Amino Acids (2008) 35: 375–382
DOI 10.1007/s00726-007-0611-3
Printed in The Netherlands

Amino Acids

Cleavage mechanism of the H5N1 hemagglutinin by trypsin and furin

X.-L. Guo[1], **L. Li**[1], **D.-Q. Wei**[1,2], **Y.-S. Zhu**[1], and **K.-C. Chou**[2]

[1] College of Life Science and Biotechnology, Shanghai Jiao Tong University, Shanghai, China
[2] Gordon Life Science Institute, San Diego, California, U.S.A.

Received August 29, 2007
Accepted September 23, 2007
Published online January 31, 2008;

Summary. The cleavage property of hemagglutinin (HA) by different proteases was the prime determinant for influenza A virus pathogenicity. In order to understand the cleavage mechanism, molecular modeling tools were utilized to study the coupled model systems of the proteases, i.e., trypsin and furin and peptides of the cleavage sites specific to H5N1 and H1 HAs, which constitute models of HA precursor in complex with cleavage proteases. The peptide segments 'RERRRKKR ↓ G' and 'SIQSR ↓ G' from the high pathogenic H5N1 H5 and the low pathogenic H1N1 H1 cleavage sites were docking to the trypsin and furin active pockets, respectively. It was observed through the docking studies that trypsin was able to recognize and cleave both the high pathogenic and low pathogenic hemagglutinin, while furin could only cleave the high pathogenic hemagglutinin. An analysis of binding energies indicated that furin got most of its selectivity due to the interactions with P_1, P_4, and P_6, while having less interaction with P_2 and little interactions with P_3, P_5, P_7, and P_8. Some mutations of H5N1 H5 cleavage sequence fitted less well into furin and would reduce high pathogenicity of the virus. These findings hint that we should focus at the subsites P_1, P_4, and P_6 for developing drugs against H5N1 viruses.

Keywords: Trypsin – Furin – H5N1 hemagglutinin – Cleavage mechanism – Pathogenicity

Introduction

The genome of influenza A virus contains 8 single-stranded RNA segments. The hemagglutinin (HA) is a glycoprotein which is responsible for binding of virus to cell and mediating fusion of the viral and cellular membranes after endocytosis (Wiley and Skehel, 1987; Martin et al., 1998). The HA is synthesized as a precursor HA0 (Kido et al., 1993; Hatta et al., 2001). To be infectious, HA0 must be cleaved into two subunits, HA1 and HA2. HA0s of all influenza A viruses are cleaved at an arginine residue adjacent to the conserved glycine, which becomes the N-terminus of HA2 (Steinhauer, 1999). The cleavage activation of HA0 can be achieved by two classes of proteases. One is lysine or arginine favoring trypsin-like proteases, which exist in a limited number of cells or tissue types. The other is ubiquitous subtilisin-like proteases which only recognize the sequence of polybasic residues (Kawaoka and Webster, 1988). The distribution of these proteases in the host and the sequence of the HA cleavage site appear to be the prime determinant for the virus pathogenicity.

The crystal structures of the high pathogenic H5N1 H5 (2FK0) and the low pathogenic H1N1 H1 (1RD8) (Stevens et al., 2006) are available in PDB database. The structure of H5 is generally aligned very well with that of H1 except for a few specific sites (Stevens et al., 2006), especially the cleavage site. The most common sequence of the high pathogenic H5N1 H5 cleavage site is 'RERRRKKR ↓ G', consisting of polybasic amino acids with positive charges. It can be recognized and cleaved by both trypsin-like proteases and subtilisin-like proteases. While the low pathogenic H1N1 H1 (1RD8) cleavage sequence is 'SIQSR ↓ G', with only one basic amino acid. It can be cleaved only by trypsin-like proteases. In this research, segments 'RERRRKKR ↓ G' and 'SIQSR ↓ G' from the high pathogenic H5N1 H5 and the low pathogenic H1N1 H1 cleavage sites, respectively, were docking to different types of proteases to study the catalytic interactions and the cleavage mechanism, and give insights into the high pathogenicity of the H5N1 virus.

Materials and methods

During the drug-discovery course many useful hints can be gained through molecular docking studies (see, e.g., Chou et al., 2003, 2006; Du et al., 2004, 2005, 2007a, b; Sirois et al., 2004; Chou, 2004c; Wei et al., 2005, 2006a, b, 2007; Zhang et al., 2006; Wang et al., 2007a, b, c). The molecular docking (Morris et al., 1996) with the Metropolis algorithm,

also known as 'Monte Carlo simulated annealing' (Chou and Carlacci, 1991), was used to finding the most favorable binding interaction. During each constant temperature cycle of Monte Carlo simulated annealing, random changes are made to the ligand's current position, orientation and conformation. The new state is then compared with its predecessor. If the energy is lower than the previous, this new state is immediately accepted. Otherwise, it is accepted probabilistically. This probability depends upon the energy and cycle temperature.

The software package MOE (Chemical Computing Ltd.) was used to conduct the docking studies. The docking energy was calculated with the electrostatic and van der Waals potential fields, and the CHARMM22 (Buck et al., 2006) force field parameters were used in this study. Two receptors were selected for the docking studies. One was 2AGE (Radisky et al., 2006), chosen from many available crystal structures of trypsin. The other was 1P8J (Henrich et al., 2003), the unique crystal structure of furin proteinase from the mouse.

The nonapeptide RERRRKKR↓G and the hexapeptide SIQSR↓G were used as the ligands, where the ↓ represents the cleavage site (Chou, 1996). Following the definition (Chou, 1993; Schechter and Berger, 1967), the corresponding amino acids were sequentially symbolized by subsites ……, P_5, P_4, P_3, P_2, P_1, and $P_{1'}$ and the matching positions in proteases by ……, S_5, S_4, S_3, S_2, S_1, and $S_{1'}$. The cleavage point was on the peptide bond between P_1 and $P_{1'}$. Four groups of docking were made between two receptors and two ligands. For each docking, a total of 100 conformations were obtained, and the one with the lowest energy was used for further energy minimizing calculation to create the final conformation.

Results

Cleavage mechanism by trypsin protease

The nonapeptide RERRRKKR↓G was used as the ligand to dock to the receptor 2AGE (Radisky et al., 2006). The complex structure is given in Fig. 1. Catalysis takes place in a cleft where a specificity pocket (S_1 pocket) protrudes into the interior of the protease. The transparent molecular surface represents the hydrophobic and hydrophilic surface of S_1 pocket which is constructed by residues Asp189, Ser190, Cys191, Gln192, Gly193, Asp194, Ser195, Ser214, Tyr215, Gly216, Ser217, Gly219, Cys220, Ala221a, Pro225, Gly226, Val227, and Tyr228 (Ibrahim et al., 2004). The S_1 pocket accommodates long side chains of the basic amino acid P_1-Arg. At its bottom, a negatively charged aspartate residue (Asp189) forms a tight salt bridge with the substrate which is very important for its specificity (Schmidt et al., 2003). A catalytic triad (His57, Asp102, and Ser 195) constitutes the core of the reactive centre, surrounds the susceptible bond between P_1-Arg and $P_{1'}$-Gly, and passes protons to break the susceptible bond (Ma et al., 2005). For the stabilization of the charges developing on the catalytic intermediates an oxyanion hole is formed by the main chain N–H of residues 192–195 known as a 'nest'.

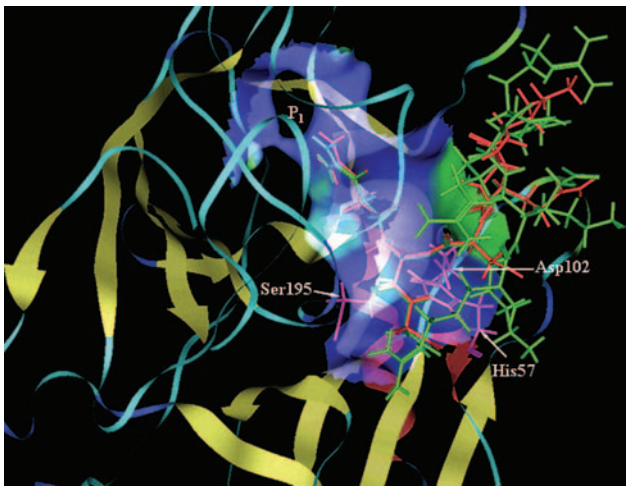

Fig. 1. The complex structures are obtained by docking peptides to trypsin. Trypsin is rendered in ribbon drawing, the nonapeptide RERRRKKR↓G is in *green* stick drawing, and the hexapeptide SIQSR↓G is in *red* stick drawing. The hydrophobic and hydrophilic surfaces of the S_1 pocket are colored in *green* and *blue*, respectively (for interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper)

Hydrogen bonding plays an important role in receptor-ligand binding interaction (see, e.g., Chou, 2004a, b, 2005a, b). It has been observed in this study that eight hydrogen bonds are formed between the receptor and the ligand: four hydrogen bonds are between the bottom of the S_1 pocket and the top of the P_1-Arg side chain (see Fig. 2a), one hydrogen bond is in the oxyanion hole (see Fig. 2c), one hydrogen bond is between ε-ammonium of Gln192 and the carbonyl oxygen of P_2-Lys, the other two hydrogen bonds are between the hydroxyl of Tyr151 and the guanidyl of P_8-Arg. Five of eight hydrogen bonds are formed on the P_1-Arg, with most of interaction with trypsin, consistent with the strict substrate requirement of trypsin for P_1 rather than other subsites.

The P_1-Arg side chain stands by Asp189-Ser190-Cys191 segment, passes curved Gln192-Gly193-Asp194-Ser195 and Ser214-Tyr215-Gly216-Ser217-Gly219-Cys220-Ala221a loops, and extends into the bottom of the S_1 pocket constructed by the Pro225-Gly226-Val227-Tyr228 segment. The guanidinium group of this side chain is perfectly framed by oxygens of the Asp189 carboxylate, the Gly219 carbonyl and the ser190 hydroxyl, for four efficient hydrogen bonds (Fig. 2a). These hydrogen bonds proved that residues Asp189, Ser190, and Gly219 are responsible for the substrate attraction and trypsin specificity.

A close view of interaction between the catalytic triad and the susceptible bond may provide useful information for in-depth understandings of the catalytic mechanism. The catalytic triad serves to make the active site Ser195 nucleophilic by modifying the electrostatic environment of the Ser195. As shown in Fig. 2b, there are interactions

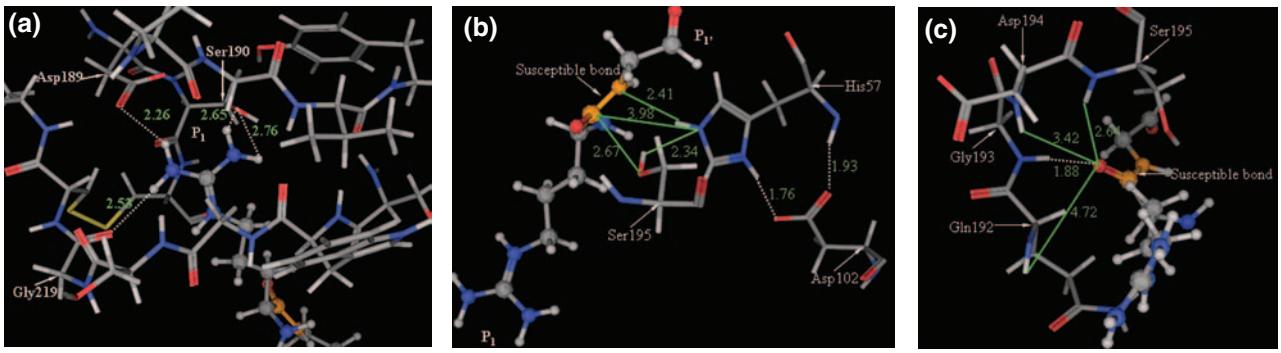

Fig. 2. Close views of interactions between trypsin and the nonapeptide, where the residues of trypsin are in stick drawing, and the ligand in ball and stick drawing: (**a**) interactions between the bottom of the S_1 pocket and the side chain of P_1-Arg; (**b**) interactions between the catalytic triad and the susceptible bond; (**c**) interactions between the oxyanion hole and the carbonyl oxygen. The *white dotted* lines represent the hydrogen bonds (for interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper)

among the catalytic triad. Two hydrogen bonds (white dotted lines) are formed between His57 and the carboxylate of Asp102 with distances of 1.93 Å and 1.76 Å. The distance of only 2.67 Å is available for the Ser195-OH attacking the carbonyl carbon in the event of catalysis. Also the distance of 2.34 Å is available for the nitrogen of the His57 accepting the hydrogen from the Ser195-OH. In the subsequent step, the interaction between the carbonyl carbon and the nitrogen of the His57 with a water molecule coming into the reaction will allow a larger distance of 3.98 Å.

As shown in Fig. 2c, the carbonyl oxygen inserts into the oxyanion hole which is formed by the mainchain N–H of residues Gln192, Gly193, Asp194, and Ser195. The distances between the carbonyl oxygen on the susceptible bond and the mainchain N–H of residues 192–195 in trypsin are 4.72 Å, 1.88 Å, 3.42 Å, and 2.64 Å, respectively. The atoms from the Gly193 and the carbonyl oxygen can form a hydrogen bond for stabilization because their distance is 1.88 Å.

The hexapeptide SIQSR ↓ G was also used as the ligand to dock to the receptor 2AGE (see Fig. 1). We can see from the figure that both of two P_1-Arg long side chains insert into the S_1 pocket and the susceptible bond in the red hexapeptide SIQSR ↓ G is also surrounded by the catalytic triad with the red P_1-Arg and $P_{1'}$-Gly well overlapped by the green ones.

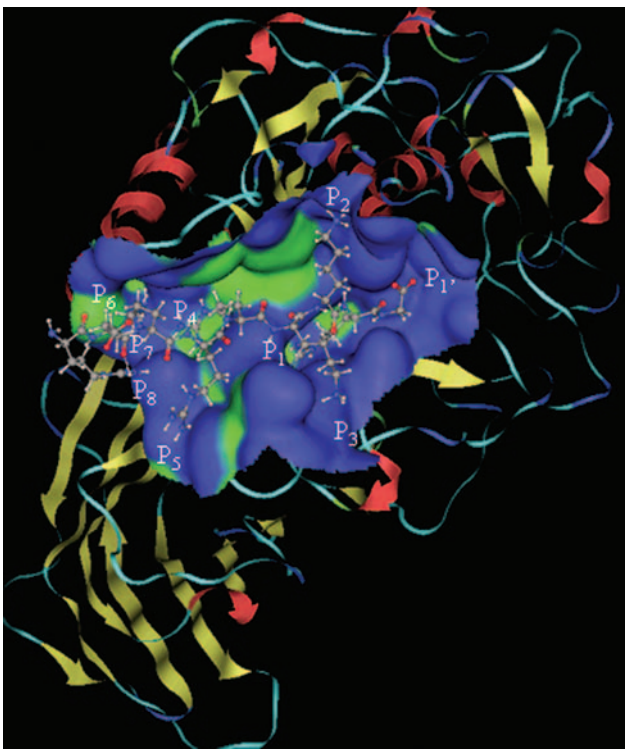

Fig. 3. The complex structure obtained by docking the nonapeptide to furin, where furin is rendered in ribbon drawing, the nonapeptide in ball and stick drawing. The hydrophobic and hydrophilic surfaces of the binding pocket are colored in *green* and *blue*, respectively (for interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper)

Cleavage mechanism by furin protease

The nonapeptide RERRRKKR ↓ G was used as the ligand to dock to the receptor 1P8J (Henrich et al., 2003). The complex structure of furin with the ligand nonapeptide RERRRKKR ↓ G is given in Fig. 3. Catalysis takes place in canyon-like crevice, with the His194-Asp153-Ser368 catalytic triad arranged in its center. The binding pocket carries overwhelming hydrophilic surface, in accordance with the preference of furin for polybasic peptide. The catalytic mechanism of all serine proteases is similar, with a serine as a nucleophile attacking the susceptible bond. A close view of the interactions between the catalytic triad and the susceptible bond is shown in Fig. 4a, where the catalysis is similar to trypsin as discussed above,

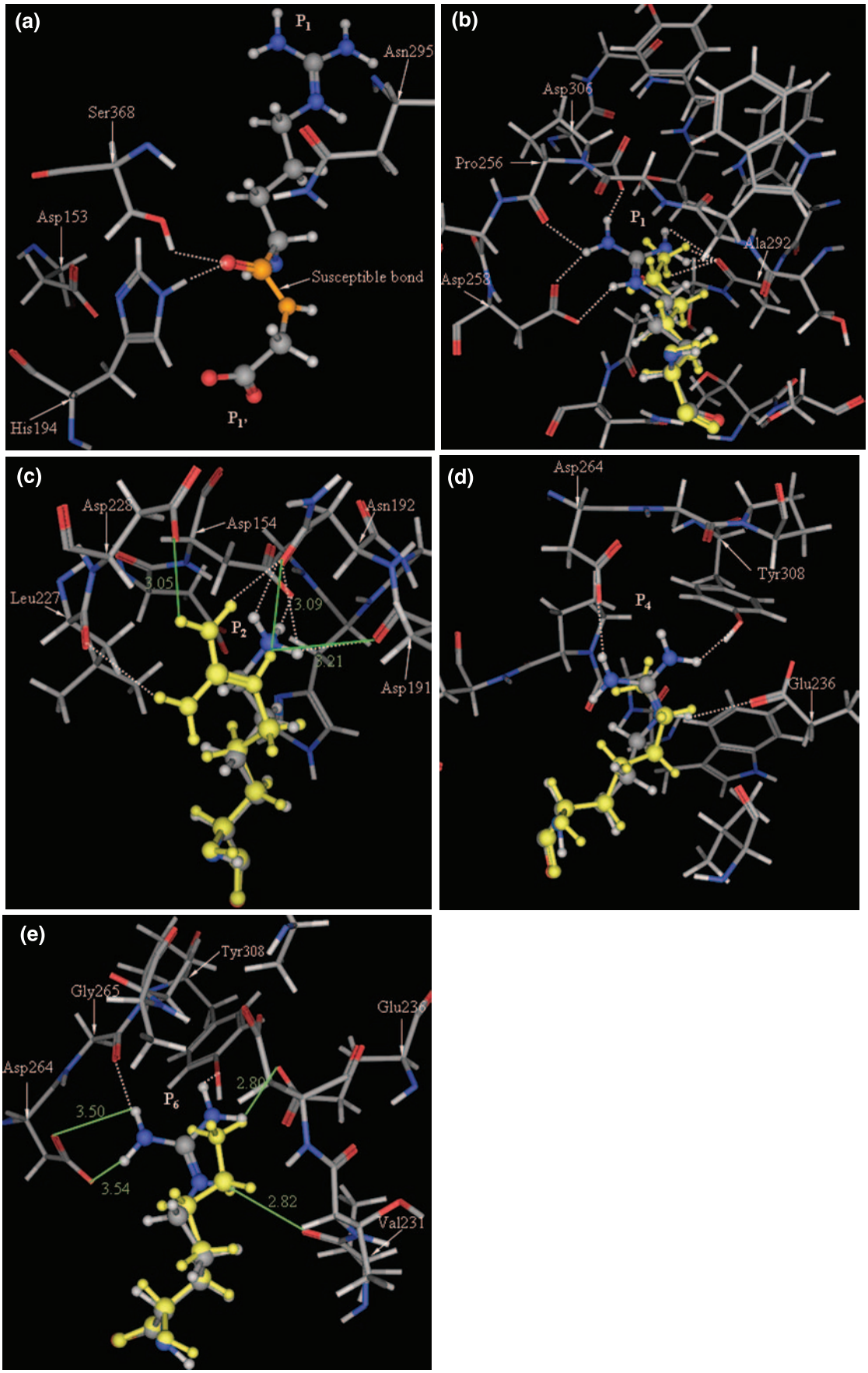
(a)
P1
Asn295
Ser368
Asp153
Susceptible bond
His194
P1'
(b)
Asp306
Pro256
P1
Ala292
Asp253
(c)
Asp228
Asp154
Asn192
3.05
Leu227
3.09
P2
3.21
Asp191
(d)
Asp264
Tyr308
P4
Glu236
(e)
Tyr308
Gly265
Glu236
Asp264
2.80
P6
3.50
3.54
2.82
Val231

Table 1. Energy (kcal/mol) contributions of each subsite

Subsite	E (binding)	E (electrostatic)	E (van der Waals)
$P_{1'}$-Gly	−10.46	−8.85	−1.61
P_1-Arg	−38.26	−27.15	−11.11
P_2-Lys	−19.30	−15.08	−4.22
P_3-Lys	−11.80	−8.61	−3.19
P_4-Arg	−23.97	−13.21	−10.76
P_5-Arg	−8.75	−4.26	−4.49
P_6-Arg	−25.09	−15.83	−9.26
P_7-Glu	−2.02	−1.79	−0.23
P_8-Arg	−9.37	−9.91	0.54
Total	−149.02	−104.69	−44.33

with the catalytic triad surrounding the susceptible bond. The carbonyl oxygen also inserts into an oxyanion hole formed by the carboxamide nitrogens of Asn295 and Ser368.

The binding free energy of the total interactions is −149.02 kcal/mol by taking into account the electrostatic and van der Waals potential fields. Energy contributions of each subsite are listed in Table 1, from which we can obtain the negative energy contribution in the order as given below:

$$P_1\text{-Arg} > P_6\text{-Arg} > P_4\text{-Arg} > P_2\text{-Lys} > P_3\text{-Lys} > P_{1'}\text{-Gly} > P_8\text{-Arg} > P_5\text{-Arg} > P_7\text{-Glu}.$$

Therefore, P_1-Arg, P_6-Arg, P_4-Arg and P_2-Lys are most important for binding interactions with dominant negative energy contributions (Table 1). These sites deserve further studies.

The S_1 binding pocket with overwhelming hydrophilic surface is constructed by residues S253-W254-G255-P256-E257-D258, W291-A292-S293-G294-N295, D306-G307-Y308-T309, and T367-S368. As shown in Fig. 4b, the P_1-Arg side chain passes the curved Ser253-Trp254-Gly255-Pro256-Asp257-Asp258 segment, and extends into a flat groove lined by the carboxylates of Asp258 and Asp306 and the carbonyls of Ala292 and Pro256. Six efficient hydrogen bonds are formed between the P_1-Arg side chain and the oxygens of these groups. Any other side chain, including that of the other basic amino acid lysine, fits less well into the S_1 pocket of furin. By an initial inspection, it appears that the binding of the P_1-Lys is identical to the binding of the P_1-Arg with their long basic side chains inserting into the S_1 pocket. However, it was observed by a closer inspection that the P_1-Lys binds to the S_1 pocket in a shallower way, only interacting with the carbonyl of Ala292 while losing the primarily interaction with Asp258, Asp306 and Pro256 due to the occupancy by lysine. The substrate substitution of lysine for arginine at P_1 position results in a docking energy weakening of 7.1 kcal/mol, consistent with the K_{cat}/K_m reduction of more than 160-fold according to the kinetic studies, suggesting the strict substrate requirement of furin for P_1-Arg (Holyoak et al., 2004; Wheatley and Holyoak, 2007).

The S_2 binding pocket is a surface crevice constructed by residues D153-D154, D191-N192-R193-H194-G195 and L227-D228. The ε-ammonium group of P_2-Lys side chain is surrounded by oxygens of the Asp154 carboxylate, the Asn192 carboxmide and the Asp191 carbonyl, for forming four efficient hydrogen bonds (Fig. 4c). Although substrate substitution of arginine for lysine at P_2 position results in a little weakening of 1.4 kcal/mol in the docking energy, the interaction mode changes, with the guanidinium group of P_2-Arg framed by the carbonyls of Leu227 and Asp191, the carboxylate of Asp228 and the carboxamide of Asn192. Only two hydrogen bonds are formed with oxygens of the Asn192 carboxamide and the Leu227 carbonyl. Kinetic studies showed that the value of K_{cat}/K_m decreased slightly according to this conservative substitution, even substitution of an alanine resulted in a 10-fold reduction in K_{cat}/K_m (Krysan et al., 1999). Thus, although the S_2 pocket lacks of strict selectivity for substrate, the hydrophilic surface and positive charges of this pocket are in favor for the lysine accommodation without excluding arginine.

The P_3-Lys extends into a bulk solvent, with its ε-ammonium group well contacting with the carboxlate of

Fig. 4. A close view of the interactions between furin and the nonapeptide, where the residues of furin are in stick drawing, the ligand in ball and stick drawing, and the white dotted lines represent the hydrogen bonds. (**a**) Interactions between the catalytic triad and the susceptible bond. (**b**) Interactions between P_1 residues and the S_1 binding pocket. The S_1 residues S253-W254-G255-P256-E257-D258, W291-A292-S293-G294-N295, D306-G307-Y308-T309, and T367-S368 are in stick drawing. Residues P_1-Arg and P_1-Lys are in ball and stick drawing, where the arginine is displayed in elementary color and the lysine in *yellow*. (**c**) Interactions between P_2 residues and the S_2 binding pocket. The S_2 residues D153-D154, D191-N192-R193-H194-G195 and L227-D228 are in stick drawing. Residues P_2-Lys and P_2-Arg are in ball and stick drawing, where the lysine is displayed in elementary color and the arginine in *yellow*. (**d**) Interactions between P_4 residues and the S_4 binding pocket. The S_4 residues V231, E236, W254-G255-P256-E257, D264-G265-P266, and Y308 are in stick drawing. Residues P_4-Arg and P_4-Lys are in ball and stick drawing, where the arginine is displayed in elementary color and the lysine in *yellow*. (**e**) Interactions between P_6 residues and the S_6 binding pocket. The S_6 residues E230-V231-T232-D233, E236, D264-G265-P266-A267, A270, and Y308 are in stick drawing. Residues P_6-Arg and P_6-Lys are in ball and stick drawing, where the arginine is displayed in elementary color and the lysine in *yellow* (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper)

Glu257 by a hydrogen bond. The geometry of the S_3 consists with the lack of preference at this site; however, long basic side chains are more favorable by the carboxylate of the surface-located Glu257.

The S_4 binding pocket is a cleft constructed by residues V231, E236, W254-G255-P256-E257, D264-G265-P266, and Y308. The guanidinium group of P_4-Arg is favorably framed by the carboxylates of Glu236 and Asp264 and the hydroxyl of Tyr308, for three efficient hydrogen bonding (Fig. 4d). P_4-Arg is ordered in a similar fashion of P_1-Arg, indicating an important dependence on P_4 recognition for the specificity of furin. Substrate substitution of lysine for P_4-Arg results in a docking binding energy weakening of 3.2 kcal/mol, with disappearance of three hydrogen bonds. Kinetic studies showed that the value of K_{cat}/K_m decreased 30-fold according to this conservative substitution, and substitution of an alanine resulted in a 2500-fold reduction in K_{cat}/K_m (Krysan et al., 1999).

The P_5-Arg runs across the loop Pro256-Glu257-Asp258-Asp259-Gly260-Lys261-Thr262-Val263-Asp264 and extends into a broad gap between Glu257 and Asp264, with its guanidyl group attracting by the carboxlate of Asp264. However, the geometry of the S_5 with the low energy contribution reveals the lack of preference at this site.

The S_6 binding pocket is a hollow constructed by residues E230-V231-T232-D233, E236, D264-G265-P266-A267, A270, and Y308. The guaninium group of P_6-Arg is well framed by the carbonyls of Gly265 and Val231, the carboxylates of Asp264 and Glu236 and the hydroxyl of Tyr308. Two efficient hydrogen bonds are formed with oxygens of the Gly265 carbonyl and the Tyr308 hydroxyl (Fig. 4e). P_6-Arg is ordered in a perfect fashion with the second largest energy contribution, playing an important role as P_1-Arg and P_4-Arg in substrate specificity. The P_6-Lys binds to the S_6 pocket in a shallower way, with its ε-ammonium group attracted by the Glu236 carboxylates and Tyr308 hydroxyl. Substrate substitution of lysine for P_6-Arg results in a docking binding energy weakening of 3.5 kcal/mol, with disappearance of hydrogen bonds.

The P_7-Glu is a negligible contributor to the energetics of substrate recognition, indicating the loose specificity in this position.

The P_8-Arg mainly senses a negative surface with its guanidyl group attracted by the carboxlate of Asp264 for two efficient hydrogen bonding. This position with the low energy contribution consists with the lack of preference at S_8; however, long basic side chains are more preferable by the carboxlate of the surface-located Asp264.

The docking result of the hexapeptide SIQSR↓G and furin with a weak binding free energy of −78.35 kcal/mol, indicates that the hexapeptide SIQSR↓G can not fit the substrate specificity of furin.

Discussion

The docking results proved that trypsin can recognize and cleave both the high pathogenic and low pathogenic hemagglutinin, while furin can only process the high pathogenic hemagglutinin with polybasic amino acids. According to energy contributions, furin generates most of its selectivity through interactions with P_1, P_4, and P_6, with interactions at P_2 being less important and little preference at P_3, P_5, P_7, and P_8. The S_1, S_4, and S_6 pockets are specifically designed to accommodate arginine, with lysine substitution fitting less well in different degrees. The substrate specificity for S_1, S_2, and S_4 in this model is consistent with the former furin studies with tetrapeptide inhibitors (Henrich et al., 2003, 2005; Holyoak et al., 2004; Rozan et al., 2004). While the P_6-Arg with the second largest energy contribution seems more important to the substrate specificity compared with the Kinetic studies using hexapeptidyl methylcoumarinamides (Krysan et al., 1999).

More and more cleavage sequences of the H5N1 HA have been mutated in their evolutions. Those mutations are classified into five types according to the number of basic amino acids: Seq.1 has the full number of basic amino acids; Seq.2 has the P_8-Arg replaced by an aliphatic one; Seq.3 has P_6-Arg replaced by an aliphatic one; Seq.4 has a basic amino acid deletion; Seq.5 has two basic amino acids mutations. Their definitions are given below:

Seq.1: RERRRKKRG, RER**K**RKKRG, RE**K**RRKKRG, RERR**K**KKRG, RERRRK**R**RG, RERRRKK**K**G;
Seq.2: **G**ERRRKKRG, **I**ERRRKKRG;
Seq.3: RE**G**RRKKRG, RE**I**RRKKRG;
Seq.4: RERR**-**KKRG, RERRR**-**KRG, REKR**-**KKRG.;
Seq.5: GEGRRKKRG

Different mutations result in docking energy weakening in different degrees (see Table 2). For the type of Seq.1, RERRRKK**K**G, RERR**K**KKRG and RE**K**RRKKRG fit less well into furin with the substitution of lysine for P_1-Arg, P_4-Arg or P_6-Arg, while the other mutation RERRRK**R**RG induces little change because of the similar preference for lysine and arginine at P_2 position. Also, Seq.2 shows little effect according to the little energy contribution of P_8 position. For the type of Seq.3 with short P_6 residues, the energy is weakened slightly, while

Table 2. The binding energies (kcal/mol) by docking different H5N1 HA cleavage sequences to furin

Cleavage sequences of H5N1 HA	E (binding)	E (electrostatic)	E (van der Waals)
RERRRKKRG	−149.02	−104.69	−44.33
RERRRKK**K**G	−141.93	−94.10	−47.83
RERRRK**R**RG	−147.62	−98.02	−49.60
RERR**K**KKRG	−145.87	−104.18	−41.69
RE**K**RRKKRG	−145.53	−100.08	−45.45
GERRRKKRG	−148.95	−99.84	−49.11
REI**I**RRKKRG	−146.95	−97.61	−49.34
GEGRRKKRG	−130.30	−92.52	−37.78

the substrate conformation changes, with the P_5-Arg threading into the S_6 pocket opposite to P_6 side chains. The situation of Seq.4 is similar to that of Seq.3, the P_5-Arg also threads into the S_6 pocket. While the last type with two basic amino acids mutations found in 2006 results in a large energy weakening of 18.7 kcal/mol. Most of today's H5N1 viruses possess the cleavage site of Seq.2 and Seq.4. If frequent mutation happens in proper positions, i.e., Seq.5, the substrate would fit less well and even reduce the high pathogenicity of the H5N1 virus.

Conclusion

In our studies, molecular insights of HA precursor in complex with cleavage proteases are provided that are very important for understanding the cleavage mechanism of trypsin and furin. The reason of high pathogenicity of H5N1was addressed through investigating a series of relevant binding interactions. Also, it was observed that some mutations of H5N1 H5 cleavage sequence fitted less well into furin and hence would reduce the high pathogenicity of the H5N1 virus.

Acknowledgements

This work was supported by the grant from Ministry of Science and Technology of the P.R.C under the contact number 2006AA02Z305, the grants from Chinese National key research program of basic science under the contract number 2005CB724303 and the grants from Chinese National Science Foundation under the contract numbers 20773085 and 30770502. Additional supports were from the Virtual Laboratory for Computational Chemistry of CNIC, and the Supercomputing Center of CNIC, Chinese Academy of Sciences.

References

Buck M, Bouguet-Bonnet S, Pastor RW, MacKerell AD Jr (2006) Importance of the CMAP correction to the CHARMM22 protein force field: dynamics of hen lysozyme. Biophys J 90: L36–L38

Chou KC (1993) A vectorized sequence-coupling model for predicting HIV protease cleavage sites in proteins. J Biol Chem 268: 16938–16948

Chou KC (1996) Review: prediction of HIV protease cleavage sites in proteins. Anal Biochem 233: 1–14

Chou KC (2004a) Insights from modelling the 3D structure of the extracellular domain of alpha7 nicotinic acetylcholine receptor. Biochem Biophys Res Commun 319: 433–438

Chou KC (2004b) Molecular therapeutic target for type-2 diabetes. J Proteome Res 3: 1284–1288

Chou KC (2004c) Review: structural bioinformatics and its impact to biomedical science. Curr Med Chem 11: 2105–2134

Chou KC (2005a) Coupling interaction between thromboxane A2 receptor and alpha-13 subunit of guanine nucleotide-binding protein. J Proteome Res 4: 1681–1686

Chou KC (2005b) Modeling the tertiary structure of human cathepsin-E. Biochem Biophys Res Commun 331: 56–60

Chou KC, Carlacci L (1991) Simulated annealing approach to the study of protein structures. Protein Eng 4: 661–667

Chou KC, Wei DQ, Du QS, Sirois S, Zhong WZ (2006) Review: progress in computational approach to drug development against SARS. Curr Med Chem 13: 3263–3270

Chou KC, Wei DQ, Zhong WZ (2003) Binding mechanism of coronavirus main proteinase with ligands and its implication to drug design against SARS (Erratum: ibid., 2003, Vol. 310, 675). Biochem Biophys Res Commun 308: 148–151

Du QS, Sun H, Chou KC (2007a) Inhibitor design for SARS coronavirus main protease based on ''distorted key theory''. Med Chem 3: 1–6

Du QS, Wang S, Wei DQ, Sirois S, Chou KC (2005) Molecular modelling and chemical modification for finding peptide inhibitor against SARS CoV Mpro. Anal Biochem 337: 262–270

Du QS, Wang SQ, Chou KC (2007b) Analogue inhibitors by modifying oseltamivir based on the crystal neuraminidase structure for treating drug-resistant H5N1 virus. Biochem Biophys Res Commun 362: 525–531

Du QS, Wang SQ, Wei DQ, Zhu Y, Guo H, Sirois S, Chou KC (2004) Polyprotein cleavage mechanism of SARS CoV Mpro and chemical modification of octapeptide. Peptides 25: 1857–1864

Hatta M, Gao P, Halfmann P, Kawaoka Y (2001) Molecular basis for high virulence of Hong Kong H5N1 influenza A viruses. Science 293: 1840–1842

Henrich S, Cameron A, Bourenkov GP, Kiefersauer R, Huber R, Lindberg I, Bode W, Than ME (2003) The crystal structure of the proprotein processing proteinase furin explains its stringent specificity. Nat Struct Biol 10: 520–526

Henrich S, Lindberg I, Bode W, Than ME (2005) Proprotein convertase models based on the crystal structures of furin and kexin: explanation of their specificity. J Mol Biol 345: 211–227

Holyoak T, Kettner CA, Petsko GA, Fuller RS, Ringe D (2004) Structural basis for differences in substrate selectivity in Kex2 and furin protein convertases. Biochemistry 43: 2412–2421

Ibrahim BS, Shamaladevi N, Pattabhi V (2004) Trypsin activity reduced by an autocatalytically produced nonapeptide. J Biomol Struct Dyn 21: 737–744

Kawaoka Y, Webster RG (1988) Sequence requirements for cleavage activation of influenza virus hemagglutinin expressed in mammalian cells. Proc Natl Acad Sci USA 85: 324–328

Kido H, Sakai K, Kishino Y, Tashiro M (1993) Pulmonary surfactant is a potential endogenous inhibitor of proteolytic activation of Sendai virus and influenza A virus. FEBS Lett 322: 115–119

Krysan DJ, Rockwell NC, Fuller RS (1999) Quantitative characterization of furin specificity. Energetics of substrate discrimination using an internally consistent set of hexapeptidyl methylcoumarinamides. J Biol Chem 274: 23229–23234

Ma W, Tang C, Lai L (2005) Specificity of trypsin and chymotrypsin: loop-motion-controlled dynamic correlation as a determinant. Biophys J 89: 1183–1193

Martin J, Wharton SA, Lin YP, Takemoto DK, Skehel JJ, Wiley DC, Steinhauer DA (1998) Studies of the binding properties of influenza hemagglutinin receptor-site mutants. Virology 241: 101–111

Morris GM, Goodsell DS, Huey R, Olson AJ (1996) Distributed automated docking of flexible ligands to proteins: parallel applications of AutoDock 2.4. J Comput Aided Mol Des 10: 293–304

Radisky ES, Lee JM, Lu CJ, Koshland DE Jr (2006) Insights into the serine protease mechanism from atomic resolution structures of trypsin reaction intermediates. Proc Natl Acad Sci USA 103: 6835–6840

Rozan L, Krysan DJ, Rockwell NC, Fuller RS (2004) Plasticity of extended subsites facilitates divergent substrate recognition by Kex2 and furin. J Biol Chem 279: 35656–35663

Schechter I, Berger A (1967) On the size of the active site in protease. I. Papain. Biochem Biophys Res Commun 27: 157–162

Schmidt A, Jelsch C, Ostergaard P, Rypniewski W, Lamzin VS (2003) Trypsin revisited: crystallography AT (SUB) atomic resolution and quantum chemistry revealing details of catalysis. J Biol Chem 278: 43357–43362

Sirois S, Wei DQ, Du QS, Chou KC (2004) Virtual screening for SARS-CoV protease based on KZ7088 pharmacophore points. J Chem Inf Comput Sci 44: 1111–1122

Steinhauer DA (1999) Role of hemagglutinin cleavage for the pathogenicity of influenza virus. Virology 258: 1–20

Stevens J, Blixt O, Tumpey TM, Taubenberger JK, Paulson JC, Wilson IA (2006) Structure and receptor specificity of the hemagglutinin from an H5N1 influenza virus. Science 312: 404–410

Wang JF, Wei DQ, Li L, Zheng SY, Li YX, Chou KC (2007a) 3D structure modeling of cytochrome P450 2C19 and its implication for personalized drug design. Biochem Biophys Res Commun (Corrigendum: ibid, 2007, Vol. 357, 330) 355, 513–519

Wang SQ, Du QS, Chou KC (2007b) Study of drug resistance of chicken influenza A virus (H5N1) from homology-modeled 3D structures of neuraminidases. Biochem Biophys Res Commun 354: 634–640

Wang SQ, Du QS, Zhao K, Li AX, Wei DQ, Chou KC (2007c) Virtual screening for finding natural inhibitor against cathepsin-L for SARS therapy. Amino Acids 33: 129–135

Wei DQ, Du QS, Sun H, Chou KC (2006a) Insights from modeling the 3D structure of H5N1 influenza virus neuraminidase and its binding interactions with ligands. Biochem Biophys Res Commun 344: 1048–1055

Wei DQ, Sirois S, Du QS, Arias HR, Chou KC (2005) Theoretical studies of Alzheimer's disease drug candidate [(2,4-dimethoxy) benzylidene]-anabaseine dihydrochloride (GTS-21) and its derivatives. Biochem Biophys Res Commun 338: 1059–1064

Wei DQ, Zhang R, Du QS, Gao WN, Li Y, Gao H, Wang SQ, Zhang X, Li AX, Sirois S, Chou KC (2006b) Anti-SARS drug screening by molecular docking. Amino Acids 31: 73–80

Wei H, Zhang R, Wang C, Zheng H, Chou KC, Wei DQ (2007) Molecular insights of SAH enzyme catalysis and their implication for inhibitor design. J Theor Biol 244: 692–702

Wheatley JL, Holyoak T (2007) Differential P1 arginine and lysine recognition in the prototypical proprotein convertase Kex2. Proc Natl Acad Sci USA 104: 6626–6631

Wiley DC, Skehel JJ (1987) The structure and function of the hemagglutinin membrane glycoprotein of influenza virus. Annu Rev Biochem 56: 365–394

Zhang R, Wei DQ, Du QS, Chou KC (2006) Molecular modeling studies of peptide drug candidates against SARS. Med Chem 2: 309–314

Authors' address: Dong-Qing Wei, College of Life Science and Biotechnology, Shanghai Jiao Tong University, Shanghai, 200240, China, E-mail: dqwei@sjtu.edu.cn