Conformational Changes in Protein Loops and Helices Induced by Post-Translational Phosphorylation

Eli S. Groban1,2, Arjun Narayanan1,2, Matthew P. Jacobson1,2*

1 Department of Pharmaceutical Chemistry, University of California San Francisco, San Francisco, California, United States of America, 2 Graduate Group in Biophysics, University of California San Francisco, San Francisco, California, United States of America

Post-translational phosphorylation is a ubiquitous mechanism for modulating protein activity and protein-protein interactions. In this work, we examine how phosphorylation can modulate the conformation of a protein by changing the energy landscape. We present a molecular mechanics method in which we phosphorylate proteins in silico and then predict how the conformation of the protein will change in response to phosphorylation. We apply this method to a test set comprised of proteins with both phosphorylated and non-phosphorylated crystal structures, and demonstrate that it is possible to predict localized phosphorylation-induced conformational changes, or the absence of conformational changes, with near-atomic accuracy in most cases. Examples of proteins used for testing our methods include kinases and prokaryotic response regulators. Through a detailed case study of cyclin-dependent kinase 2, we also illustrate how the computational methods can be used to provide new understanding of how phosphorylation drives conformational change, why substituting Glu or Asp for a phosphorylated amino acid does not always mimic the effects of phosphorylation, and how a phosphatase can “capture” a phosphorylated amino acid. This work illustrates how computational methods can be used to elucidate mechanisms of post-translational phosphorylation, which can ultimately help to bridge the gap between the number of known sites of phosphorylation and the number of structures of phosphorylated proteins.

Citation: Groban ES, Narayanan A, Jacobson MP (2006) Conformational changes in protein loops and helices induced by post-translational phosphorylation. PLoS Comput Biol 2(4): e32. DOI: 10.1371/journal.pcbi.0020032

Introduction

Post-translational phosphorylation is a ubiquitous mechanism for cellular regulation, playing a role in such diverse processes as signal transduction, transport, cytoskeletal regulation, and metabolism. A variety of amino acids can be phosphorylated, but serine, threonine, and tyrosine are the most important sites of phosphorylation in eukaryotes, whereas histidine and aspartate play the central role in the “two-component” signaling pathways of prokaryotes. Several thousand sites of post-translational phosphorylation are now known [1], and this number will continue to grow quickly. Estimates of the fraction of proteins that are phosphorylated in vivo range as high as 30% [1]; higher values are associated with particular stages of the cell cycle or responses to external stimuli. Protein kinases catalyze post-translational phosphorylation, and many kinases are themselves regulated by phosphorylation, leading to complex signaling and regulatory networks. Kinases are targets of aggressive drug development efforts [2] aimed at treating cancer and other diseases such as diabetes.

Despite the huge amount of research related to post-translational phosphorylation, the detailed role that specific sites of post-translational phosphorylation play in the function of individual proteins remains poorly understood in most cases. Structural information is particularly limited, due in part to the difficulty of obtaining sufficient purified protein in a specific modification state. X-ray crystallography has determined atomic-resolution structures for a few tens of phosphorylated proteins [3], whereas nuclear magnetic resonance (NMR) experiments have elucidated structure and dynamics of a similar number of phosphorylated peptides and small proteins in solution [4–10]. Electron paramagnetic resonance (EPR) and circular dichroism experiments can also provide information on conformational change due to phosphorylation [11], but do not provide atomic detail. The number of known phosphorylation sites will almost certainly continue to grow much more quickly than the number of structurally well-characterized phospho-proteins.

We believe that computational studies can play a central role in elucidating principles and mechanisms of post-translational modification, and ultimately help to bridge the gap between the number of known sites of phosphorylation and the number of structures of phosphorylated proteins. Only a small number of computational modeling studies have...
Many proteins are chemically modified after they are synthesized in the cell. These post-translational modifications can modulate the ability of a protein to perform chemical reactions and to interact with other proteins. At the cellular level, for example, these chemical modifications are critical for allowing the cell to respond to its environment and control its division. One of the most common mechanisms by which proteins can be modified is by phosphorylation—the addition of a phosphate group to an amino acid side chain of the protein. Thousands of proteins are known to be modified by phosphorylation, but only for a small minority of these do we have any detailed understanding of how the chemical modification regulates the function of the protein. The authors describe a computational method that can make testable predictions about the structural changes that occur in a protein induced by post-translational phosphorylation. Their results show that the method can produce structural models of the phosphorylated proteins with near-atomic accuracy, and provide insight into the energetics of conformational switches driven by phosphorylation. As such, the computational method complements experiments aimed at understanding the mechanisms of protein regulation by phosphorylation.

Figure 1. Phosphorylation Can Perturb the Energy Landscape of a Protein to Cause Changes in Conformation and Dynamics

This figure is a visual representation of such changes. In this work, we predict the structural change of proteins due to phosphorylation by locating the new global energy minimum of the energy surface. DOI: 10.1371/journal.pcbi.0020032.g001

phosphorylated protein? The critical tools we need to understand and predict the structural effects of post-translational phosphorylation are as follows:

One tool that is needed is an energy function that captures the essential physics, especially for the modified residues. We use a molecular mechanics energy function consisting of the OPLS-AA force field and a Generalized Born implicit solvent model. The use of a physics-based energy function is important for two reasons. First, there are too few structures of phosphorylated proteins to make knowledge-based energy functions feasible. Second, conformational changes induced by phosphorylation are largely driven by the electrostatic perturbation induced by the phosphate group. We use the energy function not only to predict conformational changes induced phosphorylation, but also to better understand the key physical effects underlying these conformational changes, especially in our case study of cyclin dependent kinase 2 (CDK2).

Another necessary tool is sampling algorithms capable of exploring critical degrees of freedom. After phosphorylating a protein in silico, we need to explore the new energy landscape and identify the new global energy minimum, if in fact it is significantly different than that of the unphosphorylated protein. In principle, molecular dynamics could be used for this purpose, and in some cases, this type of strategy has been successful, in studies by ourselves and others [18–30]. However, the timescales for converting from the non-phospho to the phosphorylated form are unknown, and in fact are not physically meaningful because the in silico phosphorylation is alchemical, i.e., we do not attempt to mimic a kinase actually performing the phosphorylation. For relatively large conformational changes, or those involving high-energy barriers, e.g., Pro peptide bond isomerization, the relevant timescale could be quite long—microseconds or more. Instead, we use a strategy in which we have adapted algorithms that we previously developed for homology modeling, i.e., sampling methods for side chains [37,38], loops [39–41], and helices [42]. The essence of the approach is to combine dihedral angle sampling methods, which enable large energy barriers to be surmounted, with direct minimization to enumerate many local minima. We discuss our strategy in detail in Materials and Methods.

The key conclusion of this study is that our molecular mechanics-based methods appear to be capable of reproducing conformational changes induced by post-translational phosphorylation, with near-atomic resolution in most cases considered here, which are limited to relatively modest conformational changes and not, e.g., more drastic order-disorder transitions. This work thus represents a significant step toward a broadly applicable method for predicting
structural effects of phosphorylation. Through a detailed case study of CDK2, we also illustrate how the computational methods can be used to provide new understanding of how phosphorylation drives conformational change, why substituting Glu or Asp for a phosphorylated amino acid does not always mimic the effects of phosphorylation, and how a phosphatase can “capture” a phosphorylated amino acid.

Results/Discussion

We first present an in-depth study of activation loop phosphorylation in CDK2, both to introduce the computational methods and to highlight the ability of these methods to gain new insights into how phosphorylation drives conformational changes. We then present a broader survey of our ability to predict phosphorylation-driven conformational changes.

Case Study: CDK2

CDK2 is a member of the cyclin-dependent kinase protein family [43], whose members play a central role in cell cycle regulation. CDK2 is activated through binding of cyclin A and post-translational phosphorylation of a threonine residue on the activation loop, which lies close to the catalytic site. Compared to other phosphorylated proteins, there is a wealth of structural information for CDK2 in both its phosphorylated and unphosphorylated forms [3]. In this case study, we examine the ability of our method to: (1) reconstruct and predict the active conformation of the activation loop in CDK2 bound to cyclin A [44]; (2) reconstruct the cyclin A dependence of CDK2 activation [44,45]; (3) discriminate between two potential phosphate localization sites in the CDK2/kinase-associated phosphatase (KAP) complex [46]; and (4) examine the effects of substituting Thr160 with a Glu, which results in only partial activation relative to phosphorylation of Thr160 [47].

A structural alignment and superposition of the active and inactive [48] forms of CDK2 bound to cyclin A reveals that the global Cα root mean square deviation (RMSD) between the two structures is only 0.5 Å if residues 152–163 are removed from the calculation. The phosphorylatable threonine residue itself moves approximately 10 Å upon phosphorylation, as measured at the ω oxygen. In the unphosphorylated structure, Thr160 is well solvated and somewhat disordered (B-factors > 50). In the phosphorylated structure, pThr160 localizes to interact with a cluster of Arg residues (50, 126, and 150). We will refer to this and similar conformations as “active” conformations of the loop.

As described in Materials and Methods, we use a hierarchical loop prediction algorithm [40] based upon dihedral angle backbone sampling, rotamer-based side chain optimization, an all-atom force field and a Generalized Born solvation model [49,50] to predict the structural consequences of phosphorylation on loops and their surroundings. The hierarchical prediction algorithm allows us to quickly prune the conformational space and focus sampling on energetically favorable regions of conformation space (Figure 2). Unlike Monte Carlo and molecular dynamics sampling schemes, the algorithm itself has no knowledge of the starting conformation of the loop. The entire conformational space of the loop is sampled with varying sampling resolution in a hierarchical manner. This method has proven successful in recreating crystallographic conformations of unphosphorylated loops to loop lengths of 12 residues. In this paper, we expand on this loop prediction method to predict the structures of phosphorylated loops.

Before attempting to predict the phosphorylated structure from the unphosphorylated structure, we first ensure that we
can predict the conformation of the phosphorylated loop with high accuracy when all other portions of the phosphorylated structure are taken from the crystal structure. We refer to this test as “loop reconstruction.” This tests the suitability of the energy function for identifying the correct conformation of the phosphorylated loop and the suitability of the sampling function for generating native-like structures. Much of our analysis will focus on the accuracy with which the phosphate group is positioned, which is measured by the distance of the phosphorus atom to its crystallographic position. We also provide other measures of accuracy for the overall loop, including a backbone RMSD measure (calculated based on the N, Cα, and C atoms), and the RMSD calculated over all heavy atoms, including side chains. For all of these measures of accuracy, the predicted structure is first aligned to the reference crystal structure by a least-squares superposition of all Cα atoms, and the RMSD calculated over all heavy atoms, including side chains. For all of these measures of accuracy, the predicted structure is first aligned to the reference crystal structure by a least-squares superposition of all Cα atoms, excluding the loop being simulated. Thus, structural differences outside of the loop region being predicted can affect the reported measures of accuracy.

For the complex of phosphorylated CDK2 with cyclin A, we are able to reproduce the conformation of the activation loop with less than 2 Å RMSD overall. The phosphate group is placed with particularly high accuracy, less than 1 Å error as measured by the position of the phosphorous atom (Figures 2 and 3, and Table 1). We also wish to highlight the differences between Figure 2B and 2C, which plot the energy of different local minima identified during the four-stage prediction algorithm versus the backbone RMSD (Figure 2B) and the error in the phosphate position (Figure 2C). In the energy versus backbone RMSD plot, an energy “funnel” is clearly evident. That is, the lowest energy identified decreases as the loop, as a whole, approaches native-like conformations. On the other hand, the phosphate group itself appears to respond to a narrow, deep energy well. That is, the phosphate locates the Arg cluster very early in the simulation (during the first of the four stages of the loop prediction algorithm), and remains there while the rest of the loop adjusts to the new position of the phosphate group.

Next, we examine our ability to predict the phosphorylated loop conformation starting from the unphosphorylated structure. Naively attempting to predict the conformation of the phosphorylated activation loop from the unphosphorylated structure, holding the nearby side chains fixed, results in a poor prediction, with an error in the position of the phosphorus atom of approximately 9 Å and a backbone RMSD of approximately 7 Å (Table 1). The reason for this failure is clear. A comparison of the phosphorylated and unphosphorylated structures reveals that the backbone of the phosphorylated conformation of the activation loop passes directly through the side chain of Tyr179 in the unphosphorylated form (Figure 4). This single misplaced side chain can prevent the loop from assuming the correct activated conformation. Thus, in this case, introduction of the phosphate group not only changes the activation loop conformation, but also rearranges the side chain hydrogen bonding network well beyond the loop itself.

Table 1. Results from the CDK2 Case Study

| Case        | Deviations from Phosphorylated Crystal Structure (Å) |
|-------------|-----------------------------------------------------|
|             | Loop Only | Loop + Surroundings |
|             | P | BB | P | BB |
| Reconstruction | 0.6 | 1.7 | 1.7 | 1.8 |
| Prediction   | 8.8 | 6.7 | 0.8 | 2.9 |
| CDK2/KAP     | 1.1 | 2.3 | 0.5 | 1.3 |

The case study is summarized using two measures: the deviation of the phosphate group from its position in the crystal structure of phosphorylated CDK2 (“P”), and the overall backbone RMSD of the predicted loop compared to the same structure (“BB”). For the CDK2/KAP case, the reference structure is that of phosphorylated CDK2 in complex with KAP (1fq1), whereas the other cases are compared to the crystal structure of phosphorylated CDK2 bound to cyclin A (1JST). “Loop Only” refers to predicting only the loop in question; these results are provided only for comparison. “Loop + Surroundings” refers to the protocol in which we predict the loop residues as well as side chains of residues within 4.5 Å of any atom in the loop; these are the results that should be used to evaluate the success of the method. The different prediction cases are as follows. “Reconstruction” refers to predicting residues 152–163 in the crystal structure of phosphorylated CDK2 in complex with cyclin A. “Prediction” refers to the prediction of residues 152–163 after in silico phosphorylation of unphosphorylated CDK2/cyclin A (1fin). Finally, “CDK2/KAP” refers to predicting residues 155–165 in the structure of the phosphorylated activation loop of CDK2 when in complex with its phosphatase, KAP. It is important to note that the prediction methodology being used has no knowledge of the starting structure of the loop.

DOI: 10.1371/journal.pcbi.0020032.t001
Thus, sampling of loop conformations must be coupled with, at least, the sampling of side chain conformations in the vicinity of the loop to permit accurate prediction. As described in Materials and Methods, we have adopted a strategy in which side chains near the loop are removed during the loop build-up, and side chains both on the loop and its vicinity are optimized and energy minimized simultaneously for the representative loop from each cluster. Applying this method to the loop reconstruction test increases the error by a small amount. Overall, the error in the phosphorus position increased by about 1 Å, and the backbone RMSD for the activation loop increased an insignificant amount (Figure 5 and Table 1). We note that Figure 5A clearly shows two major ‘‘basins’’ of attraction for the phosphorylated amino acid. The one with the lower energy places the phosphate in contact with the Arg cluster; the higher energy state places the phosphate in solution. These two conformations, which we will call “active” and “inactive,” respectively, will be examined further below.

When predicting the structural change upon phosphorylation of CDK2, this strategy permits Tyr179, as well as the three arginine residues that form salt bridges with pThr160, to re-optimize around the loop. The prediction with the simultaneous optimization of the loop and surrounding side chains results in a phosphorous atom error of 0.8 Å and a backbone RMSD of 2.9 Å (Figures 3 and 3B, Table 1). The phosphorus atom position is quite accurate, whereas the remainder of the loop may require the optimization of other structural elements to assume the active conformation. The prediction clearly captures the qualitative change in conformation upon phosphorylation, however. Additionally, we performed the above calculations again, except this time substituting the three Arg residues of the arginine cluster with Lys residues (unpublished results). In this case, the phosphate does not localize to the new “Lys cluster,” and we predict that the kinase would not be active. The preference of phosphorylated side chains for interacting with Arg, relative to Lys, has been noted previously, and we will explore this issue further in a separate publication.

CDK2 Phosphorylation in the Absence of Cyclin Binding

Cyclin A binding is required for the full activation of CDK2 because phosphorylated CDK2 without cyclin A exhibits only 0.3% of the activity of the fully active, phosphorylated, cyclin A–bound structure [3]. The crystal structure of phosphorylated CDK2 in the absence of cyclin A shows that the activation loop is disordered [45]. The loop thus likely adopts several dynamically interconverting conformations with the pThr fully solvated, instead of bound by the arginine cluster. The activation loop in unphosphorylated CDK2 in the absence of cyclin A is ordered, albeit with somewhat elevated B-factors (roughly 60).

The results of predicting the loop in the phosphorylated, cyclin A–unbound structure [45], showed that the phosphate did not localize to the Arg cluster, which is qualitatively consistent with phosphorylation not contributing to activation (results not shown). However, the low-energy predicted loop conformations showed the phosphate localized at a different position, and the results did not provide any evidence that the loop should in fact be disordered. This is
due to a deliberate bias in the loop prediction algorithm, which was originally designed to predict well-structured loops, not ensembles of structures characterizing a disordered loop. A full solution will require an algorithm that obeys detailed balance and thus correctly treats the loop conformational entropy; we have recently developed a Monte Carlo version of the loop sampling algorithm that accomplishes this goal, and we will report results in due course. As a preliminary step, we have performed the calculations with one key modification to the existing loop prediction algorithm. Specifically, the algorithm constrains the loop conformations to be relatively close to the body of the protein using a simple distance cutoff; more than 99% of loops in the Protein Data Bank (PDB) satisfy this criterion [40]. However, this restriction prevents disordered loops from exploring conformations in which the loop is well solvated and makes few contacts to the remainder of the protein. By simply turning off this screening, the loop prediction for phosphorylated CDK2 in the absence of cyclin A does show evidence of diverse low-energy conformations (Figure 6B), in the sense that the 20 lowest-energy conformations are highly diverse; the phosphorylated Thr160 is well solvated in most of these. Turning off the screening does not significantly affect the results for the in silico phosphorylation of CDK2 with cyclin A bound: pThr160 still localizes strongly to the Arg cluster (Figure 6A) in all 20 of the lowest-energy conformations. Thus, in this case, the computer simulations can provide at least a qualitative prediction that phosphorylation of CDK2 in the absence of cyclin binding leads to disorder. More work is clearly required to test this capability in other cases of phosphorylation-induced disorder; the remainder of our results focuses exclusively on cases in which the phosphorylated loops are well ordered.

Substitution of Glu for the Phosphorylated Thr in the Activation Loop of CDK2

Among proteins that require phosphorylation for activation, there are many examples of a constitutively active mutant in which the phosphorylatable residue is substituted with either aspartate or glutamate [51–55]. Connell-Crowley et al., however, show that in the CDK2/cyclin A complex, a T160E mutation confers no additional activity to the complex [47]. To examine why, we use the active form of CDK2/cyclin A to create a CDK2/cyclin A T160E mutant in silico, and then use our prediction protocol to predict the conformation of the activation loop, residues 152–163 in the mutant protein. Using the protocol in which side chains surrounding the loop are optimized concurrently with the loop itself, the lowest-energy structure places the Cβ atom of Glu160 9 Å away from its position in the wild-type active loop (Figure 7). The side
Table 2. Energy Differences (kcal/mol) between “Active” and “Inactive” Conformations of the Activation Loop

| Energy Component | pThr160 | T160E | CDK2/KAP |
|------------------|---------|-------|----------|
| Covalent         | 15.3    | 16.4  | 10.8     |
| Lennard-Jones    | 23.8    | 2.4   | 30.1     |
| Solvation        | 240.6   | 161.2 | 94.6     |
| Total            | −18.5   | 2.9   | 24.6     |

The energy differences are broken down by different components of the molecular mechanics energy: covalent (bonds, bond-angles, and torsions), the Coulombic electrostatic term, Lennard-Jones, and the solvation free energy. The active form is defined by pThr160 being localized near the Arg cluster. In the pThr160 case, this is the lowest-energy-predicted structure, i.e., as represented in Table 2. In the Thr160Glu (T160E) substitution, the lowest-energy structure has the Glu side chain pointed out into solution; we refer to this as the “inactive” conformation. The second highest-energy structure has the Glu side chain in a position analogous to the pThr side chain, and we refer to this as the “active” structure for T160E. For CDK2/KAP, active/inactive take on different meanings. The active conformation is analogous to the other active structures, i.e., the phosphate group is localized at the Arg cluster. The inactive conformation corresponds to lowest-energy conformation, in which the phosphate is correctly localized near the N-terminus of a helix in KAP.

DOI: 10.1371/journal.pcbi.0020032.t002

Phosphorylated CDK2 in Complex with Its Phosphatase

Finally, we examine conformational changes in the activation loop necessary for dephosphorylation to occur. KAP is responsible for the dephosphorylation, and subsequent inactivation, of CDK2 [3,46]. A crystal structure of the CDK2/KAP complex shows that pThr160 localizes to the N-terminus of a helix in the KAP protein, and interacts in that position with a single arginine side chain. The phosphate group must be stable enough in this position to allow the phosphatase to “pull” pThr160 from the Arg cluster on CDK2. Given the highly favorable electrostatic interactions between pThr160 and the three Arg side chains in the cluster, this did not seem intuitively obvious.

Nonetheless, the loop predictions in the CDK2/KAP complex did in fact place the phosphorus atom within 0.5 Å of its position in the crystal structure (Table 1); the lowest energy structure with pThr160 placed in contact with the Arg cluster is approximately 25 kcal/mol higher in energy. Table 2 helps to explain why. The phosphate placed close to the Arg cluster does in fact have much more favorable electrostatic interactions, as measured by the Coulombic term in the molecular mechanics energy function (111 kcal/mol difference). However, the solvation term strongly favors the pThr160 in its correct position in contact with KAP, due largely to the improved solvent accessibility of the Arg cluster after the pThr160 is removed from contact. In addition, the pThr160 is not as buried in the KAP complex. Finally, the covalent terms again indicate significantly higher internal strain in the activation loop when the pThr160 is in contact with the Arg cluster, and much of this strain is relieved when the loop adopts the phosphatase-bound conformation (this might be referred to as a spring-loaded mechanism).

Predicting Conformational Changes of Loops in Response to Phosphorylation

Encouraged by our results on CDK2, we have performed similar calculations on a larger, more diverse test set (Table 3). Our results show that in all cases of loop reconstruction, we reproduce the structure of the phosphorylated loop with high accuracy (Table 4). We reconstruct seven of the loops to within 1 Å backbone RMSD with respect to the original crystal structure and predict the phosphorus atom to within 0.5 Å of its crystallographic position. Although our predictions are not error free, the relatively small magnitude of the error gives us some confidence that we can predict the correct conformation of a phosphorylated loop using our hierarchical sampling methodology, all-atom force field, and Generalized Born solvation model.

In cases in which suitable unphosphorylated and phosphorylated structures of the same protein exist, we sought to extend this methodology to the more realistic situation in...
Table 3. Proteins Used in Our Test Set

| Protein | Phosphorylated Structures (PDB ID) | Phosphorylated Residue(s) | Residues Predicted | Unphosphorylated Structures (PDB ID) |
|---------|-----------------------------------|--------------------------|-------------------|-------------------------------------|
| Hpr     | 1FU0 [73]                         | Ser46                    | 42–48             | 1PTF [74]                           |
| SpoIIAA | 1H4X [75]                         | Ser57                    | 49–58             | 1H4Z [75]                           |
| Psp     | 1J9T [76]                         | Asp11                    | 10–18             | 1LT0 [77]                           |
| CDK2    | 15T [44]                          | Thr160                   | 152–163           | 1FN1 [48]                           |
| ERK2    | 2ERK [78]                         | Thr183, Tyr185           | 172–186           | 1ERK [79]                           |
| FixJ    | 1DSW [57]                         | Asp54                    | 73–99             | 1DCK [62]                           |
| Spo0A   | 1QMP [80]                         | Asp55                    | 54–63             |                                     |
| GLM     | 1MK6                               | Ser74                    | 64–75             |                                     |
| PMM     | 1KS5 [81]                         | Ser108                   | 105–114           |                                     |
| Pim1    | 2BIK                               | Ser261                   | 251–262           |                                     |
| PSIIγ   | 1CM8 [82]                         | Thr183, Tyr185           | 177–189           |                                     |
| β-PGM   | 1LYH [83]                         | Asp48                    | 7–16              |                                     |
| LCK     | 3LCK [84]                         | Tyr394                   | 389–402           |                                     |

Unphosphorylated structures are listed only for the cases in which we attempt to predict the phosphorylated structure starting from the unphosphorylated structure. In the case of FixJ, the predicted region contains a helix (ten residues) and two surrounding loops (eight and five residues), and the phosphorylated amino acid is not located within this region. We include this case to highlight the ability to extend our methods to regions larger than loops and their immediate surroundings.

β-PGM, β-phosphoglucomutase; GLM, probable glutaminase from Bacillus subtilis; LCK, human lymphocyte kinase; Pim1, proto-oncogene serine/threonine protein kinase pim1; PMM, phosphomannomutase; Psp, phosphoserine phosphatase; Spo0A, stage 0 sporulation factor A.

DOI: 10.1371/journal.pcbi.0020032.t003

Table 4. Reconstruction of Phosphorylated Loop Conformations

| Protein | Loop Length | Loop Only | Loop + Surroundings |
|---------|-------------|-----------|---------------------|
|         | P Error     | BB RMSD   | Heavy RMSD          | Sampled BB RMSD Range | P Error | BB RMSD | Heavy RMSD | Sampled BB RMSD Range |
| Hpr     | 7           | 2.9       | 0.4                 | 1.7                 | 0.2–3.3 | 2.8     | 0.3       | 0.8          | 0.2–3.5       |
| SpoIIAA | 10          | 0.7 (2.1) | 0.3 (0.4)           | 1.4                 | 0.2–3.8 | 0.6 (2.2) | 0.3 (0.8) | 1.2          | 0.2–6.6       |
| Psp     | 9           | 0.2       | 0.2                 | 0.5                 | 0.2–3.2 | 0.3     | 0.2       | 1.0          | 0.2–4.0       |
| CDK2    | 12          | 0.6       | 1.7                 | 3.5                 | 1.2–8.4 | 1.7     | 1.8       | 1.7          | 1.0–12.5      |
| ERK2    | 15          | 0.3/0.3   | 3.8                 | 5.2                 | 1.8–10.8 | 2.3/1.0 | 4.7       | 3.2          | 2.6–14.2      |
| FixJ    | 8/10/5      | 0.1       | 0.5                 | 1.2                 | 0.4–4.7 | 0.3     | 0.8       | 1.0          | 0.4–4.7       |
| Spo0A   | 10          | 0.3       | 0.2                 | 0.8                 | 0.2–3.1 | 0.8     | 0.4       | 0.9          | 0.2–4.1       |
| GLM     | 12          | 0.1       | 0.2                 | 1.0                 | 0.2–6.2 | 0.2     | 2.1       | 1.6          | 0.4–5.8       |
| PMM     | 10          | 0.3       | 0.4                 | 1.0                 | 0.3–5.4 | 0.3     | 1.0       | 1.1          | 0.4–5.9       |
| Pim1    | 12          | 1.2       | 1.2                 | 2.9                 | 1.2–14.9 | 1.3     | 1.4       | 2.2          | 0.4–14.8      |
| PSIIγ   | 12          | 0.4/0.5   | 2.3                 | 2.8                 | 1.1–5.7 | 0.7/5.6 | 1.4       | 1.6          | 1.2–8.2       |
| β-PGM   | 10          | 0.9       | 0.6                 | 1.3                 | 0.2–7.1 | 0.6     | 0.2       | 0.6          | 0.2–8.8       |
| LCK     | 14          | 1.2       | 3.0                 | 3.3                 | 1.6–10.1 | 2.5     | 2.7       | 2.4          | 1.8–11.9      |

“Reconstruction” refers to the ab initio prediction of loop residues with the rest of the protein in the phosphorylated conformation. “Loop Only” refers to the prediction of only the loop in question, whereas “Loop + Surroundings” refers to the prediction of the loop in question in addition to surrounding side chains with at least one atom within 4.5 Å of any atom in the loop. The accuracy of the prediction is assessed by three measures: the error in the predicted phosphorus atom position (“P Error”), a backbone RMSD measure (“BB RMSD”), and the RMSD for all heavy atoms included in the prediction (“Heavy RMSD”). See the text for details. The “Sampled BB RMSD Range” column indicates the range of backbone RMSDs of sampled loops, emphasizing that our ab initio build-up procedure samples many different conformations of the loop to be predicted, and the final prediction is the lowest-energy conformation sampled. In the case of FixJ, the predicted region contains a helix (ten residues) and two surrounding loops (eight and five residues), and the backbone and heavy atom RMSDs are calculated over the entire 23-residue region. For SpoIIAA, the predictions using a phosphate group with a –2 charge (as in the other test cases) gave poor results (in parentheses), whereas using a protonated phosphate group gave much better results; see text for details. Abbreviations used for the protein names are defined in Table 3.

DOI: 10.1371/journal.pcbi.0020032.t004
implies that the predominant charge state is generally of the phosphorylated amino acids is approximately 6, which phosphate group leads to excellent results. In current work, we are performing the loop predictions with a protonated phosphate (the pH used in the crystallization conditions is 6.5). In fact, the proximity of Asp56 is likely to shift the pKa higher and favor field from the rest of the protein. In this case, the close macromolecules due to desolvation and/or the electrostatic pKa's of titratable groups can be significantly shifted in confirm this assumption. However, it is well known that the pKa's of titratable groups can be significantly shifted in physiological conditions, and the good results we obtained for the other test cases using a phosphate group with a –2 charge (as in the other test cases) gave poor results (in parentheses), whereas using a protonated phosphate group gave much better results; see text for details. ERK2, as described in the text, undergoes a domain reorientation as well as a loop movement, which is likely the cause of the poor predictions. The improvement of the CDK2 prediction when optimization of surrounding side chains is performed in concert with the loop prediction highlights the need to incorporate side chain flexibility in the calculation to successfully predict the phosphorylated loop conformation from the unphosphorylated structure.

| Protein      | Loop Length | Loop Only | Heavy | Sampled BB | Loop + Surroundings | Heavy | Sampled BB |
|--------------|-------------|-----------|-------|------------|---------------------|-------|------------|
|              |             | P Error   | BB RMSD | RMSD       | Range               |      | Range      |
| HPy          | 7           | 1.0       | 0.4    | 1.3        | 0.4–6.1             | 1.0   | 0.5        | 1.1        | 0.4–7.0 |
| SpolI AA     | 10          | 0.3 (6.1) | 0.7 (5.7)| 1.7        | 0.6–7.1             | 0.9 (6.5)| 0.8 (3.9)| 1.5        | 0.5–6.7 |
| Psp          | 9           | 2.1       | 0.5    | 0.8        | 0.4–5.9             | 0.4   | 0.4        | 1.0        | 0.4–5.9 |
| CDK2         | 12          | 8.8       | 6.7    | 8.1        | 3.2–9.2             | 0.8   | 2.9        | 3.6        | 2.0–11.0 |
| ERK2         | 15          | 15/14.8   | 8.8    | 10.7       | 3.6–10.9            | 15.4 / 3.0 | 6.7 | 5.0        | 3.6–13.1 |
| FixJ         | 8/10/5      | 1.3       | 1.1    | 1.5        | 0.4–4.8             | 1.0   | 1.6        | 1.7        | 0.4–4.8 |

“Prediction” refers to the ab initio prediction of loop residues with the rest of the protein held in the unphosphorylated conformation. “Loop Only” refers to the prediction of only the loop in question; these results are provided only for comparison. “Loop + Surroundings” refers to the prediction of the loop in question in addition to surrounding side chains with at least one atom within 4.5 Å of any atom in the loop; these are the results that should be used in evaluating the success of our method. Definitions of the columns are provided in Table 4, and abbreviations used for the protein names are defined in Table 3. In the case of FixJ, the predicted region contains a helix (ten residues) and two surrounding loops (eight and five residues), and the backbone and heavy atom RMSDs are calculated over the entire 23-residue region. For SpolI AA, the predictions using a phosphate group with a –2 charge (as in the other test cases) gave poor results (in parentheses), whereas using a protonated phosphate group gave much better results; see text for details. ERK2, as described in the text, undergoes a domain reorientation as well as a loop movement, which is likely the cause of the poor predictions. The improvement of the CDK2 prediction when optimization of surrounding side chains is performed in concert with the loop prediction highlights the need to incorporate side chain flexibility in the calculation to successfully predict the phosphorylated loop conformation from the unphosphorylated structure.

DOI: 10.1371/journal.pcbi.0020032.t005

Protein Phosphorylation In Silico

Table 5. Prediction of Phosphorylated Loop Conformations

To test our ability to model conformational responses that extend beyond a single loop and its surroundings, we consider the repacking of helices that is known to occur upon phosphorylation of response regulator proteins of prokaryotes [3]. The response regulator proteins are the second part of the two-component signaling system that bacteria use to sense and respond to their environment. In most two-component systems, a sensor histidine kinase autophosphorylates a His upon sensing signal and transfers this phosphate to an Asp on a response regulator. Upon phosphorylation, certain structural changes occur, most notably a helix shift and loop rearrangement, which ultimately allow the response regulator to activate transcription of a set of genes.

We used our existing helix prediction algorithm to model the structural changes of the response regulator protein FixJ (from Sinorhizobium meliloti [57]), specifically, the loop-helix-loop region that undergoes a helix shift and loop rearrangements. Another response regulator protein, Spo0A, has crystal structures in both the phosphorylated and unphosphorylated forms, but the crystal structure for the unphosphorylated form of Spo0A is a domain-swapped dimer, solved at a pH of 4.5. As discussed in Materials and Methods, the helix prediction algorithm employs rigid body sampling for the helix (residues 87–94) combined with the hierarchical loop sampling algorithm for predicting the connecting loops [42] (residues 79–86 and 95–99). The overall backbone RMSD for reconstructing this region in the phosphorylated crystal structure is 0.5 Å, with most of the error arising from the first flanking loop, which is longer, closest to the phosphate, and implementing a new version of our molecular mechanics algorithms that allows automatic identification of the optimal protonation state for phosphorylated amino acids and other titratable groups (especially His), using a thermodynamic cycle analogous to that employed by the numerous Poisson-Boltzmann based pKa prediction algorithms [56].

Predicting Conformational Changes of Helices in Response to Phosphorylation

PLoS Computational Biology | www.ploscompbiol.org April 2006 | Volume 2 | Issue 4 | e320246
and insulin receptor tyrosine kinase. In order to test the prediction of the protein structure, such as a hinge-bending movement or structures that were less than 15 residues in length. We exclude protein structures determined by X-ray crystallography (with better resolution) from in silico phosphorylation of the unphosphorylated structure (unpublished results). The reconstruction (starting from the unphosphorylated structure) is in red, and the prediction (starting from the unphosphorylated structure) is in green. Figure prepared with Chimera [72]. DOI: 10.1371/journal.pcbi.0020032.g008

Figure 8. Helix Reconstruction and Prediction of FixJ
The loop-helix-loop region of FixJ was predicted starting from either the phosphorylated structure (blue) or the unphosphorylated structure (unpublished results). The reconstruction (starting from the phosphorylated structure) is in red, and the prediction (starting from the unphosphorylated structure) is in green. Figure prepared with Chimera [72].

Conclusion
We have described our initial efforts to predict and understand how the structure of a protein is modulated by post-translational phosphorylation. We believe that this work has practical significance in that we demonstrate that it is possible to make testable predictions concerning the structure of phosphorylated proteins, given the structure of the unphosphorylated protein and a known site of phosphorylation. In this work, we have restricted our efforts to predicting relatively modest, localized conformational changes, and we have assumed knowledge of which portions of the protein undergo significant conformational change (those portions closest to the phosphorylated amino acid). Despite these limitations, this modeling technology can be used to create hypotheses about mechanisms of regulation by phosphorylation that can then be tested experimentally, e.g., by site-directed mutagenesis. Applications of this type are underway. In addition, our in-depth case study of CDK2 illustrates how the computational methods can be used to obtain new insights into the energetic subpinnings of phosphorylation-induced conformational change even in a well-studied system.

Materials and Methods
Dataset selection. We searched the PDB [58] for phosphorylated protein structures determined by X-ray crystallography (with better than 2.5 Å resolution) that are phosphorylated on well-ordered loop structures that were less than 15 residues in length. We exclude structures in which phosphorylation causes a large, global rearrangement of the protein structure, such as a hinge-bending movement or domain rearrangement, as is the case with glycosyn phosphorylase [59] and insulin receptor tyrosine kinase [60,61]. In order to test the limitations of our method, we include one test case, ERK2, in which phosphorylation causes a domain rearrangement [3]. We also include a prokaryotic response regulator, FixJ, in which phosphorylation of an Asp induces a significant conformational change in the orientation of a helix [57,62]; this case is successfully treated with an extension of our methods described below. The test set is listed in Table 3.

We determined the loop length that we would predict for each phosphorylated structure using visual inspection. In the cases in which crystallographic structures are available for both the unphosphorylated and phosphorylated protein, these structures were superimposed and the residues to be predicted were defined as the portion of the loop that deviated in the superposition. For the reconstruction of phosphorylated structures without knowledge of the unphosphorylated form, the loop residues to be optimized were determined using a combination of visual inspection of secondary structure and crystallographic B-factors.

Molecular mechanics energy function. All energy calculations use the OPLS-AA force field [57,63] and the Surface Generalized Born (SGB) model of solvation [49,50]. The molecular mechanics energy function represents electrostatics by a relatively simple model of fixed atomic partial charges interacting through the Coulomb approximation. The solvent model captures key effects of desolvation with relatively modest computational expense. Despite the simplicity of the energy function (i.e., it neglects polarizability contributions to electrostatics, and implicit solvent models have well-known limitations), it performs well in predicting conformations of phosphorylated loops (see Results).

The force field parameters for the phosphorylated amino acids were generated by an automated atom-typing algorithm provided in the Impact software package. The atomic partial charges for the phosphorylated amino acid side chains were adjusted slightly from the default values by performing quantum chemistry calculations. The partial charges for phosphoserine (pSer) and phosphothreonine (pThr) were taken from previous work by Wong et al. [65], whereas charges for phosphotyrosine (pTyr) and phosphoaspartate (pAsp) were determined in their −2 and −1 charge states by performing quantum mechanical calculations with the software program Jaguar [60]. Methyl- and benzy1-phosphates were used to represent the pTyr phosphate, and acetyl phosphate was used for pAsp. Geometry optimization of the phosphate ion was carried out at the HF/6–31G** level, incorporating a condensed-phase environment via a self-consistent reaction field (SCRF) algorithm [67,68]. Single point calculations were performed at the LMP2/cc-pvDZ level, also with SCRF treatment of solvation. Electrostatic potential fitting was used to determine the partial charges. The atomic partial charges for all four phosphorylated amino acids are provided in Tables S1–S4.

Loop prediction methodology. This study uses the method of Jacobson et al. [40] for predicting loop conformations. In brief, the loop prediction methodology uses an ab initio dihedral sampling scheme to enumerate conformations of the loop backbone that are free from steric clashes. Other methods have employed similar dihedral angle sampling schemes, including ICM [60], Chimera [72], and the work of DePristo, et al [71]. Unlike Monte Carlo and Molecular Dynamics sampling schemes, the algorithm itself has no knowledge of the starting conformation of the loop, and therefore, does not start predictions based on a starting structure of the loop. These closed backbone conformations are clustered, and a single member of each cluster is then selected for side chain addition and optimization, followed by complete energy minimization. The lowest energy structure is selected as the output of the loop prediction algorithm. The algorithm also permits explicit treatment of crystal packing. We have used this capability in this work (in the “loop reconstruction” cases, as described below), but did not identify any clear crystal-packing artifacts relevant to the conformations of the phosphorylated loops.

The Jacobson et al. paper [40] describes a hierarchical refinement procedure in which multiple iterations of the loop prediction algorithm are used to reduce errors caused by insufficient sampling. The parameters in this scheme have been slightly modified in this study due to the inclusion of some rather long loops (up to 15 residues) in our test set. The first stage allows for unrestricted sampling. After this stage is complete, the top ten lowest energy structures are passed to a first refinement stage in which more extensive sampling is performed around these low-energy basins. Specifically, loop conformations in this stage are then selected for side chain addition and optimization, followed by complete energy minimization. The lowest energy structure is selected as the output of the loop prediction algorithm. The algorithm also permits explicit treatment of crystal packing. We have used this capability in this work (in the “loop reconstruction” cases, as described below), but did not identify any clear crystal-packing artifacts relevant to the conformations of the phosphorylated loops.
maximum Cα deviation is restricted to 5 Å. Finally, the five lowest-energy structures from this stage are subjected to a third and final round of refinement, in which the maximum Cα deviation is restricted to 2.5 Å. In all, this procedure provides a rank-ordered list consisting of at least 15 loops and their associated energies in the context of the full protein. The final prediction is the loop conformation with the lowest energy.

Most published tests of loop prediction methods, including the Jacobson et al. paper [40], evaluate their success by the ability to recognize phosphorylated loops in a native protein structure. In these tests, all portions of the protein other than the loop in question remain in the native conformation during the simulation. Success of a prediction methodology in such a test is an important prerequisite for more realistic applications. We perform loop reconstruction tests in this work to assess the ability of the molecular mechanics energy function to identify correct conformations of phosphorylated loops, and to ensure that the sampling methods are sufficient to generate near-native conformations. However, predicting conformational changes induced by phosphorylation, i.e., by phosphorylating a residue in silico, is quantitatively more challenging. In the cases we consider, the sites of phosphorylation are located on loops, and most of the conformational change is localized to that loop.

However, there is always some degree of conformational rearrangement in the vicinity of the loop, especially in the configurations of side chains contacting it. Similarly, predicting the conformation of a loop in a homology model is more challenging than reconstructing a loop in a native protein structure because the environment surrounding any given loop in a homology model contains errors that can affect the loop prediction accuracy. We address this issue by performing rotamer optimization and minimization of side chains in the immediate vicinity of the loop concurrently with the optimization of the side chains on the loop itself. In the test set presented here, this strategy performs well in predicting local structural changes, despite the fact that there are also some changes in backbone conformation in the surroundings. We speculate that small changes in the conformations of the surrounding side chains can compensate for not explicitly allowing backbone relaxation. We also use this strategy in the control studies of reconstructing phosphorylated loops in which, interestingly, it sometimes improves the accuracy.

Helix prediction methodology. The helix prediction algorithm used in this paper is based on the work of Li et al [42]. Briefly, the helix backbone is treated as a rigid body and sampled in six degrees of freedom (three random and three rotations). To address this issue, two flanking loops are sampled using the loop prediction algorithm described above. Again, the method broadly samples the possible configurations of the helix and surrounding loops, independent of any specific structural context and then hierarchically samples more finely around low-energy basins. As with the loop prediction algorithm, side chains on the loop-helix-loop region, and the surroundings if desired, are sampled using a rotamer-based optimization algorithm.

Development of "rotamer" libraries for phosphorylated residues. The loop and helix prediction algorithms require the use of a rotamer library for sampling side chain conformations. However, the number of phosphorylated residues in the PDB is insufficient to construct these libraries by statistical analysis of observed side chain conformations. Instead, we obtained a rotamer library by exhaustively exploring the energy landscape of the side chains of phosphorylated amino acid dipeptides, and retaining all conformations below an energy threshold that ensures inclusion of all experimentally observed rotamers. We use the same energy function as in the rest of the work, i.e., OPLS-AA/SG. The pSer and pThr side chains are sampled at 10° resolution, whereas pTyr and pAsp are sampled at 30° resolution, due to the larger number of rotatable bonds. The total number of rotamers is 802 for pSer, 501 for pThr, 858 for pTyr, and 1,008 for pAsp. These values do not include the rotation of the phosphate group about the phosphoestor bond, which is sampled uniformly at the same resolution as the other bonds. The rotamer libraries are provided in Tables S3–S8.

Supporting Information

Table S1. Partial Atomic Charges Employed for pThr

Table S2. Partial Atomic Charges Employed for pSer

Table S3. Partial Atomic Charges Employed for pTyr

Table S4. Partial Atomic Charges Employed for pAsp

Table S5. Rotamer Library Used for Sampling pThr

Table S6. Rotamer Library Used for Sampling pSer

Table S7. Rotamer Library Used for Sampling pTyr

Table S8. Rotamer Library Used for Sampling pAsp

Accession Numbers

The Protein Data Bank (http://www.rcsb.org/pdb) accession numbers for the proteins discussed in this paper are as follows: β PGM (PDB ID: 1LVH); CDK2, phosphorylated (PDB ID: 1JST); CDK2, unphosphorylated (PDB ID: 1FN); CDK2/KAP complex (PDB ID: 1FQ1); cyclin A (PDB ID: 1B39); ERK2, phosphorylated (PDB ID: 1ERK); ERK2, unphosphorylated (PDB ID: 1E7R); Fj1, phosphorylated (PDB ID: 1D5W); Fj1, unphosphorylated (PDB ID: 1TCK); GLM, (PDB ID: 1MKI); Hpr, phosphorylated (PDB ID: 1F0U); Hpr, unphosphorylated (PDB ID: 1FTF); LCK, (PDB ID: 3LCK); Pim1, (PDB ID: 2BIK); PMM, (PDB ID: 1K53); Psp, phosphorylated (PDB ID: 1F7); Psp, unphosphorylated (PDB ID: 1L7O); P38 γ, (PDB ID: 1CM8); SpoIIAA, phosphorylated (PDB ID: 1H4X); and SpoIIAA, unphosphorylated (PDB ID: 1H4Z); Spo0A, (PDB ID: 1QMP).

Acknowledgments

Author contributions. MPJ analyzed the data. ESG, AN, and MPJ wrote the paper. ESG, AN, and MPJ performed the experiments. ESG, AN, and MPJ conceived and designed the experiments. ESG, AN, and MPJ performed the experiments. ESG, AN, and MPJ analyzed the data. ESG, AN, and MPJ wrote the paper.

Funding. ESG acknowledges support from the Burroughs Wellcome Fund, the National Institutes of Health grant GM08284. This work was supported by National Science Foundation CAREER grant MCB-0346399, and by start-up funds provided by Howard Hughes Medical Institute Biomedical Research Support Program grant #5309246 to the University of California, San Francisco School of Medicine.

Competing interests. MPJ is a member of the Scientific Advisory Board of Schrodinger, Inc., a company that distributes scientific software, including software that he has written. However, MPJ states that he believes that the work described in this manuscript has no significant commercial value, since the study is concerned with a fundamental problem in biophysics (i.e., it is unrelated to drug discovery, which is the company’s focus).

References

1. Kreegipuu A, Blom N, Brunak S (1999) PhosphoBase, a database of phosphorylation sites, release 2.0. Nucleic Acids Res 27: 2577–2579.

2. Bridges AJ (2001) Chemical inhibitors of protein kinases. Chem Rev 101: 2541–2571.

3. Johnson LN, Lewis RJ (2001) Structural basis for control by phosphorylation. Chem Rev 101: 2209–2242.

4. Keane NE, Chavanieu A, Quirk PG, Evans JS, Levine BA, et al. (1994) Structural determinants of substrate selection by the human insulin-receptor protein-tyrosine kinase. Eur J Biochem 226: 525–536.

5. Quirk PG, Patchell VB, Colyer J, Drago GA, Gao Y (1996) Conformational effects of serine phosphorylation in phospholamban peptides. Eur J Biochem 236: 85–91.

6. Quirk PG, Patchell VB, Gao Y, Levine BA, Perry SV (1995) Sequential phosphorylation of adjacent serine residues on the N-terminal region of cardiac troponin-I: Structure-activity implications of ordered phosphorylation. FEBS Lett 370: 175–178.

7. Chavanieu A, Keane NE, Quirk PG, Levine BA, Calas B, et al. (1994) Phosphorylation effects on flanking charged residues: Structural implica-

PLoS Computational Biology | www.ploscompbiol.org April 2006 | Volume 2 | Issue 4 | e320248
tions for signal-transduction in protein-kinases. Eur J Biochem 224: 115–123.

8. Thoeye A, Lindemann A, Kinzel V, Reed J (1999) Direct effects of phosphorylation on preferred backbone conformation of peptides: A nuclear magnetic resonance study. Biophys J 76: 76–87.

9. Thoeye A, Pipkorn R, Bosseweger D, Kinzel V, Reed J (2001) Influence of myristoylation, phosphorylation, and deamidation on the structural behavior of the catalytic subunit of CAMP-dependent protein kinase. Biochemistry 40: 225–231.

10. Costaudo G, Everard-Todeschi N, Gharbi-Benarous J, Benarous R, Girault JP (2001) Conformational analysis by NMR and molecular modeling of the 41–69 hydrophilic domain of HIV-1 encoded v-Pr55 on an oligo-catalytic antigen. J Mol Biol 312: 803–817.

11. Hwang I, Thorgeirsson T, Lee J, Kustu S, Shin YK (1999) Physical evidence for phosphorylation-dependent conformational change in the enhanced binding activity of NtrC. Proc Natl Acad U S A 96: 4880–4885.

12. Smart JL, McCammon JA (1999) Phosphorylation stabilizes the N-termini of alpha-helices. Biopolymers 49: 225–233.

13. Shen TY, Wong CF, McCammon JA (2001) Atomic Brownian dynamics simulation of peptide phosphorylation. J Am Chem Soc 123: 9107–9111.

14. Hamelberg B, Shen T, McCammon JA (2005) Phosphorylation effects on cis/trans isomerization and the backbone conformation of serine-proline motifs. Accelerated molecular dynamics analysis. J Am Chem Soc 127: 1969–1747.

15. Luo R, David L, Hung H, Devaney J, Gilson MK (1999) Strength of solvent-exposed salt-bridges. J Phys Chem B 103: 727–736.

16. Maier MC, Feng J, Am SS, Troxel JY, Henschel-Edman A, et al. (1998) Structural examination of the influence of phosphorylation on the binding of fibronectinpeptide A to bovine thrombin. Biochemistry 37: 5888–5902.

17. Zhou Y, Abagyan R (1998) How and why phosphorytrosine-containing peptides bind to SH2 and PTB domains. Structure 6: 513–522.

18. Feng MH, Philippopoulos M, MacKerrell AD, Lim C (1996) Structural characterization of the phosphorytrosine binding region of a high-affinity SH2 domain-phosphopeptide complex by molecular dynamics simulation and chemical shift calculations. J Am Chem Soc 118: 11265–11277.

19. Stultz CM, Levin AD, Edelman ER (2002) Phosphorylation-induced conformational changes in a mitogen-activated protein kinase substrate. Implications for tyrosine hydroxylase activation. J Biol Chem 277: 47653–47661.

20. Roche P, Mouawad I, Perahia D, Samama JP, Kahn D (2002) Molecular modeling of the catalytic domains of protein phosphatases 1 and 2A with Zn2+ cofactors. J Mol Biol 323: 1236–1254.

21. Wozniak-Celmer E, Oldziej S, Ciarkowski J (2001) Theoretical models of the catalytic subunit of CAMP-dependent protein kinase. Biophys J 8: 931–939.

22. Roche P, Mouawad I, Perahia D, Samama JP, Kahn D (2002) Molecular modeling of the catalytic domains of protein phosphatases 1 and 2A with Zn2+ cofactors. J Mol Biol 323: 1236–1254.

23. Mendieta J, Gago F (2004) In silico activation of Src tyrosine kinase reveals structural basis of oncogenic activation caused by point mutations in the kinase domain of the MET proto-oncogene. Modeling studies. Proteins 44: 32–43.

24. Klose KE, Weiss DS, Kustu S (1993) Glutamate at the site of phosphor-rylation of nitrogen-regulatory protein NTRC mimics aspartyl-phosphate and activates the protein. J Mol Biol 232: 67–78.

25. McCabe TJ, Fulton D, Roman LJ, Sessia WC (2000) Enhanced electron flux and reduced calmodulin dissociation may explain “calcium-independent” eNOS activation by phosphorylation. J Biol Chem 275: 6125–6128.

26. Georgescu RE, Alexov EG, Gunner MR (2002) Combining conformational flexibility and continuum electrostatics for calculating pK(a)s in proteins. Biophys J 83: 1731–1748.

27. Birck C, Mourey L, Gouet P, Fabry B, Schumacher J, et al. (1999) Structural basis of oncogenic activation caused by point mutations in the kinase domain of the MET proto-oncogene. Modeling studies. Proteins 44: 32–43.

28. Mendieta J, Gago F (2004) In silico activation of Src tyrosine kinase reveals structural basis of oncogenic activation caused by point mutations in the kinase domain of the MET proto-oncogene. Modeling studies. Proteins 44: 32–43.
tyrosine kinase domain of the human insulin receptor. Nature 372: 746–754.

62. Gouet P, Fahy B, Guillet V, Birck C, Mourey L, et al. (1999) Structural transitions in the FisJ receiver domain. Structure Fold Des 7: 1517–1526.

63. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. J Phys Chem B 105: 6474–6487.

64. Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc 118: 11225–11236.

65. Wong SE, Bernacki K, Jacobson M (2005) Competition between intra-molecular hydrogen bonds and solvation in phosphorylated peptides: Simulations with explicit and implicit solvent. J Phys Chem B 109: 5219–5258.

66. Schrodinger LLC (2002) Jaguar 5.0 [computer program]. Portland (Oregon): Schrodinger, L.L.C.

67. Marten B, Kim K, Cortis C, Friesner RA, Murphy RB, et al. (1996) New model for calculation of solvation free energies: Correction of self-consistent reaction field continuum dielectric theory for short-range hydrogen-bonding effects. J Phys Chem 100: 11775–11788.

68. Tannor DJ, Marten B, Murphy R, Friesner RA, Stikoff D, et al. (1994) Accurate first principles calculation of molecular charge distributions and solvation energies from ab initio quantum mechanics and continuum dielectric theory. J Am Chem Soc 116: 11875–11882.

69. Abagyan R, Totrov M (1994) Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. J Mol Biol 235: 983–1002.

70. Bruccoleri RE, Karplus M (1987) Prediction of the folding of short polypeptide segments by uniform conformational sampling. Biopolymers 26: 137–168.

71. DePristo MA, de Bakker PIW, Lovell SC, Blundell TL (2003) Ab initio construction of polypeptide fragments: Efficient generation of accurate, representative ensembles. Proteins 51: 41–55.

72. Petersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, et al. (2004) UCSF Chimera—A visualization system for exploratory research and analysis. J Comput Chem 25: 1605–1612.

73. Audette GF, Engelmann R, Hengstenberg W, Deutscher J, Hayakawa K, et al. (2000) The 1.9 A resolution structure of phospho-serine 46 HPt from Enterococcus faecalis. J Mol Biol 303: 545–553.

74. Jia Z, Vondoselaar M, Hengstenberg W, Quill JW, Delbaere LT (1994) The 1.6 Å structure of histidine-containing phosphotransfer protein HPt from Streptococcus faecalis. J Mol Biol 256: 1341–1355.

75. Seavers PR, Lewis RJ, Brammigan JA, Verschueren KH, Mursudov GN, et al. (2001) Structure of the Bacillus cell fate determinant SpoIIAA in phosphorylated and unphosphorylated forms. Structure 9: 605–614.

76. Cho H, Wang W, Kim R, Yokota H, Damo S, et al. (2001) BeF(3)(-) acts as a phosphate analog in proteins phosphorylated on aspartate: Structure of a BeF(3)(-) complex with phosphoserine phosphatase. Proc Natl Acad Sci U S A 98: 8525–8530.

77. Wang W, Cho HS, Kim R, Jancarik J, Yokota H, et al. (2002) Structural characterization of the reaction pathway in phosphoserine phosphatase: Crystallographic “snapshots” of intermediate states. J Mol Biol 319: 421–431.

78. Canagarajah BJ, Khokhlatchev A, Cobb MH, Goldsmith EJ (1997) Activation mechanism of the MAP kinase ERK2 by dual phosphorylation. Cell 90: 859–869.

79. Zhang F, Strand A, Robbins D, Cobb MH, Goldsmith EJ (1994) Atomic structure of the MAP kinase ERK2 at 2.3 Å resolution. Nature 367: 704–711.

80. Lewis RJ, Brammigan JA, Muchova K, Barak I, Wilkinson AJ (1999) Phosphorylated aspartate in the structure of a response regulator protein. J Mol Biol 294: 9–15.

81. Regni C, Tipton PA, Beamer LJ (2002) Crystal structure of PMM/PPGM: An enzyme in the biosynthetic pathway of P. aeruginosa virulence factors. Structure 10: 269–279.

82. Bellon S, Fitzgibbon MJ, Fox T, Hsiao HM, Wilson KP (1999) The structure of phosphorylated p38gamma is monomeric and reveals a conserved activation-loop conformation. Structure Fold Des 7: 1057–1065.

83. Lahiri SD, Zhang G, Dunaway-Mariano D, Allen KN (2002) Caught in the act: The structure of phosphorylated beta-phosphoglucomutase from Lactococcus lactis. Biochemistry 41: 8351–8359.

84. Yamaguchi H, Hendrickson WA (1996) Structural basis for activation of human lymphocyte kinase Lck upon tyrosine phosphorylation. Nature 384: 484–489.