Mapping the putative binding site for uPA protein in Esophageal Cancer-Related Gene 2 by heteronuclear NMR method

Yong Geng a, Yingang Feng a, Tao Xie a, Yuanyuan Dai b, Jinfeng Wang a,*, Shih-Hsin Lu c

a National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, 15 Datun Road, Beijing 100101, China
b Department of Clinical Pharmacology, Cancer Institute, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing 100021, China
c Department of Etiology and Carcinogenesis, Cancer Institute, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing 100021, China

A R T I C L E   I N F O

Article history:
Received 15 July 2008
and in revised form 25 August 2008
Available online 18 September 2008

Keywords:
ECRG2
Kazal-type domain
Serine proteinase inhibitor
Urokinase-type plasminogen activator
Reactive site loops

A B S T R A C T

Esophageal Cancer-Related Gene 2 (ECRG2) is a novel member of the KAZAL-type serine proteinase inhibitor family and plays an important role in the inhibition of human esophageal cancer cell proliferation. The previous studies have shown that ECRG2 can bind the urokinase-type plasminogen activator (uPA)/plasmin system and inhibit its activity. In this study, the strategy of cloning, overexpression, and purification of ECRG2 for obtaining a properly folded ECRG2 with accurately formed disulfide bonds was established. The heteronuclear NMR experiments were performed with isotope labeled ECRG2 to investigate the binding interface of the protein with uPA. The sequence regions of ECRG2 for uPA binding were determined. Analysis indicates that the uPA-binding loops of ECRG2 are in correspondence with the reactive site loops for binding of serine proteinase in turkey ovomucoid third domain (OMTKY3). The structural similarity of ECRG2 to OMTKY3 was identified and a model for ECRG2 was proposed.

Available online 18 September 2008

ESCREG2, Esophageal Cancer-Related Gene 2; ESCC, esophageal squamous cell carcinoma; MT2A, metallothionein 2A; MMPs, matrix metalloproteinases; ECM, extracellular matrices; TATI, tumor-associated trypsin inhibitor; PSTI, pancreatic secretory trypsin inhibitor; uPA, urokinase-type plasminogen activator.

© 2008 Elsevier Inc. All rights reserved.

Abbreviations used:
ECRG2, Esophageal Cancer-Related Gene 2; ESCC, esophageal squamous cell carcinoma; MT2A, metallothionein 2A; MMPs, matrix metalloproteinases; ECM, extracellular matrices; TATI, tumor-associated trypsin inhibitor; PSTI, pancreatic secretory trypsin inhibitor; uPA, urokinase-type plasminogen activator.

* Corresponding author. Fax: +86 10 64872026.
E-mail address: jfw@sun5.ibp.ac.cn (J. Wang).

0003-9861/5 - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.abb.2008.08.023
Materials and methods

Cloning, overexpression, and purification of ECRG2

Analysis of ECRG2 sequence has indicated that the segment of the N-terminal 19 residues behaviors as a trans-membrane peptide. For preparing the recombinant ECRG2 protein, these 19 amino acids were truncated in the construction of the expression vector. Thus, peptide ECRG2(20–85) (called as ECRG2 hereafter), was expressed and purified for present study.

A DNA fragment encoding ECRG2 was amplified by PCR using the forward primer (ccgggcgggggccgggggacgccggccc) and backward primer (aaagaagtgctaccttcaacgattttctagc) which introduce an EcoRI site and a sequence encoding a factor Xa cleavage site at the 5’ end and a HindIII site at the 3’ end of the gene, respectively. The fragment was inserted into the T7 RNA polymerase-based expression vector pET-32a. Another DNA fragment encoding a linker peptide (SGS(GSGSGSG) including EcoRI and EcoRV sites was synthesized and inserted into the vector pET-32a. The resulting expression vector (Supplementary material (SM) Fig. 1) for thioredoxin (Trx) fusion protein ECRG2, namely pET-32a-Trx-(linker-Xa)-ECRG2, was used for the expression and purification of ECRG2.

The pET-32a-Trx-(linker-Xa)-ECRG2 plasmid was transformed into the oxidative cytoplasm of Escherichia coli Origami (DE3) strain. The E. coli Origami (DE3)/pET-32a-Trx-(linker-Xa)-ECRG2 strain was grown overnight in 100 ml LB media (with 50 µg/ml ampicillin and 15 µg/ml kanamycin) at 37°C. The overnight culture was used to inoculate another 1000 ml LB media. When the culture OD600 reached 0.6, the fusion protein expression was initiated by adding IPTG to a final concentration of 1 mM. After 20 h of additional cultivation at 16°C, cells were harvested by centrifugation at 4000 rpm for 30 min. The cell was then resuspended in 20 ml of buffer A (50 mM Tris/HCl, pH 8.0, 300 mM NaCl, 10 mM imidazole) with 0.5 mM PMSF, 0.2 mg/ml lysozyme, 0.2 mg/ml DNasel, and 2 mg/ml benzamidine. The cells were lysed by three 30 s sonication steps at 4°C (Labsonic U sonicator at 150 W intensity) followed by overnight freezing at ~80°C. After thawing, the extract was centrifuged for 60 min at 40,000 rpm and 4°C to separate the soluble and insoluble portions.

The soluble fusion protein Trx-(linker-Xa)-ECRG2 was purified using a metal-chelating affinity chromatography. For this, the supernatant was collected and loaded onto a Ni2+ -chelating column (10 ml) pre-equilibrated with buffer A. The column was extensively washed first with the buffer A, followed by buffer A containing 10 mM imidazole, and finally with factor Xa buffer (50 mM Tris–HCl, pH 8.0, 140 mM NaCl, 5 mM CaCl2). Then the factor Xa (1 U/mg of Trx-Xa-ECRG2) was loaded on the column for cleavage of the Trx fusion protein, and the solution in the column was circulated by a pump for 16 h at room temperature. Afterwards, the column was washed again with 50 ml of the buffer A. The ECRG2 containing fractions were concentrated to 5 ml by an ultrafiltration. The collected solution was loaded onto a Superdex 50 column pre-equilibrated with 0.1 mM NH4HCO3 buffer. Then the column was washed with 0.1 mM NH4HCO3 buffer. The collected fraction containing ECRG2 was lyophilized. Eluted fraction of protein was >99% pure as verified by SDS–PAGE. To substantiate the identity of purified ECRG2, mass spectrum was recorded on an Applied Biosystems Voyager MALDI-TOF mass spectrometer. Uniformly 15N- and 13C-single labeled and 15N/13C-double labeled ECRG2 were obtained by growth in minimal media containing 15NH4Cl and/or [13C]-glucose as the sole nitrogen and/or carbon sources. The purity of proteins was checked by SDS–PAGE to ensure a single band.

NMR spectroscopy

The NMR samples contained 1.0 mM 15N-, 13C-, or 15N/13C-labeled ECRG2, 100 mM KCl, and 0.01% NaN3 in 50 mM deuterated acetate buffer (pH 5.0). All NMR experiments were run on a Bruker DMX 600 MHz spectrometer equipped with a triple-resonance cryo-probe at 293 K. The 2D 1H-15N HSQC and 3D 1H-15N–13C HNCO, HNCA, HN(CO)CA, HN(CA)CO, HNCCAB, CBCA(CO)NH, and HBHA(CO)NH, and 3D 1H–15N TOCSY-HSQC and NOESY-HSQC experiments were adopted for the resonance assignments. All NMR data were processed and analyzed with FELIX software (Accelrys Inc.). 1H chemical shifts were referenced to the internal DSS (2,2-dimethyl-2-silamantane-5-sulphonate). 15N chemical shifts were referenced indirectly [11].

Results

Recombinant ECRG2 for NMR experiments

In this study, the stable-isotope-labeled ECRG2 is required for analysis the interaction of ECRG2 with uPA. Hence, the pET-32a-Trx-(linker-Xa)-ECRG2 expression plasmid was obtained, and the desired fusion protein Trx-(linker-Xa)-ECRG2 was expressed in the soluble form. As is well known, missing or wrong disulfide bridges often cause formation of insoluble aggregates when extracellular proteins containing cysteine residues are expressed in the conventional E. coli strains [12]. However, using E. coli Origami (DE3) strain which carries a trxB-/gor522-double mutation in the expression of the putative serine proteinase inhibitor allows disulfide formation in its oxidative cytoplasm [13,14]. Actually, expression of ECRG2 in the oxidative cytoplasm of E. coli Origami (DE3) provided a soluble form of the protein which can provide a properly folded ECRG2.

Moreover, the constructed plasmid and the expression and purification procedures can facilitate folding of ECRG2 with the correctly formed disulfide bonds. In the fusion protein Trx-(linker-Xa)-ECRG2, there is a linker peptide containing SGSGSGSGSG and a factor Xa cleavage site. This linker peptide makes a relatively large space between the two proteins, facilitating the proper folding of ECRG2 in the fusion protein and the proper cleavage of ECRG2 from the fusion protein using factor Xa. Factor Xa is a protease cleaving the fusion protein without the additional amino acid at the N-terminus of the target protein. Thus, the recombinant ECRG2 contains 66 residues (Ser20-Cys85) (Fig. 1). The mass of ECRG2 determined by MALDI-TOF MS analysis is 7192.77 (SM Fig. 2).

A recent report [15] has proposed that the 13C chemical shifts can predict disulfide formation, providing simple rules for predict-
ing the redox state of cysteines. As was indicated, if the $^{13}\text{C} \beta$ chemical shift is greater than 35.0 ppm, the redox state can be assigned to oxidized. The assignment of ECRG2 provided the $^{13}\text{C} \alpha$ and $^{13}\text{C} \beta$ chemical shifts of residues Cys32 ($^{13}\text{C} \alpha$: 53.47 ppm, $^{13}\text{C} \beta$: 40.14 ppm), Cys45 ($^{13}\text{C} \alpha$: 56.55 ppm, $^{13}\text{C} \beta$: 39.10 ppm), Cys53 ($^{13}\text{C} \alpha$: 54.56 ppm, $^{13}\text{C} \beta$: 39.35 ppm), Cys64 ($^{13}\text{C} \alpha$: 60.19 ppm, $^{13}\text{C} \beta$: 35.44 ppm), Cys67 ($^{13}\text{C} \alpha$: 56.25 ppm, $^{13}\text{C} \beta$: 35.19 ppm), and Cys85 ($^{13}\text{C} \alpha$: 56.35 ppm, $^{13}\text{C} \beta$: 38.81 ppm). Apparently, all six cysteines have the $^{13}\text{C} \alpha$ chemical shifts larger than 35.0 ppm. Therefore, the recombinant ECRG2 can folded properly with correctly formed disulfide bonds, namely Cys32-Cys67, Cys45-Cys64, and Cys53-Cys85, as indicated in Fig. 1.

Secondary structures of ECRG2

The conformation of ECRG2 was explored by the far-UV CD spectrum (see Supplementary material). Analysis of the CD spectrum of ECRG2 provided the estimation of the secondary structural components: 14% of $\alpha$-helix, 29% of $\beta$-sheet, 22% of turn, and 34% of loop or coil contents. The consensus chemical shift index (CSI) of ECRG2 predicts that residues Cys64-Lys72 turn, and 34% of loop or coil contents. The consensus chemical shift spectrum of ECRG2 provided the estimation of the second-order ($^{13}\text{C}$) chemical shifts of residues Cys32 ($^{13}\text{C} \alpha$: 56.25 ppm, $^{13}\text{C} \beta$: 35.19 ppm), Cys67 ($^{13}\text{C} \alpha$: 56.25 ppm, $^{13}\text{C} \beta$: 35.19 ppm), and Cys85 ($^{13}\text{C} \alpha$: 56.35 ppm, $^{13}\text{C} \beta$: 38.81 ppm). Apparently, all six cysteines have the $^{13}\text{C} \alpha$ chemical shifts larger than 35.0 ppm. Therefore, the recombinant ECRG2 can folded properly with correctly formed disulfide bonds, namely Cys32-Cys67, Cys45-Cys64, and Cys53-Cys85, as indicated in Fig. 1.

**Binding of ECRG2 with uPA protein**

The recombinant ECRG2 contains five proline residues. Thus, 61 residues were considered in the analysis of interaction between ECRG2 and uPA protein using 2D $^1\text{H}-^1\text{H}$ HSQC spectrum of $^{15}\text{N}$-labeled ECRG2 (Fig. 4A). The distribution of amide resonances in Fig. 4A represents a native folded conformation of the recombinant ECRG2. About 92% amide resonances were assigned. There are two sets of cross peaks in Fig. 4A, one is strong and the other is weak. The set of strong peaks is from correctly folded ECRG2 whereas the set of sharp unassigned peaks can be considered as resonance signals from species of recombinant ECRG2 with mis-pairing disulfide bonds. The intensity ratio of two cross peaks for Gly75 can be used for quantitatively estimating the amount of correctly folded ECRG2 in the sample solution. The estimation of the volumes of two cross peaks for Gly75 provided 83.75% and 16.25% for strong and weak cross peaks, respectively. Since the sample solution used for 2D $^1\text{H}-^1\text{H}$ HSQC NMR experiments contains total of 0.5 mM ECRG2, the amount of correctly folded ECRG2 is estimated to be 0.42 mM. Binding of uPA protein with ECRG2 was detected using 2D $^1\text{H}-^1\text{H}$ HSQC spectrum. Cross peaks in the 2D HSQC spectrum may shift upon the binding of ECRG2 with uPA protein. Fig. 4B shows the spectrum obtained for 0.5 mM $^{15}\text{N}$-labeled ECRG2 in the presence of 0.5 mM uPA (LIVZON Pharmaceutical Group, Inc., China) (ECRG2 to uPA = 1:1). Clearly, the fast exchange between...
 resonances for bound and free proteins was observed, suggesting a weak binding affinity. The cross peaks for residues in the sequence regions Val41-Ile43, Ile47-Tyr49, and Tyr60-His65 as well as Val52, Cys67, and Glu69 showed explicit shifts. Of them, the cross peaks for residues Ala42, Ile47, Glu63, Cys64, and Cys67 shifted largely. However, the cross peaks for residues Glu21, Ala23-Leu25, Lys28-Lys29, Asp31-Ser33, Tyr35, Lys37-Tyr38, Leu50, Cys53-Thr59, Thr68, Leu71-Gly75, Val77, Phe79-His81, and Gly83-Ser84 did not shift or showed very little shift. This suggested that the residues in the sequence regions Val41-Ile43, Ile47-Tyr49, and Tyr60-His65 of ECRG2 must be in contact with uPA protein.

When 1 mM uPA (ECRG2 to uPA = 1:2) was added into the ECRG2 sample, vast majority cross peaks in the 2D 1H–15N HSQC spectrum were disappeared, only the set of sharp unassigned resonances in Fig 4A and a few strong resonances were remained (Fig. 4C). This can be explained by the fact of binding between ECRG2 and uPA. Due to a higher molecular weight of uPA, the binding of uPA to ECRG2 may cause the formation of a larger protein complex, resulting in the disappearance of the resonances for ECRG2 in the complex with uPA. The remaining sharp unassigned resonances (Fig. 4C) indicate that ECRG2 with mis-pairing disulfide bonds did not interact with uPA. On the other hand, this corroborated that only the species of ECRG2 with correctly formed disulfide bonds can bind to uPA. Nevertheless, the binding of uPA to ECRG2 is weak as indicated above. It should be mentioned that chemical shift changes can also arise from conformational changes, but since the residues showing resonance shifting are much localized (and since all peaks disappear on binding) this is less likely.

Thus, the segments Val41-Ile43, lle47-Tyr49, and Tyr60-His65 form a putative binding site of uPA on ECRG2.

**Discussion**

**Structural resemblance between ECRG2 and OMTKY3**

Turkey ovomucoid third domain (OMTKY3) is a canonical inhibitor of serine proteinase which belongs to the Kazal family of proteinase inhibitors [16]. OMTKY3 is a 56-residue polypeptide, which contains three covalent disulfide bonds. The sequence alignment (Fig. 5A) shows that ECRG2 has 40% sequence homologous to OMTKY3. The solution structure of OMTKY3 has been determined (pdb ID: 1TUR), showing one α-helix (Asn33-Ser44), one antiparallel β-sheet constructed by three β-strands (Arg21-Ser26, Lys29-Gly32, and Thr49-Gly54), and four reverse turns: Cys8-Tyr11, Gly25-Asn28, Val42-Asn45, and Ser44-Leu47. Six cysteine residues in the Kazal motif form three pairs of disulfide bonds: Cys8-Cys38, Cys16-Cys35, and Cys24-Cys56 (Fig. 1) [17]. Residues Lys13-Arg21 and Gly32-Asn33 as well as Asn36 in OMTKY3 are in the consensus region of contact with serine proteinases. The reactive site peptide bond is Leu18-Glu19 [18].

Based on the reported 3D structure of OMTKY3, the estimated secondary structure content of OMTKY3 contains approximately 21% α-helix, 29% β-sheet, and 16% turn. These values are in close proximity to the above estimated contents of secondary structural components for ECRG2. The consensus CSI and NOE connectivities (SM Fig. 4 and Figs. 2 and 3) further indicated that the predicted secondary structural regions in ECRG2 are in correspondence with the secondary structural regions in OMTKY3 (Fig. 5A). Therefore, it can be proposed that the 3D structure of ECRG2 resembles those of OMTKY3. The structure of ECRG2 was modeled by software Modeler [19] using OMTKY3 as a template, having an OMTKY3-like tertiary folding (Fig. 5B).

**Interaction of uPA protein with ECRG2**

As was indicated, Kazal-type proteinase inhibitors bind tightly to their cognate enzymes because their reactive site loops are geometrically complementary to the enzyme active sites [20,21]. OMTKY3 inhibits those serine proteinases that have the chymotryptic specificity. Several groups have reported the binding of OMTKY3 to a bacterial serine proteinase, streptomyces griseus pro-
OMTKY3 binds tightly to the active site of SGRB [22,23]. Similar to the interaction of OMTKY3 with serine proteinase, uPA exhibited binding to ECRG2 in serum-free conditioned medium from either Pg/pcDNA3.1-ECRG2 cells or Pg/pcDNA3.1 [10]. As is well known, uPA is a trypsin-like serine proteinase involved in tissue remodeling and cell migration, and inhibits aggressiveness of cancer cells, possibly through the down-regulation of uPA/plasmin activity [24,25]. The molecule of uPA comprises three domains: an N-terminal growth factor domain (GFD), a kringle domain (KD), and a serine proteinase domain. The serine proteinase domain contains the catalytic site [25,26]. In the binding of OMTKY3 with serine proteinase, the active site of serine proteinase binds to the reactive site loops Lys13-Arg21 and Gly32-Asn36 of OMTKY3 [18]. Analysis of the NMR data indicated that segments of residues Val41-Ile43, Ile47-Tyr49, and Tyr60-His65 of ECRG2 are the uPA-binding regions. The sequence locations of these binding regions in ECRG2 are very similar to those of the reactive site loops in OMTKY3 (Fig. 5A and B). In other words, the sequence regions for binding of uPA protein in ECRG2 are in correspondence with the reactive site loops for binding of serine proteinase in OMTKY3. Apparently, segments Val41-Ile43, Ile47-Tyr49, and Tyr60-His65 can be considered as the putative reactive site loops for uPA binding in ECRG2. Therefore, similar to the binding of OMYKY3 with serine proteinase, ECRG2 can bind with serine proteinase domain of uPA. However, the present research cannot provide the information about the binding site on the serine proteinase domain of uPA. This requires the further study of the interaction between ECRG2 and serine proteinase domain of uPA, and determination of the complex structure of ECRG2 with the serine proteinase domain of uPA protein.

In conclusion, the expression of ECRG2 in the oxidative cytoplasm of E. coli Origami (DE3) resulted in soluble, properly folded protein with accurately formed disulfide bonds. The structural similarity of ECRG2 to OMTKY3 was identified. The experimentally determined putative reactive site loops for binding of uPA in ECRG2 are in correspondence with the reactive site loops for binding of serine proteinase in OMTKY3.

Acknowledgment

This research was supported by the National Natural Science Foundation of China (NNSFC 30770434).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.abb.2008.08.023.

References