

Az első uracil-DNS nukleáz szerkezeti és funkcionális jellemzése.

Szakmai beszámoló
Beke Angéla Ph.D.

74200 azonosítójú NKTH-OTKA-H07 B kategóriájú projekthez

A jelen beszámoló tartalmazza az elért eredmények rövid, tételszerű felsorolását, majd a publikálatlan kísérletes eredmények részletes bemutatását, továbbá a mellékletekben két beküldött kéziratot. Az első kéziratból a terjedelem korlát miatt a „Materials and Methods” fejezet kihagyásra került, ugyanígy a második kézirat „Supplementary” anyaga.

A pályázatban célul tűztem ki a nemrég uracil-DNS specifikus nukleázként azonosított fehérje (UDE) részletes jellemzését enzimatis aktivitása, szerkezete és funkciója szempontjából. Az UDE-t eredetileg késői stádiumú *Drosophila* lárvából izoláltam, a rekombináns His-taggal ellátott fehérjének uracil-DNS-re specifikus degradáló aktivitását mértem uracil bázisokat random tartalmazó plazmid szubsztráton [1]. Munkahipotézisünkben feltételeztük, hogy ecetmuslicában az UDE enzim szerepet játszik a teljes átalakulás során zajló sejthalál folyamatokban. A lárva állapotok során ugyanis a lárvális szövetekből a kanonikus uracil-DNS metabolizmus két kulcsfontosságú enzime, a preventív szerepet játszó dUTPáz és a javításért felelős UNG (fő uracil-DNS glikoziláz), egyaránt hiányzik [2]. Így, a lárvális genomban lehetőség nyílik uracil felhalmozódására, mely uracilok aztán a lárvaállapot végén forrópontokként szolgálhatnak az apoptotikus folyamatokat indító/kísérő DNS-degradáció számára. A lárvaállapotok végétől a bábállapotban kifejeződő UDE eddig példa nélkül álló enzimaktivitását jól el tudtuk helyezni a vázolt hipotézisben. Tekintettel szerkezeti és funkcionális újszerűségére, feltételezett érdekes szerepére és jövőbeli lehetséges alkalmazásaira, célul tűztük ki a fehérje részletes jellemzését.

A pályamunkában leírt szisztematikus kutatást azonban nem sikerült maradéktalanul megvalósítani, számos nem várt akadályba ütköztünk, ugyanakkor számos meglepő új eredménnyel gazdagodtunk, melyek a jövőben lehetővé teszik a részletes funkcionális jellemzés megvalósítását, valamint elérhető közelségbe hozzák a fehérje szerkezetének meghatározását is.

A projektben megvalósított eredmények röviden:

1. Beállítottam **nukleotid-szintű felbontást engedő gélelektroforetikus módszert** a fehérje aktivitása nyomán keletkező termékek részletes analizésére. Sajnos azonban egyelőre nem sikerült megfelelő méretű/szekvenciájú oligonukleotid szubsztrátot találni.
2. Finomítottam a rekombináns **His-UDE tisztítási eljárását**, minek során sikerült megszabadulni az addigi nukleinsav szennyezéstől.
3. Kimutattam, hogy a „szennyező” **nukleinsav** a várakozással ellentétben nem DNS, hanem egy **RNS** molekula.
4. Kimutattam, hogy az **UDE által kötött RNS**, többé-kevésbé homogén, **1000bp** körüli molekulamérettel jellemezhető.
5. Kimutattam, hogy az UDE által kötött RNS tartalmazza az UDE **saját mRNS**-ét.
6. Ezek után eukarióta expressziós rendszerben is kifejeztem a rekombináns fehérjét. SF9 rovarsejtekben jól expresszáldott, de a feltárás során nagy mennyiség csapadékban maradt, a tisztított fehérje nagy arányban tartalmazott **riboszómális fehérje-szennyezőket**, továbbá RT-PCR segítségével kimutatható volt az **UDE mRNS**-ének jelenléte a fehérje preparátumban.
7. *Drosophila* **báb extraktban szintén kimutattam az UDE RNS-kötését**. Ez alapján feltehető, hogy az UDE RNS-kötése nem csupán a rekombináns fehérjére jellemző artefaktum.

8. Kimutattam a rekombináns His-UDE RNS-kötött formája és a nukleinsav-mentes UDE közti **konformációs és stabilitásbeli különbségeket**. (CD, fluorimetriás felfűtés, thermofluor)

9. Limitált proteolízissel kimutattuk, hogy a **DNS és RNS kötőhelyek részben különböznek**, részben átfednek az UDE fehérje felszínén.

10. Vizsgáltam a rekombináns His-UDE **RNS- és DNS-kötő képességét** gélshift analízis segítségével. Az eredmények finomítása jelenleg folyamatban van.

Az UDE által kötött RNS azonosításáról és az RNS-, ill. DNS-kötésének összehasonlító elemzéséről jelenleg egy kézirat készülében van. A vonatkozó eredmények részletes kifejtését ábrákkal illusztrálva a beszámoló alább tartalmazza.

11. A fehérje **doménanalízisét** limitált proteolízissel vizsgáltuk többféle enzimmel, ill. egy kémiai hasítószerezrel (hidroxilaminnal) is. Kimutattuk az **1A és az 1B relatív önállóságát**. Ezt követően az **1A és 1B fragmensek de novo szerkezetmodellézését** kollaborációban elvégeztük.

12. Western bloton megerősítettem, hogy a vöröslisztbogárban a predikcióval összhangban az **UDE-nak egy csonkolt formája** van jelen, melyből hiányzik az 1A motívum.

13. A csonkolt enzimre jó modellként szolgáló hidroxilaminnal hasított rekombináns UDE **megőrizte DNS-kötő képességét** és uracil-plazmidon detektálható **aktivitását**. Ezért feltételezhető, hogy az 1A motívum/domén nem esszenciális része a fehérjének, ill. a homológ 1B helyettesíteni tudja egy dimerizálódás révén.

14. Az UDE feltételezett **dimerizálódását** géliszűrővel támasztottuk alá.

A fenti eredményeket részleteiben a beszámolóhoz csatolt, jelenleg revízió alatt álló kézirat tartalmazza (1. melléklet), melyben a projekt azonosító száma feltüntetésre került.

15. A plazmidon detektálható aktivitás részletes jellemzéséhez megfelelő oligonukleotid-szubsztrát megtalálása érdekében több **kötődési tesztet** is kifejlesztettünk, ill. fejlesztünk: fluorimetriás, thermofluor, gélshift titrálások és többféle footprint eljárás.

16. Két konkrét tesztet (nagy specificitású kimotripszines limitált emésztés ill. triptofán fluorimetriás titrálás) azonosítottunk, mely **különbséget** detektál az uracilos és a kontroll **DNS kötődésében**.

17. A megfelelő oligonukleotid szubsztrát megtalálása továbbra is elengedhetetlennek tűnik a fehérje sikeres **kristályosításához** is. Az apo fehérjét, valamint uracilt tartalmazó 15, ill. 30 nt hosszú egyes és kettős szálú oligonukleotidokkal komplexált formáit, továbbá rekombináns úton előállított fragmenseit eddig **közel kétezer körülményben** próbáltuk kristályosítani, felhasználva nemzetközi együttműködések (SPINE-2, ESFRI, INSTRUCT (Oxford, Hamburg)) keretében elérhető robotizált csúcstechnikát is. Mindaddig nem kaptunk megfelelő méretű és minőségű fehérjekristályt.

Az aktivitás és az atomi szintű szerkezet jellemzésében ezideig a fenti limitált eredményeket tudtuk csupán felmutatni. A felmerülő problémákat a jövőben mindenekelőtt a megfelelően kötődő szubsztrát keresésével/megtalálásával kívánjuk kiküszöbölni.

18. **Új antitestet is fejlesztettünk** a már nukleinsav-mentesen előállított His-UDE fehérje ellen, amely lehetőséget adott a fehérje fiziológiás formájának hatékonyabb immunoprecipitálására.

19. Az új poliklonális antitest segítségével kimutattam és azonosítottam a fehérjének egy **második izoformáját**, ami az első tömegspektrometriás adatok szerint a C-terminálison csonkolt forma, elveszítve nemcsak a nukleáris lokalizációs szignálját, de feltehetőleg a negyedik konzervált motívumát is.

20. Rekombináns úton előállított UDE fragmensek segítségével részletesen jellemeztem a **két antitestre jellemző domináns epitópokat**, melyek magyarázatot adtak a két poliklonális antitest meglepő izoforma-specificitására.

21. Jellemeztem az **újjonnan azonosított izoforma expressziós mintázatát** az egyedfejlődés során, és kimutattam, hogy a hosszú forma bábállapotban történő megjelenését mintegy előkészíti a rövid forma mennyiségének 3. lárva stádium elejétől megfigyelhető fokozatos emelkedése.

22. Az SF9 rovarsejtben termelt rekombináns UDE-n detektáltunk egy **foszforilációs helyet**. A foszforiláció pontos lokalizálása, ill. jelentőségének tisztázása, valamint a fiziológiásan jelenlévő két izoforma foszforiláltságának vizsgálata jelenleg folyamatban van.

Az izoformákra, azok szabályozására és poszt-transzlációs módosításukra vonatkozó eredményeket egy újabb kéziratban tervezzük közölni még ebben az évben. A vonatkozó eredmények részletes kifejtését ábrákkal illusztrálva a beszámoló alább tartalmazza.

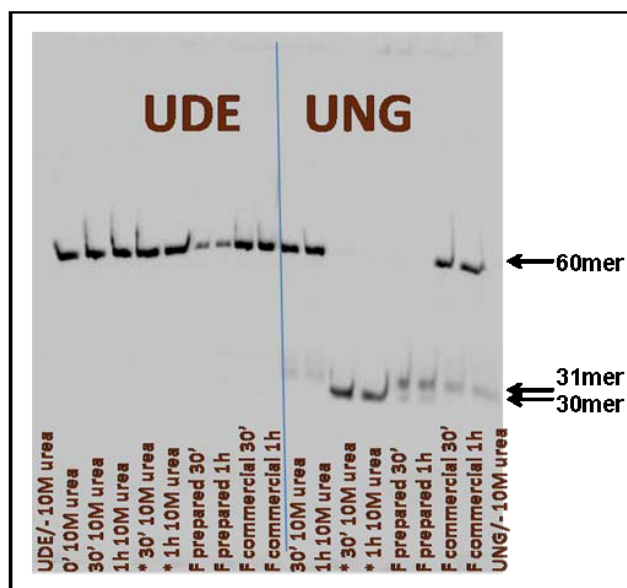
23. A munkahipotézis más, teljes átalakulással fejlődő rovarokra való kiterjeszhetőségét is vizsgáljuk. Ennek során a vezetésem alatt dolgozó egyetemista és gimnazista diákok Western blot segítségével többféle rovarmintát vizsgálnak. Eddig a **vörös lisztbogárból és molyból** sikerült **UDE homológ fehérjét** kimutatni.

24. Az UDE sejtbeli és egyedfejlődésben betöltött szerepét több úton is vizsgáljuk a csoportban. A beszámolóhoz csatolt 2. kéziratban leírjuk az **UDE siRNS** technikával történő **csendesítése** által okozott részben letális fenotípust.

A beküldött kézirat, melyben a projekt azonosító száma feltüntetésre került, a 2. számú mellékletben található.

A még nem közölt eredmények részletes bemutatása

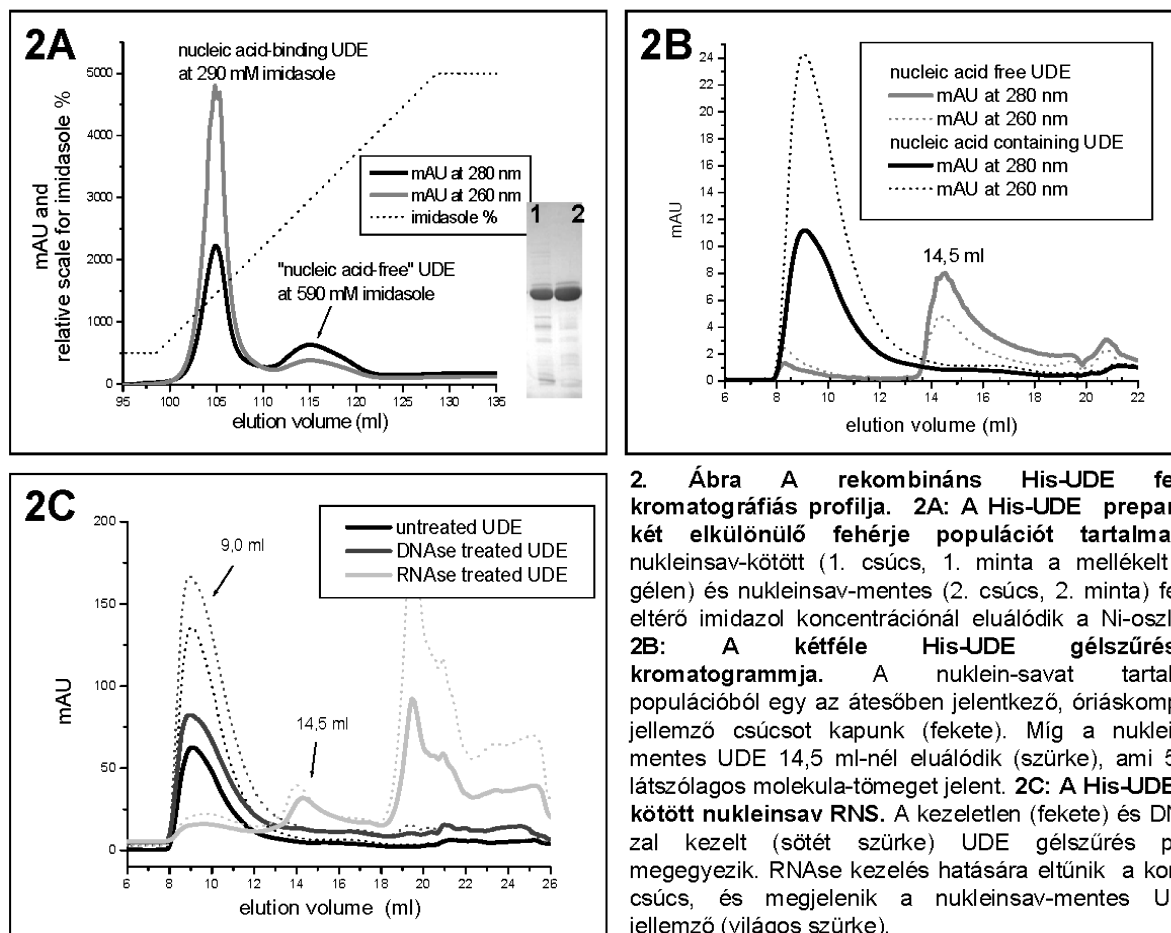
A pályázatban tervezettek szerint beállítottam egy nukleotid-szintű felbontást engedő gélelektroforetikus módszert a fehérje aktivitása nyomán keletkező termékek részletes analizésére. Szusztrátként 5' végén Cy3 fluoreszcens jelzővel ellátott, uracilt meghatározott helyen tartalmazó, egyes és kettős szálú, 30mer, ill. 60mer oligonukleotidokat használtam (1 ábra). Kontroll kísérletben vizsgáltam az UNG enzim aktivitása révén az említett szubsztrátokon keletkező bázismentes helyek spontán hasadását az alkalmazott körülmények között (1 ábra). Sajnos eddig nem sikerült egyik szubsztráton sem az UNG reakciótól egyértelműen megkülönböztethető UDE



1. ábra: **Nagyfelbontású gélelektroforetikus eljárás.** A reakciókhoz 60 bp hosszú kettős szálú oligonukleotidot használtam, melyben az 5' végén fluoreszcensen jelölt szál a 31. pozícióban uracilt tartalmazott. A mintákat 7M ureát tartalmazó, 20% TBE-PAGE segítségével analizáltam. Többféle mintafelvívő kódtét, ill. reakció leállítási módot alkalmaztam. Sajnos a rekombináns His-UDE az adott körülményben, az adott szubsztráton nem aktív. Tanulság azonban, hogy az UNG által okozott AP helyek nagyon labilisak és nemcsak az irodalomban ismert lúgos kezelés (*) hatására hasadnak el, de az általam alkalmazott formamidós mintakódtétben (*F prepared*) hődenaturálva is. Részleges hasadás tapasztalható a 10M ureás ill. a vásártolt formamidós loading dye (*commercial*) alkalmazása esetén is. Ezért, a megfelelő szubsztrát/körülmény megtalálása után is szükséges az adatok körültekintő interpretálása, ill. független módszerrel történő megerősítése.

aktivitást detektálni. A felmerült problémát több úton kíséreltem megoldani: a fehérje előállításának/kezelésének optimalizálásával, a reakcióhoz adott esetleges effektor hatású anyagok adagolásával, ill. a szubsztrát oligonukleotidok szekvenciájának változtatásával. Az aktivitás jellemzésére a közeljövőben további erőfeszítéseket teszek.

A His-taggel ellátott rekombináns fehérje preparátumokat mindig nagy mennyiségű nukleinsav szennyezte. Feltételeztük, hogy esetleg ez interferálhat az aktivitás tesztünkkel, ezért az addig manuálisan kivitelezett, Ni-agaróz affinitás oszlopon történő tisztítást felváltottam gradiens elúciót, ill. folyamatos, több hullámhosszon történő monitorozást lehetővé tevő műszeres (Akta Purifier, GE Healthcare Ltd.) elválasztásra.



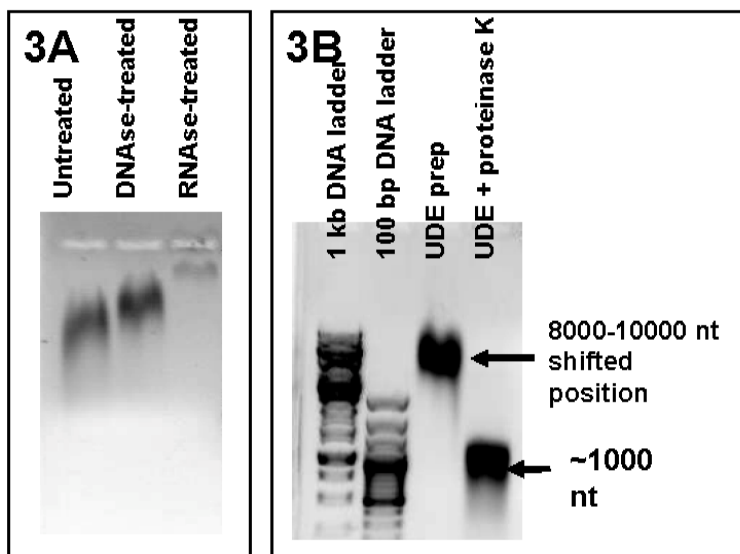
2. Ábra A rekombináns His-UDE fehérje kromatográfiás profilja. **2A:** A His-UDE preparátum két elkülönülő fehérje populációt tartalmaz. A nukleinsav-kötött (1. csúcs, 1. minta a mellékelt SDS gélen) és nukleinsav-mentes (2. csúcs, 2. minta) fehérje eltérő imidazol koncentrációnál eluálódik a Ni-oszlopról. **2B:** A kétféle His-UDE géliszűrésének kromatogramma. A nukleinsavat tartalmazó populációból egy az átesőben jelentkező, óriáskompleyre jellemző csúcsot kapunk (fekete). Míg a nukleinsav-mentes UDE 14,5 ml-nél eluálódik (szürke), ami 57 kD látszólagos molekula-tömeget jelent. **2C:** A His-UDE által kötött nukleinsav RNS. A kezeletlen (fekete) és DNase-zal kezelt (sötét szürke) UDE géliszűrés profilja megegyezik. RNase kezelés hatására eltűnik a komplex csúcs, és megjelenik a nukleinsav-mentes UDE-ra jellemző (világos szürke).

A gradiens elúció során az UDE két elkülönülő csúcsban eluálódott, 2-300mM imidazolnál egy nukleinsavas frakciót, 600mM imidazolnál pedig egy többé-kevésbé nukleinsav-mentes frakciót kaptam (2A ábra). A két frakció további géliszűrése alátámasztotta, hogy a nukleinsav minden bizonnyal az UDE-hoz kötött állapotban van jelen (2B ábra). A nukleinsav tartalmú frakció DNase-kezelés utáni géliszűrés kromatogramma nem változott, míg RNase-kezelés hatására a komplex csúcs mérete drasztikusan lecsökkent, ugyanakkor megjelent a nukleinsav-mentes UDE-ra jellemző csúcs (2C ábra).

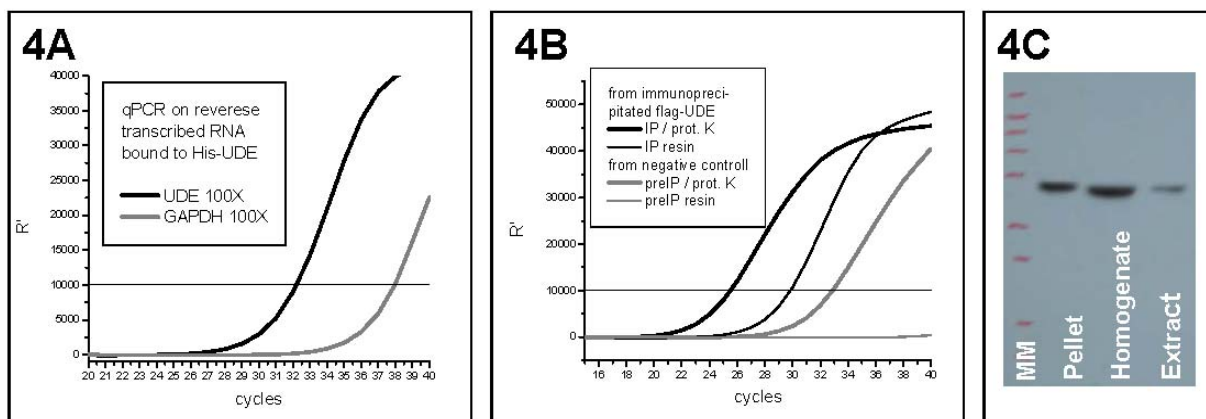
A kötött nukleinsavat sikerült agaróz gélen is kimutatni, ami viszonylagos méretbeli homogenitásra is utal (3A ábra). Az RNase-os, ill. DNase-os emésztés itt is megerősítette a kötött nukleinsav RNS voltát (3A ábra), ill. a minta proteináz K-s emésztésével megszüntetve a látszólagos molekula-tömeg eltolódást az RNS valódi mérete 1000 nt körülnek adódott (3B ábra).

A detektált méret és a bakteriális overexpressziós rendszer sajátosságai miatt feltételeztük, hogy ez az RNS az UDE saját mRNS-e lehet, amit q-PCR kísérletben ki is mutattunk (4A. ábra). Kíváncsi voltam, hogy ez az RNS-kötő tulajdonság csupán az overexpressziós rendszerben előállított rekombináns fehérjére jellemző artefaktum, vagy az UDE-nek valóban jellemző tulajdonsága. Eukarióta expressziós rendszerben is kifejeztem a rekombináns fehérjét. SF9 rovarsejtekben jól expresszálódott, de a feltárás során nagy mennyiség csapadékban maradt (4C. ábra), a tisztított fehérje nagy arányban tartalmazott riboszómális

fehérjeszennyezőket (tömegspektrometriás adatok), továbbá q-PCR segítségével kimutatható volt az UDE mRNS-ének jelenléte a fehérje preparátumban (4B. ábra).

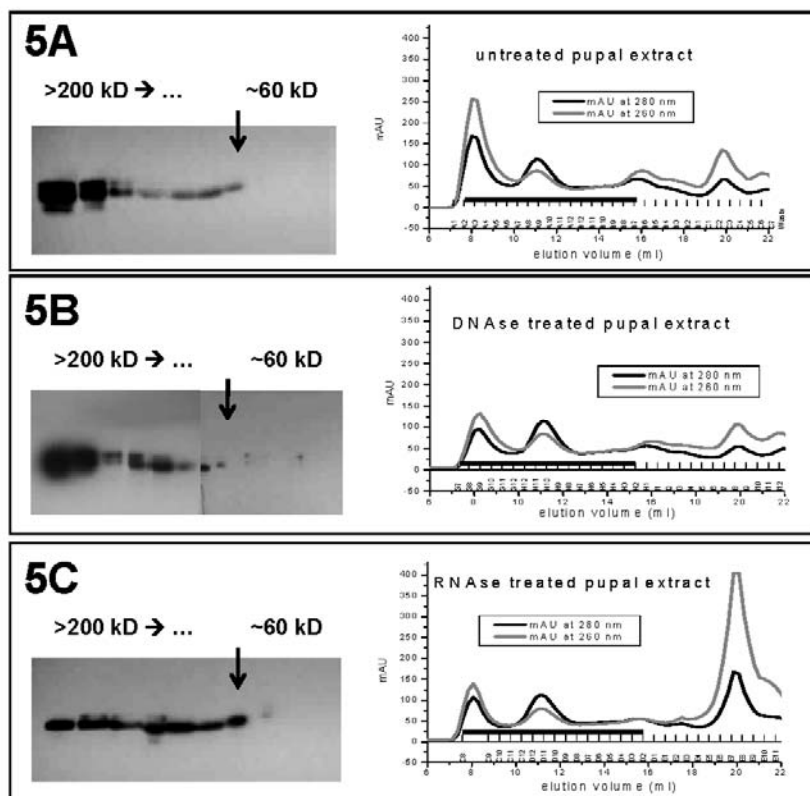


3. Ábra A rekombináns His-UDE fehérje nukleinsav tartalma. 3A: A His-UDE preparátumban lévő nukleinsav RNS. A nukleinsavas UDE agaróz gélen detektálható, DNase-rezisztens, ugyanakkor RNase-szenzitív jelet ad. 3B: A His-UDE által kötött RNS kb. 1000 nt hosszú molekula. Az UDE fehérjét proteináz K-val emésztve a visszamaradó RNS ugyanolyan intenzitással detektálható, de megszűnik a fehérjével való komplexálódás okozta eltolódás. Az 1kb-os és a 100 pb-os DNS marker alapján, a látszólagos mérete 1000 nt körüli.



4. Ábra Az UDE fehérje köti a saját mRNS-ét. 4A: Q-PCR a His-UDE preparátumon. A His-UDE RNS-ét oligoT primerrel cDNS-sé konvertáltam, majd UDE-ra (fekete), ill. GAPDH (szürke) háztartási fehérje génjére specifikus primerekkel q-PCR-t végeztem. Ez alapján a kötött RNS nagy része az UDE saját RNS-e lehet. 4B: Q-PCR az eukarióta rendszerben termelt flag-UDE preparátumon. Az SF9 sejtextraktból immunoprecipitált flag-UDE felét gyantával együtt (vastag fekete), másik felét proteináz K-val történő elúció után (vékony fekete) használtam reverz transkripcióhoz, majd qPCR-hez. Negatív kontrollként a preimmun szérummal végzett immunoprecipitálásból indultam ki (szürkék). 4C: Az eukarióta rendszerben termelt flag-UDE csökkent szolubilitása. A Western bloton látható, hogy a homogenizátumban jelenlévő flag-UDE nagy része centrifugálás után a csapadékban marad. Nagyobb só koncentrációval a szolubilitás csupán kis mértékben emelhető (itt nem közölt adat). A homogenizálás hatékonyságát mikroszkóppal ellenőriztük.

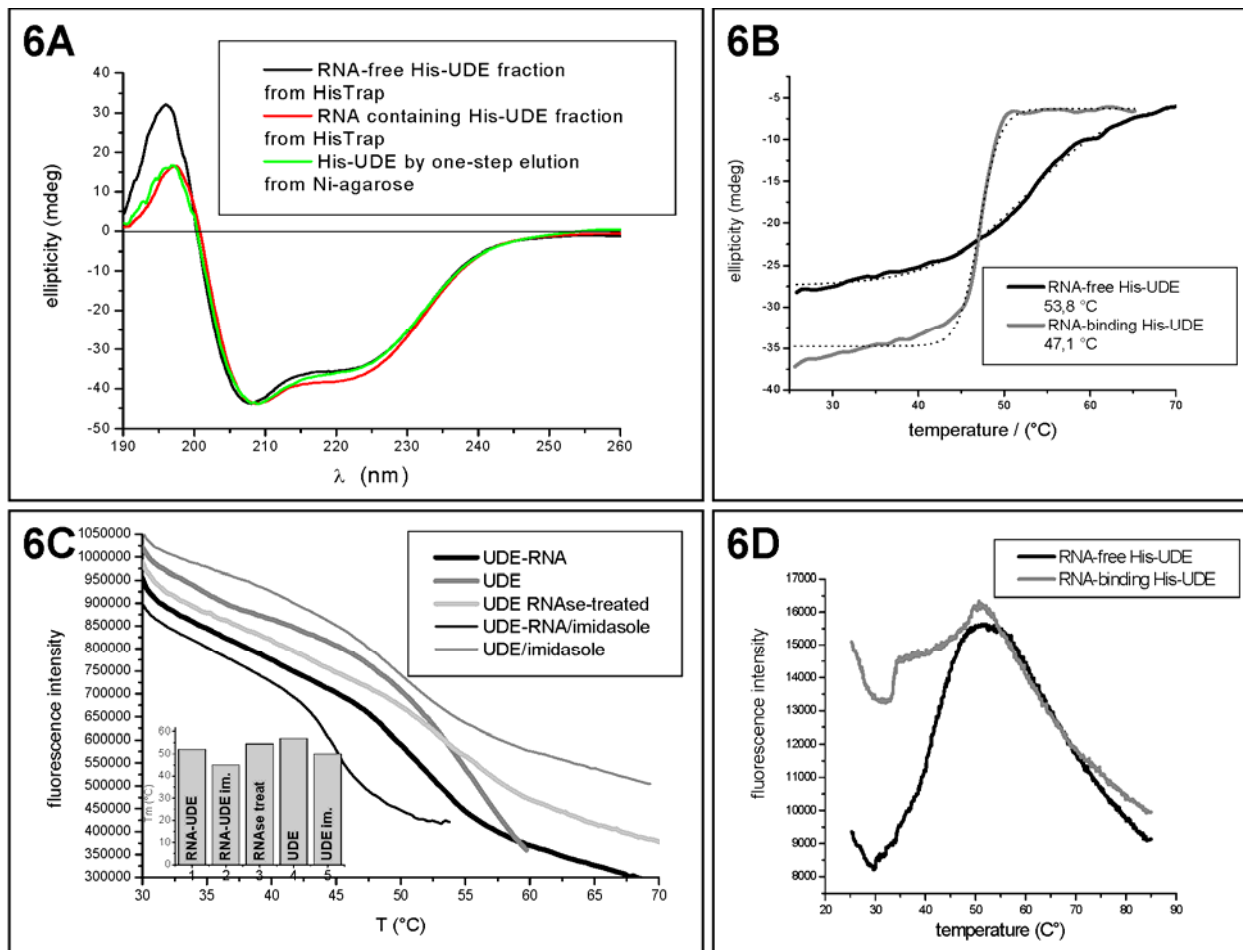
Drosophila báb extrakt kezeletlen, RNase-, ill. DNase-kezelt aliquotjait gélzúréssel frakcionáltam, a frakciókat Western bloton analizáltam, mely eredmény szintén alátámasztotta, hogy az UDE-re fiziológias állapotban is jellemző egy nagymértékű komplexálódás, melyben RNS molekula/ák is fontos szerepet játszanak (5. ábra). Ez alapján feltehető, hogy az UDE RNS-kötése nem csupán a rekombináns fehérjére jellemző artefaktum. *Felmerült az RNS-kötés jelentőségének, szerepének kérdése. Mennyiben befolyásolja az RNS-kötés az aktivitást, vagy a lokalizációt, ill. képes-e az UDE fehérje valamiképp szabályozni saját RNS-ének további processzálását. Ezen kérdéseket a jövőben további vizsgálat tárgyává tesszük.*



5. Ábra: *Drosophila* báb extraktban a fiziológiás UDE-re nagy mértékű komplexálódás jellemző
5A: Western blot (balra) a kezeltlen bábextrakt gélszűrt frakcióiból (jobbra) Látható, hogy a fiziológiás UDE zöme a nagy méretű komplexeket tartalmazó áteső frakciókban található.
5B: Western blot a DNase-zal kezelt báb extrakt gélszűréséből. Szignifikáns változás nem látható az UDE frakciók közti eloszlásában a kezeltlenhez viszonyítva.
5C: Western blot a RNase-zal kezelt báb extrakt gélszűréséből. Az UDE eloszlása a különböző méretekhez rendelhető frakciók közt szignifikánsan megváltozik, csökken a komplexálódás mértéke.

Mivel a két forma a kromatográfiai eljárások során két elkülönülő csúcsban eluálódik, felmerült a kérdés, hogy van-e valamilyen konformációs különbség köztük, ill. feltételezhető-e köztük egyfajta dinamikus egyensúly. Ezért vizsgáltam a rekombináns His-UDE RNS-kötött és RNS-mentes formája közti konformációs és stabilitásbeli különbségeket. A két forma távoli UV-ban rögzített CD spektruma kismértékű eltérést mutat (6A ábra), mely különbség finomabb elemzése jelenleg folyamatban van. Fluorimetriás és CD-s felfűtési görbék alapján az RNS-kötött forma mintegy 5°C-kal alacsonyabb olvadáspontú, mint a nukleinsav-mentes UDE, ugyanakkor plazmid DNS-sel komplexált forma további stabilizálódást eredményez (6B ábra).

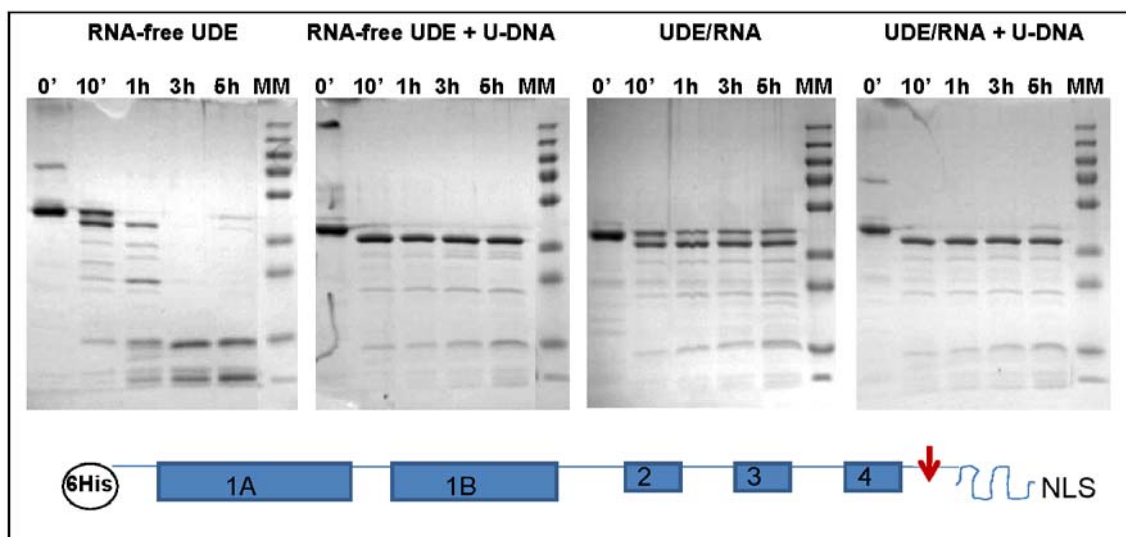
Az ún. thermofluor mérésekben a fokozatos felfűtéssel fokozatosan denaturálódó fehérje növekvő hidrofób felszínéhez kötődő SyproOrange fluoreszcens festék növekvő fluoreszcenciáját követjük, a kapott szigmoid görbe inflexiós pontja megadja a T_m-et. A thermofluor mérést 96 lyukú lemezekben q-PCR készülékben végeztem, ami lehetőséget adott számos oldatkörülmény (pH, puffer, sókoncentráció, EDTA) és hozzáadott kémiai ágensek (uracil, dezoxiuridin, ATP, fémionok) stabilitásra gyakorolt esetleges hatásának vizsgálatára is. Ezek a mérések megmutatták, hogy az alacsony pH-k nem kedvezőek, a túl nagy vagy túl kicsi sókoncentráció bizonyos körülmények közt szintén kedvezőtlen, ill. a hozzáadott szabad uracil képes stabilizálni a fehérjét. A thermofluor mérés során azonban mindvégig mintegy 10°C-kal alacsonyabb T_m-et kaptam az RNS-mentes UDE-ra, mint a CD-ben vagy a triptofán fluoreszcenciát követve. A SyproOrange, mint kémiai ágens, destabilizáló hatását kizártam fluorimetriás és CD kontroll kísérletekben. A thermofluor sajnos alkalmatlan az RNS-kötött forma olvadáspontjának meghatározására, mert ez esetben a kiindulási állapotban is jelentősen nagyobb a fluoreszcens jel a tiszta UDE-éhoz képest (6C ábra), ami viszont arra is utal, hogy a két forma közt jelentős konformációs különbségek lehetnek. Feltételezésünk szerint az RNS-kötés egy kevésbé stabil, nagyobb hidrofób felszíneket jelentő, „nyitott” UDE konformációt stabilizálhat, míg az RNS-mentes UDE frakció egy kompaktabb, stabilabb, „zárt” konformációt képviselhet.



6. Ábra: Az RNS-His-UDE és az RNS-mentes His-UDE konformációs és stabilitásbeli különbségei

6A: Az RNS-UDE és UDE távoli UV tartományban felvett CD spektruma. Szolid mértékű eltérés: a 198 nm körüli pozitív sáv megemelkedett az RNS-UDE formában (piros), a 208 nm-nél és 222 nm-nél az α -hélixekre jellemző negatív sávok relatív aránya változik az RNS-mentes UDE-hoz (fekete) képest. A Ni-oszlopról manuálisan egy lépésben eluált UDE spektruma (zöld) kis mértékben mindkettőtől eltér. **6B:** **Hődenaturáció CD görbéje.** Fokozatos felfűtés során az ellipticitás változását 208 nm-en követve. Az RNS-UDE (szürke) kooperatívabb jellegű, de alacsonyabb olvadáspontot mutató görbét ad, mint az RNS-mentes UDE (fekete). **6C:** **Hődenaturáció fluorimetriás görbéi.** Triptofán fluoreszcencia kevésbé adott kooperatív átmenetet tükröző görbéket. A tisztítás során alkalmazott imidazol átmenetileg destabilizálja mindkét UDE frakciót (vékony görbék, mellékelt oszlopdiaagramon 2, 5 vs 1, 4). Dialízis után mintegy 7°C-kal emelkedik mindkét frakció olvadáspontja (vastag fekete és középszürke, oszlopdiaagramon 1, 4). Az RNS-UDE (fekete, 1) mintegy 5°C-kal alacsonyabb olvadásponttal jellemezhető, mint az RNS-mentes (középszürke, 4). Az RNS-UDE RNase-os kezelése (világos szürke, 3) csökkenő kooperativitást és növekvő Tm értékeket eredményez. **6D:** **Thermofluor görbék.** A fehérje hidrofób felszínéhez kötődő SyproOrange festék fluoreszcenciáját követjük a felfűtés során. A görbe kezdeti növekvő szakaszára szigmoid görbe illeszthető, mely inflexiós pontja megadja a Tm-et. Az RNS-UDE (szürke esetén ezek a görbék értékelhetetlenek, mivel szobahőn is magas jelet adnak. Ez egyben a kétféle UDE eltérő konformációjára is utal.

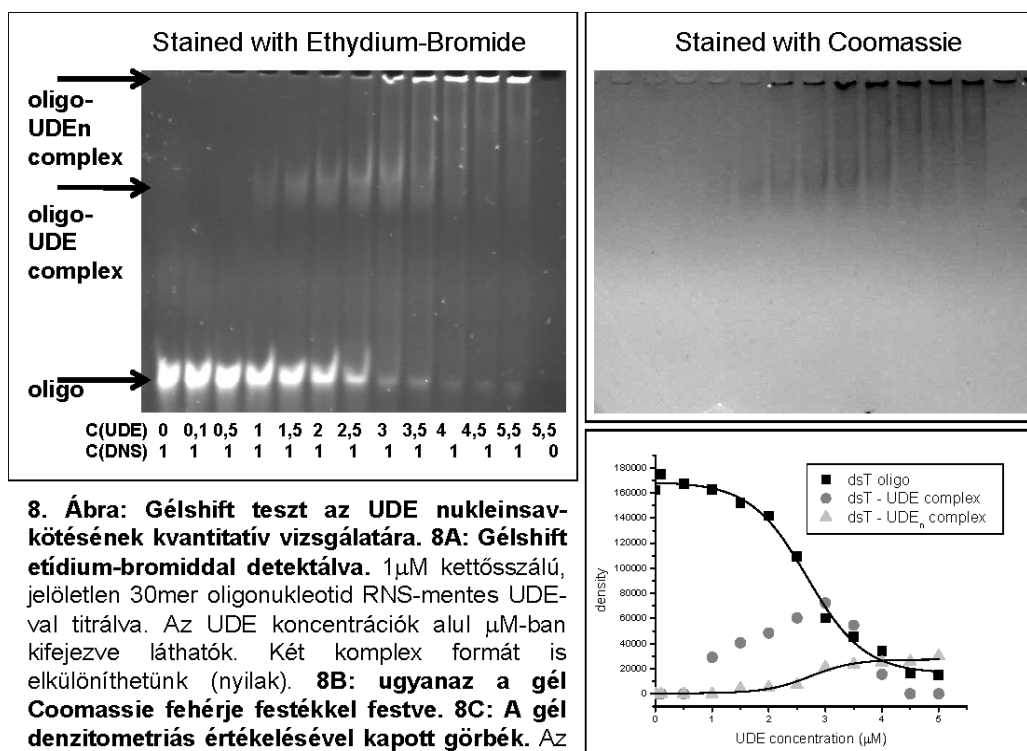
Limitált proteolízissal kimutattuk a DNS- és RNS-kötés által okozott védő hatás eltérő jellegét. A fragmensek tömegspektrometriás azonosításával megállapítottuk, hogy az RNS a fehérje C-terminális részén kiterjedtebben véd, mint a DNS (7. ábra). A DNS mégis képes kiszorítani az UDE által kötött RNS-t a kötőhelyről (7. ábra). Ez alapján részben átfedő kötőhelyeket feltételezhetünk.



7. Ábra: Limitált emésztés Asp-N endoproteinázzal. Az RNS-mentes UDE emésztési profiljához (első) képest a hozzáadott DNS (második) jelentősen véd a proteolizissal szemben: a sematikus ábrán piros nyíllal jelzett helyen történő első hasítás után gyakorlatilag megáll a fehérje további emésztése. Az RNS-UDE emésztési profiljában (harmadik) még ez a hely is védett valamelyest, ami az RNS-kötés kiterjedtebb/részben eltérő kölcsönható felszínére utal. Az RNS-UDE-hoz az RNS-sel összemérhető mennyiségű uracil-tartalmú plazmid DNS-t adva (negyedik) visszakapjuk az UDE + DNS emésztési profilját, ami arra utal, hogy a DNS kiszorította az RNS-t a kötő helyről.

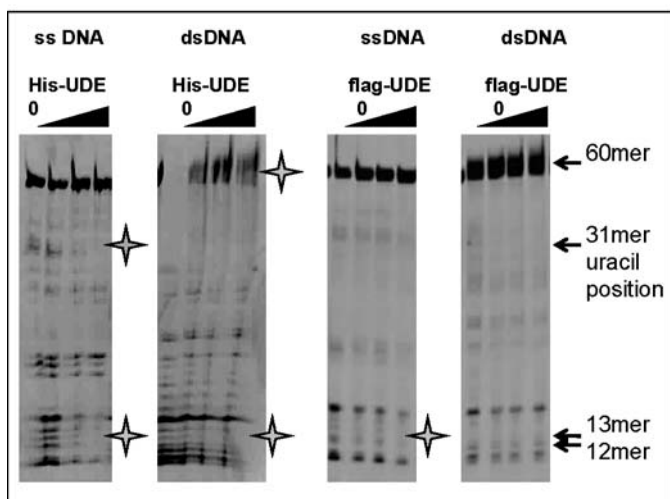
A rekombináns His-UDE RNS- és DNS-kötő képességét gélshift analízis segítségével fogom összehasonlítani, mivel oligonukleotidokkal sikerült a fluorimetriás titrálásnál pontosabb, nagyobb felbontású (külön detektálhatók a különböző sztöchiometriájú komplexek), kisebb anyagigényű módszert, kidolgoznom (8AB ábra). A gélkép denzitomet-rálásával számszerű-síthető a komplexek keletkezése, ill. a szabad oligonukleotid fogyása. Kimutattam,

hogy egy 30mer oligonukleotid legalább két UDE-t képes kötni, a komplexebb formára vonatkozó telítési görbe szigmoid (8C ábra), ami a kötőhelyek kooperativitására, ill. az UDE a DNS-en történő oligomerizálására utalhat. Az eredmények finomítása, ill. az RNS-es mérések jelenleg folyamatban vannak. Az UDE által kötött RNS azonosításáról és az RNS-, ill. DNS-kötésének összehasonlító elemzéséről jelenleg egy kézirat készülöben van.



8. Ábra: Gélshift teszt az UDE nukleinsav-kötésének kvantitatív vizsgálatára. 8A: Gélshift etídium-bromiddal detektálva. 1µM kettősszalú, jelöletlen 30mer oligonukleotid RNS-mentes UDE-vel titrálva. Az UDE koncentrációk alul µM-ban kifejezve láthatók. Két komplex formát is elkülöníthetünk (nyilak). **8B: ugyanaz a gél Coomassie fehérje festékkel festve.** **8C: A gél denzitometriás értékelésével kapott görbék.** Az oligonukleotid fogyására és az oligo-UDEn komplex keletkezésére egyaránt szigmoid görbe illeszthető.

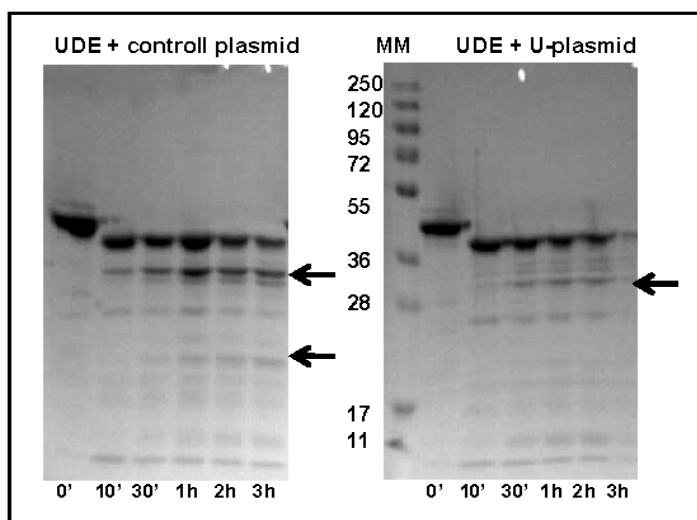
A megfelelően kötődő oligonukleotid megtalálása nemcsak az aktivitás részletes jellemzéséhez, de minden valószínűség szerint a kristályosítás sikeréhez is elengedhetetlen, tekintettel az eddig elvégzett igen nagyszámú, és változatosan variált kristályosítási kísérletek sikeretelenségére. Ezért a jelen munka fő irányát szükséges volt ilyen irányba módosítani. Több módszerrel is vizsgáltuk és vizsgáljuk az UDE oligonukleotid kötését: a fluorimetriás, thermofluor titrálások limitált felbontóképességűek a releváns komplexek kialakulásának követésére, ezért a továbbiakban a gélshift és a footprint szélesebb körű alkalmazását is megkezdjük. A gélshift alkalmasnak tűnik a látszólagos kötődési állandó meghatározásán túl a különböző sztöchiometriájú komplexek meghatározására is (vö. 8. ábra).



9. Ábra: DNase-I footprint. 4 μM (His-UDE esetén), ill. 8 μM (flag-UDE esetén), 5' végen Cy3 festékkel fluoreszcensen jelölt, a 31. helyen uracilt tartalmazó, egyes (ssDNA) és kettős (dsDNA) szálú, 60mer oligonukleotidokat 0-16, ill. 0-24 μM növekvő koncentrációjú UDE-val inkubáltam, majd 1 unit DNase-I enzimmel 30 (His-UDE esetén) ill. 60 (flag-UDE esetén) másodpercig emésztettem. A reakciót EDTA, ill. denaturáló formamidos mintakóttél hozzáadásával állítottam le. A mintákat 20X20 cm, 7M urea, 20% TBE-PAGE-n választottam el. A Cy3 festék fluoreszcenciáját UV lámpa felett detektáltam. Úgy tűnik, hogy az UDE az egyesszálú oligonukleotidokat jobban védi, ill. a kötődés helye (csillaggal jelölt tartományok) az oligo két vége felé található. Mivel ez a fehérje ezen az oligonukleotidon nem aktív, feltehető, hogy a detektált kötődés nem megfelelő. A továbbiakban a módszert az aktivitás szempontjából releváns kötődés azonosítására kívánjuk használni.

A különböző footprint technikákkal pedig jól behatárolható a kötődés helye a DNS-en. Eddig DNase-I-gyel végeztem ilyen kísérleteket (9. ábra), amiből úgy tűnik, hogy a 60mer oligonukleotidon (amin az UDE nem is aktív) a fehérje az 5' és a 3' végekhez közel köt, és nem a középén található uracil környékén. A módszer komoly hátránya egyrészt, hogy a DNase szekvencia preferenciája miatt nem kapunk egyenletes létrát, másrészt, hogy az enzim sztérikus akadályok miatt nem tudja nukleotid pontossággal megközelíteni az UDE fehérje kötő helyét. A hátrányok kiküszöbölhetők bázis-, ill. gerincmódosító kémiai ágensek alkalmazásával (etiláció, metiláció, bázismentes hely interferencia), ilyen assay-k beállítására hamarosan sor kerül.

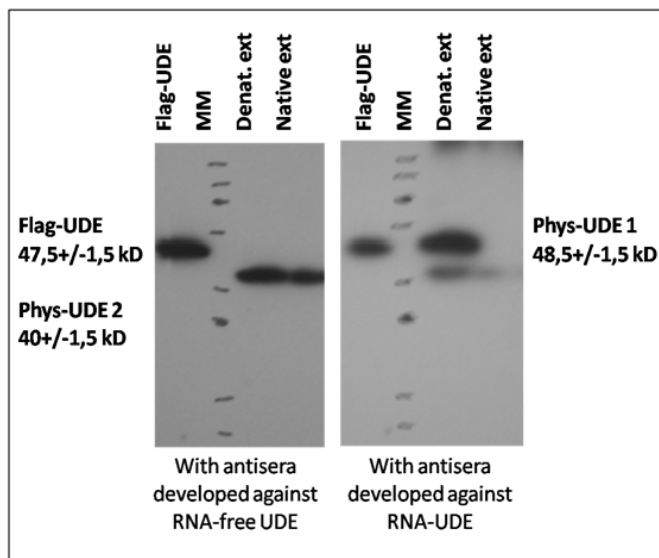
Fontos előrelépést jelenthet továbbá, ha bármilyen szignifikáns különbséget tudunk detektálni adott szekvenciájú és méretű uracilos ill. kontroll oligonukleotid kötődése esetén. Két konkrét tesztet azonosítottunk, mely különbséget detektál az uracilos és a kontroll szubsztrát kötődésében. Egyrészt Pukáncsik Mária PhD hallgató fluorimetriás titrálás során kimutatta, hogy a kontroll, ill. az uracil-tartalmú 30mer oligonukleotidok hasonló erősséggel kötődnek ugyan, de U-DNS kötődés esetén elmarad a kontrollban megfigyelhető eltolódás nagyobb hullámhossz felé. Másrészt nagy specificitású kimotripszinnel emésztve az uracilos, ill. kontroll plazmiddal komplexált fehérjét, eltérő emésztési



10. Ábra: Nagy specificitású kimotripszinnel végzett limitált emésztés során különbség detektálható az uracil-DNS és kontroll DNS kötődése által okozott védés profilja között. A nyilak nyilvánvaló eltérésekre mutatnak. A különbségként jelentkező fragmentumok tömegspektrometriás azonosítása folyamatban van.

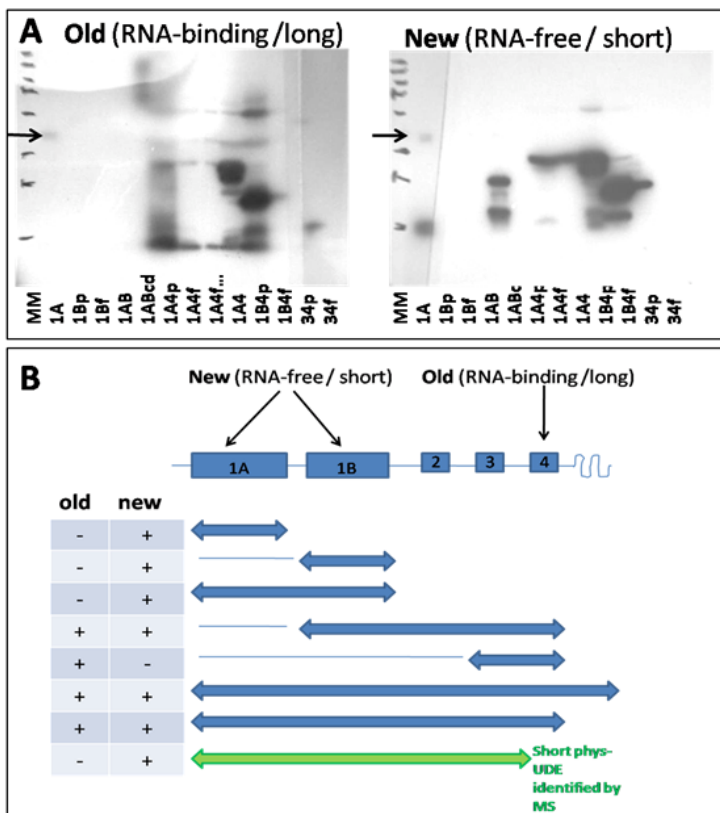
mintázatot kaptam (10. ábra), mely szintén bizonyos konformációs változásokra utal az uracil felismerése során. A hasítóhelyek azonosítása jelenleg folyamatban van.

A pályázat másik fő iránya volt a fiziológiásan jelenlevő feltételezett poszt-transzlációs módosítások jellemzése, ill. aktivitásra/funkcióra gyakorolt hatása.



11. ábra: Két poliklonális anti-UDE antiszérum izoforma specificitása. A Western bloton az RNS-mentes UDE ellen fejlesztett antiszérum (balra) egy, a rekombináns flag-UDE-nál mintegy 7 kD-nal kisebb látszólagos molekulatömegű izoformát (phys-UDE2) ismer fel. Ugyanazt a membránt újra festve az RNS-UDE ellen fejlesztett antiszérummal (jobbra) előtűnik a denaturáló módon előállított mintában (Denat. ext) egy a rekombinánsnál valamivel nagyobb látszólagos molekulatömegű sáv (phys-UDE1). A jobb oldali membránon a rövid izoforma azért adhat jelet, mert az újrafestés előtt az előző blot nem lett visszafestve. Fontos megjegyezni, hogy a hosszú forma szolubilitása sokkal rosszabb, a natív extraktban nem ad jelet.

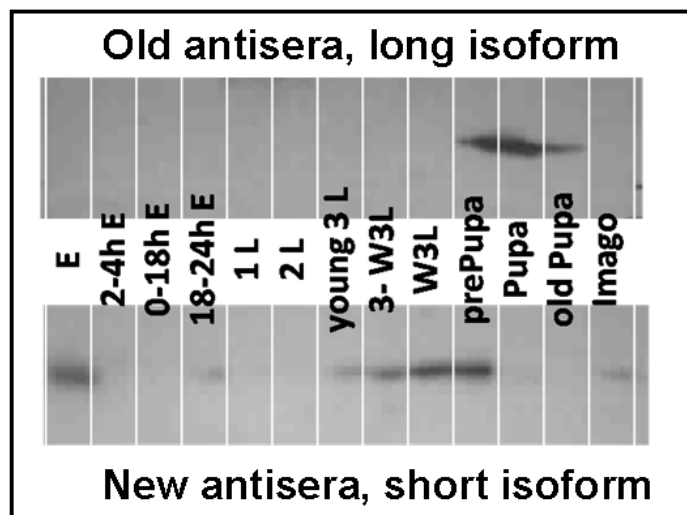
Eredetileg uracil-DNS affinitás kromatográfiával terveztem a fiziológiás formát izolálni erre a célra, de mivel nagy mennyiségben elő tudtam állítani immáron nukleinsav-mentes fehérjét, úgy döntöttünk, hogy érdemesebb lesz új poliklonális antitestet fejleszteni, ami nagyobb eséllyel alkalmazható immunoprecipitálásra is. Az így előállított új antitest meglepő módon az UDE-nek egy új izoformáját ismerte fel (11. ábra).



12. ábra: A két, izoforma-specifikus, poliklonális antiszérum domináns epitópjainak körülírása rekombináns úton előállított UDE fragmensek segítségével. **A:** A rekombináns fragmensek különböző preparátumain végzett Western blot a hosszú formára specifikus régi antitesttel (balra), ill. a rövid izoformára szelektív új antitesttel (jobbra). **B:** összefoglaló táblázat, mely alapján leszűkíthető a két antiszérum domináns epitópjainak lokalizációja. A rövid izoformát immunoprecipitálás után tömegspektrometriával analizáltuk. Az első adatok szerint a C-terminálison rövidül, mely rövidülés feltehetőleg érinti a 4. motívumot is (zöld nyíllal jelzett). Az eredményeket összevetve jelenleg a következő hipotézis állítható fel: A régi antitest domináns epitópjá valahol a 4. motívum környékén lehet, ezért nem ismeri fel a fiziológiás rövid izoformát. Az új antitest domináns epitópjá az egymással is homológ 1A és 1B motívumnál lehet. Feltételezésünk szerint ebben a régióban elképzelhető valamilyen poszt-transzlációs módosítás a hosszú formában, ami interferálhat az új antitesttel való kölcsönhatásával. Ez a módosítás a rekombináns UDE-ről hiányzik.

Immunoprecipitálással izoláltam ezt az izoformát, ami az első tömegspektrometriás adatok szerint a C-terminálison csonkolt forma, elveszítve nemcsak a nukleáris lokalizációs szignálját, de feltehetőleg a negyedik konzervált motívumát is. Ezután rekombináns úton előállított UDE fragmensek segítségével részletesen jellemeztem a két poliklonális antitestre jellemző domináns epitópot, melyek magyarázatot adtak a két poliklonális antitest meglepő izoforma-specificitására (12. ábra).

Továbbá jellemeztem az újonnan azonosított izoforma expresszióját az egyedfejlődés során, és kimutattam, hogy a hosszú forma bábállapotban történő megjelenését mintegy előkészíti a rövid forma mennyiségének 3. lárva stádium elejétől megfigyelhető fokozatos emelkedése (13. ábra). Ez a mintázat, valamint megelőző RT-PCR ill. Northern blot eredményeink alapján azt feltételezzük, hogy a rövid izoforma poszt-transzlációs enzimátikus hasítás eredményeképp jön létre, ami az amúgy esetleg veszélyes aktivitású UDE-t hatástalanítja a fejlődés korábbi szakaszaiban. Ez a poszt-transzlációs szabályozás oldódhat fel valamiképp a 3. lárvastádium végén, ill. a prepupa szakaszban. A szabályozás pontos mechanizmusának leírására a jövőben további kísérleteket tervezünk. Az SF9 rovarsejtben termelt rekombináns UDE-n előkísérletben kimutattunk egy foszforilációs helyet. A foszforiláció jelentőségének tisztázása, valamint a fiziológiásan jelenlévő két izoforma foszforiláltságát jelenleg vizsgáljuk.



13. ábra: A két izoforma eltérő expressziós mintázata a *Drosophila melanogaster* egyedfejlődése során. A rövid izoforma, mely minden valószínűség szerint nem tartalmazza a nukleáris lokalizációs szignált, sem a 4. motívumot, expressziós mintázata (alsó blot) nagyban eltér a bábállapokra specifikus hosszú formáétól (felső blot). A rövid forma jelen van embrióban (E), azon belül is a késői embrióban (18-24h E), illetve a 3. lárva stádium elejétől, mintegy előkészíti a hosszú forma megjelenését a prepupa szakaszban; bábállapotban (pupa) eltűnik, ahol a hosszú forma nagy mértékben expresszálódik, majd a kifejlett légyben (imago), a hosszú forma eltűnésével újra megjelenik. Hipotézisünk szerint a rövid forma a hosszú poszt-transzlációs hasításával keletkezhet, amely hasítás feltehetőleg a fehérjeszintézist követően közvetlenül megtörténik, és az UDE aktivitás poszt-transzlációs szabályozása részét képezi.

Az izoformákra, azok szabályozására és poszt-transzlációs módosításukra vonatkozó eredményeket egy újabb kéziratban szeretnénk közölni még ebben az évben.

A fentiekben bemutatott eredmények jelentős előrelépést jelentenek ennek az érdekes és minden ízében újszerű fehérjének a vizsgálatában. A munka korántsem befejezett, az aktivitás részletes jellemzése és a nagyfelbontású atomi szerkezet meghatározása még mindig várat magára. A projekt megvalósítása során született eredmények kirajzolják a további kutatás irányát: megfelelő oligonukleotid szubsztrát keresése, az RNS-kötés szerepe, az izoformák regulációjának leírása, poszt-transzlációs módosítások azonosítása. Így eljuthatunk az aktivitás részletes jellemzéséhez és a szerkezet meghatározásához, mely tudás elengedhetetlen az UDE fehérje jövőbeli felhasználásához.

Hivatkozott irodalom:

1. Bekesi, A., et al., *A novel fruitfly protein under developmental control degrades uracil-DNA*. *Biochem Biophys Res Commun*, 2007. **355**(3): p. 643-8.
2. Bekesi, A., et al., *Developmental regulation of dUTPase in Drosophila melanogaster*. *J Biol Chem*, 2004. **279**(21): p. 22362-70.

1. számú melléklet: kézirat revízió alatt a “FEBS Journal”-ben

Domain organization of a novel uracil-DNA degrading factor reveals potential conserved DNA binding surfaces and rationalizes a physiologically observed truncation

Mária Pukáncsik¹, Angéla Békési¹, Eva Klement², Éva Hunyadi-Gulyás², Katalin F. Medzihradszky^{2,3}, Jan Kosinski^{4,5}, Janusz M. Bujnicki⁴ and Beáta G. Vértessy¹

¹Institute of Enzymology, Biological Research Center, Hungarian Academy of Sciences, Budapest, Hungary,

²Proteomics Research Group, Biological Research Center, Hungarian Academy of Sciences, Szeged, Hungary,

³Department of Pharmaceutical Chemistry, University of California, San Francisco, USA,

⁴Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell Biology, Warsaw, Poland,

⁵Institute of Biochemistry and Biophysics PAS, Warsaw, Poland

Abstract

Uracil in DNA is usually considered to be a mistake, however, it may be used for signalling in *Drosophila* development *via* recognition by a novel uracil-DNA degrading factor (UDE). The UDE protein has no detectable similarity to any other uracil-DNA binding factors and has no structurally or functionally described homologues; therefore its characterization is indispensable for understanding this recognition process. Here, combination of theoretical and experimental analyses reveals domain organisation and DNA-binding pattern of UDE. Sequence alignments and limited proteolysis with different proteases show extensive protection by DNA at the N-terminal duplicated conserved Motif 1A/1B segment, and a well-folded domain within the C-terminal half encompassing conserved Motifs 2-4. Fold prediction models for Motif 1A and 1B suggest similar α -helical bundles and two conserved positively charged surface patches that may bind DNA. Circular dichroism spectroscopy also reinforce presence of α -helices in UDE. Full functionality of a physiologically occurring truncated isoform in *T. castaneum* lacking one copy of the N-terminal conserved Motif 1 is revealed by activity assays of a representative truncated construct of *D. melanogaster* UDE. Gel filtration and analysis of predicted structural models suggest a possible dimerization mechanism for preserving functionality of the truncated isoform.

Introduction

The nucleobase uracil is not a normal constituent of DNA, although as compared to thymine (i.e. 5-methyl-uracil), it provides the same Watson-Crick interaction pattern for adenine and is actually used as the adenine-counterpart base in RNA. Despite its usual absence, there are two physiological ways for uracil to appear in DNA: cytosine deamination and thymine replacement. Cytosine to uracil transitions *via* hydrolytic deamination are among the most frequently occurring spontaneous mutations. These generate premutagenic U: G mispairs [1,2]. Thymine replacement by uracil can occur if the cellular dUTP/dTTP ratio increases since most DNA polymerases will incorporate either U or T against A, based solely on availability of the corresponding building block nucleotides [3,4]. Thymine-replacing uracil moieties are not mutagenic, as these provide the same genomic information, but may perturb binding of factors that require the 5-methyl group on the uracil ring for recognition.

There are also two mechanisms to ensure uracil-free DNA: prevention and excision. dUTPases prevent uracil incorporation into DNA by removing dUTP from the DNA polymerase pathway [5]. Uracil in DNA, either produced by cytosine deamination or uracil misincorporation, is excised by uracil-DNA glycosylases (UDGs) during the base excision repair (BER) pathway [6,7]. Among the different UDGs, the protein product of the *ung* gene (termed UNG) is by far the most efficient in catalyzing uracil excision [8].

UNG is responsible for most of the repair process since its mutations in *E. coli*, mouse and human was found to induce considerable increase in uracil content [9-11].

Null mutations in the dUTPase gene (*dut*) result in non-viable phenotype that can be rescued by a second null-mutation in the *ung* gene. The *dut-ung-* phenotype presents mass uracil incorporation into DNA [9,12]. Interestingly, an analogous situation with simultaneous lack of dUTPase and UNG activities arises in *Drosophila* larvae under physiological conditions. On one hand, the *ung* gene coding for the major UDG enzyme is not present in the *Drosophila* genome [13]. On the other hand, it has been shown that dUTPase level is under detection limit in larval tissues and the enzyme is present exclusively in the imaginal discs [14]. Simultaneous lack of UNG and dUTPase may lead to accumulation of uracil-substituted DNA in fruitfly larval tissues. A specific protein termed uracil-DNA degrading factor (UDE), was identified in *Drosophila* late larvae and pupae that recognizes and degrades uracil-DNA [15], strengthening the hypothesis that *Drosophila melanogaster* may use uracil-DNA as a signal to turn on metamorphosis related cell death [15] [16,17].

UDE, the recently described uracil-DNA degrading factor, constitutes the first member of a new protein family recognizing uracil-DNA. It has no glycosylase activity and its sequence does not show any appreciable similarity to other nucleases or uracil-DNA recognising proteins [15] (Figure 1). Homologous protein sequences were found only in translated genomes of other pupating insects, however, no structural or functional data have been published on any of these putative proteins. In all these homologous sequences, four distinct conserved sequence motifs could be identified (Motifs 1-4), among which the first motif is substantially longer and is usually present in two copies (Motifs 1A and 1B). Comparison of these motifs to motifs in uracil-DNA glycosylases does not offer any clue regarding the structure and function of UDE, since no apparent similarity could be observed (Figure 1B)[18]. Therefore, structural analysis of UDE is indispensable to define the function of this novel protein family. Investigation of this protein may offer new insights into the physiological role and catalytic mechanism of nucleases.

Towards this goal, in the present study we probed the domain organisation of UDE from *D. melanogaster*, expressed as a recombinant protein (Dm_{rc}UDE), by limited proteolysis and revealed that a specific truncated fragment lacking the N-terminus may fold into a stable conformation. Interestingly, we also identified such a truncated physiologically occurring UDE isoform from the pupating insect, *Tribolium castaneum* (TcUDE) [19]. The TcUDE isoform lacks one copy of the N-terminal duplicated first motif. We generated the respective segment from Dm_{rc}UDE by chemical cleavage with hydroxylamine and found that this truncated segment preserves catalytic specificity and activity. The structural results therefore offer an explanation for the physiological existence of the truncated isoform. *De novo* modelling was performed using ROSETTA and a 3D structural model was constructed for the tandemly duplicated N-terminal motifs 1A and 1B. The model suggests that both motifs constitute similar three-helical bundles with the same topology and relative orientation. High content of helical secondary structure in UDE was also independently confirmed by circular dichroism. The predictions together with the domain organisation studies offer a model of DNA-binding to an extended surface on the protein along the conserved motifs.

Results

Identification of a physiologically occurring truncated isoform of UDE

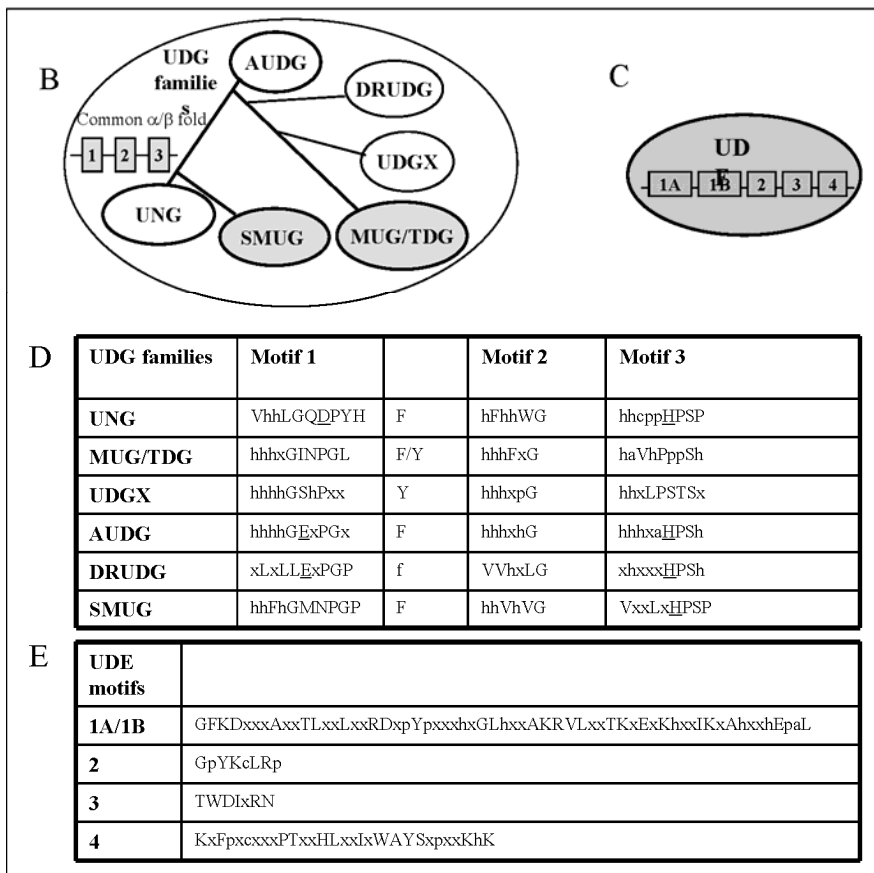
BLAST searches indicated that UDE has homologues only in genomes of pupating insects (Figure 1A.) The multiple sequence alignment shows four conserved motifs (Figure 1A and 1D). The first extended UDE motif is present in two highly similar copies. The UDE homologue from *T. castaneum* contains only one copy of motif 1, suggesting that lack of the first motif may still result in a functional protein (Figure 1A).

```

D. mel. 40 MAR--GDSPFFPKMEALSTKLEEREMQVRLTVRGLLGAQVSTFKARELEKI MA XGVVPSKRE ENGGAS ENNAKT -DDEDFVE---TVPDGFDAARAPELIIA 150
D. ana. 1 MRE--GEFFPKMEALSTKLEEREMQVRLTVRGLLGAQVSTFKARELEKI MA XGVVPSKRE ENGGAS ENNAKT -DDEDFVE---TVPDGFDAARAPELIIA 111
D. ere. 1 MRE--GEFFPKMEALSTKLEEREMQVRLTVRGLLGAQVSTFKARELEKI MA XGVVPSKRE ENGGAS ENNAKT -DDEDFVE---TVPDGFDAARAPELIIA 111
D. sec. 1 MRE--GDSPFFPKMEALSTKLEEREMQVRLTVRGLLGAQVSTFKARELEKI MA XGVVPSKRE ENGGAS ENNAKT -DDEDFVE---TVPDGFDAARAPELIIA 111
D. yak. 1 MRE--GDSPFFPKMEALSTKLEEREMQVRLTVRGLLGAQVSTFKARELEKI MA XGVVPSKRE ENGGAS ENNAKT -DDEDFVE---TVPDGFDAARAPELIIA 111
D. per. 2 MRE--GEFFPKMEALSTKLEEREMQVRLTVRGLLGAQVSTFKARELEKI MA XGVVPSKRE ENGGAS ENNAKT -DDEDFVE---TVPDGFDAARAPELIIA 111
D. pse. 1 MRE--GEFFPKMEALSTKLEEREMQVRLTVRGLLGAQVSTFKARELEKI MA XGVVPSKRE ENGGAS ENNAKT -DDEDFVE---TVPDGFDAARAPELIIA 111
D. wil. 1 MRE--GEFFPKMEALSTKLEEREMQVRLTVRGLLGAQVSTFKARELEKI MA XGVVPSKRE ENGGAS ENNAKT -DDEDFVE---TVPDGFDAARAPELIIA 111
D. gri. 1 MRE--GEFFPKMEALSTKLEEREMQVRLTVRGLLGAQVSTFKARELEKI MA XGVVPSKRE ENGGAS ENNAKT -DDEDFVE---TVPDGFDAARAPELIIA 111
D. vir. 101 MRE--GEFFPKMEALSTKLEEREMQVRLTVRGLLGAQVSTFKARELEKI MA XGVVPSKRE ENGGAS ENNAKT -DDEDFVE---TVPDGFDAARAPELIIA 231
D. moj. 1 MRE--GEFFPKMEALSTKLEEREMQVRLTVRGLLGAQVSTFKARELEKI MA XGVVPSKRE ENGGAS ENNAKT -DDEDFVE---TVPDGFDAARAPELIIA 111
A. aeg. 1 MRE--GEFFPKMEALSTKLEEREMQVRLTVRGLLGAQVSTFKARELEKI MA XGVVPSKRE ENGGAS ENNAKT -DDEDFVE---TVPDGFDAARAPELIIA 111
A. gam. 2 MRE--GEFFPKMEALSTKLEEREMQVRLTVRGLLGAQVSTFKARELEKI MA XGVVPSKRE ENGGAS ENNAKT -DDEDFVE---TVPDGFDAARAPELIIA 112
C. pip. 1 MRE--GEFFPKMEALSTKLEEREMQVRLTVRGLLGAQVSTFKARELEKI MA XGVVPSKRE ENGGAS ENNAKT -DDEDFVE---TVPDGFDAARAPELIIA 111
A. mel.* 6 STE--FEESFPEFQGLQVTLALDGDITQVWVAFFVFAVSTLQVTRERELEKI MA XGVVPSKRE ENGGAS ENNAKT -DDEDFVE---TVPDGFDAARAPELIIA 276
D. mor. 49 MQRDDEQDQFQFKAARLRLLEERELNPLVTVRGLLGAQVSTFKARELEKI MA XGVVPSKRE ENGGAS ENNAKT -DDEDFVE---TVPDGFDAARAPELIIA 161
T. cas. -----NREEDF---TVPDGFDAARAPELIIA 28
motif 1A
D. mel. RRFDFCKLAKLIGERRVLSVDFRREIATATREVVQVLEDFEFKFAERLDFDRAVLEAVVLEL-----DFDRLARFLEAAVGGFAFVHMTWPFADDTTHQVLS 244
D. ana. RRFDFCKLAKLIGERRVLSVDFRREIATATREVVQVLEDFEFKFAERLDFDRAVLEAVVLEL-----DFDRLARFLEAAVGGFAFVHMTWPFADDTTHQVLS 225
D. ere. RRFDFCKLAKLIGERRVLSVDFRREIATATREVVQVLEDFEFKFAERLDFDRAVLEAVVLEL-----DFDRLARFLEAAVGGFAFVHMTWPFADDTTHQVLS 225
D. sec. RRFDFCKLAKLIGERRVLSVDFRREIATATREVVQVLEDFEFKFAERLDFDRAVLEAVVLEL-----DFDRLARFLEAAVGGFAFVHMTWPFADDTTHQVLS 225
D. yak. RRFDFCKLAKLIGERRVLSVDFRREIATATREVVQVLEDFEFKFAERLDFDRAVLEAVVLEL-----DFDRLARFLEAAVGGFAFVHMTWPFADDTTHQVLS 225
D. per. RRFDFCKLAKLIGERRVLSVDFRREIATATREVVQVLEDFEFKFAERLDFDRAVLEAVVLEL-----DFDRLARFLEAAVGGFAFVHMTWPFADDTTHQVLS 225
D. pse. RRFDFCKLAKLIGERRVLSVDFRREIATATREVVQVLEDFEFKFAERLDFDRAVLEAVVLEL-----DFDRLARFLEAAVGGFAFVHMTWPFADDTTHQVLS 225
D. gri. RRFDFCKLAKLIGERRVLSVDFRREIATATREVVQVLEDFEFKFAERLDFDRAVLEAVVLEL-----DFDRLARFLEAAVGGFAFVHMTWPFADDTTHQVLS 225
D. wil. RRFDFCKLAKLIGERRVLSVDFRREIATATREVVQVLEDFEFKFAERLDFDRAVLEAVVLEL-----DFDRLARFLEAAVGGFAFVHMTWPFADDTTHQVLS 225
D. vir. RRFDFCKLAKLIGERRVLSVDFRREIATATREVVQVLEDFEFKFAERLDFDRAVLEAVVLEL-----DFDRLARFLEAAVGGFAFVHMTWPFADDTTHQVLS 244
D. moj. RRFDFCKLAKLIGERRVLSVDFRREIATATREVVQVLEDFEFKFAERLDFDRAVLEAVVLEL-----DFDRLARFLEAAVGGFAFVHMTWPFADDTTHQVLS 224
A. aeg. RRFDFCKLAKLIGERRVLSVDFRREIATATREVVQVLEDFEFKFAERLDFDRAVLEAVVLEL-----DFDRLARFLEAAVGGFAFVHMTWPFADDTTHQVLS 225
A. gam. RRFDFCKLAKLIGERRVLSVDFRREIATATREVVQVLEDFEFKFAERLDFDRAVLEAVVLEL-----DFDRLARFLEAAVGGFAFVHMTWPFADDTTHQVLS 229
C. pip. RRFDFCKLAKLIGERRVLSVDFRREIATATREVVQVLEDFEFKFAERLDFDRAVLEAVVLEL-----DFDRLARFLEAAVGGFAFVHMTWPFADDTTHQVLS 227
A. mel. RRLDKVHLVDFVHRAKLVDFDFRREIATATREVVQVLEDFEFKFAERLDFDRAVLEAVVLEL-----LAKVLEEDVFLAVV-----EVDFDGFDAARAPELIIA 390
D. mor. RRFDFCKLAKLIGERRVLSVDFRREIATATREVVQVLEDFEFKFAERLDFDRAVLEAVVLEL-----DFDRLARFLEAAVGGFAFVHMTWPFADDTTHQVLS 272
T. cas. RRFDFCKLAKLIGERRVLSVDFRREIATATREVVQVLEDFEFKFAERLDFDRAVLEAVVLEL-----DFDRLARFLEAAVGGFAFVHMTWPFADDTTHQVLS 145
motif 1B
E. mel. RGLKLELDFVDFRREIATATREVVQVLEDFEFKFAERLDFDRAVLEAVVLEL-----LAKVLEEDVFLAVV-----EVDFDGFDAARAPELIIA 390
D. ana. RGLKLELDFVDFRREIATATREVVQVLEDFEFKFAERLDFDRAVLEAVVLEL-----LAKVLEEDVFLAVV-----EVDFDGFDAARAPELIIA 316
D. ere. RGLKLELDFVDFRREIATATREVVQVLEDFEFKFAERLDFDRAVLEAVVLEL-----LAKVLEEDVFLAVV-----EVDFDGFDAARAPELIIA 316
D. sec. RGLKLELDFVDFRREIATATREVVQVLEDFEFKFAERLDFDRAVLEAVVLEL-----LAKVLEEDVFLAVV-----EVDFDGFDAARAPELIIA 316
D. yak. RGLKLELDFVDFRREIATATREVVQVLEDFEFKFAERLDFDRAVLEAVVLEL-----LAKVLEEDVFLAVV-----EVDFDGFDAARAPELIIA 316
D. per. RGLKLELDFVDFRREIATATREVVQVLEDFEFKFAERLDFDRAVLEAVVLEL-----LAKVLEEDVFLAVV-----EVDFDGFDAARAPELIIA 316
D. pse. RGLKLELDFVDFRREIATATREVVQVLEDFEFKFAERLDFDRAVLEAVVLEL-----LAKVLEEDVFLAVV-----EVDFDGFDAARAPELIIA 316
D. gri. RGLKLELDFVDFRREIATATREVVQVLEDFEFKFAERLDFDRAVLEAVVLEL-----LAKVLEEDVFLAVV-----EVDFDGFDAARAPELIIA 316
D. wil. RGLKLELDFVDFRREIATATREVVQVLEDFEFKFAERLDFDRAVLEAVVLEL-----LAKVLEEDVFLAVV-----EVDFDGFDAARAPELIIA 316
D. vir. RGLKLELDFVDFRREIATATREVVQVLEDFEFKFAERLDFDRAVLEAVVLEL-----LAKVLEEDVFLAVV-----EVDFDGFDAARAPELIIA 316
D. moj. RGLKLELDFVDFRREIATATREVVQVLEDFEFKFAERLDFDRAVLEAVVLEL-----LAKVLEEDVFLAVV-----EVDFDGFDAARAPELIIA 316
A. aeg. RGLKLELDFVDFRREIATATREVVQVLEDFEFKFAERLDFDRAVLEAVVLEL-----LAKVLEEDVFLAVV-----EVDFDGFDAARAPELIIA 316
A. gam. RGLKLELDFVDFRREIATATREVVQVLEDFEFKFAERLDFDRAVLEAVVLEL-----LAKVLEEDVFLAVV-----EVDFDGFDAARAPELIIA 316
C. pip. RGLKLELDFVDFRREIATATREVVQVLEDFEFKFAERLDFDRAVLEAVVLEL-----LAKVLEEDVFLAVV-----EVDFDGFDAARAPELIIA 316
A. mel. KNLGVVDFVDFRREIATATREVVQVLEDFEFKFAERLDFDRAVLEAVVLEL-----LAKVLEEDVFLAVV-----EVDFDGFDAARAPELIIA 449
D. mor. HALFKLVDFVDFRREIATATREVVQVLEDFEFKFAERLDFDRAVLEAVVLEL-----LAKVLEEDVFLAVV-----EVDFDGFDAARAPELIIA 444
T. cas. TELKLVDFVDFRREIATATREVVQVLEDFEFKFAERLDFDRAVLEAVVLEL-----LAKVLEEDVFLAVV-----EVDFDGFDAARAPELIIA 214
motif 4

```

Figure 1. Sequence alignments for UDE homologues and conserved motifs in UDE and members of the UDG superfamily. Panel A: Conservation characteristics within UDE homologues found by Tblastp and NCBI genomic blast in genomes of pupating insects are presented. Grey background comprise conserved motifs, red letters indicate strictly conserved residues. Species represented are *D.mel.*: *Drosophila melanogaster* (CG18410), *D.ana.*: *Drosophila ananassae* (GF 23181-PA), *D.ere.*: *Drosophila erecta* (GG12323-PA), *D.sec.*: *Drosophila sechellia* (GM23464-PA), *D.yak.*: *Drosophila yakuba* (GE10776-PA), *D.per.*: *Drosophila persimilis* (GL13863-PA), *D.pse.*: *Drosophila pseudoobscura* (GA14925-PA), *D.wil.*: *Drosophila willistoni* (GK13169-PA), *D.gri.*: *Drosophila grimshawi* (GH18464-PA), *D.vir.*: *Drosophila virilis* (GJ23671-PA), *D.moj.*: *Drosophila mojavensis* (GI24043-PA), *A.aeg.*: *Aedes aegypti* (ENSANGP00000015628), *A.gam.*: *Anopheles gambiae* (EAA05227.2), *C.pip.*: *Culex pipiens* (EDS29350.1), *A.mel.*: *Apis mellifera* (XP_001122207), *B.mor.*: *Bombyx mori* (protein sequence was reconstructed from two contigs using NCBI ORF prediction on contig 6156 GI:46642882 and contig 413277 GI:46731897), *T.cas.*: *Tribolium castaneum* (XP_971843) (sequence identifiers in parantheses, cf also [43]).



Panel B: Evolutionary relationship and organization of conserved motifs among UDG proteins [18]. Grey background indicates uracil-DNA recognizing proteins present in *Drosophila melanogaster*. **Panel C: Organization of conserved motifs in UDE.** Panels D and E: Consensus sequences of UDG (D) and UDE (E) motifs. Upper-case letters represent conserved residues, lower-case letters refer to residues with conserved characteristics (h, hydrophobic; a, aromatic; p, polar/charged). Non-conserved positions are shown by x. A conserved F/Y residue, overlapping with the uracil ring, is invariably present C-terminal to Motif 1 in UDGs. Underlined D/E residues in UDG Motif 1 are involved in catalysis; underlined histidine in UDG Motif 3 is suggested to stabilize reaction intermediates. Note lack of detectable similarities between UDE and UDG motifs.

To confirm the *in silico* prediction of the UDE-like protein product in *Tribolium*, extracts of the insect larvae were investigated on Western blot using the polyclonal antiserum produced against recombinant Dm_{rc}UDE. As expected from the high sequence similarity, the antiserum did recognise the *Tribolium* protein (TcUDE) as well (Figure 2). The blot clearly indicates that larval extract from *T. castaneum* contains a single protein that reacts with the UDE-specific antibody. This positive band is found at a position corresponding to a much lower molecular mass as compared to the recombinant Dm_{rc}UDE and to the physiological form of DmUDE. The altered position of TcUDE was in agreement with the genomic data (cf. Figure 1A) and led to the conclusion that the physiologically occurring TcUDE lacks the N-terminal segment. These results suggest that an isoform of UDE lacking Motif 1A may fold on its own, and may form a functional protein.

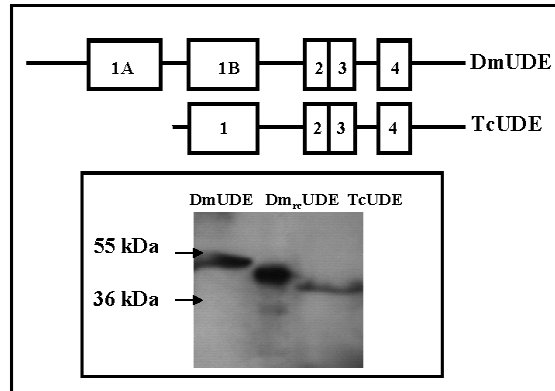


Figure 2. Immunodetection of UDE homologue from *T. castaneum*. Western blot indicates that polyclonal anti-DmUDE antiserum recognizes the UDE homologue from *T. castaneum* that appeared at a lower position as compared to physiological DmUDE or recombinant Dm_{rc}UDE. Lane 1, *D. melanogaster* larval extract; lane 2, purified recombinant Dm_{rc}UDE; lane 3, *T. castaneum* larval extract.

Domain organisation

To delineate domain organisation of the UDE protein more precisely, limited proteolysis experiments were performed. Three proteases with different specificities were used. Experiments were conducted with Dm_{rc}UDE alone and also in the presence of added DNA to study potential DNA-binding-induced conformational changes.

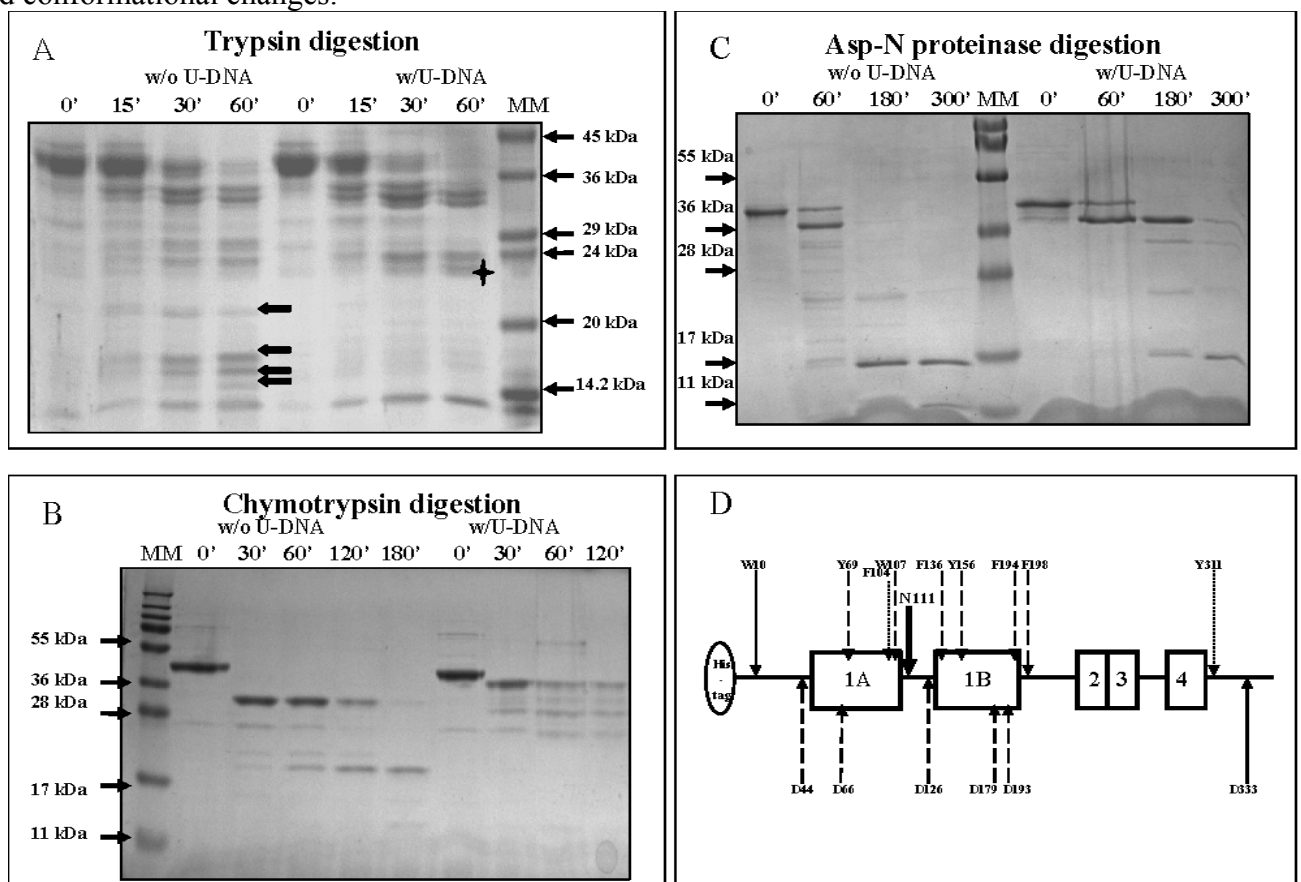


Figure 3. Initial domain analysis of *D. mel.* UDE by limited proteolysis. (A) Tryptic digestion pattern. Arrows indicate fragments preferentially produced in the absence of DNA, star shows the detected position of stable fragment persisting in the presence of DNA. **(B) and (C) Limited digestion patterns using high specificity chymotrypsin and Asp-N endoproteinase.** Timescale of limited digestion and the presence or absence of added ligand are indicated at the top of the gel. MM stands for molecular markers (data in kDa). **(D) Summary of cleavage sites identified by mass spectrometry.** Top row, chymotryptic sites, bottom row, Asp-N sites. Solid arrows indicate cleavage sites similarly observable both in the presence and absence of DNA. Dashed arrows indicate sites protected in the presence of DNA. Dotted arrows indicate cleavage sites detected only in the presence of DNA. Cleavage site of hydroxylamine is marked with bold arrow.

Trypsin was selected first, as the UDE protein contains many potential tryptic cleavage sites (i.e. lysine and arginine residues) scattered throughout the sequence. Figure 3A indicates fast initial fragmentation leading to loss of 5-7 kDa fragments from either N- or C-termini, or both. This initial fragmentation is not affected by the presence of DNA. Flexibility of the N-terminal segment (residues 1-47) is also suggested by the drastic overrepresentation of basic residues leading to an extremely high pI (11.5) for this segment. At later stages of proteolysis, DNA protection is evident as a specific fragment (indicated by a star) persists stably in the presence of DNA, while this fragment is rapidly degraded in the absence of DNA. Several smaller fragments (cf. arrows) are produced in relatively great amount in the absence of DNA while these peptides are practically absent in the presence of DNA. Data suggest the presence of an inner folded core that is suggested to participate in DNA binding, based on DNA-binding induced stabilization. The large number of potential tryptic cleavage sites prevented straightforward identification of the fragments, observed on SDS-PAGE, by mass spectrometry.

For further characterization and localization of flexible segments undergoing conformational transitions upon DNA binding to UDE, two additional sets of experiments were conducted, using highly specific chymotrypsin [20] and Asp-N endoproteinase. These enzymes have considerably less potential cleavage sites in the protein. In both cases, protection by DNA is evident again (Figure 3B,C). Figure 3B shows that in the absence of DNA, initial chymotryptic cleavage removes a segment of about 9.6 kDa from UDE, while in the presence of DNA, the removed peptide is much smaller, around 3 kDa. Mass spectrometric analysis of the initially cleaved fragments revealed that the C-terminus remained intact and therefore the two peptide bonds most sensitive to chymotrypsin could be localized at the N-terminus at Trp10 and Tyr69 (W¹⁰, Y⁶⁹) in the presence and in the absence of DNA, respectively (cf Figure 3D). DNA binding is therefore associated with significant protection at the Tyr69–Arg70 (Y⁶⁹–R⁷⁰) peptide bond located within the conserved motif 1A. In addition, DNA-binding induced conformational changes are also reflected at the peptide bonds Phe104–Glu105 (F¹⁰⁴–E¹⁰⁵) and Tyr311–Ile312 (Y³¹¹–I³¹²) that become exposed in the presence of DNA (cf Figure 3D).

To characterize the potential involvement of the C-terminal region of UDE in DNA-binding, Asp-N endoproteinase was also used for limited proteolysis as the C-terminus of the protein is rather rich in Asp residues (Figure 3C,D). When digested by Asp-N endoproteinase, the primary cleavage removes a short fragment of about 3.4 kDa, independently of the presence of DNA. This loss is in good agreement with a C-terminal cleavage (at Asp333 (D³³³), leading to loss of 2.6 kDa), while cleavage at the first N-terminal Asp residue (Asp44 (D⁴⁴)) would remove a 6.6 kDa peptide that is much larger than estimated from the gel electrophoretic analysis. It is evident that in the absence of DNA, additional cleavages can also occur yielding 23-25 kDa and 17 kDa polypeptides, as observed on SDS-PAGE. Binding of DNA induces significant protection against all these sites, except at the Asp333 (D³³³) site that shows the same highly exposed character for Asp-N endoproteinase digestion either in the presence or absence of DNA.

Results of proteolytic experiments are summarized in Figure 3D. DNA binding provides significant protection against proteolytic cleavage along most of the protein. The C-terminal part of UDE containing 2-4 motifs is expected to be compactly folded because we did not identify any preferred cleavage sites by

either of the proteases either in the presence or absence of DNA. In agreement with the physiological isoform in *Tribolium* (TcUDE), proteolysis data suggest that the truncated protein lacking Motif 1A may constitute a well-folded segment, especially in the presence of DNA.

To further strengthen this suggestion, we selected a chemical agent, hydroxylamine that cleaves peptide bonds exclusively between asparagine and glycine [21]. There is only one such dipeptide in DmUDE at residues Asn111–Gly112 ($N^{111}-G^{112}$) (Figure 3D), located between Motifs 1A and 1B. Figure 4A shows that in agreement with the previously determined exposed character of the linker segment between Motifs 1A and 1B, hydroxylamine cleaved the protein into an N-terminal Met1–Asn111 (M^1-N^{111}) and a C-terminal Gly112–Glu355 ($G^{112}-E^{355}$) fragment, as verified by mass spectrometry. The molecular masses of the cleavage products are 14 kDa and 27 kDa as calculated from the sequence, while values of 16 kDa and 32 kDa were estimated from the SDS-PAGE gels. The C-terminal fragment closely corresponds to the physiological TcUDE isoform. The presence of the N-terminal His-tag on Dm_{rc}UDE allowed straightforward separation of N- and C-terminal hydroxylamine-cleaved segments by Ni-NTA chromatography (Figure 4A).

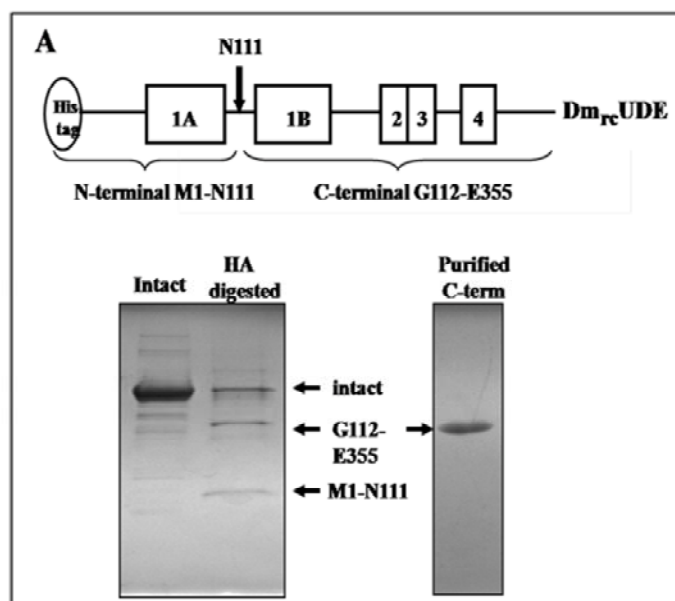
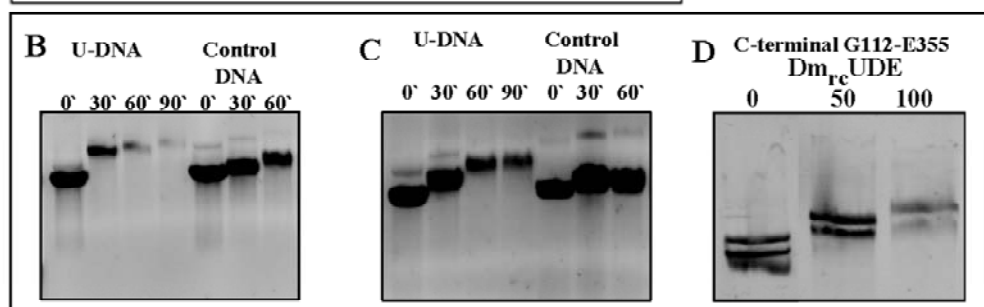


Figure 4. A: Production and characterization of the truncated UDE isoform. Panel A. Cleavage with hydroxylamine generates the expected fragments.

Schematic representation marks the single cleavage site at residue N125 between the 1A and 1B motifs with an arrow. Gel images show gelectrophoretic analysis of hydroxylamine cleavage and purification of the C-terminal motif to homogeneity. HA, hydroxylamine.

Panels B,C: Truncated UDE lacking Motif 1A preserves U-DNA degrading activity.

Uracil-DNA and control DNA were incubated for the indicated time intervals with truncated Dm_{rc}UDE (G112-E355 segment) either in the presence of 1 mM MgCl₂ (panel B) or 1 mM EDTA (panel C). In the presence of EDTA, the catalytic activity of the enzyme is significantly lower, but still present.



D: Electrophoretic mobility shift assay.

The concentration of Dm_{rc}UDE G112-E355 segment used in the experiment is listed in µg/ml at the top of the lanes. 20 µg/ml DNA was used in all mixtures.

To check if removal of Motif 1A alters the specific function of the protein, we performed catalytic assay and electrophoretic mobility shift assays with the purified Gly112–Glu355 ($G^{112}-E^{355}$) C-terminal fragment. Figure 4B,C shows that the C-terminal segment preserves catalytic activity and specificity for uracil-substituted DNA that does not depend on the presence or absence of available divalent metal ions. The gel shift result indicates the DNA-binding capability of the C-terminal fragment and also demonstrates the specific DNA-cleaving activity (Figure 4D).

To determine whether the absence of the N-terminus has any effect on quaternary structure organisation of UDE, the native molecular mass for the full length protein and the C-terminal fragment was determined by analytical gel filtration (Figure 5). The full length protein eluted at a position corresponding to 52 kDa, somewhat larger than the full-length monomer calculated molecular mass of 41.3 kDa. This alteration may indicate partial rapid equilibrium dimerization. Interestingly, the C-terminal G112–E355

fragment eluted from the gel filtration column at practically the same position corresponding to 52 kDa. Since the calculated molecular mass of the monomeric G112–E355 fragment is 27 kDa, the elution profile strongly suggests that this fragment forms a dimer.

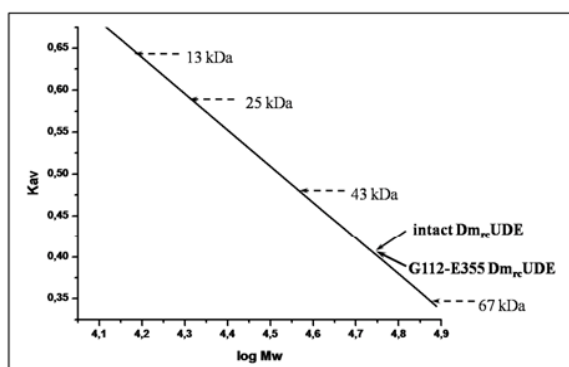


Figure 5. Determination of UDE oligomer status by analytical gel filtration. Dashed arrows indicate molecular mass markers: RNase (13 kDa), chymotrypsin (25 kDa), ovalbumin (43 kDa), BSA (67 kDa). Solid arrows indicate elution position of intact Dm_{rc}UDE UDE and its C-terminal G112-E355 fragment. Note that these two proteins have very similar native molecular masses as estimated by gel filtration.

Structure prediction of UDE reveals pseudosymmetrical arrangement of two alpha-helical bundles

For structural prediction, DmUDE full length sequence was submitted to GeneSilico Metaserver [22] which is the gateway providing unified interface to several servers for secondary and tertiary structure predictions. The analysis of predictions of domain composition suggested that UDE contains an N-terminal helical region of approximately 30 residues and at least three structural domains corresponding to Motifs 1A, 1B and the C-terminus, encompassing Motifs 2, 3 and 4. The very C-terminus of 40-50 residues as well as the loop connecting Motifs 1A and 1B (between residues 109-137) are predicted to be mostly disordered. All three domains are predicted to be mainly helical, although the secondary structure predictions for the third domain were uncertain as there was no agreement between alternative servers.

The fold recognition analysis did not reveal any confident matches to known protein structures suggesting that UDE 3D structure may exhibit a novel fold. Therefore, to predict at least partially the tertiary structure of UDE, we performed *de novo* modeling of the region encompassing Motifs 1A and 1B using ROSETTA program [23]. In total about 500,000 different low resolution models (aka decoys) were generated and 10% of the lowest energy structures were clustered based on their similarity. The representatives of the best clusters were refined with ROSETTA full atom refinement protocol and scored with Model Quality.

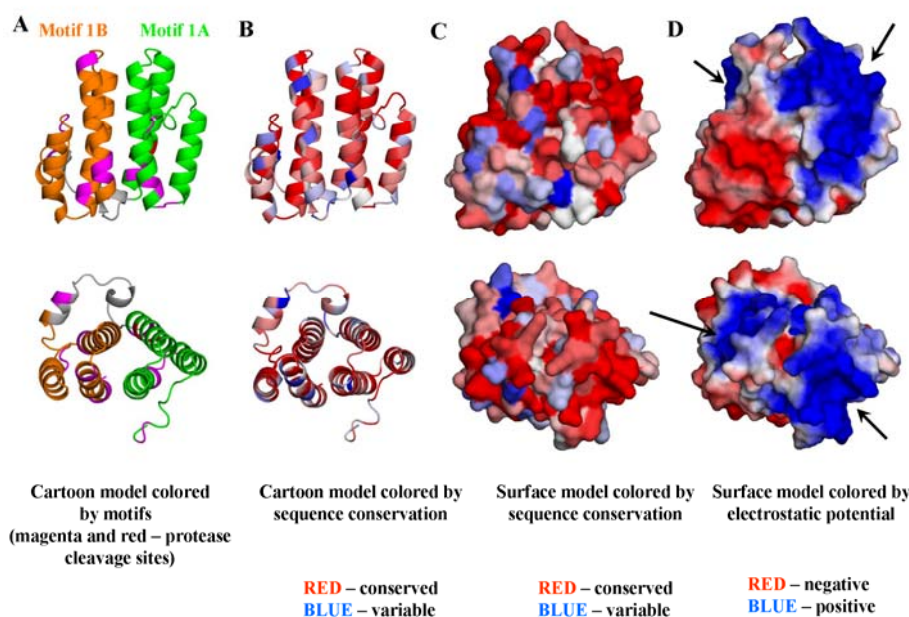


Figure 6. Structural model of *D.melanogaster* UDE duplication fragment. Structures are shown in two views: front (*upper panel*) and top (*bottom panel*). **A:** Cartoon representation. Duplicated motifs 1A and 1B are colored green and orange, respectively, and non-conserved linker is colored gray. Peptide bonds protected from proteolytic cleavage upon DNA binding are colored magenta. The peptide bond between residues 104 and 105 cleaved only upon DNA binding is colored red. Note that the duplicated fragments are only approximately symmetrical as the model is of low resolution and local conformation of the backbone is uncertain. **B and C:** Sequence conservation mapped onto the ribbon diagram (B) or the molecular surface (C) (conserved residues are colored salmon and red, variable residues are colored blue). **D:** Electrostatic potential mapped onto the molecular surface (positively and negatively charged regions are colored blue and red, respectively). Arrows indicate the positively charged conserved patches that may accommodate DNA.

Assessment Programs (MQAPs) ProQ [24] and MetaMQAP [25]. Evaluation of the largest clusters revealed that both Motifs 1A and 1B constitute similar three-helical bundles with the same topology and relative orientation of the helices. Nevertheless, the top clusters differed in relative orientation of the two helical bundles to each other (data not shown). Among these clusters, one single cluster contained the specific topology that exhibited pseudosymmetrical orientation of the two motifs. Importantly, members of this cluster exhibited low energy levels and were well scored by MQAPs. Figure 6 depicts the predicted model in several different orientations. The two homologous Motifs (Motif 1A and Motif 1B) form a four-helix bundle interaction surface (Figure 6A, B). On the surface of the model, a well-conserved positively charged surface is well defined. This may serve as the nucleic acid binding surface, in agreement with the limited proteolysis data.

Estimation of secondary structural elements by circular dichroism spectroscopy

To verify structural predictions, circular dichroism (CD) spectroscopy measurements were carried out, as CD spectra in the far-UV wavelength (190-240 nm) range are much indicative of different secondary structural elements [26]. Spectra of the intact protein and that of the C-terminal fragment G112-E355 showed double minima at 208 and 222 nm, characteristic for the presence of α -helices (Figure 7). Quantitative evaluation of the spectral data was performed by the k2d program [27]. The estimated percentages of protein secondary structures from CD spectra reveal a majority of α -helices, with low amount of β -structure.

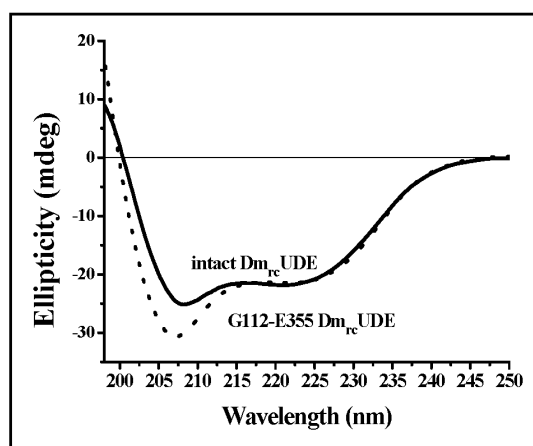


Figure 7. CD-spectra of intact UDE (solid line) and C-terminal fragment (dashed line) confirm the presence of α -helices.

Discussion

The potential signalling role of deoxyuridine moieties in genomes of pupating insects was first suggested by Deutsch and co-workers [16] based on the lack of uracil-DNA glycosylase activity in these insects. The hypothesis stating that uracil-DNA might be present transiently in larval stages and its degradation at the end of larval stages may contribute to cell death during metamorphosis was much debated due to independent findings from several laboratories, which showed the presence of uracil-DNA glycosylase activity in some developmental stages of *Drosophila* [28-32]. This debate could be resolved by the fully annotated *Drosophila* genome that clearly indicated the lack of the major UDG gene *ung* but presence of several, catalytically much less efficient UDG genes. Absence of dUTPase from larval stages [14] and the recent discovery of a strictly regulated uracil-DNA degrading factor (UDE) [15] reinforced the hypothesis on the possible role of uracil-DNA in *Drosophila* and suggested a role for UDE in programmed cell death during metamorphosis. Functional analysis of UDE identified this protein as a novel uracil-recognising factor [15] with no similarities to either UDGs [18] or the ExoIII/Mth212 nuclease [33]. Multiple sequence alignments of UDE homologues from all available pupating insect genomes indicated conserved motifs present in most species with the same distribution (Figure 1).

The UDE homologue in *T. castaneum* lacks one copy of the N-terminal duplicated first motif (Figures 1, 2, cf also Figure 9 for a flowchart of the different approaches). TcUDE showed reactivity with the antiserum produced against recombinant Dm_{rc}UDE, suggesting that the truncated TcUDE isoform constitutes a well folded UDE-like protein. It was also observable on the blot that the physiological forms of the proteins from both *Drosophila* and *Tribolium* extracts were detected at much higher electrophoretic positions than expected from the calculated molecular mass values: molecular mass data estimated from the electrophoretic experiment are 53.3 and 41.7 kDa for DmUDE and TcUDE homologues respectively, whereas sequence-based theoretical masses are 39.9 and 27 kDa. Recombinant Dm_{rc}UDE, expressed in *E.*

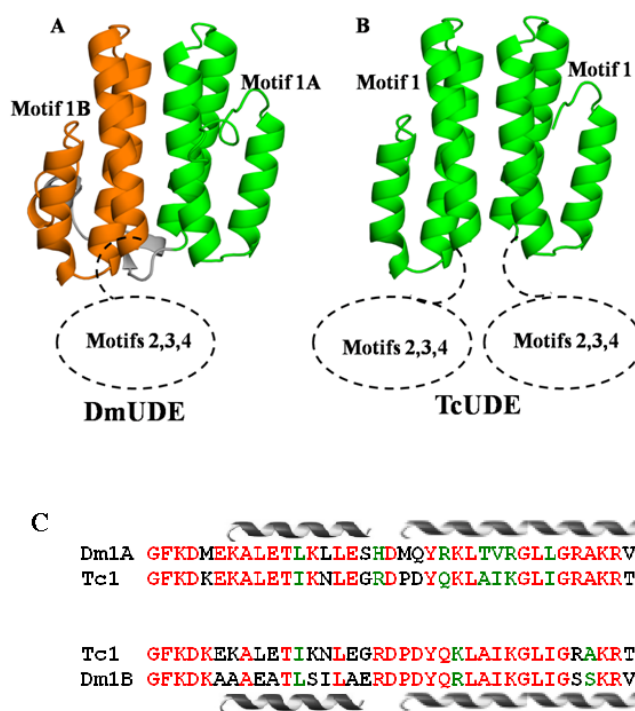
coli, did not show such large deviation from its expected position (calculated molecular mass is 41.3 kDa, electrophoretic estimation is 44.2 kDa). The large shift of the physiological samples to the higher apparent molecular mass positions on SDS-PAGE may be indicative of some post-translational modification, the identification of which is in progress.

Limited proteolysis experiments indicated that DNA-binding may occur along the conserved motifs. It induces a protein conformational transition resulting in significant protection against multiple proteases at peptide bonds scattered throughout the duplicated Motifs 1A and 1B (Figure 3). Secondary structure prediction revealed that the duplicated fragments at the N-terminal end of UDE are mainly helical, and this observation has been confirmed also by circular dichroism measurements (Figure 7). According to the low resolution structural model of the fragment encompassing Motifs 1A and 1B the duplicated motifs together form a conserved helical bundle (Figure 6AB). The pseudodimer contains large conserved segment on the surface, which is composed of two positively charged patches separated by small region of negative potential (Figure 6CD, arrows). These patches may correspond to DNA binding and/or catalytic site. However, based on this low resolution model the DNA binding mode and catalytic site residues cannot be confidently predicted. Interestingly, many protease cleavage sites that are protected after DNA binding are located on the opposite non-conserved and negatively charged side of the structure (Figure 6A). Therefore, these sites are not likely to be directly sterically protected from the cleavage by bound DNA. Instead, they probably lay in a region that is partially flexible or disordered in the absence of DNA.

The primary structure of TcUDE implied that the lack of one copy of Motif 1 does not necessarily perturb formation of a functional protein. This hypothesis was confirmed by producing the respective truncated isoform with chemical cleavage from Dm_{rc}UDE (Figure 4). We therefore conclude that the physiological form of TcUDE could have the same unique function and

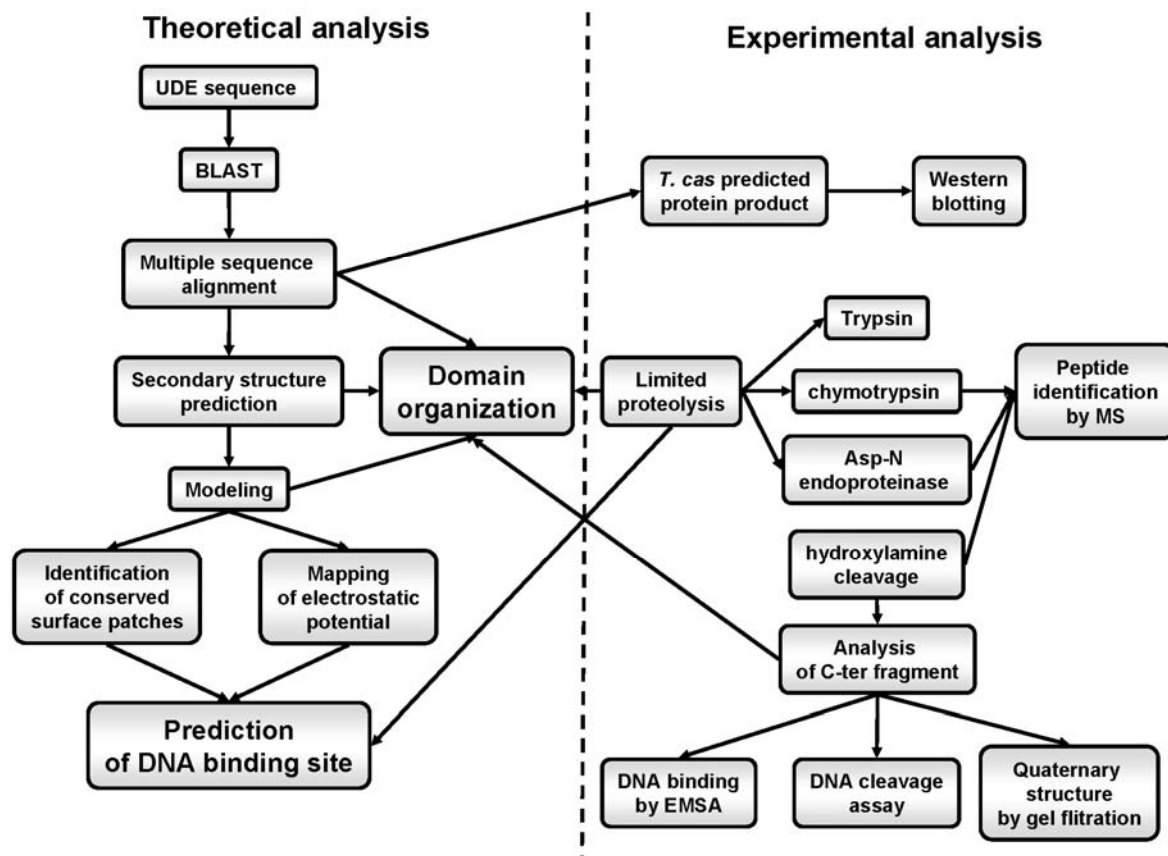
the same putative physiological role. Native molecular mass estimation by gel filtration indicated that the truncated Dm_{rc}UDE, representing TcUDE, forms a homodimer (Figure 5). A structural model for this homodimer is suggested in Figure 8. The two homologous Motifs 1A and 1B in the Dm_{rc}UDE monomer may lead to the formation of a partial pseudodimer, and this interaction pattern may be preserved in the suggested TcUDE dimer (Figure 8C). The low resolution predicted model prevented direct assessment of interaction pattern between the two copies of Motif 1 (1A/1B). Still, numerous hydrophobic interactions are very probable between the conserved apolar residues, and H-bonds can also easily form between the polar side chains within the three-helical bundles. In the dimerized modules, the nucleic acid binding surface can also be formed in a manner much similar to the full-length UDE proteins.

Conclusion



UDE significance is twofold: i) it may be developed into a manifold molecular biotechnological tool [34], and ii) its targeting may yield species-specific insecticides to be used against e.g. malaria mosquitoes. Here, we employed a multidisciplinary set of theoretical and experimental approaches (schematically described in Figure 9) to reveal structural and functional characteristics. The present data allow insight into the domain structure and nucleic acid binding site of this novel DNA-degrading protein in the context of sequence motifs previously not described in nucleases or uracil-recognition proteins.

Figure 9. Flowchart scheme of bioinformatics and experimental approaches



Acknowledgements

This work was supported by grants from Hungarian Scientific Research Fund (K68229, NKTH-OTKA H07-BEL74200); Howard Hughes Medical Institutes (#55005628 and #55000342); Alexander von Humboldt Foundation; GVOP-3.2.1.-2004-05-0412/3.0; and JÁP_TSZ_071128_TB_INTER from the National Office for Research and Technology, Hungary; FP6 STREP 012127; FP6 SPINE2c LSHG-CT-2006-031220; and TEACH-SG LSSG-CT-2007-037198 from the EU.

Materials and methods (TRUNCATED!!!)

Recombinant protein expression [15], Western blotting, Limited proteolysis, Hydroxylamine digestion [21], Mass spectrometry [35-37], Catalytic assay [15], Electrophoretic mobility shift assay of DNA binding, Analytical gel filtration, Circular dichroism spectroscopy, Bioinformatics analyses [38],[22,39],[24],[25],[40],[41],[42].

References

1. Lindahl, T., *Instability and decay of the primary structure of DNA*. Nature, 1993. **362**(6422): p. 709-15.
2. Krokan, H.E., F. Drablos, and G. Slupphaug, *Uracil in DNA--occurrence, consequences and repair*. Oncogene, 2002. **21**(58): p. 8935-48.
3. Pearl, L.H. and R. Savva, *The problem with pyrimidines*. Nat Struct Biol, 1996. **3**(6): p. 485-7.
4. Mosbaugh, D.W., *Purification and characterization of porcine liver DNA polymerase gamma: utilization of dUTP and dTTP during in vitro DNA synthesis*. Nucleic Acids Res, 1988. **16**(12): p. 5645-59.
5. Vertessy, B.G. and J. Toth, *Keeping Uracil Out of DNA: Physiological Role, Structure and Catalytic Mechanism of dUTPases*. Acc Chem Res, 2008.

6. Krokan, H.E., R. Standal, and G. Slupphaug, *DNA glycosylases in the base excision repair of DNA*. *Biochem J*, 1997. **325 (Pt 1)**: p. 1-16.
7. Dogliotti, E., et al., *The mechanism of switching among multiple BER pathways*. *Prog Nucleic Acid Res Mol Biol*, 2001. **68**: p. 3-27.
8. Kavli, B., et al., *hUNG2 is the major repair enzyme for removal of uracil from U:A matches, U:G mismatches, and U in single-stranded DNA, with hSMUG1 as a broad specificity backup*. *J Biol Chem*, 2002. **277**(42): p. 39926-36.
9. Lari, S.U., et al., *Quantitative determination of uracil residues in Escherichia coli DNA: Contribution of ung, dug, and dut genes to uracil avoidance*. *DNA Repair (Amst)*, 2006. **5**(12): p. 1407-20.
10. Nilsen, H., et al., *Uracil-DNA glycosylase (UNG)-deficient mice reveal a primary role of the enzyme during DNA replication*. *Mol Cell*, 2000. **5**(6): p. 1059-65.
11. Kavli, B., et al., *B cells from hyper-IgM patients carrying UNG mutations lack ability to remove uracil from ssDNA and have elevated genomic uracil*. *J Exp Med*, 2005. **201**(12): p. 2011-21.
12. el-Hajj, H.H., L. Wang, and B. Weiss, *Multiple mutant of Escherichia coli synthesizing virtually thymineless DNA during limited growth*. *J Bacteriol*, 1992. **174**(13): p. 4450-6.
13. Gerald Rubin, A.S., Roger Hoskins, Hugo Bellen, Susan Celniker, and Gary Karpen laboratories. *Drosophila Genome Project, Release 4*. 2004 [cited; Available from: <http://www.fruitfly.org>].
14. Bekesi, A., et al., *Developmental regulation of dUTPase in Drosophila melanogaster*. *J Biol Chem*, 2004. **279**(21): p. 22362-70.
15. Bekesi, A., et al., *A novel fruitfly protein under developmental control degrades uracil-DNA*. *Biochem Biophys Res Commun*, 2007. **355**(3): p. 643-8.
16. Deutsch, W.A., *Why do pupating insects lack an activity for the repair of uracil-containing DNA? One explanation involves apoptosis*. *Insect Mol Biol*, 1995. **4**(1): p. 1-5.
17. Dudley, B., A. Hammond, and W.A. Deutsch, *The presence of uracil-DNA glycosylase in insects is dependent upon developmental complexity*. *J Biol Chem*, 1992. **267**(17): p. 11964-7.
18. Aravind, L. and E.V. Koonin, *The alpha/beta fold uracil DNA glycosylases: a common origin with diverse fates*. *Genome Biol*, 2000. **1**(4): p. RESEARCH0007.
19. Richards, S., et al., *The genome of the model beetle and pest Tribolium castaneum*. *Nature*, 2008. **452**(7190): p. 949-55.
20. Keil, B., *Specificity of proteolysis* 1992, Berlin-Heidelberg-NewYork: Springer-Verlag 335.
21. Bornstein, P. and G. Balian, *Cleavage at Asn-Gly bonds with hydroxylamine*. *Methods Enzymol*, 1977. **47**: p. 132-45.
22. Kurowski, M.A. and J.M. Bujnicki, *GeneSilico protein structure prediction meta-server*. *Nucleic Acids Res*, 2003. **31**(13): p. 3305-7.
23. Rohl, C.A., et al., *Protein structure prediction using Rosetta*. *Methods Enzymol*, 2004. **383**: p. 66-93.
24. Wallner, B., H. Fang, and A. Elofsson, *Automatic consensus-based fold recognition using Pcons, ProQ, and Pmodeller*. *Proteins*, 2003. **53 Suppl 6**: p. 534-41.
25. Pawlowski, M., et al., *MetaMQAP: a meta-server for the quality assessment of protein models*. *BMC Bioinformatics*, 2008. **9**: p. 403.
26. Whitmore, L. and B.A. Wallace, *Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases*. *Biopolymers*, 2008. **89**(5): p. 392-400.
27. Andrade, M.A., et al., *Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network*. *Protein Eng*, 1993. **6**(4): p. 383-90.
28. Deutsch, W.A. and A.L. Spiering, *A new pathway expressed during a distinct stage of Drosophila development for the removal of dUMP residues in DNA*. *J Biol Chem*, 1982. **257**(7): p. 3366-8.
29. Deutsch, W.A., *Enzymatic studies of DNA repair in Drosophila melanogaster*. *Mutat Res*, 1987. **184**(3): p. 209-15.
30. Morgan, A.R. and J. Chlebek, *Uracil-DNA glycosylase in insects. Drosophila and the locust*. *J Biol Chem*, 1989. **264**(17): p. 9911-4.
31. Green, D.A. and W.A. Deutsch, *Repair of alkylated DNA: Drosophila have DNA methyltransferases but not DNA glycosylases*. *Mol Gen Genet*, 1983. **192**(3): p. 322-5.
32. Breimer, L.H., *A DNA glycosylase for oxidized thymine residues in Drosophila melanogaster*. *Biochem Biophys Res Commun*, 1986. **134**(1): p. 201-4.
33. Georg, J., et al., *The Methanothermobacter thermoautotrophicus ExoIII homologue Mth212 is a DNA uridine endonuclease*. *Nucleic Acids Res*, 2006. **34**(18): p. 5325-36.
34. Békési, A., Felföldi, F., Pukáncsik, M., Zagyva, I., Vértessy, G.B., *Uracil-DNA nuclease: protein enzyme possessing nuclease activity specific for uracil containing nucleic acid, process for its preparation and methods of use*, in *United States Patent and Trademark Office*. 2005, 11/160040: USA.
35. Varga, B., et al., *Active site closure facilitates juxtaposition of reactant atoms for initiation of catalysis by human dUTPase*. *FEBS Lett*, 2007. **581**(24): p. 4783-8.
36. Nemeth-Pongracz, V., et al., *Flexible segments modulate co-folding of dUTPase and nucleocapsid proteins*. *Nucleic Acids Res*, 2007. **35**(2): p. 495-505.
37. Dubrovay, Z., et al., *Multidimensional NMR identifies the conformational shift essential for catalytic competence in the 60-kDa Drosophila melanogaster dUTPase trimer*. *J Biol Chem*, 2004. **279**(17): p. 17945-50.
38. Larkin, M.A., et al., *Clustal W and Clustal X version 2.0*. *Bioinformatics*, 2007. **23**(21): p. 2947-8.
39. Bradley, P., et al., *Free modeling with Rosetta in CASP6*. *Proteins*, 2005. **61 Suppl 7**: p. 128-34.
40. Landau, M., et al., *ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures*. *Nucleic Acids Res*, 2005. **33**(Web Server issue): p. W299-302.
41. Baker, N.A., et al., *Electrostatics of nanosystems: application to microtubules and the ribosome*. *Proc Natl Acad Sci U S A*, 2001. **98**(18): p. 10037-41.
42. DeLano, W.L., *Pymol*. 2002, DeLano Scientific: San Carlos, CA, USA.
43. Zhang, G., et al., *Identification and characterization of insect-specific proteins by genome data analysis*. *BMC Genomics*, 2007. **8**: p. 93.

2. számú melléklet: kézirat beküldve

Uracil-DNA in *Drosophila*: interpretation and developmental involvement

Villő Muha, Angéla Békési, Barbara Hodoscsek, Gábor Merényi and Beáta G. Vértessy*

Institute of Enzymology, Hungarian Academy of Science, Budapest, Hungary

* to whom correspondence should be addressed. E-mail: vertessy@enzim.hu

Abstract: Uracil in DNA is usually considered to be a mistake that has to be corrected to prevent mutagenic errors. This adverse notion on uracil in DNA is, however, justified only for those uracils that are products of cytosine deamination, while thymine-replacing uracils present the normal coding message in the Watson-Crick base pairs. We show that *Drosophila* cells interpret uracil-substituted DNA as normal DNA. Importantly, this unusual trait is under strong developmental control in the fruitfly organism, and applies only for animals before pupation. The normal progress of metamorphosis is drastically perturbed by silencing of the protein genes of dUTPase and uracil-DNA degrading factor, which are involved in uracil-DNA metabolism. Results suggest that uracil-DNA may be involved in developmental signaling in *Drosophila*.

Textbooks say that there are two major differences between the chemical composition of DNA and RNA: i) at the sugar level (deoxyribose vs. ribose) and ii) at the base level (thymine vs. uracil). Uracil, however, frequently appears in DNA either from cytosine deamination or thymine-replacing incorporation. These appearances are usually ephemeral due to catalytic action of two key factors: dUTPase, responsible for prevention of uracil incorporation into DNA and the major isoform of uracil-DNA glycosylase (UNG) that excises uracil from DNA. Recent results showed that UNG is not encoded in *Drosophila melanogaster* genome (1) and dUTPase is down-regulated in larval tissues (2). Lack of these two enzymes in genetically engineered model organisms led to accumulation of uracil-substituted DNA (U-DNA) (3) and therefore a similar situation can be envisioned in *Drosophila* larvae under physiological spatial-temporal control.

We first asked if such chemically unusual DNA may be tolerated and interpreted in *Drosophila* cells. U-DNA specific cell response was provoked by transfecting cells with exogenous plasmid U-DNA and by increasing dUMP content of genomic DNA with 5'-fluorouracil treatment.

Upon transfection of both human and *Drosophila* cells with normal DNA plasmid, the expression of the fluorescent protein encoded by the plasmid can be readily observed, according to the usual expectation (Fig.1A, Fig.S1). The same plasmid can be produced in a uracil-substituted version (U-plasmid, containing approx 3000 uracil/million bases, cf with levels less than 10 uracil/million in normal cells/organisms) in double mutant (dut-, ung-) *E. coli* cells (3). Human cells transfected with U-plasmid do not show any appreciable fluorescent signal, indicating that uracil-substituted DNA was not interpreted, probably due to its degradation. In *Drosophila* cells, however, transfection with U-plasmid leads to expression of the fluorescent protein at levels comparable to that observed in cells transfected with normal plasmid.

The drug 5'-fluorouracil, frequently used as an inhibitor of thymidylate metabolism, leads to perturbation of nucleotide levels thereby inducing major increase in the uracil content within DNA (4). Fig. 1B and Supplementary T1 report that 5'-fluorouracil destroys human cells with a half-maximal dose ranging between 0.5-300 μ M, while *Drosophila* cells tolerate this drug very well: no inhibition of growth is observed.

These results can be explained only by assuming that the genetic information stored in uracil-substituted DNA serves as a cognate code for transcription in *Drosophila* cells. Such ability of the fruitfly cells is most probably due to lack of UNG so that U-DNA does not get rapidly degraded.

Having established that *Drosophila* cells in *in vitro* culture may tolerate and do correctly interpret uracil-substituted DNA, we wished to assess the physiological response of the fruitfly to silencing of key factors involved in uracil-DNA metabolism. The ubiquitous enzyme responsible for keeping uracil out of

DNA is dUTPase. Efficient silencing could be achieved using well-described transgenic approaches (5) (Fig.S2), lowering dUTPase protein levels under detection limit (Fig.S3A). We observed that this silencing did not perturb normal life and development of embryo and larvae. At early pupal stage, however, dUTPase silencing led to lethality – no silenced animals could develop into imago (Fig.S3B). Morphologic observations indicated serious adverse effects: failure of head eversion and developmental arrest (Fig.1C). This result argued that although dUTPase is dispensable in larval tissues (2), the enzyme is essential for normal development during metamorphosis. Furthermore, it also indicated that tolerance of uracil-DNA (synthesized in lack of dUTPase) may be stage-specific in *Drosophila*.

Recently, we have described the existence of a novel uracil-DNA degrading factor (UDE, CG18410) that is expressed at the latest larval/pupal stage (6). The uracil-DNA degrading action of this protein that becomes apparent at the onset of metamorphosis may explain the stage-specific tolerance of the fruitfly towards dUTPase silencing and also bode a potential role for UDE in *Drosophila* development. To check this conjecture, a transgenic fly exhibiting UDE silencing was produced. Here, silencing led to significant decrease of UDE protein level (Fig.S3C). The UDE-silenced strain behaved very much similar to the dUTPase-silenced strain (Fig.S3D). It showed no effect until the latest larval stages, but about 55 % of the animals failed to complete metamorphosis.

The extraordinary situation of tolerance and interpretation of uracil-DNA may not be exclusively present in *Drosophila*. The major factor that allows this tolerance is quite probably the lack of the *ung* gene that is by far the most powerful catalyst responsible for cleaving uracil out of DNA. Interestingly, we observed that the absence of *ung* gene and appearance of *ude* tend to occur simultaneously among *holometabolous* insects (Fig.1D).

We conclude that uracil-substituted DNA can be tolerated and interpreted in *Drosophila*, serving an exception that proves the rule. Uracil-DNA metabolism may have a signaling role in the fruitfly, as its perturbation *via* silencing of dUTPase and UDE leads to developmental arrest at pupation.

References and Notes

1. M. D. Adams *et al.*, *Science* **287**, 2185 (Mar 24, 2000).
2. A. Bekesi *et al.*, *J Biol Chem* **279**, 22362 (May 21, 2004).
3. S. U. Lari, C. Y. Chen, B. G. Vertessy, J. Morre, S. E. Bennett, *DNA Repair (Amst)* **5**, 1407 (Dec 9, 2006).
4. H. A. Ingraham, B. Y. Tseng, M. Goulian, *Mol Pharmacol* **21**, 211 (Jan, 1982).
5. G. Dietzl *et al.*, *Nature* **448**, 151 (Jul 12, 2007).
6. A. Bekesi *et al.*, *Biochem Biophys Res Commun* **355**, 643 (Apr 13, 2007).
- 7 Materials, methods, supporting figures and tables are available as supporting material on *Science Online*.
8. Funding was provided by OTKA K68229; OTKA-A08 CK78646, HHMI #55000342; Alexander von Humboldt-Stiftung; NKTH JÁP_TSZ_071128_TB_INTER; FP6 LSHG-CT-2006-031220, NKTH-OTKA H07-BEL74200

Supporting Online Material

Materials and Methods
Tables S1 and S2
Figures S1 to S5

Figure 1.

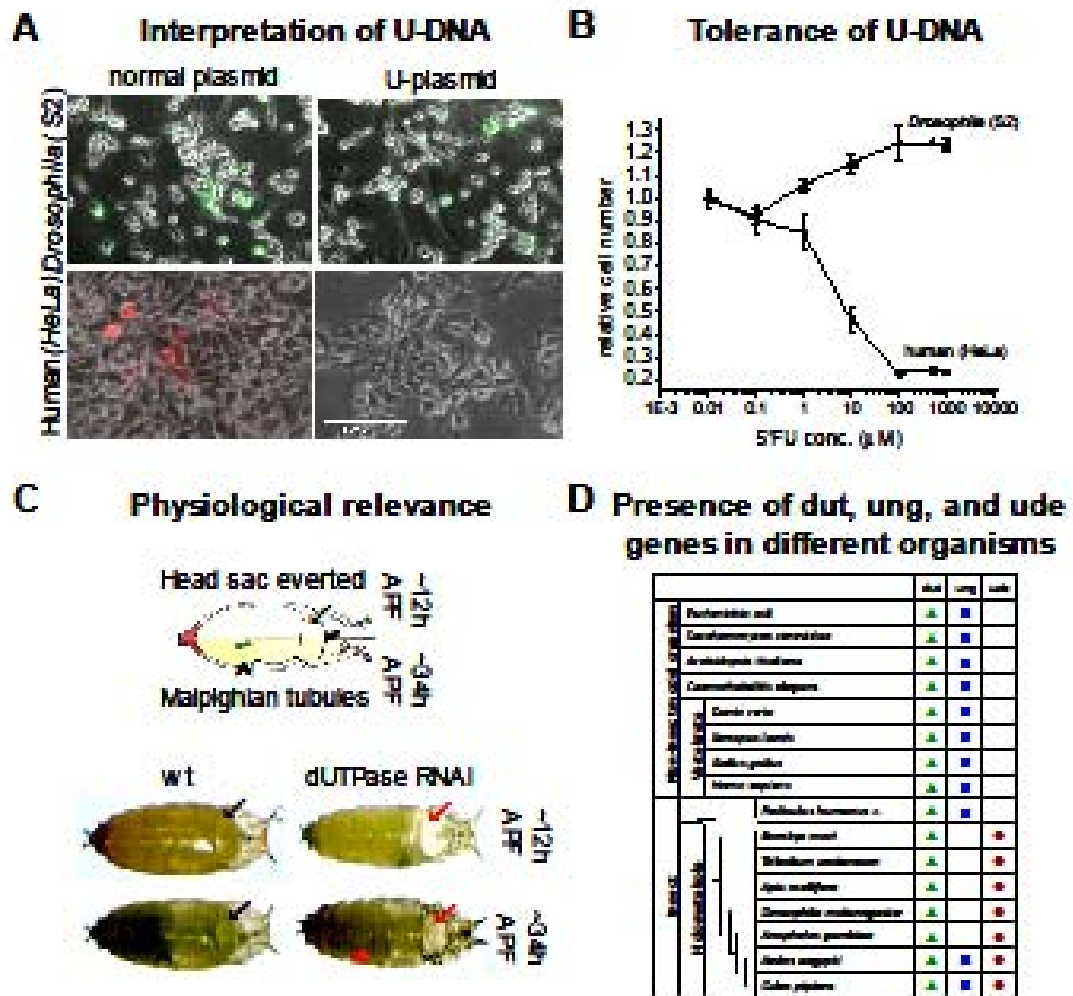


Fig.1 U-DNA interpretation in *Drosophila* models.

(A) *Drosophila* S2 and human HeLa cells transfected with normal plasmid or U-plasmid. *Drosophila* S2 cells were transfected with yellow and human HeLa cells with red fluorescent protein (using plasmids pRmYFP and pdsRedMonomer, respectively, cf Supporting Material). Plasmid with high uracil-content (U-plasmid) can drive the expression of fluorescent protein in *Drosophila*, but not in human cells. (B) Dose-response curve of 5'-fluorouracil (5'FU) treatment of *Drosophila* S2 cells and human HeLa cells. *Drosophila* S2 cells tolerate high doses of 5'-fluorouracil, a thymidylate synthase inhibitor, reflecting that UDNA may be well tolerated in *Drosophila*. (C) Developmental deficiencies in *Drosophila* pupae upon silencing of dUTPase. Incomplete head sac eversion (°) was detected at ~12h after puparium formation (APF). Malpighian tubules (□) do not appear at ~34h APF indicating that further development is partially arrested. (D) Occurrence of factors involved in U-DNA metabolism, recognition and repair.