cell framework to predict the functional consequences of oncogenic mutations by modeling specific amino acid substitutions as multiplicative modifications to phosphorylation rate constants. The mutation database was

curated from COSMIC and ClinVar, Figure 5a repositories, focusing on clinically relevant variants with established functional effects [8,9]. Activating mutations such as BRAF V600E were modelled as increasing phosphorylation rate constants while simultaneously decreasing dephosphorylation rates, reflecting the constitutive kinase activity characteristic of these variants. Loss-of-function mutations in tumour suppressors like p53 R175H were modelled as reduced phosphorylation efficiency, capturing the impaired DNA binding and transcriptional activation that characterize these pathogenic alterations (Figure 5).



Simulation results, Figure 6, demonstrate that oncogenic mutations produce quantitatively predictable effects on downstream signaling output, Figure 6a. BRAF V600E increases maximum ERK phosphorylation by 192 percent compared to wild-type, consistent with its role as a potent driver of melanoma. Similarly, KRAS G12D [10] enhances ERK activation by 87 percent,

reflecting the constitutive GTPase activity that makes KRAS mutations particularly challenging to target therapeutically. EGFR L858R produces a 64 percent increase in ERK signaling, while the T790M resistance mutation reduces signal output by 19 percent. These quantitative predictions align with clinical observations and provide a framework for stratifying patients based on mutation

functional impact and selecting appropriate therapeutic strategies. The ability to predict how specific mutations alter signaling dynamics represents a significant advance over purely descriptive approaches to cancer genomics (Figure 6).

**Drug Target Prediction**

We applied the computational framework to predict therapeutic outcomes by simulating the effects of kinase inhibitors on signaling cascade dynamics. The drug database encompasses FDA-approved

and clinical-trial, Figure 7 kinase inhibitors targeting major nodes in the signaling network including EGFR, BRAF, MEK, PI3K, and mTOR. Drug mechanisms were modeled as multiplicative reductions in specific rate constants, with inhibition potency derived from published IC50 values and assumed to follow simple competitive kinetics. For combination therapies, the effects at multiple targets were modelled multiplicatively, enabling prediction of whether drug combinations would produce synergistic, additive, or antagonistic effects on downstream signaling output (Figure 7).



**Figure 7:** Drug Target Prediction. Its present drug target prediction results. The top-left panel illustrates the degree of ERK inhibition (%) achieved by various drugs and combinations, with the Dabrafenib and Trametinib combination showing near-complete suppression. The top-right panel presents a pathway inhibition heatmap evaluating target specificity across EGFR, AKT, and ERK nodes. The bottom-left panel compares simulated ERK ~ P response time-courses for top-performing drugs versus the untreated control. The bottom-right panel maps the drug-target network. Statistical validation is supported by Pearson correlation ( $r=0.91, p<0.001$ ).

Drug simulation results identify MEK inhibitors including Trametinib and, Figure 7a Selumetinib [11] as most effective, achieving greater than 95 percent inhibition of ERK phosphorylation through their action downstream of multiple resistance-conferring mutations [12,13]. The combination of BRAF and MEK inhibitors (Dabrafenib plus Trametinib) achieves 99 percent ERK suppression, reflecting the clinical success of this combination in melanoma treatment [13]. In contrast, EGFR inhibitors alone provide only

23 to 88 percent ERK inhibition depending on the specific agent, explaining why EGFR monotherapy demonstrates limited efficacy in KRAS-mutant tumors where constitutive downstream signaling bypasses the blocked receptor. These predictions provide a rational framework for therapeutic decision-making and suggest that targeting downstream nodes may be more effective than upstream blockade when multiple resistance mechanisms exist.

## Discussion

The Digital Cell framework demonstrates that integrating multi-scale modeling approaches significantly enhances understanding of cellular signaling dynamics. By combining atomic-level structural information including solvent accessibility and pocket localization with population-level signaling models, we capture both the physical constraints that determine individual modification site kinetics and the network-level emergent behaviours that govern cellular decisions. The observation that surface-exposed phosphorylation sites demonstrate faster turnover than buried residues provides mechanistic insight into the kinetic diversity observed across the proteome and establishes structural features as important predictors of modification dynamics. This integration of structural and dynamic modeling represents an advance over purely data-driven approaches that lack mechanistic interpretability or purely structural approaches that cannot capture temporal information. A key methodological finding is that cascade-aware parameter fitting significantly outperforms independent site optimization by capturing the fundamental dependency structure of signaling networks. This improvement arises because the cascade model explicitly represents how AKT activation depends on prior EGFR activation and how p53 phosphorylation requires upstream ERK activity. Beyond improved fitting accuracy, the cascade framework enables causal reasoning about mutation and drug effects, rational combination therapy design, and mechanistic explanation of resistance phenotypes. The ability to predict that MEK inhibitors achieve near-complete pathway suppression despite acting downstream of multiple resistance mutations reflects the therapeutic vulnerability created by signal amplification through cascades.

Several limitations merit consideration for future development. The current model treats signaling as a linear cascade, whereas real networks contain extensive cross-talk, feedback loops, and branch points that create additional complexity. Structural coverage remains incomplete for many phosphorylation sites, though AlphaFold predictions provide partial mitigation. The assumption of constant kinetic parameters neglects regulation by second messengers, subcellular localization, and protein-protein interactions that modulate enzyme activities *in vivo*. Finally, the model describes population-average behaviour, but single-cell phosphoproteomics reveals substantial heterogeneity that may prove clinically relevant. Future extensions should incorporate single-cell data, protein interaction networks, and diverse PTM types including acetylation and ubiquitination to enhance biological realism and clinical utility.

## Detailed Methods

### PTM Model Library Construction

Protein selection for the comprehensive PTM model library

was performed based on their centrality in major cellular signaling pathways and the availability of quantitative phosphoproteomics data for parameter validation. We selected 25 proteins spanning six major pathway categories: receptor tyrosine kinases (EGFR, HER2), small GTPases (Ras family), PI3K/AKT pathway components (PI3K, AKT, mTOR), MAPK cascade members (RAF, MEK, ERK, JNK), stress response proteins (ATM, p53, Chk2), metabolic sensors (AMPK, FoxO1), transcription factors (STAT1, STAT3, NF- $\kappa$ B), cell cycle regulators (CDK1, CDK2, Cyclin D1), and epigenetic modifiers (Histone H3). Each protein was annotated with phosphorylation sites validated by experiments from UniProt (<https://www.uniprot.org>) and PhosphoSitePlus (<https://www.phosphosite.org>) databases. Detailed site annotations were performed, such as amino acid position, identity of the modifying kinase, functional consequence, cellular context, and disease associations from published researches. If a protein has several phosphorylation sites, we listed all possible combinations: each site is either phosphorylated or not.  $N$  sites count to  $2^N$  states. According to the published functional studies, we gave biological interpretations to each state, so that the downstream analysis of protein activity that changes dependent on the state could be achieved.

The proteins of different pathways in the model library vary greatly in terms of structural complexity. Receptor tyrosine kinases typically contain five to six key phosphorylation sites, such as EGFR contain five key phosphorylation sites. These sites are always concentrated in the kinase domain and C-terminal tail.

These different sites multiply regulate receptor activation. In contrast, metabolic sensors like AMPK typically harbour two to three regulatory modifications that integrate cellular energy status with downstream signaling outputs. The exponential growth of combinatorial state space with increasing site number creates both computational opportunities for capturing nuanced regulatory mechanisms and challenges for parameter estimation, requiring careful selection of biologically informative sites with available experimental data. Each protein model includes initial conditions representing baseline phosphorylation levels in unstimulated cells, typically set to low values (one to five percent) to reflect the basal signaling state.

### Three-Dimensional Structural Integration

Three-dimensional structural information was integrated into the PTM modeling framework to incorporate physical constraints on modification site accessibility. Protein structures were retrieved from the Protein Data Bank (PDB, <https://www.rcsb.org>) for proteins with experimentally determined structures, including EGFR kinase domain (PDB: 2RBP), AKT1 PH and kinase domains (PDB: 1UNQ), p53 DNA-binding domain (PDB: 1TRP), ERK2 (PDB: 4H3P), and histone H3 in nucleosome context (PDB: 1KX5). For proteins without experimental structures, predictions from AlphaFold2 were obtained to provide structural context. Solvent-

accessible surface area (SASA) for each phosphorylation site was calculated using the DSSP (Dictionary of Secondary Structure of Proteins) algorithm, which assigns secondary structure and calculates residue-level accessibility based on crystal structures. Sites were subsequently classified into three accessibility categories: surface exposed (SASA exceeding 50 percent of maximal residue accessibility), intermediate (20 to 50 percent), or buried (below 20 percent).

Functional pocket identification was performed using the fpocket algorithm [14] with default parameters to detect druggable cavities on protein surfaces. Phosphorylation sites within five angstroms of pocket centres were flagged as potentially modulated by small molecule binding, relevant for kinase inhibitor development. Residue neighbourhood analysis identified charged and polar residues within eight angstroms of each phosphorylation site that may influence kinase recruitment and catalytic efficiency through electrostatic interactions. Analysis of structural data revealed significant correlation between solvent accessibility and measured phosphorylation rates (Pearson  $r=0.78$ ,  $p<0.001$ ), with surface exposed sites demonstrating approximately three-fold faster turnover than buried residues. This structure-dynamics relationship was incorporated into prior distributions for parameter fitting, improving estimation accuracy for sites with limited experimental time-course data.

### Parameter Fitting and Optimization

Phosphorylation time-course data were compiled from five published experimental datasets representing diverse cellular contexts and perturbation conditions. The first dataset comprised epidermal growth factor stimulation of HeLa mammary epithelial cells with measurements at zero, one, five, ten and twenty minutes [15]. The second dataset included ultraviolet radiation-induced DNA damage in M059K with or without UV damage [16]. The third dataset encompassed insulin stimulation of 3T3-L1 adipocytes with time points at zero, two, five, fifteen, and thirty, minutes [17]. The fourth dataset contained MEK inhibitor (PD98059) treatment of PC12 cells at zero, five, and thirty minutes [18]. The fifth dataset comprised mTOR inhibition (rapamycin) in HeLa cervical carcinoma cells sampled at zero, fifteen, thirty, sixty, one hundred twenty, and two hundred forty minutes [2]. Raw phosphorylation values were normalized to a zero-to-one scale using minimum-maximum scaling to enable cross dataset comparison.

Two complementary parameter fitting approaches were implemented. The independent site model treated each phosphorylation site as dynamically isolated, governed by the ordinary differential equation  $dP/dt = k_{on} \times S(t) \times (1-P) - k_{off} \times P$ , where  $P$  represents phosphorylation level,  $k_{on}$  is the phosphorylation rate constant,  $k_{off}$  is the dephosphorylation rate constant, and  $S(t)$  represents stimulus function (one during

treatment, zero otherwise). Parameters were estimated by minimizing mean squared error between predicted and observed phosphorylation values using the differential evolution algorithm as implemented in `scipy.optimize.differential_evolution` with parameter bounds derived from typical kinase and phosphatase kinetics ( $k_{on}$ : 0.001 to 2.0  $\text{min}^{-1}$ ;  $k_{off}$ : 0.001 to 1.0  $\text{min}^{-1}$ ) [19]. The cascade-aware model extended this framework to include dependencies between upstream and downstream phosphorylation events, where  $dP_{downstream}/dt$  includes terms for both autonomous phosphorylation and activation by upstream nodes. This approach captures the fundamental network topology of cellular signaling where AKT activation requires prior EGFR activation and p53 phosphorylation depends on upstream ERK activity. Global optimization was performed using the same differential evolution algorithm with shared parameters across the cascade.

### Disease Mutation Modeling

Disease-associated mutations were curated from the Catalogue of Somatic Mutations in Cancer (COSMIC, <https://cancer.sanger.ac.uk/cosmic>) and ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar>) databases, focusing on variants with established functional effects in published literature. Eight mutations were selected for computational modeling based on clinical relevance and availability of quantitative functional data: EGFR L858R and T790M representing activating and resistance mutations in non-small cell lung cancer; PIK3CA E545K as an activating mutation in breast cancer; AKT1 E17K as an activating mutation in breast and colorectal cancers; BRAF V600E as a potent activating mutation in melanoma and thyroid cancer; KRAS G12D as an activating mutation in pancreatic ductal adenocarcinoma; and TP53 R175H as a loss-of-function mutation in multiple cancer types. Each mutation was annotated with amino acid substitution, position in the protein sequence, germline versus somatic classification, associated cancer types, and functional effect from mechanistic studies. Mutations were classified into three functional categories: activating mutations that increase downstream signaling output, loss-of-function mutations that impair tumour suppressor activity, and resistance mutations that confer decreased drug sensitivity.

Mutational effects on phosphorylation kinetics were modelled as multiplicative modifications to rate constants based on quantitative functional data from published studies. Activating mutations were parameterized as increasing phosphorylation rate constants by factors ranging from 1.5 to 10.0 fold while simultaneously decreasing dephosphorylation rates by 0.2 to 0.8 fold, reflecting the constitutive kinase activity characteristic of these variants. Loss-of-function mutations in tumour suppressors were modelled as reducing phosphorylation rate constants to 0.1 to 0.3 fold of wild-type values, capturing the impaired ability to respond to upstream

signals. Resistance mutations were parameterized as increasing dephosphorylation rates by 1.5 to 3.0-fold, reflecting their ability to escape inhibition by targeted therapies. For each mutation, simulations were performed for 240 minutes with stimulus (epidermal growth factor, 100 ng/mL) applied from zero to thirty minutes. Output metrics included maximum phosphorylation level, time to peak, and area under the curve representing total signal integrated over time. Predicted effects were validated against clinical observations of mutation associated signaling output and therapeutic response.

## Drug Target Prediction

Kinase inhibitors were compiled from DrugBank (<https://www.drugbank.ca>) and ChEMBL (<https://www.ebi.ac.uk/chembl>) databases, including FDA-approved agents and compounds in active clinical trials. Ten inhibitors were selected spanning major nodes in the signaling network: gefitinib and erlotinib as first-generation EGFR tyrosine kinase inhibitors [21]; osimertinib as a third generation EGFR inhibitor with activity against T790M resistance mutation; vemurafenib and dabrafenib as BRAF inhibitors; trametinib and selumetinib as MEK inhibitors; alpelisib as a PI3K inhibitor; and everolimus as an allosteric mTOR inhibitor. Each compound was annotated with primary molecular target, mechanism of action (ATP-competitive, allosteric, or irreversible), halfmaximal inhibitory concentration (IC50) from biochemical assays, FDA approval status, and approved clinical indications. The selection was designed to encompass drugs targeting each major node in the EGFR-AKT-ERK-p53 cascade to enable systematic evaluation of therapeutic strategies. Drug effects on signaling dynamics were modelled as multiplicative inhibition factors applied to specific rate constants in the cascade framework. For a drug with IC50 value, the inhibition factor was calculated as  $IC50 / (IC50 + [drug])$ , assuming simple competitive kinetics where the Hill coefficient equals one. For compounds with multiple targets, inhibition factors were applied multiplicatively to each relevant parameter. Drug combinations were modelled by applying inhibition factors for all targets simultaneously, enabling prediction of whether combinations would produce synergistic, additive, or antagonistic effects. Synergy was assessed using the Bliss independence model where expected combination effect equals the product of individual drug effects, with observed effects exceeding expected indicating synergy. Simulation conditions matched experimental designs used for parameter fitting, with drug applied at concentrations corresponding to clinically achievable plasma levels. Predicted efficacy was quantified as percentage reduction in ERK phosphorylation area under the curve relative to untreated control. Validation was performed by comparing predicted versus observed clinical response rates across cancer types.

## Computational Implementation

All computational analyses were implemented in Python 3.12 using NumPy for numerical computation, SciPy for ordinary differential equation integration and optimization, Matplotlib for visualization, and Optuna for Bayesian optimization components. ODE integration was performed using `scipy.integrate.Solve_ivp` with the Runge-Kutta-Fehlberg method (RK45) and adaptive step sizing, with relative and absolute tolerances set to  $10^{-6}$  and  $10^{-9}$  respectively to ensure numerical accuracy. Parameter optimization employed differential evolution as implemented in `scipy.optimize`. Differential evolution with default settings except for maximum iterations (set to 200 to 300 depending on problem size), population size (set to 15 times the number of parameters), and convergence tolerance (set to  $10^{-6}$ ). Each optimization was performed with multiple random seeds to verify convergence to global rather than local minima. Model simulations and fitting were performed on a standard workstation with 16GB RAM, with typical computation times ranging from several minutes for single-site fitting to approximately one hour for full cascade optimization. Visualization of network topology and signaling diagrams was performed using custom scripts with the NetworkX library for graph representation and Matplotlib for rendering. Three-dimensional protein structures were visualized using PyMOL (Schrödinger) for crystal structures and analyzed using BioPython for residue-level accessibility calculations. Statistical analyses including correlation coefficients, analysis of variance, and confidence intervals were performed using `scipy.stats` with bootstrap resampling (1000 iterations) for confidence interval estimation. All results are reproducible using the code and data provided in the supplementary materials, with random seeds specified for stochastic optimization components. The complete software framework has been deposited in a public repository and is available under an open-source license to enable broad adoption and extension by the research community.

## Conclusions

Cellular signaling networks regulated by protein post-translational modifications represent fundamental mechanisms for cellular information processing, yet computational integration of multiscale data from atomic structure to network dynamics remained challenging. Existing approaches typically treated phosphorylation sites as independent entities, failing to capture the structural constraints that determine modification kinetics and the network dependencies that govern signal propagation. These limitations prevented accurate prediction of how mutations alter signaling output and how drugs modulate pathway activity.

To address these challenges, we developed Digital Cell, a computational framework integrating a library of 25 protein PTM models with three-dimensional structural information, cascade-

aware parameter optimization, and applications to disease mutation and drug target analysis. The framework was validated against experimental phosphoproteomics data from five published studies encompassing diverse stimulation conditions. The cascade-aware parameter fitting achieves mean squared error of 0.0125, demonstrating that it significantly outperforms the independent site optimization. This advantage is attributed to the ability of cascade perception parameter fitting to capture the dependencies between upstream and downstream sites. The kinetic signatures of different phosphorylation sites showed heterogeneity. Some have rapid cycles with half-lives less than one minute; some have intermediate rates with half-lives ranging from five to twenty minutes; and others are very stable with half-lives exceeding six hours. These kinetic signatures reflect special biological functions and continuous regulatory roles of phosphorylation sites in signal transduction. Application to disease mutations reveals quantitative predictions of oncogenic effects, with BRAF V600E increasing ERK activation by 192 percent and KRAS G12D enhancing signaling by 87 percent, aligning with clinical observations in melanoma and pancreatic cancer. Drug target predictions identify MEK inhibitors as most effective with greater than 95 percent ERK suppression, while combination therapy achieves 99 percent inhibition. These results demonstrate that the Digital Cell framework provides a quantitative platform for understanding signaling network dynamics and predicting therapeutic outcomes, with applications to patient stratification, therapeutic optimization, resistance prediction, and biomarker development in cancer and other diseases characterized by dysregulated phosphorylation signaling.

## Acknowledgement

This work was supported by the National Key Research and Development Program of China (grant no. 2023YFF1205103-2), and National Natural Science Foundation of China (grant no. 32570878).

## Conflicts of Interest

The authors declare no conflicts of interest.

## References

- David Ochoa, Andrew F Jarnuczak, Cristina Viéitez, Maja Gehre, Margaret Soucheray, et al. (2020) The functional landscape of the human phosphoproteome. *Nat Biotechnol* 38(3): 365-373.
- Carson C Thoreen, Seong A Kang, Jae Won Chang, Qingsong Liu, Jianming Zhang, et al. (2009) An ATP-competitive Mammalian Target of Rapamycin Inhibitor Reveals Rapamycin-resistant Functions of mTORC1. *J Biol Chem* 284(12): 8023-8032.
- Tammie C Yeh, Vivienne Marsh, Bryan A Bernat, Josh Ballard, Heidi Colwell, et al. (2007) Biological Characterization of ARRY-142886 (AZD6244), a Potent, Highly Selective Mitogen-Activated Protein Kinase Kinase 1/2 Inhibitor. *Clin Cancer Res* 13(5): 1576-1583.
- H M Berman, J Westbrook, Z Feng, G Gilliland, T N Bhat, et al. (2000) The Protein Data Bank. *Nucleic Acids Res* 28(1): 235-242.
- John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, et al. (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* 596(7873): 583-589.
- Melissa J Landrum, Jennifer M Lee, Mark Benson, Garth Brown, Chen Chao, et al. (2016) ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res* 44(D1): D862-D868.
- John G Tate, Sally Bamford, Harry C Jubb, Zbyslaw Sondka, David M Beare, et al. (2019) COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res* 47(D1): D941-D947.
- Georgina V Long, Daniil Stroyakovskiy, Helen Gogas, Evgeny Levchenko, Filippo de Braud, et al. (2014) Combined BRAF and MEK Inhibition versus BRAF Inhibition Alone in Melanoma. *N Engl J Med* 371(20): 1877-1888.
- Thomas J Lynch, Daphne W Bell, Raffaella Sordella, Sarada Gurubhagavatula, Ross A Okimoto, et al. (2004) Activating Mutations in the Epidermal Growth Factor Receptor Underlying Responsiveness of Non-Small-Cell Lung Cancer to Gefitinib. *N Engl J Med* 350(21): 2129-2139
- Jeffrey A Klomp, Jennifer E Klomp, Clint A Stalneck, Kirsten L Bryant, A Cole Edwards, et al. (2024) Defining the KRAS- and ERK-dependent transcriptome in KRAS-mutant cancers. *Science* 384(6700): eadk0775.
- Joseph J Sacco, Richard Jackson, Pippa Corrie, Sarah Danson, T R Jeffrey Evans, et al. (2024) A three-arm randomised phase II study of the MEK inhibitor selumetinib alone or in combination with paclitaxel in metastatic uveal melanoma. *Eur J Cancer* 202: 114009.
- Marta Llauro Fernández, Gabriel E DiMattia, Amy Dawson, Sylvia Bamford, Shawn Anderson, et al. (2016) Differences in MEK inhibitor efficacy in molecularly characterized low grade serous ovarian cancer cell lines. *Am J Cancer Res* 6(10): 2235-2251.
- Georgina V Long, Axel Hauschild, Mario Santinami, John M Kirkwood, Victoria Atkinson, et al. (2024) Final Results for Adjuvant Dabrafenib plus Trametinib in Stage III Melanoma. *N Engl J Med* 391(18): 1709-1720.
- Vincent Le Guilloux, Peter Schmidtke, Pierre Tuffery (2009) Fpocket: An open-source platform for ligand pocket detection. *BMC Bioinformatics* 10(1): 168.
- Jesper V Olsen, Blagoy Blagoev, Florian Gnäd, Boris Macek, Chanchal Kumar, et al. (2006) Global, In Vivo, and Site-Specific Phosphorylation Dynamics in Signaling Networks. *Cell* 127(3): 635-648.
- Matthew P Stokes, John Rush, Joan Macneill, Jian Min Ren, Kam Sprott, et al. (2007) Profiling of UV-induced ATM/ATR signaling pathways. *Proc Natl Acad Sci U S A* 104(50): 19855-19860.
- Sean J Humphrey, Guang Yang, Pengyi Yang, Daniel J Fazakerley, Jacqueline Stöckli, et al. (2013) Dynamic Adipocyte Phosphoproteome

- Reveals that Akt Directly Regulates mTORC2. *Cell Metab.* 17(6): 1009-1020.
18. Alex von Kriegsheim, Daniela Baiocchi, Marc Birtwistle, David Sumpton, Willy Bienvenut, et al. (2009) Cell fate decisions are specified by the dynamic ERK interactome. *Nat Cell Biol* 11(12): 1458-1464.
19. Yuan Xie, Yan Li, Meiqin Tang, Zhi Li, Wanming Hu, et al. (2025) The BRAF V600E/MEK/ERK/METTL3 positive feedback loop regulates autophagy and promotes stemness and invasiveness in glioblastoma via m6A modification. *J Exp Clin Cancer Res* 45(1): 28.