# BlockMaster: Partitioning Protein Kinase Structures Using Normal-Mode Analysis†

**Marina Shudler and Masha Y. Niv***

*The Institute of Biochemistry, Food Science and Nutrition, The Hebrew University of Jerusalem, Rehovot 76100, Israel*

*Received: January 30, 2009; Revised Manuscript Received: April 26, 2009*

Protein kinases are key signaling enzymes which are dysregulated in many health disorders and therefore represent major targets of extensive drug discovery efforts. Their regulation in the cell is exerted via various mechanisms, including control of the 3D conformation of their catalytic domains. We developed a procedure, BlockMaster, for partitioning protein structures into semirigid blocks and flexible regions based on residue−residue correlations calculated from normal modes. BlockMaster provided correct partitioning into domains and subdomains of several test set proteins for which documented expert annotation of subdomains exists. When applied to representative structures of protein kinases, BlockMaster identified semirigid blocks within the traditional N-terminal and C-terminal lobes of the kinase domain. In general, the block regions had elevated helical content and reduced, but significant, coil content compared to the nonblock (flexible) regions. The specificity-determining regions, previously used to derive inhibitory peptides, were found to be more flexible in the tyrosine kinases than in serine/threonine kinases. Two blocks were identified which spanned both lobes. The first, which we termed the "pivot" block, included the αC−β4 loop in the N-terminal lobe and part of the activation loop in the C-terminal lobe and appeared in both the active and inactive conformations of the kinases. The second, which we termed the "loop" block, differed between the active and inactive conformations. In the structures of active kinases, this block included part of the activation loop in the C-terminal lobe and the αC helix in the N-terminal lobe, representing a known interaction that stabilizes the active conformation. In the inactive structures, this block included G loop residues instead of the αC residues. This novel inactive "loop" block may stabilize the inactive conformation and thus downregulate kinase activity.

## Introduction

**Protein Kinases and Their Activation Dynamics.** Protein kinases use ATP to phosphorylate protein substrates and release ADP as a byproduct. These enzymes are involved in a myriad of signaling pathways, and the regulation of their catalytic activity is crucial for normal functioning of the cell. They are dysregulated in many health disorders and represent the biggest family of drug targets alongside G-protein-coupled receptors.[1] The introduction of the BCR-Abl kinase inhibitor imatinib mesylate (Gleevec, Novartis) revolutionized the treatment of chronic myeloid leukemia.[2] Currently, 11 kinase inhibitors are FDA-approved for the treatment of cancer, and hundreds of kinase inhibitors are in different stages of development.[3]

Because of the fundamental importance of protein kinases in cell signaling, it is not surprising that organisms have developed multiple layers of kinase activity regulation. These control mechanisms include, among others, intrinsic propensity of the catalytic domain to assume an active conformation,[4,5] regulation via phosphorylation,[6,7] allosteric activation by other proteins,[8] and regulation through specificity of interactions with the substrate.[9] Important open issues that are pertinent to kinase activation dynamics include elucidating the dynamic processes that are involved in activation[10−12] and trapping the different intermediate conformations assumed by the kinase catalytic domains en route to activation.[5,13,14]

**Studying Conformational Dynamics.** Multiple novel techniques are emerging to address the formidable challenge of investigating protein conformational dynamics at the molecular level.[15] While experimental advances (including cryo-EM, small-angle X-ray scattering, and fluorescence spectroscopy) can expand our knowledge of protein dynamics, they lack spatial resolution of the structures. Such resolution can be provided by molecular dynamics (MD) simulations or by a combination of MD with experimental techniques.[16−18] However, conventional MD simulations are too short to access physiologically relevant time scales. Novel approaches are being developed to overcome this difficulty, including enhanced sampling techniques such as replica exchange, potential-smoothing protocols and reaction path methods.[16,19,20] An important category of enhanced MD simulations relies on spatial coarse-graining, in which residues or small groups of atoms are treated as single particles.[17]

One successful and computationally simple approach to coarse-graining approximations is the normal-mode analysis (NMA), in which molecular motions are decomposed into orthogonal vectors representing vibrational motions (normal modes). In the elastic network model (ENM) version of the NMA, a harmonic potential with a single force constant accounts for pairwise interactions between all Cα atoms that are within a certain cutoff distance.[21] The ENM has been shown to capture key features of many biological systems as compared to all-atom potential NMA and to experimental data, such as crystallographic B-factors[21] and H/D exchange rates;[22] it has also been shown to agree well with atomistic simulation methods.[23,24] However, despite the immense usefulness and success of NMA, there are intrinsic limitations in the harmonic approximation of the equilibrium conformation[25] as it is not adequate for modeling nonequilibrium dynamics. Indeed, in both equilibrium and

† Part of the "Robert Benny Gerber Festschrift".

* To whom correspondence should be addressed. E-mail: niv@agri.huji.ac.il.

10.1021/jp900885w CCC: $40.75 
Published on Web 06/01/2009

nonequilibrium states, many motions are of large amplitude and highly anharmonic; even along the harmonic modes, the amplitudes of protein motions remain unknown. Finally, and perhaps most importantly, there is no rigorous algorithm for the identification of the biologically relevant modes of motion among the low-frequency normal modes and no time resolution of the dynamic processes.

**"Divide and Conquer" Approaches to Protein Dynamics.** Among the approaches that can overcome these drawbacks are a Newtonian propagation rotation−translation block molecular dynamics (RTB-MD),[26,27] and the multibody order (N) dynamics [MBO(N)D] scheme.[28] These methods produce approximate Newtonian dynamics of coupled rigid bodies (blocks) which form a supersystem of interest (e.g., a large protein molecule). A key step in block dynamics is the identification of rigid or semirigid blocks. Several methods exist for the identification of protein domains (i.e., independently folding units)[29,30] and motion hinges.[31,32] Fewer methods are aimed at identifying rigid substructures or subregions within a domain. Among these methods is the FIRST (floppy inclusion and rigid substructure topography) algorithm, which uses graph theory to identify rigid substructures,[33] the TLSMD (translation/libration/screw motion description) algorithm, which analyzes the distribution of atomic displacement parameters in a crystal structure to generate an optimal description of the protein chain as a succession of residue groups,[34] and methods like RAPIDO (rapid alignment of proteins in terms of domains),[35] which use structural alignment of two structures to identify groups of equivalent atoms whose interatomic distances are constant.

In the current work, we hypothesized that structural regions that remain semirigid during the functional dynamics of a protein are comprised of residues that have highly correlated motion (in other words, move together) in the lowest-frequency normal modes. Therefore, our criterion for inclusion of a residue into a rigid block was high motional correlation with the rest of the residues in the same block. The correlation was calculated from the normal modes as described in the Methodology Section. The use of normal-mode-based motional correlation between residues as a basis for partitioning proteins into substructures has been employed by Yesylevskyy and co-workers in several recent works.[36−39] Our procedure, which we term "BlockMaster", differs from this approach mainly in that the residues can remain unassigned to any block and therefore can be rendered completely flexible in subsequent treatment of dynamics. We describe the BlockMaster procedure and show that it is successful in partitioning structures of proteins for which decomposition into domains or subdomains has been studied in the literature. We then use BlockMaster to partition the catalytic domains of representative protein kinases. The accepted annotation of the protein kinase catalytic domain partitions it into an N-terminal lobe (N-lobe) and C-terminal lobe (C-lobe).[40] Neither of these lobes is rigid, and the details of the intradomain structure are the subject of the current work. We chose four representative kinases and analyzed the active and inactive conformations for each of them. The results of our analysis are presented below.

## Methodology Section

**Calculation of the Correlation Matrix.** The correlations between motions of atoms were calculated using NMA. As previously,[41,42] NMA was performed using NOMAD-Ref.[43] Correlation between residues $i$ and $j$ was calculated as the average correlation summed over lowest-frequency nontrivial normal modes. NOMAD-Ref uses Tirion's ENM,[21] in which the standard potential energy function is replaced by the following expression

$$E_{\mathrm{p}} = \sum_{r_{ij}^0 < R_{\mathrm{C}}} f(r_{ij} - r_{ij}^0)^2 \tag{1}$$

where $r_{ij}$ denotes the vector connecting Cα atoms $i$ and $j$, the 0 superscript indicates the initial configuration of the Cα atoms, $R_{\mathrm{C}}$ is the spatial cutoff for interconnections between Cα atoms, set to 10 Å (except for test case #8, in which a cutoff of 15 Å was used), and $f$ is the phenomenological force constant arbitrarily set to 100 kcal/(mol·Å$^2$). The normalized eigenvectors $U_i$ of the Hessian matrix due to eq 1 are the normal modes, and the eigenvalues $\Omega_i$ are the squares of the respective frequencies.

$$C_{ij} = \frac{\sum_{l=7}^{106} \frac{U_{il}U_{jl}}{\Omega_l}}{\left(\sum_{m=7}^{106} \frac{U_{im}U_{im}}{\Omega_m}\right)^{1/2}\left(\sum_{n=7}^{106} \frac{U_{jn}U_{jn}}{\Omega_n}\right)^{1/2}} \tag{2}$$

Correlation between residues $i$ and $j$ (eq 2), as was defined by ref 44 and used, for example, in refs 41, 44, and 45 was calculated (in MATLAB) as the average correlation summed over 100 lowest-frequency nontrivial (7−106) normal modes. That was enough for convergence of the correlation matrix (results not shown).

**Identification of Rigid Blocks: BlockMaster (implemented in Java).** ***Preprocessing.*** A graph network was constructed in which the Cα atoms are the nodes and the Euclidean distance between each pair of nodes is the "length" of the edge. The edges exist only between nodes with distances smaller or equal to 10 Å (default distance cutoff used in the ENM calculation in NOMAD-Ref); these nodes are termed neighbors. From the correlation matrix calculated using eq 2, a distCorrelation matrix is computed as follows: $\mathrm{distCorrelation}_{ij} = 1 - C_{ij}$. The value of each element in the distCorrelation matrix varies between 0 (best correlation) and 2 (worst correlation).

***Block Creation.*** The pairs of neighbor nodes which are not yet assigned to a block are ordered by their distCorrelation value. For each new block $i$, the lowest distCorrelation pair initiates the block if the distCorrelation is equal to or lower than a predefined distCorrelation threshold. A node is then added to block $i$ if the following conditions are fulfilled: (a) the node has not yet been assigned to any block; (b) it is a neighbor of one of the nodes in block $i$; (c) the node has the lowest distCorrelation sum value to all of the nodes in block $i$; and (d) the distCorrelation value to each node in block $i$ is not higher than a predefined distCorrelation threshold. This loop is continued as long as appropriate candidates exist. After that, the procedure is repeated to initiate a new block $i + 1$. If the newly created block is smaller than the predefined minimum block size (which in our simulations was 4), the block is dismissed, and its nodes remain unassigned to a block. We tested the performance of the method using predefined distCorrelation threshold values ranging from 0.04 to 2.0 using eight established test cases which are detailed below. The agreement with the partition established in the literature was calculated at different distCorrelation threshold values, as detailed in Supporting Information and shown in Supporting Figure 1. The 0.4 value was chosen for use in the analysis of protein kinases because it was the most stringent threshold among those that gave satisfactory agreement with the literature for the test cases. (The maximal average agreement was obtained for a threshold value of 0.7, and some of the calculations were repeated at this value and confirmed the robustness of our conclusions).

**TABLE 1: Test Case Structures**

| # | protein name | PDBs (chains) used | ref |
|---|---|---|---|
| 1 | lysine-, arginine-, ornithine-binding protein | 1LST (A), 2LAO (A) | 36 |
| 2 | glutamine-binding protein (GLNBP) | 1WDN (A), 1GGG (A) | 36 |
| 3 | phosphoglycerate kinase | 13PK (B), 16PK (A) | 36 |
| 4 | calmodulin | 1CLL (A), 1CDL (A) | 36 |
| 5 | 5-enolpyruvyl-shikimate-3-phosphate synthase | 1RF5 (A), 1RF6 (A) | 36 |
| 6 | GroEL | 1OEL (A−G), 1WE3 (A−N) | 50 |
| 7 | adenylate kinase (ADK) | 1ANK (A−B), 4AKE (A−B) | 51 |
| 8 | citrate synthase | 1CTS (A), 2CTS (A) | 51 |

The subsequent analysis of BlockMaster output and the images in Supporting Information Figures S2−S13 were obtained using MATLAB; Figure S1 (Supporting Information) was prepared in EXCEL. Figures 1, 3 and 4 were prepared using Accelrys Software Inc. (http://accelrys.com/).

***Preparation of the Structures for Analysis.*** The structures were downloaded from the Protein Data Bank.[46] Missing loops were modeled via the ModLoop server.[47] Cα atom coordinates of each structure were used for subsequent analysis.

***Calculation of Physicochemical and Secondary-Structure Statistics.*** Secondary-structure analysis was performed using the DSSP program[48] for each structure file. The information from DSSP was parsed and combined with the appropriate protein sequence. The different DSSP assignments (described in http://swift.cmbi.ru.nl/gv/dssp/) were combined as follows: H, G, I = helix; E = strand; T, S, B, unassigned = coil. The number of residues in secondary structures in blocks and in flexible regions was calculated for each chain in each PDB file, and the average and standard deviation were calculated.

Amino-acid-type information for physicochemical analysis was extracted from each studied protein, for the block and for the flexible regions. The amino acids were divided into the following groups: aliphatic [AVLIMC], aromatic [HWYF], polar [NQST], negatively charged [ED], positively charged [KR], and special conformation [GP], as in our previous work.[49]

## Results

**BlockMaster Partitions Protein Structures in Agreement with Expert Opinion.** The idea behind the BlockMaster structure partitioning procedure is that residues that move together (have correlated motion) in the lowest-frequency normal modes can be treated as a semirigid block. It is well-known that some regions in the protein can be highly flexible, and we therefore do not expect each residue to necessarily participate in a block, as it may remain a flexible independent unit (or block of size 1). In order to assign residues to semirigid blocks, we have calculated normal-mode-derived correlations between residues and used a greedy algorithm to build the blocks, as described in the Methods Section. We chose two groups of proteins to assess the applicability of BlockMaster to the problem of identifying semirigid substructures in proteins. The first group consisted of two-domain proteins, in which each domain is an independent folding unit (ref 36; entries 1−5 in Table 1). The partitioning into domains was obtained by RAPIDO.[35] The second group (entries 6−8 in Table 1), more relevant to our task of partitioning the kinase catalytic domain, included single-domain proteins which had been divided into subdomains by experts.

We compared the distribution of blocks obtained by BlockMaster to the accepted partitioning. To visualize the results, all of the residues in the block belonging to one domain (subdomain) were colored in that domain's (subdomain's) color. If more than one domain was represented in a block, the color was assigned according to the domain to which the majority of residues belong. If the domains were represented equally in one block, residues of this block were colored yellow. Residues unassigned to any block were colored gray. BlockMaster results for all of the test cases compared well with the literature and are detailed in Supporting Information. Figure 1 shows the agreement for two examples, GroEL[50] (case 6 in Table 1) and adenylate kinase[51] (case 7 in Table 1).

Our analysis of the test cases suggested that BlockMaster partitions proteins into blocks that fall within the established domains and/or subdomains and can therefore be used for analysis of structures for which partitioning has not been established.

**BlockMaster Partitioning of Protein Kinases.** Protein kinases can phosphorylate serine and threonine residues (S/T kinases) or tyrosine residues (Tyr kinases).[52] These enzymes act as molecular switches that can adopt at least two extreme conformations, an "on" state that is maximally active and an "off" state that has minimum activity. Thus, upon activation, they adopt catalytically active "on" conformations that are structurally very similar. The "off" states of protein kinases are not subject to the active states' chemical constraints, and different classes of kinases have developed distinct "off" states, from which the adoption of the catalytically active conformation

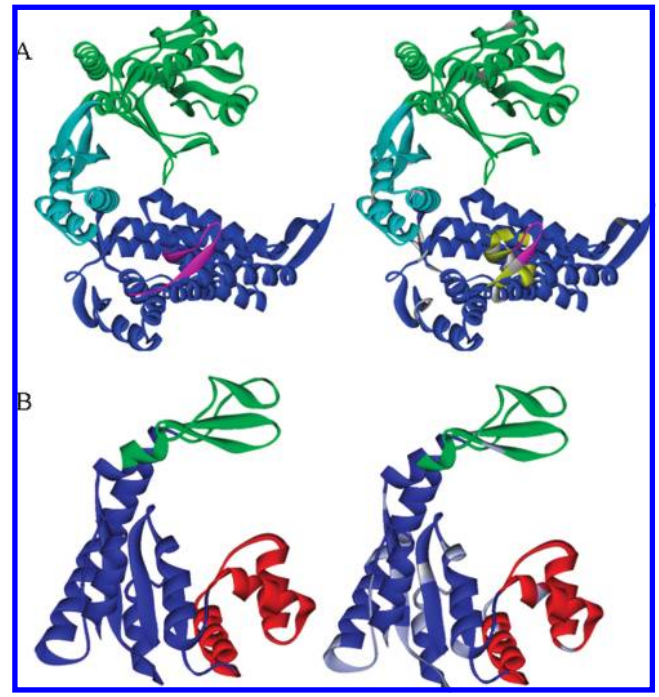


**Figure 1.** Illustration of BlockMaster validation. (A) Partitioning of GroEL: left panel, from the literature;[50] right panel, by BlockMaster. (B) Partitioning of adenylate kinase: left panel, from the literature;[51] right panel, by BlockMaster.

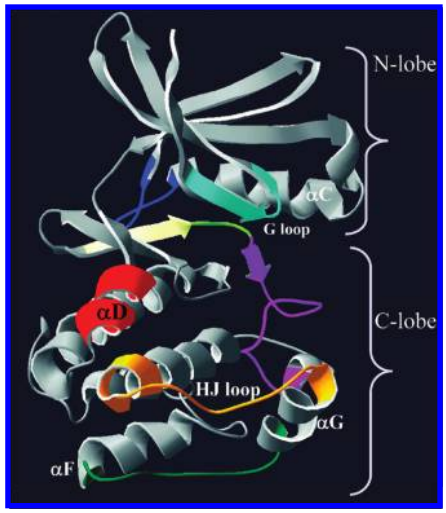


**Figure 2.** Protein kinase catalytic domain structure. The N-lobe and C-lobe are shown. The αD helix is colored red, and the HJ loop[59] is colored orange. The loop connecting αG and αF helices is colored dark green. The region preceding the DFG motif is colored yellow, the motif is colored light green, and the rest of the activation loop is colored magenta. The G loop is colored cyan, and the αC−β4 loop is colored blue. The regions are mapped onto the 1ATP.pdb structure of protein kinase A (PKA), and the figure was prepared with SwissPDBViewer, http://www.expasy.org/spdbv/.

can be impeded in different ways. Protein kinases consist of a smaller N-terminal lobe (N-lobe) and a larger C-terminal lobe (C-lobe), with the ATP-binding site located between them. The N-lobe is composed of five β-strands and a helix referred to as αC, whereas the C-lobe is predominantly helical (see Figure 2). The N-lobe harbors the phosphate-binding loop, called the G loop, which contains a conserved glycine-rich sequence motif (GXGXφG). A centrally located loop in the C-lobe is known as the "activation loop," typically 20 to 30 residues in length. This region provides a platform for the peptide substrate. In most kinases, this loop is phosphorylated when the kinase is active. Phosphorylation of the activation loop stabilizes it in an open and extended conformation that is permissive for substrate binding.[40]

In the following analysis, our goal was to identify semirigid blocks in protein kinases and to compare them in the active and inactive conformations, as well as to compare the blocks appearing in the Tyr kinases versus those in the S/T kinases.

We chose two representative Tyr kinases, Abl,[53] the target for the well-established antioncogenic drug imatinib, and insulin receptor kinase (IRK),[54] and two representative S/T kinases, Akt/PKB,[55] a novel drug target in cancer, and CDK2,[56] part of the family of key regulators of the cell cycle. The structures used in the analysis are summarized in Table 2.

To explore whether existing methods clearly divide protein kinases into subregions, we have used several methods. In particular, we have partitioned protein kinase catalytic domains with RAPIDO, which requires two structures of one chain or one structure of multiple chains as an input. The partitioning critically depended on the input structures. For example, analysis of chain A and C of CDK2 1FIN active structures gave a single rigid block (no partition at all), while alignment of 1FIN chain A and 1JST chain A resulted in two rigid blocks and a very small flexible block, and alignment of 1FIN chain A and 1HCL chain A generated two other rigid regions and three flexible regions. RAPIDO analysis of PKB structures 1O6K chain A and 1O6L chain A generated one rigid region, while 1O6K chain A and 1MRY chain A generated five rigid regions and four flexible ones. The same situation was seen in other kinase catalytic domains.

HingeProt[32] (http://bioinfo3d.cs.tau.ac.il/HingeProt/) using slowest normal mode 1 partitioned kinase structures CDK2 (1FIN, 1QMZ, 1JST chain A), IRK (1IR3, 1IRK), and ABL (1OPJ, 2G2F, 2G2H, 2G2I) into two rigid parts, which agree (with a deviation of 3−4 residues) with the accepted annotation of partitioning protein kinase catalytic domain to the N-lobe and C-lobe. The rest of the CDK2 structures (1HCL, 1HCK, 1JST chain C) and the PKB structures (1MRY, 1O6K, 1O6L) are partitioned differently. In slowest mode 2, each protein was partitioned differently.

Another method that we examined for partitioning of kinases is TLSMD.[34] This method uses experimental information on isotropic and anisotropic B-factors to partition individual protein structures into sequential blocks. Regions that are close in structure but nonconsecutive in sequence cannot be assigned to the same block, and no differentiation between flexible and rigid substructures is obtained. The results obtained for kinases showed different partitions for different structures, manifested in a different number of rigid parts, their size, and rigid region boundaries.

Thus, there is no trivial partition of protein kinase catalytic domains into substructures, except for the general division into N-terminal and C-terminal lobes (see Figure 2). Since the results obtained with BlockMaster on the test case proteins 1−8 were reliable, we proceeded to analyze representative protein kinase structures using this new procedure.

**Semirigid Blocks and Flexible Regions in Representative Protein Kinase Structures.** The protein kinases broke up into 32−45 blocks, most of them comprising 5−8 amino acids each. In total, about 70% of the residues in kinase structures were found in blocks. The percentage of aliphatic residues in blocks (37 ± 2%; the statistics comes from analyzing all of the chains listed in Table 2) was higher than that in flexible regions (26 ± 5%). There was a higher percentage of special-conformation residues (G,P) in the flexible regions (15 ± 3%) than that in blocks (9 ± 2%). We also found that 25 ± 4% of the block residues reside in coils versus 57 ± 13% of the flexible residues. In contrast, the helical content of blocks was 57 ± 3%, versus 29 ± 10% in flexible regions (see also Supporting Information). These trends were in accordance with the expectation that the semirigid regions would correspond to the hydrophobic core and have more pronounced secondary structure, while the flexible regions would have residues that ensure conformational flexibility and less secondary structure. Our results are in agreement with Flores and co-workers,[57] who surveyed a large data set of residues residing in hinges (a subgroup of what we define as flexible regions). According to Flores et al.[57] (page 18), amino acids glycine and serine are more likely to occur in hinges, whereas phenylalanine, alanine, valine, and leucine are less likely to occur. We performed additional calculation with these groups of amino acids. Indeed, the percentage of AVLF residues is bigger in blocks than in flexible regions, and the percentage of GS is smaller in blocks than in the flexible regions. Flores et al. also find (page 9; Figure 3) that residues in alpha helices are less likely to occur in hinges, and turn and coil residues are more likely to be in hinges.

Notably, our results were obtained without explicitly representing the physicochemical properties of the residues or hydrophobic, electrostatic, and other terms of the potential since a simplified Hookean potential was used for all residues. The fact that the partition of the residues into the flexible and the

**TABLE 2: Protein Kinase Structures Analyzed in This Work**

| number | kinase type | kinase name | PDB (chains) used | conformation |
|---|---|---|---|---|
| 1 | S/T kinase | CDK2 | 1FIN (A,B), 1JST (A,B), 1QMZ (A,B) | active |
| | | | 1HCL (A), 1HCK (A) | inactive |
| 2 | S/T kinase | Akt/PKB | 1O6K (A), 1O6L (A) | active |
| | | | 1MRY (A) | inactive |
| 3 | Tyr kinase | Abl | 2G2F (A,B), 2G2H (A,B), 2G2I (A,B) | active |
| | | | 1OPJ (A,B) | inactive |
| 4 | Tyr kinase | IRK | 1IR3 (A) | active |
| | | | 1IRK (A) | inactive |

rigid subgroups agrees with their properties stems from the fact that the residue's character dictates the folded structure, and this structure, in turn, determines the dynamics.[58] Thus, from analysis of the shape-determined motions of the biomolecules, one implicitly analyzes the underlying forces that have determined these shapes.

Many of the flexible residues were the same for the different protein kinases that we analyzed. The regions that were either flexible in all of the analyzed structures or participated in blocks in all of the analyzed structures are shown on a representative kinase in Figure 3.

Several conclusions could be drawn from inspection of these results. (I) the C-lobe has more consensus rigid blocks than the N-lobe. This is in agreement with the finding that the blocks have increased helical content since the C-lobe is predominantly helical. (II) The main difference between the S/T kinases and Tyr kinases analyzed in this work is the higher flexibility of the Tyr kinases in the substrate-binding region, namely, in the loop connecting helices G and F (colored dark green in Figure 2), in the HJ loop (colored orange in Figure 2), and at the C-terminal end of the αD region (colored red in Figure 2). The latter two regions were described in detail in our previous work and used as the basis for deriving kinase-specific inhibitory peptides.[59,60,61] We were previously unaware of flexibility differences in these regions in the S/T and Tyr kinases or upon activation. Interestingly, a concurrent computational study analyzed the dynamical changes induced by binding of ATP in a Go model of PKA, an S/T kinase. The HJ loop residues 235−240 in PKA had changed fluctuations upon ATP binding (corresponding to an activation process).[62] These differences in the substrate binding region flexibility require further analysis and are amenable to experimental test.[62] (III) The inactive structures have more flexible regions (colored gray and red in Figure 3) than the active ones. This is in accordance with the notion that the active conformation needs to be organized in the precise form that enables catalysis, while the inactive state is not under any particular evolutionary pressure for a specific conformation.[40] The main exception is the G loop, which was more rigid in at least some of the inactive structures. The G loop is discussed in the next section in more detail.

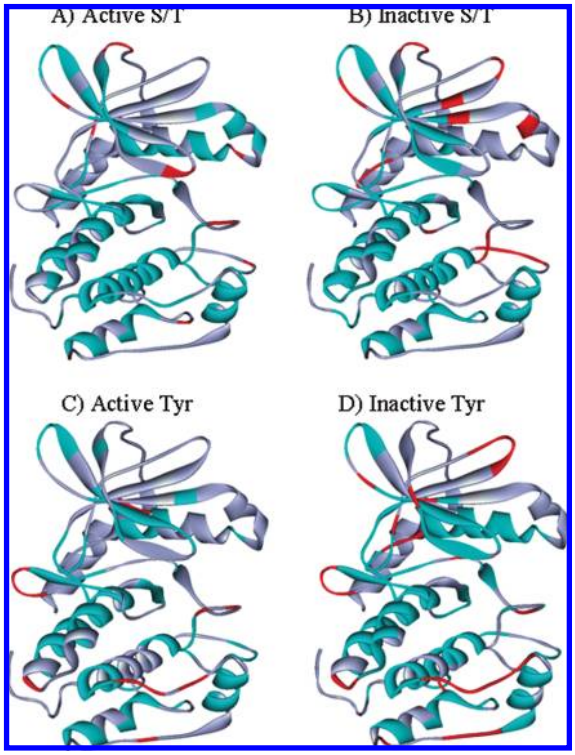


**Figure 3.** Calculated consensus flexibility projected onto the 1ATP (active PKA) structure. (A) Residues that were not assigned to a block in any of the active S/T kinases (see Table 2 for the list) are shown in red, residues assigned to a block in all of the S/T kinases are shown in cyan, and the rest are shown in gray. (B) Same color coding for residues in inactive S/T kinases. (C) Same color coding for residues in active Tyr kinases and (D) in inactive Tyr kinases.

**Blocks Shared by N-Lobe and C-Lobe.** As mentioned earlier, the protein kinase structure is traditionally divided into an N-lobe and a C-lobe (Figure 2). A change in interlobe orientation is an important hallmark of activation.[40] The lobes themselves are far from being rigid since the orientation of the αC helix within the N-lobe and that of the activation loop in the C-lobe are key aspects of kinase activation.[40,63] Furthermore, interactions between the lobes (in particular, between the αC helix and the activation loop) are important for maintaining the active conformation.[10,53] We found (see Supporting Information) that in all of the structures analyzed, the blocks reside on either the N-lobe or the C-lobe with two exceptions; these special blocks, which we termed the "pivot" and "loop" blocks, are described below.

***The Pivot Block.*** This block appeared in all of the structures that we analyzed except for the inactive Abl structure and involved residues from the αC−β4 loop in the N-lobe (residues 99−105 in the canonical PKA numbering, PDB code 1ATP.pdb) and residues preceding the DFG motif in the C-lobe (residues 180−183 in the canonical PKA numbering, light green in Figure 2). In addition, in some structures, C-lobe residue 153 and residues 171−174 (PKA numbering) also participated in this block (see Supporting Information). Structural changes originating from the activation segment are believed to be translated into distortions of the DFG motif that prop open the two lobes.[63] It is therefore interesting that the region preceding the DFG and sometimes even including D participated in a block with the N-lobe, perhaps providing a pivot point around which the structure opens. We expected that the importance of the pivot block would also be manifested in the sequences of the protein kinases. A sequence-based approach for identifying functional sites, the evolutionary trace (ET) method, aims to imitate experimental mutational analyses by using the sequence

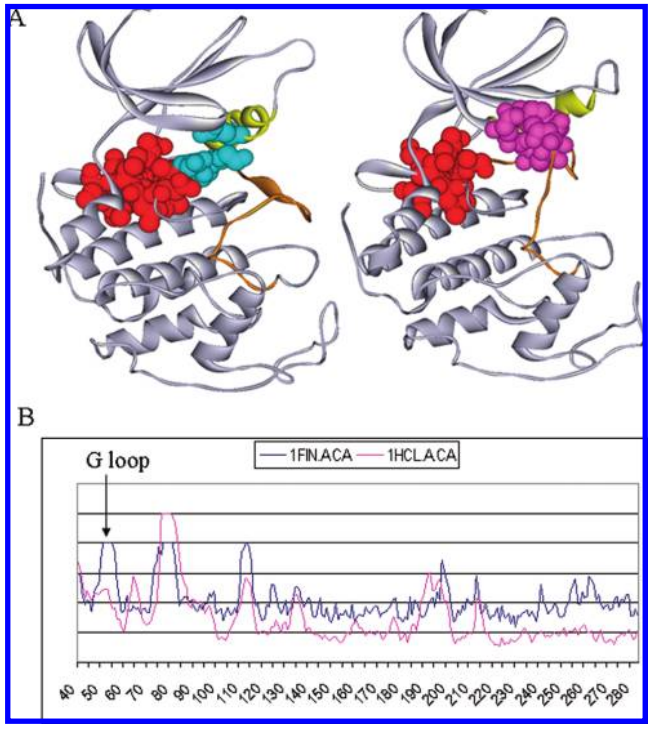


**Figure 4.** Special blocks. (A) Left panel: Active (1FIN) structure of CDK2, with the "pivot" block shown in red spheres and the active "loop" block in cyan spheres. The activation loop is shown in orange and the αC helix in yellow ribbon. Right panel: Inactive (1HCL) structure of CDK2. Colors as above, except for magenta, which represents here the inactive "loop" block. (B) Comparison of B-factors of the active (1FIN) and inactive (1HCL) CDK2 structures. The residue numbers in the canonical PKA numbering appear on the *x* axis.

variations and functional divergences found in nature. Interestingly, residues in the pivot block defined here were found to be functionally important in the ET analysis of protein kinase sequences (see Figure 4C in ref 64).

***The Loop Blocks.*** *The Active Conformation Loop Block.* Kornev and co-workers[10] compared active and inactive conformations of kinases and proposed a model in which a hydrophobic "spine", consisting of four residues ($L^{95}$, $L^{106}$, $F^{185}$, and $Y/H^{164}$ in PKA numbering), is dynamically assembled in active kinases and disassembled in inactive conformations. Because the "spine" is mediated through side chains, it could not be identified by using the current version of BlockMaster, which is based on Cα atoms only. (A full-atom version of BlockMaster is currently under development.) However, we did find that the DFG motif and following activation loop residues (184−188 in PKA numbering) in the C-lobe participate in a block with αC residues (88−96 in PKA numbering) in the N-lobe. Interestingly, this block appeared only in the active conformation of kinases CDK2, PKB, and IRK, in accordance with the established notion that interaction between the αC helix and the activation loop stabilizes the active conformation.[40,63]

*The Inactive Conformation Loop Block.* In the inactive conformations, the loop block described above is replaced by another block (Figure 4A). The activation loop residues (184−188 in PKA numbering) of CDK2, PKB, and IRK now participate in a block with the G loop residues (50 −55 in PKA numbering) from the N-lobe. Indeed, the G loop had a high B-factor in the active structures, which was significantly reduced in the inactive structures of PKB and CDK2 (Figure 4B).

We hypothesized that this loop block plays a role in stabilizing the inactive state. If this is the case, regulatory sites are likely to be located in this block. Indeed, in CDK2, phosphorylation of a tyrosine in the G loop causes flanging of the activation loop, as illustrated in structure 2CJM.pdb.[65] However, this is not enough to activate CDK2. In fact, the phosphomimicking mutation of the tyrosine or a neighboring threonine in the G loop result in inactive CDK2 (see references in ref 65), confirming the established notion that opening of the activation loop is necessary but not sufficient for kinase activity. The C-lobe part of the inactive conformation loop block involves another tyrosine (Y159), which is followed by threonine (T160). Phosphorylation of T160 is a hallmark of CDK2 activation. In the case of PKB, no phosphorylatable residues are involved in the loop block, but Cys296 is in the block and forms a disulfide bridge with Cys310. This bridge is regulated through *S*-nitrosylation and influences activation.[66,67] We could not find established regulatory elements in the loop block of IRK; although IRK possesses threonine in the activation loop part of the block and a serine in the G loop part, these residues were not predicted to be phosphorylated using the NetPhos (http://www.cbs.dtu.dk/services/NetPhos/) or Phospho.ELM (http://phospho.elm.eu.org/) servers. Overall, the stabilization of the inactive state of kinases has been less studied than stabilization of the active state, but very recently, a Go-model-based dynamics of the ATP-binding induced conformational change in PKA (a protein kinase that was not included in our study) also showed that fluctuations of the G loop and activation loop are positively correlated.[62] Further investigation of the inactive conformation loop block in structures of other kinases[68] is currently underway.

## Summary and Discussion

In view of the importance of protein flexibility in cell function and in drug design,[69] we introduced the BlockMaster procedure, which enables rapid identification of semirigid substructures within protein domains. Use of this procedure highlighted potential regulatory substructures in protein kinases, some of which have not been explicitly discussed in the literature.

BlockMaster partitions protein structures based on the correlation of motions calculated from NMA. The advantage of this procedure is that it relies on the intrinsic motion of the structure in question, as opposed to many other methods that require two different conformations of the same protein as input. In the current work, we chose to use the ENM for normal mode correlation to represent the intrinsic motions of the proteins. The ENM has been shown to be successful in many different studies of protein dynamics.[23,58,70] However, there are obvious limitations to this approximation, which cannot explore the role of side chains. Thus, our future plans include BlockMaster analysis using full-atom normal modes and comparison with the ENM results. In addition, the correlation matrix can also be calculated from MD simulations.

The overall differences between the flexible and semirigid regions in active and inactive conformations of kinases, together with the example of a state-specific "loop" block, highlight the following issue: in different conformational states, different parts of the protein behave as semirigid blocks. While these findings provide important and novel insights into the mechanisms of stabilization of the different conformational (and therefore functional) states of the enzyme, they also have an important technical consequence. The naive "freezing" of blocks identified in one conformation would prevent undergoing the conformational change needed to achieve the other conformation. A possible solution to this problem could be "on-the-fly" application of the BlockMaster procedure to intermediate snapshots

along the trajectory. However, the feasibility of such an approach requires proof of principle and will be part of our future work.

While the investigation of protein kinase dynamics continues to be the focus of our studies, BlockMaster scripts are available upon request and are immediately applicable to other proteins.

**Acknowledgment.** We thank Assaf Gottlieb, Prof. David Horn, and the members of our lab for helpful discussions. This work was supported, in part, by the U.S.−Israel Binational Science Foundation Grant #2007296.

**Supporting Information Available:** Additional parameters and results. This material is available free of charge via the Internet at http://pubs.acs.org.

## References and Notes

(1) Bozulic, L.; Morin, P., Jr.; Hunter, T.; Hemmings, B. A. *Sci. STKE* **2007**, *2007*, pe8.
(2) Quintas-Cardama, A.; Kantarjian, H.; Cortes, J. *Nat. Rev. Drug Discovery* **2007**, *6*, 834.
(3) Zhang, J.; Yang, P. L.; Gray, N. S. *Nat. Rev. Cancer* **2009**, *9*, 28.
(4) Segawa, Y.; Suga, H.; Iwabe, N.; Oneyama, C.; Akagi, T.; Miyata, T.; Okada, M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 12021.
(5) Jacobs, M. D.; Caron, P. R.; Hare, B. J. *Proteins* **2008**, *70*, 1451.
(6) Young, M. A.; Gonfloni, S.; Superti-Furga, G.; Roux, B.; Kuriyan, J. *Cell* **2001**, *105*, 115.
(7) Sanders, M. J.; Grondin, P. O.; Hegarty, B. D.; Snowden, M. A.; Carling, D. *Biochem. J.* **2007**, *403*, 139.
(8) Pellicena, P.; Kuriyan, J. *Curr. Opin. Struct. Biol.* **2006**, *16*, 702.
(9) Kobe, B.; Kampmann, T.; Forwood, J. K.; Listwan, P.; Brinkworth, R. I. *Biochim. Biophys. Acta* **2005**, *1754*, 200.
(10) Kornev, A. P.; Haste, N. M.; Taylor, S. S.; Eyck, L. F. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 17783.
(11) Yang, S.; Roux, B. T. *PLoS Comput. Biol.* **2008**, *4*, e1000047.
(12) Ozkirimli, E.; Yadav, S. S.; Miller, W. T.; Post, C. B. *Protein Sci.* **2008**, *17*, 1871.
(13) Verkhivker, G. M. *Proteins* **2007**, *66*, 912.
(14) Bukhtiyarova, M.; Karpusas, M.; Northrop, K.; Namboodiri, H. V.; Springman, E. B. *Biochemistry* **2007**, *46*, 5687.
(15) Henzler-Wildman, K.; Kern, D. *Nature* **2007**, *450*, 964.
(16) Adcock, S. A.; McCammon, J. A. *Chem. Rev.* **2006**, *106*, 1589.
(17) Sherwood, P.; Brooks, B. R.; Sansom, M. S. P. *Curr. Opin. Struct. Biol.* **2008**, *18*, 630.
(18) Shan, Y.; Seeliger, M. A.; Eastwood, M. P.; Frank, F.; Xu, H.; Jensen, M.; Dror, R. O.; Kuriyan, J.; Shaw, D. E. *Proc. Natl. Acad. Sci. U.S.A* **2009**, *106*, 139.
(19) Tozzini, V. *Curr. Opin. Struct. Biol.* **2005**, *15*, 144.
(20) Elber, R. *Curr. Opin. Struct. Biol.* **2005**, *15*, 151.
(21) Tirion, M. M. *Phys. Rev. Lett.* **1996**, *77*, 1905.
(22) Bahar, I.; Wallqvist, A.; Covell, D. G.; Jernigan, R. L. *Biochemistry* **1998**, *37*, 1067.
(23) Rueda, M.; Chacón, P.; Orozco, M. *Structure* **2007**, *15*, 565.
(24) Emperador, A.; Carrillo, O.; Rueda, M.; Orozco, M. *Biophys. J.* **2008**, *95*, 2127.
(25) Ma, J. *Structure* **2005**, *13*, 373.
(26) Essiz, S.; Coalson, R. D. *J. Chem. Phys.* **2006**, *124*, 144116.
(27) Essiz, S. G.; Coalson, R. D. *J. Chem. Phys.* **2007**, *127*, 104109.
(28) Chun, H. M.; et al. *J. Comput. Chem.* **2000**, *21*, 159.
(29) Jones, S.; Stewart, M.; Michie, A.; Swindells, M. B.; Orengo, C.; Thornton, J. M. *Protein Sci.* **1998**, *7*, 233.
(30) Kundu, S.; Sorensen, D. C.; Phillips, G. N. *Proteins: Struct., Funct., Bioinf.* **2004**, *57*, 725.
(31) Flores, S. C.; Gerstein, M. B. *BMC Bioinf.* **2007**, *8*, 215.
(32) Emekli, U.; Schneidman-Duhovny, D.; Wolfson, H. J.; Nussinov, R.; Haliloglu, T. *Proteins: Struct., Funct., Bioinf.* **2007**, *70*, 1219.
(33) Jacobs, D. J.; Rader, A. J.; Kuhn, L. A.; Thorpe, M. F. *Proteins* **2001**, *44*, 150.
(34) Painter, J.; Merritt, E. A. *Acta. Crystallogr., Sect. D* **2006**, *62*, 439.
(35) Mosca, R.; Schneider, T. R. *Nucleic Acids Res.* **2008**, *36*, W42.
(36) Yesylevskyy, S. O.; Kharkyanen, V. N.; Demchenko, A. P. *Biophys. Chem.* **2006**, *119*, 84.
(37) Yesylevskyy, S. O.; Kharkyanen, V. N.; Demchenko, A. P. *Proteins* **2008**, *71*, 831.
(38) Yesylevskyy, S. O.; Kharkyanen, V. N. *Proteins*, **2008**.
(39) Yesylevskyy, S. O.; Kharkyanen, V. N.; Demchenko, A. P. *Biophys. J.* **2006**, *91*, 670.
(40) Huse, M.; Kuriyan, J. *Cell* **2002**, *109*, 275.
(41) Niv, M. Y.; Filizola, M. *Proteins* **2008**, *71*, 575.
(42) Niv, M. Y.; Skrabanek, L.; Filizola, M.; Weinstein, H. *J. Comput.-Aided Mol. Des.* **2006**, *20*, 437.
(43) Lindahl, E.; Azuara, C.; Koehl, P.; Delarue, M. *Nucleic Acids Res.* **2006**, *34*, W52.
(44) Wang, Y. M.; Jernigan, R. L. *Biophys. J.* **2005**, *89*, 3399.
(45) Van Wynsberghe, A. W.; Cui, Q. *Structure* **2006**, *14*, 1647.
(46) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235.
(47) Fiser, A.; Sali, A. *Bioinformatics* **2003**, *19*, 2500.
(48) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577.
(49) Niv, M. Y.; Skrabanek, L.; Roberts, R. J.; Scheraga, H. A.; Weinstein, H. *Proteins* **2008**, *71*, 631.
(50) Mosca, R.; Brannetti, B.; Schneider, T. R. *BMC Bioinfor.* **2008**, *9*, 352.
(51) Snow, C.; Qi, G.; Hayward, S. *Proteins* **2007**, *67*, 325.
(52) Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S. *Science* **2002**, *298*, 1912.
(53) Azam, M.; Seeliger, M. A.; Gray, N. S.; Kuriyan, J.; Daley, G. Q. *Nat. Struct. Mol. Biol.* **2008**, *15*, 1109.
(54) Hubbard, S. R.; Wei, L.; Hendrickson, W. A. *Nature* **1994**, *372*, 746.
(55) Yang, J.; Cron, P.; Good, V. M.; Thompson, V.; Hemmings, B. A.; Barford, D. *Nat. Struct. Biol.* **2002**, *9*, 940.
(56) Schulzegahmen, U.; Brandsen, J.; Jones, H. D.; Morgan, D. O.; Meijer, L.; Vesely, J.; Kim, S. H. *Proteins: Struct., Funct., Genet.* **1995**, *22*, 378.
(57) Flores, S. C.; Lu, L. J.; Yang, J. L.; Carriero, N.; Gerstein, M. B. *BMC Bioinfor.* **2007**, *8*, 20.
(58) Tama, F.; Brooks, C. L. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 115.
(59) Niv, M. Y.; Rubin, H.; Cohen, J.; Tsirulnikov, L.; Licht, T.; Peretzman-Shemer, A.; Cna'an, E.; Tartakovsky, A.; Stein, I.; Albeck, S.; Weinstein, I.; Goldenberg-Furmanov, M.; Tobi, D.; Cohen, E.; Laster, M.; Ben-Sasson, S. A.; Reuveni, H. *J. Biol. Chem.* **2004**, *279*, 1242.
(60) Wexler, I. D.; Niv, M. Y.; Reuveni, H. *Biotechniques* **2005**, *39*, 575.
(61) Mack, E.; Ziv, E.; Reuveni, H.; Kalman, R.; Niv, M. Y.; Jorns, A.; Lenzen, S.; Shafrir, E. *Diabetes Metab. Res. Rev.* **2008**, *24*, 577.
(62) Hyeon, C.; Jennings, P. A.; Adams, J. A.; Onuchic, J. N. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 3023.
(63) Nolen, B.; Taylor, S.; Ghosh, G. *Mol. Cell* **2004**, *15*, 661.
(64) Yao, H.; Mihalek, I.; Lichtarge, O. *Proteins: Struct., Funct., Bioinfor.* **2006**, *65*, 111.
(65) Welburn, J. P. I.; Tucker, J. A.; Johnson, T.; Lindert, L.; Morgan, M.; Willis, A.; Noble, M. E. M.; Endicott, J. A. *J. Biol. Chem.* **2007**, *282*, 3173.
(66) Lu, X. M.; Lu, M.; Tompkins, R. G.; Fischman, A. J. *J. Mass Spectrom.* **2005**, *40*, 1140.
(67) Sun, J. H.; Steenbergen, C.; Murphy, E. *Antioxid. Redox Signaling* **2006**, *8*, 1693.
(68) Jacobs, M. D.; Caron, P. R.; Hare, B. J. *Proteins: Struct., Funct., Bioinfor.* **2008**, *70*, 1451.
(69) Teague, S. J. *Nat. Rev. Drug Discovery* **2003**, *2*, 527.
(70) Bahar, I.; Rader, A. J. *Curr. Opin. Struct. Biol.* **2005**, *15*, 586.

JP900885W