

**Figure S1.** The N-terminal parts of protein alignments demonstrating the presence of extension in *Trimastix* protein relative to the prokaryotic homologues.

### Aconitase

```

Trimastix      -----MSTPLLQTPLSLYDQTFP-----PYERQLAKLAQVRAKLGRQDITLAEKVLFTHVADVEALS--TAPGTYL
Nasonia      MSYCTRILQAQKVASLAAEVQQRCFSTSPITFAAAKVAAMSKFDSTAY---LPYDKLDENIKIVK-KRLDRPLTLSEKILYSHL-DEPQKQDIVRGTSYL
Dictyostelium -----MNSLVKGISKVRSTRSFSTVMSPLPKNK---LNYEGIDAKLKQFR-LHHNKPLTLAEKIYGHY-EDPSTK-VERGITYL
Monosiga      -----MLRVNASRAALVGVRAFSTSRVAASVPMRFDAP---MPYDTMRKRRLDIVR-GRLLNRPLTLSEKVLVYSHL-DEPETQDIERGVSYL
Phytophthora -----MSSSCGVFVPLSVLDPKGSPTIDQTYEKMEKNLYTVRKTNGNKPLTLAEKIYGHY-DDATTK-AKRGETYL
Ignavibacterium -----MTANFDMIKK---VYSSYKQKLADAR-KVLNRPMPTYAEKVLVYTHLWEKQ-TREYVVRKDVV
Melioribacter -----MAANFDMIKR---VYDNFENKVDKAR-KVVGRLPTYAEKVLVYHAALEEP-TREYVVRKDVV
Leptosira     -----MAFDIEMIRA---RYAKIGDLVTKAR-NVVGRLPTLLEKILYSHLWEGEPKAAEYKQKSV
Flavobacteria -----MAFDIEMIEK---VYANMVARVDKAR-ELVGKPLTLLEKILYSHLWDETSTNVYKRGVDV
Spirosoma     -----MAFDLDMIQR---VYANLGERVEAAR-KAVGKPLTLSEKILYSHLFAQTPTQALERKQDV
    
```

### HydE

```

Trimastix      -----MRSTPIPDDGVSDVRVKIWNFGGNSLSKRPESEAKAIP
Chlamydomonas MAHSLSAHSRQAGDRKLGAGAASSRPSCPSSRRIVRVAHASASAKATPDVVDLPPAHARAATAAANRRARAMASAEAAETLGDGFLGKGGSPGATA
Trichomonas    -----MNTLIVPTSRHLPSFSEALEVAISGK
Trichomonas    -----MQLNTLNPPTRELPSFSEALKVAVAGE
Caldicellulosiruptor -----MREEKMKVKEILDKAYNEH
Ruminococcus   -----MKRLIDKHLHEQ
Thermotoga     -----MTGREILEKLERR
Thermoanaerobacter -----MKVKEILEKAYHES
Symbiobacterium -----MIRQEVLLALIARAEPAGA
Clostridium    -----MKIKDIIDKAYVES

Trimastix      DLSFDELKMLLQ---KHTPETDAALFRYSNSIANQLFGNKCYVRALVEFVSNVCRKNCYCGIRKDMKGVRRYITPDEHELMALAKFAYENKFTSLVLQSG
Chlamydomonas  NLDREQVLGVLEAVVRRGDLNLERALYSHANAVTNKYCGGGVYVYRGLVEFVSNICQNDSCSYCGIRNNOKEVWRYTMPVEEVVEVAKWALENGIRNIMLQGG
Trichomonas    KMSHDHIVRLLN---AREKSEVDALHEAAEVEVTKRNFHNEVSVRGIVEFVSNACEKNCHYCGVNSYEP---FVIPEAAILDCCEFMWKKGYRNLVLQSG
Trichomonas    KMSHDHIVRLLN---PRTKDEIDALHSAAGAVTQRFVHNEVSVRGIVEFVSNACEKNCHYCGVNSYEP---FLIPEAAILNSCKFMWAKGYRNLVLQSG
Caldicellulosiruptor HLTKEIKFLLN---SEGEK-ELLFEYADRVKRYVGDVYLRGLIEFSSYCKNDCFYCGLRSSNKEAQRVYRMQDEIVEVAKRAYDMGYRNVVLQSG
Ruminococcus  MLTAEFFHRLLA---NRSEETQYARELANQVRQEVGNKIYVYRGLIEFTNYCKNDCCYCGIRRSNQAQRVRLTEEDILECCRQYQYEIFRFTVVLQGG
EFTREVLKEALS---INDRGFNEALFKLADEIRKRYVGDVYLRGLIEFSSYCKNDCFYCGLRSSNKEAQRVYRMQDEIVEVAKRAYDMGYRNVVLQSG
Thermotoga     NLTKEIKLLM---AEGEDK-DLLEKTAADRVKRYVGDVYLRGLIEFSSYCKNDCFYCGLRSSNQAQRVYRMQDEIVEVAKRAYDMGYRNVVLQSG
Thermoanaerobacter DLSRDELTCLLS---LGGEEQ-KALFAAADRVRARVMGDVYLRGLIEFSSYCKNDCFYCGIRRSNQAQRVYRMQDEIVEVAKRAYDMGYRNVVLQSG
Symbiobacterium
Clostridium    NLSQEEIVEILK---NKDEYNIKYLFNKAETTEKYCGHEVNIIRGIIEFVSNYCRNCVSYCGLNVTNNNGIKRYRMSKEEIVLVAKEAYEAGYKTLVLQSG
    
```

### HydF

```

Trimastix      MLSRSLLLRGSLLDSLQYMPFSAASTTVQKPLPFFSKLGLRTPHAAPSDLSLEPARAAGPGELOLRNTNIGIFGRMNAKGSSTLINAITQQNTSIVDAH
Trichomonas    -----MLTISIGRYFAKKGNDLPRTHISMVGFNMACKSTIMNAITQQPTSIVDST
Carboxydibrachium -----MNTTPTASRLHIGIFGRRNACKSSLINALTNQEIATVSDV
Desulfotomaculum -----MENTPFRGNRLHIAIFGRRNACKSSLINALTSQDIATVSDI
Dorea          -----MSLNSTPSSERTHIGIFGRRNACKSSLINALTGQNLATVSDV
Thiorhodococcus -----MLDTPKSLRLHIGIFGRRNACKSSLINALLGYPLSIVSDV
Halanaerobium -----MNSTAKGDRPHIAVFGRRNVKSSLINLRFNDLALVSSQ
Desulfotalea  -----MVFMQTVPKSMRHLGIFGRNTVQKSTFLNFLVDQDVAIASSQ
    
```

### Serine hydroxymethyltransferase

```

Trimastix      -----MLASAIQVPTALVAHASTGNIPLKEADPELYEIEKERRQSDGLELIASENFVSRAVLDCLGSCMTNKYSE
Zea mays      MAMATALRKLKANALRRQPLSRITPLYMASLPAATEERSGTTWTQKLNAPLEEVDPEIADIEHEKARQWKGLELIPSENFSTSVSMQAVGSMVTKNKYSE
Danio         -----MLTTLTRQITRPLCRVQRSSAAVCVRADGSWTGQESLSQDDPEMWDLLLKEKDRQCRGLELIASENFCSSRAALEAAGSCLNKNKYSE
Mus          -----MVSFSLLRITRPLQRCGQLVCMARAQHSKVAQTQAGEATGGWTGQESLSDSDPEMWEELLQREKDRQCRGLELIASENFCSRAALEALGSCLNKNKYSE
Naegleria    -----MQTTKRVSVVTSSSLRRTLASGQKRNFSSGIVNNGVNTKWAQQLNASLKDVPPELFDIEHEKRNQYKGLQIPSENFSTSKAVLDALGSMVTKNKYSE
Thermosinus  -----MNVLAGIDPEIAQAIDLERQQRQNKLELIASENFVSKAVMEAGQSVLTKNYAE
Geobacter    -----MSVLETFDPAVAEVIQETERQEYNLELIASENFVSPAVLEAQQSVLTKNYAE
Thermosipho  -----MWENVKTDPEIYDVILKEWERQYGLLELIASENFASLAVIEAMGSVLTKNYAE
Deferribacter -----MSLYEAVKNIDPQYDALMKELNRQETHIELIASENFVSKAVLEAQQSIMTKNYAE
Pelosinus    -----MKLKEFDPELAEATELEKRNQNKLELIASENFVTPMVEAMGTVLTNKYAE
Hippea       -----MSVLKDFDPVYQAIENEKKRQMYGLELIASENLVSEAVLEAQQSIMTKNYAE
    
```

### Ornithine transcarbamylase

```

Trimastix      MSAPVATPTTRSSRQALRNAKALIAKIRKLN-ALFEKDFLRTWDKSQDDIQTVLLVAQVLEELVKNIDTRVFFHGGVAVSNFRDNSTRTRFSFQFAASL
Ruminococcus  -----MKTLDQYIDKLSLNFKEMYNNDFFFTWTKTDELEAVFTVADALRFRENNISTKVFESGGLGISIFRDNSTRTRFSFASACNL
Escherichia   -----MMKIVNELIKDINSLTS-HLHEKDFLLTWEQTPDELKQVLDVAALKALRAENISTKVFNSGGLGISIFRDNSTRTRFSFASACNL
Anaerococcus  -----MTQDIKKLIDELQELDHEKLFNGDFHTWTKSQDELKAIWKTADILRAMREENISPFIKIFDSSGGLGISIFRDNSTRTRFSFASACNL
Klebsiella    -----MKTVNELIKDINKLNS-HLSEKDFLLTWEQSPDELKQVLDVAALKALRAENIATKVFNSGGLGISIFRDNSTRTRFSFASACNL
Vibrio        -----MKMNELIKEISQLES-KLHDKDFLLTWEQSPDELEKRVLTKTAQVLTMRGENIATNVFNNGLGISIFRDNSTRTRFSFASAINM
Yokenella     -----MKTVNELIKDINTLNS-HLHEKDFLLTWEQSPDELKQVLDVAALKALRAENIATNVFNNGLGISIFRDNSTRTRFSFASACNL
Sebaldella    -----MDKMKQFEEIARLEKLDYKMDLNDFFLWTKSTDELNSVFAVADALRNLRQNIISTKVFESGGLGISIFRDNSTRTRFSFASACNL
    
```

## Cpn60

```
Trimastix -----MQLFSTAAPALVRLASSAKEIKFGDEGRAAILKGVDTLAKAVATTLGPKGRVTLID-----DEF-EPRIITKDGVT
Arabidopsis MYRVLSKLSSTIGSSTSRKLVSGRIISRNYYAAKDISFGIGARAAMLQGVSEVAEAVKVTMGPKGRNVIIIE-----SSYGGPKITKDGVT
Saccharomyces -----MLRSSVVRSRATLRPLLRAYSSHKELKFGVEGRASLLKGVETLAEAVAATLGPGRNVLIE-----QFPGPKITKDGVT
Nyctotherus -----MRGLTNEVSKQATRSLISTPMRAIGTGKDISFGIEARKSLLDGC DKLADAVQLTLGPKGRNVVLD-----KGYGI PKITKDGVT
Trichomonas -----MSLIEAAKHFTRAFAKARDLKFSGSDARDHLLLGVEKLADAVVSTLGPGRNVMIIE-----LPYGPVKVTKDGVT
Cryptosporidium -----MTSVVVKIMDDGAQESKGE LARLQSIVGAIAGDLLKTTFGPKGMDKILQPIKEGPIDSTPIVTNDGAT
Entamoeba -----MLSSSSHYNGKLLSLNIDCRENVLSGIKKVADAVSVTLGPKGRVTIID-----QPYGNARVTKDGVS
Giardia -----MLQHYTSTVISEDARSGLLRGIKTIADVATTLGPRGRAVILADG---SASGTTKVTKDGVS
Ehrlichia -----MANMVVTGEQLDKSIREVVRILEDVAGCTAGPKGLTVAIS-----KPYGAPEVTKDGYK
Brucella -----MAAKDVKFGRTAREKMLRQVGDILADAVKVTTLGPKGRNVVIE-----KSFCAPIRITKDGVS
```

## H-protein of GCS

```
Trimastix -----MQR LFSVVPVAVGLSFLARFAGEKFFYTKDHEWVDES---GLVGISDYAQKHLGQIVYVDLPE
Homo MALRVVRSVRALLCTLRVAVPSPAAPCPPRFWQLGVGAVRTLRTGPALLSVRKFTKHEWVTTENG-IGTVGISNFAQEALGDVVYCSLPE
Dictyostelium -----MLKTLRFQTRAFQGNLNIKRNFCTRYTNDHEWVTSLSQNYRLGITDFAQKQLGDIVFEIPIQ
Arabidopsis -----MACRLFWASRVASHLRISVAQRGFSVVLKDLKYADSHWVKIDGN-KATFGITDHAQDHLGDVVYVELPD
Trypanosoma -----MRRALFCPAATVAATIRFYTRFFTDSEWVEHGDG-IATIGITAHAEQENLGDVVYVALPN
Trichomonas -----MISTLCNCSRNFTKLYAKTHEYIDVEGT-IGKIGLSEFAMKMLGKATFVDV-Q
Vibrio -----MESTLKFSSHEWIRDNGDGTVTGVDHQAQGLLDVVFDLPE
Bartonella -----MSKVYFTQDHEWLNIEGE-IATVGI THYAQQLGDLVFDLPE
Acidiphilium -----MTETRFKSDHEWVRLDGD-VATVGITDHAQSALGDVVVELPE
```

## P1-protein of GCS

```
Trimastix -----MQLNLSRAAHTASDILPYVAHTPEVRAQ
Solanum MERARKLANRAILKRLVSQSKQRSNEIPSSSLYRPSRVSSLSPYTFQARNNAKSFNTQQARSISVEALKPSDTPRRHNSATPEEQTK
Mus -----MQLCARAWGLRLGRGAGGGHRLARGTGLSWAQRSDSSGGGGGGGGDRGAAGASRLLELLPRHDDFSRRHIGPGDKDRRE
Saccharomyces -----MLRTRVTALLCRATVRSSTNYVSLARTRSFHSQSILLKTAATDITSTQYSRIFNPD LKNIDRPLDFTARRHLGSPSPDVKK
Paramecium -----MLRSIRLIKRVANF SRLANPHEVNP KIPDFEKSSDYIESRFIGSEFPQQVKE
Trypanosoma -----MYRSAFRSLRCSTVKYARWSAAGNYVNRHVGP TAAADTEV
Pyrococcus -----MAKH YIPN-SAHKEE
Legionella -----MPYIPIHTPNDTKE
Desulfovibrio -----MPYTPHTPEEIRQ
Xanthobacter -----MRYLPLTPEDRAE
```

## T-protein of GCS

```
Trimastix -----MLSKLLSLPTTGILERFSSHVATKKTTPFYDLHMKFGGDVDFCGYYLPIKIYANS--DIGIEHMNTR
Dictyostelium -----MNTSGFLKSLKASTRASNKCFSSSSTELKKTSLNQLHKELGGKMPVFCGWEMPVQYPA---GVLKEHMHCR
Chlorella MRRQAAVQLLRALGRAGSDAAASAKHSNSCAAAGLLATRGFADDASLKKTVLYDYHVANGGKMPVFCGWSEMPIQYKD---SIIDSTLHCR
Capsaspora -----MLCRSATAVAASIVARSAARSTAVRCGSTAATGPSKRTALYDFHVAKGGKMPVFCGWDMVPVQYSDI--GI IASHHHTR
Thalassiosira -----MLKSTATALLRRAKRTSILPSTSSRSLASSTNEEPLVKTSLYNLHKELGGDMVPVFCG YELPVLYKDNNGVMEKHLWCR
Vibrio -----MTQELLKTPHALHVEAGAKMPVFCG YDMPVQYKL---GVKKEHLHTR
Parvibaculum -----MSEPAVQTEILKTTPLHALHVELGAKMPVFCG YDMPVQYPD---GVLAEHLHTR
Halomonas -----MSELKQTPHALHLKLGAKMPVFCG YDMPVQYPL---GVKKEHEHTR
Oceanospirillum -----MSELSKTPLYDLHVELGGKMPVFCG YFEMPVQYPL---GVKKEHVHTR
```