Crystal Structure of Human Maspin, a Serpin with Antitumor Properties

REACTIVE CENTER LOOP OF MASPIN IS EXPOSED BUT CONSTRAINED*

Received for publication, August 30, 2004, and in revised form, October 14, 2004
Published, JBC Papers in Press, October 22, 2004, DOI 10.1074/jbc.M409957200

Maher Al-Ayyoubi‡, Peter G. W. Gettins‡, and Karl Volz§§

From the Departments of ‡Biochemistry and Molecular Biology and §Microbiology and Immunology, University of Illinois at Chicago, Chicago, Illinois 60612-7334

Maspin, a member of the serpin superfamily, has tumor suppressing activity against breast and prostate cancer. Maspin inhibits tumor growth by blocking cell invasion, and its reactive center loop (RCL) is thought to mediate this activity. To understand this function on the molecular level, we have solved the three-dimensional structure of Maspin to 3.1 Å resolution. The molecular structure shows the characteristic features of the serpin fold, but the RCL of Maspin is unique in length, composition, and placement. Although the RCL of Maspin is accessible and cleavable by some proteinases, it functions in the uncleaved, constrained conformation observed here. These structural results will contribute to our understanding of the mechanism by which Maspin suppresses tumors.

Maspin (mammary serine proteinase inhibitor) was first identified as a tumor suppressor protein in mammary epithelial cells (1). Its expression levels correlated negatively with tumor progression in cases in which the expression was highest in normal epithelial cells and absent in most mammary carcinoma cell lines (2, 3). Maspin was also found to have tumor suppressor activity against prostate tumor cells (4). It blocks tumor cell invasion by inhibiting cell migration (1, 5) while stimulating cell adhesion (6), possibly through an apoptosis-dependent mechanism (7). Maspin also has antiangiogenic activity that contributes to its tumor suppression ability (8). Intentional overexpression of Maspin in mammary carcinoma cell lines leads to a decreased cell motility, invasion, and metastasis (9, 10). Induction of Maspin expression in prostate cancer cells is under the control of the p53 tumor suppressor protein (11, 12). Maspin levels are regulated in part through epigenetic processes, indicating a role for DNA methylation in the control of cell-type-specific Maspin expression (13–15).

Maspin (SERPINB5) is a member of the serpin superfamily, related to the ov-serpins, a centrally conserved clade that includes, among others, equine leukocyte inhibitor, plasminogen activator inhibitor-2, and ovalbumin (1, 16). All true inhibitory serpins rely on an exposed reactive center loop (RCL) to inhibit their target proteinase, in which the proteinase binds and cleaves the RCL and becomes incorporated into a serpin-proteinase complex after a large scale conformational transformation (17). Maspin is an unusual serpin in that its RCL is necessary for activity, but there is no cleavage or rearrangement of the molecule (18, 19); its RCL must remain intact (20). Indeed, the RCL of Maspin alone is sufficient to inhibit cell migration and induce cell adhesion (6).

The RCLs of intact serpins have been observed in a variety of conformations, ranging from the “canonical” extended β-strand that is ready for recognition by its target proteinase (e.g. α1-PI) (21, 22) to a compact, three-turn α-helix as seen in the non-inhibitory serpin ovalbumin (23). Because of the demonstrated importance of the RCL of Maspin in conferring the antitumor properties of the molecule, we determined the three-dimensional structure of Maspin and herein describe the structural relationship of its RCL to the rest of the molecule.

EXPERIMENTAL PROCEDURES

Crystallization, Data Collection, Structure Determination, and Refinement—Human Maspin cDNA was kindly provided by Dr. Olga Volpert of Northwestern University and cloned into a pQE30 vector (Qiagen) for expression in Escherichia coli strain sg13009. Maspin with an amino-terminal His tag was overexpressed in the form of inclusion bodies at 37 °C. The cells were grown for 5 h after isopropyl 1-thio-β-D-galactopyranoside induction and then spun down and sonicated. The inclusion bodies were isolated by centrifugation at 10,000 × g for 20 min, dissolved in 8 M guanidine hydrochloride, and 50 mM Tris, pH 8.0, and renatured by a 20× dilution into an ice-cold solution of 150 mM NaCl, 50 mM sodium phosphate, pH 7.4. The refolded protein was dialyzed twice against 250 mM NaCl, 50 mM sodium phosphate, pH 7.4, and then concentrated and purified with nickel-nitrioltriacetic acid metal chelate (Qiagen) and Q-Sepharose (Amersham Biosciences) column chromatography circular dichroism and matrix-assisted laser desorption ionization time-of-flight mass spectroscopy verified that the product was Maspin, folded and uncleaved. The wild-type form of the Maspin protein had very poor solubility properties as measured by dynamic light scattering and did not show any promise in preliminary crystallization screens. Suspecting that the high cysteine content of Maspin was the cause, all eight cysteines of the molecule were changed to serine/alanine (C205, C34A, C183S, C205S, C214S, C287S, C323S, and C373S) by site-specific mutagenesis using the QuickChange kit (Stratagene), and the cysteine-free mutant form was overexpressed and purified as before. With the protein concentration at 18 mg/ml, optimized sparse matrix (24) crystallization trials (Screen 1, Screen 2, and polyethylene glycol/ion kits from Hampton) yielded crystals under the conditions of 45% polyethylene glycol 400 in 0.2 M sodium citrate and 0.1 M Tris, pH 8.21. Rod-shaped crystals grew to an average dimension of 50 × 50 × 500 μm within 1 week.

Diffraction experiments showed that the crystals were tetragonal, of either space group P42_2_1 or I422, with basically the same unit cell
parameters of $a = b = 147.0$ Å and $c = 118.0$ Å. The two crystal forms were optically identical. A Matthews' coefficient (25) of 3.9 Å$^3$/kDa indicated two molecules/asymmetric unit in the primitive form and one in the body-centered form, corresponding to 68% solvent for each. The crystals diffracted very well despite their high solvent content, which can be explained by the unusual molecular packing (see below).

The 45% polyethylene glycol 400 content in the Maspin crystals sufficiently protected them from ice formation during crystal cooling so that no additional cryoprotectant was necessary. Diffraction data were measured on a MAR225 charge-coupled device detector at the Southeast Regional Collaborative Access Team synchrotron radiation source of the Advanced Photon Source, Argonne National Laboratory. A total of seven data sets were collected from seven crystals. The data were processed with the HKL2000 software (26). For the P4212 form, the best quality data set had an $R_{merge}$ of 8.0% from 116,205 unique reflections, 97.5% complete to 3.1 Å. For the I422 form, the best set had an $R_{merge}$ of 12.1% for 62,800 observations of 11,544 unique reflections, 96.1% complete to 3.1 Å (Table I).

The Maspin structure was solved using molecular replacement. Initial phases were obtained using a model constructed from the three ov-serpin homologues: equine leukocyte inhibitor, plasminogen activator-inhibitor-2, and ovalbumin (1HLE, 1BY7, and 1OVA, respectively) (Fig. 1). The I422 crystal form was solved and refined first because it had only one molecule/asymmetric unit and was later used to solve the P4212 crystal form. The molecular replacement and refinement calculations were carried out using crystallography NMR software (27).

Manual rebuilding was performed on an Octane workstation (Silicon Graphics) using the program QUANTA (28). Statistics for the final data sets and refinements are shown in Table I.

### RESULTS AND DISCUSSION

The Entire Maspin Structure Is Well Defined—The two different crystal forms of human Maspin provided three independent final models of the molecule (the P4212 crystal form had two molecules/asymmetric unit). The backbone of each model was complete (Fig. 2), containing all 375 amino acids of Maspin plus three additional residues on the amino terminus from the partially interpretable electron density of the His tags. The RCLs had interpretable density (Fig. 3), with all $\phi$/$\psi$ angles in $\alpha$ or $\beta$ geometries. Least squares $\sigma$-carbon superpositions of the full-length models in all combinations yielded an average root mean square deviation of 0.37 Å, indicating that the three

### TABLE I

| Crystal form | P42,2 | I422 |
|--------------|-------|------|
| Resolution (Å)$^a$ | 3.1 (3.2–3.1) | 3.1 (3.2–3.1) |
| Number of reflections | 23,502 | 8,000 (28.2) |
| Completeness (%) | 97.5 (95.6) | 12.1 (64.4) |
| $R_{merge}$ | 8.00 (28.2) | 25.2 |
| Number of protein atoms | 5,924 | 63 |
| Number of solvent atoms | 152 | |
| Root mean square deviations | 0.007 | 0.008 |
| Bond lengths (Å) | 1.29 | 19.4 |
| Bond angles (°) | 1.36 | 20.0 |
| $R_{free}$ (%)$^b$ | 25.2 | 24.7 |
| $R_{free}$ (%)$^c$ | 24.7 |

$^a$ Numbers in parentheses represent highest resolution shell. $^b$ $R_{merge} = \Sigma |F_o| - |F_c|/\Sigma |F_o|$, where $F_o$ and $F_c$ are the observed and calculated structure factors, respectively, and the summations are over all unique reflections used in the refinement.

$^c$ $R_{free}$ is calculated as for $R_{cryst}$ except the summation is over a test set of 1% of the unique reflections omitted from refinement.

![Fig. 1. Sequence alignment of Maspin, equine leukocyte elastase inhibitor, ovalbumin, and plasminogen activator inhibitor-2. Identical residues are shown in boldface. Numbering and secondary structure elements are shown for Maspin.](image-url)
models were essentially equivalent at this resolution. Molecule A of the P42,2 crystal form will be used as the representative model in all further discussions.

**β-Sheet C of Maspin Provides Stabilizing Interactions with the RCL**—Most intact RCLs of inhibitory serpins have limited interactions with their serpin bodies, presumably for maximum conformational flexibility and proteinase accessibility. Maspin, the first example of a serpin with an RCL of defined function other than proteinase inhibition, is different. The RCL of Maspin is shorter by four residues and lies closer to the serpin core of the molecule. It is positioned farther “back” than all of the other known serpin RCL structures, bent toward the amino end of s2C (Fig. 4). The RCL is held in place through a number of stabilizing bonding interactions with amino acid side chains on the surface of β-sheet C. Specifically, the Glu-335 side chain has three potential interactions: with Lys-173, the amide backbone nitrogen of Thr-203 of s3C, and Lys-268 of s2C. The side chain of residue Gln-343 extends back to bond to the Oγ of Thr-203 of s2C. There are also additional antiparallel β-sheet backbone hydrogen bonds on the P’ stem with strand s2C, starting at residue His-344. These stabilizing interactions together with the shorter length of the RCL cause the RCL of Maspin to be more rigid in position and conformation than the RCLs of inhibitory serpins.

**The RCL of Maspin Is Accessible**—Reactive center loops can adopt a variety of conformations, ranging from the extended β-strand as seen for α1PI (21, 22) to α-helical as in ovalbumin (23). The RCL of Maspin is an intermediate case, with most of the backbone in a β-strand configuration except for three residues, Pro-337, Arg-340, and Ile-341, all of which have right-handed helical conformations (Fig. 4). The half-turn helix at 340–341 affords maximum solvent exposure of the arginine and isoleucine residues. Although there was no electron density for the Arg-340 side chain past Cβ, the positions of the backbone atoms dictate that the Arg-340 side chain points away from the body of the molecule, into solution. The arginine-isoleucine pair is accessible and cleavable by trypsin and thrombin (20, 18).

**Hydrophobic Residues Are Prominent in the RCL of Maspin**—As shown in Fig. 4, the exposed portion of the RCL of Maspin has a large hydrophobic content. Not only are the side chains of Ile-334, Val-336, Ile-341, and Leu-342 entirely accessible by solvent, but the hydrophobic surfaces of Pro-337 and adjacent residues are also completely exposed. As discussed below, the Maspin molecules associate intermolecularly through these hydrophobic surfaces of the RCLs.

**RCLs of Maspin Associate to Form Octamers**—The Maspin molecules pack as octamers in the crystalline lattice, formed by dimers of tetramers, with 422 symmetry (Fig. 5A). The intermolecular contacts are predominantly from the RCLs. The side by side contacts within each tetramer depend entirely on residues Gly-331 to Glu-335 of one RCL associating with residues in the hydrophobic crevice between hH and s1B of the adjacent molecule. The strongest hydrophobic interaction is the insertion of the Ile-334 side chain into each neighboring crevice. The forces that seem to complete the tetramer formation and drive the octamer assembly are the hydrophobic associations between groups of residues Ile-341 and Leu-342 on the most exposed surfaces of the α-helical turns in the RCLs. As shown in Fig. 5B, the eight turns all face each other and form a tight hydrophobic cluster in the center of the octamer. The only other intermolecular interactions are minor salt links and hydrogen bonds between the s3C and s4C strands of molecules in the opposing tetramers (Fig. 5).

The RCL of Maspin has been identified as the sole locus of activity for the induction of cell matrix adhesion and inhibition of cell invasion (6), so it is most likely that the hydrophobic nature of the RCL is functionally important. It is unknown whether Maspin occurs as a monomer or larger oligomers in vivo. In this work, dynamic light scattering measurements showed that the cysteine-free mutant form of Maspin aggregates at a concentration of 1.0 mg/ml with an estimated molecular mass of 378 kDa, approximating that of a Maspin octamer. The RCL-dependent oligomerization of Maspin characterized herein could explain the polymerization phenomena observed previously by others (18, 29).

**Maspin Has Non-standard Hinge and Breach Regions**—In-
Inhibitory serpins undergo a large conformational change, including insertion of the RCL into the existing β-sheet A. The mechanism involves the “hinge” portion of the RCL (P15–P9) and the “breach” region, a partial separation between the parallel β-strands s3A and s5A at the top center of β-sheet A. All of the inhibitory serpins retain a consensus sequence in the hinge region that consists of a conserved serine/threonine at P14 and small volume side chains such as alanine or glycine occupying the P11–P9 sites. The hinge region (P15–P9) of Maspin differs by having a glycine at P14, and the larger volume side chains Ile, Glu, and Val in P11–P9. Furthermore, P8 is a proline. The presence of these bulky residues makes it highly unlikely that Maspin acts as a proteinase inhibitor using the standard serpin loop insertion mechanism (30, 18). The structural results here affirm these interpretations. Additional evidence is that the breach region of the Maspin molecule was closed. This point of initial insertion for the RCL is always partially open in inhibitory serpins. Even in the non-inhibitory serpins, ovalbumin and pigment epithelium-derived factor (PEDF), the parallel β-strand hydrogen bond pattern is disrupted in the breach region. The bonding pattern of that region is most complete in Maspin (up to the Gly-169 nitrogen to Glu-325 oxygen–hydrogen bond), consistent with the higher stability and complete absence of a loop insertion process.

The RCL of Maspin Is Compatible with the Serpin Body of Ovalbumin—The evidence that the RCL of Maspin conveys the regulation of cell invasion by the molecule was obtained through chimeric RCL exchange experiments with ovalbumin (6). The sections exchanged were residues 330–345 of Maspin with residues 344–361 of ovalbumin (see Fig. 1). In comparing

---

**Fig. 4. Stereodiagram of RCL of Maspin and portions of β-sheet C.** The orientation is approximately the same as in Fig. 3.

**Fig. 5. Stereodiagram of the molecular packing of Maspin.** The molecules pack as octamers with 422 symmetry. The central intermolecular associations are mediated by the RCLs, predominantly by hydrophobic contacts. A, stereodiagram of the α-carbon backbones of the Maspin octamer. The 4-fold axis is approximately vertical. The box encloses the RCLs at the center of the octamer. B, stereodiagram shows details of the core interactions enclosed in the box in A.
the superimposed Maspin and ovalbumin structures, the termini of these two sections overlap almost perfectly, and even though the two RCLs are quite different in length, composition, and conformation, they are capable of sharing some of the same stabilizing interactions with side chains on $\beta$-sheet C. These fortuitous structural commonalities probably contributed to the success of the RCL exchange experiments by minimizing folding problems for the chimeras.

**Maspin and PEDF May Utilize Different Molecular Surfaces**—Maspin and PEDF are both antiangiogenic serpins (8, 31). As discussed previously, the RCL of Maspin is critical for its regulation of cell matrix adhesion and tumor cell invasion (6). The part of Maspin that confers its antiangiogenic activity has not yet been identified. It may not be the RCL because mutants in the RCL are still able to block angiogenesis (8). PEDF does not require its RCL for antiangiogenesis or for any known function. The known activities of PEDF's (neurotropism, antiangiogenesis, and vascular permeability) have been ascribed to helices C and D and the surrounding area (8). PEDF does not require its RCL for antiangiogenesis (52).

**Collagen-binding Surface of Maspin Is Structurally Ill-defined**—Maspin has been shown to bind collagen types I and III (38). This is thought to enhance cell adhesion in the suppression of tumor cell migration and angiogenesis. The twohybrid studies suggest that the collagen-binding region of Maspin is located between residues Tyr-84 and Tyr-112, and this region is thought to constitute a general collagen-binding motif (38). These interpretations are not supported by the structural results. The Tyr-84 to Tyr-112 sequence of the Maspin molecule is extended, defining the secondary structure component $\alpha$2A-loop-h$\alpha$. $\alpha$2A is buried largely behind h$\alpha$ and its accompanying loop, and the remaining accessible surfaces of $\alpha$2A and h$\alpha$ are not contiguous. Furthermore, the exposed surface of $\alpha$2A-loop-h$\alpha$ does not contain the complement of acidic residues that are expected for a collagen-binding region. The concentrations of acidic residues in Maspin are in the approximate regions of hG and hC, but there is no well defined acidic collagen-binding motif as found in PEDF (33, 35, 39–40).

**Conclusions and Implications**—The RCL of Maspin, the primary component of the molecule responsible for its antitumor properties, assumes a fixed conformation. Its central Arg-340 residue and neighboring hydrophobic surfaces are likely to be the principal determinants of the unique activities of Maspin. The antiangiogenic surface of Maspin remains to be identified. A Maspin-based anticancer strategy might be feasible through the development of small molecule mimics of the functional groups on the RCL of Maspin and the surrounding area.

---

2 S. Filleur, K. Volz, T. Nelius, T. A. Zaichuk, H. Huang, Y. Mirochnik, M. S. Aymerich, S. P. Becerra, Y. Yap, D. Veliceasa, E. H. Shroff, and O. V. Volpert, submitted for publication.