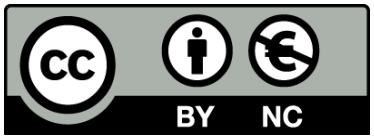
UNIVERSITAT DE BARCELONA

Anàlisi estructural i funcional de l'alcoholdeshidrogenasa de Drosophila

Ricard Albalat Rodríguez



ANÀLISI ESTRUCTURAL I FUNCIONAL DE L'ALCOHOLDESHIDROGENASA DE *Drosophila*

Memòria presentada per Ricard Albalat i Rodríguez per optar al títol de Doctor en Biologia.
Departament de Genètica. Facultat de Biologia. Universitat de Barcelona.
Programa de Genètica. Bienni 1987-1989.
Barcelona, Gener de 1994.

Ricard Albalat i Rodríguez

Vist i plau de la directora de la Tesi.

Dra. Roser Gonzàlez Duarte
Catedràtica de Genètica.

Seria injust no deixar constància de tota la gent que m'ha ajudat a fer aquest treball, per això, vull donar les gràcies,

A la Roser Gonzàlez i Duarte, directora d'aquest treball, per totes les coses que ha fet per mi.

A tota la gent que en un moment o altre han estat en el grup de "Mol", Sivilia Atrian, Susanna Balcells, Mònica Bayés, Neus Cols, Bru Cormand, Joan Fibla, Laura Gort, Daniel Grinberg, Javier Llanos, Teresa Luque, Gemma Marfany, Amalia Martinez, Lluis Ribas, Marc Valls, Lluïsa Vilageliu, Neus Visa, ja que sense el seu ajut segur que aquest treball mai s'hagués pogut fer.

Al grup d'"Epi", Jaume Baguñà, José Ramón Bayascas, David Bueno, Ferran Burgaya, Salvador Carranza, Andreu Casali, Joan Collet, Montse Corominas, Lluis Espinosa, Jordi Garcia, Agustí Miralles, Anna Muñoz, Maria Ribas, Marta Riutort, Rafael Romero, Emili Saló, Florenci Serras, Joan Anton Vela, perque han estat sempre disposats a donar-me un cop de mà.

Al grup d'"Evo" en general i més concretament a Josep M. Comeron, Francesc Mestres (Capità Jean-Luc Picard) i Lluis Serra per la seva disposició a atendre els meus dubtes.

A la Isabel Gil, al Xavi Lacasta, a la M. Antonia Martinez, a l'Amparo Mondéjar i a la "Salo" ja que la seva col.laboració m'ha facilitat molt la feina.

Als meus pares i al meu germà, per haver-me fet costat en tot moment. A la meva germana per haver llegit i corregit aquesta Tesi. Al meu nebot, l'Eduard i a tota la meva família encara que pensin que caço mosques.

A l'Anna.

ÍNDEX

1. PRÒLEG

1. PRÒLEG

Ben segur que quan, al segle XV, l'arquebisbe de l'església anglicana James Ussher establí que la Creació de l'Univers s'havia produït a les 8 h de la tarda del 22 d'octubre de l'any 4004 abans de J. C., no podia imaginar-se que cinc segles més tard la teoria del *big bang* establiria que el nostre univers es més vell, uns 12.000 milions d'anys més vell.

En aquell moment, just en el *big bang*, l'univers tenia un tamany nul i estava infinitament calent. A mesura que l'univers s'expandia, la temperatura de la radiació decreixia. Al voltant d'uns cent segons després del *big bang,* la temperatura s'hauria reduït fins a uns mil milions de graus; llavors, i durant unes hores, els protons i els neutrons haurien començat a combinar-se per produir els nuclis d'àtoms de deuteri i posteriorment, d'heli. Durant el milió d'anys següent, l'univers hauria continuat la seva expansió fins que la temperatura hagués disminuït fins a uns pocs milers de graus i els electrons i els nuclis haguessin començat a combinar-se per formar àtoms. Les regions lleugerament més denses començarien a col.lapsar-se fins que el moviment de rotació, adquirit globalment o per parts individuals, fos suficientment ràpid per compensar l'atracció gravitatòria. D'aquesta manera haurien nascut les galàxies.

Amb el temps, el gas hidrogen i l'heli es disgregarien en núvols que començarien a col.lapsar-se degut a la força de la gravetat. La temperatura del gas augmentaria fins estar prou calent per iniciar les reaccions de fusió nuclear. L'hidrogen es convertiria en heli i la calor

produïda augmentaria la pressió, impedint que els núvols es continuessin contraient, romanent estables durant molt de temps en forma d'estrelles. Tantmateix, les estrelles amb més massa necessitarien estar més calentes per compensar la seva major atracció gravitatòria i consumirien ràpidament el seu combustible. Possiblement les regions centrals d'aquestes estrelles es col.lapsarien fins a un estat d'alta densitat, com una estrella de neutrons o un forat negre. Les regions externes podrien ser expulsades amb una gran explosió, denominada supernova, i tornarien a formar part del gas de la galàxia, proporcionant nova materia per a una segona generació d'estrelles. El nostre Sol és una d'aquestes estrelles de segona o tercera generació, format ara fa uns cinc mil milions d'anys. No obstant, una petita part de la matèria es va anar acumulant independentment, formant remolins que es mourien uns contra els altres com rodes engranades; en aquests punts, les partícules de pols xocarien i donarien com a resultat, primer els planetesimals, i, posteriorment, els planetes. Entre aquests, hi trobem la Terra.

La Terra estava inicialment molt calenta. Quan es va refredar i es va formar l'escorça, el gasos que s'havien format sota la superfície es varen escapar per les escletxes i van formar l'atmosfera. Entre aquests gasos hi havia aigua, la qual originà els oceans com a conseqüència de milions d'anys de pluges torrencials. L'atmosfera primitiva molt probablement no contenia oxigen (aquest es va produir posteriorment com a resultat de l'activitat fotosintètica) i actualment es creu que aquesta atmosfera era quasi neutra. Les fonts energètiques més significatives en aquesta Terra primitiva eren la radiació solar, les descàrregues elèctriques, les erupcions volcàniques i els manantials sobrecalentats del fons oceànic.

En aquestes condicions varen poder aparèixer les primeres molècules biològicament significatives ara fa uns quatre mil milions d'anys.

Entre aquestes primeres molècules és possible que aparegués alguna cosa semblant a l'RNA, un polímer amb unes qualitats molt especials. Aquest tipus d'àcid nucleïc es podia replicar, copiant-se la cadena complementària, era capaç de realitzar diverses funcions catalítiques i podia emmagatzemar informació. Un segon tipus de polímers, molt importants biològicament, són els polipèptids; aquestes molècules són funcionalment més versàtils que els àcids nucleïcs degut a la seva major variabilitat (hi poden haver fins a 20 aminoàcids diferents en una proteïna, i només 4 bases en l'àcid nucleïc). La relació que actualment tenen els àcids nucleïcs i les proteïnes és ben coneguda, i la transmissió d'informació dels primers cap a les segones segueix el codi genètic. Encara que no hi ha cap idea clara de com es va originar aquest codi, ni de com varen ser adoptats inicialment els codons, es coneix que la polimerització no enzimàtica de l'RNA és més senzilla si s'alternen purines i pirimidines. Això generaria codons de dos tipus PuPyPu i PyPuPy que codificarien per pèptits ordenats amb aminoàcids hidrofòfics i hidrofílics, que donarien estructures en làmines β, amb un costat hidrofòbic i amb un altre hidrofílic. Aquestes làmines podrien apilar-se i formar estructures semblants a les membranes cel.lulars. Finalment, dins dels processos més inicials de l'evolució, el canvi d'RNA a DNA va suposar que tot mantenir la informació es guanyés estabilitat, qualitat necessària per aconseguir formes més complexes i eficients, com els sistemes cel.lulars.

Les primeres cèl.lules varen aparèixer fa uns tres mil milions d'anys. Si més no, d'aquesta època són uns objectes esfèrics fosilitzats trobats a Austràlia i a Sudàfrica d'aspecte similar a les actuals cianobactèries. El primitiu *progenota,* nom que s'ha donat al precursor de les cèl.lules actuals, va començar a evolucionar cap als tres llinatges cel.lulars que actualment existeixen: les arquebactèries, les eubactèries i els eucariotes.

Aquests tres llinatges van evolucionar independentment, amb més o menys *èxit.* Un primer pas evolutiu va ser la formació de colònies cel.lulars. Si aquestes colònies haguessin restat simplement com una col.lecció de cèl.lules totalment independents, no haguessin tingut molts avantatges respecte les cèl.lules aïllades. No obstant, l'existència de les colònies va permetre un avenç important: l'especialització cel.lular. Aquest tipus d'organització únicament es va donar en els eucariotes, mentre que en les arquebactèries i en les eubactèries no es va produir.

En aquesta carrera per l'*èxit* evolutiu van anar apareixent els diferents phyla animals. Els primers, les esponges o porífers, sorgiren com exemples de colònies cel.lulars, on el grau de diferenciació era molt incipient. Per contra, els celenteris o cnidaris ja varen ser organismes clarament pluricel.lulars amb dues capes de cèl.lules amb destins diferents: l'ectoderm i l'endoderm. La simetria dels celenteris era radial, és a dir, els organismes eran simètrics respecte a qualsevol pla que passés per un determinat eix, anomenat eix de simetria.

Els platelmints varen desenvolupar una tercera capa de cèl.lules, el mesoderm. Un dels tipus cel.lulars que es desenvolupen a partir del mesoderm són les fibres musculars. Això va permetre als platelmints

moure's més fàcilment i amb major eficàcia que els celenteris. Com a conseqüència del moviment, es va produir una polarització de determinats òrgans sensorials a la part de l'animal que precedeix a la resta (el cap) i, conseqüentment, el sistema nerviós es va concentrar a prop d'aquesta zona. Va aparèixer llavors el primer cervell. Amb els platelmints va aparèixer també per primer cop la simetria bilateral. Però els platelmints tenien un problema: l'aliment arribava a les seves cèl.lules simplement per absorció i, per tant, aquesta limitació en l'eficàcia de la distribució dels recursos, feia que el seu tamany estigués sempre dins d'uns màxims determinats. Encara que podien ser molt llargs, no podien ser gaire gruixuts i per això se'ls coneix com "cucs plans".

Els nemàtodes, al solucionar aquest problema, van poder augmentar el seu volum. A més a més, els nemàtodes van inventar l'anus i, per tant, van poder millorar el procés digestiu. A partir dels nemàtodes van aparèixer tres nous phyla amb noves aportacions: els moluscs, els equinoderms i els anèlids.

Encara que els nemàtodes podien tenir volum, els faltava un sistema de suport-protecció que fos eficaç. Els moluscs van desenvolupar una closca forta i rígida anomenada exoesquelet. També els equinoderms van desenvolupar un sistema de suport-protecció. En aquest cas no era exterior, sinó que es trobava dins de l'animal, és a dir, era un endoesquelet. No obstant, tant l'exoesquelet com l'endoesquelet van suposar un augment important del pes, van impedir, en gran mesura, el moviment que fins llavors s'havia aconseguit.

El tercer dels phyla, els anèlids, no van seguir la mateixa estratègia. Aquests varen inventar la segmentació: l'individu es forma a partir d'unitats bàsiques que es repeteixen al llarg de l'animal. Això va permetre una organització, una flexibilitat i una eficàcia més gran que no pas les que podem trobar en qualsevol altre organisme no segmentat.

Lògicament, però, el pas següent en l'*èxit* evolutiu es trobaria en els phyla que poguessin combinar ambdues estratègies, cosa que es va fer almenys dues vegades. Així, de la combinació de la segmentació + l'exoesquelet van sorgir els artròpodes, i de la combinació de segmentació + endoesquelet, els cordats.

És difícil determinar quin dels dos phyla és el que ha tingut més *èxit*, doncs, encara que des d'un punt de vista antropocèntric els cordats, dins els quals trobem els mamífers, que inclouen l'home, sembla que hagin guanyat la partida, hem de recordar que els artròpodes, que inclouen els insectes, són el phylum més extés i més diversificat que hi ha a la biosfera. Potser, doncs, no hi hagi res d'estrany que en el present treball conflueixi un representant de cadascun d'aquests phyla: la que ha estat estudiada, la *Drosophila* dels artròpodes; i el qui l'ha estudiat, qui ho descriu, per part dels cordats. No sé si, tampoc en aquest cas, quedarà clar qui ha estat el vençut i qui el vencedor.

Però ja fa molt temps que no apareixen nous phyla ni estructures radicalment noves. Està tot inventat? Se li han acabat els recursos a l'Evolució? Respecte a això hi ha dues idees que crec interessants. La primera suggereix que, de la mateixa manera que es va produir el pas de la cèl.lula aïllada a la colònia cel.lular i finalment a l'organisme

pluricel.lular on cada una de les seves cèl.lules ha adquirit una determinada funció, la nova evolució passa per l'organització dels individus independents en comunitats o conjunts d'organismes (ramats, bandades, etc.) i d'aquests a les societats organitzades, on cada individu té una determinada especialitat, com la societat humana en el cas dels cordats, o com les societats que formen determinats insectes (les formigues, les abelles, etc.), en el cas dels artròpodes.

La segona idea parteix del fet que l'evolució s'ha produït gràcies a que hi ha hagut un fluxe d'informació al llarg del temps. El suport físic d'aquesta informació han estat els gens. Per tant, aquests fins ara han estat l'únic mitjà de transport del qual es disposava. Però actualment hi ha un altre tipus d'informació, les idees que es generen en la societat humana i que també es transmeten al llarg del temps. Aquesta informació utilitza un nou medi de transport que s'ha denominat "meme". De la mateixa manera que els gens passen d'un individu a un altre al llarg de les generacions, els memes també es propaguen d'un cervell a un altre al llarg del temps i poden ser, doncs, objecte d'evolució.

Aquest treball és el resultat de'una barreja d'una mica de tot el que hem parlat: Energia, Àtoms i Molècules, Àcids Nucleïcs i Proteïnes, Bactèries, Artròpodes, Cordats, Gens i, espero que també, algun Meme.

2.INTRODUCCIÓ

2. INTRODUCCIÓ

Les tècniques de manipulació i de seqüencia del DNA han provocat una veritable revolta a l'hora d'estudiar i entendre l'evolució. Així, l'anàlisi comparatiu de les estructures de DNA, RNA o proteïnes de diferents organismes esdevé en la majoria dels casos el mètode més acurat i resolutiu emprat en l'estudi de l'evolució i especiació dels éssers vius. Així doncs, s'han pogut resoldre qüestions no aclarides per la metodologia clàssica en absència d'un registre fòsil adient, o quan els caràcters morfològics no eren suficientment resolutius.

Uns dels àcids nucleïcs més utilitzats per a aquesta mena d'anàlisi han estat els RNAs, en especial els rRNAs. A partir d'aquests s'ha establert una filogènia general dels organismes(Olsen, 1988; Woese 1981), alhora que ha permés confirmar la hipòtesi endosimbionta sobre l'origen de les mitocòndries i els cloroplasts (Kuntzel and Kochel, 1981; Margulis and Schwartz, 1982; Walsby, 1986).

Els estudis de les seqüències del DNA codificant també han estat utilitzades en la determinació de les relacions entre espècies, i en determinats mecanismes evolutius com la construcció dels gens, la pèrdua o el guany dels introns o la correspondència entre els exons i els dominis proteïcs (Gilbert, 1978; Traut, 1988).

Finalment, seguint la metodologia de determinació de l'estructura primària de les proteïnes s'han comparat moltes seqüències proteïques i s'han analitzat els dominis funcionals i estructurals. Així s'ha pogut demostrar que existeix una clara relació entre la funció i la conservació

d'una seqüència determinada (Jörnvall et al. 1981). A més, avui dia es disposa de la tecnologia adient per comprovar si els residus conservats d'una proteïna juguen el paper essencial postulat. La mutagènesi dirigida és una eina cabdal per dur a terme canvis específics en el DNA que es traduiran en alteracions puntuals en la proteïna.

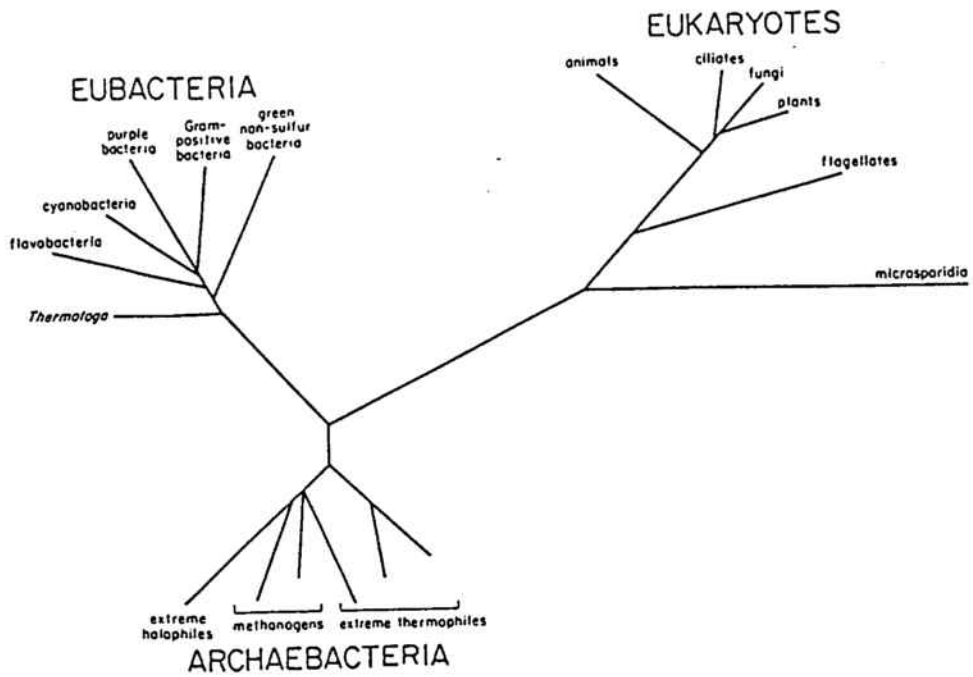


Figura 1. Arbre filogenètic universal sense arrel basat en les comparacions del RNA ribosomal on apareixen els tres regnes primaris: Eubacteris, Archeobateris i Eucariotes (extret de Woese, 1987).

Però, quan ens proposem analitzar un sistema gen-proteïna en un grup d'espècies, hem de triar un model adient. Per un costat, si el que volem és determinar la història evolutiva de determinades espècies, el sistema a utilitzar dependrà de la major o menor divergència que esperem que tinguin els organismes que volem estudiar. Per un altra, el nombre

d'espècies analitzades ha de ser el major possible per tal d'evitar que conclusions esbiaixades donin una visió parcial de la història evolutiva.

Per tant, a l'iniciar aquest treball, i pel tipus d'anàlisi que voliem realitzar, el sistema Alcoholdeshidrogenasa tenia per a nosatres un interés i un atractiu especial degut a determinades característiques que explicarem tot seguit.

2.1. EL GEN ALCOHOLDESHIDROGENASA EN EL GÈNERE *Drosophila*

CARACTERÍSTIQUES GENERALS

2.1.1. En la majoria de les espècies l'enzim ADH és codificat per un gen de còpia única. El gen alcoholdeshidrogenasa (Adh) en la major part d'espècies de *Drosophila* és, a nivell funcional, un gen de còpia única. A *D.melanogaster* aquest gen va ser aïllat per Goldberg (1980) i caracteritzat per Benyajati et al., 1980, i es localitza genèticament al braç esquerre del cromosoma 2, en la posició 50.1 (Grell et al., 1965). A nivell citològic, el locus Adh està comprés entre les bandes 35B1 i 35B3 (Goldberg et al. 1983). A més, l'anàlisi mutacional d'aquesta regió genòmica sembla indicar que el gen Adh podria trobar-se totalment inclòs dins d'un altre gen, Osp (outspread) (Chia et al., 1985). El gen Adh a *D.pseudoobscura* ha estat localitzat en la banda 36 del cromosoma U, element considerat homòleg al 2L de *D. melanogaster* (Loukas et al., 1979; Schaeffer i Aquadro, 1987). En altres espècies del grup *obscura* (*D.guanche, D.madeirensis* i *D.subobscura*) es varen trobar múltiples llocs d'hibridació, posant-se així de manifest l'existència de seqüències genòmiques homòlogues a l'Adh. Amb sondes homòlogues s'ha pogut determinar la localització del gen en espècies pertanyents a d'altres subgèneres i, per tant, filogenèticament més allunyades de *D.melanogaster* com *D. immigrans* o *D. lebanonensis* (Visa et al., 1991; Papaceit i Juan., 1993).

El locus Adh al grup *repleta* ha sofert un fenomen de duplicació de manera que aquestes espècies tenen dos gens Adh funcionals a més d'un

pseudogen (Ψ-Adh) (Batterham et al., 1984; Fisher i Maniatis, 1985; Atkinson et al., 1988; Menotti-Raymond et al., 1991).

2.1.2. **El gen Adh a *D.melanogaster* consta de 3 exons i està regulat per dos promotors diferents.** El gen Adh complet té una longitud aproximada de 1,5 Kb. La part codificant està dividida en tres exons de 99, 405 i 267 parells de bases respectivament, separats per dos petits introns de 65 i 70 nucleòtids.

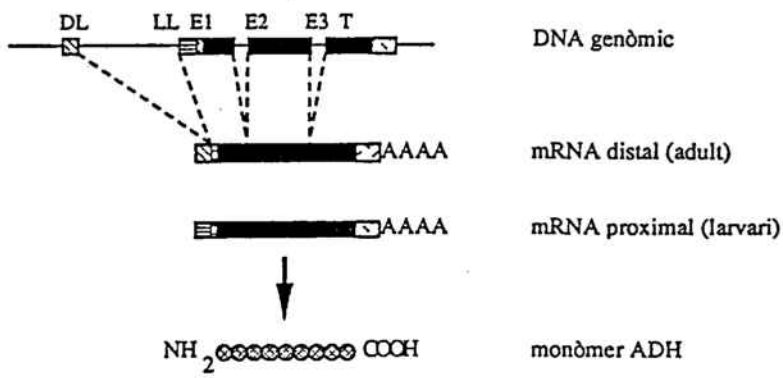


Figura 2. Estructura del gen Adh a *D.melanogaster*. El gen té 3 exons (E1, E2 i E3) i està regulat per dos promotors, distal (DP) i proximal (LP). DL, leader distal; LL, leader larvari; T, trailer (extret de Marfany, 1991a)

La transcripció del gen està dirigida per dos promotors diferents, que segons la distància a què es troben del triplet d'iniciació (ATG) s'anomenen proximal i distal. Els transcrits que es produeixen a partir de cadascun dels promotors són diferents en els extrems 5' no traduïts, però ambdós dicten la síntesi de la mateixa proteïna (Benyajati et al., 1983). A partir del promotor distal s'obté un trànscrit que conté un intró addicional de 654 pb (anomenat intró de l'adult) que posteriorment és processat, de manera que els trànscrits larvari i adult comparteixen els

37 nt anteriors al primer codó. Ambdós promotors consten d'una caixa TATA consensus, d'acord amb les descrites pels gens eucariotes (Breathnach i Chambon, 1987). El senyal de poliadenilació consensus es troba a uns 120 pb en direcció 3' del codó de terminació.

El gens Adh mantenen aquest tipus d'estructura dins el gènere *Drosophila*, amb petites variacions en les mides dels introns, especialment l'intró de l'adult i el primer exó. El primer exó a *D. melanogaster* i espècies properes té 99 pb, i només 93 pb a la resta. Això suposa que la proteïna codificada varií en dos aminoàcids (el segón i tercer) segons les espècies. Finalment, l'estructura trobada dins les espècies del grup *willistoni,* on el gen Adh ha perdut el segon intró, constituiria fins ara una única excepció dins del gènere (Anderson et al., 1993).

2.1.3. **L'expressió del gen Adh varia al llarg del desenvolupament.** Ben aviat es varen determinar dos màxims d'activitat ADH a *D. melanogaster*: en el tercer estadi larvari i en l'individu adult. Així, l'activitat s'incrementava al llarg de tot l'estadi larvari per decréixer bruscament en la fase de pupa i tornar a incrementar-se després de l'eclosió. Aquest patró d'activitat encara que majoritari presenta algunes variacions entre espècies, especialment en l'activitat de la pupa (Batterham et al., 1983; Marfany, 1991a).

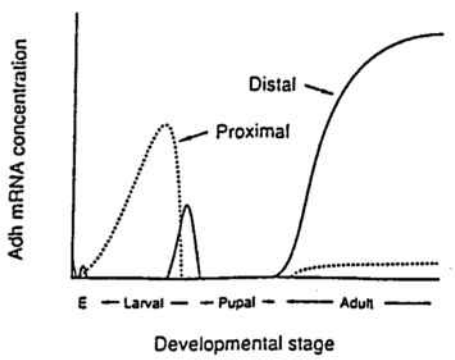


Figura 3. Expressió del gen Adh al llarg del desenvolupament de *Drosophila*. Línea contínua, expressió a partir del promotor distal; línea puntejada, a partir del promotor proximal (extret de Corbin i Maniatis, 1989a).

En les espècies del grup *repleta*, en què, com ja hem indicat, existeixen dos gens, cadascun amb el seu promotor, el patró d'expressió mimetitza les altres espècies, on l'expressió de l'Adh-1 correspondria a la del promotor proximal i la de l'Adh-2 a la del distal (Fisher i Maniatis, 1985).

2.1.4. **L'expressió del gen Adh és específica de teixit.** El gen Adh no s'expressa en tots els teixits, sinó que presenta una distribució específica. Aquesta distribució és consistent amb les funcions atribuïdes a aquesta proteïna, sent elevada en les cèlules del cos gras i variable en d'altres teixits com l'intestí mig, tubs de Malpigi i tractes genitals o reproductors (Savakis et al., 1986; Locket i Ashburner, 1989; Visa, 1991).

REGULACIÓ DE L'EXPRESSIÓ

2.1.5. **La interferència transcripcional podria ser la causa de l'alternança dels promotors.** Per explicar l'alternança d'ús dels dos promotors, a *D. melanogaster* s'ha proposat el mecanisme d'interferència transcripcional (Corbin i Maniatis, 1989b): Dos *enhancers* diferents participarien en la regulació de l'expressió. L'AAE (Adh Adult Enhancer) localitzat entre -590 i -128 pb del promotor distal. L'ALE (Adh Larval Enhancer) localitzat entre -5000 i -660 pb del promotor distal. Segons el model proposat, els factors de transcripció d'unió a l'ALE i al promotor proximal estarien presents durant l'estadi larvari i adult, mentre que els factors de l'AAE i del promotor distal serien específics de l'estadi adult. La unió d'aquests darrers iniciaria la transcripció a partir del promotor distal, *interferint* la transcripció a partir del promotor proximal (Corbin i Maniatis, 1989b)

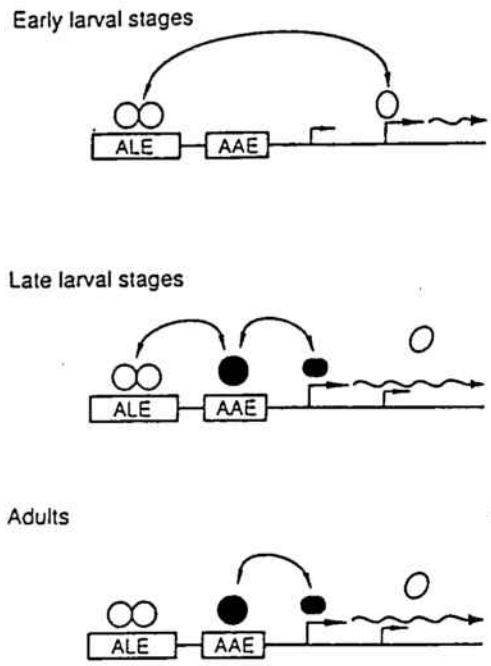


Figura 4. Model de interferència transcripcional en l'alternança de l'ús dels promotors. AAE, enhancer de l'adult; ALE, enhancer larvari; cercles i ovals blancs, factors d'unió a l'ALE i al promtor proximal respectivament; cercles i ovals negres, factors d'unió a l'AAE i al promotor distal respectivament (extret de Corbin i Maniatis, 1989b).

2.1.6. **Hi ha molts factors que regulen l'expressió transcripcional del gen.** Moltes regions que intervenen en la regulació del gen Adh han estat caracteritzades per l'aïllament del factor que se'ls uneix, per experiments de digestió amb DNasa, o bé mitjançant mapatge per delecció. Per tant, en determinades ocasions una mateixa regió ha estat descrita per mètodes diferents i ha rebut denominacions diferents. Així, hi ha un biaix en la quantitat d'informació acumulada: pel que fa a la del promotor distal és molt superior a la del proximal. Les regions d8, d7, d6, d4, d1, p8, p2, p1, boxA o p0 i *A-rich region* (Moses et al., 1990; Rowan i Dickinson, 1988; Heberlein et al., 1985) i els factors BBF-2, AEF-2, C/EBP, DEP1, DEP2, Adf-2, Adf-1 (Abel et al., 1992; Falb i Maniatis, 1992; Ayer i Benyajati, 1992; Benyajati et al., 1992) són, respectivament, algunes de les regions reconegudes pels factors descrits que participen en aquesta regulació.

EVOLUCIÓ MOLECULAR

2.1.7. La regió genòmica del gen Adh ha sofert diversos fenòmens de duplicació. La regió genòmica del gen Adh ha estat implicada en diversos processos de reordenacions cromosòmiques que han tingut destins evolutius diferents.

L'ADH de *Drosophila* pertany a una família de deshidrogenases que no presenten relació estructural amb les alcoholdeshidrogenases de llevat, plantes superiors i organismes vertebrats. S'hipotetitza que procedeix d'un gen ancestral comú per a totes elles i de funció desconeguda. En l'actualitat les semblances estructurals situen aquesta deshidrogenasa de *Drosophila* conjuntament amb les prostaglandinesdeshidrogenases de mamífers (Krook et al., 1990; Persson et al., 1991), la proteïna 25kD de *Sarcophaga peregrina* (Matsumoto et al., 1985) la P6 de *Drosophila*, (Rat et al., 1991) la ribitoldeshidrogenasa de *Klebsiella* (Dothie et al., 1985) i l'Adh-dup a *Drosophila* entre d'altres.

A més a més, en el cas del grup *repleta* existirien dos gens Adh i un pseudogen que haurien evolucionat a partir d'una mateixa seqüència i haurien divergit respecte la seva funció.

Finalment, s'han descrit fenòmens de retrotranscripció del gen Adh en tres espècies molt properes del grup *obscura* i s'ha demostrat que els retrotranscrits s'han reincorporat al genoma en posicions diferents i, posteriorment, s'han multiplicat per fenòmens de transposició.

2.1.8. El gens Adh i Adh-dup estan emparentats evolutivament. Per a nosaltres, el cas de l'Adh-dup és el que ha presentat més interés.

Aquest va ésser descrit per primera vegada a *D. pseudoobscura* (Schaeffer i Aquadro, 1987) i posteriorment trobat també a *D. mauritiana, D. simulans* (Cohn i Moore, 1988), *D. melanogaster* (Kreitman i Hudson, 1991), *D.ambigua* (Marfany i Gonzàlez-Duarte, 1991b), *D. subobscura* (Marfany i Gonzàlez-Duarte, 1992a), *D. guanche* i *D. madeirensis* (Marfany i Gonzàlez-Duarte, 1993). Totes aquestes espècies tenen en comú que pertanyen al subgènere *Sophophora*. L'Adh-dup, generat per una duplicació, es troba situat a uns 400 nucleòtids de distància cap a l'extrem 3' del gen Adh i presenta unes homologies de seqüència tant a nivell nucleotídic com aminoacídic superiors a les que es podrien esperar per atzar. A més a més, ambdós gens presenten una distribució d'exons-introns equivalent. Els exons 1 i 2 en totes les espècies descrites tenen 96 i 405 nt. respectivament. L'exó 3 presenta una variació important de longitud deguda a diferències en l'extrem 3' codificant, encara que supera sempre els 300 nt de longitud. Dels dos introns que separen aquests exons, el primer (aproximadament 250 nt) és més gran que el segon i molt més gran que el primer intró de l'Adh mentre que el segón té la mida esperada, 60 nt. El gen Adh-dup presenta, a més a més, seqüències consensus a 5' i 3' de la regió codificant com són caixes TATA i CAT, i el senyal de poliadenilació, respectivament.

2.1.9. **En el grup *repleta* probablement s'han produït dues duplicacions del gen Adh al llarg de l'evolució.** Un segon fenomen de duplicació es va produir dins del grup *repleta* (subgènere *Drosophila*). Segons sembla deduir-se de les espècies que la presenten, aquesta es va produir en els orígens del grup (Batterharm et al., 1984). S'han seqüenciat els gens de *D. mulleri* (Fisher i Maniatis, 1985), *D.*

mojavensis (Atkinson et al., 1988), *D. hydei* (Menotti-Raymond et al., 1991) i *D. navojoa* (Sullivan et al., 1990). A partir de la comparació d'aquestes seqüències s'ha proposat l'existència inicial d'un sol gen (amb una estructura possiblement similar a la que actualment presenta *D. melanogaster*) que hauria patit una primera duplicació. El gen situat més a 5' hauria evolucionat perdent funcionalitat per esdevenir actualment un pseudogen. El gen més a 3' s'hauria tornat a duplicar, formant-se els dos gens que actualment coneixem com Adh-1 i Adh-2 (Atkinson et al., 1988).

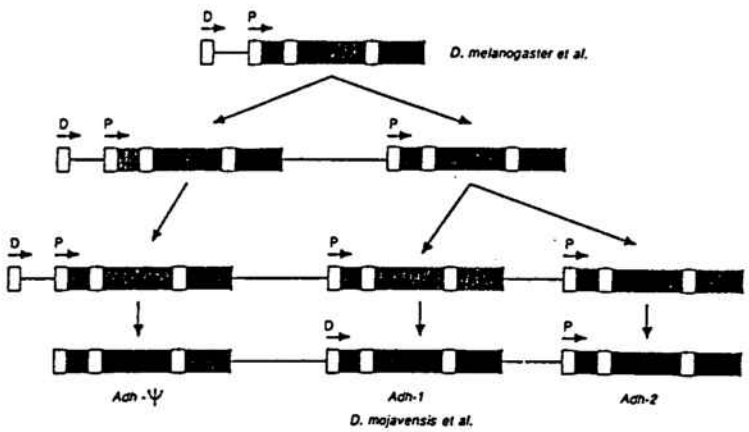


Figura 5. Evolució del gen Adh dins el subgrup repleta. P i D, promotors proximal i distal respectivament (extret d'Atkinson et al., 1988).

Si el model fos correcte podriem esperar poder trobar alguna espècie en la què la segona duplicació no s'hagués produït i, per tant, únicament presentés el pseudogen i un gen funcional. Això és precisament el que s'ha trobat a *D. mettleri*.

2.1.10. **El gen Adh ha estat retrotranscrit i transposat.** Dins del subgènere *Sophophora* i més concretament en el grup *obscura* han estat aïllades i caracteritzades seqüències que tindrien el seu origen en la retrotranscripció de l'mRNA del gen Adh (Marfany i Gonzàlez-Duarte,

1992b). Encara que això no és un fenomen de duplicació estricta, sí que ha suposat una amplificació de la seqüència d'aquest gen dins el genoma de *Drosophila*. Aquest tipus de fenomen ha estat trobat a *D. subobscura, D. madeirensis* i a *D. guanche* i la comparació a nivell de seqüències de DNA permet afirmar que ha estat recent dins de la història evolutiva. A més a més, la seqüència retrotranscrita i insertada ha estat multiplicada (transposada) diverses vegades, multiplicació que també ha afectat les regions veïnes (Marfany and Gonzàlez-Duarte, 1992b).

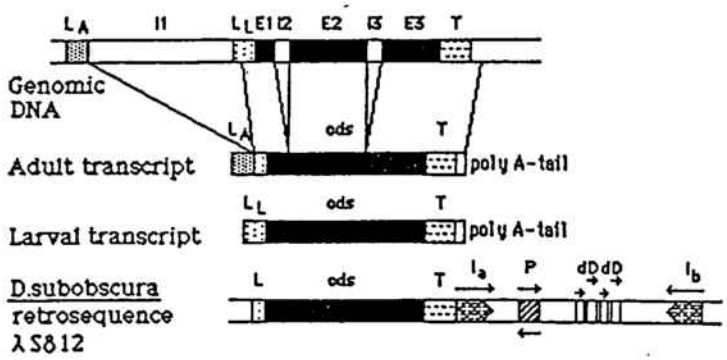


Figura 6. Origen de les retroseqüències del gen Adh a *D.subobscura*. L, leader; LA, leader de l'adult; LL, leader de la larva; E1, E2, E3, primer, segon i tercer exó; I1, I2 i I3, primer, segon i tercer intró; cds, seqüència codificant; T, trairler; Ia i Ib, repeticions invertides; P, seqüència palindròmica; d i D, dues repeticions directes invertides (extret de Marfany i Gonzàlez-Duarte, 1992b).

2.2. LA PROTEÏNA ALCOHOLDESHIDROGENASA

2.2.1. L'enzim ADH: característiques estructurals i cinètiques. L'enzim alcoholdeshidrogenasa (ADH) de *Drosophila* (EC. 1.1.1.1.) és un dímer compost per dues subunitats iguals, d'aproximadament 27.000 Da cadascuna i sense cap ió metàl.lic associat a la seva estructura. L'enzim pertany a la família de les deshidrogenases de cadena curta formades per aproximadament uns 250 aminoàcids, a diferència de les deshidrogenases de cadena mitja o llarga que contenen uns 380 o 750 aminoàcids, respectivament.

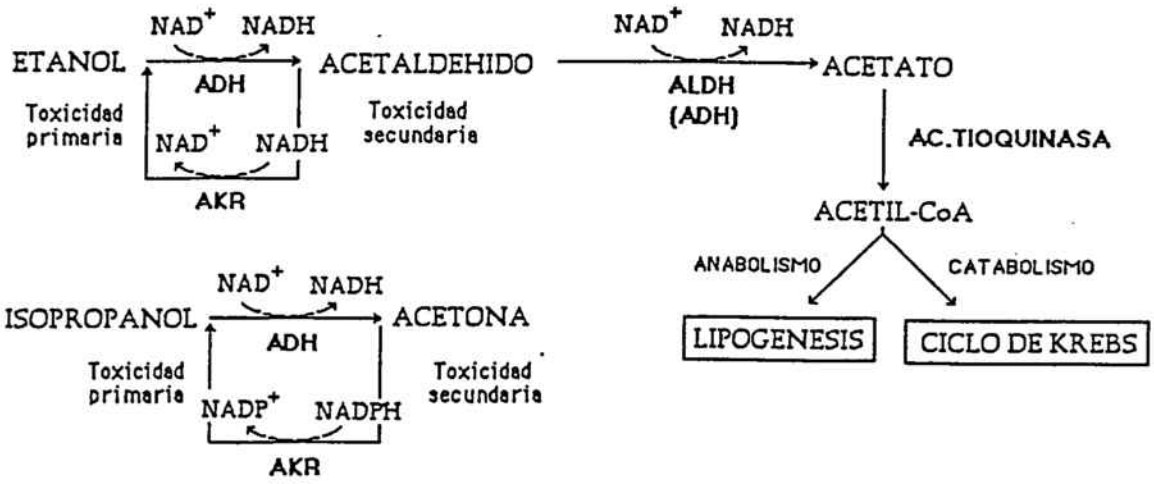


Figura 7. Metabolisme dels alcohols primaris i secundaris a *Drosophila* (extret de Atrian i Gonzàlez-Duarte, 1986).

L'enzim catalitza l'oxidació d'alcohols primaris i secundaris a aldèhids i cetones respectivament, i redueix el cofactor NAD^+ a NADH. S'han determinat les propietats bioquímiques i cinètiques de l'enzim en moltes

espècies (Thatcher, 1980; Vilageliu i Gonzàlez-Duarte, 1980; Juan i Gonzàlez-Duarte, 1981; Atrian i Gonzàlez-Duarte, 1982, 1985) i s'ha comprovat que catalitza més eficaçment l'oxidació dels alcohols secundaris que la dels primaris, sense que s'hagi trobat cap explicació per a aquest fenomen, ja que els alcohols secundaris no són components freqüents en els hàbitats de *Drosophila*.

2.2.2. **L'ADH té una funció metabòlica i detoxificant.** La funció primordial de l'enzim a *Drosophila* és la metabolització dels alcohols, produits en processos de fermentació, que es troben en el seu hàbitat, conjuntament amb la capacitat detoxificant en front d'elevades concentracions d'aquests compostos en el medi. Es creu que també podria estar implicada en la síntesi de l'hormona juvenil o en el metabolisme de feromones (Madhavan et al., 1973; Winberg et al., 1982a i b).

Així, l'etanol pot ser utilitzat com a font nutritiva transformant-lo en acetaldehid. S'ha comprovat que l'ingesta d'etanol produeix, a través del pas per acetat, un increment de triacilglicèrids en el còs gras (Geer et al., 1985), encara que la via metabòlica implicada en aquest procés no es coneguda amb profunditat (David et al., 1978; Garcin et al., 1983; Atrian i Gonzàlez-Duarte, 1986).

2.2.3. **Les comparacions de les constants cinètiques entre espècies és útil per a l'anàlisi del mecanisme enzimàtic.** El primer nivell per entendre qualsevol procés enzimàtic és caracterizar les constants cinètiques de la reacció catalitzada: Km i kcat. Els valors de la Km reflexen l'afinitat que existeix entre l'enzim i el substracte. A mesura que s'incrementen els valors de la Km, decreix l'afinitat de

l'enzim pel substracte doncs la concentració del substracte ha de ser més elevada perquè l'enzim tingui la meitat dels llocs actius ocupats. A més a més, sota condicions freqüentment emprades d'excés de substracte respecte l'enzim, la Km equival a la constant de dissociació del complexe Enzim-Substracte, valors de Km elevats indiquen que aquesta unió és feble, mentre que valors baixos correspondrien a unions E-S fortes. La segona constant cinètica, la kcat, indica el valor de la taxa de recanvi d'un enzim quan la concentració dels llocs actius és coneguda. Aquests valors per a la majoria dels enzims amb el seus substractes fisiològics, es troben entre 1 i 10^4 per segon. La kcat ha estat utilitzada com una mesura de l'eficàcia de l'enzim i el cocient kcat/Km pot considerar-se com una estima de la velocitat "in vivo" de la reacció, on generalment la concentració del substracte està lluny dels valors de la Km.

L'enzim ADH de diferentes espècies de *Drosophila* i de variants al.lèliques de *D.melanogaster* ha estat purificat, seqüenciat i caracteritzat bioquímicament (Batterham et al., 1983; Thatcher, 1977, 1980; Chambers et al., 1981a, b, 1984; Winberg et al., 1982a, b, 1983, 1985, 1986; Juan i Gonzàlez-Duarte, 1980; 1981; Chambers, 1984; Gibson et al., 1980; Vilageliu i Gonzàlez-Duarte, 1984; Atrian i Gonzàlez-Duarte, 1985; Villarroya et al., 1989; Hovik et al., 1984; Winberg i McKinley-McKee, 1988a,b). L'anàlisi d'aquests resultats ha permès extreure algunes conclusions (Winberg i McKinley-McKee, 1992): 1) En tots els casos el pH òptim de funcionament de l'enzim es troba al voltant de 9.0; 2) Malgrat les diferències en l'estructura primària de la regió postulada d'unió a l'alcohol, l'especificitat pel substracte ha estat mantinguda, probablement per a la conservació de la topologia al llarg de l'evolució; 3) Els paràmetres

cinètics indiquen que determinades posicions (14, 82, 192, 214, etc...) estan directament o indirectament relacionades amb la unió del cofactor; 4) Molt possiblement una tirosina ha de participar en la catàlisi de la reacció. Les tirosines 63, 152 i 178 es troben conservades en totes les ADHs de *Drosophila.*

2.2.4. La mutagènesi dirigida i l'homologia amb altres deshidrogenases indica que el domini d'unió al cofactor es troba a l'extrem N-terminal de l'enzim: Els enzims de la família de les deshidrogenases de cadena curta a què, com recordem, pertany l'ADH, catalitzen reaccions redox amb la participació de l'NAD^+ com a cofactor. Les prediccions d'estructura secundària han localitzat la regió d'unió del cofactor en la regió N-terminal (aminoàcids 1-140) (Benyajati et al, 1981; Thatcher i Sawyer, 1980) i, malgrat les homologies de les seqüències aminoacídiques conegudes d'unió a l'NAD^+ són molt baixes, l'estructura tridimensional hauria de ser molt similar en tots els casos. Així, aquest domini consta de sis fulls β i sis hèlixs α distribuïts en dues meitats simètriques, cadascuna de les quals és considerada com un *motiu d'unió mononucleotídic* o *Rossmann fold* (Rossmann et al., 1974).

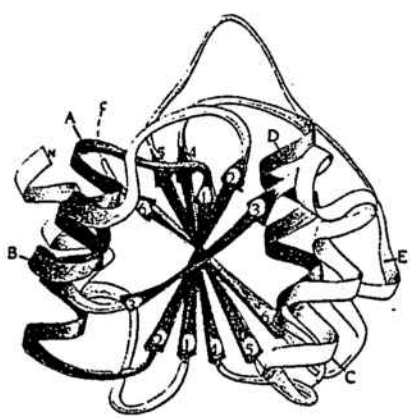


Figura 8. Domini d'unió de l'NAD^+ en la LADH on s'indiquen els sis fulls β (1...6) i cinc helix α (A...E) (extret de Branden i Tooze, 1991).

En aquesta regió hi ha uns determinats residus clau que fan possible la predicció, a partir de la seqüència, d'aquelles regions polipeptídiques que estan involucrades en la unió de l'NAD^+. Una regió en particular, el motiu $\beta1$-αA-$\beta2$, que presenta una estructura molt conservada, ha estat utilitzada per identificar aquestes regions d'unió en els enzims dels quals es desconeix la seva estructura tridimensional.

Aquest motiu abraça prop de 30 aminoàcids i conté tres residus de glicina (Gly-X-Gly-X-X-Gly), sis residus hidrofòbics conservats que formen el nucli hidrofòbic entre l'helix i les cadenes β, i, finalment, un aspàrtic conservat en l'extrem carboxílic de la $\beta2$. Aquesta regió, rica amb Gly, intervé en el posicionament de la part central de l'NAD^+. L'aspàrtic conservat té una funció especial, la d'establir un pont d'hidrogen amb el 2'-OH de la ribosa de l'adenosina. Aquest pont no es pot establir amb l'NADP doncs aquest té un grup fosfat unit al 2'-OH de la ribosa, i d'aquesta manera l'enzim pot discriminar en l'utilització de l'NAD^+ o de l'NADP com a cofactor.

```
    1 . T C A V F G L G G V G L S V I M G C K A A G A A - - R I I G V D I
    2 . K I T V V G V G A V G M A C A O S I L M K D L A D - E V A L V D V
    3 . K I G I D G F G R I G R L V L R A A L S C G A Q - - V V A V N D P
    4 . D Y L V I G G G S G G L A S A R R A A E L G A - - - R A A V V E S
    5 . Q V A I I G A G P S G L L L G Q L L H K A G I - - - D N V I L E R
    6 . V I F V A G L G G I G L D T S K Q L L K R D L K - - N L V I L D R
    7 . T F A V Q G F G N V G L H S M R Y L H R F G A K - - C V A V G E S
    8 . K V C I V G S G D W G S A I A K I V G G N A A - - - Q L A Q F D P
    9 . T V G V L G S G H A G T A L A A W F A S R H V P T A L W A P A D H
1 0 . H V T V I G G G L M G A G I A Q V A A A T G H - - - T V V L V D Q
1 1 . R V V V I G A G V I G L S T A L C I H E R Y H S - - V L Q P L D V
```

Figura 9. Seqüències dels aminoacids de la regio $\beta1-\alpha A-\beta2$ de 11 deshidrogenases diferents (extret de Wierenga et al., 1983). (1), Alcoholdeshidrogenasa de fetge de cavall; (2), Lactatdeshidrogenasa; (3), Gliceraldehid-3-fosfatdeshidrogenasa; (4), Glutatióreductasa; (5), p-hidroxibenzoathidroxilasa; (6), Alcoholdeshidrogenasa de *Drosophila*; (7), Glutamatdeshidrogenasa; (8), Glicerol-3-fosfatdeshidrogenasa; (9), Nopalinsintasa; (10), L-3-hidroxiacil-CoA-deshidrogenasa; (11), D-aminooxidasa. Les fletxes indiquen aminoàcids característics de la regió.

Totes les ADHs seqüenciades a *Drosophila* presenten en el seu extrem N-terminal una seqüència que comparteix aquestes característiques. Aquesta regió ha estat analitzada per mutagènesi dirigida i s'ha pogut demostrar el seu paper en la formació correcta del lloc d'unió de l'NAD^+ a l'enzim (Chen et al., 1990, 1991).

Aquest domini d'unió de l'NAD^+ mereix una atenció especial a *D.lebanonensis*. La seqüència aminoacídica de l'ADH (Vilarroya et al., 1989) presenta 4 canvis únics (Ala13, Phe33, Leu45 i His60) respecte a la resta de les ADH de *Drosophila*. A més, conté 3 canvis en posicions que tan sols han estat substituïdes un altre cop en les ADHs de *Drosophila*. Totes aquestes posicions pertanyen a la regió d'unió amb l'NAD^+ i possiblement són en part responsables de les diferències d'activitat d'aquest enzim. De fet, la proteïna ADH de *D.melanogaster* té una kcat 1,5 vegades superior que la kcat de *D.lebanonensis*, a més d'una termoestabilitat superior. No obstant, les Km per a l'NAD^+ i per al propan-2-ol són similars en ambdues espècies.

Especialment interessants són les posicions Ala13 (Gly14 a *D.melanogaster*) i Asn56 (Thr57 a *D.melanogaster*). La primera d'elles perquè s'ha demostrat mitjançant mutagènesi dirigida, l'important paper que aquest aminoàcid juga dintre de la Rossmann fold on, mutants Gly14Asp ó Val i Gly14Ala, produeixen enzims totalment inactius o amb pèrdues d'un 30% de l'activitat, respectivament. En efecte, el mutant Gly14->Ala de *D.melanogaster* presenta una disminució de la seva kcat d'aproximadament 1,5 vegades, a més d'una alteració de la seva Km per l'NAD^+ que la fa 3 vegades més gran que la de l'enzim salvatge.

La segona posició, Asn56, és interessant en enzims que tinguin com a substractes molècules petites del tipus dels alcohols. Aquest residu probablement es troba dins del tercer full β del domini d'unió del cofactor, flanquejat per l'Asn63, conservada en totes les deshidrogenases de cadena curta, i pels aminoàcids His/Tyr60 i Leu/Ile45, dos dels canvis exclusius de *D.lebanonensis*. A més a més, encara que aquesta no sigui una posició conservada dins de les deshidrogenases de cadena curta, sí que es troba conservada en totes les alcoholdeshidrogenases, incloent les alcoholdeshidrogenases de cadena mitja (posició 238).

2.2.5. **Els aminoàcids entre les posicions 150 i 160 formen part del domini catalític de l'enzim.** Una altra regió d'homologia entre les deshidrogenases de cadena curta la trobem entre les posicions 150 i 160 de la seqüència de l'ADH. Els residus Tyr152 i Lys156 es troben conservats en totes les seqüències i han estat considerats com els aminoàcids involucrats en la transferència catalítica de protons (Krook et al., 1990). Ja que les deshidrogenases de cadena curta no precisen ions metàl.lics per a la seva activitat, la cadena lateral d'un aminoàcid ha de jugar un paper clau en el manteniment del grup hidroxil en la posició correcta per permetre la transferència d'electrons al coenzim.

Diverses dades (Winberg et al., 1982a i b, Krook et al., 1990 i Ensor and Tai, 1991) obtingudes amb la 15-hidroxiprostaglandinadeshidrogenasa de placenta humana suggereixen que la Tyr152 podria ser la responsable d'aquest manteniment. A més a més, l'anàlisi cinètica de la catàlisi de l'ADH i la seva dependència del pH indica que la unió del substracte amb el complex enzim-cofactor es manté gràcies a la ionització d'un radical amb un pKa proper a 7,6 (Winberg i McKinley-McKee, 1988a).

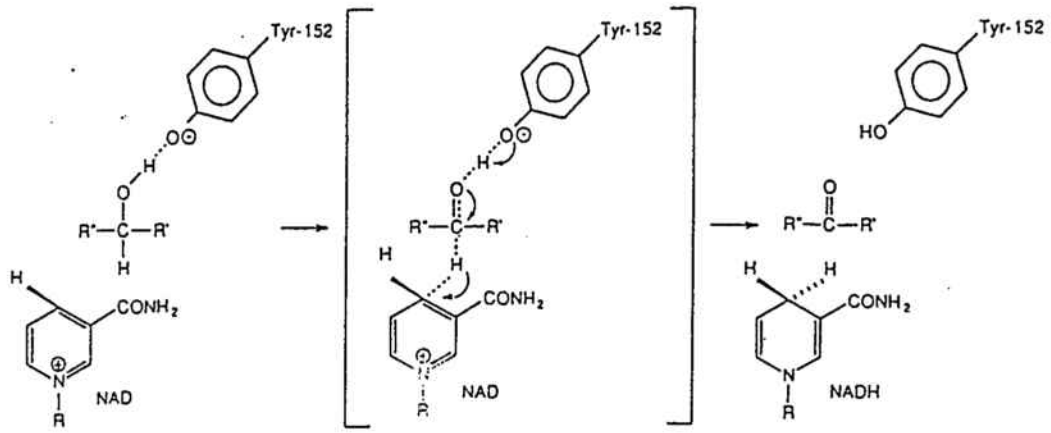


Figura 11. Representació del possible centre actiu de l'Adh de *Drosophila* on la Tyr152 facilitaria la transferència d'electrons durant l'oxidació de l'alcohol (extret de Chen et al.,1993).

S'ha proposat que la forma ionitzada d'aquest grup s'uneix al grup alcohol del substracte, produint l'alcoholat i permetent així la subsegüent transferència electrònica a l'NAD^+ (Winberg i McKinley-McKee, 1992). Una situació d'aquest tipus la podriem trobar si considerem que la cadena lateral de la Tyr152 està ionitzada i presenta un pKa perturbat per a la presència de la Lys156 que podria proporcionar el medi alcalí adequat per a la catàlisi. El grup ε-NH2 seria necessari per produir una disminució de la pKa aparent, facilitant així la reacció enzimàtica (Cols et al. 1993).

2.2.6. **L'extrem C-terminal de la proteïna participa en la catàlisi.** Ha estat descrit que l'ADH, de la mateixa manera que l'esteroideshidrogenasa, és sensible a la proteolisi amb la proteasa Glu-C i la proteasa Asp-N i que aquesta sensibilitat està associada a la inactivació de l'enzim. El lloc d'atac d'aquestes proteases és el Glu243 a *D.melanogaster*, situat a l'extrem C-terminal de la proteïna i s'ha suggerit

l'existència d'una associació entre el lloc actiu de l'enzim i aquesta sensibilitat (Krook et al., 1992).

A més, ha estat proposada una hipòtesi del mecanisme de reacció de l'ADH on l'extrem C-terminal i més concretament l'His250 jugarien un paper fonamental (Ribas, 1992). Segons aquest model, el cofactor s'uniria inicialment a l'enzim i modificaria d'alguna forma el centre actiu permetent l'entrada del substracte i el tancament d'aquest centre per el "loop" C-terminal de la proteïna. A les hores l'His250 actuaria captant un protó del grup hidroxil que, conjuntament amb la transferència d'un segon protó a l'NAD^+, provocarien l'oxidació de l'alcohol. La reacció acabaria amb l'obertura del "loop", l'alliberament dels productes de la reacció i la pèrdua del protó captat per l'His250. L'enzim estaria aleshores en disposició de repetir de nou la reacció.

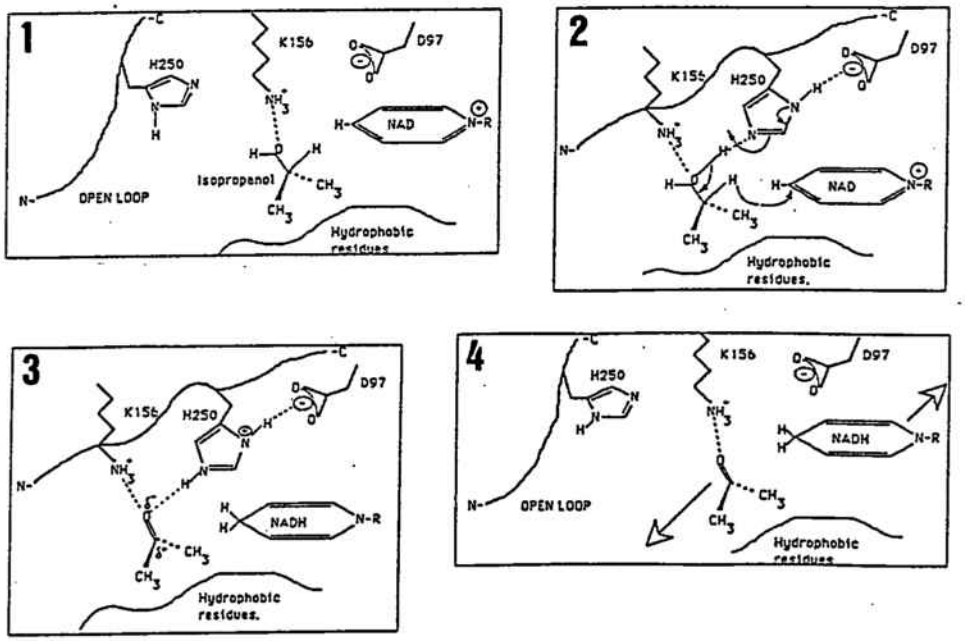


Figura 11: Model del mecanisme de reacció de l'ADH en què el "loop" de l'extrem C-terminal participa activament en l'oxidació de l'alcohol (extret de Ribas, 1992)

Encara que aquest model no deixa de ser una hipòtesi, el que sí sembla molt probable és que l'extrem C-terminal participi almenys en proporcionar l'ambient hidrofòbic necessari per a la reacció d'oxidació de l'alcohol i, per tant, qualsevol modificació d'aquest extrem pot comportar variacions de l'activitat enzimàtica de la proteïna. Els residus, Trp251, Lys249 (Gln249 en alguns casos), Trp247 i Ile245 (Val245 en alguns casos) per les seves característiques i el seu elevat grau de conservació podrien participar també activament en la funcionalitat d'aquest extrem C-terminal.

3. OBJECTIUS

3. OBJECTIUS

Els objectius generals d'aquest treball han estat l'estudi de l'evolució de la regió genòmica del gen Adh a *Drosophila* i l'anàlisi, a partir de la mutagènesi dirigida, de l'activitat enzimàtica de la proteïna. Per fer-ho, s'han assolit els següents objectius parcials:

1. Aïllament, caracterització i seqüenciació del gen Adh de les espècies *Drosophila lebanonensis* (Subgènere *Scaptodrosophila*) i *Drosophila immigrans* (Subgènere *Drosophila*).

2. Aïllament, caracterització i seqüenciació del gen Adh-dup de *D. lebanonensis* i *D.immigrans*.

3. Anàlisi de les seqüències i comparació amb els gens Adh i Adh-dup d'altres espècies.

4. Mutagènesi de l'aminoàcid Tyr152, possible centre actiu de l'enzim, de les posicions Ala13 i Asn56, que pertanyen a la regió d'unió amb el cofactor, i de l'extrem C-terminal de la proteïna.

7. Expressió i purificació de la proteïna mutada en llevat i en bacteris.

8. Estudi de l'activitat dels enzims mutats.

4. RESULTATS

4.1. DNA I PROTEÏNA

The Adh in *Drosophila*: Chromosomal location and restriction analysis in species with different phylogenetic relationships.

Chromosoma (1991) 100:315–322

CHROMOSOMA

The *Adh* in *Drosophila*: Chromosomal location and restriction analysis in species with different phylogenetic relationships

N. Visa, G. Marfany, L. Vilageliu, R. Albalat, S. Atrian, and R. Gonzàlez-Duarte

Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Diagonal 645, E-08071 Barcelona, Spain

Received August 9, 1990
Accepted January 11, 1991 by H. Jäckle

Abstract. Restriction analysis of the genomic region containing the *Adh* gene and in situ hybridization assays were performed in six *Drosophila* species belonging to three different subgenera: *D. ambigua, D. subobscura, D. madeirensis* and *D. guanche* (sg. *Sophophora*); *D. immigrans* (sg. *Drosophila*); and *D. lebanonensis* (sg. *Pholadoris*). In agreement with previous observations, comparison of restriction maps of the *Adh* region shows that *D. subobscura* and *D. madeirensis* are very closely related. Partial homology is also observed with the rest of the *obscura* group species. Nevertheless, no resemblance at the restriction map level is detected when more distantly related species are compared. In *D. ambigua, D. immigrans* and *D. lebanonensis* in situ hybridization assays reveal a single chromosomal location for *Adh*, which in *D. lebanonensis* appears to be sex linked. In contrast, in *D. subobscura, D. madeirensis* and *D. guanche* multiple sites of hybridization with homologous and heterologous probes are observed. For example, in *D. subobscura* and *D. madeirensis* the functional *Adh* gene is located on the *U* chromosome and additional homologous retrosequences are found on the *E* chromosome.

Introduction

The analysis of related genes and proteins from different species has proved to be of crucial relevance in evaluating evolutionary changes. Phylogenies of some *Drosophila* species have been proposed through the analysis of DNA structure of homologous genes, DNA cross hybridization assays and, at the protein level, following allozyme polymorphisms or analysis of immunological features. The polytene chromosomes of *Drosophila* and some other *Diptera* also allow comparative evolutionary analysis and, in cases where external morphologies are similar, the cytogenetic relationship between species has been decisive (Ruiz and Fontdevila 1982). Such comparisons are much more informative when species are closely related and form hybrids, as the homology between chromosome elements is apparent as a result of chromosome pairing in polytene salivary gland cells. When species are distantly related, however, comparison of chromosome elements is difficult and indirect methods have to be used.

The original haploid karyotype of the genus *Drosophila* consists of five large chromosomes (elements *A–E*) and a dot chromosome (*F*), which have remained quite intact through the evolution of the genus. These earlier observations (Muller 1940; Sturtevant and Novitski 1941) have been confirmed by several comparative studies (Loukas et al. 1979; Pinsker and Sperlich 1984; Krimbas and Loukas 1984; Steinemann et al. 1984). Although the chromosomal elements have retained the essential identity, some structural rearrangements have been described. Paracentric inversions and fusions, which are obvious mechanisms for altering gene order within each chromosomal element, have been reported to occur relatively frequently in *Drosophila*, but translocations and pericentric inversions are thought to be rare (Clayton and Guest 1986). However, there is evidence for the establishment of both translocations and pericentric inversions in natural populations. The location of some pteridine loci appears to reflect the occurrence of a pericentric inversion (Gregg and Smucker 1965). In addition, the location of the 5S RNA gene cluster is different in *D. hydei* and *D. melanogaster* (Alonso and Berendes 1975) and the 18/28S RNA genes of some Hawaiian *Drosophila* species are autosomal rather than *X* linked (Stuart et al. 1981).

In situ hybridization with DNA probes is a powerful method for cytological mapping of specific genes and tracing karyotypic changes during evolution (Steinemann 1982; Steinemann et al. 1984; Loukas and Kafatos 1986; Whiting et al. 1989). The aim of the present study is to determine the chromosomal location and structural organization of the genomic region containing the alcohol dehydrogenase gene (*Adh*) in several *Drosophila* spe-

Offprint requests to: R. Gonzàlez-Duarte

cies, in order to evaluate the evolutionary changes undergone by this structural region in a group of differently related species. Four species, *D. subobscura, D. madeirensis, D. guanche* and *D. ambigua*, belong to the *obscura* group (sg. *Sophophora*). The other two belong to different subgenera: *D. immigrans* (sg. *Drosophila*) and *D. lebanonensis* (sg. *Pholadoris*).

In *D. melanogaster* as well as in other *Drosophila* species the *Adh* gene has three exons interrupted by two introns and its expression is under the control of two different age-specific promoters (Goldberg 1980; Benyajati et al. 1983). Different organization of this genomic region has been described in species of the *repleta* group (Fischer and Maniatis 1985). In addition, at the 3′ flanking region of *Adh* an open reading frame (3′ORF) showing a considerable degree of homology with *Adh* has been reported (Schaeffer and Aquadro 1987).

We determined the restriction maps of the region containing the *Adh* locus for each of the aforementioned species and recombinant clones carrying part of this region were used as homologous or heterologous probes for in situ hybridization assays.

Materials and methods

Fly stocks. Flies from natural populations caught in Spain were maintained in our laboratory for several years. The following species were used in our study: *D. ambigua* (Vidrà, Barcelona, 1985), *D. guanche* (Canary Islands, 1976), *D. madeirensis* (Madeira Islands, 1984), *D. subobscura* (strain H27, with standard chromosomal arrangements, provided by R. de Frutos, Dept. of Genetics, University of Valencia, Spain), *D. immigrans* (Bordils, Girona, 1979) and *D. lebanonensis* (Gandesa, Tarragona). A second strain of *D. lebanonensis* provided by the Umeå Stock Center (Sweden) was used for chromosome preparations. An isogenic strain of *D. melanogaster* (Adh^F, provided by Brian Clark, Dept. of Genetics, University of Nottingham, UK) was used as a control.

Preparation of genomic DNA and construction of libraries. Total genomic DNA from each species was isolated using the guanidine isothiocyanate method initially described for RNA extraction (Chirgwin et al. 1979) with minor modifications. High molecular weight total DNA was partially digested with MboI, then fractionated in a sucrose gradient to obtain DNA fragments of 15–20 kb. The libraries were constructed by ligating this prepared DNA into the BamHI site of an EMBL 4 phage vector (Frischauf et al. 1983) for *D. subobscura, D. madeirensis, D. guanche* and *D. ambigua*, or into λ-Charon 35 (Wilhelmine and Blattner 1983) for *D. immigrans* and *D. lebanonensis*. Between 150000 and 200000 recombinant plaques from each species were screened for sequences homologous to the *D. melanogaster Adh* gene. A 2.7 kb HindIII-EcoRI restriction fragment of sAC1 (Goldberg 1980) that contains the complete *Adh* gene and the 3′ adjacent region was used as a probe. It was labelled either by nick-translation (Rigby et al. 1977) or by random-hexamer priming (Feinberg and Vogelstein 1983) with [α-^{32}P]dCTP (NEN-DuPont) and hybridized to phage DNA on nitrocellulose filters (Hybond-C, Amersham) in 46% formamide at 42° C overnight in the presence of 10% sodium dextran sulphate. Nonspecifically bound probe was removed with one wash in 2×SSC, 0.1% SDS at room temperature for 10 min, then two washes in 2×SSC, 0.1% SDS at 65° C for 10 min each, and two further washes in 1×SSC, 0.1% SDS at 65° C for 10 min each (1×SSC is 0.15 M NaCl, 0.015 M sodium citrate, pH 7.5). After the wash steps, the filters were autoradiographed and DNA from positive recombinant phages was isolated according to Maniatis et al. (1982).

Restriction analysis of positive clones. Fragments from positive clones of each species were subcloned into pUC18 or Bluescript vectors (Table 2) and characterized by mapping with the following restriction enzymes: BamHI, EcoRI, HindIII, PvuII, PstI, SalI and XbaI (Boehringer Mannheim). Single and double digestions were performed (Maniatis et al. 1982). *Adh* was located on the restriction map by transferring digested and electrophoresed DNA to nitrocellulose membranes (Hybond-C, Amersham) using the transfer method of Southern. Labelling of the *Adh* probe, hybridization, washing of the filters and autoradiography, were the same as for plaque hybridization. Sequencing was performed by the dideoxy method of Sanger et al. (1977) and according to conventional techniques (Maniatis et al. 1982).

In situ hybridization. Salivary glands dissected from third instar larvae were fixed and squashed in 45% acetic acid. Coverslips were removed after freezing the preparations in liquid nitrogen and slides were immediately dehydrated in ethanol and air dried. Subsequently, they were incubated in 2×SSC at 65° C for 30 min, dehydrated in ethanol and air dried. Just before hybridization the slides were immersed in 0.07 *N* NaOH for 2 or 3 min to denature chromosomal DNA, washed in 2×SSC (3×5 min), dehydrated and air dried. DNA probes were labelled by nick-translation (Rigby et al. 1977) using Bio-11-dUTP or Bio-14-dATP (Enzo and BRL). For ten slides, 1 μg of biotinylated probe was dissolved in 200 μl solution containing 50% formamide, 10% dextran sulphate, 0.4 μg salmon sperm DNA in 2×SSC. This hybridization mixture was incubated at 90° C for 10 min and applied to the squashes. The slides were then coverslipped, sealed and incubated in a moist chamber at 80° C for 10 min. After hybridization at 37° C for 16–20 h the slides were washed in 2×SSC (5×10 min). The first wash was at room temperature, the following two at 42° C and the last two at room temperature. Then the slides were incubated for 5 min in PBS containing 0.1% Triton X-100 and washed in PBS (2×5 min). Detection of biotin was carried out by incubating the slides with a streptavidin-peroxidase complex (Detek I-hrp Signal Generating System, Enzo) for 90 min. The slides were then washed in 2×SCC (2×5 min), then in PBS containing 0.1% Triton X-100 (5 min) and finally in PBS (5 min). Peroxidase activity was detected using diaminobenzidine as a chromogen. The preparations were stained with Giemsa, air dried and mounted in DPX.

Chromosome analysis. Chromosomes were examined and photographed in Zeiss photomicroscope. Chromosome nomenclature and identification of bands were based on the following salivary gland chromosome maps: *D. immigrans*, Le Calvez (1953); *D. lebanonensis*, Berendes and Thijssen (1971); *D. subobscura*, Kunze-Mühl and Müller (1958); *D. ambigua*, Mainx et al. (1953); *D. guanche*, Moltó et al. (1987); *D. madeirensis*, Papaceit (1988).

Results

In situ hybridization

Preliminary in situ hybridizations were carried out using the 2.7 kb restriction fragment $sAC_{H/E}$ from plasmid sAC1 as a probe. This fragment, containing the *D. melanogaster Adh* gene, was also used to screen genomic libraries of the six species under study. DNA fragments from positive clones were subcloned in different plasmids and used as homologous probes for in situ hybridization assays (Fig. 1). Sequence analysis revealed that some of the probes contained the functional *Adh* gene whereas others had homologous non-functional sequences (Table 1 and Fig. 2). Results obtained for each species with homologous as well as heterologous probes are summarized in Table 2.

Table 1. Comparison of positions and sequences of the *Adh* genomic region of the *Drosophila* species under study

Region	Parameter	Species					
		D. lebanonensis	*D. immigrans*	*D. ambigua*	*D. subobscura*	*D. guanche*	*D. madeirensis*
Adh							
Adult TATA box	Sequence Position	TATTTAA −1152	PS	TATTTAA −1026	TATTTAA −836	TATTTAA −831	TATTTAA PS
Larval TATA box	Sequence Position	CATAAATA −24	TATAAATA −25	TATAAATA −34	TATAAATA −32	TATAAATA −32	TATAAATA −32
First exon	Nt length Position	93 +45–+137	93 +59–+151	93 +67–+159	93 +65–+157	93 +65–+157	93 +64–−156
Second exon	Nt length Position	405 +199–+603	405 +215–+619	405 +226–+630	405 +226–+630	405 +227–+631	405 +224–+628
Third exon	Nt length Position	267 +670–+936	267 +684–+950	267 +697–+963	267 +699–+965	267 +695–+961	267 +696–+962
Polyadenylation signal	Position	+1064	PS	+1133	+1106	+1108	+1103
3′ORF							
CCAAT box	Sequence Position	PS	PS	GGCAATAAGG +1074	GGCAATAAGG +1053	GGCAATAAGG +1056	GGCAATAAGG +1050
TATA box	Sequence Position	PS	PS	TTAATTAAAA +1207	TTAATTAAAA +1162	TTAATTAAAA +1170	TTAATTAAAA +1159
First exon	Nt length Position	PS	PS	96 +1321–+1416	96 +1289–+1384	96 +1303–+1398	PS
Second exon	Nt length Position	PS	PS	405 +1709–+2113	405 +1670–+2074	405 +1522–+1926	PS
Third exon	Nt length Position	PS	PS	345 +2181–+2525	339 +2137–+2475	PS	PS
Polyadenylation signal	Position	PS	PS	+2536	+2498	PS	PS

PS, partially sequenced; Nt, nucleotide

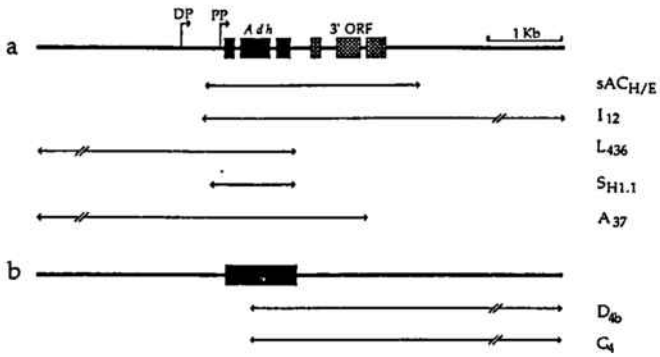


Fig. 1a, b. Probes used for in situ hybridization. Subcloned genomic fragments are represented under the general diagram. **a** The following probes contain the functional *Adh* gene and contiguous sequences: *Drosophila melanogaster* ($sAC_{H/E}$), *D. immigrans* (I_{12}), *D. lebanonensis* (L_{436}), *D. subobscura* ($S_{H1.1}$) and *D. ambigua* (A_{37}). *Solid boxes* represent the coding regions of the *Adh* gene, and *hatched boxes* represent the putative coding regions of the 3'ORF. DP and PP indicate the position of the distal promoter and proximal promoter respectively. **b** The probes from *D. madeirensis* (D_{4b}) and *D. guanche* (G_4) contain an *Adh* pseudogene that is represented as a *solid box*

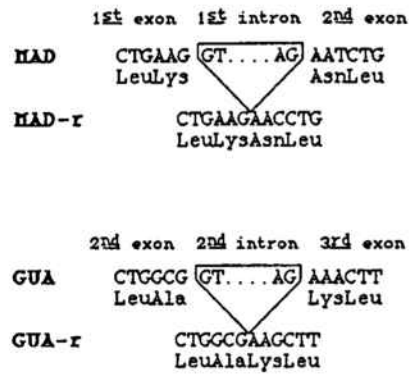


Fig. 2. Comparison of genomic sequences at some splicing junctions between the functional *Adh* gene of *Drosophila madeirensis* and *D. guanche* and their respective retrosequences. MAD, functional *Adh* gene of *D. madeirensis*. MAD-r, *Adh* retrosequence of *D. madeirensis*. GUA, functional *Adh* gene of *D. guanche*. GUA-r, *Adh*, retrosequence of *D. guanche*

The $sAC_{H/E}$ probe failed to give detectable hybrids in *D. guanche, D, immigrans* and *D. lebanonensis* but yielded positive results with *D. ambigua, D. subobscura* and *D. madeirensis*. In these three species hybridization was observed in the *B* chromosome element, homologous to the *2L* chromosome arm of *D. melanogaster* (Table 3). In addition, a second hybridization site was found in division 64 of the *E* chromosome of *D. subobscura*. The two *Adh* hybridization sites found in this species with probe $sAC_{H/E}$ were in agreement with those found using the homologous probe $S_{H1.1}$. However, in this latter case, the hybridization signal observed in the *U* chromosome was very faint, whereas the band in the *E* chromosome (Fig. 3c) was much more intense.

The 9 kb genomic insert of *D. madeirensis* in plasmid D_{4b} did not contain the functional *Adh* gene. DNA sequence analysis of this genomic clone revealed the features of a retrosequence (Fig. 2). When this probe was used to localize its homologous region in *D. madeirensis* chromosomes it hybridized to division 64 in chromosome *E* (Fig. 3f). In contrast, probe $sAC_{H/E}$ was found to hybridize at division 36 near the centromeric region of the *U* chromosome (Fig. 3c).

Three hybridization sites on *D. guanche* were observed when probe G_4 was used (Fig. 3d). Two of them were at division 98 of chromosome *O* and the third at division 62 of chromosome *E*.

D. ambigua showed only one *Adh* homologous region, at division 36 of chromosome *E*, when probes A_{37} and $sAC_{H/E}$ were used (Fig. 3a). A genomic clone containing this region was sequenced and agrees with the sequence expected for the functional gene (Table 1).

Table 2. Recombinant *Adh* probes and hybridization sites in salivary gland chromosomes of several *Drosophila* species

Probe			Hybridization site		
Species	Plasmid	Insert size (kb)	Species	Chromosome	Site
D. melanogaster	$sAC_{H/E}$[a]	2.7	*D. ambigua*	E_R	36
			D. madeirensis	*U*	36
			D. subobscura	*U*	36
				E	64
D. subobscura	$S_{H1.1}$ [Bluescript][b]	1.1	*D. subobscura*	*U*	36
				E	64
D. madeirensis	D_{4b}[pUC_{18}]	9.0	*D. madeirensis*	*E*	64
D. guanche	G_4[pUC_{18}]	9.0	*D. guanche*	*O*	98[c]
				E	62
D. ambigua	A_{37}[pUC_{18}]	4.5	*D. ambigua*	E_R	36
			D. subobscura	*E*	64
				U	36
D. immigrans	I_{12}[pUC_{18}]	5.0	*D.immigrans*	*4*	G_4
D. lebanonensis	L_{436} [Bluescript]	3.6	*D. lebanonensis*	*X*	1

Plasmid vectors containing the inserts are indicated in brackets
[a] $sAC_{H/E}$ is a HindIII-EcoRI restriction fragment from plasmid sAC1
[b] Plasmid $S_{H1.1}$ contains two copies of the 1.1 kb insert
[c] Two hybridization sites in the same division

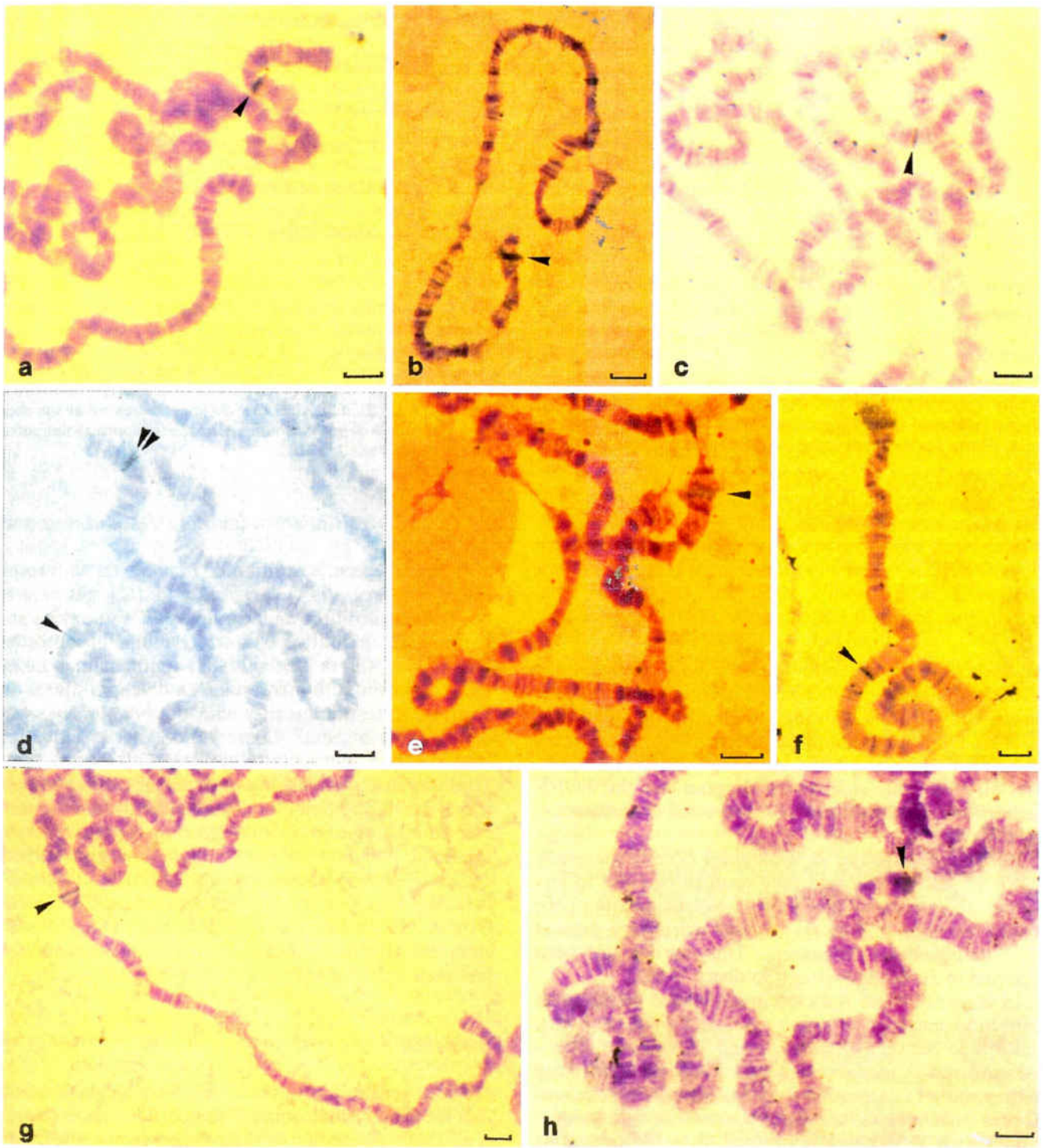


Fig. 3a–h. In situ hybridizations to *Drosophila* salivary gland polytene chromosomes. **a** Hybridization of probe A_{37} to a single site on the E_R chromosome arm of *D. ambigua*. **b** Probe $sAC_{H/E}$ hybridizes to the *U* chromosome of *D. subobscura*. **c** Additional hybridization site on the *E* chromosome of *D. subobscura* with probe $S_{H1.1}$. **d** Probe G_4 hybridizes to three chromosomal sites in *D. guanche*. **e** Hybridization of probe $S_{H1.1}$ to *D. madeirensis* chromosome *U*. **f** Probe D_{4b} hybridizes to the *E* chromosome of *D. madeirensis*. **g** Chromosome four of *D. immigrans* showing the single hybridization site obtained with the homologous probe I_{12}. **h** Probe L_{436} hybridizes to the *X* chromosome of *D. lebanonensis* . *Arrowheads* indicate hybridization sites. Bars represent 10 μm

Chromosome homologies between *D. immigrans, D. lebanonensis* and the other species have not yet been determined. Homologous genomic probes, I_{12} and L_{436} respectively, were used to localize the *Adh* gene in these two species. All the results obtained were consistent and showed that the *Adh* gene is at the G4 site of *D. immigrans* chromosome *4* (Fig. 3g) and at division 1 of the *X* chromosome of *D. lebanonensis* (Fig. 3h).

Table 3. Homologies between the chromosomes of the *Drosophila* species analysed

Species	Element				
	A	*B*	*C*	*D*	*E*
D. melanogaster[a]	*X*	(2L)	*2R*	*3L*	*3R*
D. pseudoobscura[b]	*XL*	(4)	*3*	*XR*	*2*
D. subobscura	*A*	(U)	(E)	*J*	*O*
D. madeirensis	*A*	(U)	(E)	*J*	*O*
D. guanche	*A*	(U)[c]	(E)	*J*	(O)
D. ambigua	*A*	(E)	*JL*	*JR*	*U*
D. immigrans[d]	*X*	*2*	*3*	(4)	*5*
D. lebanonensis[d]	(X)	*2R/2L*	*3R/3L*	*4R/4L*	

Data are derived from Steinemann et al. (1984) and our own results. Chromosomes containing *Adh* homologous regions are circled

[a] From O'Donnell et al. (1977)

[b] From Schaeffer and Aquadro (1987)

[c] Tentative results (see Discussion)

[d] Chromosome homologies with the other species are unknown

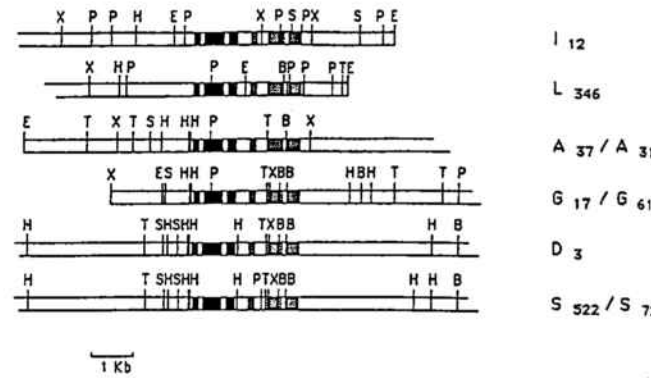


Fig. 4. Restriction map of the *Adh* region of the six *Drosophila* species under study: *D. immigrans* (I_{12}), *D. lebanonensis* (L_{346}), *D. ambigua* (A_{36}/A_{311}), *D. guanche* (G_{17}/G_{611}), *D. madeirensis* (D_3) and *D. subobscura* (S_{522}/S_{722}). Seven restriction endonucleases were used: BamHI (*B*), HindIII (*H*), EcoRI (*E*), PvuII (*P*), PstI (*T*), SalI (*S*), and XbaI (*X*). *Solid boxes* represent the coding regions of the *Adh* genes, and *hatched boxes* represent the putative coding regions of the 3'ORFs

Restriction analysis

Seven restriction enzymes were used to construct the restriction map of the *Adh* genomic region in the species under study (Fig. 4). This region extends over an average of 8 kb DNA and comprises the regulatory sequences of *Adh* with the distal and proximal promoters (Benyajati et al. 1983), the *Adh* structural gene and the 3'ORF (Schaeffer and Aquadro 1987) as revealed by sequencing data (Table 1).

D. lebanonensis and *D. immigrans* do not share any restriction site, neither do they seem to be close to any of the other species. Only a PvuII restriction site in the second exon of *Adh* in *D. lebanonensis* is also present in *D. ambigua* and *D. guanche*. This site has also been reported in *D. melanogaster* (Goldberg 1980).

In contrast, there is a close resemblance between the restriction maps of *D. subobscura*, and *D. madeirensis*, and less so between *D. guanche* and *D. ambigua*. *D. subobscura* and *D. madeirensis* seem to be particularly close as they share 14 restriction sites. In fact, all *D. madeirensis* sites determined in this study are present in *D. subobscura*, which shows two additional sites. Sequences at the flanking regions are also conserved, as three out of four distant sites are present in both species. Close proximity between these two species at the phenotypic and cytogenetic levels has been reported (Krimbas and Loukas 1984). Our results support this resemblance at the DNA level.

D. guanche shares with *D. madeirensis* and *D. subobscura* the four 3'ORF sites (PstI, XbaI, BamHI, BamHI), the two HindIII sites close to the proximal *Adh* promoter and the SalI site located between the distal and proximal promoters. This species also contains the PvuII site in the second *Adh* exon which is present in other *Drosophila* species, although it is not found in *D. madeirensis* and *D. subobscura*.

The phylogenetic position of *D. ambigua* with respect to *D. subobscura*, *D. madeirensis* and *D. guanche* has been established following the analysis of allozyme polymorphisms, quantitative characters and chromosomal homologies (Cabrera et al. 1983; Krimbas and Loukas 1984). According to our results, *D. ambigua* shares only 5 out of 12 restriction sites with the other three species of the *obscura* group. Conservation of restriction sites in the 3'ORF is particularly evident in *D. subobscura*, *D. madeirensis* and *D. guanche*, whereas *D. ambigua* only shows partial homology. Nevertheless it does share the two HindIII restriction sites close to the *Adh* proximal promoter common to the four *obscura* group species analysed. It also shows another HindIII site close to the distal promoter of *D. madeirensis* and *D. subobscura*. Overall, our results on the *Adh* genomic region are in complete agreement with the phylogenetic distances established for these species by Loukas et al. (1984).

Discussion

Previous genetic analysis has revealed a single functional *Adh* locus on chromosome *U* for *D. subobscura* (Loukas et al. 1979). According to these earlier results the hybridization site on chromosome *U* would correspond to the functional *Adh* gene and consequently, the homologous *Adh* region on chromosome *E* might be due to nonfunctional *Adh* sequences (*Adh* pseudogenes). Of all the species analysed, *D. madeirensis* and *D. subobscura* are the most closely related. Both previous analyses and the present data strongly support this relationship. Accordingly, most restriction sites in the *Adh* gene and flanking regions are conserved. When a heterologous probe ($sAC_{H/E}$) was hybridized to the *D. madeirensis* chromosomes a positive band was observed at division 36 of the *U* chromosome. The strong homology described for

this region of F_1 hybrids of *D. subobscura* and *D. madeirensis* (Krimbas and Loukas 1984), the fact that the $sAC_{H/E}$ probe contains functional *Adh* sequences and the cytological location of *Adh* in *D. subobscura* all suggest that the functional *Adh* gene of *D. madeirensis* is located on division 36 of the *U* chromosome. On the other hand, the almost perfect homology between the *E* chromosomes of the two species suggests non-functional *Adh* sequences on the *D. madeirensis E* chromosome. As the $sAC_{H/E}$ probe failed to give a positive signal on this site, probe D_{4b} containing one copy of an *Adh* pseudogene was used. Then, the existence of *Adh* homologous sequences in the expected chromosomal region was clearly confirmed.

In *D. guanche* one hybridization signal was located on the *E* chromosome in a position homologous to that observed in both *D. subobscura* and *D. madeirensis*. Also two more very close sites appeared near the telomeric end of the *O* chromosome. In this case we cannot discard the possibility that the hybridization signal is due to homology with other regions of the probe or to retrosequence-type elements. Furthermore, preliminary results obtained with a probe containing the functional *Adh* gene of *D. guanche* indicate that this gene is located close to the centromeric region of the *U* chromosome. It is most unfortunate that many attempts to define the *U* region involved in hybridization were seriously impaired by the permanently distorted morphology of this chromosomal region, which moreover is close to a break point (Moltó et al. 1987).

The chromosome maps of *D. lebanonensis* have been described by Berendes and Thijssen (1971). The *Adh* gene has been mapped with an homologous probe, L_{436}, and appears to be *X* linked. This is quite an unexpected result because the *Adh* gene in *Drosophila* has never been located on chromosome *X*, so it deserves further consideration. In *D. melanogaster* the chromosomal region containing the *Adh* gene has been found to display a high proportion of weak points, chromosomal breaks, ectopic pairing and other features usually associated with intercalary heterochromatin (Zhimulev et al. 1982). In other species, *Adh* is located near the centromeric region (*D. subobscura, D. madeirensis, D. immigrans*) and so it is again enclosed in a heterochromatic surrounding. Shifts in the amount or position of heterochromatic regions are not infrequent in the evolution of *Drosophila* (Stone 1955). A rearrangement involving a heterochromatic region would then explain the actual location of *Adh* in *D. lebanonensis*. Alternatively, rearrangements of larger chromosomal portions could be involved. In *Drosophila* several examples of centric fusions between the *X* chromosome and an autosome have been illustrated (Stone 1955). For example, in *D. americana* an *X*-autosome fusion involving the *B* element has been reported (Lucchesi 1978) and such a rearrangement might have occurred in *D. lebanonensis* as well. However, taking into consideration the fact that the *X* chromosome of *D. lebanonensis* is a telomeric element, an additional pericentric inversion would have to be assumed to account for its actual configuration. Preliminary results do not support any difference in ADH activity between adult males and females but this would not be surprising if "autosomal genes become readily dosage compensated when entire chromosome arms become newly translocated to the X chromosome during karyotype evolution" (Birchler et al. 1990).

The structural information gathered on the *Adh* gene in *Drosophila* reveals two different organization patterns. The one applying to *D. melanogaster* and to many of the species tested, consists of a single coding region controlled by two different promoters. The other, is made up of two different promoters controlling two different structural sequences tandemly arranged and coding for a very similar enzyme (Fischer and Maniatis 1985). Genomic duplications can generate tandem copies of a gene and this seems to be the case for the 3'ORF adjacent to *Adh*, initially described in *D. pseudoobscura* (Schaeffer and Aquadro 1987). The presence of the 3'ORF in all our six species supports previous claims that it corresponds to an ancient duplication. Furthermore, we here report three species of the *obscura* group that display several chromosomal regions homologous to *Adh*. Nevertheless, previous electrophoretic characterizations of the six species under study have clearly shown that in all cases there is only one active ADH (Vilageliu and Gonzàlez-Duarte 1984; Winberg et al. 1986; Hernández et al. 1988), which is in agreement with the existence of a single functional *Adh* locus.

The wealth of information gathered from the localization and characterization of individual genes in *Drosophila* is providing increasing evidence that the complex mechanisms involved in the genome evolution of vertebrates are also relevant to the dynamics of invertebrate genomes.

Acknowledgements. We are obliged to Dr. M. Izquierdo for encouragement and valuable advice regarding methods. We are indebted to Dr. M. Papaceit for helpful discussions, to M.M. Duro for technical help and to R. Rycroft for revising the English version of the manuscript. This work was partially supported by grants PB88-0188 from DGICYT to RGD, AR-87 from CIRIT (Generalitat de Catalunya) to GM and AR-89 from CIRIT to NV and RA.

References

Alonso C, Berendes HD (1975) The location of 5S (ribosomal) RNA genes in *Drosophila hydei*. Chromosoma 51:347–356

Benyajati C, Spoerel N, Haymerle H, Ashburner M (1983) The messenger RNA for alcohol dehydrogenase differs in its 5' end in different developmental stages. Cell 33:125–133

Berendes HD, Thijssen WThM (1971) Developmental changes in genome activity in *Drosophila lebanonensis casteeli* Pipkin. Chromosoma 33:345–360

Birchler JA, Hiebert JC, Paigen K (1990) Analysis of autosomal dosage compensation involving the alcohol dehydrogenase locus in *Drosophila melanogaster*. Genetics 124:677–686

Cabrera VM, González AM, Larruga JM, Gullon A (1983) Genetic distance and evolutionary relationships in the *Drosophila obscura* group. Evolution 37:675–689

Chirgwin JM, Przybyla AE, MacDonald RJ, Rutter WJ (1979) Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry 18:5294–5299

Clayton FE, Guest WC (1986) Overview of chromosomal evolution in the family Drosophilidae. In: Ashburner M, Carson HL, Thompson JN (eds) The genetics and biology of Drosophila, 3E. Academic Press, New York, pp 1–58

Feinberg AP, Vogelstein B (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132:6–13

Fischer JA, Maniatis T (1985) Structure and transcription of the *Drosophila mulleri* alcohol dehydrogenase genes. Nucleic Acids Res 13:6899–6917

Frischauf A-M, Lehrach H, Poustska A, Murray N (1983) Lambda replacement vectors carrying polylinker sequences. J Mol Biol 170:827–842

Gregg TG, Smucker LA (1965) Pteridines and gene homologies in eye colour mutants of *Drosophila hydei* and *D. melanogaster*. Genetics 52:1023–1034

Goldberg DA (1980) Isolation and partial characterization of the *Drosophila* alcohol dehydrogenase gene. Proc Natl Acad Sci USA 77:5794–5798

Hernández JJ, Vilageliu Ll, Gonzàlez-Duarte R (1988) Functional and biochemical features of alcohol dehydrogenase in four species of the *obscura* group of *Drosophila*. Genetica 77:15–24

Krimbas CB, Loukas M (1984) Evolution of the *obscura* group *Drosophila* species. I. Salivary gland chromosomes and quantitative characters in *D. subobscura* and two closely related species. Heredity 53:469–482

Kunze-Mühl E, Müller E (1958) Weitere Untersuchungen über die chromosomale Struktur und die natürlichen Strukturtypen von *Drosophila subobscura* Coll. Chromosoma 9:559–570

Le Calvez J (1953) Carte cytologique des chromosomes géants des glandes salivaires de *Drosophila immigrans* Sturt. Chromosoma 6:170–174

Loukas M, Kafatos FC (1986) The actin loci in the genus *Drosophila*: establishment of chromosomal homologies among distantly related species by in situ hybridization. Chromosoma 94:297–308

Loukas M, Krimbas CB, Mavragani-Tsipidou P, Kastritsis CD (1979) Genetics of *Drosophila subobscura* populations. VIII. Allozyme loci and their chromosome maps. J Hered 70:17–26

Loukas M, Krimbas CB, Vergini Y (1984) Evolution of the *obscura* group *Drosophila* species. II. Phylogeny of ten species based on electrophoretic data. Heredity 53:483–493

Lucchesi JC (1978) Gene dosage compensation and the evolution of sex chromosomes. Science 202:711–716

Mainx F, Koske Th, Smital E (1953) Untersuchungen über die chromosomale Struktur Europäischer Vertreter der *Drosophila obscura* Gruppe. Z Indukt Abstammungs Vererbungslehre 85: S 354–372

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning: A laboratory manual, 1st edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

Moltó MD, De Frutos R, Martinez-Sebastián MJ (1987) The banding pattern of polytene chromosomes of *Drosophila guanche* compared with that of *D. subobscura*. Genetica 75:55–70

Muller HJ (1940) Bearings of the *Drosophila* work on systematics. In: Huxley J (ed) The systematics. Clarendon Press, Oxford, pp 185–268

Note added in proof. After this manuscript was completed, the location of *D. guanche Adh* on the centromeric region of the U chromosome was confirmed.

O'Donnell J, Mandel HC, Krauss M, Sofer W (1977) Genetic and cytogenetic analysis of the ADH region in *Drosophila melanogaster*. Genetics 86:553–566

Papaceit M (1988) Estudio de las diferencias genéticas entre *D. subobscura* Collins y *D. madeirensis* Monclús. PhD Thesis, University of Barcelona

Pinsker W, Sperlich D (1984) Cytogenetic mapping of enzyme loci on chromosomes J and U of *Drosophila subobscura*. Genetics 108:913–926

Rigby PWJ, Dieckmann M, Rhodes C, Berg P (1977) Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. J Mol Biol 113:237–251

Ruiz A, Fondevila A (1982) The evolutionary history of *Drosophila buzzatii*. III. Cytogenetic relationships between two sibling species of the *buzzatii* cluster. Genetics 101:503–518

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Schaeffer SW, Aquadro CF (1987) Nucleotide sequence of the Adh gene region of *Drosophila pseudoobscura*: evolutionary changes and evidence for an ancient gene duplication. Genetics 117:61–73

Steinemann M (1982) Analysis of chromosomal homologies between two species of the subgenus *Sophophora: D. miranda* and *D. melanogaster* using cloned DNA segments. Chromosoma 87:77–88

Steinemann M, Pinsker W, Sperlich D (1984) Chromosome homologies within the *Drosophila obscura* group probed by in situ hybridization. Chromosoma 91:46–53

Stone WS (1955) Genetic and chromosomal variability in *Drosophila*. Cold Spring Harbor Symp Quant Biol 20:256–269

Stuart WD, Bishop JC, Carson HL, Frank MB (1981) Location of the 18/28S ribosomal RNA genes in two Hawaiin *Drosophila* species by monoclonal immunological identification of the RNA-DNA hybrids in situ. Proc Natl Acad Sci USA 78:3751–3754

Sturtevant AH, Novitski E (1941) The homologies of the chromosome elements in the genus *Drosophila*. Genetics 26:517–541

Vilageliu Ll, Gonzàlez-Duarte R (1984) Alcohol dehydrogenase from *Drosophila funebris* and *Drosophila immigrans*: molecular and evolutionary aspects. Biochem Genet 22:797–815

Whiting JH Jr, Pliley MD, Farmer JL, Jeffery DE (1989) In situ hybridization analysis of chromosomal homologies in *Drosophila melanogaster* and *Drosophila virilis*. Genetics 122:99–109

Wilhelmine AML, Blattner FR (1983) Lambda Charon vectors (Ch 32, 33, 34 and 35) adapted for DNA cloning in recombination-deficient hosts. Gene 26:171–179

Winberg J-O, Hovik R, McKinley-McKee JS, Juan E, Gonzàlez-Duarte R (1986) Biochemical properties of alcohol dehydrogenase from *Drosophila lebanonensis*. Biochem J 235:481–490

Zhimulev IF, Semeshin VF, Kulichkov VA, Belyaeva ES (1982) Intercalary heterochromatin in *Drosophila*. I. Localization and general characteristics. Chromosoma 87:197–227

Adh and Adh-dup sequences of *Drosophila lebanonensis* and *Drosophila immigrans*: interspecies comparisons.

Gene, 126 (1993) 171–178

GENE 06988

Adh and *Adh-dup* sequences of *Drosophila lebanonensis* and *D. immigrans*: interspecies comparisons

(Alcohol dehydrogenase; gene duplication; evolution; nucleotide substitution rate)

Ricard Albalat and Roser Gonzàlez-Duarte

Departament de Genètica, Facultat de Biologia. Universitat de Barcelona, E-08028 Barcelona, Spain

Received by G. Bernardi: 21 July 1992; Revised/Accepted: 30 October/4 November 1992; Received at publishers: 10 December 1992

SUMMARY

We have cloned and sequenced the *Adh* genomic region of *Drosophila lebanonensis* (subgenus *Scaptodrosophila*) and *D. immigrans* (subgenus *Drosophila*). This region, which contains *Adh*, encoding the alcohol dehydrogenase enzyme, and *Adh-dup* (duplicate of *Adh*), has been compared with the same fragment from *D. subobscura* (subgenus *Sophophora*). Even though the flanking regions and introns of both genes have been affected by high substitution rates, the consensus sequences have been clearly identified. Although the overall homology of the coding regions was 76–78% among the species compared, there were differences in the exon distribution of the nucleotide substitutions when *Adh* or *Adh-dup* were compared, thus showing that these two genes differ in their evolutionary pattern.

INTRODUCTION

The ADH system has been extensively analyzed. Because of its particular features and the amount of information gathered at different levels, it constitutes a good model with which to study gene structure and regulation and assess phylogenetic relationships in the *Drosophila* genus (Sullivan et al., 1990).

In *D. melanogaster*, and in other *Drosophila* species, *Adh* consists of three exons separated by two introns. The gene is controlled by two developmentally regulated promoters, adult and larval, and the gene product is tissue-specific. Other species of the genus show a different genomic organization for this region: species of the *repleta* group have two *Adh*, each one with its own promoter and one pseudogene. Recently, we have found other *Adh* pseudogenes which had originated from a retrotranscriptional event in some species of the *obscura* group (Marfany and Gonzàlez-Duarte, 1992b).

Moreover, other rearrangements have taken place in the *Adh* genomic region. A highly conserved sequence was previously described in the 3′ flanking region of *Adh* (Schaeffer and Aquadro, 1987). This sequence, referred to as *Adh-dup*, contained an ORF and shared high homology with *Adh*. However, *Adh-dup* has been reported only in the *Sophophora* subgenus. We now show that it can also be found in the *Scaptodrosophila* and *Drosophila* subgenera. Its presence in species originating among the early radiations of the genus suggests that *Adh-dup* probably constitutes a common element in most *Drosophila* species.

Previous studies in our laboratory showed that ADH activity was present in *D. immigrans* and *D. lebanonensis*. The analysis of these species revealed a direct correlation between ADH activity and alcohol tolerance: *D. lebanonensis*, which was found abundantly in wine cellars, showed a high level of enzyme activity, whereas

Correspondence to: Dr. R. Gonzàlez-Duarte, Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Av. Diagonal, 645, E-08028 Barcelona, Spain. Tel. (34-3) 402-14-94; Fax (34-3) 411-09-69; e-mail: d3gergd0@lcfq.ub.es

Abbreviations: aa, amino acid; ADH, alcohol dehydrogenase; *Adh*, gene encoding ADH; *Adh-dup*, duplicate of the *Adh* sequence; bp, base pair(s); *D.*, *Drosophila*; d.f., degree(s) of freedom; Ka, nonsynonymous substitution rate; kb, kilobase(s) or 1000 bp; Ks, synonymous substitution rate; Myr, 10^6 years; nt, nucleotide(s); oligo, oligodeoxyribonucleotide.

D. immigrans, with low activity, was practically absent in alcohol-rich environments. Differences in ADH activity were due to remarkable variations in the quantity of enzyme, while the specific activity was comparable (Vilageliu and Gonzàlez-Duarte, 1984).

The aim of the present study was to determine the nt sequence of a 3828-bp segment of the *Adh* region of *D. lebanonensis* (subgenus *Scaptodrosophila*) and a 3142-bp segment of *D. immigrans* (subgenus *Drosophila*). In order to analyze the molecular evolution, the sequences were compared with each other and with the same region of *D. subobscura* (subgenus *Sophophora*) (Marfany and Gonzàlez-Duarte, 1992a). It is interesting that two related genes, adjacently located and contained in such a short DNA segment, appear to show different evolutionary patterns.

RESULTS AND DISCUSSION

(a) Clone characterization

The λCharon35 genomic library screens yielded several phage clones homologous to the *D. melanogaster* sAC1 probe (Goldberg, 1980) of *D. lebanonensis* and *D. immigrans*. The *Adh* region was characterized (Visa et al., 1991) and then sequenced for a total of 3828 bp for *D. lebanonensis* and 3142 bp for *D. immigrans*. Both sequences were compared with the same region of *D. subobscura*. The codon positions aligned exactly among all species. Nucleotides in flanking regions and introns appeared to be subject to higher substitution rates, rendering comparisons difficult due to alignment ambiguities. In these cases, it was possible to identify only the consensus sequences.

(b) Analysis of *Adh* nt and deduced aa sequences

By analogy with other species, in the 5'-flanking sequences the *Adh* promoters and transcription start points for *D. lebanonensis* and *D. immigrans* were identified. In *D. immigrans* the sequences of the larval and the adult TATA box were identical to those of all other *Drosophila* species reported to date (Fig. 1). However, the distance between the two promoters was shorter than in any other species, only 559 nt. In *D. lebanonensis* the two TATA boxes were 1114 nt apart, and the larval TATA box (CATAAATA) started with an unusual nt due to a T→C substitution. In addition, the *boxA* sequence, previously characterized in *D. mulleri* as equivalent to the *D. melanogaster p0* binding site, was also identified in both species, and clear blocks of sequence conservation were shown. This *boxA* seems essential for the correct expression of the proximal promoter and for tissue specificity. Of the other regulatory sequences defined in *D. melanogaster* as *p1*, *p2* or *d1*, none was obvious in the sequence comparisons of *D. immigrans* and *D. lebanonensis*. However, assays with Schneider cell cultures transfected with different *D. lebanonensis* and *D. immigrans* genomic constructs obtained from the regions that we have characterized were successful (data not shown). So, the basic elements for the expression of *Adh* had to be present in these genomic fragments.

Within the coding regions, we compared 765 positions and found 171 nt substitutions between *D. immigrans* and *D. lebanonensis*, 174 between *D. immigrans* and *D. subobscura* and 170 between *D. lebanonensis* and *D. subobscura* (Table I). Thus, 22–23% of the sites compared was different among these three species. Substitutions were more frequent in synonymous positions: while only 25% of nt sites were silent, changes in these positions amounted to 63% of the overall figure. Taking into consideration that the third codon position was frequently silent, changes appeared to accumulate there rather than be equally distributed. The rest of the nt sites were nonsynonymous (75%) and represented only 36% of the total changes (about 62).

Silent substitutions were randomly distributed between exons when comparing *D. lebanonensis* versus *D. immigrans*, *D. lebanonensis* versus *D. subobscura* and *D. immigrans* versus *D. subobscura* ($\chi^2=0.196$, $\chi^2=0.162$ and $\chi^2=0.915$, d.f.$=2$, $P<0.05$). However, nonsynonymous substitutions did not appear to be randomly distributed. When *D. immigrans* was compared with *D. lebanonensis* or *D. suboscura*, a specific pattern of distribution of substitutions was found ($\chi^2=10.76$ and $\chi^2=7.64$, d.f.$=2$, $P<0.05$): the second exon showed fewer changes than expected, whereas there was an excess in the first exon. Biochemical data seem to support that the second exon contains some essential information for the catalytic domain of the ADH enzyme (Albalat et al., 1992) and that it could also be relevant for the correct folding of the protein. So, this region is bound to be under high selective constraints.

More than 50% of aa replacements between *D. lebanonensis*, *D. immigrans* and *D. subobscura* were conservative, which is significantly greater than what could be expected by random changes (15%). *D. immigrans* has six aa

Fig. 1. The nt sequence of the *Adh* region of *D. immigrans* (a) and *D. lebanonensis* (b). The deduced aa sequence is shown below. The adult and larval TATA box for *Adh* and the CAT and TATA box for *Adh-dup* are underlined. The *BoxA* sequence for *Adh* is shown in italics. Direct Z1 and complementary Z2 oligo sequences for PCR amplification are indicated by open boxes. Sequence data have been deposited with GenBank under accession Nos. M97638 (*D. immigrans*) and M97637 (*D. lebanonensis*).

(a)

(b)

TABLE I

Comparison of the *Adh* and *Adh-dup* nt sequences of *Drosophila lebanonensis* (DL), *D. immigrans* (DI) and *D. suboscura* (DS)

	Sites and nt exchanges[a]					
	DI/DL *Adh*	DI/DL *Adh-dup*	DI/DS *Adh*	DI/DS *Adh-dup*	DL/DS *Adh*	DL/DS *Adh-dup*
Number of compared sites	765	810	765	822	765	810
Total changes	171	195	174	187	170	194
Total positions (%)	22	24	23	23	22	24
Number of silent changes	108.5	133.0	112.0	119.0	108.5	135.0
Total changes (%)	63.5	68.2	64.4	63.6	63.8	69.6
Number of replacement changes	62.5	62.0	62.0	68.0	61.5	59.0
Total changes (%)	36.5	31.8	35.6	36.4	36.2	30.4
Exon1						
Number of nt	93	96	93	96	93	96
Total nt (%)	12.2	11.9	12.2	11.6	12.2	11.9
Number of changes in exon1	28	15	31	15	20	22
Total changes (%)	16.4	7.7	17.8	8.0	11.8	11.3
Number of silent sites	22.8	20.7	23.7	21.5	23.5	20.5
Silent changes	12.5	15.0	17.0	12.0	12.5	18.0
Number of changes in exon1 (%)	44.6	100.0	54.8	80.0	62.5	81.2
Corrected percent divergence	82.5	93.9	97.3	73.7	68.2	90.8
Number of replacement sites	70.2	75.3	69.3	74.5	69.5	75.5
Replacement changes	15.5	0.0	14.0	3.0	7.5	4.0
Number of changes in exon1 (%)	55.4	0.0	45.2	20.0	37.5	18.2
Corrected percent divergence	26.6	0.0	24.4	3.8	11.7	5.3
Exon2						
Number of nt	405	405	405	405	405	405
Total nt (%)	52.9	50.0	52.9	49.3	52.9	50.0
Number of changes in exon2	81	90	86	87	92	92
Total changes (%)	47.7	46.2	49.4	46.5	54.1	47.4
Number of silent sites	99.3	87.5	109.5	90.3	99.3	89.2
Silent changes	57.0	66.0	61.0	59.5	59.5	64.5
Number of changes in exon2 (%)	70.4	73.3	70.9	68.4	64.7	70.1
Corrected percent divergence	85.3	95.7	94.3	96.7	91.6	93.0
Number of replacement sites	305.7	317.5	303.7	314.7	305.7	315.8
Replacement changes	24.0	24.0	25.0	27.5	32.5	27.5
Number of changes in exon2 (%)	29.6	26.7	29.1	31.6	35.3	29.9
Corrected percent divergence	8.3	8.2	8.8	9.4	11.4	9.2
Exon3						
Number of nt	267	309	267	321	267	309
Total nt (%)	34.9	38.1	34.9	39.1	34.9	38.1
Number of changes in exon3	62	90	57	85	58	80
Total changes (%)	36.3	46.2	32.8	45.5	34.1	41.2
Number of silent sites	63.0	68.2	62.5	68.0	62.8	67.8
Silent changes	39.0	52.0	34.0	47.5	36.5	52.5
Number of changes in exon3 (%)	62.9	57.8	59.6	55.9	62.9	65.6
Corrected percent divergence	91.4	99.0	77.7	100.0	91.8	99.1
Number of replacement sites	204.0	240.8	204.5	253.0	204.2	241.2
Replacement changes	23.0	38.0	23.0	37.5	21.5	27.5
Number of changes in exon3 (%)	37.1	42.2	40.4	44.1	37.1	34.4
Corrected percent divergence	12.2	18.1	12.7	17.5	11.4	12.7

[a] The sequences have been compared using the GCG Sequence Analysis Software Package (GCG, Madison, WI). Compared sites correspond to the three exon nt positions of *Adh* (765) and *Adh-dup* (810 and 822). The number of silent and replacement nt changes is given as percentage of total changes.

changes in positions which had remained unaltered in any other species: N^{96}, V^{244} (conservative), G^{24}, L^{102}, I^{175} and D^{207} (nonconservative). *D. lebanonensis* also had six unique changes: those affecting positions A^{13}, L^{45}, H^{60} (conservative), F^{33}, N^{56} and I^{106} (nonconservative). Nevertheless, the homology of the hydrophilicity profiles of the three ADHs (data not shown) and the biochemical features of these enzymes seem to suggest that they all

share common features at the level of the tertiary structure.

(c) *Adh-dup* expression; PCR analysis

The high level of nt sequence conservation downstream from *Adh* was explained under the hypothesis of a protein-coding gene. Moreover, the homology between *Adh* and this putative gene suggested that it may represent an ancestral duplication of *Adh*; thus, it was named *Adh-dup*. In order to determine whether it was transcribed, we purified total RNA from *D. lebanonensis* (Jowett, 1986). Reverse transcriptase reaction was performed with 1.5–2.0 μg of total RNA using the MoMuLV enzyme (BRL) and the Z2 oligo (Oligo Etc., Wilsonville, OR) (Fig. 1). The Z1 oligo (Fig. 1) and *Taq* polymerase (Perking Elmer or Boehringer Mannheim) were added for the PCR amplification (Kawasaki, 1990). The sequences of Z1 and Z2 were included in exons 2 and 3, respectively. Therefore, the processed RNA (mRNA) lacking the intron would be shorter than the corresponding DNA. The PCR product was subcloned into the pBluescriptSK plasmid and sequenced. The sequence obtained agreed with the proposed exon-intron structure (Schaeffer and Aquadro, 1987). The assay was performed with larval, pupal and adult RNA, and in all cases transcriptional products were detected.

(d) Analysis of *Adh-dup* nt and deduced aa sequences

The *Adh-dup* nt sequences of *D. lebanonensis, D. immigrans* and *D. subobscura* were also compared. As was shown with *Adh*, this gene consisted of three exons separated by two introns. In all the species, the first and second exon were 96 and 405 nt long, respectively. The length of the third exon varied with the species: 309 nt in *D. lebanonensis*, 321 nt in *D. immigrans* and 339 nt in *D. subobscura*. The exon sizes were similar to those of *Adh* (93, 405 and 267 nt, respectively). The first intron of *Adh-dup* appeared to be much longer than that of *Adh* for all the species except *D. immigrans*. The intron-1 length of *Adh-dup* was 258 for *D. lebanonensis*, 285 for *D. subobscura* but only 59 for *D. immigrans*. Taking into consideration that the first intron of *Adh* has approx 60 nt, this may correspond to the length of the ancestor segment, which acquired its present size after several insertion events. The second intron had 60, 59 and 62 nt for *D. lebanonensis, D. immigrans* and *D. subobscura*, respectively, and showed a similar length to the second intron of *Adh*.

Some regulation consensus sequences were identified at the 5′-flanking sequence of *Adh-dup*. A putative TATA box, TAATTAAA, was present in all reported species. We also found a CAT box consensus sequence overlapping the 3′-flanking sequence of *Adh* (Fig. 1). However, as with *Adh*, neither the 5′- or 3′-flanking sequences nor the introns could be properly aligned. Then, only the coding nt positions of *Adh-dup* were compared among the three *Drosophila* species (Table I). We found 195 nt changes (24% of total positions) between *D. immigrans* and *D. lebanonensis*, 187 nt changes (23%) between *D. immigrans* and *D. subobscura* and 194 nt changes (24%) between *D. lebanonensis* and *D. subobscura*. Again, silent substitutions were more frequent that nonsynonymous changes and appeared concentrated mainly at the third codon position. Silent substitutions were randomly distributed among the three *Adh-dup* exons ($\chi^2=0.03$, *D. immigrans-D. lebanonensis*; $\chi^2=0.49$, *D. immigrans-D. subobscura*; and $\chi^2=0.55$, *D. subobscura-D. lebanonensis*; d.f.$=2$, $P<0.05$), whereas replacement changes were not random between *D. immigrans* and *D. lebanonensis* or *D. subobscura* ($\chi^2=17.84$, $\chi^2=8.33$, respectively, d.f.$=2$, $P<0.05$). A random distribution of replacements was found between *D. lebanonensis* and *D. subobscura* ($\chi^2=2.55$, d.f$=2$, $P<0.05$). In *D. immigrans* nonsynonymous substitutions in the first or second exon were lower than expected, whereas the third exon showed an accumulation of nt changes. This trend was also present in the *D. lebanonensis-D. subobscura* pair, although it was not significant. Given the sequence coding homology between *Adh* and *Adh-dup*, the fact that the replacement substitutions in the first *Adh-dup* exon were very few (none, *D. lebanonensis-D. immigrans;* three, *D. immigrans-D. subobscura* and four, *D. lebanonensis-D. subobscura*) is particularly striking and may reflect differential selection pressure (Marfany and Gonzàlez-Duarte, 1991).

Between 30% and 39% of all aa replacements of *Adh-dup* among *D. lebanonensis, D. immigrans* and *D. subobscura* was conservative. A high degree of similarity in the hydrophilicity profiles was also observed for ADH-DUP in the three species. The aa differences were concentrated mainly at the C-terminal end, where replacements and length variations accumulated. In contrast, experiments of proteolysis and chemical modification have shown that a segment close to the C terminus of ADH is required for enzyme activity and that any alteration in this region seriously impairs the function of the protein (Krook et al., 1992).

(e) Conclusions

(*1*) In the sequence comparisons the number of changes found among the three species under study, *D. immigrans, D. lebanonensis* and *D. subobscura*, were very similar in all cases. This feature is in agreement with the phylogenetic trees obtained from the distance matrix values (Table II), which shows that the three subgenera, *Scaptodrosophila, Drosophila* and *Sophophora* (*D. lebanonensis, D. immigrans* and *D. subobscura*, respectively) diverged nearly at the same time in the evolutionary history of the genus.

TABLE II

Interspecies divergence matrices

Species[a]	Divergence for *Adh* exons[b] (Ka\Ks)									
	MAU	ORE	MEL	PSE	AFF	MUL2	MUL1	LEB	IMM	SUB
MAU		0.185	0.075	0.642	1.220	0.892	0.838	0.862	1.075	0.738
ORE	0.018		0.173	0.533	1.049	0.902	0.822	0.761	0.943	0.662
MEL	0.009	0.020		0.616	1.187	0.896	0.863	0.820	1.029	0.809
PSE	0.056	0.059	0.052		1.110	0.798	0.834	0.876	0.911	0.376
AFF	0.131	0.139	0.135	0.122		0.916	0.947	1.362	0.903	1.339
MUL2	0.117	0.122	0.119	0.110	0.063		0.126	1.091	0.859	1.068
MUL1	0.104	0.108	0.109	0.100	0.053	0.018		1.032	0.903	1.052
LEB	0.110	0.113	0.108	0.100	0.120	0.094	0.094		1.051	1.073
IMM	0.115	0.121	0.123	0.114	0.089	0.089	0.077	0.117		1.100
SUB	0.052	0.054	0.047	0.022	0.133	0.114	0.105	0.113	0.117	

Species[a]	Divergence for *Adh-dup* exons[c] (Ka\Ks)					
	MAU	MEL	PSE	IMM	LEB	SUB
MAU		0.137	1.229	1.666	2.146	1.298
MEL	0.005		1.306	1.621	2.218	1.371
PSE	0.038	0.043		1.408	1.527	0.291
IMM	0.099	0.098	0.103		1.716	1.298
LEB	0.091	0.091	0.095	0.105		1.846
SUB	0.048	0.058	0.023	0.110	0.105	

[a] MAU, *D. mauritiana;* ORE, *D. orena*; MEL, *D. melanogaster*; PSE, *D. pseudoobscura;* AFF, *D. affinidisjuncta*; MUL1 and MUL2, *D. mulleri*; LEB, *D. lebanonensis*; IMM, *D. immigrans*; SUB, *D. subobscura.*

[b] The *Adh* exons have been used to calculate the divergence among several species of the *Drosophila* genus and *D. immigrans* and *D. lebanonenis.* Ks (top right) and Ka (bottom left) have been determined according to Li et al. (1985).

[c] The *Adh-dup* exons have been used to calculate the divergence among several species of the *Drosophila* genus and *D. immigrans* and *D. lebanonenis.* Ks (top right) and Ka (bottom left) have been determined according to Li et al. (1985).

(*2*) The *Adh* and *Adh-dup* exons have been used to calculate the divergence among several species of the *Drosophila* genus and *D. immigrans* and *D. lebanonensis.* The Ks and Ka values between each pair of species have been determined according to Li et al. (1985). Very similar results were obtained with the method of Perler and Efstratiadis (1980) (data not shown).

Then, phylogenetic trees based on the Ks and Ka matrices were constructed according to Fitch and Margoliash (1967), using the FITCH program of PHYLIP (Fig. 2). Species positions on the trees appeared to be similar with either Ks and Ka values except for *D. immigrans.* This species became closer to *D. affinidisjuncta* when Ks estimations were utilized for the tree construction. It has been reported that substitution rates in *D. affinidisjuncta* are higher than in most *Drosophila* species of the *Sophophora* and *Drosophila* subgenus (Sullivan et al., 1990), and this could well account for the tree

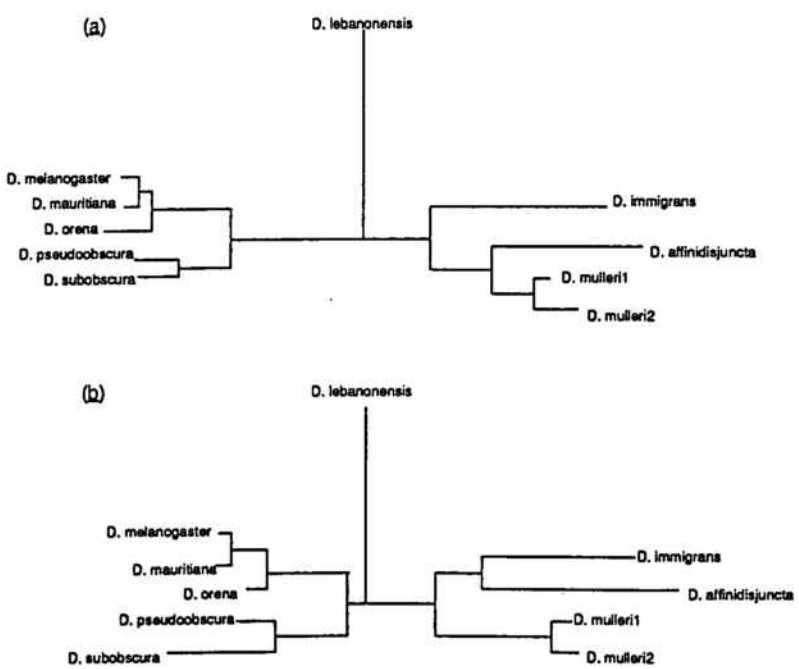


Fig. 2. Phylogenetic tree of some *Drosophila* species belonging to three different subgenera: *Scaptodrosophila (D. lebanonensis), Drosophila, (D. mulleri, D. affinidisjuncta* and *D. immigrans), Sophophora, (D. melanogaster, D. mauritiana, D. orena, D. pseudoobscura* and *D. subobscura)*: (a) tree constructed with Ka values matrix; (b) tree constructed with Ks values matrix. The branches of the trees are unconstrained, so substitution rates for species pairs may be variable.

differences. We also found that *D. lebanonensis* is closer to the *Sophophora* or to the *Drosophila* cluster depending on whether the tree is based on Ks or Ka values, respectively.

The estimation of evolutionary rates has been determined according to Kimura (1980), and the neighbour-joining method (Saitou and Nei, 1987) was used to reconstruct the phylogenetic tree. When this tree was compared to those based on the Ks and Ka matrices, a clear similarity was observed, particularly with the one obtained with Ka, in which *D. affinidisjuncta* appeared to be closer to *D. mulleri* than to *D. immigrans*.

The *Adh-dup* sequence has been described in a few species of the *Sophophora* subgenus. Moreover, up to now it has not been characterized in any species of the *Drosophila* subgenus. Evolutionary inferences drawn from the Ks and Ka matrices have a more limited scope than those concerning *Adh*. Information about other *Adh-dup* genes belonging to additional *Drosophila* subgenus is clearly needed to comprehend the evolution of this genomic region.

(*3*) When comparing *Adh* and *Adh-dup*, the number of substitutions was similar ($\chi^2=2.17$, d.f.$=2$, $P<0.05$), although *Adh-dup* showed greater silent/replacement rates than *Adh* ($\chi^2=6.8$, d.f.$=2$, $P<0.05$). So, there are differences in the selective constraints acting on these two adjacent genes.

(*4*) The divergence time for the *Drosophila* species analyzed has been calculated considering the substitution rate 5.5×10^{-9} substitutions/synonymous site year (Miyata et al., 1980, Hayashida and Miyata, 1983). This estimate, besides being widely used in mammals, has also been considered in some evolutionary analyses of *Drosophila Adh* (Atkinson et al., 1988). The values obtained for each pair of species differed slightly when considering *Adh* or *Adh-dup*: *D. immigrans-D. lebanonensis*, 53–68 Myr; *D. immigrans-D. subobscura*, 52–60 Myr and *D. lebanonensis-D. subobscura*, 53–69 Myr, respectively. These estimates are quite consistent with predictions established from biogeographical and immunological data (Throckmorton, 1975; Beverley and Wilson, 1982; 1984, respectively), whereas other estimates of substitution rates for *Drosophila* genes (Sharp and Li, 1989; Moriyama and Gojobori, 1992) give values 2–5 times higher, which would not be in accordance with those predictions. In that respect, these discrepancies among the divergence values should be viewed with caution.

ACKNOWLEDGEMENTS

We are grateful to M. Riutort, F. Mestres and A. Vilarroya for helpful advice on the use of computer programs to construct the phylogenetic tree. We are indebted to W. Li who kindly provided us with his computer program and to R. Rycroft for revising the English version. This research was supported by a grant from the Dirección General de Investigación Científica y Técnica, DGICYT PB88-0188. R.A. was in receipt of an FPI fellowship.

REFERENCES

Albalat, R., Gonzàlez-Duarte, R. and Atrian, S.: Protein engineering of alcohol dehydrogenase. FEBS Lett. 308 (1992) 235–239.

Atkinson, P.W., Mills, L.E., Starmer, W.T. and Sullivan, D.T.: Structure and evolution of the *Adh* genes of *Drosophila mojavensis*. Genetics 120 (1988) 713–723.

Beverley, S.M. and Wilson, A.C.: Molecular evolution in *Drosophila* and high diptera, I. Micro-clomplement fixtation studies of a larval hemolymph protein. J. Mol. Evol. 18 (1982) 251–264.

Beverley, S.M. and Wilson, A.C.: Molecular evolution in *Drosophila* and high diptera, II. A time scale for fly evolution. J. Mol. Evol. 21 (1984) 1–13.

Chambers, G.K.: The *Drosophila* alcohol dehydrogenase gene-enzyme system. Adv. Genet. 25 (1988) 39–107.

Fitch, W.M. and Margoliash, E.: Construction of phylogenetic trees. Science 155 (1967) 279–284.

Goldberg, D.A.: Isolation and partial characterization of the *Drosophila* alcohol dehydrogenase gene. Proc. Natl. Acad. Sci. USA 77 (1980) 5794–5798.

Hayashida, H. and Miyata, T.: Unusual evolutionary conservation and fragment DNA segment exchange in class I genes of the major histocompatibility complex. Proc. Natl. Acad. Sci. USA 80 (1983) 2671–2675.

Jowett, T.: Preparation of nucleic acids. In: Roberts, D.B. (Ed.), *Drosophila*. A Practical Approach. IRL Press, Oxford, UK, 1986, pp. 275–286.

Kawasaki, E.S.: Amplification of RNA. In: Innis, M.A., Gelfand, D.H., Sninsky, J.J. and White, T.J. (Eds.), PCR Protocols. Academic Press, San Diego, 1990, pp. 21–27.

Kimura, M.: A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16 (1980) 111–120.

Krook, M., Prozorovski, V., Atrian, S., Gonzàlez-Duarte, R. and Jörnvall, H.: Short-chain dehydrogenases: proteolysis and chemical modification of prokaryotic 3α/20β-hydroxysteroid, insect alcohol, and human 15-hydroxyprostaglandin dehydrogenases. Eur. J. Biochem. 209 (1992) 233–239.

Li, W.H., Wu, C.I. and Luo, C.C.: A new method for estimating synonymous and nonsynonymous rates of nucleotide substitutions considering the relative likehood of nucleotide and codon changes. Mol. Biol. Evol. 2 (1985) 150–174.

Marfany, G. and Gonzàlez-Duarte, R.: The *Adh* genomic region of *Drosophila ambigua*: evolutionary trends in different species. J. Mol. Evol. 32 (1991) 454–462.

Marfany, G. and Gonzàlez-Duarte, R.: The *Drosophila subobscura Adh* genomic region contains valuable evolutionary markers. Mol. Biol. Evol. 9 (1992a) 261–277.

Marfany, G. and Gonzàlez-Duarte, R.: Evidence for retrotranscription of protein-coding genes in *Drosophila subobscura* genome. J. Mol. Evol. 35 (1992b) 492–501.

Miyata, T., Yasunaga, T. and Mishida, T.: Nucleotide sequence divergence and functional constraint in mRNA evolution. Proc. Natl. Acad. Sci. USA 77 (1980) 7328–7332.

Moriyama, E.N. and Gojobori, T.: Rates of synonymous substitution and base composition of nuclear genes in *Drosophila*. Genetics 130 (1992) 855–864.

Perler, F. and Efstratiadis, A.: The evolution of genes: the chicken preproinsulin gene. Cell 20 (1980) 555–566.

Saitou, N. and Nei, M.: The neighbour-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4 (1987) 406–425.

Schaeffer, S.W. and Aquadro, C.F.: Nucleotide sequence of the *Adh* gene region of *Drosophila pseudoobscura*: evolutionary changes and the evidence for an ancient gene duplication. Genetics 117 (1987) 61–73.

Sharp, P.M. and Li, W.H.: On the rate of DNA sequence evolution in *Drosophila*. J. Mol. Evol. 28 (1989) 398–402.

Sullivan, D.T., Atkinson, P.W. and Starmer, W.T.: Molecular evolution of the alcohol dehydrogenase genes in the genus *Drosophila*. In: Hecht, M.K., Wallace, B. and Macintyre, R.J. (Eds.), Evolutionary Biology, Vol. 24. Plenum, New York, 1990, pp. 107–147.

Throckmorton, L.: The phylogeny, ecology, and geography of *Drosophila*. In: King, R.C. (Ed.), Handbooks of Genetics, Vol. 3. Plenum, New York, 1975, pp. 421–469.

Vilageliu, L. and Gonzàlez-Duarte, R.: Alcohol dehydrogenase from *Drosophila funebris* and *Drosophila immigrans*: molecular and evolutionary aspects. Biochem. Genet. 22 (1984) 797–815.

Visa, N., Marfany, G., Vilageliu, L., Albalat, R., Atrian, S. and Gonzàlez-Duarte, R.: The *Adh* in *Drosophila*: chromosomal location and the restriction analysis in species with different phylogenetic relationships. Chromosoma 100 (1991) 315–322.

Primary structure analysis of *Drosophila* alcohol dehydrogenase.

Biochem. Genet. (Life Sci. Adv.) 1992, 11: 19 – 29

Primary structure analysis of *Drosophila* alcohol dehydrogenase

S. Atrian*, G. Marfany, R. Albalat, and R. Gonzàlez-Duarte

Departament de Genètica Facultat de Biologia, Universitat de Barcelona Av. Diagonal 645. 08028, Barcelona, Spain.

SUMMARY

Drosophila alcohol dehydrogenase is a member of the short chain dehydrogenase family and, therefore, it is a non-metallo enzyme, active as a dimer of two identical subunits of 253 to 255 aminoacids. The comparison of 36 primary sequences for this enzyme, belonging to 35 different *Drosophila* species, reveals a low degree of identity (57.13%), taking into account that ADH plays a selective role in alcohol assimilation and detoxification. This value is significantly low if compared with the 50% identity that all long chain dehydrogenases, from plants to mammals, share. Alignments of the primary structures reveal that fully conserved residues are not dispersed at random, but form seven conserved boxes with a degree of identity equal to or greater than 75%, that can be identified as putative essential regions for the function and/or structure of the enzyme. Among strickly conserved residues*Trp, Gly, Asp, Asn* and *Phe* stand out, fully coincident with those described for the complete short chain dehydrogenase family. In spite of the quick rate of amino acid replacement, variable positions show a high level of conservative replacements. Thus, *Drosophila* ADH's have a high degree of sequence homology which is in total agreement with its essential metabolic funtion.

INTRODUCTION

Drosophila alcohol dehydrogenase (ADH, alcohol, NAD^+ oxidoreductase, E.C. 1. 1.1.1) is one of the gene/enzyme systems for which the largest amount of genetical, biochemical and structural information has been gathered (cf. reviews: van Delden, 1982; Sofer and Martin, 1987; Chambers, 1988; Chambers, 1991). However, it still constitutes the focus of relevant interdisciplinary studies from crystallographic, structural, biochemical, functional, genetical and evolutionary approaches. *Drosophila* alcohol dehydrogenase catalyzes the oxidation of primary and secondary alcohols to the corresponding aldehydes and ketones, coupled to the reduction of NAD^+ as coenzyme. The enzyme may constitute up to 1% of the total protein content of flies, and its role in the detoxification and metabolic assimilation of alcohols found in the feedingniches exploited by *Drosophila* is generally accepted.

Drosophila ADH is a member of the short chain dehydrogenase family (Jörnvall et al., 1981), which is active as a dimer composed of identical subunits of 255, 254 or 253 amino acids, depending on the species considered (Villaroya and Juan, 1991). ADH polypeptides are encoded by a single gene controlled by two alternative promoters in all the species so far analyzed, except those belonging to the *repleta* group, where a duplication created two functional genes (Sullivan et al., 1990a).

Although many genomic sequences and primary structures are known and some secondary structures have been predicted for this gene and its product, no tertiary structure has yet been obtained, and only indirect evidence is available for the location and role assignment of specific amino acids in active domains (Chen et al., 1990; Chen et al., 1991; Prozorovski et al., 1992; Krook et al., 1992; Albalat et al., 1992). The tridimensional structure is known for only one short chain dehydrogenase, 20ß-hydroxysteroid dehydrogenase (Ghosh et al., 1991), but 3-D conformation for the *Drosophila* enzyme can be quite different, as both enzymes share one of the lowest degree of similarity among the family members (Krozowski, 1992). In these circumstances, the analysis of all determined *Drosophila* ADH sequences becomes an invaluable tool for two main purposes: first, the identification

*Author to whom correspondence should be addressed.

of fully conserved regions and residues, which can then be postulated as functionally significant, and second, the provision of clues for site-directed mutagenesis studies.

We now present the analysis of 36 *Drosophila* ADH sequences. Six of them were obtained in our laboratory and, of these, sequences for *D.immigrans*, *D.madeirensis* and *D.guanche* are published for the first time. The large number of sequences analyzed and the heterogeneity of the taxonomical groups to which they belong provides a broader and more informative picture than previous studies (Sullivan et al., 1990b; Villarroya and Juan, 1991; Winberg and McKinley-McKee, 1992) and complements those carried out for the whole short chain dehydrogenase family (Persson et al. 1991; Krozowski, 1992).

Subgenus	Group	Species	Sw #	GenEMBL #
Drosophila	*Hawaiian* picture-wing	*D.heteroneura*	P21898	M60683/M63287
		D.silvestris	P23278	M60686/M63291
		D.differens	P22245	M60687/M63303
		D.planitibia	P23277	M60685/M63392
		D.affinidisjuncta	P21518	M37262
		D.picticornis	P23361	M60684/M63392
	repleta	*D.hydei-1*	P23236	X58694
		D.hydei-2	P23237	X58694
		D.mettleri	P22246	X57300
		D.arizonae	---	X62741
		D.mulleri-2	P07160	X03048
		D.navojoa	P12854	X15585
		D.buzzatii	---	X62743
		D.mayaguana	---	X62742
		D.wheeleri	P24267	X62851
		D.mojavensis-2	P09369	X03630
		D.mulleri-1	P07161	X03048
		D.mojavensis-1	P09370	M37276
	immigrans	*D.immigrans*	---	***
Sophophora	*obscura*	*D.persimilis*	---	M60997
		D.pseudoobscura	P07158	Y00602
		D.guanche	---	***
		D.madeirensis	---	***
		D.subobscura	---	M55545
		D.miranda	---	M60998
		D.ambigua	---	X54813
	melanogaster	*D.melanogaster-F*	P00334	M17827
		D.melanogaster-S	P00334	M17835
		D.sechellia	---	X04672
		D.simulans	P07163	M19263
		D.mauritiana	P07162	M19264
		D.teissieri	---	X54118
		D.yakuba	---	X54120
		D.orena	P07159	M33488
		D.erecta	---	X54116
Scaptodrosophila	*victoria*	*D.lebanonensis*	P10807	X54814

Table 1. *Drosophila* species used in this study. Species are listed according to their systematic placement in the genus *Drosophila* (subgenus and group) and the similarity of their ADH sequences. Accession numbers for the SwissProt Data Bank (protein sequence) and GenEmbl Data Bank (DNA sequence) are given. (---) indicates information not available in form of protein sequence. (***) indicates DNA sequences obtained in our laboratory which are neither published nor available through Data Banks.

MATERIAL AND METHODS

Drosophila ADH primary structures.

The ADH genomic regions of *D.lebanonensis* (Albalat and Gonzàlez-Duarte, 1990), *D.immigrans*, *D.ambigua*, (Marfany and Gonzàlez-Duarte, 1991) *D.subobscura*, (Marfany and Gonzàlez-Duarte, 1992), *D.madeirensis* and *D. guanche* were sequenced in our laboratory and the protein primary structure was deduced from the translation of the respective coding regions. Other protein and DNA sequences were retrieved from GenEmbl (Higgins et al., 1992) and SwisProt (Bairoch and Boeckmann, 1992) Data Banks, and their accession numbers are listed in Table 1. Files available from these code numbers contain all additional information concerning authors, general features of each sequence and reference data if already published.

Sequence comparisons and analysis.

Programs used are included in the GCG Software Package of the University of Wisconsin (v7.1) (Devereux et al., 1984), supported by a VAX-VMS computer running with OS v.5/4-3. Gene sequences where arranged and translated using ASSEMBLE, TRANSLATE and REFORMAT applications. Protein data, secondary structure predictions and amino acid composition were obtained with PEPTIDESORT, PEPTIDE-STRUCTURE, PLOTSTRUCTURE and PEPPLOT. Sequence alignments were performed with PILEUP, a program which uses the Needleman and Wunsch (1970) algorithm to score pairwise similarities, and were displayed with PRETTY and FIGURE. PLOTSIMILARITY was used to represent the average similarity among sequences in the multiple sequence alignments and DISTANCE to calculate the absolute number of residues conserved per each pair of ADH sequences.

RESULTS AND DISCUSSION

Comparison of protein sequences.

The 36 *Drosophila* ADH primary structures available to date were compared by progressive pairwise alignments, which yielded the general scheme shown in Table 2. Initial methionine is shown in all species although it has been claimed that only *D.lebanonensis* ADH keeps it in the mature form of the enzyme (Villaroya and Juan, 1991). As it is well known, an insertion of six nucleotides (TTTACC/T) at the beginning of the ADH coding region represents the addition of two extra amino acids in the corresponding polypeptide of the *Drosophila* species belonging to the *melanogaster* group. Thus, a two-position gap has been introduced in order to reflect the genetic information and to introduce no distortion in the alignment of the N-terminal region of the protein. Therefore, ADH monomer length may take three possible values: 255 for the member of the *melanogaster* group, 254 for *D.lebanonensis* and 253 for the remaining of species. If we do not consider the initial *Met*, nor third or fourth positions, 145 out of 253 residues are strictly conserved, which represents 57.13% of the ADH polypeptide. No significant differences are found if this value is split into the exons of the gene, yielding 50%, 58% and 58% respectively for the first, second and third exons. Assuming that there is a strong selective constraint for functional amino-acid replacements, we could deduce no evident correlation between functional domains of the protein and any particular exon division. Nevertheless, strictly conserved positions do not show a random dispersion along the sequence; instead, they form clusters whose location and significance will be discussed below.

Conserved boxes in *Drosophila* ADH's. Comparison to conserved domains in the SCDH family.

Conserved residues in the *Drosophila* ADH sequences are shown in the *identity* row of Table 2 and represented in Fig.1, together with a plot of the index of similarity of the 36 sequences, calculated from a modification of the Dayhoff values (Schwartz and Dayhoff, 1979) and the hydrophaty curve, according to the Kyte and Doolittle algorithm (1982).

Box 1 (aa11 to aa23, [FVAxLGGIGxTTS]) shows 84% internal identity and a net hydrophobic composition. It is widely accepted as a Rossmann fold nucleotide-binding domain, consisting of alternating alfa-helices and beta-sheets with highly conserved *Gly* and *Asp* residues. Secondary structure predictions and site-directed mutagenesis of these amino acids seem to confirm their predicted role. However, some authors claim that the *Drosophila*

```
                    1         10        20        30        40        50        60        70
D.heteroneura MV..IANSNI IFVAGLGGIG LDTSREIVKS GPKNLVVLDR IDNPAAIAEL KALNPKVTIT FYPYDVTVPL AETKKLLKTI
D.silvestris  MV..IANSNI IFVAGLGGIG LDTSREIVKS GPKNLVVLDR IDNPAAIAEL KALNPKVTIT FYPYDVTVPL AETKKLLKTI
D.differens   MV..IANSNI IFVAGLGGIG LDTSREIVKS GPKNLVVLDR VDNPAAIAEL KALNPKVTVT FYPYDVTVPV AETKKLLKTI
D.planitibia  MV..IANSNI IFVAGLGGIG LDTSREIVKS GPKNLVVLDR VDNPAAIAEL KALNPKVTVT FYPYDVTVPV AETKKLLKTI
D.affinidisjuncta MV..IANSNV IFVAGLGGIG LDTSREIVKS GPKNLVVLDR VDNPAAIAEL KALNPKVTVT FYPYDVTVPL AETKKLLKTI
D.picticornis MV..IANSNI IFVAGLGGIG LDTSREIVKS GPKNLVLLDR IDNPAAIAEL KALNPKVTVT FYPYDVTVPV DETKKLLKTI
D.hydei-1     MA..IANKNI IFVAGLGGIG LDTSREIVKS GPKNLVILDR IDNPAAIAEL KAINPKVTVT FYPYDVTVSV AESTKLLKVI
D.hydei-2     MA..IANKNI IFVAGLGGIG LDTSREIVKS GPKNLVILDR IDNPAAIAEL KAINPKVTIT FYPYDVTVSV AESTKLLKVI
D.mettleri    MA..IANKNI IFVAGLGGIG LDTSREIVKS GPKNLVILDR VENPTAIAEL KAINPKVTIT FYPYDVTVSV AESTKLLKTI
D.arizonae    MA..IANKNI IFVAGLGGIG FDTSREIVKS GPKNLVILDR IENPAAIAEL KALNPKVTVT FYPYDVTVSV AETTKLLKTI
D.mulleri-2   MV..IANKNI IFVAGLGGIG FDTSREIVKS GPKNLVILDR IENPAAIAEL KALNPKVTVT FYPYDVTVSV AETTKLLKTI
D.navojoa     MA..IANKNI IFVAGLGGIG FDTSREIVKS GPKNLVILDR IENPAAIAEL KALNPKVTVT FYPYDVTVPV AETTKLLKTI
D.buzzatii    MA..IANKNI IFVAGLGGIG FDTSREIVKS GPKNLVILDR IENPAAIAEL KALNPKVTVT FYPYDVTVPV AETTKLLKTI
D.mayaguana   MV..IANKNI IFVAGLGGIG FDTSREIVKS GPKNLVILDR IENPAAIAEL KALNPKVIVT FYPYDVTVPV AETTKLLKII
D.wheeleri    MV..IANKNI IFVAGLGGIG FDTSREIVKS GPKNLVILDR IENPAAIAEL KALNPKVTVT FYPYDVTVSV AETTKLLKTI
D.mojavensis-2 MA..IANKNI IFVAGLGGIG FDTSREIVKS GPKNLVILDR IENPAAIAEL KALNPKVTVT FYPYDVTVSV AETTKLLKTI
D.mulleri-1   MA..IANKNI IFVAGLGGIG FDTSREIVKS GPKNLVILDR IENPAAIAEL KALNPNVTVT FYPYDVTVPV AETTKLLQKI
D.mojavensis-1 MA..IANKNI IFVAGLGGIG FDTSREIVKK GPKNLVILDR IENPAAIAEL KALNPKVTVT FYLYDVTVSV AESTKLLQKI
D.immigrans   MA..IANKNI IFVAGLGGIG LDTSKGIVKA GPKNLVILDR IENPAAIAEL KAINPKVTVT FYPYDVTVPL AETKKLLTTI
D.persimilis  MS..LTNKNV VFVAGLGGIG LDTSRELVKR NLKNLVILDR IDNPAAIAEL KAINPKVTIT FYPYDVTVPV AETTKLLKTI
D.pseudoobscura MS..LTNKNV VFVAGLGGIG LDTSRELVKR NLKNLVILDR IDNPAAIAEL KAINPKVTIT FYPYDVTVPV AETTKLLKTI
D.guanche     MS..LTNKNV VFVAGLGGIG LDTSRELVKR DLKNLVILDR IDNPAAIAEL KAVNPKVTVT FYPYDVTVPV AETTKLLKTI
D.madeirensis MS..LTNKNV VFVAGLGGIG LDTSRELVKR DLKNLVILDR IDNPAAIAEL KAINPKVTVT FYPYDVTVPV AETTKLLKTI
D.subobscura  MS..LTNKNV VFVAGLGGIG LDTSRELVKR DLKNLVILDR IDNAAAIAEL KAINPKVTVT FYPYDVTVPV AETTKLLKTI
D.miranda     MS..LTNKNV VFVAGLGGIG LDTSRELVKR NLKNLVILDR IDNPAAIAEL KAINPKVTVT FYPYDVTVPI AETTKLLKTI
D.ambigua     MS..LTNKNV IFVAGLGGIG LDTSRELVKR DLKNLVILDR IDNPAAIAEL KAINPKVTVT FYPYDVTVPI SETTKLLKTI
D.melanogaster-F MSFTLTNKNV IFVAGLGGIG LDTSKELLKR DLKNLVILDR IENPAAIAEL KAINPKVTVT FYPYDVTVPI AETTKLLKTI
D.melanogaster-S MSFTLTNKNV IFVAGLGGIG LDTSKELLKR DLKNLVILDR IENPAAIAEL KAINPKVTVT FYPYDVTVPI AETTKLLKTI
D.sechellia   MAFTLTNKNV IFVAGLGGIG LDTSKELLKR DLKNLVILDR IENPAAIAEL KAINPKVTVT FYPYDVTVPI AETTKLLKTI
D.simulans    MAFTLTNKNV IFVAGLGGIG LDTSKELLKR DLKNLVILDR IENPAAIAEL KAINPKVTVT FYPYDVTVPI AETTKLLKTI
D.mauritiana  MAFTLTNKNV IFVAGLGGIG LDTSKELVKR DLKNLVILDR IENPAAIAEL QAINPKVTVT FYPYDVTVPI AETTKLLKTI
D.teissieri   MAFTLTNKNV VFVAGLGGIG LDTSKELVKR DLKNLVILDR IENPAAIAEL KAINPKVTVT FYPYDVTVPI AETTKLLKTI
D.yakuba      MAFTLTNKNV VFVAGLGGIG LDTSKELVKR DLKNLVILDR IENPAAIAEL KAINPKVTVT FYPYDVTVPI AETTKLLKTI
D.orena       MAFTLTNKNV IFVAGLGGIG LDTSKELVKR DLKNLVILDR IENPAAIAEL KAINPKVTVT FYPYDVTVPI AETTKLLKTI
D.erecta      MAFTLTNKNV IFVAGLGGIG LDTSKELVKR DLKNLVILDR IENPAAIAEL KAINPKVTVT FYPYDVTVPI AETSKLLKTI
D.lebanonensis MD..LTNKNV IFVAALGGIG LDTSRELVKR NLKNFVILDR VENPTALAEL KAINPKVNIT FHTYDVTVPV AESKKLLKKI

Identity ------N-N- -FVA-LGGIG -DTS----K- --KN-V-LDR --N--A-AEL -A-NP-V--T F--YDVTV-- -E--KLL--I
```

```
                    80        90        100       110       120       130       140       150
D.heteroneura FDKLKTVDLL INGAGILDDT QIERTIAVNF TGTVNTTTAI MDFWDKRKGG PGGVVANICS VTGFNSIYQV PVYSASKAAA
D.silvestris  FDKLKTVDLL INGAGILDDT QIERTIAVNF TGTVNTTTAI MDFWDKRNGG PGGVIANICS VTGFNSIYQV PVYSASKAAA
D.differens   FDELKTVDLL INGAGILDDN QIERTIAVNF TGTVNTTTAI MDFWDKRKGG PGGVIANICS VTGFNSIYQV PVYSASKAAA
D.planitibia  FDKLKTVDLL INGAGILDDN QIERTIAVNF TGTVNTTTAI MDFWDKRKGG PGGVIANICS VTGFNSIYQV PVYSASKAAA
D.affinidisjuncta FDKLKTVDLL INGAGILDDN QIERTIAVNF TGTVNTTTAI MDFWDKRKGG PGGVVANICS VTGFNSIYQV PVYSASKAAA
D.picticornis FDKLKTVDLL INGAGILDDH QIERTIAVNF AGTVNTTTAI LDFWDKRKGG PGGVVANICS VTGFNSIYQV PVYSASKAAA
D.hydei-1     FDKLKTVDLL INGAGILDDY QIERTIAVNF AGTVNTTTAI MAFWDKRKGG PGGVIANICS VTGFNAIYQV PVYSASKAAA
D.hydei-2     FDKLKTVDLL INGAGILDDY QIERTIAVNF AGTVNTTTAI MAFWDKRKGG PGGVIANICS VTGFNAIYQV PVYSASKAAA
D.mettleri    FDKLKTVDLL INGAGILDDY QIERTIAVNF AGTVNTTTAI MAFWDKRKGG PGGVIANICS VTGFNSIYQV PVYSASKAAA
D.arizonae    FDKLKTVDLL INGAGILDDY QIERTIAVNF TGTVNTTTAI MSFWDKRKGG PGGVIANICS VTGFNAIYQV PVYSASKAAA
D.mulleri-2   FDKLKTVDLL INGAGILDDY QIERTIAVNF TGTVNTTTAI MSFWDKRKGG PGGIIANICS VTGFNAIYQV PVYSASKAAA
D.navojoa     FDKLKTVDLL INGAGILDDY QIERTIAVNF TGTVNTTTAI MSFWDKRKGG PGGVIANICS VTGFNAIYQV PVYSASKAAA
D.buzzatii    FDKLKTVDLL INGAGILDDY QIERTIAVNF TGTVNTTTAI MAFWDKRKGG PGGVIANICS VTGFNAIYQV PVYSASKAAA
D.mayaguana   FDKLKTVDLL INGAGILDDY QIERTIAVNF TGTVNTTTAI MSFWDKRKGG PGGVIANICS VTGFNAIYQV PVYSASKAAA
D.wheeleri    FDKLKTVDIL INGAGILDDY QIERTIAVNF TGTVNTTTAI MSFWDKRKGG PGGIIANICS VTGFNAIYQV PVYSASKAAA
D.mojavensis-2 FDKLKTVDLL INGTGILDDH QIERTIAVNF TGTLNTTTAI MSFWDKRKGG PGGVIANICS VTGFNAILPV PVYSASKAAA
D.mulleri-1   FDQLKTVDLL INGAGILDDY QIERTIAVNF TGTVNTTTAI MSFWDKRKGG PGGVIANICS VTGFNAIYQV PVYSASKAAA
D.mojavensis-1 FDQLKTVDLL INGAGILDDH QIERTIAVNF TGTVNTITAI MSFWDKRKGG PGGVIANVCS VTGFNAIYQV PVYSASKAAA
D.immigrans   FAKLKTVDLL INGAGILNDH QIELTIAINF TGLVNTTTAI MDFWDKRKGG PGGAIANICS VTGFNAIYQV PVYSASKAAV
D.persimilis  FAQVKTIDVL INGAGILDDH QIERTIAVNY TGLVNTTTAI LDFWDKRKGG PGGIICNIGS VTGFNAIYQV PVYSGSKAAV
D.pseudobscura FAQVKTIDVL INGAGILDDH QIERTIAVNY TGLVNTTTAI LDFWDKRKGG PGGIICNIGS VTGFNAIYQV PVYSGSKAAV
D.guanche     FAQIKTIDVL INGAGILDDH QIERTIAVNY TGLVNTTTAI LDFWDKRKGG PGGIICNIGS VTGFNAIYQV PVYSGSKAAV
D.madeirensis FAQIKTIDVL INGAGILDDH QIERTIAVNY TGLVNTTTAI LDFWDKRKGG PGGIICNIGS VTGFNAIYQV PVYSGSKAAV
D.subobscura  FAQIKTIDVL INGAGILDDH QIERTIAVNY SGLVNTTTAI LDFWDKRKGG PGGIICNIGS VTGFNAIYQV PVYSGSKAAV
D.miranda     FAQLKTIDVL INGAGILDDH QIERTIAVNY TGLVNTTTAI LDFWDKRKGG PGGIICNIGS VTGFNAIYQV PVYSGTKAAV
D.ambigua     FAQLKTVDVL INGAGILDDY QIERTIAVNY TGLVNTTTAI LDFWDKRKGG PGGIICNIGS VTGFNAIYQV PVYSGSKAAV
D.melanogaster-F FAQLKTVDVL INGAGILDDH QIERTIAVNY TGLVNTTTAI LDFWDKRKGG PGGIICNIGS VTGFNAIYQV PVYSGTKAAV
D.melanogaster-S FAQLKTVDVL INGAGILDDH QIERTIAVNY TGLVNTTTAI LDFWDKRKGG PGGIICNIGS VTGFNAIYQV PVYSGTKAAV
D.sechellia   FAKLKTVDVL INGAGILDDH QIERTIAVNY TGLVNTTTAI LDFWDKRKGG PGGIICNIGS VTGFNAIYQV PVYSGTKAAV
D.simulans    FAKLKTVDVL INGAGILDDH QIERTIAVNY TGLVNTTTAI LDFWDKRKGG PGGIICNIGS VTGFNAIYQV PVYSGTKAAV
D.mauritiana  FAKLKTVDVL INGAGILDDH QIERTIAVNY TGLVNTTTAI LDFWDKRKGG PGGIICNIGS VTGFNAIYQV PVYSGTKAAV
D.teissieri   FAQLKTIDVL INGAGILDDH QIERTIAVNY TGLVNTTTAI LDFWDKRKGG PGGIICNIGS VTGFNAIYQV PVYSGTKAAV
D.yakuba      FAQLKTIDVL INGAGILDDH QIERTIAVNY TGLVNTTTAI LDFWDKRKGG PGGIICNIGS VTGFNAIYQV PVYSGTKAAV
D.orena       FAKLTTVDVL INGAGILDDY QIERTIAVNY TGLVNTTTAI LDFWDKRKGG PGGIICNIGS VTGFNAIYQV PVYSGTKAAV
D.erecta      FAKLTTVDVL INGAGILDDY QIERTIAVNY TGLVNTTTAI LDFWDKRKGG PGGIICNIGS VTGFNAIYQV PVYSGTKAAV
D.lebanonensis FDQLKTVDIL INGAGILDDH QIERTIAINF TGLVNTTTAI LDFWDKRKGG PGGIIANICS VTGFNAIHQV PVYSASKAAV

Identity P----T-D-L ING-GIL-D- QIE-TIA-N- -G--NT-TAI --FWDKR-GG PGG---N--S VTGFN-I--V PVYS--KAA-
```

Table 2.

```
                 160        170        180        190        200        210        220        230
D.heteroneura    LSFTTSLAKL AHITGVTVYS INPGITKTVL VHKFNSWLNV EPRVAELLLE HPTQTTLQCA QNFVKAIEAN QNGAIWKLDL
D.silvestris     LSFTTSLAKL AHITGVTVYS INPGITKTVL VHKFNSWLNV EPRVAELLLE HPTQTTLQCA QNFVKAIEAN QNGAIWKLDL
D.differens      LSFTTSLAKL AHITGVTAYS INPGITKTVL VHKFNSWLSV EPRVAELLLE HPTQTTLQCA QNFVKAIEAN QNGAIWKLDL
D.planitibia     LSFTTSLAKL AHITGVTAYS INPGITKTVL VHKFNSWLSV EPRVAELLLE HPTQTTLQCA QNFVKAIEAN QNGAIWKLDL
D.affinidisjuncta LSFTTSIAKL AHITGVTAYS INPGITKTVL VHKFNSWLSV EPRVAELLLE HPTQTTLQCA QNFVKAIEAN QNGAIWKLDL
D.picticornis    LSFTMSIAKL AHITGVTAYS INPGITKTIL VHKFNSWLNV EPRVAELLLE HPTQTTLQCA QNFVKAIEAN QNGAIWKLDL
D.hydei-1        LSFTNSLAKL APITGVTAYS INPGITKTTL VHKFNSWLDV EPRVAELLLE HPTQTSLQCA QNFVKAIEAN QNGAIWKLDL
D.hydei-2        LSFTNSLAKL APITGVTAYS INPGITKTTL VHKFNSWLDV EPRVAELLLE HPTQTTLQCA QNFVKAIEAN KNGAIWKLDL
D.mettleri       ISFTNSLAKL APITGVTAYS INPGITKTPL VHKFNSWLDV EPRVAELLLE HPTQTTLQCA QNFVKAIEAN KNGAIWKLDL
D.arizonae       LSFTNSLARL APITGVTAYS INPGITKTTL VHKFNSWLDV EPRVAELLLE HPTQTTLQCA QNFVKAIEAN QNGAIWKLDL
D.mulleri-2      LSFTNSLAKL APITGVTAYS INPGITKTTL VHKFNSWLDV EPRVAELLLE HPTQTTLQCA QNFVKAIEAN QNGAIWKLDL
D.navojoa        LSFTNSLAKL APITGVTAYS INPGITKTTL VHKFNSWLDV EPRVAELLLE HPTQTSLECA QNFVKAIEAN QNGAIWKLDL
D.buzzatii       LSFTNSLAKL APITGVTAYS INPGITKTTL VHKFNSWLDV EPRVAELLLE HPTQTTLQCA QNFVKAIEAN QNGAIWKLDL
D.mayaguana      LSFTNSLAKL APITGVTAYS INPGITKTTL VHKFNSWLDV EPRVAELLLE HPTQTTLQCA QNFVKAIEAN QNGAIWKLDL
D.wheeleri       LSFTNSLAKL APITGVTAYS INPGITVTPL VHKFNSWLDV EPRVAELLLE HPTQTRLQCA QNFVKAIEAN QNGAIWKLDL
D.mojavensis-2   LSFTNSLARL APITGVTAYS INPGITKTTL VHKFNSWLDV EPRVAELLLE HPTQTTLQCA QNFVKAIQAN QNGAIWKLDL
D.mulleri-1      LSSTNSLAKL APITGVTAYS INPGITKTPL VHKFNSWLDV EPRVAELLLE HPTQTSLQCA QNFVKAIEAN QNGAIWKLDL
D.mojavensis-1   LSFTNSLAKL APITGVTAYS INPGITKTTL VHKFNSWLDV EPRVGELLLE HPTQTSLECA QNFVKAIEAN QNGAIWKLDL
D.immigrans      VSFTSSLAKL APITGVIAYS INPGITKTTL VHKFNSWLDV EPRVAELLDE HPTQTTTQCS QNFVKAIEAN QNGAIWKLDL
D.persimilis     VNFTSSLAKL APITGVTAYT VNPGITRTTL VHKFNSWLDV EPRVAEKLLE HPTQTSQQCA ENFVKAIELN KNGAIWKLDL
D.pseudobscura   VNFTSSLAKL APITGVTAYT VNPGITKTTL VHKFNSWLDV EPRVAEKLLE HPTQTSQQCA ENFVKAIELN KNGAIWKLDL
D.guanche        VNFTSSLAKL APITGVTAYT VNPGITKTTL VHKFNSWLDV EPRVAEKLLE HPTQTSQQCA ENFVKAIELN KNGAIWKLDL
D.madeirensis    VNFTSSLAKL APITGVTAYT VNPGITKTTL VHKFNSWLDV EPRVAEKLLE HPTQTSQQCA ENFVKAIELN KNGAIWKLDL
D.subobscura     VNFTSSLAKL APITGVTAYT VNPGITKTTL VHKFNSWLDV EPRVAEKLLE HPTQTSQQCA ENFVKAIELN KNGAIWKLDL
D.miranda        VNFTSSLAKL APITGVTAYT VNPGITRTTL VHKFNSWLDV EPLVAEKLLE HPTQTSQQCA ENFVKAIELN KNGAIWKLDL
D.ambigua        VNFTSSLAKL APITGVTAYT VNPGITKTTL VHKFNSWLDV EPNVAEKLLE HPTQTSQQCA ENFVKAIELN KNGAIWNLDL
D.melanogaster-F VNFTSSLAKL APITGVTAYT VNPGITRTTL VHTFNSWLDV EPQVAEKLLA HPTQPSLACA ENFVKAIELN QNGAIWKLDL
D.melanogaster-S VNFTSSLAKL APITGVTAYT VNPGITRTTL VHKFNSWLDV EPQVAEKLLA HPTQPSLACA ENFVKAIELN QNGAIWKLDL
D.sechellia      VNFTSSLAKL APITGVTAYT VNPGITRTTL VHKFNSWLDV EPQVAEKLLA HPTQPSLACA ENFVKAIELN QNGAIWKLDL
D.simulans       VNFTSSLAKL APITGVTAYT VNPGITRTTL VHKFNSWLDV EPQVAEKLLA HPTQPSLACA ENFVKAIELN QNGAIWKLDL
D.mauritiana     VNFTSSLAKL APITGVTAYT VNPGITRTTL VHKFNSWLDV EPQVAEKLLA HPTQPSLACA ENFVKAIELN QNGAIWKLDL
D.teissieri      VNFTSSLAKL APITGVTAYT VNPGITRTTL VHKFNSWLDV EPQVAEKLLA HPTQPSLACA ENFVKAIELN QNGAIWKLDL
D.yakuba         VNFTSSLAKL APITGVTAYT VNPGITRTTL VHKFNSWLDV EPQVAEKLLA QPNQPSLACA QNFVKAIELN QNRAIWKLDL
D.orena          VNFTSSLAKL APITGVTAYT VNPGITRTTL VHKFNSWLDV EPQVAEKLLA HPIQSSLACA ENFVKAIELN ENGAIWKLDL
D.erecta         VNFTSSLAKL APITGVTAYT VNPGITRTTL VHKFNSWLDV EPQVAEKLLA HPTQTSLSCA ENFVKAIELN ENGAIWKLDL
D.lebanonensis   VSFTNSLAKL APITGVTAYS INPGITRTPL VHTFNSWLDV EPRVAELLLS HPTQTSEQCG QNFVKAIEAN KNGAIWKLDL

Identity         ---T-S-A-L A-ITGV--Y- -NPGIT-T-L VH-FNSWL-V EP-V-E-L-- -P-Q----C- -NFVKAI--N -N-AIW-LDL
```

```
                 240        250
D.heteroneura    GRLDAIEWTK HWDSGI
D.silvestris     GRLDAIEWTK HWDSGI
D.differens      GRLDAIEWTK HWDSGI
D.planitibia     GRLDAIEWTK HWDSGI
D.affinidisjuncta GRLDAIEWTK HWDSGI
D.picticornis    GRLDAIEWTK HWDSGI
D.hydei-1        GRIEAIEWTK HWDSGI
D.hydei-2        GRLDAIEWTK HWDSGI
D.mettleri       GRLDAIEWTQ HWDSHI
D.arizonae       GTLEAIEWTK HWDSHI
D.mulleri-2      GTLEAIEWTK HWDSHI
D.navojoa        GTLEAIEWTK HWDSHI
D.buzzatii       GRLEAIEWTK HWDSGI
D.mayaguana      GRLEAIEWTK HWDSHI
D.wheeleri       GTLEAIEWTK HWDSHI
D.mojavensis-2   GTLEAIEWTK HWDSHI
D.mulleri-1      GTLEAIEWTK HWDSHI
D.mojavensis-1   GTLEAIEWTK HWDSHI
D.immigrans      GRLDAVNWTK HWDSGI
D.persimilis     GTLEPITWTQ HWDSGI
D.pseudoobscura  GTLEPITWTQ HWDSGI
D.guanche        GTLEPITWTK HWDSGI
D.madeirensis    GTLESITWTK HWDSGI
D.subobscura     GTLESITWTK HWDSGI
D.miranda        GTLEPITWTQ HWDSGI
D.ambigua        GTLEPITWTQ HWDSGI
D.melanogaster-F GTLEAIQWTK HWDSGI
D.melanogaster-S GTLEAIQWTK HWDSGI
D.sechellia      GTLEAIQWTK HWDSGI
D.simulans       GTLEAIQWTK HWDSGI
D.mauritiana     STLEAIQWTK HWDSGI
D.teissieri      GTLEAIQWSK HWDSGI
D.yakuba         GTLEAIQWSK HWDSGI
D.orena          GTLEAIKWSK HWDSGI
D.erecta         GTLEAIQWSK HWDSGI
D.lebanonensis   GTLEAIEWTK HWDSHI

Identity         --L----W-- HWDS-I
```

Table 2. The amino acid sequence of 36 *Drosophila* alcohol dehydrogenases, They were aligned using PILEUP (GapWeight 3.0, GapLengthWeight 0.1) and displayed with PRETTY, so that they are sort in clusters of similarity. Initial methionine is shown although it is only present in the mature enzyme of *D.lebanonensis*, thus residue # 1 corresponds to the position immediatelly after this amino acid. A two residues gap has been introduced in the third and fourth positions to compensate the insertion presented by the melanogaster group of species. Below each block of sequences, the identity row shows those positions stricly conserved in all species, and residues conserved in the short chain DH family are boxed.

ADH-NAD$^+$ binding pocket motif [GLGGIG] matches the fingerprint of the cofactor site for long chain DH's better than the short chain DH family motif, [TxxxGxG], suggesting a possible hybrid origin for the insect enzyme (Krozowski, 1992). It is worth noting that one species, *D.lebanonensis*, actually lacks the first *Gly* residues of the hexamer, replaced by an *Ala* residue, which reduces the number of fully conserved *Gly* in Drosophila ADH's to only two (G17 and G19). Position 20, the only non-conserved in the box, shows only two possibilities, either a *Leu* or *Phe* residue.

Box 2, (aa87 to aa109, [DxLINGxGILxDxQIExTIAxN]), 75% of internal identity, includes, in addition, highly conservative substitutions in the non-identical positions (see Table 3). It first half shows a marked hydrophobicity whereas its second half shows a marked hydrophilicity. Thus, this segment is likely to have some predominant role in the maintenance of the correct shape of essential regions of the molecule. *Asp*87 and *Gly*94 are conserved along all the short chain deshidrogenase family.

Box 3, (aa117 to aa132, [TAIxxFWDKRxGGPGG], 81% of identity, is flanked by two of the short chain dehydrogenase family conserved residues, *Asn*114 and *Gly*132. This clearly hydrophilic segment would constitute the longest stretch of fully conserved amino acids, eleven in all, if not for one species, *D.silvestris* with an *Asn* instead of a *Lys* in position 127. Therefore, its relevance as a functional or

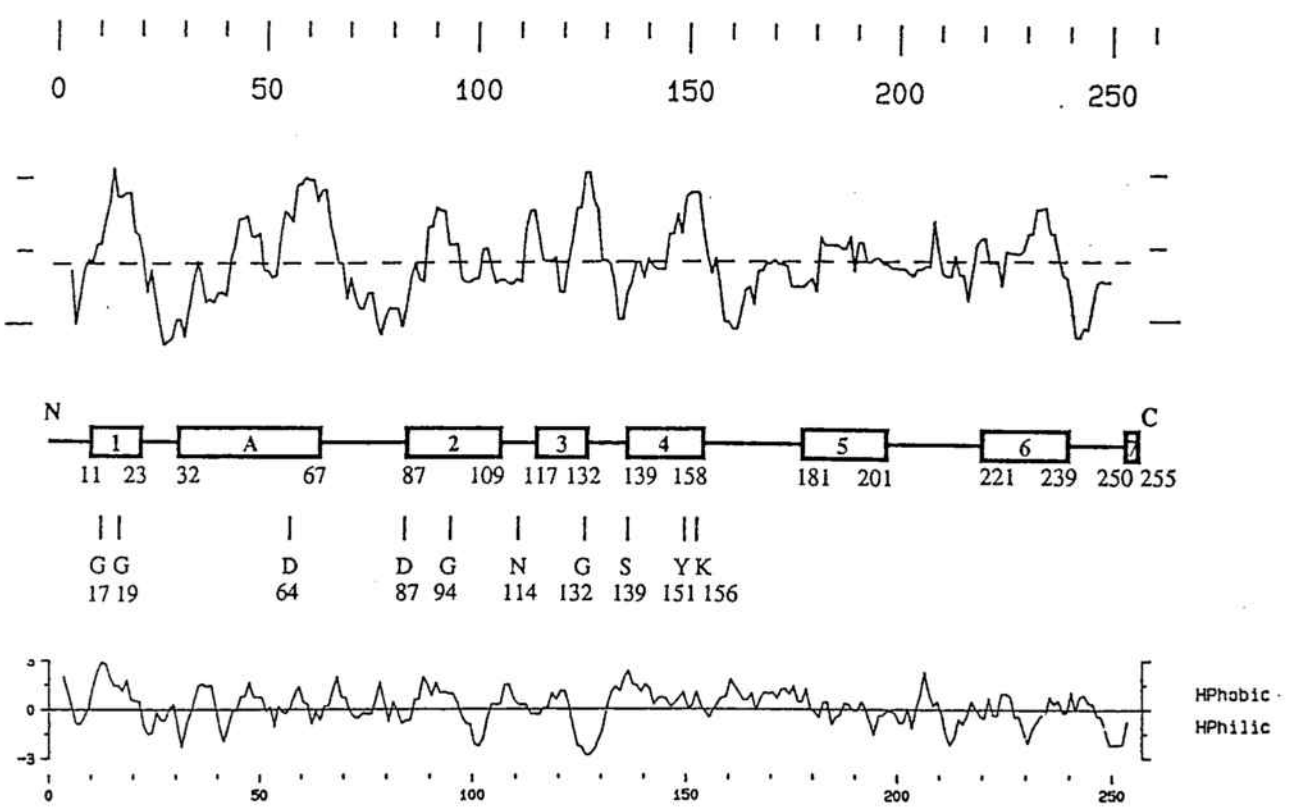


Fig. 1. Plots of similarity and average hydrophaty of *Drosophila* ADH's. Profile of similarity was calculated and represented with PLOTSIMILARITY/IDENTITY(window=8) which uses the Dayhoff table of aminoacid resemblance indices, rescaled by dividing each value by the sum of its row and column, and normalized to a mean of 0, and a standard deviation of 1.0. Perfect matches are set to 1.5. In the *identity* option, the program plots a measure of the level of identity between all sequences aligned; the identical comparisons are given a value of 1.0; others are given a value of 0.0. Boxes of conserved residues are shown below the similarity profile, together with amino acids invariable in all short chain dehydrogenases. PEPPLOT was used to calculate the average of a residues specific hydrophobicity index over a window of 6 residues according to Kyte and Doolittle (1982) hydrophaty algorithm.

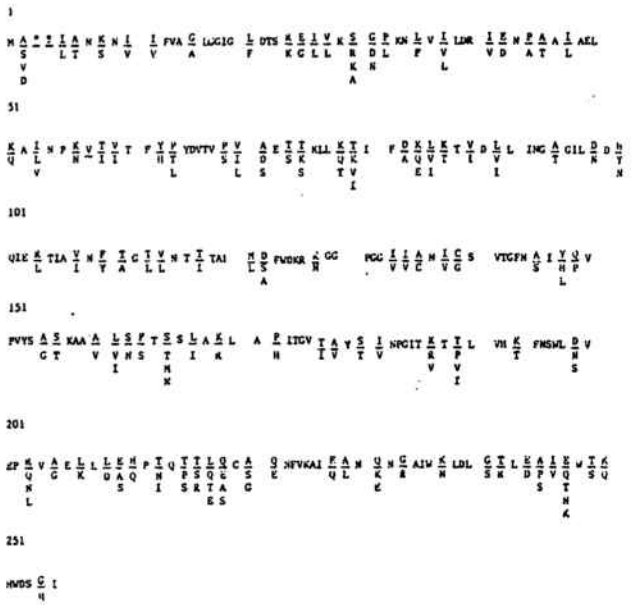


Table 3. Consensus sequence of Drosophila ADH. All amino acids found in each particular position are shown. Those above the hyphen are the most frequent, so that they form a "consensus" sequence along with the invariable residues.

structural component should be beyond any doubt.

Box 4, (aa139 to aa158, [SVTGFNxIxx VPYSxxKAA], 75% identity, comprises three short chain dehydrogenase conserved residues, of which *Tyr*152 and *Lys*156 seem to be involved in the catalytic centre of the enzyme: both chemical modification and SDM data (Krook et al., 1992; Albalat et al.,1992) confirmed *Tyr*152 as essential for the ADH enzymatic activity. The hydrophobic nature of this region is in fully agreement with these results.

Box 5, (aa181 to aa201, [NPGITxTxLVHx FNSWLxVEP], 77% identity and **Box 6**, aa221 to aa239, [NFVKAIxxNxNxAIWxLDL], 74% identity, are two hydrophobic regions separated by an hypervariable segment, as clearly shown in Table 3.

Finally, **Box 7**, (aa250 to aa255, [HWDSxI]), 83% identity, the C-terminus of the enzyme, displays a non negligible degree of conservation, which would indicate some essential contribution to the enzyme architecture and/or function. This is in agreement with the fact that controlled chemical cleavage of the C-term tail of the enzyme results in a loss of enzymatic activity (Krook et al., 1992).

Apart from this highly conserved*boxes*, it is worth noting the residues conserved in **Box A** (aa32 to aa 67), 61% identity, including *Asp*64, invariable in short chain dehydrogenases, and the subsets *LysLeuLeu* (74-76) and *IleThrGlyVal* (172-175). The catalytic role previously attributed to cysteines has now been ruled out following chemical modification (Krook et al., 1992) and site directed mutagenesis results (Chen et al, 1990). In fact, sequence analysis reveals that the first *Cys* has a non-fixed position in the ADH molecule (135 or 138) and the second one, *Cys*218, is included in the hypervariable C-terminal region, this heterogeneity being hardly consistent with any functional constraint.

Amino acid substitutions.

Table 3 shows the amino acid replacements found in the non-conserved positions of the polypeptide sequence. Residues above the hyphen are those present in the majority of species or, in the case of similar percentages, those found in the *Drosophila* subgenus species, which is though the ancestor root of the genus. (Throckmorton, 1975).

In those positions where amino acid substitution leads to more than one residue, all possibilities are indicated in a column, in decreasing order of frequency. As stated before, 145 out of 255 positions of the polypeptide are conserved. Of the 110 variable positions, 77 show replacement by one alternative amino acid (this meaning two possibilities for that position), 24 show replacement by two alternative amino acids (thus, 3 possibilities), 8 show replacement by three alternative amino acids (thus, 4 possibilities) and in one case we found four changes, thus 5 possibilities for that position. These data correlate well with a Poisson distribution (λ=0.592),

which indicates that amino acid substitutions can be considered as rare events, with a constant chance of occurring over time and space.

Relative substitution rates are indicated in Table 4. The average content of each amino acid was calculated from the composition of the 36 sequences. For a random position, there is 50% of probability of being conserved, if we assume no functional constraints. So, we would predict higher values of relative conservation for those residues playing important roles in the protein architecture and/or function. Five amino acids are conserved above the average rate: *Trp* (100%), *Asn* and *Asp* (92% + 66%), *Phe* (77%) and *Gly* (72%), data in full agreement with those found in the analysis of whole short chain dehydrogenase family (Persson et al., 1991).

Table 5 shows the kind of substitutions accounting for the 42.69% variable positions along the polypeptide chain (cf. Table 3).It is evident that most of the substitutions are fully conservative, taking as a quantitative reference the amino acid "resemblance" values from Feng et al. (1975), which include structural features and genetic interconvertability weightings. Of the 190 theoretically possible substitutions (n(n-1)/2, n=20), only 64 are found, accounting for 151 events of amino acid change detected when considering the 255 positions of all 36 sequences. It is worth noting 14 V/I, 9 L/I and 6 V/L replacements in the hydrophobic-relatively small group of amino acids, 9 S/T among the neutral-small amino acids and 5 Q/E among the hydrophilic-relatively small ones. Other common replacements found are T/I (6 times), Q/K (5) and G/A, S/A and T/A (4 times each). Four of the replacements found only once correspond to pairs of amino acids with a low score of similarity (1 in the scale of 6): G/H, S/M, L/E and L/D. The G/H substitution occurs in the conserved C-terminal end (position 254) in ten of the species analyzed , and it is surprising that the only change in this conserved block is this slightly conservative one. S/M change occurs in position 164 of *D.picticornis*, L/E in position 217 of *D.lebanonesis* and L/D in 208 of *D.immigrans*. Conservation is clearly related with the evolutionary functional constraint operating in the molecule, and radical replacements involving amino acids with very different structural and functional properties must either have taken place in non-essential regions from a functional or structural point of view, or have been "compensated" by another substitution in a neighbouring position in the third dimensional structure.

	c	x	%c/x	
Ala	12	23.8	50%	
Cys	1	2.0	50%	
Asp	8	12.1	66%	*
Glu	5	10.1	50%	
Phe	7	9.0	77%	*
Gly	13	18.1	72%	*
His	2	4.0	50%	
Ile	13	18.1	52%	
Lys	6	17.9	33%	
Leu	14	24.4	57%	
Met	---	----	----	
Asn	15	16.3	92%	*
Pro	6	10.5	57%	
Gln	2	6.1	32%	
Arg	3	5.6	53%	
Ser	5	11.0	45%	
Thr	12	25.1	47%	
Val	13	22.0	59%	
Trp	5	5.0	100%	*
Tyr	3	5.7	52%	

Table 4. Conserved residues in *Drosophila* ADH's. The absolute content of residues conserved is shown in column "c" for each amino acid. "x" represents the average content for each amino acid and "%c/x" is the percentage of each amino acid which remains fully conserved in *Drosophila* ADH's.

Phylogenetic trees.

To visualize the similarity relationships among the species compared we used the UPGMA method of Needlman and Wunsch, as described in the Material & Methods section. Establishment and analysis of phylogenetic relationships are beyond the scope of this paper. Nevertheless, if we assume that a similar functional selection has operated on ADH no matter of which species, we can consider the degree of similarity as a measure of the degree of phylogenetic distance in the genus. The unrooted trees obtained from protein and DNA (coding region) comparisons are shown in Fig.2A and 2B, respectively. Both trees largely reflect the standard

Ocurrence of the of the replacement	Replacement
14	V/I ^{5}
9	L/I ^{5} S/T ^{5}
6	L/V ^{5} T/I ^{3} Q/K ^{4}
5	Q/E ^{4} T/K ^{4} A/S ^{5}
4	A/G ^{5} A/T ^{5}
3	P/T ^{4} A/D ^{4} N/K ^{4} R/K ^{5} H/Y ^{3}
2	P/A ^{5} S/N ^{5} S/D ^{3} T/N ^{4} T/R ^{3} P/L ^{3} A/V ^{5} T/V ^{3} T/L ^{2} E/D ^{5} N/D ^{5} L/K ^{4} K/L ^{2} F/L ^{4} K/S ^{3}
1	C/A ^{2} C/G ^{3} S/P ^{4} S/G ^{5} Q/P ^{3} Q/A ^{3} Q/S ^{3} E/G ^{4} E/S ^{3} E/T ^{3} N/A ^{3} N/G ^{3} D/G ^{4} N/E ^{3} H/P ^{3} H/G ^{1} F/S ^{3} R/G ^{3} R/S ^{3} H/Q ^{4} H/N ^{4} R/Q ^{3} R/N ^{2} M/S ^{1} L/A ^{2} L/Q ^{2} L/E ^{1} L/D ^{1} V/K ^{3} L/K ^{2} M/L ^{5} Y/L ^{3} Y/F ^{5}

Table 5. Nature of the amino acid replacements found in *Drosophila* ADH, listed according their occurrence. Superindices show the degree of resemblance between each pair of amino acids, following Feng et al.'s (1985) value scale, where 6 denotes full identity and 0 the maximum dissimilarity. Pairs with an index of similarity equal to or lower than 3 are in bolt.

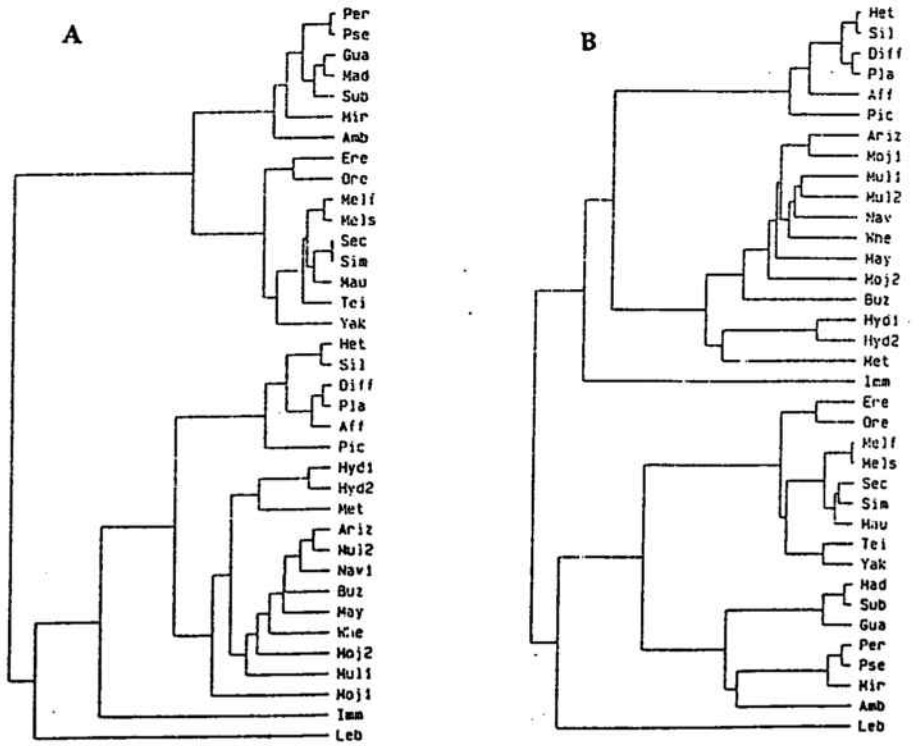


Fig 2. Dendrogram representing the clustering order obtained for the ADH protein sequence alignment (A) and the DNA -coding region-sequence alignment (B), calculated by the UPGMA method of Needlman and Wunsch. Names of species are abbreviated, see Table 1 for full names.

systematic distribution of species of *Drosophila* in subgena and groups, set up from morphological features (Throckmorton 1975), cf. Table 1). *D.lebanonensis* has always proved difficult to position (Villarroya and Juan, 1991). According to the ADH locus data, it is closer to the *Sophophora* rather than to *Drosophila* subgenus when considering the DNA tree, which fully agrees with recent considerations (Moriyama and Gojobori, 1992), but it remains associatted to the *Drosophila* subgenus in protein comparisons.

Although general phylogenetic clusters are accuratelly reflected, positions of the different species inside them show significant variations whether DNA or protein comparison is considered. From our data, it is clear that the DNA tree will provide a more faithful representation of the "classical taxonomical" classification than the protein tree. This would be affected by an important component of "convergent" evolution, perhaps forced by the environmental conditions of the ecological niche where each species develops.

SUPPLEMENTARY MATERIAL

One table with a 36x36 matrix showing the absolute number of conserved amino acids for each pair of ADH sequences is available under request to the first author.

ACKNOWLEDGEMENTS

We are grateful to R.Rycroft for revising the English version of this manuscript. This research was supported by a DGICYT (Ministerio Español de Educación y Ciencia) grant PB88-0188.

REFERENCES

Albalat,R. and Gonzàlez-Duarte, R. 1990. *Nucleic Acids Res. 18:* 6706

Albalat,R., Gonzàlez-Duarte, R. and Atrian, S. 1992. *FEBS Lett.*, (in press)

Bairoch, A. and Boeckmann, B. 1992. *Nucleic Acids Res. 20*: 2019-2022

Chambers, G.K. 1988. In: Caspari, E.W. and Scandalios, J. G. (Eds.) Advances in Genetics. Academic Press, New York, pp 39-107

Chambers, G.K. 1991. *Comp. Biochem. Physiol., 99B*: 723-730

Chen, Z.,Lu, L., Shirley,M., Lee,W.R. and Chang, S.H. 1990. *Biochemistry 29* : 1112-1118

Chen, Z., Lee,W.R. and Chang, S.H. 1991. *Eur. J. Biochem. 202* : 263-267

Devereux, J., Haeberli, P. and Smithies,O. 1984. *Nucleic Acids Res. 12:* 387-395

Feng, D.F., Johnson, M.S. and Doolittle, R.F. 1985. *J. Mol. Evol. 21*: 112-125

Ghosh, D., Weeks, C.M., Grochulski, P., Duax, W.L., Erman, M., Rimsay, R.L. and Orr, J.C. 1991. *Proc. Natl. Acad. Sci. USA 88*: 10064-10068

Higgins,D.G., Fuchs, R., Stoehr, P.J. and Cameron, G.N. 1992. *Nucleic Acids Res. 20*: 2071-2074

Jörnvall, H.,Persson, M. and Jeffery, J. 1981. *Proc. Natl. Acad. Sci. USA 78* : 4226-4230

Krook,M., Prozorovski,V., Atrian, S., Gonzàlez-Duarte,R. and Jörnvall,H. 1992. *Eur. J Biochem.* (in press)

Krozowski, Z. 1992. *Mol. Cell. Endocrinol.84* : C25-C31

Kyte, G. and Doolittle, R.F. 1982. *J. Mol. Biol. 157:* 105-132

Marfany, G., and Gonzàlez-Duarte, R. 1991. *J. Mol. Evol. 32* : 454-462

Marfany, G., and Gonzàlez-Duarte,R. 1992. *Mol Biol Evol 9* : 261-277

Needleman, S.B. and Wunsch, C.D. 1970. *J. Mol. Biol. 48:* 443-453

Persson, B., Krook,M. and Jörnvall,H. 1991. *Eur. J. Biochem. 200:* 537-543

Prozorovski, V. Krook, M., Atrian,S., Gonzàlez-
Duarte, R. and Jörnvall,H. 1992.*FEBS Lett.* 304: 46-50

Schwartz, R.M. and Dayhoff, M.O. 1979. In: Dayhoff, M.O. (Ed.) Atlas of Protein Sequence Structure, National Biomedical Research Foundation, Washington, pp 353-358

Sofer, W. and Martin, P.F. 1987. *Ann. Rev. Genet. 21*: 203-227

Sullivan, D.T., Atkinson,P.W., Bayer, C.A. and Menotti-Raymond, M.A. 1990a. In: Barker, J. S. F. (Ed.) Ecological and Evolutionary Genetics of Drosophila. Plenum Press, New York, pp 407-418

Sullivan, D.T., Atkinson, P.W. and Starmer,W.T. 1990b. In: Hecht,M.K., Wallace, B. and MacIntyre, R.J. (Eds.), Evolutionary Biology, 24, Plenum, New York, pp 107-147

Throckmorton, L. 1975. In: King, R.C. (Ed.), Handbook of Genetics, 3, Plenum, New York, pp 421-469

van Delden, W.V. 1982 . In: Kecht, M. K., Wallace, B. and Prance, G.T. (Eds.) Evolutionary Biology. Plenum Press, New York, pp 187-222

Villarroya, A. and Juan, E. 1991. *J. Mol. Evol. 32:* 421-428

Winberg, J.O. and McKinley-McKee, J.S. 1992. *Int. J. Biochem. 24*: 169-181

4.2. MUTAGÈNESI

Protein engineering of *Drosophila* alcohol dehydrogenase: The hydroxyl group of Tyr152 is involved in the active site of the enzyme.

Volume 308, number 3, 235–239 FEBS 11438 August 1992

Protein engineering of *Drosophila* alcohol dehydrogenase

The hydroxyl group of Tyr^{152} is involved in the active site of the enzyme

Ricard Albalat, Gonzàlez-Duarte and Sílvia Atrian

Department de Genètica, Facultat de Biologia, Universitat de Barcelona, Av. Diagonal 645, E 08028 Barcelona, Spain

Received 22 June 1992

Drosophila alcohol dehydrogenase is the most studied member of the family of short-chain alcohol dehydrogenases, although its tridimensional structure still remains unknown. We have engineered a *Drosophila* alcohol dehydrogenase in which tyrosine-152, an invariant residue in all members of the family, has been substituted by phenylalanine. The mutated gene has been expressed in yeast and pure mutant enzyme has been prepared by a one-step FPLC chromatographic procedure. *Drosophila* alcohol dehydrogenase-phenylalanine-152 shows no enzymatic activity. This result suggests not only that tyrosine-152 could constitute an essential building block of the active site but also that its hydroxyl group is directly involved in the redox reaction catalyzed by the enzyme.

Site-directed-mutagenesis; *Drosophila*; Alcohol dehydrogenase; Reactive residue

1. INTRODUCTION

Drosophila alcohol dehydrogenase (ADH, alcohol NAD^+ oxidoreductase; EC 1.1.1.1) is the member of the short-chain dehydrogenase family on which the largest amount of genetic, biochemical and structural information has been gathered [1]. Specific roles for several amino acids and some secondary structure predictions have been described for the *D. melanogaster* enzyme [2,3]. However, only indirect evidence is available concerning the residues involved in the active domains of ADH, as its structure is known only at the amino acid sequence level. Very recently, the first crystallographic analysis of a short chain dehydrogenase has been described [4], and the tridimensional prediction for *Drosophila* ADH is being approached through protein modelling. In the meantime, comparison of conserved residues among all short-chain dehydrogenases, and between these and the well-known long-chain ADHs, have provided valuable clues about protein sites involved in the catalytic domains [5]. It has been claimed that the secondary structure of the N-terminal region is organized into a Rossman fold nucleotide-binding domain, consisting of alternating α-helices and β-sheets with highly conserved Gly and Asp residues (Gly^{14}, Gly^{16}, Gly^{19}, Asp^{38}). Site-directed mutagenesis in these amino acids seems to confirm their predicted role in building the correct structure of the NAD^+ binding pocket [6,7]. Yet no data are available about the presumptive substrate interactive region. Primary structure comparisons have shown a highly conserved region in the middle of a polypeptide comprising amino acids Tyr^{152} and Lys^{156}. As short chain dehydrogenases do not require metal ions for their activity, an amino acid side chain has to hold the hydroxyl group in the correct position to mediate the electron transfer to the coenzyme. Biochemical data [8], identification of residues by chemical modification [9] and functional characterization of amino acids by site-directed mutagenesis of the related enzyme human-placental-15-hydroxyprostaglandin dehydrogenase [10] strongly suggest that Tyr^{152} is one of the amino acids involved in the catalytic sites of *Drosophila* ADH. To confirm this hypothesis, we have engineered a *Drosophila* ADH-Phe^{152} enzyme. We have assumed that this conservative change would not alter the overall conformation of this region and therefore changes in enzyme behaviour could not be attributed to structural disruption, but rather to an alteration of a certain chemical group essential for the catalytic function. The mutant has been expressed in a *S. cerevisiae* host/vector system previously described [11] and a one-step purification protocol by FPLC chromatography has been designed. Both crude yeast extracts and pure enzyme preparations have been assayed for ADH activity and protein content.

Correspondence address: S. Atrian, Department de Genètica, Facultat de Biologia, Universitat de Barcelona, Av. Diagonal 645, E-08028 Barcelona, Spain. Fax: (34) (3) 411-0969.

2. MATERIALS AND METHODS

2.1. *Materials*

T4 DNA ligase, *Bam*HI, *Hin*dIII, *Xba*I, dNTP's and 3,4-dichloroisocoumarin were obtained from Boehringer. *Spe*I was from New England Biolabs. *Acc*I and the DNA Sequencing Kit were from Pharmacia. Replitherm Thermostable DNA Polymerase Kit was purchased

from Epicentre Technologies. [α-^{35}S]dATP and Hybond-C nitrocellulose filters were from Amersham. PCR primers were synthesized by Oligos Etc. Inc. 1,10-Phenanthroline and E64 were obtained from Sigma. Other reported chemicals and reagents were from Sigma or Merck, and culture media reagents from Difco.

2.2. *Organisms and plasmids*

The following strains were used: *E. coli* JM109 [12] and *S. cerevisiae* WV36-201, an ADH-null strain. *D. melanogaster Adhs* intronless coding sequence was originally cloned in plasmid p3008. The phagemid pVT-U was used to express the *Drosophila Adh* gene in yeast (giving plasmid pVT-Adh). Vectors, plasmid constructions, *S. cerevisiae* strains and *Adh* expression in yeast have been previously described [11].

2.3. *DNA manipulation procedures*

Standard DNA cloning procedures were performed according to Maniatis [13] or following enzyme supplier's instructions. DNA from Minipreps was used for restriction analysis, fragment purification and bacterial transformations. DNA to transform yeast cells by the lithium salt method [14] was prepared using the Qiagen plasmid Kit for Midi preparations. Sequencing of double-stranded DNA with [α-^{35}S]dATP was carried out with the Pharmacia Sequencing Kit.

2.4. *Site-directed mutagenesis*

Site-directed mutagenesis was performed according to a cassette replacement method based on two coupled PCR reactions [15]. The first PCR reaction was carried out using the oligonucleotide 5'GGTGCCGGAGAAGACGGGCA3' as mutagenic primer (1291–1310 fragment of the non-coding strand) and the 20mer 5'TTCTGGACTTCTGGGACAA3' as upstream primer (1201–1220 fragment of the coding strand) (see complete ADH gene sequence in [16]). The second PCR reaction was carried out using the entire purified product of the first PCR as upstream and the 17mer/M13 primer (5'GTAAAACGACGGCCAGT3') as the downstream primer, taking advantage of the fact that the gene is cloned in a pUC derived vector (p3008). The first PCR amplification was performed in a final volume of 100 μl, containing 10 ng of p3008 (template DNA), 100 pmol of each primer, 2.5 μg of Repliterm Polymerase and 200 μM dNTP's. Samples were first kept at 94°C for 120 s to ensure initial denaturation, then 30 cycles of 60 s at 94°C (denaturation), 60 s at 55°C (annealing) and 60 s at 72°C (extension). Finally, 10 min at 72°C ensured that all amplified DNA was double-stranded. Products were fractionated on a 1.8% agarose gel and the band corresponding to the desired fragment was purified. Conditions for the second PCR were those described for the first step, except that the concentration of dNTP's was 400 μM, and the primer extension time was extended to 3 min per cycle, owing to the length of the segment to be amplified. Products of this second PCR were extracted once with phenol/chloroform, once with chloroform/ isoamyl alcohol, precipitated with ethanol/sodium acetate, pH 3.5 (2:0.1 v/v), resuspended in 12 μl of TE and further digested with *Bam*HI and *Xba*I. Plasmid p3008, carrying the wild type *Adh* gene, had been cleaved with the same enzymes so that the PCR fragment carrying the mutated codon could replace the original (p3008-Phe).

2.5. *Yeast cultures and protein extraction procedures*

Yeast clones were kept on yeast omission media (YOM) selective media plates and grown in liquid yeast peptone dextrose (YPD) cultures for expression of the heterologous proteins, as previously described [11]. Small scale crude extracts [11] provided enough material to assay ADH activity and to analyze protein and *Drosophila* ADH content by SDS-PAGE and immunoblotting. ADH was purified from large scale extracts prepared according to a glass beads cell disruption method [17], using as protease inhibitor the following 5 × mix: 20 mM EDTA, 20 mM PMSF, 10 μg/ml each of pepstatin, leupeptin, chymostatin, antipain, 0.5 mM 3,4-dichloroisocoumarin, 0.5 mM 1,10-phenanthroline and 0.1 mM E64C. Protein was quantified by the method of Bradford [18].

2.6. Drosophila *ADH purification*

The supernatant resulting from the disruption of 1 liter of yeast culture was first clarified by the addition of salmine sulfate to a final concentration of 0.28% (w/v) followed by centrifugation at 12,000 rpm for 20 min. The resultant supernatant was then fractionated by $(NH_4)_2SO_4$, and the protein precipitated between 35 and 55% saturation, then resuspended in 20 mM Tris-HCl buffer, pH 8.6, supplemented with 1% 2-propanol and 10^{-4} M DTT. The sample was desalted either by filtration chromatography through a 2.5 × 20 cm Sephadex- G25 column or by overnight dialysis. In the first case, fractions showing ADH activity were pooled and the solution was concentrated 10-fold using Ultrafree-CL filters Millipore (10 kDa exclusion pore). Aliquots of 500 μl of the crude ADH preparation were injected directly to a FPLC gel filtration system: Superose 12 and Superose 6 (Pharmacia) FPLC columns were connected in series to sum their resolution ranges. Columns had been equilibrated with the same Tris-HCl buffer. Samples were run at 0.3 ml/min and protein content was recorded by absorbance at 280 nm UV. Fractions showing ADH activity were kept at 4°C for further analysis.

2.7. Drosophila *ADH activity assay*

Drosophila ADH activity was measured by the increase of absorbance at 340 nm, recorded spectrophotometrically at 25°C using 2-propanol as substrate [19].

2.8. *Electrophoresis and immunoblotting*

SDS-PAGE was performed according to Laemmli [20], in 15% acrylamide gels. Proteins were visualized by both 0.1% Coomassie brillant blue and conventional silver staining. Western blotting was performed with a monoclonal antibody against *Drosophila* ADH (LLBE8) [21], following the procedures described in [22]. Quantification of protein bands on SDS-gels and immunoreacting bands on Western blots was performed by scanning densitometry in an Ultroscan XL equipped with the Gel Scan data processing Software (Pharmacia-LKB).

2.9. *Sequence analysis*

Cloning and sequencing were designed using the Sequence Analysis Software Package of the Genetics Computer Group of the University of Wisconsin (GCG) [23].

3. RESULTS AND DISCUSSION

3.1. *Mutagenesis and plasmid constructs*

Transformed *E. coli* JM109 cells were first screened for the mutant gene (p3008-Phe) by an *Acc*I digestion of their plasmid DNA, on account of the *Acc*I restriction site lost when the Tyr triplet (TAC) was replaced by the Phe triplet (TTC). Positive clones were further analyzed by DNA sequencing of the entire ADH coding region and compared with the same fragment in p3008 (Fig. 1). We thus verified the mutant sequence and established that no additional changes had been produced with respect to the *Adh*^{s} reference gene. A *Hin*dIII–*Spe*I fragment containing the mutated gene was obtained from p3008-Phe and cloned into the *Hin*dIII and *Xba*I sites of pVT-U, as previously described for wild-type *Adh* [11], yielding plasmid pVT-Adh-Phe. Yeast strain WV-201-36 was transformed with this plasmid to analyze ADH-Phe^{152} activity (Y-Phe strain). Positive and negative controls were the same yeast strain transformed with the pVT-Adh (wild-type) plasmid (Y-Wt strain) and the pVT-U plasmid without insert (Y-ϕ strain), respectively. Transformants were isolated by

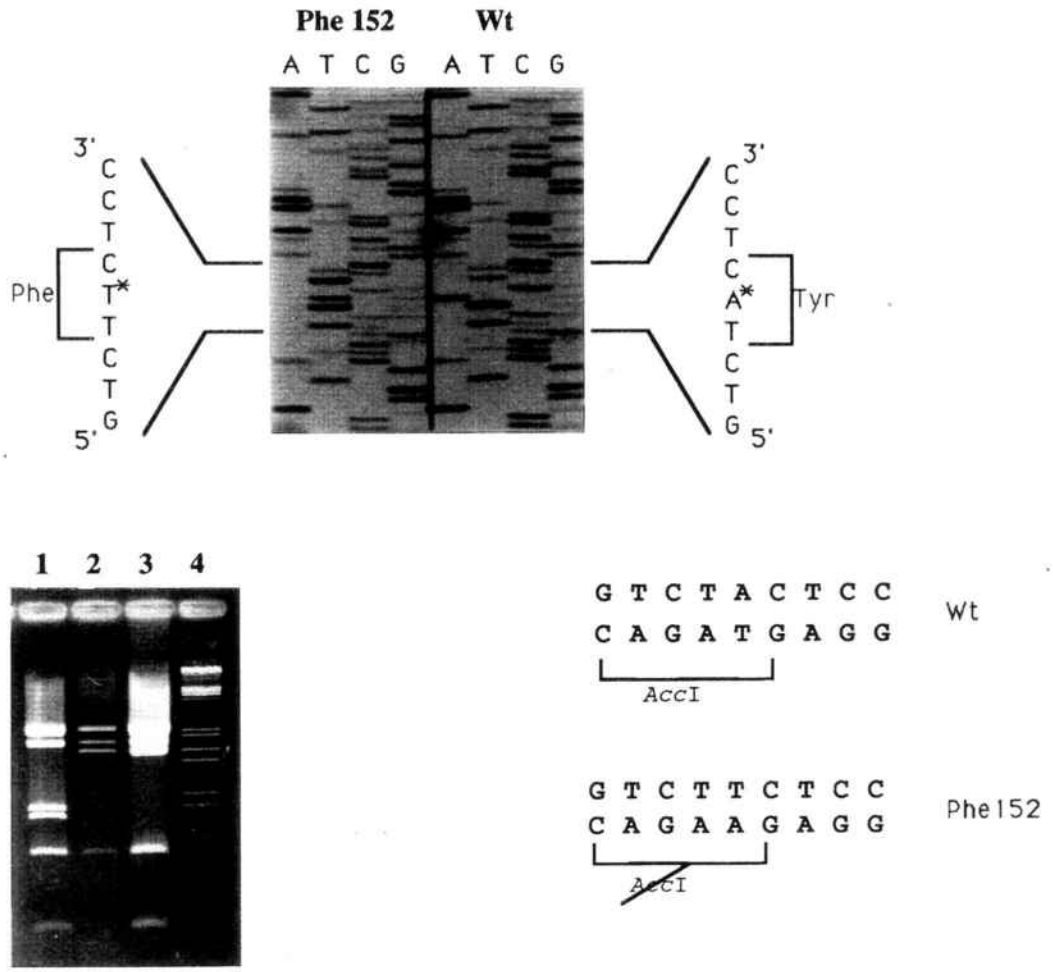


Fig. 1. DNA sequence of the wild-type *Adh* gene and the *Adh*-Phe^{152} mutated gene, showing the A→T change involved. (Below) Ethidium bromide-stained agarose gel (0.8%) of (1) 0.5 μg of plasmid pVT-Adh (wild-type gene) digested with *Acc*I, (2) 0.1 μg of plasmid pVT-Adh-Phe^{152} digested with *Acc*I, (3) 1 μg of plasmid pVT-Adh-Phe^{152} digested with *Acc*I and (4) 150 ng of λDNA digested with *Eco*RI and *Hin*dIII.

their ability to grow on selective media lacking uracil. Standard plasmid rescue procedures and *Acc*I restriction digestion were performed to verify that yeast cells carried the correct construction in each case (Fig. 1).

3.2. *Wild-type and ADH-Phe152 expression*

Crude extracts were prepared from selective 10 ml overnight cultures of Y-φ, Y-Wt and Y-Phe strains and used to quantify ADH activity. Values were considered according to total protein content of the samples, previ-

Table I

Expression of *Drosophila* ADH in yeast strain WV201-36

Transformant plasmid	Total yeast protein	ADH activity	Specific ADH activity
pVT-φ	1.401 μg/μl[a]	0.001 mU/μl[b]	0.000 mU/μg[c]
pVT-Adh (Wt)	5.420 μg/μl	1.950 mU/μl	0.360 mU/μg
pVT-Phe^{152}	1.003 μg/μl	0.001 mU/μl	0.000 mU/μg

[a]Concentration of yeast protein in 100 μl extracted from 15 ml of culture.
[b]mU of ADH activity per minute at 25°C per μg of yeast extract.
[c]mU of ADH activity per μg of total yeast protein.

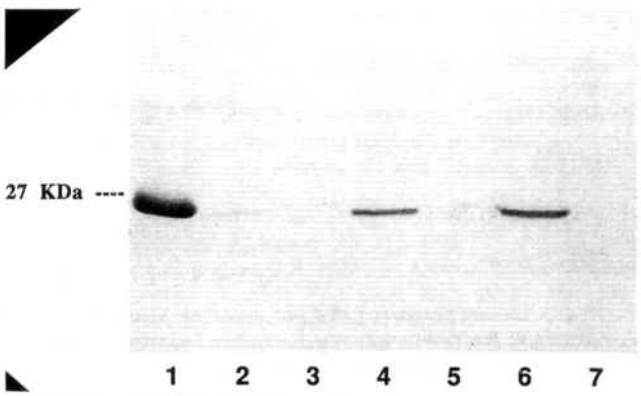


Fig. 2. Western blot analysis of *Drosophila* ADH. Control ADH, purified from *D. melanogaster* flies, was loaded in lanes (1) 0.05 mg, and (2) 1 μg. Protein extracts were from the following yeast strains (3) 20 μl of Y-φ (WV201-36 transformed with pVT-U plasmid without insert), (4) 20 μl of Y-Phe (WV201-36 transformed with pVT-Adh-Phe^{152} mutant plasmid), and (6) 5 μl of Y-Wt (WV201-36 transformed with pVT-Adh wild-type plasmid). (5) and (7) are plain sample buffer.

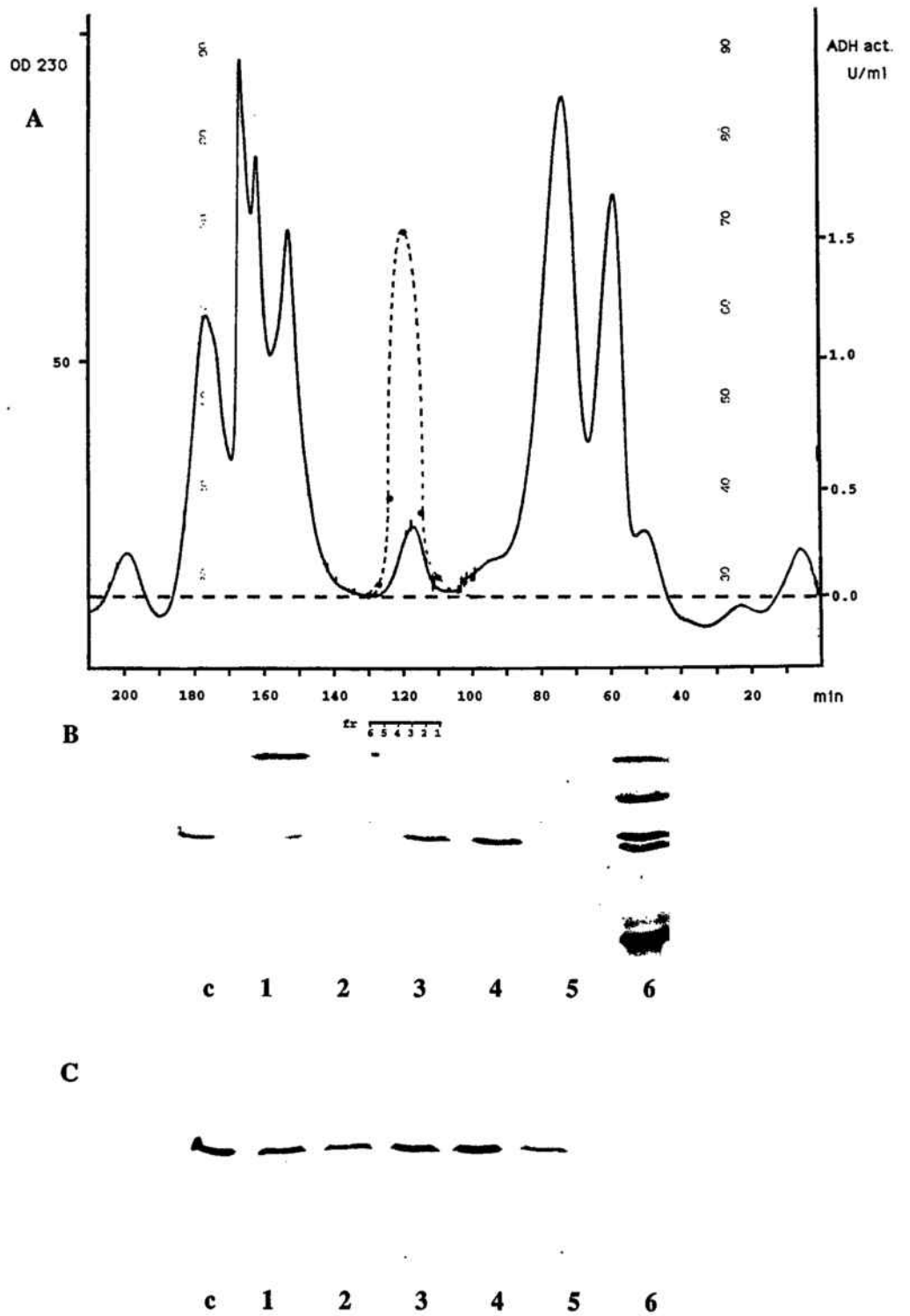


Fig. 3. Elution profile (—) of FPLC columns. The purification rationale was as follows: 500 μl of 35–55% ammonium sulfate-fractionated protein extract of the Y-Wt strain was run on the Superose 12+Superose6 series system and absorbance at 230 nm was recorded. Sample eluted between 110 and 130 min after injection was collected as 1 ml fractions and ADH activity was quantified spectrophotometrically at 340 nm (-.-.-.-). (B) SDS-PAGE of 10 μl aliquots of the fractions collected, stained for total protein with Coomassie blue: 'c' = 5 μg of control ADH (purified from fly homogenates). (1) to (5) aliquots of fractions 1–5, respectively. (6) Molecular weight markers. (C) Immunoblotting of a replica of the previous gel, showing that ADH activity corresponds to Coomassie blue stained bands.

ously quantified following Bradford. As shown in Table I, no significant ADH activity was detected in the Y-Phe extracts. Protein content from Y-ϕ and Y-Phe cultures were similar and rather low compared with that of Y-Wt. This is in agreement with the poor growth rate of these defective strains, which lack any ADH activity,

whereas *Drosophila Adh* gene complements the mutant host [11]. To assess the presence of ADH protein in the Y-Phe extracts, equivalent amounts of total protein of each strain were immunodetected, using *Drosophila* ADH monoclonal antibodies. Western blots (Fig. 2) showed that yeast cells synthesizing mutant ADH produced a similar amount of polypeptide to those expressing the wild-type gene.

3.3. *Yeast-synthezised ADH purification*

In order to extend the analysis of our preliminary results with crude extracts, a one step-purification method was performed with both Y-Wt and Y-Phe extracts from 1 liter cultures. Fig. 3A shows the profile obtained when total yeast protein of Y-Wt was chromatographed by FPLC as described in Materials and Methods. Previous assays with purified preparations of *Drosophila* ADH gave an approximate retention time of 120 min for the active dimer. In addition, fractionation of the yeast crude homogenate produced an isolated peak at the same retention time, which showed the ADH activity expected for the *Drosophila* enzyme. Coomassie blue staining of a SDS-polyacrylamide gel of the positive fractions revealed the presence of the predicted 27 kDa ADH monomer, 95% pure, following densitometry. Western blotting confirmed the ADH nature of the 27 kDa bands (Fig. 3B). When total yeast extracts from Y-Phe were chromatographed under the same conditions, an isolated peak of protein was detected at the corresponding retention time (data not shown). SDS-polyacrylamide gel stained with Coomassie blue and further immunodetected allowed the identification of the expected ADH 27 kDa band, but no enzymatic activity could be recorded in any of these fractions.

3.4. *Behaviour of the mutant ADH-Phe^{152}*

Lack of ADH-Phe^{152} catalytic activity in purified preparations supports the predicted critical role of Tyr^{152} in the substrate-binding pocket of the catalytic domain. Our results further suggest that the hydroxyl group of the Tyr side chain could play an active role in a hydrogen-bond interaction between the substrate and the established enzyme–coenzyme complex during catalysis. Besides, our data agree with those obtained from 15-hydroxyprostaglandin dehydrogenase [10], although no direct correlation could be made between the catalytic domains of the two enzymes, as they bind structurally different substrates. Some authors thus claim that different subclasses of short chain dehydrogenases should be considered [4]. Moreover, the loss of activity in the mutated 15-hydroxyprostaglandin dehydrogenase (Tyr^{151}-Ala) may have been due not only to the absence of the reactive group but also to a structural alteration of the catalytic site, whereas this second explanation could not be invoked in our case. On the other hand, the similarity on the elution profiles of *Drosophila* wild-type ADH and the mutated enzyme seems to suggest that the two can produce apparently similar dimeric forms. Some authors claim that Tyr^{152} could play a critical role in subunit binding to produce active dimeric forms [10], but our results do not seem to support this interpretation. In the absence of a tridimensional structure for *Drosophila* ADH, we postulate that the Tyr^{152} residue lies in the core of the catalytic domain and makes an essential contribution to the electron transfer route, indispensable for substrate transformation.

Acknowledgements: We are indebted to Dr. J. Fibla, from our department, for kindly providing the monoclonal antibodies used in this work. We are grateful to Dr. H. Jörnvall for helpful advice, to Dr. E. Mendez and other members of his laboratory for teaching FPLC techniques and to Dr. Casals for his helpful guidance in the use of the FPLC equipment at the University of Barcelona. We thank R. Rycroft for revising the English version of this manuscript. This work was supported by a grant from DGICYT (Ministerio de Educación y Ciencia, España) PB88-0188.

REFERENCES

[1] Chambers, G.K. (1988) in: Advances in Genetics (Caspari, E.W. and Scandalios, J.G. eds.) pp. 39–107, Academic Press, New York.
[2] Thatcher, D.R. (1980) Biochem. J. 187, 875–886.
[3] Benyajati, C., Place, A., Powers, D.A. and Sofer, W. (1981) Proc. Natl. Acad. Sci. USA 78, 2717–2721.
[4] Ghosh, D., Weeks, C.M., Grochulski, P., Duax, W.L., Erman, M., Rimsay, R.L. and Orr, J.C. (1991) Proc. Natl. Acad. Sci. USA 88, 10064–10068.
[5] Persson, B., Krook, M. and Jörnvall, H. (1991) Eur. J. Biochem. 200, 537–543.
[6] Chen, Z., Lu, L., Shirley, M., Lee, W.R. and Chang, S.H. (1990) Biochemistry 29, 1112–1118.
[7] Chen, Z., Lee, W.R. and Chang, S.H. (1991) Eur. J. Biochem. 202, 263–267.
[8] Winberg, J.O., Thatcher, D.R. and McKinley-McKee, J.S. (1982) Biochim. Biophys. Acta 704, 7–16.
[9] Krook, M., Prozorovski, V., Atrian, S., Gonzàlez-Durante, R. and Jörnvall, H. (1992) submitted for publication.
[10] Ensor, C.M. and Tai, H.H. (1991) Biochem. Biophys. Res. Commun. 176, 840–845.
[11] Atrian, S., Gonzàlez-Duarte, R. and Fothergill-Gilmore, L. (1990) Gene 93, 205–212.
[12] Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Gene 33, 103.
[13] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning, 2nd edn., Cold Spring Harbor Lab. Press.
[14] Ito, H., Fukuda, Y., Murata, K. and Kimura, A. (1983) J. Bacteriol. 153, 163–168.
[15] Landt, O., Grunert, H.P. and Hahn, U. (1990) Gene 96, 125–128.
[16] Kreitman, M. (1983) Nature 304, 412–417.
[17] Dunn, B. and Wobbe, C.R. (1990) in: Current Protocols in Molecular Biology (Ausubel, F.M. et al. eds.), Greene Publishing Associates/Wiley Interscience, New York.
[18] Bradford, M.M. (1976) Anal. Biochem. 72, 248–254.
[19] Juan, E. and Gonzàlez-Duarte, R. (1980) Biochem. J. 189, 105–110.
[20] Laemli, U. (1970) Nature 227, 680–685.
[21] Fibla, J., Enjuanes, L. and Gonzàlez-Duarte, R. (1989) Biochem. Biophys. Res. Commun. 160, 638–646.
[22] Visa, N., Fibla, J., Santa-Cruz, M.C. and Gonzàlez-Duarte, R. (1992) J. Histochem. Cytochem. 40, 39–49.
[23] Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. 12, 387–395.

Drosophila lebanonensis ADH: Analysis of recombinant wild-type enzyme and site-directed mutants.

Drosophila lebanonensis ADH: analysis of recombinant wild-type enzyme and site-directed mutants.

The effect of restoring the consensus sequence in two positions.

Ricard Albalat, Sílvia Atrian and Roser Gonzàlez-Duarte.

Departament de Genètica. Facultat de Biologia. Universitat de Barcelona.
Av. Diagonal 645. E-08028 Barcelona, Spain.

Keywords: short-chain dehydrogenases, alcohol dehydrogenase; *Drosophila lebanonensis;* site-directed-mutagenesis; catalytic efficiency.

Address for correspondence:

Roser Gonzàlez-Duarte
Department de Genètica.
Facultat de Biologia. Universitat de Barcelona.
Av. Diagonal 645.
E-08028 Barcelona, Spain.
Fax: 34-3-411-09-69
Tel: 34-3-402-14-94

Footnote: numbering of *D.lebanonensis* ADH positions is (-1) in relation to *D.melanogaster*, due to differences in the amino acid number at the N-terminus.

SUMMARY

Unique amino acid substitutions occur in *D.lebanonensis* ADH. They are found within the putative NAD^+ binding domain and affect residues that are otherwise highly conserved in all other species of the genus. To restore the consensus amino acids, we have constructed an expression system for this enzyme in *E.coli*, and engineered two mutants, $Ala^{13}Gly$ and $Asn^{56}Thr$. The biochemical and kinetic features of these "retromutants" are consistent with increased catalytic efficiency and thermal stability. Thus, results show that wild-type *D.lebanonensis* ADH can be improved by site-directed mutagenesis.

1. INTRODUCTION

Alcoholic compounds are oxidized in *Drosophila* by a member of the short-chain dehydrogenase family [1], in contrast to what happens in most eukaryotes, which have developed a medium-chain dehydrogenase to metabolize alcohols. *Drosophila* alcohol dehydrogenase (ADH) is a non-metalloenzyme, active as a dimer of two subunits of 253-255 amino acids, which shares no sequence homology with the polypeptides of the Zn-containing medium-chain alcohol dehydrogenases [2]. The three-dimensional structure of the medium-chain horse liver ADH [3] allowed the identification of functional residues involved in coenzyme and substrate binding, hydride transfer, metal coordination and monomer surface interaction. However, no tertiary structure is yet known for *Drosophila* ADH and therefore the structure/function relationships must be approached through different strategies. Putative critical residues for enzyme architecture and catalytic properties, highlighted as conserved positions among all short-chain dehydrogenases and all known *Drosophila* ADH's [2, 4, 5], have then been analyzed either by chemical modification [6] or site-directed mutagenesis [7-11]. Fragment 9-39 of the *Drosophila* polypeptide is the only alignable segment with the medium-chain ADH's [2], where the *homologous* fragment (194-224) forms the ß1-α2-ß2 motif of the dinucleotide-binding domain, or Rossmann Fold, found in all NAD^+/ FAD^+ binding enzymes [12]. The catalytic behaviour of mutants in positions Gly^{14}, Gly^{16}, Gly^{19} and Asp^{38} of the *D.melanogaster* enzyme is in agreement with the predicted involvement of this region in the binding of the cofactor and its preference for NAD^+ versus $NADP^+$ [7]. Again, sequence alignments have allowed the prediction of two key residues for substrate interaction, Tyr^{152} and Lys^{156} [2], whose substitution in *D.melanogaster* ADH yields inactive or poorly active enzymes [9-11].

Until now, all site-directed mutants have been engineered with *D.melanogaster* ADH, by far the most well-known species of the genus at all levels [13]. Nevertheless, valuable information could be obtained from other species whose ADH has been deeply characterized [5]. The fact that 110 out of 255 positions of the subunit polypeptide are not conserved in some other *Drosophila* species provides an excellent source of evolutionary tested enzyme variants and, among these, *D.lebanonensis* appears as an excellent candidate for function/structure

analysis. Specific activity of *D.lebanonensis* ADH is lower than that of other *Drosophila* ADHs, but, paradoxically these flies exploit rich alcoholic environments, and eventually outgrow *D.melanogaster* in number, probably because they accumulate larger amounts of the enzyme [14]. The amino acid sequence of *D.lebanonensis* ADH has been determined [15], the biochemical features of the enzyme have been described [16] and the ADH-coding gene has been isolated and analysed [17, 18]. Four unique amino acid substitutions with respect to all other *Drosophila* ADH's: Ala^{13}, Phe^{33}, Leu^{45} and His^{60} make this enzyme particularly interesting. These changes, as well as others only present in one additional species of all known *Drosophila* ADH's (Thr^{43}, Asn^{56} and Thr^{61}), are in the putative NAD^+ binding pocket and may therefore be considered responsible for the different catalytic properties of the enzyme. In particular, position Ala^{13} represents a singular substitution not only among *Drosophila* ADH's [5] but also among short-chain DH's [4], eukaryotic medium-chain DH's [19] and most other NAD^+/FAD^+ dependent oxidoreductases [12]. *D.melanogaster* ADH^{n11}, a $Gly^{14}Asp$ mutant generated by ethyl methane sulfonate, shows no ADH activity and low NAD^+ affinity [20]. Besides, site-directed mutants $Gly^{14}Val$ and $Gly^{14}Ala$ produce either total inactivation or a 31% decrease in activity, respectively [7]. Another position, Thr 56, appears altered only in *D.lebanonensis* (Asn^{56}) and in *D.mayaguana* (Ile^{56}). Thr/Ile replacement is a common change found in ADH interspecies comparisons [5], whereas the Thr/Asn change is much less frequent. This position lies in the assumed third ß-sheet of the NAD^+ binding domain and is one of the aligned residues with the medium-chain ADH's [2]. At the boundaries of Asn^{56}, three additional very conserved positions are found: Asn^{63}, conserved in all short-chain DH's, and $His^{60}Tyr$ and $Leu^{45}Ile$, two of the *D. lebanonensis* ADH "only" changes.

The purpose of the present study is to analyze Ala^{13} and Asn^{56} in *D.lebanonensis* ADH. In order to set up a heterologous expression system, an intronless *Adh* gene of *D.lebanonensis* was constructed using reverse-PCR, and subsequently cloned in *E.coli*. Site-directed mutagenesis was then performed on this construct to change Ala^{13} to Gly and Asn^{56} to Thr, to reconstruct evolutionary events and restore the consensus ADH residues. Recombinant wild-type ADH and $Ala^{13}Gly$ and $Asn^{56}Thr$ mutant enzymes were purified by FPLC from crude bacteria homogenates. Some of their catalytic features were studied.

2. MATERIAL AND METHODS

2.1. *Materials, organisms and plasmids*

Restriction enzymes were obtained from Boehringer-Mannheim. Taq DNA polymerase was from Promega. PCR primers were synthesized by Oligos Etc. Inc. Hybond-C nitrocellulose filters, (α-^{35}S) dATP and Ligation Kit were purchased from Amersham. Other reported chemicals and reagents were from Sigma and Merck, and culture media reagents were from Difco. Plasmid pBluescript was from Stratagene. *E.coli* JM105 and plasmid pKK223-3, used to express the *Adh* gene, were from Pharmacia-LKB Biotechnology. *D.lebanonensis* flies were from a natural population caught in Gandesa, Tarragona (Spain) and maintained in our laboratory under standard conditions for several years.

2.2. *RNA preparation and Reverse-PCR*

Total RNA of *D.lebanonensis* was purified from larvae according to Jowett [21]. cDNA was synthesized using 200 units of MoMuLV reverse transcriptase (BRL), in a final reaction volume of 20 µl, containing 1-2µg of total RNA, 100 pmols of the oligonucleotide SH5 as downstream primer, 1mM of dNTPs, 25 units of RNAase inhibitor (Boehringer) and 3 mM $MgCl_2$. Then, samples were kept for 10 min at 23°C, 45 min at 42°C and 5 min at 94°C to ensure initial hybrid denaturation. For the PCR reaction, 100 pmols of oligonucleotide SH1 (upstream primer), 2.5 units of Taq DNA polymerase and 8µl of 10x PCR buffer were added to a final volume of 100µl. A 30-cycle amplification was carried out at 94°C/30s, 50°C/60s and 72°C/60s. Finally, samples were kept at 72°C for 10 min before fragment purification.

2.3. *Cloning and expression of* D.lebanonensis Adh *in* E.coli.

The PCR product was initially subcloned in the plasmid pBluescript for restriction analysis and sequencing. In order to produce the recombinant enzyme, the *Adh* coding region of *D.lebanonensis* was cloned in pKK223-3 and the recombinant plasmid was used to transform *E.coli* JM105. Overnight cultures of 50 ml of LB-ampicillin were diluted to 500 ml of fresh LB-ampicillin and grown for 1.5h. IPTG was then added to a final concentration of 1mM and cultures were incubated at 30°C for 3 h. Cells were harvested, washed twice in 20

mM Tris-HCl pH 8.6, resuspended in 2 ml of the same buffer, sonicated three times for 15 s at 30 W and centrifuged in a microfuge for 15 min. Manipulations were performed at 4°C. Crude supernatant was used for activity assays, SDS-PAGE, immunoblotting and further purification.

2.4. *Site-directed mutagenesis by PCR*

Site-directed mutagenesis was performed by a two coupled PCR reaction method as desbribed [9], using the oligonucleotides SH2 and SH3 as mutagenic primers. PCR products were subcloned in vector pKK223-3 and *E.coli* JM105 was transformed with the recombinant constructions. Cloning products were sequenced to verify the mutant sequences and to rule out additional changes. Sequencing of double-stranded DNA with (α-^{35}S) dATP was carried out with the Pharmacia Sequencing Kit.

2.5. *Purification of recombinant ADH*

An FPLC system from Pharmacia was set up to purify the ADH enzyme. Mono Q 5/5 HR was used for ion exchange chromatography, affinity chromatography was by Blue Sepharose Hi-Trap/5 and Superose 12 was used for gel filtration. The equilibration and elution buffer was 20 mM Tris-HCl, pH 8.6 supplemented with 10^{-4}M DTT. 1M NaCl in the same buffer was used to elute Blue Sepharose and also to clean Mono Q after each run. Centricon 10 Microconcentrators (10000 MW cut-off) were used to concentrate samples. Protein content was recorded following the absorbance at 280 nm. Fractions were further tested by SDS-PAGE and immunobloting, and quantified by the method of Bradford (BioRad).

2.6. *Protein determinations: electrophoresis, immunoblotting and antibody recognition.*

SDS-PAGE was performed in 12.5% acrylamide gels and proteins were visualized by Coomassie-blue staining. Western blotting was performed with a monoclonal antibody (LLBE8) against *Drosophila lebanonensis* ADH, following [22].

2.7. *Enzymatic determinations: activity inhibition, kinetic parameters, pH profiles and thermal stability.*

ADH activity was measured spectrophotometrically by the increase in absorbance at 340 nm, using propan-2-ol as substrate and NAD^+ as cofactor in 20 mM Tris-HCl pH 8.6 [16]. For activity inhibition determinations, total protein extract from the bacterial cultures was incubated with and without monoclonanal antibody LLBE8 (anti-*D.lebanonensis* ADH) or MMBB8 (anti-*D.melanogaster* ADH) [14] for 1h at 37°C before measuring activity at 340 nm. Kinetic constants for NAD^+ were determined using 0.0625, 0.125, 0.250, 0.5, 1.0, 1.5, 2.0, 2.5 and 4.0 mM NAD^+ with a constant alcohol concentration of 12 mM. Kinetic constants for propan-2-ol were measured using 2.5, 5.0, 10, 15, 20, 40 and 120 mM alcohol with 2.0 mM of NAD^+. K_m and k_{cat} values were calculated with the program ENZFITTER on a Personal Computer. ADH activity for wild-type and mutants was also determined at pH 7.0, 8.0, 9.0 and 10.0 with 12 mM propan-2-ol and 2mM NAD^+. To evaluate the thermal stability of the enzyme, samples of purified wild-type ADH and mutant forms were incubated at 40°C. Aliquots from each sample were taken at 0', 10', 25', 40' and 60', and assayed for ADH activity with 12 mM propan-2-ol and 2 mM NAD^+. Molarities of NAD^+ and alcohols refer to final concentrations. All activity tests were performed at least twice.

2.8. *Sequence analysis.*

Cloning and sequencing were designed using the Sequence Analysis Software Package of Genetics Computer Group of the University of Wisconsin (GCG) [23]. This was also used for the analysis of DNA restriction patterns and protein comparison tables.

3. RESULTS

3.1. *Construction and expression of the* D.lebanonensis *wild-type* Adh *gene.*

The reverse-PCR protocol provided an intronless *Adh* gene suitable for an *E.coli* expression system. The MoMuLV reverse transcriptase reaction was performed on total *D.lebanonensis* RNA using as primer the oligonucleotide SH5, which comprises the last 4 coding triplets, the stop codon and a *Hind*III restriction site (Table I). After this step, only

Adh-mRNA/cDNA hybrids were obtained. Then, the direct PCR reaction was carried out using the SH1 primer containing an *Eco*RI restriction site just before the ATG triplet and the first 5 codons of the gene (Table I). Therefore, the final amplified product was the complete coding region of *D.lebanonensis Adh*, flanked by *Eco*RI and *Hind*III restriction sites at its 5' and 3' ends respectively, which allowed directional cloning in the pBluescript vector (Figure 1). Sequencing of the cloned fragment revealed that it contained the correct coding region of the *D.lebanonensis* Adh gene [17], so reverse transcription of mRNA and PCR amplification had introduced no artefactual changes.

In order to express this construct, the *Adh* coding region was then subcloned in pKK223-3, by the same *Eco*RI/*Hind*III restriction strategy as before, and the recombinant plasmid was used to transform *E.coli* JM105 cells. *Drosophila* ADH activity was detected spectrophotometrically in crude protein extracts prepared from bacterial cultures induced by IPTG. No activity was found in untransformed cells, nor in cells transformed with the pKK223-3 vector without insert. An antibody inhibition test was carried out in order to determine whether recombinant *D.lebanonensis* ADH retained the expected differential antigenic features with respect to that of *D.melanogaster*. Although mAbs MMBB8 and LLBE8 exhibit cross-reaction with both enzymes, LLBE8 specifically inhibits *D.lebanonensis* ADH, while *D.melanogaster* ADH activity remains unaltered [14]. When LLBE8 was added to the recombinant ADH protein extract, enzymatic activity fell to 18% of the control samples (protein extracts without antibody). This was in agreement with the expected antigenic behaviour for the non-recombinant enzyme.

3.2. *Site-directed mutagenesis.*

Mutagenic PCR amplifications were successfully carried out with primers SH1 and SH2 (first PCR) and SH5 (second PCR) to obtain the $Ala^{13}Gly$ mutant, and primers SH1 and SH3 (first PCR) and SH5 (second PCR) for the $Asn^{56}Thr$ mutant. Thus, in both cases, the final PCR product was the mutated coding region flanked by *Eco*RI and *Hind*III restriction sites. The presence of the desired mutation and the absence of additional changes was always verified by sequencing the PCR fragment, once subcloned in the pBluescript plasmid. Positive clones were used to subclone the *Eco*RI-*Hind*III segment in the expression vector pKK223-3.

3.3. *Purification of recombinant* D.lebanonensis *wt and mutant enzymes.*

The following purification protocol was followed to obtain pure recombinant *D.lebanonensis* ADH . *E.coli* total protein extract in 20 mM Tris-HCl pH 8.6 was injected in Blue-Sepharose Hi-Trap/5 adapted to FPLC. After 10 ml wash with the same buffer ADH was eluted with 1 M NaCl. Fractions containing ADH activity were pooled and concentrated up to a final volume of 0.5 ml. The sample was then fractionated in Superose 12. Pure ADH was recovered in 2 fractions. Calculated yield was 1 mg of pure protein per 500 ml of O/N bacterial culture. This method was used to purify wild-type ADH and the $Ala^{13}Gly$ mutant of *D.lebanonensis.*

The $Asn^{56}Thr$ site-directed mutant showed irreversible denaturation in front of 1 M NaCl. Although SDS-PAGE revealed ADH-containing fractions after the Blue-Sepharose step, neither exhaustive dialysis nor gel filtration restored activity. Besides, a similar denaturing effect was observed following incubation in NaCl of crude protein extracts from bacteria expressing this ADH mutant form. A 75% decrease in ADH activity was obtained after 3 min incubation at 4°C with 1 M NaCl; even 0.1 M NaCl led to a decrease of 65% in enzyme activity. 1 M NaCl did not affect the activity of wild-type ADH nor that of the $Ala^{13}Gly$ mutant synthesized in *E.coli.* In view of these results an alternative purification procedure was used for the ADH $Asn^{56}Thr$ mutant. 500 µl of *E.coli* protein extract in 20 mM Tris-HCl pH 8.6 was injected in MonoQ FPLC equilibrated with the same buffer. Unretained eluent contained most of the initial ADH activity, whereas a high proportion of*E.coli* proteins remained bound to the gel matrix. The ADH-containing fractions were pooled, concentrated and then fractionated in Superose 12. Overall calculated yield was 600 µg of pure ADH from 500 ml of O/N culture. When wild-type recombinant ADH was purified according to this protocol it showed a kinetic behaviour which was comparable to that obtained following the affinity chromatography described above.

Figure 2(A) shows the SDS-PAGE analysis of the purified recombinant *D.lebanonensis* ADH wild-type and mutant forms, and Figure 2(B) shows the results of a Western blot immunodetected with an ADH monoclonal antibody. The only band present in the fractions

exhibiting enzyme activity is recognized by the ADH specific mAb, and also the amount of protein synthesized is similar for the wild-type and mutant forms.

3.4. *pH profiles for ADH.*

The influence of pH on wild-type and mutant ADH's activities was tested at different pH's from 7.0 to 10.0. Recombinant ADH wild-type showed a maximum at pH 8.0, with a slight decrease at higher pH's. $Ala^{13}Gly$ and $Asn^{56}Thr$ mutants showed maximum activity at pH 10.0 and lower activity at pH 8.0 and 9.0 (Figure 3A). In all cases, however, differences never amounted to more than 20%. A decrease in activity of more than 60% appeared at pH 7.0 for wild-type and mutants, with similar pH profiles for the three proteins.

3.5. *Thermal stability of the wild-type and mutant enzymes.*

Thermal denaturation tests were performed at 40°C in duplicated samples obtained from independent preparations, to rule out the effect of specific batch contaminants. Results clearly showed differential behaviour of the mutant enzymes versus the wild-type(Figure 3B). Whereas wild-type ADH lost more than 50% of activity after 20 min at 40°C, and showed no detectable activity after 60 min incubation, both mutants retained 50% activity after 60 min at 40°C.

3.6 *Kinetic characterization of wild-type and mutated enzymes.*

The kinetic constants, K_m and k_{cat}, were calculated for the recombinant wild-type and both mutant enzymes with NAD^+ as coenzyme and propan-2-ol as substrate (Table II). In spite of the fact that the substitutions studied affect the putative NAD^+ allocating region, the coenzyme and substrate binding ability of the enzymes remained essentially the same as that of the wild-type protein, as reflected by similar K_m values. However, significant if not spectacular differences arose from the comparison of the k_{cat} parameter, as both mutants showed higher catalytic rate than the recombinant wild-type. Then, if we consider the efficiency of the enzyme, measured by the k_{cat}/K_m ratio, both substitutions led to a clear improvement of function: an increase of a 1.7-fold for position 13 and 1.3-fold for position 56.

4. DISCUSSION

D.lebanonensis ADH has two unique amino acid replacements in otherwise conserved positions among a wide range of dehydrogenases [4,19]. Their location in the polypeptide chain suggest that both Ala^{13} and Asn^{56} could be involved in the catalytic funtion of this enzyme. We have used reverse-PCR and SDM to restore the consensus amino acids, and the kinetic effects of each substitution have been evaluated. *D.lebanonensis* can tolerate high alcohol concentrations and successfully exploit ethanol/propan-2-ol-rich habitats. When competing with *D.melanogaster*, a species clearly adapted to alcohol, *D.lebanonensis* flies can eventually outgrow them in number. However, adaptation to alcohol cannot be attributed to an increased catalytic efficiency of the enzyme, as k_{cat} values of *D.lebanonensis* ADH are lower than those of *D.melanogaster* and furthermore $K_m^{alcohol}$ and K_m^{NAD+} are comparable to those of other species [24]. The unique amino acid substitutions present in the putative NAD^+ binding region could partially account for the kinetic features of the *D.lebanonensis* enzyme. Of those, position 13 (14 in *D.melanogaster*) is the first glycine of the ß-α-ß domain forming the ADP-binding Rossmann fold of the protein, and the first residue of the conserved box [$\underline{G}^{13(14)}$ L G G I G], representative of the short-chain dehydrogenase motif [$\underline{G}^{14}$ xxx G x G], also found in the medium-chain ADH's as [$\underline{G}^{199}$ xx G G x G]. Studies on *D.melanogaster* mutants in this position show that the only substitution rendering an active enzyme is Gly/Ala. A site-directed-mutant with Val^{14}[7] and an EMS-induced Asp^{14} [20] produce a dead enzyme. This is also in agreement with alanine being the only "natural" alternative: Ala^{13} in *D.lebanonensis* ADH and Ala^{199} in the medium-chain *Alcaligenes eutrophus* ADH [19].

When considering the kinetic behaviour of the $Gly^{14}Ala$ mutant form of *D.melanogaster* ADH, it showed a 31% decrease in activity [7]. This was attributed to a lower affinity for NAD^+ binding (K_m^{NAD+} increased 3-fold) and to a reduction in catalytic efficiency (k_{cat} decreased 1.47-fold) [7]. The authors claimed that Ala^{14} interfered with coenzyme binding, and that this, at its turn, affected both enzyme-substrate interaction and the ternary complex dissociation rate. We have restored glycine at position 13 of the *D.lebanonensis* enzyme. Our results clearly show that this retromutant maintains the K_m^{NAD+} whereas the k_{cat} increases by 1.7 fold. It is remarkable that the only allowed substitution in this position

(Gly/Ala) decreases the efficiency of the enzyme, irrespective of its molecular background, i.e. SDM-*D.melanogaster* Ala^{14} and wild-type *D.lebanonensis* (Ala^{13}) share a similar decrease in k_{cat} with respect to wild-type *D.melanogaster* (Gly^{14}) and SDM-*D.lebanonensis* (Gly^{13}) (Table III). In our mutant, K_m for propan-2-ol is unaltered, with only a modest k_{cat}/K_m increase, again in agreement with the *D.melanogaster* direct mutant.

Position 56 is threonine in all *Drosophila* ADH's, with the only exception of *D.lebanonensis* (Asn^{56}) and *D.mayaguana* (Ile^{56}). In contrast to the previous mutant, this position is not only non-conserved among short-chain dehydrogenases, but also enclosed in an hyper-variable region, which is difficult to align [4]. However, this threonine is conserved among all animal medium-chain alcohol dehydrogenases, position 238. $Thr^{56/238}$ could have an important role in the accommodation of the coenzyme, facilitating correct interaction with small alcohols. Other short-chain dehydrogenases, as well as class III and sorbitol dehydrogenases, which utilize much larger substrates, have an Asn or Asp residue at 238. Thus, the $Asn^{56}Thr$ mutant in *D.lebanonensis* restores the ideal amino acid for a short-chain alcohol dehydrogenase. This mutant shows a similar behaviour to $Ala^{13}Gly$; its k_{cat} value increases x1.3 with respect to *D.lebanonensis* wild-type. The kinetic effects of both substitutions are in good agreement with the size of the amino acids involved. In fact, r° for Gly is 5.0Å, and for Ala 5.5Å [25], and their substitution produces a 10% increase in their packing volume (k_{cat} ADH ^{Gly13} x1.7) , while r°for Asn (6.4Å) is only 1.38% greater than that of Thr (6.3Å) (k_{cat} ADH ^{Thr56} x1.3).

Thermal denaturation has been tested in pure enzyme preparations to check the stability of the recombinant enzyme. Both *retromutants* were significantly more stable than the wild-type enzyme. After 1 hour incubation at 40°C the former retained 50% of their activity , while the latter became totally inactive. This was fully in agreement with data obtained with *D.melanogaster* wild-type ADH which retained, under the same conditions, 60% more activity than the $Gly^{14}Ala$ mutant [7]. Again, we show that, despite the overall primary structure frame, the presence of a larger amino acid in position 13/14 disturbs the architecture of the molecule and lowers its stability.

In summary, when considering the consensus sequence for ADH in all analyzed *Drosophila* species, it appears that selection has fixed the most suitable amino acid in each position of the polypeptide chain. Thus, any alteration to this sequence would have a negative effect. By the same argument, restoring the consensus sequence in *D.lebanonensis* would produce a more efficient enzyme. This, according to our data, is really the case. Retromutants are "better" enzymes, as our results fully support an improvement in enzyme efficiency and thermal stability. It could be argued that in *D.lebanonensis* flies, the wild-type enzyme, albeit kinetically unfavourable, was fixed because the large amount of ADH synthesized compensated for a reduction in its catalytic efficiency.

On the other hand, the kinetic differences between wild-type *D.lebanonensis* ADH and the engineered mutants are not as marked as those obtained when mutating *D.melanogaster* ADH. In both *D.lebanonensis* mutants, K_m values for NAD^+ and propan-2-ol remain unaltered and k_{cat} values show a slight increase. This could be due to an inherent plasticity of the *D.lebanonensis* enzyme, in contrast to the more evolved form of *D.melanogaster*, in which further substitutions would impair the catalytic function. The ancestral phylogenetic position of *D.lebanonensis* also supports this hypothesis [26].

Acknowledgements

We are grateful to Dr.J.Fibla who kindly provided the monoclonal antibodies used in this work. We thank R.Rycroft for revising the English version of this manuscript. This work was supported by a grant from the CICYT (Plan Nacional I+D, Ministerio de Educación y Ciencia, España), BIO92-0591-CO2-02.

REFERENCES

[1] Jörnvall, H., Danielsson,O., Höög,J.-O. and Persson, B. (1993) in: Methods in Protein Sequence Analysis, (Imahori, K. and Sakiyama, F., eds.) pp.275-282. Plenum Press, New York.

[2] Jörnvall, H., Persson, M. and Jeffery, J. (1981) Proc. Natl. Acad. Sci. USA 78, 4226-4230.

[3] Eklund,,H., Nordström,B., Zeppezauer,E., Söderlund, G., Ohlsson, I., Boiwe, T., Söderberg, B., Tapia, O., Brandén, C.-I. and Akeson, A. (1976) J. Mol. Biol. 102, 27-59.

[4] Persson, B., Krook, M. and Jörnvall, H. (1991) Eur. J. Biochem. 200, 537-543.

[5] Atrian,S., Marfany,G., Albalat,R. and Gonzàlez-Duarte,R. (1992) Biochem. Genet. (Life Science Advances) 11, 19-29.

[6] Krook, M., Prozorovski, V., Atrian, S., Gonzàlez-Duarte, R. and Jörnvall, H. (1992) Eur. J. Biochem. 209, 233-239.

[7] Chen, Z., Lu, L., Shirley, M., Lee, W.R. and Chang, S.H. (1990) Biochemistry 29, 1112-1118.

[8] Chen, Z., Lee, W.R. and Chang, S.H. (1991) Eur. J. Biochem. 202, 263-267 .

[9] Albalat,R., Gonzàlez-Duarte,R. and Atrian,S. (1992) FEBS Lett.308, 235-239.

[10] Cols, N., Marfany, G., Atrian,S. and Gonzàlez-Duarte,R. (1993) FEBS Lett. 319, 90-94.

[11] Chen, Z., Jiang, J.C., Lin, Z., Lee, W.R., Baker, M.E. and Chang, S.H. (1993) Biochemistry 32, 3342-3346.

[12] Wierenga, R.K., Terpstra, P. and Hol, W.G.J. (1986) J. Mol. Biol. 187, 101-107.

[13] Chambers, G.K. (1988) in: Advances in Genetics (Caspari, E.W. & Scandalios, J.G., eds.) pp.39-107, Academic Press, New York.

[14] Fibla,J., Enjuanes,J. and Gonzàlez-Duarte,R. (1989) Biochem. Biophys. Res. Commun. 160, 638-646.

[15] Villaroya, A., Juan, E., Egestad,B. and Jörnvall, H. (1989) Eur. J. Biochem. 180, 191-197.

[16] Winberg, J.O., Hovik, R., McKinley-McKee, J.S., Juan, E. and Gonzàlez-Duarte, R. (1986) Biochem. J. 235, 481-490.

[17] Albalat,R. and Gonzàlez-Duarte,R. (1990) Nucl. Acids Res. 18, 6706.

[18] Albalat,R. and Gonzàlez-Duarte,R. (1993) Gene 126, 171-178.

[19] Sun, H. and Plapp, B. (1992) J. Mol. Evol. 34, 522-535.

[20] Thatcher, D.R. (1980) Biochem. J. 187, 875-886.

[21] Jowett,T. (1986) in: Drosophila. A Practical Approach. (Roberts, D.B., ed.), pp.275-286. IRL Press, Oxford, U.K.

[22] Visa, N., Fibla, J., Santa-Cruz, M.C. and Gonzàlez-Duarte, R. (1992) J. Histochem. Cytochem. 40, 39-49.

[23] Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. 12, 387-395.

[24] Winberg, J.O. and McKinley-McKee, J.S. (1992) Int. J. Biochem. 24, 169-181.

[25] Levitt, M. (1976) J. Mol. Biol. 104, 59-107.

[26] DeSalle, R. (1992) Molec. Phylogenetics and Evolution 1, 31-40.

TABLE I: Primers for the Reverse Transcriptase and PCR reactions

OLIGO	MUTATION	SEQUENCE	LENGTH 1st PCR	MUTATION CODON
SH1	--	CC*GAATTC***ATG**GATTTGACCAACAAG[a] (upstream)	--	--
SH2	Ala^{13} -> Gly	ACCGCCCAGACCGGCAACGA[b] (downstream)	54 nt	GCT -> GGT
SH3	Asn^{56} -> Thr	GGAAGGTGATGGTCACCTTG[c] (downstream)	182 nt	AAC -> ACC
SH5	--	*GGAAGCTTGC***TTA**GATGTGCGAGCT[d] (downstream)	--	--

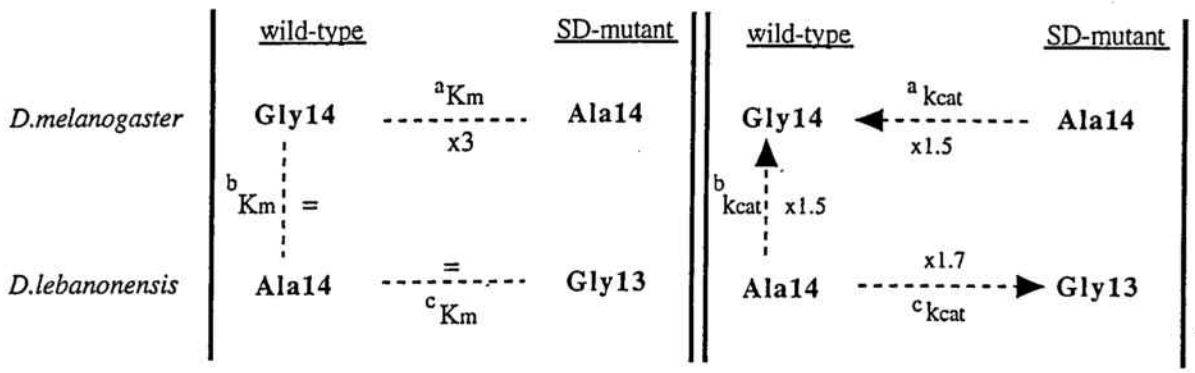
[a,d] Start and Stop codons are shown in bold. Sequences used to generate flanking restriction sites are in italics.

[b,c] Mutated nucleotides are in bold.

TABLE II: K_m and k_{cat} values for wild-type and mutant enzymes.

	Km NAD^+ (mM)	K_m propan-2-ol (mM)	k_{cat} (s^{-1})	$k_{cat}/K_{m\ NAD^+}$ ($s^{-1}mM^{-1}$)	$k_{cat}/K_{m\ prop}$ ($s^{-1}mM^{-1}$)
WT enzyme	0.23 ± 0.02	2.41 ± 0.6	1.74 ± 0.02	7.57	0.72
$Ala^{13}Gly$ mutant	0.24 ± 0.02	2.57 ± 0.3	2.92 ± 0.03	12.17	1.14
$Asn^{56}Thr$ mutant	0.20 ± 0.02	2.50 ± 0.4	2.31 ± 0.03	11.55	0.92

TABLE III: Relationship between $K_m^{NAD^+}$ and k_{cat} values of wild-type and mutants of *D.melanogaster* and *D.lebanonensis*.



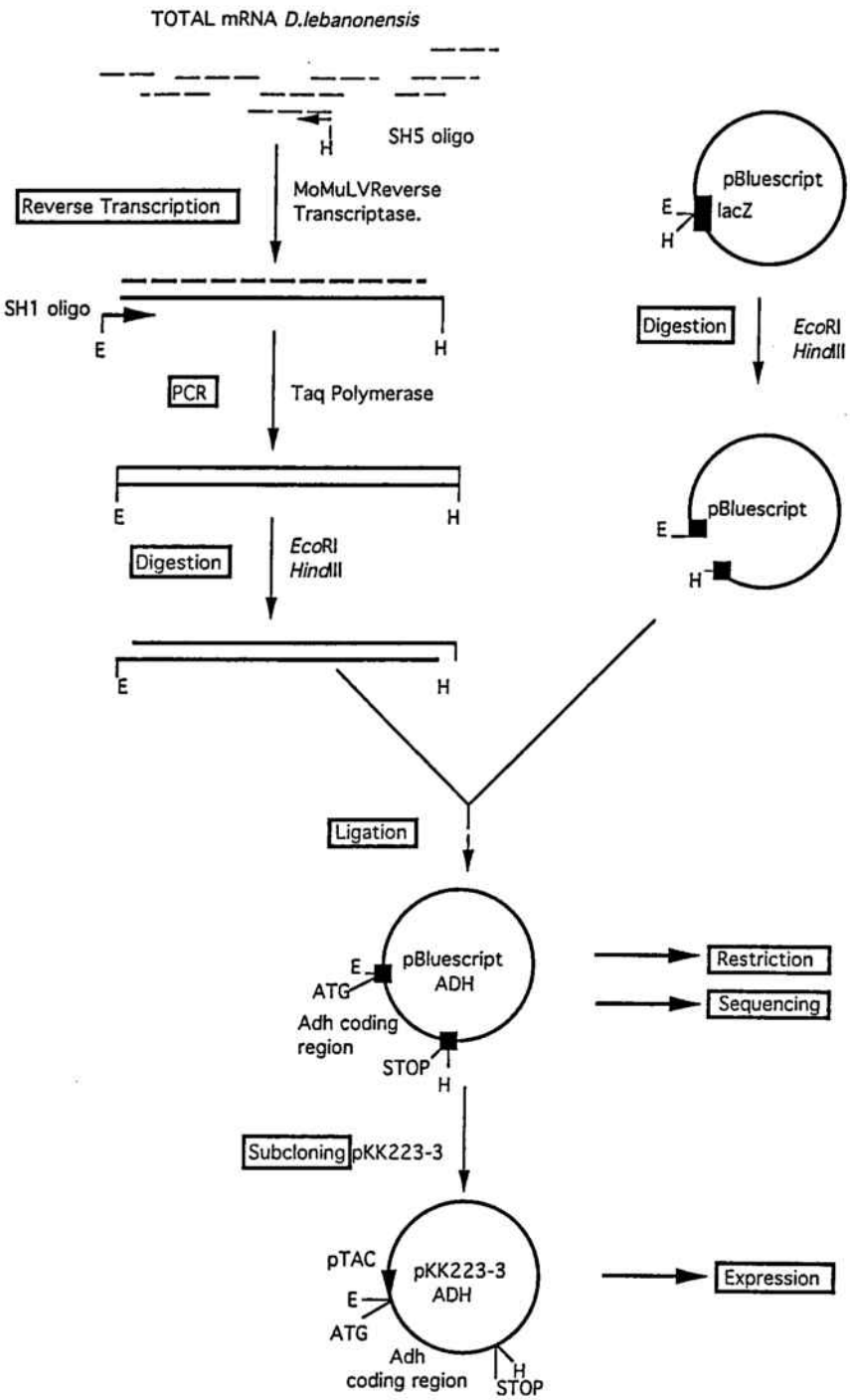
Data are: (a) from [24], (b) from [7] and (c) from our results.

LEGENDS TO THE FIGURES

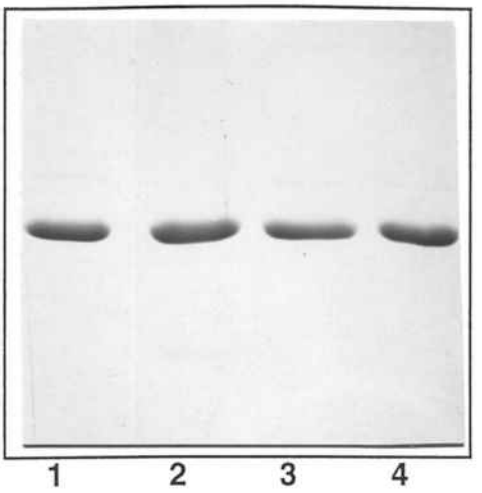
Fig. 1. Design of the expression plasmids for *D.lebanonensis* ADH. Reverse-PCR cDNA was digested with *Eco*RI and *Hind*III endonucleases and then cloned in Bluescript plasmid . Recombinant plasmids were identified by the IPTG/Xgal system. Adh-pBluescript DNA was digested and the Adh-*Eco*RI/*Hind*III fragment was subcloned in plasmid pKK223-3 for Adh expression under the pTAC promoter. E, *Eco*RI, H, *Hind*III. SH1 and SH5 are the oligonucleotides used in Reverse-PCR reactions (see TABLE I).

Fig. 2. (A) Coomassie blue-stained SDS-PAGE of purified *Drosophila* ADH and mutants. (1) 10 μg of wild-type ADH from adult flies; 10μl of recombinant ADH purified from *E.coli* JM105 transformed with (2) pKK223-ADH wild-type; (3) pKK223-Ala^{13}Gly mutant and (4) pKK223-Asn^{56}Thr mutant. (B) Western blot analysis of purified ADH using mAb LLBE8 against *D.lebanonensis* ADH. Lanes correspond to the same samples as displayed in (A).

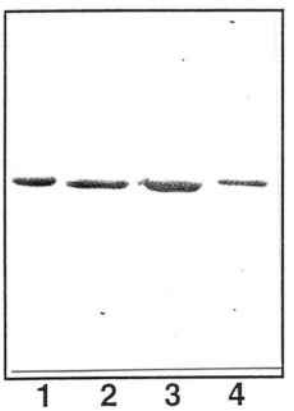
Fig. 3 (A) Effect of pH on ADH activity. Enzyme activity for wild-type and mutants was assayed under four pH's. Relative ADH activity is shown as a percentage of the highest value obtained in each case. (•) wild-type; (o) Ala^{13}Gly; (+) Asn^{56}Thr . (B) Thermal denaturation of wild-type and mutant *Drosophila* ADH. Purified preparations were incubated at 40°C and aliquots were assayed for ADH activity. Relative ADH activity is shown as a percentage of the initial activity. (•) wild-type; (o) Ala^{13}Gly; (+) Asn^{56}Thr .

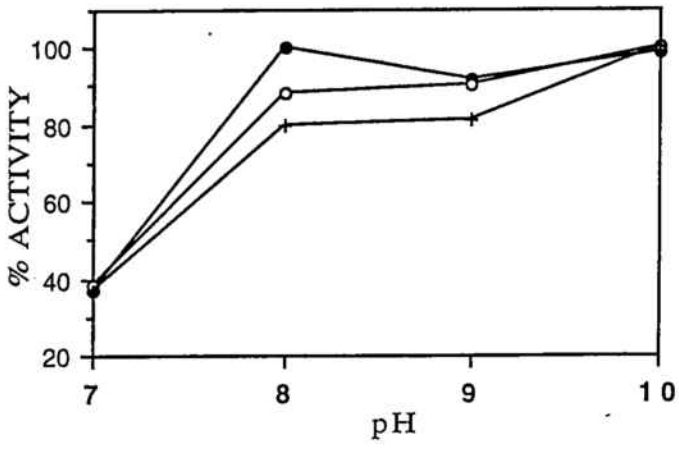

TOTAL mRNA D.lebanonensis
SH5 oligo
H
Reverse Transcription
MoMuLVReverse Transcriptase.
SH1 oligo
E
H
PCR
Taq Polymerase
Digestion
EcoRI
HindIII
pBluescript
lacZ
Digestion
EcoRI
HindIII
pBluescript
Ligation
pBluescript ADH
ATG
Adh coding region
STOP
Restriction
Sequencing
Subcloning pKK223-3
pKK223-3 ADH
pTAC
ATG
Adh coding region
STOP
Expression

A



B




% ACTIVITY
100
80
60
40
20
7
8
9
10
pH

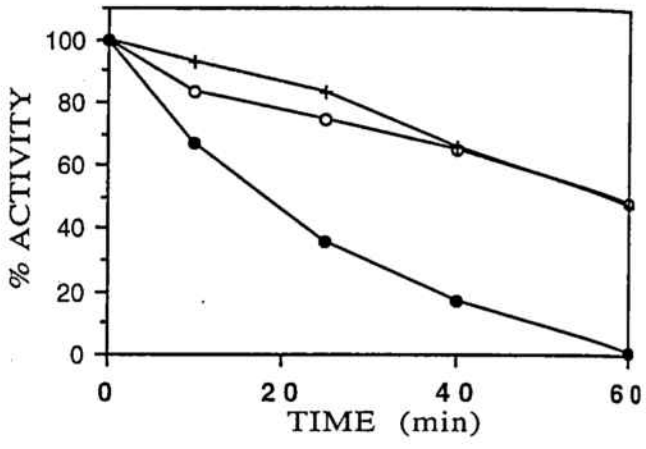

% ACTIVITY
100
80
60
40
20
0
0
20
40
60
TIME (min)

Fig. 3

Funció del domini C-terminal de l'ADH en l'activitat enzimàtica.

4.2.3. FUNCIÓ DEL DOMINI C-TERMINAL DE L'ADH EN L'ACTIVITAT ENZIMÀTICA.

La participació de l'extrem C-terminal de l'ADH en la reacció de catàlisi havia estat hipotetitzada a partir de les dades experimentals obtingudes per mutagènesi química, on l'eliminació dels once darrers aminoàcids de la proteïna produeix una pèrdua d'activitat, i de resultats amb sistemes de predicció d'estructura terciària amb els quals s'havia obtingut un model sobre el mecanisme de reacció en què, l'extrem carboxil formava un "loop" que estaria implicat en la transferència electrònica.

L'anàlisi de variants d'aquesta regió ens ha permés determinar els aminoàcids possiblement implicats en la funció enzimàtica. Això s'ha dut a terme mutagenitzant el gen Adh de *D.lebanonesis*, expressant els mutants a *E.coli* i posant a punt un sistema de purificació per FPLC de l'ADH recombinant. Així, s'han pogut establir les característiques cinètiques de les variants enzimàtiques sintetitzades.

4.2.4.1. *Mutagènesi dirigida de l'enzim ADH de* D.lebanonensis

Mitjançant PCR es varen obtenir un total de vuit mutants diferents de la regió compresa entre l'aminoàcid 243 i l'extrem C-terminal. Els productes d'amplificació varen ésser clonats a *E.coli* JM105 dins el plasmidi pKK223-3. Els oligonucleòtids utilitats per a la mutagènesi i els canvis introduïts en l'estructura primària de la proteïna es resumeixen en les Taules 1 i 2.

TAULA 1: Oligonucleòtids utilitzats i mutants realitzats.

OLIGO	MUTANT	Núm.	5' SEQÜÈNCIA[a)b)] 3'
SH1	---		*ccgaattc*ATGGATTTGACCAACAAG
TS3	Ile254 -> STOP	1	*ggaagcttctaga*CTAGTGCGAGTCCCAGTGCT
XD1	His249 -> Δ	2	*cccaagctt*CTAGATGTGCGAGTCCCACTTGGTCC
XD2	His249 -> STOP	3	*cccaagcTT*ACTTGGTCCACTCAATGG
VA1	Lys248 -> Ile (STOP250)	4	*cccaagcTT*AGATGGTCCACTCAATGGCCT
VA2	Trp246 -> Ala (STOP250)	5	*cccaagcTT*ACTTGGTCGCCTCAATGGCCT
LZ1	Glu245 -> STOP	6	*cccaagcTT*AAATGGCCTCAAGGGTGCC
LZ2	Ile244 -> STOP	7	*cccaagcTT*AGGCCTCAAGGGTGCCCAG
NF1	Ala243 -> STOP	8	*cccaagcTT*ACTCAAGGGTGCCCAGGTCC

a) En cursiva, les dianes del clonatge

b) En majúscules, la zona de l'oligonucleòtid homòleg a l'extrem 3' del gen Adh

TAULA 2: Seqüència de l'extrem C-terminal de l'enzim salvatge i dels mutants.

Núm.	SEQÜÈNCIA[a)]
ADH-wt	NGAIWKLDLG TLEAIEWTKH WDSHI*[b)]
1	NGAIWKLDLG TLEAIEWTKH WDSH*
2	NGAIWKLDLG TLEAIEWTKW DSHI*
3	NGAIWKLDLG TLEAIEWTK*
4	NGAIWKLDLG TLEAIEWTI*
5	NGAIWKLDLG TLEAIEATK*
6	NGAIWKLDLG TLEAI*
7	NGAIWKLDLG TLEA*
8	NGAIWKLDLG TLE*

a) El primer aminoàcid correspon a l'Asn230 de l'ADH de *D.lebanonensis*.

b) * = STOP

4.2.4.1. *Expressió i purificació de les proteïnes:*

Després d'induir l'expressió de cadascun dels mutants, es va procedir a la sonicació del cultiu bacterià i a l'obtenció de l'extracte proteïc cru. L'ADH es va purificar utilitzant un sistema de cromatografia FPLC amb una columna d'afinitat (Blue-Sepharosa) i una d'exclusió (Superosa 12). La Figura 1 mostra l'anàlisi de la fracció que presentava activitat alcoholdeshidrogenàsica per a cadascun dels mutants i que ha estat utilitzada en els estudis cinètics subsegüents.

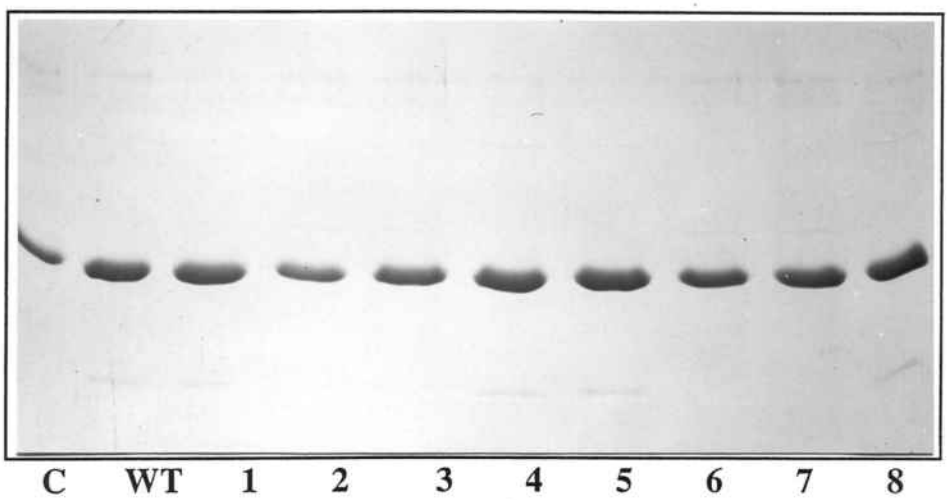


Figura 1: SDS-PAGE de la fracció amb activitat ADH de cadascun dels mutants tenyit per proteïna total amb Comassie blue: C: ADH control purificada a partir d'homogenats d'adults de *Drosophila*; WT: ADH salvatge recombinant obtinguda a partir de l'expressió del gen Adh de *D.lebanonensis* a *E.coli;* (1) fins a (8): ADH dels mutants 1 fins a 8, segons la Taula 1.

Les proteïnes purificades varen ser electrotransferides a membranes de nitrocel.lulosa per ser posteriorment immunodetectades amb dos anticossos monoclonals diferents anti-ADH de *Drosophila*: LLBE8 i

MMBB8 (Figura 2).Aquest anticossos reconeixen dos epítops diferents. L'epítop α, presuntament localitzat en l'extrem C-terminal de la molècula és reconegut per l'anticós LLBE8. L'epítop β, identificat a partir de pèptids sintètics, es troba al voltant de l'aminoàcid 192 (aminoàcids 189 a 199) i és reconegut per l'anticós MMBB8. L'epítop α és conformacional mentre que l'epítop β depén de l'estructura primària.

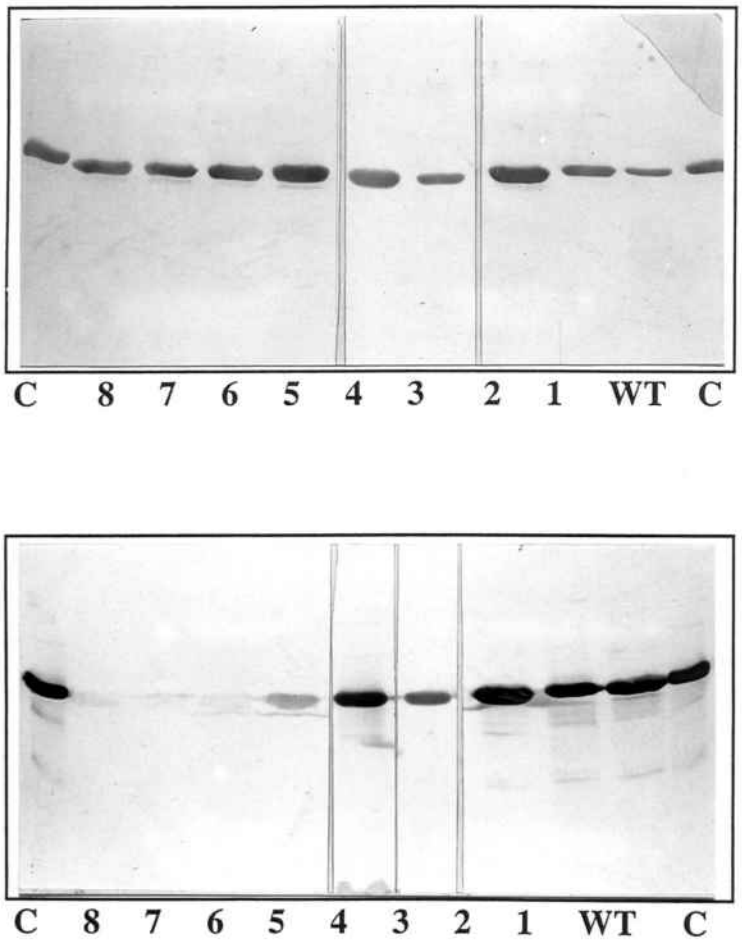


Figura 2: Anàlisi Western de l'ADH de *D.lebanonensis*. C: ADH control purificada a partir d'homogenats de mosca; WT: ADH salvatge recombinant expressada a *E.coli*; (1) Mutant 2; (2) Mutant 3; (3) Mutant 4; (4) Mutant 1; (5) Mutant 5; (6) Mutant 6; (7) Mutant 7; (8) Mutant 8, (Taula 1). Western blot amb l'anticós MMBB8 (A) i l'anticós LLBE8 (B).

4.2.4.3. *Activitat enzimàtica i constants cinètiques dels mutants*

L'activitat de l'enzim salvatge i dels mutants va ésser mesurada espectrofotomètricament seguint l'increment d'absorbància a 340 nm utilitzant propan-2-ol com a substracte i NAD com a cofactor. També es van determinar les Km per a l'NAD i el propan-2-ol i les kcat per a cadascun dels mutants (Taula 3)

TAULA 3: Km i kcat per a l'enzim salvatge i per als mutants

	MUTANT	Km NAD (mM)	xWT[a)]	Km propan-2-ol (mM)	x WT	Kcat (s^{-1})	x WT
	WT	0.23 ±0.02	---	2.41 ±0.40	---	1.74 ±0.02	---
1	Ile 254->STOP	0.39 ±0.03	1.7	11.94 ±0.52	4.9	2.06 ±0.07	1.3
2	His250->Δ	0.23 ±0.03	1.0	11.27 ±1.35	4.7	2.02 ±0.08	1.3
3	His250->STOP	0.31 ±0.03	1.3	13.40 ±1.20	5.6	2.52 ±0.15	1.5
4	Lys249->Ile (STOP250)	0.31 ±0.04	1.4	11.52 ±0.65	4.8	1.62 ±0.03	1.0
5	Trp247->Ala (STOP250)	0.38 ±0.04	1.7	12.60 ±0.83	5.2	1.50 ±0.18	1.0
6	Glu246->STOP	0.36 ±0.03	1.6	11.97 ±1.05	5.0	1.72 ±0.14	1.1
7	Ile245->STOP	0.80 ±0.06	3.5	30.35 ±1.57	12.6	1.74 ±0.04	1.0
8	Ala244->STOP	1.64 ±0.19	7.3	32.68 ±1.30	13.6	1.42 ±0.03	0.9

a) Factor multiplicatiu respecte al valor de l'enzim normal

Aquests resultats indiquen que els mutants es poden classificar en dos grups diferenciats. El primer inclou els mutants 1 fins al 6 i presenta, com a aspecte més significatiu, un increment de x5 la Km pel propan-2-ol respecte a l'enzim salvatge. En el segon -mutants 7 i 8- aquest increment seria superior a deu vegades el valor de la Km de referència.
L'especificitat per a diferents alcohols ha estat determinada respecte al propan-2-ol a 10mM de concentració. Els alcohols utilitzats com a substractes varen ser: etanol, propan-1-ol, butan-1-ol, penta-1-ol i

hexan-1-ol, com a alcohols primaris i propan-2-ol, butan-2-ol, 2-penten-3-ol i ciclohexanol, com a alcohols secundaris (Taula 4).

TAULA 4: Activitat[a)] dels enzims salvatge i mutants amb diferents alcohols

ALCOHOL	WT	1	2	3	4	5	6	7	8
ETANOL	28	3	4	23	4	33	14	nd[b)]	nd
PROPAN-1-OL	36	11	8	13	12	13	15	nd	nd
PROPAN-2-OL	100	100	100	100	100	100	100	100	100
BUTAN-1-OL	53	19	14	13	15	20	11	11	13
BUTAN-2-OL	105	145	195	170	178	247	174	257	261
PENTAN-1-OL	62	49	51	47	63	60	50	30	56
2-PENTEN-3-OL	105	154	207	201	186	290	233	330	321
HEXAN-1-OL	26	33	21	29	34	43	32	nd	nd
CICLOHEXANOL	93	127	148	131	130	153	144	163	136

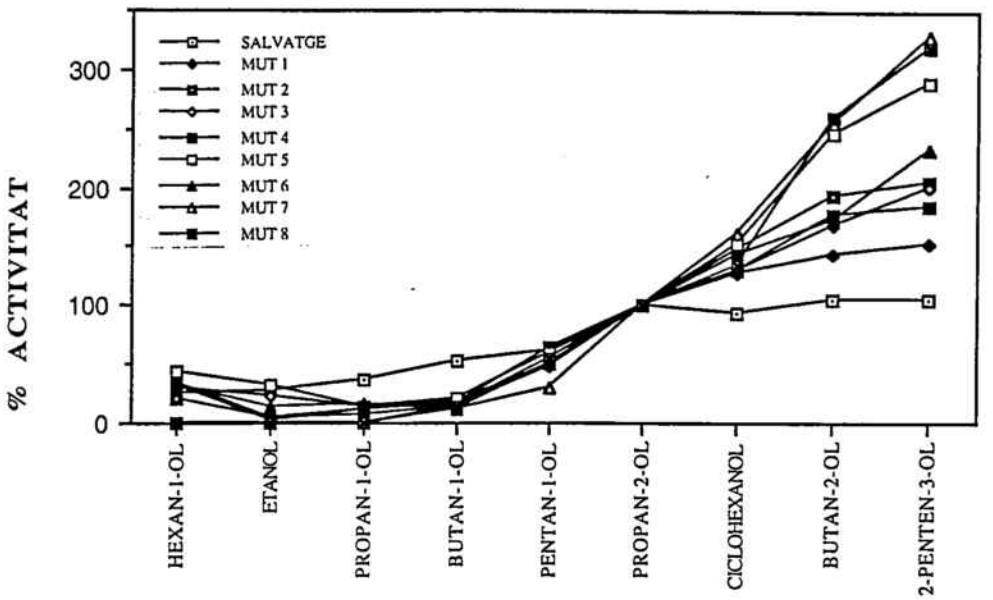
a) L'activitat ve donada com a % respecte el valor d'activitat per al propan-2-ol (100%) en cada cas.

b) nd: no detectable

Els resultats obtinguts indicaven que els alcohols secundaris són millors substractes que els primaris tant per a l'enzim salvatge com per als mutants, i que aquesta diferència s'accentuava en els enzims mutats. Així, per a l'enzim normal les diferències d'activitat entre els alcohols secundaris no són mai superiors al 15%, mentre que per als mutants superen en alguns casos el 200%. A més, semblava que hi havia una relació directa entre l'augment d'activitat i el tamany de l'acohol oxidat. Per als alcohols primaris, l'enzim salvatge presentava activitat en tots els casos, mentre que per a alguns mutants no es detectava activitat en

front de l'etanol, propan-1-ol i hexan-1-ol a la concentració d'alcohol utilitzada (Figura 3a i 3b).

a)



b)

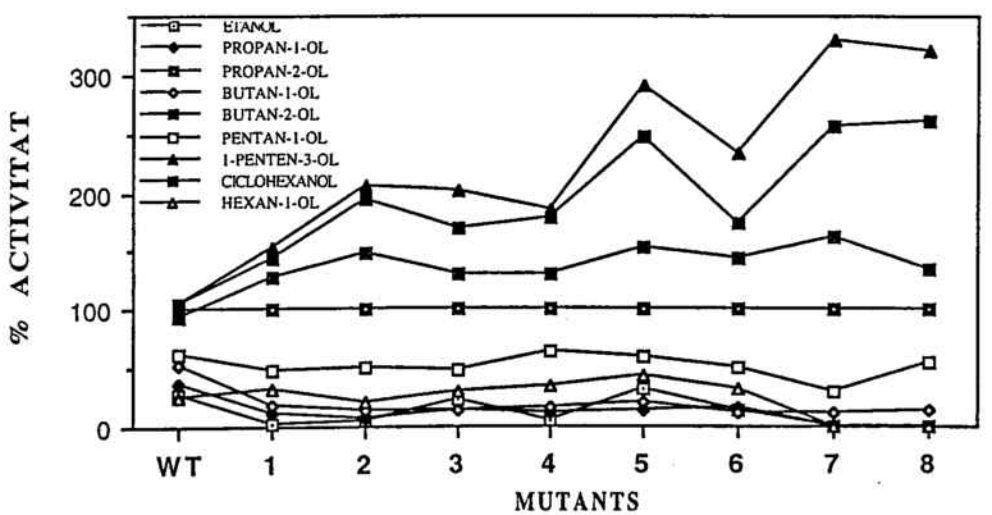


Figura 3: a) % d'activitat relativa al propan-2-ol per a cadascun dels mutants respecte als diferents alcohols. b) % d'activitat relativa al propan-2-ol per cadascun dels alcohols respecta als mutants

Aquestes diferències no eran iguals per a tots el mutants estudiats i, en general, s'accentuaven més per a aquells que havien perdut un major nombre d'aminoàcids de l'extrem C-terminal (Figura 4).

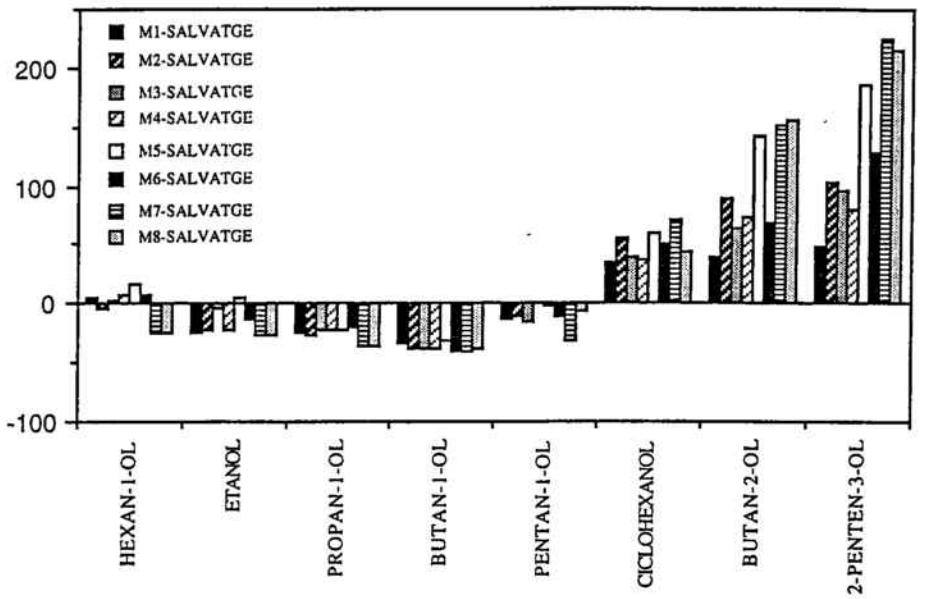


Figura 4: Diferències de les activitats relatives al propan-2-ol entre els diferents mutants i l'enzim normal respecte a cadascun dels alcohols.

5. DISCUSIÓ

5. DISCUSSIÓ

Mitjançant el present treball hem intentat aportar, des d'una perspectiva molecular, noves dades del sistema ADH de *Drosophila*. Per això, ens hem proposat analitzar la regió genòmica que compren el gen Adh a *D.lebanonensis* i a *D.immigrans* i determinar els aminoàcids importants per l'activitat enzimàtica a *D.lebanonensis*.

5.1. EL GEN Adh A *D.lebanonensis* I *D.immigrans*.

La informació genètica i bioquímica respecte el gen Adh no és comparable entre les diferents espècies del gènere *Drosophila*. Així, subgèneres com *Sophophora* i *Drosophila* han estat abastament analitzats (Sullivan et al., 1990), mentre que d'altres són poc coneguts. A més, el nombre d'espècies analitzades dins d'un subgènere no és sempre representatiu, fet que és important per obtenir, dins d'una perspectiva àmplia, conclusions generals. *D.lebanonensis* és, dins d'aquest contexte, una espècie interessant per la seva gran adaptació a medis rics en alcohol, superant inclús a l'espècie descrita com a model d'adaptació a aquest medi, *D.melanogaster*. La proteïna ADH ha estat seqüenciada (Vilarroya et al., 1989). La comparació de l'estructura d'aquesta ADH amb les altres conegudes de *Drosophila* i amb diversos membres d'una família estructuralment homòloga, la de les deshidrogenases de cadena curta (short-chain dehydrogenases), ha permés posar de manifest substitucions d'aminoàcids en posicions molt conservades i d'interés funcional.

D.immigrans, en canvi, no es troba generalment present en ambients rics amb alcohol, encara que l'enzim ADH, ben conegut a nivell bioquímic, no difereix substancialment del d'altres espècies. Aquesta espècie pertany al subgènere *Drosophila* que agrupa un nombre d'espècies elevat i no uniformament analitzat.

Aquestes són essencialment les raons per les quals ambdues espècies varen ser triades per dur a terme aquest treball.

Per aïllar la regió genòmica que compren el gen Adh d'ambdues espècies es va procedir al crivellatge de la genoteca corresponent i a l'anàlisi de restricció del fragment de DNA seleccionat amb set enzims de restricció. La comparació dels mapes de restricció de les nostres espècies amb d'altres del subgènere *Sophophora* mostrà clarament que les dianes no es mantenien. Aquesta divergència s'explica fàcilment donat que aquesta metodologia permet analitzar menys d'un 2% de les posicions nucleotídiques totals.

La regió aïllada, 7 Kb, comprenia, a més del gen Adh, una seqüència flanquejant a l'extrem 3', altament conservada, amb una pauta de lectura oberta (3'ORF), i homòloga al gen Adh (Adh-dup) (Schaeffer i Aquadro, 1987). Aquesta seqüència havia estat descrita únicament en espècies del subgènere *Sophophora* (Cohn i Moore, 1988; Kreitman i Hudson, 1991; Marfany i Gonzàlez-Duarte, 1991b; 1992a, 1993) i, per tant, la seva existència en espècies d'altres subgèneres, *Scaptodrosophila* (*D.lebanonensis*) i *Drosophila*, (*D.immigrans*) indicava que l'hipotetitzada duplicació que originà l'Adh i l'Adh-dup va tenir lloc possiblement abans de l'aparició del gènere.

Es varen seqüenciar 3.828 pb de *D.lebanonensis* i 3.142 pb de *D.immigrans*, fragments que comprenien les regions reguladores i codificants del gen Adh i Adh-dup. L'anàlisi de les seqüències va posar de manifest que l'estructura i l'organització del gen Adh en aquestes espècies era comparable a la del grup *melanogaster*. Així, trobàrem un únic gen, compost de 3 exons de 93, 405 i 267 nucleòtids cadascun (Goldberg, 1980; Benyajati et al., 1980). L'única diferència amb el gen Adh de *D.melanogaster* era la longitud del primer exó, 6 nucleòtids més curt, corresponent als aminoàcids 2n. i 3r. de la proteïna.

Pel que fa a la regió reguladora, ambdues espècies presentaven dos promotors amb situació relativa equivalent als promotors proximal i distal descrits en altres espècies del gènere. No obstant, a *D.immigrans* la distància entre les dues caixes TATA era apreciablement inferior que en d'altres espècies, únicament 559 nucleòtids. A *D.lebanonensis* era de 1.114 nucleòtids i la caixa TATA proximal presentava un canvi inesperat en el primer nucleòtid: T->C. Aquesta substitució, encara que inusual, ja havia estat descrita en diferents gens d'altres organismes (Breathnach i Chambon, 1987). Una anàlisi més minuciosa de la regió reguladora permetia identificar altres possibles regions d'unió de diferents factors de transcripció descrits en d'altres espècies del gènere.

No obstant, la comparació de les regions codificants era la que aportava més informació. Era possible alinear un gran nombre de posicions nucleotídiques i quantificar el nombre de canvis que s'havien produït al llarg de l'evolució per determinar les relacions filogenètiques entre elles. Aquest tipus d'anàlisi podia fer-se qualitativament o quantitativament. Ambdues metodologies es basen en que l'evolució

molecular és un procés divergent. Qualitativament, ens permetia reconstruir les relacions filogenètiques entre les espècies. Quantitativament, aportava informació sobre les taxes d'evolució i el temps dels processos evolutius, sota la hipòtesi que les macromolècules es comporten com rellotges moleculars (Zurckerkandl i Pauling, 1962, 1965).

Així, la comparació interespecífica de les regions codificants permetia calcular el nombre de substitucions en llocs de reemplaçament (Ka) i en llocs silenciosos (Ks). Cal distingir aquestes dues categories ja que la dinàmica evolutiva és significativament desigual (revisió a Li i Graur, 1991). Així, encara que un 22%-23% dels llocs comparats entre *D.lebanonensis, D.immigrans* i *D.subobscura* (utilitzada com a referència del subgènere *Sophophora*) eren diferents, les posicions silencioses representaven el 63% de les substitucions totals (tan sols un 25% de posicions eren silencioses), mentre que les de reemplaçament acumulaven el 37% restant (75% dels nucleòtids). A més a més, la distribució dels canvis tampoc no era la mateixa per als dos tipus de nucleòtids. Per a les posicions sinònimes la distribució era a l'atzar, mentre que per a les de reemplaçament es trobava un patró exó-depenent: l'Adh presentava un nombre de canvis en el segon exó significativament inferior a l'esperat. Aquest resultat podia relacionar-se amb alguns aspectes funcionals de la proteïna ja que, dades bioquímiques indicaven que el centre catalític de l'enzim es trobava en la regió codificada pel segon exó (Cols et al., 1993; Chen et al., 1993).

Amb els valors de Ka i Ks obtinguts i a partir de les matrius de distàncies entre les espècies comparades, es podien determinar les relacions

filogenètiques. A partir dels resultats s'observà que els tres subgèneres estudiats (*Scaptodrosophila Drosophila* i *Sophophora*) havien divergit gairebé al mateix temps dins la història evolutiva del gènere, trobant-se *D.lebanonensis* més a prop del subgènere *Drosophila* o *Sophophora* segons fóssin considerats els valors Ka o Ks respectivament. Al mateix temps, *D.immigrans* també variava de posició respecte *D.affinidisjuncta* depenent dels valors utilitzats. Aquesta anàlisi es va fer amb mètodes diferents: el càlcul del nombre de substitucions segons Li et al., 1985, Perler i Efstratiadis, 1980 i Kimura, 1980, i per construir l'arbre filogenètic seguint Fitch i Margoliash, 1967 i Saitou i Nei, 1987. És interessant remarcar que malgrat les petites diferències existents principalment en la longitud d'alguna de les branques de l'arbre, l'estructura general i la distribució de les espècies era essencialment la mateixa.

Més dificultats trobàrem a l'intentar realitzar una anàlisi quantitativa. Per calcular el temps de divergència entre les diferents espècies de *Drosophila* vàrem utilitzar com a taxa de substitució el valor de $5,5 \times 10^{-9}$ substitucions/lloc sinònim any (Miyata et al., 1980; Hayashida i Miyata, 1983; Atkinson et al., 1988). El temps de divergència entre els tres subgèneres era de 52 - 53 milions d'anys, valors concordants amb les prediccions fetes a partir de les dades biogeogràfiques i immunològiques, que dataven la separació d'aquests gèneres ara fa 40 - 60 milions d'anys (Throckmorton, 1975; Beverly i Wilson, 1982, 1984). No obstant, surgien dificultats si consideràvem altres taxes de substitució descrites a *Drosophila*. (Sharp i Li, 1989; Moriyama i Gojobori, 1992) En aquests casos, obteníem valors fins a cinc vegades superiors i, per tant,

en contradicció amb les prediccions. Això implicava que possiblement no totes les seqüències eren útils per descriure la història evolutiva de les espècies i no podien ser indiscriminadament utilitzades per a la taxonomia molecular o per datar processos evolutius (Saccone et al., 1993).

Vàrem realitzar una anàlisi similar amb el gen Adh-dup. La primera premisa necessària per fer aquest estudi era que l'Adh-dup es transcrigués. Diferents experiments de Northern fets en el nostre laboratori havien estat infructuosos en la detecció del producte d'expressió d'aquest gen. Per això, vàrem dissenyar una retrotranscripció acoblada a PCR, mètode molt potent per detectar trànscrits, encara que aquests siguin molt poc abundants (Kawasaki, 1990). Així vàrem identificar trànscrits de l'Adh-dup en diferents estadis del desenvolupament: larva, pupa i adult. Dissortadament, la tècnica no permet predir l'extensió d'aquesta expressió. Malgrat això, hipotetitzem, d'acord amb d'altres autors, que l'Adh-dup és un gen que s'expressa poc i problablement a nivell tissular/temporal restringit.

En comparar l'Adh-dup de *D.lebanonensis, D.immigrans* i *D.subobscura* observàrem que, en tots els casos, es mantenia una mateixa organització: 3 exons separats per 2 introns d'uns tamanys similars als de l'Adh excepte el primer intró i el tercer exó. El primer intró presentava una longitud sempre superior a 250 nucleòtids (60 nucleòtids en l'Adh) excepte a *D.immigrans* que té 59 nucleòtids. El tercer exó presentava tamanys variables segons les espècies: 309 nucleòtids a *D.lebanonensis*, 321 a *D.immigrans* i 339 a *D.subobscura*.

En comparar la regió codificant d'aquest gen, el 23%-24% dels nucleòtids havien canviat. També les posicions silencioses acumulaven un nombre de canvis superior respecte a les posicions de reemplaçament. La distribució per exons dels canvis silenciosos era a l'atzar, mentre que els de reemplaçament s'acumulaven en el tercer exó.

A partir dels valors de les Ka i les Ks per a l'Adh-dup es podrien construir els arbres filogenètics però, donat que el nombre d'espècies seqüenciades era relativament baix i esbiaixat, els resultats serien poc informatius. No obstant, vàrem analitzar el patró evolutiu de l'Adh-dup i vàrem observar que la relació Ks/Ka, era significativament superior per l'Adh-dup que per a l'Adh. Per tant, la selecció a que es troben sotmesos aquests dos gens, evolutivament emparentats i adjacents en el genoma, és clarament diferent.

5.2. LA PROTEÏNA ADH.

La determinació de l'estructura primària de l'ADH de *D.melanogaster* (Schwartz i Jörnvall, 1976; Thatcher, 1980) va possar de manifest la manca d'homologia amb les alcoholdeshidrogenases de llevat i de fetge de cavall, les quals estaven formades per cadenes polipeptídiques més llargues i per les que la presència d'ions metàl.lics era bàsica (Eklund et al., 1976; Jörnvall, 1970, 1977a). En la predicció de l'estructura secundària apareixia una estructura alternant de fulls ß i hèlixs α (Thatcher i Sawyer, 1980), típica dels enzims que utilitzen NAD o NADP com cofactors (Rossmann et al., 1974). A més, a diferència de les deshidrogenases descrites, el domini d'unió nucleotídic es trobava a l'extrem N-terminal de la proteïna (Eklund et al., 1976).

De 1980 ençà, moltes ADH de *Drosophila* han estat seqüenciades i la comparació de totes les estructures ha aportat una informació molt valuosa. L'alineament de seqüències de proteïnes ha desvetllat relacions que no eren obvies inicialment. Els residus funcionals, implicats en la catàlisi, es conserven dins d'una mateixa família d'enzims (McLachlan, 1971), així com els residus hidrofòbics interns o els de propietats estereoquímiques especials (Gly, Pro i Cys) (Smith i Margoliash, 1964; Jörnvall, 1977b; Taylor, 1986). Per tant, l'estudi de la funció dels residus conservats en un representant de la família aporta informació de la resta del grup. Un cop s'hipotetitzen els aminoàcids funcionals d'un enzim, la confirmació de la seva funcionalitat s'assoleix per mutagènesi dirigida, modificació química i determinació de l'estructura terciària.

La comparació de l'enzim ADH en 36 espècies de *Drosophila* va demostrar que el 57,13% de la proteïna es trobava conservada (145 dels 253 aminoàcids). A més, la distribució dels canvis no era aleatòria al llarg de la molècula ja que hi havien determinades zones que presentaven un nombre de canvis significativament inferior a d'altres regions. Aquest grau de conservació podia correlacionar-se amb diferents aspectes funcionals: la unió del cofactor a l'enzim, el manteniment de l'estructura espaial, el centre catalític de l'enzim, etc. D'aquesta manera s'han pogut establir set regions o *boxes* dins l'ADH a les quals varen assignar-se funcions diferents.

La box1 (aa11 fins a aa23) presentava un 84% de conservació i formava part del domini d'unió del cofactor. Les glicines incloses en aquesta regió corresponien a les descrites en les regions anàlogues de moltes altres deshidrogenases.

Per les seves característiques, la box2 (aa87 fins a aa108) i la box3 (aa117 fins a aa132), amb un 75% i un 81% de conservació, respectivament, participarien en el manteniment de la conformació de la proteïna activa. En aquestes regions es trobaven Asp87, Gly94, Asn114 i Gly132, aminoàcids conservats també en totes les deshidrogenases de cadena curta.

La box4 (aa139 fins a aa158) presentava un 75% de conservació. Aquesta regió contenia la Tyr152 i la Lys156, aminoàcids implicats en el centre actiu de l'enzim.

La box5 (aa181 fins a aa201) i la box6 (aa221 fins a aa239) presentaven un 77% i un 74% de conservació, respectivament. Aquestes regions es trobaven separades per un segment hipervariable on els canvis aminoacídics han estat abundants. Encara es desconeix la funció d'aquestes regions.

Finalment, la box7 (aa250 fins a aa255) presentava un 83% de conservació. L'extrem C-terminal de la molècula podria participar en la reacció d'oxidació de l'alcohol, o, si més no, tenir alguna funció conformacional important. El fet que digestions químiques controlades d'aquesta zona haguessin produït la pèrdua d'activitat enzimàtica recolçaria aquesta hipòtesi (Krook et al., 1992).

En analitzar els canvis d'aminoàcids produïts en les nostres espècies, *D.lebanonensis* i *D.immigrans*, i *D.subobscura* (com a referència del subgènere *Sophophora*) vàrem observar que la majoria eren conservatius (més del 50%) i distribuits en determinades posicions. Les seqüències proteïques de *D.lebanonensis* i *D.immigrans* presentaven per

primera vegada canvis en dotze noves posicions. No obstant, el manteniment dels perfils d'hidrofobicitat era una primera indicació de que l'estructura terciària de la proteïna es conservava en totes les espècies.

De la comparació més general, amb les 36 espècies de *Drosophila*, vàrem deduir que respecte les 110 posicions variables, en 77 ocasions el canvi es feia per únicament un altre amino àcid (dues possibilitats), en 24 ocasions, per dos (tres possibilitats), en 8 ocasions, per tres (quatre possibilitats) i en 1 ocasió, per quatre (cinc possibilitats). Aquests valors s'ajustaven a una distribució de Poisson. Cal doncs considerar que la substitució d'aminoàcids és un succés poc freqüent al llarg de l'evolució.

5.3. MUTAGÈNESI DIRIGIDA.

Una bona metodologia per l'anàlisi d'un sistema consisteix en variar sistemàticament algun del seus components i observar-ne els efectes. Per això, una aproximació a l'anàlisi funcional de l'ADH va ser, mitjançant mutagènesi dirigida, substituir els residus conservats i comparar el comportament de l'enzim salvatge i el del mutant. Tres regions presentaven un interés especial: la del centre actiu, la regió d'unió del cofactor i la de l'extrem C-terminal.

Inicialment vam utilitzar una construcció que ja existia del gen Adh de *D,melanogaster* (Atrian et al., 1990) que havia estat clonada en el vector p3008 (Shen et al., 1989). La mutagènesis es va dur a terme sobre aquesta construcció i vàrem reclonar el producte mutat en el plasmidi pVT-U dins l'hoste *S.cerevisiae* WV201-36.

El sistema d'expressió en llevat presentava una limitació important, el baix rendiment que s'obtenia quan la proteïna expressada no tenia activitat. La soca de *S.cerevisiae* era ADH-negativa i, per tant, quan es transformava amb l'Adh salvatge de *Drosophila* augmentava extraordinàriament la seva eficàcia. En canvi, quan es transformava amb el plasmidi pVT-U intacte o amb un mutant sense activitat ADH, presentava una viabilitat molt reduïda. Per això vàrem decidir canviar el sistema d'expressió, utilitzar com a hoste *E.coli* i expressar l'Adh de *D.lebanonensis*. El canvi d'espècie era degut a que: 1) *D.lebanonensis* presentava substitucions d'aminoàcids en posicions molt conservades en les ADH de *Drosophila* i en les deshidrogenases de cadena curta, i 2) els primers intents de cristalització s'havien dut a terme amb l'ADH de *D.lebanonensis* (Gordon et al., 1992). Vàrem procedir a la construcció del gen Adh de *D.lebanonensis* sense introns i a la seva expressió en el vector pKK223-3.

Els aminoàcids que vàrem substituir es trobaven en posicions ja hipotetitzades com a rellevants per al funcionament de l'enzim. Així, inicialment vàrem dirigir el nostra interés al centre actiu. Un candidat per ocupar aquest lloc, participant en la transferència de protons i per tant en l'activitat enzimàtica, era la tirosina 152. El fet que la Tyr152 estigués conservada en totes les ADH, conjuntament amb la 15-hidroxiprostaglandina deshidrogenasa (Ensor i Tai, 1991), suggerien que aquest residu jugués un paper important en l'interacció enzim-substracte, probablement anàleg a la de l'ió metàl.lic en les deshidrogenases de cadena mitja.

Els nostres resultats mostraren que el canvi de la Tyr152 per Phe produia la total inactivació de l'enzim. Aquesta pèrdua d'activitat no podia relacionar-se amb una alteració en la conformació de l'enzim, i havia de ser produïda per la participació directa del grup hidroxil de la cadena lateral de la Tyr152 en la ruta de la transferència electrònica, indispensable per a l'oxidació del substracte.

Resultats posteriors d'altres autors, substituint aquesta posició per Glu, Gln, Cys i His, han reafirmat el paper d'aquest amino àcid en el centre actiu de l'enzim (Cols et al., 1993; Chen et al., 1993).

La segona regió d'interès de l'enzim era el domini d'unió a l'NAD. Dos dels aminoàcids que la formen presenten, a *D.lebanonensis*, uns canvis únics, Ala13Gly i Asn56Thr, respecte a la seqüència consensus.

A més, l'ADH de *D.lebanonensis* presentava unes característiques cinètiques diferencials (Winberg i Mckinley-McKee, 1992). Així, el quocient kcat/Km per l'ADH de *D.lebanonensis* era menor que a *D.melanogaster*. Aquest fet estava en aparent contradicció amb l'elevada tolerància de *D.lebanonensis* davant d'hàbitats rics en alcohol. Aquesta adaptació, doncs, no era aparentment deguda a un enzim més eficaç, sinó que semblava que podia explicar-se per l'elevada quantitat de proteïna produïda per aquesta espècie.

Mitjançant mutagènesi dirigida sobre l'Adh de *D.lebanonensis* vàrem produir les substitucions Ala13->Gly i Asn56->Thr, restablint en ambdós casos l'aminoàcid més comú en les ADHs de *Drosophila*. El primer mutant si be no presentava variacions en les Km per a l'NAD i per al propan-2-ol, incrementava el valor de la kcat (x1,7) i la termoestabilitat.

Respecte a la posició 56, el mutant Asn56->Thr de *D.lebanonensis* presentava característiques cinètiques més semblants a les de l'enzim salvatge de *D.melanogaster*. Aquesta posició, situada en el tercer full β del domini d'unió de l'NAD (Jörnvall et al., 1981) es troba conservada en totes les alcoholdeshidrogenases de cadena curta del regne animal (Sun i Plapp, 1992) i podria estar relacionada amb la correcta acomodació del coenzim i del substracte, especialment dissenyat per alcohols alifàtics de petit tamany. L'enzim mutant presentava respecte al salvatge, major termoestabilitat, valors de kcat incrementats (x1,3) i Km comparables.

Una tercera regió interessant de la proteïna era l'extrem C-terminal. Tan sols a partir de l'anàlisi de l'activitat enzimàtica amb diferents alcohols, s'havia pogut determinar algunes de les característiques de la regió d'unió del substracte. Així, per exemple, la pèrdua d'activitat davant del metanol suggeria que la interecció entre el grup o grups alquil de l'alcohol i el lloc actiu de l'enzim era important per a l'activitat catalítica. A més, estudis amb enantiomorfs de diferents alcohols indicaven que existien dues regions hidrofòbiques diferents a les quals s'unirien els dos grups alquil dels alcohols secundaris (Winberg et al., 1982a).

Vàrem alterar la regió C-terminal per mutagénesi dirigida i vàrem constatar que es produia una pèrdua d'activitat enzimàtica al eliminar alguns aminoàcids d'aquesta zona. Els vuit mutants obtinguts podien agrupar-se en dues clases segons els grau de variació dels paràmetres cinètics. Les variacions més importants eren les que afectaven als valors de les Kms per al propan-2-ol. El primer grup de mutants (1 fins a 6) presentaven la Km augmentada aproximadament 5 vegades respecte al

valor per l'enzim salvatge. Aquest resultat explicava l'elevat grau de conservació observat en la longitud de totes les ADHs de *Drosophila*. Els resultats obtinguts amb els mutants 1 i 2 semblaven indicar que l'efecte observat era de tipus conformacional, lligat a la longitud d'aquesta regió, més que no pas a les característiques intrínseques dels residus. L'activitat que presentaven aquests mutants indicava a més que l'His250 no participava directament en la transferència electrònica. Possiblement, el "loop" C-terminal intervé en la unió enzim-substracte alhora que podria proporcionar l'ambient hidrofòbic adequat per a la reacció d'oxidació. A més, superat un llidar (mutants 7 i 8) es produeix un nou increment del valor de la Km, x10. El caràcter hidrofòbic de l'Ile245 que era eliminada en el mutant 7 podria explicar aquest resultat.

En comparació amb les importants variacions de les Km observades per al propan-2-ol, les altres constants cinètiques es mantenien molt més estables respecte a l'enzim salvatge. Així, les kcats dels mutants s'incrementaven lleugerament en els mutants 1, 2 i 3, i es mantenien essencialment iguals en la resta dels casos. Aquest efecte podria explicar-se si l'ADH seguís, d'acord amb hipòtesis d'altres autors, un mecanisme Theorell-Chance (Theorell i Chance, 1951) pels alcohols secundaris, on l'alliberament del cofactor reduït és el punt limitant en la velocitat de la reacció.

Les Kms dels mutants per a l'NAD també experimentaven diferents graus de variació respecte als dos grups de mutants. La variació era relativament petita en el primer grup, oscilant entre x1,3 i x1,7 vegades respecte a l'enzim salvatge. En el segon grup, l'augment era molt més significatiu: 3,5 vegades per al mutant 7 i 7,3 vegades per al mutant 8. Per

tant, encara que l'extrem C-terminal no intervingués directament en la unió amb el cofactor, l'alteració del mateix podria afectar la unió amb l'NAD.

L'especificitat en front diferents alcohols va ser determinada respecte al propan-2-ol a 10mM de concentració. Els resultats mostraven que per a l'enzim salvatge les diferències d'activitat entre els alcohols secundaris eren petites, mentre que els mutants presentaven diferències molts grans entre els diferents alcohols secundaris, que semblaven estar relacionades amb el tamany de l'alcohol i amb la zona que afectava la mutació (Figura 3). Resultats similars, però amb diferències menys clares, s'observaven també amb els alcohols primaris.

Finalment, l'anàlisi dels mutants amb els diferents anticossos monoclonals, LLBE8 i MMBB8, va servir per determinar un segon epítop per a l'ADH. L'anàlisi dels patrons de degradació de la proteïna ADH havia permés determinar l'epítop reconegut per l'anticós LLBE8, situat en l'extrem carboxílic de l'enzim (Fibla, 1990). Els nostres resultats confirmaven aquesta hipòtesi i permetien posicionar aquest epítop al voltant del aminoàcid 246. A més, el fet que aquesta regió C-terminal fos antigènica suggeria que en l'estructura terciària de la proteïna es trobava situada en la superfície de l'enzim. Aquesta situació recolçava la idea que aquest extrem podia formar un "loop" directament implicat en la unió de l'enzim i el substracte, i que aportaria les condicions d'hidrofobicitat necessàries per tal que tingués lloc la reacció.

6. CONCLUSIONS

CONCLUSIONS

1. A *D.lebanonensis* i a *D.immigrans* el gen Adh és un gen de còpia única estructurat en 3 exons i regulat per dos promotors, un proximal i un altre distal.

2. Adjacentment al gen Adh trobem una pauta de lectura oberta (3'ORF) que es transcriu i es processa. Per homologia amb l'Adh aquest gen s'anomena Adh-dup. La presència de l'Adh-dup a *D.lebanonensis* i a *D.immigrans* indica que la duplicació que va originar ambdós gens és ancestral i molt possiblement es remonta als inicis de l'aparició del gènere.

3. Les substitucions en llocs silenciosos són més freqüents que en llocs de reemplaçament. A més, aquestes últimes no es troben distribuïdes uniformement entre els exons, possiblement per restriccions funcionals.

4. L'anàlisi filogenètic de *D.lebanonensis* i *D.immigrans* respecte a la resta de les espècies del gènere suggereix que els subgèneres *Scaptodrosophila, Drosophila* i *Sophophora* varen divergir gairebé simultàniament en la història evolutiva del gènere. No obstant, la datació d'aquesta divergència depèn de la taxa de substitució considerada.

5. L'anàlisi de les substitucions sinònimes i de reemplaçament mostra que la selecció a que estan sotmesos l'Adh i l'Adh-dup és diferent malgrat estar evolutivament emparentats i localitzats adjacentment en el genoma.

6. Els canvis d'aminoàcids produïts a l'ADH són majoritàriament conservatius i es troben localitzats en determinades posicions. A més, els

resultats mostren que la substitució d'aminoàcids és un succés poc freqüent al llarg de l'evolució.

7. La distribució dels canvis aminoacídics no és aleatòria al llarg de la molècula. Aquesta distribució pot correlacionar-se amb diferents aspectes funcionals de l'enzim.

8. La Tyr152 de l'ADH participa directament en la ruta de la transferència electrònica indispensable per a l'oxidació del substracte i, per tant, és part fonamental del centre catalític de l'enzim.

9. L'Ala13 i l'Asn56 possiblement intervenen directament o indirectament en la transferència redox o en l'alliberació dels productes de la reacció. A més,*D.lebanonensis* sembla tenir un plasticitat o flexibilitat superior respecte a possibles alteracions en la seva estructura que *D.melanogaster*.

10. L'extrem C-terminal de l'ADH participa en la unió enzim-substracte alhora que podria intervenir en el mecanisme enzimàtic proporcionant l'ambient hidrofòbic adequat per a la reacció d'oxidació de l'alcohol.

11. Els resultats permeten posicionar un segon epítop de l'ADH a partir de l'aminoàcid 246. La antigenicitat de la regió C-terminal suggereix que aquesta forma part de l'estructura externa de l'enzim.

7. BIBLIOGRAFIA

BIBLIOGRAFIA

Abel T., Bhatt R. and Maniatis T., 1992. Genes Dev., **6**: 466-480.

Anderson C. L., Carew E. A. and Powell J. R., 1993. Mol. Biol. Evol., **10**: 605-618.

Atkinson P. W., Mills L. E., Starmer W. T. and Sullivan D. T., 1988. Genetics, **120**: 713-723.

Atrian S. and Gonzàlez-Duarte R., 1982. Advances in Genetics, Development and Evolution of *Diosophila*. Plenum Press, N.Y., 251-261.

Atrian S. and Gonzàlez-Duarte R., 1985. Biochem. Genet., **23**: 891-911.

Atrian S. and Gonzàlez-Duarte R., 1986. Anales de Biologia, **10**: 3-19.

Atrian S., Gonzàlez-Duarte R. and Fothergill-Gilmore L.A. 1990. Gene, **93**: 205-212.

Ayer S. and Benyajati C., 1992. Mol. Cell. Biol., **10**: 3512-3523.

Batterham P., Gritz E., Starmer W. T. and Sulliva D. T., 1983. Biochem. Genet., **21**: 871-883.

Batterham P., Chambers G., Starmer W. T. and Sullivan D. T., 1984. Evolution, **38**: 644-657.

Benyajati C., Wang N., Reddy A., Weinberg E. and Sofer w., 1980. Nucleic Acid Res. **8**: 5649-5667.

Benyajati C., Place A.R., Powers D.A., and Sofer W., 1981. Proc. Natl. Acad. Sci. USA, **78**:2717-2721.

Benyajati C., Spoerel N., Haymerle H. and Ashburner M., 1983. Cell, **33**: 125-133.

Benyajati C., Ewel A., McKeon J., Chovav M. and Juan E., 1992. Nucleic Acids Res., **20**: 4481-4489.

Beverly S.M. and Wilson A.C., 1982. J. Mol. Evol., **18**: 251-264.

Beverly S.M. and Wilson A.C., 1984. J. Mol. Evol., **21**: 1-13.

Branden C. and Tooze J., 1991. Introduction to Protein Structure. Garland Publishing, Inc. N.Y. and London.

Breathnach M. D. and Chambon P., 1987. Ann. Rev. Biochem., **50**: 349-383.

Chambers G.K., Laver W.G., Campbell S. and Gibson J.B., 1981a. Proc. Natl. Acad. Sci. USA, **78**: 3103-3107.

Chambers G.K., Wilks A.V. and Gibson J.B., 1981b. Aust. J. Biol. Sci., **34**: 625-637.

Chambers G. K., Wilks A.V. and Gibson J.B., 1984. Biochem. Genet., **22**: 153-168.

Chambers G.K., 1984. Biochem. Genet., **22**: 529-549.

Chia W., Karp R., McGill S. and Ashburner M., 1985. J. Mol. Evol., **186**: 689-706.

Chen Z., Lu L., Shirley M., Lee W.R. and Chang S.H., 1990. Biochemistry, **29**: 1112-1118.

Chen Z., Lee W.R. and Chang S.H., 1991. Eur. J. Biochem., **202**: 263-267.

Chen Z., Jiang J.C., Lin Z.G., Lee W.R., Baker M.E. and Chang S.H., 1993. Biochemistry, **32**: 3342-3346.

Cohn V. and Moore G., 1988. Mol. Biol. Evol., **5**: 154-166.

Cols N., Marfany G., Atrian S. and Gonzàlez-Duarte R., 1993. FEBS Lett., **319**: 90-94.

Corbin V. and Maniatis T., 1989a. Nature, **337**: 279-282.

Corbin V. and Maniatis T., 1989b. Genes and Dev., **3**: 2191-2200.

David J., Bocquet C., Herrewege J.V., Fouillet P. and Arens M.F., 1978. Biochem. Genet., **16**: 203-211.

Dothie J.M., Giglio J. R., Moore C. B., Taylor S.S. and Hartley B. S., 1985. Biochem. J., **230**: 569-578.

Eklund H., Nordström B., Zeppezauer E., Söderlund G., Ohlsson I., Boiwe T., Söderberg B. O., Tapia O., Bräden C.I. and Akesson A., 1976. J. Mol. Biol., **102**: 27-59.

Ensor C.M. and Tai H.H., 1991. Biochem. Biophys. Res. Comun., **176**: 840-845.

Falb D. and Maniatis T., 1992. Genes Dev., **6**: 454-465.

Fibla J., Enjuanes L. and Gonzàlez-Duarte R., 1989. Biochem. Biophys Res. Commun., **160**: 638-646.

Fibla J., 1990. Tesi Doctoral. Universitat de Barcelona.

Fisher J. A. and Maniatis T., 1985. Nucleic Acids Res., **13**: 6899-6917.

Fitch J. A. and Margoliash E., 1967. Science, **155**: 279-284.

Garcin F., Cote J., Radouco-Thomas S., Chawla S.S. and Radouco-Thomas C., 1983. Experientia, **39**: 1122-1123.

Geer B.W., Langevin M.L. and McKechnie S.W., 1985. Biochem. Genet., **23**: 607-622.

Gibson J.B., Chambers G.K., Wilks A.V. and Oakeshott J.B., 1980. Aust. J. Biol. Sci., **33**: 479-489.

Gilbert W., 1978. Nature, **271**: 501.

Goldberg, D. A. 1980. Proc. Natl.Acad. Sci. USA, **77**: 5794-5798.

Goldberg D. A., , Posakony J. W. and Maniatis T., 1983. Cell, **34**: 59-73.

Gordon E., Burry S., Sawyer L., Atrian S. and Gonzàlez-Duarte R., 1992. J. Mol. Biol., **227**: 356-358,

Grell E. H., Jacobson K. B. and Murphy J. B., 1965. Science, **149**: 80-82.

Hayashida H. and Miyata T., 1983. Proc. Natl. Acad. Sci. USA, **80**: 2671-2675.

Heberlein U, England B. and Tjian R., 1985. Cell, **41**: 965-977.

Hovik R., Winberg J.O. and McKinley-McKee J.S., 1984. Insect Biochem., **14**: 345-351.

Jörnvall H., 1970. Eur. J. Biochem., **16**: 25-40.

Jörnvall H., 1977a. Eur. J. Biochem., **72**: 425-442.

Jörnvall H., 1977b. Pyridine Nucleotide-Dependent Dehydrogenases, Ed. Sund H.: 62-71

Jörnvall H., Persson M. and Jeffery J., 1981. Proc. Natl. Acad. Sci. USA, **78**: 4226-4230.

Juan E. and Gonzàlez-Duarte R., 1980. Biochem. J., **189**: 105-110.

Juan E. and Gonzàlez-Duarte R., 1981. Biochem. J., **195**: 61-69.

Kawasaki E.S., 1990. PCR Protocols, Ed. Innis M.A., Gelfand D.H., Sninsky J.J. and White T.J. Academic Press, San Diego: 21-27.

Kimura M., 1980. J. Mol. Evol., **16**: 111-120.

Kreitman M. and Hudson R. R., 1991. Genetics, **127**: 565-582.

Krook M., Marekov L. and Jörnvall H., 1990. Biochemistry, **29**: 738-743.

Krook M., Prozorovski V., Atrian S., Gonzàlez-Duarte R. and Jörnvall H., 1992. Eur. J. Biochem., **209**: 233-239.

Kuntzel H. and Kochel H. G., 1981. Nature **293**: 751-755.

Leslie C. A. and Levine L., 1973. Biochem. Biophys Res. Commun., **52**: 717-724.

Li W. H., Wu C. I. and Lou C.C., 1985. Mol. Biol. Evol., **2**: 150-174.

Li W.H. and Graur D., 1991. Fundamentals of Molecular Evolution.

Locket T. J. and Ashburner M., 1989. Dev. Biol., **134**: 430-438.

Loukas M., Krimbar C. B., Mavragani-Tsipidou P. and Kastritsis, 1979. J. Hered., **70**: 17-26.

Madhavan H., Conscience-Egli M., Sieber F. and Ursprung H., 1973. J. Insect Physiol., **19**: 235-341.

Marfany G., 1991a. Tesis Doctoral. Universitat de Barcelona.

Marfany G. and Gonzàlez-Duarte R., 1991b. J. Mol. Evol., **32**: 454-462.

Marfany G. and Gonzàlez-Duarte R., 1992a. Mol. Biol. Evol., **9**: 261-277.

Marfany G. and Gonzàlez-Duarte R., 1992b. J. Mol. Evol., **35**: 492-501.

Marfany G. and Gonzàlez-Duarte R., 1993. Mol. Phyl. Evol., **2**: 13-22.

Margoliash E., 1963. Proc. Natl. Acad. Sci. USA, **50**: 672-679.

Margulis, L. and Schwartz K. V., 1982. Five Kingdoms. W. H. Freeman and Company.

Matsumoto N., Sekimizu K., Soma G-I., Ohmura Y., Andoh T., Nakanishi Y., Obinata M. and Natori S., 1985. J. Biochem., **97**: 1501-1508.

McLchland A. D., 1971. J. Mol. Biol., **61**: 409-424.

Menotti-Raymond M., Starmer W. T. and Sullivan D. T., 1991. Genetics, **127**: 355-266.

Miyata T. Yasunaga T. and Mishida T., 1980. Proc Natl. Acad. Sci. USA, **77**: 7328-7332.

Moriyama E.N. and Gojobori T., 1992. Genetics, **130**: 855-864.

Moses K., Heberlein U. and Ashburner M., 1990. Mol. Cell. Biol., **10**: 539-548.

Olsen G. J., 1988. Cold Spring Harbor Symp. Quant. Biol. **52**: 825-837.

Papaceit M. and Juan E., 1993. Chromosoma, **102**: 361-368.

Perler F. and Efstratiadis A., 1980. Cell, **20**: 555-566.

Persson B., Krook M. and Jörnvall H., 1991. Eur. J. Biochem., **200**: 537-543.

Rat L., Veuille M., Lepesant J., 1991. J. Mol. Evol., **33**: 194-203.

Ribas Ll., 1992. Tesis Doctoral. Univesitat d'Edinburgh.

Rossmann M.G., Moras D. and Olsen K.W., 1974. Nature, **250**: 194-199.

Rowan R.G. and Dickinson W.J., 1988. J. Mol. Evol., **28**: 43-54.

Saccone C., Lanave C. and Pessole G., 1983. J. Mol. Evol. **37**: 154-159.

Saitou N. and Nei M., 1987. Mol. Biol. Evol., **4**: 406-425.

Savakis C. and Ashburner M., 1985. Cold Spring Harbor Symp. Quant. Biol.,**50**: 505-514.

Savakis C., Ashburner M. and Willis J. H., 1986. Dev. Biol., **114**: 194-207.

Schaeffer S. W. and Aquadro C. F., 1987. Genetics, **117**: 61-73.

Schwartz M F. and Jörnvall H., 1976. Eur. J. Biochem., **68**: 159-168.

Shen N. L., Subrahmannyam G., Clark W., Martin W. and Sofer W., 1989. Develop. Gen., **10**: 210-219.

Sharp P. M. and Li W.H., 1989. J. Mol. Evol., **28**: 398-402.

Smith E. L. and Margoliash E., 1964. Fed. Proc., **23**: 1243-1247.

Sullivan D. T., Atkinson P. W. and Starmer W. T., 1990. Evolutionary Biology, ed. M.K. Hecht, B. Wallace and R.J. Macintyre. Plenum N.Y.: 107-147.

Sun H. W. and Plapp B. V., 1992. J. Mol. Evol., **34**: 522-535.

Taylor W. R., 1986. J. Mol. Biol., **188**: 233-258.

Thatcher D. R., 1977. Biochem. J., **163**: 317-323.

Thatcher D.R., 1980. Biochem. J., **187**: 875-886.

Thatcher D. R. and Sawyer L., 1980. Biochem. J., **187**: 884-886.

Theorell H. and Chance B., 1951. Acta. Chem. Scand., **5**: 1127-1144.

Throckmorton L., 1975. Handbooks of Genetics. Ed. King R. C. Plenum, NY., **3**: 421-469.

Traut, T. W., 1988. Proc. Natl. Acad. Sci. USA, **85**: 2944-2948.

Vilageliu Ll. and Gonzàlez-Duarte R., 1980. Experientia, **36**: 828-829.

Vilageliu Ll. and Gonzàlez-Duarte R., 1984. Biochem. Genet., **22**: 795-815.

Vilarroya A., Juan E., Egestad N. and Jörnvall H., 1989. Eur. J. Biochem., **180**: 191-197.

Visa N., Marfany G., Vilageliu Ll., Albalat R., Atrian S. and Gonzàlez-Duarte R., 1991. Chromosoma, **100**: 315-322.

Visa N., 1991. Tesis Doctoral. Universitat Autónoma de Barcelona.

Walsby A. E., 1986. Nature, **320**: 212.

Wierenga R.K., Drenth J. and Schulz G.E., 1983. J. Mol. Biol., **167**: 725-739.

Winberg J.O., Thatcher D.R. and McKinley-McKee J.S., 1982a. Biochim. and Biophys. Acta, **704**: 7-16.

Winberg J.O., Thatcher D.R. and McKinley-McKee J.S., 1982b. Biochim. and Biophys. Acta, **704**: 15-25.

Winberg J.O., Thatcher D.R. and McKinley-McKee J.S., 1983. Biochem. Genet., **21**: 63-80.

Winberg J.O., Hovik R. and McKinley-McKee J.S., 1985. Biochem. Genet., **23**: 205-216.

Winberg J.O., Hovik R., McKinley-McKee J.S., Juan E. and Gonzàlez-Duarte R., 1986. Biochem. J., **235**: 481-490.

Winberg J.O. and McKinley-McKee J.S., 1988a. Biochem. J., **251**: 223-227.

Winberg J.O. and McKinley-McKee J.S., 1988b. Biochem J., **255**: 589-599.

Winberg J.O. and McKinley-McKee J.S., 1992. Int. J. Biochem., **24**: 169-181.

Woese C. R., 1981. Sci. Am. **244**: 98-125.

Woese, 1987. Bacterial Evolution. Microbiol. Rev. **51**: 221-271.

Zurckerkandl E. and Pauling L., 1962. Horizons in Biochemistry. Academic Press, N.Y., 189-225.

Zurckerkandl E. and Pauling L., 1965. Evolving Genes and Proteins. Academic Press, N.Y., 97-166.