

NMR structure note: human esophageal cancer-related gene 2

Yingang Feng · Yong Geng · Tao Zhou ·
Jinfeng Wang

Received: 1 December 2011 / Accepted: 16 January 2012 / Published online: 21 April 2012
© Springer Science+Business Media B.V. 2012

Biological context

Esophageal cancer is one of the most common fatal cancers worldwide and occurs at very high frequencies in developing countries (Parkin et al. 1993). Esophageal cancer-related gene 2 (ECRG2), also named serine peptidase inhibitor Kazal-type 7 (SPINK7), was identified as a growth-regulating gene by comparing normal esophageal epithelia and primary squamous cell carcinomas tissues from high incidence families in Lin-Xian County of China using differentially displayed PCR (Su et al. 1998; Cui et al. 2010). It has been shown that ECRG2 is a tumor suppressor gene involved in the regulation of cell proliferation, induction of apoptosis, and inhibition of cancer cell migration, invasion and metastasis (Cui et al. 2003; Huang et al. 2007). ECRG2 also participates in centrosome amplification in a p53-dependent manner and has a role in maintaining chromosome stability (Cheng et al. 2008). Recently, it was found that ECRG2 regulates cell migration/invasion through urokinase-type plasmin activator (uPA)/urokinase-type plasmin activator receptor (uPAR) pathway by directly inhibiting uPA/plasmin activity, and

may represent a novel therapeutic target for cancer (Huang et al. 2007; Cheng et al. 2009, 2010; Cui et al. 2010).

The full length ECRG2 protein contains 85 residues, in which the N-terminal 19 residues form a signal peptide and the C-terminal 56 residues (residues 30–85) contain a typical x(8)-C-x(12)-C-x(7)-C-x(6)-Y-x(3)-C-x(2)-C-x(17)-C conserved region, coding a Kazal-type serine protease inhibitors domain (Figs. 1, 2) (Cui et al. 2010). The Kazal-type protease inhibitors with various physiological functions are conserved across species ranging from human to invertebrates (Laskowski and Kato 1980; Rimphanitchayakit and Tassanakajon 2010) while ECRG2 is classified as a serine peptidase inhibitor Kazal-type 7 and conserved in mammals (Fig. 1). Moreover, the Kazal-type protease inhibitors have a scissile peptide bond that is potentially cleaved by the protease with a very low reaction rate (Laskowski and Kato 1980; Rimphanitchayakit and Tassanakajon 2010). The scissile peptide bond of ECRG2 is between P1 residue (Ile47) and P'1 residue (Thr48) (In inhibitors, the number after “P” indicates the position of the residue towards the N-terminus of the scissile bond, while after “P'” indicates residues towards the C-terminus of the scissile bond). In addition, ECRG2 contains more residues between the first two cysteines besides its N-terminal signal peptide as compared with the other Kazal-type protease inhibitors (Geng et al. 2008).

The backbone and partial side chain resonance assignments for ECRG2 were determined previously by NMR technique (Geng et al. 2008). Utilizing a homology modeling and the NMR titration experiments, a putative reactive site of ECRG2 as the uPA-binding site was identified, suggesting that ECRG2 can inhibit uPA activity through a manner similar to the other Kazal-type inhibitors of serine proteases (Geng et al. 2008). In the present study, the complete NMR resonance assignments of ECRG2 were

Yingang Feng and Yong Geng contributed equally to this work.

Y. Feng · Y. Geng · T. Zhou · J. Wang (✉)
National Laboratory of Biomacromolecules,
Institute of Biophysics, Chinese Academy of Sciences,
15 Datun Road, Beijing 100101, China
e-mail: jfw@sun5.ibp.ac.cn

Y. Feng
Qingdao Institute of Bioenergy and Bioprocess Technology,
Chinese Academy of Sciences, No. 189 Songling Road,
Qingdao 266101, China

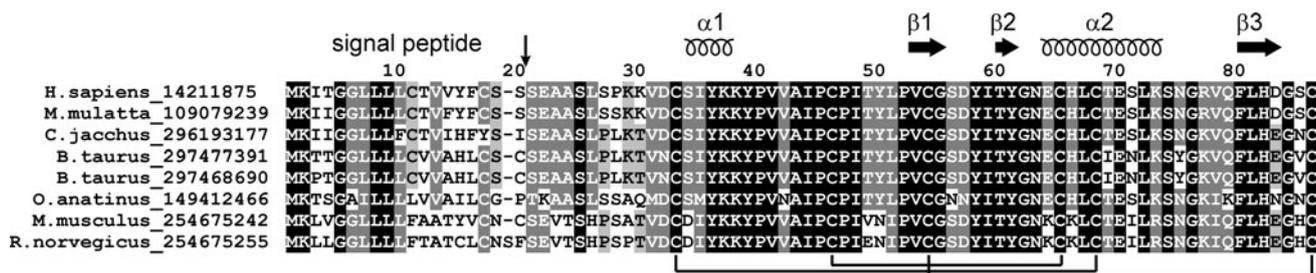
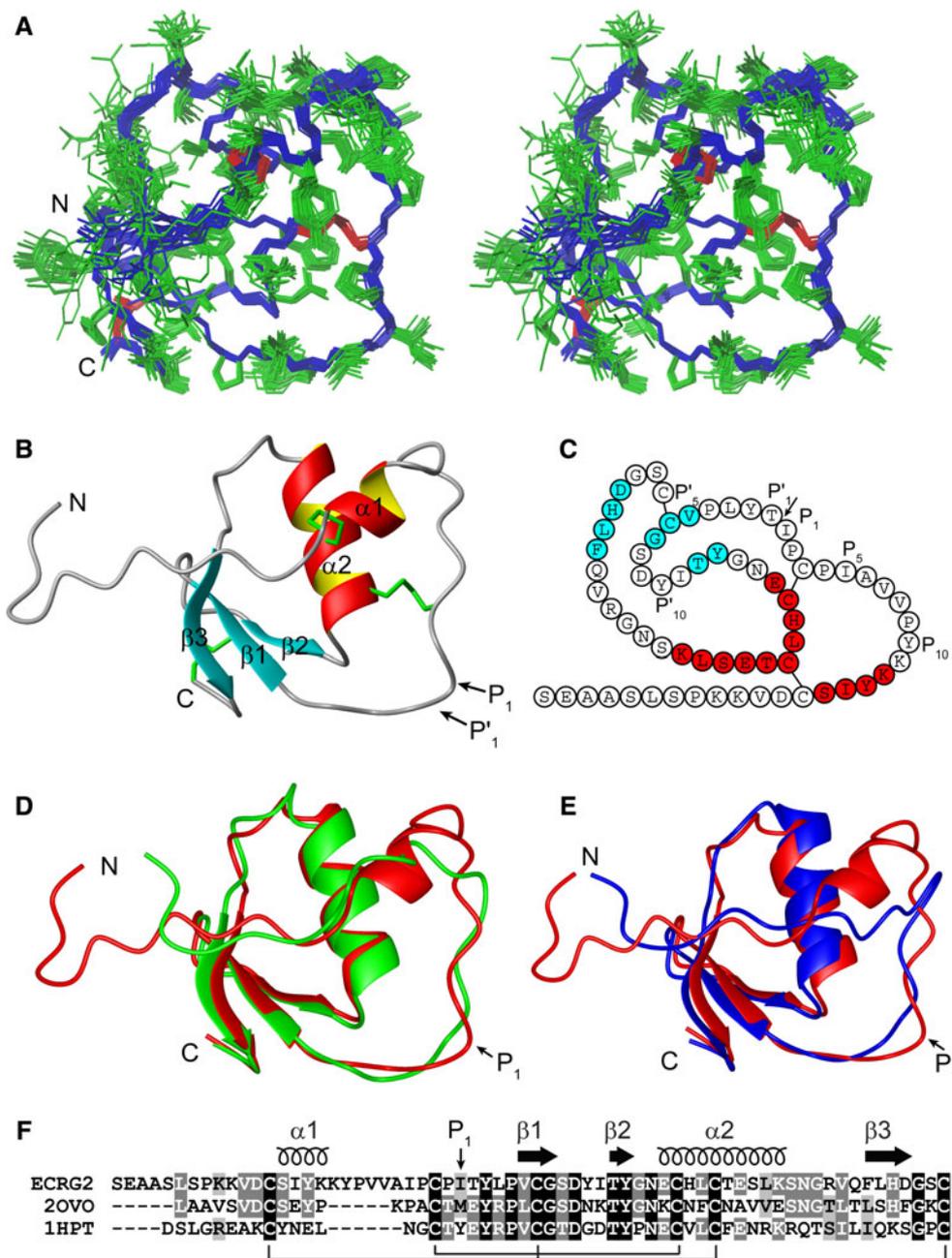


Fig. 1 Sequence alignment of mammalian ECRG2 proteins. Each pair of cysteine residues forming disulfide bonds are connected by lines. The sequence positions of secondary structures are indicated at the top of sequences. The cleavage site of signal peptide is indicated by an arrow

Fig. 2 NMR solution structure of human ECRG2. **a** Stereo view of ECRG2 structure ensemble. For clarification, residues 20–27 are not shown. Backbone are in blue, side chains of cysteines are in red, and the other side chains are in green. **b** Ribbon representation of ECRG2. The side chains of cysteine residues and disulfide bonds are shown as sticks in green. **c** The covalent primary sequence of human ECRG2. The scissile peptide bond on the reactive loop is indicated by an arrow. P1, P2, etc. denote the residues towards the N-terminus of the scissile peptide bond, while P'1, P'2, etc. denote the residues towards the C-terminus of the scissile peptide bond. The residues in α -helices and β -strands are in red and cyan, respectively. **d** Structure superimposition of ECRG2 (red) and ovomucoid (green). **e** Structure superimposition of ECRG2 (red) and PSTI (blue). **f** Structure-based sequence alignment of ECRG2, ovomucoid, and PSTI. Disulfide bonds are indicated by lines. The secondary structural elements of ECRG2 are indicated at the top of sequences



obtained, and the high quality solution structure of ECRG2 was determined by NMR spectroscopy. Analysis of the ECRG2 structure revealed a number of structural features distinct from the other Kazal-type protease inhibitors, which suggest ECRG2 to be an interesting specific inhibitor of uPA activity.

Methods and results

Protein expression and purification

Since the N-terminal 19 residues of ECRG2 form a signal peptide, these 19 amino acids were truncated in construction of the expression vector for obtaining the recombinant ECRG2 protein. The peptide ECRG2(20–85) (called as ECRG2 hereafter for simplicity) were expressed and purified as described previously (Geng et al. 2008). The NMR samples consisted of 0.5–1.0 mM ^{15}N -, ^{13}C -, or $^{15}\text{N}/^{13}\text{C}$ -labeled ECRG2(20–85) in 50 mM deuterated acetate buffer (pH 5.0) containing 100 mM KCl, 0.01 % (w/v) 2,2-dimethyl-2-silantane-5-sulphonate (DSS), 0.01 % (w/v) NaN_3 , and 10 % (v/v) D_2O .

NMR spectroscopy

All NMR experiments were run at 293 K on a Bruker DMX 600 MHz spectrometer equipped with a z-gradient triple-resonance cryo-probe. The 2D and 3D NMR experiments including ^1H – ^{15}N HSQC, ^1H – ^{13}C HSQC, ^1H – ^{13}C – ^{15}N HNC0, HN(CA)CO, HNCACB, CBCA(CO)NH, HBHA(CO)NH, H(C)CH-TOCSY, and (H)CCH-TOCSY, ^1H – ^{15}N TOCSY-HSQC and NOESY-HSQC, and ^1H – ^{13}C NOESY-HSQC were adopted for the backbone and side chain resonance assignments. The mixing times for ^{13}C – ^{13}C TOCSY and ^1H – ^1H NOESY experiments were 12 and 200 ms, respectively. All NMR data were processed and analyzed with FELIX software (Accelrys Inc.).

Structure calculation

Initial structures of ECRG2 were generated using the CANDID module of the CYANA software (Herrmann et al. 2002), and the NOE assignments given by CANDID were checked manually. The structures were refined in explicit water using CNS software (Brunger et al. 1998) and RECOORDScript (Nederveen et al. 2005). The dihedral restraints generated using TALOS+ (Cornilescu et al. 1999; Shen et al. 2009) and the disulfide bonds C32–C67, C45–C64, and C53–C85 were introduced in the structure refinements. 100 structures were calculated, and 20

structures with lowest energies were selected for final analysis. The quality of the structures was analyzed using MOLMOL (Koradi et al. 1996) and PROCHECK-NMR (Laskowski et al. 1996). All figures depicting structures were generated using MOLMOL. The Dali server (Holm and Rosenstrom 2010) and SSM server (Krissinel and Henrick 2004) were utilized for searching the structural similarity.

Structure of ECRG2

ECRG2 belongs to the Kazal-type protease inhibitor which contains six conserved cysteine residues forming three intra-molecular disulfide bonds between pair wise cysteine residues Cys32–Cys67, Cys45–Cys64, and Cys53–Cys85. The correct formation of these disulfide bonds in recombinant ECRG2 is crucial for further structural and functional studies of this protein. The obtained chemical shifts and NOE patterns in our previous study suggested that the purified ECRG2 contained three correctly formed disulfide bonds (Geng et al. 2008), and this was further confirmed by the structural determination in the present study.

The initial structures of ECRG2 were calculated with the chemical shift assignments and NOE data by an automatic method using the software CYANA. No disulfide bond restraints were used for the initial structure calculation. The structures generated by CYANA showed a Kazal fold with six correctly paired cysteine residues as they should be in a Kazal fold (data not shown). Since the initial structure calculation was performed without the disulfide bond restraints or manual NOE assignments, the obtained initial structures proved the correct folding and oxidation of the purified recombinant ECRG2 proteins. Therefore, the disulfide bond restraints were added in the subsequent structure calculations.

The solution structures of ECRG2 were then determined with NOE-derived distance, backbone dihedral angle, hydrogen bond, and disulfide bond restraints (Table 1). Residues 29–85 of ECRG2 form a compact core structure while the N-terminal region (20–28) is disordered (Fig. 2). The core structure contains two α -helices ($\alpha 1$: 33–36 and $\alpha 2$: 63–72), an anti-parallel β -sheet formed by three β -strands ($\beta 1$: 52–54, $\beta 2$: 59–60, and $\beta 3$: 79–82), and four loops connecting the secondary structural elements: $L_{\alpha 1\beta 1}$ (37–51), $L_{\beta 1\beta 2}$ (55–58), $L_{\beta 2\alpha 2}$ (60–62), and $L_{\alpha 2\beta 3}$ (73–79) (Fig. 2). The three disulfide bonds, C32–C67, C45–C64, and C53–C85 are established between the N-terminus of helix $\alpha 1$ and helix $\alpha 2$, between middle of loop $L_{\alpha 1\beta 1}$ and helix $\alpha 2$, and between strand $\beta 1$ and the C-terminus of the protein, respectively. No large hydrophobic core is formed except four small hydrophobic regions (Val52–Tyr60–

Table 1 Experimental restraints and structural statistics of ECRG2

Parameter	Value
<i>Number of distance restraints</i>	
Total NOEs	2,434
Intraresidual NOEs	893
Sequential NOEs	513
Medium-range NOEs	281
Long-range NOEs	507
Ambiguous NOEs	240
Hydrogen bonds	26
<i>Number of dihedral angle restraints</i>	
ϕ angle	45
ψ angle	49
<i>PROCHECK_NMR Ramachandran map analysis</i>	
Most favored regions	85.7 %
Additional allowed regions	13.1 %
Generously allowed regions	0.5 %
Disallowed regions	0.6 %
<i>RMSD from mean (residues 29–85) (Å)</i>	
Backbone heavy atoms	0.42 ± 0.08
All heavy atoms	0.79 ± 0.09
<i>RMSD from mean (secondary structure residues^a) (Å)</i>	
Backbone heavy atoms	0.26 ± 0.05
All heavy atoms	0.60 ± 0.04
<i>RMSD from experimental restraints</i>	
Distance restraints (Å)	0.0038 ± 0.0005
Dihedral angle restraints (°)	0.38 ± 0.08
<i>RMSD from ideal geometry</i>	
Bond lengths (Å)	0.0101 ± 0.0003
Bond angles (°)	1.05 ± 0.04
Impropers (°)	1.27 ± 0.08
<i>CNS energies</i>	
Lennard-Jones (kcal/mol)	−280 ± 9
Electrostatic (kcal/mol)	−2,505 ± 37

^a Secondary structure residues: 33–36, 52–54, 59–60, 63–72, 79–82

Leu66-Val77-Phe79, Ile34-Tyr35-Tyr38-Pro44, Pro46-Thr48-Leu50, and Lys36-Val40-Leu71) in the ECRG2 structure. These hydrophobic residues and three disulfide bonds were the key factors contributing to the stability of the structure.

Deposition of assignments and structure coordinates

The atomic coordinates and NMR-derived restraints of ECRG2 have been deposited in the Protein Data Bank (<http://www.rcsb.org/>) with accession code 2LEO. The chemical shift assignments of ECRG2 have been deposited in the Biological Magnetic Resonance Data Bank (<http://www.bmrb.wisc.edu/>) with accession number 17719.

Discussion and conclusions

ECRG2 is a Kazal-type inhibitor with distinct structural features

The structural homology search in PDB using the Dali server and the SSM server found a Kazal-type inhibitor fold of the ECRG2 structure and a large number of Kazal domain structures similar to that of ECRG2. The high structural similarity was identified for ECRG2 with ovomucoid and human pancreatic secretory trypsin inhibitor (PSTI) (PDB codes 2OVO and 1HPT) (Fig. 2d, e), providing the highest DALI Z scores of 8.0 and 6.3, and the RMSDs of 1.9 and 2.6 Å, respectively.

However, careful analysis revealed several significant differences between structures of ECRG2 and other Kazal domains. The first major difference was shown by the backbone and secondary structure in the segment between the first two cysteine residues. The residues in this segment are not conserved in Kazal domains and the conformation of this segment in ECRG2 is quite different from that in other Kazal domains. An additional short helix was formed by the residues in this segment of ECRG2 (Fig. 2d–f). Second, the reactive site is more hydrophobic in ECRG2 than in other Kazal domains. The ECRG2 reactive site loop, residues from P4 to P'4 positions (PCPITYLP), contains five hydrophobic but no charged residues. However, Ovomucoid and PSTI contain two or three hydrophobic and one to three charged residues. Third, ECRG2 has more solvent exposed hydrophobic residues (Ile, Leu, Val, Pro, Phe) except those in the reactive site loop. ECRG2 contains 9 solvent-exposed hydrophobic residues in a total of 66 residues, but ovomucoid and PSTI contain only 2 and 5 solvent-exposed hydrophobic residues in a total of 56 residues, respectively (Fig. 3a, c, e). Fourth, the electrostatic potential surface of ECRG2 is quite different from ovomucoid and PSTI (Fig. 3b, d, f). On the surface of ECRG2, there are more neutral and negatively charged residues around the reactive loop, and positively charged residues on the opposite side. However, more positively charged residues are around the reactive site of Ovomucoid and PSTI.

The distinct structure features of ECRG2 revealed in this study suggest that ECRG2 may have unique functions different from other Kazal-type inhibitors. The determined solution structure here could provide a structural basis of ECRG2-based drug development since the function of ECRG2 as an uPA inhibitor makes it to be a potential drug. However, the complicated purification procedure and poor long-term stability of ECRG2 become the shortcomings in dealing with ECRG2. To overcome these shortcomings needs the further engineering studies based on the determined ECRG2 structure.

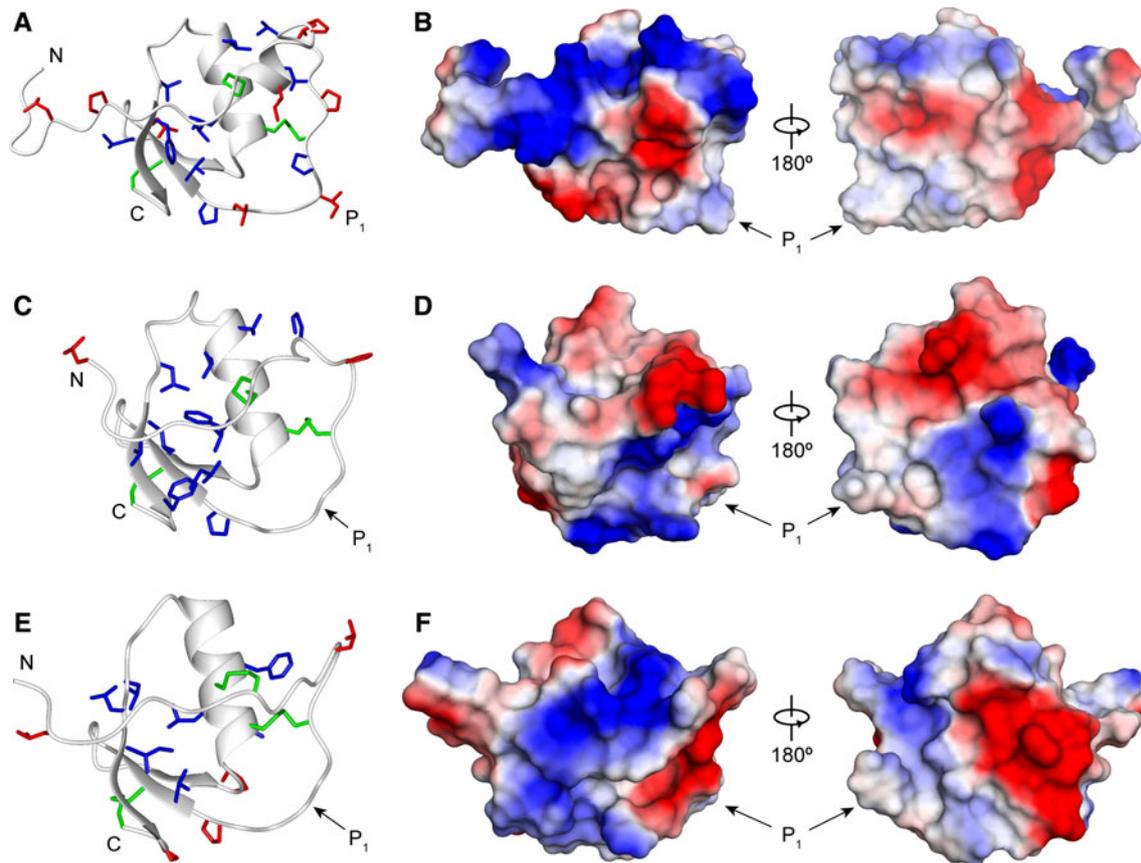


Fig. 3 Comparison of electrostatic potential surfaces and hydrophobic residues of ECRG2 (**a** and **b**), ovomucoid (**c** and **d**), and PSTI (**e** and **f**). **a**, **c**, and **e** Ribbon diagrams of ECRG2, ovomucoid, and PSTI. The solvent-accessible hydrophobic residues (ILE, LEU, PHE, PRO, and VAL) are shown as *red sticks* and buried hydrophobic

residues are in *blue*. Disulfide bonds are indicated by *green sticks*. **b**, **d**, and **f** Electrostatic potential surfaces of ECRG2, ovomucoid, and PSTI. Positively charged surfaces are in *blue*, and negatively charged surfaces are in *red*

Conclusions

ECRG2 is a Kazal-type inhibitor with distinct structural features. The determined ECRG2 structure has an additional short α -helix, hydrophobic reactive site, more solvent-exposed hydrophobic residues, and unique surface charge distribution, showing a unique structure different from the known Kazal-type inhibitors. These imply that ECRG2 may have a specific inhibitor activity which may be an attractive feature for uPA inhibitors in the future discovering of the potential cancer therapy.

Acknowledgments This work was supported by the National Natural Science Foundation of China (30770434).

References

- Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 54:905–921
- Cheng X, Shen Z, Yang J, Lu SH, Cui Y (2008) ECRG2 disruption leads to centrosome amplification and spindle checkpoint defects contributing chromosome instability. *J Biol Chem* 283:5888–5898
- Cheng X, Shen Z, Yin L, Lu SH, Cui Y (2009) ECRG2 regulates cell migration/invasion through urokinase-type plasmin activator receptor (uPAR)/beta1 integrin pathway. *J Biol Chem* 284:30897–30906
- Cheng X, Lu SH, Cui Y (2010) ECRG2 regulates ECM degradation and uPAR/FPRL1 pathway contributing cell invasion/migration. *Cancer Lett* 290:87–95
- Cornilescu G, Delaglio F, Bax A (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR* 13:289–302
- Cui Y, Wang J, Zhang X, Lang R, Bi M, Guo L, Lu SH (2003) ECRG2, a novel candidate of tumor suppressor gene in the esophageal carcinoma, interacts directly with metallothionein 2A and links to apoptosis. *Biochem Biophys Res Commun* 302:904–915
- Cui Y, Bi M, Su T, Liu H, Lu SH (2010) Molecular cloning and characterization of a novel esophageal cancer related gene. *Int J Oncol* 37:1521–1528
- Geng Y, Feng Y, Xie T, Dai Y, Wang J, Lu SH (2008) Mapping the putative binding site for uPA protein in Esophageal Cancer-Related

- Gene 2 by heteronuclear NMR method. *Arch Biochem Biophys* 479:153–157
- Herrmann T, Guntert P, Wuthrich K (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J Mol Biol* 319:209–227
- Holm L, Rosenstrom P (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res* 38:W545–W549
- Huang G, Hu Z, Li M, Cui Y, Li Y, Guo L, Jiang W, Lu SH (2007) ECRG2 inhibits cancer cell migration, invasion and metastasis through the down-regulation of uPA/plasmin activity. *Carcinogenesis* 28:2274–2281
- Koradi R, Billeter M, Wuthrich K (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graph* 14:51–55
- Krissinel E, Henrick K (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* 60:2256–2268
- Laskowski MJ, Kato I (1980) Protein inhibitors of proteinases. *Annu Rev Biochem* 49:593–626
- Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* 8:477–486
- Nederveen AJ, Doreleijers JF, Vranken W, Miller Z, Spronk CA, Nabuurs SB, Guntert P, Livny M, Markley JL, Nilges M, Ulrich EL, Kaptein R, Bonvin AM (2005) RECOORD: a recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank. *Proteins* 59:662–672
- Parkin DM, Pisani P, Ferlay J (1993) Estimates of the worldwide incidence of eighteen major cancers in 1985. *Int J Cancer* 54:594–606
- Rimphanitchayakit V, Tassanakajon A (2010) Structure and function of invertebrate Kazal-type serine proteinase inhibitors. *Dev Comp Immunol* 34:377–386
- Shen Y, Delaglio F, Cornilescu G, Bax A (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR* 44:213–223
- Su T, Liu H, Lu S (1998) Cloning and identification of cDNA fragments related to human esophageal cancer. *Chin J Oncol* 20:254–257