doi:10.1016/j.jmb.2004.07.014 *J. Mol. Biol.* (2004) **342**, 1533–1546

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

# The Structure of MBL-associated Serine Protease-2 Reveals that Identical Substrate Specificities of C1s and MASP-2 are Realized Through Different Sets of Enzyme–Substrate Interactions

**Veronika Harmat[1]†, Péter Gál[2]†, József Kardos[2,3], Katalin Szilágyi[2] Géza Ambrus[2], Barbara Végh[2], Gábor Náray-Szabó[1] and Péter Závodszky[2]***

[1]*Protein Modeling Group Hungarian Academy of Sciences Eötvös Loránd University Pázmány Péter sétány. 1A H-1117 Budapest, Hungary*

[2]*Institute of Enzymology Biological Research Center Hungarian Academy of Sciences P.O. Box 7, H-1518 Budapest Hungary*

[3]*Department of Biochemistry Eötvös Loránd University Pázmány Péter sétány. 1C H-1117 Budapest, Hungary*

**Corresponding author*

A family of serine proteases mediates the proteolytic cascades of several defense mechanisms in vertebrates, such as the complement system, blood coagulation and fibrinolysis. These proteases usually form large complexes with other glycoproteins. Their common features are their modular structures and restricted substrate specificities. The lectin pathway of complement, where mannose-binding lectin (MBL) recognizes the carbohydrate structures on pathogens, is activated by mannose-binding lectin-associated serine protease-2 (MASP-2). We present the 2.25 Å resolution structure of the catalytic fragment of MASP-2 encompassing the second complement control protein module (CCP2) and the serine protease (SP) domain. The CCP2 module stabilizes the structure of the SP domain as demonstrated by differential scanning calorimetry measurements. The asymmetric unit contains two molecules with different CCP-SP domain orientations, reflecting increased modular flexibility at the CCP2/SP joint. This flexibility may partly explain the ability of the MASP-2 dimer to perform all of its functions alone, whereas the same functions are mediated by the much larger $C1r_2$–$C1s_2$ tetramer in the C1 complex of the classical pathway. The main scaffold of the MASP-2 SP domain is chymotrypsin-like. Eight surface loops determine the S1 and other subsite specificities. Surprisingly, some surface loops of MASP-2, e.g. loop 1 and loop 2, which form the S1 pocket are similar to those of trypsin, and show significant differences if compared with those of C1s, indicating that the nearly identical substrate specificities of C1s and MASP-2 are realized through different sets of enzyme–substrate interactions.



*Keywords:* innate immunity; complement; serine protease; modular structure; hinge bending

## Introduction

The complement system is a key element of innate immunity in vertebrates. It is capable of recognizing and eliminating invading pathogenic microorganisms and altered host cells through opsonization and cell lysis. Complement is a sophisticated cascade system, where serine protease enzymes activate each other in a strictly ordered manner.

The complement serine proteases have several important features, which distinguish them from the well-characterized degradative proteases (e.g. trypsin, chymotrypsin, kallikreins, etc.). All these enzymes (except factor D) are modular proteins: the serine protease domain is preceded by several non-catalytic modules. Each protease has unusually restricted specificity for protein substrates and they

† V.H. & P.G. contributed equally to this work.

Abbreviations used: MBL, mannose-binding lectin; MASP-2, mannose-binding lectin-associated serine protease-2; SP, serine protease; CCP, complement control protein.

E-mail address of the corresponding author: zxp@enzim.hu

exert their function as part of larger glycoprotein complexes.

The first components of the classical and lectin pathways of complement activation are supramolecular enzyme complexes consisting of a recognition subunit and associated serine proteases.[1] The classical pathway C1 complex superficially resembles the initiation complex of the lectin activation pathway. The recognition subunit of the classical pathway is C1q, which resembles a bunch of six tulips consisting of N-terminal collagen-like arms and C-terminal globular heads. The globular heads bind to the activator structures (e.g. immune complexes), which results in the activation of the serine protease zymogens (C1r and C1s) associated with the collagen-like region. One C1q together with a heterotetramer of C1r and C1s proteases (C1s–C1r–C1r–C1s) form the C1 complex.[2,3] The first enzymatic event in the classical pathway is the autoactivation of zymogenic C1r. Activated C1r then cleaves zymogenic C1s, which in turn cleaves and activates C2 and C4, forming the components of the C3-convertase enzyme complex.

The recognition subunit of the lectin pathway, mannose-binding lectin (MBL) has C-terminal globular C-type lectin domains and N-terminal collagen-like stalks.[4] MBL binds to carbohydrate arrays on the surface of pathogens and triggers the activation of the complement cascade through MBL-associated serine protease-2 (MASP-2). MASP-2 is a C1s-like enzyme, performing a similar function by cleaving C4 and C2.[5,6] Unlike C1s, however, MASP-2 can autoactivate and therefore trigger the complement cascade without the assistance of any other protease.[7] It has been shown that a complex consisting of two MBL subunits and a dimer of MASP-2 molecules represents the minimal complement-fixing unit.[8] Consequently, a MASP-2 dimer is able to perform all functions mediated by the $C1r_2C1s_2$ tetramer in the C1 complex.

The serine protease (SP) domain contributes to the extremely narrow substrate specificity, in part, but the complement control protein (CCP) modules, which are connected directly to the SP domains, play an essential role as well. We have recently expressed and characterized the CCP1-CCP2-SP, CCP2-SP and SP fragments of MASP-2.[9] We showed that the SP domain is sufficient for autoactivation and can cleave the C2 substrate as efficiently as the intact molecule. For efficient C4 cleavage, however, the presence of the CCP2 module is essential. The CCP2 module may contain additional substrate binding sites for the C4 molecule. The CCP2-SP fragment can therefore be considered as the minimal catalytic unit of MASP-2.

Here for the first time, we report the 3D structure of the activated catalytic fragment (CCP2-SP) of an MBL-associated serine protease, human MASP-2, and discuss the structural background of the unusual characteristics of this enzyme. We compare the structure of MASP-2 with the structures of serine proteases participating in the complement cascade, such as C1r and C1s and in other physiological processes (examples include thrombin, trypsin and factor Xa). We pay special attention to the surface loops of the SP domain, which are the major determinants of the substrate specificity of late evolving serine proteases,[10] and to the role of the non-catalytic module in the physiological function of the protease.

## Results and Discussion

### Expression, structure determination and overall structure

The human recombinant MASP-2 CCP2-SP fragment was expressed in *Escherichia coli* BL-21 DE3 pLysS cells using the pET-17b expression vector. The recombinant construct (328 amino acid residues) contains an Ala-Ser-Met-Thr extra tetrapeptide at the N terminus, which is followed by the Ile363 residue of MASP-2. The purification and functional characterization of this fragment is described elsewhere.[9] Since human MASP-2 does not contain glycosylated side-chains, the recombinant protein produced in *E. coli* cells is identical with that isolated from natural sources.

The structure was solved by molecular replacement, and refined to 2.25 Å resolution (Table 1). At the end of refinement the $R_{work}$ and $R_{free}$ factors were 0.174 and 0.224, respectively. In all, 97% of the residues could be built in the electron density maps. Glycerol molecules and sodium ions, both present in the crystallization medium, were built in the electron density map (electron density map at the active site is shown in Figure 1). The asymmetric unit contains two molecules (denoted molecules A and B) with important differences in their conformations (Figure 2a).

The overall structure shows the spherical SP domain (240 residues) and the prolate shape CCP2 module (86 residues) attached with interdomain angle (the angle between the main axis of CCP2 and the SP surface) of 55° and 48° for MASP-2 molecule A and B, respectively. Figure 2a shows overall conformations of molecules A and B.

The SP domains of the two MASP-2 molecules are in virtually equivalent conformations, except for residues 440 and 441 (c10, c11) of the activation peptide and some surface side-chains. (Throughout the text MASP-2 numbering is used in comparisons with homologous proteins, together with chymotrypsin numbering marked with "c" for the SP domain.) Only the C-terminal residues of the cleaved activation peptide are disordered, which is typical for the activated SP structures. Structures of the SP domains of MASP-2 and structurally similar trypsin, thrombin and C1s are shown in Figure 2b.

The overall conformation of the CCP2 module with six β-strands (B1–B6, Figure 2c) shows the features typical among the CCP module structures. We observed slight twists at the two ends of the ellipsoidal CCP2 modules between molecules A

**Table 1.** MASP-2 data collection and refinement statistics

| | |
|---|---|
| *Crystal parameters* | |
| Space group | *P*1 |
| Cell constants | $a=40.95$ Å, $b=41.52$ Å, $c=102.99$ Å, $\alpha=96.4°$, $\beta=91.8°$, $\gamma=119.5°$ |
| *Data quality* ($I \geq -3\sigma(I)$) | |
| Resolution range (last resolution shell) (Å) | 35.8–2.23 (2.29–2.23) |
| $R_{merge}$ | 0.083 (0.255) |
| Completeness (%) | 94.8 (74.3) |
| *Refinement residuals (all reflections)* | |
| $R_{work}$ (no. observations) | 0.174 (25,665) |
| $R_{free}$ (no. observations) | 0.224 (1356) |
| *Model quality* | |
| RMS bond lengths (Å) | 0.006 |
| RMS bond angles (°) | 0.904 |
| RMS general planes (Å) | 0.003 |
| Ramachandran plot: residues in core/allowed/disallowed regions | 458/69/0 |
| *Model contents* | |
| Protein residues | 636 |
| Protein atoms | 4826 |
| Water molecules/ions/glycerol molecules | 362/2 $Na^+$/7 |
| Residues in dual conformations | 3 |
| Residues with disordered side-chains | 20 |
| Disordered residues | 15 |

and B (1.3–2.3 Å shifts of $C^\alpha$ atoms), which originated from conformational differences of the loops of the two modules. The overall main-chain difference between the CCP2 modules of the two MASP-2 molecules (RMS difference of $C^\alpha$ atoms 0.64 Å) is comparable to the differences between the CCP2 modules of MASP-2 and its homolog, C1s[11] (RMS difference of $C^\alpha$ atoms 0.79 Å). This refers to significant flexibility of the CCP2 module of MASP-2. The N-terminal segment and loop B4-B5 of molecule B are disordered. The structures of the CCP2 module of MASP-2, C1r[12] and C1s[11] aligned, are shown in Figure 2c.

The N-terminal end of the MASP-2 fragment is a part of the CCP1-CCP2 linker region. In molecule A it has a conformation similar to that detected in the C1r CCP1-CCP2-SP fragment structures,[13] supporting the previous assumption[14] that the configuration around the CCP1–CCP2 junction is similar to that of C1r.

## SP domain: some of the surface loops are more similar to trypsin or thrombin, than to C1s

The SP domain of MASP-2 has the conserved core structure of the chymotrypsin fold consisting of two six-stranded β-barrel domains packed against each other, with the catalytic residues Ser633 (c195),

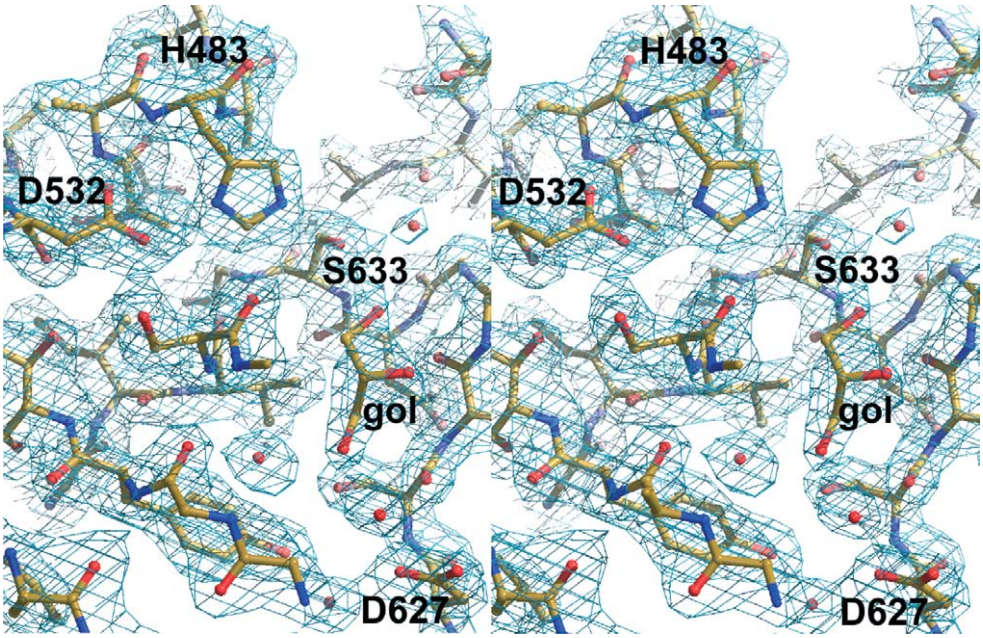


**Figure 1.** Stereo view of $\sigma_A$-scaled $2F_o-F_c$ map in the active site region of MASP-2 (molecule A). The map is contoured at the 1.1 σ level in light blue.

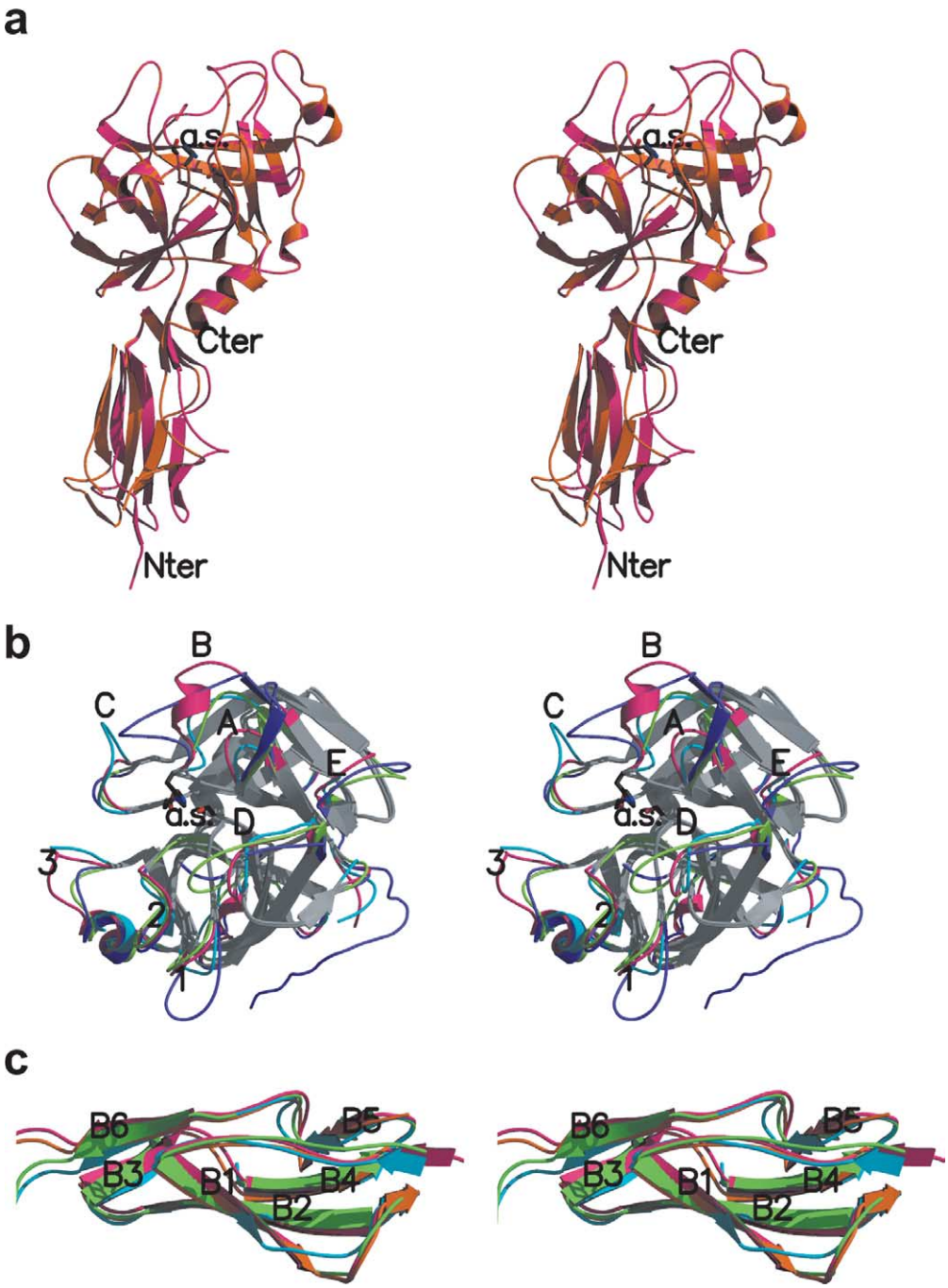


**Figure 2.** Overall structure of the MASP-2 CCP2-SP fragment. a, Stereo ribbon diagram of molecule A (magenta) and molecule B (orange) of the crystal structure, with the SP domains superimposed. The N and C termini are labeled. Residues of the catalytic triad are shown as sticks and are labeled "a.s." b, Stereo diagram of the active SP domain of molecule A (loop regions magenta) superimposed over that of C1s (pdb entry 1elv, loop regions light blue), porcine trypsin (pdb entry 1avw, loop regions green) and thrombin (pdb entry 1k21, loop regions dark blue). The structurally conserved regions of the proteins are colored gray for the sake of clarity. The loops are labeled as described by Perona & Craik.[10] The active site residues are shown as sticks. c, Stereo diagram of the CCP2 module of molecule A superimposed over that of C1s (light blue) and C1r (pdb entry 1md8, green). The β-strands are labeled. There are three regions with major conformational differences: loop B1-B2, loop B3-B4, and loop B5-B6. Although there are only minor differences in the positions of the loop region B2-B3 and residues 430–433 of the hinge region, these contribute directly to the significantly different relative module orientation observed in the case of MASP-2 compared to that of C1r and C1s.

His483 (c57) and Asp532 (c102) located at the junction of the two barrels (Figure 2b). The structure shows the elements of the catalytic apparatus in their active conformation. The oxyanion hole, as well as Asp627 (c189), the primary determinant of the S1 specificity, is in a canonical conformation in the MASP-2 crystal structure. Ser654 (c214) is hydrogen bonded to

Asp532 (c102) as seen in several enzyme structures of this family.

In late evolving enzymes the structural determinants of substrate specificity are more complex than in their early evolved counterparts and also contain loop structures that modulate substrate recognition and binding.[10] To compare the structural features of these loops with the substrate specificities of the respective enzymes we superimposed the structure of MASP-2 on other members of the chymotrypsin family including enzymes of the complement, digestion and coagulation systems, proteases with broad, as well as high substrate specificities (Table 2). Although, MASP-2 has similar functions to complement components C1r (autoactivation upon recognition of target surface by MBL/C1q) and C1s (C2 and C4 cleavage) the length, amino acid composition and conformation of most surface loops in the SP domain of MASP-2 is different from both enzymes. Surprisingly, some of the surface loops are more similar to thrombin or trypsin, than to C1r or C1s.

The substrate specificities of late evolving enzymes are defined and modulated on top of the S1 site amino acid by various loop interactions. Loops A, B, D and E are believed to contact the substrate from the leaving group side, whereas loop C interacts with the N-terminal part of the bond to be cleaved (we use the loop nomenclature proposed by Perona & Craik[10]). Loops 1 to 3 can modulate the S1 site specificity and loops 2 and 3 have the potential to alter subsite preferences. Loop A of MASP-2, C1r and C1s are similar in position, but their conformations are different, in spite of the fact that loop A of MASP-2 and C1r are of the same length. A unique characteristic of loop region B is the helix conformation of its first six residues (485–490, c59–c60d). In the case of thrombin and C1r, loop B restricts access to the active site. The shorter loop B of MASP-2 leaves the active site more exposed, similarly to that of C1s, factor IXa and digestive enzymes. Backbone conformation of loop D of MASP-2 is surprisingly similar to that of kallikrein. While loop D of C1s and C1r possess similar positions, they leave the substrate-binding groove more accessible. Loop E, which binds $Ca^{2+}$ for some members of the family, lines the substrate-binding groove from the far end at the leaving group side.

In C1s loop C is significantly longer than in MASP-2 and it restricts the access to the active site. In MASP-2 at the closest point of loop C to the active site the side-chains of Asp526 (c97) and His525 (c96) line a more open S2 site. In MASP-2 it is loop 2 that has an insertion making the substrate-binding site narrower, but from the opposite side of the S1–S3 sites. Loops 1 and 2 form the bottom and one side of the substrate specificity pocket, and they have conformation very similar to that of trypsin. While the conformation of most of loop 2 is similar to e.g. that of trypsin and coagulation enzymes, its conformation around Met658 (c218) is similar to that of chymotrypsin and kallikrein forming the site for the P3 side-chain. Loop 3 closes the substrate-binding groove from the N-terminal end with its 605–608 segment containing three proline residues. This region is of low polarity and is presumably very rigid. Loop 3 of C1s is longer (and disordered) and leaves the substrate-binding groove more open than that of MASP-2. Therefore, it is very likely that there are differences between MASP-2 and C1s in the significance of subsites in substrate recognition and specificity.

In the case of the C1r CCP1-CCP2-SP dimer structure, some residues of loop regions B and E are involved in intermolecular contacts between the CCP1 module of one monomer and the SP domain of the other. The corresponding loops of MASP-2 are different in length and also in conformation, indicating that in contrast to a recent model of MBL–MASP-2 complex,[14] the interactions and the way of dimer formation observed for the zymogen form of C1r cannot be applied directly to MASP-2.

### Importance of subsite flexibility in substrate specificity of MASP-2

MASP-2 has only a few natural substrates: zymogenic MASP-2, C2, C4, and the pseudo-substrate C1 inhibitor, suggesting that the access to the substrate-binding subsites is restricted. While all of them have Arg in the P1 position (nomenclature proposed by Schechter & Berger[15]), the surrounding residues among these molecules vary in size and polarity, requiring some flexibility of the subsites (Figure 3a). Structure determination of the SP domain complexed with a specific peptide inhibitor would be required for detailed analysis of substrate-binding subsites. However common features of substrate-binding subsites found in the crystal structures of several members of the chymotrypsin family complexed with inhibitors have been extensively studied. We superimposed such structures onto the MASP-2 SP domain in order to characterize the possible subsites (Figure 3b). We found that the S1 site is deeper than that of C1s. In MASP-2 the S2′ and S3 subsites may establish more contacts with the substrate, while the S2 and S1′ subsites are more exposed than those of C1r and C1s. Although MASP-2 shares substrates with C1s, most of its substrate-binding subsites show differences.

As described above, the conformations of loops 1 and 2 are similar to those of trypsin forming an S1 pocket deeper than that of C1r and C1s. In contrast to C1s but similarly to C1r, the access to the substrate-binding pocket is not affected by the disordered side-chain of residue c192, as the Arg630 (c192) side-chain carbon atoms are stabilized by hydrophobic contacts with Leu575 (c143). In the crystal structure a glycerol molecule is bound at the entrance of the S1 pocket (Figure 1). An electron density peak with octahedral coordination sphere nearby two negatively charged residues (Asp627 c189 and Glu662 c221) is modeled as $Na^{+}$. $Mg^{2+}$ was excluded as a possible bound ion because the

**Table 2.** Surface segments of MASP-2 significantly different ($D > 1.5$ Å) from the homologous segments of other SP domains

A. *Overall structure*

| | C1s[a] | C1r | Trypsin | Chymo-trypsin | Thrombin | Protein C | Factor VIIa | Factor IXa | Factor Xa | Kallikrein 6 | Hepsin | Matriptase | Enteroki-nase |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PDB id | 1elv | 1md8 | 1avw | 1ab9 | 1k21 | 1aut | 1cvw | 1rfn | 1mq5 | 1lo6 | 1p57 | 1eax | 1ekb |
| No. $C^{\alpha}$ | 199 | 213 | 202 | 204 | 213 | 203 | 205 | 203 | 201 | 193 | 210 | 206 | 210 |
| r.m.s.d. (Å) | 0.91 | 1.06 | 0.98 | 1.02 | 0.92 | 1.02 | 1.05 | 1.07 | 0.97 | 1.01 | 0.95 | 0.96 | 1.05 |

B. *Surface segments of MASP-2*

| MASP-2 numb. | Chymotrypsin numb. | Loop label[b] | Type of variation | Conformation similar to |
|---|---|---|---|---|
| 436–441 | 3–8 | Activation peptide | | |
| 451–454 | 22–25 | | Variable | All, except for C1s |
| 463–469 | 34–43 | A | Deletion | |
| 485–496 | 59–65 | B | Insertion | |
| 503–511 | 72–82 | E | Deletion | |
| 524–528 | 95–98 | C | Insertion | |
| 555–562 | 125–130 | | Insertion | Thrombin, factors VIIa, IXa, Xa |
| 575–582 | 143–152 | D | Deletion | C1s, kallikrein |
| 594–611 | 164–175 | 3 | Major insertion | |
| 621–625 | 185–187 | 1 | Insertion | Trypsin, hepsin, matriptase, enterokinase |
| 641–645 | 203–205 | | Insertion | C1r, thrombin |
| 657–665 | 217–224 | 2 | Insertion | Trypsin, thrombin, factors VIIa, Xa, IXa, hepsin, protease C, matriptase, enterokinase |

[a] The references for the crystal structures are 11, 12, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57 for C1s, C1r, trypsin, chymotrypsin, thrombin, protein C, factors VIIa, IXa and Xa, kallikrein 6, hepsin, matriptase and enterokinase, respectively.

[b] Loop labels as defined by Perona & Craik.[10]

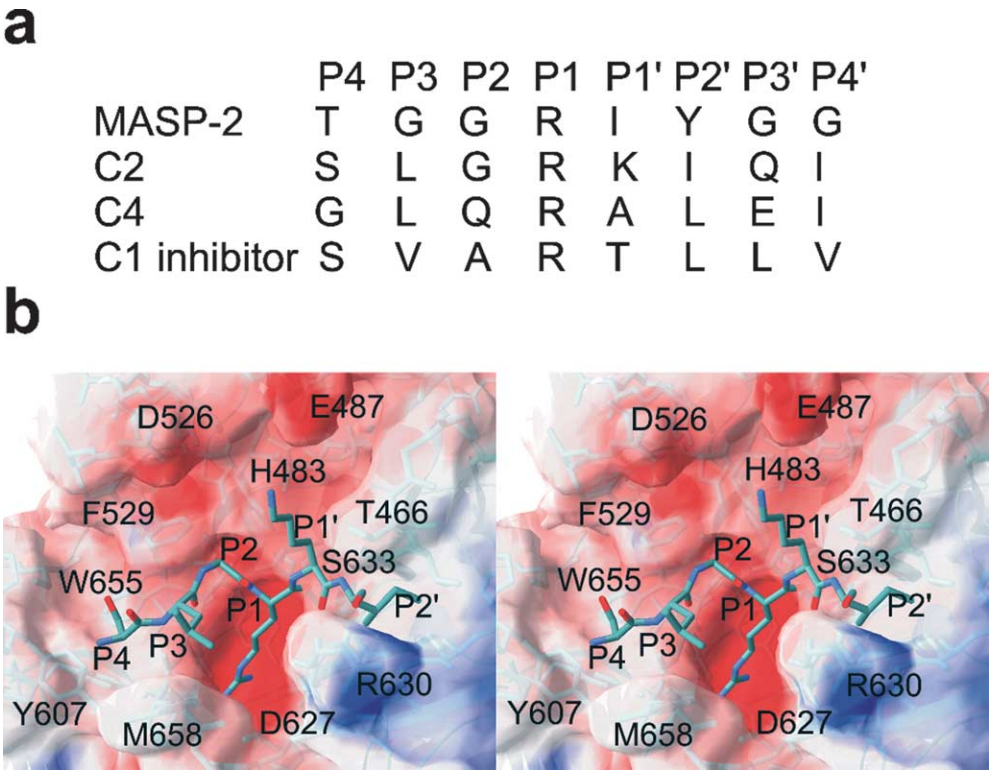


**Figure 3.** Substrate binding subsites of MASP-2. a, Sequences of the P4-P4′ segments are shown for the natural substrates: MASP-2, C2 and C4, as well as for the pseudo substrate C1 inhibitor. b, Stereo view of the molecular surface of MASP-2 substrate binding subsites region colored for electrostatic potential (red, negative; blue, positive; plotted on a −10 to +10 scale). A model peptide (drawn as sticks) representing the P4-P2′ residues of C2 is shown superimposed over the MASP-2 structure.

protein was purified and crystallized in the absence of $Mg^{2+}$ but in the presence of $Na^{+}$. The sodium ion is bound in a small cavity formed by the 657–662 (c217–c222) segment of loop 2, which possesses a conformation similar to that of chymotrypsin. It does not establish salt-bridges, but it is connected to the negatively charged residues of the protein *via* a water molecule and a serine $O^{\gamma}$ atom of its coordination sphere. This suggests that the sodium ion can be easily dissociated upon substrate binding. This is in accordance with the results of our solution kinetic studies that have shown no inhibitory effect of sodium (Figure 4).

The groove that binds the N-terminal part of the bound peptide is shallow compared to that of C1r and C1s. The S2 subsite is shallow, Phe529 (c99) is in a position similar to that of C1r and C1s. The side-chain of the P2 residue is partly exposed to the solvent, while this site is buried by loop C in C1s. Water-mediated hydrogen bonds may be

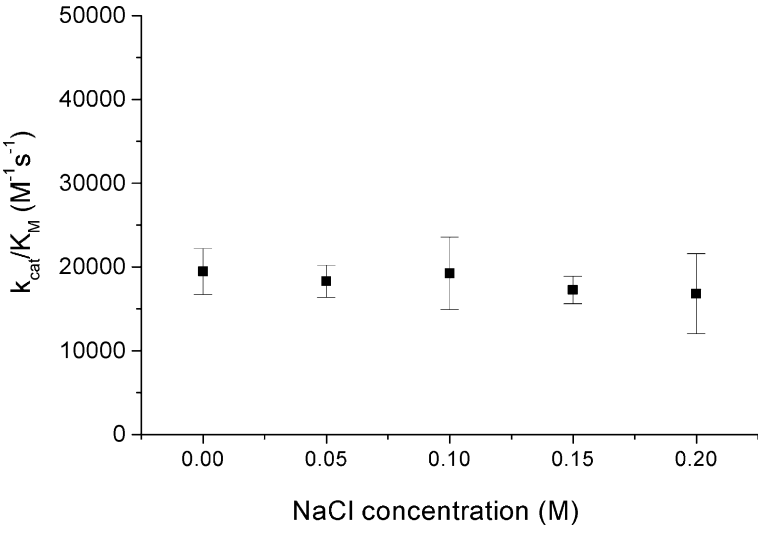


**Figure 4.** The effect of NaCl concentration on the enzymatic activity of MASP-2 CCP2-SP fragment at constant ionic strength. The enzymatic assay is described in Materials and Methods.

established by a P2 Gln side-chain and Tyr523 (c94) as well as Gln526 (c96a) side-chains of loop C. In the S3 subsite hydrophobic interactions can be established by Met658 (c218) of loop 2 with the apolar P3 side-chain of the substrate. Hydrogen bonds formed by P3 and residue Gly656 (c216) stabilize the backbone of the bound peptide.

On the leaving group side the access to the subsites is more restricted, although the S1′ site is open, like in C1r and C1s. The small P1′ side-chains of C4 and C1 inhibitor can contact Thr466 (c37), while the P1′ Lys side-chain of C2 may form a salt bridge with Glu487 (c60a) of loop B. P2′ side-chains are hydrophobic or aromatic, and are bound in a hydrophobic pocket formed by Gly631 (c193) and side-chain carbon atoms of Arg630 (c192), Leu581 (c148), Leu575 (c143) and Thr467 (c41). This subsite is a hydrophobic pocket also in C1r and C1s, although it is built up by different residues in the three enzymes. At the far end of the leaving group side loop E may contact the substrate, while that of C1s, C1r, thrombin or hepsin is partially buried by loops A or B.

## Inability to be induced by $Na^+$: structural basis is similar to that of trypsin

The structural basis of the activity-enhancing effect of sodium binding of several haemostatic proteases has been extensively studied.[16,17] The orientation of the c224 carbonyl oxygen has a critical role in $Na^+$ binding, as it is one of the oxygen atoms coordinating $Na^+$. A necessary condition of the sodium-inducibility is Phe or Tyr at the 225 position of the sequence. Enzymes with Pro at the 225 position, such as trypsin and chymotrypsin, do not show sodium inducibility, because the conformation of the 224–225 peptide moiety is different and the 224 carbonyl oxygen atom is in a wrong position for coordinating $Na^+$. The open base of the S1 site connected to the $Na^+$ site through a water channel is an additional feature of the coagulation enzymes that is not a characteristic of trypsin and chymotrypsin.

Since MASP-2 has Tyr at position c225, we examined the effect of $Na^+$ on its esterolytic activity towards synthetic substrate. The results of our solution studies show no activation effect of sodium on MASP-2 (Figure 4). The crystal structure is in accordance with these results. It shows a closed S1 site with the Gln665 (c224) carbonyl oxygen atom in the wrong position to bind $Na^+$. Surprisingly, the position of this carbonyl oxygen atom is similar to that of trypsin, and different from closer homologs of MASP-2, such as C1r and C1s. Moreover, the backbone conformation of the 661–667 (c221A–c226) segment of loop 2 is virtually identical with that of trypsin. The trypsin-like conformation of the 665–666 (c224-c225) backbone regions is stabilized by backbone–backbone interactions with loop 1 similar to those found in trypsin. The consequent steric closure of the carbonyl groups of Leu621 (c185) and Gln665 (c224), forces the latter to keep its position similar to that of trypsin.

## Semi-flexible CCP2/SP interface

The comparison of MASP-2 molecules A and B exposed differences in the conformations of the regions of the CCP2 module (loop region B2-B3 and residues 430–433 of the hinge region) that form the CCP2/SP interface. Moreover we observed alternatives of the interface interactions resulting in

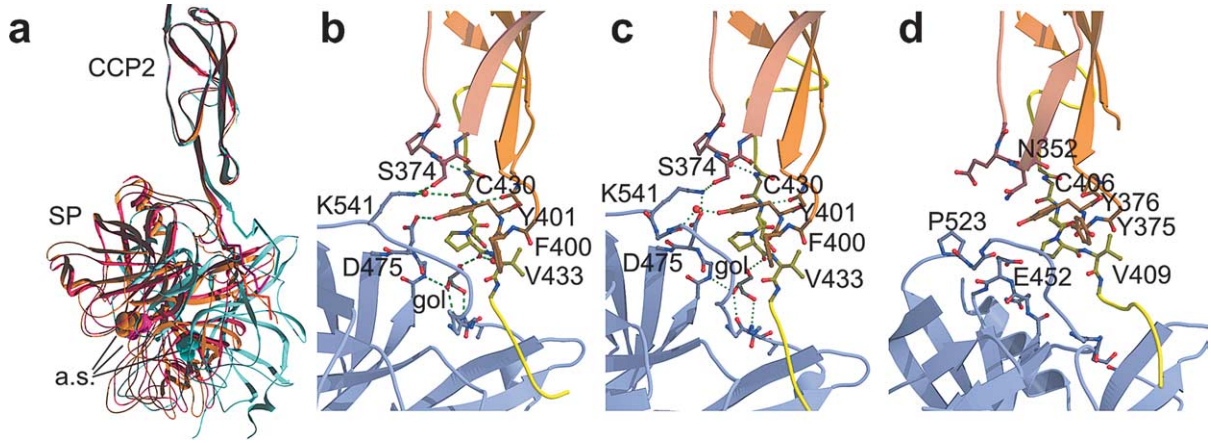


**Figure 5.** The CCP2/SP interface region. a, Backbone conformations of MASP-2 molecule A (magenta), molecule B (orange) and C1s (light blue), are shown with the CCP2 module superimposed. Due to differences in the topology of the interface contacts, the active site of MASP-2 is shifted significantly compared to that of C1s (serine c195 side-chain is shown as space filling representation). The CCP2/SP interface region of MASP-2 molecule A, molecule B, and C1s is shown in b, c and d, respectively. Residues forming interdomain contacts in MASP-2 are shown as balls and sticks, as well as the corresponding residues of C1s. In MASP-2 these residues form a hydrogen bond network (shown as green dotted lines), which is not established in C1s. For the sake of clarity, some side-chain atoms are not shown. The carbon atoms and the ribbon representation of SP domains are shown in blue, while those of the interdomain linkers are in yellow, the N-terminal loop and B1 are in rose and the rest of CCP2 module is in orange, respectively.

different domain orientations of molecules A and B (Figure 5a).

Different CCP2 module/SP domain orientations were also observed between structures of different fragments of C1r.[12] The present structure of MASP-2, however, demonstrates that different CCP2/SP orientations can also exist within the same physicochemical circumstances (i.e. same crystal), indicating that this kind of flexibility may have functional relevance. The network of hydrogen bonds and van der Waals contacts at the interface is very similar among the structures of zymogenic C1r CCP1-CCP2-SP, active form of C1r CCP2-SP and C1s. Surprisingly, in both molecules of the MASP-2 structure new patterns of interface interactions were found (Figure 5b–d), indicating that the flexible CCP2/SP junction of MASP-2 differs from that of the related complement proteases. The C1r CCP2-SP zymogen form can be considered as an intermediate conformation between that of C1s and MASP-2 structures, possessing some of the hydrogen bonds and contacts of both type, but with elongated interatomic distances.

The differences of the interface region cause rotation of the SP domain with respect to the CCP2 module within the homologous structures of MASP-2, C1r and C1s. Alignment of the CCP2 structures reveals large displacements of the catalytic residue Ser633 (c195) $C^{\alpha}$ atoms from that of MASP-2 for C1s (distances of 11.1 Å and 12.2 Å) and C1r (distances between 9.1 Å and 13.9 Å), respectively. These differences are significantly larger than those found between C1r and C1s (maximal distance 6.6 Å). However, the area of the buried accessible surface of the interface is similar for these molecules: 722 $Å^2$, 748 $Å^2$ and 751 $Å^2$, for MASP-2 molecules A, B and C1s, respectively, and 578 $Å^2$–667 $Å^2$ for C1r structures.[11-13]

Though many of the interacting residues of the interface correspond to those described for C1r and C1s, there are also striking differences. Detailed structure of the interdomain interface of MASP-2 compared to C1s is shown in Figure 5b–d. Conserved residues Phe400 and Tyr401 of the B3-B4 loop are bound between the hydrophobic CCP2-SP linker and 541–549 (c111–c119) and 474–475 (c48-c49) regions of the SP domain, as found in C1r and C1s. However, the slight bending within the linker residues 429–434 of MASP-2, results in a shift of Phe400: its side-chain is above Leu544 (c114)-Asn545 (c115), while the corresponding tyrosine is located above the dipeptide moiety of C1s with next register (Gly527 c115 and Pro528 c116). Because of the hinge bending, the Tyr401 side-chain is too far from the Ile544 (c114) backbone NH so the hydrogen bond found in the C1s and C1r structures is not established. The Tyr401 hydroxyl group of molecule B is connected to Val542 (c112) carbonyl oxygen atom through a water molecule (W325), while that of molecule A is rotated further away and is stabilized by a hydrogen bond with the Asp475 (c49) side-chain. In both molecules an interdomain hydrogen bond is formed by Lys541 (c111) and Ser374 side-chains, which is not found in either of the C1s or C1r structures.

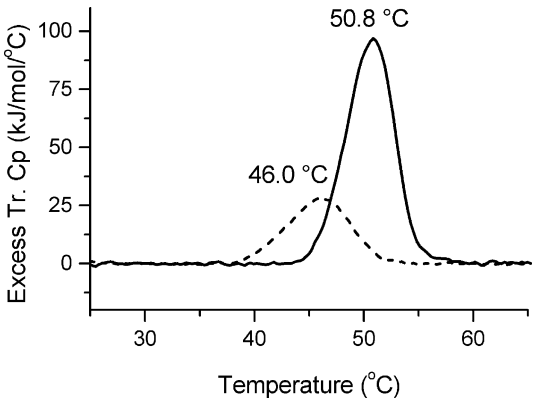


**Figure 6.** DSC melting profiles of MASP-2 fragments. Excess transition heat capacity of the SP (- - -) and CCP2-SP (—) fragments in a buffer of 20 mM Hepes (pH 7.0), 145 mM NaCl. Melting temperatures are indicated.

For both variants of MASP-2 structures a hydrogen-bonded network of water molecules stabilizes the interface. In both molecules a glycerol molecule is bound in a cavity formed by side-chains of Leu473 (c47), Tyr474 (c48), residues Glu431, Pro432, Cys434 (c1) and backbone atoms of 550 (c120)–552 (c122) (Figure 5b and c). This cavity is more accessible and it has a more hydrophobic character in C1r and C1s. While C1r and C1s have five and four proline residues, respectively, in the interface region, MASP-2 has only two, which may contribute to its higher flexibility.

As demonstrated by the differential scanning calorimetry (DSC) measurements the presence of the CCP2 module stabilizes the structure of the SP domain (Figure 6). While the fragment containing only the SP domain showed a melting transition at a relatively low temperature (46.0 °C), the CCP2-SP fragment unfolded at a substantially higher melting point (50.8 °C). Moreover the melting curve of the CCP2-SP fragment is much sharper than that of the SP domain, indicating a more cooperative unfolding process. The larger calorimetric enthalpy change and the significantly higher melting temperature of the CCP2-SP fragment indicate that the CCP2 module establishes tight interaction with the SP domain and significantly improves its stability. These results show that the CCP2 module and the SP domain form a cooperative folding unit. The results of the DSC measurements are in agreement with the crystal structure: both provide proof of the existence of strong interactions between the two domains.

It is remarkable that while the CCP2/SP interaction is tight and contributes significantly to the stability of the structure, it is also a flexible hinge region that allows the change of the relative position of the two domains without compromising the stability. It is interesting to note that the SP domains of C1r, C1s and MASP-2 contain only two disulfide

bridges each, whereas the trypsin-like enzymes of vertebrates contain usually four to six disulfide bridges.[18] Human trypsins and the single-domain factor D complement protease contain five and four disulfide bridges, respectively. In the case of C1r, C1s and MASP-2 CCP-SP structures, the CCP/SP interface could have taken over the stabilizing function of certain disulfide bridges of the serine protease domain.

Other examples of multidomain proteases of the chymotrypsin family are members of the blood coagulation cascade. The epidermal growth factor (EGF)-like domain of factors VIIa, IXa, Xa and protein C preceding the SP domain participate in cofactor binding[19,20] and have extensive interdomain interface (typical values of the buried area range between 1300 and 1500 Å$^2$, with the number of interdomain contacts shorter than 3.9 Å being between 75 and 90). In the structure of hepsin, a membrane-bound protease, the buried surface between the SP and the scavenger receptor cysteine-rich (SRCR) domain is even larger (1703 Å$^2$, with 98 interdomain contacts). It is supposed that the role of the SRCR domain is the mediation of protein–protein interactions, e.g. orientating the substrate or the SP domain. In complement enzymes MASP-2, C1r and C1s, the interdomain interface is significantly smaller (the buried area is between 578 and 751 Å$^2$ with interdomain contacts of 18 to 52). The anchor site of the CCP domain on the surface of the SP domain is also apparently different from that of the SRCR and EGF domains. These facts strengthen our view that the CCP/SP interface could serve as an important hinge region for the complement proteases.

## Functional implications

MASP-2 plays a central role in the initiation of the lectin pathway of complement, since it is capable of autoactivating and cleaving C4 and C2, the precursors of the C3 convertase enzyme complex. The structure described here provides the first insight into the catalytic machinery of a protease of the lectin pathway, a constituent of innate immunity.

According to the current model of the initiation enzyme complex of the lectin pathway a homodimer of MASP-2 associated with an MBL molecule can trigger the complement cascade.[8,21] The MASP-2 dimer binds to the collagen-like stalks of MBL, like the $C1r_2C1s_2$ tetramer binds to C1q in the C1 complex. The nature of these bindings and the structure of the supramolecular complexes are probably similar, since *in vitro* experiments showed that MBL can also bind and activate the $C1r_2C1s_2$ tetramer [22,23]. The current models of MBL–MASP-2[14] and C1[12,24] complexes assume considerable flexibility of both the recognition molecules and the associated serine proteases during the activation process. Our present study strengthens these views: we observed significant flexibility at the CCP2/SP interface of MASP-2.

The crystal structure of C1s shows the CCP2 module tightly fixed at the surface of the SP module. Comparing the structures of the zymogenic and activated forms of the CCP2-SP fragments of C1r, some differences were detected between the relative positions of the two domains. In the case of MASP-2, however, we observed two distinct conformers of the same polypeptide chain, suggesting a functionally relevant interdomain flexibility. This conclusion is also supported by the fact that the two conformers are stabilized by different interactions. These facts suggest that MASP-2 possesses the most flexible CCP2/SP junction among the members of the C1r/C1s/MASPs enzyme family.

Unlike C1r, the MASP-2 molecules form dimers *via* interactions of the N-terminal CUB1-EGF-CUB2 (CUB: C1r/C1s/sea urchin Uegf/bone morphogenic protein) region rather than the C-terminal region, so the SP domains of the MASP-2 monomers are at the opposite ends of the dimer. It is likely, therefore, that significant flexibility is needed to place the distal SP domains of the MASP-2 dimer in the correct position during autoactivation. The subsequent cleavage of C2 and C4 substrates also requires significant conformational movements of the MASP-2 dimer, and especially the SP domains. In other words, the closed circular conformation of the MASP-2 dimer should be converted into an open form. It is remarkable that the MASP-2 homodimer can perform all those functions (e.g. binding to MBL, autoactivation, cleaving C4 and C2) that are mediated by the C1s–C1r–C1r–C1s tetramer in the C1 complex. Theoretically, the MASP-2 homodimer should be at least as flexible as the $C1r_2C1s_2$ tetramer. Nevertheless, the $C1r_2C1s_2$ tetramer has twice as many hinge points as the MASP-2 dimer, to produce the same level of flexibility. We can hypothesize, therefore, that the other potential hinge points of MASP-2 (e.g. the CCP1-CCP2 and the CUB2-CCP1 junctions) are more flexible than the corresponding regions of C1s and possibly even those of C1r.

As we demonstrated earlier, the CCP2 and SP domains determine the substrate specificity of MASP-2 towards natural substrates [9]. The SP domain of MASP-2 and C1s contains all necessary contact sites for efficient C2 binding and cleavage and it forms a covalent complex with C1-inhibitor. It is surprising, therefore, that most of the surface loops, which determine S1 and other subsite preferences, exhibit different conformations in the two, functionally closely related, and highly specific SP domains. It is likely that the same substrate specificity can be realized through different enzyme–substrate interactions. Building upon these differences and using the MASP-2 structure presented here it is possible to design synthetic inhibitors that would specifically block the unwanted pathological activation of the lectin pathway (e.g. in the case of ischemia-reperfusion), without interfering with other activation routes.

As for C4 cleavage, the presence of the CCP2

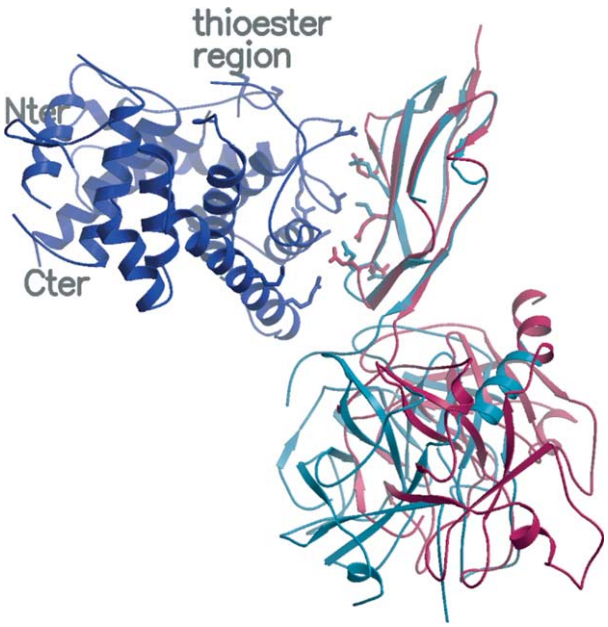


**Figure 7.** Possible binding site of MASP-2 and C1s on C4d. The model is constructed on the basis of structural homology between C3d and C4d. Superimposed CCP2 domains of MASP-2 (magenta) and C1s (pdb entry 1elv, light blue) were aligned with C4d (dark blue, pdb entry 1hzf, its C and N termini and thioester region labeled). Glu378, Arg376 and Glu397/Glu398 of MASP-2 as well as Glu356, Lys354 and Glu372/Glu373 of C1s contacts oppositely charged regions of C4d surface (Arg1041/Lys1080, Asp1054/Asp1044/Glu1083 and Lys1155/Arg1148).

module is necessary for the efficient reaction to occur. We can presume that the CCP2 module contains additional binding sites for the C4 substrate. This assumption is supported by kinetic data, since the presence of CCP2 module causes a decrease in the $K_M$ value, which indicates a stronger binding of the substrate.[9] This phenomenon closely resembles C1r[25] and C1s,[26] where the CCP modules also strongly influence the catalytic properties of the molecule. Modeling studies based on neutron scattering measurements suggest that C4 has a two-domain structure in solution.[27] Although we cannot exclude the possibility that the CCP modules establish contact with both domains of C4, in the present study we could only use the available structural data of the C4d domain.[28] The surfaces of the SP domains of MASP-2 and C1s are formed mainly by loops showing high diversity in length and conformation, whereas the surfaces of the CCP2 module of both proteins contain structurally conserved secondary elements (i.e. β-strands). Therefore, we can suggest that the potential C4 binding sites on the CCP2 modules have similar interaction patterns (Figure 7). We fitted the C4d fragment to the CCP2 modules of MASP-2 and C1s and built a speculative model representing the interactions between C4 and the proteases.

## Conclusions

The structure of MASP-2 shows both features typical of the members of the trypsin-like family of serine proteases and features that explain the unusual characteristics of this key enzyme of complement activation. The very narrow substrate specificity of this enzyme can be explained by the restricted access of subsites, which are determined by the surface loops of the SP domain. These loops, however, have different conformations than those of C1s, an enzyme with almost the same substrate specificity. These loops are the most variable regions of the serine protease structures and they can differ significantly even between two closely related enzymes. One of the most important functional differences between MASP-2 and C1s is that MASP-2 is capable of autoactivation. This fact could explain the different loop structures of the two enzymes. The SP domain alone, however, is not sufficient to determine all catalytic properties of MASP-2. The CCP2 module, which contains additional binding site(s) for the bulky C4 substrate is connected to the SP domain through a flexible module–domain junction. The intermodular flexibility must be a fundamental prerequisite of the function of this protease to fulfil its dual function.

## Materials and Methods

### Proteins

The CCP2-SP and the SP fragments of MASP-2 were prepared as described.[9] The purified proteins were concentrated to 2 mg/ml and stored at 4 °C in the presence of 0.05% (w/v) $NaN_3$.

### Crystallization and data collection

Crystals were grown by the hanging-drop method at 20 °C. Crystals were obtained by mixing 2 μl of reservoir solution and 2 μl of protein solution. The reservoir solution contained 30% (w/v) PEG 6000, 0.2 M NaCl, 10% (v/v) glycerol and 0.1 M Tris–HCl (pH 7.5). The protein solution contained 0.8 mg/ml of the active form of MASP2 CCP2-SP,[9] 140 mM NaCl and 20 mM Tris–HCl (pH 7.4). Synchrotron data were collected at LURE on the DW32 beamline and at SPring-8 on the BL41XU beamline. Due to scaling problems, the former dataset was used for structure determination. Data were processed and scaled to a resolution of 2.23 Å using Mosflm,[29] as well as programs SCALA[30] and TRUNCATE[31] of the Collaborative Computing Project Number 4.[32] The asymmetric unit contains two molecules.

### Structure determination and refinement

The structure was solved by molecular replacement, with the CCP2 and SP domains of C1s (PDB id 1elv) as search models, using the program Beast[33] of the Collaborative Computing Project Number 4. Refinement was carried out with Refmac5,[34] using restrained maximum likelihood refinement and TLS refinement.[35] Arp[36] was used for automatic solvent building. Model building was carried out using the O program.[37] Tight

non-crystallographic restraints were applied to SP domains except for some residues found in different conformations in the two molecules of the asymmetric unit. The final model contains residues 362–440 and 445–686 of molecule A, and 366–412, 416–441 and 445–686 of molecule B. The stereochemistry of the structure was assessed with PROCHECK.[38] Data collection and refinement statistics are shown in Table 1.

Figures were generated using the programs MOLSCRIPT,[39] Bobscript[40] Raster3D,[41] and Swiss-PDBViewer.[42] The GRASP program[43] was used for generating molecular surfaces and calculating electrostatic potential at surface points. Structural alignments were carried out using program O and Swiss-PDBViewer. Surface areas were calculated using the program SURFACE.[44]

### Enzyme activity assay

The enzymatic activity measurements were carried out on a Jasco V-550 UV–VIS spectrophotometer, using 1 cm cuvettes, at 37 °C, in 20 mM Hepes (pH 8.5), 5 mM $CaCl_2$. *Z*-Lys-SBzl (α-*N*-benzyloxycarbonyl-lysine thiobenzyl ester, SIGMA) and DTDP (4,4′-dithiodipiridine, SIGMA) were used as substrate and chromogenic reagent, respectively. Enzyme-free reactions were used as negative controls. All experiments were done in triplicate.

The effect of NaCl concentration on the enzymatic activity was assayed at 1.2 μg/ml enzyme concentration with the NaCl concentration ranging between 0 and 200 mM, the ionic strength was held constant (200 mM) using tetraethylammonium chloride. $k_{cat}/K_M$ values were determined directly from the observed rate constant at low substrate concentrations (30 μM) using the $\varepsilon_{324} = 19{,}800\ M^{-1}\ cm^{-1}$ value, as described.[45]

### Differential scanning calorimetry (DSC)

Calorimetric measurements were performed on a VP-DSC (MicroCal) differential scanning calorimeter. Denaturation curves were recorded between 20 °C and 70 °C at a pressure of 2.5 atm, using a scanning rate of 1 deg. C/minute. The protein concentration was set to 0.2 mg/ml. Samples were dialyzed against 20 mM Hepes (pH 7.0), 145 mM NaCl, and the dialysis buffer was used as a reference. Heat capacities were calculated as outlined by Privalov.[46]

### Protein Data Bank atomic coordinates

The atomic coordinates and structure factors were deposited in the Protein Data Bank with accession code 1q3x.

## Acknowledgements

The authors are grateful to Anita Lewit-Bentley as local contact for helping in the course of data collection at LURE. K.J. thanks Yuji Goto for his support and James Villanueva and Masahide Kawamoto for their help in the data collection at SPring-8. The authors are grateful to Richard E. Dickerson, Verne N. Schumaker and Emil Reisler at the Department of Chemistry and Biochemistry, UCLA for reading the manuscript and for their helpful comments. This work was supported by the Hungarian National Science Foundation (OTKA) grants T034994, TS004730 and T046444, the Hungarian Ministry of Health (EÜ Tanács 555/2003), Chemical Works of Gedeon Richter, National Research and Development Plan (NKFP 1/010). The financial support of the Janos Bolyai Research Fellowship of the Hungarian Academy of Sciences for H.V. is gratefully acknowledged.

## References

1. Gál, P. & Ambrus, G. (2001). Structure and function of complement activating enzyme complexes: C1 and MBL-MASPs. *Curr. Protein Pept. Sci.* **2**, 43–59.
2. Arlaud, G. J., Colomb, M. G. & Gagnon, J. (1987). A functional model of the human C1 complex. *Immunol. Today*, **8**, 106–111.
3. Schumaker, V. N., Závodszky, P. & Poon, P. H. (1987). Activation of the first component of complement. *Annu. Rev. Immunol.* **5**, 21–42.
4. Turner, M. W. (1996). Mannose-binding lectin: the pluripotent molecule of the innate immune system. *Immunol. Today*, **17**, 532–540.
5. Thiel, S., Vorup-Jensen, T., Stover, C. M., Schwaeble, W., Laursen, S. B. Poulsen, K. *et al*. (1997). A second serine protease associated with mannan-binding lectin that activates complement. *Nature*, **386**, 506–510.
6. Rossi, V., Bally, I., Thielens, N. M., Esser, A. F. & Arlaud, G. J. (1998). Baculovirus-mediated expression of truncated modular fragments from the catalytic region of human complement serine protease C1s. *J. Biol. Chem.* **273**, 1232–1239.
7. Vorup-Jensen, T., Petersen, S. V., Hansen, A. G., Poulsen, K., Schwaeble, W. Sim, R. B. *et al*. (2000). Distinct pathways of mannan-binding lectin (MBL)- and C1-complex autoactivation revealed by reconstitution of MBL with recombinant MBL-associated serine protease-2. *J. Immunol.* **165**, 2093–2100.
8. Chen, C. B. & Wallis, R. (2001). Stoichiometry of complexes between mannose-binding protein and its associated serine proteases: defining functional units for complement activation. *J. Biol. Chem.* **276**, 25894–25902.
9. Ambrus, G., Gál, P., Kojima, M., Szilágyi, K., Balczer, J. Antal, J. *et al*. (2003). Natural substrates and inhibitors of mannan-binding lectin-associated serine protease 1 and 2: a study on recombinant catalytic fragments. *J. Immunol.* **170**, 1374–1382.
10. Perona, J. J. & Craik, C. S. (1997). Evolutionary divergence of substrate specificity within the chymotrypsin-like serine protease fold. *J. Biol. Chem.* **272**, 29987–29990.
11. Gaboriaud, C., Rossi, V., Bally, I., Arlaud, G. J. & Fontecilla-Camps, J. C. (2000). Crystal structure of the catalytic domain of human complement C1s: a serine protease with a handle. *EMBO J.* **19**, 1755–1765.
12. Budayova-Spano, M., Grabarse, W., Thielens, N. M., Hillen, H., Lacroix, M. Schmidt, M. *et al*. (2002). Monomeric structures of the zymogen and active catalytic domain of complement protease C1r: further insights into the C1 activation mechanism. *Structure*, **10**, 1509–1519.
13. Budayova-Spano, M., Lacroix, M., Thielens, N. M., Arlaud, G. J., Fontecilla-Camps, J. C. & Gaboriaud, C. (2002). The crystal structure of the zymogen catalytic

domain of complement protease C1r reveals that a disruptive mechanical stress is required to trigger activation of the C1 complex. *EMBO J.* **21**, 231–239.
14. Feinberg, H., Uitdehaag, J. C. M., Davies, J. M., Wallis, R., Drickamer, K. & Weis, W. I. (2003). Crystal structure of the CUB1-EGF-CUB2 region of mannose-binding protein associated serine protease-2. *EMBO J.* **22**, 2348–2359.
15. Schechter, I. & Berger, A. (1967). On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **27**, 157–162.
16. Wells, C. M. & Di Cera, E. (1992). Thrombin is a Na(+)-activated enzyme. *Biochemistry*, **31**, 11721–11730.
17. Dang, Q. D. & Di Cera, E. (1996). Residue 225 determines the $Na^+$-induced allosteric regulation of catalytic activity in serine proteases. *Proc. Natl Acad. Sci. USA*, **93**, 10653–10656.
18. Kenesi, E., Katona, G. & Szilagyi, L. (2003). Structural and evolutionary consequences of unpaired cysteines in trypsinogen. *Biochem. Biophys. Res. Commun.* **309**, 749–754.
19. Banner, D. W., D'Arcy, A., Chene, C., Winkler, F. K., Guha, A. Konigsberg, W. H. *et al*. (1996). The crystal structure of the complex of blood coagulation factor VIIa with soluble tissue factor. *Nature*, **380**, 41–46.
20. Bode, W., Brandstetter, H., Mather, T. & Stubbs, M. T. (1997). Comparative analysis of haemostatic proteinases: structural aspects of thrombin, factor Xa, factor IXa and protein C.. *Thromb. Haemost.* **78**, 501–511.
21. Wallis, W. (2002). Structural and functional aspects of complement activation by mannose-binding protein. *Immunbiology*, **205**, 433–445.
22. Ohta, M., Okada, M., Yamashina, I. & Kawasaki, T. (1990). The mechanism of carbohydrate-mediated complement activation by the serum mannan-binding protein. *J. Biol. Chem.* **265**, 1980–1984.
23. Lu, J., Thiel, S., Wiedermann, H., Timpl, R. & Reid, K. B. M. (1990). Binding of the pentamer/hexamer forms of mannan binding protein to zymosan activates the proenzyme $C1r_2$–$C1s_2$ complex of the classical pathway, without involvement of C1q. *J. Immunol.* **144**, 2287–2294.
24. Lőrincz, Zs., Gál, P., Dobó, J., Cseh, S., Szilágyi, K., Ambrus, G. & Závodszky, P. (2000). The cleavage of two C1s subunits by a single active C1r reveals substantial flexibility of the C1s–C1r–C1r–C1s tetramer in the C1 complex. *J. Immunol.* **165**, 2048–2051.
25. Kardos, J., Gál, P., Szilágyi, L., Thielens, N. M., Szilágyi, K. Lőrincz, Zs. *et al*. (2001). The role of the individual domains in the structure and function of the catalytic region of a modular serine protease, C1r. *J. Immunol.* **167**, 5202–5208.
26. Rossi, V., Cseh, S., Bally, I., Thielens, N. M., Jensenius, J. C. & Arlaud, G. J. (2001). Substrate specificities of recombinant mannan-binding lectin-associated serine proteases-1 and -2. *J. Biol. Chem.* **276**, 40880–40887.
27. Perkins, S. J., Nealis, A. S. & Sim, R. B. (1990). Molecular modeling of human complement component C4 and its fragments by X-ray and neutron solution scattering. *Biochemistry*, **29**, 1167–1175.
28. van den Elsen, J. M. H., Martin, A., Wong, V., Clemenza, L., Rose, D. R. & Isenman, D. E. (2002). X-ray crystal structure of the C4d fragment of human complement component C4. *J. Mol. Biol.* **322**, 1103–1115.
29. Leslie, A. G. W. (1993). Autoindexing of rotation diffraction images and parameter refinement. In *Proceedings of the CCP4 Study Weekend: Data Collection and Processing 29–30 January* (Sawyer, L., Isaacs, N. & Bailey, S., eds), pp. 44–51, SERC Daresbury Laboratory, Daresbury, UK.
30. Evans, P. R. (1993). Data reduction. In *Proceedings of the CCP4 Study Weekend: Data Collection and Processing 29–30 January* (Sawyer, L., Isaacs, N. & Bailey, S., eds), pp. 114–122, SERC Daresbury Laboratory, Daresbury, UK.
31. French, G. S. & Wilson, K. S. (1978). On the treatment of negative intensity observations. *Acta. Crystallog. sect. A*, **34**, 517–525.
32. Collaborative Computational Project Number 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D*, **50**, 760–763.
33. Read, R. J. (2001). Pushing the boundaries of molecular replacement with maximum likelihood. *Acta Crystallog. sect. D*, **57**, 1373–1382.
34. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallog. sect. D*, **53**, 240–255.
35. Winn, M., Isupov, M. & Murshudov, G. N. (2001). Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallog. sect. D*, **57**, 122–133.
36. Lamzin, V. S. & Wilson, K. S. (1997). Automated refinement for protein crystallography. *Methods Enzymol.* **277**, 269–305.
37. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallog. sect. A*, **47**, 110–119.
38. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283–291.
39. Kraulis, P. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946–950.
40. Esnouf, R. M. (1999). Further additions to Molsript version 1.4, including reading and contouring of electron density maps. *Acta Crystallog. sect. D*, **55**, 938–940.
41. Merritt, E. A. & Bacon, D. J. (1997). Raster3D photorealistic molecular graphics. *Methods Enzymol.* **277**, 505–524.
42. Guex, N. & Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*, **18**, 2714–2723.
43. Nicholls, A., Sharp, K. A. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet.* **11**, 281–296.
44. Lee, B. & Richards, F. M. (1971). The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55**, 379–400.
45. Grasetti, D. R. & Murray, J. F., Jr (1967). Determination of sulfhydryl groups with 2,2- and 4,4-dithiodipyridine. *Arch. Biochem. Biophys.* **119**, 41–49.
46. Privalov, P. L. (1979). Stabilty of proteins: small globular proteins. *Advan. Protein Chem.* **33**, 167–241.
47. Song, H. K. & Suh, S. W. (1998). Kunitz-type soybean trypsin inhibitor revisited: refined structure of its complex with porcine trypsin reveals an insight into the interaction between a homologous inhibitor from *Erythrina caffra* and tissue-type plasminogen activator. *J. Mol. Biol.* **275**, 347–363.

48. Yennawar, N. H., Yennawar, H. P. & Farber, G. K. (1994). X-ray crystal structure of gamma-chymotrypsin in hexane. *Biochemistry*, **33**, 7326–7336.
49. Dullweber, F., Stubbs, M. T., Musil, D., Sturzebecher, J. & Klebe, G. (2001). Factorising ligand affinity: a combined thermodynamic and crystallographic study of trypsin and thrombin inhibition. *J. Mol. Biol.* **313**, 593–614.
50. Mather, T., Oganessyan, V., Hof, P., Huber, R., Foundling, S., Esmon, C. & Bode, W. (1996). The 2.8 Å crystal structure of Gla-domainless activated protein C. *EMBO J.* **15**, 6822–6831.
51. Kemball-Cook, G., Johnson, D. J. D., Tuddenham, E. G. D. & Harlos, K. (1999). Crystal structure of active site-inhibited human coagulation factor VIIa (des-Gla). *J. Struct. Biol.* **127**, 213–223.
52. Hopfner, K.-P., Lang, A., Karcher, A., Sichler, K., Kopetzki, E. Brandstetter, H. *et al*. (1999). Coagulation factor IXa: the relaxed conformation of Tyr99 blocks substrate binding. *Structure*, **7**, 989–996.
53. Adler, M., Kochanny, M. J., Bin, Y., Rumennik, G., Light, D. L., Biancalana, S. & Whitlow, M. (2002). Crystal structures of two potent nonamidine inhibitors bound to factor Xa. *Biochemistry*, **41**, 15514–15523.
54. Bernett, M. J., Blaber, S. I., Scarisbrick, I. A., Dhanarajan, P., Thompson, S. M. & Blaber, M. (2002). Crystal structure and biochemical characterization of human kallikrein 6 reveals that a trypsin-like kallikrein is expressed in the central nervous system. *J. Biol. Chem.* **277**, 24562–24570.
55. Somoza, J. R., Ho, J. D., Luong, C., Ghate, M., Sprengeler, P. A. Mortara, K. *et al*. (2003). The structure of the extracellular region of human hepsin reveals a serine protease domain and a novel scavenger receptor cysteine-rich (SRCR) domain. *Structure*, **11**, 1123–1131.
56. Friedrich, R., Fuentes-Prior, P., Ong, E., Coombs, G., Hunter, M., Oehler, R. & Pierson, D. (2002). Catalytic domain structures of MT-SP1/matriptase, a matrix-degrading transmembrane serine proteinase. *J. Biol. Chem.* **277**, 2160–2168.
57. Lu, D., Futterer, K., Korolev, S., Zheng, X., Tan, K., Waksman, G. & Sadler, J. E. (1999). Crystal structure of enteropeptidase light chain complexed with an analog of the trypsinogen activation peptide. *J. Mol. Biol.* **292**, 361–373.

***Edited by R. Huber***

*(Received 20 May 2004; received in revised form 5 July 2004; accepted 12 July 2004)*