A 20 Residues Motif Delineates the Furin Cleavage Site and its Physical Properties May Influence Viral Fusion

Sun Tian

Institute of Biomechanics, South China University of Technology, Guangzhou, Guangdong, 510640, China.

Abstract: Furin is a proprotein convertase that proteolytically cleaves protein precursors to yield functional proteins. Efficient cleavage depends on the presence of a specific sequence motif on the substrate. Currently, the cleavage site motif is described as a four amino acid pattern: R-X-[K/R]-R \downarrow . However, not all furin cleavage recognition sites can be described by this pattern and not all R-X-[K/R]-R \downarrow sites are cleaved by furin. Since many furin substrates are involved in the pathogenesis of viral infection and human diseases, it is important to accurately characterize the furin cleavage site motif. In this study, the furin cleavage site motif was characterized using statistical analysis. The data were interpreted within the 3D crystal structure of the furin catalytic domain. The results indicate that the furin cleavage site motif is comprised of about 20 residues, P14–P6'. Specific physical properties such as volume, charge, and hydrophilicity are required at specific positions. The furin cleavage site motif is divided into two parts: 1) one core region (8 amino acids, positions P6–P2') packed inside the furin binding pocket; 2) two polar regions (8 amino acids, positions P7–P14; and 4 amino acids, positions P3'–P6') located outside the furin binding pocket. The physical properties of the core region contribute to the binding strength of the furin substrate, while the polar regions provide a solvent accessible environment and facilitate the accessibility of the core region to the furin binding pocket. This furin cleavage site motif also revealed a dynamic relationship linking the evolution of physical properties in region P1'–P6' of viral fusion peptides, furin cleavage efficacy, and viral infectivity.

Keywords: Furin, cleavage site motif, physical property, viral fusion peptides

Background

Many secreted proteins are synthesized as precursors that require proteolytic cleavage of part of the polypeptide to become fully functional. Furin, a proprotein convertase involved in the proteolytic cleavage of many of these protein precursors, cycles between the trans-Golgi network and the cell surface.^{1,2} In the secretory pathway, furin recognizes the cleavage motif on the protein precursors and cleaves the substrates, resulting in functional proteins.

Furin substrates include both host proteins and pathogen proteins. The known host substrates of furin cleavage include precursors of extracellular matrix proteins, extracellular proteases, receptors, and hormones. Many pathogenic proteins, such as fusion proteins on the viral envelope and secreted toxins of bacteria, require furin mediated cleavage to initiate pathological infectivity. The cleavage of these furin substrates influences the molecular pathogenesis of a wide range of human diseases, including cancer, neurological disorders, and various viral infections.^{1–4} Therefore, an accurate description of the furin cleavage site motif is crucial for understanding the molecular mechanism of furin cleavage-mediated viral infections and furin cleavage-associated human diseases.

Many furin cleavage sites share the sequence pattern R-X-[K/R]-R \downarrow . However, not all furin cleavage sites include this pattern, e.g. the cleavage sites of human albumin precursor RGVFRR²⁴ \downarrow DA (gi|4502027)⁵ and human C-type natriuretic peptide precursor RLLR⁷³ \downarrow DL.⁶ In addition, not all sites comprising R-X-[K/R]-R are efficiently cleaved by furin, e.g. a mutated form of Sindbis Virus PE2 protein SGRSKR³²⁸ \downarrow LV (position 328 of Sindbis virus polyprotein P130 containing PE2 protein).⁷

In this report, a statistical analysis of 130 known furin cleavage sites was performed, and the consequence of biochemical mutations of the furin cleavage site was examined. The statistical data were interpreted within the 3D crystal structure of the furin catalytic domain solved by Henrich et al.⁸ A sequence motif of about 20 amino acids coding for a furin cleavage recognition site was revealed. The motif is divided into two regions: a core region (8 amino acids, P6–P2') and two polar regions

(cc)

Correspondence: Sun Tian, Institute of Biomechanics, South China University of Technology, Guangzhou, Guangdong, 510640, China. Tel: 00862039380608; Fax: 00862039380607; Email: sun.tian@nuolan.net

Copyright in this article, its metadata, and any supplementary data is held by its author or authors. It is published under the Creative Commons Attribution By licence. For further information go to: http://creativecommons.org/licenses/by/3.0/.

(8 amino acids, positions P7–P14; and 4 amino acids, positions P3'–P6'). In addition, specific physical properties are required at specific positions. Characterization of the furin cleavage site motif suggested a relationship between furin cleavage, the evolution of the physical properties in region P1'–P6' of viral fusion peptides, and the infectivity of dozens of viruses.

Materials and Methods

Collection of known furin cleavage sites

The furin cleavage sites were collected through both published literature via PubMed and the Swissprot database annotation.⁹ Only published and experimentally verified furin cleavage sites with positions of the cleavage sites indicated were included for statistical analysis. In total, 130 published and experimentally verified furin cleavage sites were used.¹⁰ The table of substrates, their taxons, and taxons of furins that are responsible for cleavage are publicly available on the associated website www.nuolan.net.

Statistical analysis using physical properties

- 1) Furin is an extracellular enzyme and furin cleavage generally takes place after the secreted signal peptides of a protein sequence are cleaved off. Therefore, secreted signal peptides in the original proprotein sequences were substituted with gap symbols.
- 2) The $n_{seq} = 130$ furin cleavage sites were aligned.
- Physical properties of amino acids v were retrieved from the AA index database (http:// www.genome.jp/aaindex).¹¹
- 4) An approach similar to the procedure described by Georg Neuberger¹² was used to calculate deviations *dev_{furin}* (*i*) and correlation coefficients *R_{furin}* (*i*) of physical properties *v* at a specific position *i*.
- To reduce the sequence redundancy, a sequence weight W_k was assigned to each furin cleavage site by the sum of mismatch method.¹³
- $dev_{furin}(i)$: deviation of physical property v from UNIREF database.¹⁴
- *v_{furin}* (i): mean value of physical property *v* of collected known furin cleavage sites at a specific position *i*.

• $v_{Uni \text{ Re } f}$ and $\delta_{Uni \text{ Re } f}$: mean value and standard deviation from UNIREF database.¹⁴

$$dev_{furin}(i) = \frac{v_{furin}(i) - v_{Uni \operatorname{Re} f}}{\delta_{Uni \operatorname{Re} f}}$$
(1)

$$v_{furin}(i) = \frac{\sum_{k=1}^{n_{seq}} w_k v_{furin}(k,i)}{\sum_{k=1}^{n_{seq}} w_k}$$
(2)

• To analyze the physical properties v at a specific position, position *i*, the correlation coefficients $R_{furin}(i)$ between two variables were calculated:

i) the vector of 20 observed amino acids counts $c_{furin}(a, I)$ and

ii) the vector v of 20 amino acid physical property value v_a

5) Correlation coefficients were interpreted within the 3D structure of the dec-RVKR-cmk inhibited mouse furin binding pocket (Protein Data Bank accession code IP8J).⁸ To model the 3D structure and analyze the active site, the procedures described by Guex et al. on the Swiss-Pdb Viewer site http://www.expasy. org/spdbv/were followed.¹⁵

Results and Discussion

Length of the furin cleavage recognition site motif

Many furin cleavage site motifs share a short four amino acids pattern: R-X-[K/R]-R \downarrow , which is located inside the binding pocket of furin. However, one plausible structural model of the furin cleavage site motif comprises not only a short core region located inside the furin binding pocket, but also two extended polar regions located outside the binding pocket that facilitate solvent accessibility. If this hypothesis is correct, consecutive polar and flexible regions encompassing the furin cleavage site are expected. To test this hypothesis, the mean values of two representative physical properties, EISD840101, a consensus normalized hydrophobicity scale of amino acids,¹⁶ and VINM940101, normalized flexibility parameters of amino acids,¹⁷ of 130 known furin cleavage sites

were compared with the background value of the UNIREF database.¹⁴

As shown in Figure 1, the continuous deviations demonstrate the existence of flexible and polar regions on both sides of the furin cleavage site. By setting the threshold to 10% of the sum of absolute values of the two deviations, the plot identified two consensus boundary positions, P14 and P6', although the exact boundaries of the furin cleavage site motif might vary from substrate to substrate. Therefore, the furin cleavage recognition site comprises about 20 amino acids, from position P14 to position P6'.

Physical properties of the core region

Position P1

In the data set collected from the published literature, all 130 furin cleavage sites have a positively charged arginine residue at position P1. A mutation of this arginine at position P1 diminished the detectable furin cleavage,^{18–21} indicating that arginine is required at position P1.

In fact, mutations of arginine at position P1 are routinely used as negative controls in biochemical experiments that test furin cleavage.

Positions P2, P3, and P4

Positively charged residues arginine and lysine were the residues most often found at position P2. A high correlation coefficient (0.99) for the positive charge property ZVEL_CH_P1²² was observed at this position (Table 1, R2). As shown in Figure 2, a positively charged residue at position P2 extends its side chain to form favourable interactions with negatively charged aspartic acids 153, 154, and 191 in the furin binding pocket.

Interestingly, $\sim 20\%$ of the substrates possess residues other than arginine or lysine at position P2, notably small residues such as glycine, alanine, serine, and threonine. As indicated in Figure 2, the positively charged arginine at position P1 interacts with the negatively charged ring formed by glutamic acids 257 and 331 and aspartic acids 258, 301, and 306, while the arginine at position P4 interacts with glutamic acid 236 and aspartic acid 264. A small, flexible residue at



Figure 1. Motif length and boundaries (20 amino acids, P14–P6') of the furin cleavage site. The graph illustrates the deviations of mean values of the hydrophobicity scales EISD840101¹⁶ and the flexibility parameter VINM940104¹⁷ over 51 positions (P26–P25'), encompassing the 130 known furin cleavage sites. The furin cleavage site is indicated as a vertical red line. The UNIREF average corresponds to 0% deviation and is indicated as a horizontal black baseline. The mean values of the physical properties were calculated as the deviation from the UNIREF average in percent of the UNIREF standard deviation. The plot was smoothed by running the average over three positions. Continuous deviations were observed on both sides of the furin cleavage sites. By setting the threshold to 10% of the sum of absolute values of the two deviations, the plot identified two consensus boundary positions: P14 and P6' (shown in purple), although the exact boundaries of the furin cleavage site motif might vary from substrate to substrate.

ID	Substrate position	Correlation coefficient R	Name of physical property	Description
R1	P2	0.82	ZIMJ680104	Isoelectric point ²⁵
R2	P2	0.99	ZVEL_CH_P1	Positive charged (without histidine in hydrophobic surrounding) ²²
R3	P3	0.67	ZIMJ680104	Isoelectric point ²⁵
R4	P3	0.69	ZVEL_CH_P1	Positive charged (without histidine in hydrophobic surrounding) ²²
R5	P4	0.69	ZVEL_CH_P1	Positive charged (without histidine in hydrophobic surrounding) ²²
R6	P5	0.72	ZVEL_CH_P1	Positive charged (without histidine in hydrophobic surrounding) ²²
R7	P5	0.74	ZIMJ680104	Isoelectric point ²⁵
R8	P6	0.66	ZIMJ680104	Isoelectric point ²⁵
R9	P7	-0.71	HUTJ700101	Heat capacity ²⁷
R10	P8	0.69	VASM830102	Relative population of conformational state C ²⁸
R11	P9	0.70	ISOY800104	Normalized relative frequency of bend R ²⁹
R12	P10	-0.66	WERD780101	Propensity to be buried inside ³⁰
R13	P10	0.71	VINM940103	Normalized flexibility parameters (B-values) ¹⁷
R14	P1′	0.73	ZVEL_TINY_	Tiny residues ²²
R15	P2´	0.80	ZVEL_ALI_2	Aliphatic residues ²²
R16	P3′	0.65	LEVM780106	Normalized frequency of reverse turn, unweighted ⁴⁸
R17	P3′	-0.66	HUTJ700101	Heat capacity ²⁷
R18	P4´	-0.65	WERD780101	Propensity to be buried inside ³⁰
R19	P4′	0.77	KARP850101	Flexibility parameter for no rigid neighbors ¹⁷
R20	P5′	-0.69	ROSG850101	Mean area buried on transfer ³¹

Table 1. List of selected correlation coefficients for physical properties (The selected physical properties provide the best representation and most concordant results. The complete list is available on the associated website www.nuolan.net).

position P2 might facilitate the furin cleavage process by allowing residues at positions P1 and P4 to form those specific interactions with the furin binding pocket. Indeed, furin substrate specificity assays revealed that motifs like KRTTR↓ were sometimes more efficiently cleaved than some of the canonical R-X-[K/R]-R↓ motifs.²³ Conversely, bulky and rigid aromatic residues like phenylalanine, histidine, tyrosine, and tryptophan might not be accepted at position P2 because they may impose local structural constraints. A mutation study of human albumin precursor (gi|4502027) reported that furin

cleavage efficiency was greatly reduced when the arginine at position P2 was replaced with histidine, a positively charged residue with a rigid ring structure on its side chain.⁵

At position P3, a preference for a positive charge was observed, but the correlation coefficient for the positive charge property ZVEL_CH_P1²² was much weaker (0.69; Table 1, R4). This is consistent with the structural environment surrounding position P3. As Figure 2 shows, the side chain of the residue at position P3 orients outward and does not appear to form a specific interaction with the furin binding pocket.



Figure 2. Interactions between the furin binding pocket and substrate residues at positions P1–P4. At position P1, a positively charged arginine residue forms specific interactions with a densely negatively charged ring formed by glutamic acid 257, aspartic acid 258, aspartic acid 301, aspartic acid 306 and glutamic acid 331 of the furin binding pocket. At position P2, a positively charged lysine residue tends to form interactions with negatively charged aspartic acid 154 or aspartic acid 191 of the furin binding pocket. At position P3, the side chain of valine is oriented outwards and it does not appear to form any specific interaction. At position P4, a positively charged arginine tends to form specific interactions with glutamic acid 236 and aspartic acid 264 of the furin binding pocket. This structure was modeled based on the published 3D structure of the furin catalytic domain in complex with an inhibitor (Protein Data Bank accession code IP8J).⁸ The figure was generated with Swiss–Pdb Viewer.⁴⁷

Positions P4, P5 and P6

Approximately 90% of the 130 collected furin cleavage sites have positively charged residues at position P4. Most of them are arginines, rarely lysine and histidine. As Figure 3 shows, the positively charged arginine is preferred at this position because it may interact with glutamic acid 236, aspartic acid 264, and tyrosine 308 of the furin binding pocket.

Aliphatic residues such as valine, leucine, and isoleucine are also accepted at position P4, representing about 8% of the substrates. In that case, the absence of a positive charge at substrate position P4 was observed to be compensated by the presence of a positive charge at substrate position P5 or P6, and known examples are listed in Table 2. The 3D structure of the furin binding pocket suggests the possible structural mechanism of this compensatory effect. The absence of a positively charged residue at position P4 results in

of the furin binding pocket. However, a positively charged residue at position P5 or P6 may interact with the negatively charged glutamic acid 230 or glutamic acid 257 of the furin binding pocket. As a consequence, the gain of an interaction between a positive charge and a negative charge at substrate position P5 or P6 compensates for the loss of the interaction between a positive charge and a negative charge at substrate position P4. In fact, 129 of the 130 sites contain at least one positively charged residue at position P4 or P6. The exception is one of the furin cleavage sites of Human Vitamin K-dependent protein C precursor (gi|131067),²⁴ which has a positively charged arginine at position P5. Therefore, the absence of a positively charged residue at position P4 can be compensated for by the presence of a positively charged residue at P6, or to lesser extent at position P5.

loss of favourable interactions with negatively

charged glutamic acid 236 and aspartic acid 264



Figure 3. The structural environment explains the compensatory effect for positive charge over positions P4, P5, and P6. At position P4, the positively charged residue arginine appears to interact with side chains of glutamic acid 236, aspartic acid 264, and tyrosine 308 of the furin binding pocket. If position P4 is occupied by an aliphatic residue instead of a positively charged residue, the loss of interaction between positive charge and negative charge at position P4 can be compensated for by the gain of interaction between positive charge at position P4 can be compensated for by the gain of interaction between positive charge and negative charge at position P4 can be compensated for by the gain of interaction between positive charge and acid 230 and glutamic acid 257. This structure was modeled based on the published 3D structure of the furin catalytic domain in complex with an inhibitor (Protein Data Bank accession code IP8J).⁸ The figure was generated with Swiss–Pdb Viewer.⁴⁷

Region P1-P6

The structural environment of the furin catalytic domain that surrounds region P1–P6 of the substrate is fairly negatively charged (Fig. 2 and Fig. 3),⁸ and consequently, a general trend toward basicity is favoured in this region. The significant correlation coefficients of physical properties, such as positive charge ZVEL_CH_P1²² and isoelectric point ZIMJ680104,²⁵ support this supposition (Table 1, R1–R8). Therefore, the presence of positively charged residues at positions P2–P6 should increase furin cleavage efficiency.

The effect of cysteine in region P1-P6

No cysteine was found in region P2–P6 in the known furin cleavage sites, although other small hydrophilic residues, such as serine and threonine,

and even small hydrophobic residues, such as alanine, glycine, and proline, were frequently observed. Cysteine is a fairly acidic residue. Its isoelectric point is 5.02, the third lowest of all amino acids. Therefore, it is possible that the fairly negatively charged environment of the furin binding pocket does not favour the presence of cysteine in region P2-P6. Another plausible explanation could be that furin is an extracellular enzyme and most of its substrates are located in the extracellular space, where the -SH groups of cysteines have a tendency to form disulfide bonds. Introducing cysteine might entirely alter the local structure of substrates in region P1-P6 and reduce furin cleavage. In fact, a mutation study of proalbumin reported that a mutation from lysine to cysteine at position P2 greatly diminished the cleavage of

GI number	Imber Name of the protein	
124945	human integrin alpha 4 precursor ⁴⁹	HVISKR ⁵⁹⁷ ↓
4506267	human pro-parathyroid hormone ⁵⁰	KSVKKR ³¹ ↓
1718154	human vascular endothelial growth factor C precursor ⁵¹	HSIIRR ²²⁷ ↓
4502027	human albumin precursor⁵	RGVFRR ²⁴ ↓
120611	human furin self-cleavage site ⁵²	RGVTKR ⁷⁵ ↓
131067	human vitamin K-dependent protein C precursor ²⁴	LRIRKR ⁴² ↓
12643343	human integral membrane protein 2B ⁵³	KGIQKR ²⁴³ ↓

Table 2. List of known furin cleavage sites with an aliphatic residue at position P4 and a positively charged residue at position P5 or P6.

proalbumin.⁵ The effect of introducing cysteine into region P2–P6 was also demonstrated in the case of a single nucleotide polymorphism of isoform II of Ectodysplasin-A protein (EDA). Mutated EDA with cysteine present at position P2 or position P5 was not efficiently cleaved by furin, and this mutation was found in patients with X-linked hypohydrotic ectodermal dysplasia. Loss of furin cleavage sites due to the presence of cysteine was thought to be one of the causes of this disease.²⁶

Position P1'-P2'

At position P1', a preference for small hydrophilic residues was observed. A high correlation coefficient (0.73) for the tiny residues property, ZVEL TINY,²² was detected (Table 1, R14). As illustrated in Figure 4, residues at position P1' are located in a small hydrophilic pocket formed by serine 253, asparagine 295, tyrosine 329, threonine 365, threonine 367, and serine 368 of the furin binding pocket. This small pocket can limit the volume of the residue at position P1' and the hydrophilicity explains the preference for hydrophilic residues at this substrate position. No furin cleavage site was observed with bulky, hydrophobic residues leucine, isoleucine, valine, or tryptophan at position P1'. Moreover, furin cleavage efficiency was greatly reduced when the aspartic acid at position P1' of the furin cleavage site RGV-FRR²⁴ DA of human albumin precursor was mutated to valine RGVFRR²⁴ \downarrow VA⁵

At position P1', a negatively charged residue is also accepted, and about 22% of the furin cleavage sites have large, negatively charged aspartic acid or glutamic acid. As Figure 4 shows, a negatively charged residue at position P1' may stabilize itself by extending its side chain towards positively charged histidine 194 or histidine 364.

At position P2', hydrophobic residues, particularly the aliphatic residue valine, appear to be favoured. A significant correlation coefficient (0.8) for aliphatic residues, ZVEL_ALI_2,²² was observed (Table 1, R15). The structural environment of position P2' is less clear, but the strong preference for valine suggests that the residue is located in a hydrophobic pocket.

Physical properties of the two flexible polar regions: P7–P14 and P3′–P6′

Exit from the core region and solvent accessibility: P7–P10 and P3′–P6′

On both sides of the furin cleavage site, P7-P10 and P3'-P6', changes in preferred physical properties were observed. Residues in these two regions share a tendency to be polar and flexible, which is supported by the correlation coefficients for many polarity and flexibility related physical properties: HUTJ700101 (Heat capacity),²⁷ VASM830102 (Relative population of conformational state C),²⁸ ISOY800104 (Normalized relative frequency of bend R),²⁹ WERD780101 (Propensity to be buried inside), ³⁰ VINM940103 (Normalized flexibility parameters),¹⁷ KARP850101(Flexibility parameter for no rigid neighbours),¹⁷ and ROSG850101(Mean area buried on transfer)³¹ (Table 1, R9-R13 and R16–R20). However, neither single residue type preference nor concordantly significant correlation coefficients for charge or volume were detected. The tendency toward polarity and flexibility of P7–P10 and P3'–P6' suggests that residues in these two regions expose their side chains to the solvent



Figure 4. The structural environment explains the preference for small hydrophilic residues at position P1[']. At position P1['], a serine residue of the substrate (the colour of the hydrophilic serine was changed from yellow to green for visualization purposes) is packed tightly in a small hydrophilic pocket. This structural environment should generally favor a small hydrophilic residue. The presence of positively charged histidine 194 and histidine 364 suggests that negatively charged residues are also accepted at position P1[']. This structure was modeled based on the published 3D structure of the furin catalytic domain in complex with an inhibitor (Protein Data Bank accession code IP8J).⁸ The figure was generated with Swiss–Pdb Viewer.⁴⁷

and may form linkers. These two polar linkers do not seem to interact directly with the furin binding pocket; rather, it is plausible that they extend their side chains to the solvent and form weak interactions with the polar surface of furin. Thus, the solvent accessibility of these two regions (P7–P10 and P3'–P6') facilitates binding of the core region (P6–P2').

Taken together, the data indicate that:

- the exit substrate positions from the furin binding pocket are position P7 and position P3'.
- regions P7–P10 and P3'–P6' are located outside the furin binding pocket where they form two flexible, solvent accessible linkers.

Region P11–P14

In general, there were no significant preferences for residue type or for physical properties at single positions in region P11–P14. Nonetheless, Figure 1 illustrates that deviations of hydrophobicity and flexibility do exist and that region P11–P14 is a solvent accessible transition region.

Glycosylation of substrate residues at

position P1' may interfere with viral infectivity The structural analysis revealed a relationship between glycosylation at position P1' and viral infectivity. There is a limitation on residue volume at position P1' (see section Position P1'–P2'), where small, hydrophilic residues favour furin cleavage. Surprisingly, a mutation in the PE2 protein sequence of the Sindbis virus that substituted asparagine for the wild-type serine at position P1' resulted in defective furin cleavage.⁷ According to the structural model and statistical analysis, a small, hydrophilic residue like asparagine should be acceptable at position P1'. However, N-linked glycosylation

The dilemma:

Region P3'–P6'of the furin cleavage site motif on the viral fusion peptides can not be too hydrophobic or too hydrophilic



Tune the composition of small hydrophobic amino acids such as glycine, alanine and proline.

Figure 5. The hydrophobicity of region P3'-P6' is fine-tuned by the virus and it affects the efficiency of furin cleavage and the efficiency of viral fusion.

on the amide nitrogen of the asparagine side chain would increase the volume of the asparagine to an extent that hinders the substrate from fitting into the narrow furin binding pocket. Although glycosylated asparagine at position P1' greatly reduces furin cleavage, non-glycosylated asparagine should be accepted. Indeed, non-glycosylated asparagine has been found in the final cleaved product of the PE2 protein.³²

Glycosylation is a frequently observed feature of viral proteins. Both N-linked glycosylation on asparagine and O-linked glycosylation on serine or threonine can dramatically increase residue volume. Therefore, the presence of asparagine, serine, or threonine at substrate position P1' deserves particular attention in the analysis of furin cleavage-mediated viral infection. In addition, the observation that glycosylation interferes with furin cleavage emphasizes the importance of studying the physical properties and the structural environment of a small functional motif rather than only relying on analyzing the sequence motif itself with pattern match based approaches.

Hydrophobicity of region P3'–P6' in viral substrates

Many virus-host fusion related viral envelope proteins require furin-mediated cleavage before they can initiate viral fusion. Surprisingly, many

viral substrates include a stretch of hydrophobic amino acids instead of hydrophilic residues in region P3'-P6'. Known examples include hemagglutinin precursor of influenza A virus,^{33–36} fusion glycoprotein F0 precursor of Newcastle disease virus,³⁷ fusion glycoprotein F0 precursor of human parainfluenza 3 virus,³⁸ fusion glycoprotein F0 precursor of human respiratory syncytial virus,³⁹ fusion glycoprotein F0 precursor of measles virus strain Edmonston,⁴⁰ envelope polyprotein precursor gp72 of bovine leukemia virus,⁴¹ envelope surface glycoprotein gp160 precursor of human immunodeficiency virus 1,⁴² envelope polyprotein precursor of human T-lymphotropic virus 2,43 and envelope polyprotein precursor gp140 of human immunodeficiency virus type 2 (isolate ST).⁴⁴ The average hydrophobicity scale of region P3'-P6' of those viral substrates is 0.546, while the average hydrophobicity scale of region P3'-P6' of all 130 collected furin cleavage substrates is -0.183 (the hydrophobicity scale is calculated with physical property CIDH920105 normalized average hydrophobicity scale of amino acids).⁴⁵

In theory, in order to provide a solvent accessible environment in region P3'-P6', hydrophilic residues should be favoured (see section P7-P10 and P3'-P6'). Therefore, the increased hydrophobicity of region P3'-P6' of the viral substrates might decrease the furin cleavage efficiency, thus decreasing viral infectivity. However, on the other hand, the hydrophobicity of this region is essential to the cellular function of the final cleaved products, which are responsible for the fusion of the viral envelope with the host's cytoplasmic membrane.⁴⁶ The virus solves this dilemma by including small, hydrophobic amino acids such as glycine, alanine, and proline in region P3'-P6'. Consequently, viral substrates maintain sufficient hydrophobicity for the viral fusion process without compromising furin cleavage efficiency. In the rapid evolution of viruses, a subtle balance between the infectivity of viral particles and the furin cleavage efficiency of these viral fusion peptides is achieved by fine-tuning the hydrophobic amino acid composition of region P3'-P6' (Fig. 5).

Conclusion

This study demonstrated that the furin cleavage site motif is comprised of about 20 residues (P14–P6'). Specific physical properties are required at specific positions and regions of this motif. The furin cleavage site motif is divided into two parts: 1) a core region (8 amino acids, position P6–P2') inside the furin binding pocket that

contributes to the binding strength of the substrate and 2) two polar regions (8 amino acids, positions P7–P14; and 4 amino acids, positions P3'–P6') outside the furin binding pocket that mainly contribute to the solvent accessibility of the substrate. The motif length and physical properties of the furin cleavage recognition site are summarized in Figure 6.

Acknowledgement and Disclosure

This work is an extension and revised study of author's PhD thesis. This work was self financed and completed in author's own time. The author received PhD degree from the Institute for Genomics and Bioinformatics at Graz University of Technology in June 2007. The author's PhD study was funded by GENAU Bioinformatics Integration Network PhD programme 2005–2007 (PhD programme and grant leader: Professor Zlatko Trajanoski at TUGraz). The author has presented his PhD study at GENAU BIN PhD students meeting (March 2007) and summarized his results in his PhD thesis (June 2007) at the Institute for Genomics and Bioinformatics at Graz University of Technology http://genome.tugraz. at/theses.shtml. Sun Tian thank Professor Zlatko Trajanoski.



Figure 6. Summary of motif length and physical properties of furin cleavage site.

Conflict of Interests

The author had a full time job after PhD graduation in June 2007. This work was self financed and completed in author's own time. The author has declared that no conflict of interest exists.

References

- Molloy SS, Anderson ED, 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. Seidah NG. Prat A: Precursor convertases in the secretory pathway, cytosol and extracellular milieu. *Essays Biochem*. 2002;38:79–94.
- Nakayama K. Furin: a mammalian subtilisin/Kex2p-like endoprotease involved in processing of a wide variety of precursor proteins. *Biochem J.* 1997;327(Pt 3):625–635.
- Brennan SO, Nakayama K. Furin has the proalbumin substrate specificity and serpin inhibitory properties of an in situ hepatic convertase. *FEBS Lett.* 1994;338:147–151.
- Wu C, Wu F, Pan J, Morser J, Wu Q. Furin-mediated processing of Pro-C-type natriuretic peptide. *J Biol Chem.* 2003;278:25847–25852.
- Klimstra WB, Heidner HW, Johnston RE. The furin protease cleavage recognition sequence of Sindbis virus PE2 can mediate virion attachment to cell surface heparan sulfate. *J Virol.* 1999;73:6299–6306.
- Henrich S, Cameron A, Bourenkov GP, Kiefersauer R, Huber R, Lindberg I, et al. The crystal structure of the proprotein processing proteinase furin explains its stringent specificity. *Nat Struct Biol.* 2003;10:520–526.
- Bairoch A, Apweiler R, Wu CH, Barker WC, Boeckmann B, Ferro S, et al. The Universal Protein Resource (UniProt). *Nucleic Acids Res.* 2005;33:D154–D159.
- Sun T. Sequence-analytic characterization and prediction of furin cleavage recognition site based on a simple substrate-catalytic domain binding structural model. Institute for Genomics and Bioinformatics, Graz University of Technology. 2007. PhD.
- Tomii K, Kanehisa M. Analysis of amino acid indices and mutation matrices for sequence comparison and structure prediction of proteins. *Protein Eng.* 1996;9:27–36.
- Neuberger G. A Framework for the Knowledge-Based Prediction of Short FunctionalMotifs from Amino Acid Sequence. Vienna University. 2006. PhD.
- Vingron M, Sibbald PR. Weighting in sequence space: a comparison of methods in terms of generalized sequences. *Proc Natl Acad Sci USA*. 1993;90:8777–8781.
- Suzek BE, Huang H, McGarvey P, Mazumder R, Wu CH. UniRef: Comprehensive and Non-Redundant UniProt Reference Clusters. *Bioinformatics*. 2007.
- Guex N, Diemand A, Peitsch MC, Schwede T. DeepView Swiss-PdbViewer. GlaxoSmithKline R&D and the Swiss Institute of Bioinformatics.
- Eisenberg D. Three-dimensional structure of membrane and surface proteins. Annu Rev Biochem. 1984;53:595–623.
- Vihinen M, Torkkila E, Riikonen P. Accuracy of protein flexibility predictions. *Proteins*. 1994;19:141–149.
- Raghunath M, Putnam EA, Ritty T, Hamstra D, Park ES, Tschodrich-Rotter M, et al. Carboxy-terminal conversion of profibrillin to fibrillin at a basic site by PACE/furin-like activity required for incorporation in the matrix. *J Cell Sci.* 1999;112(Pt 7):1093–1100.
- Leighton M, Kadler KE. Paired basic/Furin-like proprotein convertase cleavage of Pro-BMP-1 in the trans-Golgi network. *J Biol Chem.* 2003;278:18478–18484.
- Komada M, Hatsuzawa K, Shibamoto S, Ito F, Nakayama K, Kitamura N. Proteolytic processing of the hepatocyte growth factor/ scatter factor receptor by furin. *FEBS Lett.* 1993;328:25–29.

- Logeat F, Bessia C, Brou C, LeBail O, Jarriault S, Seidah NG, et al. The Notch1 receptor is cleaved constitutively by a furin-like convertase. *Proc Natl Acad Sci U S A*. 1998;95:8108–8112.
- Zvelebil MJ, Barton GJ, Taylor WR, Sternberg MJ. Prediction of protein secondary structure and active sites using the alignment of homologous sequences. *J Mol Biol.* 1987;195:957–961.
- Matthews DJ, Goodman LJ, Gorman CM, Wells JA. A survey of furin substrate specificity using substrate phage display. *Protein Sci.* 1994;3:1197–1205.
- 24. Drews R, Paleyanda RK, Lee TK, Chang RR, Rehemtulla A, Kaufman RJ, et al. Proteolytic maturation of protein C upon engineering the mouse mammary gland to express furin. *Proc Natl Acad Sci* USA. 1995;92:10462–10466.
- Zimmerman JM, Eliezer N, Simha R. The characterization of amino acid sequences in proteins by statistical methods. *J Theor Biol.* 1968;21:170–201.
- Chen Y, Molloy SS, Thomas L, Gambee J, Bachinger HP, Ferguson B, et al. Mutations within a furin consensus sequence block proteolytic release of ectodysplasin-A and cause X-linked hypohidrotic ectodermal dysplasia. *Proc Natl Acad Sci U S A*. 2001;98:7218–7223.
- Hutchens JO. Heat capacities, absolute entropies, and entropies of formation of amino acids and related compounds. In *Handbook of biochemistry*. Edited by Sober HA. Cleveland, Ohio: Chemical Rubber Co. 1970:B60–B61.
- Vasquez M, Nemethy G, Scheraga HA. Computed Conformational States of the 20 Naturally-Occurring Amino-Acid-Residues and of the Prototype Residue Alpha-Aminobutyric Acid. *Macromolecules*. 1983;16:1043–1049.
- 29. Isogai Y, Nemethy G, Rackovsky S, Leach SJ, Scheraga HA. Characterization of multiple bends in proteins. *Biopolymers*. 1980;19:1183–1210.
- Wertz DH, Scheraga HA. Influence of water on protein structure. An analysis of the preferences of amino acid residues for the inside or outside and for specific conformations in a protein molecule. *Macromolecules*. 1978;11:9–15.
- Rose GD, Geselowitz AR, Lesser GJ, Lee RH, Zehfus MH. Hydrophobicity of amino acid residues in globular proteins. *Science*. 1985;229:834–838.
- Heidner HW, Johnston RE. The amino-terminal residue of Sindbis virus glycoprotein E2 influences virus maturation, specific infectivity for BHK cells, and virulence in mice. *J Virol*. 1994;68:8064–8070.
- 33. Basak A, Zhong M, Munzer JS, Chretien M, Seidah NG. Implication of the proprotein convertases furin, PC5 and PC7 in the cleavage of surface glycoproteins of Hong Kong, Ebola and respiratory syncytial viruses: a comparative analysis with fluorogenic peptides. *Biochem J* 2001;353:537–545.
- 34. Walker JA, Molloy SS, Thomas G, Sakaguchi T, Yoshida T, Chambers TM, et al. Sequence specificity of furin, a proproteinprocessing endoprotease, for the hemagglutinin of a virulent avian influenza virus. *J Virol.* 1994;68:1213–1218.
- Horimoto T, Nakayama K, Smeekens SP, Kawaoka Y. Proprotein-processing endoproteases PC6 and furin both activate hemagglutinin of virulent avian influenza viruses. *J Virol.* 1994;68:6074–6078.
- 36. Stieneke-Grober A, Vey M, Angliker H, Shaw E, Thomas G, Roberts C, et al. Influenza virus hemagglutinin with multibasic cleavage site is activated by furin, a subtilisin-like endoprotease. *EMBO J*. 1992;11:2407–2414.
- 37. Gotoh B, Ohnishi Y, Inocencio NM, Esaki E, Nakayama K, Barr PJ, et al. Mammalian subtilisin-related proteinases in cleavage activation of the paramyxovirus fusion glycoprotein: superiority of furin/PACE to PC2 or PC1/PC3. *J Virol.* 1992;66:6391–6397.
- Ortmann D, Ohuchi M, Angliker H, Shaw E, Garten W, Klenk HD. Proteolytic cleavage of wild type and mutants of the F protein of human parainfluenza virus type 3 by two subtilisin-like endoproteases, furin and Kex2. *J Virol.* 1994;68:2772–2776.

- 39. Gonzalez-Reyes L, Ruiz-Arguello MB, Garcia-Barreno B, Calder L, Lopez JA, Albar JP, et al. Cleavage of the human respiratory syncytial virus fusion protein at two distinct sites is required for activation of membrane fusion. *Proc Natl Acad Sci U S A*. 2001;98:9859–9864.
- Bolt G, Pedersen IR. The role of subtilisin-like proprotein convertases for cleavage of the measles virus fusion glycoprotein in different cell types. *Virology*. 1998;252:387–398.
- Zarkik S, Decroly E, Wattiez R, Seidah NG, Burny A, Ruysschaert JM. Comparative processing of bovine leukemia virus envelope glycoprotein gp72 by subtilisin/kexin-like mammalian convertases. *FEBS Lett.* 1997;406:205–210.
- 42. Vollenweider F, Benjannet S, Decroly E, Savaria D, Lazure C, Thomas G, et al. Comparative cellular processing of the human immunodeficiency virus (HIV-1) envelope glycoprotein gp160 by the mammalian subtilisin/ kexin-like convertases. *Biochem J.* 1996;314(Pt 2):521–532.
- Hasegawa H, Tatsumi M, Ogawa-Goto K, Takahashi H, Kojima A, Iwasaki T, et al. Processing of the HTLV-II envelope precursor glycoprotein gp63 by furin is essential for cell fusion activity. *AIDS Res Hum Retroviruses*. 2002;18:1253–1260.
- Hallenberger S, Moulard M, Sordel M, Klenk HD, Garten W. The role of eukaryotic subtilisin-like endoproteases for the activation of human immunodeficiency virus glycoproteins in natural host cells. *J Virol*. 1997;71:1036–1045.
- 45. Cid H, Bunster M, Canales M, Gazitua F. Hydrophobicity and structural classes in proteins. *Protein Eng.* 1992;5:373–375.

- Weissenhorn W, Hinz A, Gaudin Y. Virus membrane fusion. *FEBS Lett.* 2007;581:2150–2155.
- 47. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*. 1997;18:2714–2723.
- Levitt M. Conformational preferences of amino acids in globular proteins. *Biochemistry*. 1978;17:4277–4285.
- Bergeron E, Basak A, Decroly E, Seidah NG. Processing of alpha4 integrin by the proprotein convertases: histidine at position P6 regulates cleavage. *Biochem J.* 2003;373:475–484.
- Hendy GN, Bennett HP, Gibbs BF, Lazure C, Day R, Seidah NG. Proparathyroid hormone is preferentially cleaved to parathyroid hormone by the prohormone convertase furin. A mass spectrometric study. *J Biol Chem.* 1995;270:9517–9525.
- Siegfried G, Basak A, Cromlish JA, Benjannet S, Marcinkiewicz J, Chretien M, et al. The secretory proprotein convertases furin, PC5, and PC7 activate VEGF-C to induce tumorigenesis. *J Clin Invest*. 2003;111:1723–1732.
- Anderson ED, VanSlyke JK, Thulin CD, Jean F, Thomas G. Activation of the furin endoprotease is a multiple-step process: requirements for acidification and internal propeptide cleavage. *EMBO J*. 1997;16:1508–1518.
- Kim SH, Wang R, Gordon DJ, Bass J, Steiner DF, Lynn DG, et al. Furin mediates enhanced production of fibrillogenic ABri peptides in familial British dementia. *Nat Neurosci*. 1999;2:984–988.