Tudor domains in proteins that interact with RNA

In the *Drosophila* embryo, the first cells to differentiate are pole cells, which arise from cytoplasmic organelles (polar granules) localized at the posterior pole of the egg. Here, the pole plasm contains two localized signals: a determinant that regulates abdominal development, and a second signal that regulates pole cell formation¹. Mutations in posterior group genes disrupt normal abdominal segmentation and pole cell formation. The results of sequence analysis presented here show that a posterior group gene, *tudor*², encodes a protein containing ten previously unrecognized repeats; these repeats will be termed 'tudor domains'. Tudor appears to possess no other domains in addition to the ten repeats, and its molecular function remains unknown.

The *Drosophila homeless* gene product (hls), which is required for RNA localization during oogenesis³, contains a



single tudor domain (Figs 1, 2) in addition to a region homologous to yeast splicing factors that are members of the DE-H family of RNA-dependent ATPases³. Mutations in the *hls* gene cause disruption of the transport and localization of specific mRNAs during oogenesis owing to changes in microtubule organisation³. The similarity of the hls sequence to splicing factors suggests that it might function during the processing of pre-mRNA whose products direct microtubule organization, which is required for mRNA transport³.

Figure 1

ゴント Consensus RQMAUTDIAPG....ALRFSAQN.....IEDGPKIEKMTTEMRQ.(16). DL V KFSQ... P100/Caeel RPUFUTEITD.....DLHFYUQD....VETGTQFQKLMENMRN.(16). EF I KFV.... P100/Human AKAP149/Human VETITVNQVN.....AGHIFTQQHT...HPTFHAIRSIDQQBYL.(16).1VI & PGA.... Tud_Drome 1 VDLYTHVDHVG....PYL WGH.....VNRDAASLISERIRN.(30).NIN P.PG..... YDUVISYVENG.....PYLFCHLK.....SSDHDIS7MMOQIER.(12). TA U RESE... Tud_Drome 2 DAVE:RFIDS.....PSNFTVQK.....VANIGKFEQLMDEMFS.(14). AP 1 KC..... Tud Drome 3 EARSISWWLS.....PFQFYIVP.....KSVSAKYDNIMRDWRE.(14). ST U RQRK... Tud Drome 4 FQALUVYTAK.....PYRUYUQP.....QAIVPSMQTLLDNWYE.(14). QI % RSS.... Tud_Drome 5 DCTVLSHCDN.....PAQFYTHP.....IDQLSKLNQLHENLQT.(13). AD T MVSV... Tud_Drome 6 CS-VISHVNG.....ICDFFIQLE....RDSKALELIELVLRE.(12). LI % LFEL... Tud_Drome 7 Tud_Drome 8 TKAISTHVEN.....TSRSSQFS....EKDSLMDISCEKSMG.(12).DDM OFAD... SECIISYGNS.....PKSFYUQM....KHNSADLDLIVKTLQS.(16).NGV Y QE.... Tud Drome 9 HNCVVVQFDG.....PMSFVVQM....ESDVPALEQMTDKLLD.(13). AL VVQFPE... Tud_Drome 10 HIs/Drome GSITCIVN.....CGKFFFQP.....QSFEECIRNMSEIFNA.(16). MM L RR.... F16d3.2/Caeel CEUAVSSIIS.....ASHFFIQQPTHPSFASLRHLDMYMGSLYG.(14). LL A-PV..... C56g2.1/Caeel F32e10.5/Caeel 1 IP10LYPCE......YYR1VEHSMPTDIGLQPKRFDRIKENLTF.(22).TVA & CI..... F32e10.5/Caeel 2 DEFMVERIEN.....ERIIVITS.....KDMLERRREAEEYLTK.(15). TS S FH..... VSIVYADS.....PKREFIRA....LADDDQYEKIGTTLAE.(72). GGIVILQDS... YK65 Caeel 1 ANAIFAAG......PTDISIRQLSL.DPMPDYMYAKIKDECAL.(10).FGGFY AFI.... YK65 Caeel 2 F22D6.6/Caeel 1 ANVERTRAES.....PSRIMURL......TNHITDSALTFRE.(12). DY L PTD.... F22D6.6/Caeel 2 GENVIWNDEEQLKEVDPVDIMIYG....NDQLLSMAEQLDTYYSN. (22).FAV & NEEKAMY 2ndary Structure EEEEEe һһнннннннн PEFFFF eEEEEE DTAL:KAYDKA.....VASSKHAL.......KNGDICETSGK.(29). DE S IWSE... SMN/Human VERALSGNGEN.....EDLIKIKK......DLQEVIELTKD.(27). DK M VWSE... SMNH/Human SMNH/Caeel VEAALLGDPTN.....VELLKIKE.....DLGEIISLQED.(25). ER I PHP.... Consensus DCQUYE KUESVRAGQ.....AEUVIDE NRETIEAVKUAQUP....AGFANEP..AG REUNIAL U22055 693-809 P100/Caeel DGE YF REEKVESPAK....THEFELDE NREVEPSTREGIES....PAFSTRVLPAQ TETAFAF 040029 658-775 P100/Human AKAP149/Human DGAMMR.QUVASYEETNE...VEIRVUDV.GYKRUKVDULRQJR....SDFVTLP..FQ.AEVLUDS X97335 710-828 DVEYRR RVVSADLEGQSM.RAEIDFVDF YNRTVDSHDLMFPK..QPKLLQNIP..LH FQVIVLG X62420 13-147 Tud Drome 1 DGHLYRAMVCAVYAQR.....YRVVVVDYANSELLSASDUFQIP....PELLEIK..PFAFFALAG X62420 412-524 Tud Drome 2 DQEWYR-EILRVDDS.....VIVRHVDF YEQNVKRHLIGHIA....EKHLEMP..RQAIKCCLKG X62420 597-707 Tud_Drome 3 DNAILR TVTACNHMMRK...VRVFCVDT SLITVTSEDIWQLE....QRFADPP..CM/HRCSFHS X62420 1018-1133 Tud Drome 4 DGNWYR RISGKDSNAAC...FEWFYIDV NTEEIKRDDIKALD....AKFYEHAS.GF WEINLPI X62420 1311-1426 Tud_Drome 5 DKCWYR-KIIDAEL.....WVLLFIDY NTDCWSDATDIKES....MWSHIEP...F LFCALPI X62420 1619-1728 Tud_Drome 6 DELWYR-QLQKELPDSR....VETLFIDVENTSTTSK..CLMLS....EEIASLP..SLEKKCSLQL X62420 1797-1907 Tud Drome 7 Tud_Drome 8 DLEFYRARILEVLEDDQ....VKVILIDY NTTVVDK.LYELP....QEFTLIK..PVAEICSMEP X62420 1981-2091 DACTYR SIKSVLDPSQG...FEVFULDY NTLVVPE...VWQLP....QEIEPIP..SL-LHCQLSK X62420 2165-2278 Tud_Drome 9 DEVFYR QIRKVLDDGK....CEVHFIDF NNAVTQQ..FRQLP...EELAKPA..RYARHCELDA X62420 2349-2460 Tud Drome 10 DSYFOR TVIRPENOSNROPMFYVRFIDV NCTLUPMOLMRLMPRELTEQYGDLP...PR FEGRUAN S79915 902-1022 HIs/Drome ILKWSREMIIKIETQL.... IFLFELDS FRKVVPSSDLRLMP....QKFAKLP..PFEIPETEDE 278062 307-424 F16d3.2/Caeel GNAWFR VIVQYFDETDE...VFWXFUDY GYSKMARQDLRQIR....TDLMSLP..FQ TEUMLAH U23177 270-388 C56g2.1/Caeel F32e10.5/Caeei 1 QGTWYR KIENVPPRGIW...WYVNLIDA MSRQVMKSDVRRLP...LAFGHYP..PM VKATIRG 041992 68-193 F32e10.5/Caeel 2 QGTMKR SCCGEEGST..... IQLLLIDY ISIEVSKSDUYNIP....NKDEVNME.PF TLVSIKS 041992 327-440 NGTWFF LAKQPPKQPPQSGQUMCHFVDV VCEKFPVAAIRLEP.PAVHPVMSIG..SMIREURMDV L23651 157-330 YK65 Caeel 1 DDRWER QUIRASKIDKQ..AVCVYLLDV AFQYVRKEAMRRLN....STSPFK.KMLUFKCKIGG L23551 444-557 YK65_Caeel 2 F22D6.6/Caeel 1 ERVYRE RIVDVCRNNEL...EKVEE:DD:VIAWYQPEC/GELD....QHYMYYE..WQ HQVSWEG 271262 54-163 F22D6.6/Caeel 2 TGERQE LIVECDTF.....AEVRELDS GRDMVLTGSLYKIH....ROHCREP..PM LRISMHG 271262 379-509 eee 2ndary Structure -----EEEEEE eeee DGCIYP TIASIDFKRET...CVVVVTGV NREEQNL..SDLLS....PICEVAN..NIEQNAQENE U18423 36-160 SMN/Human DGQCYESEIERIDEENGT...AAITFAGY SNAEVTPLLNLKPVE----- T91668 SMNH/Human DGKKVF RIDSLTPAG.....VAI----- D72470 SMNH/Caeel

Multiple alignment of tudor domain sequences. Residues are coloured (hydrophobic: ACFILMVWY, in red; small: ACGPS, in green; acidic: DE, in brown; and, basic: KR, in dark blue) whose physicochemical characteristics are conserved in at least 16 of 22 sequences. Secondary structure, predicted¹⁰ with an expected accuracy greater than 82% (upper case) or 72% (lower case), is shown beneath the alignment (E/e denotes β-strand, H/h denotes α -helix); a consensus line is shown above the alignment (h = hydrophobic, σ = small residues). Other tudor repeats appear to be partially encoded by human (EMBL accession codes: W76301, H20629/Z45201/T52693/F07472, F13399/T77659, R13795/R14717), Zea mays (T23393) and Arabidopsis thaliana (T22348/T04365) expressed-sequence tags (ESTs). For comparison, regions of the human SMN gene product and homologues partially encoded by ESTs (human: T91668, Caenorhabditis elegans: D72470 and overtapping ESTs), are also shown. A BLASTX search¹¹ with the tudorsimilar EST W76301 yields $P = 10^{-2}$ when aligned with SMIN. Internal repeats in tudor were detected using REPRO12: 18 pairs of non-overlapping regions in tudor scored over 100, including a 680-residue alignment scoring 400 (repeats 5-7 aligned with 8-10). These findings were complemented using dotplots13, BLASTP11 and MACAW14 self-comparisons (using MACAW, two-alignment blocks of the nine repeats yielded P-values of 8.8×10^{-16} and 9.6×10^{-9} [N = 2515⁹]). A profile constructed from an alignment of nine tudor domains and compared¹⁵ with databases, indicated the presence of single homologous domains in p100, F13d3.2, AKAP149 and hls (SWise scores noise plateau 4998-5446,

level < 4780). Two subsequence profiles demonstrated the presence of the remaining homologues, including the tenth repeat in tudor (tudor domain scores > 4533, noise plateau < 4480). By contrast to profile methods, $MoST^{16}$ identified only human and *C. elegans* p100, and *Drosophila* hls as containing tudor domains (P < 0.02, using an alignment block of nine tudor repeats). Species: Caeel, *C. elegans*; and Drome, *D. melanogaster*.

<u>PROTEIN SEQUENCE MOTIF</u>



Figure 2

Schematic representation of the domain organizations of selected tudor domain-containing proteins (approximately to scale). Red ovals represent tudor domains. Blue rectangles in p100 represent domains homologous to staphylococcal nuclease¹⁷. Helicase motifs (I–VI and Ia) are shown in Roman numerals; KH domains are typically single-stranded DNA- or RNA-binding motifs; RII represents a sequence that is predicted to bind regulatory subunits (RII α and RII β) of type II α cAMP-dependent protein kinase⁷; the black box represents a putative leucine-zipper motif⁷. The possibility¹⁷ of a p100 repeat-like domain fragment in the carboxy-terminal region of p100 is indicated by a question mark. Species: Caeel, *Caenorhabditis elegans*; Drome, *Drosophila melanogaster*.

A single tudor domain is present in p100 (Ref. 4), a human nuclear protein that coactivates gene transcription mediated by the Epstein-Barr virus nuclear antigen 2 (EBNA 2) by simultaneously interacting with EBNA 2 and both subunits of transcription factor TFIIE; p100 also binds single-stranded DNA⁴. Single tudor domains also occur in two sequences, a human A-kinase anchor protein (AKAP149)⁵ and Caenorhabditis elegans putative protein C56g2.1, each of which contains a KH domain with a potential for RNA-binding⁶. AKAP149 also contains a domain that binds regulatory subunits (RII α and RII β) of the type II α isoform of cAMP-dependent protein kinase (PKA!la)⁷. An alternatively spliced version of AKAP149 (S-AKAP84) participates in spermiogenesis, probably by facilitating ordering of spermatid mitochondria7.

Finally, the product of the spinal muscular atrophy-determining gene survival motor neuron (SMN)⁸ scored highly in profile searches at levels similar to or higher than various tudor repeats; similar results (not shown) were obtained in reciprocal searches. The low similarity of SMN homologues to tudor domain sequences could indicate that these are not homologues. However, the perceived role of SMN in RNA metabolism⁹ and the apparently similar localizations of SMN and p100 to unusual nuclear structures^{4,9} might suggest otherwise. Although a tudor domain in SMN remains a possibility, further evidence is essential to resolve this issue.

An insufficient amount of functional information concerning these proteins is currently available from which to ascribe putative function(s) to tudor domains. However, identification of these apparently homologous domains in developmentally important proteins, including several with putative RNA-binding functions, hints at either an RNA-binding role or a protein-binding function during RNA metabolism and/or transport.

Acknowledgements

I would like to thank T. Gibson and J. Thompson (EMBL, Heidelberg, Germany) for supplying PROPLOT, and K. Talbot (University of Oxford, UK) for helpful discussions on SMN. C. P. P. is a Wellcome Trust Career Development Research Fellow and a member of the Oxford Centre for Molecular Sciences, which is supported by the UK EPSRC, BBSRC and the MRC.

References

- 1 St Johnston, D. and Nüsslein-Volhard, C. (1992) Cell 68, 201–219
- 2 Golumbeski, G. S., Bardsley, A., Tax, F. and Boswell, R. E. (1991) *Genes Dev.* 5, 2060–2070
- 3 Gillespie, D. E. and Berg, C. A. (1995) *Genes* Dev. 9, 2495–2508
- 4 Tong, X. et al. (1995) Mol. Cell. Biol. 15, 4735–4744
- 5 Trendelenburg, G., Hummel, M., Riecken, E. O. and Hanski, C. (1996) *Biochem. Biophys. Res. Commun.* 225, 313–319
- 6 Burd, C. G. and Dreyfuss, G. (1994) Science 265, 615–621
- 7 Lin, R-Y., Moss, S. E. and Rubin, C. S. (1995) J. Biol. Chem. 270, 27804–27811
- 8 Lefebvre, S. et al. (1995) Cell 80, 155–165
- 9 Liu, Q. and Dreyfuss, G. (1996) EMBO J. 15,
 - 3555-3565
- 10 Rost, B. and Sander, C. (1993) J. Mol. Biol. 232, 584–599
- 11 Altschul, S. F., Boguski, M. S., Gish, W. and
- Wootton, J. C. (1994) Nat. Genet. 6, 119–129 12 Heringa, J. and Argos, P. (1993) Proteins Struct.
- Funct. Genet. 17, 391–411 13 Thompson, J. D., Higgins, D. G. and Gibson, T. J.
- (1994) Comput. Applic. Biosci. 10, 19–30 14 Schuler, G. D., Altschul, S. F. and Lipman, D. J.
- (1991) Proteins Struct. Funct. Genet. 9, 180–190 15 Birney, E., Thompson, J. D. and Gibson, T. J.
- (1996) Nucleic Acids Res. 24, 2730–2739 16 Tatusov, R. L., Altschul, S. F. and Koonin, E. V.
- (1994) Proc. Natl. Acad. Sci. U. S. A. 91, 12091-12095
- 17 Ponting, C. P. Protein Sci. (in press)

CHRISTOPHER P. PONTING

Fibrinolysis Research Unit, University of Oxford, The Old Observatory, South Parks Road, Oxford, UK OX1 3RH. Email: ponting@molbiol.ox.ac.uk

Next month in TiBS

The evolution of ribonucleotide reduction Peter Reichard

Information and peer-review on the Internet Christopher G. Burd, Markus Babst and Scott D. Emr

Finding protein-binding sites in DNA – the next generation Kornelie Frech, Kerstin Quandt and Thomas Werner Chaperones get in touch: the hip-hop connection Judith Frydman and Jörg Höhfeld

CH•••O hydrogen bonding in biology Markus C. Wahl and Muttaiya Sundaralingam

Regulation of RNA polymerases I and III by the retinoblastoma protein: a mechanism for growth control Robert J. White