Supplementary Information

Expression and specificity of a chitin deacetylase from the nematophagous fungus *Pochonia chlamydosporia* potentially involved in pathogenicity

Almudena Aranda-Martinez[1], Laia Grifoll-Romero[2], Hugo Aragunde[2], Enea Sancho-Vaello[2], Xevi Biarnés[2], Luis Vicente Lopez-Llorca[1], and Antoni Planas[2]*

[1]Laboratory of Plant Pathology, Department of Marine Sciences and Applied Biology, Multidisciplinary Institute for Environmental Studies Ramón Margalef, University of Alicante, PO box 99, 03080 Alicante, Spain. [2]Laboratory of Biochemistry, Institut Químic de Sarrià, Universitat Ramon Llull, Via Augusta 390, 08017, Barcelona, Spain.

Contents:

Figure S1. (A) Pc_2566 nucleotide sequence and predicted protein sequence. Signal peptide is shown in italics. Double underline shows the catalytic domain and single underline indicates the position of the CBM18 predicted modules. Circle and square boxes point putative O-glycosylation and N-linked glycosylation sites, respectively. Cysteine residues are highlighted with (~) symbols. The position of introns is indicated by arrows in the nucleotide sequence (see Fig. S2 for full genomic sequence) while (*) symbol indicates the stop codon. (B) *Pc*CDA protein domains. Modular structure of *Pc*CDA protein showing both CBM18 domains flanking the CE4 catalytic domain and signal peptide in the N-terminal end, and structure of expressed protein purified by affinity chromatography containing CE4 domain and Strep Tag peptide.

A)

```
1    ATG GCT CTG AGA ATA TCC CTC CTA TTT GCC GTG TGT GCT CTG GCT ACC GAC GGT TCC CGT
1    M   A   L   R   I   S   L   L   F   A   V   C   A   L   A   T   D   G   S   R

61   TTC ATT CGG GCA GCA GAG GCC AAT GAG ATG TCC ATG TGC GGT CCC GGT AAA GGA TCA TGC
21   F   I   R   A   A   E   A   N   E   M   S   M   C   G   P   G   K   G   S   C

121  GAA ATG GGC TCA TGT TGT TCA GAG ATG GGG TAT TGC GGC ACT ACG GAA CAG TAT TGC TCA
41   E   M   G   S   C   C   S   E   M   G   Y   C   G   T   T   E   Q   Y   C   S

181  GGC TCT CAG TGC CAG TTG GAC TAC AGC CAT ACT TGC CAC ACG CTG GTT CCC CCT CGT GGG
61   G   S   Q   C   Q   L   D   Y   S   H   T   C   H   (T)  L   V   P   P   R   G

241  AAG GAC ACT TCA GGA ATA ACC CGT CGA CAA GTC GGA AAC GTC CCT TAT GGA CCA ATG ATT
81   K   D   T   (S)  G   I   T   R   R   Q   V   G   N   V   P   Y   G   P   M   I

301  ACA GCT TGC AAA CAG CCT GGC ATG GTA GCC TTG ACA TTC GAC GAC GGT CCG TAC ATT TAC
101  T   A   C   K   Q   P   G   M   V   A   L   T   F   D   D   G   P   Y   I   Y

361  ACT ACC GAA ATC CTG AAC CTG CTT GAC CGG CTC AAG GTC AAG GCT ACT TTC TTC ATC ACG
121  T   T   E   I   L   N   L   L   D   R   L   K   V   K   A   T   F   F   I   T

421  GGT GAC AAC CGA GTC AAG GGA CAT ATT GAC GAT CCG GCG ACT CGC TGG CCA AGC ATT ATA
141  G   D   N   R   V   K   G   H   I   D   D   P   A   T   R   W   P   S   I   I

481  CGT CGC ATG TAC GAC GCG GGA CAT CAA ATT GGG AGC CAC ACC TGG ACG CAT CGT GAT CTG
161  R   R   M   Y   D   A   G   H   Q   I   G   S   H   T   W   T   H   R   D   L

541  AAC CAT ATC AAC GAG ACG GTG CGG CGA GCT GAA ATC ATT CAC AAT GAG ATG GCC ATA CGG
181  N   H   I   [N] E   T   V   R   R   A   E   I   I   H   N   E   M   A   I   R

601  AAC ATT CTT GGT TGG ATC CCC ACG TAC ATC CGG CCA CCT TTC TTG GAG TGT TCC GCC AGA
201  N   I   L   G   W   I   P   T   Y   I   R   P   P   F   L   E   C   S   A   R

661  TCT GGG TGT GAG AAG ACC ATG AGA GAC ATG GTG TAT CAC TCC ATC TCC ACC AAC TTG GAC
221  S   G   C   E   K   T   M   R   D   M   V   Y   H   S   I   S   T   N   L   D

721  ACC AAA GAC TAC ATG TAT GAC GAC CCA ACC CTC ATT CAA AGG GCA AAG GAC CTC TAC TCT
241  T   K   D   Y   M   Y   D   D   P   T   L   I   Q   R   A   K   D   L   Y   S

781  AAT GGA CTG TCA ACC AAT CCA CGG CAG AAC TCA TAC ATT GTC CTT GCA CAC GAT GTT CAT
261  N   G   L   S   T   N   P   R   Q   N   S   Y   I   V   L   A   H   D   V   H

841  GAG CAA ACC GTG TAT GAA CTG ACT CCA TAC ATG GTA AAG GTT GCT AGA GAG AGA GGC TAC
281  E   Q   T   V   Y   E   L   T   P   Y   M   V   K   V   A   R   E   R   G   Y

901  CAG CTT GTC ACG GTA GGT GAA TGC CTA GGA GAC CCC AGA GGT AAC TGG TAT AGA TCA GCG
301  Q   L   V   T   V   G   E   C   L   G   D   P   R   G   N   W   Y   R   S   A

961  GAA GAC TCG CCA AAA TCC ACC AGA AGC GAC AAA AAG AAA GAC TCA TGT CCC GGT CCA AAA
321  E   D   S   P   K   S   T   R   S   D   K   K   K   D   S   C   P   G   P   K

1021 AAG CCA ATC GTC ACG GAT ATC CCG CCC CCA AGG AAT ACG ACG AGC AAG ATA TCT CCC AAC
341  K   P   I   V   (T)  D   I   P   P   P   R   N   (T)  T   (S)  K   I   (S)  P   N

1081 CAA CGC TGC GGA GGA AGT ACA GGA TAC ATC TGT CCA GGC TCC GGG TTT GGA GAC TGC TGC
361  Q   R   C   G   G   S   T   G   Y   I   C   P   G   S   G   F   G   D   C   C

1141 TCT CAC TAT GGA TAT TGT GGA TCA ACT CCC GAG TTC TGT GGC ACA GGG TGT GAC AAT GAC
381  S   H   Y   G   Y   C   G   S   T   P   E   F   C   G   T   G   C   D   N   D

1201 TTT GGG GAC TGC AAC CCA AAT CCC CCA GGC ATC TGG GAC ACA ACG AAT GGC TTA TGT GGC
401  F   G   D   C   N   P   N   P   P   G   I   W   D   T   T   N   G   L   C   G

1261 GCT CGG TTC TCT GCG ACA TGC CTT AAT TTT GCA GGA AAG GCA TGC TGC TCC AAG TAT GGA
421  A   R   F   S   A   T   C   L   N   F   A   G   K   A   C   C   S   K   Y   G

1321 TAT TGG TAT GTA GTG CAT TCT GTG TGG TTA CTT TAT CGG TTG GAC TGA CGT GGT CGG CAG
441  Y   W   Y   V   V   H   S   V   W   L   L   Y   R   L   D   *
```
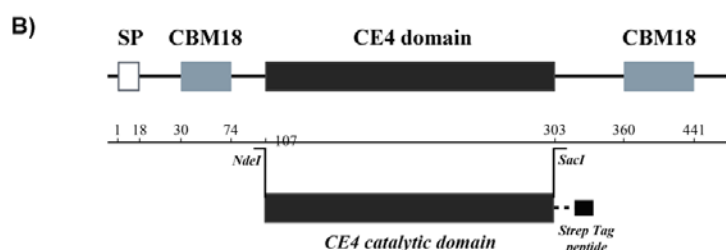
B)

SP CBM18   CE4 domain   CBM18

1 18 30 74 107   303 360 441

NdeI   SacI

CE4 catalytic domain   Strep Tag peptide

2

Figure S2. Augustus and GeneMark output for Pc_2566 gene prediction. (A) Augustus predicts 5 introns while (B) GeneMark only 4. Two first introns are coincident in size and position.

A. Augustus output for Pc_2566 gene prediction

```
unnamed-1  AUGUSTUS  gene         2596  4297  0.07  +  .  g2
unnamed-1  AUGUSTUS  transcript   2596  4297  0.07  +  .  g2.t1
unnamed-1  AUGUSTUS  start_codon  2596  2598  .     +  0  transcript_id "g2.t1"; gene_id "g2";
unnamed-1  AUGUSTUS  initial      2596  2819  0.26  +  0  transcript_id "g2.t1"; gene_id "g2";
unnamed-1  AUGUSTUS  internal     2910  2974  0.5   +  1  transcript_id "g2.t1"; gene_id "g2";
unnamed-1  AUGUSTUS  internal     3031  3653  0.64  +  2  transcript_id "g2.t1"; gene_id "g2";
unnamed-1  AUGUSTUS  internal     3859  3898  0.45  +  0  transcript_id "g2.t1"; gene_id "g2";
unnamed-1  AUGUSTUS  internal     3977  4144  0.36  +  2  transcript_id "g2.t1"; gene_id "g2";
unnamed-1  AUGUSTUS  terminal     4200  4297  0.4   +  2  transcript_id "g2.t1"; gene_id "g2";
unnamed-1  AUGUSTUS  intron       2820  2909  0.48  +  .  transcript_id "g2.t1"; gene_id "g2";
unnamed-1  AUGUSTUS  intron       2975  3030  0.46  +  .  transcript_id "g2.t1"; gene_id "g2";
unnamed-1  AUGUSTUS  intron       3654  3858  0.46  +  .  transcript_id "g2.t1"; gene_id "g2";
unnamed-1  AUGUSTUS  intron       3899  3976  0.53  +  .  transcript_id "g2.t1"; gene_id "g2";
unnamed-1  AUGUSTUS  intron       4145  4199  0.42  +  .  transcript_id "g2.t1"; gene_id "g2";
unnamed-1  AUGUSTUS  CDS          2596  2819  0.26  +  0  transcript_id "g2.t1"; gene_id "g2";
unnamed-1  AUGUSTUS  CDS          2910  2974  0.5   +  1  transcript_id "g2.t1"; gene_id "g2";
unnamed-1  AUGUSTUS  CDS          3031  3653  0.64  +  2  transcript_id "g2.t1"; gene_id "g2";
unnamed-1  AUGUSTUS  CDS          3859  3898  0.45  +  0  transcript_id "g2.t1"; gene_id "g2";
unnamed-1  AUGUSTUS  CDS          3977  4144  0.36  +  2  transcript_id "g2.t1"; gene_id "g2";
unnamed-1  AUGUSTUS  CDS          4200  4294  0.4   +  2  transcript_id "g2.t1"; gene_id "g2";
unnamed-1  AUGUSTUS  stop_codon   4295  4297  .     +  0  transcript_id "g2.t1"; gene_id "g2";
```

B. GeneMark output for Pc_2566 gene prediction

Gene #	Exon #	Strand	Exon Type	Exon Range		Exon Length	Start/End Frame
3	1	+	Initial	2596	2819	224	1 2 - -
3	2	+	Internal	2910	2974	65	3 1 - -
3	3	+	Internal	3031	3653	623	2 3 - -
3	4	+	Internal	3700	4144	445	1 1 - -
3	5	+	Terminal	4200	4297	98	2 3 - -

Figure S3. *Pochonia chlamydosporia* (Pc_2566) and *Metarhizium acridum* (NW_006916702.1) chitin deacetylase (CDA) genomic sequence alignment. Introns of *M. acridum* sequence are marked in yellow. First two introns (red bars) predicted by both Augustus and GeneMark are coincident. Third intron (green bars) from *P. chlamydosporia* was predicted taking into account the position of *M. acridum* third intron.

```
Pc_2566 genomic sequence        ATGGCTCTGAGAATATCCCTCCTATTTGCCGTGTGTGCTCTGGCTACCGACGGTTCCCGT
NW_006916702.1 Ma genomic sequence  ATGACACCAAAACTCGGCCTCCTCCTCGCTGTGTACTTCCAGAGCCTGCTAGCCTTCCGT

Pc_2566 genomic sequence        TTCATTCGGGCAGCAGAGGCCAATGAGATGTCCATGTGCGGTCCCGGTAAAGGATCATGC
NW_006916702.1 Ma genomic sequence  TTCAATCGGGAACTAGTCGCCAATGAAAAAGCAATGTGCGGTCCCGGCGAGGGATCATGT

Pc_2566 genomic sequence        GAAATGGGCTCATGTTGTTCAGAGATGGGGTATTGCGGCACTACGGAACAGTATTGCTCA
NW_006916702.1 Ma genomic sequence  GAAATGGGTTCATGCTGCTCAGAGATGGGGTTCTGCGGCACCACAGCGGAGTATTGTTCA

Pc_2566 genomic sequence        GGCTCTCAGTGCCAGTTGGACTACAGCCATACTTGCCACACGCTGTGAGTAGCAACTACG
NW_006916702.1 Ma genomic sequence  GGCTCGCAGTGTCAACTAGACTACAGTCATACTTGCGATACGCTGTGAGCATCAAA----

Pc_2566 genomic sequence        CAAGTTGTGTGATGTGATCCAGAGACCGTTGAAATGACATGACATGC-CATGCCGAGCTG
NW_006916702.1 Ma genomic sequence  ------------------------ACCAGATTGGCAAGACTTTCATGTAATTGGCTA

Pc_2566 genomic sequence        ACAAGTTGCATCCAGAGTTCCCCCTCGTGGGAAGGACACTTCAGGAATAACCCGTCGACA
NW_006916702.1 Ma genomic sequence  ACCAATAGAGTTAAGGGTGCCTCCTGATGGCAGTGACACTTCAGAAATACCCCGACCACG

Pc_2566 genomic sequence        AGTCGGAAACGTCCCTTATGGTGAGAGCCACGAAGGAAAGGGGTCGTGATTTTCGGAGGC
NW_006916702.1 Ma genomic sequence  GGTTGGGAAAGTCCCATATGGTAATTGTTCACAGAGTTTGCCGGAAGGGACACTTTGGAC

Pc_2566 genomic sequence        TAACACATCTCCCCAGGACCAATGATTACAGCCTTGCAAACAGCCTGGCATGGTAGCCTTG
NW_006916702.1 Ma genomic sequence  TAACGATTT----CAGGGCCCATGATCACATCATGCAACAATCCAGGGATGGTGGCTCTG

Pc_2566 genomic sequence        ACATTCGACGACGGTCCGTACATTTACACTACCGAAATCCTGAACCTGCTTGACCGGCTC
NW_006916702.1 Ma genomic sequence  ACGTTCGACGACGGCCCCTACATCTACACCACGGAACTTCTGGACCTTCTCGCAGCGCAC

Pc_2566 genomic sequence        AAGGTCAAGGCTACTTTCTTCATCACGGGTGACAACCGAGTCAAGGGACATATTGACGAT
NW_006916702.1 Ma genomic sequence  GAGGTCAAGGCCACCTTCTTCATCACGGGGAGACAACCGCGCCAAGGGACACATCGATGAC

Pc_2566 genomic sequence        CCGGCGACTCGCTGGCCAAGCATTATACGTCGCATGTACGACGCGGGACATCAAATTGGG
NW_006916702.1 Ma genomic sequence  CCGGCGACGGAATGGCCGAGTATCCTGCGGCGCATGTACAACGCGGGCCACCAGGTCGCT

Pc_2566 genomic sequence        AGCCACACCTGGACGCATCGTGATCTGAACCATATCAACGAGACGGTGCGGCGAGCTGAA
NW_006916702.1 Ma genomic sequence  AGTCACACGTGGACACATCGCGACTTGACTCAGGTCAACGAGACGGTGCGGCGGGGCAGAG

Pc_2566 genomic sequence        ATCATTCACAATGAGATGGCCATACGGAACATTCTTGGTTGGATCCCCACGTACATCCGG
NW_006916702.1 Ma genomic sequence  ATTATTCACAACGAAATGGCCCCTTCGCAATATCCTTGGCCGGATCCCCACGTACATCCGG

Pc_2566 genomic sequence        CCACCTTTCTTGGAGTGTTCCGCCAGATCTGGGTGTGAGAAGACCATGAGAGACATGGTG
NW_006916702.1 Ma genomic sequence  CCCCCATTCCTGGAATGTTCTACCGGTTCCGGGTGTGAAGAAACCTTGGGTGACTTGGCA

Pc_2566 genomic sequence        TATCACTCCATCTCCACCAACTTGGACACCAAAGACTACATGTATGACGACCCAACCCTC
NW_006916702.1 Ma genomic sequence  TACCACTCCATCTCGGCCAACTTGGACACCAAAGACTACATGTACGATGACCCGGCCCTC

Pc_2566 genomic sequence        ATTCAAAGGGCAAAGGACCTCTACTCTAATGGACTGTCAACCAATCCACGGCAGAACTCA
NW_006916702.1 Ma genomic sequence  ATACAGAGGTCAAAGGACCGCTATTCCAGCACCCTGTCCACAAACTCGAAAGAGAACTCA

Pc_2566 genomic sequence        TACATTGTCCTTGCACACGATGTTCATGAGCAAACCGTGTATGAACTGACTCCCATACATG
NW_006916702.1 Ma genomic sequence  TACATTGTGCTGGCTCATGATGTTCACGAACAAACGGTACACAATCTGACAGAGTACATG

Pc_2566 genomic sequence        GTAAAGGTTGCTAGAGAGAGAGGCTACCAGCTTGTCACGGTAGGTGAATGCCTAGGAGAC
NW_006916702.1 Ma genomic sequence  ATCAGTCTTGCGAGAGAAAGAGGGCTACAAGCTCGTCACGGTGGGCGAATGCCTAGGGGAC

Pc_2566 genomic sequence        CCCAGAGGTAACTGGTATAGATCAGCGGAAGACTCGCCAAAATCCACCAGAAGCGACAAA
NW_006916702.1 Ma genomic sequence  CCCGAGGAGAACTGGTACCGGTTCGGCAGAGACCTCTTCGAACGTGGCCAGGACCGTTGGC

Pc_2566 genomic sequence        AAGAAAGACTCATGTCCCGGTCCAAAAAAGCCAATCGTCACGGGATATCCCGCCCCCAAGG
NW_006916702.1 Ma genomic sequence  AAGAGGAACCTGTGCTCGGGACCGGCCAAACCGCAGGTCACATCTGTCTCACCTCCGCGG

Pc_2566 genomic sequence        AATACGACGAGCAAGATATCTCCCAACCAACGCTGCGGAGGGAAGTACAGGATACATCTGT
NW_006916702.1 Ma genomic sequence  AATACGACTACCAAAATATCGCCCAACCAACGCTGCGGTGGGAGCACAGGATATATCTGT

Pc_2566 genomic sequence        CCAGGCTCCGGGTTTGGAGACTGCTGCTCTCACTATGGATATTGGTAATTCTGTCTTTCA
NW_006916702.1 Ma genomic sequence  CCGGGGTCCGGGTTTGGGGACTGCTGCTCCCCACTGGGGATATTGGTATGACCCCTTTCAA

Pc_2566 genomic sequence        AAGCCCACAACAACCCACTCATCATGAGAAGAGCCTCTTTTTCACTGACCGTGTCTCACT
NW_006916702.1 Ma genomic sequence  GCGGT--------AGTCGTTGCCCGCGGGTTCGTTTGCCTTTTCGCTGACCTTGTCCTACC

Pc_2566 genomic sequence        AGTGGATCAACTCCCGAGTTCTGTGGCACAGGGTGTGACAATGACTTTGGGGACTGCAAC
NW_006916702.1 Ma genomic sequence  AGCGGATCAACGCCCGAGTTTTGTGGCACGGGGTGCGACAATGACTTTGGGGACTGCGAT

Pc_2566 genomic sequence        CCAAATCCCCCAGGCATCTGGGACACAACGAATGGCTTATGTGGCGCTCGGTTCTCTGCG
NW_006916702.1 Ma genomic sequence  CCATCCCCGCAGGGTGTTTTGGATACAACGAATGGCCTGTGTGGCACTCGGTACTCTGCT

Pc_2566 genomic sequence        ACATGCCTTAATTTTGCAGGAAAGGCATGCTGCTCCAAGTATGGATATTGGTATGTAGTG
NW_006916702.1 Ma genomic sequence  ACCTGTCTTCACTTTGCGGGAAAGACGTGCTGCTCAAAGTATGGATATTGGTATGTGC--

Pc_2566 genomic sequence        CATTCTGTGTGGTTACTTTATCGGTTGGACTGACGTGGTCGGCAGCGGCAATAAGACTGA
NW_006916702.1 Ma genomic sequence  CGTTTCTCGTCGCCAAGTTTATTGGTGTTCTGACAATTCAAGTAGCGGGAGTCAGGTTGA

Pc_2566 genomic sequence        GCATTGTGGTGAAGGTTGCCAGGCAAAGTACGGCCGATGCGATTAGACAGATGGTGGCGC
NW_006916702.1 Ma genomic sequence  GCATTGTGGTGAGGGGTGCCAGGACAAGTACGGTCATTGTGACTGAATGA----------

Pc_2566 genomic sequence        ATCGTTACAGAGCTGAGATGTGACGA
NW_006916702.1 Ma genomic sequence  --------------------------
```

4

Figure S4. pET22b-*Pc_2566*_CE-StrepIIC plasmid for *Pc*CDA protein expression. The plasmid contains the *Pc*CDA CE4 domain and the Strep tag peptide coding sequences followed by a stop codon before the His tag codifying peptide of the original pET22b(+) vector.
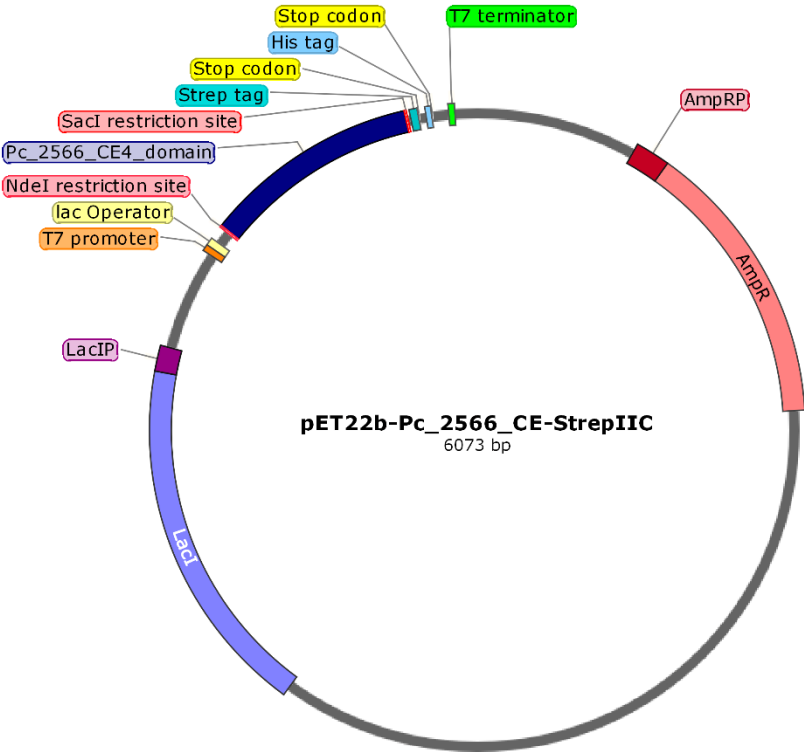
Figure S5. SDS polyacrylamide gel electrophoresis analyses of protein expression and refolding. A) Proteins from induced *E. coli* BL21 strain transformed with pET22b-Pc_2566_CE4. Lane S: supernatant (soluble fraction), Lane P: insoluble fraction (pellet) after cell disruption. B) Protein refolding and purification. Lane 1, culture supernatant (soluble fraction). Lanes 2 and 3, pellet after Triton X100 treatment. Lanes 4 and 5, samples after two washes to remove Triton X100. Lane 6, supernatant after urea treatment for protein solubilisation. Lane 7, pellet obtained after urea treatment. Lane 8, sample after protein refolding by dialysis. Arrowheads indicate expected size of *Pc*CDA catalytic domain. Original gels in Fig. S11.
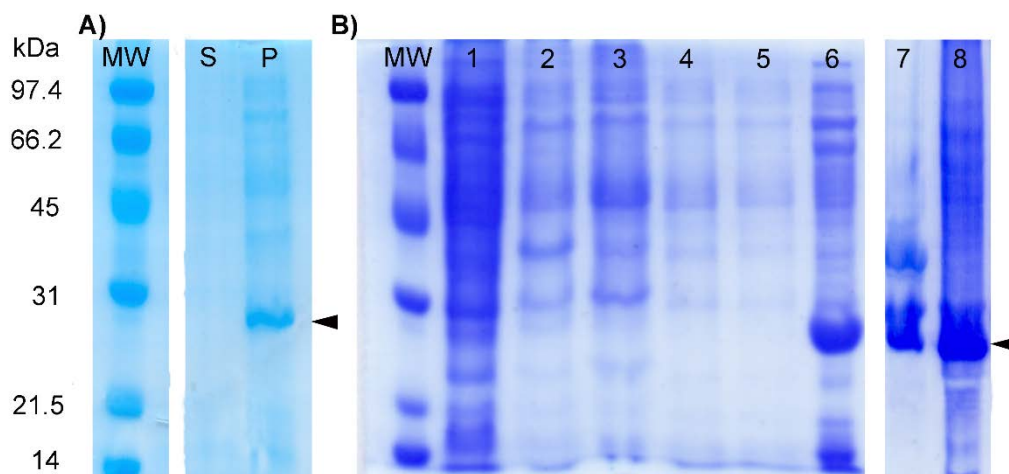
Figure S6. HPLC-MS monitoring of deacetylase activity by *Pc*CDA catalytic domain on GlcNAc$_4$. Chromatograms show the presence of the starting product (A4) and the formation of mono-deacetylated (A3D1) and di-deacetylated (A2D2) products at different reaction times: (A) 1 h, (B) 16 h, (C) 24 h and (D) 100 h. Reaction conditions: 0.2 mM substrate, 3.2 nM enzyme, 50 mM K$_2$HPO$_4$, 300 mM NaCl, pH 8.0, 37°C.
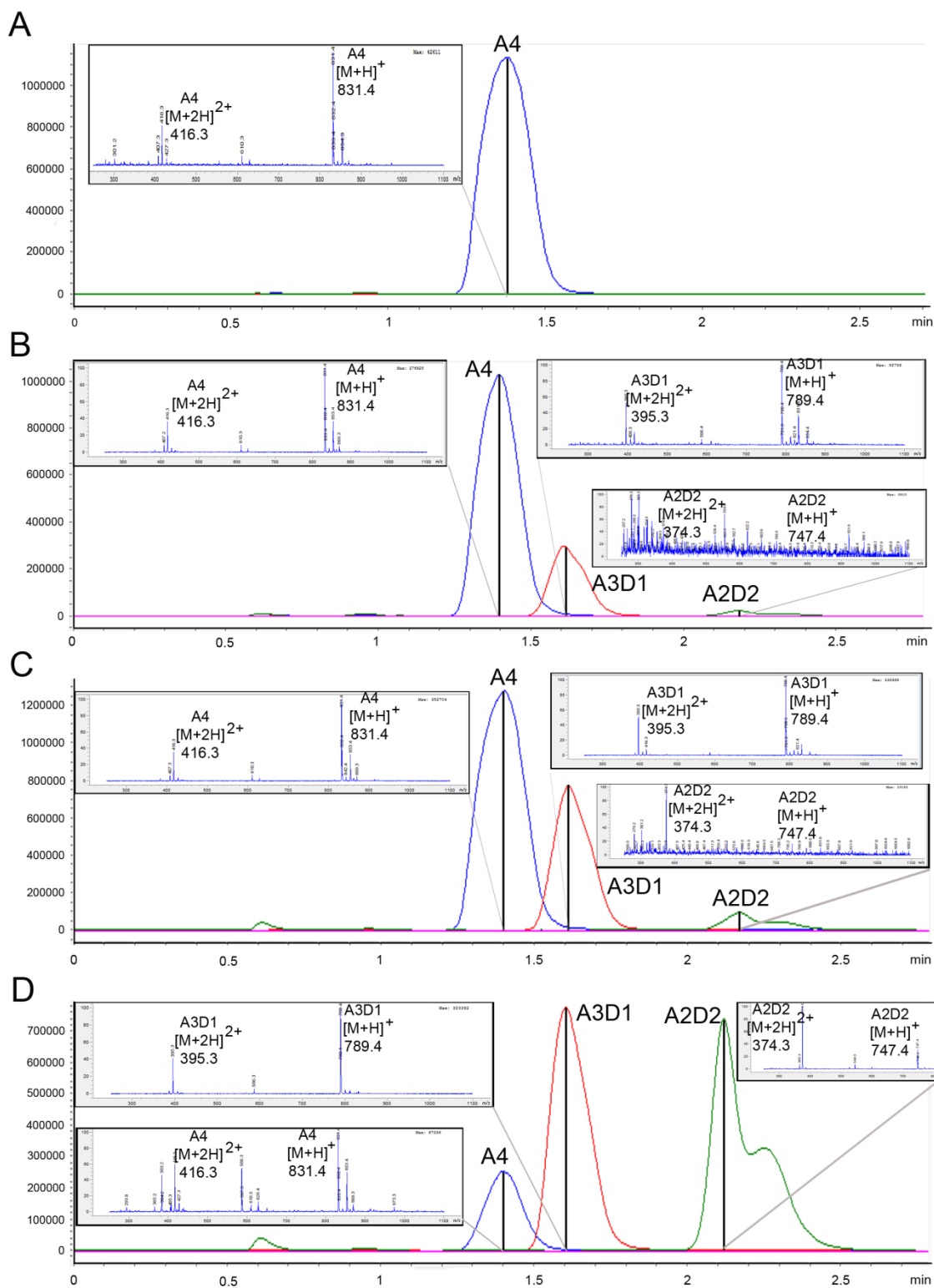
Figure S7. HPLC-MS monitoring of GlcNAc$_3$ incubated with the *Pc*CDA catalytic domain. Chromatograms show the presence of the starting product (A3) and the absence of reaction products with time: (A) 10 min, (B) 5 h, (C) 24 h and (D) 100 h. Reaction conditions: 0.2 mM substrate, 3.2 nM enzyme, 50 mM K$_2$HPO$_4$, 300 mM NaCl, pH 8.0, 37ºC.
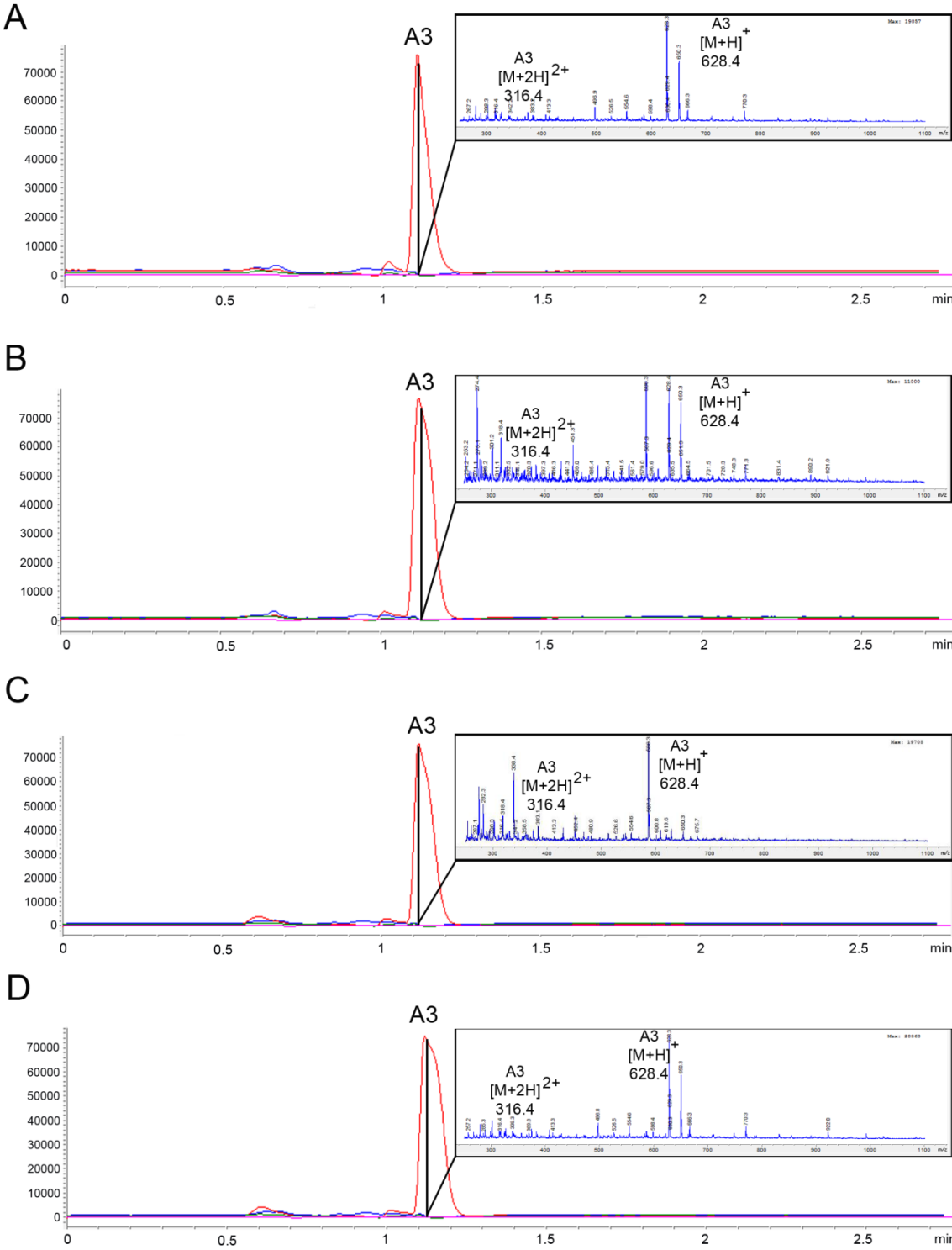
Figure S8. Sequence alignment of *Pc*CDA from *P. chlamydosporia* (catalytic domain) and *Cl*CDA from *C. lindemuthianum.*

```
Pc_2566 CDA    MALRISLLFAVCALATDGSRFIRAAEANEMSMCGPGKGSCEMGSCCSEMGYCGTTEQYCS
ClCDA          ------------------------------------------------------------


Pc_2566 CDA    GSQCQLDYSH---TCHTLVPPRGKDTSGITRRQVGNVPYGPMITACKQPGMVALTFDDGP
ClCDA          -----MHFSTLFGAAATAALAGSTNASPLARRQ---VPVGTPILQCTQPGLVALTYDDGP


Pc_2566 CDA    YIYTTEILNLLDRLKVKATFFITGDNRVKGHIDDPATRWPSIIRRMYDAGHQIGSHTWTH
ClCDA          FTFTPQLLDILKQNDVRATFFVNGNNW--ANTEAGSN--PDTIRRMRADGHLVGSHTYAH
                                    Loop 1


Pc_2566 CDA    RDLNHINETVRRAEIIHNEMAIRNILGWIPTVIRPPFLEGSARSGCEKTMRDMVYHSIST
ClCDA          PDLNTLSSADRISQMRQLEEATRRIDGFAPKYMRAPYLSGDA--GCQGDLGGLGYHIIDT
                 Loop 2                           Loop 3


Pc_2566 CDA    NLDTKDYMYDDPTLIQRAKDLYSNGLSTNPRQNSYIVLAHDVHEQTVYELTPYMVKVARE
ClCDA          NLDTKDYENNKPETTHLSAEKFNNELSADVGANSYIVLSHDVHEQTVVSLTQKLIDTLKS
                 Loop 4              Loop 5        Loop 6


Pc_2566 CDA    RGYQLVTVGECLGDPRGNWYRSAEDSPKSTRSDKKKDSCPGPKKPIVTDIPPPRNTTSKI
ClCDA          KGYRAVTVGECLGDAPENWYKAHHHHHH--------------------------------


Pc_2566 CDA    SPNQRCGGSTGYICPGSGFGDCCSHYGYCGSTPEFCGTGCDNDFGDCNPNPPGIWDTTNG
ClCDA          ------------------------------------------------------------


Pc_2566 CDA    LCGARFSATCLNFAGKACCSKYGYWYVVHSVWLLYRLD
ClCDA          --------------------------------------
```

Figure S9. Structural models of *Pc*CDA catalytic domain in complex with substrates. A) Model 2 with superimposed A3 ligand from template 4OUI (VcCDA·A3 complex). B) Simulated docking of A4 ligand to Model 1. In both, A and B, the ligand binding mode places the penultimate GlcNAc residue properly oriented for catalysis in subsite 0. C) Simulated docking of A4 ligand to Model 1 in a second binding mode, which places the reducing end GlcNAc residue in subsite -2. This binding mode was only possible in Model 1 with an extended Loop 1 conformation, but not in Model 2 with a closed Loop 1 conformation. Loops are coloured as in Figures 4 and 6.
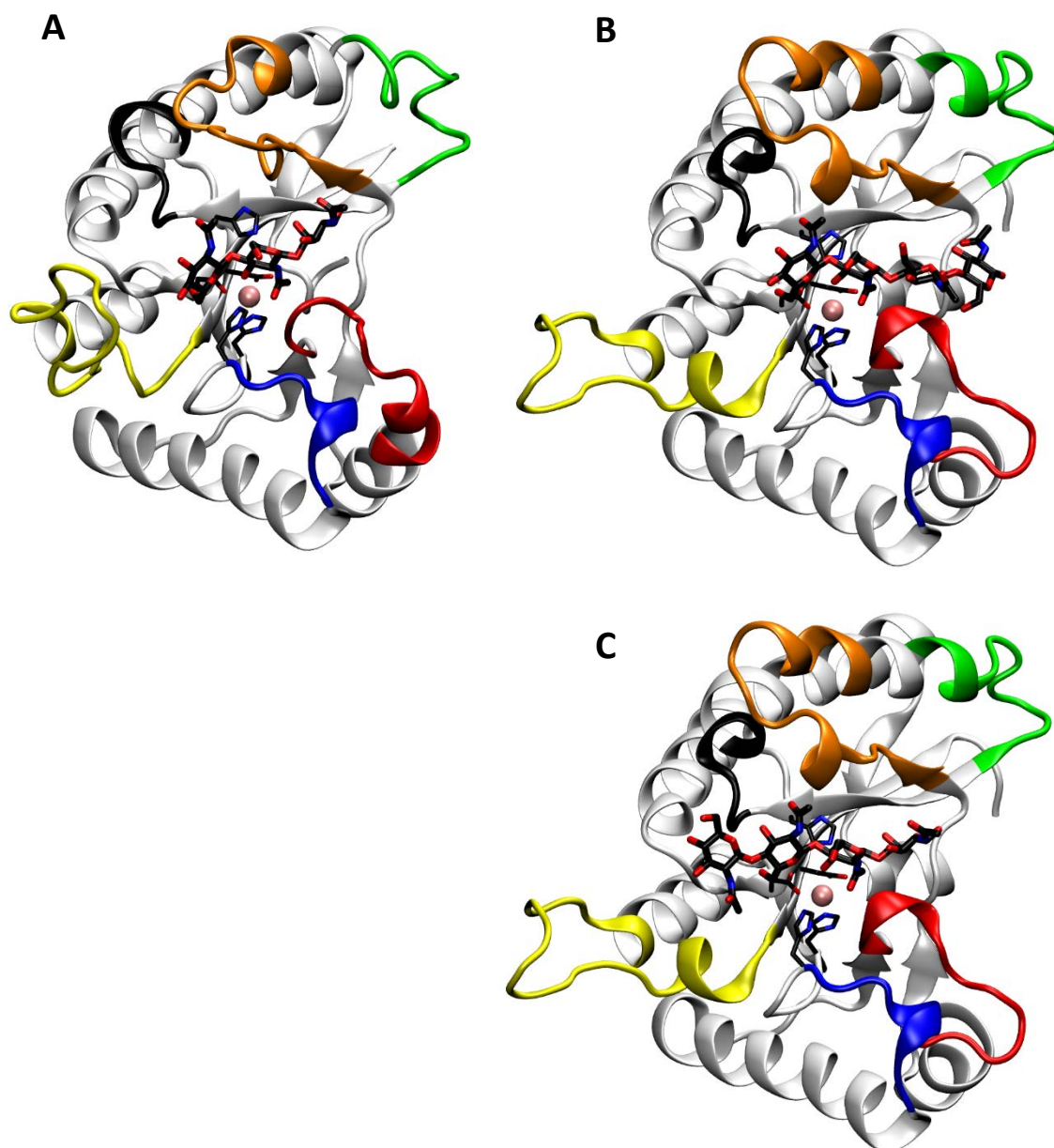
Figure S10. Superposition of Models 1 and 2 (model structures A and B in Figure S9). Both models are essentially identical along the protein structure, except for Loop 1, which appears as an ensemble of conformations in both models. Shown here in yellow, the extended (Model 1) and closed (model 2) Loop 1 conformations for the lowest energy models.
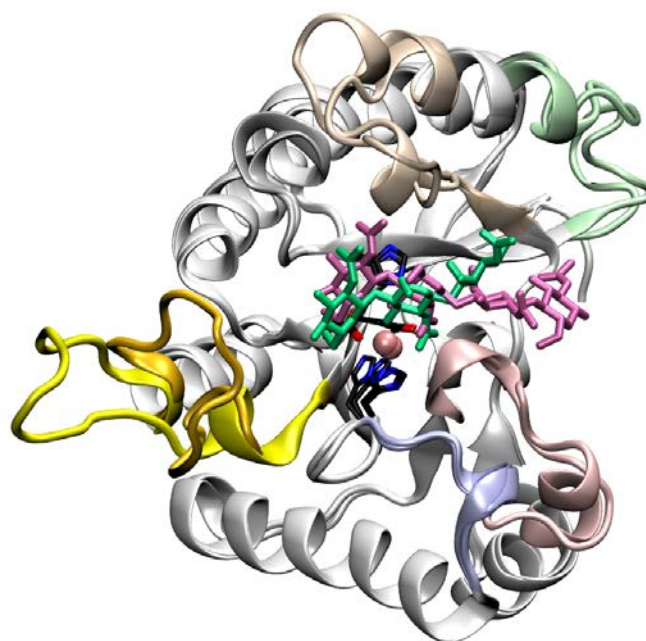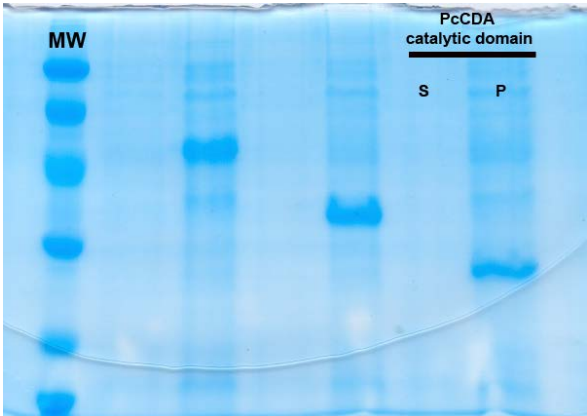
Figure S11. Original SDS-PAGE gels used in Figure S5.

A) Original gel in Figure S5.A (labelled lanes)

B) Original gels in Figure S5.B (labelled lanes)