cells

MDPI

Supplemantary Information

The Diverse Calpain Family in Trypanosomatidae: Functional Proteins Devoid of Proteolytic Activity?

Table 1. Calpain sequences retrieved from *A. deanei* genome.

Sequence ID	Assemble	Domain architecture	Predicted molecular mass (kDa)	Conserved catalytic triad (C,H,N)	Note
EPY23292.1 **CAD2213457.1**	GCA_000442575.2 GCA_903995115.1	DUF1935	14.54	------	Identical Proteins
EPY28159.1 **CAD2219314.1** **CAD2219315.1**	GCA_000442575.2 GCA_903995115.1 GCA_903995115.1	DUF1935	12.92	------	Identical Proteins
EPY33088.1 **CAD2215177.1**	GCA_000442575.2 GCA_903995115.1	DUF1935	15.25	------	Identical Proteins
EPY34498.1 **EPY43070.1** **EPY43802.1** **CAD2215188.1**	GCA_000442575.2 GCA_903995115.1	DUF1935	16.09	------	Identical Proteins
EPY37700.1 **CAD2215183.1**	GCA_000442575.2 GCA_903995115.1	DUF1935	20.03	------	Identical Proteins
EPY38860.1 **EPY39796.1** **CAD2215173.1**	GCA_000442575.2 GCA_903995115.1	DUF1935	14.39	------	Identical Proteins
EPY41474.1 **CAD2215168.1**	GCA_000442575.2 GCA_903995115.1	DUF1935	14.06	------	Identical Proteins
EPY42018.1	GCA_000442575.2	DUF1935	15.82	------	------
EPY38340.1	GCA_000442575.2	fragmented CysPc	23.97	------	Short amino acid sequence from the catalytic domain
EPY38663.1	GCA_000442575.2	fragmented CysPc, CBSW	32.88	------	Short amino acid sequence from the catalytic domain
EPY17791.1 **CAD2215189.1**	GCA_000442575.2 GCA_903995115.1	CysPc	80.68	------	Identical Proteins
EPY31064.1	GCA_000442575.2	CysPc	61.9	present	Identical Proteins

CAD22217327.1	GCA_903995115.1				
EPY33036.1	GCA_000442575.2	CysPc	53.67	------	------
EPY19782.1	GCA_000442575.2	CysPc	104.13	------	Identical Proteins
CAD22214161.1	GCA_903995115.1				
EPY35501.1	GCA_000442575.2	CysPc	41.89	------	Identical Proteins
CAD22215192.1	GCA_903995115.1				
CAD22215193.1	GCA_903995115.1	CysPc	56.70	------	------
EPY40913.1	GCA_000442575.2	CysPc	30.53	------	Identical Proteins
CAD22215397.1	GCA_903995115.1				
CAD22215575.1	GCA_903995115.1	CysPc	75.57	------	------
EPY17550.1	GCA_000442575.2	CysPc	54.99	------	Identical Proteins
CAD22222307.1	GCA_903995115.1				
EPY40095.1	GCA_000442575.2	CysPc	66.74	present	------
EPY16971.1	GCA_000442575.2	2xCysPc	50.11	------	------
EPY35473.1	GCA_000442575.2	DUF1935, CysPc	81.01	present	------
EPY34508.1	GCA_000442575.2	DUF1935, CysPc	95.30	------	Identical Proteins
EPY37930.1					
CAD22214023.1	GCA_903995115.1				
EPY27435.1	GCA_000442575.2	DUF1935, CysPc	95.86	------	Identical Proteins
CAD22214024.1	GCA_903995115.1				
EPY24669.1	GCA_000442575.2	DUF1935, CysPc	82.47	------	Identical Proteins
CAD22214025.1	GCA_903995115.1				
EPY28634.1	GCA_000442575.2	DUF1935, CysPc	128.39	------	Identical Proteins
CAD22214026.1	GCA_903995115.1				
EPY39328.1	GCA_000442575.2	DUF1935, CysPc	88.95	------	Identical Proteins
CAD22214027.1	GCA_903995115.1				
CAD22215190.1	GCA_903995115.1	DUF1935, CysPc	77.06	------	------
EPY26598.1	GCA_000442575.2	DUF1935, CysPc	77.53	------	Identical Proteins
EPY43876.1					
CAD22215191.1	GCA_903995115.1				

CAD221 5195.1	GCA_903 995115.1	DUF1935, CysPc	77.01	------	------
EPY264 55.1	GCA_000 442575.2	DUF1935, CysPc	77.03	------	------
EPY276 08.1	GCA_000 442575.2	DUF1935, CysPc	94.24	present	Identical Proteins
CAD222 0197.1	GCA_903 995115.1				
EPY298 41.1	GCA_000 442575.2	DUF1935, CysPc	80.85	------	------
EPY313 84.1	GCA_000 442575.2	DUF1935, CysPc	46.69	------	------
EPY321 39.1	GCA_000 442575.2	DUF1935, CysPc	80.20	present	Identical Proteins
CAD221 4018.1	GCA_903 995115.1				
EPY351 55.1	GCA_000 442575.2	DUF1935, CysPc	75,67	------	------
EPY359 37.1	GCA_000 442575.2	DUF1935, CysPc	80.33	present	------
EPY367 66.1	GCA_000 442575.2	DUF1935, CysPc	62.07	------	------
EPY374 55.1	GCA_000 442575.2	DUF1935, CysPc	85.88	------	------
EPY403 46.1	GCA_000 442575.2	DUF1935, CysPc	77.54	------	------
EPY418 99.1	GCA_000 442575.2	DUF1935, CysPc	83.53	------	------
EPY210 44.1	GCA_000 442575.2	CysPc, CBSW	73.48	------	Identical Proteins
CAD221 8077.1	GCA_903 995115.1				
EPY417 80.1	GCA_000 442575.2	CysPc, CBSW	123.76	present	Identical Proteins
CAD221 6952.1	GCA_903 995115.1				
CAD222 0901.1	GCA_903 995115.1	CysPc, CBSW	121.94	------	------
CAD221 3168.1	GCA_903 995115.1	CysPc, CBSW	130.64	------	------
EPY204 06.1	GCA_000 442575.2	CysPc, CBSW, CysPc	217.36	------	------
EPY222 62.1	GCA_000 442575.2	CysPc, CBSW, CysPc	174.58	------	------
EPY325 74.1	GCA_000 442575.2	CysPc, CBSW, CysPc	98.51	------	------
CAD221 3171.1	GCA_903 995115.1	CysPc, CBSW, CysPc	97.73	------	------
CAD221 5402.1	GCA_903 995115.1	CysPc, CBSW, CysPc	354.59	------	------
EPY227 07.1	GCA_000 442575.2	RPT1, CysPc, CBSW	236.37	------	Identical Proteins
CAD221 5572.1	GCA_903 995115.1				
CAD221 5393.1	GCA_903 995115.1	CysPc, CBSW, CysPc, RPT1	522.62	------	------
EPY280 67.1	GCA_000 442575.2	KISC, 3xARM, CBSW	103.81	------	Identical Proteins; Kinesin motor domain associated with CBSW
CAD222 1641.1	GCA_903 995115.1				

CAD221 5401.1	GCA_903 995115.1	RPT1, CysPc	189.04	------	------
EPY204 04.1	GCA_000 442575.2	2xRPT1, CysPc	161.06	------	------
CAD221 5396.1	GCA_903 995115.1	RPT1, RPT2, CysPc	301.59	------	------
EPY215 60.1	GCA_000 442575.2	2xRPT1	24.58	------	RPT sequences annotated as calpain-like proteins
EPY270 93.1	GCA_000 442575.2	2xRPT1	157.43	------	Identical Proteins; RPT sequences annotated as calpain-like proteins
CAD221 5574.1	GCA_903 995115.1				
EPY162 12.1	GCA_000 442575.2	2xRPT1	11.23	------	Identical Proteins; RPT sequences anotated as calpain-like proteins
CAD221 5398.1	GCA_903 995115.1				
EPY204 05.1	GCA_000 442575.2	2xRPT1	58.08	------	RPT sequences anotated as calpain-like proteins
EPY222 64.1	GCA_000 442575.2	2xRPT1	125.40	------	RPT sequences anotated as calpain-like proteins
EPY245 73.1	GCA_000 442575.2	2xRPT1	239.87	------	RPT sequences anotated as calpain-like proteins
EPY259 74.1	GCA_000 442575.2	2xRPT1	189.50	------	Identical Proteins; RPT sequences anotated as calpain-like proteins
CAD221 5400.1	GCA_903 995115.1				
EPY297 04.1	GCA_000 442575.2	RPT1, RPT2, RPT1	101.27	------	RPT sequences anotated as calpain-like proteins

The ID of the calpain sequences was retrieved from two different assemblies of the *Angomonas deanei* genome (Genebank ID 14191). The sequences were ordered based on their domain architecture. The predicted molecular mass was calculated in http://www.bioinformatics.org/sms/prot_mw.html. Sequences annotated as "calpain-like proteins" but with less than 100 amino acids or without any calpain associated domain were not included. CBSW—calpain-type alpha-beta-sandwich domain, CysPc—proteolytic core domain, KISC—kinesin domain, DUF1935—conserved N-terminal calpain-related domain from trypanosomatids, RPT—repeated domain found in de-ubiquitinating proteins, 2× and 3× indicate the number of times that a domain appears in the sequence. Fragmented CysPc stands for short amino acid sequences from the catalytic domain.

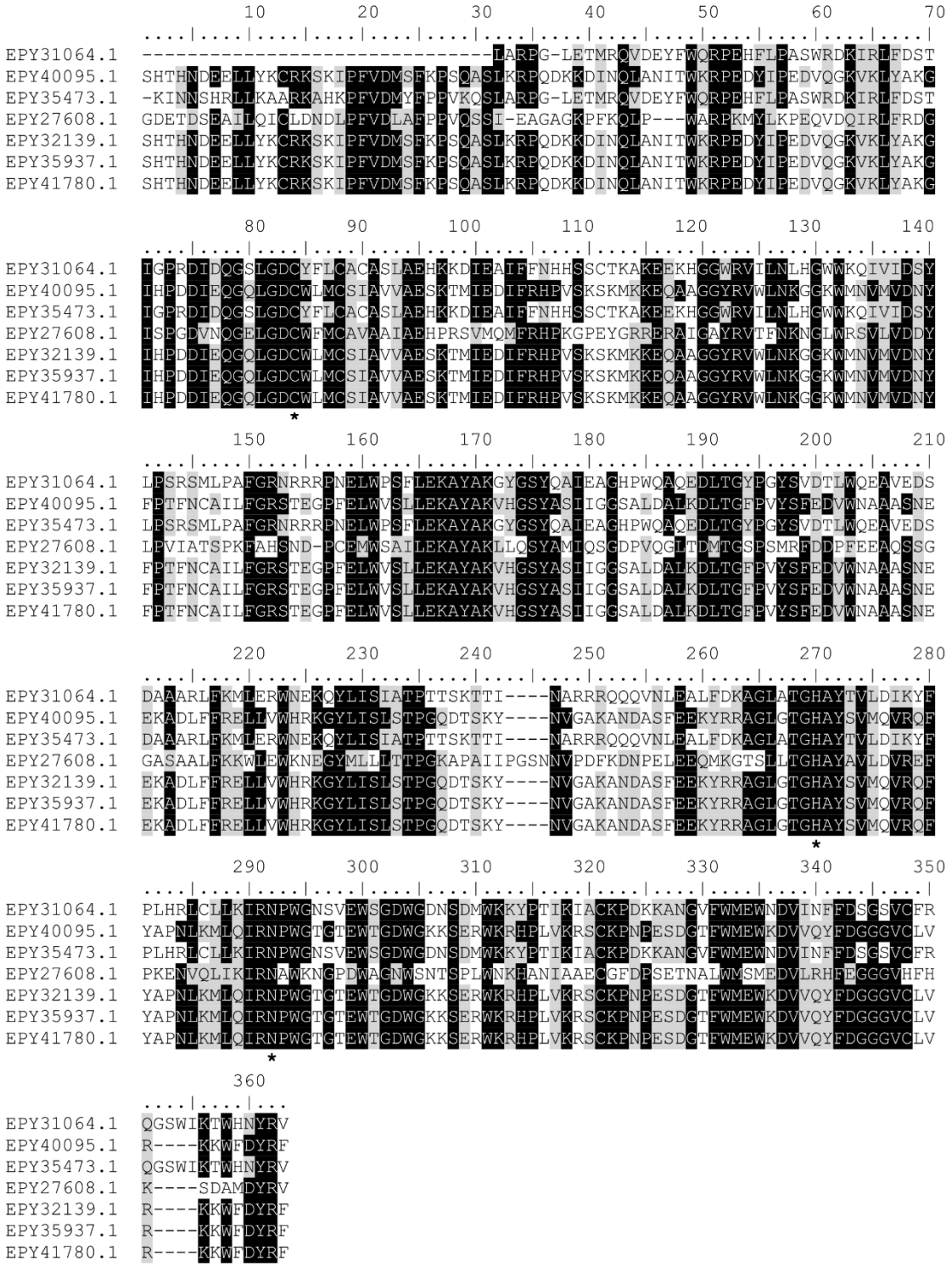

Figure S1. Multiple sequence alignment of the cysteine peptidase core (CysPc) domain of *Angomonas deanei* with the conserved catalytic triad (C,H,N). The sequence alignment was performed using ClustalW. GeneDB ID of the *A. deanei* calpain sequences are on the left. Identical and similar residues are represented in black and grey, respectively. The amino acids responsible for the proteolytic activity of conventional calpains are indicated by a *.