Supplementary Material (ESI) for Chemical Communications

Crystal structure and silica condensing activities of silicatein α/cathepsin L chimeras.

Michael Fairhead[a], Kenneth A. Johnson[b], Stephen A. McMahon[b], Thomas Kowatz[b], Lester G. Carter[b], Muse Oke[b], Huanting Liu[b], James H. Naismith*[b] and Christopher F. van der Walle*[a]

[a] *SIPBS, University of Strathclyde, 27 Taylor Street, Glasgow, Scotland. E-mail: chris.walle@strath.ac.uk*

[b] *The Centre for Biomolecular Science, University of St. Andrews, North Haugh, St. Andrews, Scotland. E-mail: naismith@st-andrews.ac.uk*

Additional supporting material

Molecular biology

The procathepsin-L gene (amino acids 18 to 333) was amplified from clone 6712564 (LGC Promochem) harbouring preprocathepsin-L (UniProtKB/TrEMBL entry P07711) by PCR using primers PCLFP01 (5′-end) and PCLRP01 (3′-end) and ligated into *Pichia pastoris* vector pPICZ-α-B (Invitrogen) which contains a gene for Zeocine™ resistance as well as an α factor secretion signal which leads to secretion of recombinant protein into the growth media. Mutations to the loop region (amino acids 285-294) were made using primers PCLFP01 and LOOPRP01 in a standard PCR protocol, digesting the product with PstI and SspI and ligating into pPICZ-α-procathepsin-L, similarly digested. All other mutations were made using the Quikchange™ method (Stratagene) and the sequences confirmed by the University of Dundee Sequencing Service.

Primer	**Sequence 5′ to 3′**	**Purpose**
PCLFP01	GCGAAGGCTGCAGGAGCTACTCTAACATTTGAT	Introduce PstI site
PCLRP01	GATGGGGTGACACATTGGCGCCTAGAGC	Introduce SacII site
CATFPCS [a]	CAGGGTCAGTGTGGTTCTTCTTGGGCTTTTAGTGC	C25S mutation
CATFPASY [a]	CAGGGTCAGTGTGGTGCTTCTTACGCTTTTAGTGCT ACTG	S24A W26Y
CATFPLNHAM [a]	GACTGTAGCAGTGAAGACTTGAATCATGCTATGCTG GTGGAAGGC	M161L, D162N, G164A, V165M mutation
LOOPRP01	CACGACCACCAACCGATGCCTAAATAGAGGTTGTT GGTCTTTATAACCGAC	^{173}ESTESDNN^{180} to ISNNQ mutation
CATFP2S [a]	TTTATTTTGAGCCAGACTGTAGCAGTTCAAGCTTGA ATCATGCTATGCTGGTGGTTG	E159S, D160S mutation
CATRP4S [a]	GTCCTTCCTGTTCTATAAAGAAGGCATTTATTTTTCG TCAGACTGTAGCAGTTCAA	E153S, P154S mutation

[a] only the forward primer is shown

Protein expression and purification

Pichia pastoris strain X-33 was transformed with plasmid using lithium chloride as described in the Invitrogen Easyselect™ *Pichia* Expression Kit manual. Transformed cells were plated onto YPG-agar plates (2% yeast extract, 2% peptone, 2% glycerol, 2% agar) containing 100 µg/mL of zeocin. After 3 days at 31 °C the resulting colonies were screened for protein expression. Cells were grown in 50 mL of YPG medium in a 250 mL baffled flask (Nalgene, UK) for 48 h at 30 °C. Cells were then transferred to 1000 mL of YPM (2% yeast extract, 2% peptone, 1% methanol) medium and grown in a 2 L baffled flask for 96 h at 20 °C, with fresh methanol added every 24 h. Wet cell biomass at the end of this period was typically 70-80 g/l. The clarified supernatant was concentrated to 40 mL, $(NH_4)_2SO_4$ added to 1 M and loaded onto a 20 mL Phenyl Sepharose 6 Fast Flow column (GE healthcare, UK) pre-equilibrated with buffer A (20 mM Tris.HCl, pH 8.2) + 1 M $(NH_4)_2SO_4$. After loading, the column was washed with buffer A + 1 M $(NH_4)_2SO_4$, the protein eluted isocratically in buffer A and desalted. The sample was then loaded onto a 5 mL Q Sepharose column and eluted with a linear gradient from 0 to 1 M NaCl. Protein containing fractions were exchanged into buffer B (20 mM CH_3COONa pH 5), loaded onto a 5 mL SP Sepharose column and eluted with a linear gradient from 0 to 1 M NaCl. Typically, 20-40 mg of pure protein were obtained per litre of culture supernatant, depending upon the construct being expressed. Processing of inactive procathepsin-L mutants to the mature form was done by incubating 20 mg of the selected mutant with 0.2 mg of activated wild type cathepsin-L in 0.1 M CH_3COONa pH 5, with 2 mM DTT for 48-96 h at room temperature. Processing was checked by SDS-PAGE and continued until a single band at 24 kDa (mature form) was present. Processed enzyme was then exchanged into 10 mM CH_3COONa pH 5 and stored at –80 ˚C.

Structural Biology

The protein was screened for crystallization against 4 commercial 96-condition screens: 1)Wizard 1 and 2 (Emerald Biosystems); 2)JCSG+, 3)Classics and 4)Pegs (GE Healthcare). Conditions which gave crystals were used to design optimization screens based on a simple stochastic method using in-house software[1]. The crystal used for data collection was obtained by the hanging drop method using protein at 10mg/ml mixed 2µl:1µl with the well solution of 450µl containing 17.36% PEG3350, 0.1M sodium acetate, pH4.5, and 0.1M lithium sulphate. Prior to exposure to X-rays, the crystal was transferred to a cryoprotecting solution containing 20% PEG3350, 0.1M sodium acetate, pH 4.5, 0.1M lithium sulphate and 20% PEG400. The crystal was flash frozen in liquid nitrogen and transported in a dry cryogenic dewar to beamline ID29 at the European Synchrotron Radiation Facility. Data were collected from a single crystal at 100 K at a wavelength of 1.27 Å. The data were indexed and merged using Denzo and Scalepack in the integrated package HKL2000[2].

The structure was solved using PHASER[3, 4] using a monomer of cathepsin (PDB 1MHW) as a model finding four monomers in the asymmetric unit. Manual modification of the protein to convert the sequence to 4SER used COOT[5]. The structure was refined using REFMAC5[6], TLS parameters, isotropic B-thermal factors and NCS restraints were employed throughout. Water molecules were added using COOT[5]. Structure quality was checked with PROCHECK[7] and MOLPROBITY[8]. The final structure and experimental data were deposited with the PDB[9] and have accession code 2VHS.

Table 1 X-ray data

Data collection	**4SER**
λ (Å)	1.27
Resolution Last shell (Å)	28 – 1.5 (1.55 - 1.5)
Spacegroup	P1
Cell (Å)	a = 56.8 b = 58.1 c = 70.2 α = 105.7 β = 105.0 γ = 105.1
Unique refl's	108723
Average redundancy	1.9 (1.7)
I/σ	20 (2.1)
Complete (%)	92 (74)
R_{merge}	0.039 (0.293)
Refinement	
R %	18.8 (26.4)
R_{free} %	22.0 (32.1)
rmsd bonds (Å) /angles (°)	0.006 / 0.985
Ramach'n favoured (%)	98
PDB accession code	2VHS

REFERENCES

1. B. Rupp, B. W. Segelke, H. I. Krupka, T. P. Lekin, J. Schafer, A. Zemla, D. Toppani, G. Snell and T. Earnest, *Acta Cryst. D*, 2002, **58**, 1514-1518.
2. W. Minor, M. Cymborowski and Z. Otwinowski, *Acta Physica Polonica A*, 2002, **101**, 613-619.
3. A. J. McCoy, R. W. Grosse-Kunstleve, L. C. Storoni and R. J. Read, *Acta Cryst. D*, 2005, **61**, 458-464.
4. L. C. Storoni, A. J. McCoy and R. J. Read, *Acta Cryst. D*, 2004, **60**, 432-438.
5. P. Emsley and K. Cowtan, *Acta Cryst. D*, 2004, **60**, 2126-2132.
6. G. N. Murshudov, A. A. Vagin, A. Lebedev, K. S. Wilson and E. J. Dodson, *Acta Cryst. D*, 1999, **55**, 247-255.
7. R. A. Laskowski, M. W. MacArthur, D. S. Moss and J. M. Thornton, *J. Appl. Cryst.*, 1993, **26**, 548-558.
8. I. W. Davis, L. W. Murray, J. S. Richardson and D. C. Richardson, *Nucleic Acids Res*, 2004, **32**, W615-W619.
9. H. Boutselakis, D. Dimitropoulos, J. Fillon, A. Golovin, K. Henrick, A. Hussain, J. Ionides, M. John, P. A. Keller, E. Krissinel, P. McNeil, A. Naim, R. Newman, T. Oldfield, J. Pineda, A. Rachedi, J. Copeland, A. Sitnov, S. Sobhany, A. Suarez-Uruena, J. Swaminathan, M. Tagari, J. Tate, S. Tromm, S. Velankar and W. Vranken, *Nucleic Acids Res*, 2003, **31**, 458-462.