Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Cell Biology

Network structures and parameters in multiscale modeling in ErbB signaling networks

Ai Shinobu[1], Ayaka Nagasato-Ichikawa[2] and Mariko Okada[2,3]

Signal transduction is a complex system governing cellular behavior across physiological and pathological contexts. Advances in systems biology have positioned cell modeling as a powerful tool for reconstructing the dynamics and trajectories of disease processes. Nevertheless, despite progress in AI-assisted model generation, parameter estimation remains a challenge, especially under data constraints. In contrast, molecular dynamics simulations offer crucial, high-resolution insights by uncovering conformational activation mechanisms and by extracting kinetic parameters; however, they face scalability limitations. This review focuses on modeling of the ErbB signaling system, highlighting recent advances at both the cellular and molecular scales. Emerging trends, such as simulation data reuse, machine learning-guided network inference, and modeling within realistic environmental contexts, are now driving a compelling integration of these molecular and cellular modeling paradigms.

Addresses
[1] Premium Research Institute for Human Metaverse Medicine (WPI-PRIMe), The University of Osaka, Suita, Osaka 565-0871, Japan
[2] Laboratory for Cell Systems, Institute for Protein Research, The University of Osaka, Suita, Osaka 565-0871, Japan
[3] Department of Bioengineering, Imperial College London, London, SW7 2AZ, United Kingdom

Corresponding author: Okada, Mariko (mokada@protein.osaka-u.ac.jp)

Current Opinion in Cell Biology xxxx, **98**:xxx

This review comes from a themed issue on **Cell Signalling (2026)**

Edited by **Alexis Barr** and **Ana Cuenda**

For a complete overview see the Issue and the Editorial

Available online xxx

https://doi.org/10.1016/j.ceb.2025.102603

Introduction

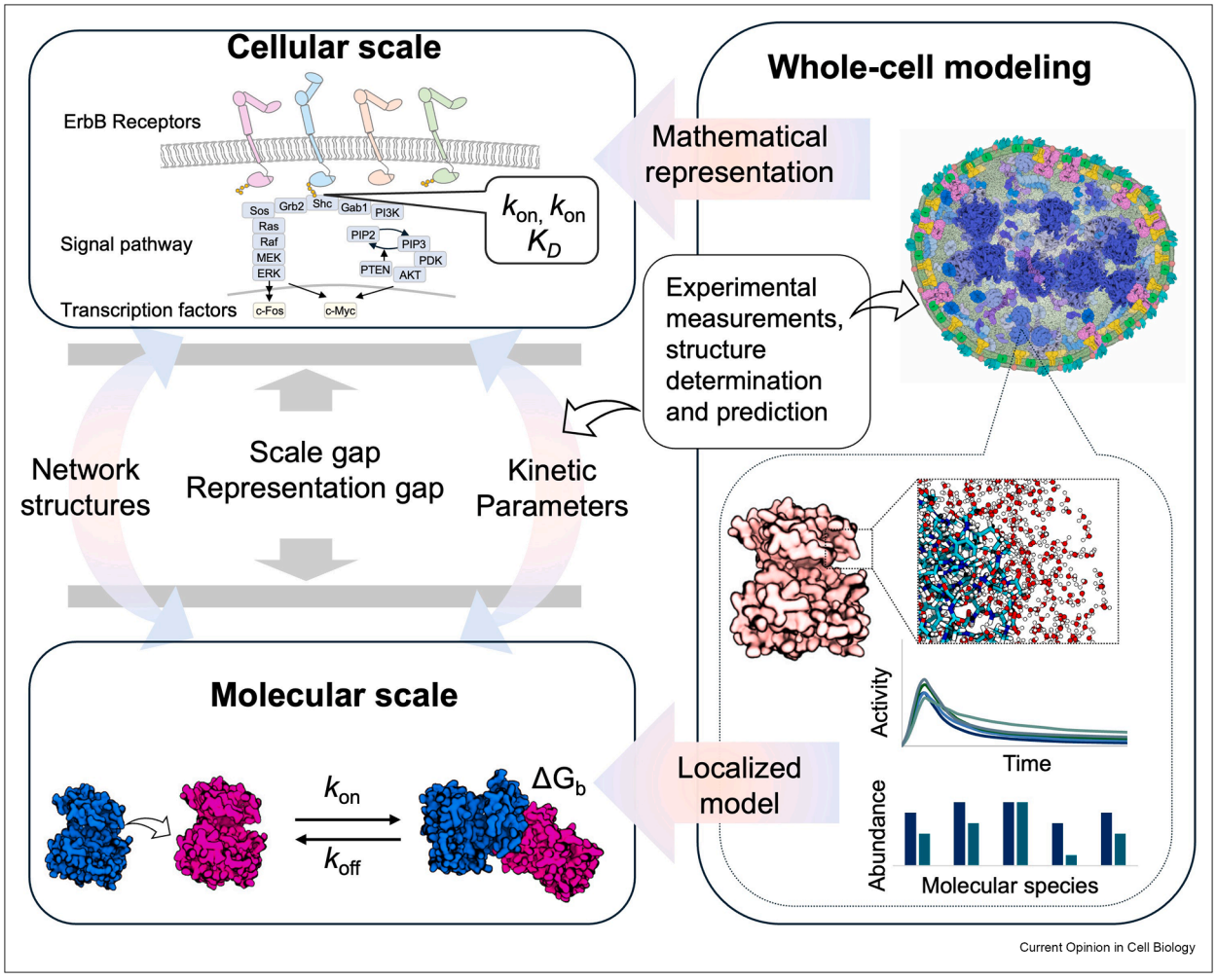
Signal transduction systems consist of intricately interconnected biochemical reaction networks modulated by post-translational modifications, protein interactions, and enzymatic activities. These systems play a crucial role in maintaining physiological homeostasis by dictating cell fate. Consequently, disruptions in these networks are closely associated with the onset and progression of numerous human diseases. To gain quantitative insight into these complex processes, research has increasingly focused on characterizing their activation dynamics.

Activation dynamics are unique to individual cells or tissues, governed by multi-scale regulatory mechanisms. This often spans from small-scale changes, such as protein phosphorylation and dephosphorylation, to global-scale modifications including transcriptionally mediated positive and negative feedback regulations [1,2]. Despite the rise of machine learning (ML) and AI technologies, mechanistic modeling with ordinary differential equations (ODEs) remains essential for understanding the underlying mechanisms of complex biological systems, particularly for reconstructing disease trajectories and predicting cell fate decisions.

Mathematical models depend critically on prior knowledge of the regulatory network structure, molecular initial values (concentration), and precise kinetic parameters, such as binding rate constants (k_{on}, k_{off}), interaction strengths (K_D), and enzyme kinetics (K_m, V_{max}) to accurately capture dynamics in numerical simulations [3,4] (Figure 1). However, these parameters are challenging to acquire comprehensively through experiments. The common practice of defining a simplified network structure and then estimating parameters via statistical methods or genetic algorithms requires extensive time-course data (e.g., phosphorylation patterns), which are often scarce in public databases. Furthermore, the identification of network structure itself is difficult, as biological networks are highly cell- and tissue-dependent.

In contrast, structure-based computational approaches, primarily molecular dynamics (MD) simulations, can estimate critical kinetic and thermodynamic parameters needed by cell-based models, including binding free energies (ΔG_b), binding and dissociation rate constants (k_{on}, k_{off}, respectively), and residence times (RT) [5]- based on high-resolution structural information of target proteins. However, MD applications face distinct challenges. Many signaling proteins lack complete structural data, such as those containing intrinsically disordered regions (IDRs) [6], and their membrane association complicates interpretations of

Figure 1



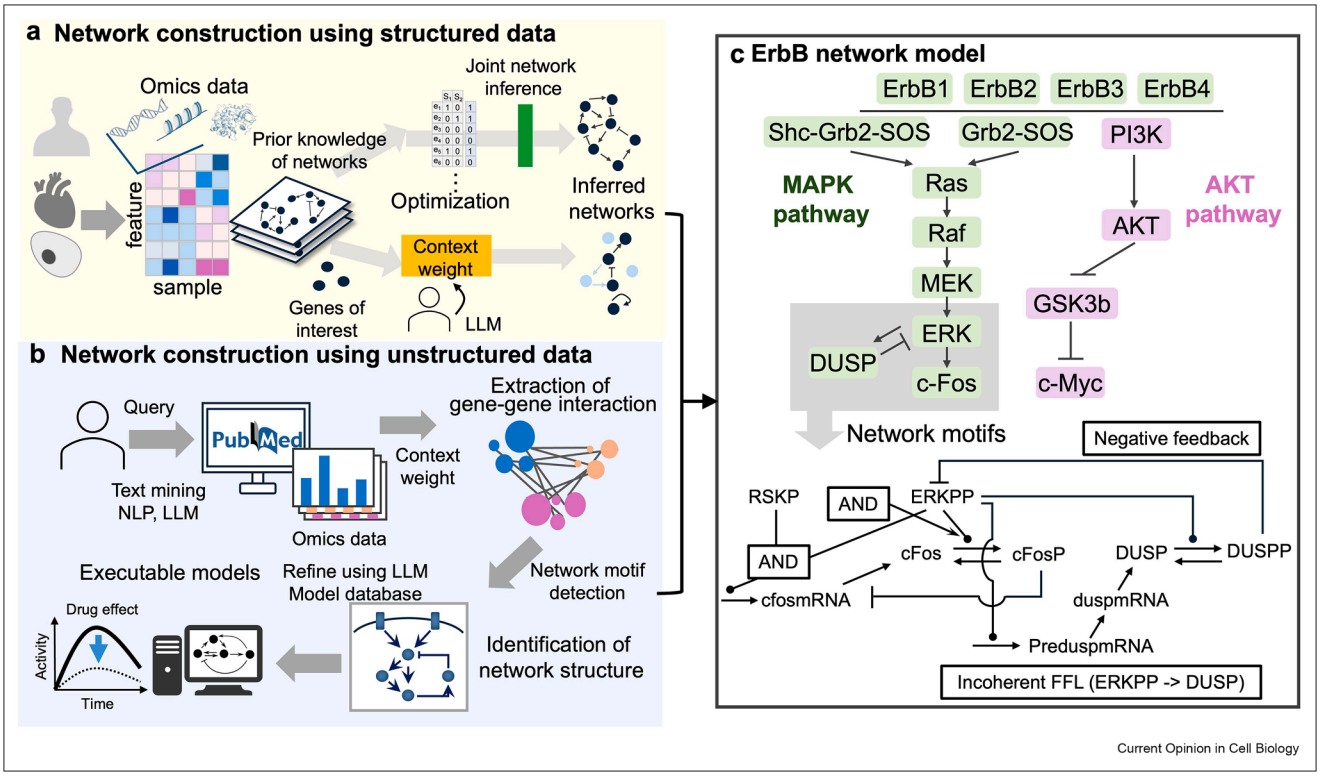
Conceptual framework connecting cellular- and molecular-scale approaches for multiscale analysis of signaling systems. At the cellular scale (**top left**), models based on ordinary differential equations (ODEs) provide mathematical representations of signaling networks, capturing kinetic relationships among molecular components. At the molecular scale (**bottom left**), simulations such as molecular dynamics (MD) generate localized models that describe individual protein interactions and conformational dynamics. The two modeling layers are connected through the exchange of kinetic and network information, while experimental data and structure-prediction methods provide external inputs supporting both levels. Together, they represent complementary approximations of biological complexity, and their integration provides a foundation for developing whole-cell models (**right**) that unify molecular and systems-level behavior.

cell- and tissue-dependent interactions. Current MD simulations also struggle to capture slow binding and activation processes observed experimentally.

These bottlenecks are, however, being gradually solved by implementing emerging computational tools. Structural prediction tools like AlphaFold [7] are a promising approach to modeling full-length proteins, while algorithmic advances are expanding the timescale and systems size accessible to MD. In parallel, natural language processing (NLP) and ML techniques are aiding the context-dependent reconstruction of biological networks for ODE and other models (Figure 2).

In this article, we focus on the ErbB receptor signaling system, which is implicated in numerous human diseases. We provide an overview of recent advances in both molecular- and cellular-level simulations, specifically addressing persistent challenges of parameter estimation and network identification. Ultimately, we explore how the integration of these multiscale

Figure 2



Integrating knowledge database and NLP to model signaling networks. **(a)** Structured data from knowledge database, combined with ML and deep learning (DL) techniques, is applied to omics data, including transcriptome and proteome data, to identify or prioritize key genes in the network. **(b)** Unstructured data, such as literature information from PubMed, can be analyzed using text mining, NLP or large language models (LLMs) to extract gene–gene interaction for network construction and model development. The resulting network structure can be further refined using LLMs or ML/DL to develop executable, context-dependent models. **(c)** An example illustrates a ErbB signaling network includes MAPK and AKT pathways with network motifs.

simulations offers crucial mechanistic insights into the signaling network and its role in human diseases.

Cellular modeling of ErbB receptor signaling

Mathematical models based on ODEs have long been used to recapitulate the complex regulation of signal transduction systems. By encoding network structures, initial conditions, and parameters, these models aim to reconstruct system dynamics and uncover underlying control principles. However, constructing such models typically assumes a practically known network structure. Conversely, there is growing demand for modeling approaches even when the network architecture is not predefined or needs to be simplified without compromising predictive accuracy. This is especially true in the quantitative systems pharmacology (QSP) modeling [8]. To address this challenge, data-driven approaches for network inference using omics data and literature information (also called as bibliome [9]) have been actively explored in recent years (Figure 2a and b).

Knowledge database and large language models (LLMs) for network and model reconstruction

A new direction in modeling is the development of automated frameworks that extract molecular interactions to systematically reconstruct context-dependent networks and executable models. One such example is CORNETO (Constrained Optimization for the Recovery of NETworks from omics) [10], which generates unified biological networks by integrating prior knowledge from databases with various omics datasets. This method optimizes network flow (e.g., favoring efficient paths) across the unified network while simultaneously identifying sample-specific subnetworks. CORNETO's strength lies in its ability to perform joint inference, making it particularly effective for inferring the best propagation pathways even when data is limited (e.g. only the receptor and a downstream transcription factor).

Another similar automated network inference method using knowledge databases is the Python package Neko [11]. Given a list of molecular entities of interest (e.g.,

differentially expressed genes (DEGs) or proteins) and a predefined interaction source (e.g., public database), NeKo allows users to select various strategies to connect molecules and genes, automatically generating Boolean equations for each entity in the network.

Similarly, Erdem et al. created SPARCED [12], a Python-based framework capable of building large-scale ODE models, which covers pathways from epidermal growth factor (EGF) and interferon receptor signaling to cell cycle and apoptosis, using minimal text input. SPARCED integrates existing models from public databases and leverages gene expression data to predict drug responses. These automated pipelines are particularly useful for experimental scientists who need to further analyze their experimental data and gain the mechanistic insights.

While knowledge bases like OmniPath [13] and SIGNOR [14] are useful for extracting predefined gene interactions for omics data mapping, they are ineffective when interactions are not yet registered. In such cases, extracting interactions from literature using NLP technologies becomes necessary. Tsutsui et al. developed a framework using the BERT model (pretrained with whole PubMed information) to extract context-dependent gene regulatory networks (GRNs), weighted by MeSH-defined disease, through network embedding with DEGs [15]. These context-weighted GRNs improved drug response prediction and facilitated model development. The group's associated tool, Text2Model [16,17], further streamlines the process by automatically converting textual gene interaction types in GRNs (or KEGG networks) into executable ODE models of ErbB receptor signaling (Figure 2c). A parallel development involves using transformer models to predict gene−drug interactions; for example, Yamagiwa et al. [18] leveraged BioConceptVec [19] and custom embeddings for drug target prediction in ErbB network. Crucially, a yearly analysis showed this method successfully predicted future drug targets in addition to known interactions, underscoring its potential for integration with omics data in drug repositioning.

Motifs and network refinement

When gene interactions are sourced from knowledge bases or literature without constraints on architecture, the resulting GRNs or mathematical models can be large, sparse, and disconnected. However, biological networks such as signaling and transcriptional networks often include recurring network motifs that operate at different temporal scales [20] (Figure 2c). These motifs are crucial for characterizing the dynamic behaviors of the network itself.

Signaling pathways involving transcription factors, for instance, often contain multiple regulatory motifs such as positive and negative feedback, feedforward loops (FFLs), or AND-gates [21] (Figure 2c). These regulatory motifs are the hallmarks of the growth signaling circuits that specify cellular outcomes. Strikingly, the self-sustained ErbB signaling pathway is commonly exploited as a mechanism for drug resistance and metastatic cell survival [22]. The automated identification of such motifs within data-driven network reconstructions allows for predictions of dynamic features prior to simulation by leveraging the functional characteristics associated with specific motifs. For example, Kadelka et al. [23] systematically evaluated and compiled a database of network motifs (like FFLs) from previously curated Boolean network models, offering a foundation for omics or literature-driven modeling approaches.

With the recent advances in cell-based modeling, which build upon structural identification methods, accurate network identification (structural identifiability) has become a prerequisite for executable model development, as it is strongly associated with parameter identifiability, the ability to determine model parameters from experimental data [24,25]. Therefore, the improvement of existing parameter estimation methods and the development of new ones will become increasingly critical for the automated modeling and simulation of biological networks.

Parameter inference problem and solution in context-dependent modeling

A serious problem in parameter estimation for signaling pathway ODE models is that the number of available time-series or dose−response data needed for parameter fitting is typically far smaller than the number of parameters in the model. Consequently, extensive efforts have focused on developing strategies for estimating parameters from limited experimental data.

The group led by Hasenauer developed a large-scale model of ErbB receptor signaling that includes over 4,100 parameters [26]. By integrating public drug response datasets from more than 100 cancer cell lines and applying their own methods, they achieved efficient parameter estimation and successfully predicted combinatorial drug responses. Furthermore, acknowledging that different experimental techniques yield data with varying quantitative accuracy, they proposed a spline-based parameter estimation method tailored for semi-quantitative datasets [27].

Another modeling approach focuses on extracting shared parameters by fitting experimental data from distinct cancer cell types. This method is founded on the assumption that molecular interactions and enzyme activities are invariant across various cell types and human tissues, allowing the parameters to be used for patient-specific modeling. For example, Imoto et al. [16] applied a shared parameter set across multiple cancer subtypes and patients, with patient-specific

dynamics arising solely from differences in gene expression values. The study confirmed the finding, first proposed by an earlier study [28], that *in silico* signaling dynamics are better prognostic biomarkers than the original gene (or protein) expression values [29]. The study further proposed the basis for developing multiscale modeling that explicitly links cellular to tissue dynamics through molecular parameters.

Molecular modeling and simulation of ErbB receptors

The previous section discussed automated construction of cell-based models and the persistent challenges of structural and parameter identifiability. Here, we focus on molecular-scale approaches. Structure-based modeling and simulation methods such as MD simulations contribute to the study of signaling systems in two complementary ways: by providing quantitative parameters for network models, and by elucidating molecular mechanisms of the underlying signaling. These molecular contributions are directly relevant to the cellular-level identifiability challenges, as quantitative descriptors of binding and dynamics are necessary for constructing reliable cell-level models, and mechanistic insights define which molecular states or interactions should be represented in the first place (Figure 1).

Quantitative parameters from molecular models

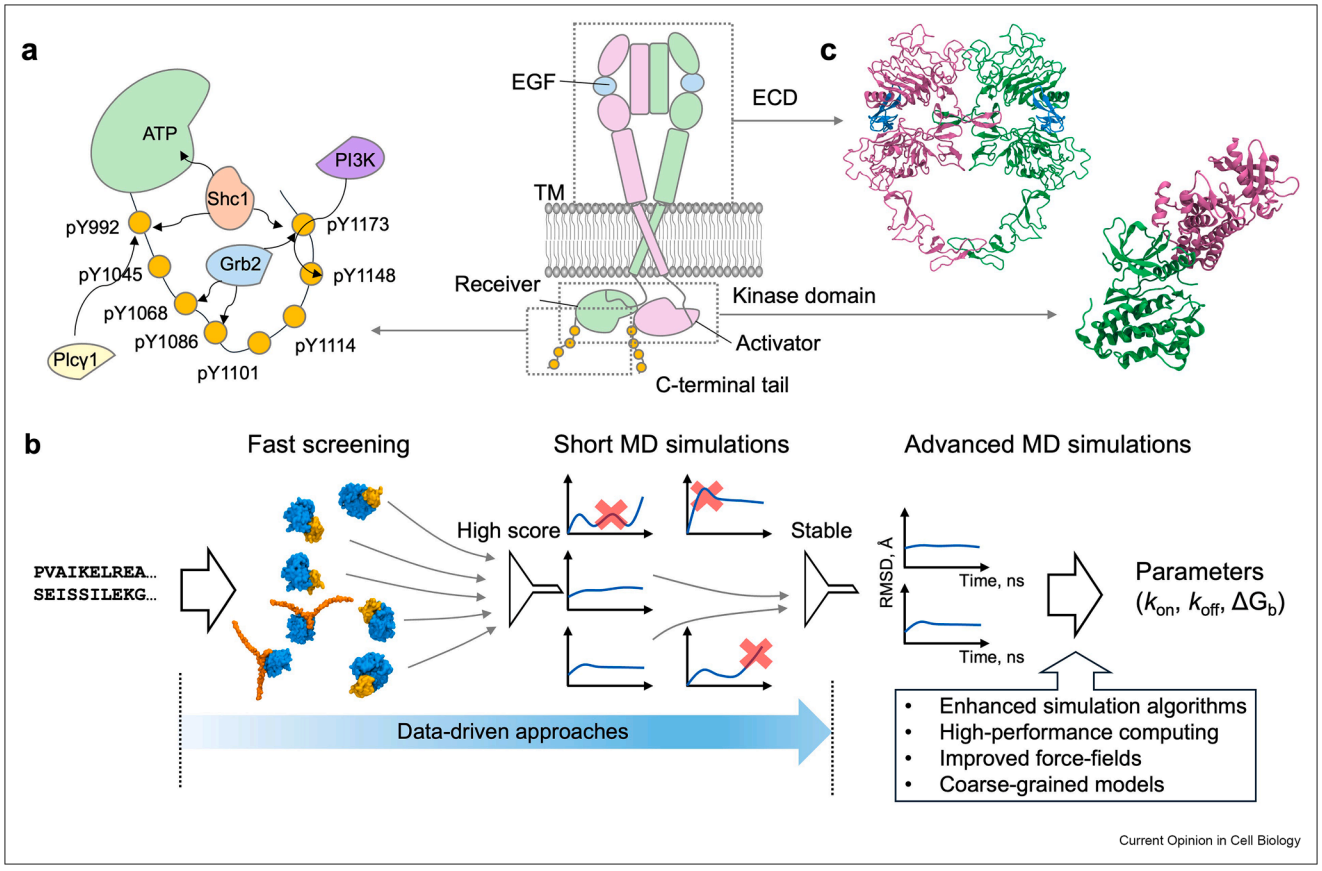
Cellular-level models rely heavily on kinetic rate constants (k_{on}, k_{off}), to accurately capture how signaling interactions evolve over time. While direct calculation of these rates from MD simulations remains difficult due to the short timescales MD can simulate, binding affinities (K_D, ΔG_b) can be used as proxies ($K_D = k_{off}/k_{on}$, and $\Delta G_b = RT \ln K_D$). Physics-based methods for calculating free-energies are now relatively well established [30], and newer ML approaches are beginning to provide faster alternatives [31]. However, a complete network representation still requires explicit determination of both k_{on} and k_{off}. Recent studies have demonstrated their estimation in model systems [32], yet extending these approaches to complex, multi-component interactions under cellular conditions remains a major challenge.

While there has been progress in extracting parameters from molecular simulations, most results remain limited to isolated protein complexes and therefore lack cellular context. A complementary bottom-up direction has been to let effective parameters emerge from multiscale simulations that embed molecules in realistic environments. Coarse-grained simulations have derived two-dimensional binding rates for receptors anchored to the cell membrane [33], and have also quantified permeability across curved membranes [34]. The most ambitious step is a whole-cell kinetic model of a minimal bacterium, which integrated structural data, kinetic measurements, and realistic geometry to simulate a complete cell cycle [35]. Although developed for a minimal cell, the study demonstrates the feasibility of linking structural detail and cellular physiology.

A central difficulty lies not only in bridging temporal and spatial scales, but also in the need to integrate different model representations. MD simulations require interactions to be predefined for specific sites and partners, whereas cellular experiments usually report ensemble-averaged behaviors of molecules within complexes. For example, a single experimental binding constant may reflect contributions from multiple adaptor-binding sites, while an MD simulation typically probes only one of these at a time (Figure 3a). For the ErbB receptor family, array-based analyses quantified SH2-domain binding to phosphorylated peptides, although these measurements were limited to partial protein structures [36]. LC/MS analyses captured intracellular interactions between ErbB receptors and adaptor proteins [37], which would need to be complemented by other methods to quantify interaction strength. While valuable, these datasets were obtained under specific experimental conditions, primarily describing one-to-one interactions. Thus, they cannot address the one-to-many interaction dynamics or capture temporal changes in interaction strength driven by protein phosphorylation or dephosphorylation. A deeper challenge is identifying which microscopic processes contribute to the quantitative parameters of protein interactions observed at the cellular level.

A practical strategy is to reduce the overwhelming space of possible interactions to a manageable subset, and then perform precise parameter estimation on that reduced set. This can be organized as a pipeline (Figure 3b): structure prediction methods such as AlphaFold [7] can serve as an initial screen, generating candidate receptor–adaptor complexes and filtering them by model confidence. Short MD simulations can then be used to screen out unstable complexes. For the remaining candidates, advanced MD methods can be used to calculate binding constants that can be incorporated into ODE models. Alternatively, ML-based approaches can replace this step by predicting candidate interactions directly from sequence and structure data [38]. In practical terms, AlphaFold predictions are fast on GPU machines (less than 1 h per complex), the short MD simulations for screening are moderate (~20 h per complex for 100-ns simulations on 8 nodes of a national supercomputer), whereas calculating parameters with advanced simulations remains the most computationally intensive step, often requiring several days per complex on hundreds of nodes, depending on the method. This demonstrates how structural predictions can be refined into a tractable set of complexes for cellular modeling. Further progress, however, will depend on improving the estimation of kinetic parameters, and equally, on representing receptor internal dynamics that shape partner choice.

Figure 3



A pipeline for comprehensive EGFR–protein binding prediction. **(a)** Schematic model of the full-length EGFR in the asymmetric dimer state (right) and interactions of adaptor proteins at the C-terminal tail (left). Phosphorylated tyrosine sites (orange circles) indicate well-known positions involved in downstream signaling. The diagram highlights several adaptor proteins (Shc1, Grb2, PLCγ1, and PI3K) and their possible binding sites along the receptor tail. **(b)** Pipeline combining structure prediction and molecular dynamics (MD) simulations for estimating interaction parameters for cellular modeling. **(c)** Structural models of EGFR. Left: Cryo-EM structure of the EGF-bound EGFR ECD dimer. Visualized from PDB ID: 8HGS; right: crystal structure of the EGFR kinase domain in the asymmetric dimer configuration. Visualized from PDB ID: 2GS2. ECD: extracellular domain, TM: transmembrane.

Molecular mechanisms from structural modeling and simulation

Recent advances in MD simulations, cryo-EM, and super-resolution imaging now provide incredibly detailed views of ErbB receptor complexes and their dynamics (Figure 3c). Multiple noncanonical assemblies coexist beyond the ligand-bound dimer, including ligand-free oligomers that stabilize active dimers [39], mutation-dependent partner selection within the ErbB family [40], and asymmetric EGFR-HER2 heterodimers in which EGFR binds the ligand and HER2 stabilizes the interface [41]. Collectively, these findings indicate that ErbB signaling is controlled by a variety of receptor structures whose distribution and drug sensitivities are affected by mutations. The critical challenge for cellular modeling is to determine the functional significance of these assemblies: namely, which of them define distinct signaling routes, and how their drug sensitivities should be represented in models. Both issues remain unresolved, and systematic strategies will be needed to translate structural diversity into quantitative rules.

One way to address the complexity of these states is to represent signaling through multiple routes. This approach has been demonstrated in the MAPK pathway, where mechanistic models separated parallel channels to explain adaptive drug resistance [42]. To accurately model the role of these ErbB assemblies, we need to gather their structural and dynamical information within environments that closely mimic a living cell.

Engineered systems such as receptors reconstitution in extracellular vesicles offer platforms where receptor organization and dynamics can be probed under more native-like conditions [43].

Within the kinase domain, simulations have shown that activation is not a simple on/off switch but proceeds through intermediate states with transition probabilities and timescales that can now be explicitly defined [44]. A related example comes from RAF kinase dimers, where stabilization of different conformations was shown to drive distinct functional outcomes [45]. These results suggest that cellular models should move beyond a simple active-inactive distinction and incorporate intermediate, mutation- and drug-dependent states.

A remaining challenge lies in interpreting the diversity of receptor assemblies and the many possible conformational intermediates within the kinase domain. For assemblies, particle-based simulators [46] can explore how different oligomer sizes form, especially when combined with experimental approaches reporting oligomer distributions. For conformational changes, enhanced sampling simulations can map the multitude of intermediates. However, translating them into a limited number of representative states for cellular modeling remains difficult and may determine how effectively structural dynamics informs our understanding of signaling.

Conclusion and outlook

This review has highlighted recent advances in cellular modeling and molecular modeling and simulation of ErbB signaling. Despite remarkable progress at both levels, current modeling frameworks still face fundamental challenges that limit their integration. Across both cellular and molecular scales, the key limitation is identifiability, the ability to uniquely determine model structure and parameters from data. In cellular models, sparse time-series and perturbation data make it difficult to resolve network topology or kinetic parameters with confidence. In molecular modeling and simulations, incomplete structures and limited sampling hinder the unique definition of conformational and kinetic states, and translating simulation results into experimentally comparable parameters remains challenging. These issues together define the persistent gap between molecular detail and system-level representation. Addressing this gap requires community-wide efforts toward standardization of methods and tools, a concept already well established in DNA and RNA sequencing workflows. Several efforts have aimed to test drug response reproducibility or provide tools for navigating databases used for parameter estimation and simulation validation [47,48]. SBML was the earliest effort to standardize biological network models [49], supported by the BioModels Database [50] as a repository promoting standardized practices. The shift toward data-driven research also extended to the field of MD modeling and simulations, driven by ML applications that require large, annotated datasets. As the field is becoming increasingly interdisciplinary, the need for standardized, reproducible workflows has grown. A recent open letter [51] highlighted the need for adopting Findable, Accessible, Interoperable, and Reusable (FAIR) principles in the MD community. Examples of FAIR tools include Jupyter-based workflows [52], searchable MD data repositories [53], and ML-guided force-field optimization frameworks [54].

While cellular- and molecular-scale approaches are often pursued separately as complementary strategies, both are in fact approximations toward a longer-term goal: whole-cell models [55], in which network structure and quantitative parameters emerge together without requiring translation across scales (Figure 1). This multiscale integration, demonstrated by early whole-cell kinetic models [35] and advances in bottom-up parameter generation, is essential for achieving a comprehensive, mechanistic understanding of ErbB signaling and its role in human disease.

Author statement

Ai Shinobu: Investigation; Methodology; Visualization; Writing – original draft; Writing – review & editing.

Ayaka Nagasato-Ichikawa: Investigation; Visualization; Writing – original draft.

Mariko Okada: Investigation; Methodology; Supervision; Writing – original draft; Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests.

Acknowledgements

We thank Dr Akira Imamoto for critical reading of the manuscript. M.O. was supported by the JST CREST [Program Number JPMJCR21N3], JST ASPIRE [Program Number JPMJAP24B1] and Provost's Visiting Professorship from Imperial College London. A.S. was supported by the World Premier International Research Center Initiative (WPI), MEXT, Japan, and by MEXT as part of the "Program for Promoting Researches on the Supercomputer Fugaku" (Simulation- and AI-driven next-generation medicine and drug discovery based on "Fugaku" [Program Number JPMXP1020230120]). A.N.I. was supported by JSPS KAKENHI [Program Number 25K18877].

Data availability

No data was used for the research described in the article.

References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest
** of outstanding interest

1. Ubersax JA, Ferrell Jr JE: **Mechanisms of specificity in protein phosphorylation**. *Nat Rev Mol Cell Biol* 2007, **8**:530–541.

2. Kholodenko BN: **Cell-signalling dynamics in time and space**. *Nat Rev Mol Cell Biol* 2006, **7**:165–176.

3. Kholodenko BN, Hancock JF, Kolch W: **Signalling ballet in space and time**. *Nat Rev Mol Cell Biol* 2010, **11**:414–426.

4. Ferrell JE: **Understanding Cell Signaling: Motifs, Recurring Themes, and the Theory of Nonlinear Dynamics,**. *Nonlinear Dyn* 2021.

5. Wang J, Do HN, Koirala K, Miao Y: **Predicting biomolecular binding kinetics: a review**. *J Chem Theor Comput* 2023, **19**: 2135–2148.

6. Wright PE, Dyson HJ: **Intrinsically disordered proteins in cellular signalling and regulation**. *Nat Rev Mol Cell Biol* 2015, **16**:18–29.

7. Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore L, Ballard AJ, Bambrick J, *et al.*: **Accurate structure prediction of biomolecular interactions with AlphaFold 3**. *Nature* 2024, **630**:493–500.

8. Derbalah A, Al-Sallami H, Hasegawa C, Gulati A, Duffull SB: **A framework for simplification of quantitative systems pharmacology models in clinical pharmacology**. *Br J Clin Pharmacol* 2022, **88**:1430–1440.

9. Grivell L: **Mining the bibliome: searching for a needle in a haystack? New computing tools are needed to effectively scan the growing amount of scientific literature for useful information: new computing tools are needed to effectively scan the growing amount of scientific literature for useful information**. *EMBO Rep* 2002, **3**:200–203.

10. ** Rodriguez-Mier P, Garrido-Rodriguez M, Gabor A, Saez-Rodriguez J: **Unifying multi-sample network inference from prior knowledge and omics data with CORNETO**. *Nat Mach Intell* 2025, **7**:1168–1186.

This study presents methods for deriving context-dependent gene networks from multi-sample omics data and prior biological knowledge. By performing inference jointly across all samples, the method improved unified network construction with optimizing network flow, such as shorter pass, while also identifying unique subnetworks within individual samples.

11. Ruscone M, Tsirvouli E, Checcoli A, Turei D, Barillot E, Saez-Rodriguez J, Martignetti L, Flobak Å, Calzone L: **NeKo: a tool for automatic network construction from prior knowledge**. *PLoS Comput Biol* 2025, **21**, e1013300.

12. Erdem C, Mutsuddy A, Bensman EM, Dodd WB, Saint-Antoine MM, Bouhaddou M, Blake RC, Gross SM, Heiser LM, Feltus FA, *et al.*: **A scalable, open-source implementation of a large-scale mechanistic model for single cell proliferation and death signaling**. *Nat Commun* 2022, **13**:3555.

13. Türei D, Korcsmáros T, Saez-Rodriguez J: **OmniPath: guidelines and gateway for literature-curated signaling pathway resources**. *Nat Methods* 2016, **13**:966–967.

14. Licata L, Lo Surdo P, Iannuccelli M, Palma A, Micarelli E, Perfetto L, Peluso D, Calderone A, Castagnoli L, Cesareni G: **SIGNOR 2.0, the SIGnaling Network Open Resource 2.0: 2019 update**. *Nucleic Acids Res* 2020, **48**:D504–D510.

15. * Tsutsui M, Arakane K, Okada M: **Literature-derived, context-aware gene regulatory networks improve biological predictions and mathematical modeling**. *bioRxiv* 2025, https://doi.org/10.1101/2025.07.26.666734.

This study introduces an NLP approach for extracting human-disease GRNs from PubMed abstracts in their entirety, and demonstrates its application in predicting drug responses and generating mathematical models. Improving prediction accuracy using MeSH terms and quantitative validation could be applied to other NLP approaches in biological studies.

16. Imoto H, Yamashiro S, Okada M: **A text-based computational framework for patient -specific modeling for classification of cancers**. *iScience* 2022, **25**, 103944.

17. Arakane K, Imoto H, Ormersbach F, Okada M: **Extending BioMASS to construct mathematical models from external knowledge**. *Bioinform Adv* 2024, **4**:vbae042.

18. Yamagiwa H, Hashimoto R, Arakane K, Murakami K, Soeda S, Oyama M, Zhu Y, Okada M, Shimodaira H: **Predicting drug-gene relations via analogy tasks with word embeddings**. *Sci Rep* 2025, **15**, 17240.

19. Chen Q, Lee K, Yan S, Kim S, Wei C-H, Lu Z: **BioConceptVec: creating and evaluating literature-based biomedical concept embeddings on a large scale**. *PLoS Comput Biol* 2020, **16**, e1007617.

20. Shoval O, Alon U: **SnapShot: network motifs**. *Cell* 2010, **143**: 326-e1.

21. Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U: **Network motifs: simple building blocks of complex networks**. *Science* 2002, **298**:824–827.

22. Sinha S, Farfel A, Luker KE, Parker BA, Yeung KT, Luker GD, Ghosh P: **Growth signaling autonomy in circulating tumor cells aids metastatic seeding**. *PNAS Nexus* 2024, **3**:gae014.

23. ** Kadelka C, Butrie T-M, Hilton E, Kinseth J, Schmidt A, Serdarevic H: **A meta-analysis of Boolean network models reveals design principles of gene regulatory networks**. *Sci Adv* 2024, **10**:eadj0822.

This study demonstrated the design principles of Gene Regulatory Networks (GRNs) through a comprehensive meta-analysis of the largest repository of published, expert-curated Boolean GRN models. Future research investigating how the overabundance of specific classes of Nested Canalizing Functions (NCFs) affects dynamical robustness indicates the potential to elucidate stable dynamics in GRNs.

24. Gábor A, Villaverde AF, Banga JR: **Parameter identifiability analysis and visualization in large-scale kinetic models of biosystems**. *BMC Syst Biol* 2017, **11**:54.

25. Raue A, Karlsson J, Saccomani MP, Jirstrand M, Timmer J: **Comparison of approaches for parameter identifiability analysis of biological systems**. *Bioinformatics* 2014, **30**: 1440–1448.

26. Fröhlich F, Kessler T, Weindl D, Shadrin A, Schmiester L, Hache H, Muradyan A, Schütte M, Lim J-H, Heinig M, *et al.*: **Efficient parameter estimation enables the prediction of drug response using a mechanistic pan-cancer pathway model**. *Cell Syst* 2018, **7**:567–579.e6.

27. Dorešić D, Grein S, Hasenauer J: **Efficient parameter estimation for ODE models of cellular processes using semi-quantitative data**. *Bioinformatics* 2024, **40**:i558–i566.

28. Fey D, Halasz M, Dreidax D, Kennedy SP, Hastings JF, Rauch N, Munoz AG, Pilkington R, Fischer M, Westermann F, *et al.*: **Signaling pathway models as biomarkers: patient-specific simulations of JNK activity predict the survival of neuroblastoma patients**. *Sci Signal* 2015, **8**:ra130.

29. Tsutsui M, Okada M: **DynProfiler: a python package for comprehensive analysis and interpretation of signaling dynamics leveraged by deep learning techniques**. *Bioinform Adv* 2024, **4**:vbae145.

30. Nagarajan A, Amberg-Johnson K, Paull E, Huang K, Ghanakota P, Chandrasinghe A, Chief Elk J, Sampson JM, Wang L, Abel R, *et al.*: **Predicting resistance to small molecule kinase inhibitors**. *J Chem Inf Model* 2025, **65**: 2543–2557.

31. Passaro S, Corso G, Wohlwend J, Reveiz M, Thaler S, Somnath VR, Getz N, Portnoi T, Roy J, Stark H, *et al.*: **Boltz-2:**

towards accurate and efficient binding affinity prediction. *bioRxiv* 2025, https://doi.org/10.1101/2025.06.14.659707.

32. Wang J, Miao Y: **Protein-protein interaction-Gaussian accelerated molecular dynamics (PPI-GaMD): characterization of protein binding thermodynamics and kinetics**. *J Chem Theor Comput* 2022, **18**:1275–1285.

33. * Peterson KJ, Slepchenko BM, Loew LM: **Mesoscale simulations of membrane-tethered reactions to parameterize cell-scale models of signaling**. *bioRxivorg* 2025, https://doi.org/10.1101/2024.12.04.626844.
This study presents a method to convert 3D binding rates into effective 2D binding rate constants by explicitly modeling membrane tethering and molecular geometry using coarse-grained spatial simulations. The approach enables accurate parameterization of membrane-localized signaling events for use in cellular-scale models.

34. Davoudi S, Ghysels A: **Defining permeability of curved membranes in molecular dynamics simulations**. *Biophys J* 2023, **122**:2082–2091.

35. Thornburg ZR, Bianchi DM, Brier TA, Gilbert BR, Earnest EE, Melo MCR, Safronova N, Sáenz JP, Cook AT, Wise KS, *et al*.: **Fundamental behaviors emerge from simulations of a living minimal cell**. *Cell* 2022, **185**:345–360.e28.

36. Jones RB, Gordus A, Krall JA, MacBeath G: **A quantitative protein interaction network for the ErbB receptors using protein microarrays**. *Nature* 2006, **439**:168–174.

37. Schulze WX, Deng L, Mann M: **Phosphotyrosine interactome of the ErbB-receptor kinase family**. *Mol Syst Biol* 2005, **1**:8. 2005.

38. Huber HA, Finley SD: **Multiscale probabilistic modeling: a Bayesian approach to augment mechanistic models of cell signaling with machine-learning predictions of binding affinity**. *bioRxivorg* 2025, https://doi.org/10.1101/2025.05.23.655795.

39. ** Iyer RS, Needham SR, Galdadas I, Davis BM, Roberts SK, Man RCH, Zanetti-Domingues LC, Clarke DT, Fruhwirth GO, Parker PJ, *et al*.: **Drug-resistant EGFR mutations promote lung cancer by stabilizing interfaces in ligand-free kinase-active EGFR oligomers**. *Nat Commun* 2024, **15**:2130.
This study combines structure-based modeling, multi-microsecond MD simulations, high-resolution in vitro spatial assays (FLImP), and in vivo functional validation to uncover how drug-resistant EGFR mutations stabilize ligand-free oligomers. The reconstructed full-length EGFR models provide a mechanistic hypothesis tested across complementary approaches, showing that these mutations promote the assembly of oligomer-obligate active dimers by engaging ectodomain and transmembrane interfaces not previously linked to activation.

40. Dey P, Gadewal N, De A: **Pathogenic HER3 dimerization domain mutations create a structural bias towards unconventional EGFR-HER3 signalling axis in breast cancer**. *Int J Biol Macromol* 2023, **242**, 124765.

41. Bai X, Sun P, Wang X, Long C, Liao S, Dang S, Zhuang S, Du Y, Zhang X, Li N, *et al*.: **Structure and dynamics of the EGFR/HER2 heterodimer**. *Cell Discov* 2023, **9**.

42. Fröhlich F, Gerosa L, Muhlich J, Sorger PK: **Mechanistic model of MAPK signaling reveals how allostery and rewiring contribute to drug resistance**. *Mol Syst Biol* 2023, **19**, e10988.

43. Gonzalez-Magaldi M, Gullapalli A, Papoulas O, Liu C, Leung AY-H, Guo L, Brilot AF, Marcotte EM, Ke Z, Leahy DJ: **Structure and organization of full-length epidermal growth factor receptor in extracellular vesicles by cryo-electron tomography**. *Proc Natl Acad Sci U S A* 2025, **122**, e2424678122.

44. * Inoue M, Ekimoto T, Yamane T, Ikeguchi M: **Computational analysis of activation of dimerized epidermal growth factor receptor kinase using the string method and Markov state model**. *J Chem Inf Model* 2024, **64**:3884–3895.
This study combines all-atom MD simulations with the string method and Markov state modeling to characterize the conformational transition between inactive and active states of asymmetric EGFR dimer. The resulting free-energy landscape captures long-timescale dynamics and identifies intermediate states, offering detailed mechanistic insight into the asymmetric dimer activation process.

45. Rukhlenko OS, Khorsand F, Krstic A, Rozanc J, Alexopoulos LG, Rauch N, Erickson KE, Hlavacek WS, Posner RG, Gómez-Coca S, *et al*.: **Dissecting RAF inhibitor resistance by structure-based modeling reveals ways to overcome oncogenic RAS signaling**. *Cell Syst* 2018, **7**:161–179.e14.

46. Michalski PJ, Loew LM: **SpringSaLaD: a spatial, particle-based biochemical simulation platform with excluded volume**. *Biophys J* 2016, **110**:523–529.

47. Niepel M, Hafner M, Mills CE, Subramanian K, Williams EH, Chung M, Gaudio B, Barrette AM, Stern AD, Hu B, *et al*.: **A multi-center study on the reproducibility of drug-response assays in mammalian cell lines**. *Cell Syst* 2019, **9**:35–48.e5.

48. Xie Z, Kropiwnicki E, Wojciechowicz ML, Jagodnik KM, Shu I, Bailey A, Clarke DJB, Jeon M, Evangelista JE, Kuleshov MV, *et al*.: **Getting started with LINCS datasets and tools**. *Curr Protoc* 2022, **2**, e487.

49. Keating SM, Waltemath D, König M, Zhang F, Dräger A, Chaouiya C, Bergmann FT, Finney A, Gillespie CS, Helikar T, *et al*.: **SBML level 3: an extensible format for the exchange and reuse of biological models**. *Mol Syst Biol* 2020, **16**, e9110.

50. Malik-Sheriff RS, Glont M, Nguyen TVN, Tiwari K, Roberts MG, Xavier A, Vu MT, Men J, Maire M, Kananathan S, *et al*.: **BioModels-15 years of sharing computational models in life science**. *Nucleic Acids Res* 2020, **48**:D407–D415.

51. Amaro RE, Åqvist J, Bahar I, Battistini F, Bellaiche A, Beltran D, Biggin PC, Bonomi M, Bowman GR, Bryce RA, *et al*.: **The need to implement FAIR principles in biomolecular simulations**. *Nat Methods* 2025, **22**:641–645.

52. Bayarri G, Andrio P, Gelpí JL, Hospital A, Orozco M: **Using interactive Jupyter notebooks and BioConda for FAIR and reproducible biomolecular simulation workflows**. *PLoS Comput Biol* 2024, **20**, e1012173.

53. Tiemann JKS, Szczuka M, Bouarroudj L, Oussaren M, Garcia S, Howard RJ, Delemotte L, Lindahl E, Baaden M, Lindorff-Larsen K, *et al*.: **MDverse, shedding light on the dark matter of molecular dynamics simulations**. *eLife* 2024, **12**:RP90061.

54. Wang X, Li J, Yang L, Chen F, Wang Y, Chang J, Chen J, Feng W, Zhang L, Yu K: **DMFF: an open-source automatic differentiable platform for molecular force field development and molecular dynamics simulation**. *J Chem Theor Comput* 2023, **19**:5897–5909.

55. Luthey-Schulten Z, Thornburg ZR, Gilbert BR: **Integrating cellular and molecular structures and dynamics into whole-cell models**. *Curr Opin Struct Biol* 2022, **75**, 102392.