

MOLECULAR SURFACE SEQUENCE ANALYSIS OF SEVERAL *E. COLI* ENZYMES AND IMPLICATIONS FOR EXISTENCE OF CASEIN KINASE-2 BACTERIAL PREDECESSOR

Ivan Torshin

Laboratory Of Kinetic And Catalysis, Chair Of Physical Chemistry, Chemical Department of Moscow State University, Moscow, 119899, Russia

TABLE OF CONTENTS

1. Abstract
2. Introduction
 - 2.1. Casein kinase 2 functions in eukaryotic cells
 - 2.2. Physiological and evolutionary implications for existence of a bacterial CK2 predecessor
 - 2.3. Aims of the present work
3. Methods. Molecular surface sequence analysis
 - 3.1. Surface accessibility calculations
 - 3.2. Finding surface atoms list
 - 3.3. Finding surface sequences
 - 3.4. Identification of CK2 PROSITE sequence patterns.
 - 3.5. Identification surface sequences and CK2 patterns overlap
4. Results. Placement of CK2 sites on molecular surfaces of several *E. coli* key proteins
 - 4.1. Alkaline phosphatase
 - 4.2. Adenylate kinase
 - 4.3. Phosphofructokinase-1
 - 4.4. D-glyceraldehyde-3-phosphate dehydrogenase
 - 4.5. Isocitrate dehydrogenase
5. Discussion. Molecular surface sequence analysis procedure and implications for existence of casein kinase 2 bacterial predecessor
 - 5.1. Protein kinase phosphorylation in bacteria
 - 5.2. Surface calculations and surface sequence identification
 - 5.2.1. Surface accessibility calculations
 - 5.2.2. Surface atom identification
 - 5.2.2.1. Surface accessibility cutoffs
 - 5.2.2.2. Ray tracing molecule representations
 - 5.2.2.3. Arbitrary empiric parameters used during surface accessibility calculations and surface atom identification
 - 5.2.2.4. Choice of dot density and minimal accessible atomic surface
 - 5.2.2.5. Choice of probe size
 - 5.2.2.6. Application of the procedures
 - 5.3. Identification of surface residues, sequences and site sequences
 - 5.3.1. Identification of surface residues and sequences
 - 5.3.2. Identification of surface CK2-sites
 - 5.4. CK2 sites in *E. coli* enzymes: summary of structural and sequence data for the enzymes analyzed
 6. Perspective
 7. References

1. ABSTRACT

Casein kinase-2 (CK2) is known as pleiotropic eukaryotic protein kinase that phosphorylates significant number of cellular proteins. Not all functions of the protein were registered up to the present time. However, it is known that this Ser/Thr-specific kinase is involved in the cell cycle progression and is essentially required for the eukaryotic cell viability. Fully automated molecular surface analysis procedure for identification of functionally significant surface residues and sequences on the base of protein spatial structure was elaborated. Using the elaborated procedure, several *E. coli* enzymes spatial structures and sequences were investigated. It was found

that most of the casein kinase 2 potential sites found in sequences of enzymes are accessible for modification. Four of the 5 structures studied have CK2 consensus sites that may definitely influence the activity of the enzyme upon phosphorylation. Some of the potential "CK2-sites" has amino acid contents characteristic for physiological substrates of casein kinase 2 in eukaryotes. The main point of the elaborated method and the structural evidence for existence of a putative casein kinase *E. coli* predecessor or a protein with similar kinase activity are discussed. Physiological, biochemical, structural and evolutionary aspects of the existence of the putative predecessor are considered.

2. INTRODUCTION

2.1. Casein kinase 2 functions in eukaryotic cells

Casein kinase 2 (CK2) is a protein serine/threonine kinase that is ubiquitously distributed in eukaryotes. This pleiotropic protein kinase is involved in the phosphorylation of many of cellular targets, as there are more than a hundred proteins are phosphorylated by the enzyme at the consensus CK2 sequence sites (1).

Casein kinase-2 is required for cell's viability and for cell cycle progression. CK2 concentration is especially elevated in proliferating tissues, either normal or transformed. One of the important properties of this kinase is its absence of regulation, as its activity is independent on intracellular concentrations of cyclic nucleotides and calcium (1).

The enzyme is involved in some principal cellular processes, particularly: DNA processing; transcriptional regulation; translational control; post-translational modifications. DNA processing: It was shown that topoisomerase II is a substrate for casein kinase 2. Also, casein kinase 2 has important non-enzymatic function as it plays a significant role in regulating human topo II alpha protein action via stabilization against thermal inactivation (2). Transcriptional regulation: it was shown that the action of casein kinase 2 on binding of transcription factors Oct-1 and Oct-2 to H2B histone is independent on the effects of protein kinase A and protein kinase C. Thus CK2 has some special still undefined role in regulation of transcription (3). Translational control: the direct interaction of CK2 subunits with ribosomal proteins also suggests that CK-2 beta or the whole enzyme molecule may be involved in ribosome assembly or translational control (4). Post-translational modifications: Protein phosphorylation is a significant mechanism in many cellular functions, including genomic regulation and control of cell proliferation (5). A major problem concerning CK2 is its apparent lack of regulation. At least 160 intracellular substrates of CK2 were identified (1).

2.2. Physiological and evolutionary implications for existence of a bacterial CK2 predecessor

1. The CK2 has enzyme rather wide substrate range and is involved in several principal intracellular processes: DNA processing; transcriptional regulation; translational control; post-translational modifications, that are essential both for eukaryotes and prokaryotes.

2. Recently made analysis of the evolutionary origins of the eukaryotic-type protein kinase superfamily allowed to identify five distinct families of known and predicted putative protein kinases with representatives in bacteria and archaea that share a common ancestry with the eukaryotic protein kinases were detected. Four of these protein families have not been identified previously as protein kinases (6).

Thus, as it is stated in a range of works on CK2 physiology, this enzyme has fundamental role in eukaryotic cellular regulation (1, 7, 8, 9). Evolutionary data (6) show

that existence of bacterial predecessors of widely documented eukaryotic protein kinases does not seem improbable. The structural evidences, presented in this work, allow to suppose the existence of a bacterial CK2 predecessor or at least a protein with a like activity in *E. coli*.

2.3. Aims of the present work

The whole analytic procedure developed is based on the main assumption that if there are functional phosphorylation sites in structures of a set of bacterial proteins, there is (are) some correspondent enzyme(s) for modification of the sites.

The starting point of the investigation was an observation, that many amino acid sequences of *E. coli* enzymes do contain consensus sites for CK2 phosphorylation. However, only on the base of presence of a sequence with a certain pattern potential functional properties could not be assessed. However, having spatial structures of the corresponding enzymes, some additional observations allow to evaluate the functional properties of each consensus sequence.

Thus, the primary aims of the present work were to define concept "functional surface of a protein molecule" free of controversy related to empirical parameters of surface residue identification and to investigate the placement of CK2 sequence sites in relation to the surfaces of several bacterial proteins. In other words, kinase sites presented on the surface of protein molecule may be called as 'functional', especially, if their modification may affect region(s) of enzyme's active or allosteric site.

3. METHODS. MOLECULAR SURFACE SEQUENCE ANALYSIS

Spatial structures of the proteins were taken from PDB databank (10), CK2 pattern description is presented in the PROSITE database (11).

During the present work several programs were elaborated and implemented as executables and MS DOS batch scripts. All the programs used for the protein structure analysis were elaborated by the author and are written in Borland Pascal 7.0 (object-oriented Pascal). The peculiarities of the implemented algorithms which are essential for obtaining the physico-chemically consistent results are considered in the 'Discussion' section of this paper. Here brief descriptions of each program used are presented. For visualization of protein structures the RasMol 2.5 program was used (12).

3.1. Surface accessibility calculations

SURFC (Surface Calculations, a program).The program performs atomic surface accessibility calculations. The program takes as inputs PDB file name and globule name, type of atomic subset (all atoms, protein atoms, heteroatoms) to be used in calculations, and optionally dot density (number of dots per atom sphere) and probe size. Outputs are a PDB file with atomic accessibilities and a file with per-residue accessibilities. The surface calculation

routine used follows the method of numerical surface calculation. In the procedure a set of surface dots for each atom-sphere is found, then inaccessible dots are deleted and the final area and dots are found. As in the most of PDB files hydrogen atoms are not presented hydrogen atoms were not used in surface calculations. In order to obtain accessibility values for solvent accessible surface, checking with probe of given radius (1.2-1.6) was used. Relatively low values of densities (100..300, number of dots per atom sphere) were used during calculations. The reasons for this are presented in the 'Discussion' section.

3.2. Finding surface atoms list

SURFAT (Surface Atoms, a program). The program takes as inputs PDB file name and globule name. Probe size that corresponds to the resolution of the surface representation and minimal surface atom accessibility may be supplied optionally. The program finds list of surface atoms and produces a text file with surface atoms and accessibilities for each atom. The program builds three ray tracing representations of the molecule and then finds the atoms composing the outer part of the molecule's surface on the base of the ray tracing representations for three perpendicular directions (X,Y,Z axes).

3.3. Finding surface sequences

SURFSQ (Surface Sequences, an MS DOS batch). SURFSQ batch is used to find the list of surface sequences. The batch script takes PDB experiment name and globule name and finds surface sequences and single surface residues for the given chain of the PDB experiment. It uses the data supplied by SURFC, SURFAT and some auxiliary programs for operation with text tables. The "surface residues" are defined as having at least one "surface" atom. The output is presented in the .SIT file format developed by the author that is based on the format of PROSITE entries. The sites found using PROSITE patterns are also stored in the .SIT format. An example of .sit file format is presented in the section dealing with sequence pattern identification.

3.4. Identification of CK2 PROSITE sequence patterns

SCANPS (Scan ProCite, a program). The program identifies patterns presented in the PROSITE database for a given sequence. Program takes name of sequence containing file as one parameter. Sequence is to be supplied in FASTA format. The output is made into a file of .SIT format where the records of "SS"-type contain direction of pattern occurrence (+1 or -1), the pattern's borders in the sequence (N1, N2), a reserved two-symbol field and amino acid sequence for the pattern. Example of .sit file format is presented below:

```
CC
CC Example of .sit file format:
CC
CL 449;
CC Chain length is presented in the "CL" record
CH A;
ST CK2_ CK2 phosph. site;
PA PS00006: [ST]-X(2)-[DE];
SS +1 36 39 __ SLSD
SS +1 131 134 __ TILE
SS +1 147 150 __ STAE
//
```

3.5. Identification surface sequences and CK2 patterns overlap

SQSOVR (Sequence Overlaps, a program). The program finds sequence overlaps between the sequences of specified pattern types, presented in .SIT file format. The program takes as parameters names of the input and output .sit files, two sequence pattern identifiers and sequence overlap gap (default 1). A text line table with overlap chain regions, sites and SSE string for the given chain region is composed and written as comment lines for the file of .sit format containing overlap regions for the two specified site types. Identification surface sequences and CK2 patterns overlaps is used as indication of a sequence sites' placement on the surface of the molecule. Default overlap gap value is 1 (at least two consequent residues in a protein sequence or one or more residues in both sequences). Additional details of the method, that are of some physico-chemical importance are presented in the correspondent section of the "Results" section.

4. RESULTS. PLACEMENT OF CK2 SITES ON MOLECULAR SURFACES OF SEVERAL *E. COLI* KEY PROTEINS

A functional site (a site that may have some biological significance) must be accessible for those cellular molecules that may interact with it. Therefore, the sequences and atoms to be modified are to be placed on the molecular surface of a given protein. A procedure for identification of surface sequences on the base of protein spatial structures that uses no arbitrary assumptions for solvent accessibility values of "surface" residues was elaborated. The procedure of identification of surface sequence sites described in this article was applied for a set of *E. coli* proteins for the purpose of functional analysis of CK2 sites occurrences.

For the analysis were taken structures of several key *E. coli* enzymes, involved in pathways, essential for cell's functioning and development: signal transduction, phosphate supply and energy supplies (glycolysis and tricarboxylic acid cycle). The enzymes are listed in the table 1.

4.1. Alkaline phosphatase

The phosphate equilibrium in *Escherichia coli* is regulated by the inorganic phosphate supplies from the surrounding medium. Derepression of the *phoA* (alkaline phosphatase) gene and enzyme's activation particularly occurs under phosphate starvation conditions (lake water) (13). The *phoA* gene is repressed by high inorganic phosphate concentrations in the medium (14). Alkaline phosphatase (APase, ALP), a Zn and Mg-binding metalloenzyme found in most organisms, the hydrolyses phosphate esters, optimally in basic conditions (15). It is found in the periplasmic space of *E. coli*, in lysosome-like vacuoles in yeasts, and is a membrane-bound glycoprotein in mammals, being attached to the membrane via a GPI anchor (16).

The results of the analysis of *E. coli* APase structure are presented in the figure 1 and in table 2. 9

Casein kinase 2 bacterial predecessor

Table 1. Some key *E. coli* enzymes with established spatial structures and with sequences containing CK2 sequence patterns

Enzyme (abbreviation)	EC number	Chain length	N subunits/ domains	N	Metabolic pathway
Alkaline phosphatase (ALP)	3.1.3.1	449	2/1		Phosphate supply
Adenylate kinase (ADK)	2.7.4.3	214	1/1		Nucleotide metabolism
Phosphofructokinase (PFK)	2.7.1.11	320	4/2		Glycolysis
Glyceraldehyde-3-phosphate dehydrogenase (GAPD)	1.2.1.12	330	4/2		Glycolysis
Isocitrate dehydrogenase (ICD)	1.1.1.42	416	2/1		Tricarboxylic acid cycle

Table 2. CK2 consensus sites and adjacent surface sequences in the structure of *E. coli* alkaline phosphatase (PDB file 1alk)

N	CK2 site	Sequence	Surf. seq.	Comment
1	36 39	N-SLSD-K	28 41	Involved in interglobular contact
2	131 134	P-TILE-M	121 134	
3	147 150	NV-STAE-L	148 148	Single surf. res. Thr 148, rel. accs. 0.11
4	190 193	G-SITE-Q	--- ---	Wholly "buried", non-functional site.
5	210 213	K-TFAE-T	211 224	
6	216 219	A-TAGE-WQ	211 224	
7	224 227	K-TLRE-E	226 231	ATP binding site sequence
8	245 248	N-SVTE-AN	247 248	
9	311 314	L-SKNE-K	311 315	
10	405 408	N-SEED-SQ	403 411	Active site LOOP 402 417; interglobular contact

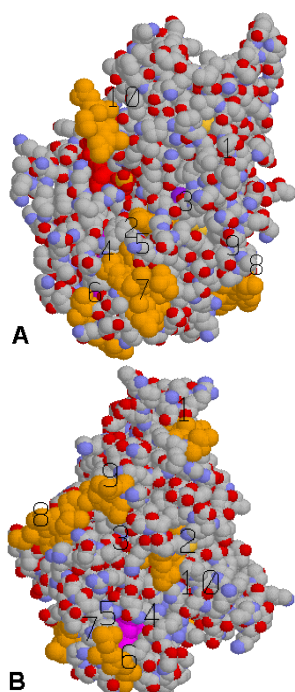


Figure 1. Casein kinase 2 consensus sites on the surface of *E. coli* alkaline phosphatase. a) Active site sequence is shown in red. Bound phosphate is shown as yellow spot surrounded by three oxygens (a-right from the shown active site sequence). CK2 site sequences (labels 5, 6 7 8 10) that may be seen as surface are shown in orange. Atoms of CK2 site sequence (label 10) are placed more or less near the active site sequence (on the 10-15 angstrom distance) and on the same loop (coil) 402-417. The only accessible residue Thr 148 in a CK2 site sequence (label 3) is shown as a small purple spot. b) Sites 1 2 4 6 8 9 are shown. Fully "buried" (inaccessible, all atoms have zero accessibility) CK2 site sequence (label 4) is shown as a large spot in magenta.

from 10 consensus CK2 sites were identified as surface. As follows from the data presented in the table 2, there are at least three potential CK2 sites, that may influence activity and stability of the enzyme. Site CK2-1 (36; 39) is involved in the interglobular contact of the APase dimer, thus dimer formation and stability may depend on its modification. Site CK2-7 is adjacent to the ATP/GTP-binding site (217; 224) with consensus pattern [AG]-X(4)-G-K-[ST] (PS00017), therefore, modification of the CK2-7 site may influence ATP-ligand binding and folding of the enzyme. Site CK2-10 is placed on the loop 402 417 in the sequence fragment adjacent to the residues of active site (Zn binding His 412) and residues involved in the interglobular interaction (Glu 406; Glu 411; Gln 416) and may influence both these functional regions of the molecule (figure 1a).

Thus, almost all (excepting one) potential CK2 phosphorylation sites, found during sequence scanning with CK2 pattern are placed on the molecule outer surface and modification of some of them may definitely influence the APase activity after potential phosphorylation.

4.2. Adenylate kinase

Adenylate kinase is an essential enzyme, responsible for recycling AMP in energetically active cells (17). This monomeric enzyme is located in cytoplasm and reversibly catalyzes interconversion of adenosine monophosphate (AMP) and adenosine triphosphate (ATP) into the two ADP molecules. Magnesium ions are required as cofactor (18). The enzyme has been crystallized and three-dimensional structure was obtained (19).

The results of the molecular surface analysis are presented in the figure 2 and in the table 3. All consensus CK2 sequences found in the sequence of the enzyme were identified as surface sites (figure 2a). For localization of substrate and product binding sites an ADK-inhibiting analog of ATP 5'-adenyl-imido- triphosphate (ANP)

Table 3. Consensus CK2 sites and adjacent surface sequences of the *E. coli* adenylate kinase molecule (PDB file 1ank).

N	CK2 site	Sequence	Surf. seq.	Comment
1	30 33	QI-STGD-M	32 37	Involved in AMP binding
2	41 44	K-SGSE-L	39 45	
3	149 152	DV-TGEE-L	147 162	
4	155 158	T-TRKD-D	147 162	Involved in ANP(ATP) binding

Table 4. CK2 putative sites and surface sequences in the molecule of *E. coli* phosphofructokinase

N	CK2 site	Sequence	Surf. seq.	Comment
1	9 12	T-SGGD-A	9 12	5-8 angstrom from molecules in active site
2	56 59	Y-SVSD-M	45 56	Involved into interprotein contact
	56 59		58 63	
3	145 148	S-TVVE-A	143 145	Involved into interprotein contact
	145 148		147 156	
4	164 167	I-SVVE-V	158 168	
5	197 200	EF-SRED-L	187 213	
6	238 241	KE-TGRE-T	237 241	8-12 angstrom from the allosteric ADP(ADP-2)

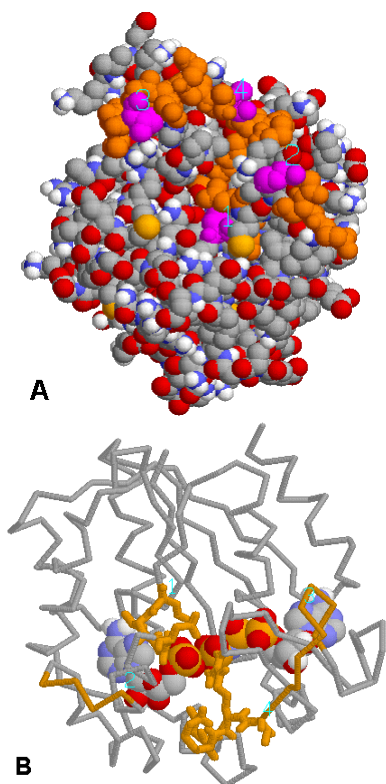


Figure 2. Casein kinase 2 consensus sites on the surface of *E. coli* adenylate kinase. a) Space-filling representation of the enzyme's atomic coordinates (PDB entry 1ank). All identified CK2 consensus sites are placed on the surface and are shown in orange, being numbered according to their occurrence in the protein's sequence. Atoms of the Ser/Thr residues to be modified are shown in magenta. b) C-alpha trace of *E. coli* adenylate kinase is shown. Consensus CK2 sequences are marked in orange. Bound ligand molecules are shown using space-filling representation. The atoms of CK2 sites placed in the vicinity of atoms of the bound ligands (ANP and AMP) are shown using wireframe mode.

was used in the work (20) instead of adenosine triphosphate. This allow to obtained structure of the enzyme-substrate complex possibly very close to transition state of the enzyme (figure 2b). Atoms of the site CK2-1 are placed on the 5-10 angstrom distance from AMP atoms, where as atoms of the CK2-4 are placed on the 3-5A distance from ANP atoms. OG1 atom of Thr 31 in the CK2-1 site forms a hydrogen bond with N7 atom of the AMP molecule, whereas Arg 156 of the CK2-4 site binds the second phosphate group in ANP by forming a hydrogen bond with O3G phosphate atom. Thus, phosphorylation of the consensus CK2 sites in *E. coli* adenylate kinase may definitely affect substrate binding and catalytic properties of the enzyme.

4.3. Phosphofructokinase-1

Phosphofructokinase (PFK) is one of the key regulatory enzymes in the glycolysis. It catalyzes the phosphorylation by ATP of fructose 6-phosphate into fructose 1,6-bisphosphate. The enzyme is allosterically activated with ADP and GDP and non-allosterically with fructose-6-phosphate (21, 22).

The results of surface sequence and CK2 sites placement analysis of the *E. coli* pfkA phosphofructokinase are presented in the figure 3 and in the table 4. All consensus CK2 sites found in the sequence of the enzyme were identified as surface sites.

The found surface CK2 site sequences are shown in figure 3ab. As atoms of the site CK2-1 are placed on the 5-10 angstrom distance from the atoms of molecules bound in active sites (figure 3a), this allows to suppose influence of possible CK2 phosphorylation on the ligand binding in active site. At the same time, atoms of sites CK2-3 and CK2-6 are placed on 8-12 angstrom distance from the ADP-2 atoms (figure 3b). Some residues of sites CK2-2 (ASP 59) and CK2-3 (VAL 147) are involved in interglobular contact AB. Thus, phosphorylation of the CK2 site sequences may influence binding of the ADP and formation/stability of interglobular contact AB.

Table 5. CK2 sites and surface sequences in the structure of glyceraldehyde- 3- phosphate dehydrogenase from *E.coli*. Coordinate dataset presented in PDB entry 1GAD was used (26)

N	CK2 site	Sequence	Surf. Seq.	Comment
1	23 26	R-SDIE-I	20 24	
2	189 192	P-SHKD-W	178 195	Placed near NADP
3	238 241	NV-SVDD-L	241 243	Involved in interglobular contact
	238 241		239 239	
4	274 277	Y-TEDD-V	272 288	
5	290 293	t-SVFD-A	291 293	
6	309 312	V-SWYD-NE	310 310	Placed near NADP

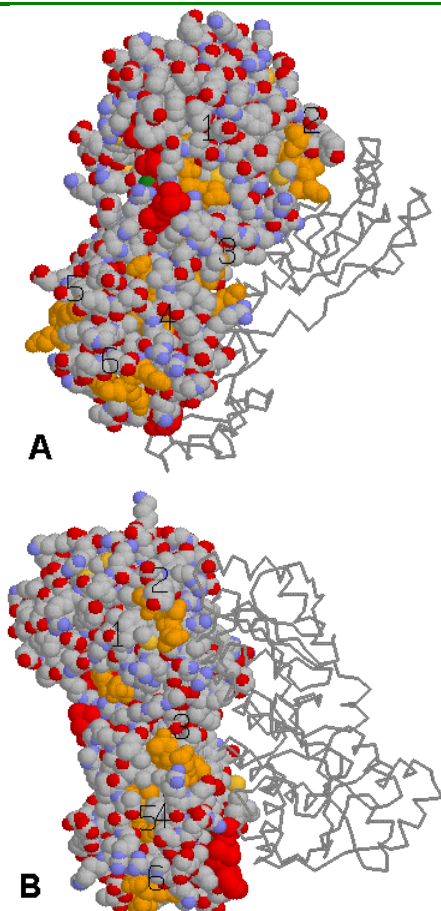


Figure 3. Casein kinase 2 consensus sites on the surface of *E. coli* phosphofructokinase. a) Two globules of the *E. coli* homotetrameric phosphofructokinase are shown. The A globule is represented using space-filling model, the B globule- using C-alpha trace. Fructose-1,6-bisphosphate (FBP) and adenosine diphosphate (ADP-1) bound in the active site of the A-globule are shown in red, and Mg atom bound to ADP is shown in green. Putative casein kinase 2 sites (labels 1-6) are shown in orange. b) Second ADP molecule (ADP-2, shown in red) is bound in a pocket formed by the globules A and B.

4.4. D-glyceraldehyde-3-phosphate dehydrogenase

Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) plays an important role in glycolysis and gluconeogenesis. It reversibly catalyses the oxidation and phosphorylation of D- glyceraldehyde- 3- phosphate into 1,3- diphospho- glycerate (23). The enzyme is located in

cytoplasm and has been found to bind to actin and tropomyosin, and may thus have a role in cytoskeleton assembly, which also may be important for positioning of the glycolytic enzymes (23) (24). Coenzyme NAD⁺ (NADP) is an essential requirement for both catalytic functions (phosphorylative activity and oxidative activity) (25). The enzymes from several bacterial sources were crystallized and the spatial structures were obtained (26, 27).

Results of molecular surface sequence analysis of CK2 sites in the glyceraldehyde- 3- phosphate dehydrogenase from *E. coli* are presented in the table 5 and in figure 4. All CK2 consensus sites were identified as surface. As it may be seen in the figure 4, consensus CK2 phosphorylation site 3 is also involved in the intersubunit contact. Some atoms of the CK2 site sequences 2,3 and 6 are placed on 7-10 angstrom distance from the NADP molecule essential for catalytic activity. Thus, covalent modification of these sites may affect the catalytic properties of the enzyme particularly due to steric hindrances and induced conformational changes after attachment of the phosphate. Sites CK2-3 (238, 241), CK2-5 (290, 293) and CK2-6 (309, 312) are also may be identified in the sequence of *Bacillus Stearotherophilus* enzyme placed in the same sequence regions. In the three-dimensional structure of the *Bacillus* enzyme (27) these three sites were also identified as surface, with atoms of serine atoms presented on the surface of the molecule.

4.5. Isocitrate dehydrogenase

Isocitrate dehydrogenase is an enzyme of the tricarboxylic acid cycle that reversibly catalyzes conversion of isocitrate into 2-oxoglutarate with release of a proton and carbon dioxide molecule. When bacteria are grown on acetate, isocitrate dehydrogenase is phosphorylated and thus inactivated (28). Conversely, when cells are cultured on a preferred carbon source, such as glucose, the enzyme is dephosphorylated and recovers full activity. The *E. coli* enzyme is cold sensitive (29). *E. coli* enzyme was crystallized and the spatial structure was obtained (30).

In the present work, both consensus CK2 sites found in the sequence: (64, 67, I-SWME-I) and (271, 274, N-TGKE-I) were identified as surface even with hydroxyl atoms of the Ser/Thr residues placed on the surface of the molecule, as it is shown in figure 5. As both CK2 consensus sites are placed relatively far from active site change in the activity of the molecule may be induced through the complex conformational change in globule after phosphate attachment.

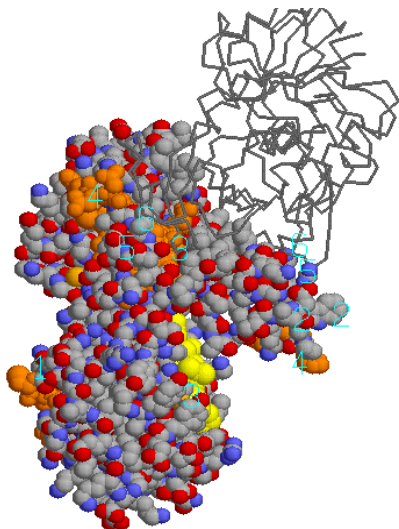


Figure 4. Casein kinase 2 consensus sites on the surface of *E. coli* D- glyceraldehyde- 3- phosphate dehydrogenase. Two subunits of the functional tetramer are presented: one using space-filling representation and another using C-alpha trace representation. NADP of the active site is shown in yellow, putative CK2 sites (labels 1-6) are shown in orange. Site 3 is involved in the intersubunit contact. Atoms of CK2 site sequences 2,3 and 6 are placed on 7-10 angstrom distance from the NADP molecule.

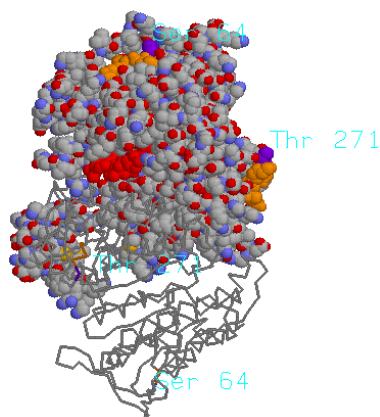


Figure 5. Casein kinase 2 consensus sites on the surface of *E. coli* isocitrate dehydrogenase. Chain "A" of the enzyme's dimer is represented using space-filling mode. C-alpha trace of the chain "B" is shown. Bound NADP of the chain "A" is shown as red spot in the center. Both the two consensus CK2 sites found in the sequence are placed on the surface and are shown in orange. Atoms of the Ser/Thr residues to be putatively phosphorylated are shown in purple.

5. DISCUSSION. MOLECULAR SURFACE SEQUENCE ANALYSIS PROCEDURE AND IMPLICATIONS FOR EXISTENCE OF CASEIN KINASE 2 BACTERIAL PREDECESSOR

5.1. Protein kinase phosphorylation in bacteria

Phosphorylation with protein kinases that has a role in signal transduction or other cellular processes is not restricted to eukaryotic cells only. Protein phosphorylation in *Escherichia coli* and some other bacterial species was

investigated in the work (31). In each of the bacterial species the presence of several phosphorylated proteins was registered. The results bring evidence that protein phosphorylation catalyzed by protein kinases is a post-translational modification widespread among prokaryotes (31).

In recent works, protein phosphorylation is considered to be a universal phenomenon among bacteria in view of the fact that it has been observed in nearly 100 different species. For example, in *E. coli* about 130 different phosphoproteins have been detected but only a dozen have been identified so far (32). Most of the bacterial protein kinases responsible for this modification of proteins share the common property of using adenosine triphosphate as phosphoryl donor. However, they differ from one another in a number of structural and functional aspects. Namely, they exhibit a varying acceptor amino acid specificity and can be classified, on this basis, in these three main groups: protein histidine kinases, protein serine/threonine kinases and protein tyrosine kinases (33).

Protein Ser, Thr and Tyr kinases play essential roles in signal transduction in organisms ranging from yeast to mammals, where they regulate a variety of cellular activities. A number of genes that encode eukaryotic-type protein kinases have also been identified in four different bacterial species, suggesting that such enzymes are also widespread in prokaryotes. Although many of them have yet to be fully characterized, several studies indicate that eukaryotic-type protein kinases play important roles in regulating cellular activities of these bacteria, such as cell differentiation, pathogenicity and secondary metabolism (34). As it was already mentioned analysis of the proteins encoded by the completely sequenced bacterial and archaeal genomes was conducted (6) has allowed to identify five distinct families of known and predicted putative protein kinases that share a common ancestry with the eukaryotic protein kinases.

Conservation of sequence corresponds to conservation of certain biofunctionally important fragments of protein spatial structure. Active sites are often found to be well-conserved during evolutionary sequence studies. As macromolecular recognition processes are thought to be based on the complementarity of molecular shapes and charge distributions (35), conservation of active site regions, that allows to identify genomic sequences as kinases of certain type may correspond to the conservation of substrate specificity. Thus some kinase consensus sequences, established on the base of analyses of eukaryotic kinases may also occur in bacterial protein sequences as functional kinase sites. In order to analyze potential functional properties of these sites that are found in sequences of many bacterial proteins a special set of molecular surface sequence analysis procedures was used.

5.2. Surface calculations and surface sequence identification

For finding most probable functional CK2 kinase sites that occur as sequence patterns it is necessary to analyze their placement in protein's spatial structure. The most simple way to do it for a given protein molecule

Casein kinase 2 bacterial predecessor

seems to be usage of a molecular viewer program for inspection of molecular surface produced using sphere model (spacefilling representation) of the protein. However, visual analysis is insufficient for finding surface residues: an atom that may be 'seen' in a molecular viewer, may have zero solvent accessibility, if the probe of definite size could not be placed in touch with it, thus many residues may be identified as "surface", whereas they are inaccessible.

Arbitrariness of the visual analysis for identification surface residues may be shown on the results of CK2 sequence site placement in the spatial structure of *E. coli* alkaline phosphatase as an example (figure 1). While the only accessible residue Thr 148 in CK2-3 sequence site is shown only as a small purple spot in the figure 1a, CK2-4 sequence with all atoms having zero accessibilities is shown as a large spot in magenta in the figure 1b.

Despite giving inadequate results in some cases due to arbitrariness of selecting surface atoms the method of visual analysis is unreliable in the case of large molecules, as it is non-automated. In the present work a complex method for identification of surface atoms/residues/ sequences on the base of given protein structure is described and it's application for several protein structures is considered. The main algorithms for identification of surface sequences in a given protein structure that were elaborated and used are: atomic surface accessibility calculation routines, "ray tracing" representation composition and analysis routines and routines for surface residues and sequences identification.

5.2.1. Surface accessibility calculations

Surface accessibility calculations are necessary for identification of potential 'active' atoms comprising the surface of the molecule. Some features of the elaborated surface calculations algorithm are considered below.

In this work the "numeric" approach of surface calculation was used. Molecule is considered as a combination of 'fused spheres' (fused sphere model), each presenting individual atoms. Each sphere is represented by a certain number of evenly distributed on the sphere's surface dots and molecule surface dots are found as set of dots of all spheres remaining after deleting the dots, placed inside crossing of spheres (36). Increasing Van der Waals radii set with a given probe radius allows to delete abundant dots inaccessible by the probe of given size. Being computationally more effective numerical calculation procedure produces results close to widespread "analytical method" (Lee and Richards) (37).

Most of the available programs for surface calculations may be classified into programs for molecular surface visualization (38, 39) and programs for high precision quantitative surface calculations of surface energy (40). As the primary purpose of surface energy calculations is finding reliable energy values for groups of atoms, the method may give somewhat enlarged accessibility values for particular atoms.

Another important problem in atomic surface accessibility calculations is that in the methods usually used the continuity of the atomic surface is not taken into account. Thus, a value of accessible surface of 0.5 sq. angstrom may correspond to 5 separate patches on surface of an atom with 0.1 sq angstrom surface each or to one continuous square-like or ribbon-like (dispersed) patch. However, the biochemical significance of a continuous square patch is higher than that of a small dispersed ones.

In order to overcome the considered problems related to surface calculations another surface accessibility calculations procedure was elaborated. The procedure is based on the "numeric" surface calculation algorithm, described in the work (36). First feature of the present method is that only radii of neighboring atoms are increased with probe size, whereas the radius of atom under calculation is remained unchanged. The second feature is that radii increase on the size of probe is performed only during counting the remaining dots, while value of surface area is calculated on the base of "standard" radii values for the default Van der Waals radii set.

For evaluation of surface properties of atoms/residues are used "absolute" and "relative" surface atom/residue accessibility values. The "absolute" value is calculated as sum of atomic solvent accessible areas (remaining dots), whereas the relative accessibility is ratio of "absolute" one to the standard accessibility for given atom/residue. The third feature of the elaborated procedure of surface calculations is that both "absolute" and "relative" accessibilities are calculated for each atom during one cycle of calculation, thus tables of standard atom/residue accessibility are unnecessary.

Usage of low dot densities (100-300) for numerical surface calculations allows easily tackle the continuous patch problem as in the case of such dot density values all the patches on the atomic surface (represented as remaining dots) would be continuous square-like patches of 0.3-0.5 sq angstrom in size.

5.2.2. Surface atom identification

The main problems of physico-chemically motivated identification of surface residues and sequences in a protein structure derived from X-ray data using solvent accessibility criteria are arbitrariness in absolute/relative surface accessibility values assigned to potential surface residues and disregarding overall molecule topology while using atom/residue accessibility cutoffs. These problems were tackled in the elaborated set of algorithms in some extent.

5.2.2.1. Surface accessibility cutoffs

In most cases, the surface atoms and residues are selected using some cut-off value for atomic (residue) solvent accessibility. For instance, in documentation to the program WHATIF (41) value of 0.1 sq. angstrom as minimal value of surface atom accessibility, sufficient for hydrogen bond formation with solvent is mentioned (it would correspond to 1-5 sq. angstrom per residue). Used alone this is rather broad

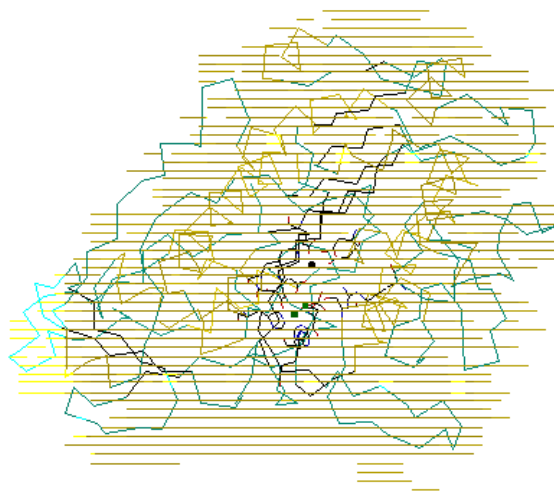


Figure 6. Subunit of *E. coli* alkaline phosphatase with superposed ray-tracing representation of the molecule for X direction in the orthogonal coordinate system. Zn atoms (green) and residues of the active site (CPK coloring) along with C-alpha trace are shown. Rays are presented as yellow lines. The gaps between the yellow lines representing rays correspond to the "opaque" zones- atoms and clusters of atoms. The surface atoms are identified as being placed in first non-empty atom boxes corresponding to the outermost segments. For the visualization program KINEMAGE was used.

criterion, since for many proteins more than 80 per cent of all protein atoms would be listed in this case. Another approach was proposed in the work (36) in which cutoff value for whole amino acid was used, in this case an atom is to be considered as "surface", if accessibility of this atom's residue more than 30-50 sq. angstrom (2.5-5 sq. angstrom per atom). Thus, the empirical criteria may be chosen more or less arbitrarily. "Fixed" cut-off values significantly depend on the method used in calculations, parameters of the given methods (such as dot density in the numerical calculations and set of Van der Waals radii) and probe size used.

On one hand, usage only of some or other cut-off value for atoms/residues to be considered as surface would not allow to find consistent set of surface atoms/residues. Therefore, for identification of surface atoms to be consistent usage of arbitrary cutoff values for surface atoms is to be avoided or at least be significantly restricted. On other hand, as it was mentioned, "surface" of a molecule is defined as part of surface accessible to solvent in the most of algorithms. However, in this definition no distinction is made between the "outer" and "inner" solvent accessible surfaces. Thus, atoms placed in a completely closed cavity (void) with a size slightly bigger than the given probe size would have non-zero solvent accessibility and therefore would be counted as "surface". At the same time these atoms do not have an external function as they could not interact with other molecules being isolated inside the globule. As it was shown in the

work (42) the primary role of voids in protein globules seems to be stabilization of molecular structure. In other words, while using surface accessibility calculations, local molecule topology may be assessed by using probes of different sizes, whereas overall molecule topology, mentioned here in terms of "outer" and "inner" molecular surfaces is not taken into consideration.

Thus, absence of definite relation between residue surface accessibility value and presence of a residue on molecule surface shows insufficiency of usage only surface atom accessibility calculations for surface atom selection.

5.2.2.2. Ray tracing molecule representations

An algorithm, that uses the idea of "ray tracing" was elaborated and used in the present work as such an additional method for finding surface atoms. Usually the method, termed as "ray casting" is implemented on specially constructed parallel computers, and is used in mechanical engineering. Also it was used for high precision surface calculations (43). Ray-casting process imitates passing of light, modeled by a grid of parallel straight lines through a three-dimensional representation of an object. The obtained ray representation of the solid object is list of ray entry and exit point locations for each point of the two-dimensional grid.

An example of such a representation demonstrating the idea is built for *E. coli* alkaline phosphatase molecule and presented in the figure 6. For the visualization KINEMAGE program was used (44). In the procedure elaborated atoms pertaining to the surface may be found as first last ones "met" at the passing of a "ray". Firstly, all atoms of the molecule are distributed between boxes, then "ray tracing" algorithm is applied and the boxes lying on the outer side of the box are found as first non-empty boxes. The accessibility values of the atoms in the selected outer boxes are checked and thus the list of surface atoms is produced. Using three perpendicular directions for tracing allows to find all surface atoms for the given molecule.

5.2.2.3. Arbitrary empiric parameters used during surface accessibility calculations and surface atom identification

As it was mentioned above, the usage of empiric parameters may significantly influence the results of calculation and is to be avoided or at least be significantly restricted. Empirical parameters of surface accessibility calculations are: *dot density* (numerical): 500-700 or slice thickness (analytical): 0.01-0.05 angstrom; *probe size*: 1.3..1.6 angstrom for water molecules in a protein structure (45). This region may be another for some other protein structure. Semi-empirical parameter of surface atom identification: *minimal area of accessible surface patch* for a surface accessible atom: 0.1..5 sq. angstrom.

5.2.2.4. Choice of dot density and minimal accessible atomic surface

The value of calculated solvent accessible surface of a given atom depends on probe size and dot

Casein kinase 2 bacterial predecessor

density (for a given set of Van der Waals radii). With a probe size fixed the maximal value of dot density may be assessed as surface area of an entire atom's Van der Waals sphere divided on the chosen value of minimal surface area cutoff for surface atoms (0.1 sq. angstrom, as proposed in (41)). As in formation of hydrogen bonds with solvent mostly oxygen and nitrogen atoms may be involved (with Van der Waals radii of 1.4 and 1.7 respectively) the values of dot density are ranging in 250-400 dots per sphere. Density values lower than these may be used for calculations with cutoffs bigger than 0.1 sq. angstrom. Using this consideration two empiric parameters- dot density and minimal surface atom accessible area may be reduced to one- minimal atomic surface accessibility.

Another moment already mentioned in the section on surface accessibility calculations is that continuous patches are to have higher functional value. As hydrogen bond formation with solvent may be considered as the least of the biochemical functions, then division of accessible atomic surface into fragments lesser than 0.1 sq angstrom would be biochemically insignificant. In other words, accessible atomic surface may be "quantificated" using value corresponding to the least possible function. In this case the minimal area of a square-like patch would have surface of 0.4 sq angstrom (0.3-0.5).

5.2.2.5. Choice of probe size

As it was shown in the work (45) relatively large volume fluctuations of atoms at the protein-water interface indicate that they have some more variable packing than corresponding atoms in the protein core. It was also shown that usage of "standard probe size" of 1.4 angstrom may be used only as a simplest model, since size of water molecules bound in the protein molecule ranges approximately from 1.25 to 1.55 angstrom (45).

For choice of probe sizes for a given structure to be more or less biochemically adequate the following "protein biochemistry rule" was proposed. In many proteins accessibility of C-atoms of the main chain (carbons in peptide bonds carboxyls) rarely exceeds 0.2 sq angstrom and in the most cases is zero. This fact is observed while using different programs for surface accessibility calculations. As these hydrophobic atoms with almost zero accessibility could not have any surfacial function this observation may be used as calibrating rule for finding minimal probe size for a given protein molecule coordinate dataset. Probe sizes calculated in this way are ranging from 1.3 to 1.6 for different protein structures, which may show usefulness of the procedure of calibrating probe size for the given protein structure instead of using some fixed probe size for all protein structures.

5.2.2.6. Application of the procedures

Introduced in this work partial algorithmic solution of the surface calculation problems, related to the task of identifying surface atoms potentially important for interactions with other biomolecules may be summarized as follows:

1. Using of ray tracing technique along with surface accessibility calculations for identification of protein surface atoms;

2. Finding value of dot density for numerical surface calculations on the base of given minimal area of surface atom; after calculating accessibilities with so chosen dot density values all atoms with non-zero accessibility are taken into account;

3. Using the principle of continuous patch of atomic accessible surface;

4. Automatic calibration of probe size for the given protein structure.

For finding list of surface atoms using the described procedures coordinate datasets for monomers were used along with coordinates of heteroatoms for a monomer. In an oligomeric molecule some sequence sites may be packed between globules and thus have a zero or close to zero accessibility, while be inaccessible from outside of the molecule. Thus, if for the surface calculations the whole coordinate set would be used, some of the sites, that are placed on the outer molecule surface in interglobular contact regions would be counted as "buried" (inaccessible). However, phosphorylation of these sites in a monomer before it's dimerization *in vivo* will certainly influence protein-protein contact formation and thus may influence the functional properties of the oligomeric protein. Therefore, in the case of oligomeric enzymes surface atom identification calculations are to be performed for separate globules.

5.3. Identification of surface residues, sequences and site sequences

5.3.1. Identification of surface residues and sequences

List of surface atoms obtained using the two independent methods: surface accessibility calculations and ray tracing procedure is used for identification of surface residues and sequences. The minimal requirement for a "surface residue" used in this work is that "surface residue" ought to have at least one "surface atom", according to the given definition of surface atoms. This requirement for surface residues used for identification of surface residues and sequences has it's base in the impossibility of evaluation globule *in vivo* dynamics on the base of static structure of a given protein produced by methods of X-ray protein crystallography. In other words, "buried" atoms in a residue that has at least one atom placed on outer molecular surface may become accessible in the globule that is subjected to constant interactions with other cell biomolecules.

Activity of an enzyme may be conserved under significant structural changes. For example, during folding studies of *Escherichia coli* alkaline phosphatase the protein was refolded *in vitro*. It was found that the enzymatic activity reaches its asymptotic value in 1h. In contrast, the structural rigidity (packing) of the hydrophobic core of the protein, monitored by the recovery of the tryptophan phosphorescence lifetime, returns to its characteristic native-like value over several days (46). These data suggest that the core of the protein undergoes continued observable structural rearrangements affecting the rigidity of the protein. As amplitude of these rearrangements could not be assessed in detail, for identification of surface residues the rule "at least one surface atom per surface residue in static

Casein kinase 2 bacterial predecessor

compact structure" was proposed.

Surface sequences are defined as comprised of adjacent (i, i+1) surface residues. Thus, every residue in a surface sequence is identified as "surface residue", that is, containing surface atoms.

5.3.2. Identification of surface CK2-sites

It was already mentioned that it is hardly possible to evaluate globule *in vivo* dynamics on the base of static structure of a given protein produced by methods of X-ray protein crystallography. This aspect touches two moments important for identification of a site surface placement: finding sequences of surface residues and identification of surface kinase sites.

Still there are no definite data on conformational changes that may induced by phosphorylation. However, for example, limited proteolysis -a kind of post-translational modifications requires significant changes in relatively long chain segments. In the work (47) on the base of structural analysis of limited proteolytic sites within native folded protein structures it was shown that significant conformational change in the target proteins are required in order to facilitate binding of the protein substrates into the active site of the attacking serine protease. The results strongly suggest that large local motions (significant changes in the conformation of at least 12 residues) are required for local proteolysis. As a prime determinant for limited proteolysis an ability to unfold locally without perturbing the overall protein conformation was proposed (47).

As protein globule "*in vivo* dynamics", which particularly implies structural changes during and after post-translational modifications, usually is not known for a given protein structure, therefore the amplitudes of chain segment displacements and consequent changes in individual atomic and residue accessibility are also unknown. Thus usage only of calculated surface accessibilities even calculated using definite principles, proposed in the present work is insufficient for determination of possibility for a site to become surface *in vivo*. However, a "surface sequence", containing tens of atoms is less probable to become completely buried during constant globule conformational changes than one "surface atom". Therefore, identification of surface sequences and CK2 patterns overlaps may be used as indication of surface placement, rather than static relative accessibility measurements for atoms or residues even without usage of more or less arbitrary cutoff values. Default overlap gap value of 1 corresponds to at least two consequent residues in a protein sequence or to one or more residues, presented in both sequences.

Overlap with surface sequence is most general rule to show possibility of displacement for a given sequence site in the 3D structure. Thus, a short site sequence like CK2 kinase pattern may have small or even zero accessibility value in the static structure, however, if there are overlaps with surface sequences, the site sequence may become accessible during *in vivo* dynamic transitions

of the globule. As surface sequences are involved in multiple interactions inside hydrogen bond network of the solvent they would have higher motility. If a site sequence does not contain atoms that are identified as "surface" (according to the definition used in the present work) and has no overlaps with surface sequences, most possible that this site would be constantly inaccessible.

5.4. CK2 sites in *E. coli* enzymes: summary of structural and sequence data for the enzymes analyzed

Protein kinase CK2 is a ubiquitous eukaryotic protein kinase responsible for the phosphorylation of Ser and Thr residues specified by acidic side chains in many proteins, including several key enzymes, cytoskeletal proteins, growth factor receptors, transcription factors and enzymes involved in many aspects of DNA metabolism (8). An essential moment concerning CK2 is its apparent lack of regulation through pathways of widely known second messengers: cyclic nucleotides and calcium (48). The enzyme is an important component of signaling pathways that control the growth and division of cells (49).

Casein kinase 2 is a protein serine/threonine kinase that phosphorylates many different proteins. The following moments are important for substrate specificity (50) of this enzyme:

1. Ser is favored over Thr under comparable conditions.
2. An acidic residue (either Asp or Glu) must be present three residues from the C-terminal of the phosphate acceptor site.
3. Asp is preferred to Glu.
4. Additional acidic residues in positions +1, +2, +4, and +5 increase the phosphorylation rate. Most physiological substrates have at least one acidic residue in these positions.
5. A basic residue at the N-terminal of the acceptor site decreases the phosphorylation rate, while an acidic one increases it.

The sequence sites corresponding to the elaborated consensus pattern [ST]-x(2)-[DE] (where S or T is the phosphorylation site for CK2; PROSITE entry PS00006) are found in most of the known physiological substrates of casein kinase 2 (11). CK2 consensus patterns occur in sequences of many *E. coli* proteins (data not presented). In the present work positions of the consensus sequences in spatial structures of several enzymes were analyzed using the procedure of molecular surface sequence analysis. The summary of the enzyme structures analyzed is presented in the table 6.

On the limited dataset of 5 enzyme structures analyzed the following approximations could be made:

1. In average, more than a half CK2 site consensus sequences are fully placed on surface sequences, while almost all CK2 consensus sequences in all protein structures studied have overlap with surface sequences. Thus, most part of CK2 site consensus sequences are fully placed on outer molecular surfaces of the analyzed key *E. coli* proteins.

Table 6. Summary data on CK2 consensus sequences in structures of several key enzymes from *E.coli*

Summary table items	ALP	ADK	GAFD	ICD	PFK
Number of consensus ck2 sequences	10	4	6	2	6
Number of fully accessible sites (with all or with more than a half of residues identified as surface)	6	3	3	2	6
Number of sites with half of the residues accessible	2	1	2	0	0
Sites with all residues buried excepting one	1	0	1	0	0
Fully buried kinase sites	1	0	0	0	0
Sites that may influence activity and (or) stability being phosphorylated	3	2	3	0	4

Table 7. Possible activity regulation mechanisms in the investigated enzymes

Mechanism/Enzyme	ALP	ADK	GPA	PFK	ICD
Modification of interglobular contact	2	0	1	2	0
Modification of active site sequence, active site environment	1	0	0	1	?
Modification of a ligand binding sequence.	1	2	2	1	0

2. 4 from 5 proteins have several CK2 consensus sites (from 1/3 to 1/2 of all sites for each protein) placed on the molecular surfaces that may affect activity or stability of the enzymes upon being modified by phosphate attachment.

Sequence environment of the consensus pattern sequences (rules 1-5 above) is important for additional evaluation of each site's activity. Applying additional optional rules for CK2 sequence pattern identification allows to propose the most probable sites to be phosphorylated by casein kinase 2. In alkaline phosphatase (table 2) the most probable sites are CK2-7 (less, rule 5) and CK2-10 (rule 4), in adenylate kinase (table 3) sites CK2-3 and CK2-4 (rule 4), in glyceraldehyde- 3- phosphate dehydrogenase sites CK2-1 (less, rule 5, but also rule 4), CK2-3, CK2-4 and CK2-6, in phosphofructokinase sites CK2-5 (rule 4) and CK2-6 (rule 5).

Another feature that seems to be important concerns the positions of the CK2 sites in the structures of the enzymes investigated. Thus, in 3 proteins from 5 some sites with phosphorylation enhancing rule 4 realized are placed on the chain segments also involved in the binding of a ligand molecule (allosteric ligand, substrate, coenzyme) or at least are adjacent to them or placed on a short distance. ATP-binding loop in alkaline phosphatase contains site CK2-7, site CK2-4 in adenylate kinase is placed on the segment involved in ATP binding, site CK2-6 in glyceraldehyde- 3- phosphate dehydrogenase is placed near NADP binding sequence. This may represent an important mechanism of influencing activities of some target bacterial proteins. The summary on possible regulation mechanisms of the enzymes investigated is presented in the table 7.

Thus, as it follows from the data presented in the table 7 the possible regulation mechanism common for all cases considered is modification of ligand binding site, whereas modification of interglobular contact sequences stands on the second place.

6. PERSPECTIVE

In the present work a complex method for identification of surface atoms, residues and sequences on the base of given protein structure is described and it's application for several protein structures is considered. On the base of structural interpretation of several bacterial enzyme structures existence of a bacterial predecessor of widely pleiotropic eukaryotic protein casein kinase 2 or a protein with a similar activity was hypothesized. The following moments allow to state this: many amino-acid sequences of *E. coli* enzymes do contain consensus sites for CK2 phosphorylation and almost all CK2 consensus sequence are placed on the surfaces of all the proteins investigated. Some of these CK2 sites may significantly influence activity of the proteins considered. Two main possible regulation mechanisms proposed are modification of ligand binding sites and modification of interglobular contact sites.

The whole *E. coli* genome contains totally about 4400 genes, among them there are about 1600 non-identified genes classified as "hypothetical unclassified unknown" (51). About 100 of the non-identified genes correspond to protein sequences with length 350 -390 residues- length typical for eukaryotic casein kinases 2. There are two possible ways of identification putative predecessor sequences in the *E. coli* genome - analysis of this set of possible protein sequences and /or scanning of the whole genome using CK2 signatures elaborated for eukaryotic proteins.

The presented results allow to assume existence of some protein that may modify functional CK2 sequence sites found in the structures of several *E. coli* enzymes. However, unless the biochemical or at least genomic evidence(s) are not found, the existence casein kinase 2 putative bacterial predecessor or at least a bacterial protein kinase having similar activity is still to be questioned.

7. REFERENCES

1. L. A. Pinna, & F. Meggio: Protein kinase CK2 ("casein kinase-2") and its implication in cell division and proliferation. *Prog Cell Cycle Res* 3, 77-97 (1997)

2. C. Redwood, S. L. Davies, N. J. Wells, A. M. Fry, & I. D. Hickson: Casein kinase 2 stabilizes the activity of human topoisomerase IIalpha in a phosphorylation-independent manner. *J Biol Chem* 27, 3635-42 (1998)

3. S. J. Grenfell, D. S. Latchman, & N. S. Thomas: Oct-1 and Oct-2 DNA-binding site specificity is regulated *in vitro* by different kinases (published erratum appears in *Biochem J* 1996 Aug 1;317(Pt 3):959) *Biochem J*, 889-93 (1996)

4. J. H. Lee, J. M. Kim, M. S. Kim, Y. T. Lee, D. R. Marshak, & Y. S. Bae: The highly basic ribosomal protein L41 interacts with the beta subunit of protein kinase CKII and stimulates phosphorylation of DNA topoisomerase IIalpha by CKII. *Biochem Biophys Res Commun* 23, 462-7 (1997)

5. P. Agostinis, L. A. Pinna, & W. Merlevede: Metabolic regulation through second-site phosphorylation. *Verh K Acad Geneesk Belg* 5, 407-27; di (1989)

6. C. J. Leonard, L. Aravind, & E. V. Koonin: Novel families of putative protein kinases in bacteria and archaea: evolution of the "eukaryotic" protein kinase superfamily. *Genome Res*, 1038-47 (1998)

7. A. Krehan, F. Meggio, R. Pipkorn, L. A. Pinna, & W. Pyerin: Identification of structural elements of subunit beta of human protein kinase CK2 participating in tight physical alpha-beta intersubunit contacts directly adjacent to a surface-oriented region. *Eur J Biochem* 25, 667-72 (1998)

8. L. A. Pinna: Protein kinase CK2. *Int J Biochem Cell Biol* 2, 551-4 (1997)

9. B. Boldyreff, F. Meggio, L. A. Pinna, & O. G. Issinger: Efficient autophosphorylation and phosphorylation of the beta-subunit by casein kinase-2 require the integrity of an acidic cluster 50 residues downstream from the phosphoacceptor site. *J Biol Chem* 26, 4827-31 (1994)

10. F. C. Bernstein, T. F. Koetzle, G. J. Williams, E. F. Meyer, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, & M. Tasumi: The Protein Data Bank. A computer-based archival file for macromolecular structures. *Eur J Biochem* 8, 319-24 (1977)

11. A. Bairoch: PROSITE: a dictionary of sites and patterns in proteins. *Nucleic Acids Res* 20 Suppl, 2013-8 (1992)

12. R. A. Sayle, & E. J. Milner-White: RASMOL: biomolecular graphics for all. *Trends Biochem Sci* 2, 374 (1995)

13. R. Ozkanca, & K. P. Flint: Alkaline phosphatase activity of *Escherichia coli* starved in sterile lake water microcosms. *J Appl Bacteriol* 8, 252-8 (1996)

14. A. Torriani: From cell membrane to nucleotides: the phosphate regulon in *Escherichia coli*. *Bioessays* 1, 371-6 (1990)

15. F. M. Hulett, E. E. Kim, C. Bookstein, N. V. Kapp, C. W. Edwards, & H. W. Wyckoff: Bacillus subtilis alkaline

phosphatase III and phosphatase IV - cloning, sequencing and comparisons of deduced amino acid sequence with *Escherichia coli* alkaline phosphatase 3-dimensional structure. *J Biol Chem* 266, 1077-1084 (1991)

6. E. Garattini, J. C. Hua, & S. Udenfriend: Cloning and sequencing of bovine kidney alkaline phosphatase cDNA. *Gene* 59, 41-46 (1987)

17. T. Rose, P. Glaser, W. K. Surewicz, H. H. Mantsch, J. Reinstein, B. Le, A. M. Gilles, & O. Barzu: Structural and functional consequences of amino acid substitutions in the second conserved loop of *Escherichia coli* adenylate kinase. *J Biol Chem* 26, 23654-2365 (1991)

18. P. D. Karp, M. Riley, S. M. Paley, A. Pellegrini-Toole, & M. Krummenacker: EcoCyc: Encyclopedia of *Escherichia coli* genes and metabolism. (<http://ecocyc.pangeasystems.com:1555/NEW-IMAGE?type=ENZYME&object=ADENYL-KIN-MONOMER>) *Nucleic Acids Res* 2, 50-3 (1998)

19. C. W. Muller, & G. E. Schulz: Structure of the complex between adenylate kinase from *Escherichia coli* and the inhibitor Ap5A refined at 1.9 Å resolution. A model for a catalytic transition state. (PDB entry 1ake) *J Mol Biol* 22, 159-177 (1992)

20. M. B. Berry, B. Meador, T. Bilderback, P. Liang, M. Glaser, & G. N. Phillips: The closed conformation of a highly flexible protein: the structure of *E. coli* adenylate kinase with bound AMP and AMPNP. *To be published PDB, 1ank* (1994)

21. R. A. Poorman, A. Randolph, R. G. Kemp, & R. L. Henrikson: Evolution of phosphofructokinase--gene duplication and creation of new effector sites. *Nature* 30, 467-9 (1984)

22. J. Heinisch, R. G. Ritzel, B. o. von, A. Aguilera, R. Rodicio, & F. K. Zimmermann: The phosphofructokinase genes of yeast evolved from two duplication events. *Gene* 7, 309-21 (1989)

23. X. Y. Huang, L. A. Barrios, K. h. von, S. Honda, D. G. Albertson, & R. M. Hecht: Genomic organization of the glyceraldehyde- 3- phosphate dehydrogenase gene family of *Caenorhabditis elegans*. *J Mol Biol* 206, 411-424 (1989)

24. A. Dugaiczky, J. A. Haron, E. M. Stone, O. E. Dennison, K. N. Rothblum, & R. J. Schwartz: Cloning and sequencing of a deoxyribonucleic acid copy of glyceraldehyde- 3- phosphate dehydrogenase messenger ribonucleic acid isolated from chicken muscle. *Biochemistry* 22, 1605-1613 (1983)

25. J. D. Hillman: Mutant analysis of glyceraldehyde 3-phosphate dehydrogenase in *Escherichia coli* *Biochem J* 179, 99-107 (1979)

26. E. Duee, L. Olivier-Deyris, E. Fanchon, C. Corbier, G. Branlant, & O. Dideberg: Comparison of the structures of wild type and a N313T mutant of *Escherichia coli* glyceraldehyde 3-phosphate dehydrogenases: implication for NAD binding and cooperativity. *To be published PDB, 1gad* (1995)

27. T. Skarzynski, P. C. Moody, & A. J. Wonacott: Structure of holo-*glyceraldehyde-3-phosphate dehydrogenase from *Bacillus Stearothermophilus* at 1.8 angstroms resolution *J Mol Biol*, 171-186 (1987)

Casein kinase 2 bacterial predecessor

28. J. H. Hurley, P. E. Thorsness, V. Ramalingam, N. H. Helmers, D. E. Koshland, & R. M. Stroud: Structure of a bacterial enzyme regulated by phosphorylation, isocitrate dehydrogenase. *Proc Natl Acad Sci U S A* 8, 8635-9 (1989)
29. A. J. Cozzone: Regulation of acetate metabolism by protein phosphorylation in enteric bacteria. *Annu Rev Microbiol* 52, 127-64 (1998)
30. B. L. Stoddard, A. Dean, & D. E. Koshland: The structure of isocitrate dehydrogenase with isocitrate, NADP, and calcium at 2.5 Angstroms resolution: a pseudo-Michaelis ternary complex *Biochemistry* 32, 9310-9325 (1993)
31. M. Dadssi, & A. J. Cozzone: Occurrence of protein phosphorylation in various bacterial species. *Int J Biochem* 2, 493-9 (1990)
32. A. J. Cozzone: Diversity and specificity of protein-phosphorylating systems in bacteria. *Folia Microbiol (Praha)* 4, 165-70 (1997)
33. A. J. Cozzone: ATP-dependent protein kinases in bacteria. *J Cell Biochem* 5, 7-13 (1993)
34. C. C. Zhang: Bacterial signalling involving eukaryotic-type protein kinases. *Mol Microbiol* 2, 9-15 (1996)
35. E. Katchalski-Katzir, I. Shariv, M. Eisenstein, A. A. Friesem, C. Aflalo, & I. A. Vakser: Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. *Proc Natl Acad Sci U S A* 8, 2195-9 (1992)
36. F. Eisenhaber, P. Lijnzaad, P. Argos, & M. Scharf: The double cubic lattice method: efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies *J Comp Chem* 16, 273-284 (1995)
37. B. Lee, & F. M. Richards: The interpretation of protein structures: estimation of static accessibility. *J Mol Biol* 5, 379-400 (1971)
38. M. L. Connolly: The molecular surface package. *J Mol Graph* 1, 139-41 (1993)
39. R. J. Zauhar: SMART: a solvent-accessible triangulated surface generator for molecular graphics and boundary element applications. *J Comput Aided Mol Des* , 149-59 (1995)
40. F. Eisenhaber: Hydrophobic regions on protein surfaces. Derivation of the solvation energy from their area distribution in crystallographic protein structures. *Protein Sci* , 1676-86 (1996)
41. G. Vriend: WHAT IF: A molecular modeling and drug design program *J Mol Graph* 8, 52-56 (1990)
42. S. J. Hubbard, & P. Argos: Evidence on close packing and cavities in proteins *Curr Opinion in Biotech* 6, 375-381 (1995)
43. J. L. Ellis, & G. Kedem: The ray-casting machine. *In Proceedings ICCD-84* 1, 533-538 (1984)
44. D. C. Richardson, & J. S. Richardson: The KINEMAGE: a tool for scientific communication *Trends in Biochem Sci* 19, 135-8 (1994)
45. M. Gerstein, J. Tsai, & M. Levitt: The volume of atoms on the protein surface: calculated from simulation, using Voronoi polyhedra. *J Mol Biol* 249 , 955-966 (1995)
46. V. Subramaniam, N. C. Bergenhem, A. Gafni, & D. G. Steel: Phosphorescence reveals a continued slow annealing of the protein core following reactivation of *Escherichia coli* alkaline phosphatase. *Biochemistry* 3, 1133-6 (1995)
47. S. J. Hubbard, F. Eisenmenger, & J. M. Thornton: Modeling studies of the change in conformation required for cleavage of limited proteolytic sites. *Protein Sci* , 757-68 (1994)
48. R. Jakobi, W. J. Lin, & J. A. Traugh: Modes of regulation of casein kinase 2. *Cell Mol Biol Res* 4, 421-9 (1994)
49. D. W. Litchfield, & B. Luscher: Casein kinase 2 in signal transduction and cell cycle regulation. *Mol Cell Biochem* 127-128, 187-99 (1993)
50. F. Marchiori, F. Meggio, O. Marin, G. Borin, A. Calderan, P. Ruzza, & L. A. Pinna: Synthetic peptide substrates for casein kinase 2. Assessment of minimum structural requirements for phosphorylation. *Biochim Biophys Acta* 97, 332-8 (1988)
51. F. R. Blattner, G. Plunkett, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, & Y. Shao: The complete genome sequence of *Escherichia coli* K-12 (A public version of the database is accessible (<http://cgsc.biology.yale.edu>) *Science* 277, 1453-1462 (1997)

Keywords: Casein kinase 2, *Escherichia Coli*, Protein structure analysis, Molecular surface analysis, Alkaline phosphatase, Adenylate kinase, Phosphofructokinase-1, D-glyceraldehyde-3-phosphate dehydrogenase, Isocitrate dehydrogenase

Send correspondence to: Dr. Ivan Torshin, Laboratory of kinetic and catalysis, Chair of Physical Chemistry, Chemical Department of Moscow State University, Moscow, 119899, Russia, Tel:095-939-4595, E-mail: TIIY@phys.chem.msu.ru, TIIY135@yahoo.com

Received 3/17/99 Accepted 3/25/99