Electronic Supplementary Material (ESI) for RSC Chemical Biology.

Supplementary Figures and Tables

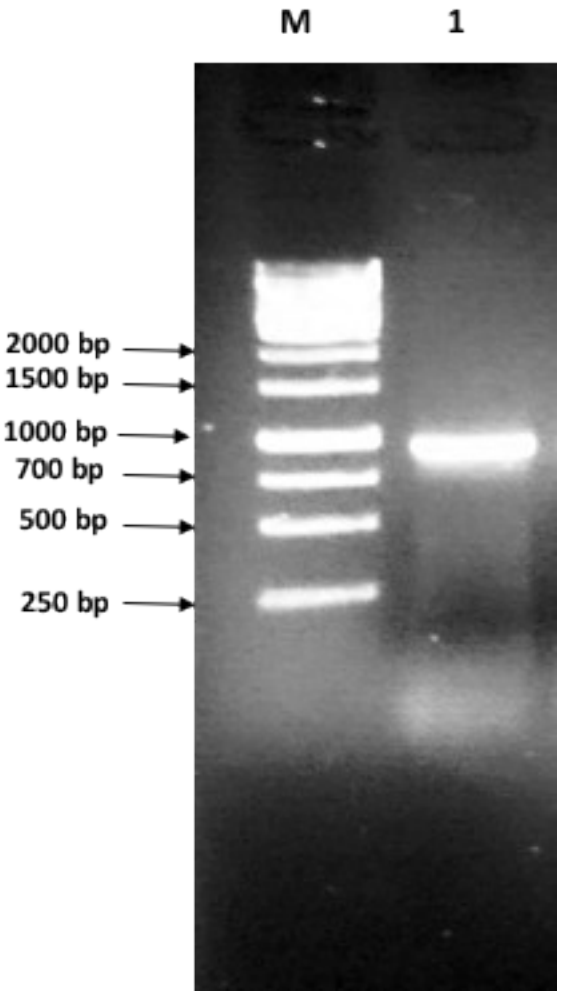


Figure S1. Agarose gel electrophoresis of the OCS nucleotide consensus region amplified from grass pea genomic DNA.
A consensus sequence region of ~1kb was identified using a multiple sequence alignment of *Ls*OCS genes. The region was PCR amplified from genomic, grass pea DNA using primers F393-OCS and R1366-OCS (see **Materials and Methods**). The PCR product was analyzed by electrophoresis on a 1.0% agarose gel. Lane M, molecular size marker 1kb (Thermo scientific©); lane 1, PCR product from *Ls*OCS (expected size of 938bp).

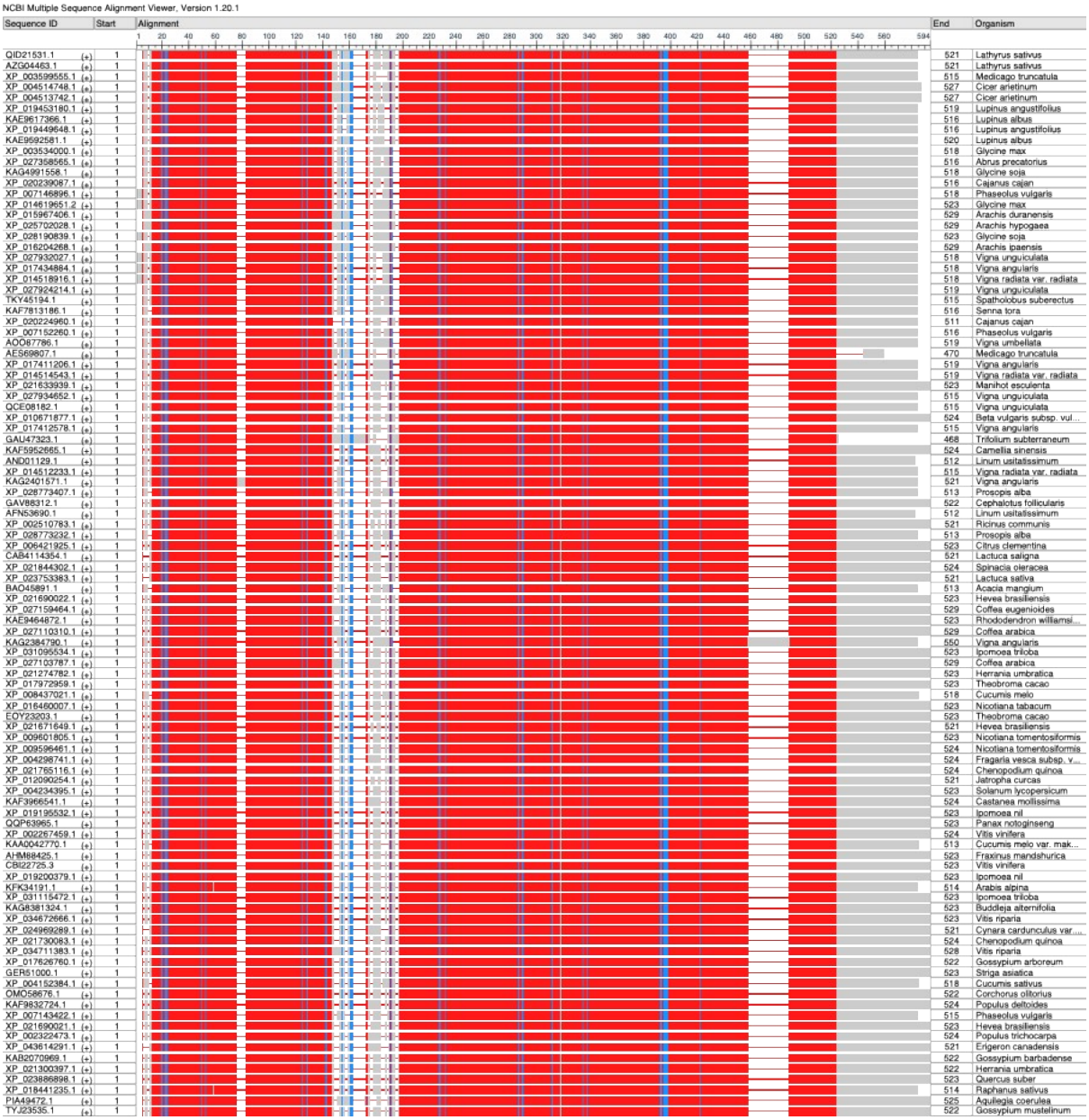


Figure S2. A graphical representation of a multiple protein sequence alignment of *Ls*OCS. A BlastP database search was performed using the protein sequence of *Ls*OCS at the NCBI website (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins) against the non-redundant (NR) database with default parameters. A multiple sequence alignment of 100 homologous proteins with sequence identity >76% is presented. Sequence regions are colored according to degree of conservation based on the residue's relative entropy threshold. Red color indicates highly conserved regions and blue indicates less conserved ones. The top sequence is that of *Ls*OCS.

S3a. His-SUMO-*Ls*OCS DNA sequence:

ATGAGCAAGCATCACCATCATTCAGGCCATCACCATACCGGACACCACCATCATTCAGGCAGTC
ATCACCATTCCGGATCTGCTGCGGGTGGCGAAGAAGATAAGAAACCGGCAGGTGGCGAAGGTGG
CGGTGCCCATATCAACCTGAAAGTGAAAGGTCAAGACGGCAACGAAGTCTTTTTCCGCATCAAA
CGTTCTACCCAGCTGAAAAAGCTGATGAACGCATACTGTGACCGTCAGTCTGTAGACATGACCG
CAATTGCTTTCCTGTTTGATGGTCGTCGCCTGCGTGCGGAACAGACCCCGGATGAACTGGAGAT
GGAAGATGGCGACGAAATCGACGCAATGCTTCATCAGACTGGTGGCGAAACCGCAACCACCCTC
ACCGGTTTACTCCAATCCGTCGCCAAAACATTCCCCTCTCGACGTGGCATCTCCCTCGCCGGAA
AATTCGACCTCACTCACTCTCATCTTAACGAATTAGTCGAATCTGCCGCAGATCATCTCATCTC
TGCCGGAATCAAACCAAACGACGTCGTCGCTCTTACTTTCCCCAACACCGTCGAGTATGTTATA
TTGTTTTTAGCTGTTATTCGAGTCCGAGCCACGGCGGCGCCTTTGAATGCAGCTTATACAGCTG
AAGAATTCGAGTTTTATCTATCTGACTCCGAATCCAAGCTTCTATTAACGCCTTTAGAAGGTAA
CAAGCCGGCGCAAGACGCGGCTTCAAAGCTCAGTATTCCTCTCGGCTCGGCTTCTCTCACGAAA
TCTGAAGAAGAAACCAAGCTTACAATCTCCCTGAAACATCCCGAGTCAGGTTTAAAATCTGACT
CAGTAAACTCGGTGGCCAAACTCATTAACGAACCATCCGACGTGGCACTTTTTCTTCACACATC
AGGTACCACGAGTCGTCCCAAGGGAGTTCCACTGACTCAACACAACCTGGTTTCGTCTGTCAAA
AACATCCAATCGGTTTACCAACTCACTGAATCAGATTCAACCGTGATCGTGCTTCCACTTTTTC
ACGTTCACGGGTTAATCGCTGGATTGCTGAGTTCACTGGGTTCCGGGGCTGCGGTGGTGTTACC
GGCGGCGGGGAGATTCTCGGCCTCAACCTTTTGGAAAGACATGATTCAATACAATGCGACGTGG
TACACAGCGGTACCTACCATACACCAGATCATACTAGATCGCCACCTAAATAACCCTGAACCGG
CTTACCCGAAACTCCGGTTTATTAGAAGCTGTAGTGCTTCGTTGGCACCGGTTATTCTAGGTCG
CTTAGAGGAATCGTTTGGGGCACCAGTTTTGGAGGCTTATGCTATGACTGAGGCTTCTCATTTA
ATGAGTTCGAATCCTTTGCCACAAAATGGGCCCCACAAAGCTGGATCGGTTGGGAAGCCCGTGG
GTCAAGAAATGGCTATACTTGATGAGTCGGGCCGGGTTTTGGAGGCTGGAGTGAATGGTGAAGT
TTGTATTAGAGGAGAAAATGTTACGAAAGGTTACAAGAATAATGAAGCAGCTAATACGGCAGCG
TTTTTGTTTGGTTGGTTTCATACGGGTGACATTGGTTACTTTGATTCTGATGGATATTTGCATC
TTGTGGGTCGGATTAAAGAGCTTATCAACAGAGGAGGAGAGAAAATATCACCAATAGAAGTGGA
TGCTGTTCTTCTATCTCATCCAGACGTAGCTCAGGCCGTTGCTTTTGGAATACCAGATCAGAAA
TATGGTGAAGAGATACATTGTGCAATCATACCAAGAGAAGGATCAAACATTGATGCTGAAGAGG
TGCTAACATTTTGCAAGAAGAATCTGGCATCTTTCAAAGTCCCCAAAAAGGTTTTCATTACTGA
TTCTCTGCCCAAGACGGCTACTGGCAAGATTTTGCGTCGTCTTGTAGCAGAACACTTTGTTTCT
AAAGTTTAA*

S3b. His-SUMO-*Ls*OCS protein sequence:

MSKHHHHSGHHHTGHHHHSGSHHHSGSAAGGEEDKKPAGGEGGGAHINLKVKGQDGNEVFFRIK
RSTQLKKLMNAYCDRQSVDMTAIAFLFDGRRLRAEQTPDELEMEDGDEIDAMLHQTGGETATTL
TGLLQSVAKTFPSRRGISLAGKFDLTHSHLNELVESAADHLISAGIKPNDVVALTFPNTVEYVI
LFLAVIRVRATAAPLNAAYTAEEFEFYLSDSESKLLLTPLEGNKPAQDAASKLSIPLGSASLTK
SEEETKLTISLKHPESGLKSDSVNSVAKLINEPSDVALFLHTSGTTSRPKGVPLTQHNLVSSVK
NIQSVYQLTESDSTVIVLPLFHVHGLIAGLLSSLGSGAAVVLPAAGRFSASTFWKDMIQYNATW
YTAVPTIHQIILDRHLNNPEPAYPKLRFIRSCSASLAPVILGRLEESFGAPVLEAYAMTEASHL
MSSNPLPQNGPHKAGSVGKPVGQEMAILDESGRVLEAGVNGEVCIRGENVTKGYKNNEAANTAA
FLFGWFHTGDIGYFDSDGYLHLVGRIKELINRGGEKISPIEVDAVLLSHPDVAQAVAFGIPDQK
YGEEIHCAIIPREGSNIDAEEVLTFCKKNLASFKVPKKVFITDSLPKTATGKILRRLVAEHFVS
KV*

Figure S3. Sequence of the cloned and expressed *Ls*OCS gene

a.The DNA sequence of *Ls*OCS (1563 bp) was N-terminally fused to a 14-His tag (highlighted in yellow) followed by a bd-SUMO linker (highlighted in green). **b.** The protein sequence of *Ls*OCS (521 aa), N-terminally fused to a 14-His tag (highlighted in yellow) and a bd-SUMO linker (highlighted in green)

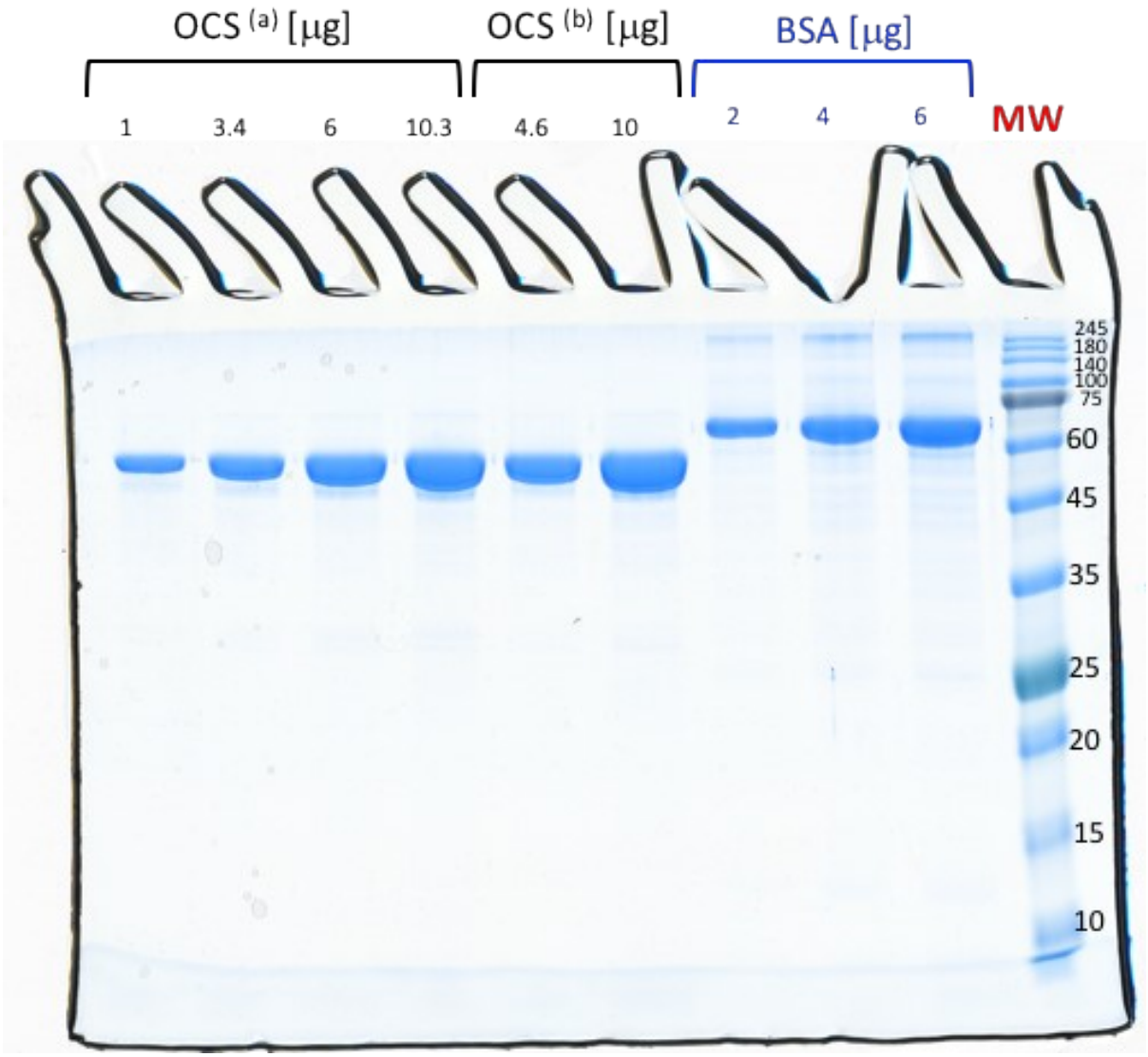
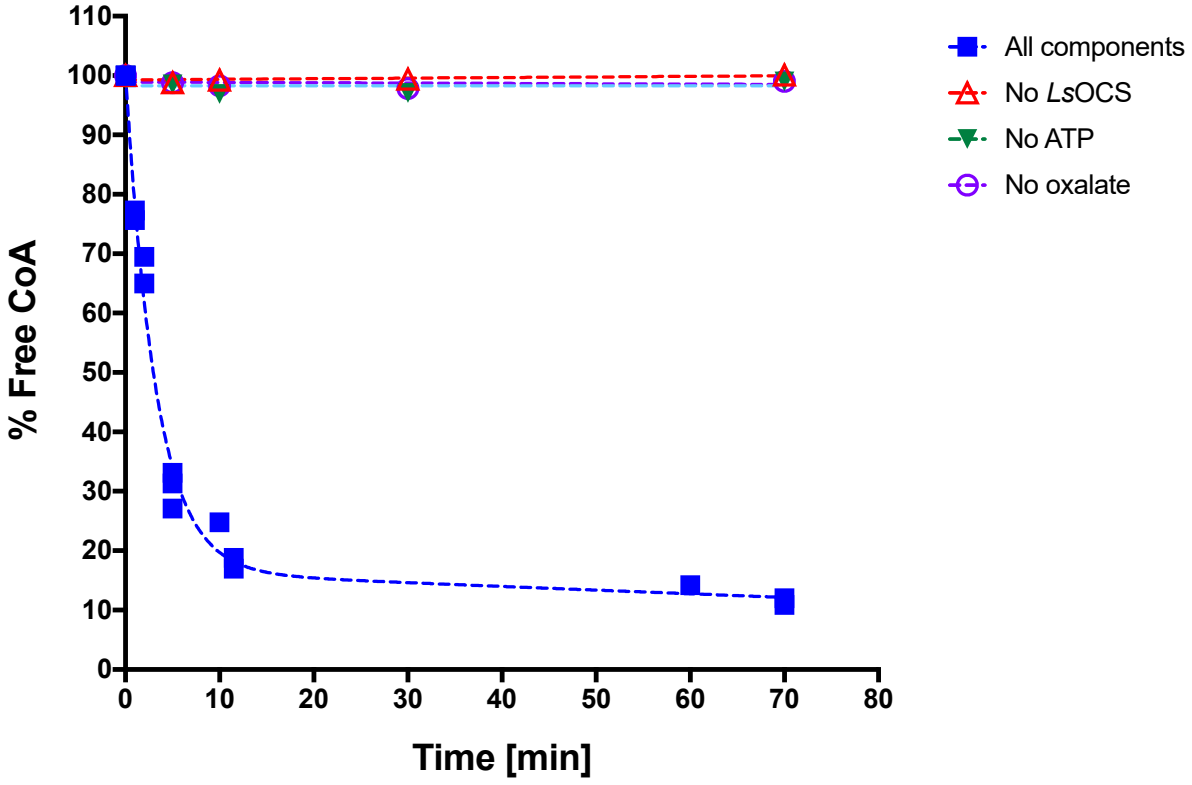


Figure S4. SDS-PAGE gel of purified *Ls*OCS.
Increasing amounts of purified *Ls*OCS protein were run on a 12% SDS-PAGE gel, and stained with Instant Blue™. The amounts, in micrograms, of *Ls*OCS protein loaded in each lane are noted above the lane (six left lanes). *Ls*OCS$^{(a)}$ and *Ls*OCS$^{(b)}$ denote two different batches of purified protein used. A bovine serum albumin (BSA) standard was run as control (three lanes on the right). MW – prestained molecular weight markers.

S5a.



S5b.

Best-fit values	
Y0	100.2
K	0.2875
Y0	95.91 to 104.4
K	0.2430 to 0.3398
R squared	0.9897

Figure S5. Time course of CoA ligation to oxalate catalyzed by *Ls*OCS.
a. A buffered solution (HEPES 125mM, pH 8) of $MgCl_2$ (2mM), ATP (10mM), CoA (1.5mM) and purified *Ls*OCS (0.25μM) was mixed with buffered NaOxalate (5mM, pH 8) and incubated at room temperature. At different time points, samples (100μl) were removed and mixed with a DTNB solution (100μl, 2mM). The absorption of the mixtures at 412nm was measured, and the concentration of free (i.e. unligated) CoA was calculated. The percent of free CoA in the ligation was plotted for each time point. **b.** The data of the reaction with all components was fit to an exponential decay curve $Y=Y_0 \bullet e^{(-kT)}$, in which Y= % of free CoA, Y_0 = % of free CoA at time zero, k = the decay constant [min^{-1}], T = time.

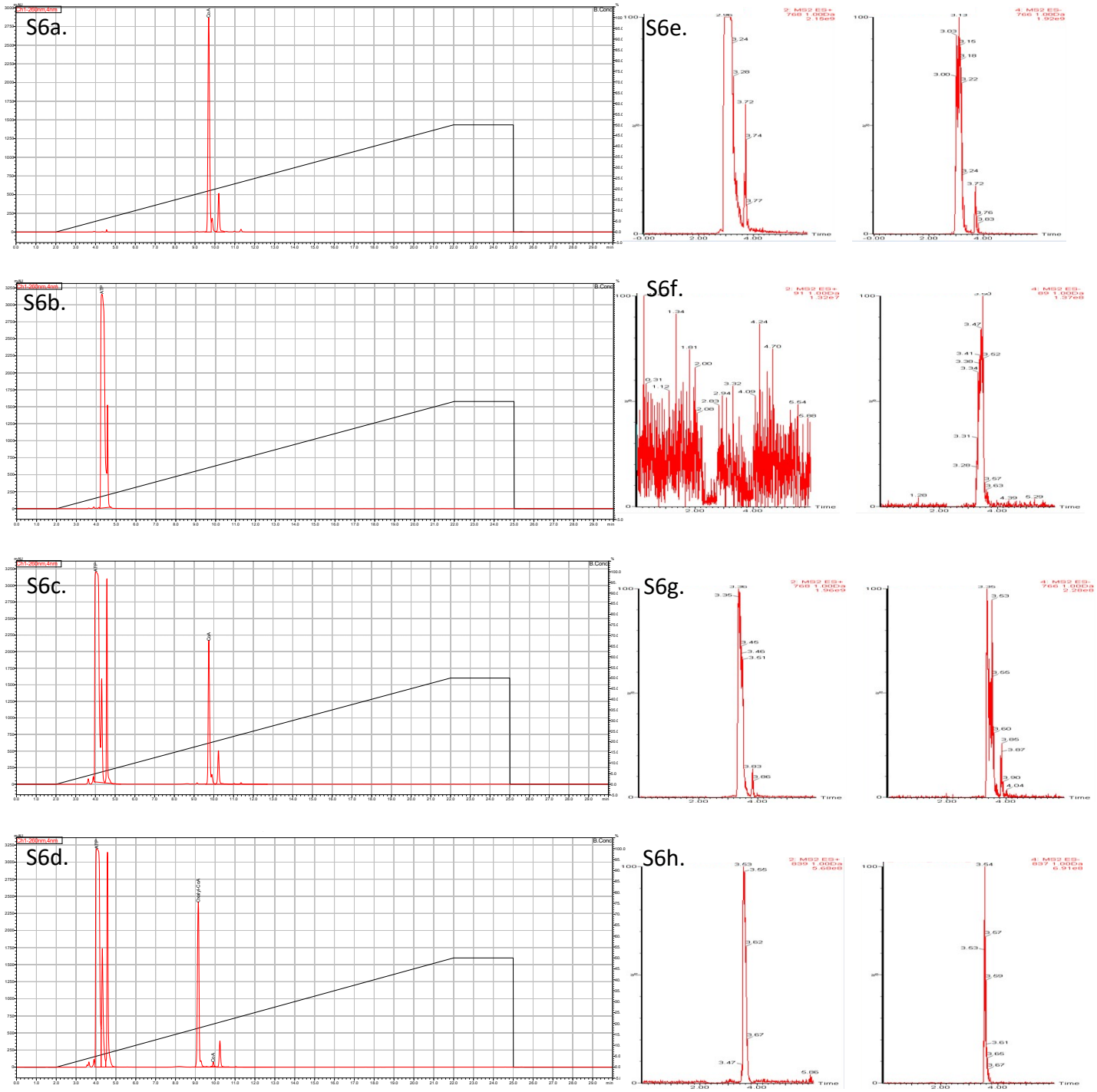


Figure S6. Analysis of substrates and products of the *in-vitro* reaction.
Samples of individual substrates, the *in-vitro* reaction mixture prior to enzyme addition and the reaction mixture following 1h of incubation with purified *Ls*OCS were separated on a C18 column, detected at 260nm and subjected to a mass spectrometry analysis (see **Materials and Methods**). Following incubation with the enzyme >95% of the free CoA were converted to oxalyl-CoA. HPLC-UVA (5a-d): **a.** CoA (21μg) **b.** ATP (24 μg) **c.** Reaction mixture (5μl) containing CoA (4mM), ATP (50mM), oxalate (40mM), MgCl

(4mM) and HEPES buffer (125mM) prior to incubation with *Ls*OCS **d.** Reaction mixture (5μl) following 1h of incubation at 37°C with purified *Ls*OCS (2μM). UPLC-MS (5e-h): **e.** CoA **f**. Oxalate. **g**. Reaction sample mix before incubation with *Ls*OCS **h**. Reaction sample mix after incubation with *Ls*OCS. Extracted ion chromatograms in positive (on left) and negative (on right) electrospray MS ionization are depicted for particular characteristic m/z values (noted at upper right corner of each panel).

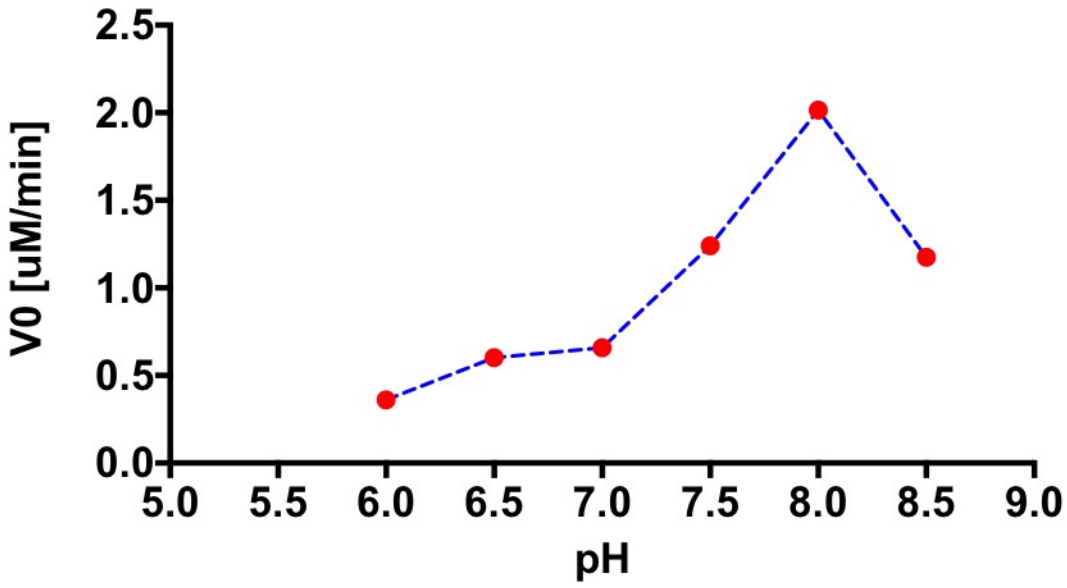
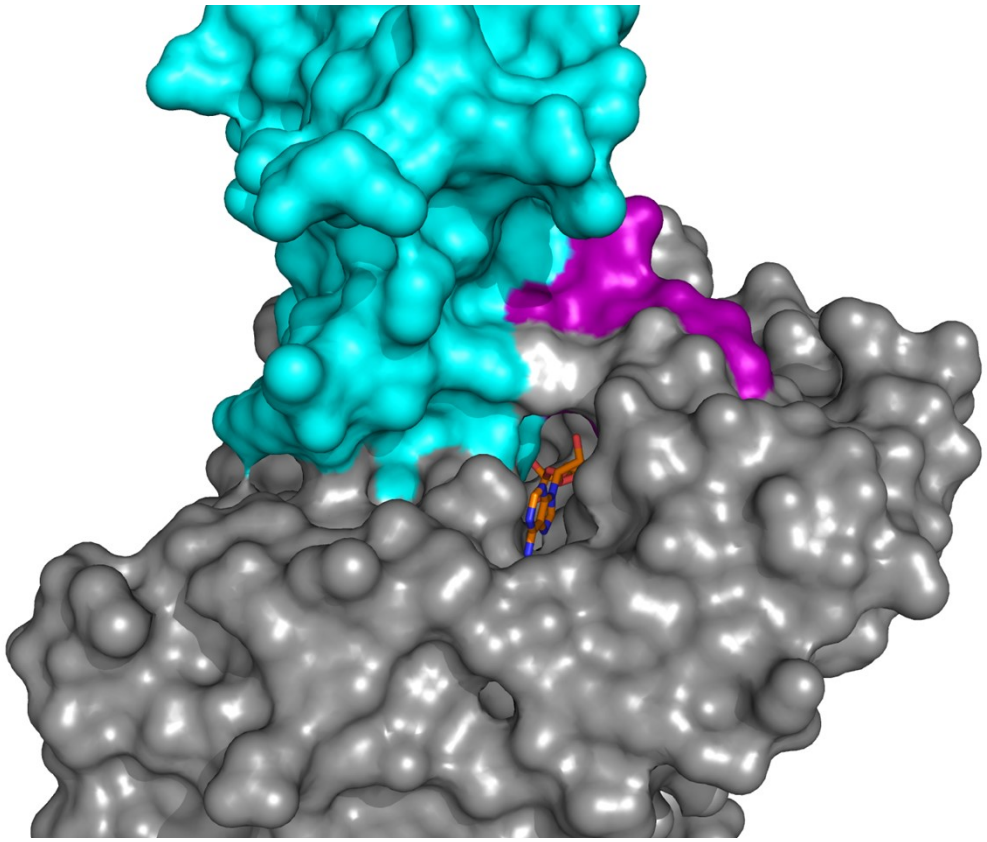
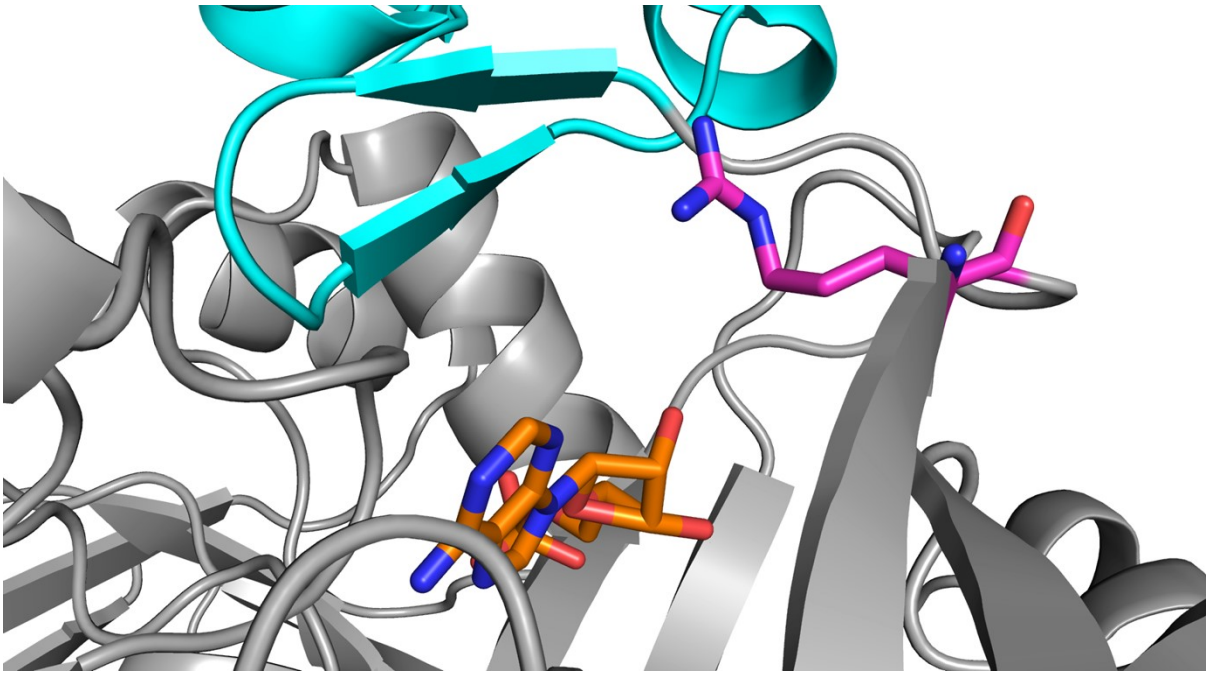


Figure S7. The activity of *Ls*OCS at different pHs. The reaction rates of *Ls*OCS (10μM) in the presence of oxalate (0.4mM), $MgCl_2$ (5mM), ATP (0.5mM) and CoA (1mM), were measured in Tris-HCl buffered solutions (50mM) at different pH values. Shown are the corresponding initial reaction rates (V0).

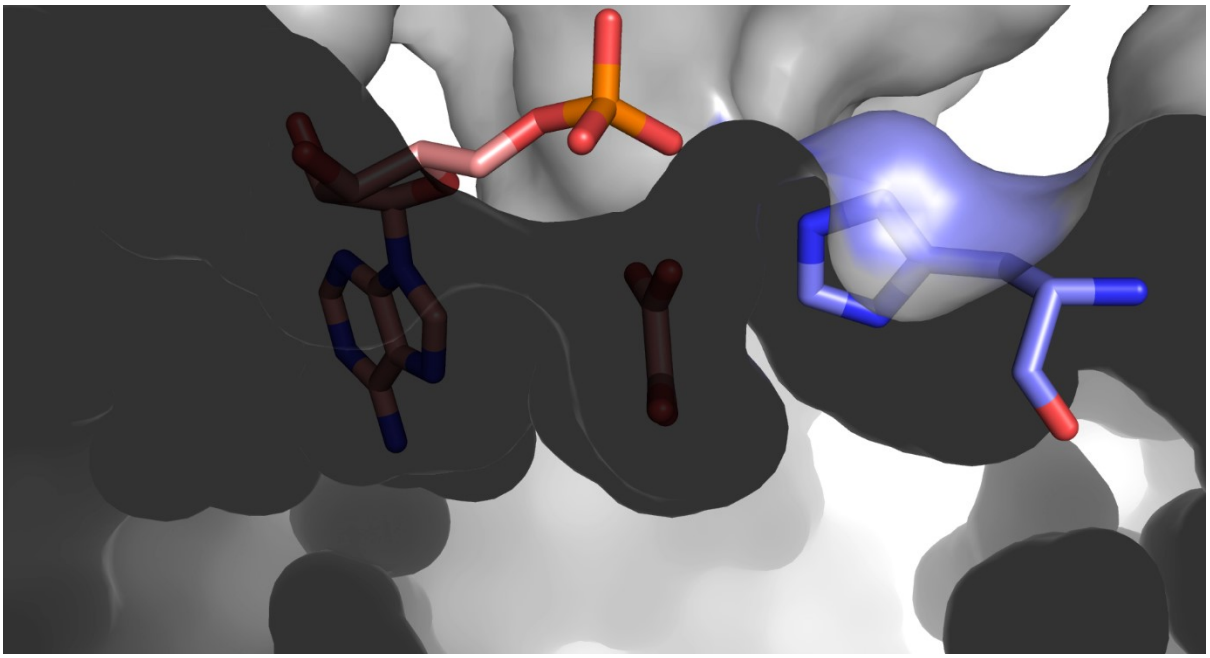


Supplementary Figure S8. The substrate binding cavity of *Ls*OCS.
*Ls*OCS crystal structure with the N- (residues 1-411) and C-terminal (residues 420-521) domains shown as grey and cyan surface plots, respectively. The hinge-loop region (^{417}IKEL^{420}) is shown in purple. The bound AMP is show as an orange, blue and red stick model, residing in a cavity formed by the N- and C-terminal regions.

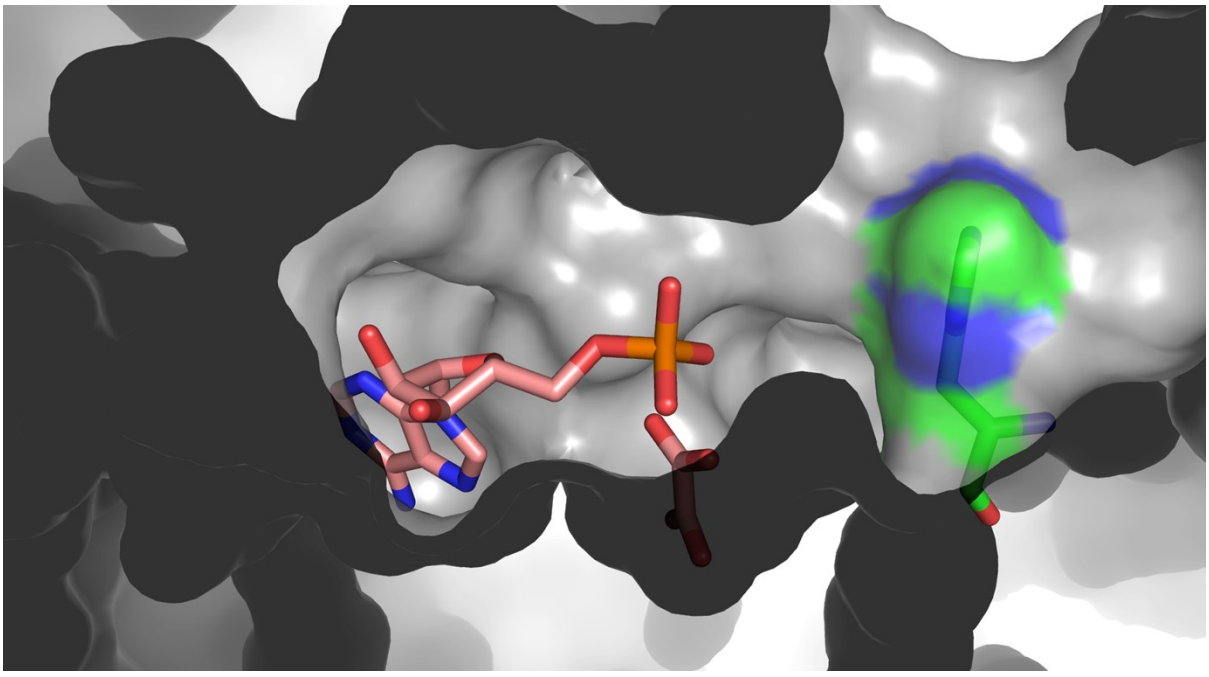


Supplementary Figure S9. The position of Arg46.
*Ls*OCS crystal structure with the N- (residues 1-411) and C-terminal (residues 420-521) domains shown as grey and cyan cartoon plots, respectively. The bound AMP is shown as an orange, blue and red stick model. Arg416 is shown as a magenta, blue and red stick model, pointing away from the AMP ribose oxygens.

S10a.



S10b.



Supplementary Figure S10. His221 blocking the pentathein tunnel of *Ls*OCS. The structure of *Ls*OCS was superimposed on that of *At*AAE3 (PDB ID: 5IE3). The AMP and oxalate ligands bound to *At*AAE3 are shown as pink, red and blue sticks. **a.** The imidazole ring of His221 of *Ls*OCS is oriented such that it blocks the entrance of the pentathein tunnel. **b.** The imidazole ring of the corresponding His214 residue of *At*AAE3,

which was crystalized in the adenylate-forming configuration, is rotated such that the entrance to the pentathein tunnel is open.