

# Substrate-Specific Reorganization of the Conformational Ensemble of CSK Implicates Novel Modes of Kinase Function

Michael A. Jamros<sup>1</sup>, Leandro C. Oliveira<sup>2</sup>, Paul C. Whitford<sup>3</sup>, José N. Onuchic<sup>3</sup>, Joseph A. Adams<sup>4</sup>, Patricia A. Jennings<sup>1\*</sup>

**1** Department of Chemistry and Biochemistry, University of California San Diego, La Jolla, California, United States of America, **2** Laboratório Nacional de Ciência e Tecnologia do Bioetanol – CTBE/CNPEM, Campinas, São Paulo, Brazil, **3** Center for Theoretical Biological Physics, Rice University, Houston, Texas, United States of America, **4** Department of Pharmacology, University of California San Diego, La Jolla, California, United States of America

## Abstract

Protein kinases use ATP as a phosphoryl donor for the posttranslational modification of signaling targets. It is generally thought that the binding of this nucleotide induces conformational changes leading to closed, more compact forms of the kinase domain that ideally orient active-site residues for efficient catalysis. The kinase domain is oftentimes flanked by additional ligand binding domains that up- or down-regulate catalytic function. C-terminal Src kinase (Csk) is a multidomain tyrosine kinase that is up-regulated by N-terminal SH2 and SH3 domains. Although the X-ray structure of Csk suggests the enzyme is compact, X-ray scattering studies indicate that the enzyme possesses both compact and open conformational forms in solution. Here, we investigated whether interactions with the ATP analog AMP-PNP and ADP can shift the conformational ensemble of Csk in solution using a combination of small angle x-ray scattering and molecular dynamics simulations. We find that binding of AMP-PNP shifts the ensemble towards more extended rather than more compact conformations. Binding of ADP further shifts the ensemble towards extended conformations, including highly extended conformations not adopted by the apo protein, nor by the AMP-PNP bound protein. These ensembles indicate that any compaction of the kinase domain induced by nucleotide binding does not extend to the overall multi-domain architecture. Instead, assembly of an ATP-bound kinase domain generates further extended forms of Csk that may have relevance for kinase scaffolding and Src regulation in the cell.

**Citation:** Jamros MA, Oliveira LC, Whitford PC, Onuchic JN, Adams JA, et al. (2012) Substrate-Specific Reorganization of the Conformational Ensemble of CSK Implicates Novel Modes of Kinase Function. *PLoS Comput Biol* 8(9): e1002695. doi:10.1371/journal.pcbi.1002695

**Editor:** Ruth Nussinov, National Cancer Institute, United States of America and Tel Aviv University, Israel

**Received:** May 29, 2012; **Accepted:** August 5, 2012; **Published:** September 20, 2012

**Copyright:** © 2012 Jamros et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported in part by National Institutes of Health Grants 5T32GM008326, DK54441, and GM67969. This work was supported by the Center for Theoretical Biological Physics sponsored by the NSF(Grant PHY-0822283) and by NSF- MCB-1051438. JNO is a CPRIT Scholar in Cancer Research sponsored by the Cancer Prevention and Research Institute of Texas. LCO was supported by the Brazilian Agencies Fundação de Amparo à Pesquisa de São Paulo (FAPESP) Grant 2011/13424-7 and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) Grant 503199/2009-5. The SSRL Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research, and by the National Institutes of Health, National Center for Research Resources, Biomedical Technology Program. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: pajennin@ucsd.edu

## Introduction

The Src family of tyrosine kinases (SFKs) is comprised of modular signaling enzymes involved in the control of cellular growth and differentiation. The members of this family contain three important structural domains: a C-terminal tyrosine kinase domain (comprised of a small and large lobe), which is preceded in sequence by the non-catalytic regulatory SH2 and SH3 domains [1,2,3,4]. While phosphorylation of the activation loop is autocatalytic, phosphorylation of the C-terminal tail is inhibitory and requires Csk [5]. Csk contains the same structural domains as SFKs, but lacks an inhibitory C-terminal tail and an N-terminal sequence for membrane localization [6]. Additionally, Csk is not regulated through phosphorylation of its activation loop. Instead, Csk is constitutively active and increased activity is coupled to its association with membrane adaptor proteins.

Each domain of Csk plays a role in interacting with various binding partners. The kinase domain binds and phosphorylates all

nine members of the Src family [7], a process that requires the binding of ATP and magnesium. The SH2 domain of Csk is responsible for binding a large number of scaffolding proteins. Since Csk is a cytosolic protein and the substrate Src is membrane localized, localization of Csk to Src requires interaction between the Csk SH2 domain and several scaffolding proteins. These scaffolding proteins include the ubiquitously expressed transmembrane protein Cbp, Caveolin-1, Paxillin, the insulin receptor substrate IRS-1, and Helicobacter pylori CagA [8,9,10,11]. Lastly, the SH3 domain has been shown to bind PKA and the phosphatase PEP [12,13].

The gamut of binding partners for Csk suggests it must be an adaptable protein in order to carry out all of its known functions. One aspect that may be central to this adaptability is nucleotide-derived conformational changes. Pre-steady-state kinetic studies suggest that a slow conformational change in Csk limits Src phosphorylation [14,15]. In addition, previous work from our lab examined the effects of nucleotide binding on Csk using hydrogen-

## Author Summary

The Src protein kinases are integral members of numerous signaling pathways involved in cellular growth and differentiation. The master regulator of the Src family is the protein kinase Csk, which adds a phosphate to the C-terminal tail, inhibiting Src Kinase function. Proper regulation of these signaling pathways by Csk is essential as unregulated activity in these pathways is correlated with the development of various cancers and autoimmune diseases. Understanding the nature of the mechanism and structure of Csk may lead to therapeutics and a better understanding of Src signaling pathways. Conformational changes associated with nucleotide binding and release have been shown to regulate the efficiency of Src down-regulation by Csk. To obtain insights into the nature of these nucleotide-induced structural changes, we examined the conformation of Csk in solution while bound to the ATP analog AMP-PNP and product ADP using a combination of small angle x-ray scattering and molecular dynamics. Surprisingly, both nucleotides induce extended conformations of Csk compared to the apo-enzyme, suggesting a novel mode of function. Further understanding of this mode of function may aid in the design of cancer therapeutics that act by regulating Src signaling pathways by modulating the function of Csk.

deuterium exchange mass spectrometry (DXMS) and found binding of the ATP analog AMP-PNP as well as ADP led to changes in the protection of multiple peptides [16,17,18]. These data suggest that nucleotide binding to Csk has conformational effects on the protein that extend beyond the nucleotide pocket in the kinase domain and into the SH2 domain, the site of adaptor protein binding and catalytic regulation. While these studies establish the existence of long-range communication across Csk, they do not provide a structural framework in which to understand these observed inter-domain relationships.

Small angle x-ray scattering (SAXS) and NMR studies combined with molecular dynamics simulations has become an increasingly powerful tool to understand the conformational ensemble properties of proteins in solution [19,20,21,22,23,24,25,26]. Using a combination of SAXS and structure-based models (SBM), or SAXS-SBM, we showed that apo Csk is comprised of an ensemble of extended and compact conformations in solution, rather than adopting a single conformational form as depicted by the X-ray structure [27]. Here, we applied these methods to Csk to ask how nucleotide binding affects the conformational landscape of this flexible multi-domain kinase. We show that upon binding of the ATP analog AMP-PNP, there is a shift in the conformational ensemble of Csk towards more extended, open conformations. When bound to the reaction product ADP, the ensemble shifts even further towards these extended conformations. These data suggest that the orientation of regulatory domains in Csk is highly sensitive to the presence of nucleotides in the kinase domain. Most importantly, rather than inducing more compact forms, occupancy of the nucleotide pocket in the kinase domain induces more open forms of Csk.

## Results

### Solution Scattering of Csk

We previously collected SAXS data on the apo form of Csk and identified an ensemble of compact and extended conformations that describe the protein in solution [27]. In the current study, we asked whether or not the binding of nucleotides to Csk might alter

this ensemble. DXMS studies comparing apo Csk to nucleotide bound Csk (AMP-PNP and ADP) displayed changes in protection of many peptides, which suggests both the binding of ATP and delivery of the  $\gamma$  phosphoryl group of ATP induce global conformational changes in Csk [16]. In order to obtain a structural description of how binding of these nucleotides alters the conformation of Csk, we collected SAXS data on AMP-PNP and ADP bound Csk.

To determine if there are significant global changes in the solution structure of these bound states, we initially compared the scattering profile of each protein using a Kratky plot (Figure 1). Kratky plots ( $I(q) \cdot q^2$  versus  $q$ ) provide a qualitative analysis of the “folded-ness” of a protein. Globular proteins display a bell shaped curve, whereas extended proteins lack a peak and will either plateau, or have elevated  $I(q) \cdot q^2$  values for higher  $q$  values [28]. The apo protein plot indicates that Csk is well-folded with flexible linkers, which is consistent with our previous SAXS analysis, as well as crystallographic data [27,29]. However, both the AMP-PNP and ADP bound protein show signs of possible disorder and local unfolding, as indicated by the higher  $I(q) \cdot q^2$  values at larger scattering angles (higher  $q$ ). While the AMP-PNP bound protein data shows moderate deviations from the apo, the ADP bound protein shows a dramatic shift towards more disorder. Additionally, the experimentally determined radius of gyration ( $R_g$ ) values

**Figure 1. Small-angle X-ray scattering curves and Kratky plot of experimental data.** A) Solution scattering data for apo Csk (black), AMP-PNP bound Csk (red), and ADP bound Csk (blue). B) Kratky plot of experimental solution scattering data colored in the same manner as A. Scattering data is normalized for ease of comparison in the plot. doi:10.1371/journal.pcbi.1002695.g001

show the apo protein ( $R_g = 39 \text{ \AA}$ ) to be more compact than the AMP-PNP bound ( $R_g = 40 \text{ \AA}$ ) and the ADP bound protein ( $R_g = 44 \text{ \AA}$ ). Taken together, these data suggest significant shifts in the conformations of Csk when nucleotide is bound to the protein.

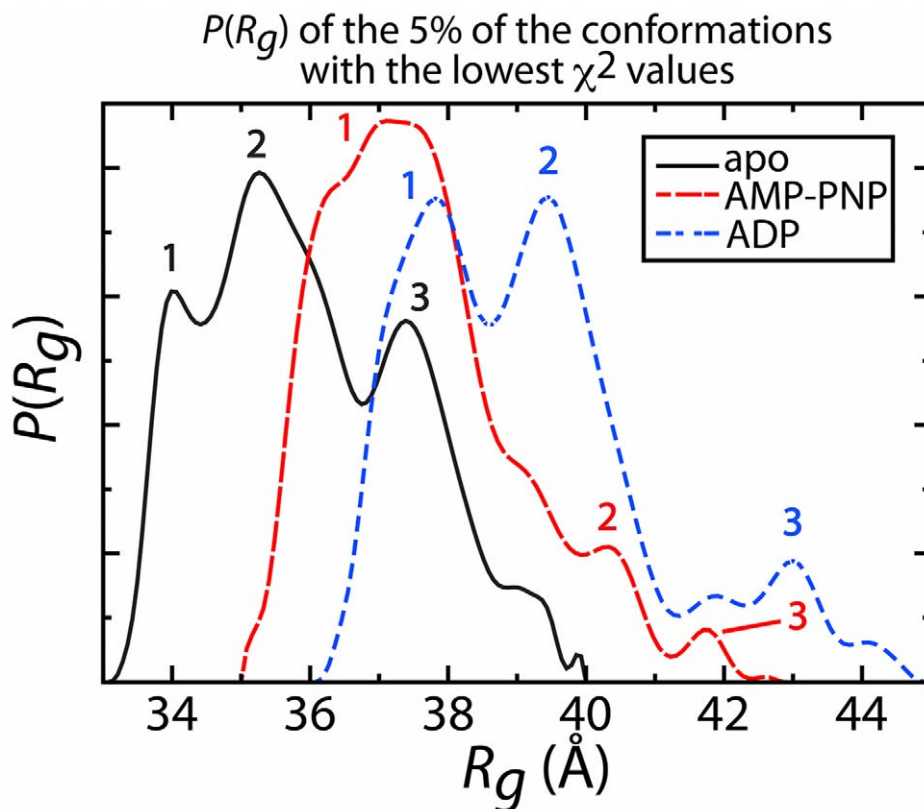
### Generation of Candidate Conformations via Molecular Dynamics

To further investigate the nature of Csk's conformational change upon nucleotide binding, we modeled the experimental scattering data using SBM. In our previous studies using SAXS-SBM we showed the crystal structure does not accurately describe apo Csk in solution, but rather it is composed of an ensemble of compact and extended conformations. The higher experimental  $R_g$  values of both the AMP-PNP and ADP bound forms of Csk suggest an ensemble of structures that are on average more extended than the apo protein. Using an all-atom SBM, we generated 40,000 candidate conformations for analysis. Theoretical scattering curves were generated for each individual conformation and then compared to the experimental scattering curve for all states of Csk using  $\chi^2$  to measure the goodness of fit. Specifically,  $\chi^2$  is a measure of the discrepancy between the theoretical and the experimental curves (see Methods for more detail). For each experimental curve we selected candidate conformations from the simulation, using the 5% of conformations with the lowest  $\chi^2$  values for each case. We then calculated  $P(R_g)$  ( $R_g$  values calculated using the `g_gyrate` module of Gromacs), the probability in  $R_g$  space, for all low- $\chi^2$  conformations (Figure 2). This process effectively applies a low- $\chi^2$  filter to the candidate

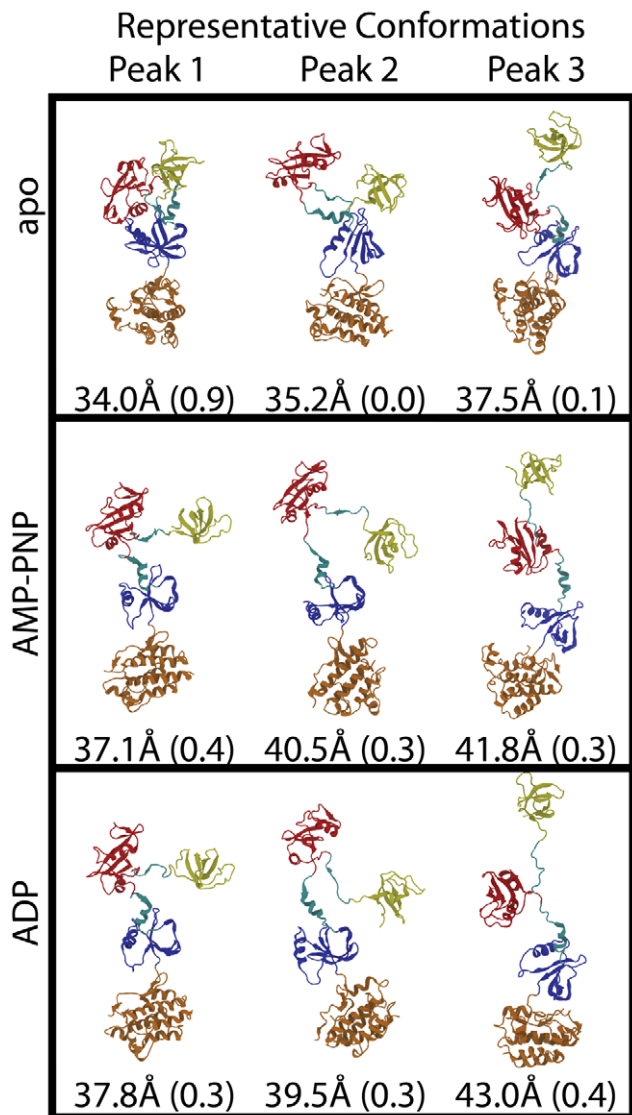
conformations, which range from  $R_g = 25.5 \text{ \AA}$  to  $R_g = 45.1 \text{ \AA}$ . In addition, we generated  $P(R_g)$  plots for the 2.5% and 7.5% of conformations with the lowest  $\chi^2$  values and the data were consistent, ensuring the robustness of the 5% selection criteria.

The low  $\chi^2$  filtered sets each possess multiple distinct peaks, representing likely conformational populations of Csk in solution. These populations most likely contribute to the experimental signal, but the ensemble distribution is not extracted from  $P(R_g)$  directly (see Methods). The apo protein prefers the most compact conformations, with peaks at  $\sim 34.0 \text{ \AA}$ ,  $\sim 35.2 \text{ \AA}$ , and  $\sim 37.5 \text{ \AA}$ , which is consistent with our previous analysis [27]. The AMP-PNP protein displays a shift towards more extended conformations with one major peak at  $\sim 37.8 \text{ \AA}$  and two minor peaks at  $\sim 39.5 \text{ \AA}$  and  $\sim 41.6 \text{ \AA}$ . Lastly, the ADP bound protein displays two major peaks at  $\sim 37.1 \text{ \AA}$  and  $\sim 40.5 \text{ \AA}$ , as well as a minor but potentially significant population at  $\sim 43.0 \text{ \AA}$ . Representative conformations from all peaks are depicted in Figure 3.

Common to all states of Csk studied here, we observe a similar major population ( $R_g \sim 37\text{--}38 \text{ \AA}$ ) that fits the experimental data well. However, each state of Csk displays conformations of differing  $R_g$  values that also describe the experimental data. The apo protein adopts more compact conformations than those observed in either of the nucleotide bound states. Upon binding of AMP-PNP, the predominant population (peak 1) covers a broad  $R_g$  range centered around a peak at  $\sim 37 \text{ \AA}$ , shifting away from the compact structures observed for the apo protein. When ADP is bound, the  $P(R_g)$  shifts even further towards higher  $R_g$  values, displaying two populations at high  $R_g$  that are not among the low  $\chi^2$  conformations for the apo and AMP-PNP bound protein.



**Figure 2. The probability of  $R_g$  ( $P(R_g)$ ) of the 5% of conformations with the lowest  $\chi^2$  values from the SBM simulations.** Apo Csk appears in black, AMP-PNP bound Csk appears in red, and ADP bound Csk appears in blue. The three most prominent populations, from which representative conformations were selected for ensemble analysis, are highlighted for each state. doi:10.1371/journal.pcbi.1002695.g002



**Figure 3. Representative conformations from the peaks of the  $P(R_g)$  plot, along with their  $R_g$  values.** Appearing next to the  $R_g$  values of each conformation is the fraction of the best-fit ensemble made up by that conformation. For all conformations the SH3 domain appears in yellow, the SH2 domain appears in red, the N-lobe of the kinase domain appears in blue, the C-lobe of the kinase domain appears in orange, and the domain-domain linkers appear in cyan.  
doi:10.1371/journal.pcbi.1002695.g003

Overall, the apo protein is the most compact form, binding of AMP-PNP shifts towards the more extended structures, and binding of ADP shifts to populations of even more extended conformations.

### Defining the Conformational Ensemble

While the  $P(R_g)$  analysis provides insight into the populations that are likely to be present in solution, it does not provide the probability of that population in solution. This is because each peak contains numerous conformations and possible bias by the simulation resulting in higher or lower sampling of a population. In order to quantify the probability of each population in solution we performed conformational ensemble analysis using conformations selected from the  $P(R_g)$  analysis. We selected four representative conformations from the three most prominent  $P(R_g)$  peaks

for each state, where all the selected conformations are distinct ( $\text{RMSD} > 2 \text{ \AA}$ ). Representative structures selected for ensemble analysis appear in Figure 3. From this pool of candidate conformations we created all possible combinations that include one conformation from each population. Each combination was considered a candidate ensemble, with which we performed ensemble analysis (see Methods). For each weighted ensemble, we generated a theoretical scattering curve, which was then compared to the experimental data (Figure 4).

For each state of Csk, we find an ensemble containing conformations from the peaks in  $P(R_g)$  that accounts well for the experimental data. The apo protein is heavily weighted towards compact conformations, where the best-fitting ensemble ( $\chi^2 = 2.6$ ) arises from a mix of 90% peak 1 ( $\sim 34 \text{ \AA}$ ) and 10% peak 3 ( $\sim 37 \text{ \AA}$ ). The absence of peak 2 ( $\sim 35 \text{ \AA}$ ) may be explained by a very small population or as a possible transient conformation between those represented at peaks 1 and 3. Both the AMP-PNP and ADP bound Csk conformational ensembles are comprised of extended structures. For the AMP-PNP bound protein, 40% of the best-fit ensemble is comprised of a conformation from the population of peak 1 ( $R_g \sim 37 \text{ \AA}$ ), with 30% each from peaks 2 ( $R_g \sim 40 \text{ \AA}$ ) and 3 ( $R_g \sim 41 \text{ \AA}$ ). This ensemble fits the experimental data with a  $\chi^2 = 2.7$ . When ADP is bound to Csk, there is a shift towards highly extended conformations with the best-fit ensemble consisting of 30% peak 1 ( $R_g \sim 38 \text{ \AA}$ ), 30% peak 2 ( $R_g \sim 39 \text{ \AA}$ ), and 40% peak 3 ( $R_g \sim 43 \text{ \AA}$ ), with  $\chi^2 = 3.8$ .

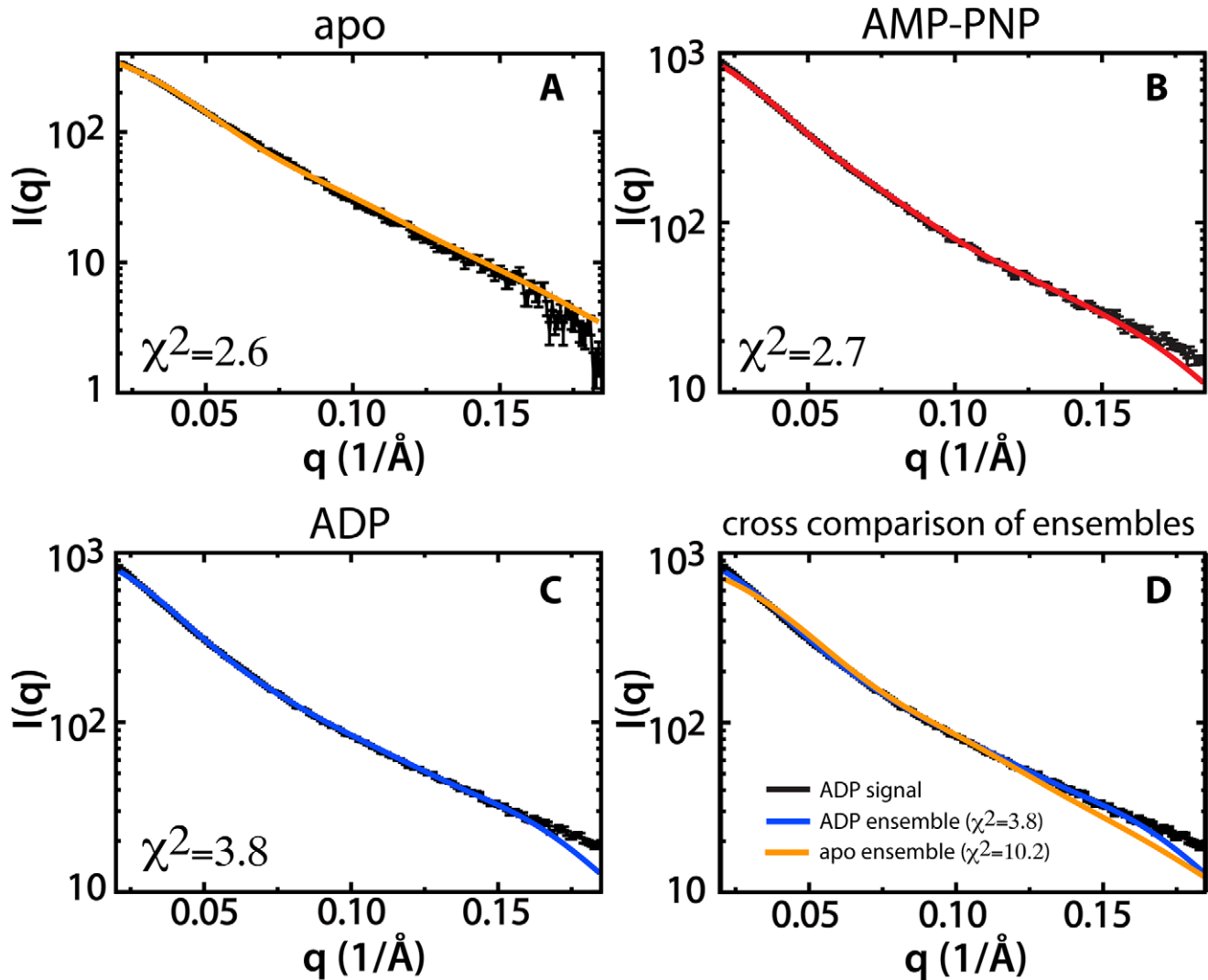
The population analysis results in a unique conformational ensemble for each state of Csk. The apo protein adopts primarily compact conformations, whereas the AMP-PNP bound protein adopts more extended conformations and the ADP bound protein adopts highly extended conformations. To confirm that each ensemble best describes that state of Csk, we fit the theoretical curves for all of the best-fit ensembles to the experimental data of all other states of Csk (Figure 4, Table 1). This comparison shows that the best-fit ensemble for each set of data describes the data better than any other ensemble does, reinforcing the notion that a unique solution is necessary for each state of Csk to describe the protein conformationally in solution.

## Discussion

### Binding of Nucleotide Shifts the Conformational Ensemble towards Extended Conformations

Using SBMs to model SAXS data, we find Csk adopts a variety of conformations in solution that range from compact ( $R_g = 34 \text{ \AA}$ ) to highly extended ( $R_g = 43 \text{ \AA}$ ). For every state of Csk examined here, the scattering data are best described by an ensemble of conformations, not a single structure. Interestingly, the  $P(R_g)$  distributions reveal a common population for the apo, AMP-PNP bound, and ADP bound protein, centered around  $37 \text{ \AA}$ . For the apo protein, this population accounts for 10% of the best-fit ensemble and is the most extended conformation present in that ensemble. When AMP-PNP is bound to Csk, this population accounts for 40% of the best-fit ensemble, with the remainder of the ensemble composed of more extended conformations. Binding of ADP shifts this ensemble even further towards extended conformations, but still contains a contribution from the population common to all states (30%). Notably, 40% of the best-fit ensemble is comprised of a conformation from peak 3 ( $43 \text{ \AA}$ ), a conformation that does not appear in the  $P(R_g)$  for either the apo protein or the AMP-PNP bound protein.

This shift in the ensemble upon nucleotide binding, while surprising for a protein kinase, is consistent with observations for other proteins. Many recent studies highlight the role of dynamics



**Figure 4. Theoretical scattering curves of the best-fitting conformational ensembles fitted to the experimental scattering curves (black) for apo Csk (A, orange), AMP-PNP bound Csk (B, red), and ADP-bound Csk (C, blue).** The  $\chi^2$  value for the ensemble fit to the experimental data appears in the lower left corner of the plot for each curve. A comparison of the fit of the apo Csk ensemble (orange) and the ADP-bound ensemble (blue) to the ADP experimental data (black) appears in D. doi:10.1371/journal.pcbi.1002695.g004

and shifts in population during kinase catalysis. The energy landscape of PKA-C, as characterized using NMR and molecular dynamics, shows the apo enzyme can explore the landscape and access open and closed conformations, while ligand binding

(nucleotide, substrate, or inhibitor) drives the enzyme to select alternate conformational states [30]. Combined crystallographic and NMR studies on arginine kinase display large substrate-induced domain motions and also show that the solution structure of the substrate-free protein is an equilibrium between substrate-bound and substrate-free forms [31]. The catalytic rates of two different forms of adenylate kinase are correlated with the timescales of motion in hinge and lid regions of the protein [32]. Similarly, allosteric effects upon ligand binding are observed in other nucleotide binding proteins. Solution NMR experiments on the nucleotide binding domain of the chaperone Hsp70 have revealed significant rotation in the subdomains when comparing the ATP and ADP bound states [33]. These changes impact the accessibility of a hydrophobic surface cleft distal from the nucleotide binding site, which is known to be essential for communication between the nucleotide binding domain and substrate binding domain. Allosteric effects are also observed upon cAMP binding to the exchange protein directly activated by cAMP (EPAC), where long-range communication between the

**Table 1. Cross-calculation of conformational ensembles.**

Best-fit Conformational Ensemble from:	$\chi^2$ calculated for experimental signal of:		
	Apo	AMP-PNP	ADP
Apo	2.6	6.4	10.2
AMP-PNP	5.1	2.7	4.6
ADP	6.0	3.2	3.8

The best-fit ensemble for each state of Csk was fit to the experimental data for all other states of Csk.

doi:10.1371/journal.pcbi.1002695.t001

phosphate binding cassette and the N-terminal helical bundle is transmitted via two intramolecular pathways [34]. Interconversion between competing conformations is also critical in the function of kinase adaptor proteins, such as Crk where proline isomerization switches the protein between an autoinhibitory conformation and an uninhibited, active conformation [35].

In the SAXS studies for Csk, the extended conformations of the nucleotide-bound states arise primarily from rearrangement of the regulatory domains. The observed extension of the SH2-kinase linker and SH3-SH2 linker may allow the domains to more easily rearrange to both make catalytically functional interdomain contacts and interact with other binding partners (discussed below). Communication between the SH2 and kinase domains is necessary for efficient catalysis as indicated by mutational studies of F183, a critical hydrophobic residue in the SH2-kinase linker [17]. When F183 is mutated to a glycine this communication is broken and catalytic efficiency decreases nearly 1000-fold. Interdomain communication between the SH3 and kinase domains also aids in catalysis. Mutations in the SH3-SH2 linker decrease catalytic activity up to five-fold by disrupting the interaction surface between the domains characterized by NMR [36].

### The Conformational Ensemble Provides Insight into Catalytic Cycling

Kinetic experiments using a short peptide substrate revealed that Csk exists in both low and high activity forms [15]. Overall, the high activity form represents only a small fraction of the total Csk, based on ‘burst’ amplitudes in pre-steady-state kinetic experiments. Interestingly, the present study also indicates that much of the enzyme in solution adopts open forms that are not expected to reflect high-activity states. In particular, the small and large kinase lobes adopt open rotomers in most calculated species, which are not expected to position ATP for highly productive phosphoryl transfer. In contrast, a higher population of Csk transiently adopts the active form when using Src rather than a peptide substrate in pre-steady-state kinetic experiments. This implies that a physiological target with increased enzyme-substrate contacts may promote a more closed, enzymatically-competent kinase domain. However, regardless of these substrate-dependent changes in conformer distribution detected in kinetic experiments, turnover is limited by a slow conformational change in Csk when using Src as a substrate [14]. The large changes observed between AMP-PNP- and ADP-bound forms in the SAXS-SBM study may correspond to the rate-limiting changes observed in the pre-steady-state kinetic experiments for Src phosphorylation. If the AMP-PNP- and ADP-bound Csk complexes reflect population distributions before and after Src phosphorylation, then the SAXS studies may effectively capture important forms of the enzyme within the catalytic cycle. Thus, the conversion of the more open forms observed in the ADP-bound state into the more closed forms in the ATP-bound complex may be correlated with the observed slow conformational changes in Csk kinetic experiments. These findings are consistent with recent work on Abl kinase, where the ADP-bound kinase domain is more open and exhibits increased flexibility [37].

### Csk Conformational Ensembles May Facilitate Signaling Processes

Csk is capable of interacting with a broad range of adaptor proteins through its SH2 domain [38]. These interactions likely require a high degree of adaptation and flexibility as Csk adopts a cytoplasmic location and then migrates to the plasma membrane in a phosphorylation-dependent manner [39]. This trafficking of

Csk between the cytoplasm and membrane phospho-receptors may be facilitated by the regulatory SH2 and SH3 domains, which can transiently detach from the kinase core and become available for recognition at the plasma membrane. Diffusion of the SH2 domain is faster than the full protein, which may allow Csk to search for an SH2 domain binding partner more quickly. Such a scenario is reminiscent of the ‘fly-casting’ model for intrinsically disordered proteins, where unstructured regions of a protein may offer an increased capture radius for substrate recognition of specific targets [40]. Similarly, linker flexibility in Csk may allow the SH2 domain to effectively search for adaptors and then draw the kinase to the membrane for Src down-regulation. The SAXS-SBM results reported in this study provide new insights into these potential motions and how they may be coupled to both signaling and substrate processing in the cell.

## Methods

### Protein Expression and Purification

The full-length Csk protein was expressed in *Escherichia coli* strain BL21(DE3) [41], and purified by Ni<sup>2+</sup> affinity chromatography [42]. The purified full-length enzyme was dialyzed against 50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 5 mM DTT with 15% (v/v) glycerol and then concentrated to 107 μM and stored at –80°C.

### Small-angle X-ray Scattering Measurements

SAXS data were collected at 25°C at beamline 4-2 of the Stanford Synchrotron Radiation Laboratory. Scattering was independent of protein concentration (0.5 mg/mL–2.5 mg/mL) indicating that interparticle ordering and aggregation are negligible. The AMP-PNP bound samples were generated through addition of AMP-PNP in a 50:1 molar ratio to protein as well as the addition of 10 mM MgCl<sub>2</sub>. Both AMP-PNP and MgCl<sub>2</sub> were added from stocks prepared in Csk dialysis buffer. The ADP samples were prepared in the same manner. Data shown were obtained with 1 mg/mL protein for the apo and AMP-PNP bound proteins and 1.8 mg/mL for the ADP bound protein. A more detailed description of SAXS methodology for the study of proteins in solution may be found elsewhere [43].

### Small Angle Scattering Data Analysis

The data were converted from a TIFF image to I(q) versus q using SasTool (<http://ssrl.slac.stanford.edu/~saxs/analysis/sastool.htm>). q is the angular dependence of the scattering profile, which can be expressed as  $q = 4\pi(\sin\theta/\lambda)$ , where  $\theta$  is half the scattering angle and  $\lambda$  is the wavelength of scattered X-rays. The radius of gyration, R<sub>g</sub>, was determined by Guinier analysis. The goodness of the linear fit of the data in the low q range indicated no significant non-specific aggregation of the protein.

### Analysis of the Structural Ensemble via Molecular Dynamics Simulations

We employed molecular dynamics simulations with an all-atom structure based model to generate an ensemble of candidate conformations (i.e. structures that may explain the scattering profile). A detailed description of the all-atom model may be found elsewhere [44]. From this ensemble, theoretical scattering profiles were generated, and these profiles were compared to the experimentally measured data.

We used an all-atom structure-based forcefield for the individual domains but only include steric interactions between domains. These models are based on the concepts of energy landscape theory [45,46,47], have a low computational cost, and provide

dynamic descriptions of proteins [48,49,50,51] and assemblies [52] that are in good agreement with experiments. The functional form of the all-atom structure-based forcefield is

$$V = \sum_{\text{bonds}} \varepsilon_r (r - r_0)^2 + \sum_{\text{angles}} \varepsilon_\theta (\theta - \theta_0)^2 + \sum_{\substack{\text{planar/improper} \\ \text{dihedrals}}} \varepsilon_\xi (\xi - \xi_0)^2 + \sum_{\text{backbone dihedrals}} \varepsilon_{\phi,bb} F_d(\phi) + \sum_{\text{sidechain dihedrals}} \varepsilon_{\phi,sc} F_d(\phi) + \sum_{\text{contacts}} \varepsilon_c \left[ \left( \frac{\sigma_{ij}}{r} \right)^{12} - 2 \left( \frac{\sigma_{ij}}{r} \right)^6 \right] + \sum_{\text{non-contacts}} \varepsilon_{nc} \left( \frac{\sigma_{nc}}{r} \right)^{12}$$

Where

$$F_d(\phi) = \left[ 1 - \cos(\phi - \phi_0) + \frac{1}{2} (1 - \cos(3(\phi - \phi_0))) \right]$$

$r_0$ ,  $\theta_0$ ,  $\xi_0$  and  $\phi_0$  are given the values found in the crystal structure (PDB code: 1K9A) [29] and  $\sigma = 2.5 \text{ \AA}$ ,  $\varepsilon_r = 100/\text{\AA}^2$ ,  $\varepsilon_\theta = 20/\text{rad}^2$ ,  $\varepsilon_\xi = 10/\text{rad}^2$  and  $\varepsilon_{nc} = 0.01$ . Each native atom-atom contact interacts *via* a Leonard-Jones 12-6 interaction, where the energetic minimum corresponds to the native distance  $\sigma_{ij}$  (as defined by the crystal structure). A contact is defined as any atom pair that is (a) separated by less than 6  $\text{\AA}$ , is (b) separated by at least 4 residues in sequence, and (c) has no atom between them (i.e. the ‘‘Shadow Algorithm’’ [53]).  $\varepsilon_{nc}$  and  $\sigma_{nc}$  define the excluded volume of each atom. Explicit representation of the atoms ensures that non-physical states, such as those with overlapping atoms or unrealistic bond angles, are disallowed. Contact and dihedral interactions were weighted as previously described [44]. In order to facilitate rapid sampling of all possible domain configurations, we removed all stabilizing inter-domain interactions and gave no configurational bias to the non-rigid dihedral angles (i.e. dihedrals not restrained by orbital hybridization) in the linkers.

The forcefield files for Gromacs [54] were generated by an online resource (<http://smog.ucsd.edu>) [55]. A timestep of 0.0005 time units was used and the simulation was coupled to a temperature bath *via* Langevin dynamics. The total simulation time was 20,000 time units, which corresponds to approximately 100  $\mu\text{s}$  [56].

## References

1. Thomas SM, Brugge JS (1997) Cellular functions regulated by Src family kinases. *Annu Rev Cell Dev Biol* 13: 513–609.
2. Boggon TJ, Eck MJ (2004) Structure and regulation of Src family kinases. *Oncogene* 23: 7918–7927.
3. Brown MT, Cooper JA (1996) Regulation, substrates and functions of src. *Biochim Biophys Acta* 1287: 121–149.
4. Abram CL, Courtneidge SA (2000) Src family tyrosine kinases and growth factor signaling. *Exp Cell Res* 254: 1–13.
5. Cole PA, Shen K, Qiao Y, Wang D (2003) Protein tyrosine kinases Src and Csk: a tail’s tale. *Curr Opin Chem Biol* 7: 580–585.
6. Brauningner A, Holtrich U, Strebhardt K, Rubsamene-Waigmann H (1992) Isolation and characterization of a human gene that encodes a new subclass of protein tyrosine kinases. *Gene* 110: 205–211.
7. Ingleby E (2008) Src family kinases: regulation of their activities, levels and identification of new pathways. *Biochim Biophys Acta* 1784: 56–65.
8. Brown MC, Turner CE (2004) Paxillin: adapting to change. *Physiol Rev* 84: 1315–1339.
9. Tobe K, Sabe H, Yamamoto T, Yamauchi T, Asai S, et al. (1996) Csk enhances insulin-stimulated dephosphorylation of focal adhesion proteins. *Mol Cell Biol* 16: 4765–4772.
10. Tsutsumi R, Higashi H, Higuchi M, Okada M, Hatakeyama M (2003) Attenuation of *Helicobacter pylori* CagA x SHP-2 signaling by interaction between CagA and C-terminal Src kinase. *J Biol Chem* 278: 3664–3670.
11. Kawabuchi M, Satomi Y, Takao T, Shimonishi Y, Nada S, et al. (2000) Transmembrane phosphoprotein Cbp regulates the activities of Src-family tyrosine kinases. *Nature* 404: 999–1003.
12. Ghose R, Shekhtman A, Goger MJ, Ji H, Cowburn D (2001) A novel, specific interaction involving the Csk SH3 domain and its natural ligand. *Nat Struct Biol* 8: 998–1004.
13. Yaqub S, Abrahamsen H, Zimmerman B, Kholod N, Torgersen KM, et al. (2003) Activation of C-terminal Src kinase (Csk) by phosphorylation at serine-364 depends on the Csk-Src homology 3 domain. *Biochem J* 372: 271–278.
14. Lieser SA, Shaffer J, Adams JA (2006) SRC tail phosphorylation is limited by structural changes in the regulatory tyrosine kinase Csk. *J Biol Chem* 281: 38004–38012.
15. Shaffer J, Sun G, Adams JA (2001) Nucleotide release and associated conformational changes regulate function in the COOH-terminal Src kinase, Csk. *Biochemistry* 40: 11149–11155.
16. Hamuro Y, Wong L, Shaffer J, Kim JS, Stranz DD, et al. (2002) Phosphorylation driven motions in the COOH-terminal Src kinase, CSK, revealed through enhanced hydrogen-deuterium exchange and mass spectrometry (DXMS). *J Mol Biol* 323: 871–881.
17. Wong L, Lieser S, Chie-Leon B, Miyashita O, Aulob B, et al. (2004) Dynamic coupling between the SH2 domain and active site of the COOH terminal Src kinase, Csk. *J Mol Biol* 341: 93–106.

Similar to previous studies [57,58,59], we compare the candidate conformations from the simulation to the SAXS data by generating theoretical SAXS profiles for each conformation. Theoretical scattering curves were generated for all candidate conformations and ensembles using the CRYSOLO software package [60]. Default parameters were used with the following exceptions: maximum order of harmonics = 50, order of Fibonacci grid = 18, number of points = 100, and maximum s-value = 0.18. Here, we measure the goodness of fit of each structure by calculating  $\chi^2$ , which is defined as,

$$\chi^2 = \frac{1}{N_q} \sum_{iq} \left( \frac{(I^{\text{exp}}(q) - I^{\text{simul}}(q))}{\sigma(q)} \right)^2$$

where  $N_q$  is the number of data points in the scattering curve,  $I(q)$  is the SAXS intensity and  $\sigma(q)$  is the experimental error of  $I^{\text{exp}}(q)$ .

## Ensemble Analysis of the Candidate Conformations

Theoretical SAXS profiles were generated for linear combinations of conformations,

$$I_{\text{combined}}(q) = \sum w_i I_i(q)$$

where  $I_i(q)$  is the intensity profile of simulated conformation  $i$  and  $w_i$  is the weight of conformation  $i$ .  $I_{\text{combined}}(q)$  was determined for all combinations of  $w_i = 0.1n$ , where  $n$  is an integer of value 0 to 10.  $\chi^2$  was then calculated between each  $I_{\text{combined}}(q)$  and the experimental profile to determine the set of  $w_i$  values that minimize  $\chi^2$ .

## Acknowledgments

X-ray scattering data were collected at the Stanford Synchrotron Radiation Laboratory, a national user facility operated by Stanford University on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences. This work is dedicated to the memory of Hiro Tsuruta. The authors would like to acknowledge Bob Wong and Christian Cole for aid in sample preparation and Elizabeth Baxter for insightful discussions.

## Author Contributions

Conceived and designed the experiments: MAJ LCO PCW JNO PAJ. Performed the experiments: MAJ LCO PCW. Analyzed the data: MAJ PCO PCW PAJ. Contributed reagents/materials/analysis tools: PCW JNO PAJ. Wrote the paper: MAJ LCO PCW PAJ JAA.

18. Wong L, Lieser SA, Miyashita O, Miller M, Tasken K, et al. (2005) Coupled motions in the SH2 and kinase domains of Csk control Src phosphorylation. *J Mol Biol* 351: 131–143.
19. Rozycki B, Kim YC, Hummer G (2011) SAXS ensemble refinement of ESCRT-III CHMP3 conformational transitions. *Structure* 19: 109–116.
20. Yang S, Blachowicz L, Makowski L, Roux B (2010) Multidomain assembled states of Hck tyrosine kinase in solution. *Proc Natl Acad Sci U S A* 107: 15757–15762.
21. Kim SJ, Dumont C, Gruebele M (2008) Simulation-based fitting of protein-protein interaction potentials to SAXS experiments. *Biophys J* 94: 4924–4931.
22. Pelikan M, Hura GL, Hammel M (2009) Structure and flexibility within proteins as identified through small angle X-ray scattering. *Gen Physiol Biophys* 28: 174–189.
23. Marquez JA, Smith CI, Petoukhov MV, Lo Surdo P, Mattsson PT, et al. (2003) Conformation of full-length Bruton tyrosine kinase (Btk) from synchrotron X-ray solution scattering. *EMBO J* 22: 4616–4624.
24. Johs A, Harwood IM, Parks JM, Nauss RE, Smith JC, et al. (2011) Structural characterization of intramolecular Hg(2+) transfer between flexibly linked domains of mercuric ion reductase. *J Mol Biol* 413: 639–656.
25. Krukenberg KA, Forster F, Rice LM, Sali A, Agard DA (2008) Multiple conformations of *E. coli* Hsp90 in solution: insights into the conformational dynamics of Hsp90. *Structure* 16: 755–765.
26. Lingel A, Weiss TM, Niebuhr M, Pan B, Appleton BA, et al. (2009) Structure of IL-33 and its interaction with the ST2 and IL-1RAcP receptors—insight into heterotrimeric IL-1 signaling complexes. *Structure* 17: 1398–1410.
27. Jamros MA, Oliveira LC, Whitford PC, Onuchic JN, Adams JA, et al. (2010) Proteins at work: a combined small angle X-RAY scattering and theoretical determination of the multiple structures involved on the protein kinase functional landscape. *J Biol Chem* 285: 36121–36128.
28. Glatter O, Kratky O (1982) *Small Angle X-ray Scattering*: Academic Press. 515 pp.
29. Ogawa A, Takayama Y, Sakai H, Chong KT, Takeuchi S, et al. (2002) Structure of the carboxyl-terminal Src kinase, Csk. *J Biol Chem* 277: 14351–14354.
30. Masterson LR, Shi L, Metcalfe E, Gao J, Taylor SS, et al. (2011) Dynamically committed, uncommitted, and quenched states encoded in protein kinase A revealed by NMR spectroscopy. *Proc Natl Acad Sci U S A* 108: 6969–6974.
31. Niu X, Bruschiweiler-Li L, Davulcu O, Skalicky JJ, Bruschiweiler R, et al. (2011) Arginine kinase: joint crystallographic and NMR RDC analyses link substrate-associated motions to intrinsic flexibility. *J Mol Biol* 405: 479–496.
32. Henzler-Wildman KA, Lei M, Thai V, Kerns SJ, Karplus M, et al. (2007) A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* 450: 913–916.
33. Bhattacharya A, Kurochkin AV, Yip GN, Zhang Y, Bertelsen EB, et al. (2009) Allosteric in Hsp70 chaperones is transduced by subdomain rotations. *J Mol Biol* 388: 475–490.
34. Mazhab-Jafari MT, Das R, Fotheringham SA, SilDas S, Chowdhury S, et al. (2007) Understanding cAMP-dependent allostery by NMR spectroscopy: comparative analysis of the EPAC1 cAMP-binding domain in its apo and cAMP-bound states. *J Am Chem Soc* 129: 14482–14492.
35. Sarkar P, Saleh T, Tzeng SR, Birge RB, Kalodimos CG (2011) Structural basis for regulation of the Crk signaling protein by a proline switch. *Nat Chem Biol* 7: 51–57.
36. Shekhtman A, Ghose R, Wang D, Cole PA, Cowburn D (2001) Novel mechanism of regulation of the non-receptor protein tyrosine kinase Csk: insights from NMR mapping studies and site-directed mutagenesis. *J Mol Biol* 314: 129–138.
37. Shan Y, Seeliger MA, Eastwood MP, Frank F, Xu H, et al. (2009) A conserved protonation-dependent switch controls drug binding in the Abl kinase. *Proc Natl Acad Sci U S A* 106: 139–144.
38. Chong YP, Mulhern TD, Cheng HC (2005) C-terminal Src kinase (CSK) and CSK-homologous kinase (CHK)—endogenous negative regulators of Src-family protein kinases. *Growth Factors* 23: 233–244.
39. Mustelin T, Tasken K (2003) Positive and negative regulation of T-cell activation through kinases and phosphatases. *Biochem J* 371: 15–27.
40. Shoemaker BA, Portman JJ, Wolynes PG (2000) Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proc Natl Acad Sci U S A* 97: 8868–8873.
41. Bougeret C, Rothhut B, Jullien P, Fischer S, Benarous R (1993) Recombinant Csk expressed in *Escherichia coli* is autophosphorylated on tyrosine residue(s). *Oncogene* 8: 1241–1247.
42. Lin X, Lee S, Sun G (2003) Functions of the activation loop in Csk protein-tyrosine kinase. *J Biol Chem* 278: 24072–24077.
43. Jacques DA, Trehwella J (2010) Small-angle scattering for structural biology—expanding the frontier while avoiding the pitfalls. *Protein Sci* 19: 642–657.
44. Whitford PC, Noel JK, Gosavi S, Schug A, Sanbonmatsu KY, et al. (2009) An all-atom structure-based potential for proteins: bridging minimal models with all-atom empirical forcefields. *Proteins* 75: 430–441.
45. Whitford PC, Sanbonmatsu KY, Onuchic JN (2012) Biomolecular dynamics: order-disorder transitions and energy landscapes. *Rep Prog Phys* 75: 076601.
46. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG (1995) Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* 21: 167–195.
47. Onuchic JN, Wolynes PG (2004) Theory of protein folding. *Curr Opin Struct Biol* 14: 70–75.
48. Chavez LL, Onuchic JN, Clementi C (2004) Quantifying the roughness on the free energy landscape: entropic bottlenecks and protein folding rates. *J Am Chem Soc* 126: 8426–8432.
49. Clementi C, Nymeyer H, Onuchic JN (2000) Topological and energetic factors: what determines the structural details of the transition state ensemble and “enroute” intermediates for protein folding? An investigation for small globular proteins. *J Mol Biol* 298: 937–953.
50. Lammert H, Schug A, Onuchic JN (2009) Robustness and generalization of structure-based models for protein folding and function. *Proteins* 77: 881–891.
51. Oliveira LC, Schug A, Onuchic JN (2008) Geometrical features of the protein folding mechanism are a robust property of the energy landscape: a detailed investigation of several reduced models. *J Phys Chem B* 112: 6131–6136.
52. Whitford PC, Geggier P, Altman RB, Blanchard SC, Onuchic JN, et al. (2010) Accommodation of aminoacyl-tRNA into the ribosome involves reversible excursions along multiple pathways. *RNA* 16: 1196–1204.
53. Noel JK, Whitford PC, Onuchic JN (2012) The Shadow Map: A General Contact Definition for Capturing the Dynamics of Biomolecular Folding and Function. *J Phys Chem B* 116: 8692–702.
54. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, et al. (2005) GROMACS: fast, flexible, and free. *J Comput Chem* 26: 1701–1718.
55. Noel JK, Whitford PC, Sanbonmatsu KY, Onuchic JN (2010) SMOG@ctbp: simplified deployment of structure-based models in GROMACS. *Nucleic Acids Res* 38 Suppl: W657–661.
56. Kouza M, Li MS, O'Brien EP, Hu C-K, Thirumalai D (2005) Effect of Finite Size on Cooperativity and Rates of Protein Folding. *J Phys Chem A* 110: 671–676.
57. Yang S, Park S, Makowski L, Roux B (2009) A rapid coarse residue-based computational method for x-ray solution scattering characterization of protein folds and multiple conformational states of large protein complexes. *Biophys J* 96: 4449–4463.
58. Forster F, Webb B, Krukenberg KA, Tsuruta H, Agard DA, et al. (2008) Integration of small-angle X-ray scattering data into structural modeling of proteins and their assemblies. *J Mol Biol* 382: 1089–1106.
59. Svergun DI (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J* 76: 2879–2886.
60. Svergun D, Barberato C, Koch MHJ (1995) CRYSOLE - A Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J Appl Crystallogr* 28: 768–773.