Direct and reversed amino acid sequence pattern analysis: structural reasons for activity of reversed sequence sites and results of kinase site mutagenesis

Ivan TORSHIN[1]

Laboratory of Kinetic and Catalysis, Chemical Department, Moscow State University, Moscow, 119899, Russia

During studies of kinase phosphorylation, not all functional kinase phosphorylation may be found using consensus sequence patterns. This type of phosphorylation is termed 'non-consensus' or 'cryptic' phosphorylation. Results presented here based on molecular dynamics of short peptides show that protein kinases may phosphorylate not only established consensus sequences (reading a sequence from N-terminus to C-terminus) but also reversed consensus sequences (reading from C- to N-terminus). Several protein sequences were analysed and corresponding biochemical data were presented. Similarity of molecular shapes of direct and reversed consensus peptides, and sequence conservation in the regions of reversed sites in the analysed proteins, indicate that at least part of the phosphorylation sites considered as 'cryptic' may be explained in terms of reversed consensus pattern occurrences.

Key words: direct consensus pattern, inverted repetitions, molecular recognition, protein kinase phosphorylation, reversed consensus pattern.

INTRODUCTION

Protein phosphorylation is important for regulation of many cellular processes. The influence of kinase phosphorylation on the functioning of a given protein may be studied using two main groups of biochemical techniques. Direct methods include analysis of phosphorylated peptides and mutagenesis of kinase sites. Using indirect methods, *in vivo* activity of a given protein is registered and the influence of some kinase phosphorylation on a target protein's activity can be investigated using specific kinase activators and inhibitors [1].

The substrate specificity of many eukaryotic kinases is restricted to specific protein subsequences containing Ser/Thr or Tyr residues. The sequence sites of interest found using sequence pattern analysis (for example, with PROSITE patterns [2]) are termed as 'consensus sites'.

When experimentally identified phosphorylated sequence sites could not be ascribed to a kinase type, modification has sometimes been termed 'non-consensus modification'. In the case of mutagenesis experiments this corresponds to the conservation of regulation by the given kinase after mutation/deletion of all consensus sites (extensive mutagenesis of kinase sites).

Protein–protein interaction of kinase with the target molecule may depend on overall surface topology and physicochemical properties of atoms on both molecular surfaces [3] or mostly on the sequence fragments corresponding to the consensus pattern [4]. As many different cellular proteins are phosphorylated, for example, by protein kinase C (PKC), the latter molecular recognition hypothesis of protein–peptide interaction seems to be more appropriate in the case of kinase substrate recognition [5].

Kinase substrate sequence patterns crucial for recognition by definite kinases are mostly relatively short sequence fragments (three to seven residues). Short sequence fragments were extensively studied in order to establish amino acid regularities for usage in protein fold prediction [6]. Many protein sequences were found to have inverted repeats with a length of at least four amino acid residues [7]. Thus some of the kinase sites also may be presented in protein sequences as 'inverted repetitions' or 'reversed sequence patterns'.

The present paper deals with structural determinants of a consensus pattern recognition and possible biological significance of reversed kinase pattern occurrences on the base of published extensive mutagenesis results.

METHODS

Molecular dynamics (MD) of direct and reversed PKC consensus tripeptides

The MD experiments with the direct and reversed consensus sequences were performed using AMMP (Another Molecular Mechanics Program), one of the most efficient molecular mechanics programs available for PC computers [8]. The modelling used all hydrogen atoms, all non-bonding and electrostatic terms without any distance cutoffs; the implicit solvent model was implemented as increased van der Waals radii for polar side-chain atoms [9].

For analysis of conformational properties of examples of direct and reversed PKC consensus sequences (tripeptides SGR, RGS and SVR, RVS, using the amino acid one-letter code), terminal atoms of the peptides were fixed at 8 Å (1 Å = 0.1 nm) distance. Sampling of conformational space [10] for each pair of the peptides was performed starting from the fully extended chains and using ten different sets of initial atom velocities for each peptide with Maxwell's distribution. The sample sets of direct and reversed consensus tripeptide conformations after 50 ps of MD were obtained and then the molecular surface shape and charge distributions of the averaged structures were compared.

Superimposition of structures for averaging was performed by using programs of the TINKER package [11]. Averaging was performed by a program written by myself (from whom details

Abbreviations used: AMMP, Another Molecular Mechanics Program; PKC, protein kinase C; TyrK, tyrosine kinase; MD, molecular dynamics; RMS, root mean square; CFTR, cystic fibrosis transmembrane conductance regulator; VIP(R), (human) vasoactive intestinal peptide (receptor).

[1] e-mail TIY@phys.chem.msu.ru or TIY135@yahoo.com

are available on request). As averaging leads to significant geometry distortions, the obtained average structures were optimized using the AMMP program by energy minimization down to root mean square (RMS) 0.1. Figures with molecular-structure representations were prepared using WebLab Viewer, available from MSI Corp. (San Diego, CA, U.S.A.).

Protein sequences and sequence pattern databases

Sequences of the proteins studied were obtained from the SwissProt database [12], whereas kinase sequence patterns and protein signatures were from PROSITE [2].

Identification of direct and reversed sequence site patterns

Program SCANPS (SCAN ProSite) was elaborated for identification of direct and reversed pattern occurrences. The program takes the sequence in FASTA format, and the output is made into a file in .SIT file format developed by myself and based on the format of PROSITE database entry [2]. For visualization of sequence site placement the program STRDIAG was used.

Site-conservation analysis was performed by scanning of all sequences of a given protein with the pattern under interest using the program SEQPATT (SEQuence PATTern).

RESULTS

Determinants of kinase substrate recognition of direct/reversed consensus sequences on the base of peptide molecular shapes were analysed for PKC consensus peptides. Sequences of proteins with established non-consensus regulation were analysed for the presence of direct and reversed PKC and tyrosine kinase phosphorylation sites. Reversed-sequence-site occurrences that do not overlap with direct consensus sequences and presence ('conservation') of a consensus site in protein sequences from other organisms are considered as indicating a functional value of reversed sites.

Molecular-shape analysis of examples of PKC consensus tripeptides

Results of conformational and surface-shape analysis for direct and reversed consensus PKC peptides are presented in Figures 1 and 2. Comparison of molecular surface properties was performed visually. Molecular surfaces of the peptides are presented at 2.5 Å resolution.

Molecular shapes RVS and SVR peptides

Shape patterns of RVS and SVR tripeptides are triangular (△) or V-like (disregarding side-chain valine atoms, which are possibly of little importance for kinase recognition). Both shapes have positively charged spots of arginine guanidine atoms placed on both sides of the triangle and adjacent negatively charged regions are also presented on both sides (Figure 1, side view). Both shapes are flattened (Figure 1, view from above), and both

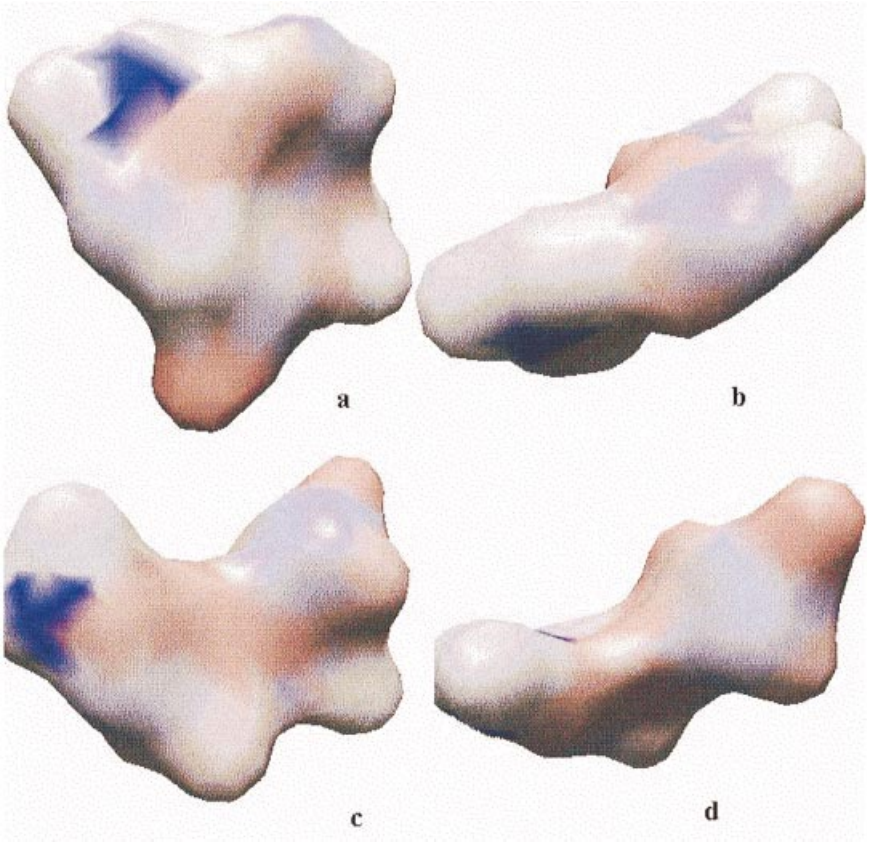

Figure 1 Molecular shapes of the SVR and RVS peptides

Positively charged regions (blue) and negatively charged ones (red) are shown. (**a**) SVR-peptide, side view; (**b**) SVR peptide, view from above; (**c**) RVS peptide, side view; (**d**) RVS peptide, view from above.

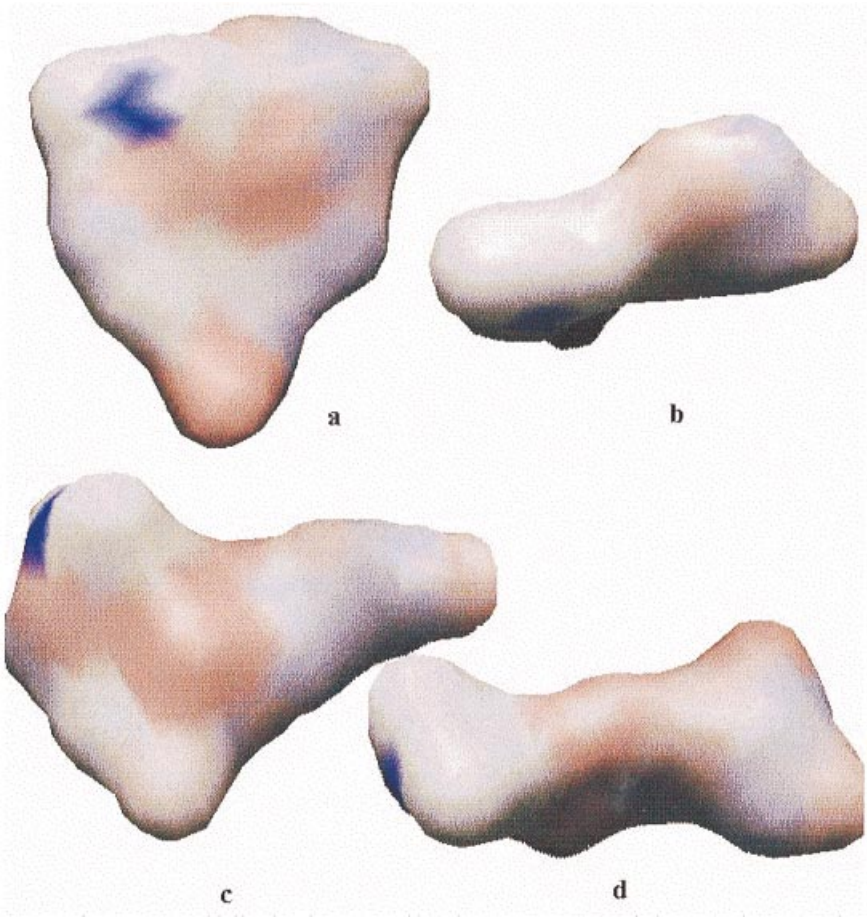

Figure 2 Molecular shapes of the SGR and RGS peptides

(**a**) SGR peptide, side view; (**b**) SGR peptide, view from above; (**c**) RGS-peptide, side view; (**d**) RGS peptide, view from above.

surfaces potentially facing the bottom of the kinase-binding cleft have sequential neutral (backbone), negative (backbone atoms) and positive (guanidine) regions.

Molecular shapes of SGR and RGS peptides

Like SVR/RVS peptides, SGR/RGS peptides also have flattened V-like molecular shapes (the SGR peptide is somewhat more compact). Both guanidine positive spots and negative spots adjacent to guanidine groups are also present. In the RXS peptides (RGS and RVS) the negative spots are formed by main-chain oxygen atoms of the first and second residues, whereas in the SXR peptides (SGR and SVR) the spots that are due to main-chain oxygen and the hydroxy group of the first residue are placed in similar molecular regions. Consequent neutral, negative and positive regions on the surfaces facing the bottom of the cleft are also presented in the SGR and RGS peptides.

The following determinants of kinase recognition common for the direct and reversed PKC pattern peptides thus may be proposed.

1. Flattened triangle (V-like) shape
2. Positive guanidine charge spots on both sides of the triangle and adjacent negative charge spots
3. Surface potentially facing the bottom of the kinase-binding cleft has consequently arranged neutral, negatively and positively charged regions.

Extensive mutagenesis analysis

PKC phosphorylation

PKC preferentially phosphorylates serine or threonine residues that are close to a C-terminal basic residue. The presence of additional basic residues at the N- or C-terminal of the target amino acid enhances the phosphorylation reaction [13]. The elaborated consensus pattern is thus (ST)-X-(RK), where S or T is the phosphorylation site.

Renal sodium/phosphate (Na^+/P_i) co-transporter 2

Two renal Na^+/P_i co-transporters (NPT1 and NPT2) are integral membrane proteins important for the reabsorption of phosphate by the kidney cells [14].

Renal brush-border membrane Na^+/P_i co-transport activity is inhibited by hormonal mechanisms, particularly those involving

Table 1 Direct and reversed PKC consensus sites [(ST)-X-(RK)] in the renal Na^+-dependent phosphate-transport proteins from rat

n	NPT1				NPT2				Comment (for NPT2)
1	+1	15	17	SFR	+1	5	7	SER	
2	+1	55	57	SNK	+1	91	93	SQK	Mirror repeat
3	+1	231	233	SEK	+1	295	297	SLR	Extracellular domain
4	+1	246	248	SGR	+1	460	462	SPR	Extracellular domain
5	+1	325	327	SRK	+1	506	508	TAK	Mirror repeat
6	+1	462	464	TTR	+1	623	625	SPR	
7	−1	167	169	RLT	−1	89	91	RLS	Mirror repeat
8	−1	248	250	RQS	−1	464	466	KLS	Extracellular domain
9	−1	458	460	KET	−1	504	506	KRT	Mirror repeat
10	−1	461	463	KTT	−1	593	595	RAT	

the activation of PKC [15]. The sequence of the NPT2 gene from rat with all PKC consensus sites removed was expressed [16]. PKC-mediated inhibition was not prevented by the removal of the PKC consensus sites. The involvement of 'cryptic' phosphorylation sites has been discussed [16].

Four reversed PKC sites were identified in the sequence of NPT2. None of the potential PKC sites occur in the regions of transmembrane helices. Results are presented in Table 1 and in Figure 3.

All the direct sites were removed during the experiments described in [16]. Thus reversed PKC sites adjacent to direct ones ['mirror repeats', like PKC (+1, 91–93, SQK) and PKC (−1, 89–91, RLS); '−1' means reversed and '+1' means direct] and having one common residue to be phosphorylated could not be the cause of the non-consensus PKC phosphorylation. Therefore, only two sites may have functional value: PKC (−1, 464–466, KLS) and PKC (−1, 593–595, RAT). As the site PKC (−1, 464–466, KLS) occurs in a potential extracellular domain, only one reversed site may be considered as potentially active, namely PKC (−1, 593–595, RAT). The reversed site PKC (−1, 593–595, RAT) found in the sequence of rat NPT2 also presented in the human NPT2, being placed at the same sequence region [PKC (−1, 595–597, RAT)]. However, this site was not found in the sequence of murine NPT2.

Excitatory amino acid transporter 1

Excitatory amino acid transporter 1 (glial glutamate transporter 1, GLAST-1) transports L-glutamate as well as L- and D-aspartate and acts as symport co-transporting Na^+. The protein is essential for terminating the postsynaptic action of glutamate [17].

Removal of all putative PKC sites of rat brain GLAST-1 by site-directed mutagenesis did not abolish inhibition of glutamate transport by the PKC activator PMA [18]. After stimulation with the inhibitor, radioactive-phosphate-labelled mutant transport protein devoid of all predicted PKC sites was detected. The data suggest that the neurotransmitter transporter activity of the GLAST-1 protein is inhibited by phosphorylation at some non-consensus PKC sites [18]. The results of the sequence-pattern analysis are presented in Figure 4 and in Table 2.

A reversed PKC site is placed in region 1–67 of potential cytoplasmic domain relatively far from other PKC sites, both direct and reversed [site PKC (−1, 24–26); Table 2 and Figure

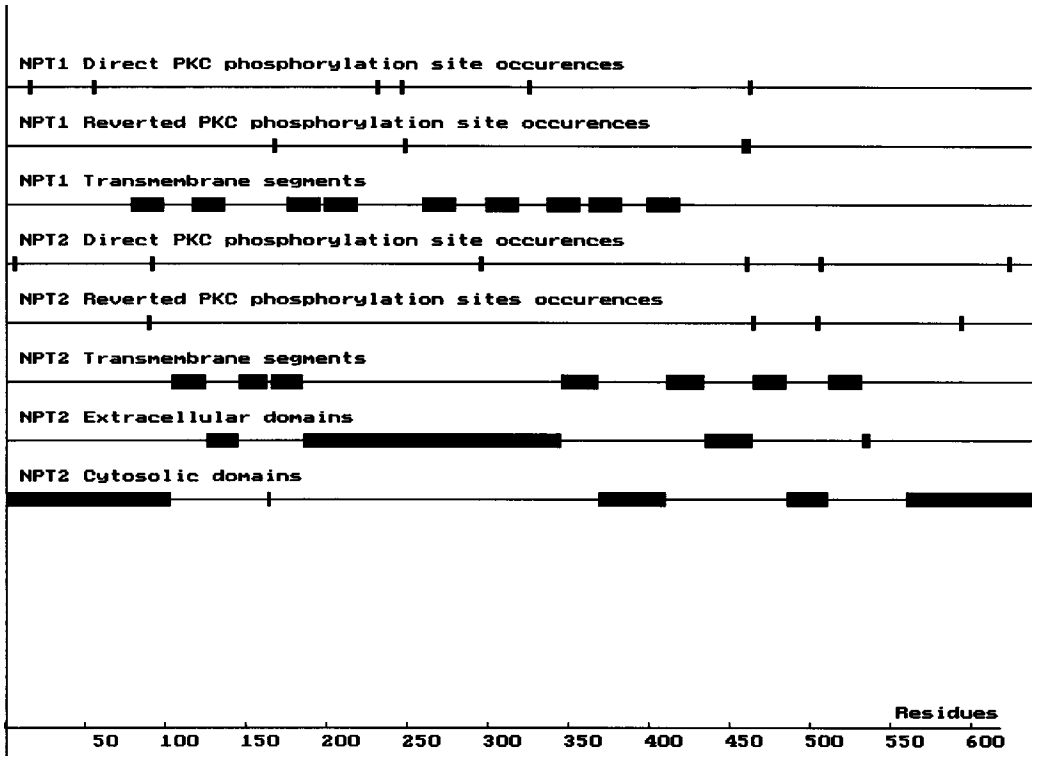

Figure 3 Direct and reversed PKC site occurrences in the sequences of the two renal sodium-dependent phosphate-transport proteins from rat

PKC sites, potential transmembrane segments and domains for both sequences are shown.

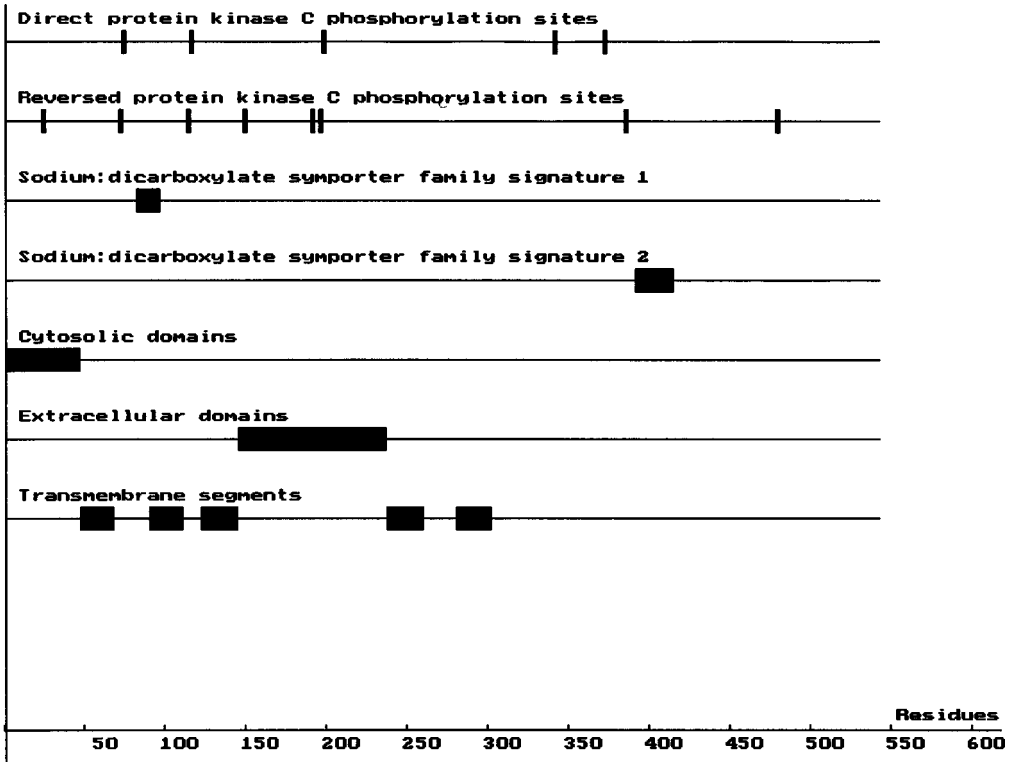

Figure 4 Sequence sites in the rat GLAST-1 protein

PKC direct and reversed sites are shown on different strings.

Table 2 Sequence sites in the mammal GLAST-1 proteins

543R rat				543R mouse				542R calf				542R human			
								+1	67	69	TLR	+1	67	69	TLR
+1	74	76	SYR	+1	74	76	SYR	+1	74	76	SYR	+1	74	76	SYR
+1	116	118	SGK	+1	116	118	SGK	+1	116	118	SGK	+1	116	118	SGK
+1	198	200	SFK	+1	198	200	SFK	+1	198	200	SFK	+1	198	200	SFK
+1	341	343	TRK	+1	341	343	TRK	+1	341	343	TRK	+1	341	343	TRK
+1	372	374	TFK	+1	372	374	TFK	+1	372	374	TFK	+1	372	374	TFK
												+1	532	534	TEK
−1	24	26	KRT	−1	24	26	KRT	−1	24	26	KRT	−1	24	26	KRT*
−1	72	74	KMS	−1	72	74	KMS	−1	72	74	RMS	−1	72	74	RMS
−1	114	116	KAS	−1	114	116	KAS	−1	114	116	KAS	−1	114	116	KAS
−1	149	151	KGT	−1	149	151	KGT	−1	149	151	KGT	−1	149	151	KGT
−1	191	193	KTS	−1	191	193	KTS								
												−1	163	165	RVT
−1	196	198	KRS	−1	196	198	KRS	−1	196	198	KRS	−1	196	198	KRS
								−1	226	228	RIT	−1	226	228	RIT
−1	385	387	RIT	−1	385	387	RIT	−1	385	387	RVT	−1	385	387	RVT*
−1	479	481	RTT	−1	479	481	RTT	−1	479	481	RTT	−1	479	481	RTT*

4]. Two other reversed PKC sites, namely PKC (−1, 385–387) and PKC (−1, 479–481), are placed in the region 380–530, also far from direct PKC sites. The reversed PKC sites that overlap with the direct ones (+1, 74–76, SYR and −1, 72–74, RMS; +1, 116–118, SGK and −1, 114–116, KAS; +1, 198–200, SFK and −1, 196–198, KRS) contain the same serine residues as the deleted adjacent sites.

For investigation of PKC site conservation in sequences from other organisms, the sequence set was scanned using the SEQPATT program. The results are presented in Table 2.

Three reversed PKC sites placed relatively far from direct PKC sites PKC (−1, 24–26), PKC (−1, 385–387) and PKC (−1, 479–481) (marked with an asterisk in Table 2) are presented in all available sequences of GLAST-1 protein. These reversed sites are

Table 3 Conservation of the reversed tyrosine kinase site in CFTR proteins from several organisms (CFTR-SQUAC-*Squalus acanthias*)

Sequence SwissPot Identifier	Protein sequence length	Site's position in sequence	Sequence of the site
CFTR-BOVIN	1481	275 283	YCWEEAMEK
CFTR-HUMAN	1480	275 283	YCWEEAMEK
CFTR-RABIT	1450	245 253	YCWEEAMEK
CFTR-SHEEP	1481	275 283	YCWEEAMEK
CFTR-SQUAC	1492	276 284	YCWEDAMEK

supposed to be some of the PKC non-consensus sites mentioned in [18].

Tyrosine kinase phosphorylation

The results of molecular modelling presented here suggest that both direct and reversed PKC consensus sites may have functional activity. They also raise the questions of whether reversed sites occur only as PKC sites, and, if not, whether reversed sites of other types have some functional value.

Registration of a protein's regulation by a kinase when consensus sites are absent in the sequence of the protein may be considered as indirect evidence of non-consensus phosphorylation. This was observed in the case of phosphorylation with tyrosine kinases.

The consensus patterns for the tyrosine kinase phosphorylation site proposed in PROSITE documentation (PDOC00007) are (RK)-X_2-(DE)-X_3-Y and/or (RK)-X_3-(DE)-X_2-Y, where Y is the phosphorylation site [19]. In the PROSITE database entry the pattern (RK)-$X_{2,3}$-(DE)-$X_{2,3}$-Y is proposed.

Tyrosine kinase phosphorylation of human cystic fibrosis transmembrane conductance regulator (CFTR)

Cystic fibrosis, a fatal hereditary disorder, is caused by mutations in the gene for the cystic fibrosis transmembrane conductance regulator (CFTR) involved in ATP-dependent chloride transport across the membranes of heart cells [20].

In [21] the role of the tyrosine kinase in the gating of the human CFTR chloride channel was investigated. The experiments were performed using the patch-clamp technique and the channel's open probability was found to be dependent on tyrosine phosphorylation and dephosphorylation by membrane-associated tyrosine phosphatases [21]. Two reversed sites of potential tyrosine kinase phosphorylation are placed closer to the N-terminal end of the protein: TyrK (−1, 89–95, YLGEVTK) and TyrK (−1, 275–283, YCWEEAMEK). As the first of these sites is placed in the region of a potential transmembrane helix (positions 81 and 103), it possibly would have no functional value. The second site is situated in a potential intracellular domain. The data on the conservation of the second TyrK site in sequences from other sources are given in Table 3.

The second site is presented in the sequences of proteins from five different organisms in the similar sequence regions. The sequence of the site is almost fully conserved [excluding E → D substitution in the protein from the spiny dogfish (*Squalus*

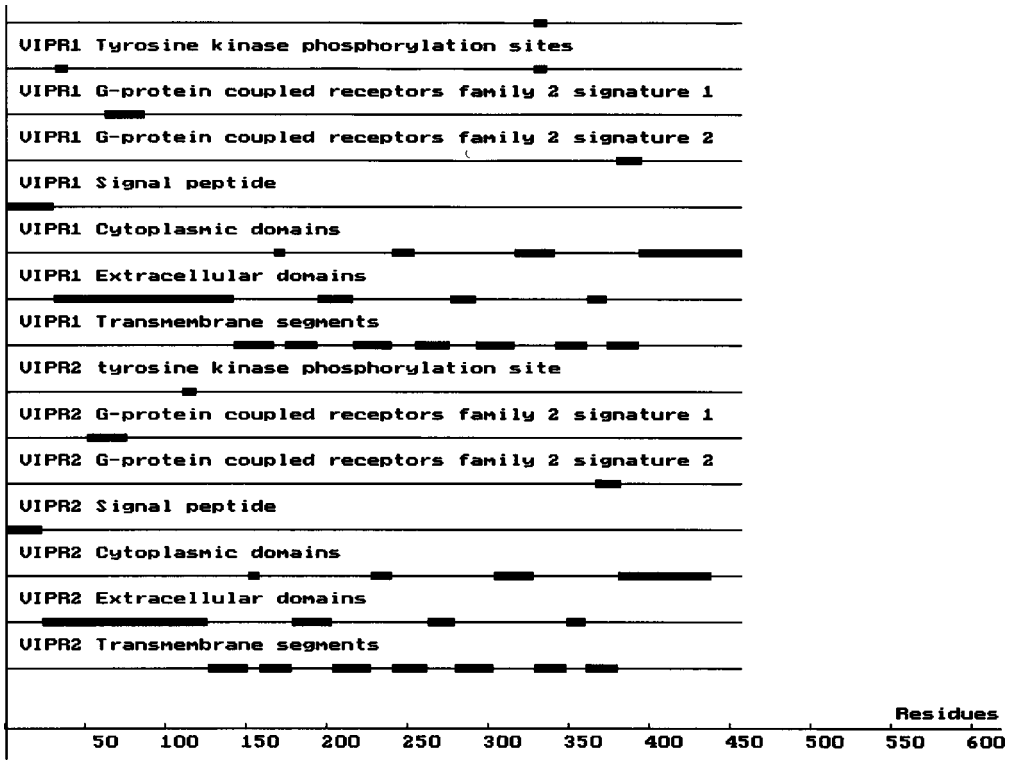

Figure 5 Potential domain structure and sequence sites of VIPR1 and VIPR2

The VIPR2 sequence is shorter than the VIPR1 sequence (438 and 457 residues respectively) owing to 'deletions' in the first extracellular domain corresponding to the region of one of the potential direct tyrosine kinase sites in VIPR1.

acanthias)]. The site was not found, however, in the sequences of murine CFTR and CFTR from *Xenopus laevis*. Thus this conserved reversed tyrosine kinase phosphorylation site may be one of the possible explanations of the registered CFTR regulation by tyrosine kinase/phosphatases.

Human vasoactive intestinal peptide receptors (VIPRs) and potential phosphorylation by tyrosine kinase of human VIPR proteins

Vasoactive intestinal peptide (VIP) is a neuroendocrine mediator distinguished by its capacity to directly affect both B and T cells. Distinct subsets of neural, respiratory and gastrointestinal cells bear specific high-affinity receptors for VIP (VIPRs). VIP and related agonists influence T-cell migration. Structurally different G-protein-associated receptors [VIPR1 (P32241) and VIPR2 (P41587)] mediate these chemotactic effects [22]. It was found that peripheral-blood lymphocyte chemotaxis to VIP and special substrates for each VIPR form was decreased by inhibition of tyrosine kinase. It may be supposed that both VIPR sequences have some tyrosine kinase sites [22].

In the sequence of human VIPR1 protein there are two potential tyrosine kinase site-directed sequences (+1, 32, 39, RLQEECDY; +1, 329, 336, RKSDSSPY), while in the sequence of the VIPR2 protein only one reversed tyrosine kinase site is presented (−1, 111, 119, YSDPEDESK) (Figure 5). The reversed site TyrK (−1, 111, 119, YSDPEDESK) found in human sequence is also presented in sequences of murine and rat VIPR2 proteins (TyrK(−1, 110, 118, YNDPEDESK) and TyrK(−1, 110, 118, YNDPEDESK) respectively).

As both the first direct TyrK site in the VIPR1 sequence and the reversed one in the VIPR2 sequence are placed in the sequence region of a potential extracellular domain (Figure 5), they may appear to be non-functional sites. However, according to recent data [23], the tyrosine residues of a protein can be phosphorylated while being present in an extracellular domain.

Dickens et al. [24] found that all tyrosine residues of the peptide RRDIYETDYYRK were rapidly phosphorylated. Therefore, the unusual pattern (RK)-X_3-(DE)-X_3-Y, the reversed occurrence of which was found in the CFTR and VIPR sequences [CFTR TyrK (−1, 275, 283, YCWEEAMEK), VIPR2 TyrK (−1, 111, 119, YSDPEDESK)] corresponds to at least one functional sequence (RRDIYETDYYRK).

Thus the opposite action of VIPR receptors registered in many *in vivo* experiments may be explained by the presence of a reversed tyrosine kinase site.

DISCUSSION

Conformational sampling of short peptides: molecular mechanics modelling of the short PKC kinase direct and reversed consensus sequences and analysis of the molecular surface shapes

The stability of a molecular complex depends on the shape complementarity of the interacting molecular surfaces. Some adjustment of molecular shapes may be required for interacting surfaces to fit together [25]. Charge distributions that define electrostatic potential of the interacting surfaces are also important. Despite similarities in topology of a binding site, differences in electrostatic potential can account for the experimentally observed differences in substrate recognition [26]. Both complementarity of shapes and complementarity of charge distributions are determined by molecule conformations and types of atoms in the given molecules. Molecules having close molecular shapes with similar charge distributions under given conditions may be both recognized and processed by the same enzyme molecule.

In order to assess in detail the possible properties of the peptides it would be necessary to perform molecular-mechanics modelling using the structure of the PKC kinase active site. However, the structure of the PKC catalytic domain has not been determined. Thus a comparison of peptides' equilibrium molecular shapes was used to establish possible determinants of kinase substrate recognition.

The restrained MD structures are superior to those obtained from distance geometry as regards local backbone conformation and side-chain positions [27]. As equilibrium conformation attained after MD is more reliable, MD simulations were used to obtain conformations of the peptides.

Choice of model peptides

The PKC consensus [(ST)-X-(RK)] contains a position that may feature any residue, and there are a number of possibilities for substituting the X position for another amino acid. The effects of definite substitutions may be considered prior to performing MD simulations. Substitution with an acid residue would inevitably lead to the formation of a hydrogen-bonding interaction between two adjacent basic and acid residues and thus a compact stable conformation may be conceived. In the case of basic residues the repulsion between two adjacent basic residues would lead to their opposite placement and extended conformation of the peptide. In the case of a neutral residue, however, no preliminary predictions of the kind could be made; therefore modelling was performed using sequences with neutral X residues (glycine and valine). In addition, it has been demonstrated [28] that PKC *in vivo* usually exhibits a preference for peptides with neutral X residues.

Solvent model and procedure for conformational sampling

In the case of short peptides the percentage of inaccessible molecular surface due to folding is very small, thus the use of explicit and implicit solvent models would give almost identical results. MD of small proteins with the solvent model just leads to an 11 % overall increase in the RMS fluctuations in comparison with 'vacuum' modelling [29].

Initial distribution of velocities is rather important for obtaining results consistent with X-ray crystallography or NMR data [10]. The structures averaged over the several MD trajectories are significantly closer to the X-ray structures than any of the individual trajectory average structures. As after 50 ps of MD the change in conformation is relatively small (RMS 0.1), the sets of direct and reversed consensus tripeptide conformations (trajectories) were obtained after 50 ps of MD.

Fixation of terminal peptide atoms

Fixation of terminal peptide atoms influences the results of the peptides' MD [30]. Consensus sequence sites are situated mostly inside a polypeptide chain and rarely on the ends; therefore immobilization of the two terminal peptide atoms is necessary. The main problem here is the choice of distance between the fixed terminal atoms: the main-chain N atom at the N-terminus and the main-chain C atom at the C-terminus. On the one hand, in the work of Zaliani [31] it was shown that peptide of a PKC pseudosubstrate binds to the kinase in significantly extended conformation. On the other hand, PKC substrate peptides tend to form helical-like conformations, being in solution in an unbound state or existing predominantly in random form [32]. The model distance between terminal atoms of a tripeptide α-helix is 7 Å; for a tripeptide β-strand the distance is 9 Å. The distance of 8 Å used in the present work was chosen as a

compromise between the fully extended conformation of the β-strand and the compact conformation of the α-helix.

Molecular-shape analysis

The modelling results show that the example peptides of the direct and reversed PKC consensus sequences form comparable recognition shapes under similar simulation conditions (101325 Pa, 300 K, implicit solvent model), which suggests biofunctional significance of the reversed PKC consensus.

Extensive mutagenesis analysis

One of the particular problems in this kind of experiment is related to the use of the kinase inhibitors *in vivo*. For example, a well-known inhibitor of tyrosine kinase, genistein, was used for analysis of the kinase influence on the CFTR channel. As inhibition was observed using this technique, it was proposed that tyrosine kinase influences the channel [33]. However, it was then found that the inhibitor directly interacts with CFTR [34]. The controversy was solved using cell-attached and excised patch clamps [21]. The channel's probability of being open was found to be dependent on tyrosine kinase dephosphorylation without the use of kinase inhibitors/activators.

Direct and reversed pattern analysis proposes a strategy for planning extensive mutagenesis experiments. Taking NPT2 as an example (Table 1): there are six direct and four reversed consensus sites, seemingly ten sites to mutate. However, two of them represent mirror repeats, three of the PKC sites were identified potentially as non-functional due to their presence on extracellular domains, and there are three separate sites. Thus out of ten potential sites, only serine/threonine in five sites are to be mutated/deleted, at least at the first stages of analysis.

PKC substrate specificity and direct/reversed PKC phosphorylation sequence patterns

An oriented peptide-library approach was used to determine the optimal substrate selectivity of PKC family members [35]. The results indicated that all PKCs 'prefer' basic residues at positions -6, -4 and -2 to the serine/threonine. Conventional PKCs also preferred basic residues at $+2$, $+3$ and $+4$, whereas novel PKCs and atypical PKCs preferred hydrophobic residues at these positions [35]. In terms of the present paper a peptide with a basic residue placed at position -2 to the serine/threonine corresponds to the reversed consensus pattern, whereas peptide with a basic residue at $+2$ corresponds to a direct consensus. Both were identified as PKC substrates [35].

Conclusions

Molecular-shape analysis of the example peptides has shown overall shape similarity of direct and reversed PKC consensus sequences. Reversed pattern analysis proposes additional potential sites of phosphorylation and may be useful for the analysis of phosphorylated peptides and analysis of kinase site mutagenesis.

Possible biological significance of reversed sequence site occurrences was assessed using data on sequence conservation and domain structure of the proteins. Another important question to consider is biological validity of reversed occurrences of consensus types other than the kinase consensus (for example, some ligand-binding sequences).

Thus the present paper shows that some of the non-consensus phosphorylation sites may correspond to reversed kinase sites and proposes structural reasons for the reversed sites' functionality.

I thank D. Lucas and H. Kendall for valuable recommendations during preparation of the paper and for looking at the paper prior to its submission. I thank Professor R. Harrison for his advice in using program AMMP for molecular mechanics simulations, and for a presubmission review of the paper.

REFERENCES

1 Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. and Watson, J. D. (1992) in Molecular Biology of the Cell, 3rd edn. (Robertson, M., ed.), pp. 551–560, Garland Publishing Inc., New York
2 Bairoch, A. (1992) Nucleic Acids Res. **2**, 2013–2018
3 Katchalski-Katzir, E., Shariv, I., Eisenstein, M., Friesem, A. A., Aflalo, C. and Vakser, I. A. (1992) Proc. Natl. Acad. Sci. U.S.A. **89**, 2195–2199
4 Saier, M. H. (1994) Res. Microbiol. **14**, 647–650
5 Kennelly, P. J. and Krebs, E. G. (1991) J. Biol. Chem. **26**, 15555–15558
6 Biro, J. (1981) Med. Hypotheses **7**, 995–1007
7 Kolaskar, A. S. and Samuel, S. L. (1991) Protein Sequences Data Anal. **4**, 105–110
8 Harrison, R. W., Chatterjee, D. and Weber, I. T. (1995) Proteins **2**, 463–471
9 Harrison, R. W. (1993) J. Comput. Chem. **14**, 1112–1122
10 Caves, L. S., Evanseck, J. D. and Karplus, M. (1998) Protein Sci. **7**, 649–666
11 Ponder, J. W. and Richards, F. M. (1987) J. Comput. Chem. **8**, 1016–1024
12 Bairoch, A. and Apweiler, R. (1999) Nucleic Acids Res. **2**, 49–54
13 Kishimoto, A., Nishiyama, K., Nakanishi, H., Uratsuji, Y., Nomura, H., Takeyama, Y. and Nishizuka, Y. (1985) J. Biol. Chem. **26**, 12492–12499
14 Li, H. and Xie, Z. (1995) Cell. Mol. Biol. Res. **41**, 451–460
15 Magagnin, S., Werner, A., Markovich, D., Sorribas, V., Stange, G., Biber, J. and Murer, H. (1993) Proc. Natl. Acad. Sci. U.S.A. **90**, 5979–5983
16 Hayes, G., Busch, A. E., Lang, F., Biber, J. and Murer, H. (1995) Pfleugers Arch. **43**, 819–824
17 Arriza, J. L., Fairman, W. A., Wendy, A., Wadiche, J. I., Murdoch, G. H., Kavanaugh, M. P. and Amara, S. G. (1994) J. Neurosci. **14**, 5559–5569
18 Conradt, M. and Stoffel, W. (1997) J. Neurochem. **6**, 1244–1251
19 Cooper, J. A., Esch, F. S., Taylor, S. S. and Hunter, T. (1984) J. Biol. Chem. **25**, 7835–7841
20 Yoshimura, K. and Anzai, C. (1996) Nippon Rinsho (Japanese Clinical Medicine) **54**, 825–833
21 Fischer, H. and Machen, T. E. (1996) Biophys. J. **71**, 3073–3082
22 Schratzberger, P., Geiseler, A., Dunzendorfer, S., Reinisch, N., Kahler, C. M. and Wiedermann, C. J. (1998) J. Neuroimmunol. **8**, 73–81
23 Hou, J., McKeehan, K., Kan, M., Carr, S. A., Huddleston, M. J., Crabb, J. W. and McKeehan, W. L. (1993) Protein Sci. **2**, 86–92
24 Dickens, M., Tavaré, J. M., Clack, B., Ellis, L. and Denton, R. M. (1991) Biochem. Biophys. Res. Commun. **17**, 772–778
25 Whitesides, G. M., Mathias, J. P. and Seto, C. T. (1991) Science **25**, 1312–1319
26 Bravi, G., Legname, G. and Chan, A. W. (1995) J. Mol. Graphics **1**, 83–88
27 Clore, G. M., Nilges, M., Brunger, A. T., Karplus, M. and Gronenborn, A. M. (1987) FEBS Lett. **21**, 269–277
28 Woodgett, J. R., Gould, K. L. and Hunter, T. (1986) Eur. J. Biochem. **16**, 177–184
29 Brooks, C. L. and Karplus, M. (1989) J. Mol. Biol. **20**, 159–181
30 Mao, B., Maggiora, G. M. and Chou, K. C. (1991) Biopolymers **3**, 1077–1086
31 Zaliani, A., Pinori, M., Ball, H. L., DiGregorio, G., Cremonesi, P. and Mascagni, P. (1998) Protein Eng. **1**, 803–810
32 Chang, D. K., Chien, W. J. and Arunkumar, A. I. (1997) Biophys. J. **72**, 554–566
33 Illek, B., Fischer, H., Santos, G. F., Widdicombe, J. H., Machen, T. E. and Reenstra, W. W. (1995) Am. J. Physiol. **268**, 886–893
34 French, P. J., Bijman, J., Bot, A. G., Boomaars, W. E., Scholte, B. J. and de Jonge, H. R. (1997) Am. J. Physiol. **273**, 747–753
35 Nishikawa, K., Toker, A., Johannes, F. J., Songyang, Z. and Cantley, L. C. (1997) J. Biol. Chem. **272**, 952–960

Received 12 April 1999/5 October 1999; accepted 18 November 1999