# Synthetic metabolism for *in vitro* acetone biosynthesis driven by ATP regeneration

by

Ekaterina Kozaeva, Manuel Nieto-Domínguez, Abril D. Hernández and Pablo I. Nikel

## Supplementary Data

**Table S1.** Oligonucleotides used in this work.

| Number | Name | Sequence (5'→3') | Application |
|---|---|---|---|
| 1 | pET28a-F | AGTGCTAGGAuCCGGCTGCTAACAAAGCC | USER cloning to build vector pET28a(+)-NtHis6x-TEV-Thl |
| 2 | pET28a-Nt-R | ATGGCCCuGAAAATAAAGATTCTCGCCGC | |
| 3 | thl-F | AGGGCCAuAAAGAAGTTGTAATAGCTAG | |
| 4 | thl-R | ATCCTAGCACuTTTCTAGCAATATTGCTG | |
| 5 | pET28a-F | ATGAGAuCCGGCTGCTAACAAAGCC | USER cloning to build vector pET28a(+)-NtHis6x-TEV-AtoDA |
| 2 | pET28a-Nt-R | ATGGCCCuGAAAATAAAGATTCTCGCCGC | |
| 6 | atoDA-F | AGGGCCAuAAAACAAAATTGATGACATTACAAGACGCCAC | |
| 7 | atoDA-R | ATCTCAuAAATCACCCCGTTGCGTAT | |
| 8 | pET28a-F | AGTAAGAuCCGGCTGCTAACAAAGCC | USER cloning to build vector pET28a(+)-NtHis6x-TEV-Adc |
| 2 | pET28a-Nt-R | ATGGCCCuGAAAATAAAGATTCTCGCCGC | |
| 9 | adc-Nt-F | AGGGCCAuTTAAAGGATGAAGTAATTAAACAAA | |
| 10 | adc-R | ATCTTACuTAAGATAATCATATATAACTTCAG | |
| 11 | pET28a-F | ATCTCTTCuGAGCACCACCACCACC | USER cloning to build vector pET28a(+)-CtHis6x-Adc |
| 12 | pET28a-Ct-R | AGAGCCCAuGGTATATCTCCTTCTTAAAG | |
| 13 | adc-Ct-F | ATGGGCTCuTTAAAGGATGAAGTAATTAAACAAA | |
| 14 | adc-R | AGAAGAGAuCTTAAGATAATCATATATAACTTCAGCTCT | |
| 15 | adc-QC-F | TTAAAGGATGAAGTAATTAAACAAATTAGCACG | Removing the NtHis6x-TEV site from vector pET28a(+)-NtHis6x-TEV-Adc |
| 16 | adc-QC-R | TCCTTTAACATGGTATATCTCCTTCTTAAAGTTAAACAAAATT | |
| 11 | pET28a-F | ATCTCTTCuGAGCACCACCACCACC | USER cloning to build vector pET28a(+)-NtHis6x-TEV-AjPPK2 |
| 2 | pET28a-Nt-R | ATGGCCCuGAAAATAAAGATTCTCGCCGCT | |
| 17 | ajppk2-F1 | AGGGCCAuGACACAGAGACAATTGCTTCAGC | |
| 18 | ajppk2-R1 | AATTGAGuTAAAATGTCTGGAATCGG | |
| 19 | ajppk2-F2 | ACTCAATuCGAAGTCCCCCAAGCCGA | |
| 20 | ajppk2-R2 | AGAAGAGAuTTAATCTGTGTCACGGTCGGCCT | |
| 11 | pET28a-F | ATCTCTTCuGAGCACCACCACCACC | USER cloning to build vector pET28a(+)-NtHis6x-TEV-SmPPK2 |
| 2 | pET28a-Nt-R | ATGGCCCuGAAAATAAAGATTCTCGCCGCT | |
| 21 | smppk2-F | AGGGCCAuGCGCTGGACGAAGCTCCTG | |
| 22 | smppk2-R | AGAAGAGAuTTAGCGCAGAAAACCTGGCCC | |

**Table S2.** Gene fragments used in this work.

| Description | Sequence (5'→3') |
|---|---|
| Sequence encoding the *thl* gene | **ATG**AAAGAAGTTGTAATAGCTAGTGCAGTAAGAACAGCGATTGGATCTTATGGAAAGT<br>CTCTTAAGGATGTACCAGCAGTAGATTTAGGAGCTACAGCTATAAAGGAAGCAGTTAA<br>AAAAGCAGGAATAAAACCAGAGGATGTTAATGAAGTCATTTTAGGAAATGTTCTTCAA<br>GCAGGTTTAGGACAGAATCCAGCAAGACAGGCATCTTTTAAAGCAGGATTACCAGTTG<br>AAATTCCAGCTATGACTATTAATAAGGTTTGTGGTTCAGGACTTAGAACAGTTAGCTT<br>AGCAGCACAAATTATAAAAGCAGGAGATGCTGACGTAATAATAGCAGGTGGTATGGAA<br>AATATGTCTAGAGCTCCTTACTTAGCGAATAACGCTAGATGGGGATATAGAATGGGAA<br>ACGCTAAATTTGTTGATGAAATGATCACTGACGGATTGTGGGATGCATTTAATGATTA<br>CCACATGGGAATAACAGCAGAAAACATAGCTGAGAGATGGAACATTTCAAGAGAAGAA<br>CAAGATGAGTTTGCTCTTGCATCACAAAAAAAAGCTGAAGAAGCTATAAAATCAGGTC<br>AATTTAAAGATGAAATAGTTCCTGTAGTAATTAAAGGCAGAAAGGGAGAAACTGTAGT<br>TGATACAGATGAGCACCCTAGATTTGGATCAACTATAGAAGGACTTGCAAAATTAAAA<br>CCTGCCTTCAAAAAAGATGGAACAGTTACAGCTGGTAATGCATCAGGATTAAATGACT<br>GTGCAGCAGTACTTGTAATCATGAGTGCAGAAAAAGCTAAAGAGCTTGGAGTAAAACC<br>ACTTGCTAAGATAGTTTCTTATGGTTCAGCAGGAGTTGACCCAGCAATAATGGGATAT<br>GGACCTTTCTATGCAACAAAAGCAGCTATTGAAAAAGCAGGTTGGACAGTTGATGAAT<br>TAGATTTAATAGAATCAAATGAAGCTTTTGCAGCTCAAAGTTTAGCAGTAGCAAAAGA<br>TTTAAAATTTGATATGAATAAAGTAAATGTAAATGGAGGAGCTATTGCCCTTGGTCAT<br>CCAATTGGAGCATCAGGTGCAAGAATACTCGTTACTCTTGTACACGCAATGCAAAAAA<br>GAGATGCAAAAAAAGGCTTAGCAACTTTATGTATAGGTGGCGGACAAGGAACAGCAAT<br>ATTGCTAGAAAAGTGC**TAG** |
| Sequence encoding the *atoDA* genes | **ATG**AAAACAAAATTGATGACATTACAAGACGCCACCGGCTTCTTTCGTGACGGCATGA<br>CCATCATGGTGGGCGGATTTATGGGGATTGGCACTCCATCCCGCCTGGTTGAAGCATT<br>ACTGGAATCTGGTGTTCGCGACCTGACATTGATAGCCAATGATACCGCGTTTGTTGAT<br>ACCGGCATCGGTCCGCTCATCGTCAATGGTCGAGTCCGCAAAGTGATTGCTTCACATA<br>TCGGCACCAACCCGGAAACAGGTCGGCGCATGATATCTGGTGAGATGGACGTCGTTCT<br>GGTGCCGCAAGGTACGCTAATCGAGCAAATTCGCTGTGGTGGAGCTGGACTTGGTGGT<br>TTTCTCACCCCAACGGGTGTCGGCACCGTCGTAGAGGAAGGCAAACAGACACTGACAC<br>TCGACGGTAAAACCTGGCTGCTCGAACGCCCACTGCGCGCCGACCTGGCGCTAATTCG<br>CGCTCATCGTTGCGACACACTTGGCAACCTGACCTATCAACTTAGCGCCCGCAACTTT<br>AACCCCCTGATAGCCCTTGCGGCTGATATCACGCTGGTAGAGCCAGATGAACTGGTCG<br>AAACCGGCGAGCTGCAACCTGACCATATTGTCACCCCTGGTGCCGTTATCGACCACAT<br>CATCGTTTCACAGGAGAGCAAA**TAATG**GATGCGAAACAACGTATTGCGCGCCGTGTGG<br>CGCAAGAGCTTCGTGATGGTGACATCGTTAACTTAGGGATCGGTTTACCCACAATGGT<br>CGCCAATTATTTACCGGAGGGTATTCATATCACTCTGCAATCGGAAAACGGCTTCCTC<br>GGTTTAGGCCCGGTCACGACAGCGCATCCAGATCTGGTGAACGCTGGCGGGCAACCGT<br>GCGGTGTTTTACCCGGTGCAGCCATGTTTGATAGCGCCATGTCATTTGCGCTAATCCG<br>TGGCGGTCATATTGATGCCTGCGTGCTCGGCGGTTTGCAAGTAGACGAAGAAGCAAAC<br>CTCGCGAACTGGGTAGTGCCTGGGAAAATGGTGCCCGGTATGGGTGGCGCGATGGATC<br>TGGTGACCGGGTCGCGCAAAGTGATCATCGCCATGGAACATTGCGCCAAAGATGGTTC<br>AGCAAAAATTTTGCGCCGCTGCACCATGCCACTCACTGCGCAACATGCGGTGCATATG<br>CTGGTTACTGAACTGGCTGTCTTTCGTTTTATTGACGGCAAAATGTGGCTCACCGAAA<br>TTGCCGACGGGTGTGATTTAGCCACCGTGCGTGCCAAAACAGAAGCTCGGTTTGAAGT<br>CGCCGCCGATCTGAATACGCAACGGGGTGATTTA**TGA** |
| Sequence encoding the *adc* gene | **ATG**TTAAAGGATGAAGTAATTAAACAAATTAGCACGCCATTAACTTCGCCTGCATTTC<br>CTAGAGGACCCTATAAATTTCATAATCGTGAGTATTTTAACATTGTATATCGTACAGA<br>TATGGATGCACTTCGTAAAGTTGTGCCAGAGCCTTTAGAAATTGATGAGCCCTTAGTC<br>AGGTTTGAAATTATGGCAATGCATGATACGAGTGGACTTGGTTGTTATACAGAAAGCG<br>GACAGGCTATTCCCGTAAGCTTTAATGGAGTTAAGGGAGATTATCTTCATATGATGTA<br>TTTAGATAATGAGCCTGCAATTGCAGTAGGAAGGGAATTAAGTGCATATCCTAAAAAG<br>CTCGGGTATCCAAAGCTTTTTGTGGATTCAGATACTTTAGTAGGAACTTTAGACTATG<br>GAAAACTTAGAGTTGCGACAGCTACAATGGGGTACAAACATAAAGCCTTAGATGCTAA<br>TGAAGCAAAGGATCAAATTTGTCGCCCTAATTATATGTTGAAAATAATACCCAATTAT<br>GATGGAAGCCCTAGAATATGTGAGCTTATAAATGCGAAAATCACAGATGTTACCGTAC<br>ATGAAGCTTGGACAGGACCAACTCGACTGCAGTTATTTGATCACGCTATGGCGCCACT<br>TAATGATTTGCCAGTAAAAGAGATTGTTTCTAGCTCTCACATTCTTGCAGATATAATA<br>TTGCCTAGAGCTGAAGTTATATATGATTATCTTAAG**TAA** |

| | |
|---|---|
| Sequence encoding the *ajppk2* gene | **ATG**GACACAGAGACAATTGCTTCAGCCGTACTGAACGAGGAACAACTTAGCCTTGATC<br>TGATTGAGGCCCAGTACGCATTAATGAACACTCGCGACCAGAGCAACGCTAAGAGTCT<br>GGTGATTTTAGTATCAGGGATCGAGCTTGCGGGCAAAGGCGAAGCAGTTAAGCAATTG<br>CGCGAATGGGTCGATCCACGTTTTCTGTACGTTAAGGCGGATCCACCGCACCTTTTCA<br>ATTTAAAGCAACCTTTTTGGCAACCCTATACTCGTTTTGTTCCGGCCGAGGGCCAAAT<br>TATGGTATGGTTCGGCAACTGGTATGGAGATTTACTTGCCACAGCAATGCATGCCAGT<br>AAACCCCTTGACGATACTCTGTTTGATGAGTATGTAAGCAACATGCGTGCGTTCGAAC<br>AAGACCTTAAAAATAACAATGTCGATGTTCTTAAGGTGTGGTTCGATTTATCATGGAA<br>ATCCCTGCAGAAGCGCTTGGACGATATGGACCCTTCAGAAGTACACTGGCACAAGCTG<br>CATGGCCTTGACTGGCGTAATAAGAAGCAGTACGATACCCTGCAGAAACTTCGTACCC<br>GCTTTACAGACGATTGGCAAATTATCGATGGAGAAGACGAGGATTTGCGCAACCATAA<br>CTTTGCGCAAGCGATCTTGACTGCCCTGCGCCACTGTCCCGAACATGAGAAGAAAGCT<br>GCCCTGAAATGGCAACAAGCCCCGATTCCAGACATTTTAACTCAATTCGAAGTCCCCC<br>AAGCCGAGGACGCCAACTATAAGTCGGAATTGAAGAAACTTACCAAGCAGGTAGCAGA<br>CGCGATGCGCTGCGACGACCGCAAAGTGGTGATCGCCTTTGAGGGAATGGACGCAGCT<br>GGCAAAGGCGGCGCGATCAAACGCATTGTGAAGAAGCTGGATCCCCGTGAATATGAAA<br>TCCACACGATTGCAGCGCCAGAAAAATATGAATTGCGCCGCCCCTACCTGTGGCGCTT<br>TTGGAGTAAACTTCAGTCAGACGATATCACGATTTTTGATCGCACATGGTACGGTCGC<br>GTATTAGTTGAGCGTGTAGAAGGATTTGCCACAGAGGTCGAATGGCAACGCGCATACG<br>CTGAAATCAACCGCTTTGAAAAAAACCTGAGTAGTTCCCAAACTGTTCTGATTAAGTT<br>TTGGTTGGCTATTGATAAAGACGAACAGGCTGCCCGTTTCAAAGCCCGCGAATCGACC<br>CCACATAAGCGCTTTAAGATCACAGAGGAGGACTGGCGTAACCGCGACAAATGGGATG<br>ACTATTTAAAAGCTGCCGCTGATATGTTCGCCCATACCGACACAAGCTACGCGCCTTG<br>GTACATCATTTCCACGAATGATAAGCAGCAGGCCCGCATTGAAGTTCTGCGCGCCATC<br>CTTAAGCAGCTTAAGGCCGACCGTGACACAGAT**TAA** |
| Sequence encoding the *smppk2* gene | **ATG**GCGCTGGACGAAGCTCCTGCTGAAGCTCGTCCCGGATCCCGCGCAGTAGAGTTAG<br>AGATTGACGGGCGCTCTCGCATCTTTGATATTGACGACCCTGACCTTCCGAAGTGGAT<br>TGACGAAGAGGCTTTCCGCTCCGATGACTATCCTTACAAAAAAAAATTAGATCGCGAG<br>GAGTACGAAGAAACTTTAACCAAGTTGCAAATTGAACTTGTGAAGGTCCAGTTTTGGA<br>TGCAAGCTACCGGCAAACGTGTGATGGCCGTGTTTGAAGGACGTGACGCAGCAGGGAA<br>AGGAGGAGCAATTCATGCCACTACGGCAAACATGAATCCGCGTAGCGCACGTGTCGTG<br>GCCCTTACCAAACCTACTGAGACAGAGCGCGGACAGTGGTACTTCCAGCGTTATGTCG<br>CAACTTTTCCTACGGCAGGAGAGTTTGTATTGTTTGACCGTTCATGGTACAACCGCGC<br>CGGAGTAGAGCCGGTTATGGGGTTCTGTACCCCTGACCAGTATGAGCAGTTCTTAAAG<br>GAGGCACCGCGCTTCGAGGAGATGATTGCAAACGAAGGGATTCACCTGTTCAAGTTTT<br>GGATTAATATCGGACGCGAAATGCAATTAAAACGCTTCCATGACCGCCGTCACGATCC<br>CCTGAAGATTTGGAAATTGTCTCCAATGGACATTGCGGCATTGAGTAAGTGGGACGAT<br>TATACTGGGAAACGCGATCGTATGTTAAAAGAGACCCATACGGAACATGGTCCATGGG<br>CTGTAATTCGTGGGAACGATAAGCGCCGTAGTCGTATCAATGTGATTCGCCATATGCT<br>GACCAAACTTGATTATGATGGAAAGGACGAAGCCGCGATTGGTGAGGTTGACGAGAAG<br>ATTCTGGGTTCGGGGCCAGGTTTTCTGCGC**TAA** |

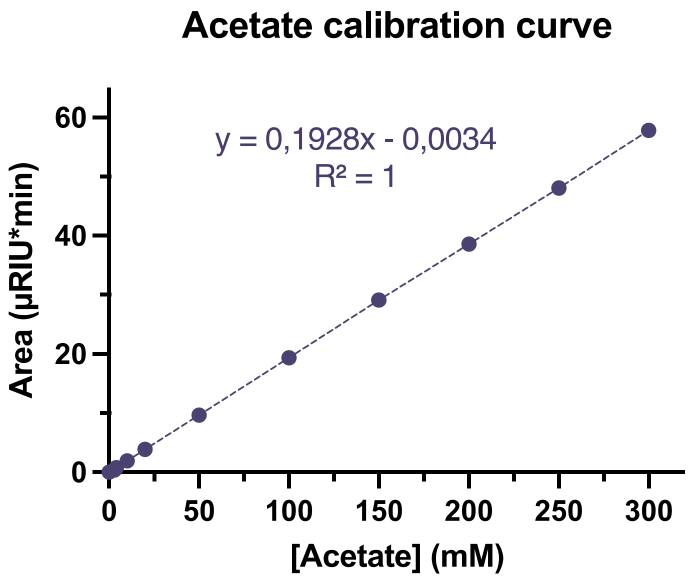
**Table S3.** Protein concetrations obtained with the culture conditions used in this work.

| Protein | **Protein concentration** (mg $L^{-1}$) | Culture medium |
|---|---|---|
| Thl | 16.8 | 2×YT |
| AtoDA | 40.1 | 2×YT |
| *Aj*PPK2 | 25 | 2×YT |
| *Sm*PPK2 | 42.4 | 2×YT |
| Adc (1) | 23.8 | 2×YT |
| Adc (2) | 112 | Auto-induction |

Calculations were performed based on the amount of the recovered pure protein(s). Theoretical extinction coefficients for $A_{280}$ of the enzymes listed above were as follows: Thl, 33,920 $M^{-1}$ $cm^{-1}$; AtoDA, 20,970 $M^{-1}$ $cm^{-1}$; *Aj*PPK2, 121,350 $M^{-1}$ $cm^{-1}$; and *Sm*PPK2, 57,410 $M^{-1}$ $cm^{-1}$.

## Supplementary Figures

**Figure S1. Example of the calculations used to determine acetate concentrations *via* HPLC analysis.**

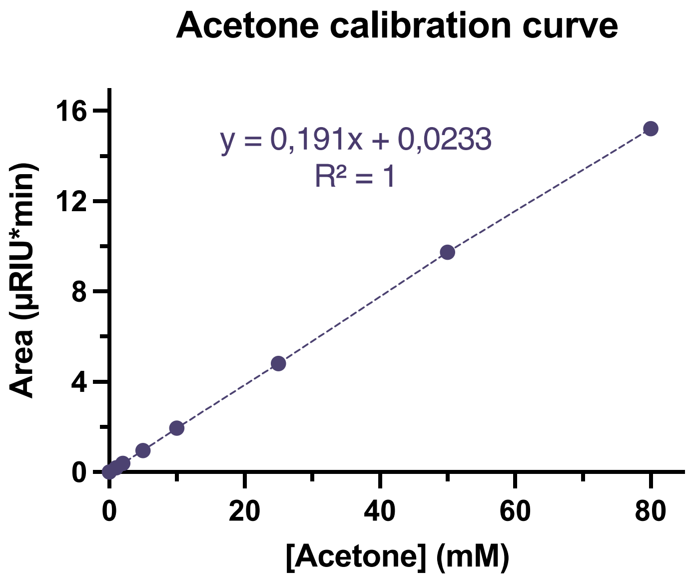



| **Sample** | **Retention time** (min) | **Area** (µRIU min$^{-1}$) | **Average of technical replicates** | **SD** | **[Acetate]** (mM) |
|---|---|---|---|---|---|
| 25 mM Ac, 10 mM ATP + Path 5% (w/v) TriClAc (1) | 15.515 | 1.602 | 1.6083 | 0.0085 | 8.568 |
| 25 mM Ac, 10 mM ATP + Path 5% (w/v) TriClAc (1) | 15.533 | 1.614 | | | |
| 25 mM Ac, 10 mM ATP + Path 5% (w/v) TriClAc (2) | 15.513 | 1.643 | 1.6411 | 0.0036 | 8.742 |
| 25 mM Ac, 10 mM ATP + Path 5% (w/v) TriClAc (2) | 15.535 | 1.638 | | | |
| 25 mM Ac, 10 mM ATP + Path 5% (w/v) TriClAc (3) | 15.515 | 1.613 | 1.6168 | 0.0043 | 8.613 |
| 25 mM Ac, 10 mM ATP + Path 5% (w/v) TriClAc (3) | 15.540 | 1.619 | | | |

Ac – Acetate
ATP – Adenosine triphosphate
Path – Enzymatic mixture, containing Acs, Thl, AtoDA, and Adc
TriClAc – Trichloroacetic acid

Each experiment was executed in 3 independent biological replicates and 2 technical duplicates.

**Figure S2. Example of the calculations used to determine acetone concentrations *via* HPLC analysis.**

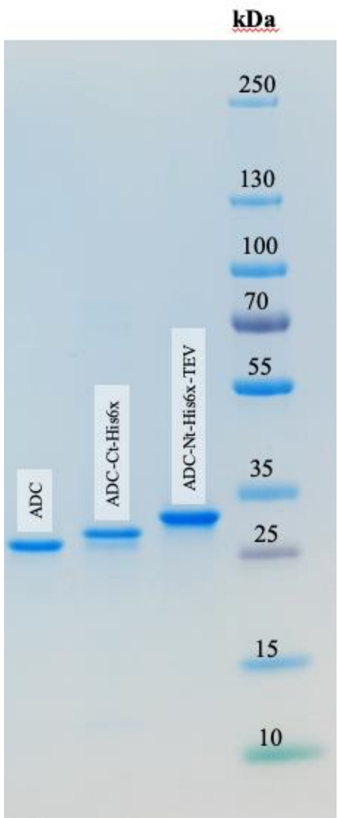



| Sample | Retention time (min) | Area (µRIU min⁻¹) | Average of technical replicates | SD | [Acetate] (mM) |
|---|---|---|---|---|---|
| 25 mM Ac, 10 mM ATP + Path 5% (w/v) TriClAc (1) | 22.800 | 1.445 | 1.4521 | 0.0101 | 7.668 |
| 25 mM Ac, 10 mM ATP + Path 5% (w/v) TriClAc (1) | 22.758 | 1.4593 | | | |
| 25 mM Ac, 10 mM ATP + Path 5% (w/v) TriClAc (2) | 22.795 | 1.4864 | 1.4964 | 0.0142 | 7.906 |
| 25 mM Ac, 10 mM ATP + Path 5% (w/v) TriClAc (2) | 22.758 | 1.5065 | | | |
| 25 mM Ac, 10 mM ATP + Path 5% (w/v) TriClAc (3) | 22.795 | 1.4979 | 1.4905 | 0.0104 | 7.874 |
| 25 mM Ac, 10 mM ATP + Path 5% (w/v) TriClAc (3) | 22.752 | 1.4831 | | | |

Ac – Acetate
ATP – Adenosine triphosphate
Path – Enzymatic mixture, containing Acs, Thl, AtoDA, and Adc
TriClAc – Trichloroacetic acid

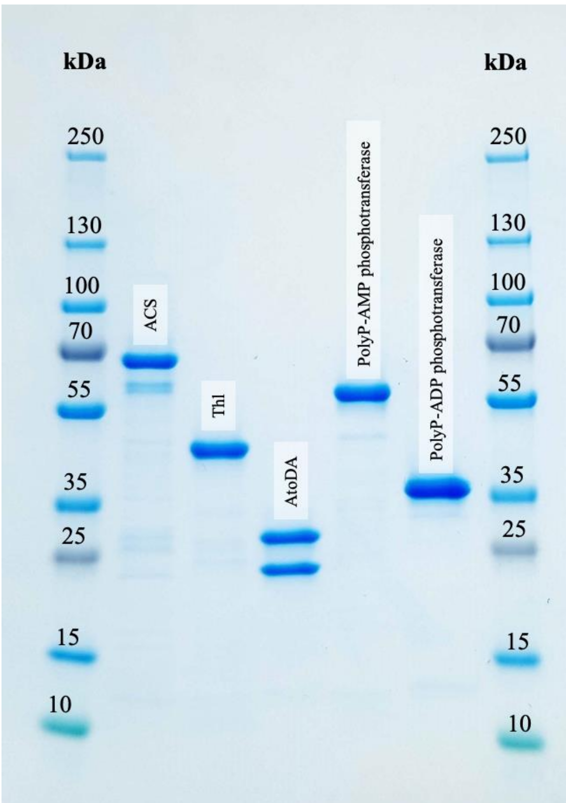
Each experiment was executed in 3 independent biological replicates and 2 technical duplicates.

**Figure S3. Evaluation of the protein purity through SDS-PAGE analysis.**




1 µg of protein was loaded in all of the lanes

- Adc ~28 kDa
- Adc_CtHis6× ~29 kDa
- Adc_NtHis6×-TEV ~30 kDa




1 µg of protein was loaded in all of the lanes, except for AtoDA (2 µg of protein)

- PolyP-AMP-transf_NtHis6×-TEV ~58 kDa
- Polyp-ADP-transf_NtHis6×-TEV ~37 kDa
- Acs ~66 kDa
- Thl_NtHis6×-TEV ~44 kDa
- AtoDA (heterodimer)
  - AtoD_NtHis6×-TEV ~26 kDa
  - AtoA ~23 kDa