

Technical Note

FurinDB: A Database of 20-Residue Furin Cleavage Site Motifs, Substrates and Their Associated Drugs

Sun Tian ^{1,2,†,*}, Qingsheng Huang ^{1,3,†}, Ying Fang ¹ and Jianhua Wu ^{1,*}

¹ Institute of Biomechanics, School of Bioscience and Bioengineering, South China University of Technology, Guangzhou 510006, China; E-Mails: huangqqss@163.com (Q.H.); yfang@scut.edu (Y.F.)

² Nuolan Net, Wuhan 430071, China

³ School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, China

† These authors contributed equally to this work.

* Authors to whom correspondence should be addressed; E-Mails: sun.tian@nuolan.net (S.T.); wujianhua@scut.edu.cn (J.W.); Tel.: +86-20-39380608 (J.W.); Fax: +86-20-39380607 (J.W.).

Received: 28 December 2010; in revised form: 17 January 2011 / Accepted: 19 January 2011 / Published: 8 February 2011

Abstract: FurinDB (freely available online at <http://www.nuolan.net/substrates.html>) is a database of furin substrates. This database includes experimentally verified furin cleavage sites, substrates, species, experimental methods, original publications of experiments and associated drugs targeting furin substrates. The current database release contains 126 furin cleavage sites from three species: mammals, bacteria and viruses. A main feature of this database is that all furin cleavage sites are recorded as a 20-residue motif, including one core region (eight amino acids, P6–P2') and two flanking solvent accessible regions (eight amino acids, P7–P14, and four amino acids, P3'–P6'), that represent our current understanding of the molecular biology of furin cleavage. This database is important for understanding the molecular evolution and relationships between sequence motifs, 3D structures, cellular functions and physical properties required by furin for cleavage, and for elucidating the molecular mechanisms and the progression of furin cleavage associated human diseases, including pathogenic infections, neurological disorders, tumorigenesis, tumor invasion, angiogenesis, and metastasis. FurinDB database will be a solid addition to the publicly available infrastructure for scientists in the field of molecular biology.

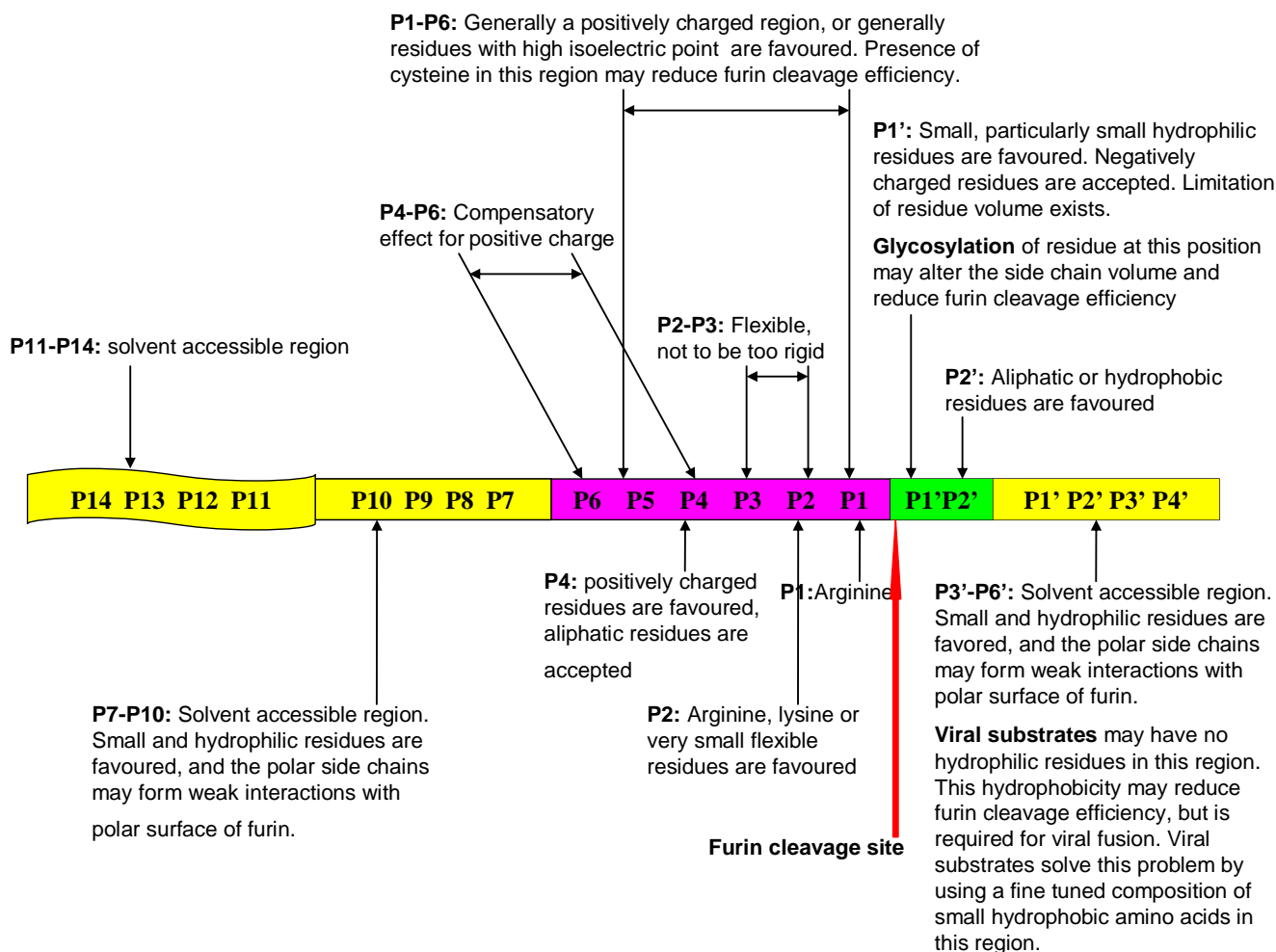
Keywords: furin; furin cleavage site; molecular database

1. Introduction

In the secretory pathway, many proteins are initially synthesized as inactive protein precursors, and proteolytic cleavage of these protein precursors by furin is an important step in the production of biologically active proteins [1,2]. Functionally, furin has two main subcellular localizations: the trans-Golgi network and the cell surface. In the trans-Golgi network, furin cleaves and activates most host substrates, including extracellular matrix proteins, signaling peptides, growth factors, serum proteins, cell surface transmembrane receptors and ion channels [1,2]. On the cell surface, furin cleaves many pathogenic substrates, such as bacterial toxins and viral fusion peptides. Maturation of functional proteins by furin cleavage is known to be involved in a wide range of human diseases, such as pathogenic infections, neurological disorders, tumorigenesis, tumor invasion, angiogenesis and tumor metastasis [1,2]. Recently, the use of the furin cleavage motif for targeted drug delivery in the treatment of tumors has been demonstrated [3].

An important step in furin cleavage is the recognition of specific cleavage site motifs presented by furin substrates. Furin's consensus cleavage site motif was widely reported to be a canonical four amino acid motif R-X-[KR]-R. In one of the author's recent work, instead of this four amino acid motif, the furin cleavage site has been characterized as a 20 amino acid motif running from the P14–P6' region, which can be divided into one core region (eight amino acids, P6–P2') and two flanking solvent accessible regions (eight amino acids, P7–P14, and four amino acids, P3'–P6') [4]. The backward numbering (P1–P14) and forward numbering (P1'–P6') starts at the conserved arginine immediately prior to the furin cleavage site. The favored physical properties of each region required for interaction with the furin binding pocket and the solvent accessibility of furin substrates have been revealed (Figure 1) [4]. In mammals, bacteria and viruses, the evolutionarily conserved physical properties of furin cleavage site motifs are evident, and the dynamic relationship between the physical properties of the P1'–P6' region and both cellular function and viral infection has been analyzed with respect to different species [4]. Using this updated understanding of the sequence motif and the molecular biology of furin cleavage, in this work, we have developed the FurinDB database. This database aims to facilitate further studies into the molecular biology of furin cleavage, the molecular mechanisms of its associated human diseases, and possible therapeutic treatments. The current version of FurinDB is a collection of 126 experimental data of known furin cleavage sites. It contains information about the amino acid residues in furin cleavage sites, the presence of multiple cleavage sites in the same protein, furin substrates, substrate species, experimental methods, original publications and associated drugs that target furin substrates. Substrates are grouped according to their biological functional categories and taxonomy.

Figure 1. Physical properties of the 20 residue furin cleavage site motif. The physical properties of this motif have been conserved in molecular evolution. Distinct physical properties at different regions in the site affect the molecular biology of furin cleavage. The dynamic relationship between the physical properties of the P1'–P6' region and viral infectivity is explained. The figure is reproduced from the author's previous publication [4].



We expect FurinDB to be a useful resource in the study of the molecular evolution and molecular biology of furin cleavage site motifs, as well as the elucidation of the molecular mechanisms and the progression of furin cleavage mediated human diseases, such as pathogenic infections, neurological disorders, tumorigenesis, tumor invasion, angiogenesis, and metastasis.

2. Results and Discussion

2.1. Database Content

Currently, the FurinDB database contains 126 cleavage sites from 110 substrates originating from three species [4], as well as 15 associated drugs that have either already been approved by the FDA or are in trials. All substrates, cleavage sites and original experimental records were collected from peer-reviewed literature [4]. The drug targets were identified by querying against Ingenuity Knowledge Base. The FurinDB database was developed using PHP 5 and MySQL database.

Each cleavage site is considered an entry, and each entry includes the following information:

1. Furin substrate sequence information: Name and description of the gene encoding the substrate protein and the substrate protein ID. External links to the NCBI protein sequence database and the NCBI Entrez gene database are also provided.
2. Furin cleavage site sequence information and location: The furin cleavage site P14–P6' is represented as a 20 amino acids motif that contains one core region (eight amino acids, positions P6–P2') packed inside the furin binding pocket; and two polar regions (eight amino acids, positions P7–P14; and four amino acids, positions P3'–P6') located outside of the furin binding pocket [4]. The position P1, immediately prior to the furin cleavage site is recorded. The presence of multiple cleavage sites on the same protein substrate is indicated.
3. Taxonomy information of substrate: Each furin substrate is assigned to one of three taxonomy groups (mammals, bacteria or viruses).
4. Functional information about the furin substrates: Each furin substrate is assigned to one of seven predefined biological functions: (1) extracellular matrix protein; (2) signal peptide, hormone or growth factor; (3) serum protein; (4) transmembrane receptor; (5) viral protein; (6) bacterial protein or (7) other [1,2].
5. Experimental methods and conditions: Methods used to measure furin cleavage and the final cleavage products, the origin of the furin enzyme and the origin of the furin substrate used in the experiment and the literature record of the original experiments. These detailed experimental methods and conditions are grouped into 10 different types [5]. External links to the PUBMED literature database are also provided.
6. Associated drugs: Information is included for each furin substrate targeted by a drug that has either entered phase I trials or has already received FDA approval.

2.2. Database Access and Web Interface

The FurinDB database is publicly accessible through a web interface [5]. The website provides three different ways to facilitate searching the database, with different queries for gene name, biological function, taxonomy, drug availability and furin cleavage site motif.

Three different search methods have been implemented in the database:

1. Query by seven functional categories, three species and if the furin substrate is a known drug target.
2. Query by standard Entrez gene name.
3. Query by the eight amino acids motif in the core region P6–P2' of the furin cleavage site, which is located inside the furin binding pocket. Either one or multiple amino acids can be input at any of the eight positions. The symbol X is interpreted to mean any amino acid.

Query results are displayed in table form (Figure 2). The core region (eight amino acids, positions P6–P2'), which is packed inside the furin binding pocket, is displayed in blue, and the arginine at position P1 immediately prior to the furin cleavage site is displayed in red.

Figure 2. Screenshot of the FurinDB database. Information of substrates, cleavage sites, taxonomies, biological functions, experimental methods and associated drugs are displayed in table form.

Furin Cleavage Site

Furin cleavage site study the presence and location of furin cleavage sites in protein sequences from different organisms: mammal, bacteria and virus.

Nuolan	FurinDB substrates	Cleavage motif	Reference
--------	--------------------	----------------	-----------

Results: 15 furin cleavage sites match the search criteria:

Protein ID	Gene name	Description	P1 Position	Upstream sequence of cleavage site P1-P14	Downstream sequence of cleavage site P1'-P6'	Organism	Function Group	Drug	Furin protein from	Substrate from	Experiment	Publication	Note	Site	Disease
16554579	COL5A1	Collagen, type V, alpha 1	1595	VIQLPLIQASRTRR	NIDASQ	mammal	Extraacellular matrix protein	Collagenase	Human	Human	1	11741999			Burn, Dupuytren contracture
9660749	COL13A1	Collagen, type XIII, alpha 1	108			mammal	Extraacellular matrix protein	Collagenase	Spodoptera frugiperda	Human	0	11013208			Burn, Dupuytren contracture
41688524	COL17A1	Collagen, type XVII, alpha 1				mammal	Extraacellular matrix protein	Collagenase	Not specified	Human	Cleavage site not by furin	Swiss-prot annotation 15047704	Cleavage site not by furin		Burn, Dupuytren contracture
81871180	COL23A1	Collagen, type XXIII, alpha 1	110	REKLDGVAKLRTVR	EAPSEC	mammal	Extraacellular matrix protein	Collagenase	Rat	Rat	4, 6	12644468			Burn, Dupuytren contracture
38570075	COL25A1	Collagen, type XXV, alpha 1	112	QKSYEHMAKIRAR	EAPSEC	mammal	Extraacellular matrix protein	Collagenase	Human	Human	4, 5	11927537			Burn, Dupuytren contracture
124945	ITGA4	Integrin, alpha 4 (antigen CD49D, alpha 4 subunit of VLA-4 receptor)	597	EAAHYHLPVHISKR	STEEFP	mammal	Extraacellular matrix protein	Natalizumab	Human	Human	1, 6	12691605			Crohn's disease, Rheumatoid arthritis, Multiple sclerosis
4504783	ITGAV	Integrin, alpha V (vitronectin receptor, alpha polypeptide, antigen CD51)	890	VAGQGERDHLITKR	DLALSE	mammal	Extraacellular matrix protein	Abciximab, CNTO 95, EMD121974	Human	Human	3, 4	10657249			Melanoma, Prostatic carcinoma, Stroke, Angina, Stenosis, Myeloproliferative syndrome, Myelodysplastic syndrome, Leukemia, Lymphoma, Brain tumor, Multiple myeloma, Lung tumor, Small intestine tumor, Plasmacytoma, Glioblastoma multiforme, Sarcoma
21542456	CXCL10	Chemokine (C-X-C motif) ligand 10	96	KNLLKAVSKERSKR	SP	mammal	Signaling peptides Hormone Growth factor	MDX-1100	probably Human(from Alexis Biochemical)	Human	1, 6	14739277			Ulcerative colitis
124494	CXCL10	Chemokine (C-X-C motif) ligand 10	96	KNLMKAFSQKRSKR	AP	mammal	Signaling peptides Hormone Growth factor	MDX-1100	probably Human(from Alexis Biochemical)	Mouse	1, 6	14739277			Ulcerative colitis
557563	TGFβ2	Transforming growth factor, beta 2	301	YRLSEQQTNRKKR	ALDAAY	mammal	Signaling peptides Hormone Growth factor	AP-12009	Human	Human	6	11390472			Glioblastoma, Pancreatic tumor, Colorectal tumor, Melanoma, Glioma, Anaplastic astrocytoma

3. Future Development

Future directions for development of the database include two parts: (1) We aim to collect additional furin substrates from the literature and extend this effort to include other members of the mammalian proprotein convertase family [6,7]. Furthermore, given the fact that many furin substrates are involved in growth, proliferation and extracellular matrix remodeling, we anticipate that more drugs, particularly anti-cancer drugs, will be developed to target furin substrates. This effort will be aided by the novel design of specific small molecular inhibitors targeting individual members of the mammalian proprotein convertase family. These specific small molecules are predicted to be derivatives of the short peptide pattern [K/R]-X-V-X-K-R [8]. Database information on furin substrate associated drugs will be regularly updated; (2) An algorithm for the prediction of location and presence of furin cleavage sites in protein sequences is currently being tested. The webserver for this prediction tool will be running along with the database, and it will also be freely available at the same web site which hosts the FurinDB.

4. Conclusions

We have developed FurinDB, a database of experimentally verified furin cleavage sites, substrates and associated drugs targeting furin substrates. All data is freely available at [5]. The database will be a valuable tool for studying the molecular biology and molecular evolution of furin cleavage sites, as well as elucidating the molecular mechanisms and progression of furin cleavage associated human diseases.

Acknowledgements

This work was supported by NSFC grant 11072080, 10772069 (to J.W.) and 10972081 (to Y.F.). Sun Tian has a full time job. Sun Tian is self-financed and works on this project in his own personal time.

References

1. Molloy, S.S.; Anderson, E.D.; Jean, F.; Thomas, G. Bi-cycling the furin pathway: from TGN localization to pathogen activation and embryogenesis. *Trends Cell Biol.* **1999**, *9*, 28–35.
2. Thomas, G. Furin at the cutting edge: from protein traffic to embryogenesis and disease. *Nat. Rev. Mol. Cell Biol.* **2002**, *3*, 753–766.
3. Hajdin, K.; D'Alessandro, V.; Niggli, F.K.; Schafer, B.W.; Bernasconi, M. Furin targeted drug delivery for treatment of rhabdomyosarcoma in a mouse model. *PLoS One* **2010**, *5*, e10445.
4. Sun, T. A 20 Residues Motif Delineates the Furin Cleavage Site and its Physical Properties May Influence Viral Fusion. *Biochem. Insights* **2009**, *2*, 9–20.
5. Furin Cleavage Site. Available online: <http://www.nuolan.net/substrates.html> (accessed on 7 February 2011).
6. Nakayama, K. Furin: a mammalian subtilisin/Kex2p-like endoprotease involved in processing of a wide variety of precursor proteins. *Biochem. J.* **1997**, *327*, 625–635.
7. Seidah, N.G.; Chretien, M. Proprotein and prohormone convertases: a family of subtilases generating diverse bioactive polypeptides. *Brain Res.* **1999**, *848*, 45–62.
8. Sun, T.; Wu, J. Comparative study of the binding pockets of mammalian proprotein convertases and its implications for the design of specific small molecule inhibitors. *Int. J. Biol. Sci.* **2010**, *6*, 89–95.

© 2011 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).