

# NLS copy-number variation governs efficiency of nuclear import – case study on dUTPases

Gergely Róna<sup>1,2</sup>, Hajnalka L. Pálincás<sup>1,3</sup>, Máté Borsos<sup>1</sup>, András Horváth<sup>1</sup>, Ildikó Scheer<sup>1,2</sup>, András Benedek<sup>1,2</sup>, Gergely N. Nagy<sup>1,2</sup>, Imre Zagyva<sup>1</sup> and Beáta G. Vértessy<sup>1,2</sup>

1 Institute of Enzymology, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Budapest, Hungary

2 Department of Applied Biotechnology and Food Sciences, Budapest University of Technology and Economics, Hungary

3 Doctoral School of Multidisciplinary Medical Science, University of Szeged, Hungary

## Keywords

dUTPase; importin; nuclear localization signal; nucleocytoplasmic trafficking; oligomer proteins; regulation; stoichiometry

## Correspondence

G. Róna and B. G. Vértessy, Department of Applied Biotechnology and Food Sciences, Budapest University of Technology and Economics, Szt Gellért tér 4, H-1111, Budapest, Hungary  
Fax: +36 14665465  
Tel: +36 14631401  
E-mails: rona.gergely@ttk.mta.hu; vertessy.beata@ttk.mta.hu

(Received 14 December 2013, revised 10 September 2014, accepted 29 September 2014)

doi:10.1111/febs.13086

Nucleocytoplasmic trafficking of large macromolecules requires an active transport machinery. In many cases, this is initiated by binding of the nuclear localization signal (NLS) peptide of cargo proteins to importin- $\alpha$  molecules. Fine orchestration of nucleocytoplasmic trafficking is of particularly high importance for proteins involved in maintenance of genome integrity, such as dUTPases, which are responsible for prevention of uracil incorporation into the genome. In most eukaryotes, dUTPases have two homotrimeric isoforms: one of these contains three NLSs and is present in the cell nucleus, while the other is located in the cytoplasm or the mitochondria. Here we focus on the unusual occurrence of a pseudo-heterotrimeric dUTPase in *Drosophila virilis* that contains one NLS, and investigate its localization pattern compared to the homotrimeric dUTPase isoforms of *Drosophila melanogaster*. Although the interaction of individual NLSs with importin- $\alpha$  has been well characterized, the question of how multiple NLSs of oligomeric cargo proteins affect their trafficking has been less frequently addressed in adequate detail. Using the *D. virilis* dUTPase as a fully relevant physiologically occurring model protein, we show that NLS copy number influences the efficiency of nuclear import in both insect and mammalian cell lines, as well as in *D. melanogaster* and *D. virilis* tissues. Biophysical data indicate that NLS copy number determines the stoichiometry of complexation between importin- $\alpha$  and dUTPases. The main conclusion of our study is that, in *D. virilis*, a single dUTPase isoform efficiently reproduces the cellular dUTPase distribution pattern that requires two isoforms in *D. melanogaster*.

## Structured digital abstract

- [ABC dUTPase](#) and [importin- \$\alpha\$  bind](#) by [molecular sieving](#) ([View interaction](#))
- [AAA dUTPase](#) and [importin- \$\alpha\$  bind](#) by [molecular sieving](#) ([View interaction](#))
- [ABC dUTPase](#) and [importin- \$\alpha\$  bind](#) by [comigration in non denaturing gel electrophoresis](#) ([View interaction](#))
- [AAA dUTPase](#) and [importin- \$\alpha\$  bind](#) by [comigration in non denaturing gel electrophoresis](#) ([View interaction](#))
- [ABC dUTPase](#) and [importin- \$\alpha\$  bind](#) by [isothermal titration calorimetry](#) ([View interaction](#))
- [AAA dUTPase](#) and [importin- \$\alpha\$  bind](#) by [isothermal titration calorimetry](#) ([View interaction](#))

## Abbreviation

NLS, nuclear localization signal.

## Introduction

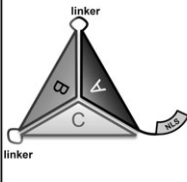



Nucleocytoplasmic trafficking in eukaryotes is an efficient means of shaping the macromolecular composition of the various cellular compartments [1,2]. Multiple levels of regulation through a set of karyopherin molecules and the RanGTP/RanGDP gradient across the nuclear envelope, as well as numerous post-translational modifications, contribute to directional transport of macromolecules and their complexes to and from the nucleus [3–6]. For import into the nucleus, proteins above a certain size (associated with a relative molecular mass > 40 kDa) usually possess short polypeptide segments, termed nuclear localization signals (NLSs), that govern their attachment to the import apparatus of karyopherins [7,8]. NLSs are usually found within flexible regions of proteins, and form specific classes of short linear recognition motifs. Complexation, i.e. non-covalent attachment of several protein molecules to each other, leads to formation of either homo- or hetero-oligomers. A subunit lacking an NLS within an oligomer may exploit the capability of an NLS present on another subunit, and hence hijack its way into the nucleus (piggyback transport) [9–11]. In homo-oligomers, in which each subunit contains one NLS, either all or some of these NLSs may be functional, thus adding another level of regulation of transport processes. This issue has not yet been covered in detail in the literature, although several sporadic observations have been reported [12–14]. In this study, our starting hypothesis was that the copy number and functionality of the NLSs in an oligomer protein may possess a regulatory capacity for efficiency of nuclear import.

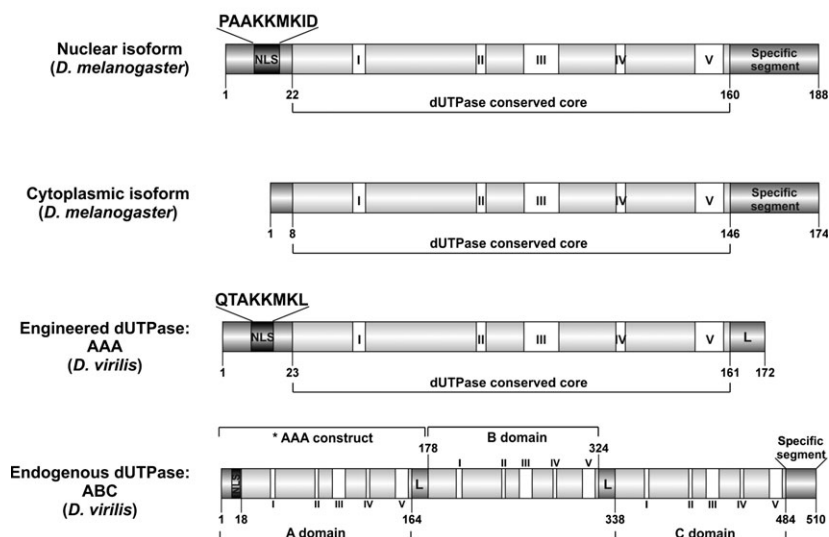
The structural basis for NLS recognition by importin- $\alpha$  is well understood [8,15]. The classical import pathway, conserved throughout eukaryotic evolution, involves accommodation of the NLS of the cargo protein into the binding groove of importin- $\alpha$  [7]. Adaptor molecules, such as importin- $\alpha$ , further recruit importin- $\beta$  via their N-terminal importin  $\beta$ -binding domain, and the resulting cargo/importin- $\alpha$ /importin- $\beta$  complex passes through the nuclear pore. Inside the nucleus, the complex dissociates upon interaction with RanGTP [16]. The cargo protein remains in the nucleus, and the karyopherins are recycled for further transport events. In an oligomeric protein with multiple NLSs, each NLS may recruit one importin- $\alpha$  molecule with a 1 : 1 NLS/importin- $\alpha$  stoichiometry. The binding affinity of the NLS for importin- $\alpha$  correlates with the steady-state nuclear accumulation of the cargo [17].

In the present work, we exploit the unusual characteristics of *Drosophila* dUTPases to evaluate the

regulatory potential of NLS copy number in nuclear import. In humans and *Drosophila melanogaster*, two dUTPase isoforms are present: one of these is nuclear, while the other is either mitochondrial (human) or cytoplasmic (*D. melanogaster*) [18–23]. dUTPases are ubiquitous and essential enzymes that are responsible for the prevention of uracil incorporation into DNA. These enzymes catalyze the hydrolysis of dUTP into dUMP and pyrophosphate, thereby sanitizing the nucleotide pool and providing the dUMP precursor for *de novo* dTTP synthesis [24,25]. These functions contribute to the role of these enzymes in genome integrity; the subcellular localization of dUTPase isoforms in nuclei and mitochondria is consistent with this role. The exact physiological role of the cytoplasmic isoform is not yet understood. Both isoforms form homotrimers, which is the characteristic and evolutionarily conserved organization in this dUTPase family [24–34]. In *Drosophila melanogaster* the nuclear isoform contains one NLS/subunit, i.e. three NLSs in one trimer, whereas the other isoform, generated by alternative splicing, does not contain any NLSs [19,23]. In *Drosophila virilis*, however, genome annotations predict a rather unique arrangement for dUTPase: three copies of the monomeric dUTPase gene segments are joined together so that the resulting predicted protein product is a pseudo-heterotrimer (Fig. 1) in which the three dUTPase domains are only slightly different (87.76% identity) and are connected through linker peptide regions [35]. This predicted protein possesses only one NLS, and has an estimated molecular mass of ~ 55.7 kDa. In the present study, we provide strong experimental evidence for the existence of this pseudo-heterotrimeric dUTPase both at the mRNA and protein levels. Hence, we termed this physiologically occurring dUTPase trimer with one NLS a ‘pseudo-heterotrimer’. This trimer is particularly useful in order to analyze the effect of NLS copy number variations. Our experimental data obtained using *in vitro* reconstituted systems of *Drosophila* dUTPases and importin- $\alpha$ , as well as immunocytochemistry on various eukaryotic cell lines and in *D. melanogaster* and *D. virilis* tissues, indicate that transport efficiency into the nucleus depends on the number of NLSs, and each NLS is capable of binding one importin- $\alpha$  molecule. We also show that, in *D. virilis* tissues, the pseudo-heterotrimeric dUTPase with one NLS is found in both the nucleus and the cytoplasm. Thus, the *D. virilis* dUTPase may efficiently provide dUTPase function in both the nucleus and the cytoplasm, resulting in a similar cellular distribution to the *D. melanogaster* dUTPase pool, which is achieved through two dedicated isoforms.

**Fig. 1.** Schematic representation of the dUTPase constructs used in this study. *D. virilis* is predicted to possess a unique dUTPase, consisting of three covalently linked dUTPase domains, forming a pseudo-heterotrimer (termed 'ABC') and containing only one NLS, which is located at the N-terminus, as an extension to the 'A' dUTPase domain. Using the sequence for the 'A' domain, an artificial dUTPase (expected to form a dUTPase homotrimer termed 'AAA') was engineered, which harbors three NLSs. *D. melanogaster* has two homotrimeric dUTPase isoforms arising from alternative splicing: a shorter, cytoplasmic isoform (termed 'Cyto') that lacks an NLS, and a longer, nuclear isoform ('Nuc') containing NLSs on all three subunits. The schematics also show the position and sequence of the NLSs together with the five characteristic conserved motifs of dUTPases (indicated by white boxes with Roman numerals). All *Drosophila* dUTPases possess a specific C-terminal end that is not present in other species. The linker region linking the ABC pseudo-heterotrimer is indicated by an 'L'.

| Drosophila dUTPase proteins   |  |   |   |
|---|--|---|---|
| Pseudo-heterotrimeric   | Homotrimeric   |   |   |
| „ABC”<br>Pseudo-heterotrimer  | „AAA”<br>homotrimer  | Cytoplasmic isoform<br>(„Cyto”)   | Nuclear isoform<br>(„Nuc”)  |
|  |  |  |  |
| Source  |  |   |   |
| Drosophila virilis  | Engineered construct   | Drosophila melanogaster   |   |
| Number of NLS/trimer  |  |   |   |
| 1   | 3  | 0   | 3   |

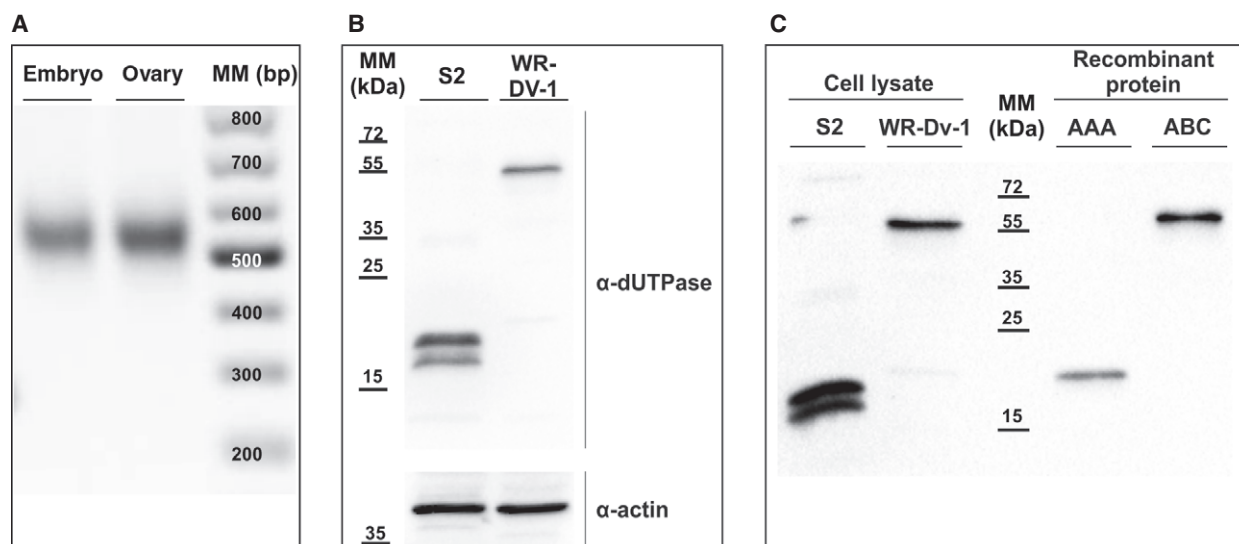


## Results and Discussion

### *D. virilis* cells contain a single dUTPase isoform, corresponding to the ABC pseudo-heterotrimer

In *D. virilis*, genome annotations indicated that only one dUTPase isoform may be present (gene symbol: Dvir\_GJ10455, gene ID: 6627974), in which three copies of the dUTPase subunit are covalently linked in one continuous polypeptide (termed ABC pseudo-heterotrimer) with one NLS (Fig. 1, see also Fig. S1 for sequence alignments). Figure 1 shows a comparison of the domain arrangement of the dUTPases from *D. melanogaster* and *D. virilis*, indicating the (potential) NLS segment (verified for *D. melanogaster* dUTPase [19]), as well as the five characteristic conserved motifs of dUTPases (indicated by Roman numerals in Fig. 1). In agreement, 3D structural investigations of dUTPases showed a homotrimeric organization of mostly  $\beta$ -pleated sheets (jelly-roll fold) [24,25];

however, the NLS segment has not been resolved in any dUTPase structure so far, presumably due to its flexibility. To check whether this annotation is in fact correct and is not perturbed by post-transcriptional and/or post-translational modifications, we investigated the mRNA and protein levels of *D. virilis* dUTPase in cell lines and in *D. virilis* tissue samples. *In silico* analysis of the *D. virilis* dUTPase gene indicated the potential presence of a splice site (Fig. S2). However, the 5' RACE results shown in Fig. 2A indicate that, at the mRNA level, both *D. virilis* embryos and ovary contain only one detectable dUTPase band. The 5' RACE products (used to detect alternative 5' end splicing variants) were sub-cloned and further analyzed by sequencing. The results clearly show that the dUTPase mRNA pool in *D. virilis* embryo and ovary contains only one detectable species (Fig. S3) that contains one NLS. At the protein level, Fig. 2B shows that, in contrast to the two isoforms of *D. melanogaster* but in agreement with the 5' RACE results, the



**Fig. 2.** Detection of possible *D. virilis* dUTPase isoforms by 5' RACE (A) and at the protein level (B). Antibody specificity is shown in (C). (A) The 5' RACE results (cf Fig. S3) clearly indicate a single dUTPase mRNA present in both embryos and ovaries. This band of cDNA was detected using the primers described in Table S1, and corresponds to the pseudo-heterotrimer (ABC) sequence containing one NLS, as verified by sequencing (Fig. S3). No alternative isoforms were detected. (B) Western blot analysis of the *D. melanogaster* Schneider S2 and *D. virilis* WR-DV-1 cell lines. In S2 cells, two dUTPase bands are visible, corresponding to apparent molecular masses of 21 and 23 kDa (in agreement with previous results [21]). In WR-DV-1 cells, only one band is visible, which corresponds to an apparent molecular mass of 59 kDa, in line with the expected calculated molecular mass of 55.7 kDa for the pseudo-heterotrimeric ABC dUTPase. An actin loading control is also shown. (C) A polyclonal antibody raised against *D. melanogaster* dUTPase also reacts with the *D. virilis* dUTPase. S2 and WR-DV-1 cell lysates were probed with the antibody, giving two specific bands for S2 representing the nuclear isoform (~ 23 kDa) and the cytoplasmic isoform (~ 21 kDa), and one specific lane for WR-DV-1 as expected. The antibody specifically recognizes the endogenous ABC pseudo-heterotrimer in the WR-DV-1 lysate as the recombinant protein gives a signal at a similar height (slightly above due to the presence of a His-tag on the recombinant form). The engineered AAA form (which has disintegrated into monomers in the gel) has a similar mass to the nuclear isoform of *D. melanogaster*.

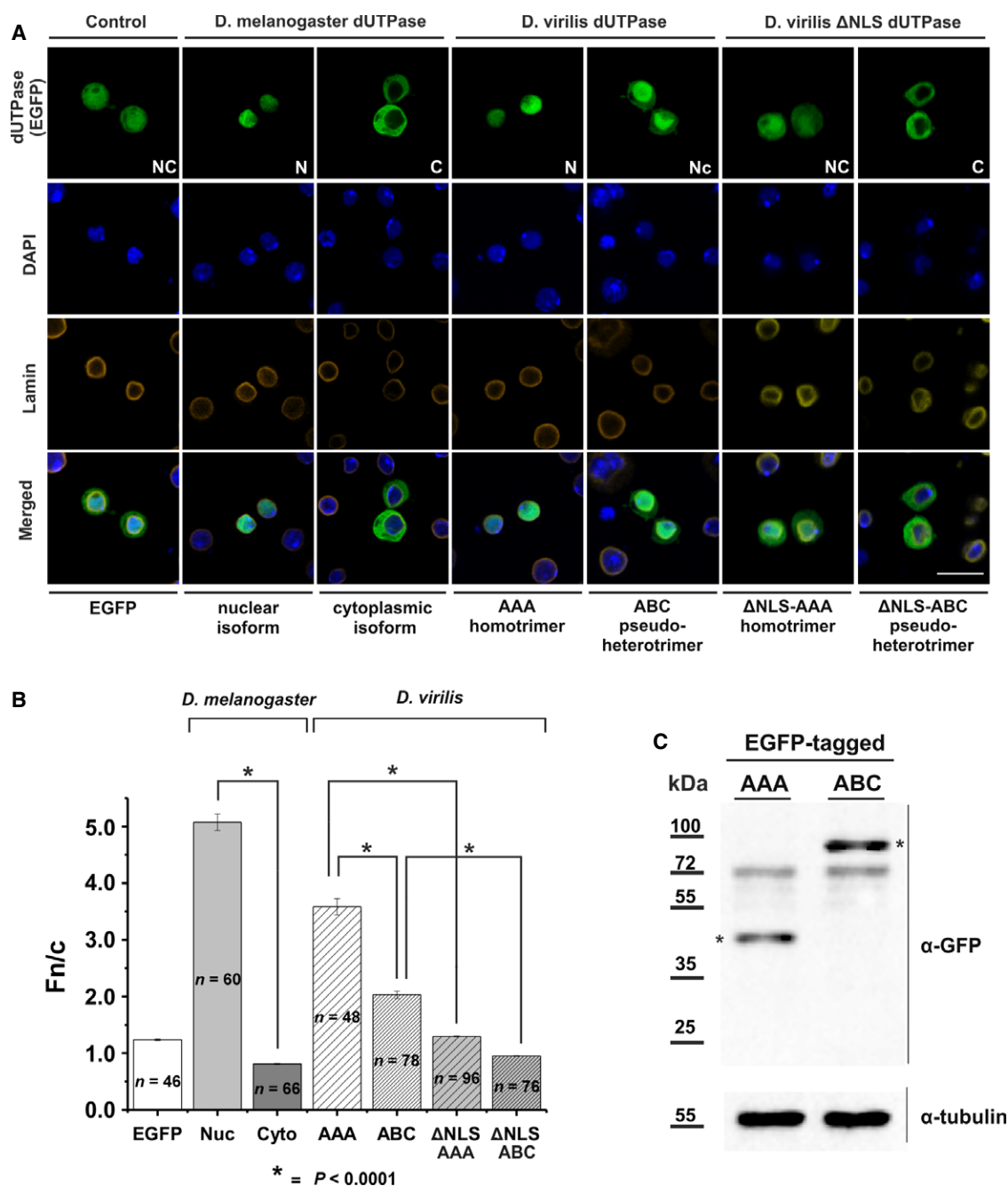
*D. virilis* cell line (WR-DV-1) contains only one dUTPase isoform, which we term ABC dUTPase. This single isoform has an apparent molecular mass of 59 kDa, in good agreement with the calculated molecular mass of the pseudo-heterotrimeric ABC dUTPase (55.7 kDa). The antibody specificity for *D. virilis* dUTPase was addressed by comparing western blots from cell lysates (from S2 and WR-DV-1 cells) and purified recombinant *D. virilis* dUTPase constructs (Fig. 2C). For these experiments, we used the polyclonal antibody raised against *D. melanogaster* dUTPase [21], which shows high specificity but also reacts with the *D. virilis* dUTPase, presumably due to the very close similarity of these two proteins (identity 73.82%, similarity 93.08%).

Based on these findings, we designed various dUTPase constructs for which the only major difference was in the number of NLSs (Fig. 1). For adequate comparisons with the ABC pseudo-heterotrimer, we also cloned the A subunit separately; this is expected to form a characteristic dUTPase homotrimer (AAA homotrimer) possessing three NLSs that does not

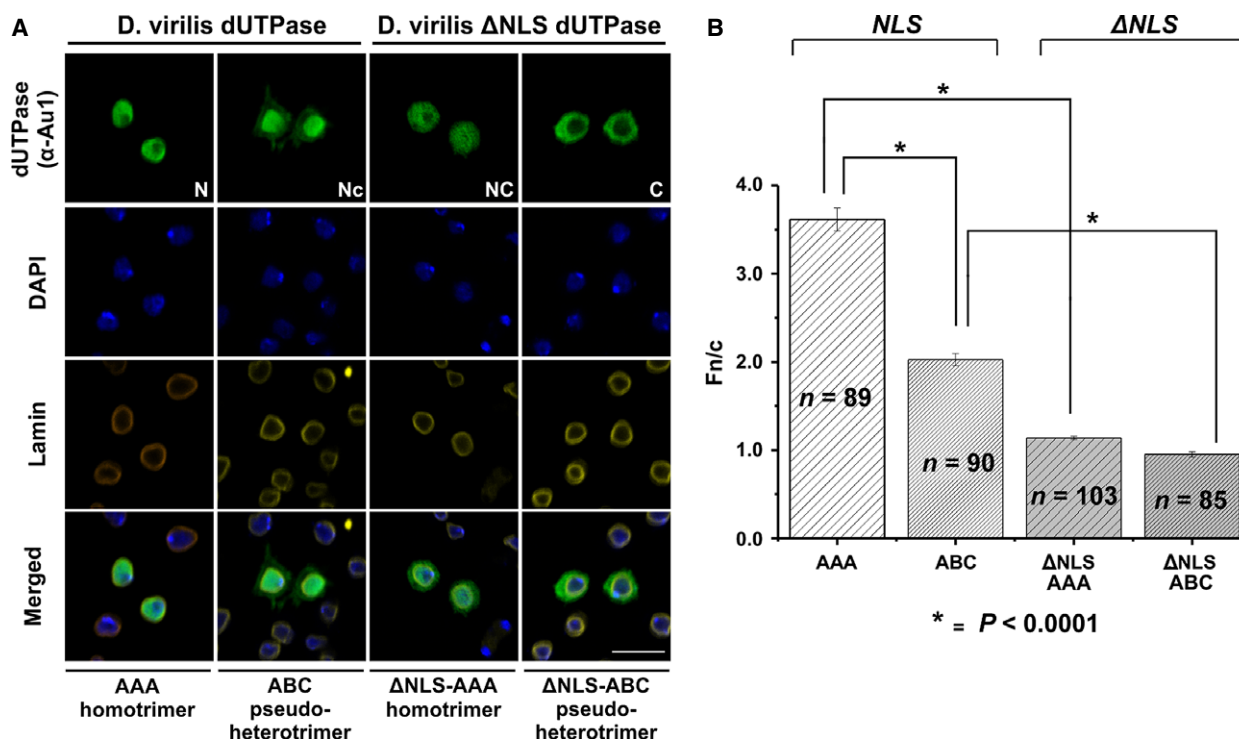
occur physiologically. In addition, we used the two physiologically present isoforms of *D. melanogaster* dUTPase, corresponding to the NLS-lacking cytoplasmic isoform and the NLS-containing nuclear isoform. These *D. melanogaster* isoforms also form homotrimers, as shown previously [21,31,32,36].

### The cellular distribution of dUTPase proteins depend on the copy number of NLS sequences in one trimer

Cellular localization studies were first performed on insect cell lines transfected with plasmids encoding EGFP-labeled dUTPase proteins (*D. melanogaster* cytoplasmic and nuclear isoforms, and *D. virilis* ABC and AAA constructs, Fig. 3). A control experiment in which EGFP was present alone showed that, according to expectations, the EGFP protein (molecular mass 26.94 kDa, no NLS present) is almost equally distributed in the cytoplasm and the nucleus, as its size allows passive diffusion through the nuclear pore. However, the EGFP-labeled *D. melanogaster* isoforms



**Fig. 3.** Effect of NLS copy number on localization of *Drosophila* dUTPase constructs in the *D. virilis* WR-Dv-1 cell line. (A) Localization was categorized into five types: N, completely nuclear; Nc, mainly nuclear; NC, homogenous distribution between the nucleus and cytoplasm; nC, mainly cytoplasmic; C, completely cytoplasmic. EGFP (green) was used as a control, and is evenly distributed throughout the cell (NC). *D. melanogaster* isoforms are strictly nuclear (N) or cytoplasmic (C) in agreement with previous studies [19]. In the case of *D. virilis*, the artificially engineered 'AAA' homotrimer construct containing three NLS sequences shows exclusive nuclear (N) localization, while the wild-type 'ABC' pseudo-heterotrimer with only one NLS is mainly present in the nucleus but is also observable in the cytoplasm (Nc). To assist in visual inspection of the localization pattern, DNA was stained with DAPI (blue) and the nuclear envelope was visualized by lamin Dm0 staining (yellow). Scale bar = 20  $\mu$ m. (B) Image analysis to quantify the relative subcellular localization of GFP-tagged constructs were also performed. Fn/c ratios (means  $\pm$  SEM) were determined as described in Experimental procedures. (C) Western blot showing the total level of the AAA and ABC EGFP-tagged constructs used during imaging. Equal protein loading is shown by the tubulin loading control. Transfected cells express the two constructs at a similar level. Asterisks indicate the specific bands (identified by the correct molecular masses). The commercially available GFP antibody also produced a faint non-specific band that appears in both lanes.

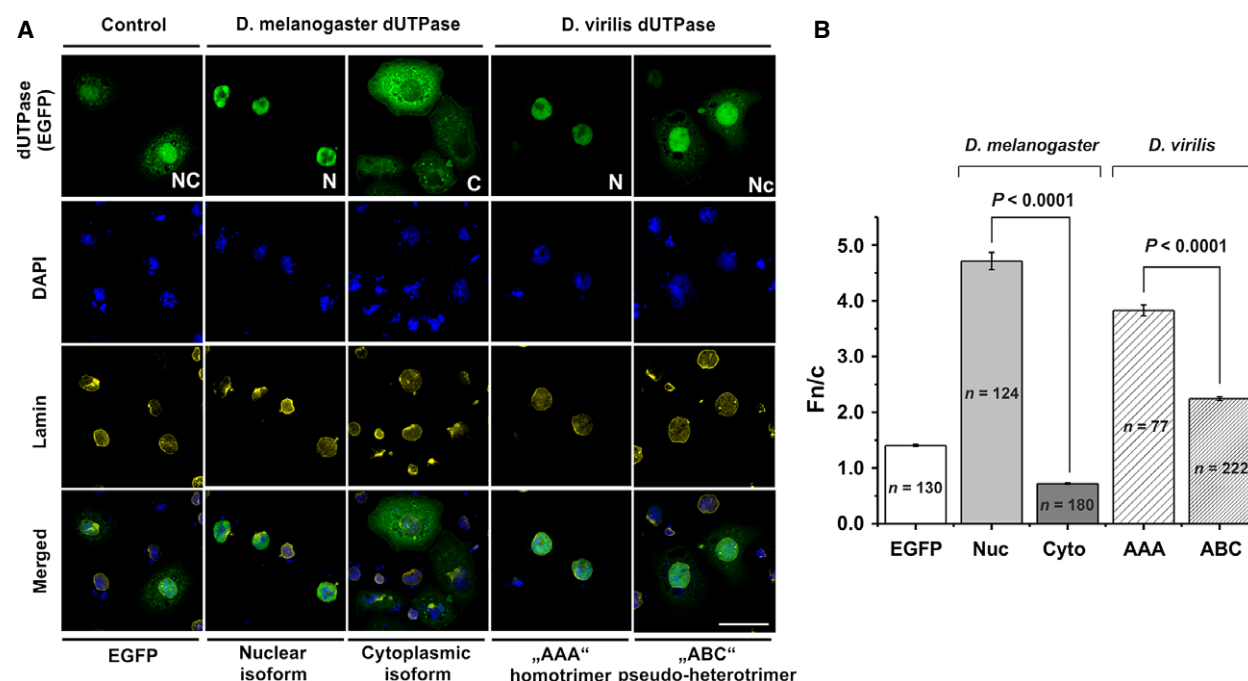


**Fig. 4.** Effect of NLS copy number on localization of *Drosophila* dUTPase AU1-tagged constructs in WR-Dv-1 cells. (A) The intracellular distribution of AU1-tagged *D. virilis* and *D. melanogaster* dUTPase constructs (green) was visualized in WR-Dv-1 cells. To assist in visual inspection of the localization pattern, DNA was stained with DAPI (blue) and the nuclear envelope was visualized by lamin Dm0 staining (yellow). Scale bar = 20  $\mu$ m. (B) Image analysis to quantify relative subcellular localization of AU1-tagged constructs was also performed. Fn/c ratios (means  $\pm$  SEM) were determined as described in Experimental procedures.

showed strict organelle-specific distribution to the cytoplasm or the nucleus, as expected for these proteins [19]. The *D. melanogaster* nuclear isoform is a homotrimer with three NLSs, and this isoform is exclusively present in the cell nucleus (Fig. 3A). The *D. virilis* AAA dUTPase protein also contains three NLSs, and is also located strictly in the nucleus. In contrast, the *D. virilis* ABC pseudo-heterotrimer, containing only one NLS, is present in both the nucleus and the cytoplasm. In addition to the direct qualitative visual observations, the cellular distribution pattern of the various EGFP-labeled proteins was also evaluated by a more quantitative method (Fig. 3B). Here, we calculated the relative amount of EGFP-related fluorescence present in the nucleus and the cytoplasm as a measure of the relative organellar distribution of the respective proteins. In the case of random distribution, i.e. no preference for either the nuclear or the cytoplasmic compartment, the mean ratio of nuclear (Fn) to cytoplasmic (Fc) fluorescence (Fn/c ratio) is  $\sim 1$ , as seen for EGFP (Fn/c ratio: 1.23) when expressed alone. The visually fully exclusive distribution of the *D. melanogaster* isoforms is also well

reflected in the Fn/c ratios of  $\sim 5.10$  and  $0.81$  for the nuclear and cytoplasmic dUTPases, respectively. There is a statistically significant difference in the cellular distributions of the *D. virilis* ABC and AAA proteins. The AAA construct with three NLSs is more predominant in the nucleus (Fn/c ratio 3.58), while the ABC protein with one NLS shows considerable cytoplasmic localization as well (Fn/c ratio 2.03). The AAA construct is associated with a somewhat smaller Fn/c ratio (Fig. 3B) compared to the nuclear *D. melanogaster* dUTPase isoform, potentially due to the different NLS sequences (Fig. 1). In fact, we have shown previously that, in the *D. melanogaster* dUTPase NLS, both the N-terminal proline residue as well as the C-terminal isoleucine/aspartate dipeptide contribute to efficient nuclear targeting [19]; these residues are absent from the *D. virilis* dUTPase NLS. The NLS-lacking constructs  $\Delta$ NLS-AAA and  $\Delta$ NLS-ABC show significant perturbation of nuclear accumulation (Fig. 3A,B), indicating that the NLS segments, identified from sequence similarities, perform as a functional NLS. Figure 3C shows that the total level of the AAA and ABC constructs is similar in the imaged samples, and





**Fig. 5.** Effect of NLS copy number on localization of *Drosophila* dUTPase constructs in S2 cells. (A) The *D. melanogaster*-derived Schneider S2 cell line was used to investigate nuclear transport efficiency of the GFP-tagged dUTPase constructs (green). Cells were counter-stained with DAPI (blue), and the nuclear envelope was visualized by lamin Dm0 staining (yellow). For better visual inspection, cells were plated onto concanavalin A-treated coverslips to flatten cells. The localization patterns are in agreement with the observations in the *D. virilis* cell line (Fig. 3). Scale bar = 20  $\mu$ m. (B) Image analysis to quantify relative subcellular localization of GFP-tagged constructs was also performed. Fn/c ratios (means  $\pm$  SEM) were determined as described in Experimental procedures.

the same image acquisition settings were used in each case when comparing different constructs in each cell type.

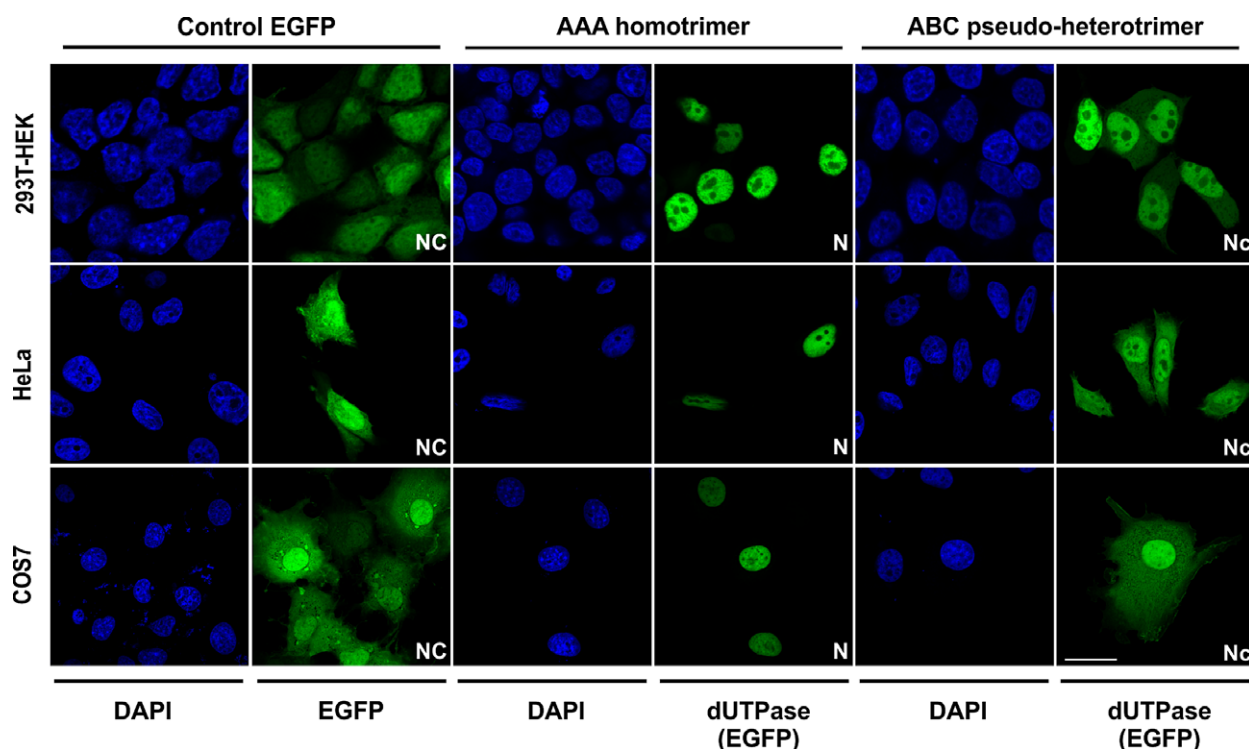
To exclude potential perturbing effects from the large GFP tag, we also used a shorter tag to investigate the cellular distribution of the AAA and ABC proteins. We selected the AU1 tag for this purpose. The results of these experiments are shown in Fig. 4. We found that the localization pattern is indistinguishable for the GFP-tagged construct versus the AU1-tagged construct, yielding almost the same Fn/c ratios for AAA-AU1 and ABC-AU1 proteins (3.61 and 2.02, respectively) as for their GFP-tagged pairs.

The *D. melanogaster* Schneider S2 cell line showed similar results for the above-mentioned constructs (Fig. 5), as did the *Spodoptera frugiperda*-derived cell line Sf9 (data not shown).

The cellular distribution patterns observed in insect cell lines were also investigated in three mammalian cell lines (Fig. 6). The high conservation of the nuclear import framework allows such studies, as shown previously [37]. The clear advantage of using the human HeLa, 293T-HEK and African green monkey COS7 cell lines in cellular localization studies is the cellular

morphology (epithelial or fibroblast-like) of these cells, enabling clearer observation of cytoplasmic compartments. Differences in the distribution of the AAA and ABC *D. virilis* dUTPases are clearly present in the three mammalian cell lines as well. The ABC protein is present in both the nucleus and the cytoplasm, whereas the AAA protein is restricted to the nucleus.

Having established the clear difference between the distribution patterns of the *D. virilis* ABC and AAA forms and the *D. melanogaster* nuclear and cytoplasmic dUTPase isoforms in plasmid-transfected cell lines over-expressing the various protein constructs, we next wished to investigate the localization of the physiologically occurring endogenous dUTPases (Fig. 7). Samples were stained for dUTPase (green), and also for DNA (blue) to outline the nucleus, lamin Dm0 (yellow) and F-actin (red). Lamin was used to highlight the nuclear envelope, whereas F-actin staining outlines the cell membrane and the cellular cytoskeleton. We observed that, in both the cell lines and the tissues, the dUTPase presence is strong in the nucleus but it is also detectable in the cytoplasm in both *D. melanogaster* and *D. virilis*. Based on the previous experiments reported in Figs 3–6, we can dissect the



**Fig. 6.** NLS copy number-dependent cellular localization of dUTPase constructs in mammalian cell lines. Three mammalian cell lines (293T-HEK, HeLa and COS7) were used in order to investigate the distribution of the GFP-tagged *Drosophila* dUTPase constructs (green). The control EGFP is evenly distributed between the nucleus (blue DAPI staining for DNA) and the cytoplasm. The AAA dUTPase construct localizes solely in the nucleus (N), while the ABC pseudo-heterotrimeric construct containing only one NLS is mainly present in the nucleus but is also observed in the cytoplasm (Nc). Scale bar = 20  $\mu$ m.

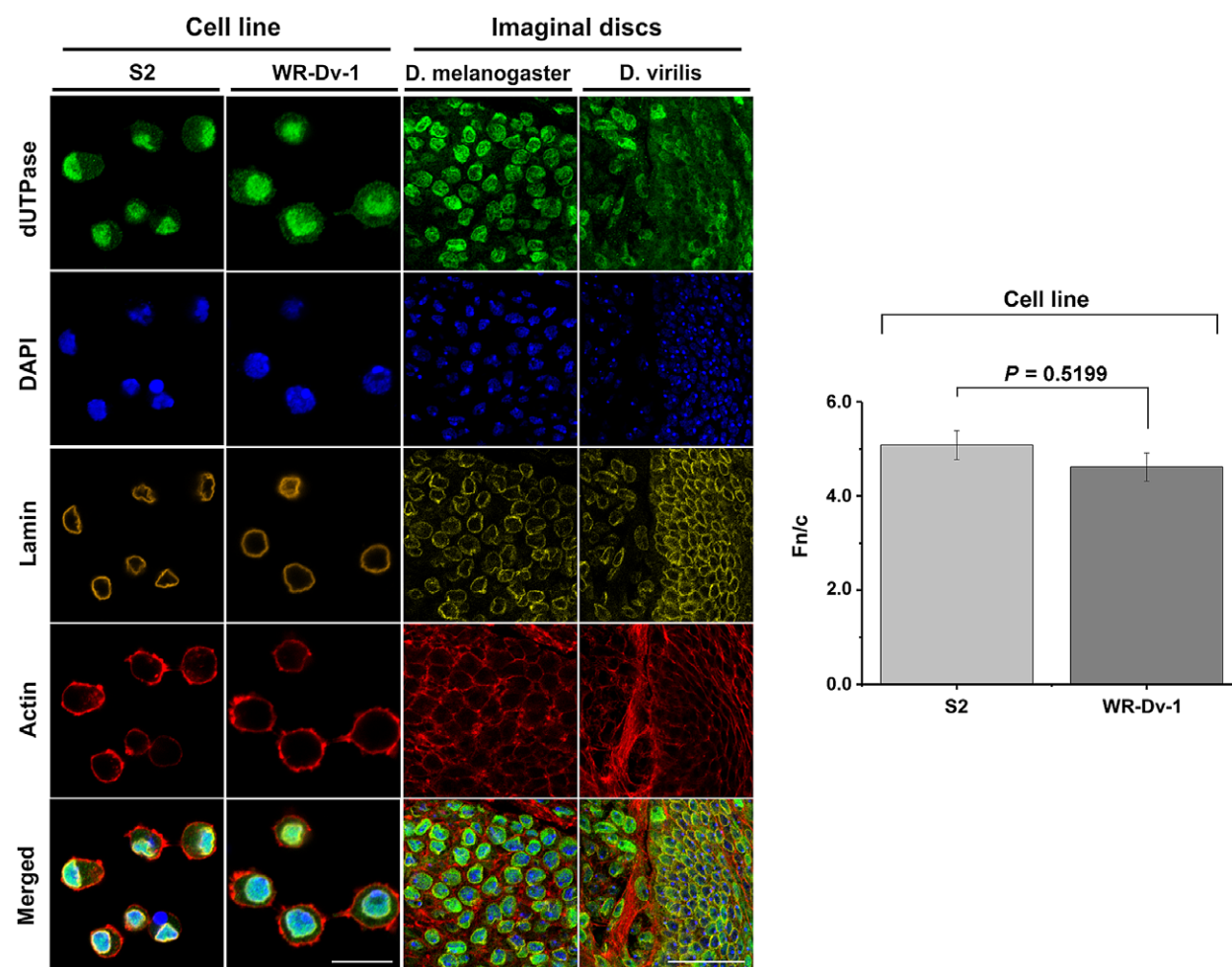
molecular reasons for this distribution. In *D. melanogaster*, two isoforms are required to result in a balanced presence of dUTPase: the nuclear and cytoplasmic isoforms, strictly dedicated to the respective compartments. However, in *D. virilis*, the single ABC dUTPase provides an enzyme presence in both the nucleus and the cytoplasm, and hence the weakened specificity of cellular distribution for the ABC dUTPase may be an advantage, allowing a nucleocytoplasmic distribution with just one protein isoform.

#### Complexation of dUTPase and importin- $\alpha$ : NLS copy number dictates stoichiometry

To investigate the molecular mechanism underlying the differences in the cellular distribution of the *D. virilis* ABC pseudo-heterotrimer and the artificial AAA homotrimer, we used biophysical methods to characterize the interaction of dUTPases with importin- $\alpha$ , a major karyopherin protein that is responsible for nuclear import of classical NLS-containing cargo proteins. The results of size-exclusion chromatography experiments clearly indicated that the AAA construct

elutes at the expected position for the homotrimer, strictly conserved and characteristic for this dUTPase family (Fig. 8). It is also evident that a molecular species associated with altered hydrodynamic behavior (probably indicating an increased molecular mass) is formed when importin- $\alpha$  and AAA or ABC dUTPases are mixed (in a 3 : 1 molar ratio for importin- $\alpha$ /AAA and a 1 : 1 molar ratio for importin- $\alpha$ /ABC). This observation is interpreted as a characteristic experimental indication of complex formation between importin- $\alpha$  and dUTPases. The size-exclusion chromatography profile for the complex of ABC dUTPase/importin- $\alpha$  markedly differs from that for AAA dUTPase/importin- $\alpha$  indicating alterations in the complexes of the various oligomers (Fig. 8A). These qualitative observations were confirmed when the size-exclusion chromatography profiles were investigated by SDS/PAGE of the chromatography fractions (Fig. 8B). In the complex with AAA dUTPase, importin- $\alpha$  shows a slightly earlier elution peak compared to the complex with ABC, potentially indicating that a complex of higher molecular mass is formed in the case of the AAA dUTPase/importin- $\alpha$  complex. In



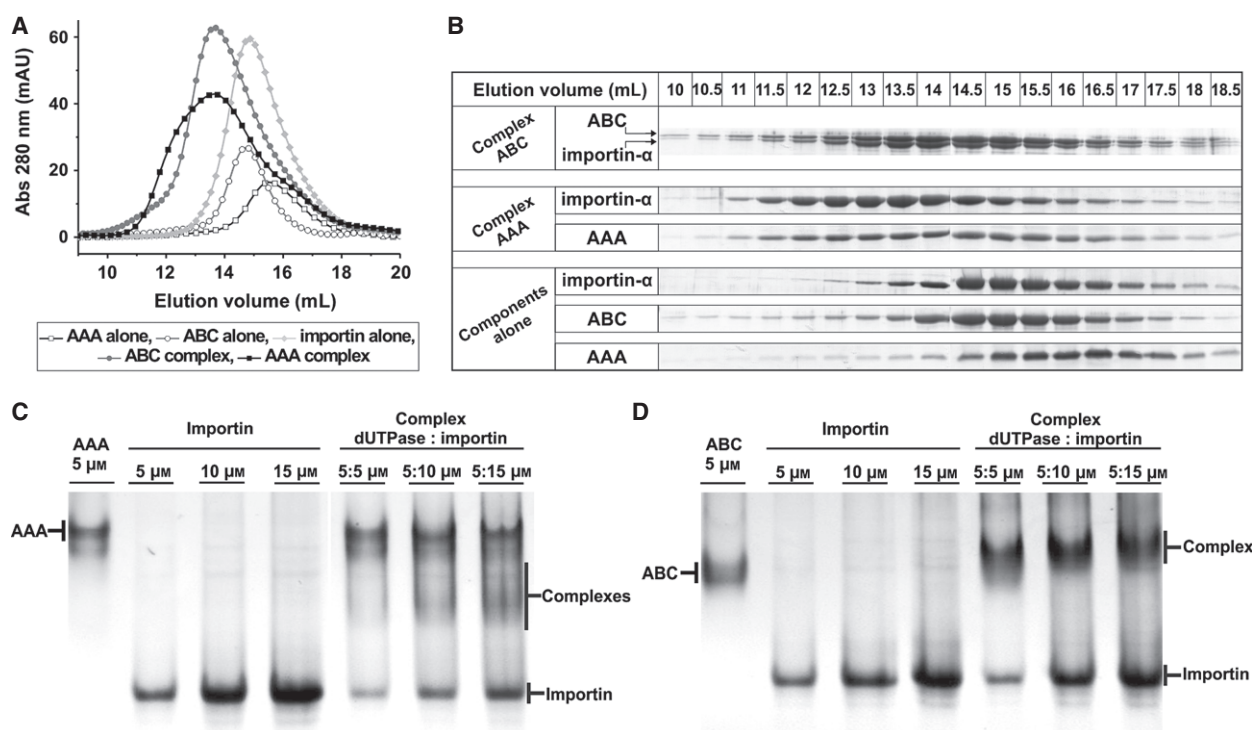


**Fig. 7.** Comparison of the localization pattern of endogenous dUTPases in *D. melanogaster* and *D. virilis* cell lines and tissues. Cellular distribution patterns were visualized using immunocytochemistry to stain endogenous dUTPase isoforms (green) in both *D. melanogaster* (S2) and *D. virilis* (WR-Dv-1) cell lines as well as in *D. melanogaster* and *D. virilis* larval wing discs. To assist in visual inspection, DNA was stained with DAPI (blue), the nuclear envelope was visualized by lamin Dm0 staining (yellow), and cell boundaries are highlighted by phalloidin-TRITC staining for F-actin (red). Scale bar = 20  $\mu$ m.

addition, the size-exclusion chromatography profile for the AAA dUTPase/importin- $\alpha$  complex is highly asymmetric, hinting the possibility of formation of multiple species with different molecular masses. This possibility was investigated in native gel experiments. Figure 8C,D show that multiple oligomeric forms are present in the AAA dUTPase/importin- $\alpha$  mixtures, but only one species is observed in the ABC dUTPase/importin- $\alpha$  mixtures.

In order to obtain quantitative thermodynamic insights into the characteristics of complex formation between importin- $\alpha$  and the various dUTPases, we performed isothermal titration microcalorimetry (Fig. 9). In these experiments, the titration data were well fitted using a 'one set of sites' model. The thermodynamic data extracted from the titration curves show that the

dissociation constant, as well as the molar enthalpy and entropy of the interactions, are very similar for both the AAA dUTPase/importin- $\alpha$  and ABC dUTPase/importin- $\alpha$  complexes (Fig. 9A,C). The protein concentrations in the isothermal titration calorimetry experiments were calculated in monomers (see Experimental procedures). One monomer of the AAA trimer contains one NLS, and the ABC pseudo-heterotrimer also contains one NLS. The isothermal titration calorimetry data clearly show that the 'n' values, indicating the stoichiometry of the complexation, are highly similar (Fig. 9C). We ascertained that physical interaction between the NLS-deleted constructs  $\Delta$ NLS-AAA and  $\Delta$ NLS-ABC and importin- $\alpha$  in these reconstituted systems is practically fully abolished based on the isothermal titration calorimetry data for these constructs



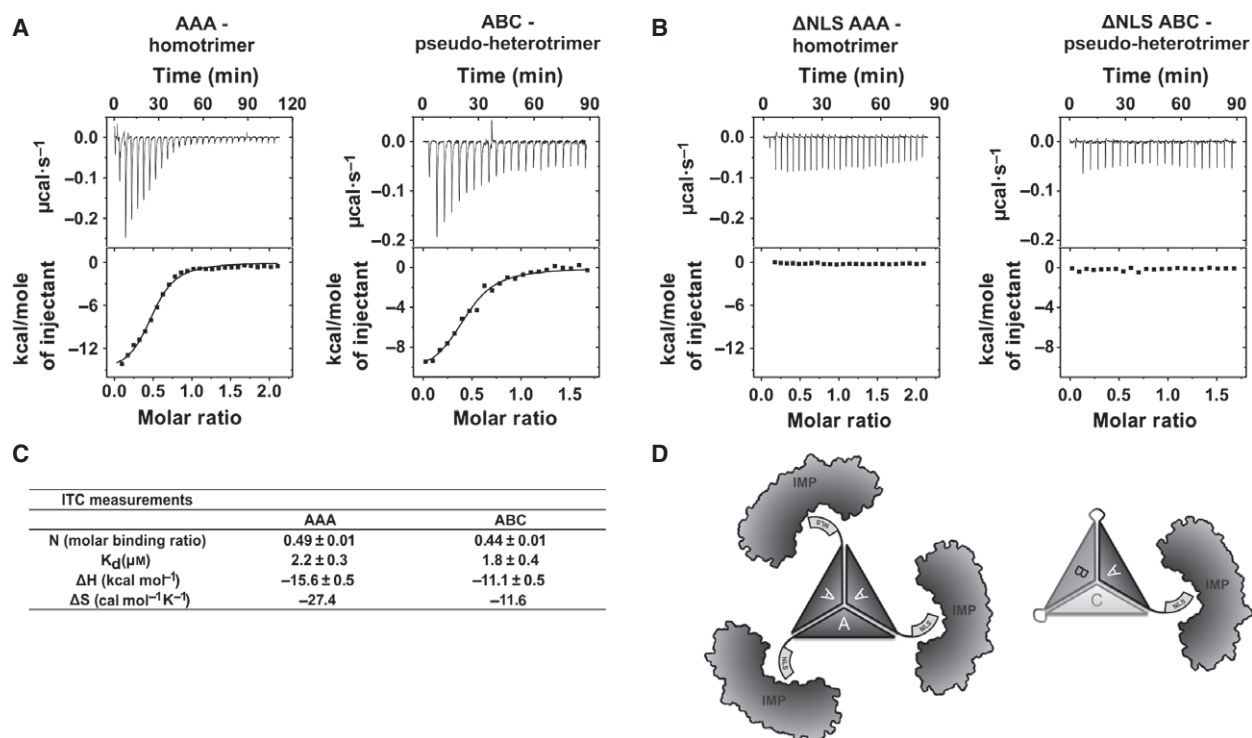
**Fig. 8.** Size-exclusion chromatography and native PAGE analysis of complex formation between distinct dUTPase constructs and importin- $\alpha$ . The results of size-exclusion chromatography (A,B) and native PAGE analysis (C,D) show complex formation between the dUTPase constructs and importin- $\alpha$ . (A) Size-exclusion chromatograms of importin- $\alpha$ , *D. virilis* AAA dUTPase, and ABC dUTPases alone, and importin- $\alpha$ /*D. virilis* AAA and importin- $\alpha$ /*D. virilis* ABC mixtures. (B) Fractions obtained by size-exclusion chromatography were analyzed by SDS/PAGE and detected by Coomassie Brilliant Blue staining. The peaks eluting when protein mixtures are applied contain both dUTPase and importin- $\alpha$ , confirming the formation of complexes. (C,D) dUTPase constructs and importin- $\alpha$  were run on native gels either alone or in mixtures, as indicated, and detected by Coomassie Brilliant Blue staining. Note the appearance of new bands when the proteins are applied in mixtures, indicating complex formation.

(Fig. 9B). In the dUTPase proteins, it is the NLS that is responsible for the interaction with importin- $\alpha$ , hence the most plausible explanation for the stoichiometry data is that the ABC pseudo-heterotrimeric dUTPase binds one importin- $\alpha$  molecule by its single NLS whereas the AAA homotrimeric dUTPase binds three importin- $\alpha$  molecules by the three NLSs (Fig. 9D).

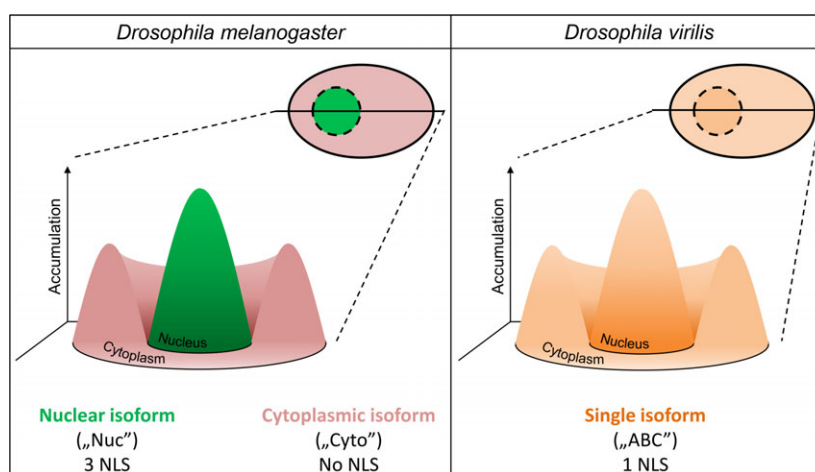
## Conclusions

We have shown that the number of NLS sequences present on *Drosophila* dUTPases correlates with the efficiency of their nuclear accumulation. When the protein lacks NLS sequences (cytoplasmic *D. melanogaster* dUTPase), it is excluded from the nucleus. In the presence of three NLSs in homotrimeric dUTPases (*D. melanogaster* nuclear dUTPase isoform and the engineered *D. virilis* AAA dUTPase), the localization of the trimers is exclusively nuclear. However, in the

case of the *D. virilis* ABC dUTPase (the endogenous form), which harbors only one NLS, the cellular distribution involves both nuclear and cytoplasmic compartments. This feature provides dUTPase in the nucleus as well as in the cytoplasm through a single *D. virilis* isoform, whereas, in *D. melanogaster*, two isoforms are required to perform this task. This cellular distribution is well explained at the molecular level by our results, which show that the NLS copy number dictates the stoichiometry of the dUTPase/importin- $\alpha$  complex. It appears that each of the three NLSs of the *D. virilis* AAA homotrimer is capable of binding one importin- $\alpha$  molecule. This may enhance nuclear accumulation since compared to a partly cytosolic species with just one NLS motif, a species with three NLSs may capture an importin- $\alpha$  molecule more efficiently and spend a greater proportion of time bound to it, thereby increasing the probability of recruiting importin- $\beta$  and being delivered across the nuclear pore complex. This is consistent with previous work in which NLS affinity



**Fig. 9.** Thermodynamic analysis of complex formation between dUTPase constructs and importin- $\alpha$  using isothermal titration microcalorimetry measurements. (A,B) Titration data for AAA dUTPase/importin- $\alpha$  and ABC dUTPase/importin- $\alpha$ , as well as for  $\Delta$ NLS AAA dUTPase/importin- $\alpha$  and  $\Delta$ NLS ABC dUTPase/importin- $\alpha$ . Note that for the  $\Delta$ NLS constructs, there is no appreciable complex formation with importin- $\alpha$ . The upper graphs present a baseline-corrected timeline; the lower graphs show the integrated data (black squares), and, where binding data are observed, the fit of the binding isotherm by an 'independent binding sites' model is also shown (solid line). Values extracted are shown in (C). Note that, for these calculations, the AAA concentration given corresponds to the concentration of AAA monomers. (D) A schematic representation of the possible complexes formed between AAA dUTPase and importin- $\alpha$  and between ABC dUTPase and importin- $\alpha$ .



**Fig. 10.** Schematic representation of localization patterns of physiological dUTPase isoforms. In both *D. melanogaster* (left) and *D. virilis* (right), dUTPase protein is present in the nucleus, as well as in the cytoplasm. To achieve this overall distribution, two organelle-specific dUTPase isoforms are required in *D. melanogaster*, whereas a single nucleocytoplasmic isoform is sufficient in *D. virilis*.

was tested in the context of steady-state nuclear accumulation [17]. The results suggested that the amount of cargo bound to the importin- $\alpha$  receptor determines the import characteristics of the cargo [17]. We therefore directly show in a well-characterized system with clear physiological relevance that NLS copy number in oligomers (avidity) [14] contributes to the regulation of nucleocytoplasmic trafficking, together with the affinity of the NLSs [17]. We also conclude that, in *D. virilis*, evolution of a distinct pseudo-heterotrimeric dUTPase allows cellular requirements to be satisfied by a single isoform (Fig. 10).

## Experimental procedures

### Cell culture

293T-HEK cells were kindly provided by Yvonne Jones (Cancer Research UK, Oxford, UK), and COS7 and HeLa cells were purchased from the American Type Culture Collection (Manassas, VA, USA). Cells were cultured in Dulbecco's modified Eagle's medium/F12 HAM (Sigma, St Louis, MO, USA) supplemented with 50  $\mu\text{g}\cdot\text{mL}^{-1}$  penicillin/streptomycin (Gibco/Life Technologies, Carlsbad, CA, USA) and 10% fetal bovine serum (Gibco) in a humidified 37 °C incubator with 5% CO<sub>2</sub> atmosphere. Schneider S2 cells (derived from *Drosophila melanogaster*) and Sf9 cells (derived from *Spodoptera frugiperda*) were purchased from Gibco. The WR-Dv-1 cell line (derived from *Drosophila virilis*) was obtained from the Drosophila Genomics Resource Center (Bloomington, IN, USA). Schneider S2 and WR-Dv-1 cells were cultured in Schneider insect medium (Sigma), and Sf9 cells were cultured in Sf-900 II medium (Gibco), both supplemented with 10% fetal bovine serum (Gibco) and 50  $\mu\text{g}\cdot\text{mL}^{-1}$  penicillin/streptomycin (Gibco) and kept in a 26 °C incubator.

### Plasmid constructs and cloning

The *D. virilis* strain was obtained from the Drosophila Species Stock Center (La Jolla, CA, USA). For bacterial expression, the *D. virilis* ABC dUTPase (pseudo-heterotrimer) gene was amplified from *D. virilis* genomic DNA, isolated using a MasterPure™ DNA purification kit (Epicentre, Madison, WI, USA), using primers DvirDF and DvirDR3. The PCR fragment was cloned into the *NheI/BglII* sites of vector pET-15b (Novagen/Merck Millipore, Billerica, MA, USA). The *D. virilis* homotrimeric AAA dUTPase was constructed by amplifying the A subunit of the ABC pseudo-heterotrimer from genomic DNA using primers DvirDF and DvirDRs1, and the fragment was cloned into the *NheI/BglII* sites of vector pET-15b (Novagen).

For mammalian cell line expression, both dUTPases were tagged by EGFP by cloning them into the *KpnI/SmaI*

restriction sites of the pEGFP-N1 vector (Clontech, Mountain View, CA, USA) after amplification from pET-15b vectors using the A\_F, A\_R and ABC\_R primers. For insect cell line expression, the EGFP-tagged *D. virilis* dUTPases (AAA and ABC) were cloned into the *KpnI/NotI* sites of the pIZ/V5-His vector (Life Technologies, Carlsbad, CA, USA). NLS-deleted constructs ( $\Delta\text{NLS-AAA}$  and  $\Delta\text{NLS-ABC}$ ) were created by QuikChange mutagenesis (Stratagene, Santa Clara, CA, USA) using the NLSdelF and NLSdelR primers on AU1- or EGFP-tagged and untagged constructs (for both AAA and ABC dUTPases), and cloned into the insect expression vector (pIZ) or the bacterial expression vector (pET-15b). AU1-tagged dUTPase constructs were generated by annealing a pair of single-stranded oligonucleotides encoding the AU1 tag (A\_Au1\_F and A\_Au1\_R), yielding compatible ends for ligation into the *BamHI* and *KpnI* sites of the pIZ-A-EGFP vector, replacing EGFP with the AU1 tag on the C-terminus of the dUTPase. For AU1 tagging of ABC dUTPase, the ORF was amplified via PCR (using primers ABC\_Au1\_F and ABC\_Au1\_R), where the reverse primer encodes the AU1 tag, and the product was cloned into the *KpnI* and *EcoRI* sites of an empty pIZ vector. The methods for creating *D. melanogaster* GFP-tagged nuclear and cytoplasmic dUTPase isoforms have been described previously [19]. The primers used in this study were synthesized by Eurofins MWG GmbH (Ebersberg, Germany), and are listed in Table S1. All constructs were verified by sequencing by Eurofins MWG GmbH.

### Plasmid transfection and immunocytochemistry

The subcellular localization of dUTPase constructs were investigated using various mammalian and insect cell lines. DNA transfections were performed using FuGENE HD reagent (Roche, Mannheim, Germany) according to the manufacturer's instructions. Briefly, sub-confluent cultures of cells grown in 24-well plates were incubated with 1  $\mu\text{g}$  DNA plus 3  $\mu\text{L}$  transfection reagent for 24 h in serum-containing medium. In the case of S2 cells, coverslips were coated with concanavalin A (Sigma) to flatten cells and aid visual inspection. Cells were fixed using 4% paraformaldehyde (pH 7.4 in PBS), washed with PBS, and permeabilized using 0.1% Triton X-100 for 5 min. Cells were then incubated in blocking solution (5% goat serum, 1% bovine serum albumin in PBS) for 2 h, followed by incubation with primary antibodies overnight at 4 °C in blocking buffer. Endogenous dUTPase was detected using the polyclonal antibody described previously [21] at 1 : 10 000 dilution, and a monoclonal antibody specific for lamin Dm0 (ADL67.10 from the Developmental Studies Hybridoma Bank, Iowa City, IA) was used at 1 : 400 dilution to stain the nuclear membrane. AU1-tag was visualized by an AU1-tag specific antibody (Novus Biologicals, Littleton, CO, USA, 1 : 1000). After extensive washing steps, cells

were incubated with secondary antibodies (1 : 1000) coupled to either Alexa Fluor 488 or Alexa Fluor 633 (Molecular Probes/Life Technologies, Carlsbad, CA, USA). To aid visual inspection of the localization pattern, cells were counterstained with 1  $\mu\text{g}\cdot\text{mL}^{-1}$  DAPI (4',6-diamidino-2-phenylindole, Sigma) and 0.5  $\mu\text{g}\cdot\text{mL}^{-1}$  phalloidin-TRITC (Sigma), and embedded in FluorSave™ reagent (Calbiochem/Merck Millipore, Billerica, MA, USA). Images were acquired using either a Zeiss LSCM 710 microscope or a Leica DM IL LED Fluo microscope equipped with a Leica DFC345 FX monochrome camera for image analysis. The same image acquisition settings were applied in each case when comparing different constructs in each cell type.

### Image analysis

Image analysis to quantify relative subcellular localization was performed by IMAGEJ 1.46j (National Institutes of Health, Bethesda, MD, USA) using single-cell measurements, measuring the mean ratio of nuclear (Fn) to cytoplasmic (Fc) fluorescence (Fn/c) within each cell. Statistical analysis of the relative subcellular localization changes was performed using INSTAT 3.05 software (GraphPad Software, San Diego, CA, USA) using the non-parametric Mann–Whitney test. Differences were considered statistically significant at  $P < 0.05$ .

### Immunohistochemistry and immunoblot analysis

Localization of endogenous dUTPase was visualized in *D. virilis* and *D. melanogaster* tissues (data are only shown for larval wing discs but localization was similar in all inspected tissues). Ovary, testis, larval imaginal wing discs, gut and salivary gland samples were collected and immediately fixed in 50% *n*-heptane and 50% PEM/formaldehyde (100 mM PIPES, 1 mM  $\text{MgCl}_2$ , 1 mM EGTA, 2.5% Tween-20, 4% PFA, pH 6.9) for 30 min, with vigorous shaking at room temperature, then washed with inactivating buffer (50 mM Tris, 150 mM NaCl, 0.5% Tween-20, pH 7.4). Blocking was performed in 5% goat serum, 1.5% BSA, 0.1% Tween-20, 1% Triton X-100, 0.001%  $\text{NaN}_3$  in PBS, pH 7.4, for 4 h at room temperature. Tissues were incubated in primary antibodies diluted in blocking buffer (anti-dUTPase, 1 : 10 000; anti-lamin Dm0, 1 : 400), at 4 °C for 16 h. Samples were further washed with blocking buffer for 8 h at room temperature. Secondary antibodies coupled to either Alexa Fluor 488 or Alexa Fluor 633 (Molecular Probes) were applied at 1 : 1000 dilution in blocking buffer for 2 h at room temperature. Tissues were stained using 1  $\mu\text{g}\cdot\text{mL}^{-1}$  DAPI and 0.5  $\mu\text{g}\cdot\text{mL}^{-1}$  phalloidin-TRITC, and embedded in FluorSave™ reagent. Images were acquired using a Zeiss LSCM 710 microscope.

For immunoblotting, sub-confluent cells in T25 flasks were transfected with 6.5  $\mu\text{g}$  DNA supplemented with

30  $\mu\text{L}$  transfection reagent for 48 h (with FuGENE HD, Roche). Cells were collected, washed twice with PBS, and resuspended in lysis buffer comprising 50 mM Tris/HCl pH 7.4, 140 mM NaCl, 0.4% NP-40, 2 mM dithiothreitol, 1 mM EDTA, 1 mM phenylmethylsulfonyl fluoride, 5 mM benzamidine and one cOmplete ULTRA™ EDTA-free protease inhibitor cocktail tablet (Roche). Cell lysis was assisted by sonication. The insoluble fraction was removed by centrifugation (20 000 *g* for 15 min at 4 °C). The protein concentration was determined by a protein assay (Bio-Rad, Hercules, CA, USA) to ensure equivalent total protein load per lane. Proteins were resolved under denaturing and reducing conditions on a 14% polyacrylamide gel, and transferred to poly vinylidene difluoride membrane (Immobilon-P; Merck Millipore). Membranes were blocked using 5% non-fat dried milk, and incubated with primary antibodies against dUTPase (1 : 100 000), and GFP (Sigma, 1 : 1000) for 1 h at room temperature. After washing the membranes, horseradish peroxidase-coupled secondary antibody was applied (Amersham Pharmacia Biotech/GE Healthcare, Little Chalfont, UK). Immunoreactive bands were visualized using enhanced chemiluminescence reagent (GE Healthcare) and recorded on X-ray films (GE Healthcare). The specificity of the primary antibody against *D. virilis* dUTPase is shown in Fig. 2C.

### Recombinant protein production

dUTPase constructs were expressed in the *Escherichia coli* Rosetta BL21 (DE3) pLysS bacterial strain (as previously described for dUTPases [38,39]), and purified using Ni-NTA affinity resin (Qiagen, Hilden, Germany). Transformed cells growing in Luria broth medium were induced at an attenuation at 600 nm of 0.6 using 0.5 mM isopropyl- $\beta$ -D-thiogalactopyranoside for 4 h at 25 °C. Cells were harvested, and lysed in lysis buffer comprising 50 mM Tris/HCl, pH 8.0, 300 mM NaCl, 0.5 mM EDTA, 0.1% Triton X-100, 10 mM 2-mercaptoethanol, 1 mM phenylmethylsulfonyl fluoride, 5 mM benzamidine and one cOmplete ULTRA™ EDTA-free protease inhibitor cocktail tablet (Roche), assisted by sonication, and cell debris was pelleted by centrifugation at 20 000 *g* for 30 min. The supernatant was applied to an Ni-NTA column, and washed with lysis buffer containing 50 mM imidazole. dUTPase was eluted using elution buffer (50 mM HEPES, pH 7.5, 30 mM KCl, 500 mM imidazole, 10 mM 2-mercaptoethanol). N-terminally truncated mouse importin- $\alpha$ 2 (accession number [NP\\_034785](#)) lacking 69 N-terminal residues (residues 70–529) was expressed recombinantly and purified as described previously [15,22].

Proteins were further purified by gel filtration on a Superdex 200HR column (GE Healthcare), in 20 mM Tris/HCl, pH 7.8, 125 mM NaCl, 1 mM  $\text{MgCl}_2$ , 2 mM dithiothreitol. The proteins were > 95% pure as assessed by SDS/PAGE.



Protein concentrations were determined by UV spectrophotometry using molar absorption coefficients (0.1%) of 0.312, 0.286 and 0.864 for the ABC and AAA dUTPases and importin- $\alpha$ , respectively (calculated molecular mass values are 57.91, 21.04 and 55.13 kDa, respectively). Throughout the text, protein concentrations are given in oligomers (homotrimers or pseudo-homotrimers), except for the isothermal titration calorimetry experiments.

### Analytical gel filtration and native PAGE analysis

Analytical gel filtration was performed on a Superdex 200HR column (GE Healthcare). Samples were applied at a concentration of 32.7  $\mu$ M for importin- $\alpha$ , 10.9  $\mu$ M for AAA dUTPase, and 32.7  $\mu$ M for ABC dUTPase (to achieve 1 : 1 stoichiometry of NLSs and importins). Fractionation was started at identical elution volumes for all samples, and was analyzed by SDS/PAGE and detected by Coomassie Brilliant Blue staining.

Native PAGE analysis was performed using a Mini-PROTEAN<sup>®</sup> Tetra Cell system (Bio-Rad), according to the manufacturer's instructions, using a discontinuous buffer system consisting of a 4% stacking gel (pH 7.8) and a 10% resolving gel (pH 9.0). Proteins were detected by Coomassie Brilliant Blue staining.

### Isothermal titration microcalorimetry

Isothermal titration calorimetry experiments were performed at 20 °C on a Microcal iTC<sub>200</sub> instrument (Malvern Instruments Ltd, Malvern, UK) [40]. Proteins were dialyzed into 20 mM HEPES (pH 7.5), 100 mM NaCl, 5 mM MgCl<sub>2</sub>, 2 mM Tris(2-carboxyethyl)phosphine, 5% glycerol. AAA dUTPase was used at 400  $\mu$ M in the syringe while  $\Delta$ NLS-AAA dUTPase was used at 550  $\mu$ M in the syringe, together with 40 and 55  $\mu$ M importin- $\alpha$  in the cell, respectively. ABC and  $\Delta$ NLS-ABC dUTPase were used at 250  $\mu$ M in the syringe, together with 25  $\mu$ M importin- $\alpha$  in the cell. The concentrations are given in monomers for the AAA dUTPase. Aliquots of 1.5  $\mu$ L were used for at least 20 injection steps (the first step of 0.5  $\mu$ L was not considered in the analysis). As a control, dUTPase constructs were also injected into the buffer to allow for consideration of mixing and dilution heat effects, and were withdrawn from the binding data (data not shown). Data analysis was performed using ORIGIN 7.5 software (Northampton, MA, USA). The binding isotherms were fitted using the independent binding sites model 'One Set of Sites'. The values extracted are shown in Fig. 9C.

### 5' race

In order to isolate potential alternative transcripts from *D. virilis* embryos and ovaries and to determine their 5' ends, 5' RACE was performed using a 5'/3' RACE kit

(Roche) according to the manufacturer's instructions. The primers used are summarized in Table S1.

Briefly, *D. virilis* embryos and ovaries were harvested and immediately stored in RNAlater<sup>™</sup> (Ambion/Life Technologies, Carlsbad, CA, USA), snap-frozen in liquid nitrogen, and stored at -80 °C until use. TRIzol reagent (Life Technologies) was used to isolate total RNA from the homogenized samples according to the manufacturer's instructions, and was further purified using an RNA clean-up kit (Macherey-Nagel, Düren, Germany) after treatment with DNase (Qiagen). The quality of RNA samples was analyzed by gel electrophoresis. Total RNA (1  $\mu$ g) was reverse-transcribed using a gene-specific primer (SP1) in a 20  $\mu$ L reaction volume. RNase Cocktail<sup>™</sup> mix (Ambion) and RNase H (NEB, Ipswich, MA, USA) were added, and the mixture was incubated at 37 °C for 15 min. The resulting RNA-free cDNA preparation was further purified using a High Pure PCR product purification kit (Roche). After poly(A) tailing, cDNA was amplified by RedTaq DNA polymerase (Sigma) using a gene-specific primer (SP2) and the Oligo dT anchor primer provided with the kit. A second nested PCR was performed using another gene-specific primer (SP3) and the PCR anchor primer provided with the kit. Final RACE products were cloned into the *SalI*/*EcoRI* sites of the pBluescript SK+ vector (Stratagene) for sequencing. Sequencing was performed by Eurofins MWG GmbH using the M13 uni(-43) primer.

### In silico splicing site prediction

DNA sequences of dUTPases and surrounding genomic regions were downloaded from FlyBase (*D. melanogaster*, CG4584; *D. virilis*, GJ10455). In silico splice site predictions were performed using Human Splicing Finder version 2.4.1 (<http://www.umd.be/HSF/>) with conservative settings [41].

### Acknowledgements

This work was supported by grants from the Hungarian Scientific Research Fund (OTKA NK84008, OTKA K109486), the Baross program of the New Hungary Development Plan (3DSTRUCT, OMFB-00266/2010 REG-KM-09-1-2009-0050), the Hungarian Academy of Sciences (TTK IF-28/2012), the Medin-Prot program of the Hungarian Academy of Sciences, and the European Commission FP7 BioStruct-X project (contract number 283570), to B.G.V. G.R. and A.H. are recipients of Young Researcher Fellowships from the Hungarian Academy of Sciences. H.L.P. is a fellow of the Multidisciplinary Medical Science PhD program, University of Szeged, Hungary. We thank the Drosophila Genomics Resource Center, which is supported by National Institutes of Health grant OD010949-10, for data availability.



## Author contributions

GR, HLP, AH, MB, GNN and BGV planned the experiments, GR, HLP, AH, MB, IS, AB, GNN and IZ performed the experiments, GR, HLP, AH, MB, IS, AB, GNN and IZ analyzed the data, and GR, HLP, AH, MB and BGV wrote the paper.

## References

- Stewart M (2007) Molecular mechanism of the nuclear protein import cycle. *Nat Rev Mol Cell Biol* **8**, 195–208.
- Gorlich D & Kutay U (1999) Transport between the cell nucleus and the cytoplasm. *Annu Rev Cell Dev Biol* **15**, 607–660.
- Moroianu J, Blobel G & Radu A (1996) Nuclear protein import: Ran-GTP dissociates the karyopherin alphabeta heterodimer by displacing alpha from an overlapping binding site on beta. *Proc Natl Acad Sci U S A* **93**, 7059–7062.
- Jans DA & Hubner S (1996) Regulation of protein transport to the nucleus: central role of phosphorylation. *Physiol Rev* **76**, 651–685.
- Fontes MR, Teh T, Toth G, John A, Pavo I, Jans DA & Kobe B (2003) Role of flanking sequences and phosphorylation in the recognition of the simian-virus-40 large T-antigen nuclear localization sequences by importin-alpha. *Biochem J* **375**, 339–349.
- Nardozi JD, Lott K & Cingolani G (2011) Phosphorylation meets nuclear import: a review. *Cell Commun Signal* **8**, 32.
- Lange A, Mills RE, Lange CJ, Stewart M, Devine SE & Corbett AH (2007) Classical nuclear localization signals: definition, function, and interaction with importin alpha. *J Biol Chem* **282**, 5101–5105.
- Fontes MR, Teh T, Jans D, Brinkworth RI & Kobe B (2003) Structural basis for the specificity of bipartite nuclear localization sequence binding by importin-alpha. *J Biol Chem* **278**, 27981–27987.
- Moutty MC, Sakin V & Melchior F (2011) Importin alpha/beta mediates nuclear import of individual SUMO E1 subunits and of the holo-enzyme. *Mol Biol Cell* **22**, 652–660.
- Moreland RB, Langevin GL, Singer RH, Garcea RL & Hereford LM (1987) Amino acid sequences that determine the nuclear localization of yeast histone 2B. *Mol Cell Biol* **7**, 4048–4057.
- Hackl W, Fischer U & Luhrmann R (1994) A 69-kD protein that associates reversibly with the Sm core domain of several spliceosomal snRNP species. *J Cell Biol* **124**, 261–272.
- Falces J, Arregi I, Konarev PV, Urbaneja MA, Svergun DI, Taneva SG & Banuelos S (2010) Recognition of nucleoplasmin by its nuclear transport receptor importin alpha/beta: insights into a complete import complex. *Biochemistry* **49**, 9756–9769.
- Hara T, Arai K & Koike K (2000) Form of human p53 protein during nuclear transport in *Xenopus laevis* embryos. *Exp Cell Res* **258**, 152–161.
- Dingwall C, Sharnick SV & Laskey RA (1982) A polypeptide domain that specifies migration of nucleoplasmin into the nucleus. *Cell* **30**, 449–458.
- Fontes MR, Teh T & Kobe B (2000) Structural basis of recognition of monopartite and bipartite nuclear localization sequences by mammalian importin-alpha. *J Mol Biol* **297**, 1183–1194.
- Guttler T & Gorlich D (2011) Ran-dependent nuclear export mediators: a structural perspective. *EMBO J* **30**, 3457–3474.
- Hodel AE, Harreman MT, Pulliam KF, Harben ME, Holmes JS, Hodel MR, Berland KM & Corbett AH (2006) Nuclear localization signal receptor affinity correlates with in vivo localization in *Saccharomyces cerevisiae*. *J Biol Chem* **281**, 23545–23556.
- Tinkelenberg BA, Fazzone W, Lynch FJ & Ladner RD (2003) Identification of sequence determinants of human nuclear dUTPase isoform localization. *Exp Cell Res* **287**, 39–46.
- Merenyi G, Konya E & Vertessy BG (2010) Drosophila proteins involved in metabolism of uracil-DNA possess different types of nuclear localization signals. *FEBS J* **277**, 2142–2156.
- Muha V, Zagyva I, Venkei Z, Szabad J & Vertessy BG (2009) Nuclear localization signal-dependent and -independent movements of *Drosophila melanogaster* dUTPase isoforms during nuclear cleavage. *Biochem Biophys Res Commun* **381**, 271–275.
- Bekesi A, Zagyva I, Hunyadi-Gulyas E, Pongracz V, Kovari J, Nagy AO, Erdei A, Medzihradsky KF & Vertessy BG (2004) Developmental regulation of dUTPase in *Drosophila melanogaster*. *J Biol Chem* **279**, 22362–22370.
- Róna G, Marfori M, Borsos M, Scheer I, Takács E, Tóth J, Babos F, Magyar A, Erdei A, Bozóky Z *et al.* (2013) Phosphorylation adjacent to the nuclear localization signal of human dUTPase abolishes nuclear import: structural and mechanistic insights. *Acta Crystallogr D Biol Crystallogr* **D69**, 2495–2505.
- Ladner RD, McNulty DE, Carr SA, Roberts GD & Caradonna SJ (1996) Characterization of distinct nuclear and mitochondrial forms of human deoxyuridine triphosphate nucleotidohydrolase. *J Biol Chem* **271**, 7745–7751.
- Persson R, Cedergren-Zeppezauer ES & Wilson KS (2001) Homotrimeric dUTPases; structural solutions for specific recognition and hydrolysis of dUTP. *Curr Protein Pept Sci* **2**, 287–300.

- 25 Vertessy BG & Toth J (2009) Keeping uracil out of DNA: physiological role, structure and catalytic mechanism of dUTPases. *Acc Chem Res* **42**, 97–106.
- 26 Mol CD, Harris JM, McIntosh EM & Tainer JA (1996) Human dUTP pyrophosphatase: uracil recognition by a beta hairpin and active sites formed by three separate subunits. *Structure* **4**, 1077–1092.
- 27 Varga B, Barabas O, Takacs E, Nagy N, Nagy P & Vertessy BG (2008) Active site of mycobacterial dUTPase: structural characteristics and a built-in sensor. *Biochem Biophys Res Commun* **373**, 8–13.
- 28 Kovari J, Barabas O, Varga B, Bekesi A, Tolgyesi F, Fidy J, Nagy J & Vertessy BG (2008) Methylene substitution at the alpha-beta bridging position within the phosphate chain of dUDP profoundly perturbs ligand accommodation into the dUTPase active site. *Proteins* **71**, 308–319.
- 29 Varga B, Barabas O, Kovari J, Toth J, Hunyadi-Gulyas E, Klement E, Medzihradsky KF, Tolgyesi F, Fidy J & Vertessy BG (2007) Active site closure facilitates juxtaposition of reactant atoms for initiation of catalysis by human dUTPase. *FEBS Lett* **581**, 4783–4788.
- 30 Nemeth-Pongracz V, Barabas O, Fuxreiter M, Simon I, Pichova I, Rumlova M, Zabranska H, Svergun D, Petoukhov M, Harmat V *et al.* (2007) Flexible segments modulate co-folding of dUTPase and nucleocapsid proteins. *Nucleic Acids Res* **35**, 495–505.
- 31 Kovari J, Barabas O, Takacs E, Bekesi A, Dubrovay Z, Pongracz V, Zagyva I, Imre T, Szabo P & Vertessy BG (2004) Altered active site flexibility and a structural metal-binding site in eukaryotic dUTPase: kinetic characterization, folding, and crystallographic studies of the homotrimeric *Drosophila* enzyme. *J Biol Chem* **279**, 17932–17944.
- 32 Fiser A & Vertessy BG (2000) Altered subunit communication in subfamilies of trimeric dUTPases. *Biochem Biophys Res Commun* **279**, 534–542.
- 33 Vertessy BG, Persson R, Rosengren AM, Zeppezauer M & Nyman PO (1996) Specific derivatization of the active site tyrosine in dUTPase perturbs ligand binding to the active site. *Biochem Biophys Res Commun* **219**, 294–300.
- 34 Lari SU, Chen CY, Vertessy BG, Morre J & Bennett SE (2006) Quantitative determination of uracil residues in *Escherichia coli* DNA: contribution of ung, dug, and dut genes to uracil avoidance. *DNA Repair (Amst)* **5**, 1407–1420.
- 35 ConsortiumF (2003) The FlyBase database of the *Drosophila* genome projects and community literature. *Nucleic Acids Res* **31**, 172–175.
- 36 Takacs E, Barabas O, Petoukhov MV, Svergun DI & Vertessy BG (2009) Molecular shape and prominent role of beta-strand swapping in organization of dUTPase oligomers. *FEBS Lett* **583**, 865–871.
- 37 Cook A, Bono F, Jinek M & Conti E (2007) Structural biology of nucleocytoplasmic transport. *Annu Rev Biochem* **76**, 647–671.
- 38 Toth J, Varga B, Kovacs M, Malnasi-Csizmadia A & Vertessy BG (2007) Kinetic mechanism of human dUTPase, an essential nucleotide pyrophosphatase enzyme. *J Biol Chem* **282**, 33572–33582.
- 39 Mustafi D, Bekesi A, Vertessy BG & Makinen MW (2003) Catalytic and structural role of the metal ion in dUTP pyrophosphatase. *Proc Natl Acad Sci USA* **100**, 5670–5675.
- 40 Pierce MM, Raman CS & Nall BT (1999) Isothermal titration calorimetry of protein-protein interactions. *Methods* **19**, 213–221.
- 41 Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M & Beroud C (2009) Human splicing finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* **37**, e67.

## Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web site:

**Fig. S1.** Sequence alignment of *D. melanogaster* and *D. virilis* dUTPases.

**Fig. S2.** *In silico* splice site prediction.

**Fig. S3.** 5' RACE sequencing results from embryos.

**Table S1.** Oligonucleotides used in the study.