Proceedings of ICBioMed 2026 Symposium: Biomedical Engineering and Smart Medicine
DOI: 10.54254/2753-8818/2026.HZ38172

AlphaFold for Phosphorylation-Associated Kinase Conformational States

Ziyao Wang

Department of Chemistry, University College London, London, UK
zccazw2@ucl.ac.uk

Abstract. Activation-loop phosphorylation is a demanding test of protein-structure prediction because it modulates both local chemistry and the relative occupancy of pre-existing kinase conformations. This critical narrative review asks whether explicit phosphoresidue representation allows AlphaFold-based approaches to capture phosphorylation-induced conformational change. Targeted searches of PubMed and official publisher databases incorporated structural, solution-biophysical, biochemical and computational evidence. Extracellular signal-regulated kinase 2 (ERK2) indicates that phosphate-stabilised contacts coexist with activation-loop motion, whereas Aurora A shows that phosphorylation can induce strong activation without a coarse aspartate–phenylalanine–glycine (DFG)-out-to-DFG-in population shift. These cases make the required prediction target a condition-dependent ensemble rather than a single phosphorylated structure. AlphaFold 2-derived multiple-sequence-alignment perturbations broaden state coverage, and AlphaFold 3 can encode modified residues, ligands and ions. However, a 109-protein benchmark showed only a small, statistically non-significant improvement in phosphorylated-state alignment when phosphorylation was supplied explicitly. Current AlphaFold models can therefore produce chemically informed candidate states, but their output frequencies are not calibrated solution populations. A four-level validation framework is proposed to separate local chemistry, state recovery, perturbation-induced population shifts and prospective functional prediction. Ensemble-sensitive validation remains necessary, while physical reweighting can help explain phosphorylation-dependent kinase regulation.

Keywords: activation loop, phosphorylation, AlphaFold 3, protein kinase, conformational ensemble

1. Introduction

Protein phosphorylation is a covalent process whereby a protein kinase transfers the γ-phosphoryl group of adenosine triphosphate (ATP) to a hydroxyl-bearing amino-acid side chain (typically serine, threonine or tyrosine). Many kinases also undergo phosphorylation within their own activation segment. It is important to distinguish these two roles: substrate phosphorylation is the reaction catalysed by a kinase, whereas activation-loop phosphorylation changes the chemical state of a kinase. Catalytic competence also involves the positioning of nucleotide, Mg^{2+}, substrate and

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2026.HZ38172

conserved catalytic elements in a coordinated manner [1]. A correct overall fold is therefore not sufficient to identify a kinase regulatory state.

The kinase domain consists of an N-terminal lobe, a larger C-terminal lobe and an ATP-binding cleft in between. The β3-lysine–αC-glutamate ion pair, histidine–arginine–aspartate motif, aspartate–phenylalanine–glycine (DFG) motif, activation loop and hydrophobic regulatory spine constitute a coupled structural network [1, 2]. DFG-in, DFG-out and intermediate conformations alter catalytic geometry and expose different inhibitor-binding pockets, but the DFG label alone does not uniquely indicate activity. The αC helix, the activation loop, and the catalytic residues can undergo partially independent conformational rearrangements. Modern kinase biophysics thus characterises regulation as a redistribution across an energy landscape as opposed to a binary conversion between an inactive and an active structure [1, 3].

AlphaFold 2 (AF2) enabled highly accurate prediction of many static protein structures [4]. Later techniques have perturbed multiple-sequence alignments (MSAs), templates or inference settings to produce alternative conformations. AlphaFold 3 (AF3) extends the molecular representation to complexes containing small molecules, ions and supported modified residues, such as phosphoresidues [5]. This development removes a significant chemical-input constraint, though it does not necessarily make either system a thermodynamic ensemble predictor.

The research question is therefore narrower than whether AlphaFold can predict a protein: can it capture the conformational change caused by activation-loop phosphorylation, instead of simply positioning a phosphate group in a plausible local geometry? This difference can be experimentally tested and is chemically significant. A recent benchmark found only modest differences among AF2, unmodified AF3 and phosphorylation-aware AF3 when models were compared with phosphorylated and non-phosphorylated structures [6]. The thesis of this review is that chemical encoding, conformational state generation and condition-dependent population weighting are three separate tasks. AlphaFold has made substantial progress on the first two; the third is the principal barrier to mechanistic prediction.

2. Scope and literature strategy

This article is a critical narrative review, not a systematic review/meta-analysis. Targeted searches were searched up to 18 September 2026 in PubMed and official publisher databases. Core Boolean combinations included ("protein kinase" AND ("activation loop" OR DFG) AND phosphorylation AND (conformation OR ensemble OR dynamics)) and ((AlphaFold2 OR AlphaFold3) AND (phosphorylation OR "modified residue" OR "conformational ensemble" OR "MSA subsampling")). Backward and forward citation chaining from the foundational kinase reviews, the AF2 and AF3 papers and the direct phosphorylation benchmark was used to identify additional primary studies.

English-language, peer-reviewed studies were prioritised when they provided one of four types of evidence: an atomic structure defining phosphosite contacts; a solution-sensitive measurement of exchange or state populations; a biochemical measurement connecting conformation to function; or a computational benchmark against experimentally characterised states. Foundational older studies were retained when they defined a mechanism that is explicitly tested by newer studies. Reports presenting only a single predicted model without a state definition or experimental comparator were not treated as primary mechanistic evidence. Extracellular signal-regulated kinase 2 (ERK2) and Aurora A were selected as chemically resolved phosphorylation cases; Abelson tyrosine kinase (ABL1) was retained as a test of multistate generation. Because the search was targeted rather than exhaustive, the synthesis supports mechanistic evaluation but cannot estimate publication frequency or a pooled effect size.

DOI: 10.54254/2753-8818/2026.HZ38172

3. Why phosphorylation is difficult to predict

Activation-loop phosphorylation changes atoms, charge and non-covalent interaction possibilities. The side-chain hydroxyl becomes a phosphate monoester, commonly represented near neutral pH with a net charge close to −2, although microscopic protonation depends on the local electrostatic environment. The added phosphorus and oxygen atoms enlarge the side chain, increase hydration and hydrogen-bond acceptor capacity, and allow salt bridges with lysine or arginine. The phosphate group can also create repulsion with acidic residues and reorganise structured water [3]. A locally attractive contact must therefore be balanced against phosphate desolvation, conformational entropy and competing interactions.

The covalent reaction can be written schematically as equation (1):

$$\text{Protein–OH} + \text{ATP} \rightarrow \text{Protein–O–PO}_3^{2-} + \text{ADP} \tag{1}$$

Here, Protein–OH denotes the hydroxyl-bearing protein side chain and ADP is adenosine diphosphate. This schematic omits proton and metal-ion bookkeeping and explains covalent modification: after phosphoryl transfer, the phosphate is attached to the side-chain oxygen by a new covalent P–O bond and remains part of the protein until a phosphatase hydrolyses it. The equation, however, does not specify a three-dimensional outcome. The same phosphate can be buried in an ion-pair network, exposed to solvent or disordered, depending on the conformation and molecular context [3].

The ensemble consequence is state dependent. The equilibrium probability of state i can be written as equation (2):

$$p_i = \frac{\exp(-G_i/RT)}{\sum_j \exp(-G_j/RT)}, \tag{2}$$

where p_i is the probability of state i, G_i is its molar Gibbs free energy, R is the molar gas constant, T is the absolute temperature, and j indexes the states. Phosphorylation changes the free energy of each state by a potentially different amount and does not necessarily create a previously unattainable structure. If the modification stabilises an active-like substate more than an inactive one, the population shifts towards the active-like substate. At 298 K, only about 1.36 kcal mol^{-1} separates states whose equilibrium populations differ tenfold. This means that an incorrect phosphate protonation state, missing Mg^{2+} ion or misplaced salt bridge can change the predicted ranking of closely competing states even when both structures appear plausible.

A rigorous prediction should thus meet three different tests. First, chemical encoding asks whether there is valid geometry and appropriate local partners of the phosphoresidue. Second, conformational sampling asks whether the experimentally relevant global and local kinase states are generated. Third, thermodynamic weighting asks whether the probabilities of those states change in the experimentally correct direction and magnitude. AF3 directly improves the first task, whereas MSA-perturbed AF2 methods can improve the second. Neither success logically establishes the third.

The target is conditional as well. The same coupled kinase network can be modified by nucleotide identity, Mg^{2+} concentration, inhibitor binding, regulatory partners, mutations, other post-translational modifications, pH and temperature [1, 3]. The relevant object is the conditional distribution in equation (3), rather than a context-free phosphorylated kinase structure:

DOI: 10.54254/2753-8818/2026.HZ38172

$$P(\text{conformation} \mid \text{sequence, phosphorylation, ligands, partners and conditions}) \tag{3}$$

Here, P denotes a conditional probability distribution. The omission of a conditioning variable in a model may return a plausible structure while answering a different biochemical question. Figure 1 separates the local chemical representation and the ensemble inference required to explain regulation.

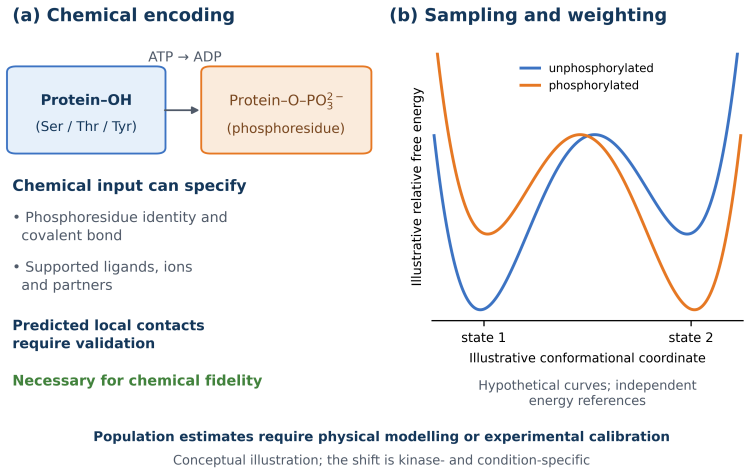

Figure 1. Chemical encoding versus conformational sampling and weighting. (a) Phosphorylation adds a covalent phosphate group. The −2 charge is a conventional illustrative assignment, not a claim that AF3 calculates protonation equilibria; local contacts are inferred from predicted structures and require validation. (b) The two curves show hypothetical relative free-energy landscapes, not experimental data or calculated free energies. Each curve has an arbitrary reference; only differences within a curve are interpreted. State 1 and state 2 are illustrative basins, not universal inactive/active labels. This conceptual illustration summarises chemical and ensemble distinctions discussed in the literature [3, 5, 6]

4. Experimental case studies define the prediction target

ERK2 demonstrates that phosphate contacts cannot be equated with a rigid active structure. Dual phosphorylation of the Thr–Glu–Tyr motif remodels the activation loop and establishes ion-pair interactions with basic residues of the kinase core, which provides an atomic explanation for the stabilisation of a catalytically competent arrangement [7]. However, relaxation-dispersion nuclear magnetic resonance (NMR) spectroscopy detected microsecond-to-millisecond motion in phosphorylated ERK2 and coupling between activation-loop exchange and motions around the catalytic core [8]. The warranted conclusion is not that phosphorylation locks ERK2 in one rigid conformation. It stabilises productive interactions in an ensemble that remains dynamic.

Aurora A exposes a second problem: the conformational coordinate chosen by the analyst can determine whether an effect is visible. Biochemical and spectroscopic studies show that both Thr288 phosphorylation and binding of the spindle factor targeting protein for Xenopus kinesin-like protein 2 (TPX2) activate Aurora A through distinct, coupled conformational changes [9, 10]. These routes are complementary rather than interchangeable. Treating TPX2 as simply a requirement for phosphorylation-dependent activity would therefore be incorrect.

DOI: 10.54254/2753-8818/2026.HZ38172

Spectroscopy resolved the distinction. TPX2 produces a pronounced shift from DFG-out towards DFG-in and stabilises a water-mediated allosteric network between αC and the active site [9]. By contrast, phosphorylation of Thr288 increases Aurora A activity by about two orders of magnitude without requiring a comparable coarse DFG-out-to-DFG-in population shift; the phosphorylated activation loop remains highly dynamic [10]. The evidence instead supports redistribution within the DFG-in basin, away from an autoinhibited DFG-in substate and towards a fully organised active substate [10]. Thus, a predictor might be able to recover the correct DFG category but fail to detect the regulatory change altogether.

Ligands provide a further layer. Measurements across Aurora A activation states found that inhibitor preferences for active- or inactive-loop conformations were retained despite phosphorylation, even though the modification altered the underlying ensemble [11]. This does not make phosphorylation irrelevant; it demonstrates that the observed binding reflects conformational selection as well as direct interactions with the ligand. A credible prediction must specify whether its state label refers to the DFG motif, the activation-loop substate, αC organisation, ligand-bound pocket geometry or a joint coordinate.

ABL1 can be used to make another, deliberately restricted inference. Its active and inactive kinase states are well characterised, including conformations of relevance to the binding of a type II inhibitor. MSA-subsampled AF2 generated multiple ABL1 states and reproduced many qualitative mutation-dependent changes in relative state populations when compared with NMR-derived evidence [12]. This result demonstrates that perturbation of evolutionary inputs can expose latent conformations of kinases. It does not demonstrate that an explicitly phosphorylated activation loop has been chemically modelled or that model counts are universal equilibrium populations. These various roles of evidence are explicit in Table 1.

Table 1. What each core kinase system can—and cannot—establish

System and context	Decisive evidence	Supported inference	Unsupported overreach
ERK2; dual Thr–Glu–Tyr phosphorylation	Crystal phosphate-contact network; NMR exchange [7, 8]	Productive contacts coexist with activation-loop dynamics.	A static active-like structure is not a rigid or calibrated ensemble.
Aurora A; phosphorylated Thr288 ± TPX2 or inhibitors	Kinetics, infrared/Förster resonance energy transfer and ensemble measurements [9-11]	Phosphorylation and TPX2 activate through different, coupled coordinates.	Phosphorylation is not "insufficient"; DFG-in/out alone is not the full mechanism.
ABL1; sequence/mutation-dependent states	NMR comparison and MSA-subsampled AF2 [12]	Evolutionary-input perturbation can generate latent kinase states.	This is not direct evidence for phosphoresidue chemistry or universal Boltzmann weights.

Note. ABL1 is not a phosphorylation case but a state-generation control.

5. What AlphaFold 2-derived multistate methods demonstrate

Standard AF2 inference was designed to predict accurate structures, not to sample a Boltzmann distribution [4]. However, alternative conformations can be elicited by changing the evolutionary information presented to the network. AF-Cluster groups homologous sequences before prediction and reported recovery of alternative conformations in selected metamorphic proteins, with NMR validation of a predicted KaiB variant state [13]. MSA subsampling varies alignment depth and

DOI: 10.54254/2753-8818/2026.HZ38172

repeated inference; in ABL1 and another test protein it produced state distributions that tracked many experimentally measured mutation effects [12].

Together, these studies establish conformational coverage: AF2's learned representation can support more than one native-like state. These studies also demonstrate that model diversity is experimentally useful when the coordinate of interest is known. Coverage is not identical to equilibrium calibration. Alignment construction, sequence depth and inference settings can influence the fraction of predictions in a state [12, 13]. A hypothetical 70:30 split among generated structures is therefore first a property of the model–protocol combination. It becomes evidence for a 70:30 solution population only after independent calibration.

Confidence scores do not close this gap. AF2 predicted local distance difference test scores estimate expected local structural accuracy and predicted aligned error reports uncertainty in relative positioning [4]. Neither is a state free energy. A rare but well-learned crystallographic state can receive high confidence, and a genuinely mobile activation loop can receive lower confidence. The choice of the highest-confidence model could thus improve local geometry while removing biologically significant alternatives.

The chemistry-specific limitation also exists. An MSA records sequence variation across evolution; it does not tell us whether one protein molecule is phosphorylated in one sample. Evolutionary information may encode that a region is regulatable or compatible with multiple states, but not the precise covalent state, nucleotide, metal or partner. MSA perturbation and explicit phosphoresidue representation are consequently complementary inputs, not competing explanations of the same phenomenon.

6. AlphaFold 3 and the representation–ensemble gap

AF3 uses a diffusion-based architecture to predict biomolecular complexes and can represent supported modified amino acids, small molecules and ions in a single input [5]. For kinase studies, this is a representational improvement: phosphothreonine or phosphotyrosine can be specified rather than approximated by an unmodified residue, and nucleotide, inhibitor or metal context can in principle be included. The general claim that AlphaFold cannot include phosphorylation is therefore no longer accurate for AF3.

This representational advance should be distinguished from molecular dynamics. The AF3 study warns that structure-prediction samples should not be interpreted as a solution-state ensemble; samples are ranked by model-confidence metrics rather than physical free energy [5]. Clashes, stereochemical issues, inappropriate ordering of flexible regions and states inconsistent with the provided ligand may also require checking of predicted complexes. General guidance accordingly treats an AlphaFold structure as a testable hypothesis, rather than a substitute for experimental structure determination [14]. In the case of a phosphosite, a predicted salt bridge is a mechanistic proposal; it is not evidence that the contact is highly occupied in solution.

A direct phosphorylation benchmark compared AF2, unmodified AF3 and phosphorylation-aware AF3 (AF3-p, the study's label for AF3 supplied with phosphorylation) against paired phosphorylated and non-phosphorylated structures from 109 proteins, most of which were from kinase families [6]. AF3-p was closer to the selected phosphorylated conformer in 73% of cases, compared with 70% for AF3 and 64% for AF2. However, the proportions aligning with phosphorylated versus non-phosphorylated conformers did not differ significantly among models ($\chi^2 = 2.28$, $p = 0.32$). All three models aligned with the largest Protein Data Bank-derived structural cluster in 98% of cases and none of the models showed significant enrichment for matching phosphorylation state beyond the random expectation (AF3-p $p = 0.44$; AF3 $p = 0.70$; AF2 $p = 0.84$) [6].

DOI: 10.54254/2753-8818/2026.HZ38172

These numbers support a narrow, accurate conclusion. Explicit phosphorylation produced only a small descriptive improvement, with no reliable phosphorylation-specific redistribution of predicted states shown by the benchmark. This outcome does not prove that AF3 never responds correctly: there are numerous phosphorylated and non-phosphorylated structures within the same cluster, and an endpoint Protein Data Bank comparison cannot directly measure solution populations. Further, it is likely that most benchmark structures were represented in AlphaFold training data, which limited prospective interpretation [6]. Despite these reservations, the absence of significant state enrichment would be an argument against the use of explicit post-translational modification input as sufficient ensemble evidence.

The most plausible synthesis is therefore a representation–thermodynamics gap. AF3 can make the molecular question better specified, but confidence-ranked structure generation does not calculate relative free energies of the competing states. Difficulties are compounded by sparse phosphoprotein structures, heterogeneous constructs, different ligands and crystallisation conditions and the possibility that regulation occurs within a coarse state category [6, 10]. The useful question is not whether AF3 can place a phosphate. It is whether matched phosphorylated and non-phosphorylated inputs produce an experimentally correct change along the biologically relevant coordinate.

7. A four-level validation framework

The reviewed evidence is organised here into a proposed four-level conceptual framework; it is not an independently validated scoring system. Level 1: local chemical validity: correct phosphoresidue bonding and stereochemistry, a defensible protonation state, plausible salt bridges and hydrogen bonds, solvent access and no serious clashes. This may be checked by chemical inspection and comparison with high-resolution structures, but establishes only the possibility of a local arrangement. Level 2: state recovery: the method must generate experimentally observed DFG, αC and activation-loop configurations. State-specific distances, dihedrals and contact patterns are more informative than global root-mean-square deviation alone [2].

Level 3 is perturbation calibration. Matched calculations must predict the direction and, ideally, the magnitude of the population change due to phosphorylation, dephosphorylation, partner binding, mutation or the addition of a ligand. Ensemble-sensitive measurements are needed: NMR exchange, Förster resonance energy transfer populations and double electron–electron resonance distance distributions provide complementary constraints on conformational states; exchange timescales require method-specific analysis [8-11]. Level 4 is prospective functional prediction. A pre-registered prediction for an unseen kinase, phosphosite or condition should correctly anticipate an activity, binding or solution-state measurement. For a mechanistic claim, evidence from the preceding levels should accompany prospective validation; a correct activity or binding prediction alone does not establish the proposed conformational mechanism.

The framework also assigns appropriate roles to methods. Candidate structures are provided by AF2 or AF3. Chemical preparation then specifies the phosphoresidue, protonation state, nucleotide, Mg^{2+}, ligand, missing atoms and solvent environment. Molecular dynamics can relax these models, while enhanced sampling and reweighting attempt to estimate transitions and relative probabilities. An example of this division of labour is found in AF2-RAVE (AlphaFold 2 with reweighted autoencoded variational Bayes for enhanced sampling): AlphaFold-derived candidates were combined with learned coordinates and physics-based sampling to identify metastable, inhibitor-compatible kinase states for conformation-selective docking [15]. Such results are promising but remain conditional on force fields, collective variables and convergence.

DOI: 10.54254/2753-8818/2026.HZ38172

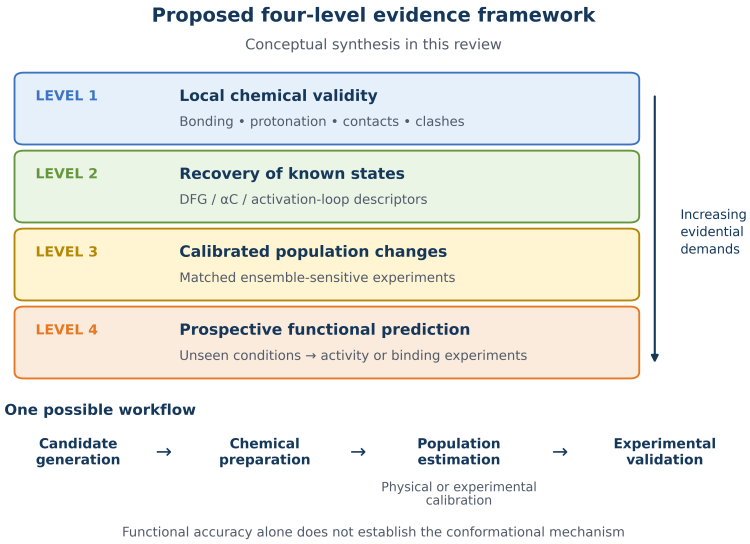

Figure 2. Proposed evidence framework for phosphorylation-aware kinase prediction. The levels distinguish chemical validity, state recovery, perturbation-induced population changes and prospective functional testing [6, 8, 10, 11, 14, 15]. This diagram illustrates the present review's synthesis, not a validated consensus standard. The lower workflow is one possible strategy: physical reweighting is not the only route to population estimation, and calculated free energies alone do not establish experimental agreement. Prospective activity or binding accuracy must be accompanied by evidence for the claimed mechanism. No new experimental validation was performed in this review

A controlled benchmark should be specified before favourable structures are selected. The phosphorylated and unphosphorylated calculations need the same constructs, ligand conditions and sampling budgets. Experimental observables or independent structural descriptors should be fixed as state definitions; state recall, direction of population shift, calibration error and uncertainty should be reported. Homology-aware train/test separation is necessary to distinguish learned structural recall from generalisation. Precise model versions, post-translational modification codes, MSAs, templates, seeds, failed predictions and analysis scripts should be reported to make this proposed evaluation reproducible [6, 12-14]. This claim ladder and the evidence workflow are summarised in Figure 2.

8. Implications for kinase inhibitor discovery

Kinase inhibitors take advantage of conformational heterogeneity. Type I inhibitors usually recognise DFG-in ATP-site geometries, whereas type II inhibitors extend into an adjacent pocket associated with a classical DFG-out geometry; other classes of inhibitors stabilise other inactive or allosteric geometries [2, 15]. These are useful categories rather than absolute rules. Binding free energy not only includes direct protein-ligand interactions, but also the energetic cost of accessing the binding-competent state. Phosphorylation, partners and disease mutations can alter that cost without changing all local contacts.

When the state panel is physically and biologically credible, multistate prediction can improve screening. AF2-RAVE combined multistate generation, enhanced sampling and induced-fit docking to recover inhibitor-compatible kinase structures in retrospective tests [15]. Aurora A nevertheless shows why phosphorylation is not a complete state label: TPX2 and inhibitors can alter the DFG distribution and ligand conformational preferences can be preserved between phosphorylation states

DOI: 10.54254/2753-8818/2026.HZ38172

[9-11]. Screening against one phosphorylated AF3 structure may thus miss a ligand-selected substate or overvalue a state favoured by model priors.

A justifiable workflow would screen a small panel of experimentally anchored or physically reweighted states and report results separately for each state. Both affinity and conformational effect should then be tested on candidate ligands since a compound can bind and also redistribute the ensemble. In this role, artificial intelligence (AI) expands structural hypotheses; this does not eliminate the chemical requirement to determine protonation, metal coordination, induced fit and state populations.

9. Research priorities and limits of the evidence

The highest priority is matched, condition-labelled data. Useful datasets would compare the same kinase constructs before and after a defined phosphorylation event with nucleotide, Mg^{2+}, ligand, partner, pH and temperature held constant. Where feasible, ensemble-sensitive observables and activity measurements should be provided alongside the atomic structures. Current Protein Data Bank pairs often confound phosphorylation with the design of constructs, bound molecules and crystallisation conditions, which prevents clean causal attribution [6]. Depositing raw NMR, Förster resonance energy transfer or double electron–electron resonance data and the forward models used to predict those observables would permit quantitative ensemble testing.

Model development should separate generation from weighting. A broad generative stage can propose diverse states; a physics-informed or experimentally calibrated stage may give context-dependent probabilities. Modified-residue treatments should report assumptions regarding protonation, and should include coupling to water, ions and multisite modifications. Benchmarks should also include negative controls in which explicit phosphorylation should not result in a significant global rearrangement. Otherwise, a method may be rewarded for generating change as opposed to predicting the correct absence, location or scale of change.

There are three limitations of this review. Its targeted narrative search cannot guarantee exhaustive coverage or support a pooled quantitative effect. The mechanistic emphasis on ERK2 and Aurora A increases biochemical depth but limits generalisation across the kinome. Finally, structural databases are conditionally sampled records instead of equilibrium experiments and the direct AlphaFold phosphorylation benchmark is largely retrospective [6]. These constraints are why the proposed framework treats prospective functional and solution-state validation as the strongest level of evidence, rather than claiming that present data settle performance for every kinase.

Claim calibration is itself an actionable research practice. A correct phosphate contact supports local chemical plausibility; recovery of DFG-out supports state coverage; an experimentally matched population shift supports regulatory inference; and prospective activity or binding supports functional relevance. Keeping these claims separate would make apparent disagreements between AI, crystallography, simulation and solution biophysics easier to diagnose and would reduce the overinterpretation of visually compelling models.

10. Conclusion

The evidence supports a qualified assessment of AlphaFold's capabilities. AlphaFold can represent a phosphorylated residue in AF3, and can sometimes generate kinase conformations associated with different regulatory states. It does not yet reliably establish the redistribution of those states caused by phosphorylation. ERK2 demonstrates that the stabilisation of phosphate contacts can be accompanied by persistent motion. Aurora A demonstrates that a robust phosphorylation-dependent

DOI: 10.54254/2753-8818/2026.HZ38172

activation can be mediated by reorganisation within the DFG-in basin rather than by the coarse DFG shift generated by TPX2. A state label must therefore match the molecular mechanism under test.

The main contribution of this review is to separate four levels of evidence: local chemical validity, recovery of known states, calibration of perturbation-induced population shifts and prospective functional prediction. Under this framework, an explicit phosphoresidue input is necessary for a chemically faithful question but is not sufficient for an ensemble-level answer. AlphaFold outputs are most defensible as candidate states that are chemically prepared, calibrated through physical or experimental approaches and tested against solution-sensitive and functional measurements.

The current evidence base remains limited by the heterogeneity of experimental conditions, the lack of matched phosphorylated/non-phosphorylated datasets and retrospective evaluation. Progress therefore depends less on producing larger collections of plausible structures than on defining the correct coordinate, controlling biochemical context and testing predicted population changes prospectively. Chemical encoding asks what structures are possible; sampling, weighting and experiment determine which structures matter.

References

[1] Endicott, J. A., Noble, M. E. M., & Johnson, L. N. (2012). The structural basis for control of eukaryotic protein kinases. *Annual Review of Biochemistry, 81*(1), 587–613. https://doi.org/10.1146/annurev-biochem-052410-090317

[2] Modi, V., & Dunbrack, R. L., Jr. (2019). Defining a new nomenclature for the structures of active and inactive kinases. *Proceedings of the National Academy of Sciences, 116*(14), 6818–6827. https://doi.org/10.1073/pnas.1814279116

[3] Nishi, H., Shaytan, A., & Panchenko, A. R. (2014). Physicochemical mechanisms of protein regulation by phosphorylation. *Frontiers in Genetics, 5*, 270. https://doi.org/10.3389/fgene.2014.00270

[4] Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature, 596*(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

[5] Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature, 630*, 493–500. https://doi.org/10.1038/s41586-024-07487-w

[6] Ramasamy, P., Zuallaert, J., Martens, L., & Vranken, W. F. (2026). Assessing the relation between protein phosphorylation, AlphaFold3 models, and conformational variability. *Protein Science, 35*(1), e70376. https://doi.org/10.1002/pro.70376

[7] Canagarajah, B. J., Khokhlatchev, A., Cobb, M. H., & Goldsmith, E. J. (1997). Activation mechanism of the MAP kinase ERK2 by dual phosphorylation. *Cell, 90*(5), 859–869. https://doi.org/10.1016/S0092-8674(00)80351-7

[8] Iverson, D. B., Xiao, Y., Jones, D. N., Eisenmesser, E. Z., & Ahn, N. G. (2020). Activation loop dynamics are coupled to core motions in extracellular signal-regulated kinase-2. *Biochemistry, 59*(29), 2698–2706. https://doi.org/10.1021/acs.biochem.0c00485

[9] Cyphers, S., Ruff, E. F., Behr, J. M., Chodera, J. D., & Levinson, N. M. (2017). A water-mediated allosteric network governs activation of Aurora kinase A. *Nature Chemical Biology, 13*, 402–408. https://doi.org/10.1038/nchembio.2296

[10] Ruff, E. F., Muretta, J. M., Thompson, A. R., Lake, E. W., Cyphers, S., Albanese, S. K., et al. (2018). A dynamic mechanism for allosteric activation of Aurora kinase A by activation loop phosphorylation. *eLife, 7*, e32766. https://doi.org/10.7554/eLife.32766

[11] Gilburt, J. A. H., Girvan, P., Blagg, J., Ying, L., & Dodson, C. A. (2019). Ligand discrimination between active and inactive activation loop conformations of Aurora-A kinase is unmodified by phosphorylation. *Chemical Science, 10*, 4069–4076. https://doi.org/10.1039/C8SC03669A

[12] Monteiro da Silva, G., Cui, J. Y., Dalgarno, D. C., Lisi, G. P., & Rubenstein, B. M. (2024). High-throughput prediction of protein conformational distributions with subsampled AlphaFold2. *Nature Communications, 15*, 2464. https://doi.org/10.1038/s41467-024-46715-9

[13] Wayment-Steele, H. K., Ojoawo, A., Otten, R., Apitz, J. M., Pitsawong, W., Hömberger, M., et al. (2024). Predicting multiple conformations via sequence clustering and AlphaFold2. *Nature, 625*(7996), 832–839. https://doi.org/10.1038/s41586-023-06832-9

DOI: 10.54254/2753-8818/2026.HZ38172

[14] Terwilliger, T. C., Liebschner, D., Croll, T. I., Williams, C. J., McCoy, A. J., Poon, B. K., et al. (2024). AlphaFold predictions are valuable hypotheses and accelerate but do not replace experimental structure determination. *Nature Methods, 21*(1), 110–116. https://doi.org/10.1038/s41592-023-02087-4

[15] Gu, X., Aranganathan, A., & Tiwary, P. (2024). Empowering AlphaFold2 for protein conformation selective drug discovery with AlphaFold2-RAVE. *eLife, 13*, RP99702. https://doi.org/10.7554/eLife.99702.3