

**Estructura tridimensional del domini
extracel·lular de 4F2hc (CD98) humana**

Joana Fort i Baixeras

**TESI DOCTORAL
Barcelona, octubre de 2006**

OBJECTIUS

OBJECTIUS

El principal objectiu d'aquesta tesi és contribuir en l'estudi de l'estructura i la funció de les HSHAT, i d'aquesta manera poder anar un pas més enllà en l'estudi de l'estructura quaternària del transportador actiu, o HAT. Així doncs ens vam disposar a purificar els dominis extracel·lulars d'rBAT i 4F2hc per tal de cristal·litzar-los i obtenir-ne l'estructura tridimensional per difracció de raigs X, així com caracteritzar-ne possibles activitats.

Tenir l'estructura tridimensional d'aquests dominis ens permetria entendre millor quina relació guarden aquests amb la família de les α -amilases, quin paper podrien tenir dins l'estructura quaternària del transportador i quina podria ser la seva posició en la part extracel·lular de la membrana en relació a la subunitat lleugera corresponent.

Podríem definir els objectius en blocs:

- Produir l'ectodomini de les subunitats pesades dels transportadors heteromèrics d'aminoàcids en un sistema procariota que rendís elevades quantitats de proteïna.
- Purificar la proteïna recombinant fins a obtenir-la en la puresa suficient per cristal·litzar.

Ja que vam assolir la producció i purificació de l'ectodomini de 4F2hc,

- Cristal·litzar i resoldre l'estructura tridimensional per difracció de raig-X de l'ectodomini de 4F2hc.
- Estudiar la possible funció d'ED-4F2hc com a lectina o glucosidasa i la possible interacció d'aquest domini amb altres proteïnes.

RESULTATS

Capítol 1. Producció i purificació dels ectodominis de 4F2hc i rBAT

Per obtenir l'estructura tridimensional per difracció de raigs X es necessiten grans quantitats de proteïna (de l'ordre de mg) d'una gran puresa i cristal·litzar-la. Així doncs la proteïna s'ha d'expressar en suficient quantitat i de manera soluble, s'ha de purificar i s'ha de cristal·litzar. Tots aquests passos requereixen un llarg procés d'optimització. A més, disposar d'una proteïna pura i soluble és una eina que permet usar diverses tècniques per estudiar-la, com el dicromisme circular, estudis d'estructura i funció, de manera que aquest capítol tracta d'explicar com hem posat a punt l'expressió i purificació dels ectodominis de les subunitats pesades dels transportadors heteromèrics d'aminoàcids, i dels problemes i dels èxits que hem trobat pel camí.

1.1. Estratègia de clonatge

Les subunitats pesades dels transportadors heteromèrics d'aminoàcids rBAT i 4F2hc tenen un gran domini extracel·lular que s'assembla a les glucosidases d'insectes i de bacteris (figura 9) (Janecek et al., 1997a; Chillaron et al., 2001). Per definir quin fragment produiríem ens vam basar en el primer motiu de la seqüència que mostra homologia amb les glucosidases (figura 7), que són dos triptòfans veïns conservats en molts representants de la família de les α -amilases com també en rBAT i 4F2hc. Aquests es troben en les nostres proteïnes (en posició W117 i W118 en totes dues) després de la cisteïna extracel·lular que forma el pont disulfur amb la subunitat lleugera i molt a prop (6 i 12 aminoàcids en rBAT i 4F2hc, respectivament) de l'hèlix transmembrana. D'aquesta manera el domini clonat en els plasmidis d'expressió ha estat des d'un aminoàcid localitzat entre aquesta cisteïna i els dos triptòfans i fins més enllà del codó de terminació de traducció (des de E111 de h4F2hc i des de W117 per hrBAT). També es va clonar el domini semblant a glucosidases de la proteïna homòloga d'rBAT en conill, ja que al nostre laboratori disposem d'un anticòs específic contra aquesta regió de la proteïna de conill.

Tots els plasmidis per a l'expressió en bacteris que es van clonar tenen un promotor

induïble per l'addició d'IPTG (isopropyl- β -D-thiogalactoside) al medi de cultiu dels bacteris.

El plasmidi que vam escollir per a l'expressió en bacteris és pTrcHis (Invitrogen). Aquest constructe conté una cua de sis histidines a N-terminal dels ectodomini, un epítot X-press i un lloc de tall per enteroquinasa, una proteasa específica (figura 9). També es van clonar els ectodomini en els plasmidis pET28 (Novagen) i pGEX6 (Amersham), que fusionen els ectodomini amb una cua de sis histidines o la proteïna de fusió GST, respectivament, a N-terminal.

El plasmidi pEF-Fc amb l'ectodomini de 4F2hc es va clonar per a l'expressió de la proteïna en el medi de cultiu de cèl·lules HEK-293T. Aquest plasmidi incorpora un pèptid senyal i un epítot HA a N-terminal de l'ectodomini i el fragment Fc de la IgG humana a C-terminal.

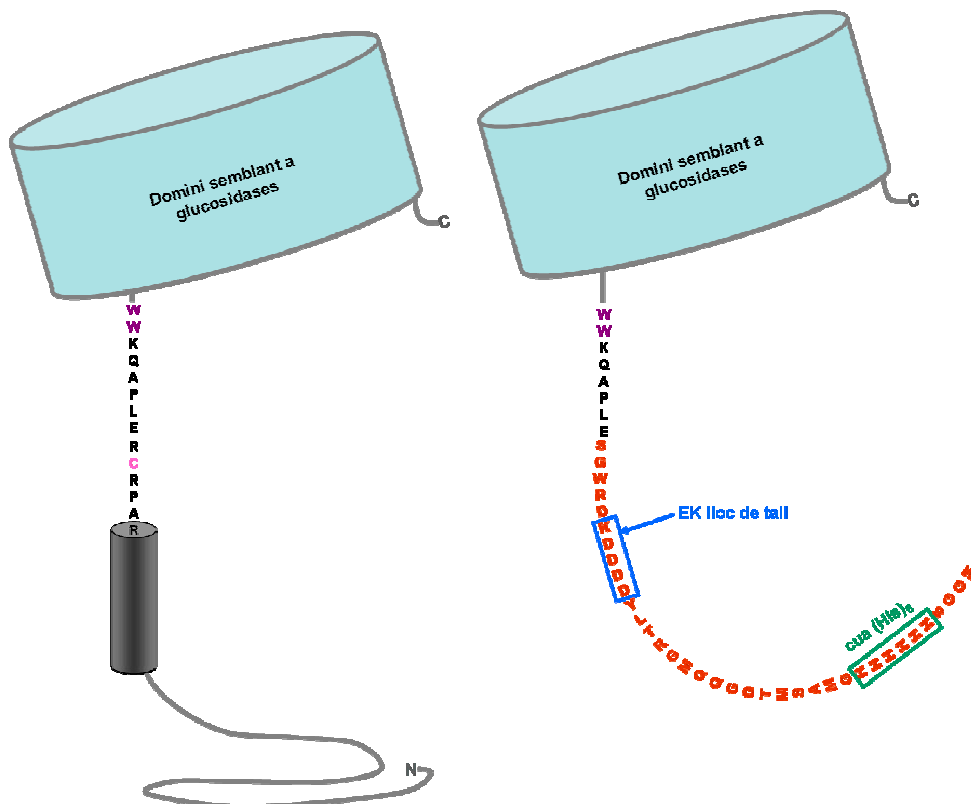


Figura 9. A l'esquerra hi ha un esquema de la proteïna 4F2hc on podem veure els dos triptòfans (W117 i W118) que tenen homologia amb les glucosidases i la cisteïna (C109) que forma el pont disulfur amb la subunitat lleugera. A la dreta hi ha la construcció de la proteïna recombinant HisED4F2hc produïda en *E. coli* amb la cua d'histidines afegida abans de l'aspàrtic 111 de 4F2hc. La caixa blava representa el lloc de reconeixement de la proteasa EK i fletxa blava el lloc de tall d'EK.

1.2. Proves d'expressió en *E. coli*

Per tal d'obtenir la major quantitat possible de proteïna soluble vam optimitzar diferents punts en l'expressió d'una proteïna en un sistema procariota: *E. Coli*. Vam fer proves amb diferents soques, diferents temps d'inducció, inducció a diferents fases del creixement a diferents temperatures i expressió de diferents constructes.

Mitjançant experiments pilot d'inducció vam estudiar la fracció on trobàvem la nostra proteïna recombinant. En el cas de HishED4F2hc després de la lisi i la centrifugació de l'extracte trobàvem la meitat de la proteïna aproximadament en la fase soluble (sobrenedant) i l'altra meitat en la fase insoluble i cossos d'inclusió (sediment o pèl·let) (figura 10). Aquesta proporció no era tan clara en el cas de HisrEDrBAT i HishEDrBAT (que tenen un comportament similar), on trobàvem més proporció de la proteïna en la fase insoluble. Per augmentar la proporció de proteïna en la fase soluble vam fer diverses proves com baixar la temperatura d'inducció a 20 °C o induir i afegir al medi un 0,1% del volum de cultiu del detergent tween-20 (Sigma). Cap d'aquestes dues proves van augmentar la proporció de proteïna que recuperàvem en la fase soluble.

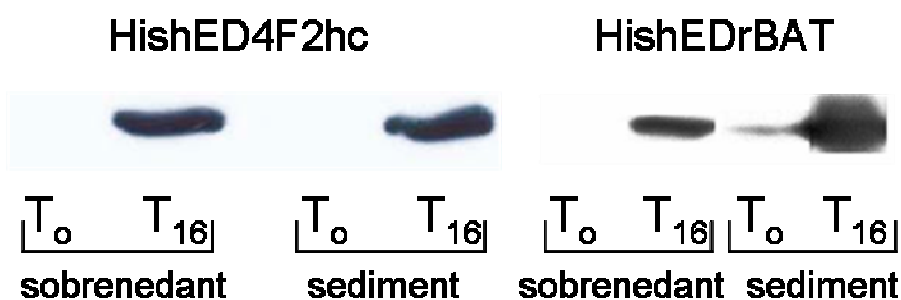


Figura 10. Western Blot antiHistidines. Les mostres corresponen a quantitats proporcionals del sobrenedant i el sediment (o pèl·let) de la mateixa centrifugació. T_x és el temps en que ha estat recollida la mostra després de la inducció en hores.

En la figura 11 es mostra el resultat d'expressions pilot en diferents soques de bacteris d'HishED4F2hc (PM=50 KDa) i d'HisrEDrBAT (PM=68 KDa). Es mostren tres experiments diferents i en tots hi ha l'expressió de les proteïnes en DH5α, que és l'usat per comparar. La soca escollida per fer les expressions a gran escala va ser

BL21 (DE3) (Novagen) ja que és una soca especial per a l'expressió de proteïnes, perquè és deficient per les proteases *Lon* i *OmpT* i per tant dóna estabilitat a la proteïna expressada, i és en la que teníem més expressió.

En tots els casos l'expressió d'HisrEDrBAT (que mostra el mateix comportament que HishEDrBAT) era molt menor que la d'HishED4F2hc i a més la proporció de proteïna insoluble era molt més elevada. Com hem comentat abans aquesta proteïna era menys soluble, i això podria ser la causa de la seva pobra expressió.

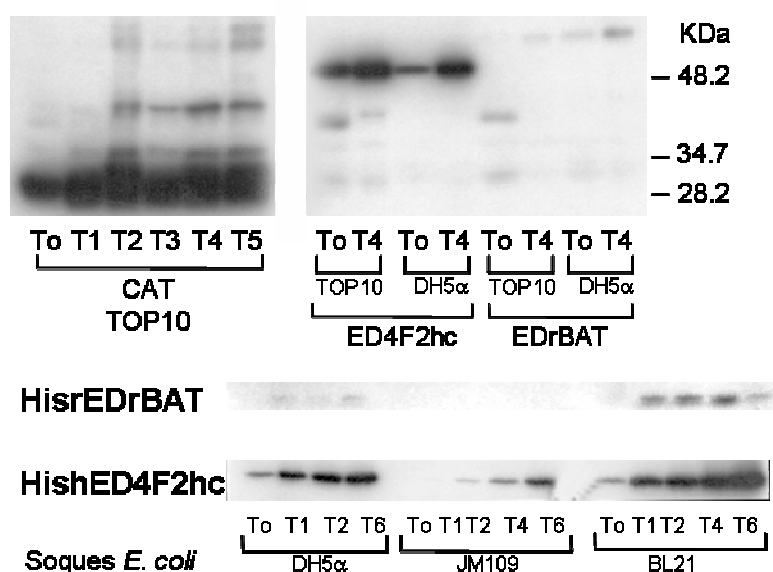


Figura 11. Western Blot antiHistidines. Les mostres han estat carregades proporcionalment de la part soluble d'1 ml de cultiu d'expressions pilot. CAT és un control positiu del vector, DH5α, TOP10, XL1Blue JM109 i BL21 DE3 són diferents soques d'*E. coli*.

Dels diferents vectors d'expressió provats, en pET28 i pGEX no es van trobar millores en l'expressió soluble dels ectodominis respecte pTrcHis i per això vam continuar amb aquest, que era el primer que vam provar i ja el coneixíem.

L'ectodomini d'rBAT, doncs, no es produïa amb prou quantitat. Per intentar millorar aquesta expressió vam clonar l'ectodomini sense la cua C-terminal (aminoàcids 652-686) perquè en aquesta regió la proteïna rBAT té tres cisteïnes que podien estar interferint en el seu plegament correcte en *E. coli*. El resultat de l'expressió d'aquest constructe tampoc va ser millor que l'ectodomini sencer.

S'han fet més proves d'expressió amb His-hEDrBAT en el nostre laboratori (Dr. Albert Rosell) com l'extracció de la proteïna dels cossos d'inclusió mitjançant desnaturalització i replegament en la columna però tots els experiments han tingut un resultat negatiu que no ens ha permès purificar-lo.

1.3. Producció de l'ectodomini de 4F2hc en *E. coli*

Després de les proves d'expressió a petita escala vam decidir continuar amb l'expressió de ED-4F2hc i escalar la producció per obtenir les quantitats necessàries per començar la cristal·lització. Al llarg del temps s'han anat optimitzant les condicions d'expressió d'aquest ectodomini amb el plasmidi pTrcHis en la soca d'expressió d'*E. coli* BL21 (DE3). Actualment partim de 4 litres de cultiu, inducció a $OD_{600} \approx 0,6$ amb 1mM d'IPTG i deixem l'expressió ON (aproximadament 16 hores). Després procedim a l'extracció de la proteïna soluble mitjançant lisi amb *French Press* (veure material i mètodes).

Posteriorment també va ser necessari produir l'ectodomini substituint les metionines per selenometionines per resoldre el problema de la fase en la resolució de l'estructura. El protocol de producció va ser semblant però el medi de creixement dels bacteris era sintètic i contenia selenometionina enlloc de metionina (veure apartat 2.3 de materials i mètodes).

1.4. Purificació de l'ectodomini de 4F2hc

Aprofitant les característiques de la proteïna produïda en *E. coli* amb el plasmidi pTrcHis la cromatografia escollida per purificar-la de l'extracte de proteïna soluble va ser la d'afinitat. Aquesta columna es carrega amb níquel, metall pel qual les histidines tenen molta afinitat. Després de fer diverses proves de purificació l'elució que donava més rendiment era directament amb una incubació amb la proteasa específica enteroquinasa. Vam utilitzar un FPLC per fer la purificació a gran escala. Aquest sistema ens permetia treballar amb volums més grans i de forma més ràpida i obtenir una proteïna d'una puresa superior.

Després de la columna d'afinitat i l'elució amb enteroquinasa la mostra es va

sotmetre a una segona cromatografia. Aquesta podia ser d'intercanvi aniònic o d'exclusió molecular. La d'intercanvi aniònic és una monoQ i s'elueix amb un gradient de sal. La d'exclusió molecular permet separar les molècules de diferents mides, però el volum final és més gran. En la figura 12 podem veure el grau de puresa que obtenim en cada pas. El rendiment final de tot el procés és de 2,5 mg per litre de cultiu d'*E. coli*.

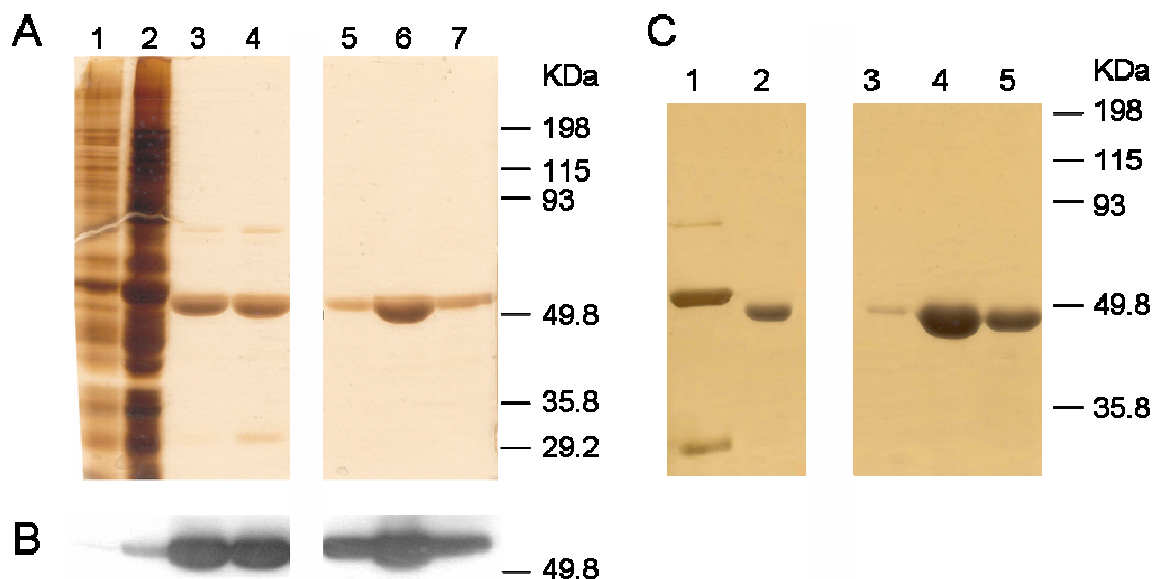


Figura 12. A) Tinció de proteïnes amb plata. **B)** Western Blot antihistidines de les mateixes mostres. Carrils 1 i 2, extractes de proteïnes del cultiu sense induir i induït respectivament; carrils 3 i 4, dos fraccions de la purificació d'afinitat amb níquel; carril 5, 6 i 7, fraccions consecutives del pic de purificació obtingut per cromatografia d'intercanvi iònic (monoQ). Densitometria: carrils 1 i 2 puresa d'un 80% i 90%, respectivament; carrils 5, 6 i 7 puresa del 100%. **C)** Gel amb mostres d'una purificació estandarditzada. Carrils: 1) proteïna eluïda de la columna d'histidines; 2) proteïna eluïda per proteòlisi de la cua d'histidines en la columna; 3, 4 i 5) elució de la columna monoQ. Densitometria: 1 puresa 70%, 2 puresa d'un 90% i 3, 4 i 5 puresa d'un 100%.

1.5. Test del plegament d'ED4F2hc per dicroisme circular i fluorescència.

Abans de posar-nos a cristal·litzar volíem assegurar-nos que la proteïna purificada estava en unes condicions properes a les natives. El fet que ED-4F2hc es produís de manera soluble ja ens donava un indici que la proteïna podia estar ben plegada. De totes maneres teníem la limitació que ens donava no conèixer una activitat concreta per testar l'ectodomini un cop expressat i purificat. Per això vam decidir fer estudis de dicroisme circular i fluorescència. Aquests estudis es van dur a terme en

col·laboració amb el Doctor Javier Turnay del laboratori de la Doctora M. Antonia Lizarbe a la Universitat Complutense de Madrid.

En aquest laboratori també es va estudiar l'aspecte d'absorció de la proteïna ED4F2hc que ens va permetre obtenir el coeficient d'extinció molar a 280 nm (ϵ^{280}) per aquesta proteïna, que és de $64150 \text{ M}^{-1} \cdot \text{cm}^{-1}$ (figura13).

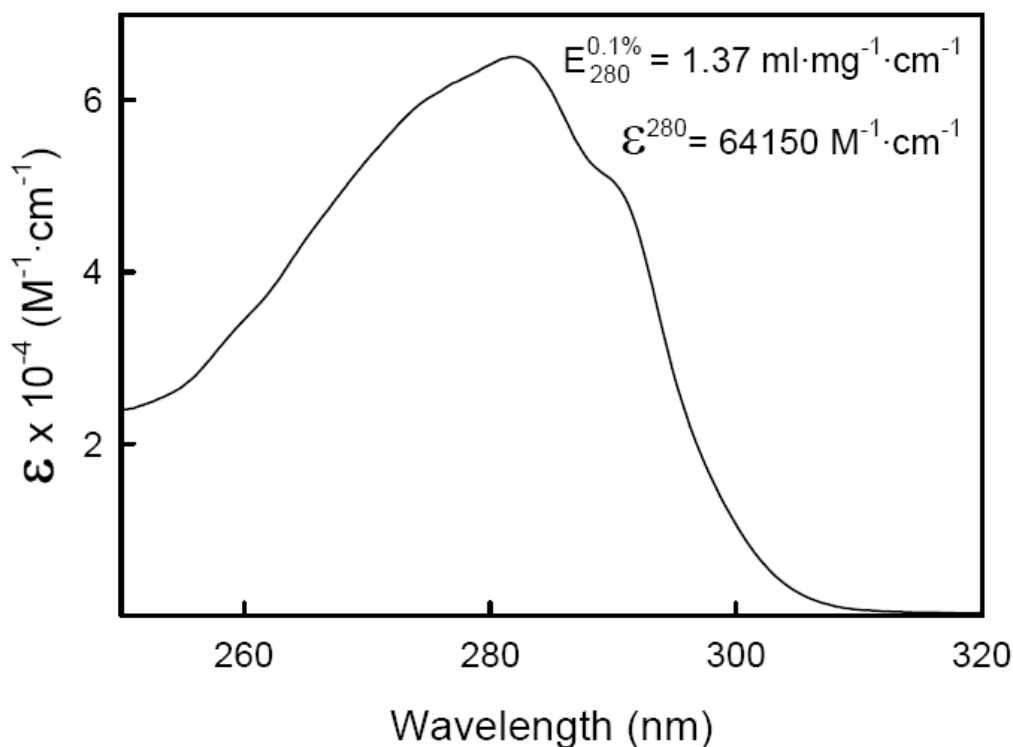


Figura13. Espectre d'absorció entre les longituds d'ona 250 i 320 i determinació del coeficient d'extinció molar d'ED-4F2hc a 0.2 mg/ml (teòric) i 0.341 mg/ml (real). La concentració real de la proteïna de la qual es va fer l'espectre es va determinar per anàlisi d'aminoàcids.

L'espectre de dicromisme circular de la proteïna HishED4F2hc ens va permetre comparar el càlcul dels percentatges de les estructures secundàries (figura 14) amb els percentatges esperats si l'estructura es correspongués a una glucosidasa com l'oligo-1,6-glucosidasa de *B.cereus* segons els dominis predits per l'alineament de seqüència (Chillaron et al., 2001). Per altra banda les corbes de desnaturalització ens demostraven que la proteïna era molt estable, amb una T_m propera als 60°C i la desnaturalització en aquestes condicions era irreversible. Aquest fet també era compatible amb una estructura complexa com un barril $(\alpha/\beta)_8$. Les corbes de desnaturalització amb 2,5mM de CaCl_2 mostren una agregació irreversible de la proteïna a uns 60°C . Amb EGTA 2,5 mM la T_m és lleugerament superior i la corba

de refredament (figura 14B corba inferior) mostra que la proteïna no es renaturalitza al disminuir la temperatura després d'haver-la desnaturalitzat.

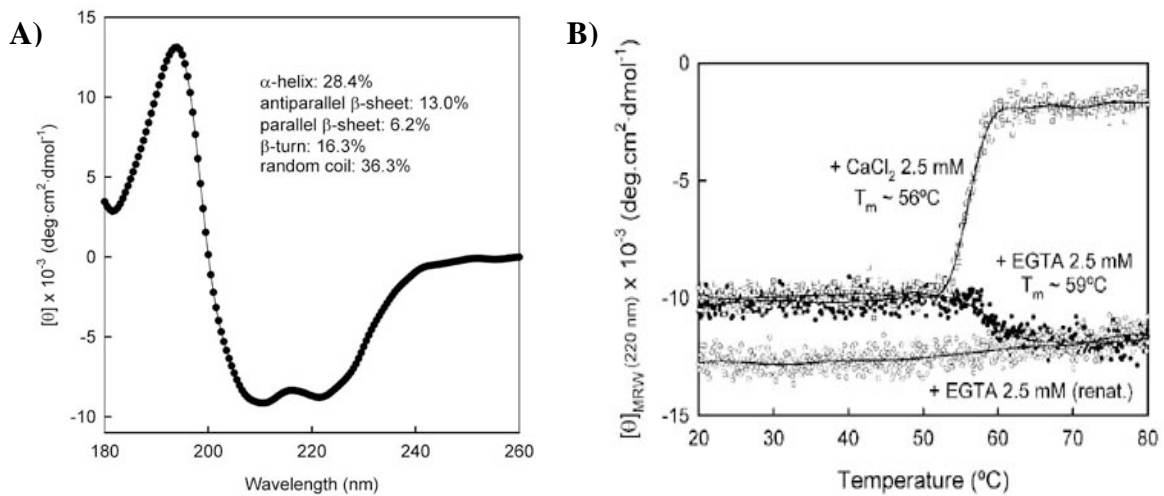


Figura 14. Gràfics dels estudis de dicroisme circular. A) Espectre de dicroisme circular en l'UV-llunyanà obtingut amb una cubeta de 0,01 cm de pas òptic i sota corrent de nitrogen (tampó HEPES 2 mM, pH 8.0) amb els percentatges d'estructura secundària calculats. B) Corbes de desnaturalització per temperatura amb 2,5 mM de CaCl₂ o 2,5 mM d'EGTA, dicroisme circular a 220nm. La línia inferior correspon a la "corba de refredament" de la mostra amb EGTA.

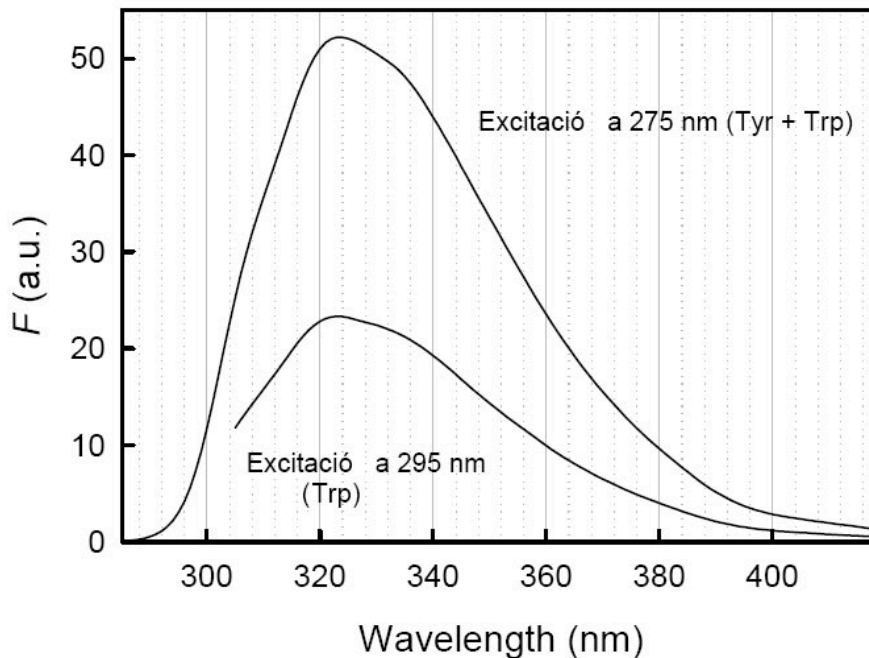


Figura15. Espectres de fluorescència de la proteïna. Amb excitació a 295 nm es veu la fluorescència només dels triptòfans mentre que a 275 la fluorescència és deguda a les tirosines i als triptòfans.

L'espectre de fluorescència mostra que el màxim dels triptòfans apareix a longitud d'ona 323 nm (figura 15). Podem veure que la contribució de les tirosines quan

l'excitació es fa a 275 nm és molt baixa, ja que el màxim a 323 nm apareix tan quan l'excitació és a 295 com quan l'excitació és a 275nm, la qual cosa indica que els dos pics són deguts als mateixos residus de triptòfan. El pic a 323 nm suggereix que la població dels triptòfans és més aviat interna, ja que és una longitud d'ona relativament baixa. No obstant això en la gràfica de fluorescència apareix un turó a les longituds d'ona 334-336 nm que suggereix la possible existència d'un grup de triptòfans en entorn més hidrofílic.

1.6. Producció i purificació de l'ectodomini de 4F2hc en cèl·lules HEK 293T

Com ja hem comentat en la introducció, les subunitats pesades 4F2hc i rBAT són glicoproteïnes de membrana de tipus II. En concret, en 4F2hc humà hi ha 4 asparagines susceptibles de ser N-glicosilades (N264, N280, N323, N405). Per tal de tenir un ectodomini de 4F2hc més semblant al natiu ens vam disposar a produir-lo en cèl·lules eucariotes. El sistema escollit va ser la producció de la proteïna recombinant en el medi de cultiu de les cèl·lules HEK 293T. Amb aquest objectiu vam utilitzar un plasmidi d'expressió en cèl·lules de mamífer (pEF). La proteïna recombinant té un pèptid senyal (SP), un epítip HA (hemaglutinina), i un lloc de tall per trombina (Thr) a N-terminal i un lloc de tall per trombina i el fragment Fc de les IgG humanes a C-terminal (figura 16).

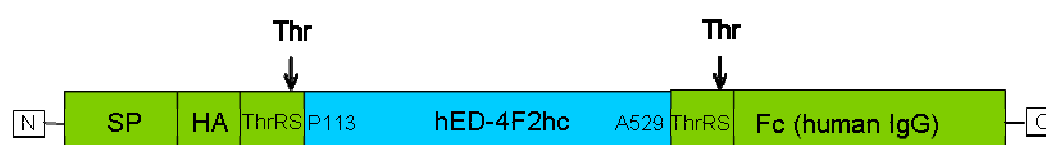


Figura16. Esquema de la proteïna recombinant que s'obté de l'expressió de pEF-ED4F2Fc.

Quan aquesta proteïna es tradueix en el reticle de les cèl·lules HEK 293T, el pèptid senyal de 21 aminoàcids de la seqüència líder de la cadena κ de ratolí (SP) que porta a N-terminal fa que aquesta sigui dirigida cap a la via de secreció. Com que la proteïna recombinant no té segment transmembrana, la proteïna no queda ancorada en la membrana, se secreta a l'exterior cel·lular i queda soluble soluble en el medi de cultiu. Després de tres dies d'expressió es recull el medi d'aquestes cèl·lules que ja

està enriquit amb la proteïna recombinant i hi posem medi fresc perquè continuï produint. Utilitzem un medi de cultiu sintètic, sense sèrum, per evitar la major part de proteïnes contaminants que ens farien més difícil la purificació de la proteïna que volem i així tampoc estimulem la divisió cel·lular i les cèl·lules no gasten energia per dividir-se.

El fragment Fc de les IgG humanes porta una cisteïna que forma un pont disulfur intercatenari, que dimeritza dues proteïnes recombinants. El resultat d'una expressió d'aquest tipus es pot veure en la figura 17.

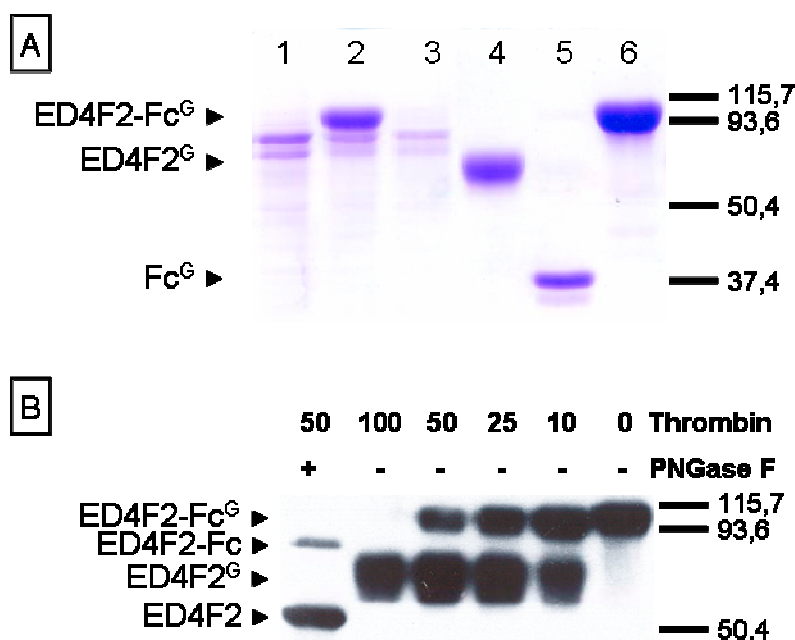


Figura 17. Purificació d'ED-4F2hc glicosilat. **A.** Tinció amb Coomassie. 1. Medi concentrat de cèl·lules HEK-293T sense transfectar 2. Medi concentrat de cèl·lules HEK-293T transfectades amb pEF-4F2-Fc 3. Rebuig de la columna d'afinitat de proteïna A 4. Elució de la columna amb Thrombina 5. Següent elució de la mateixa columna amb Citrat sòdic pH3 6. Purificació de medi concentrat amb ED4F2-Fc directament amb citrat sòdic pH3. **B.** Western Blot anti-4F2hc humà on es mostra la digestió de la proteïna de fusió ED4F2hc-Fc amb la proteasa Thrombina amb diferents Unitats/mg de proteïna i també la digestió de les dues bandes amb la glicosidasa PNGasa F, que treu totes les glicosilacions de la proteïna.

En la figura 17, veiem que la proteïna ED4F2hc-Fc glicosilada té un pes per SDS-PAGE, de 107 KDa. Com ja hem comentat, aquesta proteïna conté un pont disulfur entre les dues cadenes Fc que mostren una banda de més de 200 KDa quan correm el gel en condicions no reductores (dades no presentades). La proteïna recombinant digerida dona dues bandes corresponents a ED4F2hc glicosilat de 73 KDa i la Fc

glicosilada de d'uns 37 KDa. Quan tractem aquesta proteïna amb PNGasa F podem calcular la mida de les proteïnes sense les glicosilacions, que és de 93 KDa per ED4F2hc-Fc i 54 KDa per ED4F2hc. Així podem concloure que la mida de les glicosilacions d'ED4F2hc produït en HEK-293T és d'uns 20 KDa aproximadament.

El rendiment final d'aquest sistema de producció i purificació de ED-4F2hc glicosilat és aproximadament d'1 mg de proteïna pura (carril 4 de la figura 17) per cada 100 ml de medi que provenen de 2 plaques de cultiu de 15 cm de diàmetre amb HEK 293T transfectades transitòriament (veure materials i mètodes) i en les quals s'hi fan dos recollides de 25ml de medi a dia 4 i a dia 7 de la transfecció.

Capítol 2. Cristal·lització, difracció de raigs X i resolució de l'estructura tridimensional d'ED4F2hc

2.1. Proves de cristal·lització

Després de tenir suficient proteïna pura i a la concentració indicada (més de 5 mg/ml) vam començar els cribratges (*screenings*) de cristal·lització. La tècnica utilitzada per posar les gotes de cristal·lització va ser, en tots els casos, la de gota suspesa (*hanging drop*) i, excepte en alguns casos d'optimització, el volum de la gota era d'1 µl de solució del reservori més 1 µl de solució de proteïna d'entre 5 i 15 mg/ml.

Al principi la proteïna estava purificada amb la cua d'histidines, ja que l'eluíem de la columna d'afinitat amb un gradient d'imidazol. Després d'analitzar més de sis *screenings* generals (que representen 216 solucions de cristal·lització) amb la proteïna HisG4F2hc a 15 mg/ml no vam tenir resultats positius. El màxim que s'observava en algunes gotes eren esferòlits (una espècie d'estat sòlid que no és prou ordenat per créixer i formar un cristall). Com a conseqüència d'aquests resultats vam decidir produir proteïna sense la cua d'histidines mitjançant la proteòlisi específica amb l'enteroquinasa seguint el protocol de purificació esmentat en l'apartat 1.4 anterior i explicat amb detall a l'apartat 3.1 de material i mètodes. Amb la proteïna resultant, ED4F2hc, vam obtenir cristalls en el primer *screening*, en la condició 22 del *Crystal screening I* (veure condicions en la taula 4) s'hi van observar agulles molt fines agrupades en un punt formant feixos com els de la figura 18D.

Amb la proteïna ED-4F2hc es van posar més *screenings* obtenint formes cristal·lines en diferents condicions com es pot veure a la taula 4.

<i>Tipus de cristall</i>	<i>Condició-Screening</i>	<i>Precipitant</i>	<i>Tampó</i>	<i>Sal</i>
Feixos d'agulles	22-CSI	30% PEG4000	Tris pH8.5	0.2M Acetat de Sodi
Plaques petites	32-CSII	1,6M sulfat d'amoni	HEPES pH7,5	0,2M Clorur de sodi
Agulles petites	1-PEG ION	20% PEG3350		0.2M Fluorur de sodi
Agregats d'agulles	3-PEG ION	20% PEG3350		0.2M Fluorur d'amoni
Roseta de plaques	9-PEG ION	20% PEG3350		0.2M Clorur d'amoni
Plaques petites	23-PEG ION	20% PEG3350		0.2M Format d'amoni
Agulles petites	35-PEG ION	20% PEG3350		0.2M Sulfat d'amoni
Agulles	21-INDEX			1.8M Citrat de triamoni
Agregats d'agulles	43-INDEX	25% PEG3350	Bis-tris pH6,5	
Feixos d'agulles	72-INDEX	25% PEG3350	Hepes pH7.5	0.2M Clorur de sodi
Agulles petites	79-INDEX	25% PEG3350	Bis-tris pH6,5	0,2M acetat d'amoni
Feixos d'agulles	95-INDEX	30% PEGME		0.1M Tiocianat de potassi
Feixos d'agulles	96-INDEX	30% PEGME		0.15M Bromur de potassi

Taula 4. Mostra de condicions de cristal·lització per ED4F2hc.

Paral·lelament als *screenings* es van posar gotes al voltant de les condicions de cristal·lització per mirar de tenir cristalls més grans i més ordenats, el que s'anomena procés d'optimització. Les condicions 21 del *Crystal screening I* i 23 del *PEG ION* van ser optimitzades fins a aconseguir els 2 tipus de cristalls agulles i plaques (figura 18 A i C) i aquests cristalls optimitzats van ser congelats i difractats (veure difracció dels cristalls i resolució). Les condicions de cristal·lització optimitzades són per a les agulles: 25% PEG 4000, 0,2 M d'acetat de sodi i 0,1 M tampó TRIS PH 8.5; i per a les plaques: 25% PEG 3350, 0,2 M sulfat d'amoni, tot plegat a PH 6. En el cas de les agulles, desenvolupant les concentracions de PEG 4000 i acetat sòdic, la concentració de la proteïna, el volum i la proporció de la gota, fins a unes 325 gotes en total, vam aconseguir millorar la mida dels cristalls (fig 18 A) fins a tenir una

espècie de plaques llargues, suficientment grans en dos dimensions però molt primes en la tercera. A més no vam aconseguir baixar el número de nucleació de manera que no obteníem cristalls aïllats sinó grups de cristalls que es podien separar mecànicament (fig 18 B).

En la resta de condicions de cristal·lització dels *screenings* provats és van obtenir altres cristalls, però, o bé tenien una forma semblant als anteriorment esmentats i no es van aconseguir optimitzar (exemples figura 18 D, E i F), o bé els cristalls no difractaven a tan bona resolució (figura 18 G, H i I).

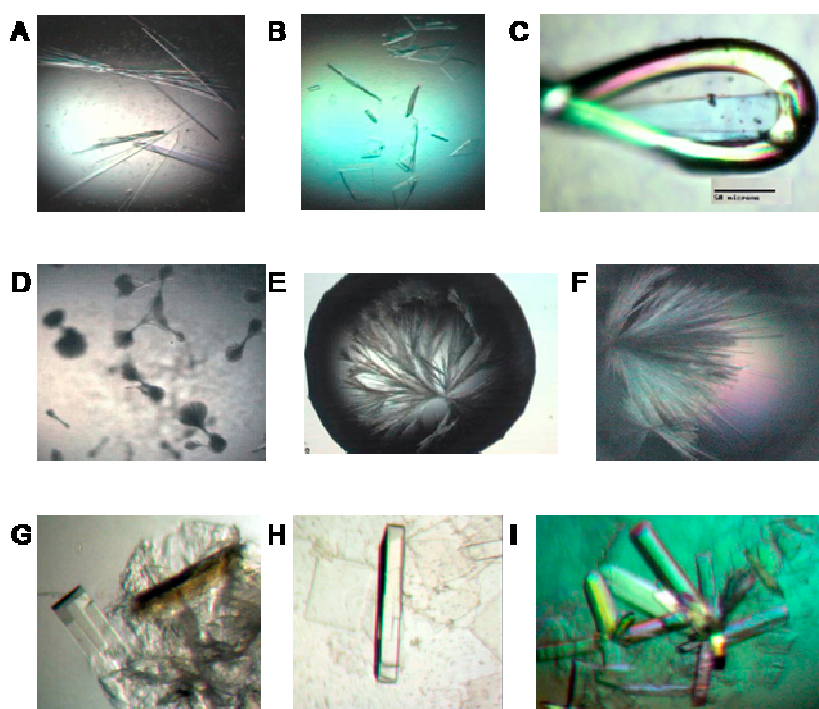


figura 18. Fotografies de diferents condicions de cristal·lització. A i B, els dos tipus de cristalls optimitzats, agulles i plaques, que van difractar a bona resolució. C, fotografia del cristall en forma d'agulla, dins d'una nansa de congelació, en el moment de ser difractat . Les fotografies C, D, E i F són exemples dels cristalls obtinguts en diferents condicions dels *screenings* de cristal·lització, sense optimitzar. Els cristalls de G, H i I són cristalls posteriors que mai van difractar a suficient resolució per poder resoldre'ls.

2.2. Proves de cristal·lització de 4F2hc glicosilat

Amb la proteïna produïda en cèl·lules 293T es van posar screenings (Crystal Screen I, Crystal screen II, PEG ION screen i Crystal screen Lite de Hampton) de

cristal·lització mitjançant un robot de cristal·lització. En dues condicions (taula 5) d'aquests screenings hi ha uns esferòlits molt grans i microcristalls que no s'han arribat a millorar (figura 19).

<i>Tipus de cristall</i>	<i>Condió-Screening</i>	<i>Precipitant</i>	<i>Tampó</i>	<i>Sal</i>
Esferòlits	26-PEG ION	20% PEG3350		0.2M Acetat de zenc
Microcristalls	45-Crystal Lite	9% PEG8000	0,1M cacodilat pH6,5	0.2M Acetat de zenc

Taula 5. Mostra de condicions de cristal·lització per ED4F2hc glicosilat.

Aquests cristalls serien interessants de difractar i veure si es poden resoldre les estructures de la proteïna amb els glúcids que té units en asparragines. A més, les dues condicions de cristal·lització tenen com a sal l'acetat de zenc, i seria interessant veure què hi fa el zenc en aquestes estructures.

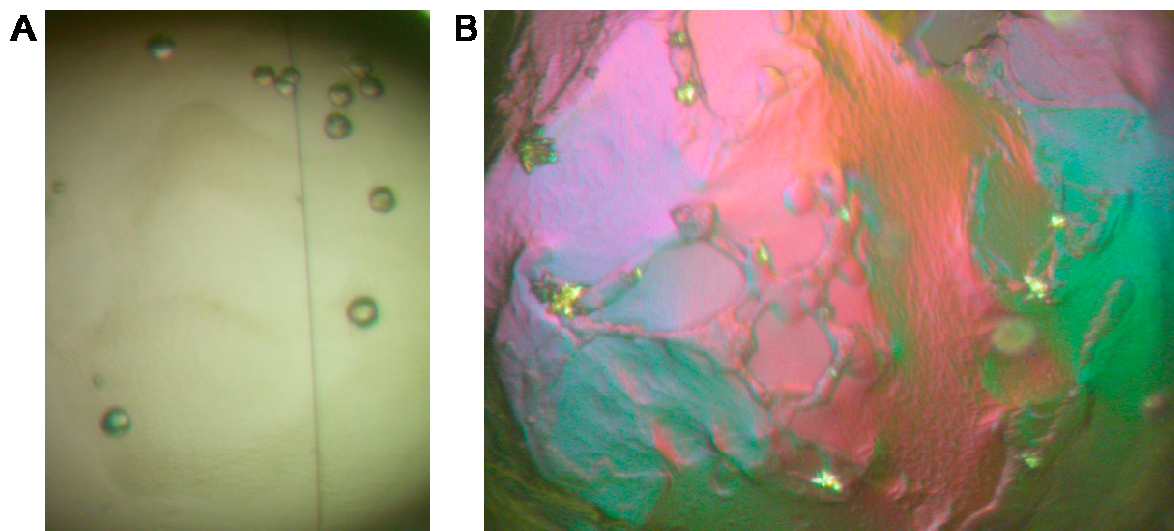


Figura 19. Dos gotes amb esferòlits (A) i microcristalls (B).

2.3. Difracció dels cristalls i processament dades

Donades les característiques dels cristalls, que ja hem comentat a l'apartat anterior, la difracció de raigs X es va fer a la línia microfocussada ID13 de l'ESFR (Grenoble), línia especial per a la difracció de cristalls petits almenys en alguna de les seves dimensions, com plaques molt primes o agulles llargues però molt primes. Els

cristalls en forma d'agulla o placa llarga (figura 18 A i B) van difractar a 1.9 Å (tot i que en l'escalat només vam poder aprofitar les dades fins a 2.1 Å) i el cristall va resultar ser de tipus monoclínic. En el cas de les plaques (figura 18C) els cristalls van difractar a 2.8 Å i eren de tipus ortoròmbic (taula 6).

Les imatges van ser analitzades i escalades internament (taula 6). Els cristalls van resultar ser del grup $P2_1$ (el monoclínic) i $P2_12_12_1$ (l'ortoròmbic) i contenien una i dues molècules per unitat asimètrica, respectivament, amb un contingut de solvent al voltant del 40% (coeficient de Mathews) (taula 6). En els dos tipus de cristalls la proteïna es trobava molt empaquetada, cosa que veiem amb el baix contingut de solvent.

El problema de la fase és el punt més crític en la resolució d'un cristall. Per resoldre'l hi ha diverses aproximacions. Nosaltres vam optar, en un primer moment, per produir la proteïna amb selenometionines en lloc de metionines i produir nous cristalls. Les condicions de cristallització en el cas dels cristalls en forma d'agulla (monoclínic) produïts amb aquesta proteïna van ser idèntiques a les de la proteïna nativa.

Amb els cristalls amb la proteïna amb selenometionina vam provar de fer MAD (*multiple-wavelength anomalous dispersion*) però no van aguantar. Aquest experiment consisteix a prendre dades de difracció a diferents longituds d'ona i és necessari irradiar més vegades el cristall. Segurament a causa de les mides reduïdes el cristall, aquest es cremava a mida que s'anava irradiant i la intensitat minvava.

Malgrat no aconseguir dades de MAD, els cristalls produïts amb selenometionina van difractar millor que els nadius (aquest canvi no necessàriament és degut a la presència del seleni; pot ser que la preparació de proteïna fos de més bona qualitat). Per aquest motiu el cristall monoclínic del qual es va resoldre l'estructura, el que tenia les dades de difracció més completes i amb més resolució, era un cristall amb selenometionines en lloc de metionines, com és pot veure en el PDB 2DH2 que conté selenis en el lloc dels sofres de les metionines.

<i>Típus de cristall</i>	<i>Agulles</i>	<i>Plaques</i>
Sistema	MONOCLÍNIC	ORTORÒMBIC
Grup espacial	P2 ₁	P2 ₁ 2 ₁ 2 ₁
Molècules per unitat asimètrica	1	2
Rang de resolució	17.83 - 2.10 Å	25.0 - 2.80 Å
Número de reflexions	20986	19304
Completesa	97.3%	94.6%
Dimensions de la cel·la	a=39.7 b=67.8 c=74.0 $\alpha=90 \beta=98.5 \gamma=90$	a=69.5 b=101.8 c=122.0 $\alpha=90 \beta=90 \gamma=90$
Contingut de solvent	41.3%	46.6%
Average $\langle I/\sigma \rangle$	14	15.74
PDBID	2DH2	2DH3
Valor R	17.8	22.0
Valor R _{free}	23.0	27.5

Taula 6. Recollida de dades i resolució de l'estructura

2.4. Resolució de l'estructura

Pel fet de no tenir dades de MAD, que permet resoldre el problema de la fase mitjançant la difracció anòmala de les selenometionines, vam intentar trobar el senyal del seleni que contenien les selenometionines per tal de resoldre la fase per reemplaçament isomorf. Però aquest senyal era massa dèbil i segurament era pel poc contingut en metionines de la proteïna (4 metionines d'un total de 425 aminoàcids).

Malgrat això, en la fluorescència de raigs X es veia un pic clar per aquest metall, el que significava que la proteïna havia incorporat el seleni, tot i que en el mapa de difracció aquest no tenia suficient intensitat per ser distingit.

Per solucionar el problema de la fase també es pot usar el mètode de reemplaçament molecular. Aquest mètode consisteix en resoldre la fase mitjançant

un model semblant al que volem trobar. Així, per utilitzar aquest mètode es necessiten estructures tridimensionals resoltes que tinguin homologia significativa amb la proteïna d'interès. En el nostre cas teníem algunes glucosidases de bacteris que estaven resoltes per raigs X. La homologia de seqüència entre aquestes proteïnes i l'ectodomini de 4F2hc era molt baixa (al voltant del 25% d'identitat) però vam observar que la conservació de l'estructura entre membres de la família era molt més alta que la identitat de seqüència, la qual cosa ens va fer suposar que també existiria aquesta conservació en la nostra proteïna. Partint d'aquesta hipòtesi es van utilitzar les estructures amb PDBID 1JI2, 1UOK i 1SMA (neopullulanase de *Bacillus Subtilis*, Oligo-1,6-glucosidasa de *Bacillus Cereus*, i α -amylase de *Thermus sp.*). Aquestes estructures tenien una identitat entre elles del 28 % entre 1UOK i les altres dues i d'un 51 % entre 1JI2 i 1SMA, com podem veure en taula 7; en canvi tenien un RMSD (*Root mean square deviation*) al voltant de 2. En comparar-les amb l'ectodomini de 4F2hc veiem que la identitat de seqüència és una mica més baixa, entre un 22 i un 25 %.

Aquests PDBs es van utilitzar per fer reemplaçament molecular amb el servidor Caspr (CNRS, Marseille) que permet utilitzar un model mixt fet entre diverses estructures i tenint en compte la seqüència de la proteïna (figura 20). Caspr va fer multialineament de seqüència amb la informació estructural i va truncar algunes zones de la proteïna problema que no tenien cap zona homòloga amb les estructures d'entrada. Després va utilitzar l'alineament de seqüència i la informació de les estructures per fer modelat de proteïnes i va treure 30 models amb la seqüència sencera d'ED-4F2hc i 30 amb la forma truncada, i 3 models amb la seqüència dels PDBs d'entrada i 3 més amb la seqüència d'aquests truncada. Aquests models els va utilitzar per fer reemplaçament molecular amb les dades del cristall monocínic. El mateix servidor va agafar els millors 15 models i va fer un primer refinament. De tot aquest procés va sortir un millor model refinat amb un R_{work} de 0.45 i un R_{free} de 0.52, encara lluny de la solució final ja que si el comparem amb el model final de tot el procés tenen un RMSD de 2.41Å.

Identitat de seqüència				
Estructura	1SMA	1UOK	1JI2	2DH2
1SMA		28.2	51.0	22.3
1UOK	28.2		27.4	22.3
1JI2	51.0	27.4		25.0
ED-4F2hc	22.3	22.3	25.0	
RMSD				
Estructura	1SMA	1UOK	1JI2	2DH2
1SMA		2.356	1.343	2.682
1UOK	2.356		2.134	2.712
1JI2	1.343	2.134		2.515
2DH2	2.682	2.712	2.515	

Taula 7. Matriu d'identitat de seqüència (en %) i de RMSD en Å (root mean square distance) entre els diferents models estructurals usats en la resolució de l'estructura de ED-4F2hc (1JI2 neopullulanase de *Bacillus Subtilis*, 1UOK Oligo-1,6-glucosidasa de *Bacillus. Cereus* i 1SMA α -amylase de *Thermus sp.* Superposició de les estructures feta amb el programa SSM (Secondary-Structure Matching, *on line* a <http://www.ebi.ac.uk/msd-srv/ssm>).

A partir d'aquest primer model calculat per Caspr vam continuar treballant el reemplaçament molecular amb MOLREP. També vam fer refinaments de cos rígid amb dos dominis per separat, el barril $(\alpha/\beta)_8$ i el C-terminal (característics de les glucosidases homòlogues) i després definint diferents motius d'estructura secundària (ex. fulls β del barril $(\alpha/\beta)_8$, hèlixs alfa del barril $(\alpha/\beta)_8$...) fins a obtenir un model pel cristall monocínic amb R 0.48, R_{free} de 0.48 i una correlació de 0.45 a 2.8 Å.

Amb aquest model vam fer reemplaçament molecular amb les dades de l'ortoròmbic. Aquest cristall mostrava 2 molècules per unitat asimètrica i després de refinar el model ortoròmbic tenia un R 0.54, R_{free} de 0.52.

Els mapes de densitat que sortien amb aquests models eren molt pobres i no tenien

suficient continuïtat per construir-hi el model. Per millorar les fases que teníem en aquells moments es va utilitzar el mitjanat entre cristalls, de manera que podíem barrejar les dades del cristall monoclínic i les dades de les dues molècules de la unitat asimètrica del cristall ortoròmbic. El mapa de densitat resultant d'aquest mètode ja permetia reconstruir moltes zones del model monoclínic que ja tenien continuïtat. Mitjançant reconstrucció manual anàvem repetint el mitjanat entre cristalls amb models cada vegada una mica millors i refinant-los, i els mapes eren cada vegada més clars i permetien un millor retraçat. Finalment es van afegir les aigües i es va refinar el model final fins a R 0.18, R_{free} de 0.23.

Amb aquest model es va tornar a fer reemplaçament molecular amb les dades del cristall ortoròmbic i les dues molècules de la unitat asimètrica van ser retraçades per separat en alguns punts i s'hi va afegir un àtom de Zn^{2+} i algunes molècules d'aigua. El model ortoròmbic va ser refinat fins a un R 0.22, R_{free} de 0.27.

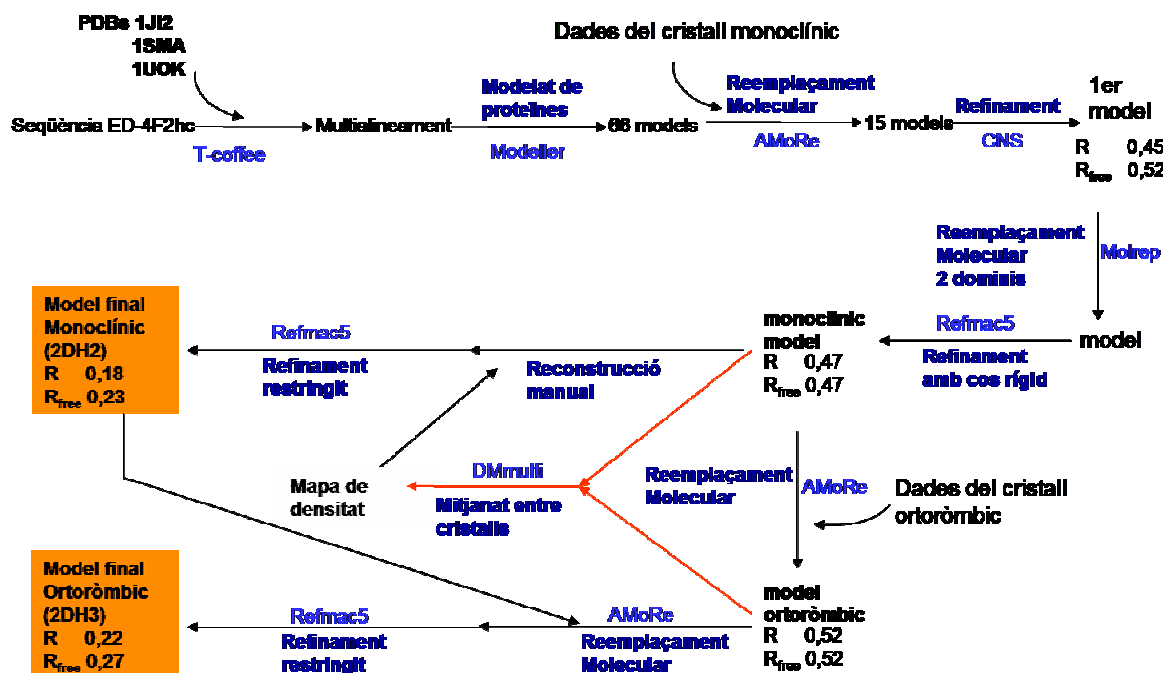


Figura 20. Esquema resum dels passos seguits per resoldre les estructures dels dos cristalls monoclínic i ortoròmbic d'ED4F2hc. En blau marí veiem les tècniques utilitzades i en blau cel els programes que hem usat per dur-les a terme. Els models finals estan dipositats en la base de dades de proteïnes PDB amb les IDs indicades. Els passos seguits fins a obtenir el 1er model es van fer mitjançant el servidor automatitzat Caspr.

Capítol 3. Característiques de l'estructura d'ED4F2hc

3.1. Descripció estructura

L'estructura de l'ectodomini de 4F2hc pot ser descrita en termes d'organització de dominis de les α -amilases. L'anomenat domini A en aquesta família és un barril TIM $(\beta/\alpha)_8$, compost per 8 fulls beta interns rodejats de 8 hèlixs alfa i en ED4F2hc el trobem a N-terminal des de l'aminoàcid triptòfan 117 a la lisina 431. Seguint aquest domini i després d'un petit fragment que els uneix, trobem l'anomenat domini C, representat per una topologia de β -sandwich, que en 4F2hc està format pel fragment d'aminoàcids aspàrtic 439 fins a l'alanina 529, i en formen part 8 fulls beta antiparal·lels (figura 21).

Els primers aminoàcids que podem veure en l'estructura adopten una conformació estesa i apunten cap a la interfase dels dominis A i C (figura 21). El barril $(\beta/\alpha)_8$ en 4F2hc comença amb una petita hèlix on hi ha els dos triptòfans 117 i 118, característica estructural molt conservada en la família de les glicosilhidrolases, com podem comprovar en l'estructura de l'oligo-1,6-glucosidasa de *B. cereus* (figura 6).

En els fragments d'aminoàcids que uneixen els fulls β i les hèlixs α següents en el barril $(\alpha/\beta)_8$ (bucles) en ED4F2hc hi trobem alguns elements d'estructura secundària poc importants. Molts membres de la família de les α -amilases tenen un altre domini en el bucle que es troba entre $A\alpha_3$ i $A\beta_4$ anomenat domini (o subdomini) B. Per alineament de seqüència sembla que l'ectodomini d'rBAT tindria aquest domini B molt semblant al de la oligo-1,6-glucosidasa de *B. cereus*, però en 4F2hc aquesta seqüència i estructura és inexistent i els bucles més importants són els que van d' $A\beta_2$ a $A\alpha_2$ i de la d' $A\beta_8$ a $A\alpha_8$. Entre el barril $(\alpha/\beta)_8$ i el tot β hi trobem una petita hèlix conservada entre les diferents estructures de la família.

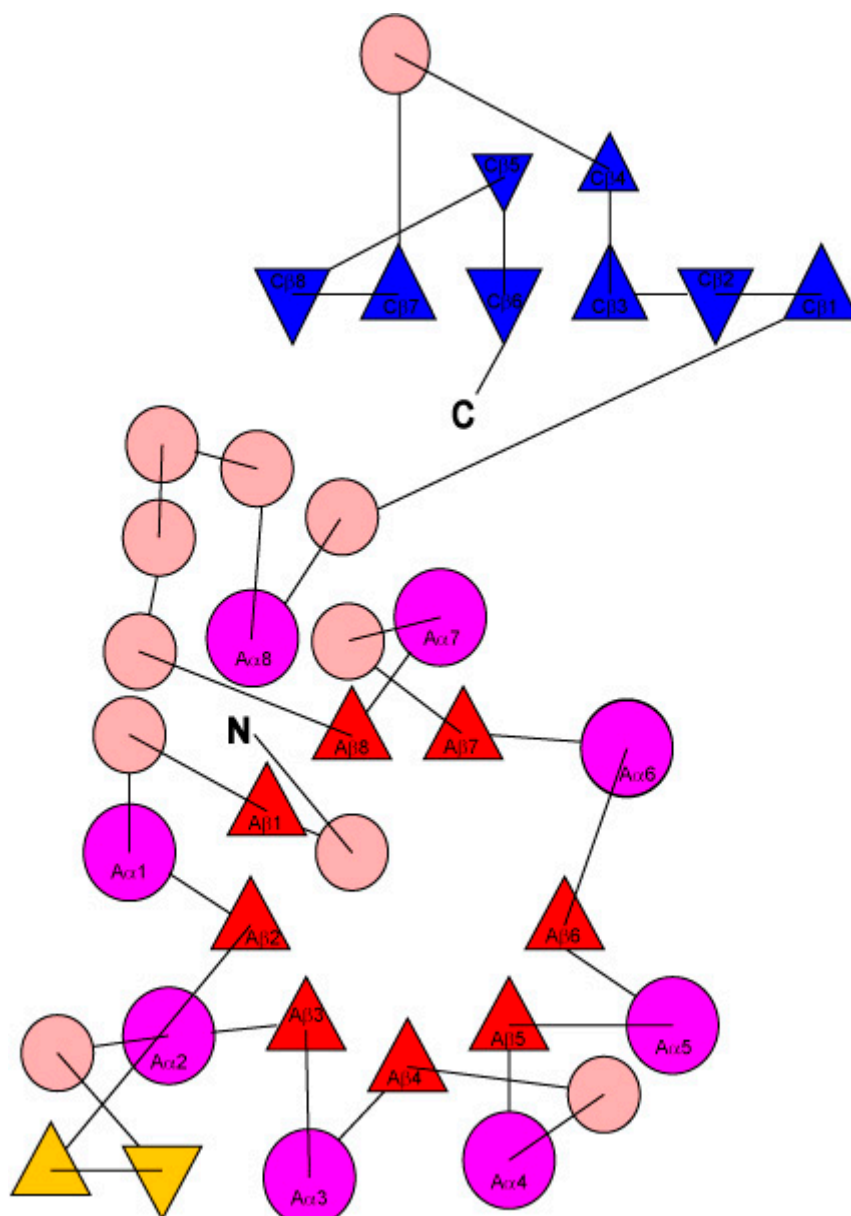


Figura 21. Topologia de l'estructura d'ED4F2hc. En vermell hi ha representats els 8 fulls beta de l'interior del barril (β/α)₈ i en rosa fúscia 8 les hèlixs alfa que les envolten. En blau marí es mostren els fulls 8 beta del domini C. En rosa pàl·lid i groc podem veure les hèlixs alfa i els fulls beta que formen part dels bucles.

Quan fem una superposició dels tres models independents que hem trobat en els cristalls (un model monocínic i les cadenes A i B en l'ortoròmbic) les tres presenten un RMSD mitjà per 409 aminoàcids alineats de 0,958 Å. Observem moltes zones idèntiques, però d'altres tenen diferències en l'estructura (figura 23). L'hèlix Aα5 (figura 23C línia de punts) presenta en la molècula B diferències importants i també veiem diferències entre les 3 molècules al *bucle* entre la Aβ6 i la Aα6. Una altra zona

de diferències és la part accessible del domini C-terminal (figura 23C línia discontinua), sobretot entre la L473 i l'A490 on trobem diferències entre les tres molècules però sobretot en la molècula A, possiblement a causa d'una coordinació de zenc que es troba en els fulls beta antiparal·lels adjacents (veure apartat 3.3 de resultats). Si ens fixem en l'N-terminal veiem que presenta diferents conformacions. El de la molècula A del cristall ortoròmbic és més curt que el de les altres dues estructures ja que en el mapa de densitat no veiem cap densitat precedint la de l'A114. Això podria indicar que aquest fragment de 5 residus (G109-P113) és mòbil en la molècula A. De fet, els residus de K105 a W108 no són visibles en cap de les tres molècules (figura 22). Alguns dels *bucles* superiors del barril $(\alpha/\beta)_8$ també presenten diferències però són fragments de molt pocs residus.

106	109	117	APRCRELPAQKWW	barril $(\alpha/\beta)_8$ h4F2hc
			DRWGSELPAQKWW	barril $(\alpha/\beta)_8$ ED4F2hc
			GSELPAQKWW	barril $(\alpha/\beta)_8$ 2DH2
			A QKWW	barril $(\alpha/\beta)_8$ 2DH3 mol A
			GSELPAQKWW	barril $(\alpha/\beta)_8$ 2DH3 mol B

Figura 22. Esquema de les seqüències N-terminals dels diferents models estructurals i proteïnes a partir de la posició 106 fins a la 118. En grana s'hi representa la seqüència de 4F2hc humà a partir de l'alanina 106. En taronja s'hi representa la seqüència N-terminal de la proteïna produïda en *E.coli* i digerida amb l'enteroquinasa. En blau hi ha la seqüència dels primers aminoàcids especificats en el model estructural del cristall monoclínic 2DH2 i en groc i verd el mateix però de les dues molècules (A i B) que trobem en l'estructura 2DH3, del cristall ortoròmbic.

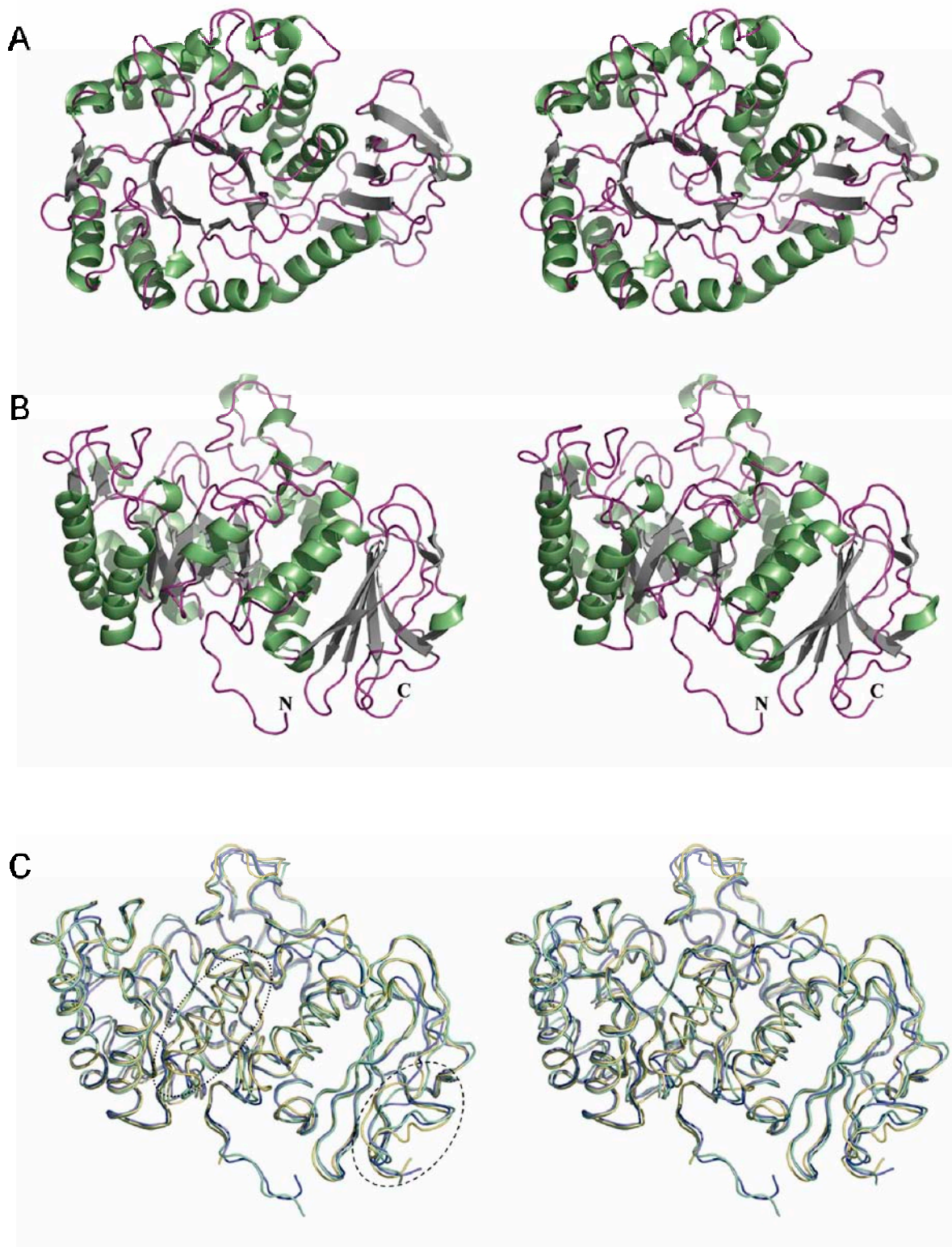


Figura 23. Representacions estèreo de l'estructura de ED-4F2hc. **A.** vista superior del barril (α/β)₈. **B.** Vista lateral **C.** Superposició de les tres molècules d'ED-4F2hc trobades en el cristall: en blau, model del cristall monoclínic; en groc i en verd, cadenes A i B del cristall ortoròmbic. Encerclats amb línies dues zones amb diferències que se citen en el text.

La superfície de l'ectodomini de 4F2hc té algunes característiques interessants de comentar. En la figura 24 podem veure com en la part superior del barril $(\alpha/\beta)_8$ (esquerra) (entenent inferior com la part N-terminal dels fulls β del barril i superior com la part C-terminal d'aquests) hi dominen les càrregues negatives (vermell), en canvi en la part inferior (dreta) hi veiem una distribució de càrregues molt més positiva (blau) deguda a les cadenes laterals d'algunes lisines i arginines que es troben en aquesta regió. En aquesta regió també es troba l'N-terminal de l'ectodomini i per tant també hauria de ser la superfície més propera a la membrana (veure figura 9). Aquestes càrregues positives podrien interaccionar amb els fosfats de la membrana de fosfolípids de la cèl·lula.

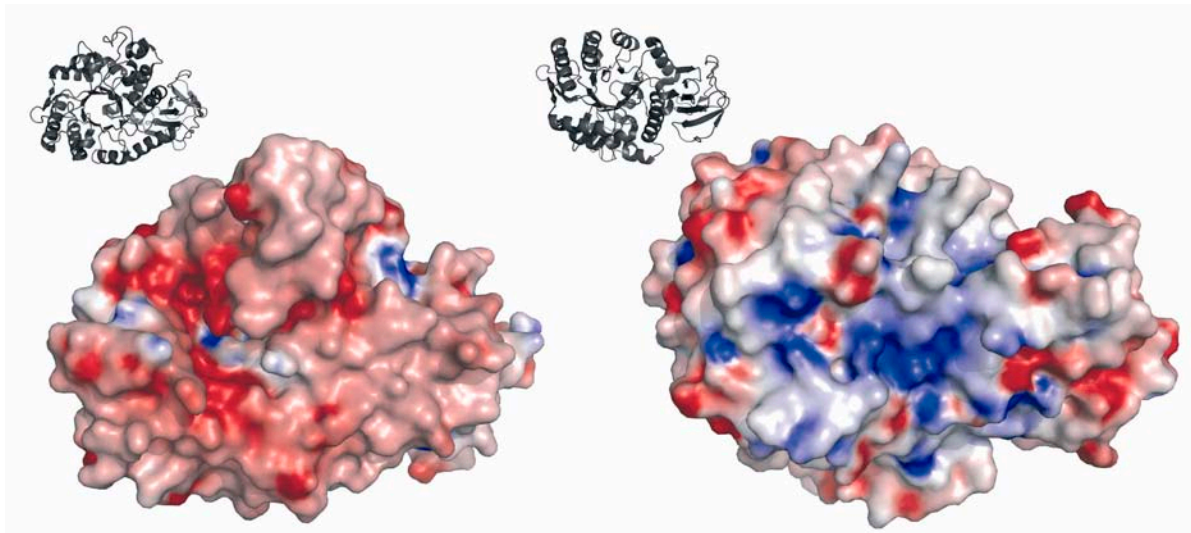


Figura 24. A la dreta visió superior del barril $(\alpha/\beta)_8$ i a l'esquerra visió inferior, des de l'N-terminal. Càlcul de potencial electrostàtic amb GRASP.

3.2. Detalls del cristall monoclínic

Com hem vist, la quantitat de solvent en el cristall monoclínic és baixa (41%) i, per tant, era de preveure un empaquetament molt compacte, com podem veure en la figura 25. L'empaquetament és tant compacte que es difícil veure la separació entre les diferents molècules, que tenen moltes zones de contactes cristal·logràfics (figura 26).

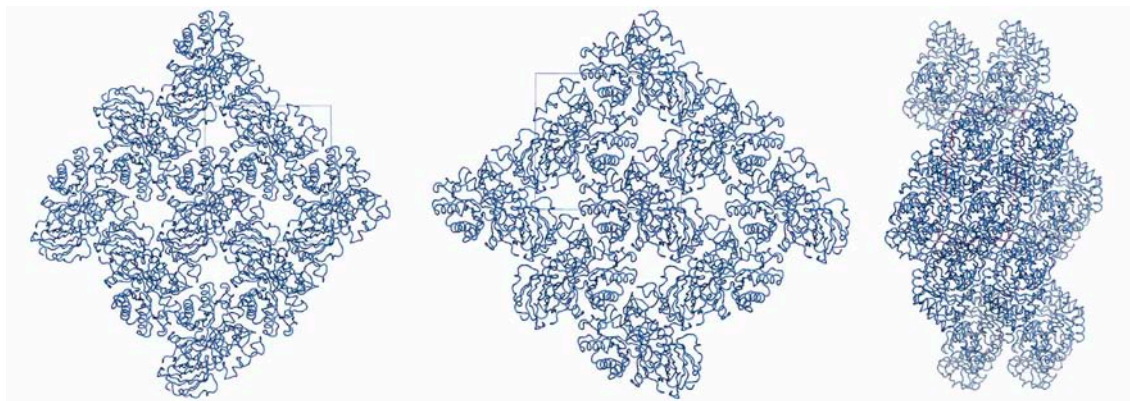


Figura 25. Disposició de les molècules d'ED-4F2hc en el cristall monoclínic. D'esquerra a dreta: visió al llarg dels eixos de simetria a, b i c.

En el cristall el 20% de la superfície accessible té contactes amb altres molècules del cristall. Aquests contactes es donen sobretot en els bucles superiors del barril (α/β)₈ (figura 26), però també en la zona inferior del domini C i en l'hèlix A α 5.

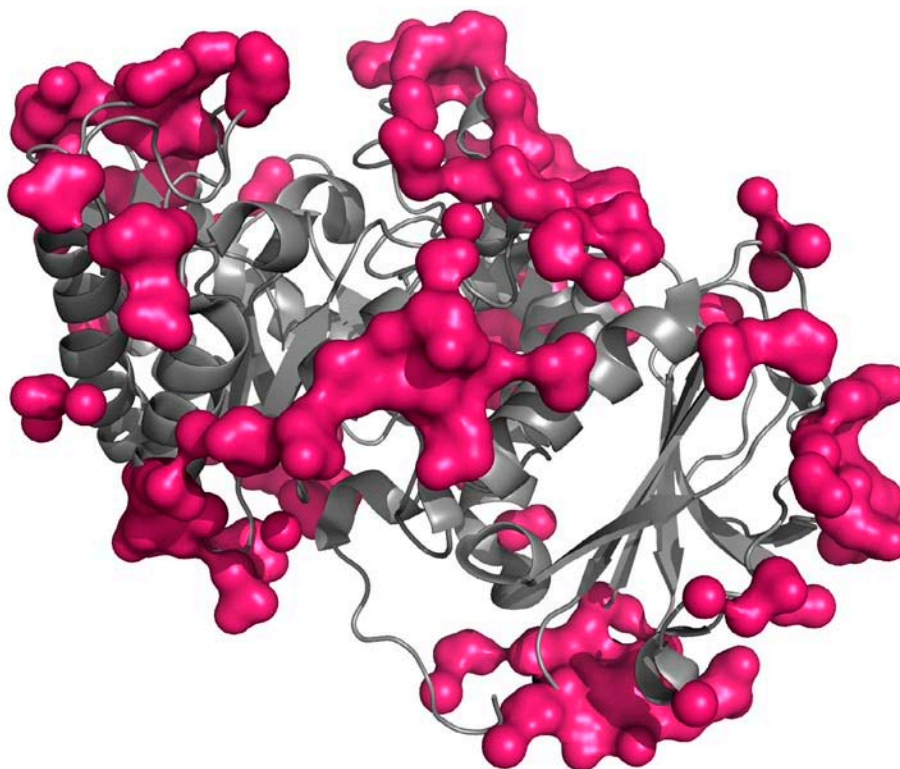


Figura 26. Superfície de contacte. En rosa fúcsia s'hi representa la superfície dels àtoms que estan en contacte amb altres molècules del cristall.

El mapa de densitat obtingut amb les dades a 2.1 Å mostra molts detalls de l'estructura. Es poden veure molt ben delimitades les cadenes laterals i també les aigües que trobem formant ponts d'hidrogen amb la proteïna (figura 27).

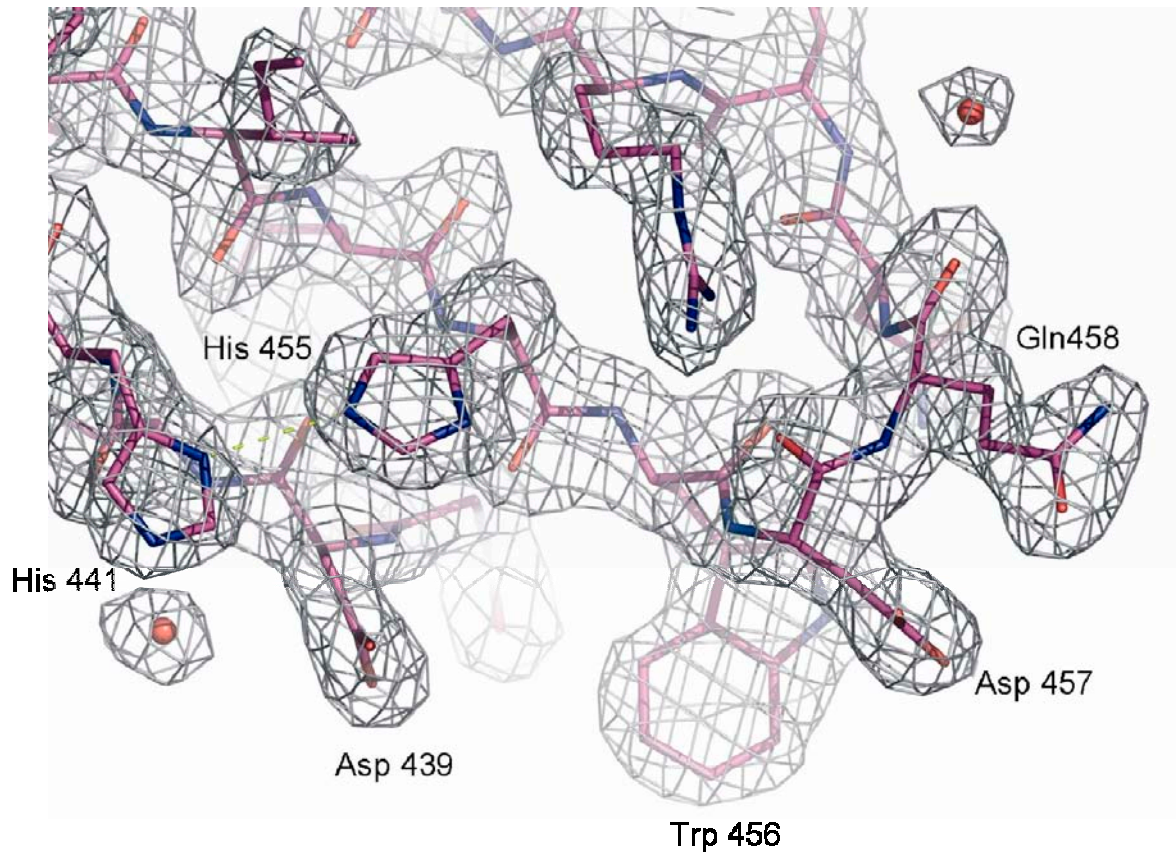


Figura 27. Detall del mapa de densitat ($\sigma=1$) cristall monoclinic a 2.1 Å. Zona inferior del domini C-terminal. S'hi poden veure les 2 histidines (His441 i His455) i l'aspàrtic (Asp 439) que en el cristall ortoròmbic estan participant en la coordinació amb zinc. La línia discontinua en groc dibuixa un pont d'hidrogen entre les dues histidines. El bucle que formen el Trp456, l'Asp457 i la Gln458 està molt conservat entre tots els ortòlegs de verèbrats de 4F2hc.

Nogensmenys hi ha algunes zones on no hi veiem densitat o on les cadenes laterals no estan tan ben definides, segurament perquè pertanyen a zones d'alta mobilitat de la proteïna. Això té sentit si ens fixem en quins són aquests fragments de l'estructura, bàsicament la cua N-terminal, els bucles exteriors del barril $(\alpha/\beta)_8$ i alguna zona exterior del domini C-terminal. Aquests llocs on la densitat no dona prou informació també es poden explicar amb un factor de temperatura, o factor B, que trobem en el pdb per a cada àtom representat (figura 28). Aquest factor depèn de la mobilitat atòmica en el cristall. Les dues hèlixs A α 5 i A α 6 tenen les zones inferiors i

superiors i també els bucles que les uneixen a la resta de l'estructura amb uns factors de temperatura alts. Una altra zona mal definida és el *bucle* entre A β 8 i A α 8 exactament des de l'aminoàcid 374 al 386. També veiem la cua N-terminal mal definida per la densitat i una zona en el domini C-terminal, la petita hèlix que es troba entre la C β 4 i la C β 5. Totes aquestes zones es troben en l'exterior de la molècula, exposades al solvent, i per aquesta raó la seva mobilitat pot ser més gran i el factor de temperatura també.

Si ens hi fixem, aquestes zones també corresponen a les zones que presenten més diferències entre els tres models d'ED-4F2hc. Tot apunta a pensar que són zones d'alta mobilitat.

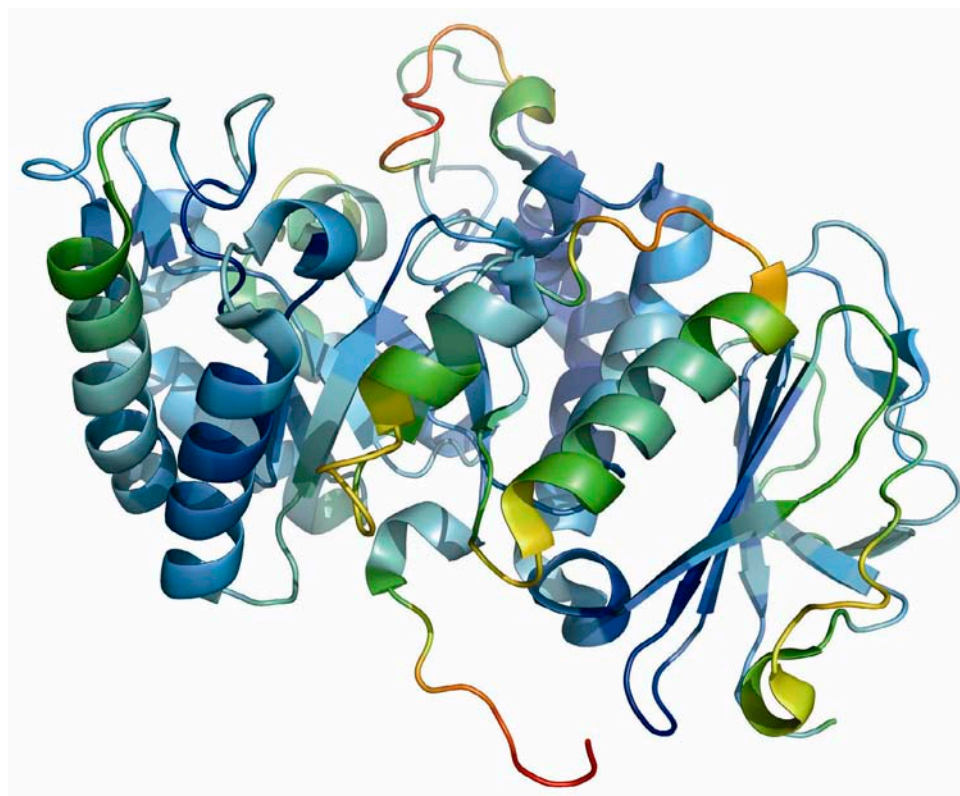


Figura 28. Estructura d'ED4F2hc amb els factors de temperatura pels C α amb una escala de colors. En aquesta figura hi estan representats els factors de temperatura al llarg de la cadena principal de la proteïna. La coloració va des del blau al vermell, passant pel verd i el groc, que correspon des dels factors B més baixos fins als més alts o, el que és el mateix, de les zones més ben definides a les menys definides.

En la figura 29 es representa el mapa de Ramachandran per al model 2DH2. la gran majoria dels residus de la molècula es troben en les seves zones característiques, mentre que uns quants s'hi troben fora. Cal destacar la Phe 218 que està

perfectament definida per la densitat electrònica en una zona que no li és característica. La resta de residus que es troben fora estan en zones mal definides per la densitat electrònica.

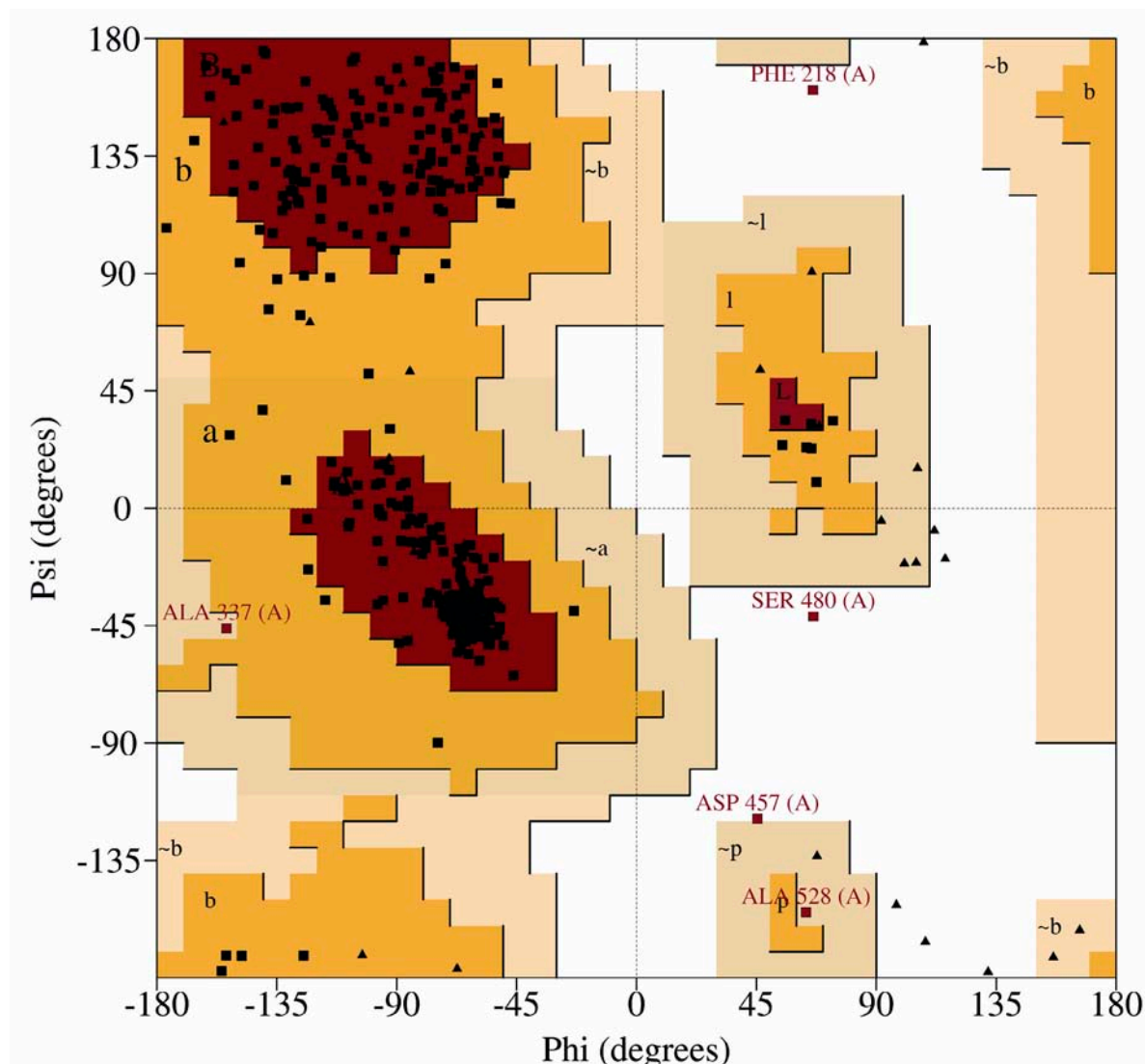


Figura 29. Mapa de Ramachandran de l'estructura d'ED4F2hc del cristall monoclínic (2DH2).

3.3. Detalls del cristall ortoròmbic

Per resoldre el cristall ortoròmbic a 2,8 Å, vam utilitzar el model que havíem obtingut en el monoclínic per fer reemplaçament molecular (figura 20). En resoldre'l vam veure que aquest cristall, a diferència de l'anterior, contenia dues molècules per unitat asimètrica. Per poder refinar l'estructura d'aquestes dues molècules es van haver de reconstruir a mà en algunes zones on el mapa de densitat d'aquest cristall no es corresponia amb el model monoclínic. El cristall també era bastant compacte,

amb un 47% de solvent i un 21% de l'àrea accessible tenia contactes en el cristall.

Les dues subunitats cristal·logràficament independents que es troben en aquest cristall estan relacionades per un eix quasi-binari i gairebé paral·lel a l'eix cristal·logràfic *c* (figura 31). Això explica que en el mapa de Patterson natiu s'observa un pic de pseudoorigen a (0.0, 21.6, 53.6) que era més important com més alta era la resolució, ja que a 7 Å el pic significava un 18% de l'origen i a 3,1 Å un 12,5%. Sorprenentment el lloc on es trenca la simetria no cristal·logràfica d'aquest eix quasi binari (amb un rmsd entre les dues molècules A i B de 1.11 Å) és principalment al voltant de l'eix de simetria local. Per consegüent no són tant els contactes cristal·logràfics sinó la presència d'un àtom de zinc que es troba coordinat entre les dues subunitats. Aquesta coordinació, que s'observa perfectament en el mapa de densitat (figura 31C) és tetraèdrica i hi participen tres residus (Asp439, His441 i His455) de la primera subunitat i només un de la segona (His441) (veure detall coordinació en la figura 31).

La presència d'aquest àtom de zinc era completament inesperada, ja que ni en la solució de cristal·lització ni al tampó de la proteïna no hi havia zinc. La concentració de zinc present en la gota de cristal·lització, doncs, havia de ser molt baixa, a nivells traça. Per aquest motiu vam fer espectroscopia de fluorescència de raigs-X per verificar la presència d'aquest metall, i no d'un altre de semblant, en un cristall del tipus ortoròmbic. En la figura 30 es mostra l'espectre de fluorescència de raigs X d'un cristall d'aquest tipus. Els pics resultants s'ajusten als pics d'absorció característics del Zn²⁺ (9.659 keV $f' = -8.1$ eV i $f'' = 3.9$ eV)

Totes aquestes observacions indiquen que l'afinitat d'aquestes molècules pel zinc ha de ser més alta de picomolar. Però si això fos així hauríem de tenir l'àtom de zinc també coordinant els tres residus en el cristall monoclínic, ja que es podria donar el cas que una molècula d'aigua estigués actuant com a quart lligand, encara que suposés una pèrdua d'afinitat. Però en el mapa de densitat del cristall monoclínic no hi ha rastre de la presència d'aquest metall, com es pot observar clarament en la figura 27.

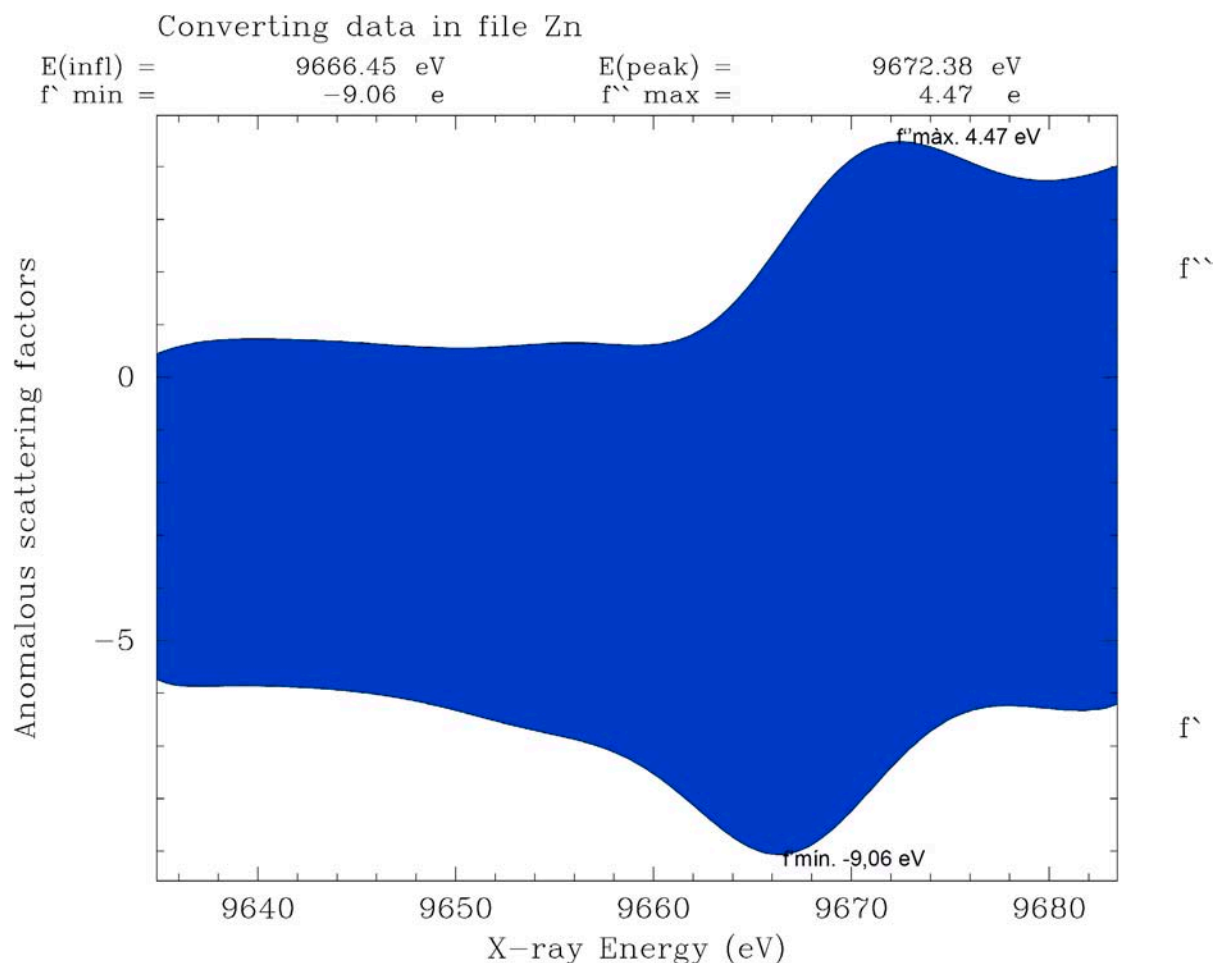


Figura 30. Espectre de fluorescència de raigs X del cristall ortoròmbic.

Aquest dímer d'ED4F2hc no es troba en solució, o almenys en una cromatografia de gel filtració d'alta resolució (superdex 200 prep grade) no es veu diferència entre la proteïna amb tampó normal o la proteïna amb EDTA afegit en el tampó. Estudis d'espectrometria de masses de la proteïna ED4F2hc en solució fets pel laboratori de la Dra. A. Lizarbe indiquen un contingut en zenc d'aproximadament el 10% de la molaritat de la proteïna, mentre que si fos un dímer hauria de ser del 50% (1 mol de zenc per cada dos mols de proteïna).

Les histidines estan, en tots tres casos, interactuant mitjançant els seus N^ε que han d'estar desprotonats. Això pot donar certa dependència de pH a aquesta interacció.

Els residus Asp 439 i His 441 estan disposats en el mateix full β (Cβ1), al costat del full antiparal·lel Cβ2 que conté la His 455. Aquest grup de residus estan en la part inferior de la molècula, entenen la part inferior com la cara N-terminal del barril (α/β)₈.

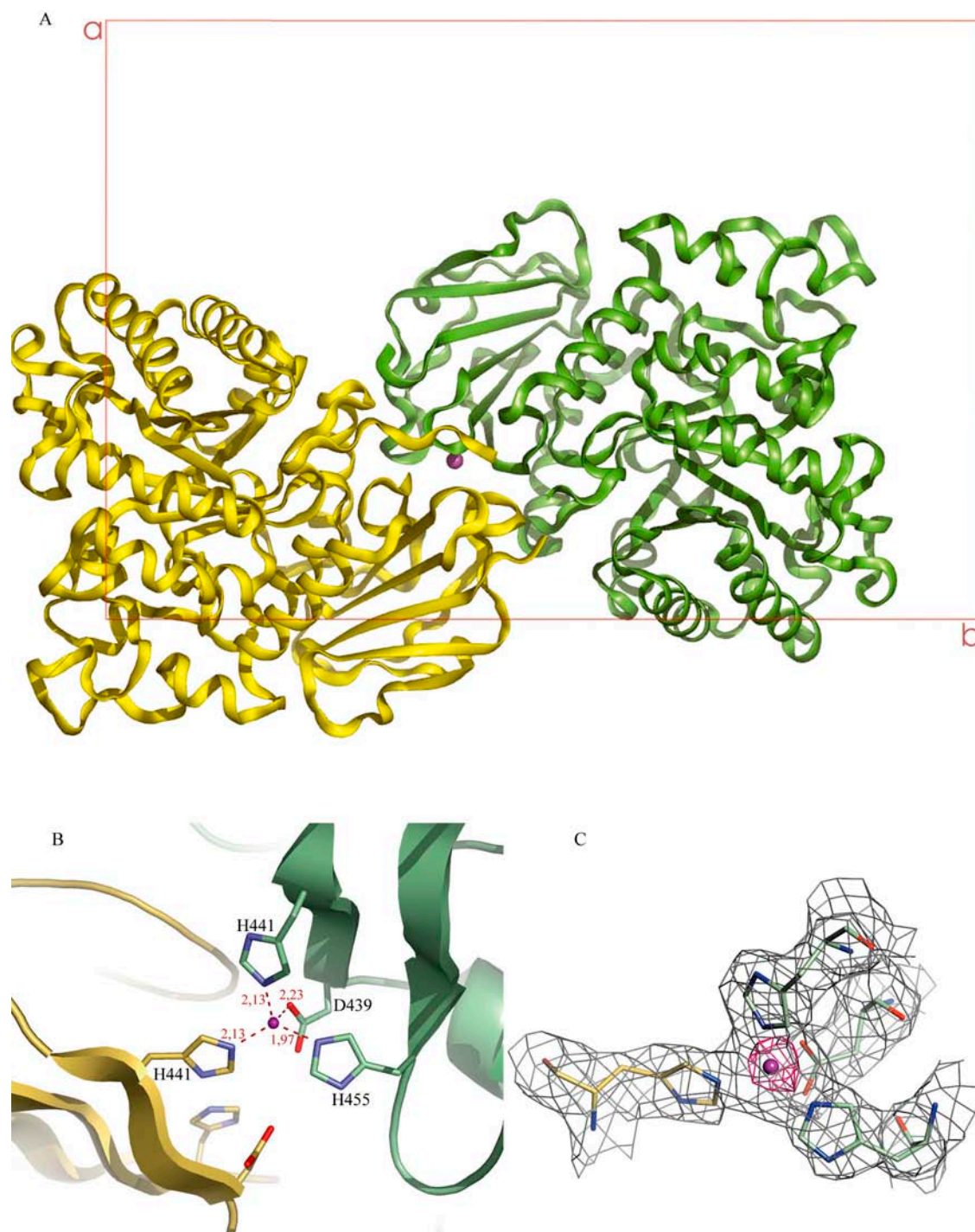


Figura 31. **A.** representació de les dues molècules de la unitat asimètrica al llarg de l'eix cristal·logràfic c, on es pot veure l'eix de la quasi simetria binària que existeix entre elles i entre les dues l'àtom de zinc en color morat. **B.** Detall dels residus que coordinen el zinc de les dues molècules. **C.** Mapa de densitats del cristall ortoròmbic on es veu clarament la coordinació tetraèdrica del zinc; en gris fosc s'hi representa el mapa de densitat a $\sigma=1$ i en rosa fúcsia el mapa a $\sigma=6$, on es veu la intensitat de la densitat d'aquest metall.

Capítol 4. Modelatge d'EDrBAT a partir de l'estructura d'ED4F2hc.

L'ectodomini d'rBAT té una identitat de seqüència, segons l'alineament de seqüència amb T-coffee, del 40% amb membres de la família de les α -amilases provenint d'insectes com la maltasa 1 de *Culex pipiens quinquefasciatus* (mosquit comú).

De les proteïnes que tenen estructures tridimensionals resoltes trobem que té una identitat de seqüència del 31% amb ED4F2hc i del 33% amb l'oligo-1,6-glucosidasa de *B. cereus* (1UOK). És sorprenent com la identitat és tan semblant entre les dues cadenes pesades de transportadors d'aminoàcids com entre rBAT i una glucosidasa de bacteris. Tot i aquesta semblança podem trobar diferències entre aquestes dues proteïnes (veure taula 8 i figura 32). En el domini barril $(\alpha/\beta)_8$ veiem una identitat del 36% amb 1UOK i més baixa en 4F2hc. Aquesta sobretot és deguda al domini protuberant B, que es troba entre $A\beta_3$ i $A\alpha_3$, ja que entre rBAT i 1UOK en aquesta regió la identitat s'eleva al 49%, i en canvi en 4F2hc aquest subdomini no existeix.

En el domini C terminal, la relació és inversa, rBAT s'assembla més a 4F2hc (34%) i bastant menys a 1UOK (17%). A més a més rBAT compta amb una cua C-terminal de 34 aminoàcids que no es troba ni en 4F2hc ni en 1UOK.

<i>domini</i>	<i>rBAT-1UOK</i>	<i>rBAT-4F2hc</i>	<i>1UOK-4F2hc</i>
barril $(\alpha/\beta)_8$ (115-561)	36	30	24
dominiB (218-286)	49	--	--
dominiC (562-654)	17	33	26

Taula 8. Identitat de seqüència entre EDrBAT,1UOK i 4F2hc per dominis. Els valors es representen en percentatge. 4F2hc no té domini B.

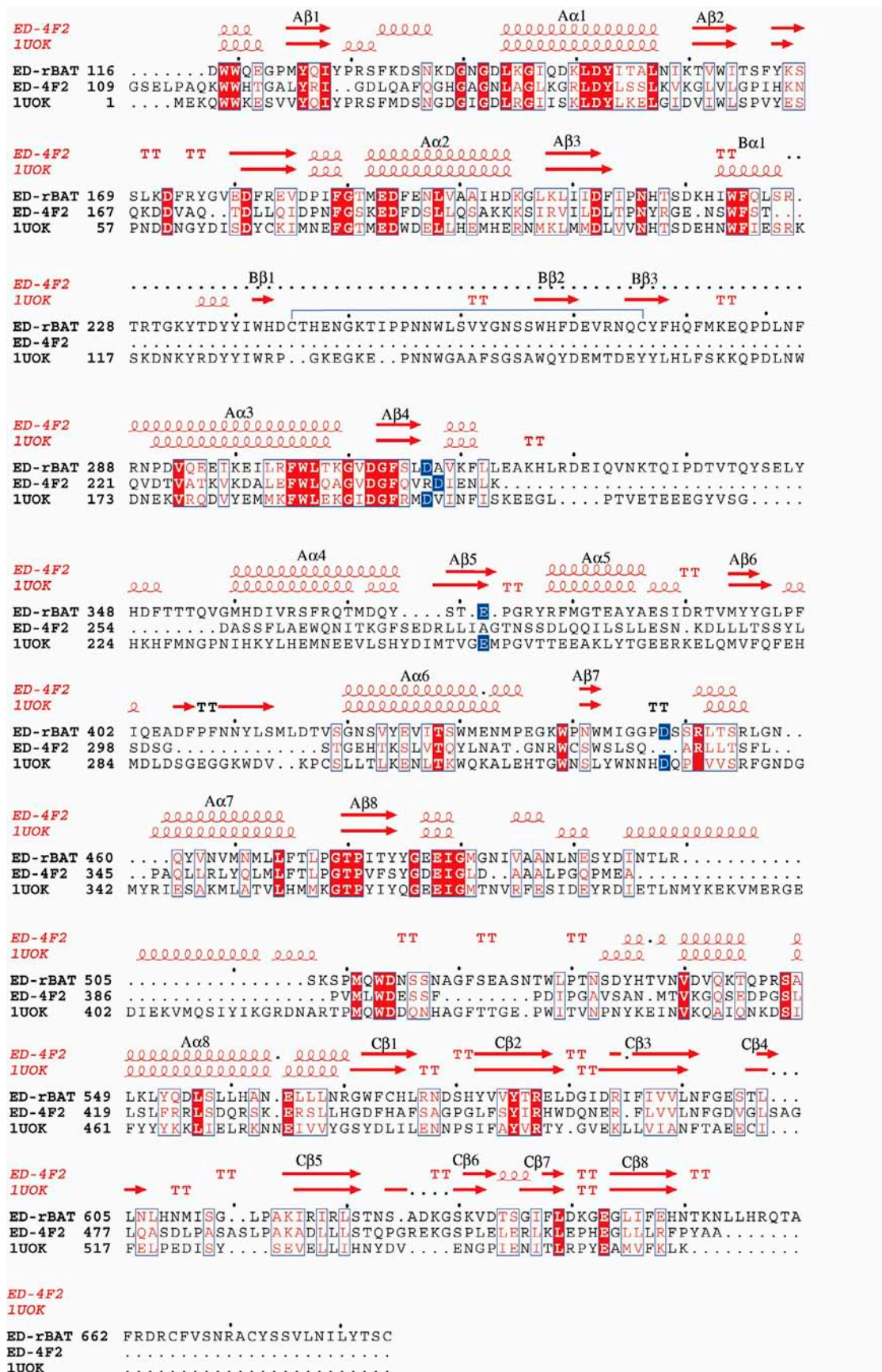


Figura 32. Alineament estructural (SSM) entre ED4F2hc i 1UOK i alineament de seqüència d'EDrBAT (T-coffee).

A partir d'aquests alineaments de seqüència vam modelar l'ectodomini d'rBAT amb el programa MODELLER i el resultat és el model de la figura 34. El model es va fer amb la seqüència d'rBATH des de l'aminoàcid 115 al 654. La cua C-terminal d'rBAT (residus 655-685) no està modelada perquè no manté cap homologia amb cap proteïna amb estructura coneguda. Per testar el model es va fer un càlcul d'energies amb la funció DOPE de Modeller (figura 33). Un perfil d'energies inferior indica que el model és millor. Podem veure com les zones que tenen més homologia segueixen el mateix perfil d'energia que el motlle i en canvi, les zones amb menys homologia, tenen el perfil d'energia més alt. Un dels llocs més ben modelats és el subdomini B, ja que com hem dit és molt semblant al de 1UOK. rBAT té en llocs estratègics d'aquest domini dues cisteïnes (Cys 242 i Cys 273) que en modelar quedaven prou properes perquè es formés un pont disulfur. Aquest pont el trobem en alguns membres de la família de les α -amilases (per exemple, en la Taka amilasa d'*Aspergillus oryzae*, maltotetrahidrolasa i α -amilasa pancreàtica de porc) (Janecek, 1997) estabilitzant aquest domini. En fer el model final aquest pont disulfur ha estat restringit i el resultat és el mostrat en la figura 34.

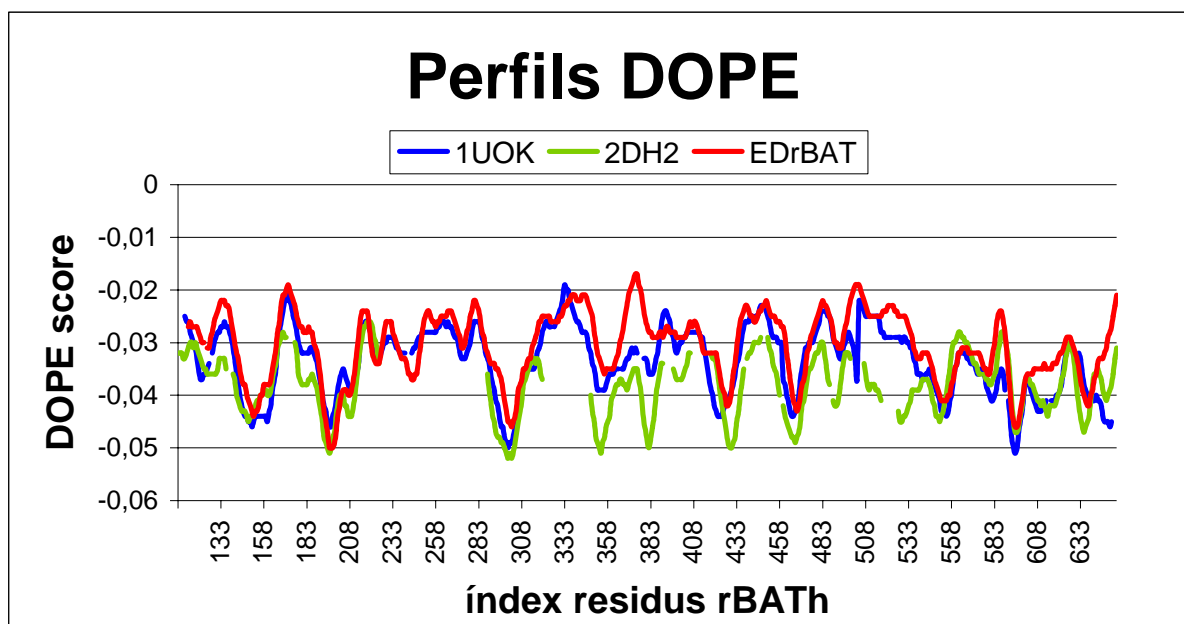


Figura 33. En el gràfic es representa el perfil dels valors d'energia calculats per MODELLER per cadascun dels residus. Es compara el perfil del model d'rBAT amb els de 2DH2 i 1UOK, models utilitzats com a motlle per MODELLER. La numeració és de rBATH. Els perfils estan tallats segons els espais de l'alineament de seqüència utilitzat per modelar.

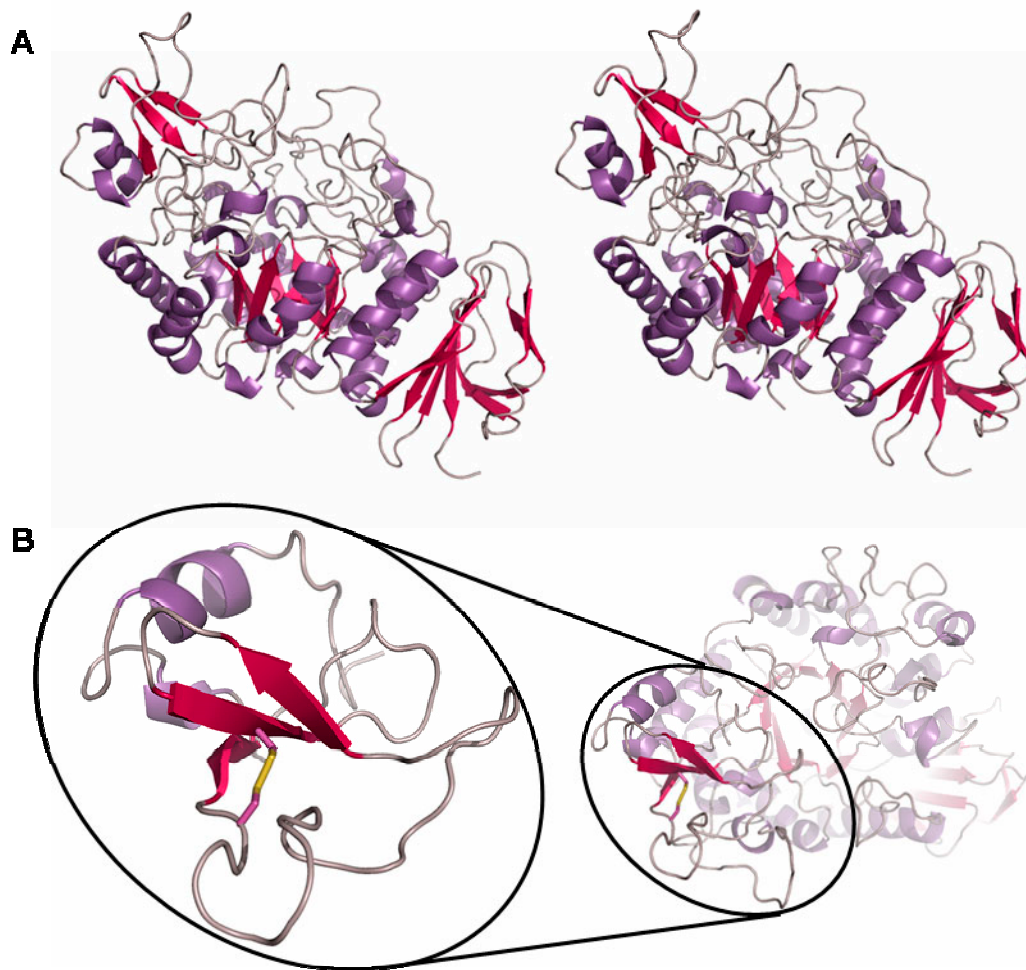


Figura 34. Model de l'ectodomini d'rBAT. **A)** Vista lateral estèreo de l'ectodomini d'rBAT. **B)** Vista superior del barril (α/β)₈ (dreta) i ampliació del subdomini B (esquerra) on es pot veure el pont disulfur que establitzaria aquesta estructura.

Capítol 5. Estudis de la funció de l'ectodomini de 4F2hc

Com ja hem comentat en la introducció la subunitat lleugera dels transportadors heteromèrics d'aminoàcids semblen ser la part catalíticament activa del transportador. La pregunta immediata és: i què hi fa la subunitat pesada? I és més, què hi fa un ectodomini d'uns 400 aminoàcids amb estructura semblant a les α -amilases exposat al costat d'un transportador? Aquestes preguntes encara estan per resoldre, però hem fet alguns experiments per investigar la possible activitat de l'ectodomini de 4F2hc. Per començar hem mirat si aquesta proteïna té activitat α -glucosidasa d'algun tipus i si, malgrat no tenir activitat catalítica, encara conserva la capacitat d'unir glúcids. Finalment hem testat la possibilitat que aquest gran domini interaccioni amb altres proteïnes candidates, segons estudis presentats a la literatura.

5.1. L'activitat α -glucosidasa

Davant l'evidència de l'homologia de seqüència i d'estructura entre l'ectodomini de 4F2hc i les α -amilases la primera pregunta després de tenir la proteïna recombinant ED4F2hc era si aquesta tenia algun tipus d'activitat d'aquesta família d'enzims. La família de les α -amilases com ja hem comentat és molt complexa i engloba moltes activitats diferents (Coutinho, 1999). Tot i tenir diferents activitats els aminoàcids catalítics estan molt conservats i consten de dos aspàrtics, un glutàmic i dues histidines (D199, D329, E255, H103 i H328 en 1UOK). En l'alineament de seqüència entre ED4F2hc i l'oligo-1,6-glucosidasa de *B.cereus* i EDrBAT (figura 32) aquests aminoàcids estan marcats en blau. Podem veure com en 4F2hc no estan conservats, només l'aspàrtic 199 podria estar en un lloc similar de la butxaca catalítica. Per intentar testar d'una manera fàcil i ràpida algun tipus d'activitat vam escollir un substrat universal per aquesta família, el methylumbelliferyl α -D-glucoside (veure fórmula en la figura 35). Aquest substrat conté una glucosa unida per un enllaç α a la metilumbeliferona, compost que quan és alliberat és fluorescent.

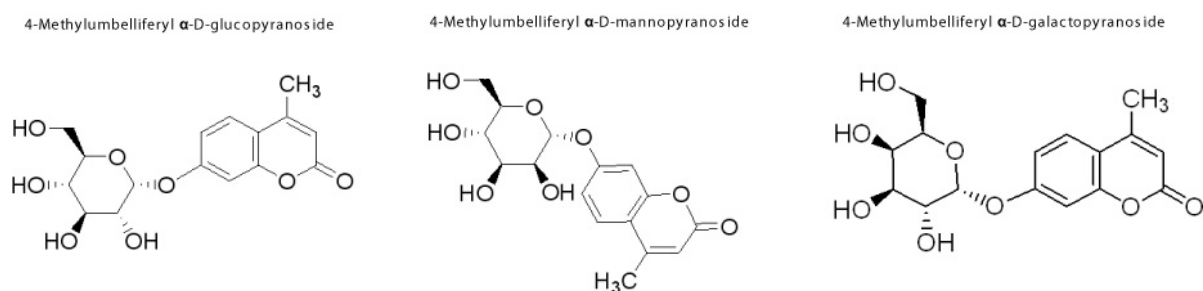


Figura 35. Fórmules dels tres compostos utilitzats en els assajos de l'activitat α -glucosidasa de ED-4F2hc.

Després de fer alguns experiments preliminars vam decidir augmentar l'espectre de l'estudi amb dos substrats més, methylumbelliferyl α -D-mannopyranoside i el methylumbelliferyl α -D-galactopyranoside (veure fórmules en la figura 35). L'experiment més representatiu es mostra en la gràfica inferior (figura 36). Malgrat tenir algunes mostres amb resultat positiu que utilitzàvem com a control positiu de l'experiment, la proteïna purificada no va demostrar mai tenir activitat encara que la testéssim a un concentració molt més elevada que els controls positius.

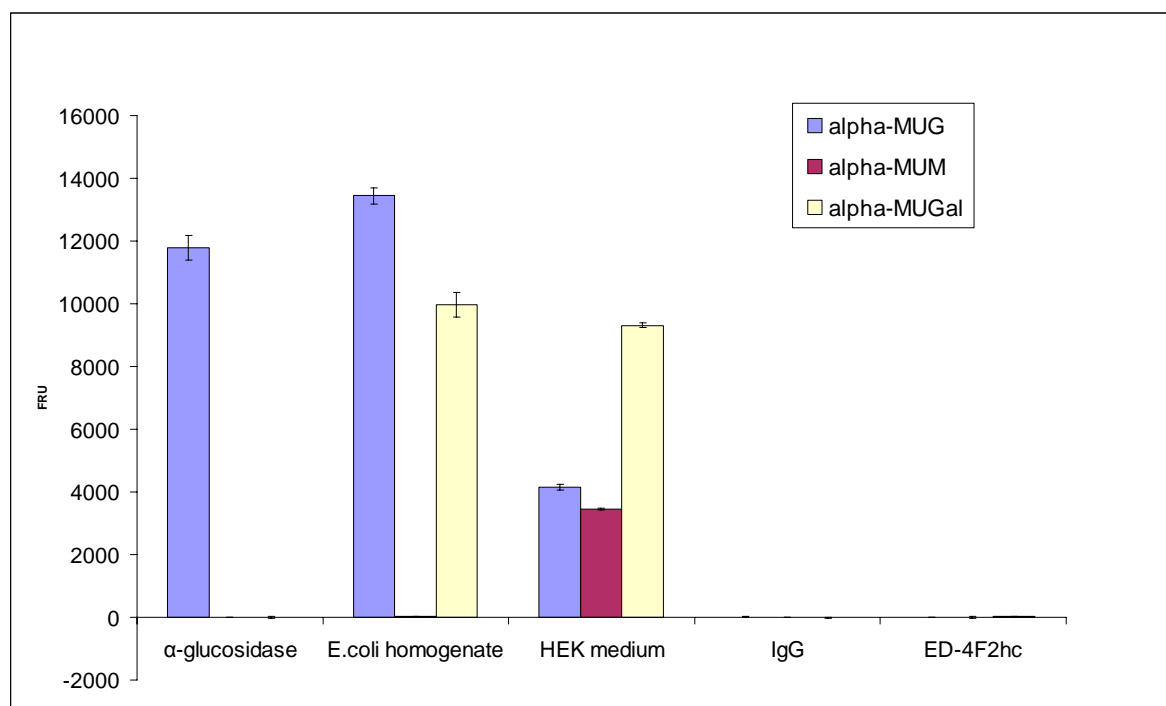


Figura 36. Activitat relacionada amb la família de les α -amilases. El tampó de reacció constava d'1mM de substrat (alpha-MUG, methylumbelliferyl α -D-glucoside; alpha-MUM, methylumbelliferyl α -D-mannopyranosid; i alpha-MUGal, methylumbelliferyl α -D-galactopyranoside) i la quantitat de proteïna indicada: 0.2 ug de α -glucosidase, 10 μ l d'homogenat d'*E. coli*, 10 μ l de medi de HEK293T concentrat, 20 μ g de IgG i 20 μ g d' ED-4F2hc purificat. Després de 7 h a 37°C la reacció es va aturar amb 2 M C_2O_3 i es va llegir a $\lambda_{excitació}=385$ nm i $\lambda_{emissió}=460$ nm. FRU són unitats relatives de fluorescència.

5.2. Interacció amb glúcids

S'ha demostrat que les proteïnes de mamífers Ym1 i YKL-40 pertanyen a la família de les quitinases però no tenen activitat catalítica i han esdevingut lectines (Chang et al., 2001; Bigg et al., 2006). Aquestes proteïnes han perdut els aminoàcids catalítics de les quitinases però conserven la capacitat d'unir glúcids.

L'ectodomini de 4F2hc tampoc conserva els aminoàcids catalítics de la família de les α -amilases i pels nostres experiments no demostra tenir cap activitat relacionada amb aquesta família. Per testar si l'ectodomini podia tenir afinitat per interaccionar amb glúcids específics vam usar glicoxips. En vam utilitzar dos de diferents. El primer era un xip comercial, de glicominds. Aquest constava de 48 glúcids diferents (veure figura 54 a materials i mètodes)

Després dels resultats negatius amb el Glycochip de Glycominds vam contactar amb el Consortium for Functional Glycomics, on vam enviar la proteïna per testar en el seu xip la unió a més de 260 glúcids. El resultat va ser, una vegada més, negatiu per atots els punts del xip (figura 37).

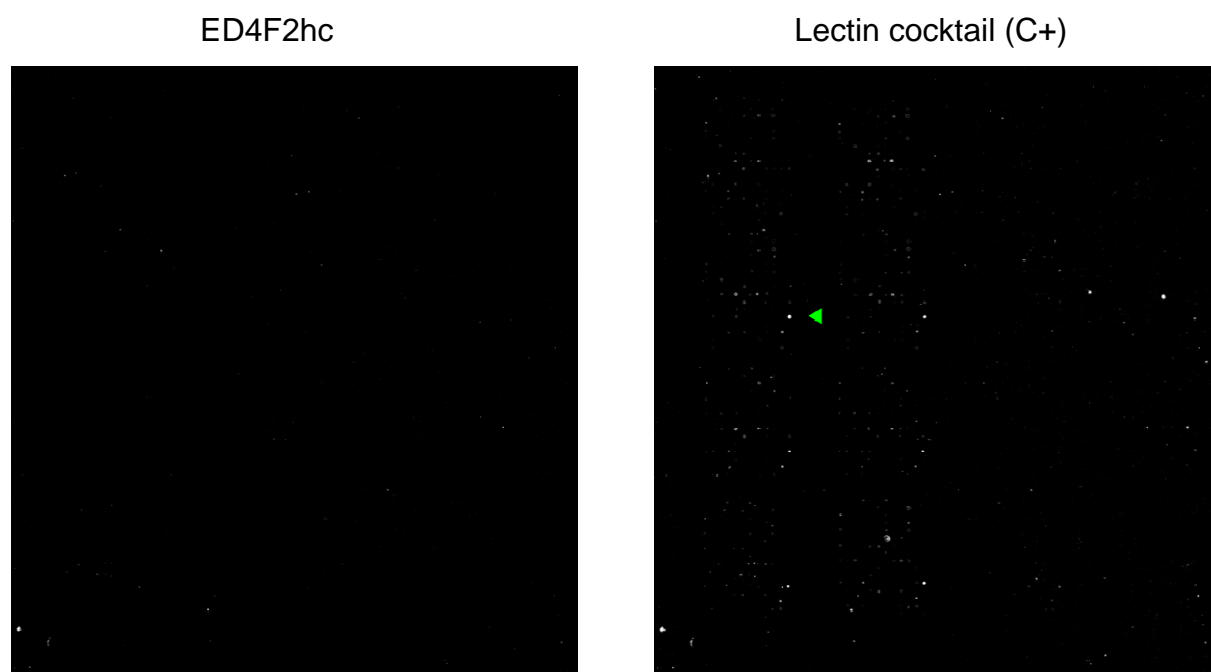


Figura 37. Resultat del glicoxip del Consortium for Functional Glycomics. A l'esquerra es mostra el resultat de la incubació del glicoxip amb ED4F2hc de bacteris a 100ug/ml i detectada usant anticòs anti-4F2hc (dilution of 1:45) i un secundari anti-mouseFITC (10ug/ml). A la dreta hi ha el resultat del mateix assaig amb un còctel de lectines com a control positiu (C+), el qual uneix sobre un 80 % dels

sucres del xip. Cada glúcid està representat en rèpliques de 6, de manera que un punt positiu hauria de trobar-se 6 vegades en el patró del xip, la qual cosa es pot veure en el control positiu però no en la mostra d'ED4F3hc.

5.3. Interacció amb proteïnes

La comparació del centre actiu d'ED4F2hc amb el de les α -amilases ens feia pensar en una possible interacció amb una proteïna en aquesta zona, ja que la nostra proteïna mostrava una butxaca més oberta i més polar que la de 1UOK (figura 38), per exemple. Les glucosidases tenen quatre capes de cadenes laterals hidrofòbiques al fons del barril $(\alpha/\beta)_8$. 4F2hc en té tres, i li manca una capa d'aminoàcids aromàtics que es troba just per sota dels residus catalítics, on té aminoàcids polars com arginina o tirosina. A més a més ED4F2hc no té el domini B sobre el barril $(\alpha/\beta)_8$, que en 1UOK tanca la fenedura catalítica.

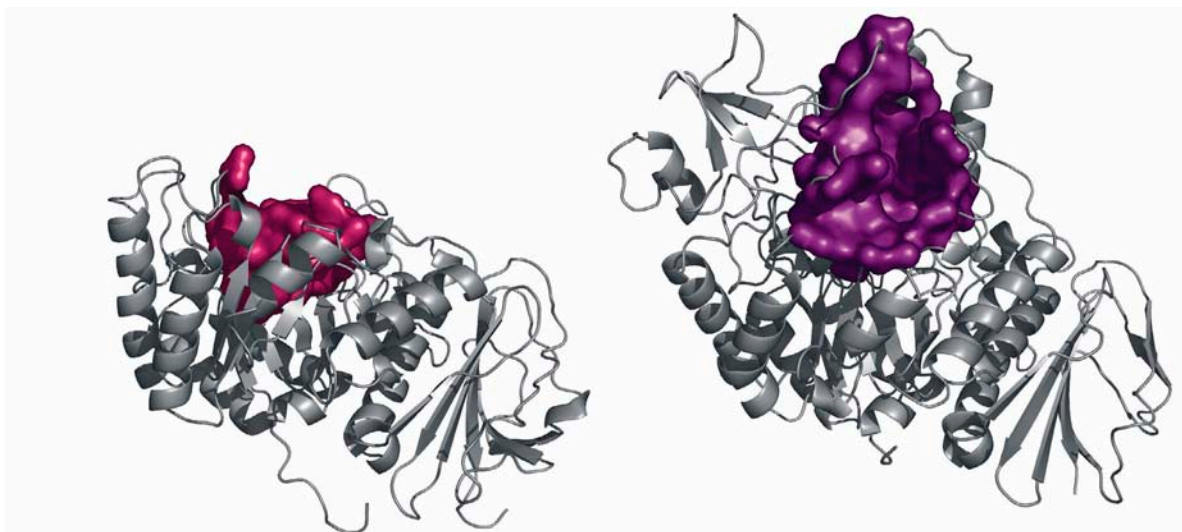


Figura 38. Superfície accessible de les fenedures catalítiques de 4F2hc (esquerra) i 1UOK (dreta). El domini B de 1UOK tanca la fenedura per sobre del barril $(\alpha/\beta)_8$ i en 4F2hc aquesta fenedura és més profunda i arriba més aball del barril $(\alpha/\beta)_8$.

En fer aquestes observacions vam proposar un model d'acoblament (*docking*) entre l'estructura de 4F2hc i la d'ICAM-1. Aquesta interacció havia estat descrita en la literatura (Liu et al., 2003). El model no pretén ser realista sinó un exemple de com una proteïna podria endinsar-se en la fenedura catalítica de 4F2hc i interaccionar

amb els seus residus (figura 39). Com ja explicarem més endavant la interacció amb ICAM-1 no ha pogut ser demostrada amb la nostra proteïna.



Figura 39. Model d'acoblament entre 4F2hc (gris fosc) i el domini 1 d'ICAM-1 (domini 1 del PDBID 1D3L) (verd).

A més de ICAM-1, en la literatura se cita un elevat nombre de proteïnes amb possible interacció amb 4F2hc amb altres proteïnes (veure apartat 3.2-3.5 de la introducció). Aprofitant que teníem proteïna recombinant de l'ectodomini de 4F2hc pura i en gran quantitat vam intentar testar la interacció directa amb algunes proteïnes candidates. Per aquest experiment vam utilitzar dues tècniques: ELISA i BIACORE. Alguns d'aquests estudis no estan finalitzats i es presenten com a resultats preliminars.

5.4. ELISA

Amb la col·laboració del laboratori del Dr. Jose Maria Casasnovas vam posar a punt la tècnica d'ELISA per testar interaccions de proteïnes amb ED-4F2hc (Jimenez et al., 2005). Amb aquesta tècnica vam testar diferents fragments dels dominis extracel·lulars dels ICAMs 1, 3 i 5 (proteïnes recombinants glucosilades facilitades pel laboratori del Dr. Casasnovas) i alguns controls, com MG i BSA. En la figura 40 s'hi representa un experiment representatiu dels resultats obtinguts amb aquesta tècnica. Les proteïnes testades són fragments dels dominis extracel·lulars N-terminals d'ICAM-1 (IC1-3D, tres dominis), i ICAM-5 (IC5-2D, dos dominis i IC5-4D, quatre dominis). MG és una proteïna bàsica, ja que aquests fragments dels ICAMs són molt bàsics i volíem descartar una interacció no específica per motius d'atracció electrostàtica.

L'experiment consistia a enganxar les proteïnes a testar en una placa especial per ELISA durant la nit. A sobre s'hi afegia l'ectodomini de 4F2hc purificat amb cues d'histidines i després d'alguns rentats es feia interaccionar amb anticòs primari anti-histidines i secundari anti-ratolí conjugat amb HRP, es revelava amb amplex rex i es llegia la fluorescència. En tots els casos hi havia com a controls negatius pous sense la proteïna test (4F2 sol) i pous sense ED4F2hc (no 4F2), de manera que podíem diferenciar les interaccions inespecífiques d'ED4F2hc amb la placa i dels anticossos amb la proteïna a testar.

Aquest experiment tenia molts problemes ja que algunes proteïnes test donaven molt senyal inespecífic (veure la barra taronja en ICAM1 3 dominis). Malgrat això en la figura 40 podem veure certa interacció específica entre ED4F2hc i ICAM1 (tres dominis), però els resultats no són prou significatius. Per aquesta raó ens vam disposar a testar més finament aquestes possibles interaccions per Ressonància de Plasmó Superficial amb el Biacore.

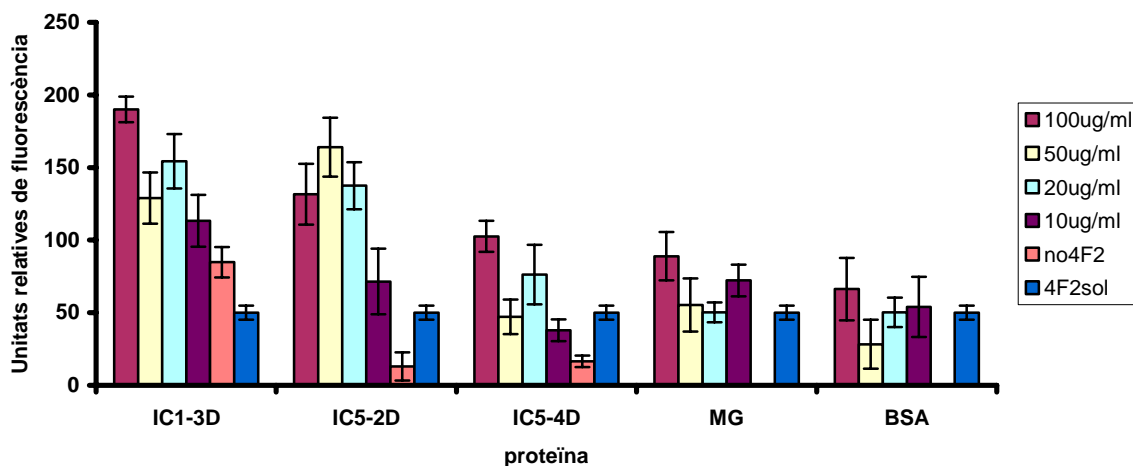


Figura 40. Gràfic de resultats d'experiments d'ELISA amb ED-4F2hc purificat. Resultats corregits pel control negatiu on no hi havia proteïnes test. La barra taronja és la línia base per a cada proteïna testada, ja que és el senyal que dona la proteïna sense haver afegit 4F2 (no 4F2).

5.5. Biacore

La tècnica Biacore es basa en un fenomen òptic: el principi de ressonància del plasmó de superfície (SPR). En il·luminar amb una llum polaritzada una interfície entre medis de diferent índex de refracció separats per una làmina de material conductor, s'aprecia una "ombra" en la reflexió. Aquest buit és degut a l'acoblament entre l'energia de la llum incident i els plasmons de superfície del film conductor que separa els dos medis. L'angle d'aquesta "ombra" amb la interfície depèn directament de l'índex de refracció del medi on es propaga l'ona evanescent, la cara oposada a les llums incident i reflectida. Aquests canvis en l'índex de refracció només es detectaran a la interfície, perquè l'ona evanescent decau de manera exponencial amb la distància a la superfície. En el Biacore, les interaccions entre biomolècules canvien la concentració del solut i, per tant, l'índex de refracció. D'aquesta manera el fenomen d'SPR permet mesurar els canvis de massa produïts en una superfície entre una molècula unida a aquesta superfície i unes altres en suspensió, en temps real, sense necessitat de marcatges i independentment de la natura de les molècules que interaccionen.

Si immobilitzem químicament una de les molècules que volem testar en un film conductor (en el nostre cas ED4F2hc), i després fem circular una solució de l'altra

proteïna per sobre, la interacció que es produeix canvia la concentració de solut i varia l'angle de ressonància, que és mesurat.

Mitjançant aquesta tècnica s'han començat a fer experiments fent passar per sobre el xip amb 4F2hc unit les següents proteïnes: ectodomini d'ICAM1, ectodomini d'ICAM5, ectodomini d'EpCAM, Galectina3, ectodomini de CD147-Fc, IgG standard i anticossos estructurals específics contra 4F2hc. Aquests anticossos eren el nostre control positiu. Respecte les altres proteïnes l'únic resultat que tenim de moment és un resultat de galectina-3 en passar sobre ED-4F2hc glicosilat (figura 41). Sobre una immobilització de 8000 RU de 4F2hc glicosilat es veu una interacció de galectina-3 (línea blava) que, en deixar de passar mostra, resta unida amb unes 200 RUs. Això no passa quan la galectina 3 la passem per sobre d'ED4F2hc produït en *E. coli* (immobilització de 5600 RU), cosa que ens podria fer pensar amb una interacció amb els glúcids de la proteïna produïda en cèl·lules HEK 293T. Per això també vam passar la galectina 3 per sobre la proteïna CD147 glicosilada (immobilització de 7400 RU) i produïda en les mateixes cèl·lules i el resultat va ser negatiu. Sembla, doncs, que existeix una interacció específica de galectina 3 amb ED4F2hc i dependent de les glicosilacions.

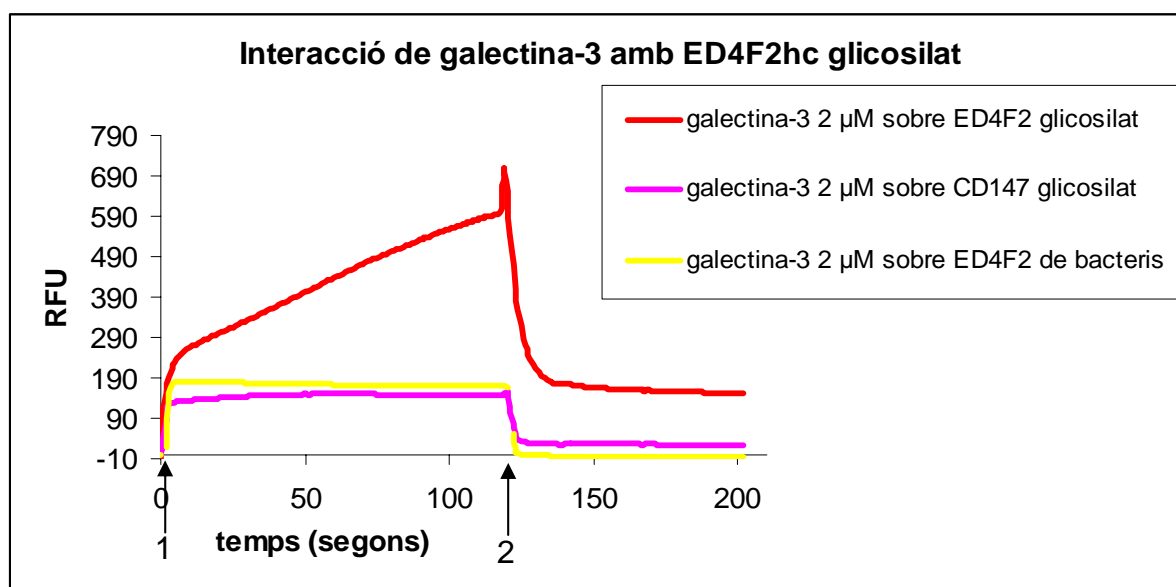


Figura 41. En el gràfic podem observar els canvis d'RU (unitats de ressonància) que es produeixen quan passem una solució amb galectina-3 2 µM sobre els canals del biacore on hi ha immobilitzats ED4F2hc produït en *E. coli* (groc), ED4F2hc glicosilat produït en HEK-293T (vermell) i CD147 glicosilat produït en HEK-293T (rosa). El temps 0 és el moment de la injecció (fletxa 1) de la solució amb galectina-3 i a temps 120 segons (fletxa 2) es deixa d'injectar i comença la fase de dissociació.

Aquests experiments s'han aturat en la fase preliminar. S'han de fer estudis més complets pels ICAMs, EpCAM, CD147, i altres proteïnes candidates.

