Ukrainica Bioorganica Acta 1 (2005) 20—24

www.bioorganica.org.ua

Comparative analysis of the similarity between primary sequences of receptor CD4 and protein kinases

Yu. Ya. Kit

Institute of Cell Biology, National Academy of Sciences of Ukraine
14/16 Dragomanov Str., 79005, Lviv, Ukraine

Summary. The protein kinase activity of a recombinant extracellular fragment of the CD4 receptor (rsCD4) has been demonstrated earlier. The catalytic domain of this protein is not yet known. Using the library of amino acid sequences of different protein kinases (KinBase) and the search program Kinom Blast Server the author has realized the investigation of homology levels between the CD4 and other known protein kinases. The comparative analysis carried out shows the CD4 to contain a sequence highly similar to some motifs of calcium/calmodulin-regulated kinases; it is located on the N-terminal part of the rsCD4 molecule. This sequence consists of ca. 150 amino acids and includes an IgV region and the main part of the first Ig-like domain of CD4. The phosphorylation site common for the protein kinase C family has been also detected within this sequence. We suggest it participates in the formation of the CD4 catalytic domain.

Key words: CD4 receptor, protein kinases, comparative analysis.

Introduction. The CD4, a surface transmembrane glycoprotein of T-helper cells, participates in several cellular processes and interactions concerning the T cell receptor (TCR)/CD3 mediated recognition of complexes of the processed MHC antigen (major histocompatibility complex) class II on the surface of B-lymphocytes and other antigen-presenting cells. This leads to stimulation of lymphocytes and induction of humoral immune responses [1—3]. The interaction of CD4 accessory molecules on T cell with MHC class II antigens on B cells is also essential for T cell activation [4—6]. The CD4 is also a major receptor protein for the human immunodeficiency virus (HIV) envelope glycoprotein gp120 [7].

We earlier demonstrated the extracellular rsCD4 domain to possess a detectable protein kinase activity [8]. In spite of our extensive attempts, we have failed to find similar regions for primary sequences of the extracellular rsCD4 domain and various protein kinases available from the GenBank, with the only exception of a weak similarity between this domain and myosin light chain kinase sequence [8]. Nowadays, when the human genome screening has been finished, there is a new opportunity for a more detailed study of this question. In this paper we describe new results obtained as a result of similarities study between the full-length primary sequences of the CD4 and various protein kinase sequences.

Methods. The rsCD4 primary sequence has been taken from the Swiss-Prot database (ACC AAH25782). To analyze the levels of sequence homology of protein kinase genes, the author has used the KinBase data as well as the Kinom Blast Server (www.kinase.com). The CD4 domain organization has been investigated using two databases — SMART (smart.embl-heidelberg.de) and Pfam (pfam.wustl.edu). To study the CD4 motifs possessing affinity to protein kinase domains or able to be phosphorylat-

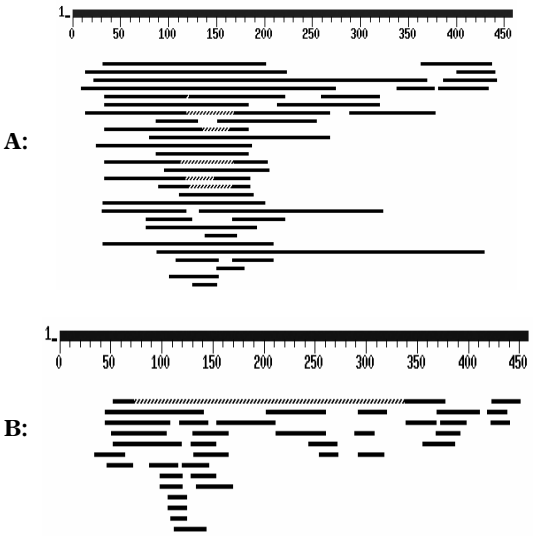
*** Corresponding author.**
Tel./fax: +38322-720087
E-mail address:
kit@biochem.lviv.ua

ed, the author has used the Scansite database and the Motif Scan program, respectively (scansite.mit.edu).

Results and Discussion. Protein kinases are enzymes that covalently attach phosphate to the side chains of proteins containing serine, threonine, or tyrosine. The process of protein phosphorylation can control their enzymatic activity, their interactions with other proteins and molecules, and determine their intracellular localization. The deciphering of the human genome sequence permits now to identify almost all human protein kinases. The protein kinase superfamily contains over a thousand members, while 518 of them are present in humans [9]. On the base of the published human genome sequences, the researchers of the SUGIN Inc. (San Francisco, USA) have obtained a new database including the analyses of protein kinases from human and model organisms (yeast, worms, and fly) [9]. This database allows the comparison of primary sequences of proteins studied and kinase domain peptide sequences of four kingdoms (yeast, worm, fly, and human).

The rsCD4 extracellular fragment has been earlier shown to posses a protein kinase activity [8]. The catalytic domain of this protein is still unknown. Later we have tried to define such domain by means of comparative analysis of similarities between this fragment primary sequence and sequences of different protein kinases. The Kinom Blast Server has been used for this aim. The analysis shows the CD4 to contain some conservative primary sequences being present at least in 101 protein kinases (Fig. 1). Among them, 7 protein kinases possess a sufficient level of homology to the CD4 (E<1). The enzymes with the highest similarity levels (unc-22, F12F3.2, Obscn, TTN) belong to calcium/calmodulin-regulated kinases and related kinase families. They are able to phosphorylate serine and threonine residues of some proteins [10]. The main feature of all these enzymes is the presence of immunoglobulin-like domains that may be involved in protein-protein and protein-ligand interactions [11]. The protein kinase unc-22 of nematode *C.elegans* has been found to possess the highest similarity to the CD4. Like the titin, a giant muscle kinase, this enzyme is known to contain unique catalytic domains [12].

A very important property being always taken into account for kinases classification is the structure of their catalytic domains. These domains are mostly conservative and contain ca. 250—300 amino acids [13]. The comparative analysis between the CD4 primary structure and sequences of protein kinase catalytic domains allows matching of 38 protein kinases



Sequences producing significant alignments

name	(group/family)	core (bits)	E (Value)
unc-22	(CAMK/MLCK)	39	5e-04
F12F3.2	(CAMK/MLCK)	38	0.001
Obscn	(CAMK/Trio)	36	0.005
TTN	(CAMK/MLCK)	35	0.008
ver-3	(TK/KIN16)	33	0.039
KDR	(TK/VEGFR)	32	0.11
pqn-25	(Other/NEK)	31	0.15
C08F8.6	(CK1/Worm6)	28	0.44
ZC3/MINK	(STE/STE20/MSN)	24	4.9
ird5	(Other/IKK)	24	6.4
C45G9.1	(CK1/TTBKL)	24	6.4
ZC2/TNIK	(STE/STE20/MSN)	24	6.4
DUN1	(CAMK/CAMK)	23	11
R03D7.5	(CMGC/GSK)	22	19

Fig. 1. The level of similarities between CD4 and protein kinases. (0—450) — a primary CD4 sequence. The lines below show the protein kinase sequences possessing similarities to CD4. A — protein kinase peptide sequence database, B — protein kinase domain database.

demonstrating some homology to this receptor (Fig. 1b). Among them, only the protein kinase C08F8.6 of the *C.elegans* possesses a sufficient homology to the rsCD4 (E = 0.44). This enzyme is related to casein kinases family able to phosphorylate some proteins on serine and threonine residues [14]. It is necessary to note that the soluble recombinant CD4 can phosphorylate casein on its serine and threonine residues; so it can be considered as a member of the casein kinases family [8]. The data obtained show the CD4 sequence to contain a region with the highest similarity to protein kinases (Fig. 1). This region is located on the N-terminal motif of this structure and consists of 150 amino acids. Earlier it has been detected the histidine to take part in the act of phosphorylation [8]. According to the CD4 primary structure, this receptor contains 4 histidines at positions 9, 132, 181, and 309. This implies the catalytic domain to be most probably located on the CD4 N-terminal region and to include His132 and/or His181.

The CD4 is known to belong to a superfamily of immunoglobulin-like proteins [15, 16]. Consequently, it must contain at least one immunoglobulin-like domain. The analysis of CD4 domain organization shows this structure to contain an immunoglobulin V-type domain (IgV) and two immunoglobulin-like domains (Ig-like 1 and Ig-like 2). They are located on the N-terminal region of the rsCD4 molecule and consist of 77 and 51 amino acids, respectively (Fig. 2).

The rsCD4 is known to participate in different protein-protein interactions; it can be phosphorylated by some protein kinases [16, 17].

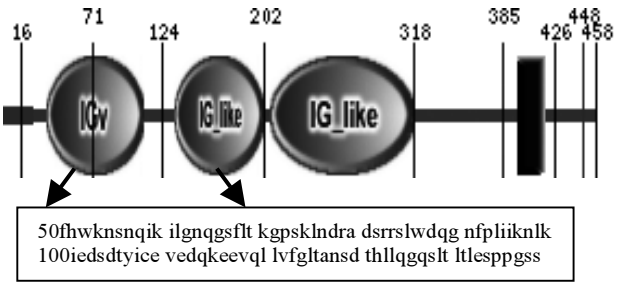


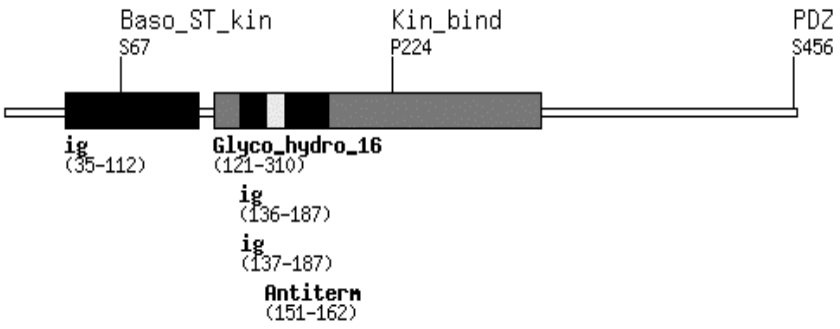
Fig. 2. The SMART diagram represents domains structure of the CD4 receptor. Amino acids position is indicated with vertical lines from N- to C-terminus of the molecule. The vertical bar shows the transmembrane segment position. Arrows determine the sequence with the highest similarity to protein kinases.

Therefore, it was very important to find the rsCD4 binding motifs and its phosphorylation sites, this research has becoming a part of our work.

A research group of the Massachusetts Institute of Technology (Cambridge, USA) has recently obtained a new database and special program permitting to search protein sequences which are likely to be phosphorylated by specific protein kinases or motifs are able to bind to domains [18]. This program is connected with the Pfam database enabling also the investigation of domains organization. These resources have been used for searches of motifs and phosphorylations sites on the rsCD4 molecule. The phosphorylation site for the protein kinases C family has been here found (Fig. 3). This site was detected within the Ig-like 1 domain; it is very probable this phosphorylation site to play a certain role in the functional activity of this domain. Two motifs with affinity to some protein kinases and to others proteins with regulatory functions have been also detected. A kinase-binding motif (Kin_bind) possessing the affinity to Erk 1 protein kinase has been found in the middle of the CD4 molecule. The Erk 1 kinase is known to participate in the processes regulating the cell proliferation, differentiation, motility, and survival [18]. Another motif able to bind the PDZ-domains has been found on the CD4 C-terminus. The PDZ (called also DHR or GLGF) domains are found in diverse membrane-associated proteins including members of the MAGUK family of guanylate kinase homologues, several protein phosphatases and kinases, neuronal nitric oxide synthase, and several dystrophin-associated proteins, collectively known as syntrophins [19].

Using the Pfam database, others kinds of domains have been detected as well. The sequences similar to glycosyl hydrolases and antiterminal domains have been found within the Ig-like 1 domain (Fig. 3). Some properties of these similar domains are given in the Table. These data demonstrate that the Ig-like 1 domain contains a region with variable biological activities and it is located nearby the sequence with highest similarity to protein kinases. It could be noted that one of these domains (an antiterminal region) possesses the

A:



B:

Motifs	Group of motifs	Sequence
PKCzeta	(Baso_ST_kin)	S67KILGNQGSFLTKGPS
Erk1 Binding	(Kin_bind)	P224EQVEFSFPLFTVEK
PDZ class1	(PDZ)	S456HRFQKTCSPI

C:

Description	Model	Length	Score	E(value)
Immunoglobulin domains	Ig)	35–112	41	1.7^{e}-0.9
		136–187	13.5	0.0038
		137–178	13.6	0.14
Glycosyl hydrolases family16	(Glyco_hydro_16)	121–310	-97.8	0.89
Antitermination protein	(Antiterm)	151–162	4.1	0.73

Fig. 3. Determination of a mutual interactions between domains, motifs and sites of phosphorylation within the CD4 receptor using the Scan Motif Scanner program. A — Graphic results of motif scanning, B — Location of motifs and phosphorylation sites, C — Pfam HMM search result.

affinity to nucleic acids [23]. Earlier it has been found that oligonucleotides bind to the CD4 [23]. Taking together, there are good reasons to believe that this antiterminal region participates in CD4-oligonucleotide interactions. Thus, it is not excluded that such kind of interaction can regulate protein kinase activity of the CD4.

Conclusion. Comparative analysis shows the rsCD4 receptor to contain a sequence located on the N-terminal part of its molecule and demonstrating a high similarity to calcium/calmodulin-regulated kinases. This sequence consists of ca. 150 amino acids and includes an immunoglobulin V-type domain and the main part of first immunoglobuline-like domains of CD4. The site of phosphorylation for members of the protein kinase C family has been detected within this sequence as well. We suggest this sequence participates in the formation of a catalytic domain of the CD4.

Table

Domains (ACC)	Properties	Reference
Ig (PF00047)	Immunoglobulin domain. Members of the immunoglobulin superfamily are found in hundreds of proteins of different functions. Examples include antibodies, a giant muscle kinase, titin and receptor tyrosine kinases. Immunoglobulin-like domains may be involved in protein-protein and protein-ligand interactions.	11
Glyco_hydro_16 (PF00722)	O-Glycosyl hydrolases (EC 3.2.1.) are a widespread group of enzymes hydrolysing the glycosidic bond between two or more carbohydrates, or between a carbohydrate and a non-carbohydrate moiety.	21
Antiterm (PF03589)	The anititermination protein is mostly found in bacteriophages, where it modifies host RNA polymerase, which then transcribes through termination sites that would have prevented expression of these genes.	22

Аналіз спорідненості первинних послідовностей рецептора CD4 та протеїнкіназ

Ю. Я. Кіт

Інститут біології клітини НАН України
вул. Драгоманова, 14/16, Львів, 79005, Україна

Резюме. Встановлено, що рекомбінантному позаклітинному фрагменту рецептора CD4 притаманна протеїнкіназна активність. Структуру каталітичного домена CD4 ще досі не визначено. Метою роботи було виявлення амінокислотних послідовностей, здатних утворити протеїнкіназний домен CD4. За допомогою бібліотеки амінокислотних послідовностей протеїнкіназ (KinBase) та пошукової програми (Kinom Blast Server) проаналізовано рівні гомології рецептора CD4 та відомих протеїнкіназ. Встановлено, що найбільша спорідненість до протеїнкіназ властива послідовності, яка містить 150 амінокислот та знаходиться в N-кінцевій частині молекули CD4 і входить до складу двох імуноглобулінових доменів. У складі послідовності виявлено сайт фосфорилювання протеїнкіназами С. Найвищий рівень гомології до цієї послідовності виявлено в родини Са/кальмодулін-залежних протеїнкіназ, здатних фосфорилювати білки по серину й треоніну. Автор припускає, що виявлена послідовність бере участь у формуванні протеїнкіназного домена CD4.

Ключові слова: рецептор CD4, протеїнкінази, порівняльний аналіз.

References

1. *Letvin N.L., King N.W., Reinherz E.L., Hunt R.D., Lane H., Schlossman S.F.* T lymphocyte surface antigens in primates // Eur. J. Immunol. — 1983. — Vol. 13, No 4. — P. 345—3447.

2. *Bierer B.E., Burakoff S.J.* T cell adhesion molecules // FASEB J. — 1988. — Vol. 2, No 10. — P. 2584—2479.

3. *Suzuki S., Kupsch J., Eichmann K., Saizawa M.K.* Biochemical evidence of the physical association of the majority of CD3 delta chains with the accessory/co-receptor molecules CD4 and CD8 on nonactivated T lymphocytes // Eur. J. Immunol. — 1992. — Vol. 22, No 10. — P. 2475—2479.

4. *Gay D., Maddon P., Sekaly R., Talle M.A., Godfrey M., Long E., Goldstein G., Chess L., Axel R., Kappler J. and Marrack P.* Functional interaction between human T-cell protein CD4 and the major histocompatibility complex HLA-DR antigen // Nature. — 1987. — Vol. 328, No 6131. — P. 626—629.

5. *Doyle C., Strominger J.L.* Interaction between CD4 and class II MHC molecules mediates cell adhesion // Nature. — 1987. — Vol. 330, No 6145. — P. 256—259.

6. *Rosenstein Y., Burakoff S.J. and Herrmann S.H.* HIV-gp120 can block CD4-class II MHC-mediated adhesion // J. Immunol. — 1990. — Vol. 144, No 2. — P. 526— 531.

7. *Maddon P.J., Dalgleish A.G., McDougal J.S., Clapham P.R., Weiss R.A., Axel R.* The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain // Cell. — 1986. — Vol. 47, No 3. — P. 333—348.

8. *Yakubov L.A., Kit Y.Y., Richter V.A., Andreeva A.Y., Karamyshev V.N., Stein C.A., Vlassov V.V.* The extracellular domain of CD4 receptor possesses a protein kinase activity // FEBS Lett. — 1998. — Vol. 431, No 1. — P. 45—48.

9. *Manning G., Whyte D.B., Martinez R., Hunter T., Sudarsanam S.* The protein kinase complement of the human genome // Science. — 2002. — Vol. 298, No 5600. — P. 1913—1934.

10. *Olson N.J., Pearson R.B., Needleman D.S., Hurwitz M.Y., Kemp B.E., Means A.R.* Regulatory and structural motifs of chicken gizzard myosin light chain kinase // Proc. Natl. Acad. Sci. USA. — 1990. — Vol. 87, No 6. — P. 2284—2288.

11. *Bork P., Holm L., Sander C.* The immunoglobulin fold // J. Mol Biol. — 1994. — Vol. 242, No 4. — P. 309—320.

12. *Tskhovrebova L., Trinick J.* Titin: properties and family relationship // Nature Reviews Molecular Cell Biology. — 2003. — Vol. 4, No 9. — P. 679—689.

13. *Hanks S.K., Hunter T.* Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification // FASEB J. — 1995. — Vol. 9, No 8. — P. 576—96.

14. Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium // Science. — 1998. — Vol. 282, No 5396. — P. 2012—2018.

15. *Clayton L.K., Hussey R.E., Steinbrich R., Ramachandran H., Husain Y., Reinherz E.L.* Substitution of murine for human CD4 residues identifies amino acids critical for HIV-gp120 binding // Nature. — 1988. — Vol. 335, No 6188. — P. 363—366.

16. *Peterson A., Seed B.* Genetic analysis of monoclonal antibody and HIV binding sites on the human lymphocyte antigen CD4 // Cell. — 1988. — Vol. 54, No 1. — P. 65—72.

17. *Кит Ю.Я., Якубов Л.А. Рихтер В.А., Власов В.В.* Фосфорилирование рецептора rsCD4 в присутствии клеток и плазмы крови человека // Доклады Академии Наук. — 1995. — Т. 340, № 3. — С. 442—446.

18. *Obenauer J.C., Cantley L.C., Yaffe M.B.* Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs // Nucleic Acids Res. — 2003. — Vol. 31, No 13. — P. 3635—3641.

19. *Roux P.P, Blenis J.* ERK and p38 MAPK-activated protein kinases: a family of protein kinases with diverse biological functions // Microbiol. Mol. Biol. Rev. — 2004. — Vol. 68, No 2. — P. 320—344.

20. *Ponting C.P., Phillips C., Davies K.E, Blake D.J.* PDZ domains: targeting signalling molecules to sub-membranous sites // Bioessays. — 1997. — Vol. 19, No 6. — P. 469—479.

21. *Henrissat B.* A classification of glycosyl hydrolases based on amino acid sequence similarities // Biochem J. — 1991. — Vol. 280, No 2. — P. 309—316.

22. *Clark A.J., Inwood W., Cloutier T. and Dhillon T.S.* Nucleotide sequence of coliphage HK620 and the evolution of lambdoid phages // J. Mol. Biol. — 2001. — Vol. 311, No 4. — P. 657—679.

23. *Stein C.A., Cleary A.M., Yakubov L.A., Lederman S.* Phosphorothioate oligodeoxynucleotides bind to the third variable loop domain (v3) of human immunodeficiency virus type 1 gp120 // Antisense Res. Dev. — 1993. — Vol. 3, No 25. — P. 19—31.