

Supplementary Material to the study entitled

„Structure and enzymatic mechanism of a moonlighting dUTPase”

Ibolya Leveles^{a,b}, Veronika Németh^a, Judit E. Szabó^a, Veronika Harmat^c, Kinga Nyíri^{a,b}, Ábris
Ádám Bendes^{a,b}, Veronika Papp-Kádár^{a,b}, Imre Zagyva^a, Gergely Róna^a, Olivér Ozohanics^d,
Károly Vékey^d, Judit Tóth^a and Beáta G. Vértessy^{a,b}

^aInstitute of Enzymology, Research Centre for Natural Sciences, Hungarian Academy of Sciences, 29 Karolina street 1113 Budapest, Hungary, ^bDepartment of Applied Biotechnology and Food Science, Budapest University of Technology and Economics, 4 Szent Gellért tér, 1111 Budapest, Hungary, ^cHungarian Academy of Sciences-Eötvös Loránd University, Protein Modeling Research Group, and Eötvös Loránd University, Institute of Chemistry, ^dInstitute of Organic Chemistry, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Budapest, 1025, Pusztaszeri ut 59-67, Hungary

*correspondence e-mail: vertessy@mail.bme.hu

Supplementary Results: Mass spectrometry

Supplementary References

Supplementary Table S1 List of mutagenic primers

Supplementary Table S2 Spectrofluorimetric data on Φ 11 phage dUTPase : nucleoside phosphate complexes

Supplementary Figure S1 Mass spectrum of Φ 11 phage dUTPase under native electrospray conditions.

Supplementary Figure S2 Phylogenetic sequence alignment of staphylococcal phage dUTPases

Supplementary Figure S3 Phylogenetic tree of staphylococcal phage dUTPases

Supplementary Results: Mass spectrometry

Electrospray mass spectrometry has been shown to be capable of providing relevant information of oligomerization (protein complexation) characteristics while also having the great advantage of not depending on molecular shape and hydrodynamic properties of the macromolecules (Grandori *et al.*, 2009, Heck, 2008, Ngounou Wetie *et al.*, 2013, Benesch *et al.*, 2007). We wished to apply this technique to learn if it may be a convenient tool for characterizing trimeric dUTPases.

The electrospray mass spectrum of Φ 11 phage dUTPase were measured on protein samples dissolved in volatile aqueous buffer (10 mM NH_4HCO_3 , pH 7.8) to approximate 'native' conditions while also allowing mass spectral measurements. Spectra presented in Fig. 2 clearly show the presence of an abundant species in the 3000-5000 m/z range. Charge distribution is narrow (mostly 16+, 17+ and 18+ charge states), which supports that the native protein conformation does survive transfer into the mass spectrometer. Mass difference between various charge states suggests that these are mostly ammonium ion adducts. Molecular mass of this species is 59,500 (\pm 100) Da, corresponding to a non-covalent trimeric form of dUTPase, as based on the monomer molecular mass of 19,853, calculated from the amino acid sequence (cf Materials and Methods). Dimers, tetramers or other oligomers are not detected (their abundance is less, or much less than 5% that of the trimer). Absence of these oligomers confirms that the trimer is not an artifact (non-specific aggregate). In the low mass range (see insert in Fig. 2), the protonated monomer species is observed with a molecular mass 19,854 (\pm 2) Da, in agreement with the amino acid sequence. Based on relative peak areas, monomer signals are only 5-10% that of the trimer, indicating high stability for the latter. In the low mass range signals corresponding to two impurities (molecular mass 19,154 and 17,425) are also observed (peaks indicated by "A" and "B" in Fig. 2). Trimers (or other oligomers) are observed only for the Φ 11 phage dUTPase monomer, and not for these impurities. This, again, supports that Φ 11 phage dUTPase is predominantly present in the solution as a trimer; and supports the conclusion that the observed trimer is not due to non-specific aggregation.

Supplementary References

- Benesch, J. L. P., Ruotolo, B. T., Simmons, D. A. & Robinson, C. V. (2007). *Chemical Reviews* 107, 3544-3567.
- Grandori, R., Santambrogio, C., Brocca, S., Invernizzi, G. & Lotti, M. (2009). *Biotechnology Journal* 4, 73-87.
- Heck, A. J. R. (2008). *Nature Methods* 5, 927-933.
- Ngounou Wetie, A. G., Sokolowska, I., Woods, A. G., Roy, U., Loo, J. A. & Darie, C. C. (2013). *PROTEOMICS* 13, 538-557.

Supplementary Table S1 List of primers used in site-directed mutagenesis experiments

Mutant	Primers
$\Phi 11\text{DUT}^{\text{F108W}}$	forward: 5' CTATATTACCCCGGGCGTGTGGGATATTAAAGGCGAAATTGATC3' reverse: 5' GATCAATTTTCGCCTTTAATATCCACACGCCCCGGGGTAATATAG3'
$\Phi 11\text{DUT}^{\text{F164W}}$	forward: 5' GCGAACGCGGCGAAAAAGGCTGGGGCAGCAGCGGCGTG3' reverse: 5' CACGCCGCTGCTGCCCCAGCCTTTTTTCGCCGCTTCGC3'
$\Phi 11\text{DUT}^{\text{E158STOP}}$	forward: 5' GAATTTGAAAGCGTGAGCTAACGCGGCGAAAAAGGC3' reverse: 5' GCCTTTTTTCGCCGCTTAGCTCACGCTTTCAAATTC3'
$\Phi 11\text{DUT}^{\Delta 101-122}$	forward: 5' GCGATTGCGAGCAACTATGGCACCTATCAGATTAACGAAG3' reverse: 5' CTTCTGTTAATCTGATAGGTGCCATAGTTGCTCGCAATCGC3'

Supplementary Table S2 Fluorescence spectral characteristics of dUTPases containing Trp residues in the apoenzyme and upon ligand binding

	hDUT ^{F158W}	$\Phi 11\text{DUT}^{\text{WT}}$	$\Phi 11\text{DUT}^{\text{F108W}}$	$\Phi 11\text{DUT}^{\text{F164W}}$
$F_{\text{max}}^{\text{dUMP}}$	0.64 ^a	1	0.96	0.897
$F_{\text{max}}^{\text{dUTP}}$	0.2 ^a	0.97	1	0.67
$F_{\text{max}}^{\text{dUPNPP}}$	0.4 ^a	0.94	0.94	0.64
$\lambda_{\text{max}}^{\text{apo}}$ (nm)	353 ^a	342	346	347
$\lambda_{\text{max}}^{\text{dUMP}}$ (nm)	347 ^a	342	346	347
$\lambda_{\text{max}}^{\text{dUTP}}$ (nm)	339 ^a	342	346	344
$\lambda_{\text{max}}^{\text{dUPNPP}}$ (nm)	343 ^a	343	346	342

^a Published in (Tóth 2007 JBC)

F_{max} values indicate maximal fluorescence intensities observed upon saturation with the corresponding ligand, as compared to the fluorescence intensity of the apoenzyme, λ_{max} values indicate the wavelength of maximum fluorescence in the emission spectrum. Note fluorescence intensity changes and blue-shift of emission maximum in the $\Phi 11\text{DUT}^{\text{F164W}}$ and the corresponding human dUTPase F158W construct, but not in the other $\Phi 11$ dUTPases.

Supplementary Figure Legends

Supplementary Figure S1 Mass spectrum of Φ 11 phage dUTPase under native electrospray conditions.

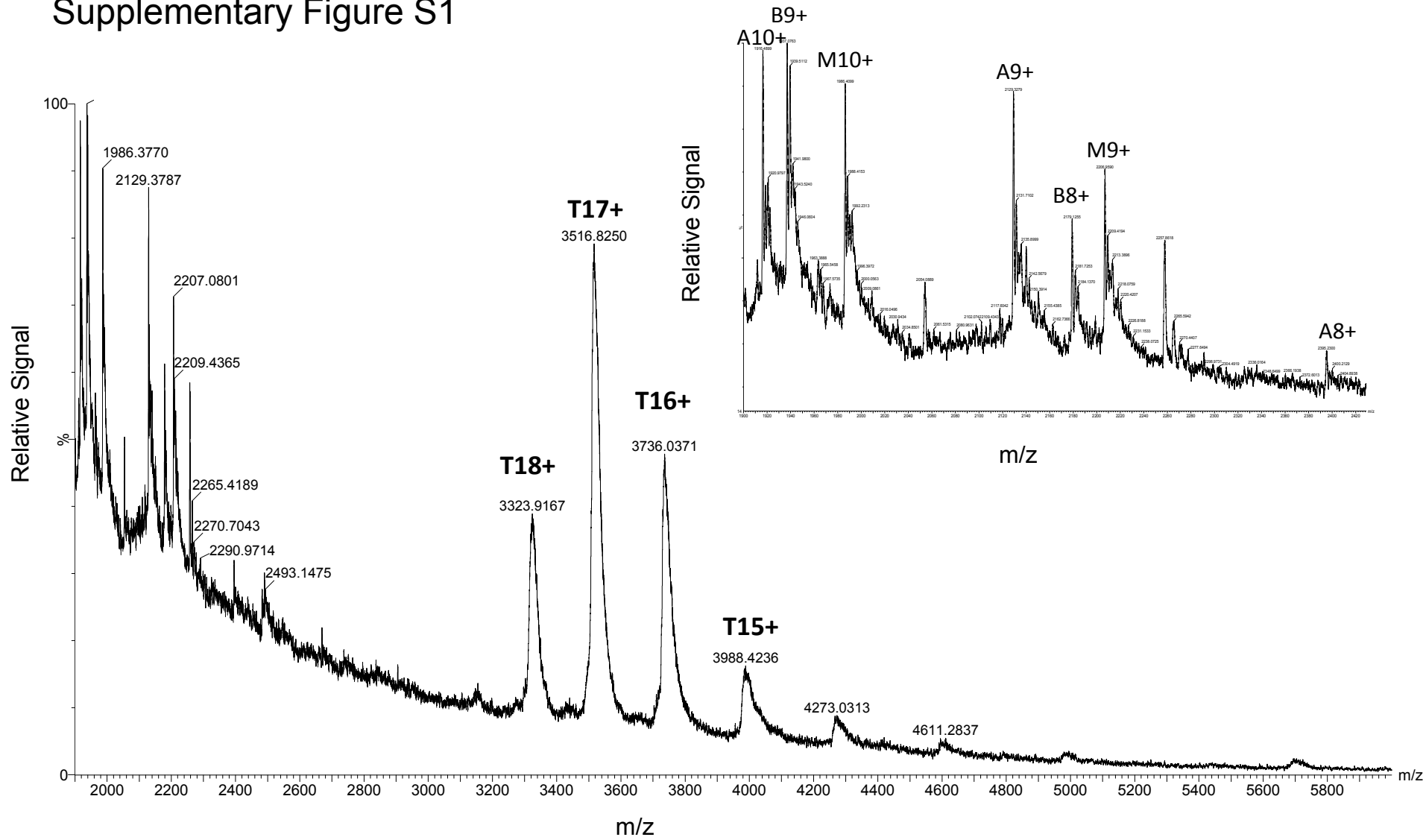
T17+, T16+ etc indicate trimer signals and 17+, 16+ etc charge states. Molecular mass of this species is determined to be 59,560 Da, corresponding to a trimer of the monomeric molecular mass of 19,854 Da. **Inset:** Monomer signals (M) and two small protein impurities (A and B) are also indicated; the numbers (e.g. A9+) indicate charge states. Molecular masses determined from the spectrum are 19,854, 19,154 and 17,425, for the dUTPase (M) and the two minor impurities (A and B), respectively.

Supplementary Figure S2 Phylogenetic alignment of staphylococcal phage dUTPases

Supplementary Figure S3 Phylogenetic tree of staphylococcal phage dUTPases

Phage dUTPase sequences were analyzed by the ClustalX 2.1 with the neighbor-joining algorithm (cf Supplementary Fig. S2). Clusters of similar sequences are shown in different colours.

Supplementary Figure S1



	I	II	III
SLT	-----MTNTLQVKLLSENARMPERNHKT	DAGYDIFSAETVVLEPQEKAVIKTDVAVSI	PEGYVGLLTSRSGVSSKTHLVIETGKIDAGYHGN
tp310-3	-----MTNTLQVKLLSENARMPERNHKT	DAGYDIFSAETVVLEPQEKAVIKTDVAVSI	PEGYVGLLTSRSGVSSKTHLVIETGKIDAGYHGN
80	-----MTNTLQVKLLSENARMPERNHKT	DAGYDIFSAETVVLEPQEKAVIKTDVAVSI	PEGYVGLLTSRSGVSSKTHLVIETGKIDAGYHGN
tp310-2	-----MTNTLQVKLLSENARMPERNHKT	DAGYDIFSAETVVLEPQEKAVIKTDVAVSI	PEGYVGLLTSRSGVSSKTHLVIETGKIDAGYHGN
tp310-1	-----MRRSRKMTNTLQVKLLSENARMPERNHKT	DAGYDIFSAETVVLEPQEKAVIKTDVAVSI	PEGYVGLLTSRSGVSSKTHLVIETGKIDAGYHGN
PVL108	-----MRRSRKMTNTLQVKLLSENARMPERNHKT	DAGYDIFSAETVVLEPQEKAVIKTDVAVSI	PEGYVGLLTSRSGVSSKTHLVIETGKIDAGYHGN
PVL-CN125	-----MRRSRKVTNTLQVKLLSENARMPERNHKT	DAGYDIFSAETVVLEPQEKAVIKTDVAVSI	ILEGYVGLLTSRSGVSSKTHLVIETGKIDAGYHGN
PVL	-----MRRSRKVTNTLQVKLLSENARMPERNHKT	DAGYDIFSAETVVLEPQEKAVIKTDVAVSI	PEGYVGLLTSRSGVSSKTHLVIETGKIDAGYHGN
42E	-----MRRSRKVTNTLQVKLLSENARMPERNHKT	DAGYDIFSAETVVLEPQEKAVIKTDVAVSI	PEGYVGLLTSRSGVSSKTHLVIETGKIDAGYHGN
52A	-----MRRSRKVTNTLQVKLLSENARMPERNHKT	DAGYDIFSAETVVLEPQEKAVIKTDVAVSI	PEGYVGLLTSRSGVSSKTHLVIETGKIDAGYHGN
29	-----MTNTLQVKLLSENARMPERNHKT	DAGYDIFSAETVVLEPQEKAVIKTDVAVSI	PEGYVGLLTSRSGVSSKTHLVIETGKIDAGYHGN
MR11	-----MTNTLQVKLLSENARMPERNHKT	DAGYDIFSAETVVLEPQEKAVIKTDVAVSI	PEGYVGLLTSRSGVSSKTHLVIETGKIDAGYHGN
71	-----MTNTLQVKLLSENARMPERNHKT	DAGYDIFSAENVVLEPQDKTVIKTDVAVSI	PEGYVGLLTSRSGVSSKTHLVIETGKIDAGYHGN
Φ11	-----MTNTLQVRLLENARMPERNHKT	DAGYDIFSAETVVLEPQEKAVIKTDVAVSI	PEGYVGLLTSRSGVSSKTHLVIETGKIDAGYHGN
Sa2usa	-----MTNILQVKLLSENARMPERNHKT	DAGYDIFSAETVVLEPQEKAVIKTDVAVSI	PEGYVGLLTSRSGVSSKTHLVIETGKIDAGYHGN
Sa3usa	-----MTNILQVKLLSENARMPERNHKT	DAGYDIFSAETVVLEPQEKAVIKTDVAVSI	PEGYVGLLTSRSGVSSKTHLVIETGKIDAGYHGN
Mu50A	MNWLELMRRTRKMTNILQVKLLSENARMPERNHKT	DAGYDIFSAETVVLEPQEKAVIKTDVAVSI	PEGYVGLLTSRSGVSSKTHLVIETGKIDAGYHGN
Mu50B	MNWLELMRRTRKMTNILQVKLLSENARMPERNHKT	DAGYDIFSAETVVLEPQEKAVIKTDVAVSI	PEGYVGLLTSRSGVSSKTHLVIETGKIDAGYHGN
Avβ	-----MTNTLQVRLLESETARMPERNHKT	DAGYDIFSAETVVLEPQEKAVIKTDVAVSI	PEGYVGLLTSRSGVSSKTHLVIETGKIDAGYHGN
Av1	-----MTNTLQVRLLESETARMPERNHKT	DAGYDIFSAETVVLEPQEKAVIKTDVAVSI	PEGYVGLLTSRSGVSSKTHLVIETGKIDAGYHGN
3A	-----MTNILQVKLLSKDARMPERNHKT	DAGYDIFSAKTVVLEPQEKAVIKTDVAVSI	PEGYVGLLTSRSGVSSKTHLVIETGKIDAGYHGN
ETA2	-----MTNTLQVKLLSENARMPERNHKT	DAGYDIFSAESVVLEPQEKAVIKTDVAVSI	PEGYVGLLTSRSGVSSKTYLVIETGKIDAGYHGN
187	-----MTNTLQVKLLSENARMPERNHKT	DAGYDIFSAETVVLEPQEKAVIKTDVAVSI	SEGYVGLLTSRSGVSSKTHLVIETGKIDAGYHGN
80α	-----MTNTLQVKLLSKNARMPERNHKT	DAGYDIFSAETVVLEPQEKAVIKTDVAVSI	PEGYVGLLTSRSGVSSKTHLVIETGKIDAGYHGN
53	-----MTNTLQVKLLSKNARMPERNHKT	DAGYDIFSAETVVLEPQEKAVIKTDVAVSI	PEGYVGLLTSRSGVSSKTHLVIETGKIDAGYHGN
NM3	-----MTNTLQVRLLENARMPERNHKT	DAGYDIFSAETVVLEPQEKAVIKTDVAVSI	PEGYVGLLTSRSGVSSKTHLVIETGKIDAGYHGN
Sa2mr252A	-----MTNTLQVRLLENARMPERNHKT	DAGYDIFSAETVVLEPQEKAVIKTDVAVSI	PEGYVGLLTSRSGVSSKTHLVIETGKIDAGYHGN
Sa3mr252B	-----MTNTLQVRLLENARMPERNHKT	DAGYDIFSAETVVLEPQEKAVIKTDVAVSI	PEGYVGLLTSRSGVSSKTHLVIETGKIDAGYHGN
47	-----MTNTLQVKLLSKNARMPERNHKT	DAGYDIFSAETVVLEPQEKAVIKTDVAVSI	PEGYVGLLTSRSGVSSKTYLVIETGKIDAGYHGN
77	-----MTNTLQVKLLSKNARMPERNHKT	DAGYDIFSAETVVLEPQEKAVIKTDVAVSI	PEGYVGLLTSRSGVSSKTYLVIETGKIDAGYHGN
85	-----MTNTLQVKLLSKNARMPERNHKT	DAGYDIFSAETVVLEPQEKAVIKTDVAVSI	PEGYVGLLTSRSGVSSKTYLVIETGKIDAGYHGN
COL	-----MTNTLQVKLLSKNARMPERNHKT	DAGYDIFSAETVVLEPQEKAVIKTDVAVSI	PEGYVGLLTSRSGVSSKTYLVIETGKIDAGYHGN
12	-----MRRNRKMTNTLQVKLLSKNARMPERNHKT	DAGYDIFSAETVVLEPQEKAVIKTDVAVSI	PEGYVGLLTSRSGVSSKTYLVIETGKIDAGYHGN
ROSA	-----MTNTLQVKLLSENARMPERNHKT	DAGYDIFSAETVVLEPQEKAVIKTDVAVSI	PEGYVGLLTSRSGVSSKTHLVIETGKIDAGYHGN
P954	-----MTNTLQVKLLSKDARMPERNHKT	DAGYDIFSAETVVLEPQEKAVIKTDVAVSI	PEGYVGLLTSRSGVSSKTYLVIETGKIDAGYHGN

Phage specific linker region

	IV	V
SLT	LGINIKND DAQV ----- YLT TNEQC FDIQGEMEN-SFVNNAK KK PF INDYYE IYK GD	KLAQLVIVPIWTP PEL KQVEEFESV SER GAKFGSSG V
tp310-3	LGINIKND DAQV ----- YLT TNEQC FDIQGEMEN-SFVNNAK KK PF INDYYE IYK GD	KLAQLVIVPIWTP PEL KQVEEFESV SER GAKFGSSG V
80	LGINIKND DAQV ----- YLT TNEQC FDIQGEMEN-SFVNNAK KK PF INDYYE IYK GD	KLAQLVIVPIWTP PEL KQVEEFESV SER GAKFGSSG V
tp310-2	LGINIKND DAQV ----- YLT TNEQC FDIQGEMEN-SFVNNAK KK PF INDYYE IYK GD	KLAQLVIVPIWTP PEL KQVEEFESV SER GAKFGSSG V
tp310-1	LGINIKND DAQV ----- YLT TNEQC FDIQGEMEN-SFVNNAK KK PF INDYYE IYK GD	KLAQLVIVPIWTP PEL KQVEEFESV SER GAKFGSSG V
PVL108	LGINIKND DAQV ----- YLT TNEQC FDIQGEMEN-SFVNNAK KK PF INDYYE IYK GD	KLAQLVIVPIWTP PEL KQVEEFESV SER GAKFGSSG V
PVL-CN125	LGINIKND DAIAS ----- NGY ITP-- GVF DIK GEID ---- LSDA IRQ--- YGT YQ INE GD	KLAQLVIVPIWTP PEL KQVEEFESV SER GAKFGSSG V
PVL	LGINIKND DAIAS ----- NGY ITP-- GVF DIK GEID ---- LSDA IRQ--- YGT YQ INE GD	KLAQLVIVPIWTP PEL KQVEEFESV SER GAKFGSSG V
42E	LGINIKND DAIAS ----- NGY ITP-- GVF DIK GEID ---- LSDA IRQ--- YGT YQ INE GD	KLAQLVIVPIWTP PEL KQVEEFESV SER GAKFGSSG V
52A	LGINIKND DAIAS ----- NGY ITP-- GVF DIK GEID ---- LSDA IRQ--- YGT YQ INE GD	KLAQLVIVPIWTP PEL KQVEEFESV SER GAKFGSSG V
29	LGINIKND DAIAS ----- NGY ITP-- GVF DIK GEID ---- LSDA IRQ--- YGT YQ INE GD	KLAQLVIVPIWTP PEL KQVEEFESV SER GAKFGSSG V
MR11	LGINIKND DAIAS ----- NGY ITP-- GVF DIK GEID ---- LSDA IRQ--- YGT YQ INE GD	KLAQLVIVPIWTP PEL KQVEEFESV SER GAKFGSSG V
71	LGINIKND DAIAS ----- NGY ITP-- GVF DIK GEID ---- LSDA IRQ--- YGT YQ INE GD	KLAQLVIVPIWTP PEL KQVEEFESV SER GAKFGSSG V
Φ11	LGINIKND DAIAS ----- NGY ITP-- GVF DIK GEID ---- LSDA IRQ--- YGT YQ INE GD	KLAQLVIVPIWTP PEL KQVEEFESV SER GAKFGSSG V
Sa2usa	LGINIKND NET ---- L-E SE DM S-- NFGR SPAGIDGKYAR LPVTDKILCMNGSYVINK GD	KLAQLVIVPIWTP PEL KQVEEFESV SER GAKFGSSG V
Sa3usa	LGINIKND NET ---- L-E SE DM S-- NFGR SPAGIDGKYAR LPVTDKILCMNGSYVINK GD	KLAQLVIVPIWTP PEL KQVEEFESV SER GAKFGSSG V
Mu50A	LGINIKND NET ---- L-E SE DM S-- NFGR SPAGIDGKYAR LPVTDKILCMNGSYVINK GD	KLAQLVIVPIWTP PEL KQVEEFESV SER GAKFGSSG V
Mu50B	LGINIKND NET ---- L-E SE DM S-- NFGR SPAGIDGKYAR LPVTDKILCMNGSYVINK GD	KLAQLVIVPIWTP PEL KQVEEFESV SER GAKFGSSG V
Avβ	LGINIKND NET ---- L-E SE DM S-- NFGR SPAGIDGKYAR LPVTDKILCMNGSYVINK GD	KLAQLVIVPIWTP PEL KQVEEFESV SER GAKFGSSG V
Av1	LGINIKND NET ---- L-E SE DM S-- NFGR SPAGIDGKYAR LPVTDKILCMNGSYVINK GD	KLAQLVIVPIWTP PEL KQVEEFESV SER GAKFGSSG V
3A	LGINIKND NET ---- L-E SE DM S-- NFGR SPAGIDGKYAR LPVTDKILCMNGSYVINK GD	KLAQLVIVPIWTP PEL KQVEEFESV SER GAKFGSSG V
ETA2	LGINIKND DIET ---- L-EI W DD G-- NFSR NVAGIDGKYAP PPVTDKILFMNGSYVINK GD	KLAQLVIVPIWTP PEL KQVEEFESV SER GAKFGSSG V
187	LGINIKNE NET ---- L-EN W V TY-- NFSR NVAGIDGKYAP PPVTDKILFMNGSYVINK GD	KLAQLVIVPIWTP PEL KQVEEFESV SER GAKFGSSG V
80α	LGINIKND HE ----- DDK MQT-- I FLRN--- ID-NEKI FEKERHLY KLGSYRIEK GER	IAQLVIVPIWTP PEL KQVEEFESV SER GAKFGSSG V
53	LGINIKND HE ----- DDK MQT-- I FLRN--- ID-NEKI FEKERHLY KLGSYRIEK GER	IAQLVIVPIWTP PEL KQVEEFESV SER GAKFGSSG V
NM3	LGINIKND DEERD GIPFLYDD DAELEDGLISILD IKGN YVDGRG ---- IRRI YQ INK GD	KLAQLVIVPIWTP PEL KQVEEFESV SER GAKFGSSG V
Sa2mr252A	LGINIKND DEERD GIPFLYDD DAELEDGLISILD IKGN YVDGRG ---- IRRI YQ INK GD	KLAQLVIVPIWTP PEL KQVEEFESV SER GAKFGSSG V
Sa3mr252B	LGINIKND DEERD GIPFLYDD DAELEDGLISILD IKGN YVDGRG ---- IRRI YQ INK GD	KLAQLVIVPIWTP PEL KQVEEFESV SER GAKFGSSG V
47	LGINIKND DEERD GIPFLYDD DAELEDGLISILD IKGN YVDGRG ---- IRRV YQ INK GD	KLAQLVIVPIWTP PEL KQVEEFESV SER GAKFGS G
77	LGINIKND DEERD GIPFLYDD DAELEDGLISILD IKGN YVDGRG ---- IRRV YQ INK GD	KLAQLVIVPIWTP PEL KQVEEFESV SER GAKFGS G
85	LGINIKND DEERD GIPFLYDD DAELEDGLISILD IKGN YVDGRG ---- IRRV YQ INK GD	KLAQLVIVPIWTP PEL KQVEEFESV SER GAKFGS G
COL	LGINIKND DEERD GIPFLYDD DAELEDGLISILD IKGN YVDGRG ---- IRRV YQ INK GD	KLAQLVIVPIWTP PEL KQVEEFESV SER GAKFGSSG V
12	LGINIKND DEERD GIPFLYDD DAELEDGLISILD IKGN YVDGRG ---- IRRV YQ INK GD	KLAQLVIVPIWTP PEL KQVEEFESV SER GAKFGSSG V
ROSA	LGINIKND MEHDG ITSLYED LD ---- DKL VNT LD IKGN YINEGEG ---- ARK VY KINK GD	KLAQLVIVPIWTP PEL KQVEEFESV SER GAKFGSSG V
P954	LGINIKND MEHDG ITSLYED LD ---- DKL VNT LD IKGN YINEGEG ---- ARK VY KINK GD	KLAQLVIVPIWTP PEL KQVEEFESV SER GAKFGSSG V

Supplementary Figure S3

