The SEP domain of p47 acts as a reversible competitive inhibitor of cathepsin L

Michael Soukenik, Anne Diehl, Martina Leidert, Volker Sievert, Konrad Büssow, Dietmar Leitner, Dirk Labudde, Linda J. Ball, Annette Lechner, Dorit K. Nagler, Hartmut Oschkinat

*Corresponding author. Fax: +493094793169.
E-mail address: oschkinat@fmp-berlin.de (H. Oschkinat).

1 Present address: Structural Genomics Consortium, University of Oxford, Botnar Research Centre, OX37LD, UK.

Abbreviations: NaP, sodium phosphate buffer; SEP, Saccharomyces cerevisiae SHP1, Drosophila melanogaster eyes closed gene, vertebrate p47; NOE, nuclear Overhauser effect; CSP, chemical shift perturbations

Abstract The solution structure of the human p47 SEP domain in a construct comprising residues G1-S2-p47(171–270) was determined by NMR spectroscopy. A structure-derived hypothesis about the domains’ function was formulated and pursued in binding experiments with cysteine proteases. The SEP domain was found to be a reversible competitive inhibitor of cathepsin L with a K_{i} of 1.5 μM. The binding of G1-S2-p47(171–270) to cathepsin L was mapped by biochemical assays and the binding interface was investigated by NMR chemical shift perturbation experiments.

Keywords: SEP; p97; p47; Cathepsin L; NMR; Protein structure

1. Introduction

AAA-type ATPases (ATPases associated with various cellular activities) and their binding partners form protein systems involved in the regulation of membrane fusion [1], postmitotic reassembly of the Golgi apparatus, ubiquitin-related processes and DNA replication. p97 is a member of this family, which includes the Clp family proteins and yeast CDC48 [2]. It forms a hexameric structure and contains two ATPase domains, belonging thus to the AAA+ superfamily. p97 binds a variety of proteins with various functions, for example VCIP135 (valosin-containing protein (p97)/p47 complex-interacting protein p135) [3] and p47. It is composed of three domains: an N-terminal domain N, which binds the conserved 370 residue eukaryotic adaptor protein p47 [2] followed by two ATPase binding domains called D1 and D2. Binding of p47 to p97 activates the ubiquitin-binding site in p47 and deactivates the ATPase activity of p97. Like many adaptor proteins, p47 is also characterised by a modular composition, comprising an N-terminal UBA domain [4,5], a SEP (Saccharomyces cerevisiae, Drosophila melanogaster eyes closed gene and vertebrate p47) domain and a C-terminal UBX domain. The latter consists of about 80 amino acids and adopts a ubiquitin-like β-grasp fold [6]. p47 binds to p97 by insertion of a conserved loop of its UBX domain into a hydrophobic pocket within the p97 N domain [2].

As yet unknown is the function of the SEP domain, which occurs frequently and mainly in single units. The co-occurrence of SEP and UBX domains in evolutionarily conserved combinations is also interesting. Almost all proteins containing a SEP domain are succeeded closely by a UBX domain, according to SMART [7]. The solution structure of the SEP domain from rat p47 using a construct containing residues 171–246 was recently determined by Yuan et al. [8]. It was suggested that the SEP domain is involved in the binding of p47 to p97 [8]. Here, we present the structure of the human p47 SEP domain, using an extended construct (residues 171–270). Furthermore, we have derived a novel, structure-based hypothesis about the function of the p47 SEP domain, implicating it as a reversible competitive inhibitor of the lysosomal cysteine protease cathepsin L. The binding interface of the SEP domain was investigated by NMR.

2. Materials and methods

2.1. Expression/purification

The DNA sequence encoding human p47(171–270) SEP domain (NM 016143) was cloned from a human foetal brain cDNA expression library (hEx1) [9] into the expression plasmid pQTEV (GenBank Accession No. AY243506). Construct residues 3–102 correspond to sequence positions 171–270 from human p47. Two N-terminal residues from the cloning site are referred to as G1 and S2. The SEP domain was expressed as His-tagged-fusion protein in Escherichia coli and purified on a MC-Poros column (Applied Biosystems) loaded with Ni^{2+}. The His tag was cleaved using TEV protease, at 4 °C, overnight. The untagged SEP domain was further purified by a final gel filtration step. Uniformly^{13}N- and^{15}C/^{15}N-labelled samples of the p47 SEP
domain were prepared by growing cells in minimal medium containing either 0.5 g/l \(^{15}\text{NH}_4\text{Cl} \) or 0.5 g/l \(^{15}\text{NH}_4\text{Cl} \) and 2 g/l \(^{15}\text{NH}_4\text{Cl} \), respectively. Deuterated \(^{15}\text{N}\)-labelled protein was prepared by growing cells in 99.8% D\(_2\)O, using non-deuterated glucose.

2.2. NMR spectroscopy/structure calculation

All NMR spectra were acquired at 300 K and recorded on a Bruker DRX 600 MHz spectrometer equipped with a cryoprobe, using the programs PASTE and PAPST [10], except for the \(^{15}\text{N}\)-edited-HSQC which was recorded on a Bruker DMX 750 MHz spectrometer. Uniform \(^{13}\text{C}\), \(^{15}\text{C}\) and \(^{15}\text{N}\)-labelled proteins with concentrations of 1 and 2 mM, respectively, in a buffer consisting of 20 mM sodium phosphate buffer (pH 7.5), 150 mM NaCl and 0.02% NaN\(_3\). At pH 5.6 in 90% H\(_2\)O/10% D\(_2\)O were used for the experiments. Spectra were processed using the XWINNMR software (Bruker Biospin). For backbone assignments, a complete series of 2D side chain-selective experiments [11] and 3D triple resonance CBCa(CO)NH/CBCaNH and HA(CAnCO)NH/HA(CAn)NH spectra were used [12]. Side chain resonances were identified from 3D HBHa(CO)NH, H(BCCaCO)NH/TOCSY and (H)CC(CO)NH-TOCSY experiments in H\(_2\)O. Inter-proton distances for structure calculation were derived from a 2D NOESY (mixing time 60 ms), a 3D \(^{15}\text{N}\) NOESY–HSQC (80 ms) both in 90% H\(_2\)O/10% D\(_2\)O and a 3D \(^{13}\text{C}\) HMQC–NOESY (80 ms) spectrum in 100% D\(_2\)O [12]. The SPARKY software was used for assignment [13].

\(^{15}\text{N}\) T\(_1\) and T\(_2\) relaxation data were measured as described previously [12]. Both T\(_1\) and T\(_2\) relaxation times were extracted from two series comprising each of 11 spectra with relaxation delays of 12, 52, 102, 152, 202, 302, 402, 602, 902, 2002 and 3002 ms for T\(_1\) measurements and 6, 10, 18, 26, 34, 42, 82, 122, 162, 202 and 242 ms for T\(_2\) measurements. Heteronuclear \(^{1}H\)-\(^{15}\text{N}\) nuclear Overhauser effect (NOE) experiments were also recorded. Structures were calculated with the program ARIA 1.2 [14] using 113 manually assigned NOEs representing secondary structure NOE patterns, 18 hydrogen bonds and 19 dihedral angles derived from specific NOE experiments [15]. After water refinement, the final ensemble was analysed by MOLMOL [16] and PROCHECK-NMR [17].

2.3. Kinetic binding experiments with cathepsins B, X and L

Human cathepsins B and X were expressed and purified as described previously [18,19]. The cDNA for human procathepsin L was amplified by PCR from a human placenta cDNA library (BD Biosciences Clontech, Palo Alto, CA) using gene-specific primers 5'-CCGC-TCGAGAAAAGAGAGGCTGAAGCT-3' and 5'-CAGTTT-3', and 5'-GATTTGCAGCCCTCAGACATGTCGCTGG-3' and Expand DNA polymerase (Roche Applied Science, Mannheim, Germany). Cathepsin L was expressed and purified according to the procedure described for cathepsin B [18]. Kinetic experiments were performed as previously described [18]. Fluorescence was monitored on a SPEX FluoroMax spectrophotofluorimeter (Jobin Yvon, Edison, NJ). All kinetic measurements were performed at pH 5.6 in the presence of 50 mM sodium citrate, 1 mM EDTA, 2 mM DTT, 0.2 M NaCl and 3% DMSO. The assays for cathepsins B and L were carried out at pH 5.5 using the substrate carbobenzoxy-LL-phenylalanine-4-methylcoumarinyl-7-amide (Cbz-FR-MCA), for carboxypeptidase A and B, respectively. The assays for cathepsins B and L were carried out at pH 5.5 using the substrate carbobenzoxy-LL-phenylalanine-4-methylcoumarinyl-7-amide (Cbz-FR-MCA), for ca-
and β2 and before α1. F243 is less than 80% conserved but also shown, since it is found at the end of this plane near the flexible C-terminus. The black-labelled residues D203, P204 and N206 are found on the domain surface in a loop preceding α1 above the diagonal plane. Residues P219, L222 and V231 lie below this plane and are mainly involved in hydrophobic contacts between the β-sheet and α2. Yuan et al. [8] determined recently the structure of the rat p47 SEP domain (1VAZ). The overall structure of the human p47 SEP domain is the same but differs slightly in the orientation of the loop between α2 and β4 (Fig. 2A, highlighted in green). We assigned manually those NOEs which determine the loop orientation (available upon request from the authors). Our construct which did not contain a His tag showed NOEs which are indicative of contacts between the loop α2/β4 and strand β4. The potential binding site is well-ordered and very similar in both human and rat structures. The Cα RMSD of two lowest energy structures in this region (residues 179–195, 198–217 and 233–241) was 1.74 Å. Interestingly, no structural differences are caused by the three point mutations between the human and rat sequences. The mutation S189T is structurally neutral and the other two mutations appear in the flexible N-terminus.

The distribution of conserved hydrophobic residues and the calculated lipophilic and electrostatic potentials on the domain surface are shown in Fig. 2B–D. Residues F191, Y201, P204, L210, I213, P219, L222, L233 and M235 cover a continuous area (Fig. 2B, left). The surface on the reverse side (Fig. 2B, right) shows residues W187 and F243 making up one conserved area and the loop containing residues D203, P204 and N206 found in a loop above a cross-sectional plane comprised by residues L210, F209, F191, L198, W187 and F243. Structural figures were generated using MOLMOL [16].
accommodating the side chains of interaction partners. This disc shaped pocket is large enough to bury a tryptophan side chain.

The electrostatic potential of the accessible surface area reveals a dipole character of the p47 SEP domain. Fig. 2 shows an acidic stretch, comprising residues D234, E236 and D237 from b4 and residues D240 and E241 from the C-terminal tail, on one side of the molecule. The basic residues R214, R215 from helix 1 and R223, R224 from helix 2 create a positive potential on the opposite side.

3.2. Dynamics of the investigated SEP construct

The SEP domain boundaries were indicated in the SMART database as S176 and T270. Despite considerable sequence homology at the C-terminus, the structured region ends at V244. To investigate the flexibility of the C-terminus, we determined 15N T1, T2 and heteronuclear NOEs (relaxation data not shown). The ratio T1/T2 drops significantly preceding residues D179 and following K245. The heteronuclear 1H–15N NOE measurements showed positive signals for the residues between D179 and K245, whereas the signals from the flexible termini were strongly negative. These data show that the N-terminal (G1-S2-p47(171–178)) and C-terminal (p47(245–270)) tails are therefore flexible in solution. The rotational correlation time τc was determined to be 6.4 ns, indicating a monomeric state for the domain under the conditions investigated.

Further experiments involving hydrogen exchange support the results of the relaxation measurements. After lyophilisation and redissolving in D2O, we immediately measured 1H–15N HSQC experiments. Both flexible tails exchanged rapidly with D2O. Amide protons from the structured regions were protected to varying extents. After 2 h, 15N-attached protons from residues 182–187, 190–192, 199, 210–214, 222, 234, 236 and 238, which are located in the core, were not exchanged. In summary, the data from relaxation experiments and D2O exchange, together with the number of assigned NOE restraints per residue, show the SEP domain to extend from residue D179 to V244.

3.3. A structure-derived hypothesis: The p47 SEP domain interacts with cysteine proteases

Neither the activity nor the cellular function of the p47 SEP domain is known to date. In light of the existing structural investigations and the biological context, we have developed a hypothesis based on the structural properties of two loops and the similarity of the overall structure of the SEP domain to inhibitors of cysteine proteases [22]. The p47 SEP domain has two very rigid and well-ordered loops (b1–b2, b3–a1), which are located close to the conserved residues of the protein core. The loop between b3 and a1 is particularly highly conserved.

The C-terminus of the domain is close to these two loops, in particular to the loop connecting b1 and b2. Such arrangements are also seen in inhibitors of cysteine proteases, such as cystatins and stefins [23]. Due to the structural similarity, we hypothesised that the SEP domain may bind to and inhibit cysteine proteases.

We therefore investigated the inhibition of three cysteine proteases, cathepsins L, B and X, by the p47 SEP domain. The experiments showed clearly that G1-S2-p47(171–270) is a reversible competitive inhibitor of cathepsin L with a $K_i$ of 1.5 μM (Fig. 3). Cathepsin B was much more weakly and cathepsin X not at all inhibited at the highest concentration of the SEP construct used (15 μM). The reduced or absent affinity of the SEP domain for cathepsins B and X ($K_i$ > 50 μM) is likely to be due to unfavourable interactions with the occluding loop in cathepsin B [24] or the mini loop in cathepsin X [25].

NMR titration experiments were used to map the binding site between cathepsin L and the 15N-labelled deuterated SEP domain. In agreement with our hypothesis, we observed chemical shift perturbations (CSP) for the amide proton and nitrogen resonances of residues S189 and G190 which are...
like manner or inhibits the activity of cathepsin L in vivo remains to be determined.

The structural coordinates have been deposited in the PDB under accession code 1SS6.

Acknowledgements: We are grateful for the financial support of this work by the BMBF-Leitprojekt “Strukturanalyse mit hohem Durchsatz für medizinisch relevante Proteine – Proteinstrukturfabrik” (Fk.01GG9812) and by the Friedrich-Baur-Stiftung (0031/2003).

References