ISSN 1993-6842. Biopolymers and cell. 2008. vol. 24. N 6. Translated from Ukrainian

BIOINFORMATICS

Bioinformatic search for potential phosphorylation sites of melusin – integrin 1-binding protein

I. V. Kroupskaya, L. M. Kapustian, L. L. Sidorik

Institute of molecular biology and genetics NAS of Ukraine
150, Zabolotnogo Str, Kyiv Ukraine, 03680

i.v.kroupskaya@imbg.org.ua

Phosphorylation is one of the most frequently occurring posttranslational modifications in proteins. It plays an essential role in transferring outside signals into a cell and regulates different cellular processes such as growth, metabolism, proliferation, motility and differentiation. Melusin is a stress response protein which strictly reacts to the threshold levels of mechanic stress and activates cardiomyocytes signaling pathways. The search for potential sites of melusin phosphorylation was performed using bioinformatic analysis of primary protein sequences. The comparative bioinformatic analysis of possible phosphorylation sites, evolutionary and structural motifs has identified Ser_{326},Ser_{329} *and* Ser_{334} *as the most likely sites for phosphorylation of melusin by protein kinase CK2 in cardiamyocytes.*

Keywords: melusin, C-domain, secondary structure, 3D-structure, cell signaling pathways, prediction phospthorylation sites.

Introduction. Melusin is a cytoplasmic protein with unique structural peculiarities synthesised in the skeletal tissues and cardiac muscle. The human melusin (integrin- 1-binding protein of 2-ITGB1BP2) is known to be one of so-called CHORD-proteins [1]. It is a multidomain protein consisting of 347 amino acid residues with the N-terminal region represented by two cysteine/histidine rich domains (CHORD-domain) able to fix Zn^{2+} ions (Fig 1). The C-terminal protein region contains so-called CS domain which is the binding site with the 1-integrin cytoplasmic domain. This domain was also discovered in -crystalline and p23 and Stg1 proteins [2] belonging to the molecular chaperone family. The C-terminal end is rich in glutamatic and aspartic amino acids and represents a region of Ca^{2+} ions fixation. Melusin interacts with the cytoplasmic domain of 1-integrine and is localized in the costameres participating in the molecular mechanism of sarcomere conjunction to the sarcolemma and intracellular matrix [1, 3]. Melusin is supposed to play a key role in the cardiac hypertrophy.

The experiments on the melusin gene inactivation in mice allowed assuming that this protein is not a critical one in the embryonic heart formation, sarcomeric organisation or heart functioning under normal conditions [2, 4]. The melusin removal breaks the hypertrophic response of the aortic ventricle and dramatically accelerates the transition to the cardiac dilatation [3, 4]. The diminishing of interaction forces between melusin, 1-integrine cytoplasmic domain

© I. V. KROUPSKAYA, L. M. KAPUSTIAN, L. L. SIDORIK, 2008

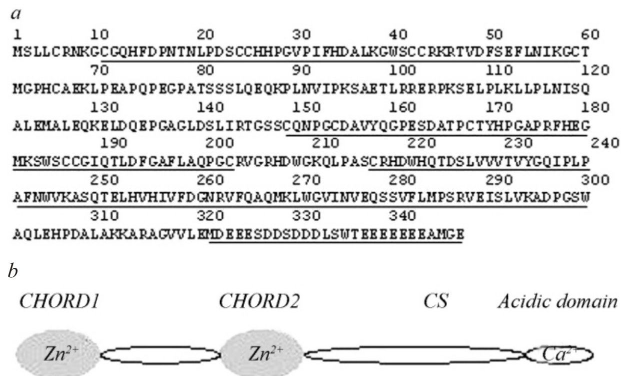

Fig. 1. Amino acid sequence (a) and domain organization (b) of human melusin: CHORD1 – 5–59; CHORD2 – 148–201; CS – 215–304; acidic domain – Ca^{2+}-binding domain 320–346

and ILK kinase in the costameric junctions causes the cardiac dysfunction and dilated cardiomyopathy (DCM) development [5]. At the same time by the melusin super expression the ventricles of transgenic mice hearts maintained the full contractile function in the case of long and constant overload [6]. Such functional properties were connected with the protection against the apoptosis and lack of stromal tissue displacement. Melusin controls the hypertrophic heart reaction in response to the various stresses e.g. high pressure. The mechanical changes in the cardiomyocytes activate several cell signaling pathways: MAPK, JAK/STAT, PI3K/AKT, AKT/GSK3 [6-8]. Melusin belongs to the proteins reacting acutely to the threshold levels of mechanical stress and activates AKT/GSK3 signaling pathways controlling the cardiomyocytes hypertrophy. This is a molecule with a unique ability to cause the compensatory hypertrophy and prevent the cardiac dilatation and cardiac failure [6]. The study on a melusin role in the cardiomyocytes signaling pathways is of great theoretical and practical interest.

Our study is devoted to the search for potential sites of melusin phosphorylation using bioinformatic analysis of the protein sequence. The prediction is based on the analysis of sites availability on the linear molecule according to the hydrophobicity level with an assumption that the hydrophobic sequences are rather hidden inside the protein molecule than exhibited outside. As a result of performed analysis it was discovered that Ser_{326}, Ser_{329} and Ser_{334} are the potential regions for phosphorylation by CK2 protein kinase.

Materials and Methods. *Phylogenetic analysis.* The primary sequence comparison was carried out applying the program CLUSTAL W [9]. The protein sequences of human melusin 1 and 2 were received from NCBI and SwissProt databases and used to search out the appropriate proteins from the other eukaryotes through the instrumentality of BLAST 2.0 on the BLAST NCBI.

Search for the phosphorylation sites. The search for potential phosphorylation regions was carried out applying the following online resources programs: ScanSite [10], NetPhos 2.0 [11], and KinasePhos 2.0 [12]. These programs, based on protein primary sequences, enable the prediction of the whole motifs for various known protein kinases applying the selective matrix, calculated as a result of protein libraries screening. They are also based on the availability analysis of sites on the protein molecule.

The protein secondary structure prediction. The protein secondary structure was predicted applying the program GORNIER4 [13] and PHD web server [14].

The protein 3D structure construction. 3D structure of melusin was simulated on the basis of CHORD containing protein-1 (2YRT) structure [15] applying the web server of the program 3D-JIGSAW (version 2.0) www. bmm.icnet.uk [16, 17]. The appropriate amino acid exposition was determined by the 3D-structure applying the program SWISS PDB Viewer.

Results and discussion. We performed the computer analysis of human melusin structure in order to reveal the potential phosphorylation sites. Special attention was paid to the Ser/Thr protein kinases participating in PI3K-AKT, AKT/GSK3 and other signaling pathways activated under conditions of various stresses and resulting in the mechanical changes in the cardiomyocytes. Using the program NetPhos 2.0 we discovered seven potential on- serine and one on-threonine phosphorylation sites in melusin, selecting these sites with prediction probability of 0.8 and higher, and considering the coincidence with the phosphorylation sites predicted by other programs: ScanSite and KinasePhose 2.0 (see the table). The most probable phosphorylation sites for protein kinase CK2 (S_{326}, S_{329}, S_{334}) were selected as a result of comparative analysis. In order to ascertain how conservative these

Table of melusin phosphorylation sites

Protein source	NetPhos 2.0	ScanSite	Kinase Phose 2.0
Homo sapiens, isoform 2	S23 S96 S107 S326 S329 S334	S326 S329 S334 T336	S83 S163 S326 S329 S334 T336
Homo sapiens, inisoform1	S65 S89 S308 S311 S316	S308 S311 S316 T318	S65 S145 S308 S311 S316 T318
Bos taurus, isoform 1	S23 S83 S96 S107 S283 S327 S330 S335	S32 S330 S335 T337	S83 S327 S330 S335 T337
Macaca mulatta, isoform 1	S65 S89 S265 S309 S312 S317	S309 S312 S317 T319	S65 S146 S309 S312 S317 T319
Macaca mulatta, isoform 2	S83 S249 S283 S326 S329 S334	S327 S330 S335 T337	S83 S164 S327 S330 S335 T337
Mus musculus	S23 S96 S283 S327 S330 S335	S327 S330 S335 T337	S83 S164 S309 S327 S330 S335 T337
Canis familiaris	S83 S107 S283 S327 S330 S335	S327 S S330 S335 T337	S90 S171 S334 S337 S342 T344
Pan troglodytes, isoform 1	S65 S78 S89 S308 S311 S318	S308 S311 S316 T318	S65 S145 S308 S311 S316 T318
Pan troglodytes, isoform 2	S23 S8 3S96 S107 2S326 S329 S334	S326 S32 S3349 T336	S83 S163 S326 S329 S334 T336
Rattus norvegicus	S106 S282 S326S329 S334	S326 S32 S3349 T336	S83 S163 S308 S326 S329 S334 T336
Sus scrofa	S23 S83 S96 S107 S162 S283 S316 S327 S330 S335	S327 S330 S335	S164 S249 S327 S330 S335 T337

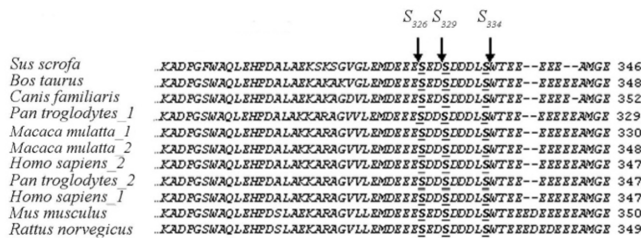

Fig. 2. Multiple alignement of the last C-terminal amino acids in melusin of various origin. Ser (*S*) localization is marked by arrows

aminoacids are, we performed the search and comparison with the eukaryotic melusin sequences known to date, namely:

Q9UKP_Homo sapiens_1,
Q5449J7_Homo sapiens,
Q32N04_ITBP2_Homo sapiens_2,
Q462R2_ITBP2_Sus scrofa,
Q9R000_INBP2_Mus musculus,
Q29RL2_Bovine,
DQ002920_Sus scrofa,
XM_590441_Bos taurus,
XM_001137370_Pan troglodytes_1,
XM_521119_Pan troglodytes_2,
XM_001069915_Rattus norvegicus,
XM_001091553_Macaca mulatta_1,
XM_001091670_ Macaca mulatta_2,
XM_853326_Canis familiaris.

It was found out that tree mentioned sites (S_{326}, S_{329}, S_{334}) are conservative for melusins of all the investigated eukaryotes (Fig. 2). Fig. 2 shows that the phosphorylation sites and surrounding amino acid residues did not change in the process of evolution which indicates the importance of these amino acids phosphorylation.

For the effective phosphorylation the aminoacids have to be available for protein kinases i.e. to be located on the protein surface. We built the models of secondary structures and performed 3D modelling for human melusin and other 12 known proteins. The analysis of predicted secondary structures for all studied proteins and corresponding 3D structures indicates that all predicted amino acids are located on the surface of the protein molecule. This is an essential confirmation that exactly S_{326}, S_{329}, S_{334} are the potential phosphorylation sites for CK2 protein kinase characterized by a high activity in the heart [18].

The protein phosphorylation as an important mechanism of post-translational modification has a significant influence on the cellular processes e.g. metabolism, differentiation, membrane transport and cell signaling pathways [19].

It is known that the C-domain of melusin is structurally similar to the C-domain of calsequestrin

a

```
MSTDSCCHHPGVPIFHDALKGWSCCRKRTVDFSEFLNIKGCTMGPHCAEK 50
eeeee e  ee       eee e  eee ee ee   e ee eeeeeeeee
LPEAPQPEGPATSSSLQEQKPLNVIPKSAETLRRERPKSELPLKLLPLNI 100
eeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeee
SQALEMALEQKELDQEPGAGLDSLIRTGSSCQNPGCDAVYQGPESDATPC 150
eee eeeeeeeeeeeeeeeeeeee eeeee eeee eee eeeeeeeee
TYHPGAPRFHEGMKSWSCCGIQTLDFGAFLAQPGCRVGRHDWGKQLPASC 200
e  ee     e  ee e  eee ee  e  eeee ee eee eeeeeee
RHDWHQTDSLVVVTVYGQIPLPAFNWVKASQTELHVHIVFDGNRVFQAQM 250
e e e eeee       ee eeee e e eeee e e eeeeeee e e
KLWGVINVEQSSVFLMPSRVEISLVKADPGSWAQLEHPDALAKKARAGVV 300
e  ee eeee e e eeee e e eeeeeee ee eeeeeeeeeeeeeee
LEMDEEESDDSDDDLSWTEEEEEEEAMGE                      350
eeeeeeeeeeeeee eeeeeeeeeeeeee
```

b

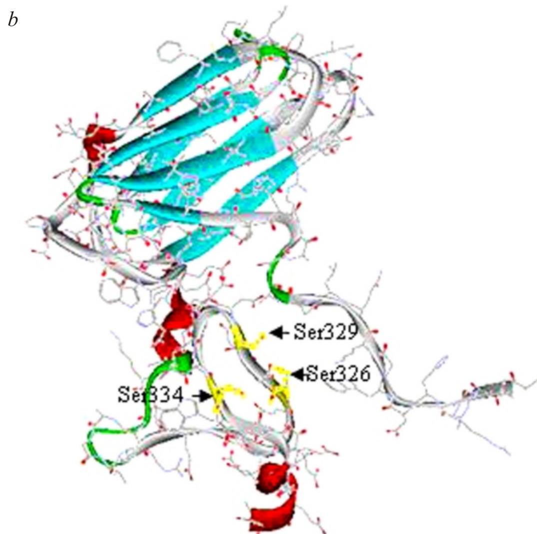

Fig. 3. Prediction of secondary structure of human melusin (*e* - exposed regions) (*a*) and predicted 3D structure of human melusin (*b*)

[20], which contains the phosphorylation sites for CK2 protein kinase [21] and fixes Ca^{2+} ions like melusin. The interaction of melusin with 1-integrine cytoplasmic domain is regulated by Ca^{2+} ions presence or absence. The attenuation of interaction between melusin, 1-integrine cytoplasmic domain and ILK in the costameres results in the disfunction and dilatation in the cardiac muscle [5]. Melusin is known to participate in the Ca^{2+} homeostasis [22], namely, in the Ca^{2+} ions concentration decrease in the sarcoplasmic reticulum of cardiomyocytes. This leads to an incorrect protein folding in response to a stress causing the cardiomyocytes apoptosis and heart failure as a consequence. We suppose that precisely the phosphorylated melusin forms can perform the function similar to that of calsequestrin i.e. to bind incorrectly folded proteins and transfer them through endoplasmic reticulum to the Golgi apparatus, ensuring the complicated cellular transport and thus participating in the cell signaling pathways. At the same time, the CS-domain and C-terminal end in melusin are also a signature for S100 proteins fixation involved in the cell Ca-regulatory function [2, 23]. The melusin phosphorylation seems to result in the post-translational change of the protein structure, whereby it binds S100 proteins.

И. В. Крупская, Л. М. Капустян, Л. Л. Сидорик

Биоинформатический поиск потенциальных сайтов фосфорилирования мелузина – интегрин- 1-связывающего белка

Фосфорилирование белков является важным механизмом посттрансляционной модификации и существенно влияет на клеточные процессы, такие как метаболизм, дифференциация, мембранный транспорт и сигнальные пути клетки. Мелузин – интегрин- 1-связывающий белок – относится к белкам, остро реагирующим на пороговые уровни механического стресса и активирующим сигнальные пути кардиомиоцитов. В данной работе проведен поиск потенциальных сайтов фосфорилирования мелузина с использованием биоинформатического анализа первичной последовательности белка. С помощью объединенного биоинформатического подхода к предсказанию сайтов фосфорилирования, а также эволюционных и структурных исследований идентифицировано, что именно Ser_{326}, Ser_{329}, Ser_{334} мелузина являются потенциальными участками для фосфорилирования протеинкиназой СК2.

Ключевые слова: мелузин, С-концевой домен, сигнальные пути клетки, предсказание сайтов фосфорилирования, вторичная структура белков, 3D-структура.

І. В. Крупська, Л. М. Капустян, Л. Л. Сидорик

Біоінформатичний пошук потенційних сайтів фосфорилювання мелузину – інтегрин- 1-зв'язувального білка

Резюме

Фосфорилювання білків є важливим механізмом посттрансляційної модифікації, що суттєво впливає на клітинні процеси, такі як метаболізм, диференціація, мембранний транспорт та сигнальні шляхи клітини. Мелузин належить до білків, які гостро реагують на порогові рівні механічного стресу та активують сигнальні шляхи кардіоміоцитів. Наші дослідження присвячені пошуку потенційних сайтів фосфорилювання мелузину з використанням біоінформаційного аналізу послідовності білка. Внаслідок об'єднаного біоінформатичного підходу передбачення сайтів фосфорилювання, еволюційних і структурних досліджень идентифіковано, що Ser_{326}, Ser_{329}, і Ser_{334} мелузина є потенційними ділянками для фосфорилювання протеїнкіназою СК2.

Ключові слова: мелузин, С- домен, сигнальні шляхи клітин, передбачення сайтів фосфорилювання, вторинна структура білків, 3D-структура.

REFERENCES

1. *Brancaccio M., Guazzone S., Menini N., Sibona E., Hirsch E., De Andrea M., Rocchi M., Altruda F., Tarone G., Silengo L.* Melusin is a new muscle specific interactor for beta(1) integrin cytoplasmic domain // J. Biol. Chem.–1999.–**274**.–P. 29282–29288.
2. *Michowski W., Lee Y-T., Chazin W. J., Kuznicki J.* Melusin binds calcyclin (S100A6) protein in a Ca^{2+}-dependent fashion // Eur. J. Biochem.–2003.–**1**, Suppl.–P. 3.7–09.
3. *Brancaccio M., Fratta L.,Notte A., Hirsch E., Poulet R., Guazzone S., De Acetis M., Vecchione C., Marino G., Altruda F., Silengo L., Tarone G., Lembo G.* Melusin, a muscle-specific integrin beta1-interacting protein, is required to prevent cardiac failure in response to chronic pressure overload // Nat. Med.–2003.–**9**, N 1.–P. 68–75.
4. *Brancaccio M., Hirsch E., Notte A., Selvetella G., Lembo G., Tarone G.* Integrin signalling: the tug-ofwar in heart gypertrophy // Cardiovasc. Res.–2006.–**70**, N 3.–P. 422–433.
5. *Hannigan G. E., Coles J. C., Dedhar S.* Integrin-linked kinase at the heart of cardiac contractility, repair and disease // Circ. Res.–2007.–**100**.–P. 1408–1414.
6. *De Acetis M., Notte A., Accornero F., Selvetella G., Brancaccio M., Vecchione C., Sbroggio M., Collino F., Pacchioni B., Lanfranchi G., Aretini A., Ferretti R., Maffei A., Altruda F., Silengo L., Tarone G., Lembo G.* Cardiac overexpression of melusin protects from dilated cardiomyopathy due to long-standing pressure overload // Circ. Res.–2005.–**96**, N 10.–P. 1087–1094.
7. *Pan J., Fukuda K., Kodama H., Sano M., Takahashi T., Makino S., Kato T., Manabe T., Hori S., Ogawa S.* Involvement of gp130-mediated signaling in pressure overload-induced activation of the JAK/STAT pathway in rodent heart // Heart Vessels.–1998.–**13**, N 4.–P. 199–208.
8. *Seko Y., Takahashi N., Sabe H., Tobe K., Kadowaki T., Nagai R.* Hypoxia induces activation and subcellular translocation of focal adhesion kinase (p125(FAK)) in cultured rat cardiac myocytes // Biochem. and Biophys. Res. Communs.–1999.–**262**, N 1.–P. 290–296.
9. *Thomson J. D., Higgins D. G., Gibson T. J.* CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice // Nucl. Acids Res.–1994.–**22**.–P.4673–4680.
10. *Obenauer J. C., Cantley L. C., Yaffe M. B.* ScanSite 2.0: Proteome-wide prediction of cell signalling interactions using short sequence motif // Nucl. Acids Res.–2003.–**31**.–P. 3635–3641.
11. *Blom N., Gammeltoft S., Brunak S.* Sequence and structure-based prediction of eukaryotic protein phosphorylation sites // J. Mol. Biol.–1999.–**294**.–P. 1351–1362.
12. *Huang H. D., Lee T. Y., Tseng, S. W., Horng J. T.* Kinase Phose 2.0: a web tool for identifying protein kinase-specific phosphorylation sites // Nucl. Acids Res.–2005.–**33**.–P. W226–229.
13. *Garnier J., Gibrat J.-F., Robson B.* GOR secondary structure prediction method version IV // Meth. Enzymol.–1996.–**266**.–P. 540–553.
14. *Rost B., Sander C., Schneider R.* PHD – an automatic mail server for protein secondary structure prediction // Comput. Appl. Biosci.–1994.–**10**.–P. 53–60.
15. *Nameki N., Saito K., Koshiba S., Kigawa T., Yokoyama S.* Solution structure of the CHORD domain of human CHORD-containing protein 1 // RIKEN Structural Genomics/Proteomics Initiative (RSGI) Release, 2008-04-08.
16. *Offman M. N., Fitzjohn P. W., Bates P. A.* Developing a move-set for protein model refinement // Bioinformatics.–2006.–**22**, N 15.–P. 1838–1845.
17. *Guex N., Peitsch M. C.* SWISS-MODEL and the Swiss-PdbViewer and the Swiss-PdbViewer: An environment for comparative protein modeling // Electrophoresis.–1997.–**18**.–P. 2714–2723.
18. *Kim S. O., Hasham M. I., Katz S., Pelech S. L.* Insulin-regulated protein kinases during postnatal development of rat heart // J. Cell Biochem.–1998.–**71**, N 3.–P. 328–339.
19. *Kuznetsov M., Janakiraman R., Troppmair M. J.* Regulating cell survival by controlling cellular energy production: novel functions for ancient signaling pathways? // FEBS Lett.–2004.–**577**, N 1–2.0.–P. 1–4.
20. *Scott B. T., Simmerman H. K., Collins J. H., Nadal-Ginard B., Jones L. R.* Complete amino acid sequence of canine cardiac calsequestrin deduced by cDNA cloning // J. Biol. Chem.–1988.–**263**, N 18.–P. 8958–8964.
21. *Cala S. E., Jones L. R.* Phosphorylation of cardiac and skeletal muscle calsequestrin isoforms by Casein Kinase II // J. Biol. Chem.–1991.–**266**, N 1.–P. 391–398.
22. *Gorza L., del Monte F.* Protein unfolding in cardiomyopathies // Heart Fail Clin.–2005.–**1**, N 2.–P. 237–250.
23. *Nowotny M., Spiechowicz M., Jastrzebska B., Filipek A., Kitagawa K., Kuznicki J.* Calcium-regulated interaction of Sgt1 with S100A6 (calcyclin) and other S100 proteins // J. Biol. Chem.–2003.–**278**, N 29.–P. 26923–26928.

UDC 577.21
Received 21.04.08