Nature of the pre-chemistry ensemble in mitogen activated protein kinases

Ranajeet Ghose[a,b,c,d]

[a]Department of Chemistry and Biochemistry, The City College of New York, 160 Convent Avenue, New York, NY 10031 (USA)

PhD Programs in [b]Biochemistry, [c]Chemistry and [d]Physics The Graduate Center of CUNY, New York, NY 10016 (USA)

Abstract

In spite of the availability of a significant amount of structural detail on docking interactions involving mitogen activated protein kinases (MAPKs) and their substrates, the mechanism by which the disordered phospho-acceptor on the substrate transiently interacts with the kinase catalytic elements and is phosphorylated, often with high efficiency, remains poorly understood. In this perspective this dynamic interaction is analyzed in the context of available biophysical and biochemical data for ERK2, an archetypal MAPK. A hypothesis about the nature of the ternary complex involving the MAPK, its substrate and ATP immediately prior to the chemical step (the pre-chemistry complex, PCC) is proposed. It is postulated that the solution ensemble (the pre-chemistry ensemble, PCE) representing the PCC comprises of several conformations that are linked by dynamics on multiple timescales. These individual conformations possess different intrinsic abilities to proceed through the chemical step. The overall rate of chemistry is therefore related to the microscopic nature of the PCE, its constituent conformational microstates, and their intrinsic abilities to yield a phosphorylated product. While characterizing these microstates within the PCE in atomic or near-atomic detail is an extremely challenging proposition, recent developments in hybrid methodologies that employ computational approaches driven by experimental data appear to provide the most promising path forward toward achieving this goal.

Introduction

The mitogen activated protein kinases (MAPKs)[1-5] phosphorylate their substrates on specific serine or threonine residues that precede a proline (a Ser/Thr-Pro motif). This family comprises of "typical" members that include the extracellular signal regulated kinases ERK1 and ERK2[6], p38 isoforms (α, β, γ and δ)[7], c-Jun N-terminal kinases JNK1, JNK2 and JNK3[8], and ERK5[9], in addition to "atypical" members, ERK3, ERK4, ERK7 and the nemo-like kinase (NLK)[10]. MAP kinases typically lie at the downstream end-point of a linear signaling triad that also comprises a MAPK kinase kinase (MAPKKK or MAP3K) and a MAPK kinase (MAPKK or MAP2K). The three-tiered MAP3K-MAP2K-MAPK cascade is upregulated in response to the activation of cell surface receptors by extracellular signals such as hormones, cytokines and growth factors[2;11;12] leading to the phosphorylation of a range of cytosolic and nuclear substrates that include downstream kinases, phosphatases, transcription factors, proteins of the nuclear pore complex, those involved in programmed cell death, to name just a few. The dysregulation of MAPK signaling leads to a variety of human diseases[13] including several forms of cancer[3;14] and a variety of neurological disorders[15;16].

While MAPKs phosphorylate on Ser/Thr residues that are part of a Ser/Thr-Pro motif[4], this sequence alone does not provide sufficient substrate affinity nor does it yield a significant degree of selectivity to distinguish it from other proline-directed kinases such as the cyclin dependent kinases (CDKs)[17]. In order to attain a high level of specificity for its native substrates, MAPKs utilize one of two so-called "docking sites", named D-recruitment site (DRS) and F-recruitment site (FRS) (Figure 1), that are distinct from the catalytic elements that form the kinase active site[18-22]. The DRS is located behind the

ATP-binding pocket and is engaged by partners that contain a D-site consensus sequence $(R/K)_{2-3}$-$(X)_{2-6}$-Φ_A-X-Φ_B ($\Phi_{A/B}$ are hydrophobic residues). The FRS is located just below the activation loop, and preferentially binds partners that contain a consensus F-site sequence (F-X-F-P). Recent biochemical and structural studies indicate that the terminal proline of the F-site may not contribute substantially to the stability of the FRS/F-site interactions[23;24]. The FRS forms fully only in the active kinase[20] upon dual phosphorylation on a conserved T-X-Y motif on the MAPK activation loop. This ensures that interaction partners containing canonical F-site sequences can only engage the active form of the enzyme. It is now becoming increasingly clear that MAPKs also bind partners that do not contain canonical D-site or F-site sequences. These non-canonical interactions also involve parts of the MAPK DRS and FRS, with both regions being simultaneously engaged in some cases, perhaps to compensate for the non-ideal interactions at each docking region[25-27]. The use of multiple interactions that are individually incapable of achieving sufficient affinity likely provides a means to expand the MAPK interactome. While the structural/dynamic basis of docking interactions have been studied extensively[20;25;27;28], the same cannot be said of the interactions of the substrate phospho-acceptor with the MAPK catalytic elements. This perspective explores the mechanisms underlying substrate phosphorylation, its relationship to the docking interactions and to the structural dynamics of the substrate phospho-acceptor. The discussion that follows relies on the large body of experimental data from biochemical and biophysical measurements on the MAPK ERK2. While ERK2 is used as a representative example, many if not most of the principles introduced here are likely to hold for most typical MAPKs, and perhaps for a larger class of kinases.

Proximity-induced Catalysis

It has been shown that almost all of the binding energy for the association of a substrate with ERK2 results from interactions between docking sequences on the former (D-site, F-site) and the corresponding recruitment sites (DRS, FRS) on the latter. The phospho-acceptor Ser/Thr-Pro motif makes little, if any, contribution to the binding energy[29]. Mutations at the ERK2 recruitment sites either completely abolish or substantially compromise interactions with several substrates containing known D-site and/or F-site sequences[30]. Additionally, mutations at the substrate phospho-acceptor motif have no appreciable impact on the binding affinity[31]. Based on these observations, in a mechanism termed *proximity-induced catalysis*[31], it was proposed that the docking interactions serve to increase the effective concentration of the substrate phospho-acceptor moiety near the kinase active site. This increase in concentration enhances the probability of encounters of the phospho-acceptor with the catalytic elements of the kinase thereby increasing the possibility of phosphorylation.

A 13-residue peptide derived from the transcription factor Elk-1 ($Elk_{387-399}$, Figure 2A, human numbering used in all cases), for which substantial biochemical and biophysical data are available[24], represents a simple but optimal system to start exploring various aspects of proximity-induced catalysis by ERK2. $Elk_{387-399}$ contains a single canonical F-site sequence (^{395}FQFP^{398}) C-terminal to a serine (Ser389) that is a known target of ERK2[32]. $Elk_{387-399}$ engages active ERK2 (dual-phosphorylated on T185 and Y187; termed ppERK2) with a moderate affinity (K_D = 8 ± 1 μM) and as expected for a canonical F-site sequence, it does not bind inactive ERK2. $Elk_{387-399}$ is a ppERK2 substrate, albeit a relatively poor one, being phosphorylated with low efficiency *in vitro*

(k_{cat} = 0.6 ± 0.1 s^{-1}, k_{cat}/K_M = 0.5 x 10^5 M^{-1} s^{-1})[24]. In comparison, the corresponding parameters[33] for an "average" enzyme-substrate pair from the BRENDA database[34] are k_{cat} ~10 s^{-1}, k_{cat}/K_M ~ 10^5 M^{-1} s^{-1}. Chemical shift perturbation (CSP) data confirm that there is no long-distance allosteric coupling between the DRS and the $Elk_{387-399}$-occupied FRS of ppERK2 in the $Elk_{387-399}$•ppERK2 complex. The region around the phospho-acceptor Ser389 shows only small CSPs (the largest are seen on Ser389 itself and on Pro390) suggesting minimal stable interactions with the ERK2 catalytic site[24].

It has been noted that a majority of sites targeted by protein kinases are located in intrinsically disordered regions (IDRs) on their substrate proteins[35]. A necessary feature of the proximity-induced catalysis mechanism, discussed above, is the persistence of these dynamics in the complex involving the kinase and its substrate and the absence of stable interactions between the phospho-acceptor and the kinase catalytic elements on the timescale of chemistry (typically 10-100 ms). Indeed, in addition to the lack of substantial CSPs in and around the phospho-acceptor Ser389 of $Elk_{387-399}$ when bound to ppERK2, the spin-spin relaxation time (T_2) for its amide ^{15}N was also found to be significantly longer than values recorded for the backbone amides of the ppERK2 core residues[24]. This suggests that Ser389 remains highly disordered on the fast, picosecond-nanosecond timescale in the $Elk_{387-399}$•ppERK2 complex.

Given the persistence of dynamics at the phosphorylation site of $Elk_{387-399}$ and its lack of a role in stabilizing the $Elk_{387-399}$•ppERK2 complex, the docked substrate can be approximated by a simple toy model shown in Figure 2B. This model assumes that the F-site sequence of $Elk_{387-399}$ and the ERK2 FRS where it is rigidly docked together comprise a single entity (infinitely strong binding) to which the phospho-acceptor (Ser389) is

tethered by a flexible linker of length L when fully extended. This model allows an estimation of the effective concentration (C_{eff}) of the phospho-acceptor near the kinase catalytic site. Assuming the linker to be a Gaussian chain with an instantaneous orientation $\Omega(r,\theta,\phi)$ with $P(\Omega;D,\xi,L)$ as the related probability distribution, C_{eff} is given by[36]

$$C_{eff} = \frac{1}{N_A}\int P(\Omega; D, \xi, L)d\Omega = \frac{1}{N_A}\left(\frac{3}{4\pi L\xi}\right)^{3/2} exp\left(-\frac{3D^2}{4L\xi}\right) \quad (1)$$

where ξ is the persistence length of the linker (a measure of its flexibility), D is the distance between the docking site and the catalytic site (both considered to be point objects) and N_A is the Avogadro number. For a given D, this model predicts a monotonic increase of C_{eff} with increasing L until a critical length $L_C = D^2/2\xi$ is reached; for $L > L_C$, C_{eff} decreases with increasing L (see Figure 2C). Figure 2D illustrates the dependence of C_{eff} on D for two different values of L, corresponding to the positions of Ser389 (separated from the F-site sequence by 5 residues; L = 17.5 Å) or Ser383 (separated by 11 residues, L = 38.5 Å) and two different values of ξ = 3.9 Å, consistent with a value measured in unfolded proteins[37] and ξ = 7.8 Å. For perspective, a 36-unit helical segment consisting of N-substituted glycines has a ξ value of 10.5 Å[38]. This analysis suggests that Ser389 should be more efficiently phosphorylated than the more distant Ser383 given the substantially lower C_{eff} in the latter case. This prediction runs counter to the experimental observation demonstrating the higher efficiency of Ser383 phosphorylation compared to Ser389[39]. In the context of the toy model, an approximately 2-fold higher rigidity of the linker connecting Ser389 to the F-site sequence compared to the corresponding Ser383 linker could yield comparable C_{eff} values. While prior phosphorylation has been shown result in

reduced flexibility in some cases[40] (prior phosphorylation on Ser383 could potentially alter the ξ value relevant to Ser389 phosphorylation), the required ~2-fold change in flexibility appears highly improbable.

The Pre-chemistry Complex (PCC) and the Chemical Step

It is evident that docking interactions and the resulting increased C_{eff} *does* play a significant role in enhancing substrate phosphorylation since in the absence of docking the recognition of the phospho-acceptor would be fully diffusion limited and thus, highly inefficient. Therefore the above exercise has some merit. However, simple arguments based on increased local concentration alone cannot explain the nuances of phospho-transfer e.g. the efficiency of phosphorylation at various target sites by ERK2 or other MAPKs. The linker, the phospho-acceptor and the kinase active site are not featureless point objects, they are comprised of chemical entities that are capable of a variety of non-covalent interactions. It is evident that the process of phosphorylation requires the formation of a suitable geometry to enable the optimal transfer of the γ-phosphate of ATP to the -OH moeity of the phospho-acceptor. This involves appropriate orientations of the -OH group of the Ser/Thr phospho-acceptor on the substrate, the γ-phosphate of ATP, the kinase catalytic base (D149 in ERK2), a basic residue (K151 in ERK2) that serves to activate the γ-phosphate, in addition to the catalytic Mg^{2+} ions. A recent QM/MM study[41] has provided clues into what this optimal conformation could be (see Figure 3). These studies also clarified the requirement for a proline residue in a *trans* configuration at the P+1 position to properly target the phospho-acceptor into the kinase active site by reducing steric clashes. Indeed, CSPs on the $Elk_{387-399}$ proline resonances support this prediction for Pro390. This contrasts Pro398 of the FQFP motif where CSPs suggest the

absence of a preferred conformation[24] required to engage ppERK2. Thus, the fallacious assumption inherent to the toy model described above is that all configurations of the phospho-acceptor (and the kinase catalytic elements) would be equally capable of producing a phosphorylated product i.e. a high C_{eff} value would necessarily imply perfect chemistry and a high phosphorylation rate. Thus, in order to proceed beyond a featureless polymer model, several issues need to be considered. These include the kinetic processes that lead to the formation of the pre-chemistry complex (PCC) involving ppERK2, the substrate and ATP (or a suitable non-hydrolysable mimic), the structure of the PCC in a given case and its deviation from the geometry that is optimal for chemistry (e.g. from the work of Turjanski[41]).

The process of binding and phosphorylation of a substrate by ppERK2 may be adequately described by Scheme I under ATP-saturating conditions. The first step in this scheme involves docking interactions between ATP-loaded ppERK2 (E) and the substrate (S) phospho-acceptor, the second involves the chemical step, and the final step involves product release (of both ADP and phosphorylated substrate; though these may occur at different rates[42] and should, in principle, be represented by a more elaborate scheme). This scheme also assumes that the chemical step and the product release steps are irreversible. While this simplified scheme is sufficient in the context of the present discussion, readers are referred to the formalism of Northrup and Hynes[43] for a more rigorous approach. The Michaelis-Menten parameters (K_M and k_{cat}) obtained from standard steady-state measurements are related to the rates shown in Scheme I by

$$k_{cat} = \frac{k_3 k_4}{(k_3 + k_4)}$$

$$K_M = \frac{k_4(k_{-2} + k_3)}{k_2(k_3 + k_4)}$$

(2)

Perturbations in any of the states shown in Scheme I with alteration of any of the corresponding rates, can therefore potentially affect both K_M and k_{cat}.

The present discussion focuses on the events after formation of the PCC (E●ATP●S; shown in red in Scheme I) through the transition state and chemistry to yield product (E●ADP●pS). The rate relevant to this process is k_3. It is to be noted that the overall rate of phosphorylation measured experimentally may, in principle, depend on any or all of the rates in Scheme I (or a more complex version thereof). However, for ERK2 mediated substrate phosphorylation, the chemical step (k_3) appears to be at least partially rate-limiting[44-46] and therefore has a significant influence on the measured phosphorylation rate. In the framework of Arrhenius theory, k_3 is determined by the activation energy (E_a) separating the PCC from the corresponding transition state along the reaction co-ordinate

$$k_3 = Aexp\left(-\frac{E_a}{kT}\right)$$

(3)

A is a pre-exponential factor. It is generally difficult to define a reaction co-ordinate in a conventional sense for the phospho-transfer reaction given the significant number of degrees of freedom that couple to the chemical step (see below). Turjanskii et. al[41] rather than using a single distance to represent the reaction co-ordinate, utilized the following linear combination of distances: d(O3β-Pγ)+d(Oγ/γ1-Hγ/γ1)-d(Pγ-Oγ/γ1)-d(Hγ/γ1-Oδ); O3β bridges the β-phosphorus to the γ-phosphorus, Pγ of ATP; Oγ/γ1 and Hγ/γ1 belong to the

–OH of the phospho-acceptor serine/threonine; Oδ is the carboxyl oxygen of the catalytic aspartate. Using this reaction co-ordinate an E_a value of ~14 kcal/mol was obtained. This value is similar (15 kcal/mol) to that calculated assuming a k_3 value of 100 s^{-1} and a pre-exponential factor of 5 ps^{-1}; the former is consistent with values expected for a "good" ERK2 substrate e.g. the transcription factor Ets-1 (discussed more extensively below)[42]. However, given the fact that the PCC is highly dynamic, an approach based on simple transition state theory is inadequate to account for the role of conformational disorder in the PCC and to describe the process that culminates with the formation of the E●ADP●pS state. Indeed, it should be realized that the each of the species represented in Scheme I, including the PCC, does not comprise of a single structure in solution but rather an ensemble of several conformational states (microstates) linked by dynamics on multiple timescales. Thus the PCC is in reality a pre-chemistry ensemble (PCE) containing several exchanging microstates.

Conformational Dynamics in the PCC and the Pre-chemistry Ensemble (PCE)

The role of dynamics in catalysis has been the source of intense debate for several years with apparently conflicting views put forward by several authors[47;48] and others who have attempted to reconcile them[49]. In the following discussion, no specific rate enhancements due to dynamics are claimed. However, it is suggested that the native i.e. equilibrium (statistical) dynamics of the PCC (encoded by the PCE) affect the ability of the system to attain the appropriate transition state or rather, transition state ensemble[50], that influences chemistry. The dynamics discussed here do not include those suggested to be involved in the crossing of the transition state or the non-statistical promoting motions proposed by Schwartz and Schramm[51-53].

The PCC possesses a range of motions represented by a set of conformational modes Q, only a small subset of which, $Q_{cat} \subseteq Q$, couple to an appropriately defined reaction co-ordinate[54]. Thus, the Q_{cat} represent conformations that allow the chemical step to proceed i.e. those for which $k_3 > 0$. In general, these conformational modes exchange with each other or with other modes Q' ($Q' \subseteq Q$; $Q' \not\subset Q_{cat}$) with rates (k_Q) that potentially range over several orders of magnitude in timescales. Exchange between conformations i.e. conformational dynamics with rates much larger than that of the chemical step (k_3 ~100 s^{-1}) do not directly influence the rate of chemistry and k_3 in such cases is determined by an effective conformation that is averaged over these fast modes. Thus, in the limiting case where all dynamics within the PCE involve rate constants $k_Q >> k_3$, both single molecule and ensemble measurements of pre-steady state kinetics using the framework of Scheme I will yield a single k_3 value averaged over the fast modes. Exchange between conformations involving free-energy barriers that are substantially higher than the thermal energy occurs with rates that are significantly slower than k_3 ($k_3 >> k_Q$). In such cases, in relation to chemistry, the system may be treated as comprising of a set of unique PCCs each with their own activation energies (in Equation (3); and in general, also with their own pre-exponential factors) and consequently a set of unique k_3 values. This distribution of k_3 values can be obtained from single-molecule measurements[55]. It is notable that a distribution of turnover rates have indeed been seen in single-molecule measurements on the catalytic domain of protein kinase A (PKA)[56] and these can reasonably be attributed to a distribution of k_3 rates due to slow dynamics. On the other hand, ensemble measurements of pre-steady-state kinetics in the context of Scheme I would yield a population-weighted k_3 value given by

$$\langle k_3 \rangle = \sum p_{conf} k_{3,conf}; \quad \sum p_{conf} = 1$$

$$\langle k_3 \rangle = \int P(k_{3,conf}) k_{3,conf} dk_{3,conf}; \quad \int P(k_{3,conf}) dk_{3,conf} = 1$$

(4)

The first line in Equation (4) corresponds to a case where the PCE can be formally parsed into discrete conformations with conformation-dependent k_3 values ($k_{3,conf}$) with the corresponding normalized populations given by p_{conf}. The second line is valid for a continuous distribution, $P(k_{3,conf})$ of $k_{3,conf}$ values. A point to note, however, is that motions with rates faster than k_3 *do* impact the ability to form the appropriate transition state and therefore affect chemistry. The existance of certain modes may prevent the formation of the ideal transition state and therefore the conformation averaged over the fast modes may deviate significantly from one that is optimal for chemistry and thus yield a sub-par k_3 value.

Thus, some of the various motional modes that exist in the PCC are conducive to the formation of an optimal transition state resulting in the highest possible k_3. Therefore, one may surmise that the PCE comprises of distinct conformational microstates (represented by Q_{cat}) that possess different intrinsic abilities for chemistry i.e. to yield E●ADP●pS (that is, in general, represented by its own distinct ensemble). The states that we term "high-activity" states yield k_3 rates that are substantially higher than those for the "low-activity" states (Figure 4)[27]. The rate of exchange between these microstates determines whether a single k_3 value (fast exchange relative to k_3) or a distribution of k_3 values (slow exchange relative to k_3) will be obtained in single molecule measurements. Therefore, a "good" substrate would possess the ability to generate a PCE in solution that efficiently populates the high-activity microstates compared to the low-activity microstates

in contrast to a “bad” substrate (Figure 4) whose PCE would predominantly populate the low-activity microstates.

It may be possible, in principle, to perturb the relative distribution of “high” and “low” activity microstates in the PCE of ERK2 (and other MAPKs) through a variety of means. One way to achieve this could be to alter the relative positioning of the docking and phosphorylation sequences on the substrate. Anecdotal data suggests that this is indeed a viable means of modulating activity given that in the case of Elk-1, mentioned above, the distal S383 appears to be a better substrate than S389 that is closer to the FRS/F-site docking interaction. However, this needs to be tested rigorously through measurements of enzyme kinetics on well-designed peptide substrates containing individual phosphorylation sites rather than in the context of peptides containing multiple sites and complex phosphorylation kinetics[39]. More direct evidence of the ability to tune the phosphorylation efficiency by the location of the phospho-acceptor relative to the docking site comes from the ERK2 substrate Ets-1.

The transcription factor Ets-1[57] is an excellent substrate of ERK2 (k_{cat}=14 ± 2 s^{-1}, k_{cat}/K_M = 6 x 10^5 $M^{-1}s^{-1}$) that interacts with the kinase without utilizing either a canonical D-site sequence or a canonical F-site sequence. Ets-1 binds ERK2 in a non-canonical bidentate fashion, partially engaging both the ERK2 DRS and the ERK2 FRS. A disordered segment on the N-terminus of Ets-1 interacts with the hydrophobic portion of the ERK2 DRS (ϕ_{hyd}, Figure 1) and the ERK2 FRS takes part in a rigid body interaction with Ets-1 PNT domain. The phospho-acceptor (Thr38) is sandwiched between these two distinct docking interactions (Figure 5). Interestingly, a partially disorded helix (α0) that bridges Thr38 to the PNT domain becomes further disordered upon interacting with

ERK2[27]. NMR relaxation measurements demonstrate that like in the case of the $Elk_{387-399}$ peptide, the region around phospho-acceptor, and indeed the Thr38 sidechain, remains disordered upon engaging ERK2. This system is well-suited to test whether the efficiency phosphorylation could be altered by moving the phospho-acceptor towards or away from the rigid PNT domain (Figure 5). Given that the N-terminus of Ets-1 remains significantly dynamic when bound to ERK2, one may expect that a shift away from the PNT domain would have a smaller effect on the catalytic parameters than a shift towards the PNT domain where a larger destabilization of the $\alpha 0$ helix would be necessary to optimally engage the ERK2 active site. As predicted by Equation (2), an alteration in the rate of the chemical step i.e k_3 (and possibly in k_4 i.e. product release) without affecting the docking interaction/s should affect both catalytic parameters. Altering the position of the phospho-acceptor with respect to the two docking interaction indeed produces the predicted effects as illustrated in Figure 5. Whether these effects indeed arise from alteration in k_3 values needs to be confirmed by carefully performed pre-steady-state kinetics measurements.

In addition to the location of the phospho-acceptor with respect to the docking site/s, it is also possible that the sequences around the phospho-acceptor site play a role in tuning the nature of the PCE. Therefore, mutation of one or more residues in this region should lead to an altered phosphorylation efficiency; an extreme case would be the replacement of the proline at the P+1 position. In Ets-1, a Pro39Ala mutation still allows phosphorylation to proceed but with a 10-fold reduction in the k_{cat}/K_M value compared to the wild-type sequence (Figure 5). The QM/MM simulations of Turjanski suggest this loss of efficiency may be attributed to the inability of a more flexible alanine to optimally and stably orient the phospho-acceptor at the MAPK active site compared to a more rigid

proline[41]. Beyond the P+1 proline, the preference for certain amino acid residues at positions ranging from P-4 to P+4 has been experimentally demonstrated for MAPKs[58]. This preference could, in principle, reflect changes in the solution PCE imposed by the range of conformations accessible to the sequence neighboring the phospho-acceptor. Indeed, computational analyses have revealed distinct sequence-to-conformation relationships in IDRs attached to folded domains[59], a situation that is applicable to a disordered loop bearing the phospho-acceptor rigidly docked to the kinase docking site/s. Another potential regulatory mechanism could involve pre-phosphorylation at other sites to alter the conformation and/or the rigidity of the loop bearing the phospho-acceptor[40]. Such a mechanism may be operative in Elk-1 which contains as many as six phosphorylatable serines or threonines that are processed at significantly different rates in the loop separating the D-site and F-site sequences[39].

Towards an Atomistic Description of the PCE

Whether any or all of the perturbations described above, and perhaps others, modulate activity by altering the relative proportion of the high-activity and low-activity states in the PCE, and indeed whether the PCE parses into such states remains to be assessed. A true test of this hypothesis will require characterization of the PCE in atomic detail. This is an extremely challenging proposition that is unlikely to be accomplished either by experimental or computational means alone. The NMR and biochemical methodology[24;27] described above that led us to postulate the presence of high-activity and low-activity microstates of the PCE involve ensemble averaged data that are potentially ambiguous and incomplete. Methods such as FRET, including at the single-molecule level[60], or DEER[61;62], though capable of providing information about

distributions within the solution ensemble, utilize only a very small number of probes. Conventional computational techniques, on the other hand, suffer from an inability to achieve the relevant timescales to allow adequate sampling of the conformational landscape, though enhanced sampling methods have served to somewhat alleviate this issue[63;64]. Also plaguing these *in silico* approaches are possible inaccuracies in the force-fields utilized in the simulations especially when applied to systems containing a significant degree of disorder[65], though recent developments are promising[66]. It is therefore clear that a hybrid approach that employs synergy between computational and experimental strategies is necessary in systems with a significant degree of disorder, as elegantly outlined by Bonomi et. al[67]. Such hybrid strategies employing experimental data from NMR[68], room-temperature X-ray crystallography[69], X-ray free-electron lasers[70] or electron microscopy[71] measurements have provided significant insight into the conformational states populated by proteins and their complexes.

In the present case, given that experimental NMR data including chemical shifts and relaxation rates can be collected in a PCC involving ERK2[24;27] and other MAPKs[72], their substrates and ATP, suggests that these data may be incorporated into a computational strategy to probe the corresponding PCE. A possible approach would be to bias the computational conformational sampling *a priori* using a subset of the NMR data (e.g. chemical shifts) while relying on additional data (e.g. spin-relaxation rates or residual dipolar couplings from NMR or distance distributions from DEER or FRET experiments) as a means to validate the ensemble determined by the hybrid approach. It is important, however, to realize that the experimental data, in addition to being incomplete, contain both random and possibly, systematic errors arising from

experimental imprecision, sample deterioration, use of imperfect models for the analysis of experimental results or hardware issues. These errors have to be accounted for in rigorous fashion when biasing the *in silico* conformational sampling. The newly proposed metadynamics metainference approach[73] attempts to achieve precisely this. In this author's opinion, this method may provide the best way forward in characterizing the MAPK PCE. However, this approach has been only been successfully tested on short peptides[74], and its ability to faithfully report on systems as challenging, both experimentally and computationally, as a MAPK PCC, is yet to be evaluated. In a subset of cases (e.g. for the PCC involving $Elk_{387-399}$, described above) it could be appropriate to reduce the size of the simulated system by focusing on specific regions e.g. the active site of ppERK2 and the loop bearing the phospho-acceptor. However, in other cases, this may be somewhat tricky due to the presence of significant long-range effects e.g. between the docking interactions and phosphorylation through coupling of the docking site/s with the kinase active site[75]. Experimental identification of such effects *a priori* e.g. through detailed chemical shift analysis[76] may help validate whether a simplified computational system is appropriate or not in a given case. Alternative approaches that, rather than using experimental data to bias the simulations, rely on selecting[77] or re-weighting[78] the conformers from independently generated ensembles (these ensembles clearly have to be large enough to be representative) *a posteriori* using experimental data, have also been proposed. These approaches may also be useful in some specific cases. It is clear, however, that whatever approach is deployed with a goal to determine the nature of the MAPK PCE in atomic or near-atomic detail, significant optimization, validation and refinement of currently available strategies will be needed. Whether these

efforts will be successful remains to be seen but it is certain that important methodological and practical insights will emerge from such endeavors.

Final Thoughts

This perspective touches on many broad topics about MAPKs and docking interactions involving them, the role of dynamics in catalysis and computational approaches to determine structural ensembles being some of them. There are several excellent treatises on each of these topics available in the literature. The aim of this perspective was not to review any of these topics exhaustively. The principal goal in this perspective was to provide a view of the MAPK PCC as a dynamic ensemble, the PCE, rather than a single static structure, and to speculate on the influence of these extensive dynamics on the overall efficiency of chemistry. While characterizing this ensemble in atomic/near-atomic detail remains a challenging proposition, current advances in hybrid methodology suggest a path forward towards achieving this goal in the future. I apologize to the numerous authors who have made seminal contributions to each of the topics mentioned earlier but whose work has not been cited.

Acknowledgements

This work is supported by the award PHY1811770 from the National Science Foundation. The author thanks Drs. Kevin Dalby (UT Austin), Andrea Piserchio (CCNY; also for making Figure 3), Sebastien Alphonse (CCNY) and Paul Robustelli (DE Shaw Research) for their critical reading of this manuscript.

References

1. Cargnello, M. & Roux, P. P. (2011). Activation and function of the MAPKs and their substrates, the MAPK-activated protein kinases. Microbiol. Mol. Biol. Rev. 75, 50-83.
2. Chen, Z., Gibson, T. B., Robinson, F., Silvestro, L., Pearson, G., Xu, B., Wright, A., Vanderbilt, C. & Cobb, M. H. (2001). MAP kinases. Chem. Rev. 101, 2449-2476.
3. Dhillon, A. S., Hagan, S., Rath, O. & Kolch, W. (2007). MAP kinase signalling pathways in cancer. Oncogene 26, 3279-3290.
4. Roskoski, R. (2012). ERK1/2 MAP kinases: structure, function, and regulation. Pharmacol. Res. 66, 105-143.
5. Tanoue, T., Adachi, M., Moriguchi, T. & Nishida, E. (2000). A conserved docking motif in MAP kinases common to substrates, activators and regulators. Nature Cell Biol. 2, 110-116.
6. Yoon, S. & Seger, R. (2006). The extracellular signal-regulated kinase: multiple substrates regulate diverse cellular functions. Growth Factors 24, 21-44.
7. Cuadrado, A. & Nebreda, A. R. (2010). Mechanisms and functions of p38 MAPK signalling. Biochem. J. 429, 403-17.
8. Weston, C. R. & Davis, R. J. (2007). The JNK signal transduction pathway. Curr. Opin. Cell Biol. 19, 142-9.
9. Nithianandarajah-Jones, G. N., Wilm, B., Goldring, C. E., Muller, J. & Cross, M. J. (2012). ERK5: structure, regulation and function. Cell Signal. 24, 2187-96.
10. Coulombe, P. & Meloche, S. (2007). Atypical mitogen-activated protein kinases: structure, regulation and functions. Biochim. Biophys. Acta 1773, 1376-87.
11. Murphy, L. O. & Blenis, J. (2006). MAPK signal specificity: the right place at the right time. Trends Biochem. Sci. 31, 268-275.
12. Pearson, G., Robinson, F., Beers Gibson, T., Xu, B. E., Karandikar, M., Berman, K. & Cobb, M. H. (2001). Mitogen-activated protein (MAP) kinase pathways: regulation and physiological functions. Endocrine Rev. 22, 153-183.
13. Kim, E. K. & Choi, E. J. (2010). Pathological roles of MAPK signaling pathways in human diseases. Biochim. Biophys. Acta 1802, 396-405.
14. Sullivan, R. J. & Flaherty, K. (2013). MAP kinase signaling and inhibition in melanoma. Oncogene 32, 2373-9.
15. Munoz, L. & Ammit, A. J. (2010). Targeting p38 MAPK pathway for the treatment of Alzheimer's disease. Neuropharmacol. 58, 561-8.
16. Alfieri, P., Cesarini, L., Mallardi, M., Piccini, G., Caciolo, C., Leoni, C., Mirante, N., Pantaleoni, F., Digilio, M. C., Gambardella, M. L., Tartaglia, M., Vicari, S., Mercuri, E. & Zampino, G. (2011). Long term memory profile of disorders associated with dysregulation of the RAS-MAPK signaling cascade. Behav. Genet. 41, 423-9.
17. Ubersax, J. A. & Ferrell, J. E. (2007). Mechanisms of specificity in protein phosphorylation. Nature Rev. Mol. Cell. Biol. 8, 530-541.
18. Sharrocks, A. D., Yang, S. H. & Galanis, A. (2000). Docking domains and substrate-specificity determination for MAP kinases. Trends Biochem. Sci. 25, 448-53.

19. Tanoue, T. & Nishida, E. (2003). Molecular recognitions in the MAP kinase cascades. Cell. Signal. 15, 455-462.
20. Lee, T., Hoofnagle, A. N., Kabuyama, Y., Stroud, J., Min, X., Goldsmith, E. J., Chen, L., Resing, K. A. & Ahn, N. G. (2004). Docking motif interactions in MAP kinases revealed by hydrogen exchange mass spectrometry. Mol. Cell 14, 43-55.
21. Reményi, A., Good, M. C. & Lim, W. A. (2006). Docking interactions in protein kinase and phosphatase networks. Curr. Opin. Struct. Biol. 16, 676-685.
22. Peti, W. & Page, R. (2013). Molecular basis of MAP kinase regulation. Protein Sci. 22, 1698-1710.
23. Sheridan, D. L., Kong, Y., Parker, S. A., Dalby, K. N. & Turk, B. E. (2008). Substrate discrimination among mitogen-activated protein kinases through distinct docking sequence motifs. J. Biol. Chem. 283, 19511-19520.
24. Piserchio, A., Ramakrishan, V., Wang, H., Kaoud, T. S., Arshava, B., Dutta, K., Dalby, K. N. & Ghose, R. (2015). Structural and dynamic features of f-recruitment site driven substrate phosphorylation by ERK2. Sci. Rep. 5, 11127.
25. Mace, P. D., Wallez, Y., Egger, M. F., Dobaczewska, M. K., Robinson, H., Pasquale, E. B. & Riedl, S. J. (2013). Structure of ERK2 bound to PEA-15 reveals a mechanism for rapid release of activated MAPK. Nature Commun. 4, 1681-1681.
26. Piserchio, A., Warthaka, M., Devkota, A. K., Kaoud, T. S., Lee, S., Abramczyk, O., Ren, P., Dalby, K. N. & Ghose, R. (2011). Solution NMR insights into docking interactions involving inactive ERK2. Biochemistry 50, 3660-72.
27. Piserchio, A., Warthaka, M., Kaoud, T. S., Callaway, K., Dalby, K. N. & Ghose, R. (2017). Local destabilization, rigid body, and fuzzy docking facilitate the phosphorylation of the transcription factor Ets-1 by the mitogen-activated protein kinase ERK2. Proc. Natl. Acad. Sci. USA 114, E6287-E6296.
28. Zeke, A., Bastys, T., Alexa, A., Garai, A., Meszaros, B., Kirsch, K., Dosztanyi, Z., Kalinina, O. V. & Remenyi, A. (2015). Systematic discovery of linear binding motifs targeting an ancient protein interaction surface on MAP kinases. Mol. Syst. Biol. 11, 837.
29. Yang, S. H., Yates, P. R., Whitmarsh, A. J., Davis, R. J. & Sharrocks, A. D. (1998). The Elk-1 ETS-domain transcription factor contains a mitogen-activated protein kinase targeting motif. Mol. Cell. Biol. 18, 710-20.
30. Burkhard, K. A., Chen, F. & Shapiro, P. (2011). Quantitative analysis of ERK2 interactions with substrate proteins: roles for kinase docking domains and activity in determining binding affinity. J. Biol. Chem. 286, 2477-2485.
31. Rainey, M. A., Callaway, K., Barnes, R., Wilson, B. & Dalby, K. N. (2005). Proximity-induced catalysis by the protein kinase ERK2. J. Am. Chem. Soc. 127, 10494-10495.
32. Fantz, D. A., Jacobs, D., Glossip, D. & Kornfeld, K. (2001). Docking sites on substrate proteins direct extracellular signal-regulated kinase to phosphorylate specific residues. J. Biol. Chem. 276, 27256-27265.
33. Bar-Even, A., Noor, E., Savir, Y., Liebermeister, W., Davidi, D., Tawfik, D. S. & Milo, R. (2011). The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. Biochemistry 50, 4402-10.

34. Schomburg, I., Jeske, L., Ulbrich, M., Placzek, S., Chang, A. & Schomburg, D. (2017). The BRENDA enzyme information system-From a database to an expert system. J. Biotechnol. 261, 194-206.
35. Iakoucheva, L. M., Radivojac, P., Brown, C. J., O'Connor, T. R., Sikes, J. G., Obradovic, Z. & Dunker, A. K. (2004). The importance of intrinsic disorder for protein phosphorylation. Nucleic Acids Res. 32, 1037-49.
36. Van Valen, D., Haataja, M. & Phillips, R. (2009). Biochemistry on a leash: the roles of tether length and geometry in signal integration proteins. Biophys J. 96, 1275-92.
37. Carrion-Vazquez, M., Oberhauser, A. F., Fowler, S. B., Marszalek, P. E., Broedel, S. E., Clarke, J. & Fernandez, J. M. (1999). Mechanical and chemical unfolding of a single protein: a comparison. Proc. Natl. Acad. Sci. USA 96, 3694-9.
38. Rosales, A. M., Murnen, H. K., Kline, S. R., Zuckermann, R. N. & Segalman, R. A. (2012). Determination of the persistence length of helical and non-helical polypeptoids in solution. Soft Matter 8, 3673-3680.
39. Mylona, A., Theillet, F. X., Foster, C., Cheng, T. M., Miralles, F., Bates, P. A., Selenko, P. & Treisman, R. (2016). Opposing effects of Elk-1 multisite phosphorylation shape its response to ERK activation. Science 354, 233-237.
40. Chin, A. F., Toptygin, D., Elam, W. A., Schrank, T. P. & Hilser, V. J. (2016). Phosphorylation increases persistence length and end-to-end distance of a segment of tau protein. Biophys. J. 110, 362-371.
41. Turjanski, A. G., Hummer, G. & Gutkind, J. S. (2009). How mitogen-activated protein kinases recognize and phosphorylate their targets: A QM/MM study. J. Am. Chem. Soc. 131, 6141-6148.
42. Callaway, K., Waas, W. F., Rainey, M. A., Ren, P. & Dalby, K. N. (2010). Phosphorylation of the transcription factor Ets-1 by ERK2: rapid dissociation of ADP and phospho-Ets-1. Biochemistry 49, 3619-3630.
43. Northrup, S. A. & Hynes, J. T. (1980). The stable states picture of chemical reactions. I. Formulation for rate constants and initial condition effects. J. Chem. Phys. 73, 2700-2714.
44. Prowse, C. N., Hagopian, J. C., Cobb, M. H., Ahn, N. G. & Lew, J. (2000). Catalytic reaction pathway for the mitogen-activated protein kinase ERK2. Biochemistry 39, 6258-6266.
45. Prowse, C. N. & Lew, J. (2001). Mechanism of activation of ERK2 by dual phosphorylation. J. Biol. Chem. 276, 99-103.
46. Waas, W. F., Rainey, M. A., Szafranska, A. E. & Dalby, K. N. (2003). Two rate-limiting steps in the kinetic mechanism of the serine/threonine specific protein kinase ERK2: a case of fast phosphorylation followed by fast product release. Biochemistry 42, 12273-12286.
47. Hammes-Schiffer, S. & Benkovic, S. J. (2006). Relating protein motion to catalysis. Annu. Rev. Biochem. 75, 519-541.
48. Kamerlin, S. C. L. & Warshel, A. (2010). At the dawn of the 21st century: Is dynamics the missing link for understanding enzyme catalysis? Proteins 78, 1339-1375.
49. Kohen, A. (2015). Role of Dynamics in Enzyme Catalysis: Substantial versus Semantic Controversies. Acc. Chem. Res. 48, 466-473.

50. Ma, B., Kumar, S., Tsai, C. J., Hu, Z. & Nussinov, R. (2000). Transition-state ensemble in enzyme catalysis: possibility, reality, or necessity? J. Theor. Biol. 203, 383-97.
51. Antoniou, D. & Schwartz, S. D. (2011). Protein dynamics and enzymatic chemical barrier passage. J. Phys. Chem. B 115, 15147-58.
52. Schwartz, S. D. & Schramm, V. L. (2009). Enzymatic transition states and dynamic motion in barrier crossing. Nature Chem. Biol. 5, 551-558.
53. Silva, R. G., Murkin, A. S. & Schramm, V. L. (2011). Femtosecond dynamics coupled to chemical barrier crossing in a Born-Oppenheimer enzyme. Proc. Natl. Acad. Sci. USA 108, 18661-5.
54. Min, W., Xie, X. S. & Bagchi, B. (2009). Role of conformational dynamics in kinetics of an enzymatic cycle in a nonequilibrium steady state. J. Chem. Phys. 131, 065104.
55. Mickert, M. J. & Gorris, H. H. (2018). Transition-state ensembles navigate the pathways of enzyme catalysis. J. Phys. Chem. B 122, 5809-5819.
56. Sims, P. C., Moody, I. S., Choi, Y., Dong, C., Iftikhar, M., Corso, B. L., Gul, O. T., Collins, P. G. & Weiss, G. A. (2013). Electronic measurements of single-molecule catalysis by cAMP-dependent protein kinase A. J. Am. Chem. Soc. 135, 7861-8.
57. Sharrocks, A. D. (2001). The ETS-domain transcription factor family. Nature Rev. Mol. Cell. Biol. 2, 827-837.
58. Songyang, Z., Lu, K. P., Kwon, Y. T., Tsai, L. H., Filhol, O., Cochet, C., Brickey, D. A., Soderling, T. R., Bartleson, C., Graves, D. J., DeMaggio, A. J., Hoekstra, M. F., Blenis, J., Hunter, T. & Cantley, L. C. (1996). A structural basis for substrate specificities of protein Ser/Thr kinases: primary sequence preference of casein kinases I and II, NIMA, phosphorylase kinase, calmodulin-dependent kinase II, CDK5, and Erk1. Mol. Cell. Biol. 16, 6486-93.
59. Mittal, A., Holehouse, A. S., Cohan, M. C. & Pappu, R. V. (2018). Sequence-to-Conformation Relationships of Disordered Regions Tethered to Folded Domains of Proteins. J. Mol. Biol. 430, 2403-2421.
60. Schuler, B., Soranno, A., Hofmann, H. & Nettels, D. (2016). Single-molecule FRET spectroscopy and the polymer physics of unfolded and intrinsically disordered proteins. Annu. Rev. Biophys. 45, 207-31.
61. Reginsson, G. W. & Schiemann, O. (2011). Pulsed electron-electron double resonance: beyond nanometre distance measurements on biomacromolecules. Biochem. J. 434, 353-63.
62. Krumkacheva, O. & Bagryanskaya, E. (2017). EPR-based distance measurements at ambient temperature. J. Magn. Reson. 280, 117-126.
63. Abrams, C. & Bussi, G. (2014). Enhanced sampling in molecular dynamics using metadynamics, replica-exchange, and temperature-acceleration. Entropy 16, 163-199.
64. Miao, Y. & McCammon, J. A. (2016). Unconstrained enhanced sampling for free energy calculations of biomolecules: a review. Mol. Simul. 42, 1046-1055.
65. Palazzesi, F., Prakash, M. K., Bonomi, M. & Barducci, A. (2015). Accuracy of current all-atom force-fields in modeling protein disordered states. J. Chem. Theory Comput. 11, 2-7.

66. Robustelli, P., Piana-Agostinetti, S. & Shaw, D. E. (2017). Developing a molecular dynamics force field for both folded and disordered protein states. Proc. Natl. Acad. Sci. USA, 115, E4758-E4766.
67. Bonomi, M., Heller, G. T., Camilloni, C. & Vendruscolo, M. (2017). Principles of protein structural ensemble determination. Curr Opin Struct Biol 42, 106-116.
68. Lindorff-Larsen, K., Best, R. B., DePristo, M. A., Dobson, C. M. & Vendruscolo, M. (2005). Simultaneous determination of protein structure and dynamics. Nature 433, 128-132.
69. Fraser, J. S., van den Bedem, H., Samelson, A. J., Lang, P. T., Holton, J. M., Echols, N. & Alber, T. (2011). Accessing protein conformational ensembles using room-temperature X-ray crystallography. Proc. Natl. Acad. Sci. USA 108, 16247-52.
70. Keedy, D. A., Kenner, L. R., Warkentin, M., Woldeyes, R. A., Hopkins, J. B., Thompson, M. C., Brewster, A. S., Van Benschoten, A. H., Baxter, E. L., Uervirojnangkoorn, M., McPhillips, S. E., Song, J., Alonso-Mori, R., Holton, J. M., Weis, W. I., Brunger, A. T., Soltis, S. M., Lemke, H., Gonzalez, A., Sauter, N. K., Cohen, A. E., van den Bedem, H., Thorne, R. E. & Fraser, J. S. (2015). Mapping the conformational landscape of a dynamic enzyme by multitemperature and XFEL crystallography. Elife 4, e07574.
71. Bonomi, M., Pellarin, R. & Vendruscolo, M. (2018). Simultaneous determination of protein structure and dynamics using cryo-electron microscopy. Biophys. J. 114, 1604-1613.
72. Kumar, G. S., Clarkson, M. W., Kunze, M. B. A., Granata, D., Wand, A. J., Lindorff-Larsen, K., Page, R. & Peti, W. (2018). Dynamic activation and regulation of the mitogen-activated protein kinase p38. Proc. Natl .Acad. Sci. USA 115, 4655-4660.
73. Bonomi, M., Camilloni, C. & Vendruscolo, M. (2016). Metadynamic metainference: Enhanced sampling of the metainference ensemble using metadynamics. Sci. Rep. 6, 31232.
74. Lohr, T., Jussupow, A. & Camilloni, C. (2017). Metadynamic metainference: Convergence towards force field independent structural ensembles of a disordered peptide. J. Chem. Phys. 146, 165102.
75. Zhou, T., Sun, L., Humphreys, J. & Goldsmith, E. J. (2006). Docking interactions induce exposure of activation loop in the MAP kinase ERK2. Structure 14, 1011-1019.
76. Selvaratnam, R., Chowdhury, S., VanSchouwen, B. & Melacini, G. (2011). Mapping allostery through the covariance analysis of NMR chemical shifts. Proc. Natl. Acad. Sci. USA 108, 6133-6138.
77. Berlin, K., Castañeda, C. A., Schneidman-Duhovny, D., Sali, A., Nava-Tudela, A. & Fushman, D. (2013). Recovering a representative conformational ensemble from underdetermined macromolecular structural data. J. Am. Chem. Soc. 135, 16595-16609.
78. Leung, H. T., Bignucolo, O., Aregger, R., Dames, S. A., Mazur, A., Berneche, S. & Grzesiek, S. (2016). A rigorous and efficient method to reweight very large conformational ensembles using average experimental data and to determine their relative information content. J. Chem. Theory Comput. 12, 383-94.

79. Seidel, J. J. & Graves, B. J. (2002). An ERK2 docking site in the Pointed domain distinguishes a subset of ETS transcription factors. Genes Dev. 16, 127-137.

Figure Legends

Figure 1. (A) Structure of the active, dual-phosphorylated form of the MAPK ERK2 shown in ribbon representation. The N- and C-lobes are shown in light and dark blue, respectively; the MAP kinase insert (including helices α1L14 and α2L14) and the C-terminal extension (including helix αL16) are shown in purple and magenta, respectively. The phosphorylated T-X-Y moiety (T185 and Y187) on the activation loop are shown in stick representation. (B) The two docking regions (recruitment sites) on ERK2 are shown. The acidic (ϕ_{chg}) and hydrophobic (ϕ_{hyd}) parts of the D-recruitment site (DRS) are shown in green and pink, respectively; a canonical D-site sequence is shown with the basic and two hydrophobic residues ($\Phi_{A/B}$) colored green and pink, respectively. Also shown in brown is the F-recruitment site (FRS) and a corresponding canonical F-site sequence. Human numbering used in all cases.

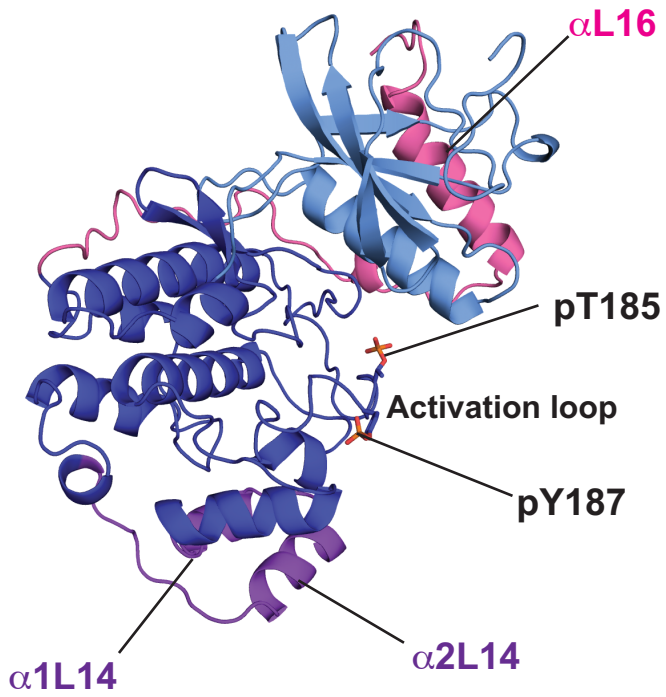
Figure 2. Schematic representation of the domain structure of the transcription factor Elk-1 showing the ETS domain, the transactivation domain (TAD), the nuclear localization signal (NLS) and the nuclear export signal (NES). The region relevant for F-site directed ERK1/2-mediated phosphorylation, including the phospho-acceptor Ser383 and Ser389 sites, is indicated. The sequence of the $Elk_{387-399}$ peptide discussed in the text is indicated by the overbar. (B) A toy model for F-site-directed phosphorylation by ERK2 is shown. The fully-extended length of the linker connecting the phospho-acceptor (in blue) with orientation $\Omega(r,\theta,\phi)$ to the F-site sequence (FQFP) is L and the distance between the FRS and the ERK2 catalytic site is D. Both sites are considered to be dimensionless point objects. (C) Variation in effective concentration (C_{eff}) with L calculated using Equation 1 for D = 11.5 Å, the approximate distance between the catalytic site and the FRS of ERK2, and two values of the persistence length (ξ). The dashed lines represent the critical values ($L_C=D^2/2\xi$) of L for each curve. The arrows indicate the values of L used in calculating the curves depicted in (D). (D) Variation of C_{eff} with D plotted for two sets of values each of L and ξ. The arrow represents D = 11.5 Å.

Figure 3. The geometry of the PCC that enables optimal phosphorylation may be represented by a set of distances: (1) between the O of the phospho-acceptor (a Ser in the case illustrated) –OH moiety and the γP of ATP (3.9 Å), (2) the Cγ of ERK2 D149 (3.9 Å), or (3) the Nζ of K151 (2.8 Å) or (4, 5) the two Mg^{2+} ions (5.7 Å and 6.1 Å , left to right); between the N atom of the Pro residue that is part of the Ser-Pro moiety and (6) the Cγ of L170 (6.8 Å) or (7) pY187 (7.0 Å) of ppERK2. From the QM/MM calculations of Turjanski et. al.[41]

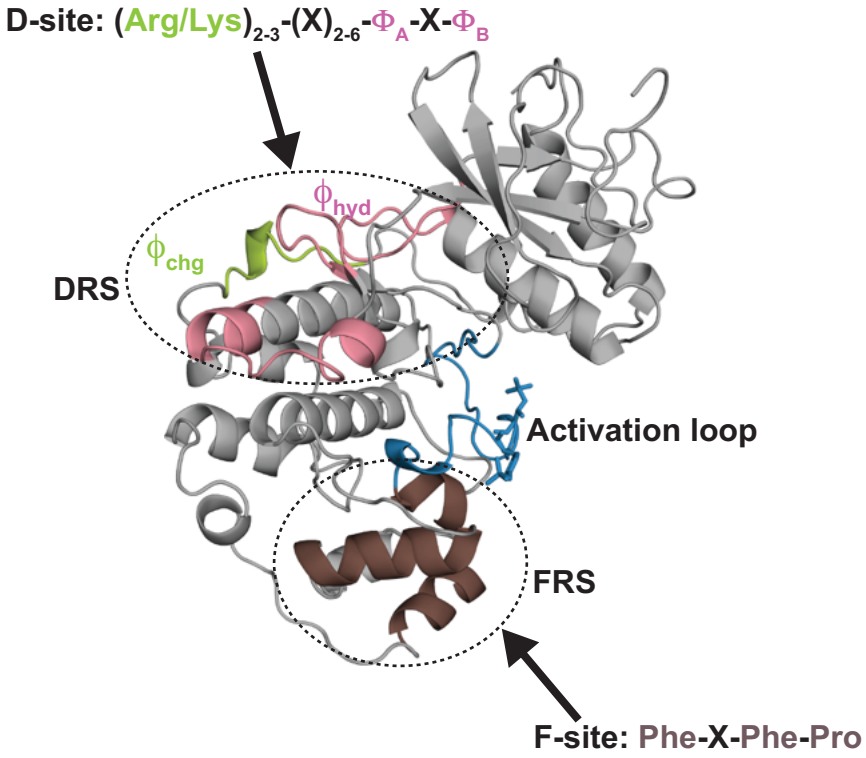
Figure 4. Schematic representation of the MAPK pre-chemistry ensemble (PCE). The substrate binds ATP-bound (the γ-phosphate is represented by the orange sphere) ERK2 using its canonical (or non-canonical) docking site to target the corresponding recruitment site on the kinase. The still dynamic phospho-acceptor localizes near the ERK2 active site generating an ensemble of structures representing the PCE. This PCE comprises of microstates in which the substrate phospho-acceptor and surrounding regions (including the Pro in the P+1 position) are in appropriate conformation along with the relevant catalytic elements of ERK2 to enable efficient chemistry (high-activity state), and those in

which they are not (low-activity state). Here the PCE is represented by only two microstates for simplicity. The PCE in reality could comprise of a discrete distribution or a continuum or multiple banded continua of microstates representing species that interconvert with rate-constants (k_Q and $k_{Q''}$) that potentially vary by several orders of magnitude (in general, some much faster and some much slower than chemistry). These states possess different inherent abilities to proceed through the chemical step (e.g. k_3 in a high-activity state is significantly higher than k_3' in a low-activity state) and yield phosphorylated product. The specific details depend on the magnitude of the exchange rates relative to the rate of chemistry.

Figure 5. The transcription factor Ets-1 (440 residues) consists of a PNT domain, a TAD and an ETS domain. The N-terminal 138 residues of Ets-1 including the PNT domain are sufficient to be recognized and phosphorylated by ERK2[79]. The binding mode of Ets-1 on ERK2 is shown schematically on the top panel. An N-terminal sequence (purple) on Ets-1 partially mimics the hydrophobic segment of a D-site sequence and docks onto the ϕ_{hyd} portion of the ERK2 DRS. The Ets-1 PNT domain (dark green) partially engages the ERK2 FRS. The phospho-acceptor sequence (Thr38-Pro39) is connected to the PNT domain by a partially disordered helix (α0, light green). The location of the positional mutants (generated using Ala spacers) of the phospho-acceptor Thr38 is shown. The labels on the left signify the location of the phospho-acceptor Thr with respect to the wild-type position – 0: wild-type; -1, -2: one, two residues away from the PNT domain (green) towards the N-terminal sequence (purple); +1, +2, +3: one, two or three residues towards the PNT domain. The corresponding steady-state catalytic parameters are shown on the right. Also shown in pink font are the catalytic parameters for a Pro39Ala mutant. The errors reported are those obtained from non-linear least squares fits for representative kinetic datasets[27]. k_{cat}/K_M values are in $\mu M^{-1}s^{-1}$.


αL16
pT185
Activation loop
pY187
α1L14
α2L14

(A)


D-site: (Arg/Lys)2-3-(X)2-6-ΦA-X-ΦB
ϕhyd
ϕchg
DRS
Activation loop
FRS
F-site: Phe-X-Phe-Pro

(B)

Figure 1

Figure 2

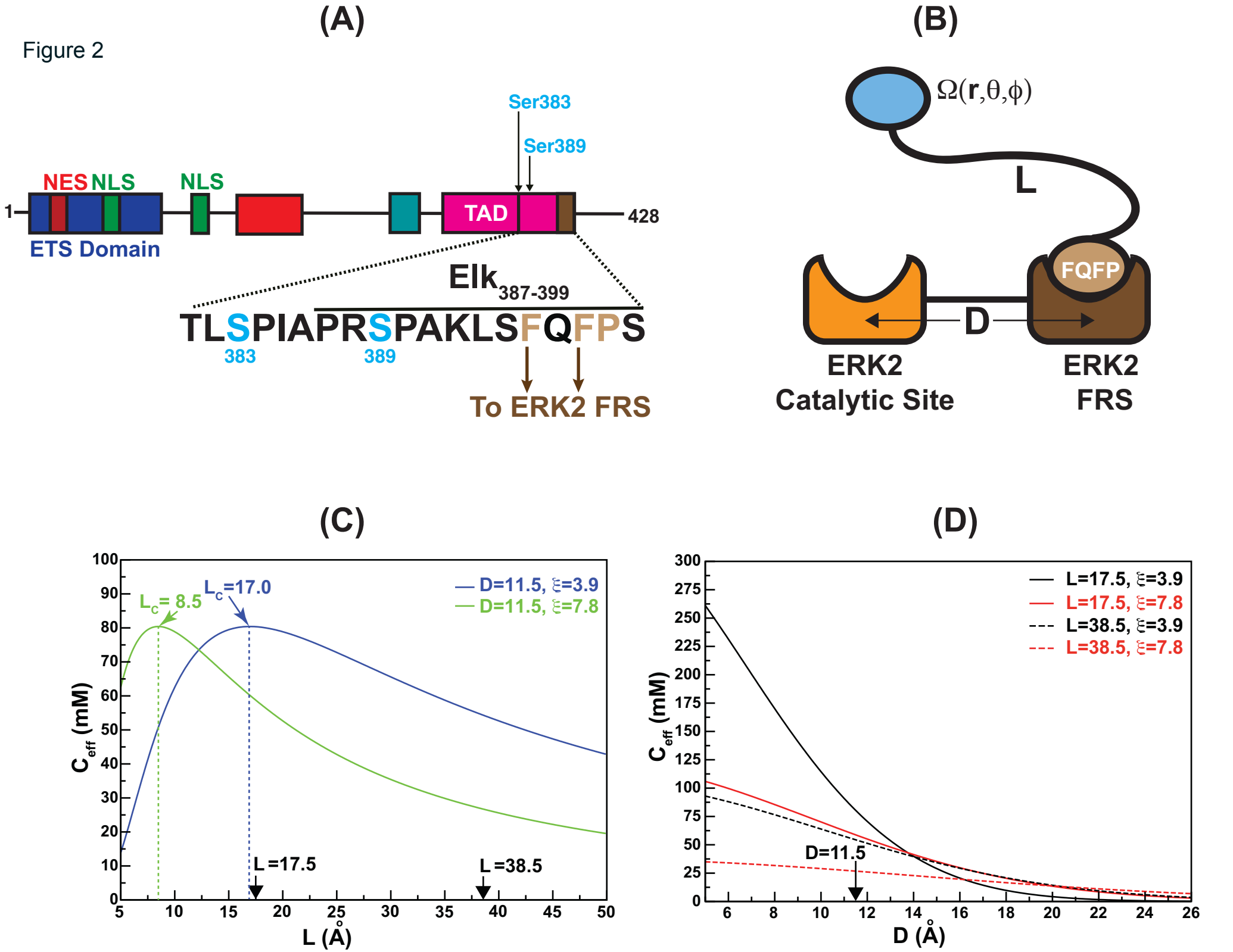

(A)
Ser383
Ser389
NES NLS
NLS
1
TAD
428
ETS Domain
Elk387-399
TLSPIAPRSPAKLSFQFPS
383
389
To ERK2 FRS
(B)
Ω(r,θ,ϕ)
L
FQFP
D
ERK2
Catalytic Site
ERK2
FRS
(C)
Lc= 8.5
Lc=17.0
D=11.5, ξ=3.9
D=11.5, ξ=7.8
Ceff (mM)
L=17.5
L=38.5
L (Å)
(D)
L=17.5, ξ=3.9
L=17.5, ξ=7.8
L=38.5, ξ=3.9
L=38.5, ξ=7.8
Ceff (mM)
D=11.5
D (Å)

Figure 3

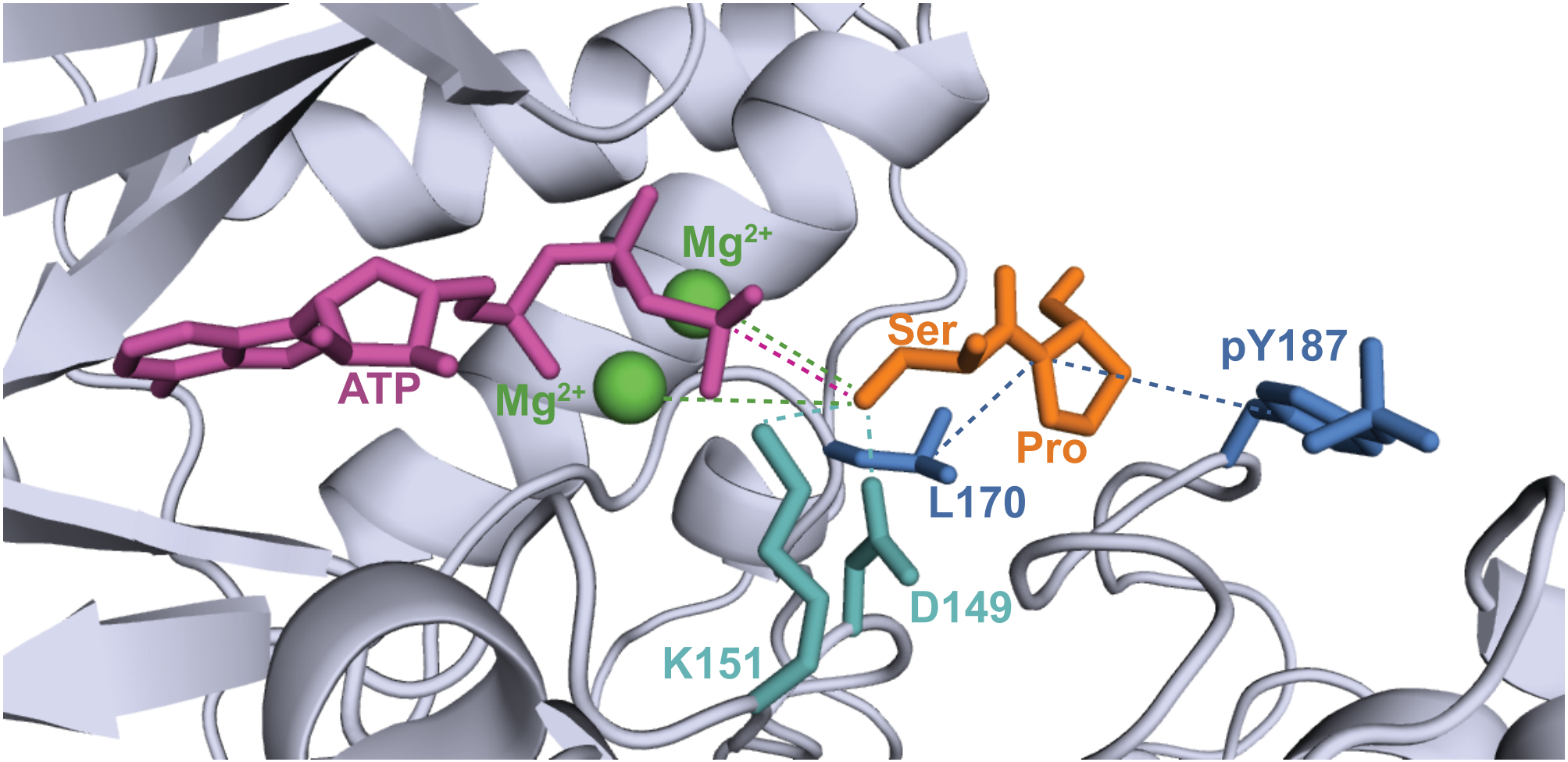


Figure 4

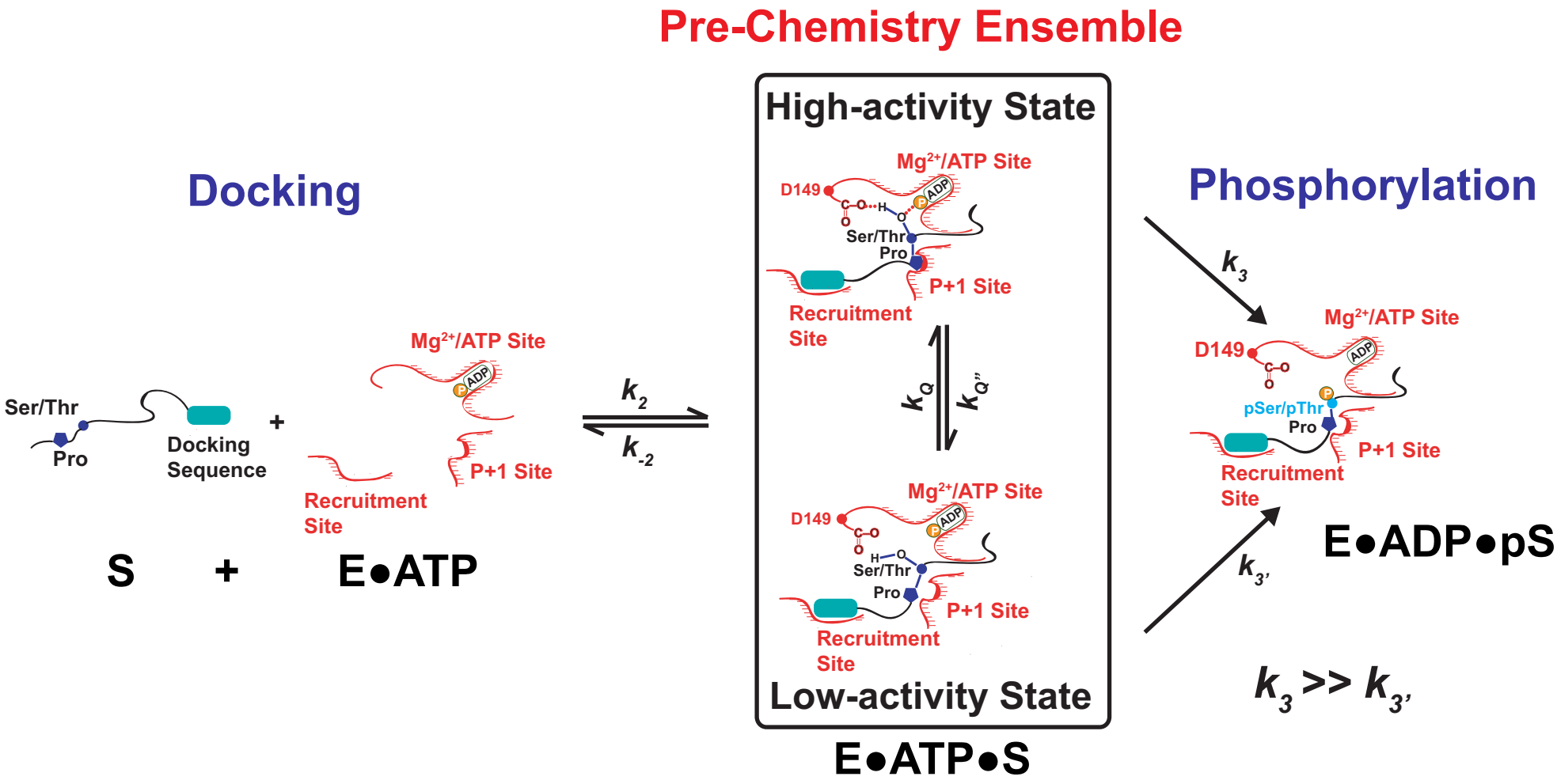
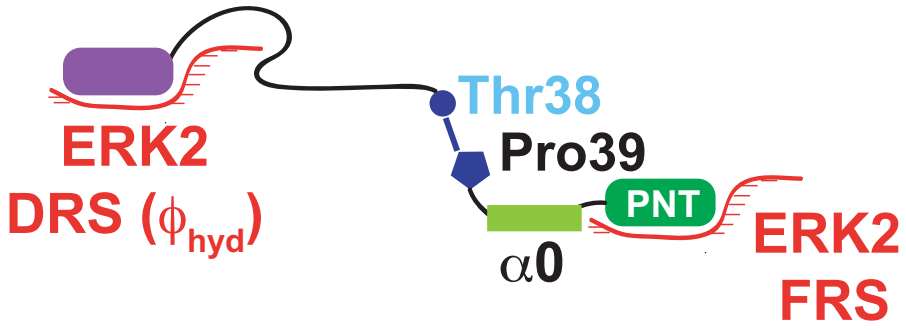
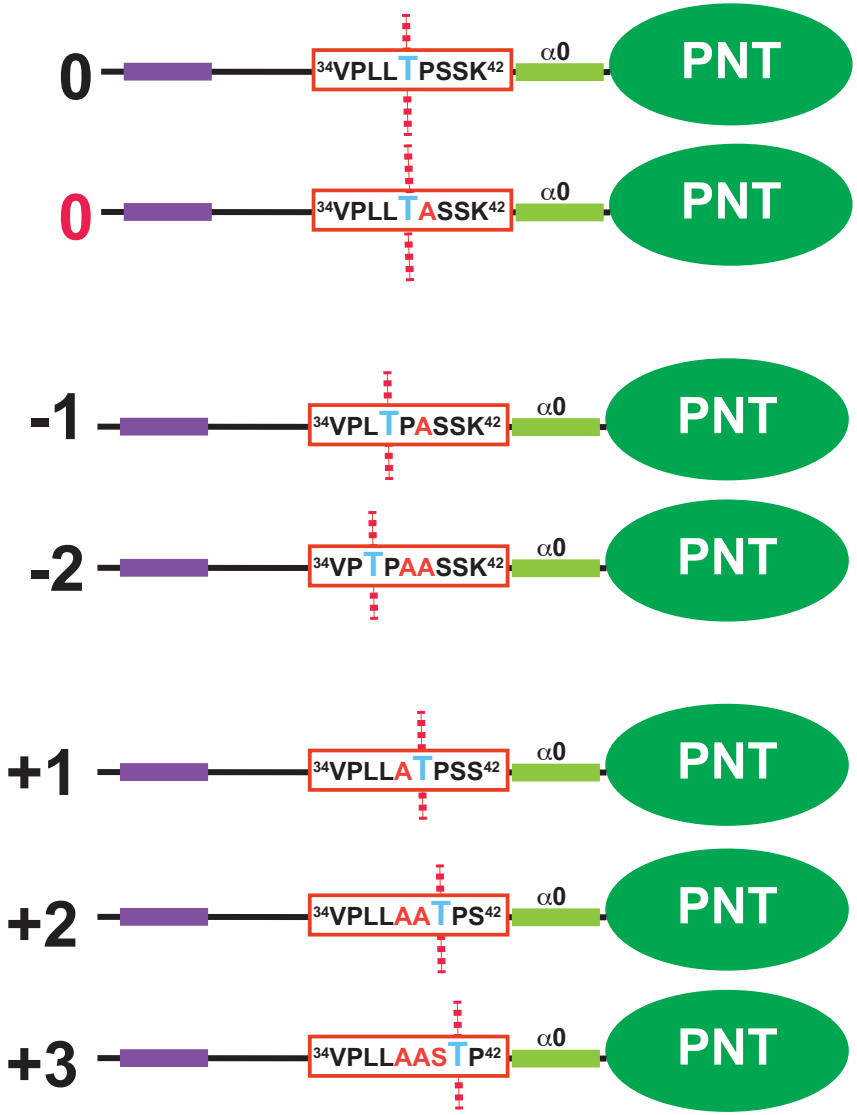

Pre-Chemistry Ensemble
Docking
Ser/Thr
Pro
Docking Sequence
+
Mg2+/ATP Site
ADP
P+1 Site
Recruitment Site
S + E•ATP
k2
k-2
High-activity State
D149
Ser/Thr
Pro
kQ
kQ''
Low-activity State
E•ATP•S
Phosphorylation
k3
pSer/pThr
k3'
E•ADP•pS
k3 >> k3'

Figure 5





	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M
0	14.0 ± 1.6	23 ± 5	0.6
0	3.5 ± 0.2	63 ± 6	0.06
-1	8.0 ± 0.1	7 ± 0.4	1.14
-2	8.4 ± 0.3	21 ± 3	0.4
+1	4.0 ± 0.3	34 ± 7	0.12
+2	5.5 ± 0.3	29 ± 3	0.19
+3	3.0 ± 0.2	37 ± 8	0.08

Scheme I