Supplementary Materials

Structural and biochemical characterization of the nucleoside hydrolase from *C. elegans* reveals the role of two active site cysteine residues in catalysis

Ranjan Kumar Singh[1,2], Jan Steyaert[1,2], Wim Versées[1,2,3]

[1] Structural Biology Brussels, Vrije Universiteit Brussel (VUB), Pleinlaan 2, 1050 Brussels, Belgium

[2] VIB-VUB Center for Structural Biology, Pleinlaan 2, 1050 Brussels, Belgium

[3] Corresponding author. Structural Biology Brussels, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussel, Belgium. Tel.: +32 2 629 18 49; fax: +32 2 629 19 63. E-mail address: wim.versees@vib-vub.be

Supplementary Information: Figures and Table

Figure S1-S3

Table S1

```
C.elegans                                   α2                                   β3          η1        TT  TT

1 yeiK               36  I.VAGNQTLDKTLINGLNVCQKLE...I..........NVPVYAGMPQPIMRQ...QIVADNIHGETGLD
  YbeK               46  S.SAGNQTPEKTLRNVLRMLTLLN.R.T..........DIPVASGAVKPLMRN...LIIADNVHGESGLD
  yaaF               35  T.VAGNVSVEKTTRNALQLLHFWN...A..........EIPLAQGAAVPLVRA...PRDAASVHGESGMA
  L.major_G1         35  T.VVGNQSLEKVTQNARLVADVAG.I.V..........GVPVAAGCTKPLVRG...VRNASHIHGETGMG
  L.donovani         35  T.VVGNQTLEKVTRNARLVADVAG.I.V..........GVPVAAGCTKPLVRG...VRNASQIHGETGMG
  P.patens           50  T.IYGNVRTPTATVNALHLLEFAG.R.E..........DIPVSEGFRTSLRGEL.KERIADFVHGADGLG
  Z.mays             33  T.IFGNCTTEHATRNALILCEKAS.H.L..........EVPVAEGSHEPLKGG..KPHVADFVHGPDGLG
  S.cerevisiae       76  T.CFGNAPPENTDYNARSLLTAMG.KAQ..........AIPVYKGAQRPLVRE...PHYAPDIHGISGLD
  C.fasciculata      35  T.VVGNQTLAKVTRNAQLVADIAG.I.T..........GVPIAAGCDKPLVRK...IMTAGHIHGESGMG
  S.musiva           42  T.IYGNATLTHTTYNTRAILKAIK.R.E..........DVGVHVGASKPFCRA...AAVAADIHGESGLD
  M.tuberculosis     34  S.TGGNIAVGQVCANNLSLLELCG.A.A..........DIPVSKGADEPLGGR...WPDHPKFHGPKGIG
  S.rimosus          35  T.VVGNQTLEKVTRNALSVARIAG.I.T..........GVPFAAGCPRPLVRT...IENAPEIHGDTGLD
  A.bicolorata       35  T.VVGNQTLEKVTRNALAVAEIAG.I.V..........GVPFAAGCPRPLVRT...IEVAADIHGESGMD
  A.deanei_G1        35  T.VVGNQTLEKVSKNALYVAEVAG.M.T..........GVRFAAGCVRPLVRP...VTTAGDIHGESGLG
  M.morganii         35  T.VVGNQTLEKVTRNALAVARIAG.I.T..........GIPFAAGCPRPLIRE...IETAPDIHGDSGLD
2 T.vivax            36  C.TDADCFVENGFNVTGKIMCLMH.N.N......MNLPLFPIGKSAA.TAVNP....FPKEWRCLAKNMD
  T.brucei           36  C.TDADCFVENGFDVTGKIMCAMH.R.L......IKTPLFPIGKSTA.TAVNA....FPTEWRFSAKNLD
  C.acanthocephali   37  V.TDADCYVEQGFNVSGKVMALMH.A.Q......EGMPLFPIAKTSF.KGVNP....FPSEWRWNAKNME
  A.deanei_G2        38  V.TDADCFVEHGFSITGKLMTLMH.A.E......EGLPLFPIGKSSL.GGVNP....FPSDWRWSAKNME
  L.major_G2         37  V.TDADCFVDQAFSVSGKLMAMMH.A.R......EATELFPVAKTSF.KGVNP....FPSEWRWSAKNMD
  S.oncopelti        39  V.TDADCYKEHAFNVTGKLMTLMH.L.E......EGTSQFPVGITSF.SGVNP....FPSEWRWSAKNME
  B.saltans          56  V.LDADCYVNDAYRVSGKLTSTVR.R.AAAAKRGVHIKPFPILKSTLSHGGHP....FPHEWRVDAKKMD
  S.warneri          34  T.IGADCYLEPSLSASVKIINRFS.N.Q..........SIQVAPSYE.RGKNP....FPKEWRMHAFFMD
  F.muscicola        36  V.TPADCYVQPAVSATRKILDLMG.G.S..........HIPVAESTV.RGINP....FPYLYRRDSFVVD
  S.radiotolerans    35  V.TPADCYIEPALDSTYKLLQLME.Q.E..........HVPLGRGDY.YGINA....FPGEWRARPEIIN
  T.campylonemoides  36  V.TPADCYAEPAVSATRKILDLME.C.S..........HIPVAQSTV.RGINP....FPRLYRRDSFVVD
  C.watsonii         35  V.TPADCYINAALNVTRKLLDLMG.C.D..........PIPVAESTV.RGINP....FPALYRRDSLIID
  M.repens           36  V.TPADCYAEPAVSATRKILDLMG.C.S..........HIPVAESSV.RGINP....FPRLYRRDSFVVD
  N.punctiforme      36  I.TPADCYIQPAVSATRKILDLMG.F.S..........HISVAESTV.RGINP....FPTLYRRDSFIVD
  H.byssoidea        36  V.TPADCYAEPAVSATRKILDLME.S.S..........HIPVALSTV.RGINP....FPRLYRRDSFIVD
3 C.elegans          38  A.VHGCVTVDQACANIKRTIRAND.R.S..........NIPVYKGAAKSILSLP.KDDTVSDFFGIDGIG
  D.melanogaster     60  TMGCGNTSRENAARNMRRILDACK.R.T..........DIPIYLGAVDALIPSL..EDEKKYFHGRDGFG
  H.contortus        35  T.VHGCVKVEQATANVARVQRANLLK.K..........LIPVYMGASEALIRKP..FVMNPFFGIDGVG
  A.aegypti          61  C.THGNTDLENSVTNAARILEGLG.R.R..........DVPLYRGASERLITPAPRRDVNRYFWGVDGFG
  C.gigas            36  C.VSGNTEVDHVCRNVLRILTICD.H.I..........EIPVYRGCTKALVDHG..K.HAKFYHGLDGLC
  N.vectensis        36  C.VGGNTSLDQVCINVMKTLECCQ.R.T..........DIPVFKGAGKPLIAK..HEPSASHFHGYDGLG
  L.anatina          36  C.VNGNTMLDQVYINTLRVLKTCD.R.M..........DIPVYKGTYKSLIGHD.LSSEGQRYHGSDGLG
  A.darlingi         39  C.VHGNTDVDNVTTNVLRILTAID.R.T..........DVPVYRGASEALITPRKRKSPEECFHGMNGFG
  P.caudatus         82  I.VGGNCDIDQIVLNAVRVLKTCN.R.L..........DIPVYRGSAVSLLGE...IPQKSEYHMLDGLG
  P.formosa          34  C.THGNTLLENVLTNTLRVLKVCN.R.L..........DIPVYEGFSKPLLDR..KQ.HASDYQGKDGLG
  O.niloticus        93  C.VHGNTSVENVCKNVLRVLQACS.H.L..........EIPVFKGAAMPILGDS.GI.SSGDFHGKDGLG
  C.variegatus       34  C.CHGNTSLDNVLKNTLRVLQVCN.R.L..........DIPVYKGFSRPLLAS..KQ.HASDYHGKDGLG
  L.polyphemus       44  C.CYGNVNLNQVCQNVMRTLVVCG.K.H..........EVPVYKGCEKALVSN...ELLGSDYHSKDGFG
  O.mykiss          103  C.VHGNTNVENVCKNTLRVLQVCN.R.L..........EIPVFRGAAKPILGN..VI.SAGNFHGQDGLG
  A.mississippiensis 35  C.CYGNTQLENVCQNVLRVLQVCN.R.L..........EIPVYQGASTPLLGA..SA.KNDYFYGIDGLG
                                                                      β3-α3
```

```
C.elegans                           α9                     α10

1 yeiK              208  .AGGP..AGELFSDIMNFTLKTQF......ENYGLAGGPVHDATCIGY
  YbeK             219  .IGNP..VSTIVAELLDFFLEYHK.....DEKWGFVGAPLHDPCTIAW
  yaaF             207  .LNRT..G....KMLHALFSHYRS.......GSMQSGLRMHDLCAIAW
  L.major_G1       210  .VGTK..PAAFMLQILDFYTKVYE......KEHD.TYGKVHDPCAVAY
  L.donovani       210  .VGTK..PAAFMLQILDFYTKVYE......KERN.TYATVHDPCAVAY
  P.patens         226  .SDSK..FGKYLYAASHFYATYHR......EAYDIDAIYLHDPATMVA
  Z.mays           208  .SQGK..YAQFLCDVCKFYLDWHT......ESYGAPVIFLHDPVSFAA
  S.cerevisiae     259  KNNSK..LRELFLELFQFFAHTYK.....DMQGFESGPPIHDPVALMP
  C.fasciculata    210  .VDTN..PARFMLEIMDYYTKIYQ......SNRYMAAAAVHDPCAVAY
  S.musiva         268  TTITT..VRKLYSQILTFFAKTYA.....DVFGLTQGPPAHDPLAVAA
  M.tuberculosis   210  .GSSP..VMRVIEDALRFYFESHE......ARGHGYLAYMHDPLAAAV
  S.rimosus        209  .VGTK..PARFVLELLDFFRDAYR......ENQGFEYPPVHDPCAVAY
  A.bicolorata     209  .VGTA..PSRFVGELLEFFASTYL......ETQGFTHPPVHDPCAVAY
  A.deanei_G1      210  .VNTN..VSSFVLQLFEFYNEMYK......KKG.LTFAAVHDPCAVAY
  M.morganii       209  .VGTA..PAQFVLELLEFFGTMYK......SAQGFDYPPVHDPCAVAY
2 T.vivax          231  ...........NFLLSILVGTMWAMCTHCELLRDGDGYYAWDALTAAY
  T.brucei         231  ..........QYLLSQMVGTMWAMSTHEEILRDGDAYYAWDALTAAY
  C.acanthocephali 233  ..........EYLLSQFVGATWASCTHVELMRPGDGYYAWDVLTAAY
  A.deanei_G2      234  ..........EFLLSQFVGSAWASCTHHELMRPGDGYFAWDVLTAAY
  L.major_G2       233  ..........KYLLSQFVGSAWASCTHFELIRPGDMYYAWDVLTAAY
  S.oncopelti      234  ..........QFLLSQFVGSAWASCTHFELIRPDDCYYAWDVLTAAY
  B.saltans        270  ..........DSLIAQFIGSSWAMCTHLELGGYGKGYYAWDVLTAAH
  S.warneri        218  ..........EFVGIDFLGVSYAAVPPLTHFFTNSTYFLWDVLTTAY
  F.muscicola      222  ..........HYPISDLAGQCYALVIP.......QDYYFWDVLATAY
  S.radiotolerans  218  ..........DYRLSNLTGQFWALTIDTIPSY.QYTYFMWDILATSY
  T.campylonemoides 222 ..........QYPISDLAGQCYALVIP.......QDYYFWDVLATAY
  C.watsonii       221  ..........QYQLSDLAGLCYALAIP.......QDYYCWDVLATSY
  M.repens         222  ..........DYPISDLAGQCYALVIP.......QDYYFWDVLATAY
  N.punctiforme    222  ..........HYPISDLAGQCYALVIP.......QDYYFWDVLATAY
  H.byssoidea      222  ..........HYPISDLAGQCYALVIP.......QDYYFWDVLATAY
3 C.elegans        223  KYGTP..LANYLSLATSIGRVKCE........ANGRQYSYCDEIAVAT
  D.melanogaster   243  AARAKEAGHPAITMLNQVEAAQWL...PMIEQYGIDTWNPCDAIAVAV
  H.contortus      220  KLDTP..LASFLSLATNIGNGVMS........KNGRQYGYCDEIAVAI
  A.aegypti        242  KE....PKNKAIVVLNQVEEVIYA........NISNWQPCDMYAAAV
  C.gigas          214  NIKTS..KAIFHSTILRTSLDKHNE.......WQFPGYSFCDPTVVAI
  N.vectensis      211  SQASN..KSKFLKSISKLSADFYKQ....RKDGFGYHYVCDLLAMGV
  L.anatina        214  ERETK..KAKFVESIMRST.......RERNKNVIKRPFHSCDSLAMAV
  A.darlingi       221  GQQA..SSNRAVQMLNEIERKVYR........KWTCWMPCDAFLVAA
  P.caudatus       264  KADTR..EARFLRLIHDKPEAFCK........KQEWGFIPCDQTAMAV
  P.formosa        210  SVKTE..KAAFIRKISELSMKKARSDEYQKEVTAGQGFNPCDVFALAA
  O.niloticus      270  AQNTS..KARFMKKIFHHSYLAAQTERLQKECVAGSGFISCDSYAVAA
  C.variegatus     210  STNTK..KAAFIKKISTLSMKKARSAEYQKEVVAGKGFNPCDVFALAA
  L.polyphemus     221  KTDTP..RGYLVQKMTEHSINRARN......IPGCNGLHLCDLFAMCV
  O.mykiss         279  AQESE..KARFMARIFRHSMEESNSERFQAEMVYGMGFVSCDSYAMAA
  A.mississippiensis 211 SLDTK..KANFMKKIHAYTLKLVQSKN...PDFLSAGFVSCDSYAMAA
                                   C-ter α9
```

Figure S1. Sequence alignment of representatives of the three homology groups of nucleoside hydrolases. 15 organisms of each group are selected and aligned separately. The first 15 (starting from Yeik), second 15 (starting from *T.vivax*) and the last 15 (starting from *C.elegans*) correspond to group I, II and III NHs respectively. Residues within an orange colored box correspond to either the β3-α3 loop (loop 1) or the C-terminus of α9 (loop2). Residues highlighted in yellow correspond to C42 of CeNH. This position corresponds to an asparagine in group I, an aspartate in group II and is variable in group III NHs. Residues highlighted in green are the conserved histidine, tryptophan and cysteine for group I, II and III NHs respectively.

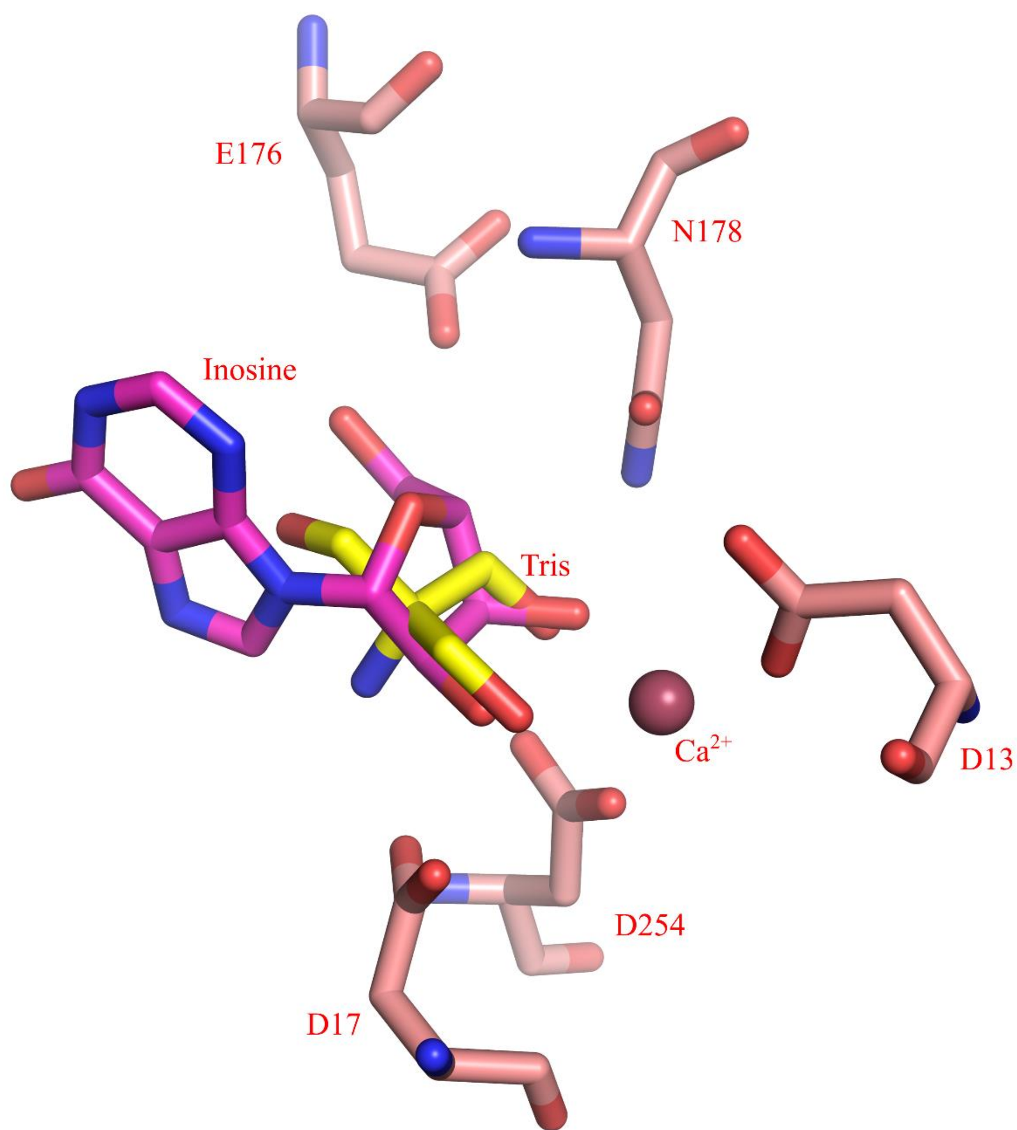

Figure S2. Overlay of the best docking conformation of inosine and the active site-bound Tris molecule observed in the CeNH structure. The residues within interaction distance of the 2', 3'and 5' hydroxyl groups of inosine and the corresponding hydroxyl groups of Tris are shown as sticks. Ca^{2+} is shown as a red sphere. The Tris and inosine molecules are shown as sticks with yellow and magenta carbons respectively.

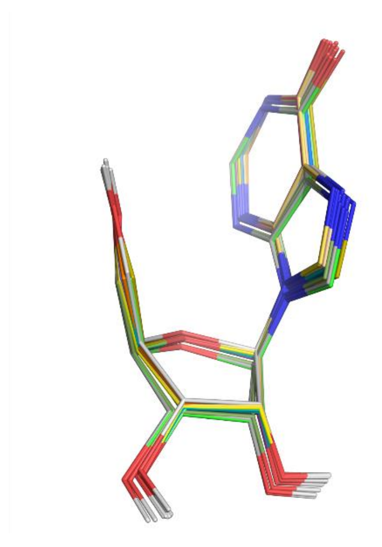

Figure S3. Superposition of 17 out of 20 minimum energy conformations of inosine docked to the CeNH active site, starting from an inosine conformation taken from the Yeik-inosine crystal structure (PDB 3B9X). 3 higher energy outliers with an entirely different conformation are rejected and are not shown here.

Table S1. SAXS data collection statistics and derived parameters.

Data-collection parameters	
X-Ray Source	BioSAXS (Rigaku)
Wavelength	1.54 Å
q range ($Å^{-1}$)*	0.0094-0.6880
Protein concentrations (mg/ml)- volume (µl)	1.5-8.0 mg/ml -70 µl
Temperature (K)	283
Structural parameters	
I(0) (relative) from P(R)	24.46 ± 0.05
R_g (Å) from P(R)	37.7 ± 0.1
I(0) (relative) from Guinier	24.56 ± 0.07
R_g (Å) from Guinier	37.7 ± 0.1
D_{max} (Å)	125 ± 5
Porod volume ($Å^3$)	277109
Molecular mass determination (kDa)	
From Porod volume (V_{porod}/1.7)	163.0
From Q_R (q=0.25 $Å^{-1}$)	135.0
From Sequence (including tag)	154.0
Data processing software	
Primary data reduction	ATSAS for BioSAXS (Rigaku)
Data Processing	PRIMUS, SCATTER, GNOM
Computation of model intensities	CRYSOL

Abbreviations: I(0), extrapolated scattering intensity at zero angle; R_g, radius of gyration calculated using either Guinier approximation (from Guinier) or the indirect Fourier transform package GNOM [from P(R)], D_{max}, maximal particle dimension; V_p, Porod volume.

*Momentum transfer $|q| = 4\pi\sin(\theta)/\lambda$.