



Universitat Autònoma de Barcelona

Departament de Bioquímica i Biologia Molecular

**Caracterització cinètica de la formació de
fibres amiloides i cerca de funcions remotes de
dominis d'activació de procarboxipeptidases
humanes**

Núria Cerdà Costa

Octubre 2008



Universitat Autònoma de Barcelona

Departament de Bioquímica i Biologia Molecular

**Caracterització cinètica de la formació de
fibres amiloides i cerca de funcions remotes de
dominis d'activació de procarboxipeptidases
humanes**

Memòria presentada per Núria Cerdà Costa, llicenciada en Bioquímica, per optar al grau de Doctora en Biotecnologia

Treball realitzat a la Unitat de Biociències del Departament de Bioquímica i Biologia Molecular de la Universitat Autònoma de Barcelona, sota la direcció de la Dra Sandra Villegas Hernández i el Dr Francesc Xavier Avilés Puigvert

Núria Cerdà Costa

vist-i-plau dels directors de tesi

Dra Sandra Villegas Hernández

Dr Francesc Xavier Avilés Puigvert

Bellaterra, Octubre del 2008

Als meus pares

AGRAÏMENTS

Vull expressar el meu agraïment a tota la gent que, gràcies a la seva ajuda directa i/o indirecta ha fet possible la realització d'aquest treball:

Primerament als meus directors de tesi, la Dra. S. Villegas i el Dr. F.X. Avilés, per la dedicació i el suport que han mostrat durant aquest temps. De la mateixa manera, vull agrair al Dr. J. Vendrell l'oportunitat d'entrar a treballar al seu grup de recerca.

A la Generalitat de Catalunya per la concessió de la beca FI que ha fet possible aquesta tesi (FI 2003 FI 00388), així com la borsa de viatge que m'ha permès fer una estada fora de Catalunya.

De la meva estada a Alemanya vull agrair al Dr. Luis Serrano, coordinador del programa d'estructures de l'EMBL, per a permetre'm realitzar una estada al seu laboratori i tota la seva ajuda a nivell científic; així com a la Dra. Alexandra Esteras per la seva assistència amb la microscòpia electrònica, i a la resta de l'"*spanish corner*".

De la meva estada a Madrid vull agrair al grup de Ressonància Magnètica Nuclear del CNIO i molt especialment al Dr. Francisco Blanco i al Dr. Jesús Prieto, per permetre'm utilitzar l'*stopped-flow* i ajudar-me a manejar-lo.

De Bilbao vull agrair al Dr. José Luis Rodríguez Arrondo i a l'Igor de l'Arada del grup d'Espectroscòpia d'Infraroig del centre mixt UPV-CSIC per la col·laboració amb els experiments d'infraroig duts a terme en aquesta tesi.

Al Dr. Baldo Oliva i al Jaume Bonet del grup de Bioinformàtica Estructural del GRIB (IMIM) per iniciar-me en el món de la bioinformàtica i ajudar-me durant tota la col·laboració.

A la Chari li vull agrair la seva gran ajuda amb els EMSA, per ensenyar-m'ho tot i realitzar la part de radioactivitat.

Als meus companys PROs per tantes coses que no hi cabrien, tant per ensenyar-me les petites coses del laboratori que són la base de la tesi com pel bon ambient que ha fet que em sentís com a casa. De la mateixa manera no em puc oblidar dels companys d'Enzimo de l'IBB, que sempre han estat el nostre 'segon labo'.

A tots els companys del Departament de Bioquímica i Biologia Molecular de Biociències, per les bones estones i per la seva ajuda en in comptables moments. Vull

fer especial menció als Ribos, que han estat en moltes ocasions com el nostre 'tercer labo', així com al Dr. Salvador Bartolomé i a l'Helena Carbó, per moltíssimes coses...

Als meus nous companys del labo, els CRI3 (i els CRI4 que sempre estan per allà), per les risoteràpies i el seus ànims en l'etapa final.

A tots els meus amics, entre ells els de la colla de Lliçà (aquí us hi fico a tots, sigueu de Sta. Eulàlia o de Sabadell) pel vostre suport moral. Vull agrair, emperò, molt especialment a les nenes de Biologia: la Marta, les Eves i la Meri, amb les quals sempre he pogut comptar.

Al Lluís li agraeixo una infinitat de coses, però suposo que hauria de recalcar la seva paciència per aguantar una tesi des del principi fins al final i mai perdre el seu somriure.

Finalment a la meva família i sobretot als meus pares per haver-me recolzat en tot moment.

*Amb el suport del Comissionat per a Universitats i Recerca del Departament
d'Innovació, Universitats i Empresa de la Generalitat de Catalunya i del Fons Social
Europeu.*

ÍNDEX

| | |
|---|--------------|
| ÍNDEX..... | xiii |
| ÍNDEX DE FIGURES I TAULES | xxi |
| ABREVIATURES | xxiii |
| | |
| I INTRODUCCIÓ | 1 |
| I.1 LES PROTEASES | 3 |
| I.1.1 Proteases i proteòlisi limitada | 3 |
| I.1.2 Tipus de proteases..... | 4 |
| I.1.2.1 Les metal·loproteases | 6 |
| I.1.3 Les metal·locarboxipeptidases..... | 8 |
| I.1.3.1 Classificació de les metal·locarboxipeptidases de la família M14..... | 8 |
| I.1.3.2 Carboxipeptidases de la subfamília A/B | 11 |
| I.1.3.3 Carboxipeptidases de la subfamília N/E | 15 |
| I.1.3.4 Estructura de les metal·locarboxipeptidases de la subfamília A/B..... | 19 |
| I.1.3.5 Segments d'activació i mecanisme d'activació de les metal·locarboxipeptidases de la subfamília A/B..... | 21 |
| I.2 PLEGAMENT I AMILOIDOGÈNESI..... | 24 |
| I.2.1 Plegament..... | 24 |
| I.2.1.1 Teories inicials | 25 |
| I.2.1.2 Embutts de plegament | 26 |
| I.2.1.3 Intermediaris | 27 |
| I.2.1.4 Plegament <i>in vivo</i> | 28 |
| I.2.2 Fibres amiloides..... | 29 |
| I.2.2.1 Fibres amiloides i amiloïdosis | 30 |
| I.2.2.2 Característiques de les fibres amiloides | 31 |
| I.2.2.3 Formació de fibres amiloides | 33 |
| I.2.2.4 Causa de la formació de fibres amiloides <i>in vivo</i> | 34 |
| I.2.2.5 Estructura de les fibres amiloides | 35 |
| I.2.2.6 Mètodes de predicció de l'amiloidogènesi | 38 |
| I.2.3 El domini d'activació de la procarboxipeptidasa A2 humana com a model d'estudi..... | 41 |
| I.2.3.1 Estudis de caracterització del plegament..... | 41 |
| I.2.3.2 Estudi de l'estat de transició del plegament d'ADA2h..... | 43 |

| | | |
|---|---|-----------|
| I.2.3.3 | Formació de fibres amiloides <i>in vitro</i> d'ADA2h | 44 |
| OBJECTIUS | | 47 |
| MATERIAL I MÈTODES | | 51 |
| M.1 APARELLS D'ÚS GENERAL | | 55 |
| M.1.1 | Productes químics..... | 56 |
| M.2 Tècniques de DNA recombinant | | 57 |
| M.2.1 | Reacció en cadena de la polimerasa, PCR (<i>polymerase chain reaction</i>) | 57 |
| M.2.1.1 | Amplificació de gens o fragments gènics | 57 |
| M.2.1.2 | Construcció d'una seqüència de DNA sintètica | 59 |
| M.2.1.3 | Introducció de llocs de complementarietat per al clonatge independent de lligació | 60 |
| M.2.1.4 | Mutagènesi dirigida mitjançant PCR | 60 |
| M.2.2 | Extracció de DNA plasmídic i purificació de DNA procedent de reaccions de PCR i de bandes de gels d'agarosa. | 62 |
| M.2.3 | Digestió amb DpnI..... | 62 |
| M.2.4 | Producció d'extrems de cadena senzilla en el clonatge independent de lligació | 63 |
| M.2.5 | Hibridació del Vector LIC amb l'insert corresponent | 63 |
| M.2.6 | Introducció de DNA forani a cèl·lules bacterianes | 64 |
| M.2.6.1 | Preparació de cèl·lules competents d' <i>E. coli</i> | 64 |
| M.2.6.2 | Transformació de cèl·lules competents d' <i>E. coli</i> | 65 |
| M.2.7 | Seqüenciació de DNA | 66 |
| M.2.8 | Transcripció <i>in vitro</i> de RNA i marcatge radioactiu | 66 |
| M.2.9 | Electroforesi d'àcids nucleics | 67 |
| M.2.9.1 | Electroforesi en gels d'agarosa | 67 |
| M.2.9.2 | Assajos de canvis de mobilitat electroforètica en gels de poliacrilamida (<i>Electrophoretic mobility-shift assays</i> , EMSA) | 68 |
| M.2.9.2.1 | Preparació del gel..... | 68 |
| M.2.9.2.2 | EMSA..... | 69 |
| M.2.10 | Determinació de la concentració d'àcids nucleics..... | 71 |
| M.2.10.1 | Determinació espectrofotomètrica | 71 |
| M.2.10.2 | Estimació per electroforesi en gels d'agarosa | 71 |

| | |
|--|-----------|
| M.2.10.3 Determinació de la concentració d'un àcid nucleic marcat radioactivament..... | 71 |
| M.3 Sistema d'expressió i purificació..... | 72 |
| M.3.1 Organismes..... | 72 |
| M.3.1.1 Soques d' <i>Escherichia coli</i> | 72 |
| M.3.1.2 Conservació de soques bacterianes..... | 72 |
| M.3.2 Medis de cultiu de bacteris..... | 73 |
| M.3.2.1 Medi líquid..... | 73 |
| M.3.2.2 Medi sòlid..... | 73 |
| M.3.2.3 Antibiòtics..... | 74 |
| M.3.3 Vectors d'expressió..... | 75 |
| M.3.4 Inducció..... | 76 |
| M.3.5 Obtenció de l'extracte cel·lular..... | 79 |
| M.3.6 Tècniques de cromatografia líquida..... | 80 |
| M.3.6.1 Cromatografia atmosfèrica de bescanvi aniònic..... | 80 |
| M.3.6.2 Cromatografia atmosfèrica d'afinitat amb histidines..... | 80 |
| M.3.6.3 Cromatografia de gel-filtració d'alta resolució (FPLC)..... | 82 |
| M.3.7 Diàlisi..... | 83 |
| M.3.8 Digestió amb factor Xa..... | 83 |
| M.4 Mètodes de proteïnes..... | 85 |
| M.4.1 Mètodes d'identificació i determinació de massa molecular..... | 85 |
| M.4.1.1 Electroforesi en SDS (SDS-PAGE)..... | 85 |
| M.4.1.1.1 Preparació del gel..... | 85 |
| M.4.1.1.2 Tinció i destinció del gel..... | 86 |
| M.4.1.2 Espectrometria de masses MALDI-TOF..... | 87 |
| M.4.2 Determinació de la concentració de proteïna..... | 87 |
| M.4.2.1 Mètodes espectrofotomètrics..... | 88 |
| M.4.2.1.1 Absorbància d'aromàtics..... | 88 |
| M.4.2.1.2 Mètode colorimètric de Bradford..... | 88 |
| M.4.2.2 Mètode electroforètic..... | 89 |
| M.4.3 Dicroisme Circular..... | 89 |
| M.4.3.1 Espectre llunyà de CD..... | 90 |
| M.4.3.2 Perfils tèrmics..... | 90 |
| M.4.3.3 <i>Stopped-flow</i> | 91 |
| M.4.4 Espectroscòpia d'absorció d'infraroig amb Transformada de Fourier..... | 92 |

| | | |
|-----------------------|---|------------|
| M.4.5 | Microscòpia electrònica de transmissió | 93 |
| M.5 | Mètodes Bioinformàtics | 95 |
| M.5.1 | Eina de cerca d'homologia: <i>Basic Local Alignment Search Tool</i> (BLAST i PSI-BLAST) | 95 |
| M.5.2 | Eina de superposició d'estructures terciàries: <i>Structural Alignment of Multiple Proteins</i> (STAMP)..... | 96 |
| RESULTATS..... | | 97 |
| 1. | CHAPTER 1: EARLY KINETICS OF AMYLOID FIBRIL FORMATION REVEALS CONFORMATIONAL REORGANISATION OF INITIAL AGGREGATES..... | 99 |
| 1.1 | SUMMARY | 101 |
| 1.2 | INTRODUCTION | 103 |
| 1.3 | MATERIALS AND METHODS | 105 |
| 1.3.1 | Materials..... | 105 |
| 1.3.2 | TANGO prediction and design of new mutants..... | 105 |
| 1.3.3 | Mutagenesis, expression and purification | 105 |
| 1.3.4 | Far-UV circular dichroism..... | 105 |
| 1.3.5 | Electron microscopy..... | 106 |
| 1.3.6 | Kinetic experiments..... | 106 |
| 1.3.7 | Formula-based predictions..... | 107 |
| 1.4 | RESULTS..... | 109 |
| 1.4.1 | TANGO analysis of ADA2h..... | 109 |
| 1.4.2 | Selection of point mutations to study the aggregation properties of ADA2h..... | 112 |
| 1.4.3 | Uncoupling between stability and aggregation..... | 112 |
| 1.4.4 | Experimental conditions to follow the aggregation process | 114 |
| 1.4.5 | Kinetic experiments..... | 116 |
| 1.4.6 | Algorithm-based predictions..... | 119 |
| 1.4.7 | β -Sheet aggregation and amyloid aggregation | 122 |
| 1.5 | DISCUSSION | 125 |
| 1.5.1 | Complex kinetics in the initial conformational transition..... | 125 |
| 1.5.2 | Stability and aggregation..... | 126 |
| 1.5.3 | Kinetics of aggregation and folding..... | 127 |

| | | |
|------------|---|------------|
| 1.5.4 | Algorithm-based predictions..... | 128 |
| 2. | CHAPTER 2: CHARACTERISATION OF DIFFERENT PROCESSES OF AMYLOIDOGENESIS BY 2D-FTIR SHOWS CONVERGENCE TO THE SAME FINAL FIBRIL..... | 131 |
| 2.1 | SUMMARY..... | 133 |
| 2.2 | INTRODUCTION..... | 135 |
| 2.3 | MATERIALS AND METHODS..... | 137 |
| 2.3.1 | Materials..... | 137 |
| 2.3.2 | Expression and purification..... | 137 |
| 2.3.3 | Infrared analysis..... | 137 |
| 2.3.4 | Transmission Electron Microscopy..... | 137 |
| 2.3.5 | Circular Dichroism..... | 138 |
| 2.4 | RESULTS..... | 139 |
| 2.4.1 | Assignment of initial infrared bands by the wild-type protein..... | 139 |
| 2.4.2 | Amyloid fibril formation by the wild-type protein..... | 142 |
| 2.4.3 | Effect of mutations in fibril formation..... | 146 |
| 2.4.4 | Two-dimensional correlation spectra..... | 148 |
| 2.4.5 | Electron microscopy..... | 150 |
| 2.4.6 | Circular dichroism..... | 153 |
| 2.5 | DISCUSSION..... | 157 |
| 2.5.1 | Assignment of the initial bands for the wild-type protein..... | 157 |
| 2.5.2 | Nucleation and Reorganisation..... | 158 |
| 2.5.3 | Influence of aggregation propensity and stability..... | 160 |
| 3. | CHAPTER 3: PREDICTION OF AN ALTERNATIVE FUNCTION FOR ADA4h AND EXPERIMENTAL VALIDATION..... | 161 |
| 3.1 | SUMMARY..... | 165 |
| 3.2 | INTRODUCTION..... | 167 |
| 3.3 | MATERIALS AND METHODS..... | 169 |
| 3.3.1 | Bioinformatics..... | 169 |
| 3.3.2 | Materials..... | 170 |
| 3.3.3 | Cloning and Expression..... | 170 |

| | | |
|--------------|--|----------------|
| 3.3.4 | <i>In vitro</i> transcription of RNA | 171 |
| 3.3.5 | RNA Electrophoretic Mobility Shift Assays (EMSA) | 172 |
| 3.4 | RESULTS | 173 |
| 3.4.1 | ADA4h particularities | 173 |
| 3.4.2 | Structural search | 174 |
| 3.4.3 | Pattern discovery | 179 |
| 3.4.4 | RNA Electrophoretic Mobility Shift Assays (EMSA) | 180 |
| 3.5 | DISCUSSION | 183 |
| 3.5.1 | Definition of RRM | 183 |
| 3.5.2 | Binding specificity of U1A | 184 |
| 3.5.3 | Moonlighting function for ADA4h | 186 |
| | Annex to chapters | 187 |
| | DISCUSSIÓ GENERAL | 189 |
| | CONCLUSIONS | 201 |
| | BIBLIOGRAFIA | 209 |
| | RESUM | 239 |
| | APÈNDIXS | 243 |
| | Apèndix A: Publicacions | 245 |
| | Apèndix B: Adreces electròniques actualitzades dels programes i bases de dades disponibles en xarxa | 247 |
| | Apèndix C: Codis PDB de les estructures esmentades en aquesta memòria | 249 |

ÍNDEX DE FIGURES I TAULES

Figures

| Figura | Pàgina | Figura | Pàgina |
|--------|--------|--------|--------|
| I.1 | 4 | 1.5 | 116 |
| I.2 | 19 | 1.6 | 120 |
| I.3 | 20 | 1.7 | 121 |
| I.4 | 22 | 1.8 | 122 |
| I.5 | 24 | 1.9 | 124 |
| I.6 | 25 | 1.10 | 125 |
| I.7 | 27 | 1.11 | 128 |
| I.8 | 32 | 2.1 | 140 |
| I.9 | 33 | 2.2 | 141 |
| I.10 | 36 | 2.3 | 143 |
| I.11 | 37 | 2.4 | 145 |
| I.12 | 42 | 2.5 | 150 |
| I.13 | 45 | 2.6 | 152 |
| M.1 | 59 | 2.7 | 154 |
| M.2 | 77 | 3.1 | 171 |
| M.3 | 78 | 3.2 | 175 |
| 1.1 | 109 | 3.3 | 176 |
| 1.2 | 111 | 3.4 | 177 |
| 1.3 | 114 | 3.5 | 178 |
| 1.4 | 115 | 3.6 | 181 |

Taules

| Taula | Pàgina |
|--------------|---------------|
| I.1 | 11 |
| I.2 | 30 |
| I.3 | 31 |
| I.4 | 43 |
| M.1 | 74 |
| 1.1 | 113 |
| 1.2 | 117 |
| 1.3 | 118 |
| 2.1 | 141 |
| 2.2 | 147 |
| 2.3 | 147 |
| 2.4 | 148 |
| 2.5 | 155 |
| 3.1 | 173 |
| 3.2 | 179 |

ABREVIATURES

| Abreviatura | Significat |
|---------------------------------|--|
| Δ | increment |
| ε | coeficient d'extinció molar |
| λ | longitud d'ona |
| α - ³² P[UTP] | uridina trifosfat amb el fosfat α substituït per l'isòtop radioactiu fòsfor 32 |
| ‡ | estat de transició |
| ³² P | isòtop radioactiu fòsfor 32 |
| 2D | bidimensional |
| 3D | tridimensional |
| Ac | acetilació |
| ACLP | proteïna aòrtica tipus carboxipeptidasa (de l'anglès <i>aortic carboxypeptidase-like protein</i>) |
| AcP | acilfosfatasa |
| AD | domini d'activació (de l'anglès <i>activation domain</i>) |
| ADA2h | domini d'activació de la procarboxipeptidasa A2 humana (veure també AD) |
| ADA4h | domini d'activació de la procarboxipeptidasa A4 humana (veure també AD) |
| ADA4hm | domini d'activació de la procarboxipeptidasa A4 humana modelat (veure també AD) |
| ADAM | proteïna desintegrina i metal·loproteasa (de l'anglès <i>a disintegrin and metalloproteinase protein</i>) |
| ADEPT | teràpia enzimàtica dirigida per anticòs (de l'anglès <i>antibody directed enzyme prodrug therapy</i>) |
| AEBP | proteïna unidora a l' <i>enhancer</i> d'adipòcits (de l'anglès <i>adipocyte enhancer-binding protein</i>) |
| Am | amidació |
| Amp | ampicil·lina |
| APP | proteïna precursora amiloide (de l'anglès <i>amyloid precursor protein</i>) |
| BSA | albúmina de sèrum boví (de l'anglès <i>bovine serum albumin</i>) |
| CCD | dispositiu de càrregues elèctriques interconnectades (de l'anglès <i>charge-coupled device</i>) |
| CCPs | carboxipeptidases citosòliques (de l'anglès <i>cytosolic carboxypeptidases</i>) |
| CD | dicroisme circular (de l'anglès <i>circular dichroism</i>) |
| cDNA | àcid desoxiribonucleic complementari (veure també DNA) |
| CNIO | Centre Nacional d'Investigacions Oncològiques |
| COS | anàlisi correlacional (de l'anglès <i>correlation analysis</i>) |
| CP | carboxipeptidasa |
| CSIC | Consell Superior d'Investigacions Científiques |
| C-terminal | carboxiterminal |
| ddH ₂ O | aigua bidestil·lada |
| DM | mutant d'ADA2h amb ambdues hèlixs α estabilitzades |
| DNA | àcid desoxiribonucleic (de l'anglès <i>deoxiribonucleic acid</i>) |

| | |
|----------------|--|
| DNasa | desoxiribonucleasa |
| DNase | desoxiribonucleasa (de l'anglès <i>deoxiribonuclease</i>) |
| dNTP | desoxiribonucleòtid trifosfat |
| DSC | calorimetria diferencial d'escombrat (de l'anglès <i>differential scanning calorimetry</i>) |
| dsDNA | DNA de doble cadena (veure també DNA) |
| DTT | ditiotreitòl |
| <i>E. coli</i> | <i>Escherichia coli</i> |
| ECI | inhibidor de carboxipeptidasa endogen (de l'anglès <i>endogenous carboxypeptidase inhibitor</i>) |
| EDTA | tetracetat d'etilè diamina |
| EF-hand | domini d'unió a calci |
| EGF | factor de creixement epidèrmic (de l'anglès <i>epidermal growth factor</i>) |
| EM | microscòpia electrònica (de l'anglès <i>electron microscopy</i>) |
| EMBL | Laboratori Europeu de Biologia Molecular (de l'anglès <i>European Molecular Biology Laboratory</i>) |
| EMSA | assaigs de canvis de mobilitat electroforètica (de l'anglès <i>electrophoretic mobility shift assays</i>) |
| <i>et al.</i> | i col·laboradors (del llatí <i>et alter</i>) |
| EtBr | bromur d'etidi (de l'anglès <i>ethidium bromide</i>) |
| ExpASy | <i>Expert Protein Analysis System</i> |
| Fn | fibronectina |
| FPLC | cromatografia líquida ràpida de proteïnes (de l'anglès <i>fast protein liquid chromatography</i>) |
| FTIR | espectroscòpia d'infraroig amb transformada de Fourier (de l'anglès <i>Fourier-Transform InfraRed spectroscopy</i>) |
| G | energia lliure (de Gibbs) |
| GRIB | Grups de Recerca en Informàtica Biomèdica |
| Ha | <i>Helicoverpa armigera</i> |
| HEPES | àcid 4-(2-hidroxiètil)-1-piperazinetanesulfònic |
| HET-s | proteïna de xoc tèrmic HSP104 de <i>Podospora anserina</i> |
| hnRNP | ribonucleoproteïna heteronuclear (de l'anglès <i>heteronuclear RNP</i>) |
| HPr | proteïna portadora de fosfats rica en histidines (de l'anglès <i>histidine-containing phosphocarrier protein</i>) |
| Hsp | proteïnes de xoc tèrmic (de l'anglès <i>heat shock proteins</i>) |
| Hydr | hidrofobicitat |
| HypF-N | domini N-terminal de la proteïna maduradora d'hidrogenases F (de l'anglès <i>N-terminal domain of the hydrogenase maturation protein F</i>) |
| H _z | <i>Helicoverpa zea</i> |
| I | intermediari |
| IBB | Institut de Biotecnologia i Biomedicina |
| IPTG | isopropil-β-tiogalactòsid |
| k | constant de velocitat |
| K | x 1000 augments (en microscòpia) |

| | |
|------------|--|
| K_i | constant d'inhibició |
| LB | Luria-Bertani |
| LCI | <i>leech carboxypeptidase inhibitor</i> (inhibidor de carboxipeptidasa de sangonera) |
| LDL | lipoproteïna de baixa densitat (de l'anglès <i>low density lipoprotein</i>) |
| LIC | clonatge independent de lligació (de l'anglès <i>ligation-independent cloning</i>) |
| ln | logaritme neperià |
| Ltd. | limitat (de l'anglès <i>limited</i>) |
| M1 | mutant d'ADA2h amb l'hèlix α -1 estabilitzada |
| M2 | mutant d'ADA2h amb l'hèlix α -2 estabilitzada |
| MALDI-TOF | espectrometria de masses de desorció/ionització mitjançant làser assistida per matriu i detecció per temps de vol (de l'anglès <i>matrix-assisted laser desorption/ionization – time of flight</i>) |
| MCS | lloc de clonatge múltiple (de l'anglès <i>multicloning site</i>) |
| MerP | proteïna periplàsmica unidora de mercuri (de l'anglès <i>mercuric-ion-binding periplasmic protein</i>) |
| mRNA | àcid ribonucleic missatger (veure també RNA) |
| mut | mutant |
| N | estat natiu |
| NA | no aplicable |
| NMR | ressonància magnètica nuclear (de l'anglès <i>nuclear magnetic resonance</i>) |
| Nna | proteïna nuclear del sistema nerviós induïda per axotomia (de l'anglès <i>nervous system nuclear protein induced by axotomy</i>) |
| N-terminal | aminoterminal |
| NTP | ribonucleòtid trifosfat |
| p | significació |
| p/v | pes/volum |
| p53 | proteïna 53 |
| PABP | proteïna unidora de seqüències poli-A (de l'anglès <i>poly(A)-binding protein</i>) |
| PAGE | electroforesi en gel de poliacrilamida (de l'anglès <i>polyacrilamide gel electrophoresis</i>) |
| PC-3 | línia cel·lular 3 de càncer de pròstata (de l'anglès <i>human prostate cancer-3</i>) |
| PCI | inhibidor de carboxipeptidasa de patata (de l'anglès <i>potato carboxypeptidase inhibitor</i>) |
| PCP | procarboxipeptidasa |
| PCR | reacció en cadena de la polimerasa (de l'anglès <i>polymerase chain reaction</i>) |
| PDB | <i>Protein Data Bank</i> |
| Pfam | <i>Database of Protein Families and HMMs</i> |
| pI | punt isoelèctric |
| PIE | element d'inhibició de la poliadenilació (de l'anglès <i>polyadenylation inhibition element</i>) |
| PIRA | disposicions paral·leles en registre (de l'anglès <i>parallel in-register arrangements</i>) |

| | |
|----------------|---|
| pol | polimerasa |
| PROSITE | <i>database of protein domains, families and functional sites</i> |
| PrP | proteïna del prió o priònica |
| PSI-BLAST | <i>Position-Specific Iterated Basic Local Alignment Search Tool</i> |
| PTB | proteïna unidora del tracte de polipirimidina (de l'anglès <i>polypyrimidine tract binding protein</i>) |
| r | coeficient de correlació |
| R | constant de gasos |
| RBD | domini canònic d'unió a RNA (de l'anglès <i>canonical RNA binding domain</i>) |
| RE | reticle endoplàsmic |
| RMN | ressonància magnètica nuclear |
| RMSD | desviació mitja quadràtica (de l'anglès <i>root mean square deviation</i>) |
| RNA | àcid ribonucleic (de l'anglès <i>ribonucleic acid</i>) |
| RNase | ribonucleasa (de l'anglès <i>ribonuclease</i>) |
| RNP | ribonucleoproteïna |
| RRM | motiu de reconeixement de RNA (de l'anglès <i>RNA recognition motif</i>) |
| SCOP | <i>Structural Classification Of Proteins</i> |
| SDS | dodecilsulfat sòdic (de l'anglès <i>sodium dodecyl sulfate</i>) |
| SH3 | domini 3 d'homologia a la proteïna Sarc (de l'anglès <i>Sarc Homology domain 3</i>) |
| snRNA | RNA nuclear petit (de l'anglès <i>small nuclear RNA</i>) |
| ssDNA | DNA de cadena senzilla (de l'anglès <i>single-stranded DNA</i>) |
| STAMP | <i>Structural Time series Analyser Modeller and Predictor</i> |
| Sup35p | factor de terminació de la traducció de <i>Saccharomyces cerevisiae</i> |
| T | temperatura |
| t | temps |
| TAE | tampó Tris-acetat-EDTA |
| TBE | tampó Tris-borat-EDTA |
| TEM | microscòpia electrònica de transmissió (de l'anglès <i>transmission electron microscopy</i>) |
| TEMED | N,N,N',N'-tetrametilè diamí |
| TFE | trifluoroetanol |
| TGF α | factor de creixement transformant α (de l'anglès <i>transforming growth factor α</i>) |
| TlowE | tampó Tris amb baixa concentració d'EDTA |
| T _m | temperatura de fusió (de l'anglès <i>melting Temperature</i>) |
| TM | marca registrada (de l'anglès <i>trademark</i>) |
| t-PA | activador tissular del plasminogen (de l'anglès <i>tissular plasminogen activator</i>) |
| Tris | Tris-hidroximetilamimetà |
| tRNA | àcid ribonucleic de transferència (veure també RNA) |
| U | estat desplegat |
| U1A | ribonucleoproteïna petita A associada a l'snRNA U1 (de l'anglès <i>U1</i>) |

small nuclear ribonucleoprotein A)

| | |
|------------------|---|
| U2AF | factor auxiliar de l'snRNA U2 (de l'anglès <i>U2 snRNA auxiliary factor</i>) |
| UniProt | <i>Universal Protein Resource</i> |
| UPV | Universitat del País Basc (del castellà Universidad del País Vasco) |
| Ure2p | proteïna de <i>Saccharomyces cerevisiae</i> reguladora transcripcional de repressió per catabòlit de nitrogen (codificada pel gen URE2) |
| UV | ultraviolada |
| v (v) | velocitat |
| v/v | relació volum/volum |
| WL | fibres amiloides en forma de cuc (de l'anglès <i>worm-like</i>) |
| WT | variant salvatge (de l'anglès <i>wild-type</i>) |
| Z _{agg} | paràmetre de propensió d'agregació de ZYGGREGATOR (veure també Z-score) |
| Z-score | paràmetre de propensió d'agregació de ZYGGREGATOR (veure també Z _{agg}) |
| ZTA1p | proteïna ζ-cristal·lina de llevat |

Taula d'abreviatures dels aminoàcids proteïnogènics

| Nom complet | Codi de tres lletres | Codi d'una lletra |
|---------------|----------------------|-------------------|
| Alanina | Ala | A |
| Arginina | Arg | R |
| Asparagina | Asn | N |
| Àcid aspàrtic | Asp | D |
| Cisteïna | Cys | C |
| Àcid glutàmic | Glu | E |
| Glutamina | Gln | Q |
| Glicina | Gly | G |
| Histidina | His | H |
| Isoleucina | Ile | I |
| Leucina | Leu | L |
| Lisina | Lys | K |
| Metionina | Met | M |
| Fenilalanina | Phe | F |
| Prolina | Pro | P |
| Serina | Ser | S |
| Treonina | Thr | T |
| Triptòfan | Trp | W |
| Tirosina | Tyr | Y |
| Valina | Val | V |

Taula d'abreviatures de bases nitrogenades, nucleòsids i nucleòtids

| Nom base nitrogenada | Nom nucleòsid | Nom nucleòtid | Codi d'una lletra |
|----------------------|---------------------|---------------------|-------------------|
| Adenina | (Desoxi-) Adenosina | (Desoxi-) Adenilat | A |
| Citosina | (Desoxi-) Citidina | (Desoxi-) Citidilat | C |
| Guanina | (Desoxi-) Guanosina | (Desoxi-) Guanilat | G |
| Timina | Desoxitimidina | Desoxitimidilat | T |
| Uracil | Uridina | Uridilat | U |

Taula dels prefixos estàndard de les unitats de mesura del sistema internacional

| Prefix | Codi d'una lletra | Valor |
|---------|-------------------|------------|
| yotta- | Y | 10^{24} |
| zetta- | Z | 10^{21} |
| exa- | E | 10^{18} |
| peta- | P | 10^{15} |
| tera- | T | 10^{12} |
| giga- | G | 10^9 |
| mega- | M | 10^6 |
| kilo- | k | 10^3 |
| hecto- | h | 10^2 |
| deca- | da | 10^1 |
| - | - | 10^0 |
| deci- | d | 10^{-1} |
| centi- | c | 10^{-2} |
| mil·li- | m | 10^{-3} |
| micro- | μ | 10^{-6} |
| nano- | n | 10^{-9} |
| pico- | p | 10^{-12} |
| femto- | f | 10^{-15} |
| atto- | a | 10^{-18} |
| zepto- | z | 10^{-21} |
| yocto- | y | 10^{-24} |

Taula d'abreviatures d'unitats

| Abreviatura | Significat |
|--------------------|----------------------------|
| A | amper |
| Å | angstrom |
| b | base |
| cc | centímetre cúbic |
| Ci | curie |
| cpm | comptes per minut |
| Da | dalton |
| deg | grau d'el·lipticitat |
| g | gram |
| g | força gravitatòria |
| h | hora |
| l / L | litre |
| M | molar |
| m | metre |
| mer | monòmer |
| min | minut |
| N | normal |
| °C | grau centígrad |
| Pa | pascal |
| pb | parell de bases |
| rpm | revolucions per minut |
| s | segon |
| U | unitats |
| U _{DO} | unitats de densitat òptica |
| V | volt |

INTRODUCCIÓ

I.1 LES PROTEASES

Les proteases són els enzims encarregats del trencament d'enllaços peptídics de pèptids i proteïnes mitjançant un mecanisme d'hidròlisi. Aquests enzims es troben referits en algunes ocasions com a peptidases o proteïnases, essent peptidasa el terme recomanat per la Unió Internacional de Bioquímica i Biologia Molecular.

Les proteases es troben distribuïdes en diferents organismes, tant en eucariotes, com a procariotes i virus, i les seves funcions són igualment diverses, actuant en la degradació inespecífica durant el recanvi proteic i el metabolisme, o en processaments proteics més específics com poden ser la maduració de les proteïnes constituents de les càpsides dels virus i l'activació de proenzims, prohormones o proneuropèptids. Aquesta àmplia funcionalitat fa que les seves possibles aplicacions siguin també molt extenses, tant en medicina com en biotecnologia.

I.1.1 Proproteases i proteòlisi limitada

Degut a l'evident perillósitat d'una activació de les proteases en llocs o moments inadequats, aquests enzims estan sotmesos a forts controls de regulació, normalment de caràcter postraduccionals: moltes proteases són sintetitzades en forma de proenzims o zimògens, formes inactives i estables. Aquestes formes han de patir un procés de proteòlisi limitada específic per part de la mateixa (autoproteòlisi) o d'una altra proteasa que elimina una part de la molècula i permet l'adquisició de la seva completa funcionalitat.

El segment proteic que inhibeix l'enzim de forma autòloga i és eliminat sol denominar-se segment d'activació o segment pro, i acostuma trobar-se en posició N-terminal respecte l'enzim actiu, tot i que hi ha exemples, com en el factor IX de coagulació o en la proinsulina (Csorba, 1991) on es troben en posició central i C-terminal, respectivament. La longitud d'aquest segment és força variable, des dels 6 residus en el cas del tripsinogen fins als 560 del plasminogen, el qual és fins i tot major que la proteïna activa (exemple: pepsinogen, Fig I.1). La grandària d'alguns segments d'activació ha portat a l'estudi de possibles funcions addicionals a la inhibició de l'enzim, encoratjat pel fet que cada cop més proteïnes presenten el fenomen *moonlighting*, pel qual una molècula té la capacitat de realitzar més d'una funció (Copley, 2003; Jeffery, 1999). Alguns treballs han demostrat la funció d'aquests segments com a xaperones co-traduccionals (Phillips i Rutter, 1996; Sørensen *et al.*, 1993), de manera que la seva presència seria necessària per tal que l'enzim adoptés

el seu plegament correcte. El fenomen de *moonlighting* ha estat demostrat també per a altres proteases, on s'han trobat funcions addicionals com la capacitat del pèptid d'activació del plasminogen per suprimir la secreció gàstrica (Bower *et al.*, 1974).

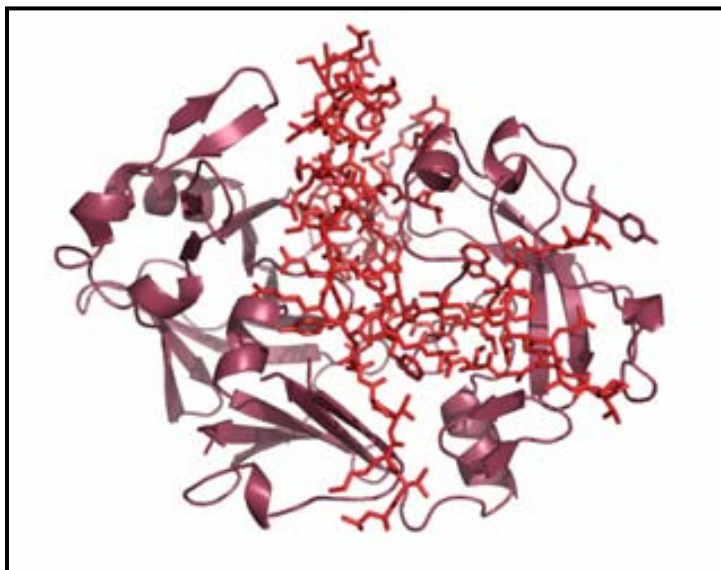


Figura I.1. Exemple d'estructura d'un proenzim: estructura tridimensional del pepsinogen. La forma activa del pepsinogen, la pepsina, és mostrada en forma de diagrama de cintes. La regió inhibidora, de 44 residus, es mostra de color vermell i en visualització de bastons.

La inhibició de l'enzim pot donar-se per varis mecanismes, per exemple creant un impediment estèric per a l'accés al centre actiu, en el cas de les procarboxipeptidases, o bé produint una distorsió del centre actiu o dels llocs de fixació del substrat, en el cas del quimotripsinogen.

I.1.2 Tipus de proteases

Les proteases solen diferenciar-se principalment en endo i exoproteases. Les endoproteases produeixen el tall proteolític a l'interior de la seqüència peptídica i les exoproteases en un dels extrems, a l'N-terminal (aminopeptidases) o al C-terminal (carboxipeptidases).

Apart d'aquesta puntualització, existeixen múltiples classificacions de proteases basades en diferents criteris: especificitat de substrat, mecanisme catalític, relacions estructurals evolutives, localització intracel·lular, funció en l'organisme, etc.

No obstant, la classificació de les proteases més utilitzada és la descrita per Rawlings i Barrett (1993), basada en el residu implicat principalment en el seu mecanisme catalític. Aquesta classificació va ser posteriorment implementada amb la creació de la base de dades MEROPS (<http://merops.sanger.ac.uk/>) (Rawlings *et al.*,

2006), la qual subclassifica els diferents tipus catalítics en famílies, definint-se una família com un conjunt de proteases homòlogues. A més a més, aquesta base de dades inclou també una classificació d'inhibidors de proteases (Rawlings *et al.*, 2004). D'aquesta manera, 8 grups diferents de proteases han estat descrits fins ara: serina proteases, treonina proteases, cisteïna proteases, aspàrtic proteases, metal·loproteases, glutàmic proteases, proteases de mecanisme catalític mixt i proteases de mecanisme desconegut.

Les **serina proteases** es caracteritzen per tenir, principalment, un residu serina al centre actiu involucrat en el mecanisme catalític. Una de les famílies més grans i importants és la de la quimotripsina (S1), on totes les proteïnes descrites fins ara són endoproteases. En aquesta família es poden trobar enzims digestius com la mateixa quimotripsina, l'elastasa i la tripsina. Aquest últim component té especial rellevància ja que es troba implicat, entre d'altres funcions, en el processament de zimògens d'altres proteases que degraden proteïnes provinents de la ingesta. En aquest grup també trobem proteïnes implicades en la coagulació sanguínia com la trombina, l'activador del plasminogen i factors de coagulació, així com d'altres proteïnes destacables com l'elastasa i els factors del complement. Una altra família remarcable és la de la carboxipeptidasa Y (S10), que conté exclusivament serina carboxipeptidases. Aquests enzims solen trobar-se en fongs, gramínies i dicotiledònies, i es caracteritzen per la presència d'un residu glutàmic que precedeix la serina catalítica i que atorga a aquests enzims un pH òptim acídic.

Els components del grup de les **treonina proteases** utilitzen el grup hidroxil de la cadena lateral de la treonina per tal de catalitzar la hidròlisi de l'enllaç peptídic. Aquest grup ha estat, juntament amb les glutàmic proteases, un dels grups descrits més recentment (Seemuller *et al.*, 1995). No s'hi troben gaires famílies, no obstant, una de les més destacades és la de les hidrolases del proteasoma (T1), en la qual les treonines reactives es troben encarades cap al lumen de la superestructura.

Les **cisteïna proteases** tenen un mecanisme catalític en el qual es troba involucrat el grup sulfhidril d'una cisteïna. El grup catalític de les cisteïna proteases engloba una gran quantitat de famílies diferents. Una de les famílies més estudiades és la de la papaïna (C1), que conté endo i exo-peptidases implicades en l'activitat proteolítica de vacuoles procariotes i de lisosomes eucariotes (un exemple en són les catèpsines). Una altra família destacada és la de les caspases (C14), on es troben proteases involucrades en la principal cascada de l'apoptosi.

Els membres del grup de les **aspàrtic proteases** presenten un residu aspàrtic al seu centre actiu, de manera que el pH òptim de la majoria és àcid. Aquest fet produeix

que la seva localització sigui, per tant, a llocs com als lisosomes o bé a l'estómac intervenint activament en la digestió. Un dels seus membres més destacats és la pepsina (família A1), ja que fou la primera proteasa descoberta i el segon enzim en ésser cristal·litzat (Northrop, 1930). Aquesta família només conté endopeptidases que presenten 2 residus aspàrtic al seu centre actiu (James, 2004). Un dels altres elements importants d'aquesta família és la renina, la qual és capaç d'actuar també a pH neutre. Altres famílies importants contenen enzims com les retropepsines (A2), o presenilines (A22).

Les **metal·loproteases** contenen un ió metàl·lic implicat en el mecanisme catalític, que normalment sol ésser un zinc. L'ió metàl·lic sol trobar-se coordinat amb tres cadenes laterals i una molècula d'aigua, la qual, en ésser activada, iniciarà l'atac nucleofílic. Aquest grup presenta un elevat nombre de famílies, entre les quals caldria destacar la representada per la metal·lopeptidasa de matriu 1 (M10), les quals participen en la renovació de la matriu extracel·lular i en importants funcions reguladores; la de les col·lagenases (M9), i la representada per la carboxipeptidasa A1 (M14).

Les **glutàmic proteases** han estat l'últim grup a incorporar-se en aquesta classificació (Fujinaga *et al.*, 2004). La resolució de l'estructura tridimensional de la proteasa model va mostrar que el residu implicat en el mecanisme catalític era un glutàmic, a diferència de la resta de proteases, creant-se, per tant, un nou grup. Al ser un grup relativament nou, només conté una família (G1), representada per la scytalidoglutàmic peptidasa, amb proteases d'origen fúngic.

L'últim grup es troba format per **proteases de mecanisme desconegut**, que vol englobar proteases que no han pogut ésser classificades degut a manca d'informació sobre la seva activitat. Algunes famílies d'aquest grup tenen com a representats el factor d'esporellació SpoIIIGA (U4) i algunes col·lagenases (U32).

1.1.2.1 Les metal·loproteases

Les metal·loproteases, tal i com s'ha descrit anteriorment, són els enzims que posseeixen un ió metàl·lic implicat en el seu mecanisme catalític. Aquest grup catalític de proteases ha estat classificat per Rawlings i Barrett (1995), basant-se en relacions estructurals evolutives. Segons aquesta classificació, les diferents famílies de metal·loproteases es divideixen quant a la manera de coordinar l'ió metàl·lic. Gràcies a la tècnica de cristal·lografia de raigs-X s'ha observat que un dels motius més importants per a la unió de l'ió metàl·lic té la forma HEXXH (clan MA de la classificació

de la base de dades MEROPS). La classificació, doncs, queda formada per quatre grups diferents:

- a) **“HEXXH + E”**: els quatre lligands de l'ió zinc són una molècula d'aigua, les dues histidines del motiu HEXXH i un residu d'àcid glutàmic. La majoria de metal·loproteases d'aquest grup corresponen a endopeptidases. Exemples d'aquest grup són la família de l'aminopeptidasa N (M1), la de la termolisina (M4), la de la carboxipeptidasa Taq de *Thermus aquaticus* (M32) i la de la glicil aminopeptidasa (M3). Aquest grup es correspon amb el subclan MA(E) de la base de dades MEROPS.
- b) **“HEXXH + H/D”**: les proteases que pertanyen a aquest grup presenten un residu d'histidina o un àcid aspàrtic com a quart grup coordinador del zinc, essent els altres tres les dues histidines del motiu HEXXH i una molècula d'aigua, com a la classe anterior. En aquest grup hi trobem exclusivament endopeptidases i constitueixen el subclan MA(M) de MEROPS. D'entre les famílies més importants hi ha la de les metal·loproteases de matriu (M10), la qual ja ha estat descrita breument amb anterioritat; i la família de l'astacina (M12), amb membres com les proteïnes ADAM (amb dominis desintegrina i metal·loproteasa), de cabdal importància en l'activació de proteïnes com la proteïna precursora amiloide (APP) o el factor de creixement transformant α (TGF α).
- c) **“diferent a HEXXH”**: aquest és un grup heterogeni on s'inclouen totes les famílies que tenen un motiu d'unió a ió metàl·lic diferent a HEXXH. No obstant, les histidines són els residus que predominantment coordinen el zinc. Es troben famílies d'aminopeptidases caracteritzades per presentar dos ions metàl·lics en el centre actiu, coordinats, sobretot, per residus àcids, com és el cas de la família leucil aminopeptidasa (M17) i de la metionil aminopeptidasa (M24). En aquest cas, juntament amb la família M52, es poden observar exemples de metal·loproteases amb ions al centre actiu diferents del zinc, com ara el cobalt i el manganès per a la família M24 i el níquel per a la família M52. En el grup amb motiu diferent a HEXXH hi trobem també la família M14, formada per carboxipeptidases, les quals han estat estudiades en aquest treball. Les carboxipeptidases presenten un motiu d'unió a zinc del tipus HXXE+H, coordinat, per tant, amb 2 histidines, un residu d'àcid glutàmic i una molècula d'aigua.

- d) “**desconegut**”: aquest grup engloba totes les famílies per les quals els lligands del metall encara són desconeguts.

I.1.3 Les metal·locarboxipeptidases

Les metal·locarboxipeptidases són els enzims que catalitzen la hidròlisi de residus C-terminals de proteïnes i pèptids, utilitzant un ió zinc en el mecanisme de catàlisi (Vendrell *et al.*, 2004). Aquestes proteïnes desenvolupen funcions variades, des de la degradació inespecífica de pèptids i proteïnes provinents de la ingesta, al processament d'hormones peptídiques i neurotransmissors.

Una de les famílies més àmpliament estudiades, sobretot pel nostre grup d'investigació, ha estat la M14 de la classificació MEROPS, de la qual a continuació realitzarem una descripció exhaustiva.

I.1.3.1 Classificació de les metal·locarboxipeptidases de la família M14

A la família M14 trobem majoritàriament metal·locarboxipeptidases, a les quals d'ara endavant ens referirem simplement com a carboxipeptidases (CPs), amb la finalitat de facilitar la lectura d'aquesta memòria. Els membres d'aquesta família tenen l'ió zinc coordinat tetrahèdicament amb dues histidines i un glutamat, presentant, doncs, un motiu del tipus **HXXE+H**. La família M14 té com a proteasa representativa la carboxipeptidasa A1 d'*Homo sapiens* (hCPA1).

La família M14 pot ésser dividida en quatre subfamílies diferents segons criteris seqüencials i estructurals (MEROPS, Rawlings *et al.*, 2006; Vendrell *et al.*, 2000):

- a) **subfamília M14A o A/B**: (clàssicament anomenades *carboxipeptidases pancreàtiques*), amb la CPA1 humana com a proteasa tipus.
- b) **subfamília M14B o N/E**: (clàssicament anomenades *carboxipeptidases reguladores*), amb la CPE bovina com a proteasa tipus.
- c) **subfamília M14C**: amb la γ -D-glutamil-(L)-meso-diaminopimelat peptidasa I i els seus ortòlegs com a úniques proteases, implicades en el metabolisme de la paret cel·lular bacteriana. Aquest enzim és una endopeptidasa, a diferència de les altres famílies que contenen només exopeptidases, de manera que no ens referirem a ella en el futur.

d) subfamília **M14D**: nom proposat per a una nova subfamília de metal·locarboxipeptidases descrites recentment (Rodríguez de la Vega *et al.*, 2007). Aquests enzims es caracteritzen, principalment, per l'absència de pèptid senyal de secreció a l'extrem N-terminal, de manera que es localitzen al citosol i al nucli cel·lular, prenent també el nom de carboxipeptidases citosòliques (CCPs) (Kalinina *et al.*, 2007). El descobriment d'aquesta família prové de l'observació d'una carboxipeptidasa implicada en la destirosinilització de la tubulina (Argarana *et al.*, 1978; Barra *et al.*, 1988; Webster, 2004), la qual sembla que correspon a l'enzim Nna1/CCP1, l'enzim representatiu d'aquesta subfamília.

En el cas de les carboxipeptidases més caracteritzades (subfamílies M14A i M14B), el grau d'identitat seqüencial és superior al 40% dins de cada subfamília i al voltant del 15-20% entre les subfamílies A/B i N/E. Les carboxipeptidases es troben al llarg de tot el regne animal, tot i que també poden trobar-se al regne vegetal, així com a organismes unicel·lulars i bacteris (Vendrell *et al.*, 2004). En humans, tot i que se n'observa l'expressió en un nombre elevat de teixits diferents, cada element es troba distribuït força restrictivament, mostrant l'especialització de la seva funció.

A continuació comentarem les subfamílies més estudiades de carboxipeptidases: la A/B i la N/E.

Els membres de la **subfamília A/B** són sintetitzats com a procarboxipeptidases (PCPs) que han d'ésser activats mitjançant un mecanisme específic de proteòlisi limitada per part d'una altra proteasa. El resultat d'aquest processament és un enzim actiu monodomini d'uns 34-36 kDa i un segment d'activació d'uns 11 kDa. Els membres d'aquesta subfamília han estat caracteritzats profundament ja que van ser les primeres carboxipeptidases en ésser descobertes. El pH òptim d'activitat dels enzims estudiats fins ara ronda la neutralitat, la qual cosa reflecteix el pH de l'entorn on desenvolupen la seva funció. Una elevada quantitat de proteïnes d'aquesta subfamília són sintetitzades al pàncrees i posteriorment excretades a l'intestí, on són activades i realitzen la funció degradativa de les proteïnes provinents de la ingesta, juntament amb altres proteases complementàries. Aquest fet va fer que les carboxipeptidases d'aquesta família fossin primerament conegudes com a carboxipeptidases pancreàtiques. No obstant, han anat incorporant-se altres carboxipeptidases a aquesta família per raons d'homologia seqüencial, que no realitzen la seva funció a l'entorn digestiu, sinó a altres llocs com a plasma (CPB2) o

l'ambient extracel·lular (CPA3 i CPA6). Degut a tot això, l'ús del terme carboxipeptidases pancreàtiques no està últimament massa recomanat.

Els membres de la **subfamília N/E** no són sintetitzats en forma de precursors inactius, de manera que la seva regulació es realitza bàsicament per un mecanisme de compartimentació força controlat i una elevada especificitat de substrat. Les seves funcions principals es troben lligades al processament d'hormones peptídiques i neurotransmissors. Les carboxipeptidases N/E tenen uns 400 residus i algunes poden tenir múltiples subunitats. Una altra característica diferencial és la possessió d'una regió d'uns 80 residus en posició C-terminal amb homologia estructural al barril- β de la proteïna transtirretina, no present als membres de la subfamília A/B. La funció d'aquest domini encara roman desconeguda, tot i que s'hipotetitza que participés en el plegament del domini carboxipeptidasa o bé que es trobés implicat en la formació d'oligòmers o en la regulació de l'activitat enzimàtica (Reznik i Fricker, 2001).

A part de la classificació seqüencial i estructural, podem trobar altres classificacions més tradicionals en base a diferents criteris. Una classificació alternativa està fonamentada en l'especificitat de les carboxipeptidases per diferents residus C-terminal. Les **formes A** presenten preferència per residus hidrofòbics en posició C-terminal, i les **formes B** per formes bàsiques. El fet que en alguns organismes es va observar la presència de dues isoformes amb especificitat tipus A (Gardell *et al.*, 1988; Pascual *et al.*, 1989), va desencadenar la seva divisió en 2 isoformes lleugerament diferents: la **isoforma A1**, amb preferència per residus C-terminal de classe alifàtica i la **isoforma A2** per residus de classe aromàtica.

Una altra classificació de les carboxipeptidases és la basada en la seva funció fisiològica. D'aquesta manera es distingeixen les carboxipeptidases pancreàtiques A1, A2 i B, involucrades en la digestió, de la resta de carboxipeptidases que realitzen funcions variades fora dels teixits digestius. Les CPs no pancreàtiques són anomenades reguladores per la seva participació en reaccions més específiques. Algunes carboxipeptidases reguladores, però, tenen una homologia seqüencial i estructural més semblant a les carboxipeptidases pancreàtiques que no pas a la resta de reguladores, com és el cas de la CPA4 i la CPB2, fet que va promoure l'ús del terme "*pancreatic-like*", la qual cosa pot portar a confusions. Per aquestes raons es recomana l'ús de la classificació estructural.

Taula I.1. Resum de les principals diferències entre les dues subfamílies de carboxipeptidases.

| Carboxipeptidases subfamília A/B | Carboxipeptidases subfamília N/E |
|--|---|
| presència de prodomini inhibidor | absència de prodomini inhibidor |
| forma activa monodomini | forma activa multidomini |
| absència de domini d'homologia a la transtirretina | presència de domini d'homologia a la transtirretina |
| longitud similar entre tots els membres | longitud força diferent entre tots els membres |
| pH òptim proper a la neutralitat | diferents pHs òptims |
| diferents especificitats | especificitat enfront residus C-terminal bàsics |
| Baixa compartimentació subcel·lular | Alta compartimentació subcel·lular |

I.1.3.2 Carboxipeptidases de la subfamília A/B

A la subfamília A/B hi trobem un gran nombre de carboxipeptidases diferents, en humans: CPA1, CPA2, CPA3 (MC-CPA), CPA4, CPB1, CPA5, CPA6, CPO i CPB2 (plasma CPB, CPU o TAFI).

Les carboxipeptidases clàssiques, les primeres en descobrir-se, foren les pancreàtiques, és a dir, les **carboxipeptidases A1, A2 i B1** (CPA1, CPA2 i CPB1, respectivament). Aquestes carboxipeptidases són sintetitzades al pàncrees com a proenzims i actuen a l'intestí, ja que és on són activades per la tripsina (Avilés *et al.*, 1993). La funció d'aquestes CPs a l'intestí és la degradació inespecífica de proteïnes provinents de la dieta de manera complementària, gràcies a les seves diferents especificitats. La CPA1 hidrolitza preferentment residus C-terminal alifàtics de cadena llarga (Anson, 1935), mentre que la CPA2 té preferència per residus aromàtics voluminosos com el triptòfan (Gardell *et al.*, 1988; Pascual *et al.*, 1989; Catasús *et al.*, 1995). En canvi, la CPB1 té més afinitat per residus bàsics en posició C-terminal (Folk *et al.*, 1960). Dins del context del procés de digestió dels vertebrats, la CPA1 continuaria la degradació dels productes de la hidròlisi portada a terme per la quimotripsina i la CPB1 per la tripsina (Puigserver *et al.*, 1986), ja que la quimotripsina i la tripsina hidrolitzen cadenes polipeptídiques deixant un residu alifàtic i bàsic en C-terminal, respectivament. D'aquesta manera, les carboxipeptidases digestives formen part d'una xarxa de proteases complexa i especialitzada que de manera cooperativa

és capaç de reduir les proteïnes ingerides a aminoàcids. Les carboxipeptidases pancreàtiques són proteïnes extracel·lulars, solubles, i no es troben glicosilades. Aquestes proteases són sintetitzades al pàncrees juntament amb un segment d'activació N-terminal d'uns 94-96 aminoàcids. Aquests zimògens són estables i es troben emmagatzemats en els grànuls pancreàtics mentre no són requerits pel procés de digestió. Les carboxipeptidases són uns dels enzims que es caracteritzen per la presència d'un segment d'activació gran, tal i com s'havia comentat en el punt 1.1.1, el qual arriba a suposar una quarta part de la molècula. S'hipotetitzava, doncs, l'existència d'alguna altra funció per a aquests dominis. El fet que les formes proenzimàtiques es trobin tan compartimentades, fa que siguin bones candidates per a la utilització d'aquests enzims com a marcadors de certes patologies. La procarboxipeptidasa B1 (PCPB1) en sèrum, per exemple, pot ser un indicador de pancreatitis aguda (Yamamoto *et al.*, 1992), i de deficiència renal crònica (Chen *et al.*, 1996).

S'ha observat que la PCPA1, en mamífers, pot trobar-se en forma monomèrica, formant complexos binaris amb les serina proteases proproteïnasa E (PPE) o quimotripsinogen C (QTC), o bé formant un complex ternari amb ambdues. Les formes B, en canvi, apareixen només com a monòmers. Aquest complex ternari ha estat descrit únicament en el sistema boví (Pascual *et al.*, 1990; Ventura *et al.*, 1997), i es disposa de la seva estructura tridimensional gràcies a la tècnica de cristal·lografia de raigs-X (Gomis-Rüth *et al.*, 1997, codi PDB: 1PYT). Encara no es té una idea contrastada del paper biològic d'aquestes estructures quaternàries de la PCPA1, tot i que es barallen unes quantes hipòtesis: control i potenciació de l'activació, modulació de l'activitat, protecció dels enzims front el pH àcid del suc pancreàtic i/o coordinació de l'aparició de les diferents activitats proteolítiques.

Les procarboxipeptidases pancreàtiques han estat durant molt de temps el principal objecte d'estudi del nostre grup. Han estat caracteritzats els proenzims de varis mamífers i insectes (Cohen *et al.*, 1956; Segundo *et al.*, 1982; Vilanova *et al.*, 1985; Bayés *et al.*, 2005), així com de les formes humanes (Pascual *et al.*, 1989). A partir d'extractes aquosos de polvoritzat pancreàtic humà van aïllar-se tres formes A i dues B. Les tres formes A es corresponen a les formes monomèriques PCPA1 i PCPA2, i la tercera és deguda al complex binari format entre la PCPA1 i la proproteïnasa E. Les dues formes B corresponen a dues formes al·lelomòrfiques del mateix enzim. Posteriorment, s'aconseguí aïllar els cDNAs de les tres procarboxipeptidases, determinant-ne les seves seqüències nucleotídiques i aminoacídiques derivades (Catasús *et al.*, 1992; Catasús *et al.*, 1995; Aloy *et al.*, 1998) i prosseguint a la seva clonació en altres organismes senzills per tal de possibilitar-ne l'expressió heteròloga (Reverter *et al.*, 1998; Barbosa Pereira *et al.*, 2002).

La **carboxipeptidasa A3** o de mastòcits (CPA3 o MC-CPA) és una proteïna sintetitzada en forma de proenzim en aquestes cèl·lules i emmagatzemada en els seus grànuls de secreció, per tal d'alliberar-se enfront d'alguna resposta. La seva funció no és coneguda amb massa detall, però se suposa que participaria en la hidròlisi de substrats peptídics en tàndem amb altres proteases (Reynolds *et al.*, 1989). Aquesta funció la realitzaria en processos on els mastòcits es troben implicats, com els asmàtics o inflamatoris; o bé en la resposta al·lèrgica. L'enzim presenta activitat de tipus A, tot i que s'ha observat que guarda un percentatge d'identitat superior enfront la forma B. La CPA3 és activada intracel·lularment, dins dels grànuls de secreció (Rath-Wolfson, 2001), i el tall es dona sobre un residu àcid, contraposadament al bàsic de les formes pancreàtiques. Aquest fet denota, per tant, l'existència d'un enzim activador alternatiu a la tripsina en aquest compartiment.

La **carboxipeptidasa A4** (CPA4) -inicialment denominada com a CPA3- s'expressa en cèl·lules canceroses de pròstata i en teixits en desenvolupament (Huang *et al.*, 1999), tot i que no es coneix la funció que hi desenvolupa. Els treballs fets fins ara d'aquesta proteïna han estat principalment dirigits a l'estudi de la seva expressió, seguint la inducció del seu mRNA en línies cel·lulars canceroses independents d'andrògens (cèl·lules PC-3). El tractament amb butirat sòdic produeix un augment sorprenent de la quantitat d'mRNA de la PCPA4, fet que queda contrarestat amb l'administració de tricostatina A, un potent inhibidor de les histona desacetilases. Aquest experiment demostra, per tant, la regulació d'aquest gen depenent de la via d'hiperacetilació de les histones. Aquesta proteasa és sintetitzada en forma de proenzim, de característiques similars a les de la resta de components de la subfamília A/B, amb les que comparteix més d'un 40% d'homologia. La CPA2, és, però, la carboxipeptidasa amb major homologia seqüencial amb la CPA4 (68%). Els estudis a nivell de proteïna són relativament recents, essent destacada l'obtenció de l'estructura terciària de la PCPA4 (García-Castellanos *et al.*, 2005).

La **carboxipeptidasa B2** (CPB2), també anomenada carboxipeptidasa R (CPR), carboxipeptidasa B de plasma (pCPB), carboxipeptidasa U (CPU) o *thrombin-activatable fibrinolysis inhibitor* (TAFI), és sintetitzada a fetge i circula per plasma en forma proenzimàtica unida a plasminogen. Els seus nivells plasmàtics es troben altament regulats. La PCPB2 comparteix una major identitat seqüencial amb la PCPB1 (42%) i s'ha demostrat experimentalment la seva especificitat enfront residus C-terminals bàsics. Presenta, com tots els membres de la subfamília A/B, un domini d'activació d'uns 95 residus, bé que en el cas de la PCPB2 aquest es troba glicosilat. La PCPB2 és activada per la trombina produïda per la cascada de coagulació, provocant-se l'alliberament del seu segment d'activació i generant un enzim altament

inestable. La CPB2 catalitza, al seu torn, l'eliminació de residus C-terminals arginina i lisina de la fibrina parcialment degradada, eliminant-ne la capacitat d'unir plasminogen i impedit la degradació de la fibrina per part de l'activador tissular del plasminogen (t-PA) (Wang *et al.*, 1998). Aquesta activitat fa que la CPB2 tingui un paper important en la inhibició de la fibrinòlisi i que connecti les cascades de coagulació i fibrinolítica. Altres funcions atribuïdes a la CPB2 són el processament de pèptids bioactius generats durant el procés d'inflamació, com les quinines i anafilotoxines (Campbell i Okada, 1989; Campbell *et al.*, 2001; Campbell *et al.*, 2002). El model de l'estructura de la CPB2 (Barbosa Pereira *et al.*, 2002) i la resolució de la seva estructura (Anand *et al.*, 2008; Marx *et al.*, 2008), han fet pal·lesa la seva similitud estructural amb la PCPB1 i s'ha detectat una zona molt dinàmica que explicaria la seva inestabilitat, permetent una regulació molt acurada dels nivells de l'enzim a plasma. Gràcies a aquestes estructures s'ha pogut elaborar una teoria sobre la inactivació de la CPB2, segons la qual l'elevada mobilitat de la regió esmentada acabaria patint, en qüestió de poc temps, un canvi conformacional irreversible que deixaria exposat un lloc de tall per la trombina. Amb aquest processament, la CPB2 esdevindria totalment inactiva, desencadenant en la seva degradació (Marx *et al.*, 2008).

La **carboxipeptidasa A6** ha estat descrita recentment a nivell proteic, tot i que havia estat identificada prèviament al genoma humà (Wei *et al.*, 2002). Els nivells de mRNA de la CPA6 s'han trobat especialment elevats al bulb olfatori adult, i, en menor mesura, a l'epidídim (Fontenele-Neto *et al.*, 2005). No obstant, aquest enzim es troba àmpliament distribuït en teixits embrionaris. La CPA6 té preferència per residus hidrofòbics grans en posició C-terminal, tot i que també és capaç d'hidrolitzar histidines (Lyons *et al.*, 2008). Estudis de localització han demostrat que aquest enzim es troba ancorat a la matriu extracel·lular i es creu que actuaria en el processament de neuropèptids en aquest compartiment. Aquesta hipòtesi es veuria recolzada per la observació de que la disrupció del gen de la CPA6 a humans es troba relacionada amb la síndrome de Duane.

Uns altres membres de la subfamília A/B també han estat identificats recentment al genoma humà: les **carboxipeptidases A5 i O**, juntament amb la carboxipeptidasa A6, esmentada en el paràgraf anterior (Wei *et al.*, 2002). Es va realitzar una predicció dels centres actius d'aquestes carboxipeptidases basant-se en models bioinformàtics (Wei *et al.*, 2002). El model de la CPA5 indica que hidrolitzaria preferentment residus C-terminals alifàtics/aromàtics amb cadenes laterals petites, la qual cosa ha estat confirmada experimentalment mitjançant expressió a petita escala de l'ortòleg de ratolí (Wei *et al.*, 2002). No obstant, la predicció per la CPO semblaria indicar que té preferència per residus àcids en posició C-terminal, ja que la posició 255, responsable

de l'especificitat de substrat, és una arginina, un residu bàsic. La CPO seria, doncs, la primera carboxipeptidasa en presentar una activitat preferent enfront d'aquest tipus de residus. Aquesta hipòtesi no ha pogut ésser comprovada experimentalment ja que aquesta carboxipeptidasa encara no ha estat estudiada a nivell proteic. Els segments d'activació de les carboxipeptidases A5 i A6 (de la qual també es va realitzar un model, Wei *et al.*, 2002), són molt semblants a la resta de membres de la subfamília A/B, en canvi, el de la CPO no ha estat encara identificat. L'expressió de CPA5 s'ha detectat principalment en testicles i pàncrees (Fontenele-Neto *et al.*, 2005) i la de CPO roman encara desconeguda.

A part de les carboxipeptidases humanes i de vertebrats en general, recentment s'han identificat i caracteritzat carboxipeptidases d'altres organismes. D'especial importància han estat les carboxipeptidases d'insectes plaga, com la CPAHa d'*Helicoverpa armigera*, una plaga del cotó, amb activitat tipus A (Bown *et al.*, 1998) i la CPBHz d'*Helicoverpa zea*, una plaga de solenàcies, la qual presenta resistència a la inhibició per part d'un inhibidor clàssic de carboxipeptidases, l'inhibidor de patata (PCI) (Bayés *et al.*, 2005). També s'han trobat carboxipeptidases a dípters hematòfags com a *Anopheles gambiae* (Edwards *et al.*, 1997), l'anomenada **MeCPA** al fong entomopatogen *Metarhizium anisopliae* i la **carboxipeptidasa T** a bacteris com *Thermoactinomyces vulgaris* (Smulevitch *et al.*, 1991) i *Streptomyces griseus* (Narahashi, 1990). Aquesta última carboxipeptidasa es caracteritza per una especificitat de substrat força ampla, ja que presenta activitat tipus A i B, capaç, doncs, d'hidrolitzar tant residus C-terminals bàsics com hidrofòbics amb eficiències comparables.

Amb la seqüenciació creixent de genomes d'organismes s'han anat identificat nous gens amb possibles productes homòlegs a les carboxipeptidases estudiades.

I.1.3.3 Carboxipeptidases de la subfamília N/E

La **carboxipeptidasa E** (CPE), també anomenada carboxipeptidasa H (CPH) o encefalina convertasa va ser la primera proteasa d'aquesta subfamília en ésser descoberta. La CPE es troba a les vesícules secretores madures de les cèl·lules neuroendocrines on processa hormones peptídiques i neurotransmissors (Fricker *et al.*, 1989). El pH òptim d'activitat de la CPE és aproximadament 5 – 6 (pH de les vesícules secretores madures), en canvi l'activitat és quasi nul·la a pH neutre. Aquest rang de pH d'activitat tan estret permet la regulació de la proteïna, ja que aquesta es troba inactiva en altres regions per on transita, com l'aparell de Golgi o l'exterior cel·lular (Greene *et al.*, 1992). En aquestes vesícules co-localitza amb les

endoproteases prohormona convertasa 1 i 2, que realitzen la seva acció sobre precursors peptídics neuroendocrins deixant lisines i arginines en posició C-terminal. La CPE, posteriorment, eliminarà aquests residus bàsics terminals per tal que aquests pèptids siguin bioactius. Com les altres CPs de la subfamília N/E, la CPE posseeix un domini enzimàtic tipus carboxipeptidasa i un domini homòleg a la transtirretina. La CPE és l'única CP de la subfamília N/E sintetitzada amb un propèptid N-terminal de 14 residus, tot i que no inhibeix o disminueix l'activitat de l'enzim. A més, la presència d'una hèlix C-terminal amfòtera de 25 residus li permet l'ancoratge a membrana.

La **carboxipeptidasa D** (CPD) realitza una funció parcialment solapada amb la de la CPE, tot i que no pot contrarestar-ne completament la manca (Fricker *et al.*, 1996). La CPD actua a la mateixa via secretora que la CPE però la seva funció no està limitada a teixits neuroendocrins. Aquesta proteïna presenta 3 dominis similars a la CPE, fet no observat a cap altra carboxipeptidasa, a més d'una regió transmembranal de 20 residus i una cua citosòlica d'uns 60 (Tan *et al.*, 1997). Els 2 primers dominis carboxipeptidasa són lleugerament diferents, el primer és òptim a pH 7 i hidrolitza principalment arginines C-terminals i el segon és òptim a pH 5 – 6 i hidrolitza preferentment lisines. Això fa que la CPD tingui un rang de pH de treball força ample permetent la seva acció a múltiples localitzacions de la via secretora i fins i tot a la superfície cel·lular (Song i Fricker, 1995). La CPD seria activa al llarg de la via secretora de l'aparell de Golgi, on es creu que desenvoluparia la seva principal funció, relacionada amb la producció de receptors i factors de creixement processats per l'endoproteasa furina i altres enzims relacionats. El tercer domini és inactiu, ja que li manquen els residus catalítics essencials per a desenvolupar no només l'activitat carboxipeptidasa, sinó qualsevol activitat tipus hidrolasa (Aloy *et al.*, 2001). Es creu que la seva acció pugui ser de domini d'unió. L'estructura del domini II de la CPD d'ànec ha estat resolta a finals dels 90 (Gomis-Rüth *et al.*, 1999), essent la primera estructura d'un membre de la subfamília N/E que s'ha pogut estudiar.

La **carboxipeptidasa M** (CPM) és una glicoproteïna unida a membrana molt abundant a pulmons i placenta, tot i que també es troba present a vasos sanguinis i a la superfície de glòbuls blancs un cop diferenciats. Formes solubles de la mateixa proteïna localitzen a fluids, com al plasma seminal, al líquid amniòtic o a l'orina. A membrana, es creu que podria participar en el control de l'activitat d'hormones peptídiques, en la modulació de receptors i en el processament o degradació de proteïnes extracel·lulars. Un estudi ha demostrat la seva capacitat per a generar des-Arg53-EGF, a més de la coincidència en la distribució de la proteasa amb el factor de creixement epidèrmic (EGF), de manera que s'ha hipotetitzat una implicació d'aquest enzim en el metabolisme de factors de creixement (McGwire i Skidgel, 1995). Aquesta

proteïna consta dels dominis carboxipeptidasa i d'homologia a la transtirretina comuns en totes les proteases membres de la subfamília N/E, i presenta, a més, un domini C-terminal que li permet la unió a la membrana (Skidgel *et al.*, 1991). La seva especificitat, com en moltes proteïnes d'aquesta subfamília, és vers residus C-terminals bàsics.

La **carboxipeptidasa N** (CPN) és sintetitzada a fetge i alliberada a sang, on es troba exclusivament en forma multimèrica de 280 kDa i a concentracions elevades. Aquesta forma multimèrica està composta per dues còpies de la subunitat catalítica de 50-55 kDa i dues còpies d'una subunitat glicosilada de 83 kDa sense activitat coneguda (Erdös, 1979). Les subunitats glicosilades protegeixen les subunitats catalítiques enfront la degradació i la seva retirada del torrent sanguini per filtració glomerular (Levin *et al.*, 1982). La CPN té elevada afinitat per residus bàsics en posició C-terminal i processa una elevada varietat de substrats peptídics, essent el principal inactivador de pèptids potents com les quinines i les anafilotoxines. Es creu que podria actuar conjuntament amb les formes solubles de la CPM (Skidgel *et al.*, 1996). Aquest enzim també és conegut amb els noms de carboxipeptidasa K (CPK), inactivador d'anafilotoxina i carboxipeptidasa B de sèrum.

Les **carboxipeptidases X1 i X2** (CPX1 i CPX2) i l'**adipocyte enhancer-binding protein** (AEBP) comparteixen entre elles una mica més d'homologia seqüencial que no pas amb altres metal·locarboxipeptidases. Aquestes proteases contenen un domini N-terminal extra d'homologia a la discoïdina i altres lectines, i s'ha vist que són inactives enfront substrats estàndard de carboxipeptidases (Xin *et al.*, 1998; Lei *et al.*, 1999; Reznik i Fricker, 2001). Comparacions dels residus dels seus setis actius amb els d'altres carboxipeptidases han revelat la manca d'un o més residus crítics per al desenvolupament de l'activitat enzimàtica, de manera que es creu que la seva activitat aniria més encaminada cap a possibles funcions d'unió. La CPX1 es troba en teixits embrionaris, de manera que la seva funció sembla estar lligada al procés de desenvolupament, possiblement mitjançant interaccions cel·lulars a través del seu domini discoïdina. La CPX2 es troba àmpliament distribuïda al cervell, on actuaria segurament com a proteïna d'unió. L'AEBP1/ACLIP és un repressor transcripcional l'expressió del qual és suprimida durant la diferenciació dels adipòcits (He *et al.*, 1995). S'ha descrit una isoforma de l'AEBP, anomenada ACLP (*aortic carboxypeptidase-like protein*), que presenta una extensió N-terminal de 380 amino àcids. L'expressió d'aquesta isoforma augmenta durant la diferenciació de les cèl·lules del múscul llis de l'aorta (Layne *et al.*, 1998).

La **carboxipeptidasa Z** (CPZ) ha estat trobada a la matriu extracel·lular de múltiples teixits embrionaris, a trofoblasts invasius de placenta i a cèl·lules malignes

d'adenocarcinoma de colon. Aquesta proteasa presenta un domini ric en cisteïnes amb un 20-35% d'identitat seqüencial amb proteïnes extracel·lulars *frizzled* de *Drosophila* i mamífers (Song i Fricker, 1997), mitjançant el qual es creu que podria interaccionar amb proteïnes del tipus Wnt. Les proteïnes Wnt són molècules molt importants en les etapes primerenques del desenvolupament, assegurant que aquest es porti a terme de manera normal. Aquestes proteïnes interaccionen amb receptors de membrana que contenen el domini *frizzled* i transmeten el senyal a l'interior cel·lular. A part d'aquest fet, s'ha observat que el gen que codifica la CPZ es troba en una regió cromosòmica associada amb desenvolupament embrionari, i la seva expressió té un patró coincident amb múltiples gens *Wnt*. Han estat proposats varis models d'actuació de la CPZ, els quals poden ser complementaris, basant-se, sobretot, en l'existència d'aquest domini *frizzled* tan conservat i l'activitat d'aquesta proteasa sobre residus C-terminalis bàsics (Reznik i Fricker, 2001).

Alguns membres de la subfamília N/E han estat trobats també a organismes invertebrats, com és el cas d'enzims *CPE-like* (Juvvadi *et al.*, 1997) i *CPD-like* (Fan *et al.*, 1999) al mol·lusc marí *Aplysia californica* i a l'insecte *Manduca sexta* (Stone *et al.*, 1994), respectivament. També han estat localitzats a bacteris com *Porphyromonas gingivalis* (Masuda *et al.*, 2002), bé que la seva funció és diferent. Igualment que en el cas de l'altra subfamília, el nombre de possibles membres creix a mesura que es finalitzen noves seqüenciacions de genomes complets d'organismes.

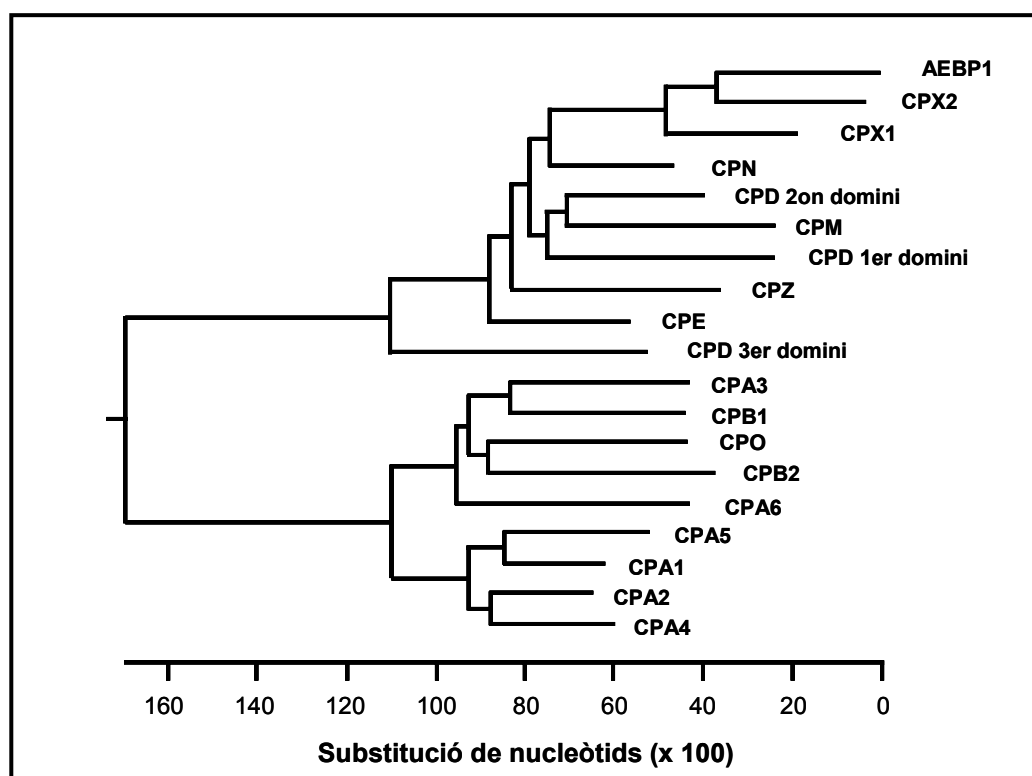


Figura I.2. Arbre filogenètic de les procarboxipeptidases humanes de les subfamílies A/B i N/E. Relació evolutiva entre les seqüències nucleotídiques dels mRNAs de procarboxipeptidases humanes. Extret de Vendrell *et al.* (2004).

I.1.3.4 Estructura de les metal·locarboxipeptidases de la subfamília A/B

La determinació de les estructures tridimensionals de metal·locarboxipeptidases és essencial per tal d'entendre el seu funcionament i dissenyar-ne qualsevol variació o molècula inhibidora. Tal i com s'ha vist en els apartats anteriors, les metal·locarboxipeptidases es troben implicades en un elevat nombre de funcions a l'organisme, de manera que anomalies en el seu funcionament poden ser causants de disfuncions, en alguns casos, de remarcada importància (Franco *et al.*, 2001; Mosnier *et al.*, 2001). Aquest fet demostra la seva gran potencialitat com a dianes terapèutiques, essent interessant el desenvolupament d'aplicacions mèdiques com l'ADEPT (*Antibody Directed Enzyme Prodrug Therapy*) (Smith *et al.*, 1997).

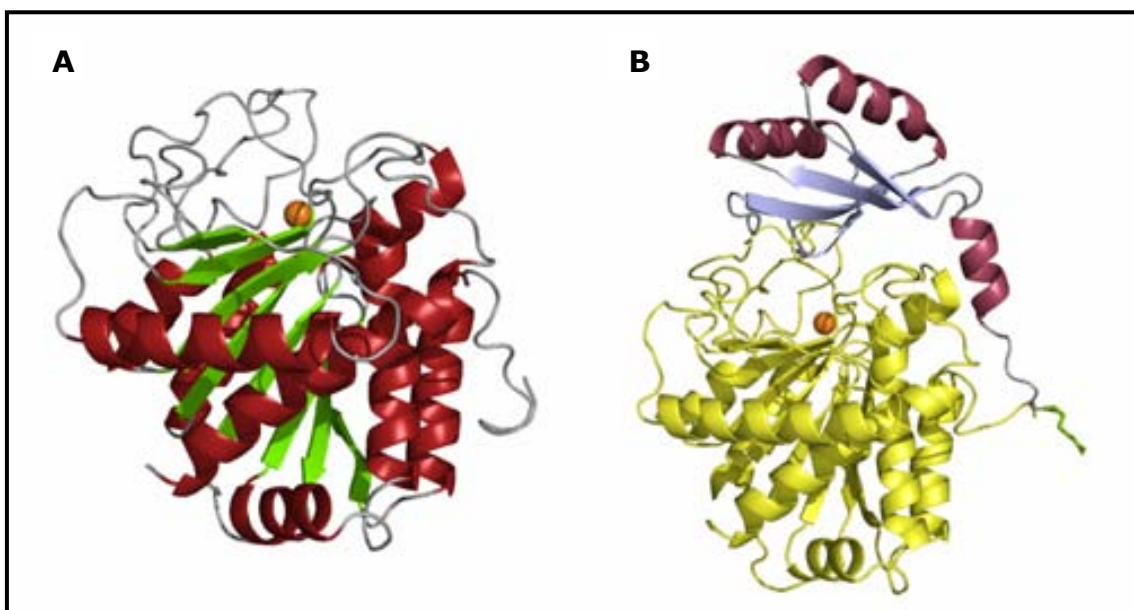


Figura 1.3. Estructura d'un domini carboxipeptidasa i una procarboxipeptidasa. (A) Diagrama de cintes de la carboxipeptidasa A1 bovina. Les hèlixs- α són mostrades en vermell i les cadenes- β en verd. (B) Diagrama de cintes de la procarboxipeptidasa B1 bovina. El segment d'activació es mostra en colors lila i blau, segons la seva estructura secundària, i el domini actiu en color groc. La cadena lateral verda correspon a l'arginina sobre la qual es produeix la proteòlisi limitada activant. En ambdues estructures l'esfera taronja representa l'ió de zinc.

L'estructura de l'enzim model de la subfamília A/B, la carboxipeptidasa A1 bovina ha estat resolta tant en la seva forma aïllada (Christianson i Lipscomb, 1986; codi pdb: 3cpa), com en complex amb inhibidors (Quioco i Lipscomb, 1971; Rees *et al.*, 1983; Christianson *et al.*, 1989; codis pdb: 4cpa i 5cpa). L'enzim activat té forma globular, i conté una fulla- β central de vuit cadenes paral·leles i antiparal·leles, amb un gir de 120° entre la primera i la darrera cadena, sobre la qual s'empaqueten vuit hèlixs- α laterals (Fig. 1.3). El centre actiu es troba dins una cavitat formada per la fulla- β , dues hèlixs- α i una regió irregular que la cobreix parcialment.

La resolució de les estructures tridimensionals de les formes proenzimàtiques revelaren la topologia dels segments d'activació, així com el seu mecanisme d'inhibició. Primerament van ésser resoltes les estructures de procarboxipeptidases d'origen porcí (Coll *et al.*, 1991; Guasch *et al.*, 1992), corresponents a les formes A i B, a més del domini globular del segment d'activació de la PCPB1 porcina mitjançant RMN (Vendrell *et al.*, 1990a; Vendrell *et al.*, 1990b; Vendrell *et al.*, 1991). Actualment es coneixen també les estructures dels proenzims A2 (García-Sáez *et al.*, 1997), A4 (García-Castellanos *et al.*, 2005), B1 (Barbosa Pereira *et al.*, 2002) i B2 humans (Marx *et al.*, 2008), a més de la PCPHa de lepidòpter (Estébanez-Perpiñá *et al.*, 2001) i la PCPB2 bovina (Anand *et al.*, 2008). Apart de les estructures cristal·logràfiques, també es coneix l'estructura del domini d'activació de la PCPA2 humana per RMN (Jiménez

et al., 2003). Tot i que els segments d'activació de les diferents formes proenzimàtiques presenten un grau d'identitat baix, la seva estructura i disposició respecte el domini catalític són molt similars. L'estructura d'aquestes regions és comentada extensament en el pròxim apartat. Altres carboxipeptidases de la subfamília A/B amb l'estructura tridimensional determinada són la forma activa de la CPA1 humana (Pallarès *et al.*, 2008); el complex ternari de la CPA amb les dues serina proteases (Gomis-Rüth *et al.*, 1997) i la CPBHz de l'insecte *Helicoverpa zea* (Bayés *et al.*, 2005).

De gran importància és també la resolució de les estructures tridimensionals dels complexos de carboxipeptidases amb inhibidors, ja que permeten entendre el mecanisme pel qual suprimeixen l'activitat de l'enzim. L'inhibidor de carboxipeptidases de patata (PCI) posseeix una cua C-terminal força rígida a l'observar-la per RMN (González *et al.*, 2003; codi pdb: 1h20), la qual s'ha vist que es col·loca a l'interior del centre actiu de l'enzim (Rees i Lipscomb, 1982; codi pdb: 4cpa).

Altres complexos resolts són els de la CPA1 bovina amb l'inhibidor de sangonera (LCI) (Arolas *et al.*, 2005; codi pdb: 2abz), o el de la CPA4 humana amb l'inhibidor endogen ECI (Pallarès *et al.*, 2005, codi pdb: 2bo9), entre d'altres.

Respecte a les altres subfamílies de metal·locarboxipeptidases, cal remarcar que els dominis actius tenen estructures tridimensionals molt similars a les de la subfamília A/B, havent estat resolta l'estructura tridimensional del domini II de la CPD d'ànec (Gomis-Rüth *et al.*, 1999) i modelada l'estructura de la carboxipeptidasa citosòlica AGBL4 (*ATP/GTP binding-like protein 4*) de *Caenorhabditis elegans* (Rodríguez de la Vega, *et al.*, 2007).

I.1.3.5 Segments d'activació i mecanisme d'activació de les metal·locarboxipeptidases de la subfamília A/B

Com ja s'ha anat exposant, les metal·locarboxipeptidases de la subfamília A/B són sintetitzades en forma de precursors inactius o zimògens, a l'igual que una bona part de les proteases conegudes. Els segments d'activació presenten una regió globular amb una topologia antiparal·lela- α /antiparal·lela- β en forma de sandvitx obert. La fulla- β té una posició central cobrint el centre actiu de la carboxipeptidasa per un costat i les hèlixs- α es troben a la superfície oposada. El segment que connecta la regió globular amb l'enzim està estructurat en una hèlix- α de longitud variable en funció dels diferents enzims i un llaç on, generalment, es troben els punts d'activació del proenzim.

Amb la resolució de l'estructura tridimensional dels proenzims de les metal·locarboxipeptidases de la subfamília A/B es va poder observar que els segments d'activació es troben formats per dues parts ben diferenciades: el domini d'activació globular (AD) i l'hèlix de connexió. El domini d'activació és una regió de plegament independent del domini catalític dins de la procarboxipeptidasa i presenta una elevada quantitat d'estructura secundària, la qual estabilitza el seu plegament. Les diferències principals entre els segments d'activació de les procarboxipeptidases resideixen sobretot a l'hèlix de la regió de connexió amb l'enzim. No obstant, al domini d'activació hi ha una deleció de 2 aminoàcids entre l'hèlix α -1 i la cadena β -2 típica de les formes B que es troba acompanyada d'una inserció de 4 residus que propicia la formació d'una hèlix 3_{10} entre les cadenes β -2 i β -3 (Fig. I.3 B i I.4). Aquesta hèlix 3_{10} es creu que és en part responsable de la manca d'activitat intrínseca observada per part de la PCPB1. Curiosament, aquestes insercions i deleccions han estat identificades al model realitzat de la PCPA6, la qual conté una hèlix 3_{10} igual que les presents a les formes B (Wei *et al.*, 2002).

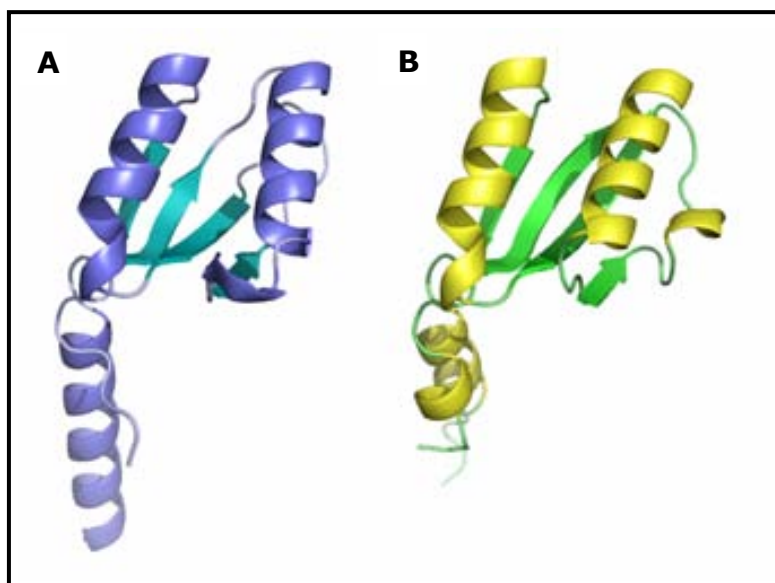


Figura I.4. Estructura dels dominis d'activació de les procarboxipeptidases A2 (A) i B1 (B). Les principals diferències es troben en l'estructuració de l'hèlix- α de la regió de connexió, molt més llarga i estructurada en les formes A, i la presència de l'hèlix 3_{10} en les formes B i en el model de la PCPA6.

La seva estructura ha estat resolta tant per cristal·lografia de raigs X formant part del proenzim, com en forma aïllada per RMN. En el cas del domini d'activació de la procarboxipeptidasa A2 humana (ADA2h), l'estructura del domini en solució presenta una petita diferència respecte a l'estructura del proenzim sencer: la segona cadena- β es troba força desestructurada (Jiménez *et al.*, 2003). Aquesta cadena conté els

residus responsables del contacte amb el centre actiu de l'enzim i, per tant, de bona part de la inhibició de l'activitat carboxipeptidasa. La presència d'aquests contactes produeix l'estructuració d'aquesta regió en la forma proenzimàtica, però la seva manca en la forma aïllada en fa impossible el seu manteniment. La petita mida del domini d'ADA2h, així com l'elevada quantitat d'estructura secundària i la manca tant de ponts disulfur com de prolines en configuració *cis* (a diferència d'ADBh) han convertit aquest domini en un bon model d'estudi de plegament. Aquests experiments són comentats en profunditat més endavant ja que aquest domini ha estat una de les proteïnes objectes d'estudi d'aquesta tesi. Els estudis de plegament van descobrir la major velocitat de plegament d'una proteïna observada fins aquell moment (Villegas *et al.*, 1995). Aquest fet, complementat amb el significant increment en el nivell d'expressió de la carboxipeptidasa al realitzar-se en forma proenzimàtica van desvelar una possible funció del domini com a xaperona co-translacional, assistint el plegament del domini catalític. La grandària i el nivell d'estructuració del segment d'activació han produït la realització d'altres treballs encarats cap la cerca d'una possible funció addicional a la inhibició de la carboxipeptidasa. Una de les primeres funcions proposades fou l'activitat com a proteïna unidora de calci, ja que s'havia descrit certa semblança d'aquests dominis amb les proteïnes de la família EF-hand (Vilanova *et al.*, 1987). No obstant, els experiments d'unió no van ser fructífers (Vilanova *et al.*, 1988) i finalment la similitud seqüencial va ésser desestimada mitjançant mètodes bioinformàtics (Vendrell *et al.*, 1989).

L'activació de les procarboxipeptidases de la subfamília A/B es dona per un mecanisme de proteòlisi limitada, normalment sobre residus bàsics del llaç final de l'hèlix de connexió dels segments d'activació. Aquesta proteòlisi es duu a terme per part d'una endoproteasa específica i suposa l'alliberament del segment d'activació sencer, perdent-se, aproximadament, una quarta part de la molècula. La proteasa responsable del tall activant a les procarboxipeptidases pancreàtiques de mamífers és la tripsina, una serina proteasa secretada a l'intestí on coincideix amb les PCPs. Ha estat descrit que els segments d'activació d'algunes procarboxipeptidases del tipus A poden comportar-se com inhibidors força potents (K_i d'ordre nM), almenys durant un cert temps després de l'activació. Una anàlisi de l'activitat enzimàtica generada durant el procés d'activació pot determinar si el segment pro està actuant com a inhibidor en *trans*, és a dir, un cop alliberat del domini catalític. Si del seguiment de la generació de l'activitat s'obté una corba sigmoide estem en un cas de pro-domini inhibidor en *trans*; si, contràriament, la forma és d'una hipèrbola, aquest fenomen no es produeix.

I.2 PLEGAMENT I AMILOIDOGÈNESI

I.2.1 Plegament

El plegament és el procés mitjançant el qual una cadena polipeptídica determinada (o estructura primària) adopta la seva estructura tridimensional (o estructura terciària) característica en condicions fisiològiques: l'anomenat estat natiu. Aquest estat és estable gràcies a la presència d'interaccions locals i de llarga distància. Les interaccions locals donen lloc a una sèrie d'estructures denominades secundàries que poden ser regulars o irregulars. Les principals estructures secundàries regulars de les proteïnes són les hèlix α i les fulles β (Branden i Tooze, 1991) (Fig. I.5).

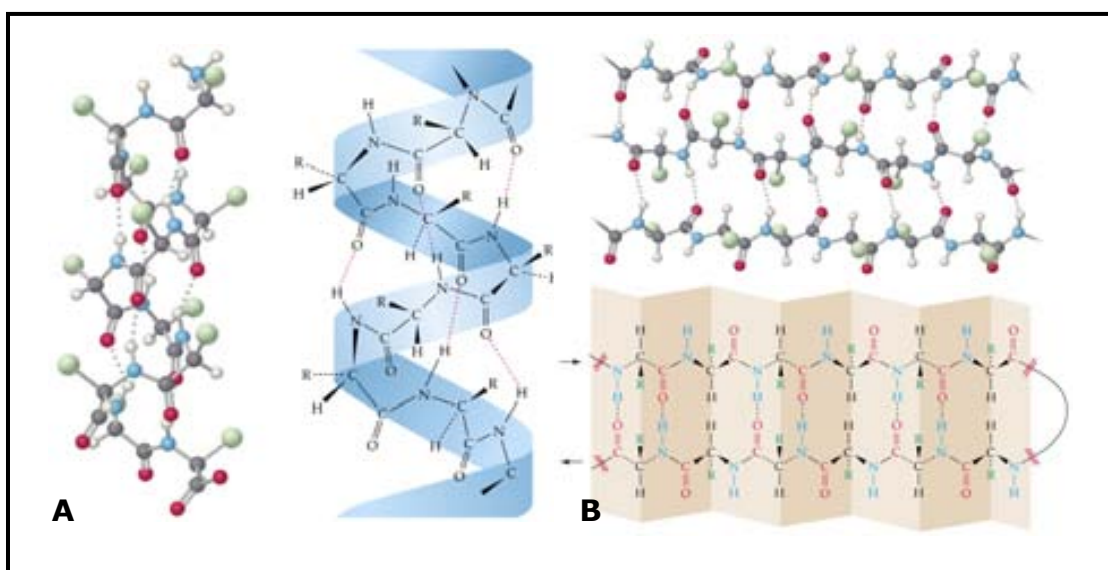


Figura I.5. Exemples d'estructures secundàries regulars. (A) Hèlix α i (B) fulla- β . Reproduït de McMurry i Fay (2004).

El procés de plegament és vital, ja que si no s'assoleix completament l'estructura nativa, la proteïna no pot exercir correctament la seva funció a l'àmbit biològic. Degut al seu caràcter fonamental, el plegament proteic ha estat una de les forces evolutives més importants, assolint-se uns nivells d'estabilitat adequats per al funcionament de la molècula, tot i que millorables per enginyeria proteica (Villegas *et al.*, 1996). Una proteïna massa estable veu dificultat el recanvi proteic, raó per la qual es creu que l'estabilitat no és màxima. S'ha de matisar, però, que l'estat natiu no és una estructura rígida, ja que es troba constituït per una multitud de subestats dinàmics, les diferències entre els quals són ínfimes, i que confereixen la flexibilitat necessària a la molècula per tal d'acomplir la seva funció biològica.

I.2.1.1 Teories inicials

A l'esquema clàssic de la reacció de plegament, la proteïna desplegada (la qual té conformació *random coil* o cabdell estadístic), ha de superar una barrera energètica important per assolir l'estat activat (estat de transició), que posteriorment patirà una pèrdua d'energia, donant lloc a l'estat natiu (Fig. I.2).

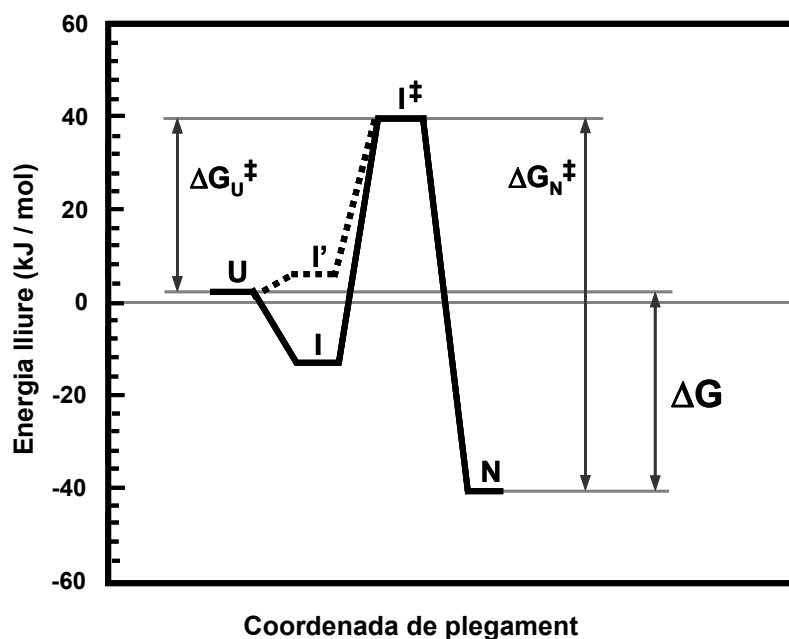


Figura I.6. Esquema de l'energètica de la reacció del plegament proteic. U: estat desplegat, I/I': intermediaris de plegament, I[‡]: estat de transició, N: estat natiu, ΔG : energia lliure de Gibbs, ΔG_U^\ddagger : energia lliure del pas desplegat al de l'estat de transició, ΔG_N^\ddagger : energia lliure del pas de l'estat de transició al natiu.

Christian Anfinsen va realitzar cap al final dels 50 un experiment clau que va iniciar l'estudi del plegament proteic. Anfinsen va aconseguir el replegament *in vitro* de la ribonucleasa A, demostrant que tota la informació necessària per a l'adquisició de l'estructura terciària nativa d'una proteïna es troba continguda en la seva seqüència aminoacídica (Anfinsen *et al.*, 1961). Aquests estudis van portar a l'assumpció de que l'estat natiu d'una proteïna es troba en el mínim d'energia lliure global del sistema en condicions fisiològiques, indicant una contribució termodinàmica al procés de plegament. Posteriorment, Levinthal va realitzar una objecció al fet que la contribució termodinàmica del plegament fos l'única, l'anomenada paradoxa de Levinthal, la qual esmena la impossibilitat de que el procés de plegament sigui a l'atzar (Levinthal, 1968; Dill i Chan, 1997). Levinthal demostrà numèricament que el temps que trigaria una proteïna en plegar-se mitjançant un procés a l'atzar seria quasi infinita; no obstant, les proteïnes pleguen normalment de manera molt ràpida, de l'ordre de segons o fins i tot

en temps inferiors. Es va proposar, llavors, que una proteïna desplegada hauria de seguir uns camins cinètics determinats (o 'vies'), de manera que només caldria assajar unes poques conformacions per tal d'assolir el plegament (Bryngelson *et al.*, 1995; Dinner *et al.*, 2000). Es creu que fins i tot, només caldria la formació de contactes nadius al voltant d'uns pocs residus clau (Fersht, 2000; Lindorff-Larsen *et al.*, 2005) i que potser l'estat natiu no residiria en un mínim energètic global, sinó en el lloc de mínima energia lliure d'entre els cinèticament possibles.

A partir d'aquí es van desenvolupar diferents models de plegament basats en la teoria de les "vies de plegament", en els quals distingim fonamentalment 2 grups: els seqüencials, en els quals es proposa primer la formació de l'estructura secundària enfront la terciària (per exemple, el model de *framework*), i els de col·lapse, en els quals l'adquisició d'ambdues estructures es produeix simultàniament (per exemple el model del col·lapse hidrofòbic). Semblava, emperò, que cap d'aquestes "vies de plegament" proposades era prou general per a explicar el procés global de plegament proteic, i no acabava de convèncer la seva poca flexibilitat. A partir d'aquests plantejaments va aparèixer la teoria dels anomenats "embuts de plegament".

I.2.1.2 Embuts de plegament

Els embuts de plegament són la nova visió del procés de plegament, els quals es refereixen als diferents estats de plegament com a 'conjunts de conformacions', enlloc de parlar de conformacions discretes, i tenen en compte múltiples rutes de plegament paral·leles (Dill i Chan, 1997). Es basen en unes superfícies tridimensionals, on la dimensió vertical representa el nombre de contactes nadius (entalpia) i les altres dimensions corresponen a l'entropia i el nombre de conformacions possibles (Fig. I.7). En algunes ocasions, l'embut també pot simplificar-se a una representació bidimensional.

La forma d'embut ve donada pel fet que se suposa que es parteix d'un conjunt molt ampli i heterogeni de conformacions desplegadas d'energia superior a l'estat natiu, el qual es va reduint a mida que disminueix tant l'energia com l'entropia durant el procés de plegament. L'embut culmina a l'estructura nativa (mínim energètic global), on el nombre de conformacions és extremadament limitat, pràcticament únic, tot i que dinàmic.

Aquest model sembla que s'adaptaria bé a plegaments característics de diferents proteïnes, ja que és força general, de manera que el que variaria serien només les formes dels embuts, pròpies de cada proteïna (de la seva seqüència) i de

cada condició. Es contempla, també, l'existència d'intermediaris, els quals, a diferència de l'estat natiu, són mínims energètics locals.

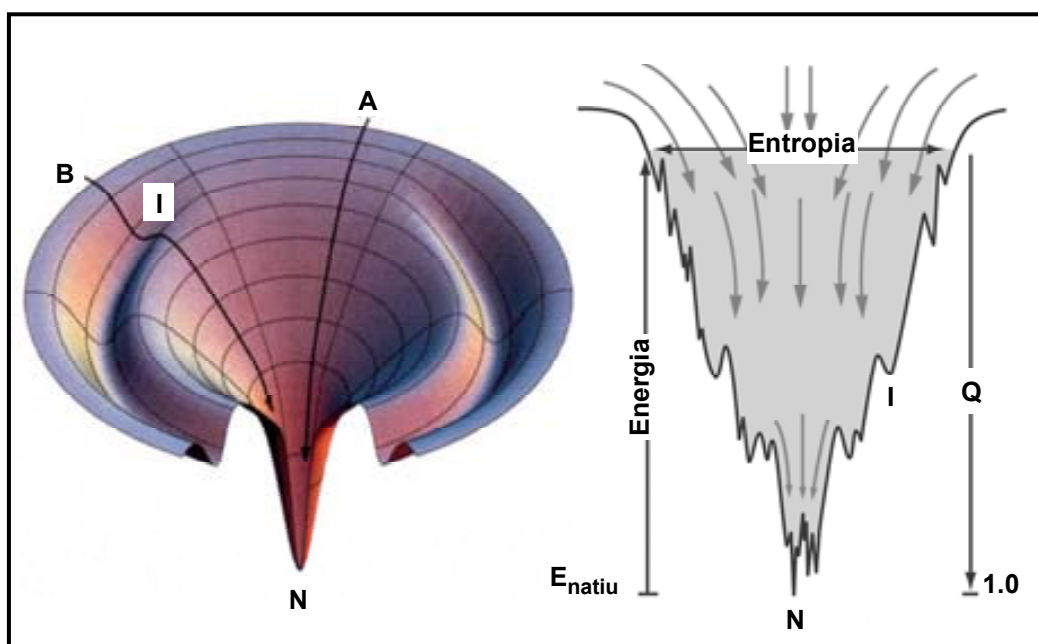


Figura I.7. Representació d'un embut de plegament tridimensional i bidimensional. N = estat natiu, I = intermediari, Q = nombre de contactes natius, E_{natiu} = energia de l'estat natiu. A i B = dos possibles camins de plegament diferents.

A partir d'ara, tot i que ens referim a estructures concretes, s'ha de tenir en compte que en realitat estem parlant dels conjunts de conformacions dinàmiques que les constitueixen.

I.2.1.3 Intermediaris

Les proteïnes de mida inferior a 100 aminoàcids solen presentar un plegament en 2 estats, és a dir, passen de la forma desplegada a la plegada, normalment de forma molt ràpida (Villegas *et al.*, 1995). Un dels primers exemples descrits fou el plegament de l'inhibidor 2 de quimotripsina (Jackson i Fersht, 1991). En canvi, les proteïnes que contenen més de 100 residus tendeixen a col·lapsar en ambients aquosos, formant unes estructures compactes amb una quantitat notable d'estructura nativa: els intermediaris. Com ja s'ha comentat, els intermediaris són mínims energètics locals, els quals es troben estructurats parcialment (Horwich, 2002). Aquests poden classificar-se de diferents maneres.

Segons la seva naturalesa els podem classificar en 2 grups:

- termodinàmics: són els que trobem en equilibri, de manera que s'acumulen i poden ésser aïllats. També denominats metaestables.
- cinètics: són detectables per canvis en la velocitat de replegament entre el principal estat de transició i l'estat natiu, però no s'acumulen i, per tant, no poden ésser aïllats.

Segons el seu paper en el procés de plegament, podem classificar-los igualment en 2 grups:

- on-pathway: intermediaris pels quals s'ha de passar forçosament per tal d'arribar a l'estat plegat (formats per interaccions natives, és a dir, formen part de la via de plegament).
- off-pathway: intermediaris que es troben fora de la via productiva de plegament (formats principalment per interaccions no natives, és a dir, no formen part de la via de plegament).

Aquestes conformacions parcialment plegades són de vital importància en el present treball, ja que a partir d'elles, en el cas de proteïnes globulars, es poden donar processos d'agregació.

1.2.1.4 Plegament *in vivo*

Totes les teories abans exposades s'han basat en experiments de replegament realitzats *in vitro*. Primerament, és extremadament difícil estudiar el procés de plegament *in vitro*, de manera que el que se segueix és el procés de replegament. El procediment habitual és tractar la proteïna amb agents i/o condicions desnaturalitzants que en promouen el desplegament i seguir posteriorment la recuperació de l'estructura nativa en eliminar l'agent i/o condició concreta.

El plegament *in vivo*, però, es creu que seria força diferent, tot i que encara és un tema força controvertit. El fet que la cadena polipeptídica es sintetitzi d'N- a C-terminal sembla que seria determinant, obligant la proteïna a començar a plegar-se tant bon punt aquesta emergeix del ribosoma per l'N-terminal (plegament co-traducciona), i dirigint el plegament de la resta de la cadena polipeptídica posteriorment (Basharov, 2003). A més, en algunes ocasions intervenen proteïnes que assisteixen el plegament proteic, les anomenades xaperones moleculars (Houry, 2001; Bukau *et al.*, 2006). La visió clàssica d'aquestes proteïnes és que s'uneixen a zones hidrofòbiques exposades de les proteïnes que no estan plegades correctament i

prevenen l'agregació d'aquestes (Ranson *et al.*, 1995). En el cas de les xaperones Hsp 60, aquestes són capaces de desplegar la proteïna utilitzant ATP i alliberar-la, donant-li una altra oportunitat de replegar-se correctament (Martin, 2000). La gran quantitat d'energia que gasta la cèl·lula en evitar l'agregació proteica és una mostra de la importància d'un plegament correcte.

Una visió més recent del plegament *de novo*, és la coordinació de les diferents xaperones/xaperonines cel·lulars que aniria des de les que interaccionen directament amb el ribosoma, com el factor desencadenant a bacteris (*trigger factor*), fins a les que interaccionen directament amb el proteasoma, com el foldosoma o sistema Hsp 90 a eucariotes (Ferbitz *et al.*, 2004; Kaiser *et al.*, 2006).

Les Xaperones/Xaperonines clàssiques no poden ser considerades com a catalitzadors, ja que no acceleren la velocitat de la reacció de plegament. En canvi, altres proteïnes intervenen en el procés de plegament actuant com a veritables catalitzadors: les prolil-isomerases, que catalitzen la isomerització cis-trans de les prolines (Gothel i Marahiel, 1999), i les disulfur isomerases (Gruber *et al.*, 2006), que catalitzen la correcta formació dels ponts disulfur intra i intercatenaris. Algunes de les Xaperones/Xaperonines clàssiques mostren addicionalment aquestes activitats enzimàtiques, com el mateix factor desencadenant, que presenta activitat prolil-isomerasa.

1.2.2 Fibres amiloides

La cèl·lula, doncs, té molts recursos per tal de dirigir la proteïna cap a la seva forma tridimensional correcta, a més de sistemes que assegurin la degradació si el plegament no s'ha assolit amb èxit. No obstant, en algunes ocasions es produeixen plegaments erronis de determinades proteïnes originant-se patologies conegudes com malalties conformacionals (Taula 1.2). Aquests plegaments incorrectes poden acabar donant oligòmers solubles, agregats inespecífics o uns agregats específics recentment molt ben caracteritzats, les fibres amiloides. En aquest darrer cas, parlem d'amiloïdosis.

Taula I.2. Exemples de malalties conformacionals. RE: reticle endoplàsmic. Reproduït parcialment de Dobson (2001).

| Malaltia | Proteïna involucrada | Lloc de plegament |
|---|---|-------------------|
| Fibrosi cística | Regulador transmembrana de la fibrosi cística | RE |
| Malaltia de Huntington | Huntingtina | Citosol |
| Anèmia falciforme | Hemoglobina | Citosol |
| Alzheimer | β -amiloide / presenilina | RE |
| Malaltia de Parkinson | α -sinucleïna | Citosol |
| Scrapie (ovelles) / malaltia de Creutzfeldt-Jacob i kuru (humans) | Proteïna priònica (Prp) | RE |
| Amiloïdosis familiars | Transtirretina / lisozim | RE |
| Retinosi pigmentària | Rodopsina | RE |
| Càncer | p53 | Citosol |

I.2.2.1 Fibres amiloides i amiloïdosis

Les fibres amiloides són uns agregats proteics ordenats i molt inerts, els quals mostren una sèrie de característiques comuns. No obstant, diferents tipus de seqüències polipeptídiques són capaces de formar-ne. Aquestes fibres es troben en forma de dipòsits extracel·lulars a teixits afectats per malalties amiloidòtiques, com poden ser l'Alzheimer, les encefalopaties espongiformes o les amiloïdosis sistèmiques (Taula I.2). En alguns casos s'han observat dipòsits d'aquestes característiques intracel·lularment, com a la malaltia de Parkinson, encara que patològicament no es consideren dins del grup de les amiloïdosis. Una d'aquestes proteïnes amiloidòtiques, la proteïna del prió (PrP), exhibeix propietats infectives, és a dir, és capaç d'induir la malaltia a un altre individu sense la presència de cap estructura d'àcid nucleic associada (Prusiner, 1982). Recentment el nombre de proteïnes conegudes amb capacitat amiloïdogènica ha anat augmentant, sobretot amb el descobriment de proteïnes no relacionades aparentment amb cap patologia o fenotip diferencial, però amb característiques típiques de les fibres amiloides. En aquests casos, les fibres es formen un cop les proteïnes són exposades a unes condicions determinades no fisiològiques (Taula I.3). Per a la denominació d'aquest tipus de fibres, la Societat Internacional de l'Amiloïdosi recomana l'ús del terme *amyloid-like* (Westermarck *et al.*, 2005), tot i que en aquest treball s'emprarà indistintament el terme amiloide amb la

finalitat de facilitar la lectura. Algunes d'aquestes proteïnes no patogèniques, com Sup35p i HET-s, són convertides en fibres amiloides pels seus propis organismes (*Saccharomyces cerevisiae* i *Podospira anserina*, respectivament), produint-ne l'adquisició d'un determinat fenotip. Per tant, les conformacions fibrilars de Sup35p i HET-s tenen propietats funcionals, actuant com a elements genètics extracromosomals, ja que són capaces d'autopropagar-se i transmetre's (Chien *et al.*, 2004).

El creixent nombre de proteïnes sense cap relació seqüencial ni estructural entre elles, però amb la capacitat de formar fibres tan semblants, va portar a pensar que la formació de fibres amiloides podria ésser una propietat inherent de les cadenes polipeptídiques (Guijarro *et al.*, 1998).

Taula I.3. Proteïnes amiloïdogèniques no patogèniques. Exemples de proteïnes formadores de fibres *amyloid-like* no relacionades amb malalties conegudes. Reproduït parcialment de Dobson (2001).

| Proteïna | Tipus d'estructura nativa |
|------------------------------|---------------------------|
| domini SH3 de la PI3 quinasa | β |
| acilfosfatasa | α/β |
| ADA2h | α/β |
| α -lactalbúmina | $\alpha+\beta$ |
| lisozim | α/β |
| citocrom c_{552} | α |
| apo-mioglobina | α |

I.2.2.2 Característiques de les fibres amiloides

El creixent nombre de proteïnes que adoptaven característiques amiloidòtiques va demostrar la necessitat d'una sèrie de criteris per tal de considerar un agregat proteic com a tal. Primerament la morfologia típica de les fibres amiloides ha de poder observar-se al microscopi electrònic de transmissió (TEM); les fibres comparteixen un diàmetre d'entre 7 i 13 nm i es troben generalment formades per un conjunt d'uns 2 a 6 protofilaments d'entre 2 i 5 nm de diàmetre entortolligats entre ells (Serpell *et al.*, 2000) (Fig. I.8).

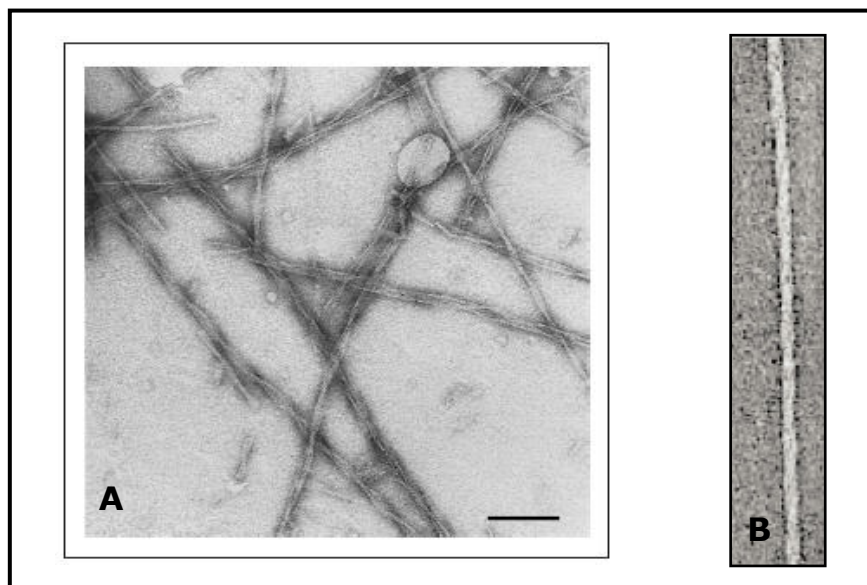


Figura 1.8. Morfologia de les fibres amiloides. (A) Observació de fibres amiloides al microscopi electrònic de transmissió per tinció amb acetat d'uranil i (B) mitjançant criomicroscòpia electrònica. Reproduït de Jiménez *et al.* (2002).

Una altra característica ha de ser la capacitat d'unió de fluoròfors específics, el Congo Red i la Tioflavina-T (LeVine, 1993), produint un *red shift* a l'espectre d'absorció del primer (provocant un desplaçament de l'espectre d'absorció cap al color vermell, és a dir, cap a major longitud d'ona), i augmentant el màxim de fluorescència en el cas del segon. Les fibres, a més, han de presentar un patró típic d'estructura β -creuada (*cross- β*) per difracció de raigs X, indicatiu d'una disposició de les fulles- β ordenades en direcció perpendicular a l'eix de la fibra (Sunde i Blake, 1997). L'elevada quantitat d'estructura secundària- β d'aquests agregats, obre les portes a estudis que utilitzin tècniques capaces de detectar-la, com poden ser el dicroisme circular o l'espectroscòpia d'infraroig. El caràcter inert de les fibres amiloides les fa resistents a la proteòlisi, de manera que estudis comparatius de resistència a proteases poc específiques entre la proteïna nativa i la forma agregada també són útils tant per a la confirmació de que es tracta d'un amiloide com per a la identificació de les parts involucrades en la formació de la fibra.

I.2.2.3 Formació de fibres amiloides

Les fibres amiloides representen un estat d'organització d'estructura final diferent a l'estat natiu. El procés de plegament és independent per cada molècula proteica i ve definit per l'adquisició d'una sèrie de contactes intramoleculars. No obstant, el procés d'agregació es dona per l'adquisició d'una sèrie de contactes intermoleculars entre diverses molècules de la mateixa proteïna, essent, per tant, depenent de concentració. En algun moment, i per causes que en molts casos es desconeixen, la proteïna es desviaria de la seva via de plegament normal, originant un estat final també estructurat i estable. Aquest fet suggereix que el plegament i l'agregació són processos que es troben en competència, i que l'entorn dirigiria l'estructura peptídica cap a un o altre camí. Dins el context dels embuts de plegament, podria considerar-se un embut doble, de manera que la proteïna 'escaparia' de la superfície energètica del plegament, i entraria en una altra superfície energètica on la fibra amiloide seria el component de menor energia (Fig. I.9) (Jahn i Radford, 2005).

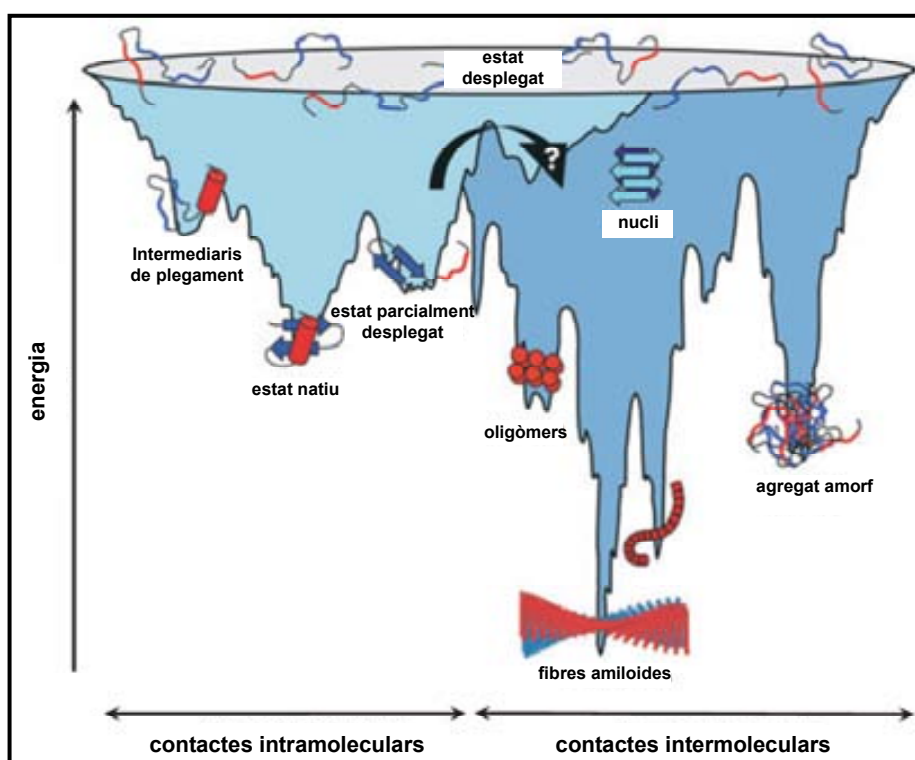


Figura I.9. Paisatge energètic del plegament proteic i l'agregació. L'embut esquerre representa el procés de plegament i l'embut de la dreta el de l'agregació. Poden observar-se els estats intermediaris de l'embut de plegament (intermediaris *on-pathway*) i els de l'agregació (intermediaris *off-pathway*). La fletxa negra representa l'intermediari desconegut per on la proteïna passaria d'un embut a l'altre. Reproduït de Jahn i Radford (2005).

La competència entre plegament i agregació queda reflectida també dins de la mateixa molècula, tal i com Chiti i col·laboradors van observar per a la proteïna acilfosfatasa (AcP). Aquesta proteïna no és patogènica però és capaç de formar fibres amiloides en presència de trifluoroetanol (TFE), i es va demostrar que les zones implicades en la formació de l'estat de transició del plegament eren diferents de les zones implicades en el procés d'agregació (Chiti *et al.*, 2002).

El procés de formació dels agregats fibril·lars es podria definir en 2 fases diferents:

- Fase de nucleació: els intermediaris amiloïdogenics s'agrupen donant nuclis d'agregació. Sol apreciar-se com una fase de latència mitjançant determinades tècniques, com la fluorescència amb Tioflavina-T o la microscòpia electrònica. És una etapa molt lenta respecte la totalitat del procés, ja que es troba desafavorida termodinàmicament. No s'observen agregats fibril·lars al microscopi.
- Fase de creixement: un cop el nucli ha assolit una grandària crítica, comença una fase ràpida on es dona la incorporació de molècules al nucli pre-format, que acabaran donant la fibra. Poden visualitzar-se agregats fibril·lars al microscopi.

Una de les grans qüestions respecte a les fibres amiloides és la identitat de les espècies citotòxiques. Estudis recents semblen indicar que els agregats pre-fibril·lars (fins i tot de proteïnes no patògenes) podrien ser els responsables de la major part d'efectes deleteris, i que els estats fibril·lars serien formes de segrest dels anteriors, desenvolupant, per tant, un paper protector (Bucciantini *et al.*, 2002). No obstant, en alguns casos, la simptomatologia de la malaltia amiloidòtica ve clarament donada per la gran acumulació de material fibril·lar, com per exemple, en l'amiloïdosi associada a diàlisi (Floege i Ketteler, 2001).

1.2.2.4 Causa de la formació de fibres amiloides *in vivo*

Segons el que s'ha vist, només un nombre reduït de proteïnes forma fibres amiloides *in vivo*, i normalment només es dona en casos d'edat avançada del pacient. La causa de tot això és que a la cèl·lula hi ha molts mecanismes que afavoreixen el correcte plegament de la proteïna i limiten l'agregació d'aquesta. Exemples d'aquests mecanismes són les xaperones moleculars i el sistema de degradació proteic del proteasoma.

La formació de fibres amiloides *in vivo* és la conseqüència d'errades en algun/s d'aquest/s sistema/es, les causes de les quals poden ser:

- Incapacitat per part de les xaperones de reconèixer la proteïna desplegada.
- Desplegament parcial de la proteïna en compartiments on les xaperones hi són absents.
- Actuació d'algunes xaperones (com Hsp70 i Hsp40) atenuant la formació d'agregats pre-fibril·lars i afavorint la formació d'estats fibril·lars (Wacker *et al.*, 2004).
- Disfunció del sistema del proteasoma cel·lular.

No obstant, com ja s'ha comentat, en alguns casos la formació de fibres amiloides no desencadena un procés patològic, sinó que obeeix a un procés fisiològic. Aquest fet ha estat observat en varis organismes i fins i tot a mamífers (Fowler *et al.*, 2006; Fowler *et al.*, 2007)

1.2.2.5 Estructura de les fibres amiloides

Les fibres amiloides posseeixen com a característica comú un nucli en forma de columna amb un patró de difracció de raigs X anomenat *cross- β* (veure apartat 1.2.2.2) (Sunde i Blake, 1997). Darrerament han emergit una elevada quantitat d'estudis i models atòmics sobre la manera mitjançant la qual les cadenes polipeptídiques s'arrangen donant lloc a l'estructura del protofilament. Actualment, cap dels models proposats és capaç d'explicar totes les propietats de les fibres descrites fins ara, encara que és possible que no hi hagi un únic model vàlid (Nelson i Eisenberg, 2006). Els diferents models es troben íntimament lligats a models de formació de fibres amiloides. En termes generals, es poden distingir 3 tipus de models atòmics, resumits a la Fig. 1.10:

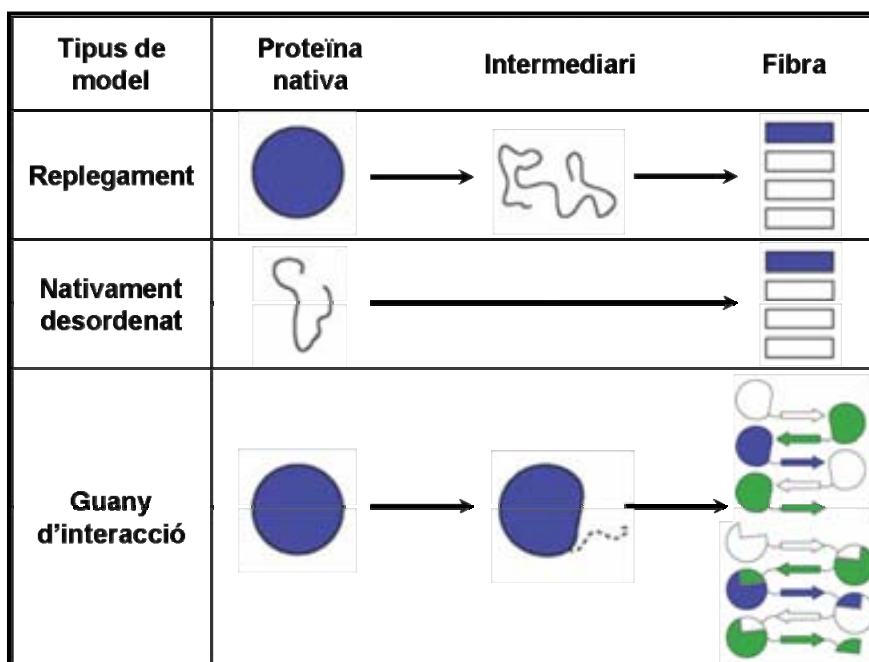


Figura I.10. Models d'estructura amiloide. Diferents models proposats per a l'adquisició d'estructura de la columna *cross-β* amiloide. Adaptat de Nelson i Eisenberg (2006).

Els **models de replegament** defensen que la proteïna ha de patir un desplegament total o parcial de la seva conformació nativa seguit d'un replegament que la portarà, finalment, a la forma fibril·lar. En el cas de proteïnes formadores d'amiloides *in vitro*, la formació d'aquesta forma intermediària amb cert grau de desplegament vindria assistida per condicions parcialment desnaturalitzants. En aquest tipus de condicions, el procés d'agregació es veuria afavorit en la seva competició vers el plegament natiu. Aquests models proposen que el fet que l'estat fibril·lar tingui moltes característiques semblants per a proteïnes molt diferents implica que aquest estigui definit per interaccions entre l'esquelet polipeptídic, el qual és comú per a totes les proteïnes. L'efecte de les cadenes laterals específiques, és a dir, de la seqüència aminoacídica, seria tan sols modulador de la velocitat de formació i l'estabilitat de l'estat fibril·lar. A les estructures fibril·lars freqüentment proposades per a aquests models, les cadenes- β dels monòmers es trobarien interaccionant entre elles de manera paral·lela per ponts d'hidrogen amb un cert desfasament, donant lloc a hèlixs- β levògires. Disposicions antiparal·leles, especialment per a seqüències curtes, també serien possibles. Exemples de proteïnes per les quals s'ha proposat aquest model són la insulina (Jiménez *et al.*, 2002) i la proteïna priònica (Govaerts *et al.*, 2004).

Els **models nativament desordenats** són aplicats a proteïnes desestructurades en el seu context fisiològic. En aquest cas, la totalitat o una part de la proteïna ha de guanyar estructura per tal de formar la columna *cross-β*. Les estructures proposades

inclouen un nucli de cadenes- β paral·leles en forma de ziga-zaga formant una estructura anomenada 'estructura- β paral·lela súper prisada' (*parallel superpleated β structure*), com es proposa per a la proteïna Ure2p (Kajava *et al.*, 2004).

Els **models de guany d'interacció** es basen en un canvi conformacional d'una regió de la proteïna nativa que deixa exposada una superfície que abans era inaccessible. Aquesta superfície interaccionarà de manera específica amb altres molècules alterades donant lloc a la fibra, i la resta de la proteïna mantindrà la seva forma nativa. Aquest tipus de models es pot subdividir alhora en 4 classes, 2 de les quals comprenen apilaments que no haurien de comportar forçosament la formació d'un nucli *cross- β* , i per tant no seran comentats. Les altres dues classes estan basades en columnes *cross- β* amb cremalleres estèriques i disposicions paral·leles en registre (*parallel in-register arrangement*, PIRA, Fig. I.11).

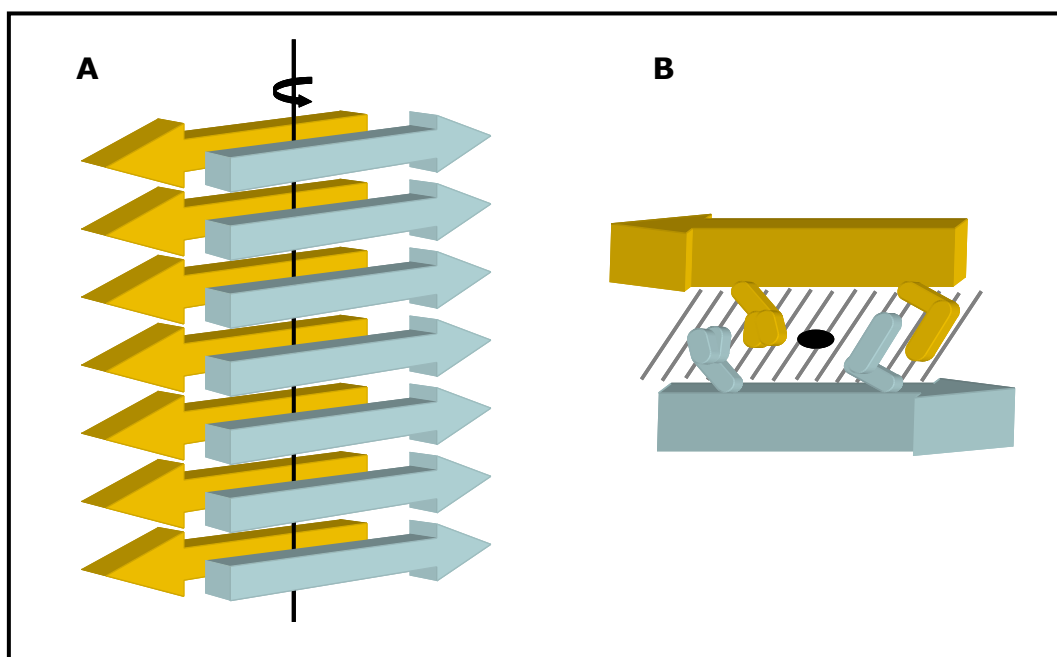


Figura I.11. Columnes *cross- β* PIRA. (A) Disposició de les fulles- β . Els fragments peptídics es troben en una conformació estesa, formant una cadena- β perpendicular a l'eix de la fibra i realitzant ponts d'hidrogen amb els pèptids immediatament superior i inferior. Aquesta disposició dona lloc a fulles- β (en aquesta figura paral·leles, tot i que s'han observat també fulles- β antiparal·leles). (B, vista superior d'A) 'Cremallera estèrica'. Aquestes fulles- β interaccionen entre si mitjançant la interfície seca (zona ombrejada), formada per interdigitació de cadenes laterals polars. El conjunt d'interfícies seques al llarg de la fibra formen l'anomenada 'cremallera estèrica'.

A la primera classe, amb el pèptid GNNQQNY del prió de llevat Sup35p com a proteïna model, la columna *cross- β* està formada a partir de les superfícies ara exposades de la proteïna completa (Nelson *et al.*, 2005). Les cadenes- β d'aquestes superfícies s'apilen de forma paral·lela formant ponts d'hidrogen i s'encaren a una fulla- β idèntica. Aquestes fulles- β encarades formen una interfície seca, ja que es dona

una interdigitació de les cadenes laterals de manera altament complementària, que exclou l'aigua. Degut a les seves característiques, el conjunt d'interfícies seques de la columna *cross-β* ha estat anomenat cremallera estèrica. Aquesta estructura va ésser obtinguda a partir de la resolució atòmica de microcristalls amb característiques molt semblants a les de les fibres amiloides. Posteriorment, la resolució atòmica d'altres pèptids va revelar també la presència de cremalleres estèriques, bé que diferents, ja que l'orientació de les molècules podia variar segons el pèptid. Aquesta variabilitat va donar lloc a una classificació de cremalleres estèriques (Sawaya *et al.*, 2007).

A la segona classe, una superfície exposada entre dos dominis produeix l'intercanvi dels dominis següents amb una altra molècula (*domain swapping*) (veure Fig. I.10). La columna *cross-β* serà formada per la zona entre dominis donant una estructura igual a la del pèptid del prió del llevat. Un exemple d'aquesta estructura es troba en una fibra amiloide de ribonucleasa A dissenyada racionalment (Sambashivan *et al.*, 2005).

No obstant, aquests models es limiten a descriure l'estructura de la fibra basant-se en l'estructura del monòmer dins l'agregat fibril·lar, però no en caracteritzen el mecanisme d'elongació. Aquest mecanisme està essent objecte d'estudi recentment, proposant-se un model força general per a l'elongació de les fibres de β_2 -microglobulina a pH àcid, el qual està basat en l'addició seqüencial de monòmers per tal de formar un nucli estructural de la mida d'un hexàmer (Xue *et al.*, 2008).

I.2.2.6 Mètodes de predicció de l'amiloidogènesi

Independentment del tipus de model de formació de la columna fibril·lar, la majoria d'autors semblen recolzar la idea que la capacitat amiloïdogènica no es troba al llarg de tota la cadena polipeptídica, sinó localitzada en zones determinades i, generalment, força curtes (López De La Paz *et al.*, 2002; Ventura *et al.*, 2004; Esteras-Chopo *et al.*, 2005). Igualment, els diferents models abans exposats, per bé que diferentment, donen importància a la seqüència aminoacídica, atribuint-li, almenys, la capacitat de controlar la velocitat d'agregació. Degut a tot això, últimament s'estan duent a terme molts esforços de caire teòric per a la identificació de les zones proteiques més rellevants en el procés d'agregació, així com per a la predicció de les velocitats d'aquest mateix procés.

Fonamentalment, podem distingir 3 tipus de mètodes per a la predicció de les regions amiloïdogèniques i/o les velocitats d'agregació:

Els **mètodes empírics o fenomenològics** es basen en els resultats obtinguts per unes quantes proteïnes i els seus mutants per tal d'englobar les propietats de les

seves seqüències, juntament amb les condicions experimentals, en equacions o algorismes. Les equacions elaborades podran ser utilitzades per inferir la tendència a agregar d'altres proteïnes problema (Pastor *et al.*, 2005). Chiti i col·laboradors han desenvolupat una fórmula que permet la predicció de la tendència d'agregació de cadenes polipeptídiques desestructurades basada en 3 factors simples i generals: hidrofobicitat, tendència a patir una transició conformacional d' α a β , i càrrega neta de tota la cadena polipeptídica (Chiti *et al.*, 2003). L'empresa Zyentia Ltd. ha replantejat aquesta fórmula, elaborant una escala de tendències d'agregació (la qual engloba els diferents paràmetres abans esmentats) per a cada un dels 20 amino àcids proteïnogènics (Pawar *et al.*, 2005; Tartaglia i Vendruscolo, 2008). El programa, Zyggregator™ (anteriorment a www.zyggregator.com, i actualment a <http://www-vendruscolo.ch.cam.ac.uk/zyggregator.php>), utilitza una finestra d'uns 7 residus per escombrar la seqüència i detectar les regions amb una tendència a agregar superior a la mitja de la cadena polipeptídica; a més a més, Zyggregator™ permet la predicció de les velocitats absolutes d'aquest procés. Un estudi semblant és la base del programa AGGREGSCAN (<http://bioinf.uab.es/aggregscan/>, Conchillo-Solé *et al.*, 2007), el qual atorga un valor de tendència d'agregació a cada aminoàcid basant-se en els resultats obtinguts per la substitució de la fenilalanina 19 del pèptid A β ₁₋₄₂ de l'Alzheimer pels 20 aminoàcids proteïnogènics (Sánchez de Groot *et al.*, 2005; de Groot *et al.*, 2006). Ambdós mètodes han obtingut bones correlacions per a la identificació de regions amiloidogèniques sobretot de pèptids i proteïnes nativament desestructurades, tot i que Zyggregator™ ha estat recentment millorat amb la incorporació de la predicció de la propensió de plegament, per tal de predir més acuradament les propensions d'agregació a proteïnes estructurades (Tartaglia *et al.*, 2008).

TANGO és un algorisme que permet identificar les regions tendents a agregar de proteïnes, a més de predir els canvis en la tendència d'agregació produïdes per mutacions a la seqüència polipeptídica (Fernández-Escamilla *et al.*, 2004). Aquest mètode assumeix que l'única estructura secundària d'un agregat proteic és una cadena- β i que les regions involucrades en l'agregació es troben totalment enterrades. L'algorisme llavors calcula per cada residu el percentatge d'ocupació entre els principals estats conformacionals en competència: l'hèlix- α , el gir- β , la fulla- β , l'estat natiu si no es tracta d'un pèptid/proteïna nativament desestructurat, i l'agregat- β . El càlcul de cada residu el realitza en base a diferents termes energètics: hidrofobicitat, energia de solvatació, interaccions electrostàtiques i ponts d'hidrogen; així com els efectes dels paràmetres de la solució (factors extrínsecs): pH, força iònica... Les regions que consisteixen en com a mínim 5 residus consecutius predits amb un mínim

d'un 5% d'estructura agregat- β cadascun són presentades com a zones propenses a agregar.

Els **mètodes basats únicament en identificació de seqüència** parteixen d'estudis combinatorials encarats a la identificació de seqüències inherentment amiloidogèniques (López de la Paz i Serrano, 2004). La seqüència trobada, STVIIIE, és capaç d'induir la formació de fibres a proteïnes que no havien mostrat la presència de fibres amiloides en les mateixes condicions (Esteras-Chopo *et al.*, 2005). A partir d'aquesta seqüència s'ha elaborat un programa de predicció basat en el grau de semblança d'una seqüència polipeptídica al patró STVIIIE (López de la Paz i Serrano, 2004).

Per últim, els **mètodes basats en la identificació d'estructura**, parteixen de l'estructura del nucli de les fibres amiloides de l'hexapèptid NNQQNY del prió de llevat Sup35p (apartat anterior; Nelson *et al.*, 2005). Aquest mètode es basa en modelar les estructures de les seqüències d'interès formades per 6 residus sobre l'estructura del parell de fulles- β adjacents observada per l'hexapèptid de Sup35p. Posteriorment es calculen les energies relacionades i es seleccionen només els hexapèptids que presenten una energia inferior a un llindar determinat (Thompson *et al.*, 2006). Un altre mètode estructural es basa en el desacoblament de l'estructura secundària presentada i la predita en algunes proteïnes, ja que s'ha observat que algunes proteïnes amiloidogèniques contenen una hèlix- α en una regió que hauria d'ésser una cadena- β segons les prediccions (Kallberg *et al.*, 2001).

El nombre de mètodes predictius per a l'amiloidogènesi creix a una velocitat sorprenent, però la majoria de mètodes són ineficaços a l'hora de determinar les tendències d'agregació de proteïnes globulars, degut a la presència d'estructura compacta. Per tal de realitzar comparacions es pot intentar eliminar la presència d'estructura utilitzant agents desnaturalitzants; no obstant, nombrosos estudis semblen indicar la presència de certa quantitat d'estructura *native-like* fins i tot en condicions desnaturalitzants força agressives (Neri *et al.*, 1992; Schwalbe *et al.*, 1997; García *et al.*, 2001).

I.2.3 El domini d'activació de la procarboxipeptidasa A2 humana com a model d'estudi

L'ús de proteïnes model altament caracteritzades ha estat molt important en l'estudi de les propietats amiloidogèniques, sobretot amb la interrelació entre els processos de plegament i agregació. En molts casos també s'utilitzen pèptids que comprenen certes parts d'una seqüència proteica com a sistemes model, ja que simplifiquen els estudis enfront la totalitat de la proteïna, tot i permetent una anàlisi detallada dels paràmetres biofísics implicats en l'agregació (Pastor *et al.*, 2005). El domini d'activació de la carboxipeptidasa A2 humana (ADA2h) ha estat una de les primeres proteïnes no patògenes en descriure's la seva capacitat de formar fibres amiloides a pH baix (Villegas *et al.*, 2000) i l'objecte d'estudi de gran part d'aquesta tesi. Experiments anteriors van permetre la caracterització profunda de la via de plegament de la proteïna a pH neutre, així com la seva millora mitjançant enginyeria de proteïnes. Els resultats d'aquests estudis, de gran importància per al desenvolupament del present treball, són explicats en detall en els següents apartats.

I.2.3.1 Estudis de caracterització del plegament

ADA2h és una proteïna de 81 residus amb un plegament tipus β -sandwich (α/β), on trobem 2 hèlix- α empaquetades contra una fulla- β de 4 cadenes (Fig. I.4 i I.12). Es disposa de la seva estructura tridimensional cristal·litzada en forma de proenzim (García-Sáez *et al.*, 1997), així com en dissolució (per RMN, Jiménez *et al.*, 2003). Entre ambdues estructures, tal i com s'ha exposat anteriorment, la major diferència és la desestructuració parcial de la segona cadena- β en absència del domini actiu de l'enzim.

D'aquesta proteïna es va caracteritzar la cinètica de plegament mitjançant diferents tècniques, veient que a pH neutre seguia un model de plegament en dos estats, és a dir, sense la presència d'intermediaris termodinàmics ni cinètics (Villegas *et al.*, 1995). Sorprenentment, la velocitat de plegament exhibida per ADA2h era la més alta observada per a cap proteïna fins aquell moment, requerint-se l'ús d'aparells de cinètica ràpida com l'*stopped-flow*. Posteriorment, es van realitzar treballs d'estabilització de les 2 hèlixs- α (Villegas *et al.*, 1996), dissenyant-se dos mutants diferents mitjançant el programa de predicció de la tendència a formar hèlix- α AGADIR (Muñoz i Serrano, 1997) (Fig. I.12):

- M1: mutant de l'hèlix 1 amb els canvis N25K, Q28E, Q32K i E33K
- M2: mutant de l'hèlix 2 amb els canvis Q60E, V64A, S68A i Q69H

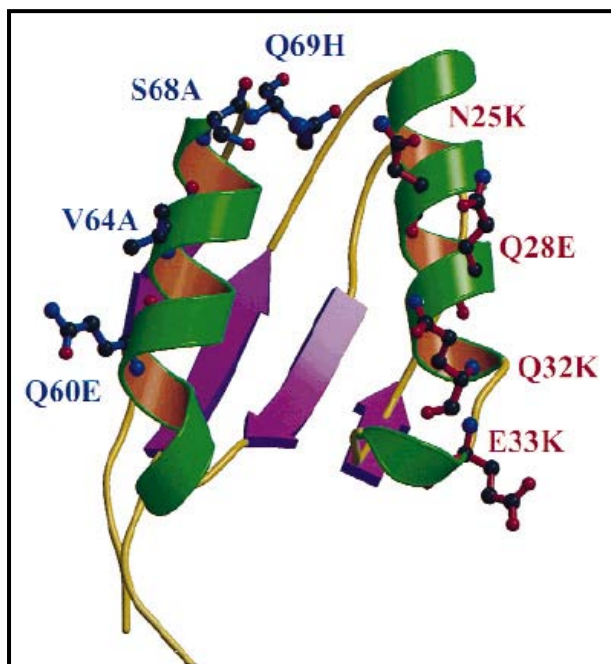


Figura I.12. Estabilització de les hèlixs- α d'ADA2h. Mutacions realitzades a ADA2h per tal d'estabilitzar l'hèlix α -1 (en vermell) i l'hèlix α -2 (en blau). Reproduït de Villegas *et al.* (2000).

Es va estudiar l'estabilitat d'aquests 2 mutants respecte la de la variant salvatge (WT) utilitzant les mateixes tècniques i es va comprovar que s'havia aconseguit un augment de l'estabilitat en ambdós casos, essent més important en el cas d'M2 (Taula I.4). Això suggeriria que l'estabilitat proteica no és òptima en la proteïna salvatge, la qual cosa té sentit si pensem que s'ha de produir la desestructuració de l'estat natiu per tal de permetre el recanvi proteic a l'àmbit cel·lular (Villegas *et al.*, 1996). També es va voler comprovar si es podia accelerar la velocitat natural de plegament, utilitzant els 2 mutants de les hèlixs (M1 i M2) i un doble mutant que contenia les 2 hèlixs- α estabilitzades (DM) (Taula I.4). Els resultats van mostrar que la velocitat de plegament tampoc havia estat optimitzada durant el procés evolutiu (Viguera *et al.*, 1997), obtenint-se velocitats encara més extremes que per la varietat salvatge.

Taula I.4. Valors termodinàmics i cinètics d'ADA2h i els mutants amb les hèlixs- α estabilitzades. *La T_m del DM no va poder ser estimada degut a que el procés de desplegament tot just havia començat a 363K (Viguera *et al.*, 1997)

| Proteïna | ΔG (Kcal mol ⁻¹) | T_m (K) | k_u (s ⁻¹) | k_f (s ⁻¹) |
|----------------------------------|---|--------------|-----------------------------|-----------------------------|
| Salvatge (<i>wild-type</i>) | 4,5 | 351,0 ± 0,2 | 0,350 ± 0,010 | 743 ± 79 |
| M1 | 5,8 | 358,2 ± 0,5 | 0,054 ± 0,002 | 1047 ± 94 |
| M2 | 6,0 | 358,5 ± 0,5 | 0,130 ± 0,003 | 2993 ± 287 |
| DM | 7,5 | * | 0,014 ± 0,001 | 4390 ± 0,02 |

Com a resultat d'aquest estudi comparatiu també es va poder comprovar que els paràmetres termodinàmics tenen caràcter additiu, ja que els valors del DM eren resultat de la suma de cada un dels mutants parcials. Aquests mutants han estat estudiats amb d'altres aproximacions com el DSC (Fernández *et al.*, 2000) i l'espectrometria de masses (Villanueva *et al.*, 2000; Villanueva *et al.*, 2002; Villanueva *et al.*, 2003).

I.2.3.2 Estudi de l'estat de transició del plegament d'ADA2h

La caracterització de l'estat de transició del plegament d'ADA2h va aportar informació sobre quines interaccions havien estat formades amb anterioritat a aquest i quines es consolidarien posteriorment. Amb aquesta finalitat es van dissenyar una sèrie de mutants puntuals d'ADA2h, on a cada un es trencava una interacció nativa determinada i s'observava l'efecte d'aquesta desestabilització sobre les energies entre els diferents estats de la via de plegament, basant-se en el mètode dels valors ϕ descrit per Alan Fersht (Fersht, 1995; Fersht i Sato, 2004). Les relacions d'energies generen valors de ϕ entre 0 i 1, els quals representen els valor extrems d'interaccions no formades i completament establertes a l'estat de transició, respectivament (Villegas *et al.*, 1998). Un valor proper a 1 indica que la mancança d'una interacció concreta en un mutant genera una gran diferència respecte a la forma salvatge a nivell de l'estat de transició, de manera que la interacció nativa hauria d'haver estat formada abans d'aquest estat activat. Valors propers a 0 indiquen el contrari, i valors intermitjos són indicatius de formacions parcials d'interaccions (sempre que s'obtingui una bona correlació de Brønsted, Villegas *et al.*, 1998). Amb aquesta aproximació es va aconseguir fer un mapa indicatiu del grau d'estructuració de l'estat de transició, demostrant que el nucli de plegament està format per les cadenes β -1 i β -3 juntament

amb l'hèlix α -2; trobant-se, per contra, l'hèlix α -1 parcialment desestructurada i les altres fulles- β (2 i 4) totalment desplegadas (Villegas *et al.*, 1998). El fet que l'hèlix 2 va ser la que, mitjançant estabilització, va permetre un clar augment de la velocitat de replegament (Viguera *et al.*, 1997) suggereix que per augmentar la velocitat de plegament s'han d'estabilitzar estructures que ja estan formades a l'estat de transició (Villegas *et al.*, 1998).

I.2.3.3 Formació de fibres amiloides *in vitro* d'ADA2h

L'estudi del plegament d'ADA2h a pH 3 va revelar la capacitat de formació de fibres amiloides d'aquesta proteïna exposada a certes condicions, observant-se la presència d'aquests agregats per microscòpia electrònica (Villegas *et al.*, 2000). Partint d'aquesta informació, es van dur a terme una sèrie d'experiments per tal de comprovar que s'acomplien les característiques típiques de les fibres amiloides esmentades en el punt I.2.2.2, així com estudis sobre les transicions estructurals en variar el pH i la temperatura.

Seguint la desnaturalització tèrmica a pH 3 per dicroisme circular, es va poder determinar en el cas del WT una transició d'estructura, observant-se com la senyal a 215 nm (mínim de l'espectre de CD per a fulla- β) augmentava en funció de la temperatura i encara més amb el refredament posterior, suggerint un canvi conformacional a la proteïna (Fig. I.13). No obstant, el més interessant va ser el fet que aquesta transició no s'observava en el mutant amb ambdues hèlixs estabilitzades, el doble mutant (i igualment en M1 i M2). Aquest mutant, un cop acabat el procés de desnaturalització-renaturalització tèrmica, presenta un espectre de CD α + β , igual que el WT (i aquest mateix mutant) a pH 7. Per microscòpia electrònica tampoc s'advertia la presència de cap agregat fibril·lar per a la variant DM. Això va portar a hipotetitzar una possible relació entre l'estabilitat i la tendència de formació de fibres amiloides, plantejant que si una estabilització en disminuïa la tendència, una desestabilització hauria de produir-ne un augment.

Posteriorment, van realitzar-se altres estudis sobre l'amiloidogènesi d'ADA2h emprant espectrometria de masses MALDI-TOF. Mitjançant aquesta tècnica es va poder calcular la cinètica d'incorporació dels monòmers a les estructures fibril·lars, el qual és un procés lent: $0.16 \times 10^{-3} \text{ s}^{-1}$ (Villanueva *et al.*, 2003). Els monòmers, al llarg del temps desapareixen de la fracció soluble de la mostra, ja que passen a formar part dels agregats. Utilitzant un control intern que no agrega en les mateixes condicions (en aquest cas, PCI), es pot calcular aproximadament la quantitat relativa de proteïna d'interès present a la solució comparant les intensitats dels pics corresponents als

seus pesos moleculars. Si els monòmers de la proteïna estudiada desapareixen de la fracció soluble, el seu pic disminuirà en comparació amb el de l'altra proteïna. En el mateix treball (Villanueva *et al.*, 2003) es van ampliar els assaigs de la resistència a la proteòlisi d'ADA2h realitzats a Villegas *et al.* (2000), ja que és una altra de les característiques que presenten aquests agregats proteics.

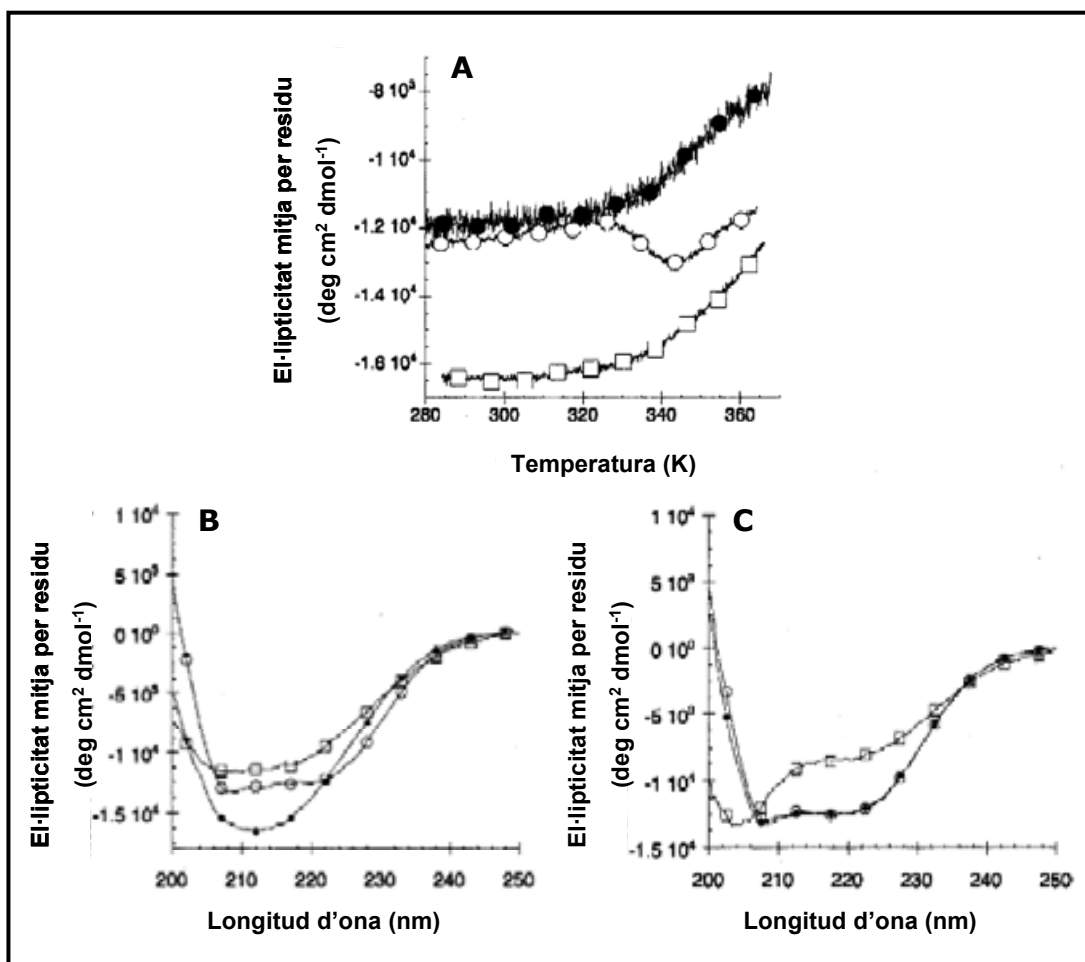


Figura 1.13. Seguiment de la transició $\alpha \rightarrow \beta$ a β d'ADA2h per temperatura i pH baix mitjançant CD. (A) Desnaturalització tèrmica per CD: DM desnaturalització i renaturalització seguida a 222 nm (cercles plens); WT seguida a 214 nm, desnaturalització (cercles buits) i renaturalització (quadrats buits). (B i C) Espectres inicials i finals de la desnaturalització tèrmica, WT (B) i DM (C). A ambdós, espectre inicial a 25 °C (cercles buits), espectre a 95 °C (quadrats buits), espectre final a 25 °C (cercles plens). Reproduït de Villegas *et al.* (2000).

OBJECTIUS

Objectius generals de la tesi

Aquesta tesi té com a punt de partida un dels objectes d'estudi del grup: les carboxipeptidases, bé que es troba centrada només en una part d'aquests enzims: el domini d'activació. Aquest domini és el responsable del manteniment de l'enzim en un estat inactiu, prevenint el dany que pot produir una proteasa quan presenta activitat fora del seu lloc d'actuació.

L'objectiu general d'aquesta tesi ha estat:

- Estudi del procés d'agregació i de possibles funcions alternatives dels dominis d'activació de les procarboxipeptidases humanes A2 i A4, respectivament.

La tesi es troba dividida en tres capítols, representatius de cada procés concret estudiat.

Objectius concrets del primer capítol de la tesi

El primer capítol de la tesi es troba centrat en l'estudi del procés d'agregació d'ADA2h i una bateria de 28 mutants puntuals mitjançant dicromisme circular.

Els objectius concrets d'aquest primer capítol han estat:

- Construcció d'11 nous mutants puntuals d'ADA2h que complementin la bateria de mutants puntuals existent. Expressió i purificació en quantitats suficients de tots els mutants, tant dels nous com dels prèviament creats.
- Posada a punt de les condicions òptimes per a l'assoliment del mateix grau de desplegament per a totes les variants d'ADA2h.
- Estudi cinètic dels events primerencs del procés de formació de fibres amiloides d'ADA2h a pH àcid a partir de l'estat desplegat, mitjançant *stopped-flow* acoblat a dicromisme circular. Comparació entre la variant salvatge (*wild-type*, WT) i els diferents mutants puntuals.
- Comparació dels diferents resultats experimentals obtinguts amb prediccions d'agregació d'algorismes informàtics.

Objectius concrets del segon capítol de la tesi

El segon capítol de la present tesi es troba centrat també en el procés de formació de fibres amiloides per part d'ADA2h, tot i que aquest cop partint de la conformació nativa de la proteïna.

Els objectius concrets d'aquest capítol han estat:

- Estudi cinètic de l'agregació de la variant WT a partir de l'estat natiu mitjançant espectroscòpia d'infraroig.
- Comparació entre la variant WT i 3 mutants puntuals d'ADA2h amb diferents velocitats d'agregació.
- Comprovació de la morfologia de les estructures finals mitjançant microscòpia electrònica de transmissió.

Objectius concrets del tercer capítol de la tesi

Per últim, el tercer capítol ha aprofundit en l'estudi d'una possible funció alternativa per al domini d'activació de la procarboxipeptidasa A4, realitzant-se primerament una cerca bioinformàtica per tal d'elaborar una hipòtesi i, posteriorment, comprovant-la experimentalment.

Els objectius concrets del capítol final d'aquesta tesi han estat:

- Cerca de proteïnes amb seqüència similar a ADA4h i diferent funció a les bases de dades, mitjançant PSI-BLAST.
- Comparació de l'estructura terciària entre ADA4h i les proteïnes trobades emprant el programa STAMP, hipotetitzant una possible funció alternativa per a ADA4h. Comparació dels residus clau per a la realització d'aquesta nova funció.
- Comprovació experimental dels resultats bioinformàtics.

MATERIAL I MÈTODES

En el següent bloc es comenten el material i els mètodes generals emprats per l'autor en la realització de la present tesi. Altres mètodes que han ajudat en la realització d'aquesta tesi però han estat utilitzats per col·laboradors no es troben descrits, tot i que apareixen breument comentats als apartats de material i mètodes de cada capítol.

M.1 APARELLS D'ÚS GENERAL

- Agitador vòrtex Heidolph, model Reax 2000 (Alemanya).
- Agitadors magnètics Agimatic-S i Agimatic-N de JP Selecta (Espanya).
- Aparell de flux detingut (*stopped-flow*) Pi-Star-180 d'Applied Photophysics (Regne Unit).
- Autoclaus Autester-E i Autester 437-P de JP Selecta (Espanya).
- Balances Mettler AJ100, PJ300 i H64 (Alemanya).
- Bany termostàtic JP Selecta, model Tectron Bio-Medic 60 (Espanya).
- Bomba peristàltica Pharmacia, model P-1 (Suècia).
- Cabina de flux laminar Faster, model TWO-30 (Itàlia).
- Centrífugues Beckman models J2-HS i J2-21 (EE.UU.) amb rotors JA-20 i JA-14.
- Comptador de centelleig LKB Wallac, model 1211 Minibeta (EE.UU.)
- Congelador -80 °C Forma Scientific, model 8325 (EE.UU.).
- Cubetes d'electroforesi per a gels d'acrilamida Bio-Rad, model Mini-Protean II (EE.UU.).
- Cubetes per gels d'agarosa Miniphor Submarine Electrophoresis Unit LKB, model 2013 (Suècia) i cubeta Pharmacia, model GNA-100 (Suècia).
- Espectrofotòmetre UV-Vis Cary100 Bio, Varian (EE.UU.).
- Espectròmetre d'infraroig Nicolet 5700, Thermo Scientific (EE.UU.).
- Espectròmetre de masses Bruker model Ultraflex (Alemanya).
- Espectropolarímetres Jasco-710 i Jasco-715 (Japó).
- Estufa incubadora de cultius Sanyo, model Mir-152 (Japó).
- Font d'alimentació per a electroforesi EPS 200 de Pharmacia Biotech (Suècia) i Power Pac 300 de Bio-Rad (EE.UU.).
- FPLC Amersham Biosciences, model Äkta Purifier, detector UV-900 i bombes P-900 (Suècia).
- Incubadors d'aire amb agitació orbital Braun, model Certomat S (Alemanya) i incubador Infors AG CH-4103 (Suïssa).
- Macrocentrífuga Heraeus model Megafuge 1.0R, Thermo Scientific (EE.UU.)
- Microcentrífuga IEC, model Micro Max (EE.UU.) i microcentrífugues refrigerades Sigma, models 2MK i 2K15 (Alemanya).
- Microscopi electrònic de transmissió Hitachi, model H-7000 (Japó).
- Microscopi electrònic de transmissió Philips Morgagni (Holanda).

pHmetre Crison Instruments S.A., model Micro pH 2001 (Espanya) amb un elèctrode Ingold (Suïssa).

Sonicador Branson Digital Sonifier 250 (EE.UU.).

Termociclador Ericomp, model Power Block (EE.UU.).

Sistema d'exposició de radioactivitat Imaging Screen-K de Kodak (EE.UU.).

Sistema de visualització per làser Personal Molecular Imager Bio-Rad (EE.UU.).

Transil·luminador de llum ultraviolada Fotodyne Inc. (EE.UU.).

M.1.1 Productes químics

Els productes químics utilitzats han estat principalment de Sigma Chemical Co. (EE.UU.). Altres productes més específics són comentats al llarg del text.

M.2 TÈCNIQUES DE DNA RECOMBINANT

M.2.1 Reacció en cadena de la polimerasa, PCR (*polymerase chain reaction*)

La tècnica de PCR (reacció en cadena de la polimerasa) és una de les tècniques actualment més emprades en Biologia Molecular. És una tècnica ràpida i senzilla que permet l'amplificació de DNA *in vitro*, inspirada en el sistema de replicació del DNA cel·lular (Mullis i Faloona, 1987). La seva utilitat va des de l'amplificació de gens de DNA amb l'objectiu de clonar-los, fins al cribatge de clons en busca dels que han incorporat el gen d'interès, o a la mutagènesi dirigida de proteïnes.

Aquest sistema es basa en les hibridacions específiques de seqüències complementàries de DNA un cop la doble cadena motlle ha estat separada degut a l'elevada temperatura. La incorporació de polimerases termoestables no desnaturalitzables en les condicions emprades (Saiki *et al.*, 1988), així com d'aparells termocicladors capaços de realitzar canvis cíclics de temperatura a elevada velocitat han permès millorar el rendiment del procés.

M.2.1.1 Amplificació de gens o fragments gènics

Per tal d'escollir el gen que es vol amplificar cal dissenyar i sintetitzar uns oligonucleòtids encebadors específics per a la seqüència d'interès (també anomenats *primers*). Amb la finalitat d'obtenir el millor rendiment possible i minimitzar els productes inespecífics, cal tenir en compte una sèrie de factors respecte al disseny dels encebadors:

- La temperatura de fusió (T_m) d'ambdós encebadors ha de ser similar entre ells, a més d'igual o superior a 50 °C. La T_m pot calcular-se segons la fórmula:

$$T_m = 4(G+C) + 2(A+T) \quad (\text{fórmula 1})$$

on A, T, C i G indiquen, respectivament, el nombre de nucleòtids d'adenina, timina, citosina i guanina en la regió de solapament. La temperatura d'hibridació sol considerar-se 5 °C per sota de la T_m dels oligonucleòtids.

- La posició 3' dels oligonucleòtids, el lloc d'unió de la polimerasa, és òptima amb la presència d'una C o una G, ja que entre aquestes bases s'estableix

una unió més forta que no pas entre A i T, aportant estabilitat per a la formació del complex amb l'enzim.

- Han d'intentar minimitzar-se les estructures secundàries dels encebadors així com la possible formació d'hibridacions entre ambdós.

Un altre factor important és assegurar-se que el gen a introduir queda clonat en fase, de manera que s'expressi la proteïna corresponent al marc de lectura correcte. Al present treball, els encebadors han estat sintetitzats per la casa comercial Roche.

Si els encebadors han estat dissenyats correctament, aquests s'uniran durant la fase d'hibridació a la seqüència complementària específica del DNA motlle, un cop la doble cadena parental ha estat desnaturalitzada per calor. Seguidament, la polimerasa s'uneix al DNA doble cadena i produeix l'elongació de la seqüència en posició 3' respecte l'encebador. Els productes d'aquesta primera reacció podran actuar en posteriors cicles com a motlle, de manera que es dona un augment exponencial del producte al llarg dels cicles.

La barreja de reacció emprada per a la clonació d'ADA4h va ser:

| | |
|-----------|---|
| 5 µl | tampó de reacció 10x sense MgCl ₂ (Biotools) |
| 1,5 µl | MgCl ₂ 50 mM |
| x µl | DNA motlle (1 ng) |
| 10 µl | de cada encebador a concentració 10 µM |
| 10 µl | barreja de nucleòtids trifosfat (dNTPs) 1 mM |
| (12-x) µl | aigua bidestil·lada estèril (fins a 50 µl) |

Just després del pas de *Hot start* s'afegeixen les polimerases:

1,5 µl polimerasa Pfu 1 U/µl (Biotools)

i el programa del termociclador utilitzat ha estat:

| | | | |
|------------|------------|-------|------------------|
| 1 cicle | 5 min | 95 °C | <i>Hot Start</i> |
| ~30 cicles | 30 s | 95 °C | Amplificació |
| | 60 s | * | |
| | 1 min 20 s | 72 °C | |
| 1 cicle | 10 min | 72 °C | Extensió final |

*La temperatura d'hibridació depèn dels encebadors dissenyats. En el cas de la clonació d'ADA4h, la temperatura assajada ha estat de 62 °C. El *Hot Start* permet assegurar-se de que totes les cadenes motlle es troben separades, i l'extensió final ajuda a l'acabament de les possibles elongacions que no havien finalitzat.

M.2.1.2 Construcció d'una seqüència de DNA sintètica

En determinades ocasions es desitja o bé construir una seqüència de DNA *de novo* molt curta o bé modificar-ne una extensivament. En aquests casos és més adient construir la molècula de DNA amb dos oligonucleòtids (un directe i un revers) parcialment solapats i acabar la síntesi de la doble cadena mitjançant el fragment *klenow* de la DNA polimerasa I d'*E.coli*. En la present tesi, aquest mètode s'ha emprat en la construcció del DNA codificant per a la sonda de RNA amb la qual es va assajar la unió amb ADA4h. Aquesta sonda està basada en la segona forqueta del snRNA U1, contenint dues repeticions del motiu d'unió a la proteïna U1A, i per això l'anomenarem sonda bipolar.

Oligonucleòtids dissenyats:

directe:

5'-**TAATACGACTCACTATAGGGAGGGTTAACATTGCACTCCGTTGTCCATCCCAAAAAAAAAA**-3'

revers:

5'-GGACAACGGAGTGCAATGTAAACCCTTTTTTTTTTGGGATGGACAACGG-3'

En negreta es mostra la regió corresponent al promotor de T7, essencial per a la transcripció *in vitro* d'aquesta molècula de DNA (veure apartat M.2.8), i en cursiva es mostra la regió parcialment solapada.

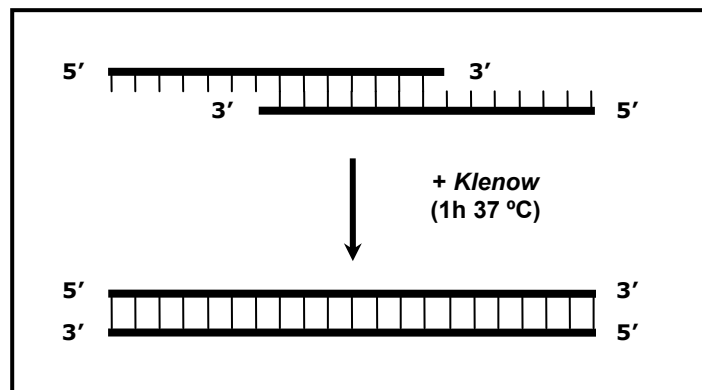


Figura M.1. Esquema de la construcció d'una seqüència de DNA sintètica. Hibridació dels 2 oligonucleòtids parcialment solapats i acabament de la síntesi de la doble cadena mitjançant el fragment *klenow* de la DNA polimerasa I d'*E.coli*.

La barreja de reacció emprada per a la construcció del DNA sintètic va ser:

2,5 µl tampó de reacció 10x (Fermentas)
 10 µl de cada encebador a concentració 100 µM
 0,5 µl barreja de nucleòtids trifosfat (dNTPs) 25 mM
 2 µl fragment *klenow* DNA polimerasa I 2 U/µl (Fermentas)

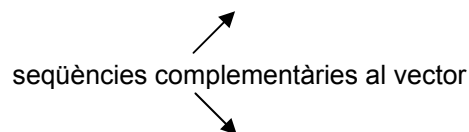
Aquesta barreja ha d'ésser incubada 1h a 37 °C. L'enzim és inactivat posteriorment incubant la barreja a 65 °C durant 10 min.

M.2.1.3 Introducció de llocs de complementarietat per al clonatge independent de lligació

Recentment es troba en augment el nombre de clonatges independents de restricció i lligació, com els sistemes LIC (*ligation-independent cloning*) (Aslanidis i de Jong, 1990) de la casa comercial Novagen. En aquest sistema, els oligonucleòtids es dissenyen amb una seqüència de 12 nucleòtids a l'extrem 5' complementària a una seqüència de cadena simple del vector. D'aquesta manera, el fragment amplificat contindrà aquesta seqüència, la qual, un cop tractada amb la DNA polimerasa del fag T4 en presència de dGTP, quedarà en forma de cadena senzilla i, per tant, serà capaç d'hibridar amb la seva cadena complementària present al vector. Aquests vectors, a més, asseguren el correcte clonatge direccional de l'insert (veure més endavant, apartat M.3.3, Fig. M.3).

Encebadors per a ADA4h:

directe: 5' **GGTATTGAGGGTCGCC**CGAGAGAAGTTTTTTGGGGACCAAGTTTTGAGG 3'



seqüències complementàries al vector

revers: 5' **AGAGGAGAGTTAGAGCC**TTACTACTGCAGGTCCTCAATTGTCAGTGG 3'

M.2.1.4 Mutagènesi dirigida mitjançant PCR

Variacions de la reacció de PCR permeten la introducció d'alguna mutació a un DNA que codifica per a una proteïna. Un mètode recent permet l'amplificació de les dues cadenes del vector sencer alhora que es realitza la mutagènesi, de manera que s'eliminen els passos de lligació (Weiner *et al.*, 1994). La clau del procés es troba en el

disseny de dos encebadors complementaris (un directe i un revers) que alineen amb la zona on es vol produir la mutació i contenen el canvi nucleotídic el qual es traduirà en el canvi aminoacídic desitjat. Un exemple són els encebadors emprats per a la mutació de l'aminoàcid aspàrtic 79 d'ADA2h a alanina (abreviat D79A):

directe 5'-CTATTCCATCATGATTGAAGCCGTCCAGTAATAGAAGC-3'
 revers 5'-AGCTTCTATTACTGGACGGCTTCAATCATGATGGAATAGGC-3'

El canvi del codó GAC per GCC produeix el pas d'aspartat a alanina mitjançant un sol canvi nucleotídic.

Per a la realització d'aquestes reaccions de PCR mutagèniques s'ha emprat el kit QuikChange™ Site-Directed Mutagenesis de Stratagene.

Una barreja de PCR conté (segons les instruccions del Manual QuikChange™):

5 µl tampó de reacció 10x sense MgCl₂ (Biotools)
 2 µl MgCl₂ 50 mM
 x µl DNA motlle (5-50 ng)
 y µl ~ 125 ng de cada encebador
 10 µl barreja de nucleòtids trifosfat (dNTPs) 1 mM
 (30,3-x-2y) µl aigua bidestil·lada estèril (fins a 50 µl)

Just després de realitzar l'etapa de *Hot Start* s'afegeixen les polimerases:

2 µl polimerasa Taq 1 U/µl (Biotools)
 0,66 µl polimerasa Pfu 1 U/µl (Biotools)

I s'incuben al termociclador segons el programa:

| | | | |
|------------|------------------------|-------|------------------|
| 1 cicle | 5 min | 95 °C | <i>Hot Start</i> |
| ~12 cicles | 30 s | 95 °C | Amplificació |
| | 60 s | * | |
| | 1 min / kb plasmidi | 72 °C | |
| 1 cicle | 10 min | 72 °C | Extensió final |

*La temperatura d'hibridació depèn dels encebadors característics per a cada mutant d'ADA2h, tot i que en el nostre cas, en tots ells l'òptima era al voltant de 58 °C, temperatura a la qual es va realitzar aquesta etapa. L'etapa d'extensió a 72 °C en el nostre cas ha tingut una durada de 6 minuts.

M.2.2 Extracció de DNA plasmídic i purificació de DNA procedent de reaccions de PCR i de bandes de gels d'agarosa.

Per a extreure el vector d'*E. coli* s'ha utilitzat el kit d'Amersham *GFX™ Micro Plasmid Prep Kit*, basat en el mètode de la llisi alcalina (Birnboim i Doly, 1979). Per a purificar tant DNA procedent de reaccions de PCR com de gels d'agarosa s'ha utilitzat el kit *GFX™ PCR DNA and Gel Band Purification Kit* de la mateixa empresa. Aquests dos kits han permès la obtenció d'un DNA força pur, apte per a experiments posteriors.

M.2.3 Digestió amb DpnI

Un altre fonament important del mètode emprat de mutagènesi dirigida per PCR (Weiner *et al.*, 1994) és l'eliminació de cadenes parentals, les quals transformarien més fàcilment a *E. coli* degut a que aquests vectors ja es troben estabilitzats. D'aquesta manera, s'evita la disminució de l'eficiència de transformació del DNA desitjat.

Les cadenes parentals difereixen de les creades per PCR en la metilació, ja que les metilases (els enzims encarregats de la metilació a la cèl·lula) no són presents a la reacció de PCR, de manera que les cadenes sintetitzades *in vitro* no es trobaran metilades. Es fa ús, llavors, de l'enzim de restricció DpnI, el qual reconeix i talla DNA metilat, disminuint significativament la presència de DNA parental i permetent un increment de l'eficiència de la transformació posterior.

A la barreja de reacció s'hi afegeix:

| | |
|--------|---------------------------------------|
| 20 µl | DNA purificat |
| 2,5 µl | tampó corresponent a aquest enzim 10x |
| 1,5 µl | DpnI 10 U/µl (Roche) |
| 1 µl | aigua bidestil·lada estèril |

Es deixa incubant 2 hores a 37 °C.

És convenient realitzar un control on només s'hi afegirà DNA parental, el qual ens donarà una idea de l'eficiència de tall de DpnI.

M.2.4 Producció d'extrems de cadena senzilla en el clonatge independent de lligació

Per a la introducció d'un insert als vectors independents de lligació, cal que aquest insert es trobi acompanyat d'una seqüència complementària a la del vector en forma de cadena simple. Degut al procés de la PCR, la seqüència introduïda als encebadors es troba en forma de cadena doble, i per tal, és necessari sotmetre l'insert a una reacció amb la DNA polimerasa del fag T4 només en presència de dGTP. L'insert provinent de PCR ha de trobar-se lliure de dNTPs, el més convenient és purificar la banda corresponent a l'insert d'un gel d'agarosa. L'insert purificat, a més, ha de trobar-se en el tampó TlowE (Tris-HCl 10 mM, EDTA 0,1 mM pH 8,0) per a assegurar l'èxit de la reacció.

Aquesta reacció utilitza l'activitat exonucleasa 3'→5' de la T4 DNA polimerasa per tal de crear els segments de cadena senzilla específics i complementaris als del vector LIC.

A la barreja de reacció s'hi afegeix:

| | |
|-------------|--|
| x µl | 0,2 pmol producte PCR-LIC en tampó TlowE |
| 2 µl | tampó de T4 DNA polimerasa 10x |
| 2 µl | dGTP 25 mM |
| 1 µl | DTT 100 mM |
| 0,4 µl | T4 DNA polimerasa a 2,5 U/µl (Novagen) |
| (14,6-x) µl | aigua bidestil·lada estèril (fins a 20 µl) |

Es deixa incubant 30 min a 22 °C; passat aquest temps s'inactiva la polimerasa incubant-la a 75 °C durant 20 min.

Al final de la reacció, les seqüències introduïdes als encebadors queden en forma de cadena senzilla, la qual és complementària als fragments de cadena senzilla presents al vector.

M.2.5 Hibridació del Vector LIC amb l'insert corresponent

Els vectors independents de lligació no requereixen de la presència d'una lligasa per tal d'aconseguir la introducció de l'insert dins el vector. Una reacció d'hibridació entre l'insert i el vector, els quals contenen extrems de cadena senzilla, és suficient per tal d'obtenir un bon rendiment (veure també apartat M.3.3, Fig. M.3).

La reacció d'hibridació es construeix de la següent manera:

1 μ l vector pET-30 Xa/LIC
2 μ l insert LIC tractat amb la T4 DNA polimerasa
(0,02 pmol)

Incubar 5 min a 22 °C i llavors afegir:

1 μ l EDTA 25 mM

Incubar una altra vegada 5 min a 22 °C.

El bacteri serà l'encarregat de produir la formació dels enllaços covalents entre l'insert i el vector, un cop la construcció hagi estat introduïda dins l'organisme hoste.

M.2.6 Introducció de DNA forani a cèl·lules bacterianes

El vector amb l'insert d'interès ha d'ésser introduït dins de bacteris per tal d'estabilitzar el vector o bé poder obtenir producció proteica. Hi ha diferents mètodes per aconseguir-ho (conjugació, transducció...), però la més emprada és la transformació degut a la seva senzillesa i capacitat d'ésser controlada. Per tal que es pugui donar a terme aquest procés, han de crear-se uns porus a la paret cel·lular dels bacteris que permetran l'entrada dels àcids nucleics.

M.2.6.1 Preparació de cèl·lules competents d'*E. coli*

Les cèl·lules competents tenen la paret bacteriana debilitada, de manera que la introducció de DNA al seu interior és molt més efectiva. La preparació es pot realitzar mitjançant diferents mètodes, però el més emprat és el del CaCl₂ (clorur càlcic) (Mandel i Higa, 1970):

1. S'agafa una colònia de soca mare d'una placa i es fa un inòcul en 5 ml de medi LB estèril.
2. S'incuba unes 12 hores a 37 °C i agitació forta, per a promoure un cultiu saturat.
3. Es realitza un inòcul 1:100 en medi LB i s'incuba en condicions similars a les del cultiu anterior fins que assolim la fase exponencial del creixement (DO₅₅₀ ~ 0,4).
4. Es reparteix el cultiu en 2 tubs de polipropilè estèrils i prèviament refredats, i es deixa en un bany de gel uns 15 min com a mínim.
5. Es centrifuguen els cultius a 3200 g durant 15 min, i a una temperatura de 4 °C.

6. S'elimina el sobrenedant per decantació i s'asseca el precipitat per inversió durant 10 min.
7. El precipitat es ressuspèn suaument en un volum 1/2 de CaCl₂ 100 mM estèril prèviament refredat, i es deixa en gel 1 hora com a mínim.
8. Es centrifuga altre cop en les mateixes condicions i es torna a assecar el precipitat per inversió.
9. Es ressuspèn el precipitat en un volum 1/20 de CaCl₂ 100 mM, deixant-ho un mínim d'1 hora en gel.
10. Es fan alíquotes de 100 µl de les competents, les quals es podran guardar a -80 °C si s'afegeix glicerol a una concentració final del 15%.

M.2.6.2 Transformació de cèl·lules competents d'*E. coli*

La transformació és un mètode d'introducció de DNA dins de cèl·lules competents. En aquest cas, s'ha utilitzat el mètode del xoc tèrmic, basat en canvis bruscs de temperatura que afavoreixen l'entrada de les molècules de DNA.

El protocol seguit ha estat el següent:

1. Descongelar una alíquota de cèl·lules competents posant-les en un bany de gel durant uns 20 min aproximadament.
2. Afegir a l'alíquota uns 50 ng de plasmidi.
3. Incubar en gel durant 30 min.
4. Xoc tèrmic:
 - col·locar 2 min en un bany a 42 °C.
 - deixar reposar 5 min en gel.
5. Afegir 1 ml de medi LB i incubar 1 hora a una temperatura de 37 °C i agitació suau (inferior a 100 rpm).
6. Plaquejar 100 µl de la solució en un medi sòlid amb un antibiòtic o un marcador apropiat que ens permeti seleccionar els clons que hagin incorporat el DNA desitjat. Deixar reposar la placa 10 min.
7. Fer un pols de centrifuga amb la resta de cèl·lules tractades amb el xoc tèrmic (uns 30 s a 6000 g).
8. Extreure la quantitat de medi necessària per a que quedin només 100 µl a l'eppendorf.
9. Ressuspendre el pellet en aquests 100 µl.
10. Plaquejar la totalitat del contingut de l'eppendorf de la mateixa manera que s'ha fet en el punt 6 (transformació 10 cops més concentrada).

En aquest protocol es realitzen 2 plaquejades, ja que si *a priori* no es coneix l'eficiència de transformació per a les condicions utilitzades (soca, vector, estat del vector en el moment de la transformació...), podria passar que fent només la primera plaquejada (punt 6) no s'obtingués cap transformant, o fent només la concentrada

(punt 10) no es poguessin distingir les colònies. Si ja s'ha fet una transformació en les mateixes condicions, la que menys ens convingui d'ambdues pot ser obviada.

És recomanable realitzar controls de transformació, un control negatiu sense afegir DNA a les cèl·lules competents, i un altre de positiu afegint un vector estabilitzat o bé amb el plasmidi parental (sense la mutació, en el cas de la mutagènesi dirigida). Aquest plasmidi parental s'ha utilitzat com a control a la PCR sense afegir-hi encebadors, i s'ha digerit amb DpnI. El resultat ens donarà una idea de la quantitat de colònies amb DNA parental a les plaques transformades amb DNA procedent de PCR mutagèniques.

M.2.7 Seqüenciació de DNA

La seqüenciació de DNA ha estat realitzat per l'empresa Sistemas Genómicos S.L. de Paterna, València, utilitzant el mètode de Sanger (Sanger *et al.*, 1977). Les mostres que s'envien provenen d'extraccions de DNA com s'ha descrit en el punt M.2.2.

M.2.8 Transcripció *in vitro* de RNA i marcatge radioactiu

La transcripció *in vitro* permet obtenir un RNA desitjat d'una forma ràpida, senzilla i controlable. Per a poder-se dur a terme cal disposar d'una molècula de DNA que contingui el promotor de la polimerasa emprada en la reacció de transcripció (veure apartat M.2.1.2). Aquesta reacció pot ser realitzada en presència d'un nucleòtid marcat radioactivament per tal de visualitzar la molècula de RNA mitjançant un gel de poliacrilamida, per exemple. La reacció s'ha dut a terme segons les instruccions del kit MEGAscript[®] T7 d'Ambion.

La barreja de reacció emprada per a l'obtenció de la sonda de RNA bipolar ha estat:

| | |
|---------|--|
| 1 µl | tampó de transcripció 10x (Ambion) |
| 0,6 µl | barreja de nucleòtids ATP, GTP i CTP cadascun a 7,5 mM* |
| 0,6 µl | nucleòtid UTP a 0,75 mM* |
| 2 µl | nucleòtid [α - ³² P]UTP (40 µCi)* (GE Healthcare) |
| x µl | DNA motlle (2,5 pmol) |
| 0,7 µl | barreja d'enzims T7 (Ambion) |
| (5,1-x) | aigua bidestil·lada estèril lliure de RNasa (<i>RNase-free</i>) (fins 10 µl) |

*Si no es desitja marcar radioactivament la molècula de RNA (per obtenir una sonda freda, per exemple), cal afegir la mateixa quantitat de nucleòtid UTP que de la resta de nucleòtids i obviar l'addició d'UTP marcat radioactivament.

La barreja de reacció és incubada durant 2 h a 37°C i posteriorment el DNA és degradat mitjançant l'addició d'1 µl de DNasa I lliure de RNasa a 10 U/µl, incubant-ho ara 15 min a 37°C.

S'ha d'anar amb cura al treballar amb RNA i cal utilitzar sempre guants i reactius lliures de RNasa, ja que aquests enzims es troben fàcilment a les mans de l'experimentador i a superfícies d'objectes, i són capaços de degradar les molècules de RNA de manera molt ràpida.

Els nucleòtids no incorporats poden separar-se de la sonda de RNA marcada radioactivament mitjançant les columnes Bio-Spin P-30 de Bio-Rad lliures de RNasa.

Tot el treball amb substàncies radioactives s'ha realitzat a les instal·lacions de radioactivitat de la Unitat de Biociències del Departament de Bioquímica i Biologia Molecular.

M.2.9 Electroforesi d'àcids nucleics

M.2.9.1 Electroforesi en gels d'agarosa

La separació de les molècules de DNA en un camp elèctric és possible gràcies a la seva càrrega negativa. Aquesta càrrega ve determinada pels grups fosfat presents a l'estructura de la macromolècula, els quals es troben en proporció 1:1 respecte les bases nitrogenades. D'aquesta manera, la quantitat de càrrega és proporcional a la grandària de la molècula de DNA.

La majoria de molècules de DNA tenen una grandària força notable, d'aquí a que es trobi força estandaritzat l'ús de l'agarosa com a matriu d'aquests gels electroforètics. L'agarosa és una substància que genera porus de major grandària que la poliacrilamida i això la fa més apta per a produir la separació de molècules d'àcid nucleic. El diàmetre del porus pot ésser controlat variant la concentració d'agarosa, permetent separar més o menys determinades mides de molècules.

Segons el que es vulgui separar, es recomanen certs percentatges d'agarosa (Sambrook *et al.*, 1989):

| % agarosa (p/v) | Rang de separació efectiu (kb) |
|--------------------|-----------------------------------|
| 0,7 | 0,8 - 10 |
| 0,9 | 0,5 - 7 |
| 1,2 | 0,4 - 6 |
| 2,0 | 0,1 -2 |

El protocol que se segueix per a cada percentatge, però, és el mateix:

1. Dissoldre la quantitat d'agarosa corresponent al percentatge d'agarosa que es vol obtenir en tampó TAE 1x.
2. Escalfar la solució per a que es dissolgui l'agarosa, la qual cosa s'aconsegueix quan la solució pren un color transparent.
3. Quan la solució d'agarosa està prou freda (que pugui tocar-se amb la mà), afegir EtBr (bromur d'etidi) per tal d'assolir-ne una concentració final de 0,5 µg/ml i abocar l'agarosa a l'espai habilitat per a la polimerització.
4. Un cop solidificat es carreguen les mostres al gel juntament amb el colorant Orange G 10x i es posa a córrer en una cubeta amb tampó TAE 1x a un voltatge constant d'uns 75 V.

Composició dels tampons utilitzats en l'electroforesi d'agarosa:

Tampó de càrrega 10x:

50% Glicerol
punta d'espàtula Colorant Orange-G

Aquest tampó conté el colorant Orange-G, el qual migra com una molècula de DNA d'aproximadament 50 pb i permet monitoritzar el progrés de l'electroforesi.

Tampó d'electroforesi TAE (Tris-acetat EDTA) 50x (1 L):

| | |
|--------|---------------------|
| 242 g | Tris base |
| 57 ml | àcid acètic glacial |
| 100 ml | EDTA 0,5 M (pH 8) |

M.2.9.2 Assajos de canvis de mobilitat electroforètica en gels de poliacrilamida (*Electrophoretic mobility-shift assays, EMSA*)

M.2.9.2.1 Preparació del gel

Les matrius de poliacrilamida són més adequades que les d'agarosa per a resoldre molècules d'àcid nucleic petites (menors d'unes 200 pb) ja que l'entramat

produït pel polímer és molt més dens. El principi de la separació, emperò, és el mateix per a ambdós tipus de gels.

Els percentatges d'acrilamida, han d'escollir-se segons la mida de les molècules que es vulguin separar (Sambrook *et al.*, 1989):

| % monòmer d'acrilamida (p/v) | Rang de separació efectiu (pb) |
|------------------------------|--------------------------------|
| 3,5 | 1000 - 2000 |
| 5,0 | 80 - 500 |
| 8,0 | 60 - 400 |
| 12,0 | 40 - 200 |
| 15,0 | 15 - 150 |
| 20,0 | 6 - 100 |

Tampó d'electroforesi TBE (Tris-borat EDTA) 10x (1 L):

| | |
|-------|-------------------|
| 108 g | Tris base |
| 55 g | àcid bòric |
| 40 ml | EDTA 0,5 M (pH 8) |

La recepta per a la polimerització de quatre gels al 5% d'acrilamida és la que segueix:

| | |
|---------|--|
| 5,8 ml | solució al 30% d'acrilamida/bisacrilamida 37,5:1 |
| 3,5 ml | TBE 10x |
| 25,7 ml | aigua bidestil·lada |
| 250 µl | persulfat amònic al 10% |
| 30 µl | TEMED |

Un cop la barreja s'ha abocat al motlle electroforètic es deixa reposar fins que hagi polimeritzat, adquirint forma sòlida.

M.2.9.2.2 EMSA

Els assajos de canvi de mobilitat electroforètica permeten avaluar la possible unió d'una proteïna a un àcid nucleic. La unió produeix una reducció en la càrrega neta negativa de l'àcid nucleic, acoblat a un augment de mida, i aquest pateix una disminució en la velocitat de migració en el gel respecte les molècules d'àcid nucleic no unides a proteïnes. Les mostres han d'ésser incubades amb un tampó d'unió adequat.

Composició del tampó d'unió 5x (Tang i Curthoys, 2001):

| | |
|--------------|--|
| 50 mM | HEPES pH 7,4 |
| 125 mM | acetat potàssic |
| 12,5 mM | acetat magnèsic |
| 2,5% | Igepal CA 630 |
| 25% | glicerol |
| 5 U/ μ l | inhibidor de RNases SupersasIn (Ambion)* |
| 5 mM | DTT* |

*És convenient afegir aquests reactius tot just al final de la preparació del tampó, degut a la seva ràpida degradació.

En el present treball s'ha assajat la unió d'una molècula de RNA, anomenada sonda bipolar, a la proteïna ADA4h, seguint el protocol:

1. Escalfar la sonda de RNA a 95 °C durant 10 min i deixar reposar 5 min en gel. Aquest pas permet la pèrdua de l'estructura secundària no desitjada de la sonda de RNA i li permet adquirir la conformació correcta.
2. Incubar durant 20 min a temperatura ambient diferents quantitats de proteïna ADA4h amb 25 fmol de sonda bipolar de RNA radioactiva en presència de tampó d'unió 1x i un volum de reacció final de 10 μ l.
3. Substituir en els casos necessaris ADA4h per les proteïnes control (negatiu, BSA; positiu, proteïna ζ -cristal·lina de llevat), i afegir quan calgui les molècules de RNA competidores per a la unió (tRNA de llevat) i la sonda bipolar freda.
4. Carregar les mostres al gel amb tampó d'aplicació d'electroforesi 6x i córrer el gel en tampó TBE 1x aplicant un voltatge constant d'uns 60-70 V durant 90 min.
5. Treure el gel de la cubeta i submergir-lo en solució de fixació durant 30 min.
6. Exposar el gel radioactiu amb el sistema *Imaging Screen-K* de Kodak durant una nit.
7. Visualitzar amb el sistema *Personal Molecular Imager* de Bio-Rad al Laboratori d'Anàlisi i Fotodocumentació, d'Electroforesi, Autoradiografies i Luminescència de la Universitat Autònoma de Barcelona.

Tampó d'aplicació d'electroforesi 6x:

| | |
|------|--------------------|
| 0,5% | Blau de bromofenol |
| 80% | Glicerol |

Solució de fixació:

| | |
|-----|-------------|
| 20% | Etanol |
| 10% | Àcid acètic |

M.2.10 Determinació de la concentració d'àcids nucleics

M.2.10.1 Determinació espectrofotomètrica

Consisteix a enregistrar l'espectre d'absorbància de la regió de 200 a 360 nm. Una unitat d'absorbància a 260 nm equival aproximadament a 50 µg/ml de DNA de doble cadena i a 40 µg/ml de RNA i DNA de cadena senzilla.

La puresa dels àcids nucleics es pot valorar amb el coeficient $U_{D_{260/280}}$, que ha de ser proper a 1,8. Valors inferiors indiquen contaminació per proteïnes.

Cal recordar que l'absorbància del RNA ha d'ésser mesurat amb una cuveta lliure d'RNases.

M.2.10.2 Estimació per electroforesi en gels d'agarosa

Quan es treballa amb quantitats molt petites de DNA, és millor estimar la seva concentració a partir de la fluorescència emesa per la tinció amb EtBr, el qual s'intercala entre les bases nitrogenades.

Es realitza una electroforesi en gel d'agarosa i es visualitza per il·luminació amb llum ultraviolada. La quantitat de DNA present a la banda de la dissolució problema s'estima per comparació de la seva intensitat amb les bandes corresponents als marcadors, de concentració coneguda.

M.2.10.3 Determinació de la concentració d'un àcid nucleic marcat radioactivament

La concentració d'un àcid nucleic que ha estat marcat radioactivament pot ésser calculada mitjançant comptatge per centelleig, basant-se en el fet que el nucleòtid radioactiu es troba en quantitats limitants. A partir d'això es calcula el rendiment d'incorporació del nucleòtid radioactiu, relacionant la radioactivitat incorporada a la sonda respecte a la radioactivitat total en comptes per minut (cpm). Posteriorment, s'aplica el rendiment obtingut com a factor corrector sobre la quantitat de RNA esperada si s'hagués obtingut un 100% d'incorporació.

Una petita quantitat de mostra (1 µl) es diposita sobre un tros de paper Whatman, es deixa assecar i es submergeix en un vial que conté líquid de centelleig. Seguidament es procedeix a la mesura de cpm mitjançant un comptador de centelleig LKB Wallac, model 1211 Minibeta que es troba dins de la instal·lació de radioactivitat.

M.3 SISTEMA D'EXPRESSIÓ I PURIFICACIÓ

M.3.1 Organismes

M.3.1.1 Soques d'*Escherichia coli*

La soca d'*Escherichia coli* utilitzada per a l'expressió heteròloga del domini d'activació de la procarboxipeptidasa A2 humana (ADA2h) i els seus mutants va ser **XL1-BLUE** (Bullock *et al.*, 1987).

Genotip: *recA1*, *endA1*, *gyrA96*, *thi-1*, *hsdR17*(r_K^- , m_K^+), *supE44*, *relA1*, λ^- , *lac^-*, [F' *proAB*, *lacI*^qZ Δ M15, Tn10(Tet^r)]

És d'especial interès la resistència a la tetraciclina, codificada al transposó Tn10, que permet seleccionar la soca enfront possibles contaminacions d'altres soques utilitzades al laboratori.

La soca utilitzada per a l'expressió d'ADA4h va ser **BL21(DE3)** (Studier i Moffatt, 1986):

Genotip: F⁻, *ompT*, *hsdS β* (r_{β} - m_{β} -), *dcm*, *gal*, (DE3) tonA

Aquesta soca resulta molt útil degut a les deleccions de les proteases *ompT* i *lon*, així com al seu doble sistema de control de l'expressió. La soca BL21(DE3) conté el gen de la RNA polimerasa del fag T7 inserat al seu genoma, sota el control del promotor de l'operó *lac*. Aquest sistema, juntament amb la utilització d'un vector amb el gen d'interès sota el control del promotor T7, permet la inexistència d'expressió en condicions basals. En absència d'inductor, la RNA polimerasa T7 no pot ésser expressada i, per tant, no es pot donar la transcripció del gen heteròleg, ja que aquest només pot ser transcrit mitjançant aquesta RNA polimerasa.

M.3.1.2 Conservació de soques bacterianes

Les soques bacterianes poden conservar-se en plaques de medi sòlid durant un període no superior a un mes, desades a una temperatura de 4 °C.

Si el que es vol és la conservació durant llargs períodes de temps, és recomanable realitzar glicerinatats al 15% que ens permetran un manteniment de la soca durant un mínim de dos anys. Amb aquesta finalitat es barregen en un vial de criocongelació 750 μ l de cultiu saturat amb 250 μ l de glicerol al 60% prèviament

esterilitzat per autoclau, col·locant-ho ràpidament en gel i congelant-ho el més aviat possible a $-80\text{ }^{\circ}\text{C}$.

M.3.2 Medis de cultiu de bacteris

Els medis de cultiu de bacteris estan constituïts per una sèrie de productes que aporten els nutrients necessaris i uns factors de selecció, els antibiòtics.

M.3.2.1 Medi líquid

El medi de cultiu utilitzat per al creixement d'*E. coli* ha estat el Luria-Bertani (LB), seguint el protocol descrit (Sambrook *et al.*, 1989).

Per a la preparació d'1 L:

| | |
|--------|-----------------------|
| 10 g | de triptona |
| 5 g | d'extracte de llevat |
| 10 g | de NaCl |
| 900 ml | d'aigua bidestil·lada |

S'ajusta el pH a 7 i s'enrasa fins a 1 L amb aigua bidestil·lada. Posteriorment s'esterilitza mitjançant un cicle d'autoclau.

M.3.2.2 Medi sòlid

Per a l'aïllament de colònies d'*E. coli* és adequat l'ús de plaques amb medi sòlid. Se segueix un protocol semblant al medi líquid, ja que conté els mateixos components, i s'afegeix agar al 15% (p/v), el qual solidifica juntament amb el medi un cop la barreja ha retornat a temperatura ambient després d'haver-se escalfat.

Per a la preparació de 500 ml de solució (20 plaques aproximadament):

1. Es preparen 500 ml d'LB segons el protocol del punt M.3.2.1.
2. S'afegeixen al medi 7,5 g d'agar i es barreja una mica (no s'aconseguirà una solució homogènia degut a que l'agar és insoluble en medi LB a aquesta temperatura).
3. S'autoclava el preparat anterior (amb la calor s'homogeneïtza l'agar).
4. Es deixa refredar una mica la solució i s'afegeix la quantitat desitjada d'antibiòtic si es volen preparar plaques per a seleccionar.
5. Mentre la solució encara estigui calenta, es reparteix en unes 20 plaques i es deixa refredar. Es treballa en condicions d'esterilitat.
6. Desar les plaques a $4\text{ }^{\circ}\text{C}$.

En aquestes condicions, les plaques es poden mantenir durant un període aproximat d'un mes.

M.3.2.3 Antibiótics

Els antibiòtics són unes substàncies químiques que tenen la capacitat o bé d'evitar el creixement o bé de produir la mort de bacteris, degut a que interfereixen en funcions cel·lulars vitals. Algunes proteïnes bacterianes confereixen resistències a determinats antibiòtics, les quals solen trobar-se codificades en gens localitzats a plasmidis. Els plasmidis emprats en biologia molecular per tal d'introduir gens foranis a bacteris contenen també un gen que confereix resistència a un antibiòtic. D'aquesta manera, la introducció d'un antibiòtic al medi de cultiu crea una pressió selectiva positiva sobre les cèl·lules bacterianes que contenen el vector d'interès, degut a que tenen assegurada la supervivència i la proliferació.

Alguns dels antibiòtics més emprats i les seves principals concentracions d'ús són:

Taula M.1. Principals antibiòtics i concentracions de treball. Antibiòtics utilitzats en biologia molecular i en expressió heteròloga de proteïnes. Extret de Sambrook *et al.* (1989). Tots els stocks poden ser guardats a -20 °C en absència de llum. *Dissoldre en etanol 100%. †Dissoldre en etanol 50%.

| Antibiòtic | Concentració de treball (µg/ml) | Mecanisme d'acció |
|-----------------|---------------------------------|--|
| Ampicil·lina | 60 | Evita la síntesi de pèptidglicà |
| Carbenicil·lina | 60 | Evita la síntesi de pèptidglicà |
| Cloramfenicol* | 170 | Unió a la subunitat 50s del ribosoma bacterià, inhibint-lo |
| Kanamicina | 50 | Unió a la subunitat 30s del ribosoma bacterià, inhibint-lo |
| Streptomicina | 50 | Unió a la subunitat 30s del ribosoma bacterià, inhibint-lo |
| Tetraciclina† | 50 | Unió a la subunitat 30s del ribosoma bacterià, inhibint-lo |

L'ampicil·lina ha estat àmpliament emprada, però recentment va perdent terreny d'ús respecte la kanamicina. L'ampicil·lina actua sobre la paret bacteriana, i l'enzim que en confereix resistència, la β -lactamasa, ha d'ésser excretat al medi per tal d'anul·lar l'efecte d'aquest antibiòtic. La conseqüència d'aquest mètode de resistència és que cèl·lules sense el gen de resistència però properes a d'altres que el posseeixen són capaces de sobreviure gràcies a la presència de l'enzim que degrada l'antibiòtic al medi. D'altra banda, la kanamicina actua a l'interior cel·lular, així com l'enzim que en

confereix resistència, i per tant només la supervivència dels bacteris que contenen el gen de resistència està assegurada.

D'entre els antibiòtics de la taula M.1, en el present treball s'han utilitzat principalment l'ampicil·lina i la kanamicina (per a ADA2h en ptz18U i ADA4h en pET30 Xa/LIC, respectivament). Els antibiòtics carbenicil·lina i tetraciclina només s'han emprat en passos en els quals s'ha volgut assegurar la màxima qualitat del procés. Aquests antibiòtics permeten una selecció de major qualitat per al vector ptz18U (ADA2h), en el cas de la carbenicil·lina, i per a la soca XL-1-BLUE (ADA2h), en el cas de la tetraciclina.

Els antibiòtics dissolts en aigua s'esterilitzen per filtració (els dissolts en etanol ja són estèrils) i són adicionats al medi de cultiu esterilitzat per autoclau un cop la seva temperatura ha disminuït fins a uns 50°C.

M.3.3 Vectors d'expressió

El vector que s'ha usat per a l'expressió d'ADA2h ha estat el **ptz18U** (antiga Pharmacia) (Mead *et al.*, 1986), un fagsmidi de 2860 pb, el qual permet seleccionar ràpidament els clons que el contenen gràcies a la seva resistència a l'ampicil·lina, Amp^R (Fig. M.2). Un altre dels seus avantatges és l'elevada quantitat de proteïna expressada, fet comprovat anteriorment amb ADA2h (Villegas *et al.*, 1996; Villegas *et al.*, 1998; Villegas *et al.*, 2000).

Aquest vector, a més, conté la regió N-terminal de la β -galactosidasa, la qual permet l' α -complementació si s'utilitza en una soca adequada, com és XL1-BLUE. L' α -complementació es basa en el fet que vector i soca contenen 2 fragments diferents de la proteïna β -galactosidasa, de manera que si aquests coexisteixen en un organisme, permeten el correcte plegament i la funció de la proteïna mitjançant interaccions no covalents. El lloc de clonació múltiple del vector (*multicloning site*, MCS) trenca la continuïtat del fragment de la proteïna que es troba en aquest, de manera que no pot expressar-se i s'evita l'activitat de la β -galactosidasa. Si s'afegeix un substrat de la β -galactosidasa que dóna color si és hidrolitzat, X-gal, a aproximadament 0,05 mg/ml, es pot distingir entre les colònies d'una placa que contenen l'insert (blanques) de les que no el contenen (blaves), seleccionant juntament amb ampicil·lina per seleccionar prèviament només les cèl·lules han incorporat el plasmidi.

A més a més, s'ha utilitzat el vector **pET-30 Xa/LIC** de 5448 pb de la casa Novagen (Fig. M.2). Aquest vector conté un promotor de T7, el qual només pot ésser reconegut per la polimerasa T7 codificada a la zona insertada del genoma de la soca

d'*E. coli* BL21(DE3), de manera que es disminueixen els nivells d'expressió basal del gen d'interès en absència de l'inductor. A més a més, el promotor del fag T7 és de naturalesa molt forta i assegura una elevada transcripció del gen d'interès. El vector també conté vàries seqüències traduïbles que faciliten la purificació de la proteïna i la seva detecció mitjançant anticossos específics, com 2 cues d'histidines (H•Tag[®]) i el pèptid de fusió S•Tag[™] (Fig. M.3 A). La presència de la cua d'histidines en posició C-terminal a la proteïna traduïda podrà obtenir-se si no s'introdueix un codó de finalització posterior a l'insert. De la mateixa manera, aquest vector conté dianes per a les proteases de plasma trombina i factor Xa, permetent la separació de la proteïna clonada de les diferents cues. L'ús de la trombina dona com a resultat la proteïna d'interès juntament amb l'S•Tag[™] i l'ús del factor Xa resulta en la proteïna recombinant amb només la cua d'histidines C-terminal (si l'encebador revers no contenia el codó stop i l'insert es troba en fase).

Aquest vector es comercialitza obert, amb dos fragments de ssDNA que flanquegen la regió d'inserció del gen. S'incorporen, al gen d'interès, les seqüències complementàries a les que conté el vector, de manera que es pugui donar la hibridació entre vector i insert en aquestes regions, i es permeti la clonació independent de lligació (Aslanidis i de Jong, 1990). Aquesta estratègia permet una gran simplificació del procés de clonació. Aquest vector conté un gen de resistència a la kanamicina, per tal de poder ésser seleccionat.

M.3.4 Inducció

La inducció de l'expressió dels dominis d'activació es produeix quan la soca ha assolit la fase exponencial del creixement, la qual cosa es comprova espectrofotomètricament observant l'absorció del cultiu a una longitud d'ona de 550 nm fins que aquest arriba a una densitat òptica d'aproximadament 0,4. La inducció es fa amb IPTG, un substrat capaç d'unir-se al repressor *lac*, produint l'expressió del/s gen/s que es trobi/n sota control del promotor *lac*. Quan el cultiu es troba en creixement exponencial s'hi afegeix una relació 1:1000 d'IPTG 40 mg/ml sobre el cultiu (concentració final a 40 µg/ml, per tant 0,17 mM d'IPTG).

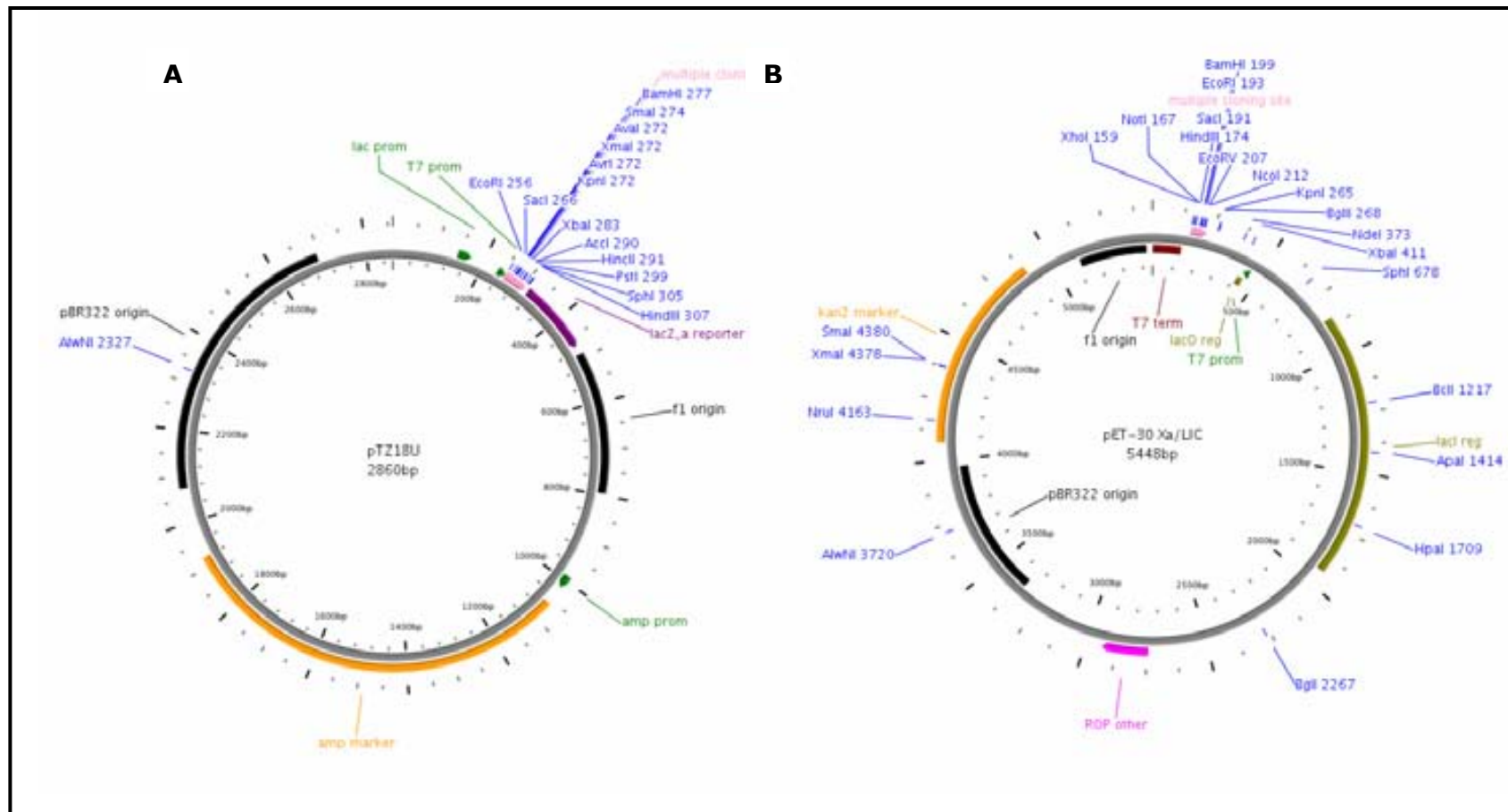


Figura M.2. Mapa dels vectors d'expressió utilitzats en el present treball. El pTZ18U (A) ha estat utilitzat en l'expressió d'ADA2h i el pET-30 Xa/LIC (B) en l'expressió d'ADA4h. Per a l'elaboració d'aquests mapes s'ha emprat el programa web PlasMapper versió 2.0 (<http://wishart.biology.ualberta.ca/PlasMapper/>) (Dong *et al.*, 2004).

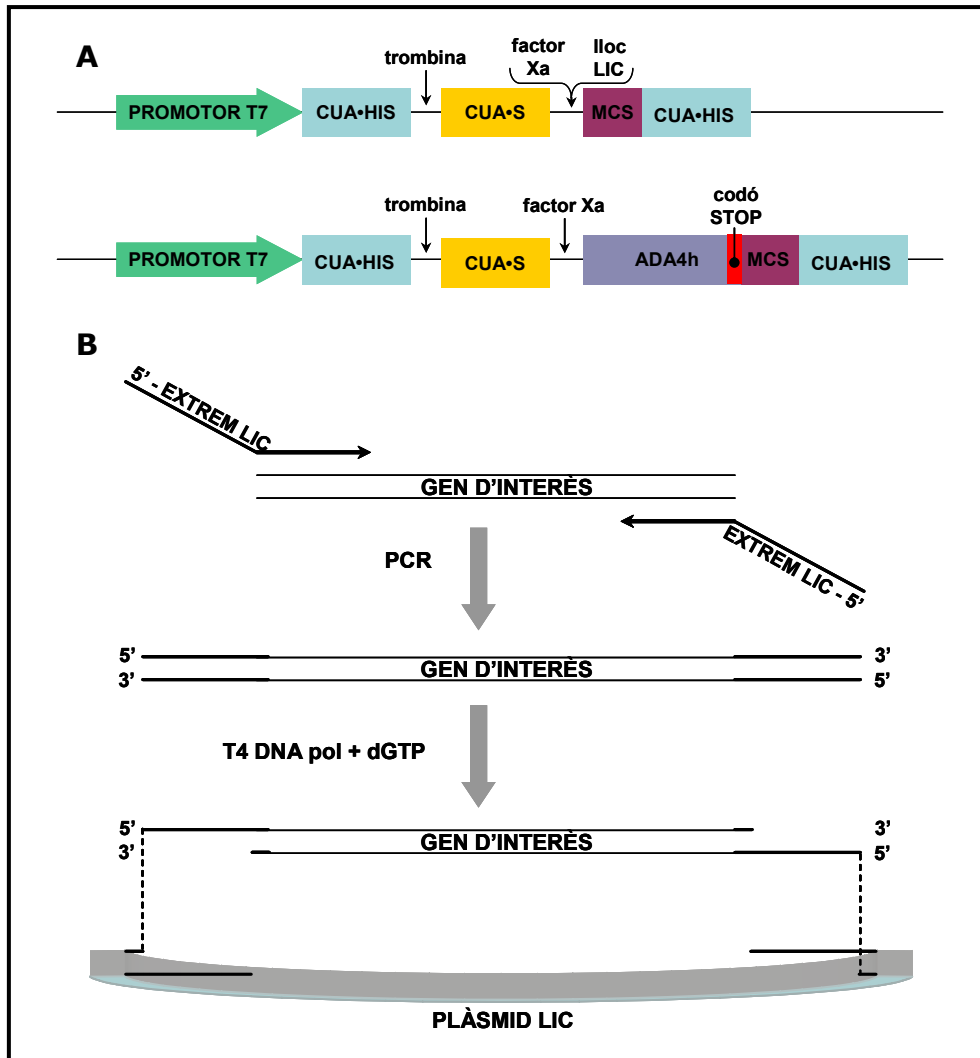


Figura M.3. Clonatge al vector pET-30 Xa/LIC. (A) Esquema de les opcions de clonatge del vector pET-30 Xa/LIC i opció escollida per a ADA4h. Cua·His, H·Tag[®]; Cua·S, S·Tag[™]; MCS, *multicloning site*. (B) Estratègia general del clonatge en un vector LIC.

M.3.5 Obtenció de l'extracte cel·lular

El primer pas de purificació de proteïna és l'obtenció de l'extracte cel·lular. El procés de purificació de proteïna varia segons una gran quantitat de factors, com per exemple:

- propietats de la proteïna expressada
- organisme utilitzat
- lloc d'acumulació de la proteïna expressada heteròlogament

En el cas d'ADA2h, s'ha utilitzat un mètode descrit prèviament (Jiménez *et al.*, 2003), bé que una mica modificat. ADA2h es localitza intracel·lularment en forma soluble, de manera que, un cop sintetitzada, aquesta pot recuperar-se fàcilment, tot i que caldrà separar-la de les nombroses proteïnes intrínseques d'*E. coli*. El mètode es basa, sobretot, en el fet que ADA2h és la proteïna més petita expressada en quantitats notables en aquesta soca d'*E. coli* (amb una massa molecular de 9212 kDa per la forma WT), amb suficient diferència de grandària respecte les altres com per separar-la per cromatografia de gel-filtració en FPLC (*fast protein liquid chromatography*).

Primerament es realitza una centrifugació per tal de recuperar les cèl·lules del medi de cultiu (3500 g, 30 min, 4 °C). Les cèl·lules es ressuspenen a 4 °C en un tampó de baixa força iònica (Tris 10 mM pH 8) i es sotmeten a 4 cicles de congelació amb nitrogen líquid seguits de descongelació a 37 °C. Posteriorment es tracta l'extracte amb DNasa I a concentració final 20 µg/ml durant 10 minuts a temperatura ambient i es torna a centrifugar per tal de separar la part soluble de la cèl·lula (48000 g, 60 min, 4 °C), on es troba la nostra proteïna, de les parts membranoses.

Per a ADA4h, no obstant, les cèl·lules van ressuspendre's en tampó d'unió de la cromatografia d'afinitat (veure M.3.6.2) i es van sotmetre al mateix tractament que l'emprat per a ADA2h, però es va afegir també Tritó X-100 a concentració final 0,1% per tal de millorar la disgregació de les membranes, juntament amb la DNasa I. A més a més es va incloure un pas de sonicació (2 cops 1 min seguits de 2 cops 1 min 30s, al 100% de potència), emprant la sonda més gruixuda, previ a la separació de les parts soluble i membranoses de les cèl·lules bacterianes (realitzada igual que per a ADA2h).

M.3.6 Tècniques de cromatografia líquida

M.3.6.1 Cromatografia atmosfèrica de bescanvi aniònic

Les cromatografies atmosfèriques pre-empaquetades Sep-pak de Waters tenen una resolució força baixa, però ens permeten separar una certa quantitat de proteïnes de la mostra, així com d'altres molècules presents a l'extracte (lípid, etc.) de manera que aquesta pugui ésser sotmesa a una cromatografia més resolutiva. El flux de la columna pot augmentar-se mitjançant l'acoblament a una bomba peristàtica P-1 de Pharmacia.

S'ha utilitzat la columna Sep-pak Vac 35 cc Accell™ Plus QMA per a la purificació d'ADA2h. Es fa servir Tris 50 mM pH 8 per tal d'equilibrar la columna, un tampó semblant al que es troba a la proteïna. Aquest tampó també s'usa per realitzar un primer rentat de dos volums de columna per a eliminar totes les molècules no unides a la reïna. L'elució es produeix per augment de la força iònica, en 3 fases diferents d'1 volum de columna cadascuna: Tris 50 mM acetat amònic 0,15 M (pH 8), Tris 50 mM acetat amònic 0,4 M (pH 8) i Tris 50 mM acetat amònic 0,8 M (pH 8). La proteïna elueix a una concentració d'acetat amònic 0,4 M.

M.3.6.2 Cromatografia atmosfèrica d'afinitat amb histidines

La cromatografia d'afinitat amb histidines se serveix de la capacitat de la cadena lateral d'aquests residus (el grup imidazol) per unir fortament ions metàl·lics com el níquel. D'aquesta manera, la introducció de 6 residus histidina a un dels extrems de la proteïna assegura una unió forta de la proteïna al níquel, el qual pot ser immobilitzat en columna mitjançant una reïna quelant. En el nostre cas, la reïna utilitzada ha estat la Chelating Sepharose Fast-Flow de GE Healthcare, amb una capacitat d'unió d'aproximadament 40 mg de proteïna amb una cua de 6 histidines per ml de reïna. La cua d'histidines introdueix un fet diferencial respecte a les altres proteïnes presents al sistema de producció, ja que aquestes últimes, en no tenir un nombre tan gran d'histidines exposades no podran restar unides a la columna i seran eluïdes en les primeres etapes de la cromatografia. Aquesta tècnica permet, doncs, obtenir la proteïna de manera quasi pura en tan sols un sol pas. La recuperació de proteïnes unides no específicament a la reïna, així com la de la pròpia proteïna d'interès es duu a terme mitjançant l'aplicació de tampons amb concentracions creixents d'imidazol.

Per tal de controlar el flux d'aquesta columna s'ha emprat, a l'igual que la cromatografia anterior, una bomba peristàltica P-1.

Un altre dels grans avantatges d'aquesta tècnica és la capacitat de donar-se lloc la unió d'afinitat fins i tot en presència d'agents caotrópics a elevada concentració, com la urea. Això permet separar primerament la majoria de proteïnes contaminants de la d'interès, per tal de que es pugui produir un relegament amb interferències mínimes.

Aquesta cromatografia ha estat emprada per a la purificació d'ADA4h amb una cua d'histidines en posició N-terminal, a partir de la fracció soluble.

Els tampons utilitzats per a la cromatografia, tots a un pH de 8, són:

Tampó de càrrega:

50 mM sulfat de níquel

Tampó d'unió:

20 mM Tris - HCl pH 8

0,5 M NaCl

5 mM imidazol

Tampó de rentat:

20 mM Tris - HCl pH 8

0,5 M NaCl

60 mM imidazol

Tampó d'elució:

20 mM Tris - HCl pH 8

0,5 M NaCl

1 M imidazol

És important el fet que la mostra ha de trobar-se en tampó d'unió a l'hora d'aplicar-la a la columna, per tal que la unió entre la cua d'histidines i el níquel sigui màxima i disminuir el nombre d'unions inespecífiques al metall.

El protocol serà expressat mitjançant volums de columna, tot i que s'utilitzarà el terme volum amb l'objectiu de simplificar:

1. Netejar bé la reïna amb ddH₂O per tal d'eliminar totes les restes d'etanol emprades en la seva conservació.
2. Aplicar 5 volums de tampó de càrrega per immobilitzar el níquel a la reïna quelant.

3. Equilibrar la columna amb 10 volums de tampó d'unió.
4. Aplicar la mostra en tampó d'unió.
5. Aplicar 15 volums de tampó d'unió a la columna per tal d'eliminar tota la mostra que no hagi quedat unida a la reïna.
6. Seguidament, es passaran 10 volums de tampó de rentat per a desfer les possibles unions inespecífiques d'altres proteïnes amb la reïna.
7. Posteriorment, s'aplicaran 10 volums de tampó de rentat-2, amb la mateixa finalitat que l'anterior (punt 6), però ara amb condicions més restrictives.
8. Per tal d'eluir la proteïna, es realitzarà un gradient de 60 mM a 1 M d'imidazol emprant el tampó de rentat i el tampó d'elució. Un cop determinada la concentració d'imidazol requerida per a l'elució de la proteïna aquesta pot ésser emprada posteriorment al rentat. Per a ADA4h ha estat de 300 mM d'imidazol.

La mostra ha d'ésser sotmesa a diàlisi enfront tampó d'unió, per tal d'eliminar l'excés d'imidazol i tornar a ésser unida a la níquel-sefarosa.

Aquest tipus de cromatografia també pot realitzar-se en *batch*, col·locant la quantitat de reïna necessària dins un recipient que pugui ésser aplicat a una centrífuga, com un tub cònic. Les centrifugacions han de realitzar-se a baixa força (50 g), permetent separar la reïna del tampó aplicat en cada cas. Aquesta modalitat sol utilitzar-se quan s'ha de donar la digestió de la proteïna de fusió per part d'una proteasa específica, per tal d'eliminar les cues que en faciliten la purificació.

M.3.6.3 Cromatografia de gel-filtració d'alta resolució (FPLC)

En els últims passos de la purificació solen emprar-se columnes altament resolutives que ens permeten, de forma senzilla i ràpida, separar la proteïna d'interès de la resta de contaminants. El tipus de cromatografia escollida depèn de les propietats de la proteïna que es vol purificar. A la purificació dels dominis d'activació s'aprofita el fet que aquests seran les proteïnes més petites presents en quantitats notables al sistema d'expressió, i per això s'ha escollit una cromatografia de gel filtració de volum d'exclusió molt petit, de manera que la resta de proteïnes quasi no entren en el rang de separació de la columna i elueixen juntes. La columna preparativa emprada ha estat una HiLoad™ 26/60 Superdex 75 Prep Grade d'Amersham Biosciences (ara GE Healthcare) per a FPLC (*Fast Protein Liquid Chromatography*). L'acoblament de columnes a aquests aparells permet l'automatització del procés, així com la possibilitat de treballar a pressions superiors a l'atmosfèrica, produint una

optimització dels fluxos i, per tant, del temps. La mostra ha de concentrar-se prèviament (fins a uns 2 ml), ja que és important en el cas d'aquest tipus de cromatografia que totes les molècules a separar parteixin del mateix punt, per tal de minimitzar la difusió. Amb aquesta finalitat s'utilitzen els concentradors per a centrifuga Amicon Ultra-15 de tamany de porus de 3500 Da de la casa Millipore. S'equilibra la columna amb 2 volums de tampó fosfat sòdic 50 mM (pH 7), i, un cop carregada la mostra, s'elueix amb 1 volum del mateix tampó a un flux de 2 ml/min i una pressió màxima de 0,5 MPa.

La proteïna, un cop eluïda, i degut al tipus d'experiments que han de realitzar-se, convé concentrar-la fins a 2 mg/ml, com a mínim, i guardar-la a -80 °C.

M.3.7 Diàlisi

Els mètodes de diàlisi permeten l'extracció de les sals presents al tampó d'una mostra o bé l'intercanvi de tampó. Amb aquesta finalitat s'han utilitzat les membranes de cel·lulosa Spectra/por 3 de mida d'exclusió de 3500 Da de Spectrum Labs. Per a diàlisis de volums inferiors a 100 µl s'han emprat els mini-dialitzadors *Slide-A-Lyzer*[®] *Mini Dialysis Unit*, amb la mateixa mida d'exclusió, de la casa Pierce.

Ambdós tipus de diàlisis s'han realitzat a 4 °C sota agitació suau i enfront un volum de tampó de com a mínim 500 cops el volum de la mostra durant 6 hores.

M.3.8 Digestió amb factor Xa

El factor Xa és una proteasa capaç de reconèixer i tallar específicament una seqüència proteica. La diana d'aquesta proteasa sol trobar-se en alguns vectors que contenen cues (també anomenades *tags*), permetent la separació d'aquestes de la proteïna d'interès, al final del seu procés de purificació. Això evita qualsevol interferència que la cua pogués ocasionar en els experiments realitzats amb la proteïna clonada.

La seqüència reconeguda pel factor Xa és:



després de la qual talla.

Un dels sistemes més adients per al tall i separació de les cues de fusió consisteix en immobilitzar la proteïna de fusió en la reïna d'afinitat corresponent a la

cua i tractar-la posteriorment amb la proteasa específica, de manera que la cua restarà unida a la reïna i la proteïna d'interès eluirà.

La reïna amb la proteïna de fusió immobilitzada es renta amb 10 volums del tampó de digestió del factor Xa: Tris-HCl 50 mM, NaCl 100 mM, CaCl₂ 5 mM, pH 8. La quantitat de clorur sòdic pot ésser augmentada, en principi fins a 500 mM, si es sospita que la proteïna resultant pot ser susceptible d'agregar. Un cop han passat uns 9 volums de columna s'afegeix la quantitat necessària de factor Xa i s'incuba durant 16 h a 21 °C. La quantitat de factor Xa òptima pot ésser calculada tenint en compte que 1 U de factor Xa és capaç de digerir 50 µg de proteïna de fusió en més del 95% en 16 h d'incubació a 21 °C.

M.4 MÈTODES DE PROTEÏNES

M.4.1 Mètodes d'identificació i determinació de massa molecular

M.4.1.1 Electroforesi en SDS (SDS-PAGE)

M.4.1.1.1 Preparació del gel

El mètode electroforètic es basa en la separació de proteïnes en aplicar un camp elèctric extern.

L'electroforesi en presència del detergent SDS ha estat una eina àmpliament emprada en l'anàlisi de proteïnes. El detergent forma una pel·lícula al voltant de la proteïna i en produeix el desplegament, degut a que el SDS té una elevada densitat de càrregues negatives que es repel·len entre elles, i fan adoptar a la proteïna la conformació on aquestes es troben més allunyades (Shapiro *et al.*, 1967). Amb l'ús d'aquest detergent les càrregues intrínseques de la proteïna queden emmascarades, i, degut a que la quantitat de detergent unit és proporcional a la grandària de la molècula (1,4 g SDS / 1 g proteïna), aquestes se separaran en el si del gel segons la seva massa molecular. L'electroforesi es realitza sobre un suport d'acrilamida, una substància que polimeritza gràcies a l'addició de persulfat amònic i N,N,N',N'-tetrametilè diamí (TEMED) creant una malla bidimensional amb porus de diàmetre controlable, segons el percentatge d'acrilamida utilitzat.

El mètode discontinu és segurament el més emprat actualment (Laemmli, 1970). Aquest procediment es basa en un gel per a l'electroforesi format per dues fases (anomenades normalment gels) amb percentatges d'acrilamida i pH diferents:

- gel concentrador: Forma un entramat força lax, amb un percentatge d'acrilamida al voltant del 5% i un pH inferior al del gel separador. Aquest gel permet la concentració de la mostra carregada, proporcionant un punt de partida igual per a totes les proteïnes en entrar al gel separador.
- gel separador: amb percentatge d'acrilamida variable. En aquest gel es donarà la separació de les molècules segons el seva massa molecular.

Per a la preparació dels gels, s'utilitzen 3 solucions diferents:

| | |
|-----------|-------------------------------------|
| solució A | 40% Acrilamida/bisacrilamida 37,5:1 |
| solució B | Tris-HCl 1,5 M SDS 0,4% (pH 8,8) |
| solució C | Tris-HCl 0,46 M SDS 0,4% (pH 6,8) |

Aquestes solucions són vàlides per a la preparació de gairebé qualsevol gel en dues fases, amb diferents percentatges d'acrilamida per al gel separador.

Per a preparar un parell de gels al 15% d'acrilamida:

| | | | |
|------------------|---|---------|----------------------|
| gel separador | { | 3,75 ml | solució A |
| | | 2,5 ml | solució B |
| | | 3,75 ml | aigua bidestil·lada |
| | | 40 µl | persulfat amònic 10% |
| | | 10 µl | TEMED |
| gel concentrador | { | 0,45 ml | solució A |
| | | 1 ml | solució C |
| | | 2,55 ml | aigua bidestil·lada |
| | | 44,4 µl | persulfat amònic 10% |
| | | 4 µl | TEMED |

Les mostres es carreguen en el gel amb tampó d'aplicació d'electroforesi 3x, després de ser tractades 5 min a 95-100 °C per tal de desplegar completament les cadenes polipeptídiques i permetre la unió homogènia del SDS:

| | |
|--------------------|--|
| Tris | 180 mM |
| Glicerol | 30% |
| SDS | 9% |
| Blau de bromofenol | 0.3% |
| β-mercaptoetanol | 15% (no convé afegir-lo a una solució <i>stock</i> , ja que sol evaporar-se amb relativa rapidesa) |

Un cop preparat el gel, aquest es col·loca en una cubeta adient, aplicant-li un amperatge constant de 20 mA/gel, i es deixa fins que el front del colorant del tampó d'aplicació hagi desaparegut per la part inferior del gel. Es retira llavors el gel de la cubeta i es disposa a la seva tinció.

M.4.1.1.2 Tinció i destinció del gel

Els gels d'acrilamida es tenyeixen utilitzant el colorant blau de Coomassie (Coomassie-Blue).

El tampó de tinció conté:

| | |
|-----|-------------------|
| 45% | metanol |
| 45% | aigua destil·lada |
| 10% | àcid acètic |

Sobre la qual s'afegeix:

| | |
|------------|-----------------------------------|
| 0,2% (p/v) | Colorant Coomassie Brilliant Blue |
|------------|-----------------------------------|

El gel es col·loca en un recipient adequat i s'hi aboca la solució de tinció, deixant-lo en agitació suau una estona fins que el gel adopti una tonalitat blavosa. Aquest colorant té un límit de detecció de 0,3 – 1 µg de proteïna.

La destinció dels gels es realitza mitjançant l'ús d'àcid acètic al 10%, deixant-lo en aquesta solució el temps necessari per a una bona apreciació de les bandes.

M.4.1.2 Espectrometria de masses MALDI-TOF

L'espectrometria de masses MALDI-TOF (*matrix assisted laser desorption/ionisation - time of flight*) és una tècnica que ens permet la determinació de la massa molecular d'una proteïna, així com la identificació d'una proteïna de massa molecular coneguda entre una barreja de diferents proteïnes.

La mostra a analitzar ha de barrejar-se amb una matriu, generalment un àcid orgànic dèbil, com pot ser l'àcid sinapínic, a una relació molar mostra/matriu d'entre 100:1 a 50000:1. Un cop la mostra s'ha assecat pot introduir-se dins de l'espectròmetre de masses, on majoritàriament la matriu rebrà l'energia d'un làser, produint-ne l'evaporació i, per tant, la transferència de la mostra a la fase gasosa. La matriu actua, a més, com a agent protonant afavorint la ionització de la mostra amb la qual està mesclada, però evitant-ne la fragmentació. Les molècules seran accelerades en un camp elèctric i arribaran al detector de temps de vol (TOF) segons la seva relació massa/càrrega (m/z). La placa de suport de la mostra pot ser també de diferents materials, utilitzant-se principalment en aquest treball la *polished steel* (acer polit). Les mostres han estat analitzades amb un espectròmetre de masses Bruker model Ultraflex al Servei de Proteòmica i Bioinformàtica de l'IBB.

M.4.2 Determinació de la concentració de proteïna

M.4.2.1 Mètodes espectrofotomètrics

M.4.2.1.1 Absorbància d'aromàtics

La determinació espectrofotomètrica de la concentració de proteïna és una de les més acurades i més fàcils d'utilitzar. Tot i que es poden utilitzar reactius que treballen a longituds d'ona corresponents a l'espectre del visible (com el reactiu de Bradford), coneixent el coeficient d'extinció molar de la proteïna (ϵ) a la longitud d'ona de màxima absorbància per part dels aminoàcids aromàtics, es pot determinar la concentració d'aquesta en una solució on es trobi pura, mitjançant la llei de Lambert-Beer:

$$C = \frac{A_{280}}{\epsilon_{280} \cdot l} \quad (\text{fórmula 2})$$

on C és la concentració de proteïna a la mostra, A_{280} és la mesura d'absorbància a 280 nm de la mostra problema i l és el pas de llum de la cubeta emprada, en cm.

ϵ es pot calcular pel mètode descrit per (Gill i von Hippel, 1989) a 280 nm, mesurant la mostra en condicions natives i desnaturalitzades (amb un agent caotròpic, com la urea o el clorur de guanidini) a 280 nm i substituint a la fórmula:

$$\epsilon_{\text{experimental, 280}} = \epsilon_{\text{teòric, 280}} \cdot \frac{A_{\text{nativa, 280}}}{A_{\text{desnaturalitzada, 280}}} \quad (\text{fórmula 3})$$

Si es vol obtenir un resultat més acurat, abans d'aplicar la llei de Lambert-Beer es pot mesurar i restar el valor d' A_{320} al de A_{280} ja que el primer es correspon a la dispersió de la llum deguda a la turbolesa de la solució, i el càlcul de la diferència permet eliminar la seva contribució a l'espectre.

Per a ADA2h els coeficients d'extinció molar són (Villegas *et al.*, 1995):

$$\begin{aligned} \epsilon_{280}^M &= 7062,32 \text{ u.a.} \cdot \text{M}^{-1} \cdot \text{cm}^{-1} \\ \epsilon_{280}^{0.1\%} &= 0,78 \text{ u.a.} \cdot \text{mg/ml}^{-1} \cdot \text{cm}^{-1} \end{aligned}$$

M.4.2.1.2 Mètode colorimètric de Bradford

El mètode de Bradford (Bradford, 1976) s'utilitza per a la quantificació de proteïnes en dissolució. Aquest mètode està basat en la interacció del reactiu de Bradford, el qual conté blau de Coomassie, amb l'enllaç peptídic de les proteïnes.

El protocol emprat ha estat el següent:

1. Realitzar una sèrie de dilucions amb albúmina de sèrum boví (BSA), amb concentracions creixents de proteïna compreses entre 0 i 100 $\mu\text{g/ml}$, i aplicar-ne 200 μl en tubs eppendorf.
2. Aplicar 200 μl de diverses dilucions de la nostra proteïna, també en tubs eppendorf.
3. Afegir a cada eppendorf, de manera ordenada, 800 μl de reactiu de Bradford (Bio-Rad), diluït prèviament 4 vegades amb aigua destil·lada.
4. Llegir l'absorbància a 595 nm amb un espectrofotòmetre, emprant el mateix ordre amb el que ha estat afegit el reactiu, i entre 10 i 60 minuts després de l'inici de la reacció.

Aquest protocol pot ésser modificat per a permetre la lectura en plaques multipouet.

M.4.2.2 Mètode electroforètic

Una altra manera de conèixer la concentració de proteïna (també es podria aplicar en el cas de DNA) és el mètode electroforètic, el qual pot ser útil per a determinar concentracions de la nostra proteïna en una barreja que en conté d'altres, prèvia separació electroforètica.

El mètode es basa en la densitometria de bandes, de manera que es realitzaran comparacions de les bandes problema entre elles, per a conèixer-ne la concentració relativa, o bé la comparació amb una banda patró, la qual cosa implica la introducció a l'electroforesi d'una solució que contingui la proteïna d'interès a concentració coneguda. La densitometria, en aquest cas, s'ha dut a terme amb l'aparell *GS-700 Imaging Densitometer* de Bio-Rad al Laboratori d'Anàlisi i Fotodocumentació, d'Electroforesi, Autoradiografies i Luminescència de la Universitat Autònoma de Barcelona.

M.4.3 Dicroisme Circular

El dicroisme circular (CD) és una tècnica força completa, que permet tant l'anàlisi de l'estructura secundària (emprant longituds d'ona que pertanyin a l'ultraviolat llunyà), com d'estructura terciària (amb longituds d'ona de l'ultraviolat proper).

El dicroisme circular és una tècnica espectroscòpica fonamentada en la llum polaritzada circularment, la qual incideix sobre la mostra. La proteïna absorbeix aquesta llum i la desvia un cert angle segons la seva estructura.

Una de les principals característiques de les fibres amiloides és el seu alt contingut en estructura secundària- β , la qual cosa fa que el CD llunyà sigui una eina valuosa per a l'anàlisi d'aquests tipus d'agregats.

S'han realitzat uns experiments seguint una sèrie de protocols que van posar-se a punt amb el CD, un Jasco J-710.

M.4.3.1 Espectre llunyà de CD

Els espectres llunyans de CD són útils per a tenir una visió global del plegament secundari de la proteïna, i observar-ne, per tant, canvis de conformació.

D'especial importància és la comprovació abans de fer l'espectre, de que el voltatge de la longitud d'ona inferior no excedeix els 600 V, ja que a partir d'aquests valors les mesures no es poden considerar fiables. Si el voltatge és superior, es pot optar per utilitzar una cubeta de pas de llum inferior, augmentar la longitud d'ona inferior de l'espectre o bé augmentar l'ample de banda fins a uns 4-8 nm. A la majoria dels casos aquest fet no es pot conèixer *a priori*, i el que cal és provar diferents condicions fins a trobar el voltatge òptim per a tenir un bon senyal. També es molt importat el tampó que s'utilitza, en aquest cas fosfat 50 mM (pH 7 i pH 2), i glicina 25 mM (pH 3), ja que son tampons de baixa absorció en l'espectre llunyà.

Els espectres s'han realitzat a una resolució de 0,1 nm, un ample de banda d'1 nm, resposta d'1 s i una velocitat de 100 nm/s. S'han acumulat 20 espectres en els experiments de fibril·lació a 2 concentracions d'ADA2h i 10 en el cas de comprovacions de conformació de les cinètiques de canvi de conformació.

Si es vol conèixer la contribució relativa de cada estructura secundària a l'espectre de CD s'ha de fer una deconvolució, mitjançant algun dels programes informàtics disponibles a servidors, com DICHROWEB (<http://www.cryst.bbk.ac.uk/cdweb/html/home.html>).

M.4.3.2 Perfils tèrmics

Els perfils tèrmics de CD llunyà permeten l'observació de l'evolució de l'espectre complet o d'un punt de l'espectre d'una proteïna enfront la temperatura, podent-se relacionar amb els canvis produïts en l'estructura secundària. Sol ésser molt emprat en les desnaturalitzacions tèrmiques, tot i que en el nostre cas ha estat utilitzat per tal de

trobar la temperatura mínima on tots els mutants d'ADA2h es trobessin en el mateix grau de desplegament.

El protocol és força semblant al de la cinètica enfront de temps:

1. Dissoldre la proteïna en urea 8 M fins a obtenir una concentració de proteïna 40 μ M i 4 M d'urea en tampó fosfat sòdic 50 mM pH 7.
2. Col·locar la solució en una cubeta de pas de llum de 0,1 cm.
3. Termostatitzar durant 5 minuts la mostra a 25 °C.
4. Realitzar un espectre inicial, entre 250 i 210 nm.
5. Seguir l'el·lipticitat a 215 nm entre 25 °C i 98 °C, a 40 °C/h, una resolució de 0,2 °C i una resposta de 8 s.
6. Un cop assolida la temperatura desitjada, mantenir-la 10 min.
7. Adquirir un espectre com al punt 4.
8. Retornar a 25 °C, amb els mateixos paràmetres que al punt 5.
9. Realitzar un espectre final per a observar-ne la conformació.

M.4.3.3 Stopped-flow

L'*stopped-flow* és un aparell que permet la mesura de cinètiques ràpides, que es produeixen en una escala de temps comparable al temps de mescla d'un experiment convencional. L'*stopped-flow* utilitzat ha estat un PiStar-180 de l'empresa Applied Photophysics, el qual té integrat un detector que permet processar el senyal de dicroisme circular, a més d'un altre detector de fluorescència.

La barreja es produeix mitjançant la injecció de la mostra de la proteïna desnaturalitzada i el desencadenant de la cinètica sobre el mateix lloc de detecció, de manera que es pot iniciar l'adquisició de dades instantàniament. Això assegura que els primers moments de la cinètica siguin també registrats.

El programari de l'aparell permet l'anàlisi de les dades obtingudes.

El protocol seguit ha estat el següent:

1. Dissoldre la proteïna en urea 8 M fins a obtenir una concentració 44 μM de proteïna i 4 M d'urea en fosfat sòdic 75 mM (pH 7) per partir d'un estat desnaturalitzat.
2. Preparar una solució HCl a 0,85 N en urea 4 M.
3. Col·locar la solució del punt 1 en la xeringa de 2,5 ml i la del punt 2 en una de 250 μl del sistema d'injecció de l'*stopped-flow*.
4. Emplenar tot el sistema amb la mostra per assegurar que quan realitzem la injecció s'injecti el que es desitja.
5. Termostatitzar durant 5 min a 60 °C (temperatura a la qual es va comprovar que ADA2h es trobava desplegada en presència de 4 M d'urea).
6. Injectar un volum de 100 μl totals i adquirir la cinètica a 220 nm durant uns 100 s, mesurant 2000 punts i acumulant 10 injeccions per tal de reduir el soroll. L'adquisició no es va poder realitzar a 215 nm per problemes de saturació de l'aparell. Es va utilitzar una amplada de banda de 8 nm per tal de minimitzar aquests problemes.
7. Adquisició d'un espectre final per a comprovar si s'ha produït un canvi en la conformació de la proteïna.

Seguint aquest protocol, la concentració final de proteïna, al donar-se la injecció, serà de 40 μM , i la d'urea, 4 M, de manera que les condicions seran molt semblants a les de l'experiment de dicromisme circular anterior realitzat amb un dicrògraf convencional.

M.4.4 Espectroscòpia d'absorció d'infraroig amb Transformada de Fourier

L'espectroscòpia d'infraroig és una tècnica àmpliament utilitzada en l'obtenció d'informació sobre l'estructura secundària proteica, tot i que pot ser emprada també per a l'anàlisi d'altres molècules. Aquesta tècnica es basa en el fet que l'absorció de llum infraroja per part de la mostra depèn de la vibració dels enllaços que la formen. En el cas de les proteïnes, la zona de l'espectre de l'infraroig més utilitzada és la compresa entre els números d'ona 1700 i 1600 cm^{-1} , corresponents a la forma de vibració I de l'enllaç amida.

L'espectroscòpia d'infraroig requereix d'una elevada concentració de proteïna per tal que la relació senyal / soroll sigui suficientment elevada i, per tant, significativa. A més a més, és força recomanable l'ús d'aigua deuterada (D_2O) en les mostres, ja

que la banda corresponent a la vibració de l'enllaç O-H de l'aigua emmascara la zona d'absorció de l'enllaç amida, evitant la visualització de les conformacions proteiques.

Per a l'estudi de l'agregació amiloide d'ADA2h a partir de l'estat natiu es va dialitzar durant tota la nit la proteïna a concentració 544 μM (5 mg/ml) enfront tampó Glicina pD 3. Posteriorment, la mostra es va col·locar dins d'una cel·la definida per dues finestres separades per 50 μm . Es va realitzar una rampa de temperatura de 20 $^{\circ}\text{C}$ a 90 $^{\circ}\text{C}$ en 10 min i posteriorment la temperatura va ésser mantinguda a 90 $^{\circ}\text{C}$ durant 30 min. Es van adquirir espectres amb una resolució nominal de 2 cm^{-1} fins que aquests van començar a evidenciar la deshidratació de la mostra. Paral·lelament es van dur a terme espectres de vapor i cinètiques de referència que només contenien el tampó on la mostra es trobava dissolta, i van ser restats durant el tractament matemàtic de les dades amb el programa RAMOPN. Les deconvolucions i les àrees van ésser calculades amb el mateix programa, i les correlacions 2-D i 3-D van ésser realitzades amb el programa Matlab versió 7.0.

M.4.5 Microscòpia electrònica de transmissió

La microscòpia electrònica ens permet magnificar objectes fins a 1 milió de vegades amb un límit de resolució d'uns 2 nm. Aquesta gran magnificació s'aconsegueix emprant un feix d'electrons. Aquest feix és modulats mitjançant una sèrie de lents electromagnètiques, de manera que aquest incideixi sobre la mostra que es vol visualitzar. Alguns electrons seran capaços d'atravessar la mostra donant-ne una imatge augmentada. Tot el sistema ha de trobar-se en el buit, per tal que els electrons no col·lisionin amb les molècules d'aire i siguin, per tant, desviats.

Les mostres es preparen sobre reixetes de coure recobertes d'una capa de carboni, les quals han estat tractades amb un aparell de descàrrega lumínica (*glow-discharge*) per tal d'eliminar-ne el caràcter hidrofòbic i permetre la correcta distribució de la mostra. Un cop realitzat aquest tractament, les reixetes han d'acabar de preparar-se abans de 30 minuts, ja que passat aquest temps, els efectes de la descàrrega lumínica desapareixen.

En el present treball hem utilitzat la tinció negativa per a la visualització de fibres amiloides, seguint el protocol de López De La Paz *et al.* (2002):

1. Realitzar dilucions de la mostra tot just abans d'aplicar-la a la reixeta, per tal de poder observar les estructures d'interès en condicions òptimes.
2. De manera ràpida aplicar 5 μl de la dilució sobre la capa de carboni de les reixetes de microscòpia. És important la rapidesa d'aquesta etapa, sobretot quan es volen observar estructures susceptibles d'ésser dissoltes en diluir.
3. L'excés de mostra es retira amb una tira de paper *Whatman* tallada molt fina. És important que el paper no toqui la capa de carboni de la reixeta.
4. Agafant la reixeta amb unes pinces i invertint-la, es posa en contacte suaument amb una gota d'uns 10 μl d'acetat d'uranil a l'1%, al qual s'ha eliminat el producte no dissolt mitjançant un filtre de 0,22 μm . Es manté la reixeta en aquesta posició durant uns 20 - 30 s, tot realitzant moviments circulars per tal que tota la mostra quedi el més uniformement tenyida.
5. L'excés d'acetat d'uranil és retirat amb una mica de paper *Whatman* de la mateixa manera que al punt 3.
6. La mostra es deixa assecar i pot ser visualitzada immediatament o bé guardada en condicions de buit per a ésser utilitzada més endavant.

Complementàriament, es poden incorporar rentats amb aigua ddH₂O per tal de netejar la mostra o eliminar l'excés d'acetat d'uranil.

Les mostres han estat analitzades mitjançant un microscopi electrònic de transmissió Philips Morgagni a l'EMBL i un Microscopi Hitachi, model H-7000 al Servei de Microscòpia de la Universitat Autònoma de Barcelona.

M.5 MÈTODES BIOINFORMÀTICS

M.5.1 Eina de cerca d'homologia: *Basic Local Alignment Search Tool* (BLAST i PSI-BLAST)

BLAST (<http://www.ncbi.nlm.nih.gov/blast/Blast.cgi>) és un programa de lliure accés de la xarxa el qual permet la cerca de similituds entre les seqüències nucleotídiques o aminoacídiques desitjades i les bases de dades disponibles a la xarxa. A partir de la comparació entre seqüències, el programa assigna un valor anomenat E (E-value) segons el grau de similitud entre aquestes. A menor valor E, major similitud entre seqüències. A més del valor E, el programa permet la visualització dels alinaments obtinguts entre la seqüència problema i cada una de les seqüències identificades com a similars per part del programa.

BLAST té diferents variants que poden ser emprades depenent de les seqüències de partida i de l'anàlisi concret que es vol realitzar. A més a més, cada variant té una sèrie d'opcions que permeten millorar l'adaptació de la cerca a les condicions concretes per a cada cas. Una de les variants més importants de BLAST és PSI-BLAST (Position specific iterative BLAST) (Altschul *et al.*, 1997), el qual permet la cerca d'homologies més remotes que mitjançant BLAST.

Per a la cerca d'homologies remotes per al domini d'activació de la procarboxipeptidasa humana A4 (ADA4h) s'ha utilitzat aquesta última variant (PSI-BLAST), mitjançant el suport Linux, el qual permet augmentar el ventall d'opcions disponibles del programa respecte la versió web. En aquest cas, primerament s'ha realitzat una cerca de PSI-BLAST sobre la base de dades *Uniprot* (Bairoch *et al.*, 2005), emprant una sola iteració, i augmentant el valor E fins a 5,0 per tal de permetre la incorporació de seqüències més remotes. La matriu obtinguda amb aquesta recerca va ser llavors introduïda en la següent cerca, ara amb el valor E restablert a 3,0. Ambdues cerques van realitzar-se emprant un filtre per tal d'evitar la presència de carboxipeptidases en els resultats, ja que no permetien la obtenció de nova informació, a més de filtres que evitaven l'obtenció de proteïnes amb funció hipotètica (no demostrada) o desconeguda.

M.5.2 Eina de superposició d'estructures terciàries: *Structural Alignment of Multiple Proteins (STAMP)*

STAMP és també un programari lliure que realitza la comparació d'estructures terciàries entre dues o més proteïnes, utilitzant les coordenades dels seus carbonis- α (Russell i Barton, 1992) que obté dels arxius pdb d'aquestes. STAMP produeix un seguit de comparacions estructurals, i proporciona a l'usuari la que té el menor valor de desviació (rmsd, *root mean square deviation*) entre les estructures donades, és a dir el millor encaix entre les estructures. A més a més, aquest programa permet la generació de fitxers PDB que permeten la visualització de la superposició.

Aquest programa ha estat emprat en la present tesi per tal d'evaluar la similitud entre les estructures terciàries d'ADA4h (codi pdb: 2boa) i una proteïna similar a l'obtinguda mitjançant l'anàlisi de PSI-BLAST (U1A; codi pdb: 1m5k). La generació d'un arxiu PDB de la superposició ha permès, a més a més, la visualització dels possibles llocs d'interacció entre ADA4h i l'estructura d'RNA que es trobava unida a U1A.

RESULTATS

**CHAPTER 1:
EARLY KINETICS OF AMYLOID FIBRIL FORMATION REVEALS
CONFORMATIONAL REORGANISATION OF INITIAL
AGGREGATES**

1.1 SUMMARY

Understanding the initial steps of protein aggregation leading to the formation of amyloid fibrils remains a challenge. Here, the kinetics of such a process is determined for a misfolding protein model, ADA2h. The double nature of the very early kinetics suggests a step model of aggregation, where the denatured polypeptide folds into an aggregated β -intermediate that subsequently reorganises into a more organised β -sheet-rich structure that finally results in amyloid fibre formation. To determine the regions of the protein involved in amyloidosis, we have analysed a series of mutants previously made to study ADA2h folding. Using the algorithm TANGO, we have designed mutants that should enhance or decrease aggregation. Experimental analysis of the mutants shows that the C terminus of the molecule (comprising the last and edge β -strand) is the major contributor to amyloid fibril formation, in good agreement with theoretical predictions. Comparison with proteins with similar topology reveals that family folds do not necessarily share the same principles of protein folding and/or aggregation.

1.2 INTRODUCTION

Amyloidoses are an increasing set of conformational diseases, including Alzheimer's disease and transmissible spongiform encephalopathies, associated with the occurrence of highly ordered protein aggregates in different tissues (Dobson, 1999; Rochet & Lansbury, 2000; Dobson, 2001). In the last decade, proteins apparently unrelated to any known disease have been found to convert into amyloid fibrils *in vitro*, showing all the physicochemical characteristics of those related to diseases (Guijarro *et al.*, 1998; Konno *et al.*, 1999; Villegas *et al.*, 2000; Fändrich *et al.*, 2001). Nowadays, it is apparent that the ability to form amyloid fibrils is a general property of polypeptide chains (Dobson, 1999; Dobson, 2001), modulated by the nature of the side-chains and the charge (López De La Paz, *et al.*, 2002; Chiti *et al.*, 2003; López de la Paz & Serrano, 2004). It has been described recently that the cellular toxicity produced by amyloid aggregates of proteins related specifically to neurodegenerative diseases is reproduced, to a similar extent, by proteins unrelated to diseases (Stefani, 2003; Stefani & Dobson, 2003). It is important to note that early specific aggregates, more than mature amyloid fibrils, seem to account for amyloid-induced cell death and the subsequent degeneration of the central nervous system in neurodegenerative diseases (Goldberg & Lansbury, 2000; Bucciantini *et al.*, 2002; Walsh *et al.*, 2002; Caughey & Lansbury, 2003; Selkoe, 2003). Hence, understanding the initial steps of the process leading to the formation of amyloid fibrils is a challenge. The use of protein models from which detailed folding and stability studies are available is a crucial step in resolving general rules underlying the molecular basis of amyloid-based diseases. For the sake of clarity, it is important to differentiate β -sheet aggregation from amyloid aggregation. The former does not produce toxic aggregates and results in the deposition of amorphous material, while the latter produces small toxic aggregates and, at the end of the process, results in the deposition of fibrillar material (Pastor *et al.*, 2008). While both processes share the formation of extensive β -sheet structure, not all aggregates will result in amyloid formation (López de la Paz *et al.*, 2005). Studies of β -sheet aggregation processes have led to the development of different algorithms that predict the effect of sequence variation on β -sheet aggregation (Chiti *et al.*, 2003; DuBay *et al.*, 2004; Fernández-Escamilla *et al.*, 2004; Linding *et al.*, 2004; Pawar *et al.*, 2005). Also, studies on a series of peptide variants have produced a sequence fingerprint that picks up short sequence stretches that have a tendency to form amyloid fibrils (López de la Paz & Serrano, 2004; López de la Paz *et al.*, 2005). In fact, recent evidence supports the idea that the amyloid formation process can be triggered by

short stretches of amino acid residues in the protein. (Ventura *et al.*, 2004; Esteras-Chopo *et al.*, 2005).

So far, theoretical predictions of the effect of point mutations on the aggregation tendency of a protein have been made for few proteins (Chiti *et al.*, 2003; Fernández-Escamilla *et al.*, 2004; Linding *et al.*, 2004). To validate their utility, it is important to test them on more model proteins. In this work, we have chosen as target protein the activation domain of human procarboxypeptidase A2, ADA2h, which forms amyloid fibrils *in vitro* under certain experimental conditions (Villegas *et al.*, 2000) and its folding pathway driving to the native state has been characterised extensively (Villegas *et al.*, 1995; Villegas *et al.*, 1996; Viguera *et al.*, 1997; Villegas *et al.*, 1998; Fernández *et al.*, 2000; Villanueva *et al.*, 2000; Villanueva *et al.*, 2002; Villanueva *et al.*, 2003). This 81-residue domain folds into two α -helices packing against a four-stranded β -sheet, as derived from the crystal structure of the complete proenzyme (García-Sáez *et al.*, 1997), and from the NMR structure (Jiménez *et al.*, 2003). The comparison of both structures shows a slight destabilisation of β -strand 2 in the isolated domain due to the lack of specific contacts within the enzyme moiety. Its folding pathway at neutral pH follows a two-state model (Villegas *et al.*, 1995), and a transition state nucleated by the central β -strands (β 1 and β 3) and a part of α -helix 2 (Villegas *et al.*, 1998). We have stated that the rational redesign of the α -helices of ADA2h by introducing favourable local interactions stabilising the native fold (Villegas *et al.*, 1996; Viguera *et al.*, 1997) is a strategy to preclude amyloid fibril formation, with subsequent application in gene therapy (Villegas *et al.*, 2000).

Here, we have performed a systematic analysis of the aggregation rates of about 30 mutants of ADA2h, and compared these experimental results with TANGO (Fernández-Escamilla *et al.*, 2004; Linding *et al.*, 2004), and with other theoretical approaches (Chiti *et al.*, 2003; DuBay *et al.*, 2004; Pawar *et al.*, 2005). The regions susceptible to β -aggregation, as predicted by TANGO, have been synthesised and characterised *in vitro*. The extremely fast kinetics of aggregation from the denatured state of this domain and its variants has required stopped-flow measurements, and constitutes a description of the very initial steps of protein aggregation leading to the formation of amyloid fibrils. The results are discussed on the light of the detailed information available for this domain and other amyloidogenic polypeptides.

1.3 MATERIALS AND METHODS

1.3.1 Materials

ADA2h variants were expressed in *Escherichia coli* strain BL21(DE3). Urea and all other reagents were purchased from Sigma-Aldrich. Peptides were synthesized by Thermo Electron Company and oligonucleotides were synthesized by Roche Molecular Biochemicals. Dialysis membranes were obtained from Spectrum Laboratories Inc.

1.3.2 TANGO prediction and design of new mutants

The online software TANGO was used for the design of new mutations: T6V, Q11K, K63A, E67K, Q69E, Y73F, S74K, E78A, E78K, D79A and D79K.

1.3.3 Mutagenesis, expression and purification

Mutagenesis was carried out with the QuikChange protocol (Stratagene) with a modified pTZ18U vector as a template (Bruix *et al.*, 1993). Expression and purification of said mutants were realized by improving the cell disruption of a previously described method (Villegas *et al.*, 1998). The protein was extracted from the cells with four freeze/thaw cycles and DNA viscosity was eliminated adding Sigma DNase I to the cell extract at a final concentration of 20 µg/ml. Design of peptides Based on Fig. 1.1, four peptides corresponding to the aggregation-prone regions were designed: Peptide 1 spans residues 2–11 (RSLETFVG DQ), peptide 2 spans residues 54–64 (VPFVNVQAVKV), peptide 3 spans residues 62–71 (VKVFLESQGI) and peptide 4 spans residues 69–79 (QGIAYSIMIE). The peptides were synthesised protected at both ends (with Ac at the N terminus and Am at the C terminus).

1.3.4 Far-UV circular dichroism

For peptides, the lyophilised powder was made to 500 µM with 7.5 mM sodium phosphate (pH 2.0), 0.01% (v/v) trifluoroacetic acid, stirred overnight at room temperature, then centrifuged at 11000 g to remove insoluble particles. The

supernatant was analysed immediately by far-UV CD in a Jasco J-715 spectropolarimeter using a 0.1 cm path length. A 1:10 (v/v) dilution of these supernatants was analysed using a 1 cm path length quartz cuvette. Measurements were averaged for 20 scans recorded at 100 nm min⁻¹. For protein variants, different concentrations (20 μM, 40 μM, 60 μM, and 200 μM) were assayed in a Jasco J-715 spectropolarimeter in appropriate path length cuvettes. In the study of aggregation from the native state, the proteins were recorded initially in 50 mM sodium phosphate (pH 7.0) at 25 °C. Then, concentrated samples (2 mg/ml) were dialysed overnight against 25 mM glycine (pH 3.0) at 4 °C, made to 20 μM and 200 μM with the dialysis buffer, and heated at 95 °C for 15 min in order to favour acquisition of the β-conformation (Villegas *et al.*, 2000). Measurements were averaged for 20 scans recorded at 100 nm min⁻¹. Thermal denaturation of 20 μM variants in 50 mM sodium phosphate buffer (pH 7.0), 4 M urea, was performed from 25 °C to 98 °C at a speed of 40 °C/h, and followed at 215 nm.

1.3.5 Electron microscopy

Peptides 1 and 4 were induced to form amyloid fibrils at a concentration of 50 μM in 7.5 mM sodium phosphate (pH 2.0), 0.01% trifluoroacetic acid, and were analysed in a Philips Morgagni with CCD camera as described (López De La Paz *et al.*, 2002). For protein experiments, fibril formation of WT ADA2h at 40 μM in 75mM sodium phosphate (pH 7), 4 M urea was induced by pH-jump as in the kinetic experiments. Two controls were realised, at 90 °C, under conditions similar to those used previously (Villegas *et al.*, 2000), and at 60 °C in the absence of urea.

1.3.6 Kinetic experiments

Samples were dissolved initially in 75 mM sodium phosphate (pH 7.0), 4 M urea and jumped to pH 2 with HCl. The mixing was carried out using a 2.5 ml for the samples and a 0.250 ml Hamilton syringe for the HCl embodied in an Applied Photophysics PiStar-180 stopped-flow. Samples were first equilibrated at 60 °C for 5 min to ensure that all mutants have the same extent of residual structure. Detection was performed by CD spectroscopy following the acquisition of conformation at 220 nm, since the signal at 215 nm was too noisy. In order to further decrease the noise, five 200 μl shots were averaged and an 8 nm band width was used. Initial and final CD

spectra were recorded, to ensure the initial random coil and the final β -conformation in a Jasco J-710 spectropolarimeter, averaging 20 scans recorded at 100 nm min^{-1} . The fitting to a double decay was performed with $p < 0.0001$.

1.3.7 Formula-based predictions

The formula developed by Dobson and col. (Chiti *et al.*, 2003) was compared with the experimental data obtained in this work:

$$\ln(v_{\text{mut}} / v_{\text{wt}}) = A\Delta\text{Hydr} + B(\Delta\Delta G_{\text{coil-}\alpha} + \Delta\Delta G_{\beta\text{-coil}}) + C\Delta\text{Charge}$$

$$\Delta\text{Hydr.} = \text{Hydr}_{\text{wt}} - \text{Hydr}_{\text{mut}}$$

$$\Delta\Delta G_{\beta\text{-coil}} = 13.64 (T_{\beta}^{\text{wt}} - T_{\beta}^{\text{mut}})$$

$$\Delta\Delta G_{\text{coil-}\alpha} = RT \ln (T_{\alpha}^{\text{wt}} / T_{\alpha}^{\text{mut}})$$

$$\Delta\text{Charge} = |\text{Charge}_{\text{mut}}| - |\text{Charge}_{\text{wt}}|$$

The general A, B, and C parameters described in the original publication are 0.63, 0.20, and -0.49, respectively.

A modification of the algorithm published by (Pawar *et al.*, 2005) is available online, <http://www.zyggregator.com/>. ZYGGREGATOR™ is a trademark by Zyentia Ltd.

1.4 RESULTS

1.4.1 TANGO analysis of ADA2h

Fig. 1.1 A shows the profile obtained when submitting the sequence of wild-type (WT) ADA2h to TANGO at pH 2. The region encompassing the C terminus of the molecule (especially β -strand 4), is predicted as the region of the protein mainly involved in the aggregation process, although a slight susceptibility is predicted at the very N terminus. In order to validate the TANGO prediction, four peptides spanning the regions predicted as related to aggregation have been synthesized (Fig. 1.1 B). Peptide 1 spans residues 2–11 (RSLET FVGDQ), peptide 2 spans residues 54–64 (VPFVNVQAVKV), peptide 3 spans residues 62–71 (VKVFLESQGI) and peptide 4 spans residues 69–79 (QGIAYSIMIE).

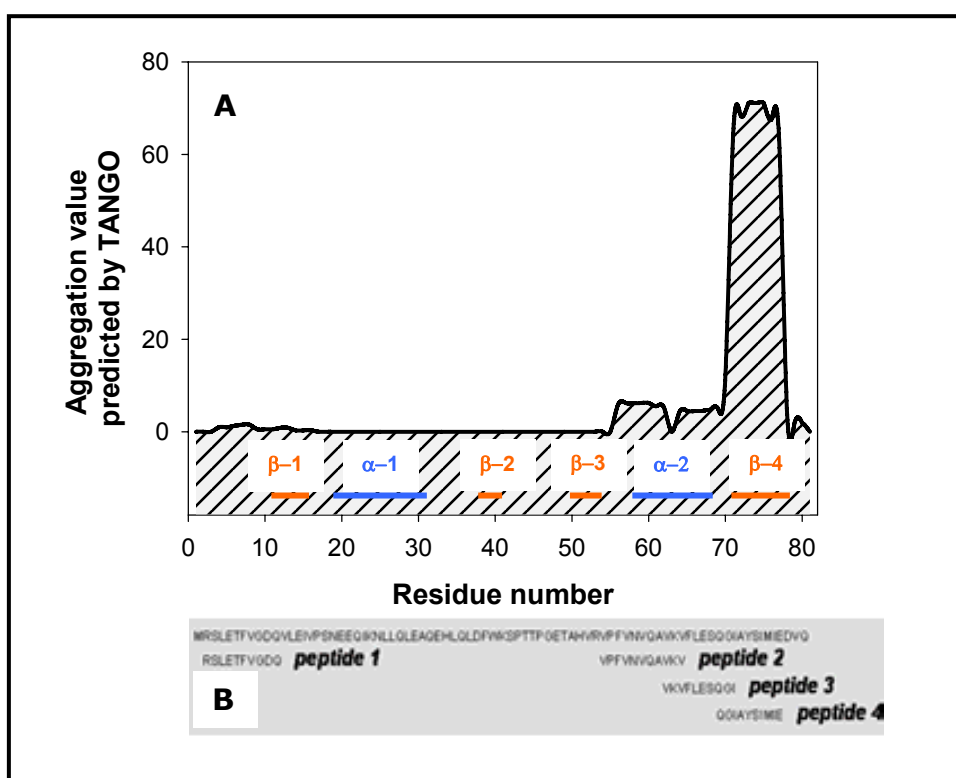


Figure 1.1: Aggregation profile of the WT-ADA2h sequence, secondary structure location, and peptides designed. (A) The relative values predicted by TANGO indicate those regions more susceptible to influence in the aggregation rate of the molecule. (B) Peptides designed to evaluate the possible residual structure in the denatured state of the domain and its potential aggregation.

This design allows us to study the N terminus (peptide 1) and the C terminus by overlapped peptides (peptides 2, 3 and 4). TANGO total aggregation prediction for these peptides at pH 2 is as follows: peptide 1, 1.0; peptide 2, 36.3; peptide 3, 0.0; and

peptide 4, 242.0. Peptides 2 and 3 in the profile have very similar heights, but peptides 2 and 3 have different scores. The prediction for peptide 3 is no aggregation at all, due mainly to the presence of K and E at relative distance $i, i+4$ favouring the α conformation. However, the profile shows some tendency to aggregation in this region, since it takes into account the tendency of adjacent sequence; in this case, the strongest aggregating peptide 4. Thus, we will expect no aggregation for peptides 1 and 3, slight aggregation or poor solubility for peptide 2 and strong aggregation for peptide 4. All peptides were dissolved at a concentration of 500 μM , stirred, centrifuged to remove insoluble particles and the supernatant was immediately analysed by far-UV CD. Peptide 1 was easily dissolved at pH 2 and, in agreement with TANGO, does not aggregate, as shown by the CD spectrum taken at 50 μM and 500 μM (Fig. 1.2 A).

The CD spectra at both concentrations correspond to a random coil conformation. Regarding the C-terminal peptides, we found that peptides 2 and 4 had very low solubility (less than 10 μM), while peptide 3 was quite soluble (around 200 μM). CD analysis of the soluble fraction of peptides 2 and 4 showed a random coil spectrum for the former and a β -sheet spectrum for the latter (Fig. 1.2 A). The CD spectrum of peptide 3 corresponds to a clear random coil conformation and no quantitative difference was found between concentrations of 50 μM and 500 μM , although a slightly red-shifted minimum is observed in the concentrated sample (Fig. 1.2 A). This again agrees very well with the TANGO prediction profile (Fig. 1.1 A). Electron microscopy analysis shows fibrils of amyloid-like morphology for peptide 4 (Fig. 1.2 B), thus indicating that aggregation in this case leads to amyloid fibril formation.

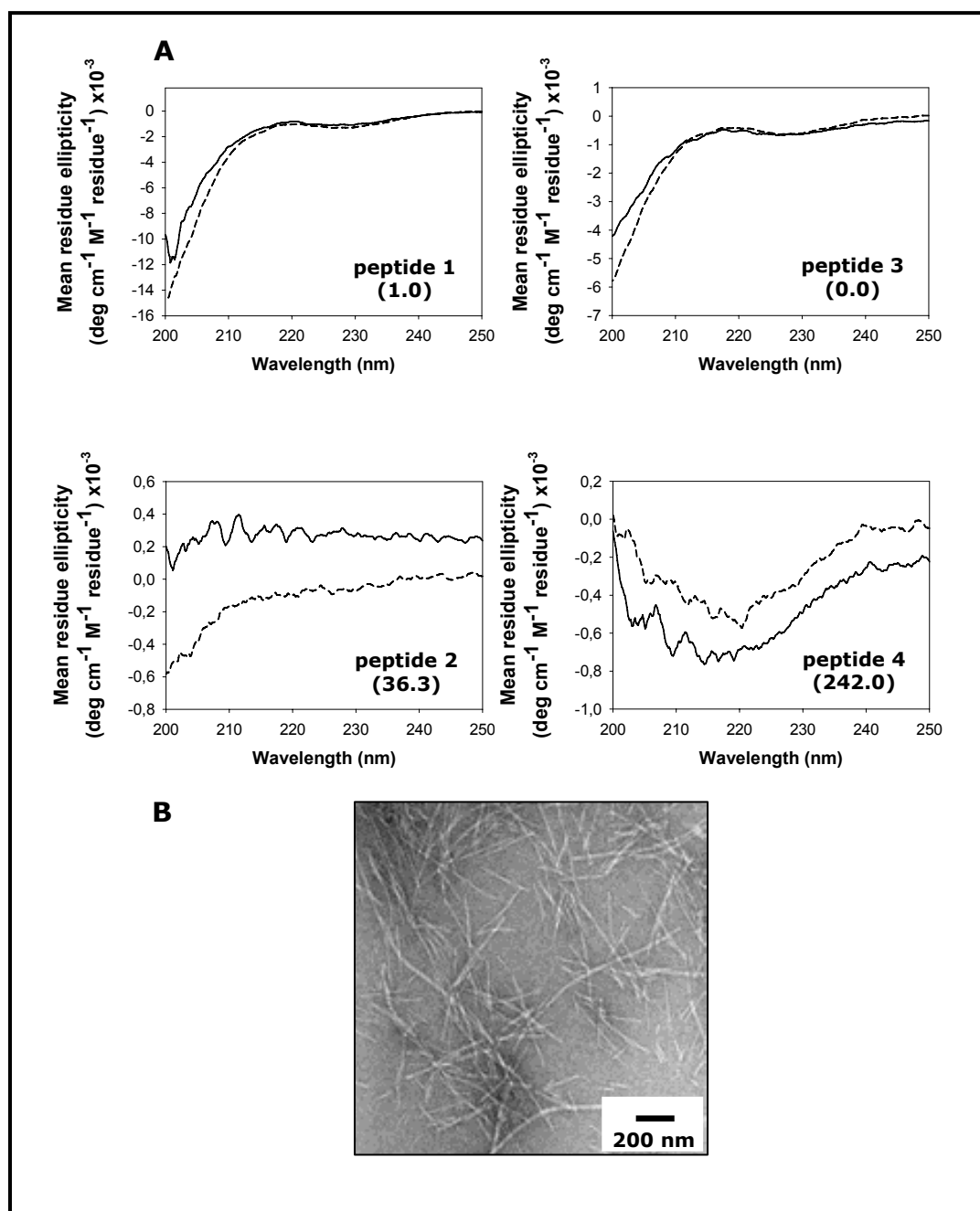


Figure 1.2: Peptide analyses. (A) far-UV CD-spectra of peptides 1 to 4 at pH 2. The values of TANGO prediction are indicated within parentheses. Continuous line, diluted peptide (50 μM for peptide 1); Dashed line, concentrated peptide (500 μM for peptide 1). (B) Electron micrography of peptide 4 under aggregating conditions at 44 K.

1.4.2 Selection of point mutations to study the aggregation properties of ADA2h

In a previous study to analyse the folding pathway of ADA2h, we generated a set of nondisruptive mutants (V12A, E14A, I15V, E20G, I23V, L26V, A31G, D38A, A50G, H51A, V52A, N58A, Q60G, V64G, F65A, I71V and I75A) (Villegas *et al.*, 1998). In order to better delineate the regions responsible for the aggregation of ADA2h we have here designed with TANGO a new set of mutants predicted to increase (T6V, K63A, Y73F and D79A), to decrease (E67K, Q69E, S74K, E78A, E78K and D79K) or not change (Q11K) the aggregation tendency of ADA2h. Table 1.1 shows the predicted value of aggregation and its normalisation by dividing by the WT value (6.62) of all of the mutants presented here. All of these mutants, with the exception of K63A, could be expressed and purified. Mutant K63A, predicted as the most aggregation prone by TANGO, could not be obtained in its soluble form because of its extremely high propensity to aggregate under all of the conditions assayed.

1.4.3 Uncoupling between stability and aggregation

Under conditions in which a protein is unfolded, Chiti and co-workers have shown that there is no direct correlation between protein stability and amyloid aggregation (Chiti *et al.*, 2003); thus, mutants that destabilize the protein do not necessarily induce aggregation. To find if this is the case for our protein model, we have analysed under equilibrium conditions the aggregation state of the WT protein and its variants by CD at two different concentrations. The conditions used are those described to induce amyloid formation in the WT protein (Villegas *et al.*, 2000). Fig. 1.3 shows an example comparing the behaviour of the WT and the destabilised mutant V64G (the rest of the mutants are not shown). The CD spectrum of the WT at pH 3 shows a β -conformation that increases at higher concentration as described (Villegas *et al.*, 2000). On the other hand, the destabilised mutant V64G has a native-like α + β conformation at pH 3 and 20 μ M, and requires a rather high concentration to adopt a pure β -conformation. This demonstrates that there is no direct correlation between destabilisation and the ability to form amyloid fibrils, as shown for acylphosphatase (AcP) (Chiti *et al.*, 2003).

Table 1.1: Predicted value of aggregation and its normalisation by dividing by the WT value of all of the mutants presented in this work. From these values three groups of predicted aggregation rates can be made: those aggregating equal to the WT (no label), those showing a decreased tendency (*italic*) and those showing an increased tendency (**bold**). (^a) This mutant was not obtained because of its high propensity to aggregate.

| VARIANT | TANGO | TANGO/WT |
|-------------------------|-------|----------|
| WT | 6.62 | 1.00 |
| T6V | 7.94 | 1.20 |
| Q11K | 6.62 | 1.00 |
| V12A | 6.62 | 1.00 |
| E14A | 8.37 | 1.26 |
| I15V | 6.61 | 1.00 |
| E20G | 6.61 | 1.00 |
| I23V | 6.61 | 1.00 |
| L26V | 6.61 | 1.00 |
| A31G | 6.61 | 1.00 |
| D38A | 6.61 | 1.00 |
| A50G | 6.61 | 1.00 |
| H51A | 6.61 | 1.00 |
| V52A | 6.61 | 1.00 |
| N58A | 9.86 | 1.49 |
| Q60G | 6.86 | 1.04 |
| K63A^a | 18.07 | 2.73 |
| <i>V64G</i> | 6.41 | 0.97 |
| <i>F65A</i> | 6.42 | 0.97 |
| <i>E67K</i> | 6.38 | 0.96 |
| <i>Q69E</i> | 6.48 | 0.98 |
| I71V | 7.08 | 1.07 |
| Y73F | 8.25 | 1.25 |
| <i>S74K</i> | 0.55 | 0.08 |
| I75A | 6.61 | 1.00 |
| <i>E78A</i> | 4.55 | 0.69 |
| <i>E78K</i> | 1.2 | 0.18 |
| D79A | 9.63 | 1.45 |
| <i>D79K</i> | 4.69 | 0.71 |

1.4.4 Experimental conditions to follow the aggregation process

Although interesting, the above set of experiments does not allow us to measure the kinetics of aggregation, since the WT protein is initially in the native state, while some of the destabilised mutants could be partly unfolded under the same conditions. Thus, we looked for conditions where all of the mutants and the WT were denatured to the same extent before initiating the aggregation process. To do so, we used the WT and the more destabilised mutant, as measured under native conditions (I75A) (Villegas *et al.*, 1998). After exploring several combinations of temperature and denaturants, we found that in the presence of 4 M urea and 60 °C, at pH 7, both variants were denatured to the same extent (Fig. 1.4); this was also the case for all of the mutants used in this work. Thus, we decided to use these conditions to compare the aggregation kinetics of the different proteins.

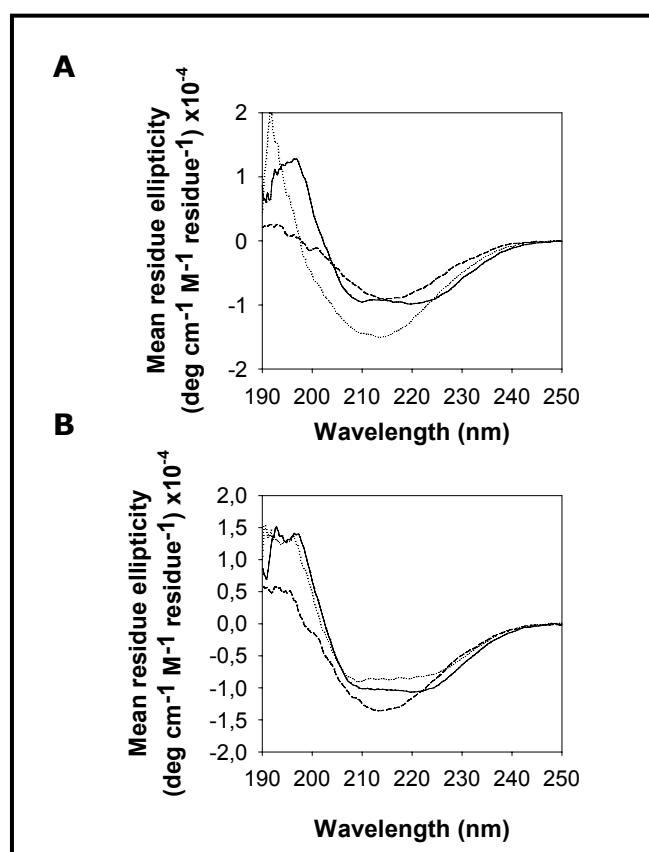


Figure 1.3: Comparison between the behaviour of the WT and the destabilised mutant V64G at different pHs and concentrations. (A) WT. (B) V64G. The aggregation was drastically induced from the native state (see materials and methods section for details). Continuous line, 20 μM of protein pH 7.0; dotted line, 20 μM of protein pH 3.0; broken line, 200 μM of protein pH 3.0.

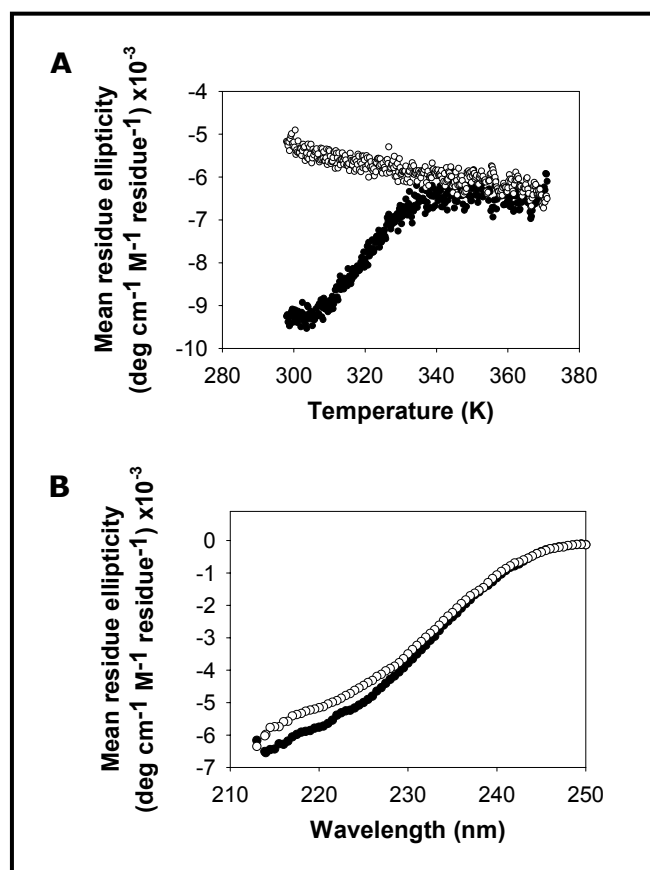


Figure 1.4: Conditions to denature variants to the same extent. The WT (filled circles) and the more destabilised mutant in native conditions, I75A (open circles), were used. (A) Thermal denaturation at pH 7.0, 40 μ M of protein, and 4 M urea. (B) Initial spectrum at pH 7.0, 40 μ M of protein, 4 M urea and 60 $^{\circ}$ C.

1.4.5 Kinetic experiments

Stopped-flow measurements allowed us to follow the very initial steps of amyloid aggregation by pH jump (Fig. 1.5).

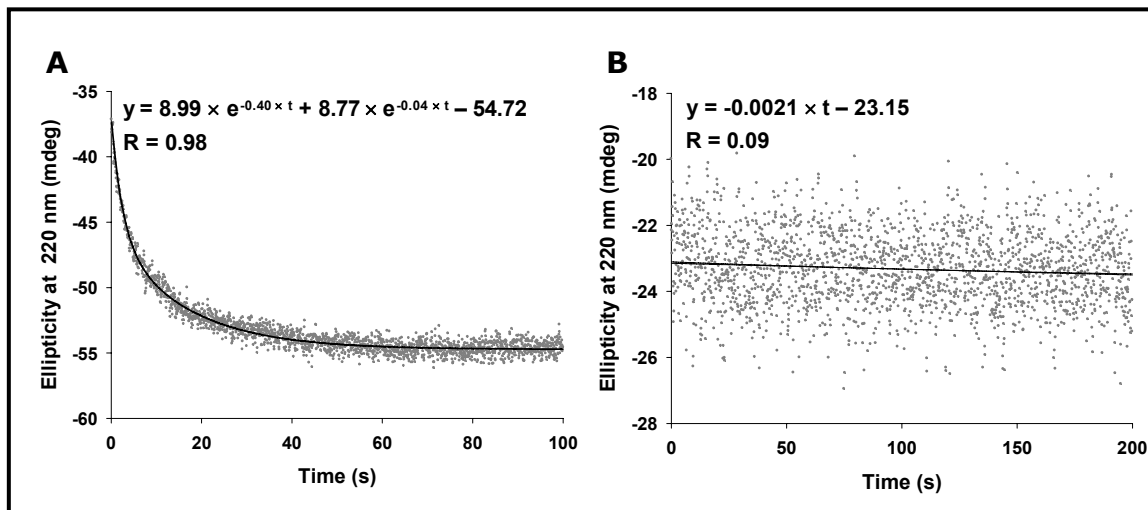


Figure 1.5: Examples of stopped-flow measurements fitted to a double exponential. (A) WT stopped-flow kinetics. (B) V64G stopped-flow kinetics. V64G was fitted to a single exponential due to its inability to aggregate.

The kinetics of the mutants that fold into the β -conformation were fitted to a double-exponential equation, and the six independent experiments of each mutant were averaged (Table 1.2).

The occurrence of two rate constants indicates different populations of molecules. When comparing the rate constants obtained at 40 μM and 60 μM protein for a rapid aggregating mutant (N58A) and a slow mutant (I23V), k_1 increases with concentration (around 40%), while k_2 does not change (Table 1.3).

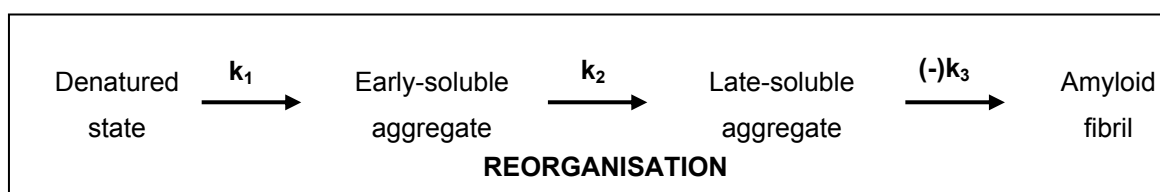
Table 1.2: Fitting of the kinetics obtained by stopped-flow. The mutants that fold into β -conformation after the pH jump are fitted to a double exponential equation with a $p < 0.0001$. NA, no clear aggregation was observed and a minor kinetics was fitted to a linear equation due to its extremely slow decrease.

| VARIANT | k_1 | STANDARD | k_2 | STANDARD |
|-------------|--------|----------|--------|----------|
| WT | 0.4023 | 0.0523 | 0.0394 | 0.0155 |
| T6V | 0.4537 | 0.0746 | 0.0227 | 0.0126 |
| Q11K | 0.0580 | 0.0054 | 0.0053 | 0.0009 |
| V12A | 0.0597 | 0.0167 | 0.0058 | 0.0015 |
| E14A | 0.4600 | 0.0218 | 0.0653 | 0.0015 |
| I15V | 0.1252 | 0.0151 | 0.0072 | 0.0005 |
| E20G | 0.1127 | 0.0103 | 0.0093 | 0.0010 |
| I23V | 0.1395 | 0.0269 | 0.0094 | 0.0010 |
| L26V | 0.1483 | 0.0353 | 0.0103 | 0.0013 |
| A31G | 0.2902 | 0.0495 | 0.0314 | 0.0050 |
| D38A | 0.0992 | 0.0381 | 0.0224 | 0.0024 |
| A50G | 0.1261 | 0.0128 | 0.0085 | 0.0010 |
| H51A | 0.2224 | 0.0564 | 0.0180 | 0.0145 |
| V52A | 0.1373 | 0.0225 | 0.0083 | 0.0005 |
| N58A | 0.8624 | 0.1399 | 0.0529 | 0.0185 |
| Q60G | 0.0239 | 0.0062 | 0.0034 | 0.0006 |
| V64G | 0.0021 | 0.0426 | | NA |
| F65A | 0.0010 | 0.0555 | | NA |
| E67K | 0.0017 | 0.0462 | | NA |
| Q69E | 0.3026 | 0.0375 | 0.0293 | 0.0048 |
| I71V | 0.0708 | 0.0132 | 0.0222 | 0.0055 |
| Y73F | 1.2936 | 0.6584 | 0.0355 | 0.0043 |
| S74K | 0.0027 | 0.0211 | | NA |
| I75A | 0.0031 | 0.0153 | | NA |
| E78A | 0.0561 | 0.0106 | 0.0067 | 0.0005 |
| E78K | 0.0013 | 0.0244 | | NA |
| D79A | 0.4854 | 0.0636 | 0.0297 | 0.0036 |
| D79K | 0.0023 | 0.0285 | | NA |

Table 1.3: Comparison of k_1 and k_2 at different concentrations. The rate constants of a fast (N58A) and a slow mutant (I23V) were compared in order to elucidate its dependence on concentration.

| mutant | k_1 40 μM | k_1 60 μM | k_1 difference (%) | k_2 40 μM | k_2 60 μM | k_2 difference (%) |
|-------------|------------------------|------------------------|----------------------|------------------------|------------------------|----------------------|
| N58A | 0.8624 | 1.1854 | 37 | 0.0529 | 0.0501 | 5.3 |
| I23V | 0.1395 | 0.2041 | 46 | 0.0094 | 0.0107 | 13.8 |

Unfortunately, a careful analysis of the concentration-dependence of $k_{1\text{mut}}/k_{1\text{wt}}$ could not be performed because concentrations of protein higher than 40 μM rendered kinetics too fast to be measured accurately, whereas at lower concentration the transition to the β -conformation was not clearly produced. In any case, k_1 is dependent on protein concentration and constitutes the formation of a polymeric ensemble, whereas k_2 is independent of protein concentration and constitutes a conformational reorganisation within the aggregate. When allowing the kinetics to continue for several hours, a loss of signal is observed with a rate constant k_3 , which could reflect a further, very slow conformational rearrangement or the formation of highly insoluble material like fibrils. The occurrence of k_3 has been measured for the WT by mass spectrometry, as well as the resistance to proteolysis of the very end product (Villanueva *et al.*, 2003). From the data published, an estimation of $0.16 \times 10^{-3} \text{ s}^{-1}$ for k_3 could be extracted. This extremely slow kinetics cannot be measured simultaneously with the apparatus used in this study, with k_1 (0.40 s^{-1}) and k_2 (0.04 s^{-1}). Thus, the more likely interpretation is as follows (Scheme 1):



In the case of the non-aggregating mutants (V64G, F65A, E67K, S74K, I75A, E78K and D79K), the CD spectrum after the pH-jump is not that of a pure β -conformation and the amplitude of the measured kinetics is extremely small in the time scale of the experiment (around 5 mdeg). Thus, k_1 , obtained by fitting the data to a linear equation, must be regarded as an estimate rather than as an accurate determination of the kinetics. It is important to note that for these mutants, whenever the conditions are forced (i.e. concentration of protein), a final β -conformation is found (see Fig. 1.3 for V64G).

1.4.6 Algorithm-based predictions

The ratios of aggregation tendency for each mutant predicted by TANGO/aggregation tendency of the WT (Fernández-Escamilla *et al.*, 2004; Linding *et al.*, 2004) are correlated with the natural logarithm of the ratio $k_{1 \text{ mutant}}/k_{1 \text{ WT}}$. In Fig. 1.6 A, all the mutants assayed in this work are included in the analyses and $r = 0.65$ is obtained. This correlation does not take into consideration the mutant K63A, which is predicted by TANGO to be the strongest aggregating mutant, and that could not be analysed due to its high tendency to aggregate. The mutants that are out of the correlation are mainly V64G, F65A, E67K and I75A; four out of seven non-aggregating mutants. However, as mentioned before, k_1 of non-aggregating mutants must be regarded as an estimate, since the signal-to-noise ratio is too small to allow accurate measurements. There is a large amount of mutants that are predicted to have the same tendency to aggregate as the WT but experimentally they have a reduced aggregation. In the case of mutants predicted to have a tendency to aggregate different from that of WT the correlation improves up to $r = 0.90$ (Fig. 1.6 B).

Another prediction can be obtained when using the formula developed by Dobson and co-workers (Chiti *et al.*, 2003). The formula is based on (A) hydrophobicity, (B) α and β propensities, and (C) charge constants (see Materials and Methods in this chapter for more detailed information). In Fig. 1.6 C, all the mutants assayed in this work are included in the correlation and $r = 0.57$ is obtained. When only mutants predicted by TANGO to have a tendency to aggregate different from that of the WT are included, the correlation drops to $r = 0.54$ (Fig. 1.6 D).

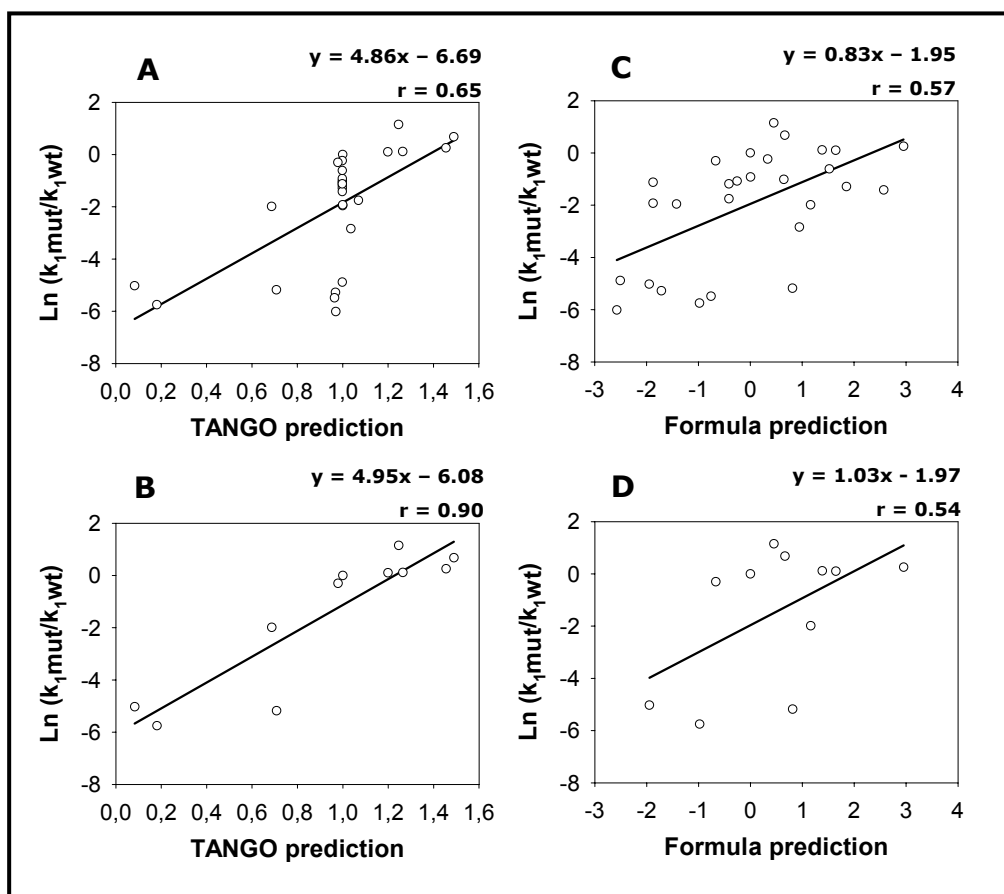


Figure 1.6: Correlation between experimental k_1 and predicted tendencies by TANGO and equation by Dobson (A) Correlation for all of the mutants used in this work with TANGO prediction. (B) Correlation for these variants with predicted tendencies clearly different from that of the WT (T6V, E14A, N58A, Q69E, Y73F, S74K, E78A, E78K, D79A, D79K and WT) with TANGO prediction. (C) Correlation for all of the mutants used in this work with the formula. (D) Correlation for these variants with TANGO predicted taxes clearly different from that of the WT with the formula.

An implemented formula has been published recently, in which extrinsic factors such as pH, ionic strength and concentration are taken into account (DuBay *et al.*, 2004). Although fairly valuable for determining a priori the rate of a process, with all its biomedical and biotechnological implications, in the case of the study presented here, these extrinsic factors are equal for all of the experimental data, and therefore this formula has not been applied.

A third algorithm-based prediction, also based on the formula mentioned above (Chiti *et al.*, 2003) but implemented with hydrophobic patterning, has been published (Pawar *et al.*, 2005). However, this version is not available on line and we have used a modified version provided by Zyentia Ltd. (ZYGREGATOR™). Two parameters, Z_{agg} of each mutant/ Z_{agg} WT and absolute rate of each mutant/absolute rate of WT have been correlated with our experimental data (Fig. 1.7).

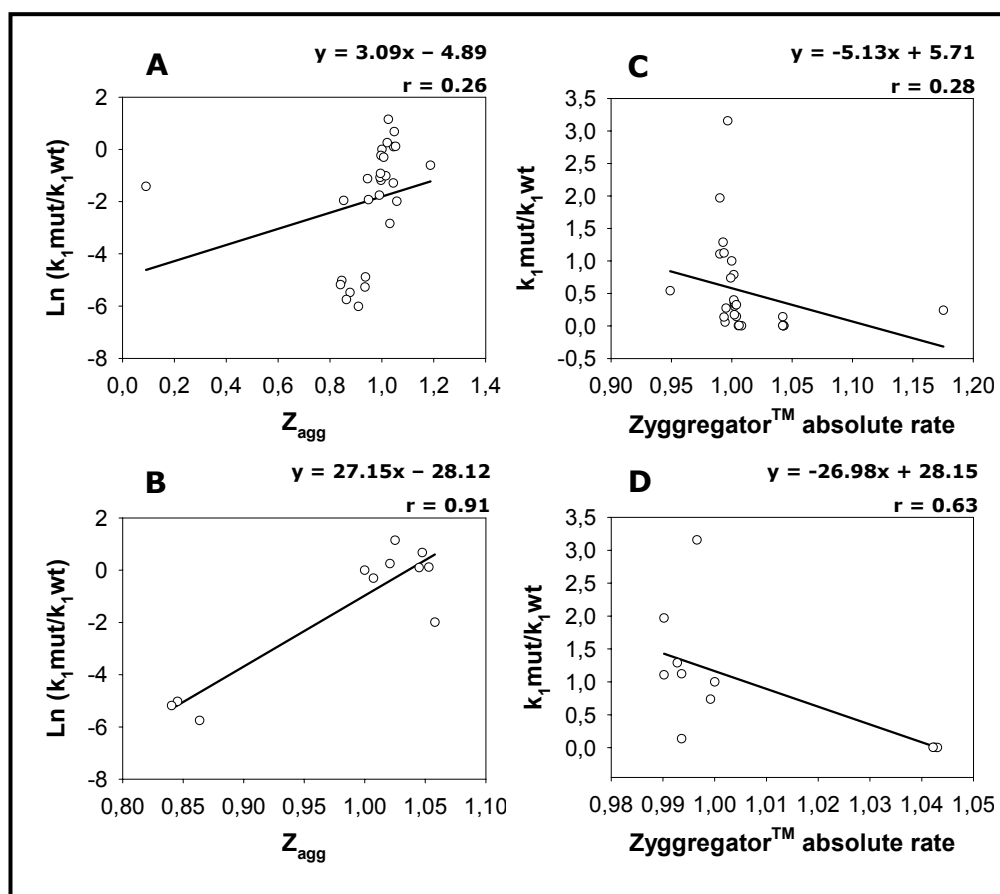


Figure 1.7: Correlation between predicted tendencies by ZYGREGATOR™ (ZYENTIA Ltd.) and experimental k_1 . (A) Correlation for all of the mutants used in this work with Z-score. (B) Correlation with Z-score for these variants with TANGO predicted taxes clearly different from that of the WT (T6V, E14A, N58A, Q69E, Y73F, S74K, E78A, E78K, D79A, D79K and WT). (C) (D) similar to A and B, but using the absolute rates of aggregation predicted by the algorithm.

For Z_{agg} the correlations are 0.26 and 0.91, when using all the mutants and those different from the WT, respectively (Fig. 1.7 A and B). The absolute rates correlate with $r = 0.28$ and $r = 0.63$ (Fig. 1.7 C and D). It is important to note that Zyentia Ltd. have modified the original algorithm for their own purposes, and that the aggregation propensity profile obtained as described following the instructions in the original publication should be different (Pawar *et al.*, 2005). Thus, the best correlation for our experimental data is obtained when using the algorithm TANGO, rather than with other approximations. This is seen clearly when comparing the profiles predicted by TANGO and by ZYGREGATOR™ (Zyentia Ltd.) (Fig. 1.8). These profiles are quite different, especially in the region encompassing $\beta 2$. Interestingly, TANGO was developed and calibrated with aggregation data obtained under equilibrium conditions, while ZYGREGATOR™ was developed to predict aggregation kinetics. Since we are focusing on the initial phase of aggregation, it is possible that the correlation by

ZYGGREGATOR™ shall improve when the overall kinetics (especially $(-k_3)$, not measured in this work) are taken into account.

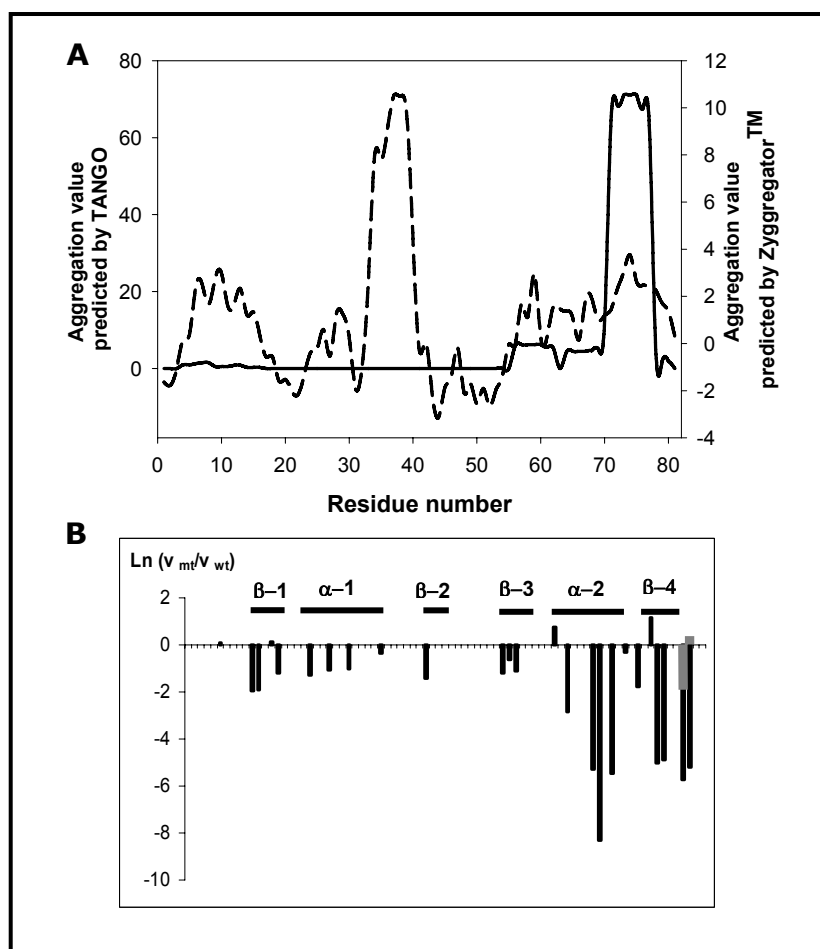


Figure 1.8: Comparison of the profiles predicted by TANGO and ZYGGREGATOR™ (ZYENTIA Ltd.) and experimental profile of ADA2h. (A) Prediction by the algorithms. Continuous line, TANGO. Dashed line, ZYGGREGATOR™ (ZYENTIA Ltd). (B) Experimental kinetics of amyloid fibril formation. All the mutants have been included in the profiles and in consequence those positions where we have two mutants have an additional bar in grey (E78K, D79K).

1.4.7 β-Sheet aggregation and amyloid aggregation

The ability of ADA2h to form amyloid fibrils was shown earlier (Villegas *et al.*, 2000). However, in the present work we initiated the kinetics in the presence of 4 M urea, which could interfere with the pathway of amyloid formation. We designed experimental conditions to differentiate β-sheet aggregation from amyloid aggregation. Three samples of 40 μM ADA2h were prepared and treated in different ways: Sample A contains 4 M urea and was incubated for 30 min at 60°C; sample B is sample A, but

without urea; sample C is sample B but treated at 90°C. In all the cases, amyloid formation was induced using the same protocol as for the kinetic experiments. A sample containing urea was not incubated at 90°C for 30 min because of the possibility of having strong carbamylation of the sample. Fig. 1.9 shows the electron micrographs (EM) of these samples obtained at different times. At time zero (right after lowering pH), little aggregated material is visible in the EM samples. In all three cases there is abundant amyloid material formed after the 30 min of heating. However, the samples exhibit slight differences in behaviour depending on the experimental conditions. At 60°C (with or without urea) a large amount of amyloid material covering the EM grid homogeneously can be found after heating for 30 min. However, in the sample treated at 90°C (sample C), there appear to be fewer fibrils at the same timepoint. In the samples without urea, fibrils seem to cluster with time, forming clumps after incubation for three days. The CD spectra of these samples at different time-points also indicate that: (i) urea seems to help in quickly obtaining more defined β -sheet spectra, although it cannot be excluded that the lower definition of the spectra in the absence of urea results from light-scattering problems generated by the clustering of the fibrils; and (ii) incubation at 60°C is more efficient than at 90°C (data not shown). From these results, it is clear that ADA2h forms amyloid fibrils under the conditions used in this work. It has been demonstrated that aggregation and amyloid formation are not coupled directly (Rousseau *et al.*, 2006a; Rousseau *et al.*, 2006b). Amyloid formation requires β -sheet aggregation, but not all β -sheet aggregates lead to fibre formation. In the case of ADA2h, we see initial formation of β -sheet aggregates that later result in fibril formation. EM analysis of peptide 4 (Fig. 1.2 B) shows fibre formation, suggesting that the strongly aggregating region at the C terminus of the protein could also be the region involved in amyloid formation.

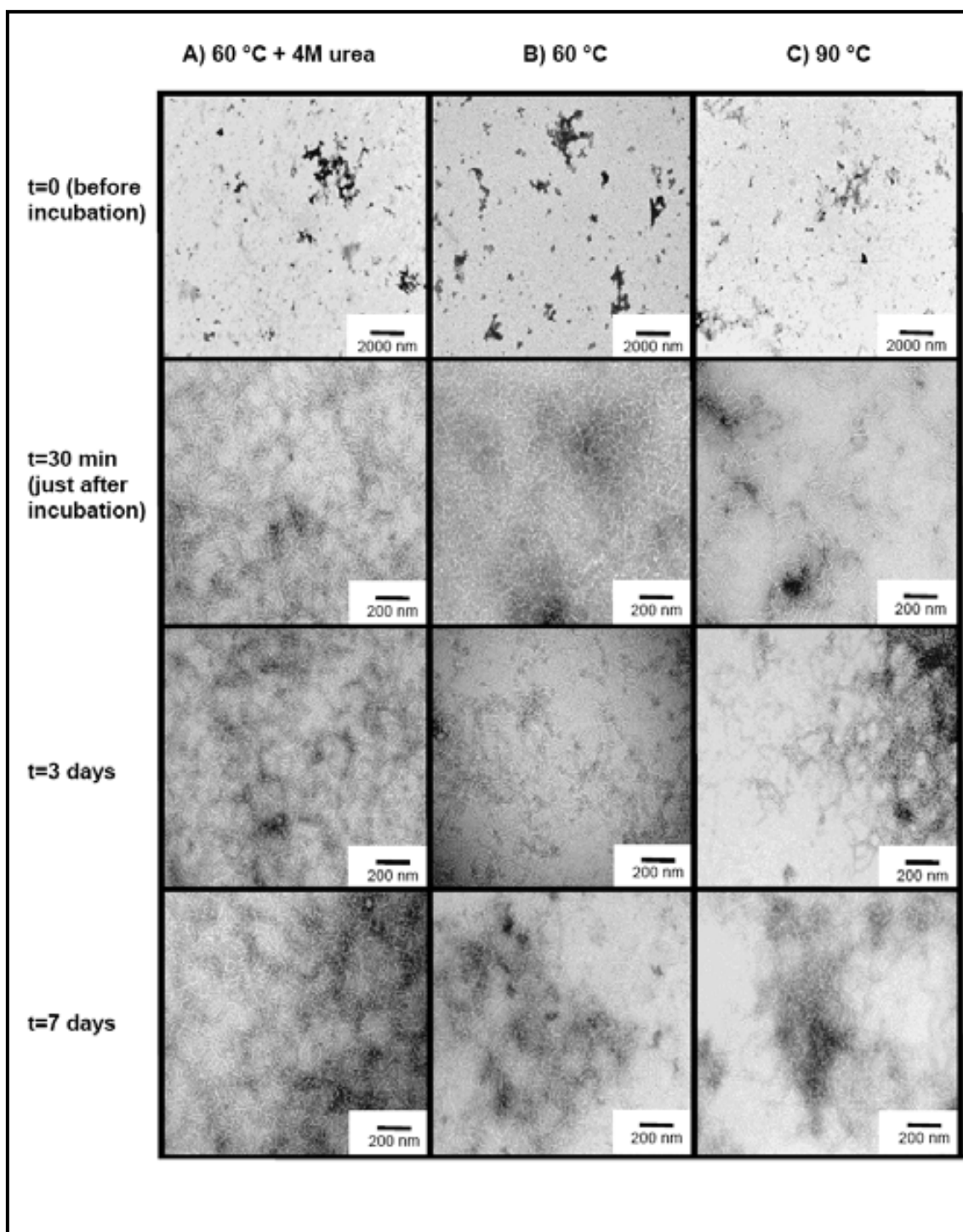


Figure 1.9: Time course of ADA2h fibril formation. Electron micrographs were taken at different times of incubation (0, 30 min, 3 days, and 7 days). (A) Sample with 4 M urea incubated at 60 °C for 30 min. (B) Sample without urea incubated at 60 °C for 30 min. (C) Sample without urea incubated at 90 °C for 30 min.

1.5 DISCUSSION

1.5.1 Complex kinetics in the initial conformational transition

We have followed the very early kinetics of β -acquisition from the denatured state of ADA2h by stopped-flow when amyloid aggregation is induced by pH-jump. The double-exponential nature of the fitting of the data for all the mutants prone to aggregation indicates multiple aggregated states. Indeed, the simple explanation for the apparent first-order kinetics as well as concentration-dependence observed here may reflect the nucleation-dependent reaction but, in such a case, there should be a very rapid nucleation event at the burst phase of kinetic measurement followed by the attachment of monomers to the nuclei. Thus, k_1 is the rate constant for the transition from the random coil ensemble to an early aggregate nucleated at the burst phase, whereas k_2 should correspond to a conformational rearrangement of the aggregate (Fig. 1.10).

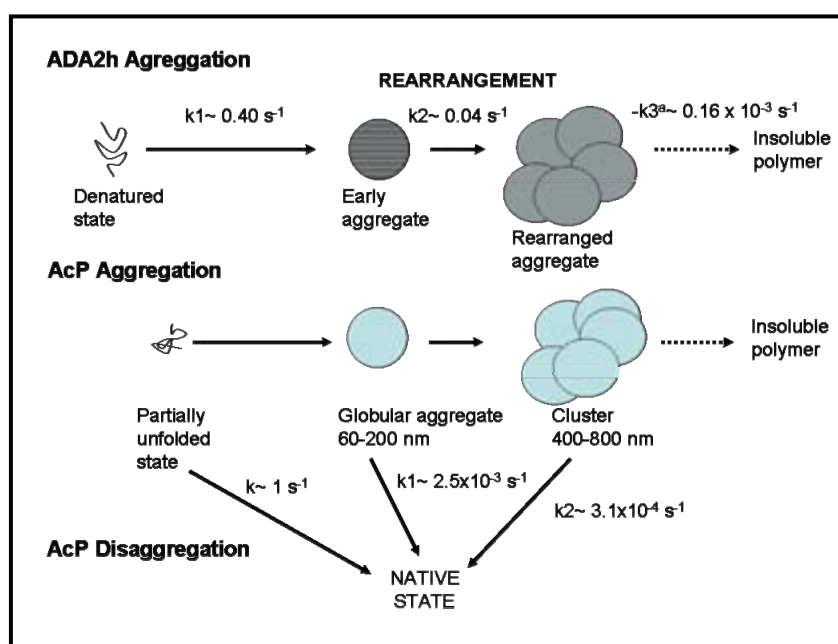


Figure 1.10: A representation of the aggregation for ADA2h and comparison with the data for AcP. (A) Aggregation process of ADA2h followed by stopped-flow measurements. (B) Aggregation process of AcP deduced from disaggregation experiments (partially reproduced from Calamai *et al.* (2005).

Our hypothesis is corroborated by very accurate data reported recently by Dobson *et al.*, who found multiple aggregated states by reverse protein aggregation (Calamai *et al.*, 2005). They studied the disaggregation process and determined the rate constants for the transition to the native state from different aggregated states (Fig.

1.10). It is noteworthy that they found a double-exponential process with rate constants of $2.9 (\pm 0.5) \times 10^{-3} \text{ s}^{-1}$, and $3.4 (\pm 0.5) \times 10^{-4} \text{ s}^{-1}$, for AcP, a protein topologically similar to ADA2h (see below). In this work, WT ADA2h aggregates with rate constants of $0.4 (\pm 0.05) \text{ s}^{-1}$, and $0.04 (\pm 0.01) \text{ s}^{-1}$, where k_1 is the fast phase and k_2 is the slow phase. They postulated that the aggregation of AcP from the partially unfolded state goes through an initial globular aggregate of 60 – 200 nm that further assembles into clusters of 400 – 800 nm and, in turn, assemble into larger structures. It is possible that the globular aggregate and the clusters are analogous to our ensembles of early-soluble aggregate and late-soluble aggregate. Unfortunately, ADA2h aggregation kinetics is so fast that a deep structural characterisation of the early kinetic ensembles involved in ADA2h amyloid aggregation is not possible in the time scale of the experiments. The rearrangement we are presenting for the soluble aggregates that originated from the denatured state of ADA2h is similar to that proposed for AcP and other homologs, where aggregation from the folded state is found (Plakoutsi *et al.*, 2004; Plakoutsi *et al.*, 2005). The main difference is that they proposed that the reorganisation takes place within the native-like aggregates, while we find it in the non-native-like aggregates. Thus, it is possible that conformational rearrangement during the process driving to aggregation is a general phenomenon for all of the ensembles involved.

1.5.2 Stability and aggregation

It has been suggested that amino acid replacements that destabilise the native state of a protein favour aggregation (Hurle *et al.*, 1994; McCutchen *et al.*, 1995; Booth *et al.*, 1997; Chiti *et al.*, 2000; Ramírez-Alvarado *et al.*, 2000; Siepen & Westhead, 2002; Smith *et al.*, 2003). It has been demonstrated also that increasing the stability of the native state is a strategy to avoid amyloid fibril formation (Villegas *et al.*, 2000). However, the intrinsic tendency of a protein sequence to aggregate is another factor that affects the aggregation process. Natively unfolded proteins do not aggregate because their sequences have been selected to avoid aggregation (Rousseau *et al.*, 2006a; Rousseau *et al.*, 2006b) It is the interplay between stability and the tendency of the sequence to aggregate that determines the aggregation outcome for a protein. In this respect, there should not be a direct relationship between the destabilising effect of a mutation and the aggregation of the protein as shown by (Chiti *et al.*, 2003) and by us in this work. Also, mutations that affect the kinetics of refolding and/or unfolding do not necessarily need to correlate with aggregation, since the same mutation could also

affect the tendency of the sequence to aggregate. Thus, in order to obtain the sequence contribution to aggregation one needs to start from the denatured state so that protein stability is eliminated as a factor.

1.5.3 Kinetics of aggregation and folding

A simple kinetics is observed by Fourier transform infrared spectroscopy and by light-scattering when raising the pH to induce ordered aggregates of BSA (Militello *et al.*, 2004). This is the case also for the 2,2,2-trifluoroethanol-induced aggregation of AcP followed by the increase of fluorescence when binding thioflavin T, Congo red or by CD spectroscopy (Chiti *et al.*, 2002). The rate constant for the WT AcP aggregation is around $9.3 \times 10^{-5} \text{ s}^{-1}$, whereas that for ADA2h is around 0.4 s^{-1} , as estimated from the accurate stopped-flow measurements in this work. That means that AcP aggregates in ten to 11 days and ADA2h in 2 s. The refolding of ADA2h driving to the native state is reported as one of the most rapid for proteins, some 0.001 s in water (Villegas *et al.*, 1995), whereas AcP is the slowest, folding some 4000 times more slowly (Chiti *et al.*, 1999). Baker and colleagues have suggested that the contact order (the average distance between residues forming an interaction in the native state normalized to the total number of residues) is a major determinant of the folding rate (Plaxco *et al.*, 1998). Inverse linear correlation was found when comparing the overall contact order of ADA2h, and other proteins displaying the same topology (AcP, U1A, HPr, and MerP) with the natural logarithm of their folding rate constants (Chiti *et al.*, 1999). This was not the case when studying HypF-N, a homologue of AcP that folds some two to three orders of magnitude faster, where the hydrophobic content was considered to be the main contributor to the rapid kinetics of folding (Calloni *et al.*, 2003). ADA2h and AcP, although sharing only 13% sequence identity, are structurally very similar in terms of regions involved in nucleating aggregation (this work and Chiti *et al.*, 2002) and in nucleating the transition state (Villegas *et al.*, 1998; Chiti *et al.*, 2003). Both molecules also aggregate mainly by C-terminal β -strands, and the extent of structure in the transition state is quite comparable. Precisely, the structure of ADA2h most susceptible to aggregation is strand β_4 , an edge strand that is not nucleating the structure of the transition state (Villegas *et al.*, 1998). This is also in concordance with kinetic partitioning (Chiti *et al.*, 2002). However, this does not seem to be a universal case for this fold, since ADA2h, AcP, spliceosomal U1A, HPr, MerP and HypF-N, folds that show similar topology when submitted to TANGO do not behave the same way (Chiti *et al.*, 1999; Plakoutsi *et al.*, 2005). Moreover, ADA1h, a close homologue (paralogue) to

ADA2h, has no tendency to aggregation at any part of its sequence according to TANGO prediction, in agreement with the fact that the experimental trials under a wide range of conditions have failed to obtain the amyloid form. Aggregation by the C terminus is not a general rule in all six sequences sharing topology (Fig. 1.11). It is important to note that AcP is not predicted to aggregate by its C terminus because the prediction was run at pH 2 (experimental conditions used in this work), whereas experimental characterisation of its aggregation has been performed at pH 5.5 (Chiti *et al.*, 2002). It remains to be seen if the kinetics of aggregation for this protein family follow a direct relationship with their folding kinetics or if the results found for ADA2h and AcP are just a coincidence.

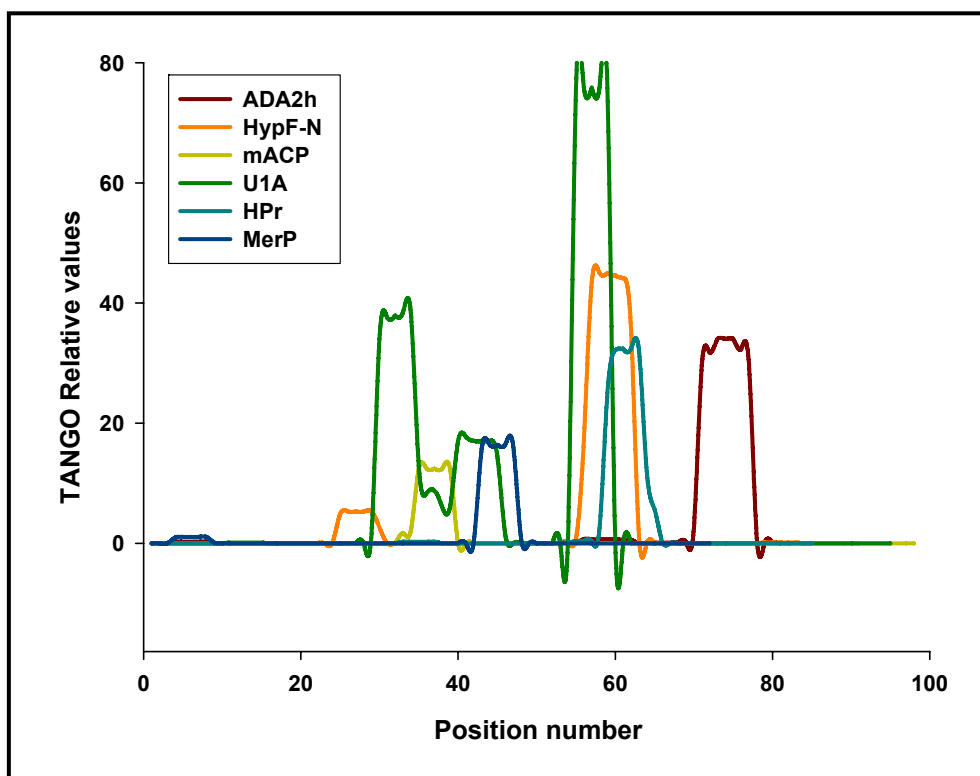


Figure 1.11: Aggregation profile of some proteins topologically similar to ADA2h. Prediction at pH 2.0, 75 mM sodium phosphate, 25 °C. The relative values predicted by TANGO indicate those regions more susceptible to influence within the aggregation rate of the molecule.

1.5.4 Algorithm-based predictions

TANGO predicts correctly the behaviour of all the peptides analysed here. Regarding the behaviour of the mutants and the correlation with TANGO predictions,

we found a very good correlation between mutants predicted to aggregate more, or less, than the WT and the experimental data. This included mutant K63A, which is predicted to be very strongly aggregating and that was fully insoluble. However, TANGO predictions were not good for mutants predicted to have the same tendency to aggregate as the WT. These mutants are all localized mainly in the region encompassing residues 15–52 and, in general, they showed slightly reduced aggregation. This could be due to an underestimation of the aggregation tendency of these sequences by the algorithm. This is reasonable, since the ZYGGREGATOR™ (ZYENTIA Ltd.) profile indicates this region to be the more prone to aggregation, especially β 2, probably overestimating the aggregation tendency of these sequences. Unfortunately, we have only one mutant in β 2, D38A, since previous mutations, F39L, K41A and S42G, were too destabilised to render sufficient amounts of pure mutant proteins (Villegas *et al.*, 1998). Interestingly, TANGO was developed and calibrated with aggregation data obtained under equilibrium conditions, while the approximation by Dobson was developed to predict aggregation kinetics. The fact that TANGO can predict the kinetics of aggregation so well suggests that the aggregation mechanism must be quite general and simple.

**CHAPTER 2:
CHARACTERISATION OF DIFFERENT PROCESSES OF
AMYLOIDOGENESIS BY 2D-FTIR SHOWS CONVERGENCE TO
THE SAME FINAL FIBRIL**

2.1 SUMMARY

Understanding the process of amyloidogenesis is one of the clues to the future cure of misfolding-based diseases, such as Alzheimer's disease, Spongiform Encephalopathies, and other important disorders affecting humans. In this work we have used one of the deepest characterised models for folding and misfolding, the Activation Domain of human procarboxypeptidase A2 (ADA2h). The WT and three mutants affecting the kinetics of aggregation have been studied by IR from the folded state at acidic pD, showing the disappearance of structured features prior to a dramatic increase of the amyloid-characteristic band upon temperature induction. TEM demonstrated that actual amyloid fibrils are formed under the conditions used in this work. The three-step kinetics of the process observed for the WT is clearly affected by the stability of each mutant, although the final state reached is the same. Our conclusion is that this domain nucleates prior to the conformational reorganisation that renders the final amyloid fibril, independently of the aggregation tendency and stability of each variant.

2.2 INTRODUCTION

The process a polypeptide chain undergoes to form amyloid fibrils is of critical interest due to its implications on the conformational diseases referred as amyloidoses. The occurrence of these highly ordered aggregates not only has been observed in disease-related proteins, but also in apparently non-pathogenic species, suggesting that they could represent a different conformational state in all polypeptide chains (Dobson, 2003). Recent investigations have led to the hypothesis that protein aggregation is in direct competition with protein folding, describing a double funnel (Jahn & Radford, 2005), where different sequence stretches are governing each process (Chiti *et al.*, 2002). In short, intramolecular contacts between residues would lead to the formation of the native structure, whereas intermolecular contacts would underlie the acquisition of an aggregated structure, whether ordered (amyloid fibrils) or not (amorphous aggregates) (Jahn & Radford, 2005).

Several studies have been centred on the modulation of amyloid fibril propensity caused by the polypeptide sequence (Chiti *et al.*, 2003; Cerdà-Costa *et al.*, 2007), leading to the creation of prediction formulas or programs which can shed some light on the possible consequences of mutations on proteins (Chiti *et al.*, 2003; Fernández-Escamilla *et al.*, 2004; Pawar *et al.*, 2005). Apart from aggregation propensity, structure conversion through the amyloid formation process is also necessary to complete comprehension of the entire picture. Eisenberg *et al.* (Nelson & Eisenberg, 2006) proposed three general models accounting for monomer structural reorganisation in fibril formation. Whereas two of the conversion models (the refolding and the natively disordered models) involve the reorganisation of the entire polypeptide chain, in the gain-of-interaction model only a small conformational change takes place, exposing a previously hidden surface which will interact with the same surfaces from other monomers.

Two-dimensional correlation analysis (2D-COS) (Harrington *et al.*, 2000; Noda, 2007) coupled to Infrared Spectroscopy (IR) allows a more detailed analysis of the secondary structure conversion from the native monomer to the amyloid fibril. In this technique the different spectra obtained upon application of an external perturbation are correlated in order to increase the resolution and have a better insight in the relationships between bands (Iloro *et al.*, 2004; Iloro *et al.*, 2005).

In the present work, two-dimensional infrared correlational spectroscopy has been applied to the study of the amyloid formation process from the native state of the activation domain of human procarboxypeptidase A2 (ADA2h) and three mutants with

different aggregation propensities, using temperature as the perturbing agent. ADA2h has been used as a model protein for amyloid-like fibril formation at acidic pH (Villegas *et al.*, 2000; Villanueva *et al.*, 2003), as there is a deep understanding of its folding mechanism at neutral pH (Villegas *et al.*, 1995; Villegas *et al.*, 1996; Viguera *et al.*, 1997; Villegas *et al.*, 1998; Fernández *et al.*, 2000; Villanueva *et al.*, 2000; Villanueva *et al.*, 2002). Moreover, the creation of a large battery of single-residue variants has provided detailed information about the amyloid-forming propensity changes upon mutation, and the importance of the sequence context and residual structure (Cerdà-Costa *et al.*, 2007). The most noticeable finding was a conformational reorganisation of the aggregation nucleus in the very early kinetics of the formation of amyloid fibrils by ADA2h from the denatured state.

The temperature-induced aggregation kinetics from the native state at acidic pH for the wild-type (WT) variant, a slowly aggregating mutant (I23V), a rapid aggregating mutant (N58A) and a non-prone to aggregation mutant (F65A) (Cerdà-Costa *et al.*, 2007), followed by IR allows us to shed some light on the conformational differences of the overall process upon mutation. TEM and CD characterisation reinforce the obtained results.

2.3 MATERIALS AND METHODS

2.3.1 Materials

ADA2h variants were expressed into a BL21(DE3) *E.coli* strain using a modified pTZ18U vector (Villegas *et al.*, 1998). Oligonucleotides were synthesized by Roche Molecular Biochemicals. Slide-A-Lyzer[®] 3,500 MWCO Mini Dialysis Units were purchased from Pierce. All other reagents were obtained from Sigma-Aldrich.

2.3.2 Expression and purification

Expression and purification of all variants were realized as described previously (Cerdà-Costa *et al.*, 2007), and the protein obtained was concentrated up to 5 mg/ml (544 μ M) in 50 mM sodium phosphate pH 7.

2.3.3 Infrared analysis

The four variants were dialyzed overnight against 25 mM glycine (pD 3.0) at 4 °C, using Slide-A-Lyzer[®] 3,500 MWCO MINI Dialysis Units and D₂O as a solvent.

The spectra were recorded in a Nicolet 5700 spectrometer (Thermo, Madison, Wisconsin) using a peltier mount and excavated cells with a 50 μ m path (Biotools, Wacounda, Illinois.) The spectra were acquired using the Series software running under OMNIC (Thermo, Madison, Wisconsin). The sample at pD 3 was heated from 20 °C to 90 °C in 10 min, and maintained afterwards at 90 °C for 30 min. Typically, 305 spectra were averaged for each minute. The series obtained was ratioed against a background obtaining the absorbance spectra with a nominal resolution better than 2 cm^{-1} . Data treatment has been described elsewhere (Arrondo & Goñi, 1999).

2.3.4 Transmission Electron Microscopy

100 μ l of all variants were dialysed overnight against 25 mM Glycine pD 3 and heated in the same mode as in the infrared experiments, in order to trigger the

aggregation process. Samples were taken before the heating and at times 30 min and 2 days after the induction of specific aggregation. The samples were diluted 1/20 in MilliQ water and quickly adsorbed to glow-discharged carbon-coated grids, due to the fast dissociation of ADA2h amyloid fibrils. The material was stained using the method described elsewhere (López De La Paz *et al.*, 2002) and used previously with ADA2h (Cerdà-Costa *et al.*, 2007).

The samples were visualised with a Hitachi H-7000 microscope in the Servei de Microscòpia of Universitat Autònoma de Barcelona.

2.3.5 Circular Dichroism

The protein variants were diluted to 20 μM with 50 mM phosphate buffer (pH 7) for the native CD measurements. The same variants were also taken to 2 mg/ml and dialysed overnight against Glycine 25 mM (pH 3) at 4 °C. The dialysed samples were then taken to 200 μM and 20 μM with the dialysis buffer, heated at 95 °C for 5 min, in order to properly activate protein aggregation, and left overnight at room temperature. 0.2 cm and 0.02 cm path length cuvettes were used for the circular dichroism measurements of 20 and 200 μM , respectively, and 20 scans recorded at 100 nm min⁻¹ were averaged for each variant.

To calculate the relative contribution of each secondary structure, spectra were deconvolved by the analysis program CDSSTR from the DICHROWEB server (Lobley *et al.*, 2002; Whitmore & Wallace, 2004) using the reference set SP175 (Lees *et al.*, 2006).

2.4 RESULTS

2.4.1 Assignment of initial infrared bands by the wild-type protein

The structure and fibril formation process of the activation domain of human procarboxypeptidase A2 (ADA2h) has been studied by infrared spectroscopy looking at the changes produced in the amide I' band arising mainly from the C=O stretching vibration of the peptide bond. The 3D plot of the deconvolved spectra during the fibril formation induced by heating ADA2h-WT at acidic pD is shown in Figure 2.1 A. It can be seen that there are two major bands whose intensity increase, located around 1682 and 1619 cm^{-1} , with variations in several bands if the spectra corresponding to different points of the process are compared (see inset). The most important change, which is the appearance of a prominent band at 1619 cm^{-1} , has also been described in other fibril formation processes (Bouchard *et al.*, 2000; Zandomeneghi *et al.*, 2004), as well as in ADA2h at acidic pD (Villegas *et al.*, 2000).

The spectrum before inducing fibril formation (Fig. 2.2 A) shows seven bands at 1682, 1671, 1660, 1647, 1632, 1619 and 1606 cm^{-1} , the latter corresponding to amino acid side chains. The percentage of amide component is shown in Table 2.1. The three-dimensional structure of ADA2h has been solved in its isolated state by NMR (Jiménez *et al.*, 2003) and within the complete proenzyme by crystallography (García-Sáez *et al.*, 1997), and corresponds to an α/β protein. The major infrared band at 1647 cm^{-1} can be attributed to its two α -helices with some component of unordered structure, whereas the 1632 cm^{-1} corresponds to the low frequency component of the antiparallel β -sheet. This component of the antiparallel β -sheet is accompanied by its high frequency counterpart, located at 1682 cm^{-1} . The minor band at 1671 cm^{-1} could be assigned to β -turns. The band at 1660 cm^{-1} which represents a 15% of the amide component has been attributed in literature to coiled coils (Reisdorf & Krimm, 1996), 3_{10} helices (Arrondo & Goñi, 1999) or β -turns (Byler & Susi, 1986) and, taken into account the 3D-structure of ADA2h, it could reflect in part the end of the helices, specially the C-terminal of α -helix 1 that folds as a 3_{10} helix (García-Sáez *et al.*, 1997), together with a component of β -turns (Villegas *et al.*, 2000).

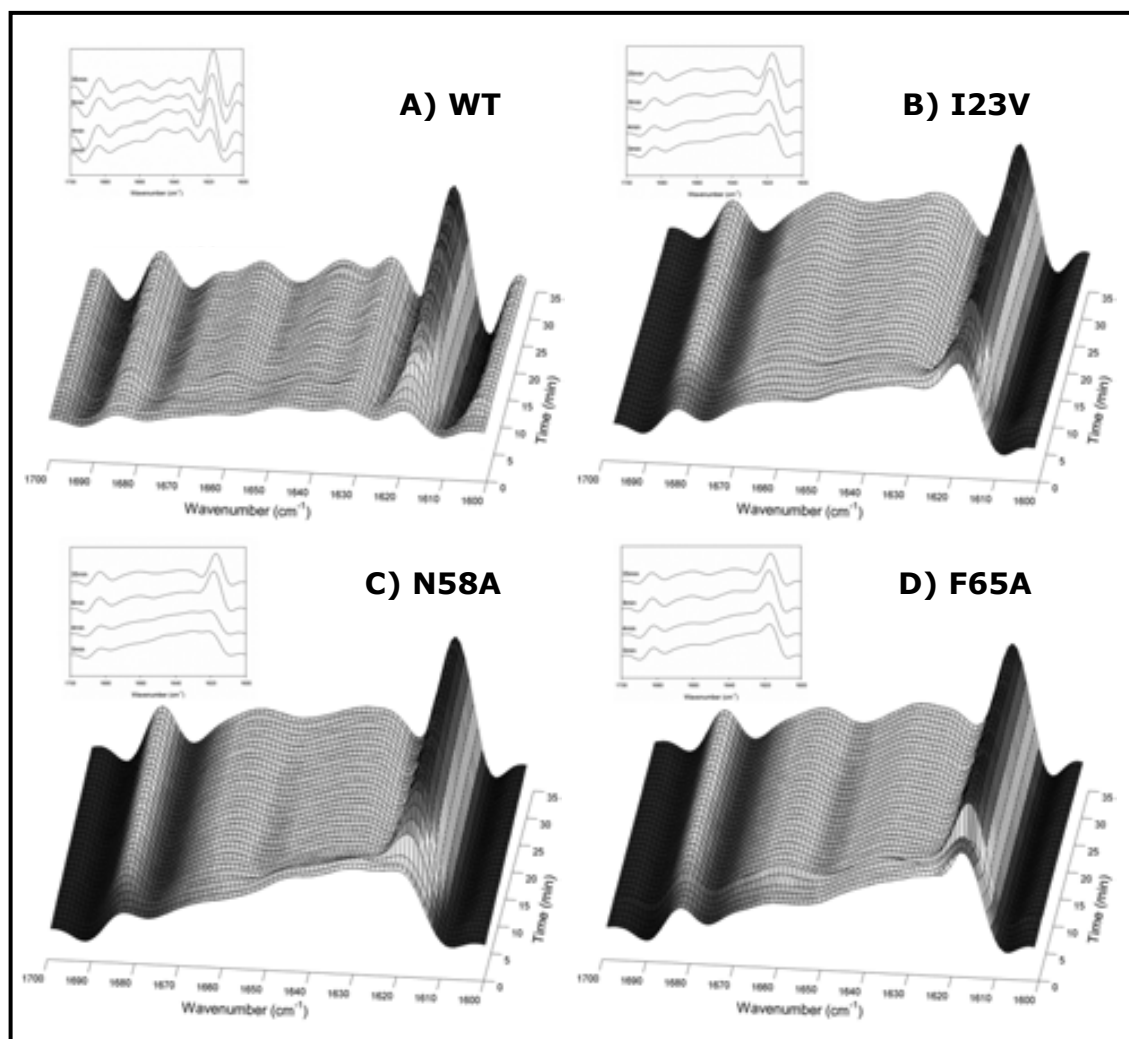


Figure 2.1. Kinetics of the amyloid fibril formation for 4 variants of ADA2h followed by FT-IR. 3-D plots of the evolution of the deconvoluted amide I' spectra of ADA2h WT (A) and the mutants I23V (B), N58A (C) and F65A (D) at acidic pH, using heat as the inducer. The spectra at the most significant times are represented in the inset, from bottom to top: times 0, 4, 8 and 35 min.

Table 2.1. Analysis of the amide I' band for the WT. Changes in position and area of the main bands of the spectra of ADA2h WT at the most remarkable times of the aggregation kinetics.

| Initial | | t = 4 min | | t = 8 min | | t = 35 min | |
|------------------------------|----------|------------------------------|----------|------------------------------|----------|------------------------------|----------|
| Position (cm ⁻¹) | Area (%) | Position (cm ⁻¹) | Area (%) | Position (cm ⁻¹) | Area (%) | Position (cm ⁻¹) | Area (%) |
| 1682.4 | 5.0 | 1683.4 | 4.1 | 1683.7 | 2.1 | 1683.4 | 4.4 |
| 1671.2 | 8.2 | 1671.6 | 11.4 | 1672.4 | 16.7 | 1672.9 | 9.1 |
| 1660.1 | 15.2 | 1658.7 | 17.1 | 1658.0 | 18.8 | 1659.0 | 16.7 |
| 1646.4 | 30.3 | 1644.8 | 28.9 | 1642.2 | 26.4 | 1641.6 | 23.2 |
| 1632.2 | 27.0 | 1631.2 | 24.1 | 1629.0 | 17.5 | 1628.0 | 21.2 |
| 1619.3 | 14.3 | 1619.3 | 14.4 | 1617.6 | 18.5 | 1616.9 | 25.4 |

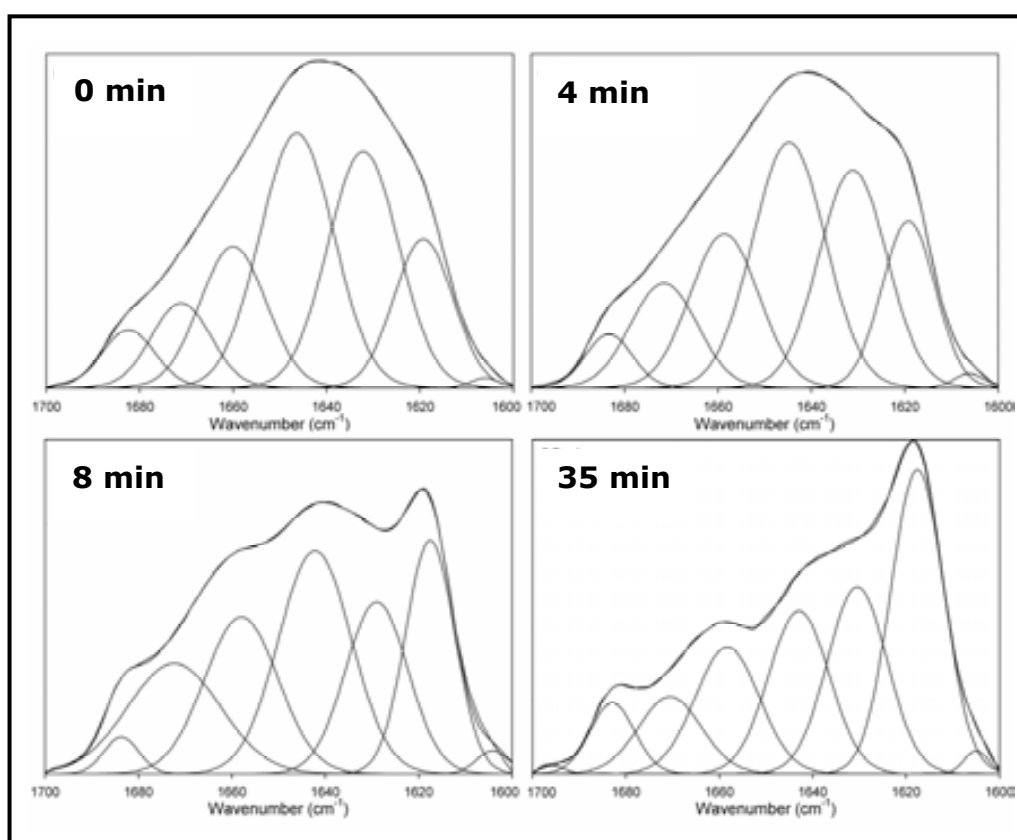


Figure 2.2. Band decomposition of the amyloid formation kinetics for ADA2h WT. Band decomposition at the most relevant times of the aggregation kinetics (0, 4, 8 and 35 minutes) of the WT variant of ADA2h at acidic pH.

2.4.2 Amyloid fibril formation by the wild-type protein

The analysis of the amide I' band corresponding to the protein before fibril induction and during discrete points of the process is shown in Fig. 2.2 and Table 2.1.

To obtain a better insight in the process of fibril formation, the time course of significant band positions and areas can be plotted (Fig. 2.3). Changes in band position indicate either conformational changes or differences in the environment. Variations in area will indicate the inter conversion of structures. During heat induction, the number of bands does not change but their position and percentage of area are different, producing a changed amide I' bandshape. Figure 2.3 A shows the changes in position of the different components. Looking at the changes associated with the band around 1619 cm^{-1} , it can be seen that the initial band at 1619.3 cm^{-1} shifts upwards to 1620.1 cm^{-1} during the first minute and rapidly returns to around 1619 cm^{-1} . After this period, a new frequency change is produced downwards to 1616.9 cm^{-1} where the position is stabilised from time 10 min. These shifts, a small one forwards and a bigger one downwards, indicates that the sample was slightly aggregated at the beginning of the experiment and that temperature disrupted these aggregates. This aggregation might be quite different from that of amyloid fibrils, -although is also rich in β structure- since the transmission electron micrograph at time zero does not show the typical amyloid architecture (Fig. 2.6, see later).

Concerning the area of band around 1619 cm^{-1} , a clear three-step transition is observed, indicating inter-conversion of structures (Fig. 2.3 B). The interpretation is as follows:

Initial protein (slightly aggregated) → disaggregated protein → nucleated aggregate → reorganised aggregate (late fibril).

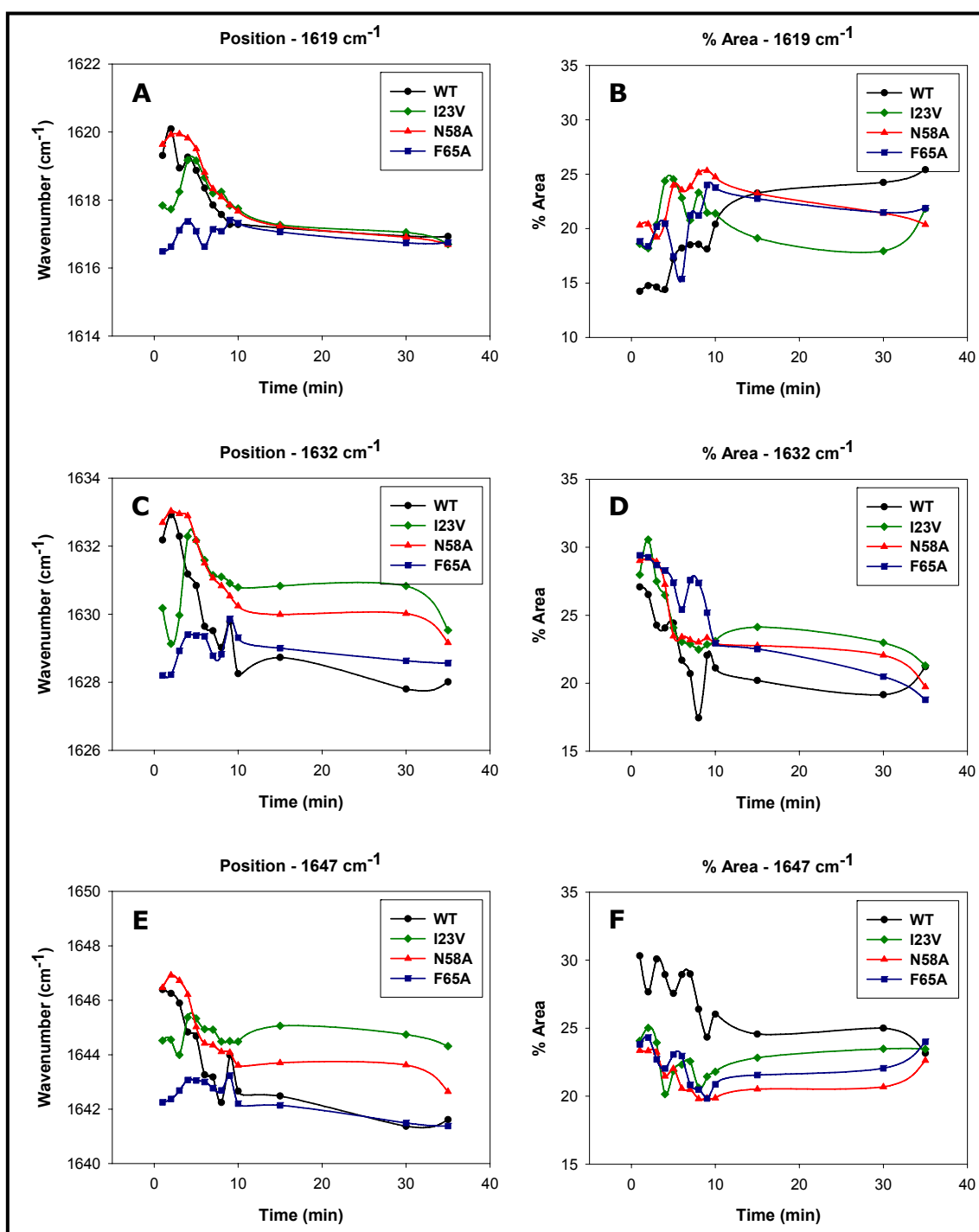


Figure 2.3. Variations in position and area of the main bands in the amyloid formation kinetics. The changes in position of the most significant bands, 1619 cm^{-1} (A), 1632 cm^{-1} (C) and 1647 cm^{-1} (E), are plotted against time for each mutant. The changes in area of these bands; 1619 cm^{-1} (B), 1632 cm^{-1} (D) and 1647 cm^{-1} (F); are also represented along the kinetics. Although the bands do change their position along the time, they have been named only with the most relevant wavenumber for the sake of clarity. See tables 2.1 – 2.4 for further information regarding positions and areas of these bands.

During the first step (time 0-4 min) the area of the 1619 cm^{-1} band remains the same, indicating a disaggregation process rather than an unfolding of the protein (Fig. 2.3 B, Table 2.1). Second step (time 4-8 min) shows an increase in area from 14.4 to 18.5% (Table 2.1), pointing out that the structure of the protein must change upon nucleation of the fibril. In the third step, from time 8 min, there is a dramatic change in area up to 25.4% (Table 2.1) corresponding to a conformational reorganisation that leads to the final amyloid fibril. The reorganisation of the initial nucleus of the fibril formed by ADA2h has been previously reported by following the early kinetics of the process from the unfolded state by stopped-flow coupled to circular dichroism (CD) (Cerdà-Costa *et al.*, 2007). The above mentioned shift downwards to 1616 cm^{-1} indicates a stronger interaction within the reorganised and late fibril.

Concerning structured features, the initial band at 1647 cm^{-1} shows also a shift downwards during the process to 1642 cm^{-1} , indicating the complete transition of α -helices to unordered structure (Fig. 2.3 E). In this case, the α -helical component decreases enhancing the contribution of unordered structure until time 8 min, when the reorganisation of the nucleus begins. The evolution of its area rapidly decreases up to min 8, with a discrete and reversible shift upwards at time 9 min, and remains similar during the reorganisation of the fibril (Fig. 2.3 F). The band at 1632 cm^{-1} , main component of the antiparallel β -sheet, behaves similarly in terms of evolution of its location, shifting downwards during the process (Fig. 2.3 C). Its area also shows the inter-conversion of structures, but in this case the shape is different to that by 1647 cm^{-1} band, pointing clearly min 8 as the moment where the conformational reorganisation takes place (Fig. 2.3 D). The β -sheet band experiments a clear decrease in area from 27% to 17.5% in the first 8 min, and is followed by a small increase just up to 21.2% (Table 2.1). It is known that β -sheets of some native proteins are different that those in amyloid-like fibrils and that a substantial reorganisation is necessary to form the fibril (Zandomenighi *et al.*, 2004). However, this is not the case for $\beta_2\text{m}$, whose native β -sheets are almost identical to aggregated β -sheets (Jahn *et al.*, 2008). Taken all together, we can conclude that for regular secondary elements there is an unfolding process simultaneously to the nucleation of the fibril, followed by a reorganisation of the nucleus at time 8 min that allows the growth of the fibril.

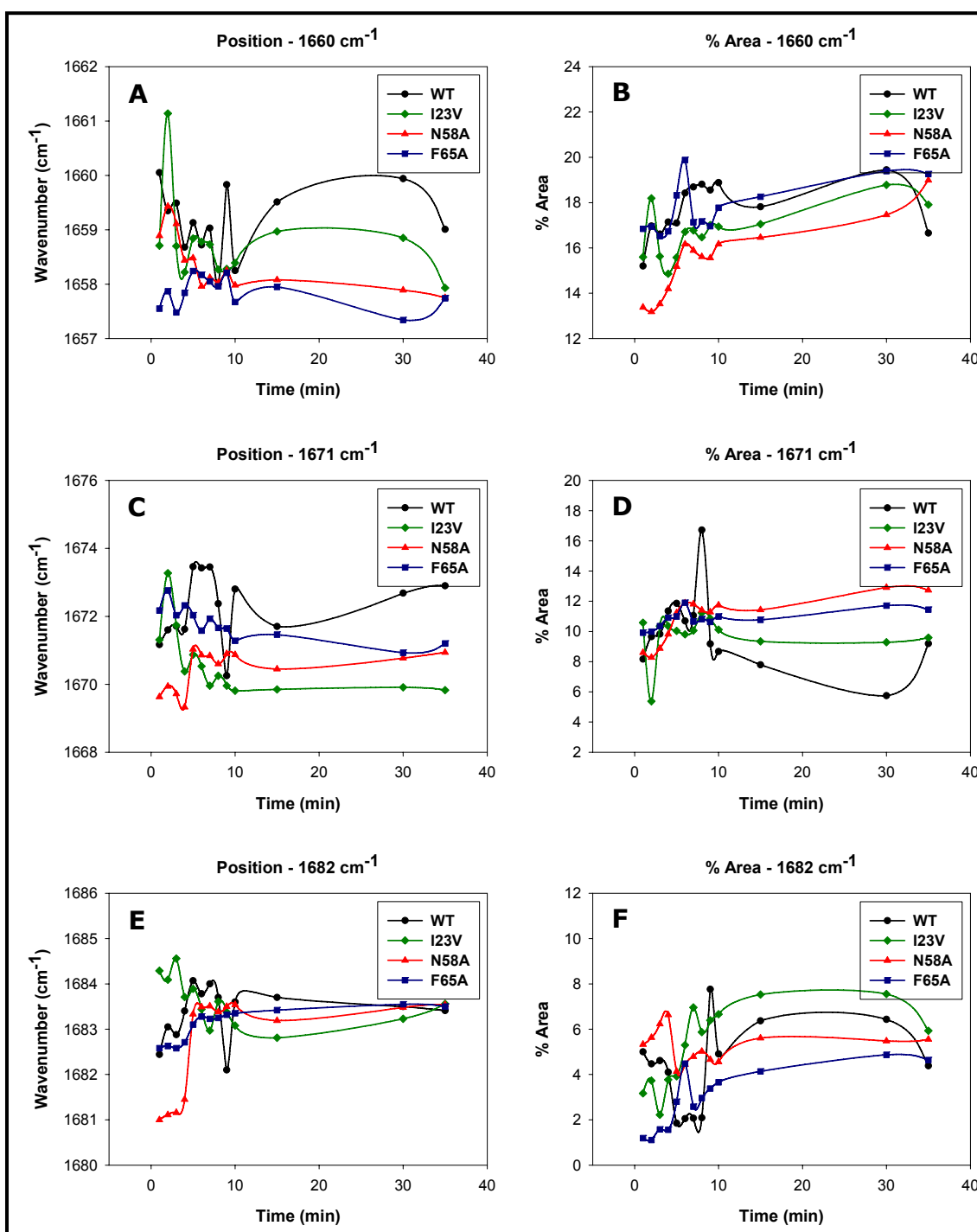


Figure 2.4. Variations in position and area of the other bands in the FT-IR spectra during the amyloid formation kinetics. The changes in position of the rest of the bands of the FT-IR spectra, 1660 cm⁻¹ (A), 1671 cm⁻¹ (C) and 1682 cm⁻¹ (E), are plotted against time for each mutant. The changes in area of these bands; 1660 cm⁻¹ (B), 1671 cm⁻¹ (D) and 1682 cm⁻¹ (F); are also represented along the kinetics. In this case, the bands have also been named only with the most relevant wavenumber for the sake of clarity. See tables 2.1 – 2.4 for further information regarding positions and areas of these bands.

Concerning minor bands, 1682, 1671, and 1660 cm^{-1} , similar phenomena are also observed at time 4 and 8 min, corroborating the relevance of min 8 in the process (Fig. 2.4). 1682 and 1671 cm^{-1} bands behave almost identical in terms of location, indicating that their environment is changing simultaneously. In addition, areas behave somehow complementary suggesting that conformational adaptation of β -strands and β -turns to the fibre mutually influence each other. In the case of the band located at 1660 cm^{-1} it is quite plausible that a transition from a mixture of 3_{10} helix plus β -turns to reorganised β -turns is what we are detecting. Taken into account that the major component is that of an antiparallel β -sheet, it makes sense to assign this component to new β -turns connecting intramolecular β -chains.

2.4.3 Effect of mutations in fibril formation

Several mutants of ADA2h were previously constructed to destabilise its fold (Villegas *et al.*, 1998), others were recently designed to change the kinetics of amyloid fibril formation, and the early kinetics of aggregation of all of them was studied (Cerdà-Costa *et al.*, 2007). For the present work we have chosen a rapid one, N58A, a slower one, I23V, and a non-aggregating one, F65A. It is important to note that when studying aggregation concentration is a key factor. In our previous work we used a protein concentration of 40 μM instead of 544 μM presented in this study, and we pointed out that even the least prone to aggregation variant shall aggregate upon increasing concentration (Cerdà-Costa *et al.*, 2007). This increase of more than 10 fold in the present work is a consequence of the different approach undertaken in both studies (CD coupled to stopped-flow and IR).

The ability to form fibrils by these variants has also been studied by IR spectroscopy, using the same conditions as for the native protein. The amide I' band has been analysed during the processes and are presented at the relevant steps for the wild-type in Tables 2.2 – 2.4. At first glance, it can be seen that whereas in the periods of disaggregation and nucleation the processes are not the same in the different mutants, the band positions and areas corresponding to the reorganisation and subsequent growing do not actually vary (Fig. 2.3), indicating that the mutations affect the propensity of the protein to aggregate, but not the fibril structure.

Table 2.2. Analysis of the amide I' band for I23V. Changes in position and area of the main bands of the spectra of I23V at the most remarkable times of the aggregation kinetics.

| Initial | | t = 4 min | | t = 8 min | | t = 35 min | |
|------------------------------|----------|------------------------------|----------|------------------------------|----------|------------------------------|----------|
| Position (cm ⁻¹) | Area (%) | Position (cm ⁻¹) | Area (%) | Position (cm ⁻¹) | Area (%) | Position (cm ⁻¹) | Area (%) |
| 1684.3 | 3.2 | 1683.7 | 3.8 | 1684.3 | 3.2 | 1683.7 | 3.8 |
| 1671.3 | 10.6 | 1670.4 | 10.4 | 1671.3 | 10.6 | 1670.4 | 10.4 |
| 1658.7 | 15.6 | 1658.2 | 14.8 | 1658.7 | 15.6 | 1658.2 | 14.8 |
| 1644.5 | 24.1 | 1645.4 | 20.1 | 1644.5 | 24.1 | 1645.4 | 20.1 |
| 1630.2 | 27.9 | 1632.3 | 26.5 | 1630.2 | 27.9 | 1632.3 | 26.5 |
| 1617.8 | 18.6 | 1619.2 | 24.4 | 1617.8 | 18.6 | 1619.2 | 24.4 |

Table 2.3. Analysis of the amide I' band for N58A. Changes in position and area of the main bands of the spectra of N58A at the most remarkable times of the aggregation kinetics.

| Initial | | t = 4 min | | t = 8 min | | t = 35 min | |
|------------------------------|----------|------------------------------|----------|------------------------------|----------|------------------------------|----------|
| Position (cm ⁻¹) | Area (%) | Position (cm ⁻¹) | Area (%) | Position (cm ⁻¹) | Area (%) | Position (cm ⁻¹) | Area (%) |
| 1681.0 | 5.3 | 1681.5 | 6.6 | 1683.4 | 5.0 | 1683.5 | 5.6 |
| 1669.6 | 8.6 | 1669.3 | 9.8 | 1670.6 | 11.4 | 1670.9 | 12.7 |
| 1658.9 | 13.4 | 1658.4 | 14.2 | 1658.0 | 15.6 | 1657.8 | 19.0 |
| 1646.5 | 23.4 | 1646.2 | 21.5 | 1644.1 | 19.8 | 1642.6 | 22.6 |
| 1632.7 | 29.0 | 1632.9 | 27.3 | 1630.8 | 23.0 | 1629.2 | 19.7 |
| 1619.6 | 20.3 | 1619.8 | 20.6 | 1618.1 | 25.2 | 1616.7 | 20.4 |

Table 2.4. Analysis of the amide I' band for F65A. Changes in position and area of the main bands of the spectra of F65A at the most remarkable times of the aggregation kinetics.

| Initial | | t = 4 min | | t = 8 min | | t = 35 min | |
|------------------------------|----------|------------------------------|----------|------------------------------|----------|------------------------------|----------|
| Position (cm ⁻¹) | Area (%) | Position (cm ⁻¹) | Area (%) | Position (cm ⁻¹) | Area (%) | Position (cm ⁻¹) | Area (%) |
| 1682.6 | 1.2 | 1682.7 | 1.6 | 1683.3 | 2.9 | 1683.5 | 4.6 |
| 1672.2 | 9.9 | 1672.3 | 10.9 | 1671.7 | 10.8 | 1671.2 | 11.4 |
| 1657.6 | 16.9 | 1657.8 | 16.7 | 1657.9 | 17.2 | 1657.7 | 19.3 |
| 1642.3 | 23.8 | 1643.1 | 22.0 | 1642.7 | 20.5 | 1641.4 | 24.0 |
| 1628.2 | 29.4 | 1629.4 | 28.3 | 1628.8 | 27.4 | 1628.6 | 18.8 |
| 1616.5 | 18.8 | 1617.4 | 20.5 | 1617.1 | 21.2 | 1616.8 | 21.9 |

Thus, if the band position at 1619 cm⁻¹ is considered (Fig. 2.3 A), the pattern described for the wild-type is similar for the mutants I23V and N58A, with slight changes in the wavenumber, but in the case of the F65A mutant the initial wavenumber is clearly shifted-down. Initial area of each variant, F65A 18.8%, N58A 20.3%, I23V 18.6% and WT 14.3% (Fig. 2.3 B, Tables 2.1 – 2.4), inversely correlates with stability (2.8; 4.0; 4.3; 4.4; Villegas *et al.*, 1998); except for N58A, the variant most prone to aggregation. The shape of the evolution of areas is completely different, although a similar percentage of area is reached at the end (Fig. 2.3 B). These results would indicate that the mutations produce conformational changes in the process; however, since influence of the environment remains the same, they do not alter the pattern of the final fibril.

2.4.4 Two-dimensional correlation spectra

To reveal interaction between bands in the changes occurring during the process of fibril formation, excluding the disaggregation time (min 0-4), the 2D-COS spectra have been plotted (Fig. 2.5). In the synchronous map, the auto-peaks (diagonal) correspond to changes in intensity and are always positive. The cross-correlation peaks indicate a relationship between the two bands involved (Arrondo *et al.*, 2004; Iloro *et al.*, 2004; Iloro *et al.*, 2005; Noda, 2007). In the wild-type protein two clear autopeaks are present in the diagonal (Fig. 2.5 A): one predominant located at 1616

cm^{-1} , indicative of the amyloid fibril and the other centred around 1647 cm^{-1} , associated with partially unfolded α -helices. When the $1616/1647 \text{ cm}^{-1}$ cross-correlation peak is considered, a negative correlation is found between them, thus pointing out that the growing of the amyloid fibril is mainly due to the disappearance of this secondary structure. Even if not any autopeak is seen at 1682 cm^{-1} , the other component of the amyloid fibril (and also the minor component of the native anti-parallel β -sheet), the presence of the less intense cross-correlation peak $1682/1647 \text{ cm}^{-1}$ indicates that the change in intensity of this band is smaller. No intensities seem to be associated with the low frequency -and main- β -sheet component which would agree with the small change in area in the overall process, as seen in Table 2.1. However, the above main mentioned cross-correlation peak, $1616/1647 \text{ cm}^{-1}$, is quite broad, comprising from native β -sheets (1632 cm^{-1}) to β -turns (1675 cm^{-1}), thus reflecting the mentioned observation that β -sheets of some native proteins are different that those in amyloid fibrils and that a substantial reorganisation is necessary to form the fibril (Zandomeneghi *et al.*, 2004).

The synchronous maps for the mutants resemble that of the wild-type (Fig. 2.5 B, C, D), although their cross-correlation peaks appear broader when compared to it. This is a consequence of the broadening of almost all bands present in their spectra (see Fig. 2.1 A-D, inset). It is noteworthy the fact that the mutants show a strong autopeak around 1625 cm^{-1} , not present in the wild-type. This autopeak is the reflection of the minimum preceding the aggregated β -sheet band at 1616 cm^{-1} , which is strongly decreasing in each of them (see also Fig. 2.1 A-D). This can be specially observed in I23V, where the 1616 cm^{-1} band does not increase very much and therefore, not a corresponding autopeak can be found. Moreover, the mutants all present an autopeak at 1682 cm^{-1} , albeit small, which is a reflection of the increase of this band during the formation of the amyloid fibril (see Fig. 2.1).

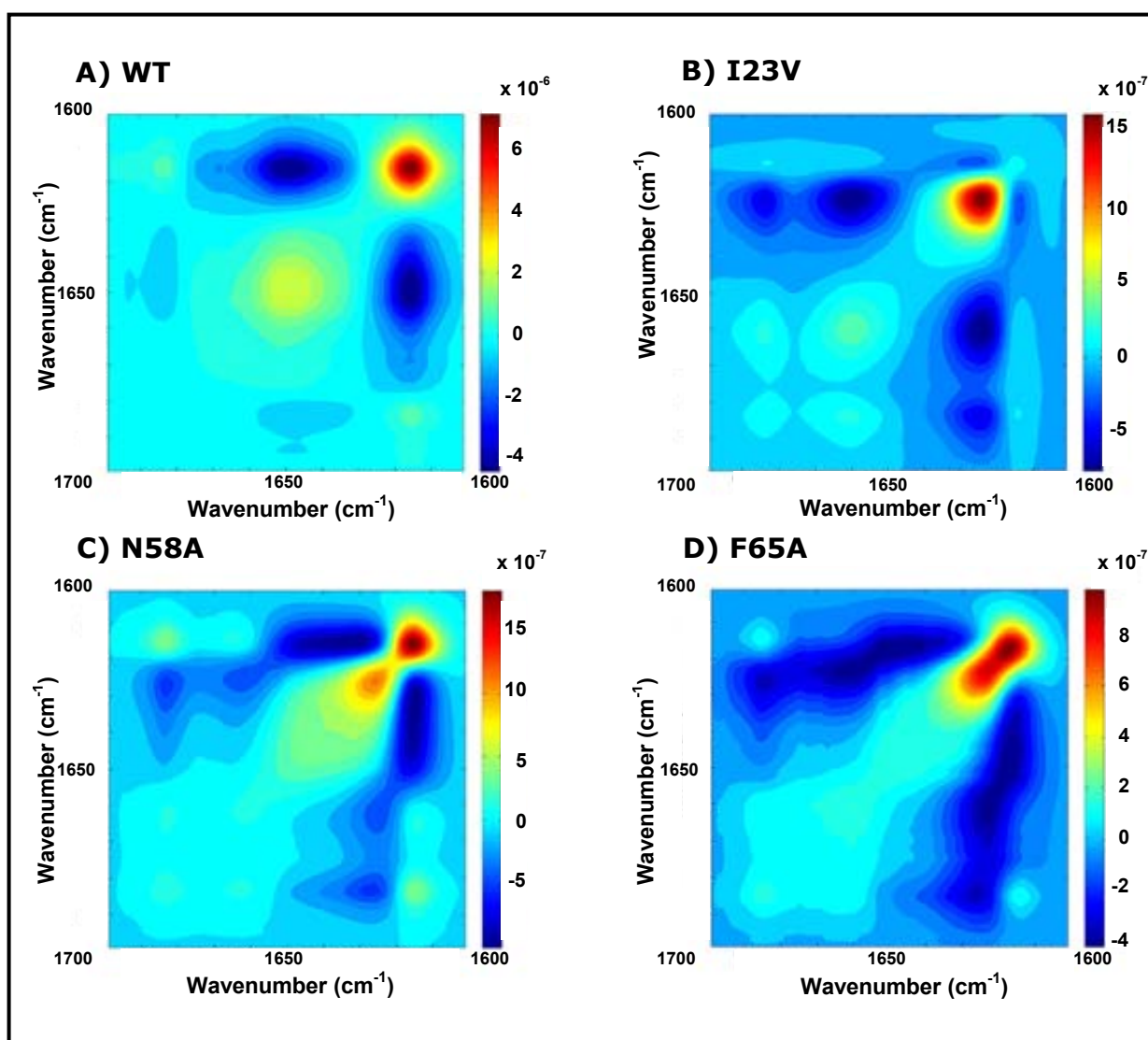


Figure 2.5. Synchronous maps for the aggregation kinetics. The synchronous maps for all the mutants are represented using the two-dimensional correlation analysis. (A) WT; (B) I23V; (C) N58A; (D) F65A. The positive correlations are shown in different shades of red according to their intensities, and the negative ones in blue.

2.4.5 Electron microscopy

Electron micrographs demonstrated that the process leads to the formation of amyloid fibrils and not just to an amorphous aggregation of the protein (Fig. 2.6). At time 0 min, N58A, F65A and WT-ADA2h show some worm-like fibrils as those described by Radford for the assembly of spherical or pre-fibrillar material that follow a non nucleated-dependent kinetics (Gosal *et al.*, 2005; Smith *et al.*, 2007), in contrast to the nucleated-dependent kinetics driving to amyloid fibrils (Harper & Lansbury, 1997).

However, no such worm-like structures were clearly observed for I23V (see discussion). Here, we have depopulated this worm-like state by temperature and allowed beginning the process of fibril formation. The fact that the process is different for each variant, as seen by the evolution of band areas in the IR (see Fig. 2.3), indicates that these worm-like structures would be conformationally different. The worm-like structures observed for the variants, with the exception of I23V, are approximately 80 - 100 nm long and 8 - 9 nm wide. It is noteworthy that F65A worm-like fibrils seem to be more organised than the ones obtained for the other variants (see discussion).

At 30 min (time required to prepare the TEM samples) after heating the samples in the same way as for the IR experiments, all of the variants show organised fibrillar material, especially defined for the most prone to aggregation mutant, N58A. The fibres are straighter, and very similar for all variants. At longer periods the fibrils are better defined for all of the variants, demonstrating that the reached structure is the same for all four processes studied.

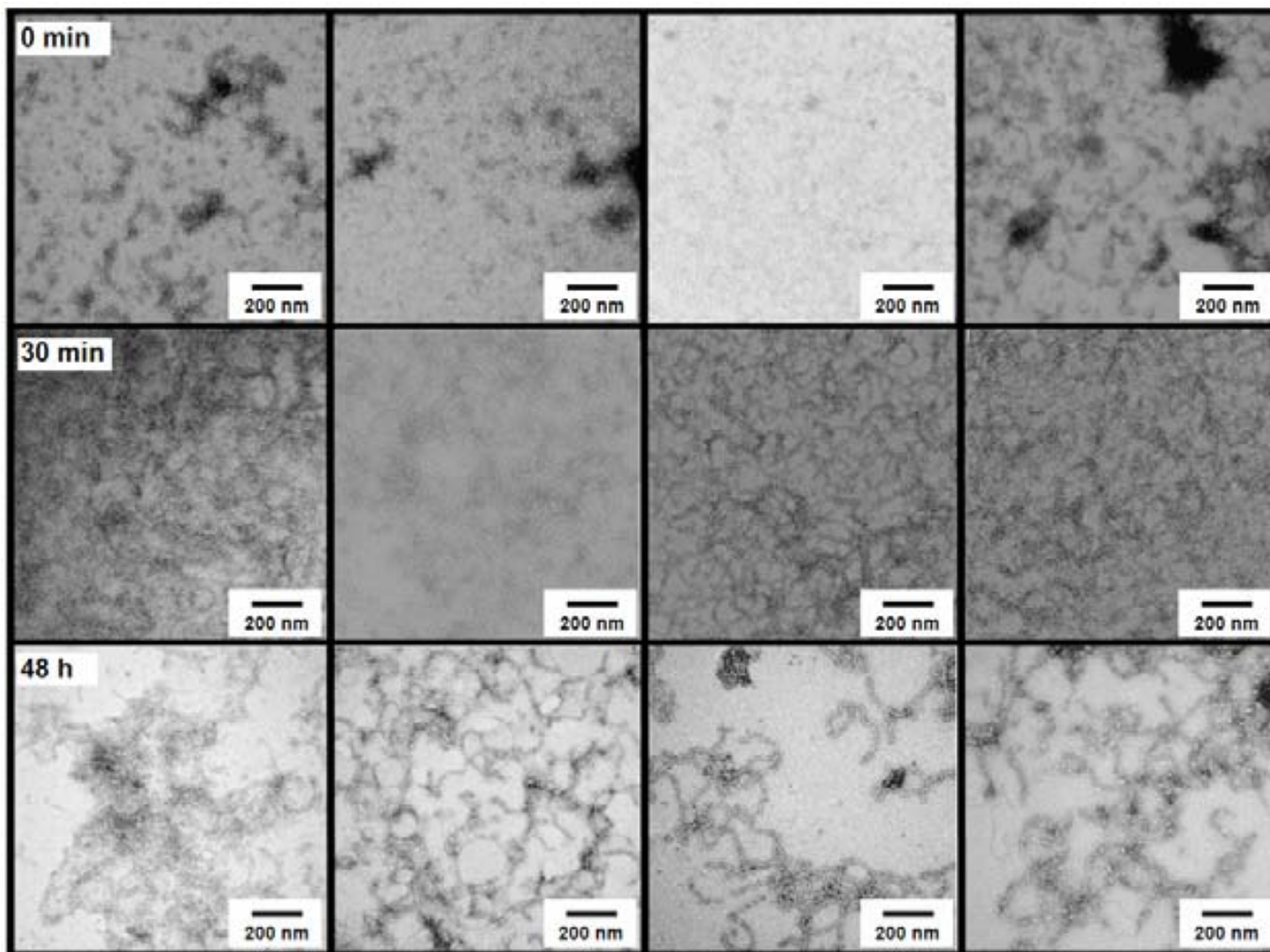


Figure 2.6. Transmission electron microscopy of the amyloid formation process. Transmission electron microscopy images of ADA2h WT (A) and the mutants I23V (B), N58A (C) and F65A (D) at different times (0, 30 min and 48 h) after the finalisation of the heating process. The time 0 micrograph has been taken at pH 3 prior to heating.

2.4.6 Circular dichroism

Fig. 2.7 shows the far-UV CD-spectra of the variants at neutral and acidic pH, measured at two different concentrations, 20 μM and 200 μM , for pH 3. At pH 7 there is no aggregation, thus the spectra obtained at 200 μM are identical to those at 20 μM . It is important to note that the samples at pH 3 were induced to form amyloid fibrils by heating at 95 °C for 5 min and, in consequence, do not correspond to the conditions of the FTIR experiment at time zero. In any case, the concentration used in the IR, 544 μM , is too high to be studied by CD, especially in the very far-UV region (210-190 nm) where the lamp is saturated and no reliable information can be obtained. Thus the CD data here presented must be regarded as an approximation, rather than to a reproduction, of the IR final spectra.

The $\alpha+\beta$ structure of the wild-type is maintained for all three mutants at neutral pH. The deconvolved CD spectra (Table 2.5) shows an α -helix content of above 31-32% for all of the variants, with the exception of F65A that slightly decreases to 27% favouring an equivalent increase in β -sheet content. This shift is a consequence of its instability rather than for corresponding to a mutation in α -helix-2, since mutation of the N-cap of this helix (N58A) or mutation within α -helix-1 (I23V) does not affect secondary structure content. It is important to indicate that the side-chain of F65 interacts with the backbone of V64, one of the main components of the hydrophobic core (Villegas *et al.*, 1998).

At acidic pH and upon temperature induction, the WT shows a loss in helical content enhancing the weight of β -sheet conformation, as previously described (Villegas *et al.*, 2000). At 20 μM , an 8% of helical content is lost in favour to an increase of β -sheet content, whereas β -turns and random coil remain the same. However, at 200 μM there is a dramatic loss of 24% of helical content in favour of an increase of β -sheet content (18%) and also of random coil conformation (6%). The effect of acidic pH is more remarkable for mutants I23A and N58A than for the WT, but especially for N58A, the most prone to amyloid fibril formation. N58A shows very similar spectra at both concentrations, reaching high β -sheet content at low concentration. Moreover, as it was the case for the WT, the random coil conformation is enhanced upon raising concentration. This is also true for mutant I23V. For these three variants, WT, I23V, and N58A, there is an increase of random coil conformation when raising concentration, and not just a pure transition to β -sheet content. This is in concordance with both the increment in area of the initial band at 1647 cm^{-1} and its shift downwards to 1641 cm^{-1} .

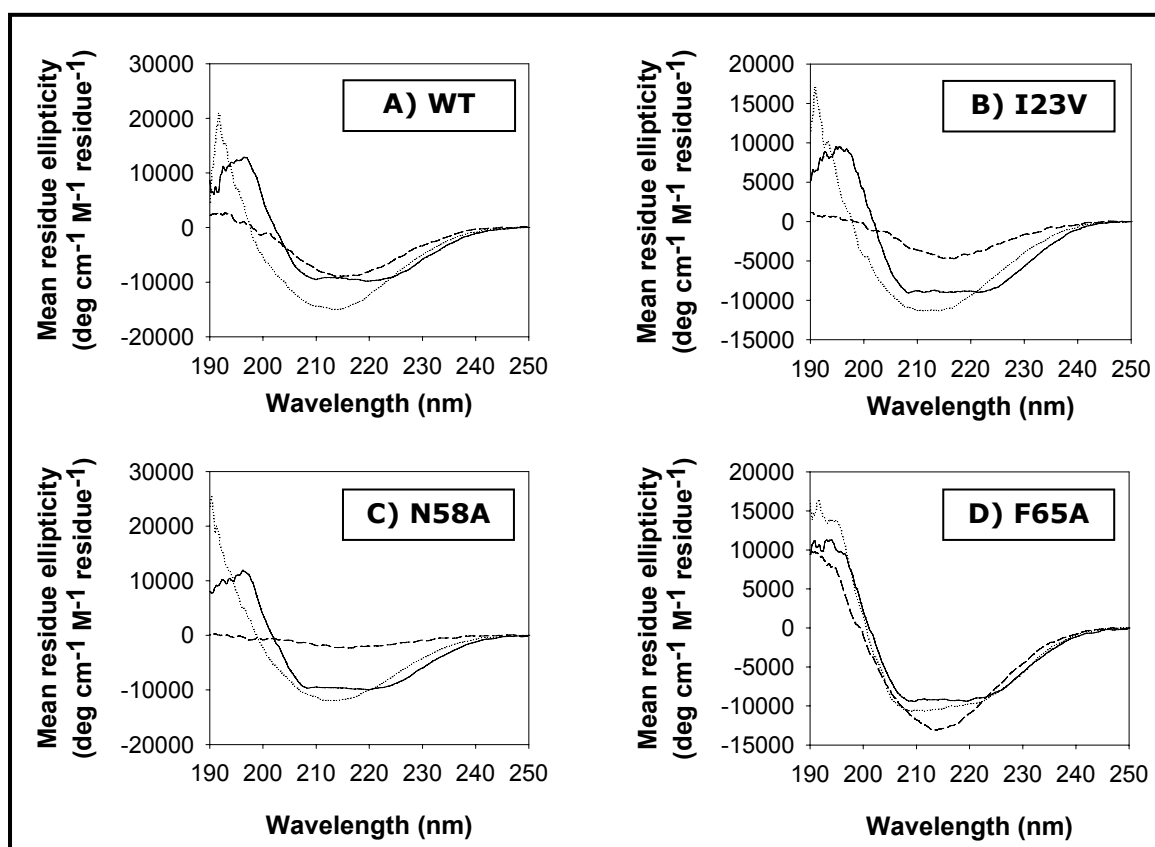


Figure 2.7. Comparison between the native and the amyloid-induced conformations of ADA2h WT and the destabilised mutants. Conformations of the WT and the mutants I23V, N58A and F65A analysed with circular dichroism spectroscopy under native conditions, at pH 7 (continuous line); and after inducing amyloidogenesis at pH 3, at a protein concentration of 20 μM (dotted line) and at a protein concentration of 200 μM (dashed line).

However, F65A, the one non-prone to aggregation (Cerdà-Costa *et al.*, 2007) and rather unstable (Villegas *et al.*, 1998), shows at acidic pH and 20 μM an spectrum similar to that at neutral pH. Upon raising concentration, the helical content decreases just a 5% in favour of a gain in β-sheet, and random coil conformation remains the same in all three conditions studied. Thus, is quite possible that mutation on this residue, interacting with part of the hydrophobic core in the WT, generates a β-conformation with some similarities to the one in the final fibril, that will be more populated at the concentration required for the IR experiment, and could account for the small change in position of band 1619 cm⁻¹ observed for this mutant.

Table 2.5. Content of secondary structure of ADA2h at native and amyloid-induced conditions. Deconvolution of the circular dichroism spectra of ADA2h variants at 20 μM and neutral pH and at 20 and 200 μM at acidic pH. (*) The spectrum of F65A at 200 μM and pH 3 is too noisy in the far-UV region due to lamp saturation, making its deconvolution unreliable.

| | α -helix (%) | β -sheet (%) | β -turn (%) | random (%) |
|---|------------------------|-----------------------|----------------------|---------------|
| WT pH 7, 20 μM | 32 | 20 | 12 | 36 |
| WT pH 3, 20 μM | 24 | 29 | 11 | 36 |
| WT pH 3, 200 μM | 8 | 38 | 11 | 43 |
| I23V pH 7, 20 μM | 31 | 20 | 13 | 36 |
| I23V pH 3, 20 μM | 17 | 35 | 12 | 36 |
| I23V pH 3, 200 μM | 6 | 42 | 12 | 40 |
| N58A pH 7, 20 μM | 32 | 20 | 12 | 36 |
| N58A pH 3, 20 μM | 9 | 43 | 11 | 37 |
| N58A pH 3, 200 μM | 7 | 39 | 12 | 42 |
| F65A pH 7, 20 μM | 27 | 25 | 12 | 36 |
| F65A pH 3, 20 μM | 27 | 25 | 12 | 36 |
| F65A pH 3, 200 μM(*) | 21 | 31 | 11 | 37 |

2.5 DISCUSSION

2.5.1 Assignment of the initial bands for the wild-type protein

Band assignment is not a straightforward procedure because of the sensitivity of IR spectroscopy to combined structural and environmental factors (Arrondo *et al.*, 2003). Despite the hallmark β -conformation of amyloid fibrils (Serpell *et al.*, 1999), it has been reiteratively shown by IR that the amide I' peaks cluster between 1630 and 1640 cm^{-1} for β -sheet proteins whereas for amyloid-like fibrils extends from 1611 to 1630 cm^{-1} (Zandomeneghi *et al.*, 2004). This structural difference makes FTIR a very suitable technique to follow fibril formation processes from all types of protein folds.

Initially, a band at 1619 cm^{-1} has a 14 % of the amide component. This indicates that this portion of the protein is initially aggregated. This aggregation, although contributing slightly to the spectra, could not be skipped at the acidic pH of the experiment, especially at the required protein concentration for the IR (5 mg/ml, 544 μM). Nevertheless, the electron micrograph obtained in the same conditions (Fig. 2.6 A, $t = 0$) does not show the presence of any structured amyloid fibril but rather some irregular worm-like structures, similar to those seen previously for β_2 -microglobulin (Gosal *et al.*, 2005). These less structured fibrils presumably originate the initial band at 1619 cm^{-1} , similarly to the 1622 cm^{-1} band previously seen for β_2 -microglobulin worm-like fibrils (Jahn *et al.*, 2008), whereas the final position at 1616 cm^{-1} would correspond to the more structured amyloid fibrils. This highlights the different nature of these aggregates compared to that characteristic of the matured amyloid fibrils, and might represent a ground state from a different aggregation pathway (Gosal *et al.*, 2005). It is noteworthy that inclusion bodies and thermally induced nonfibrillar aggregates also present a band located around 1625-1610 cm^{-1} (Fink, 1998).

The bands at 1682 and 1632 cm^{-1} correspond to the high and low frequency components of the antiparallel β -sheet, respectively. There is some disagreement in literature concerning the assignment of band components between 1670-1690 cm^{-1} . This region has classically been assigned as the high frequency component of antiparallel β -sheets (Susi, 1969; Byler & Susi, 1986; Holloway & Mantsch, 1989), but taken into account that more than one band component can be detected it seems more realistic to assign this region to combined contribution of β -sheets and turns (Olinger *et al.*, 1986; Prestrelski *et al.*, 1991). Some examples of more than one component in this region are deuterated transthyretin (1675 and 1682 cm^{-1} , Cordeiro *et al.*, 2006),

deuterated lectin (1676, and 1691 cm^{-1} , Chehín *et al.*, 1999), undeuterated fibronectin (1680, 1691 cm^{-1} , Hu *et al.*, 2006), and undeuterated LDL (1673, 1683, 1693 cm^{-1} , Gallego-Nicasio *et al.*, 2003). It is known that deuteration of the sample slightly shifts downwards the maxima of the components (Susi, 1969). For deuterated ADA2h, there is an additional band at position 1671 cm^{-1} that kinetically behaves in terms of position as that showed by the high frequency component of the antiparallel β -sheet (1682 cm^{-1}). Therefore, it makes sense to attribute this band to β -turns. In any case, the uncertainty in this assignment is not critical for the current study.

The typical IR-band for α -helices locates around 1650 cm^{-1} (Arrondo *et al.*, 1993). Some studies have pointed out that the solvent-exposed α -helices give peak frequencies that are lower than the characteristic frequencies for α -helices within proteins (Reisdorf & Krimm 1996; Manas *et al.*, 2000; Silva *et al.*, 2002). For ADA2h it has been located at 1647 cm^{-1} . This can be the case because the induction must be performed at acidic pH, where the secondary structure is not the same than at neutral pH. The far-UV CD-spectra show a clear loss in helicity at acidic pH when comparing with neutral pH.

The band at 1660 cm^{-1} has been attributed in literature to coiled coils (Reisdorf & Krimm, 1996), 3_{10} helices (Arrondo & Goñi, 1999) or β -turns (Byler & Susi, 1986) and is initially assigned in this work as the end of the α -helices, specially the C-terminal of α -helix1 of ADA2h, together with β -turns.

2.5.2 Nucleation and Reorganisation

It has recently been reported that the amide I' band maximum of native β -proteins depends on the average number of strands per sheet, decreasing wavenumbers with increasing number of strands (Zandomeneghi *et al.*, 2004). This feature is also suggested by *ab initio* calculations (Kubelka & Keiderling, 2001). A location at 1615 cm^{-1} was reported previously for WT-ADA2h amyloid fibrils that were induced at 90 °C for 30 min (Villegas *et al.*, 2000). In this study, designed to follow the process, the temperature used was 90 °C and a final band of 1616 cm^{-1} was reached. Albeit this small difference in the position of the band, the sample incubated at longer periods than this performed for the IR experiment shows by TEM the unequivocally presence of amyloid fibrils.

The initial position of the band in the wild-type spectrum shifts upwards and downwards very rapidly (3-4 min) revealing a decrease in the number and architecture of the intermolecular β -strands of the initial worm-like aggregate. Once the protein is

disaggregated, the band shifts downwards revealing an increasing number of β -strands implied in the structure, until the final fibril is reached. It must be noted, however, that the slope of this shifting is more pronounced between time 4-8 min than from time 8 min and on. This three-step kinetics is better observed when considering areas that reflect the interconversion of structures and correspond to 1- disaggregation of initial short, worm-like fibrillar material, 2- nucleation, and 3- conformational reorganisation necessary to form the final fibril. This reorganisation was already described for ADA2h when studying the early kinetics of β -structure acquisition from the unfolded state by CD induced by pH-jump (Cerdà-Costa *et al.*, 2007). There are several proteins where a conformational reorganisation of the nucleus has been reported in order to achieve ordered fibrillation (Serio *et al.*, 2000; Zandomeneghi *et al.*, 2004; Plakoutsi *et al.*, 2005; Cordeiro *et al.*, 2006). In addition, there is also a recent computer simulation that indicates that hydrophobic residues are sequestered away from the solvent during the nucleation step and that during fibril formation the oligomers undergo a process of reorganisation driven by inter-chain hydrogen bonding interactions that induce the formation of β -sheet rich assemblies (Cheon *et al.*, 2007). This mechanism perfectly fits with the data here presented for ADA2h.

For other components, the three-step transition is not as clear as for the amyloid hallmark band, but all of them show time 4 min, and especially time 8 min, as the more critical moments in the process. The low frequency component of the β -sheet, initially located at 1632 cm^{-1} , shifts downwards to 1627 cm^{-1} , in good agreement with the above mentioned differential location of IR bands for some native β -sheet proteins and amyloid fibrils (Zandomeneghi *et al.*, 2004), and reveal the accommodation of initial β -sheet into the fibril. Concerning α -helices, the behaviour of their band is identical to that of the low frequency component of the β -sheet, shifting downwards from 1647 to 1641 cm^{-1} and revealing that the complete transition of α -helices to unordered structure is concomitant to the accommodation of initial β -sheet into the amyloid fibril.

Therefore, from the results obtained, it can be concluded that the kinetics of fibril formation consists of a first step, that do not produce changes in the bands associated with fibril formation but rather the disaggregation of a short, worm-like fibrillar material, a second step where fibril is nucleated as protein partially unfolds, and finally, a reorganisation of the protein is produced until the amyloid structure is achieved. This final structure has a major component of β -sheet, with bands compatible with an antiparallel distribution, plus unordered structure, a component at 1660 cm^{-1} , which could be new β -turns, and another at 1671 cm^{-1} compatible with β -turns directly related to the low frequency component of the β -sheet.

Apart from the results discussed, the 2D-COS maps allow the correlation of bands, and shed some light to the dynamics of the process. The synchronous map shows that the major interaction is between bands by partially unfolded α -helices and the amyloid fibril. The spectra of the mutants present broadened bands, which could reflect the fact that these variants are not as structured as the wild-type protein, in concordance with their lower stability (Villegas *et al.*, 1998).

2.5.3 Influence of aggregation propensity and stability

In a previous work, we determined the concentration-dependent rate constant (k_1 , nucleation) and the independent-concentration rate constant (k_2 , reorganisation) for a large battery of single-residue variants, and demonstrated that it is the interplay between stability and the propensity of a sequence to form ordered aggregates what determines the aggregation outcome for a protein (Cerdà-Costa *et al.*, 2007). In the work here presented, wild-type (WT, $k_1 = 0.40 \text{ s}^{-1}$), a slowly aggregating mutant (I23V, $k_1 = 0.14 \text{ s}^{-1}$), a rapid aggregating mutant (N58A, $k_1 = 0.86 \text{ s}^{-1}$) and a non-prone to aggregation mutant (F65A, $k_1 = 0.00 \text{ s}^{-1}$), have been chosen to study the process of amyloid fibril formation by IR. The induction by temperature has been performed from the folded state at acidic pD in order to evaluate the combined contribution of stability and aggregation propensity.

Initial area of band around $1619\text{-}1616 \text{ cm}^{-1}$ of each variant inversely correlates with stability for all of the variants, except for N58A, the more prone to aggregation, demonstrating that the destabilisation of the native state with respect to the aggregation intermediate ensemble is coupled with an increasing propensity to non-fibril aggregation. In the case of mutant N58A, its tendency to aggregate originates a higher initial area of this hallmark band.

At time zero all of the variants, with the exception of I23V, form clear worm-like structures as seen by TEM. The lack of these structures for I23V at time zero could be related to its ability to form non-native interactions during the transition state of its folding reaction (Villegas *et al.*, 1998). It is remarkable the definition of such structures for the most unstable mutant, F65A, reinforcing the idea that worm-like structures correspond to a ground state that is easily reached from partially unfolded proteins. This ground state is in direct competition and favoured respect to amyloid fibril formation in our initial experimental conditions (Gosal *et al.*, 2005), until the thermal denaturation facilitates the formation of amyloid fibrils, presumably by allowing the exposition of buried parts of the domain important for driving the nucleation process.

This reflects that ADA2h would follow the refolding model among the monomer structural reorganisation models proposed by Eisenberg *et al.* (Nelson & Eisenberg, 2006).

In spite of the differences observed for wild-type ADA2h and its mutants, their aggregation processes all culminate in the formation of amyloid-like fibrils with equivalent characteristics.

**CHAPTER 3:
PREDICTION OF AN ALTERNATIVE FUNCTION FOR ADA4h
AND EXPERIMENTAL VALIDATION**

3.1 SUMMARY

The observation that activation segments of procarboxypeptidases are rather long compared to pro-regions of other zymogens points out the possibility that they could play additional functions apart from preserving the activation of the proenzyme and helping in its folding process. In the present work, we have compared the overall pro-domain tertiary structure with several proteins belonging to the same fold in SCOP by using structure and sequence comparisons. The best score has been obtained between ADA4h and human U1A protein from the U1 snRNP. Structural alignment has revealed the existence of RNP 1 and RNP 2 related sequences in ADA4h. After modelling ADA4h on U1A, the new structure was used to extract a new sequence pattern characteristic for important residues at key positions. The new sequence pattern allowed scanning protein sequences to predict the RNA binding function for 32 sequences undetected by Pfam. Unspecific and specific EMSA experimentally validate the ability of ADA4h to bind RNA, although competitor tRNA abolished most of the interaction. The fact that key residues in preventing activity within the proenzyme are also key residues for RNA binding might induce ADA4h domain to evolve from the canonical RNP1 and 2 sequences.

3.2 INTRODUCTION

Carboxypeptidases (CPs) are proteases that hydrolyze C-terminal peptide bonds from peptides and proteins. A relatively large number of different carboxypeptidases exists, covering a variety of specificities and complementing each other when acting in the same location (Vendrell *et al.*, 2004).

The most used classification of proteases can be found at the MEROPS database (Rawlings *et al.*, 2006) in which metallo-carboxypeptidases belong to the M14 family. The M14 family can be divided into two different subfamilies according to their sequence homology, the N/E subfamily and the A/B subfamily.

The A/B subfamily has been the most studied one, with the bovine carboxypeptidase A (CPA) as the reference model. It comprises enzymes produced as zymogens, having an N-terminal activation segment (or pro-region), about 95 residues long, which inhibits the enzyme activity until it is released due to limited proteolysis. They are usually synthesized in the pancreas and transported to the gut where they participate in food processing; nonetheless, some members like CPU and CPA3 are originated in other tissues and act in alternative processes such as blood coagulation (Hendriks, 1998) or anaphylactic and inflammatory responses (Springman *et al.*, 1995), respectively.

The activation segment, or pro-region, found in the A/B subfamily is remarkably long, constituting one fourth of the entire proenzyme. It is composed by an activation domain (AD), with a globular fold independent from the catalytic domain, and a connecting α -helix that covalently holds them together.

The sequence similarity of the different CPs pro-regions is quite low (20 to 50%), being the residues involved in the enzyme inhibition the most conserved ones; however, they share an almost identical $\alpha+\beta$ topology formed by two α -helices and four β -strands arranged in an open sandwich conformation. As a result, all the activation domains with a resolved 3D structure belong to the same family in the Structural Classification Of Proteins (SCOP) database (Pancreatic carboxypeptidase, activation domain, Murzin *et al.*, 1995; <http://scop.mrc-lmb.cam.ac.uk/scop/>).

The fact that the activation segments are relatively long compared to pro-regions of other zymogens opened some questions about the possibility that they could play other functions apart from preserving the activation of procarboxypeptidases. Folding studies performed in ADA2h showed an extremely fast two-state kinetics (Villegas *et al.*, 1995), suggesting that ADs could assist in the proenzyme folding. Supporting this hypothesis, recombinant expression of various forms of CPs in the methylotrophic

yeast *Pichia pastoris* have been proved unsuccessful, whereas the proenzyme forms can be produced at high yields.

One of the first set of experiments studied the possible Ca^{2+} binding properties, since some structural similarity with the EF-hand protein family had been found (Vilanova *et al.*, 1988). No significant results regarding Ca^{2+} binding were obtained and, later on, the similarity was also discarded using bioinformatic sequential and structural approaches, although some homologies with the complement factor B were proposed (Vendrell *et al.*, 1989).

In the present work, we have compared the overall pro-domain tertiary structure with several proteins belonging to the same fold in SCOP using structure and sequence comparisons. The sequence alignment derived from the structure comparison was used to extract a sequence pattern characteristic for important residues at key positions. The sequence pattern was further used to scan protein sequences to detect RNA binding proteins. Subsequently, we tested experimentally the RNA binding capability of ADA4h in order to confirm our prediction.

3.3 MATERIALS AND METHODS

3.3.1 Bioinformatics

The search of remote homologs of ADA4h was performed with PSI-BLAST (Altschul *et al.*, 1997) using the BLOSUM45 matrix and three iterations. The most similar sequences were obtained from Uniprot database (Bairoch *et al.*, 2005) on the first iteration. The first non-carboxypeptidase protein was included in the second iteration in order to obtain remote homologs of ADA4h and the third iteration was used for convergence.

The 3D structure of ADA4h was extracted from the structure of the human procarboxypeptidase A4 (code: 2boa; García-Castellanos *et al.*, 2005) retrieved from the PDB (Berman *et al.*, 2000). ADA4h belongs to the *pancreatic carboxypeptidase, activation domain* family, in the fold of the *ferredoxin-like*, from SCOP (Murzin *et al.*, 1995). All proteins with similar fold were retrieved from PDB and all protein domains with these folds were extracted from them. The program Structural Time series Analyser Modeller and Predictor (STAMP) (Russell & Barton, 1992) was used to superimpose AD folds and compare their structure by means of the root mean square deviation (RMSD).

The capability of ADA4h of changing the orientation of some of its loops to simulate the function of a different family of the same *ferredoxin-like* fold was tested via comparative modeling using MODELLER (Sali & Blundell, 1993) and the selected templates from the *ferredoxin-like* fold.

Inside this fold there are many proteins with functions unrelated to the carboxypeptidase precursor. Among them, many belong to several superfamilies with nucleotide recognition function. Particularly, one superfamily whose function is related with RNA binding and that can be identified by an RNA recognition motif (RRM). This motif is defined in PROSITE (Hulo *et al.*, 2006) by means of two sequence signatures named RNP1 and RNP2, separated by about 30 residues and it is supposed to bind single-stranded RNA. Due to the structural similarity of ADA4h and the proteins of the families that contained this motif, we were able to identify similar sequence patches (RNP1 and RNP2) in ADA4h separated by approximately 30 residues. Previous work has shown the possibility to infer functional relationships within the same SCOP fold by distant homology (Aloy *et al.*, 2002) and the importance of functional loops to imprint a particular function (Espadaler *et al.*, 2006). Therefore, we hypothesize a new motif based on the sequences RNP1 and RNP2 of RRM that could also include the

sequence patches of ADA4h. The new motif was tested by searching with ScanProsite (de Castro *et al.*, 2006) proteins containing the new motif in Uniprot database.

3.3.2 Materials

The MEGAscript[®] T7 Kit for transcription, the RNase inhibitor Superscript[®] and the RNase decontamination solution RNaseZAP[®] were obtained from Ambion. The RNA purification columns Micro Bio-Spin were purchased from Bio-Rad, the [α -³²P]UTP from GE Healthcare, and the C-UVette[®] DNase RNase Protein-free from Eppendorf. Primers were synthesized by Roche Molecular Biochemicals; all other chemicals were obtained from SIGMA.

3.3.3 Cloning and Expression

ADA4h was cloned into the pET-30 Xa/LIC vector from Novagen resulting in an N-terminal His-tagged protein. The recombinant vector was then transformed into the *Escherichia coli* strain BL21 (DE3), and the obtained clones were screened for positives.

Cells were grown and induced as described previously (Viguera *et al.*, 1997) and ADA4h was purified from the intracellular soluble fraction using a nickel-sepharose. A 60 mM imidazole wash was applied once the sample was loaded and the column was re-equilibrated, followed by an elution step with 300 mM imidazole. The elution fractions containing His-tagged ADA4h were dialyzed overnight against bind buffer, and were again applied to a nickel-sepharose in a batch manner. Nickel sepharose-bound protein was digested overnight at room temperature with the necessary amount of Factor Xa, in order to eliminate the His-tag. The protein was recovered in the supernatant and kept at 0 °C in 50 mM Tris-HCl 0.5 M NaCl pH 8.0. The identity of the protein was confirmed using MALDI-TOF mass spectrometry and N-terminal sequencing.

3.3.4 *In vitro* transcription of RNA

An RNA probe carrying two copies of the U1A-binding motif from stem-loop II of U1 snRNA hairpin (Allain *et al.*, 1997) was synthesized by *in vitro* transcription and named as bipolar-stem loop:

gggAggg**UUAACA**UUgCACUCCgUUg**UCCA**UCCCAAAAAAAAAA**gggUUAACA**UU
gCACUCCgUUgUCC

The online program OligoAnalyzer 3.0 from Integrated DNA Technologies (<http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/Default.aspx>) was used to test secondary structure (Fig. 3.1).

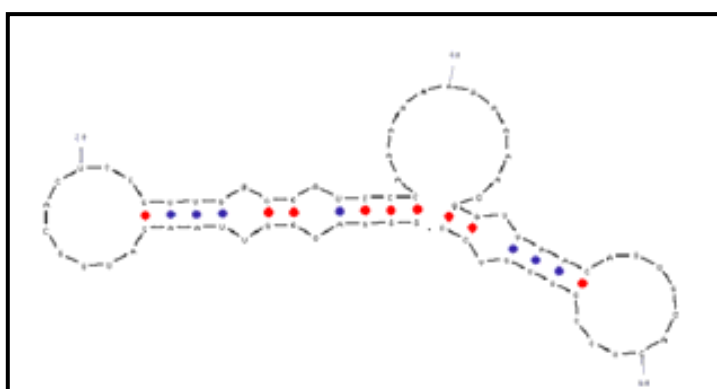


Figure 3.1: Two-dimensional structure of the RNA probe designed to test for ADA4h RNA-binding ability. The structure was generated with OligoAnalyzer 3.0 from Integrated DNA Technologies®.

In order to obtain the mentioned RNA probe, a direct primer carrying the T7 RNA polymerase promoter (shown in bold) was designed partially-overlapping with the reverse one:

5'-**TAATACGACTCACTATA**GGGAGGGTTAACATTGCACTCCGTTGTCCATCCCAAAAAAAAA-3'
 5'-GGACAACGGAGTGCAATGTTAACCTTTTTTTTTTTGGGATGGACAACGG-3'

Both primers, 1 nmol of each one, were incubated with 4 U of Klenow and 0.6 mM dNTPs at 37 °C for 1h, and the enzyme was inactivated afterwards at 65 °C for 10 min. The DNA was then analyzed in a 2% agarose gel and quantified.

In vitro transcription reactions were performed essentially as indicated in the MEGashortscript® T7 Kit. The reaction was performed in 10 µl, using a 1:10 ratio for UTP respect to other NTPs, adding 40 µCi of [α -³²P]UTP, and incubated at 37 °C for 2 h. A second *in vitro* transcription in absence of [α -³²P]UTP was carried out as indicated in the kit instruction manual and used as cold probe. The amount of DNA template

used in both reactions was 125 nM. In both cases, after the transcription reaction 10 U of RNase-free DNaseI were added and the mixture was incubated for 15 min at 37 °C. The obtained RNA was then purified using Bio-Rad Micro Bio-Spin columns P-30 Tris RNase-Free. The unlabeled RNA was quantified by UV-spectrometry using C-UVette® DNase RNase Protein-free and the radioactive probe by scintillation counting.

3.3.5 RNA Electrophoretic Mobility Shift Assays (EMSA)

25 fmol of labeled RNA were incubated in a final volume of 10 µl for 20 min at room temperature with the indicated amounts of protein (Fig. 3.6) in a buffer containing 10 mM HEPES pH 7.4, 25 mM potassium acetate, 2.5 mM magnesium acetate, 0.5% Igepal CA 630, 5% glycerol, 1 mM dithiothreitol and 1 U/µl of Superscript II (Tang & Curthoys, 2001). Yeast tRNA were also added in competition assays (0.5 µg of tRNA; ratio probe:competitor 1:2 x10⁷).

RNA probes were previously heated at 95 °C in order to disrupt their secondary structure and left 5 min on ice so that their proper conformation was finally achieved. EMSA were performed using one-phase 5% polyacrylamide gels with TBE buffer (90 mM Tris-HCl pH 8.5, 110 mM boric acid, 2 mM EDTA) as running buffer and were left at 60-70 V for 90 min. The gel was then fixed with 20% ethanol, 10% acetic acid, for 30 min, introduced in a plastic-bag and visualized using a Personal Molecular Imager (Bio-Rad).

3.4 RESULTS

3.4.1 ADA4h particularities

First of all, a computational characterisation of the available activation domains (ADs) was carried out using online tools such as ProtParam (Gasteiger *et al.*, 2005) from the ExPASy webpage (<http://www.expasy.org/tools/protpar-ref.html>) (Table 3.1). All domains show a conserved length, although its composition is quite different. The pI for all domains is below neutrality; nonetheless, ADA4h presents a fairly basic pI (9.30) compensated by a highly acidic connecting helix. Thus, the entire activation segment retains an acidic pI, similarly to the other procarboxypeptidases. Apart from ADA4h, only the ADs of the recently described procarboxypeptidases A5 and A6 and the AD from PCPU approach neutrality (Barbosa Pereira *et al.*, 2002; Wei *et al.*, 2002).

Table 3.1. Summary of the theoretical isoelectric points for the different activation domains of the A/B subfamily carboxypeptidases. The theoretical pI have been calculated using the program ProtParam from the ExPASy webpage.

| Activation domain | Theoretical isoelectric point |
|-------------------|-------------------------------|
| ADA1h | 4.83 |
| ADA2h | 4.37 |
| ADA3h | 5.23 |
| ADA4h | 9.30 |
| ADA5h | 6.79 |
| ADA6h | 6.10 |
| ADBh | 5.42 |
| ADUh | 6.04 |

The basic character of ADA4h and its aggregation tendency made its purification rather difficult.

3.4.2 Structural search

We found a significant sequence alignment (e-value $1e-17$) using Position-Specific Iterated Basic Local Alignment Search Tool (PSI-BLAST) between ADA4h and U1 SNP1-associating protein (Q03714), a U1 type yeast protein involved in mRNA splicing.

The structure of ADA4h was assigned to the SCOP family *pancreatic carboxypeptidase, activation domain* within the ferredoxin-like fold. On the other hand, the U1 SNP1-associating protein (Q03714) was not found in the Protein Data Bank (PDB), therefore we modeled its structure using MODELLER and the structure of one of the members of the U1 small nuclear ribonucleoprotein A (U1A) family, the human U1 small nuclear ribonucleoprotein A (P09012) which structure is found with code 1M5Kc in the PDB (Rupert & Ferré-D'Amare, 2001), as template. The model was evaluated with ProsaII (Sippl, 1993) and was energetically favourable, suggesting that both sequences share the same structure (see supplementary material). Consequently, the domain of U1 SNP1-associating protein was assigned to the *canonical RNA binding domain (RBD)* SCOP family, also on the *ferredoxin-like* fold.

The structure of human U1 small nuclear ribonucleoprotein A (U1A) was superimposed with ADA4h. This superimposition allowed us to allocate the RNA molecule, co-crystallized with U1A in 1M5Kc, and to assume the putative binding between RNA and the activation domain of human procarboxypeptidase A4 (Fig. 3.2 B).

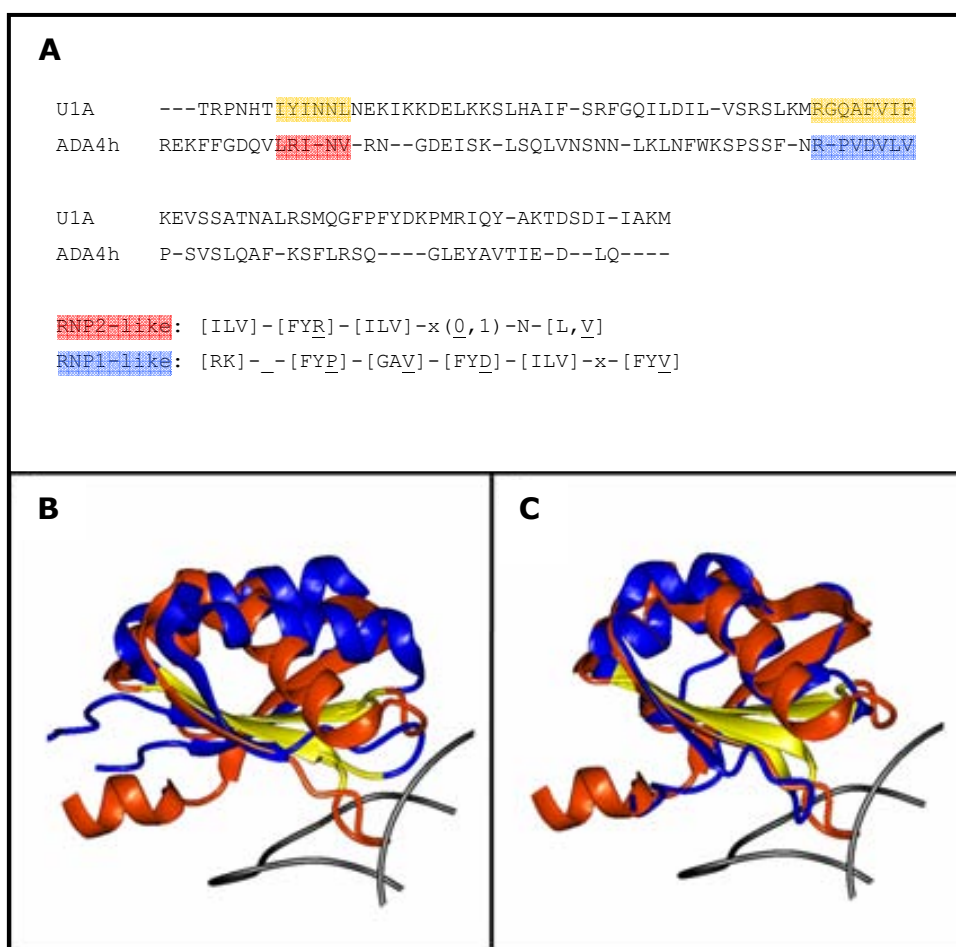


Figure 3.2. Structural superposition of ADA4h and U1A. (A) Sequence alignment derived from the structural superposition. The RNP regions of U1A are emphasized in yellow. RNP1 and 2 regions stand out in red and blue for ADA4h. The PROSITE pattern description is shown below the alignment and modifications from the original pattern are underlined. (B) Superposition of structures obtained from the structural alignment of the activation domain of human procarboxypeptidase A4 (in blue) with the RNA-binding protein U1A (in red). RNP regions of both structures are highlighted in yellow. (C) Structural alignment of RNA-binding protein U1A (in red) and the structural model of ADA4h (in blue) using the structure of U1A in 1M5Kc as template. RNP regions of both structures are highlighted in yellow.

Although the superimposition shows a putative binding region between ADA4h and RNA with two sequence patterns similar to the RNA Recognition Motif (RRM) of U1A (Fig. 3.2 A), there are conformational restrictions to accommodate the backbone chain of ADA4h to perform similar binding. The major differences between both backbone conformations are found in loops 1 and 3. We postulate that the orientation of these loops is dependent on the partner molecule that interacts with ADA4h, where the actual conformation of ADA4h binds the catalytic domain within procarboxypeptidase A4 while on the interaction with RNA it adopts a conformation similar to U1A. In agreement with this hypothesis, the main motion of the binding region was also correlated with the RNP1 and RNP2 sequences patches in loops 1 and 3 in a

previous work (Gargallo *et al.*, 2003). In order to test this hypothesis, we modeled ADA4h using the conformation of U1A as template and the structural alignment of the superimposition where the putative RNA-binding regions found on ADA4h corresponded to the RRM patterns of U1A. The new conformation of ADA4h was evaluated with the pseudo-energies of Prosall (Sippl, 1993). According to these statistical potentials, the modified orientation of loops 1 and 3, whose flexibility has been described before (Mittermaier *et al.*, 1999) was energetically possible but not favourable (Fig. 3.3).

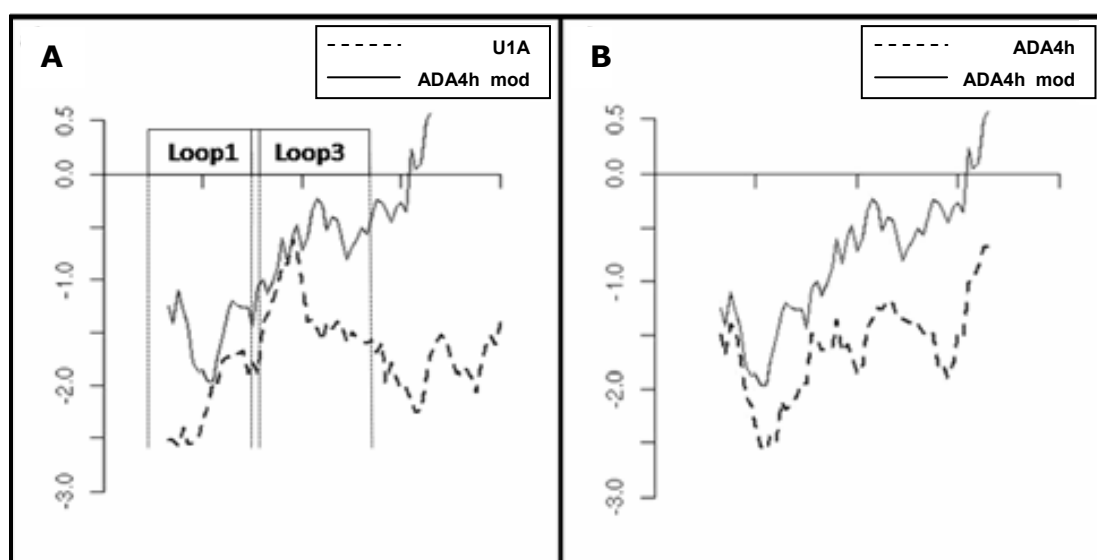


Figure 3.3. Evaluation of the structure of ADA4h modeled with the conformation of U1A. (A) Comparison between the pseudo-energy of U1A (dashed line) and the sequence of ADA4h thread on the structure of U1A (solid line), labeled as modeled ADA4h (ADA4h mod). Labels loop1 and loop3 contains the region of the evaluated model corresponding to each loop and confirms that the conformational change is possible. (B) Comparison between the original structure of ADA4h (dashed line) and the sequence of ADA4h thread on the structure of U1A (solid line).

We further compared the disposition of the side-chains of ADA4h in the binding region with the RRM residues of U1A to check whether they may perform similar binding with RNA (Fig. 3.4 A and B and 3.5 A and B). The comparison shows a comparable set of interacting residue-base pairs: in particular Tyr 13 of U1A interacts through a π interaction with the base C40 of RNA that can be emulated by the guanidine group of Arg 11 of ADA4h. Also, Phe 56 of U1A has an aromatic stacking with A41 of RNA (Nagai *et al.*, 1995), that would be weakened if the interaction is carried out by Asp48 of ADA4h (see further discussion).

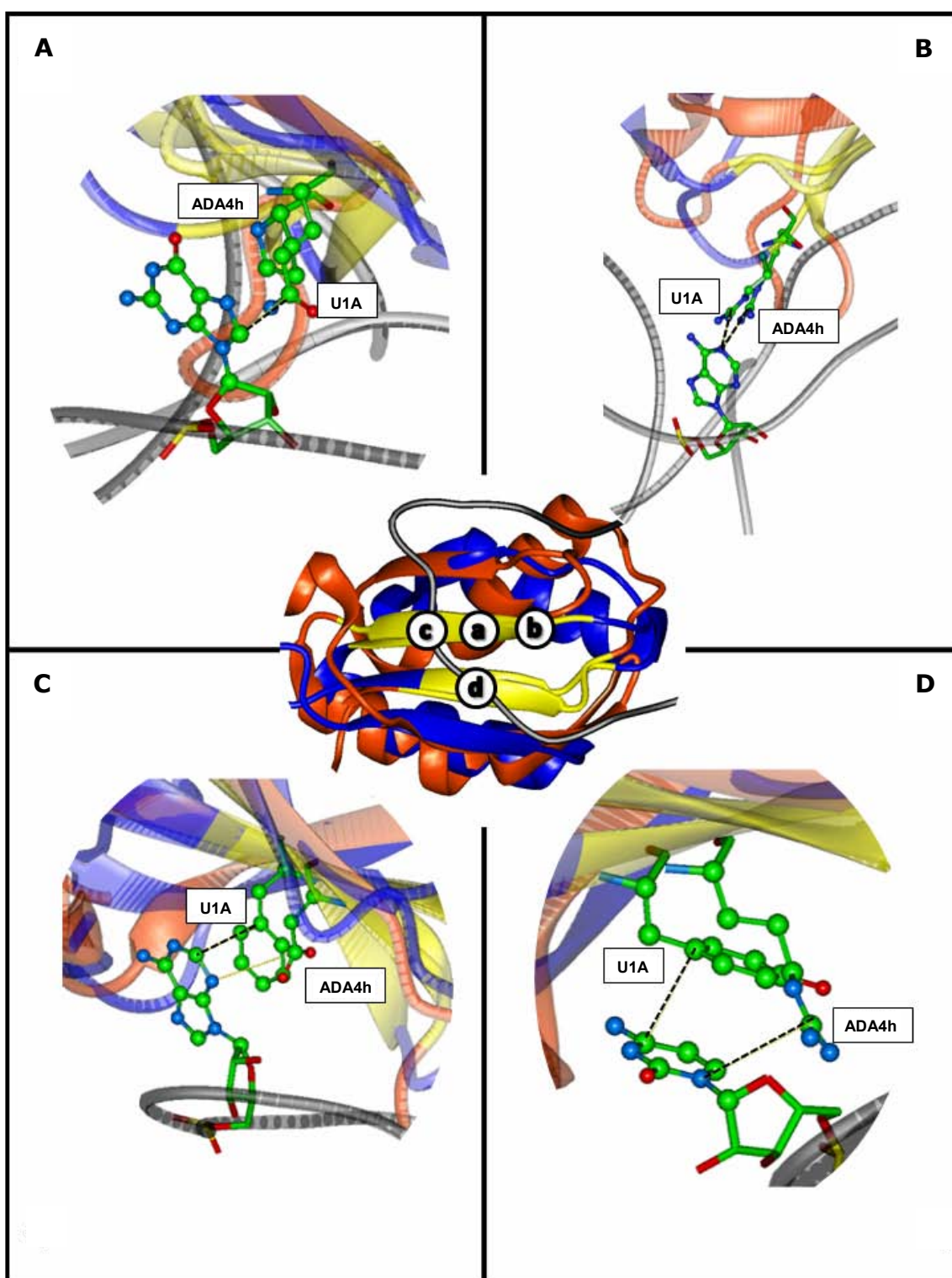


Figure 3.4. Interactions between RNA and U1A that can be done by ADA4h. (A) The stacking interaction performed by Gln54 of U1A (RNP1) with G39 of the RNA (3.15\AA) can not be carried out by Pro45 from ADA4h. (B) The hydrogen bond interaction of Arg52 of U1A (RNP1) with A36 (3.67\AA) is also performed by Arg45 in ADA4h (4.52\AA). (C) The interaction of Phe56 of U1A (RNP1) with A41 of the RNA (3.68\AA) can not be performed by Asp48 in ADA4h (4.37\AA). (D) The π interaction between Tyr13 of U1A (RNP2) and C40 of the RNA (3.52\AA) can be carried out by Arg11 of ADA4h (4.09\AA).

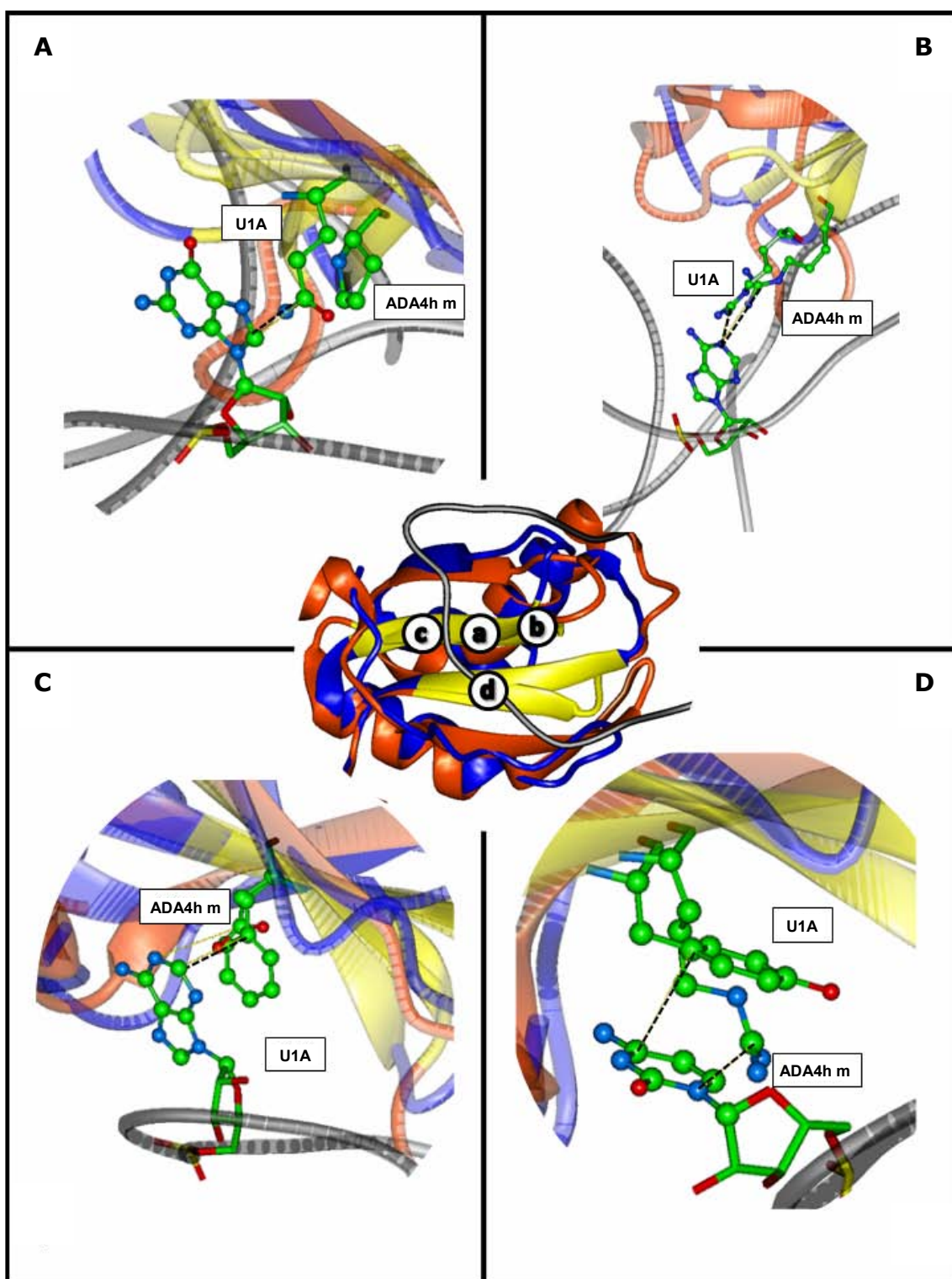


Figure 3.5. Interactions between RNA and U1A that can be done by the sequence of ADA4h threaded on the structure of U1A (ADA4h m). The change of loop orientation of the original structure of ADA4h improves some putative interactions with RNA. (A) The stacking interaction carried out by Gln54 of U1A (RNP1) with G39 of the RNA (3.15Å) can not be done by Pro45 from ADA4h. (B) The hydrogen bond interaction between Arg52 of U1A (RNP1) and A36 of the RNA (3.67Å) can be carried out by Arg45 in ADA4h (5.06Å). (C) The stacking of Phe56 of U1A (RNP1) with A41 of the RNA (3.68Å) can not be performed by Asp48 in ADA4h (4.01Å). (D) The π interaction carried out by Tyr13 of U1A (RNP2) with C40 (3.52Å) can still be done by Arg11 of ADA4h (2.23Å).

3.4.3 Pattern discovery

From the structural alignment, a pattern was identified performing the RNA recognition function. The RRM identified in 1M5Kc has been described as a combination of two regions (RNP1 and RNP2) in PROSITE. According to the sequence comparison from the structural alignment, these two regions present variants of the original sequences of U1A in the ADA4h sequence (Fig. 3.2 A). Furthermore, both regions are structurally located in the same position according to the structural alignment (Fig. 3.2 B and 3.2 C). Consequently, we propose a new RRM pattern that would characterise a putative binding of RNA.

The new RRM pattern was defined by variations on the sequences of RNP1 and RNP2 regions (Fig. 3.2 A). A PROSITE SCAN (de Castro *et al.*, 2006) search was performed on Uniprot database to find all sequences matching the new motif.

Table 3.2 shows the distribution of functions of the 70 new sequences found on Uniprot using the new RRM motif but not the original PROSITE RRM motif.

Table 3.2. Functional distribution of the new sequences found with the modified RNA recognition motif. The first six functions, excluding procarboxypeptidase precursor, are related with nucleotide recognition. It must be noted that 75% of the new sequences found with the modified RRM are related with this function.

| Function found | Frequency (%) |
|-----------------------------|---------------|
| Carboxypeptidase precursor | 2 |
| Nucleolin | 34 |
| Splicing factor | 10 |
| RNA binding | 16 |
| Nucleotide binding | 13 |
| DNA helicase | 2 |
| Polysaccharide biosynthesis | 1 |
| Dyoxigenase | 4 |
| Transmembrane component | 1 |
| Hypothetical protein | 17 |

A total of 48 sequences out of this 70 were aligned with the RMM_1 motif from Pfam (Finn *et al.*, 2006) under an e-value threshold of 10^{-5} . Moreover, the RRM_1 motif aligned with our motif only for 35 of these 48, while for 13 the alignment was located on a different position of the sequence. Consequently, the new pattern described by us predicts the RNA binding function for 32 sequences undetected by Pfam and may

imply a different binding behavior for 13 known RNA-binding proteins, while 50% confirmed the expected function of our new pattern.

Besides, 52 sequences out of these 70 perform functions involving RNA binding according to the information extracted from Swissprot. This implies that we have found an explanation for 4 sequences that could not be described by using Pfam. Finally, of the remaining 18 proteins, 5 contributed to functions unrelated with nucleotide binding and 13 had unknown function. ADA4h was among the 5 proteins unrelated with RNA-binding, according to Uniprot labelled functions. The corroboration of our prediction would imply that: 1) a new RMM motif type for RNA binding would be possible; and 2) ADA4h would have a moonlighting undetected function.

3.4.4 RNA Electrophoretic Mobility Shift Assays (EMSA)

The purified ADA4h protein was tested for RNA binding activity using the U1A binding-motif of stem-loop II of U1 small nuclear RNA (snRNA) (Allain *et al.*, 1997). A 70 mer RNA probe, named as bipolar-stem loop, was designed carrying two copies of the target motif. The presence of the correct secondary structure shown in Fig. 3.6 was checked by using the online program OligoAnalyzer 3.0 from Integrated DNA Technologies®. A free energy of -18.9 kcal/mol is calculated for this secondary structure.

The recently found yeast ζ -crystalline (ZTA1p), an RNA binding protein (Fernández *et al.*, 2007), and bovine serum albumin (BSA) were used as positive and negative controls, respectively (Fig. 3.6). A small amount of radioactivity was found in the shifted band which concomitantly increases with the concentration of protein (lanes 3-5) and was not present in the negative control (lane 2) neither when the cold probe was added as competitor in a ten fold excess (lane 1).

Another experiment with competitor tRNA was carried out in order to check the ADA4h specificity for the bipolar-stem loop (Fig. 3.6), and, although the shifted band is not as clear as in the EMSA without competitor tRNA (Fig. 3.6), still some shift can be observed (see lanes 3-4). The shifting also decreases for yeast ζ -crystalline as expected (lane 5).

These results confirm that ADA4h has RNA binding ability, although does not show a strong specificity for the bipolar-stem loop designed in this work.

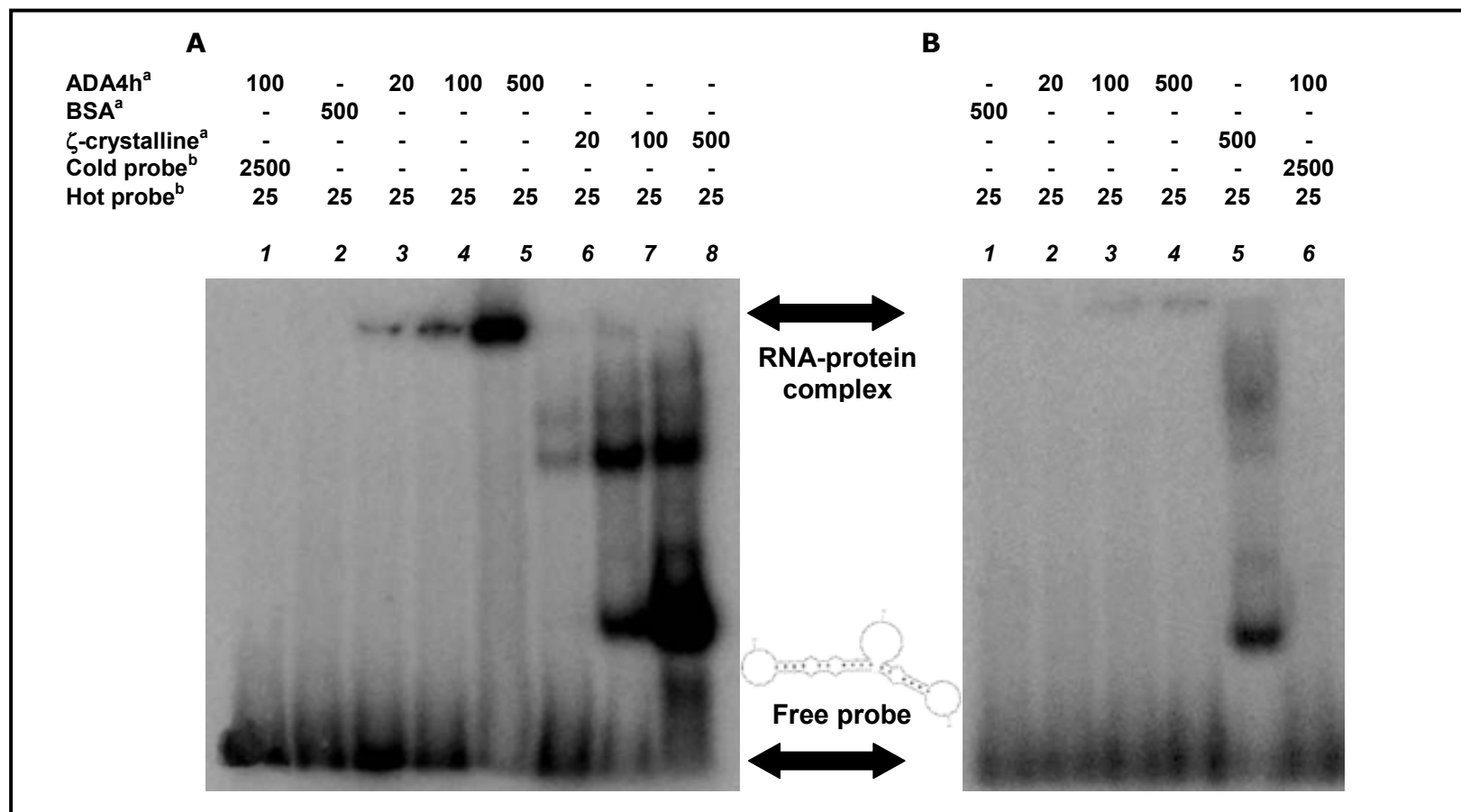


Figure 3.6. Unspecific and specific Electrophoretic Mobility Shift Assay (EMSA). (A) Unspecific EMSA. Where (^a) ng. (^b) fmol. Lane 1: ADA4h + cold probe. Lane 2: protein negative control (BSA). Lanes 3-5: increasing amounts of ADA4h + hot probe. Lane 6: increasing amounts of positive control protein (yeast ζ-crystallin) + hot probe. (B) Specific EMSA. (^a) ng. (^b) fmol. All lanes contains 0.5 μg of competitor tRNA. Lane 1: protein negative control (BSA). Lanes 2-4: increasing amounts of ADA4h + hot probe. Lane 5: protein positive control (yeast ζ-crystallin) + hot probe. Lane 6: ADA4h + cold probe. In both cases, part of the free probe was left to escape the gel in order to appreciate the faint retarded band. The secondary structure of the bipolar-stem loop is indicated between them. The structure was generated with OligoAnalyzer 3.0 from Integrated DNA Technologies®.

3.5 DISCUSSION

3.5.1 Definition of RRM

The RNA Recognition Motif (RRM) is one of the most abundant in eukaryotes, and is estimated to be present in 2% of the human gene products (Maris *et al.*, 2005). Initial biochemical studies on poly(A)-binding protein (PABP) and hnRNP protein C defined a consensus RNA binding domain of about 90 residues containing a central sequence of eight conserved residues (Adam *et al.*, 1986; Swanson *et al.*, 1987). The sequence [RK]-G-[FY]-[GA]-[FY]-[ILV]-X-[FY] was referred as RNP1 when a second consensus sequence, [ILV]-[FY]-[ILV]-X-N-L, was found in the N-terminus (Dreyfuss *et al.*, 1988). RNP2 is less conserved and shorter than RNP1. RNP1 contains mainly aromatic and basic residues whereas RNP2 is rather hydrophobic.

The search for structures related to ADA4h fold has rendered as the best score the U1A RNP (1M5Kc). RNP2 of U1A fits perfectly in the consensus with the sequence IYINNL. In the case of ADA4h, the RNP2 aligned sequence corresponds to: LRI-NV. The differences respect to the consensus sequence are labeled in italics and underlined. In position 2, there is an Arg instead of Phe/Tyr, although the aliphatic nature of Arg could also allow some extent of hydrophobic packing. Moreover, this Arg11 is interacting with Pro46 by a hydrogen bond, as does Tyr13 with Gln54 in the U1A protein (Jessen *et al.*, 1991). Position 4 is deleted in ADA4h sequence, lacking the consensus Asn but it is followed by another Asn. In position 6 we have a Val instead of a Leu, as is the case for Y14 and U2AF35 RRM s (Maris *et al.*, 2005). Albeit these differences, the sequence conserves four out of six residues and is quite similar to the RNP2 consensus sequence.

RNP1 sequence of U1A is RGQAFVIF, where Gln does not correspond to Phe neither Tyr. This particular Gln, Gln54, stacks on G42 of the Polyadenylation Inhibition Element (PIE) RNA in the crystal structure of the complex, performing the same function as any Phe or Tyr (Allain *et al.*, 1997). In the case of ADA4h, the corresponding sequence is R-PVDVLV, showing the lack of residue 2. Although just 3 residues out of 8 are strictly conserved, it could be considered that Val47 and Asp48 are not so different to [GA]-[FY] of the consensus sequence. In fact, an equivalent Val is found in PABP and in U2AF35 RNP1 sequences, and also an equivalent Asp is found in ALY RNP1 sequence (Maris *et al.*, 2005). Finally, residue at position 8 is not conserved as is the case for the polypyrimidine tract binding protein (PTB) RNP1 sequences A and C, LA, PABP and Nucleolin RNP1 sequence B (Maris *et al.*, 2005).

Taking all these criteria into account the sequence of RNP1 of ADA4h approaches to the consensus except for the lack of Gly in position 2.

It is important to note three different general features of RNA binding domains. Firstly, although some conserved aromatic residues are always found at the interface of the complex, the topology of the bound RNA is characteristic of each complex solved and the sequence-specificity can not easily be predicted. Secondly, RRMs that bind RNA show a wide spectrum of affinities. And finally, some RRMs are specific for protein binding rather than for RNA binding, suggesting that some RRM might have evolved from RNA to protein recognition (Kielkopf *et al.*, 2004).

According to these feature, Gly is a relevant residue of several nucleotide recognition motifs such as RRM (DeAngelo *et al.*, 1995; Kim *et al.*, 1997). Nevertheless, in our predicted RRM pattern this Gly residue is lost. Therefore, we have to deeply analyse the importance of Gly on the recognition and binding of RNA. Firstly, it is clear that the fact that Gly has not a side-chain indicates that this residue has to be involved in the flexibility and length of the pattern, improving the interaction or facilitating the specificity. Would that mean that the loss of Gly implies the loss of the RNA binding activity? The fact that our modified RNP1 motif was aligned several times (50% correctly aligned) with the RRM Pfam's domain shows that this is not the case. Moreover, the prediction of the putative function for new proteins and the experimental validation for one of them, clearly answers the question. Therefore, the role of Gly in the traditional RRM motif is to facilitate the binding, but this can still be plausible by conservation of other key residues, in particular positively charged and aromatic residues.

3.5.2 Binding specificity of U1A

U1A protein tertiary structure has been obtained in solution in complex with one of its specific RNA, the PIE from its pre-mRNA (Allain *et al.*, 1996; Allain *et al.*, 1997). The determinants for binding have been analyzed in detail and compared to those of the previously known crystal complex with stem loop II from U1-snRNA (Oubridge *et al.*, 1994). The complex with stem loop II was also solved by nuclear magnetic resonance (NMR). In both cases intermolecular recognition required conformational changes in the protein, shifting loop 3 and the C-terminal extension, and in the RNA, kinking it by induced fitting.

The general fold for U1A is a four-stranded β -sheet flanking on one side by two α -helices. The complex for PIE shows β -sheet surface interacting with the bases

whereas the phosphate backbone points towards the solvent, away from the protein. This implies that electrostatic interactions are not determinant for binding and that hydrophobic contact must be regarded as the main force fitting together both molecules. This is also the case for stem loop II.

PIE and stem loop II have a 7 nucleotide sequence that interacts with the β -sheet and the C-terminal extension of U1A. The RNP1 and RNP2 sequences of the β -sheet interact with this ssRNA in loop conformation. These interactions are also shown by other RNA protein complexes, and are mainly responsible for unspecific contacts. The fact that ADA4h contains RNP1 and RNP2 related sequences could account for the experimentally observed binding to bipolar-stem loop constructed in this work. Tyr13, Gln54 and Phe56 of U1A are stacking against the hydrophobic bases, and correspond to Arg11, Pro46 and Asp48 in ADA4h. Arg11 preserves the stacking interaction carried out by Tyr13, while the stacking of Gln54 with G39, which is likely to be specific for the interaction of U1A with PIE RNA (Allain *et al.*, 1997), can not be formed with Pro46. The stacking interaction of Phe56 with A41 can also be carried out by Asp48, although weakened. On the other hand, we hypothesize that the replacement of A41 by guanine would yield a putative salt bridge with Asp48 (Fig. 3.4 and 3.5). Finally, two hydrophobic interactions are important for U1A, Ser46 and Lys88; Pro40 and Glu75 of ADA4h align to them, respectively.

Concerning the contacts with the dsRNA structure, loops 1 and 3 are the ones interacting, and have been pointed out as the key residues for specificity. An important region in U1A is an unusual cluster of 6 basic residues. Lys20 is found aligned to Asn16 of ADA4h but is preceded by Arg15, a residue more basic than Lys. Lys22 and Lys23 have no basic counterpart in ADA4h. Although nor Arg47 neither Lys50 align with Lys38 of ADA4h, all they are located in loop 3 near to the RNA-binding interface. Finally, the most important basic residue in U1A and characteristic of RNP1, Arg52, perfectly aligns to Arg45 in ADA4h. Specificity in these loops is not a common feature of all RRM, since its size and distribution of basic amino acids is quite variable (Burd & Dreyfuss, 1994). Thus, it is fairly possible that ADA4h, with a cluster of 3 basic residues in loops 1 and 3, uses this mechanism for molecular discrimination to a different extent than U1A does.

Loop 3 is the most variable segment between RRM, and in the case of U1A complex is determinant for specificity. However, important residues in U1A have his counterpart in ADA4h: Arg52 (Arg45 in ADA4h) contributes with several hydrogen bonds and packs against Leu49 (Phe43 in ADA4h) and the pair CG closing the structure of the RNA. The residues positioning the dsRNA portion are Ser46 (Pro40 in ADA4h), Ser48 (Ser42 in ADA4h), Leu49 (Phe43 in ADA4h), and Arg52 (Arg45 in

ADA4h). This can account for the inability of 0.5 μg of competitor tRNA to totally abolish the complex between ADA4h with 25 fmol of bipolar-stem loop RNA, a ratio of $2 \times 10^7:1$, in the EMSA.

3.5.3 Moonlighting function for ADA4h

Apart from the relevance in the folding process, the known function for the activation domains of carboxypeptidases is the inhibition of the enzyme in the proenzyme form. The contacts in the crystal structure between ADA4h and its enzyme moiety involve one residue of RNP2 (Arg11), three residues corresponding to β -strand 2 (Asn35, Phe36, and Trp37) and one located in loop 3 (Lys38) of the domain (García-Castellanos *et al.*, 2005). The fact that important residues for inhibition are located in the same region determinant for RNA binding might induce ADA4h domain to evolve from the canonical RNP1 and 2 sequences. This hypothesis is based on the suggestion that some RRM might have evolved from RNA to protein recognition (Kielkopf *et al.*, 2004), as previously mentioned. It is likely that among the activation domain of procarboxypeptidases, ADA4h is the one that better conserves RNA binding properties. Apart from the structural approach presented in this work pointing out ADA4h as related to RNA binding proteins, it is noteworthy that ADA4h shares about 20% of homology with U1A, whereas ADA2h and ADB only share 11 and 3%, respectively, as determined using the multiple sequence alignment program CLUSTALW (Chenna *et al.*, 2003).

hPCPA4 is a procarboxypeptidase mainly expressed in cancerous prostate, brain and during development. Its gene is subject to genomic imprinting and its regulation is due to histone hyperacetylation, as observed in sodium butyrate treatment of androgen-independent prostate cancer PC-3 cell lines (Huang *et al.*, 1999), that lead to growth inhibition. Very low expression has been found in pancreas or other normal tissues. Recently, some proteins involved in different processes such as metabolism have been proved to bind RNA, participating in regulatory mechanisms (Ciesla, 2006). Although further experiments have to be performed, this could also be the case for ADA4h as pointed out in this work, where fold homology has revealed as a suitable tool for predicting protein function.

Annex to chapters

- The first chapter of this thesis has been done in collaboration with Dr. Luis Serrano's group at the European Molecular Biology Laboratory (EMBL). The 17 mutants corresponding to the previous studies of the transition state of the folding pathway of ADA2h had been constructed by Dr. Sandra Villegas. The electron microscopy assays have been performed by Dr. Alexandra Esteras Chopo at EMBL.
- The second chapter has been done in collaboration with Dr. José Luis Rodríguez Arrondo at the Unidad de Biofísica, Centro Mixto CSIC/UPV-EHU, in Bilbao. The infrared spectroscopy experiments have been carried out in tandem with Igor de la Arada at UPV.
- The bioinformatics part of the third chapter of this thesis has been executed in collaboration with Dr. Baldo Oliva's group from the Laboratori de Bioinformàtica Estructural (GRIB), at Universitat Pompeu Fabra – Institut Municipal d'Investigació Mèdica (IMIM), in Barcelona, in tandem with Jaume Bonet. The RNA binding assays were performed in collaboration with Dr. M. Rosario Fernández from the Yeast metabolism and molecular biology group, at Universitat Autònoma de Barcelona.

DISCUSSIÓ GENERAL

La present tesi ha estat centrada principalment en l'estudi de dos dominis d'activació de procarboxipeptidases humanes, enzims extensament caracteritzats pel nostre grup de recerca, tant des d'un punt de vista funcional-evolutiu com de la seva utilitat com a model d'agregació proteica. Tot i que l'objecte d'estudi sempre han estat dominis d'activació, aquesta tesi conté dues parts lleugerament independents: els estudis de l'agregació del domini d'activació de la procarboxipeptidasa A2 humana (ADA2h) per dicroisme circular i espectroscòpia d'infraroig, presentats en el primer i el segon capítols d'aquesta tesi, respectivament, i l'estudi d'una possible funció alternativa per part d'ADA4h, presentat en el tercer capítol.

Els dominis d'activació (AD) de les procarboxipeptidases humanes només es troben a la subfamília A/B de procarboxipeptidases, on s'hi inclouen les procarboxipeptidases pancreàtiques, els enzims més caracteritzats de tota la família (Vendrell *et al.*, 2004). Aquests dominis són estructures relativament grans, d'uns 80 residus, que inhibeixen el domini catalític de l'enzim mitjançant contactes amb el seu centre actiu. L'activitat d'aquestes carboxipeptidases ha d'adquirir-se mitjançant el tall produït per una proteasa específica a l'hèlix que connecta el domini d'activació amb el domini catalític, amb la consegüent eliminació del primer. Tots els dominis presenten una estructura tridimensional molt conservada del tipus antiparal·lela- α /antiparal·lela- β en forma de sandvitx obert, amb només pocs trets diferencials entre les formes A i B (Reznik i Fricker, 2001; Vendrell *et al.*, 2004). No obstant, les seqüències aminoacídiques dels dominis d'activació es troben més allunyades entre si que no pas les seves estructures (Aloy *et al.*, 1998). Els dominis estudiats en la present tesi són ADA2h, el qual forma part d'una procarboxipeptidasa de tipus pancreàtic, i ADA4h, que forma part d'una carboxipeptidasa amb funció encara desconeguda, però suposadament relacionada amb processos cancerígens i d'apoptosi.

ADA2h és un domini que ha estat profundament caracteritzat en treballs anteriors, tant a nivell d'estructura (García-Sáez *et al.*, 1997; Jiménez *et al.*, 2003), com a nivell de plegament (Villegas *et al.*, 1995; Villegas *et al.*, 1996; Viguera *et al.*, 1997; Villegas *et al.*, 1998; Villegas *et al.*, 2000). En estudis posteriors, es va poder comprovar la seva capacitat de formar fibres amiloides riques en conformació- β si el pH del seu medi era suficientment disminuït (pH 2 – 3), representant, doncs, una 'fugida' del seu camí de plegament normal (Villegas *et al.*, 2000). No obstant, els mutants creats per a estabilitzar ambdues hèlixs- α de la proteïna eren incapaços de formar aquest tipus d'estructures.

Aquest últim estudi va representar el punt de partida de la present tesi, amb la finalitat d'ampliar els coneixements de la formació de fibres amiloides per part d'ADA2h. L'agregació d'ADA2h ha estat estudiada des del punt de vista cinètic mitjançant dues tècniques biofísiques especialitzades principalment en el seguiment de les estructures secundàries de les proteïnes: el dicromisme circular (CD) i l'espectroscòpia d'infraroig amb transformada de Fourier (FTIR). Aquestes dues tècniques permeten una visió global dels canvis produïts a nivell d'estructura secundària, ja que les característiques específiques de cada una permeten obtenir informació complementària més que no pas redundant. Les diferències principals entre aquests dos estudis han sigut l'estat conformacional inicial, la concentració proteica i el mètode d'inducció de l'agregació. Els experiments de CD van ésser desenvolupats a partir d'un estat extensivament desnaturalitzat d'ADA2h, mitjançant urea 4 M i temperatura relativament elevada (60 °C). D'aquesta manera s'eliminava gran part de la contribució de l'estabilitat i l'estructura, i es maximitzava, en canvi, la contribució de la seqüència polipeptídica i les seves propietats físico-químiques. Es permetia, llavors, la comparació dels resultats obtinguts amb mètodes predictius basats en les propietats de la seqüència primària (Chiti *et al.*, 2003; Fernández-Escamilla *et al.*, 2004; Tartaglia i Vendruscolo, 2008). Oposadament, el punt de partida conformacional escollit per als experiments de FTIR va ésser el més proper possible a l'estat natiu en les condicions de l'assaig, per tal de poder observar l'evolució de l'estructura al llarg de tot el procés. Una aproximació similar ja havia estat desenvolupada en els estudis de desplegament d'altres proteïnes pel grup del Dr. José Luis Rodríguez Arrondo, de la Universitat del País Basc, amb el qual es va col·laborar per a la realització d'aquests experiments (Iloro *et al.*, 2004; Iloro *et al.*, 2005). La concentració d'ADA2h emprada ha estat notablement diferent en ambdues aproximacions, de l'ordre de 40 μ M (0,37 mg/ml) en els experiments de CD i de 544 μ M (5 mg/ml) en els de FTIR. En aquest últim cas, la raó principal d'utilitzar concentracions proteiques tan elevades ha estat la menor sensibilitat de la tècnica de FTIR. De la mateixa manera, la selecció diferencial del mètode d'inducció de l'agregació també ha estat conseqüència de la gran concentració proteica necessària per al FTIR i de les limitacions de manipulació de la mostra en aquesta tècnica. Per als assaigs de CD la inducció va ésser realitzada per salt de pH, és a dir, per disminució de pH de la mostra mitjançant l'addició d'una petita quantitat d'àcid clorhídric concentrat. En el cas del FTIR, l'elevada concentració proteica produeix una acceleració dràstica de la velocitat de l'agregació, no compatible amb la col·locació de la mostra dins de la cel·la i el posterior tancament d'aquesta per a la seva anàlisi. Degut a això, el pH de la mostra va ésser disminuït suaument mitjançant diàlisi enfront tampó glicina a pD 3 i la inducció es va donar dins de l'aparell de FTIR

per augment de temperatura. Un protocol molt similar a aquest havia estat ja emprat en els primers estudis de l'agregació d'ADA2h (Villegas *et al.*, 2000). Aquests trets distintius eren suficients per a produir processos d'agregació sensiblement diferents. Una de les principals conseqüències dels diferents tractaments de la mostra per a l'anàlisi en ambdues tècniques podia observar-se en els estats d'agregació inicials observats per microscòpia electrònica de transmissió (TEM). Mentre que els agregats amorfs eren les espècies principalment presents al començament de les cinètiques d'agregació per CD, degut a l'estabilització de l'intermediari tendent a la agregació per condicions desnaturalitzants; unes fibril·les curtes i en forma de cuc (referides normalment com *worm-like*, WL) semblaven ser les pobladores majoritàries a l'inici de les cinètiques seguides per FTIR. A més a més, la banda de l'espectre inicial corresponent a les hèlixs- α d'ADA2h es trobava desplaçada cap a números d'ona menors, molt propers a la zona específica pertanyent a conformacions desestructurades. Per tant, en aquesta segona aproximació cinètica, la major concentració proteica, i el pH àcid des de l'inici dels experiments, evitaven el manteniment d'un estat conformacional inicial natiu, a diferència del que es desitjava. Agregats WL semblants han estat visualitzats per part del grup de Radford i col·laboradors per a una proteïna amiloidogènica, la β_2 -microglobulina, emprant condicions de medi lleugerament diferents a les que promouen l'amiloidogènesi (Gosal *et al.*, 2005; Jahn *et al.*, 2008).

El seguiment de les cinètiques d'adquisició d'estructura- β per CD va revelar l'existència de dues etapes ben diferenciades, ja que les dades podien ésser ajustades a funcions exponencials decreixents dobles. El fet d'observar dues velocitats diferents indicava la presència de múltiples poblacions d'agregats al llarg de la coordenada de reacció. Aquesta pluralitat d'estats agregats també ha estat observada per a la proteïna acilfosfatasa (AcP), la qual presenta un plegament similar a ADA2h (Calamai *et al.*, 2005). No obstant, en el treball anterior les velocitats van ésser inferides a partir de l'estudi del procés invers, és a dir, de desagregació d'AcP partint de dos estats agregats diferents. Tant les cinètiques d'ADA2h presentades en aquesta tesi, com les d'AcP, emperò, assenyalaven que la segona etapa es corresponia amb una reorganització conformacional del primer estat agregat, ja que la segona velocitat que la governa és pràcticament independent de la concentració proteica (Plakoutsi *et al.*, 2004; Calamai *et al.*, 2005). Aquesta reorganització també va poder ésser detectada per a ADA2h en l'etapa final de la cinètica de FTIR, un cop les grans conversions d'estructures havien esdevingut i es produïa l'evolució de l'estructura cap a l'estat final. Un cop acabades les cinètiques seguides per CD i FTIR es va visualitzar per TEM el

material assajat, observant-se fibres de tipus amiloide en ambdós casos, demostrant, doncs, que es tractava del seguiment d'un procés de formació de fibres amiloides i no simplement d'una agregació amorfa o d'un increment dels agregats WL en el experiments a alta concentració. És destacable la formació de fibres amiloides en condicions experimentals dràsticament diferents, fet que apunta que encara que el punt de partida sigui diferent, s'arriba al mateix estat d'agregació ordenada.

A les cinètiques d'FTIR es va realitzar el seguiment de tot l'espectre corresponent a la banda de l'enllaç amida-I proteic, podent-se observar, a cada temps, les diferents estructures secundàries d'ADA2h i la seva evolució. La generació contínua d'espectres permetia obtenir informació sobre els canvis globals produïts a tot el domini al llarg del procés d'agregació, contraposadament als estudis de CD, on només se seguia una longitud d'ona al llarg del temps, la corresponent a l'estructura- β . El seguiment d'espectres complets no podia ésser efectuat pel CD degut a que la major lentitud d'aquest aparell en l'elaboració d'un espectre complet hagués evitat la detecció acurada de l'eminent velocitat de la primera etapa del procés d'agregació. D'entre tot l'espectre de l'amida-I, la banda que aporta més informació sobre la formació de la fibra amiloide correspon al número d'ona inicialment situat a 1619 cm^{-1} . Experiments recents amb la proteïna amiloidogènica β_2 -microglobulina han demostrat, a més, l'associació de números d'ona propers a 1619 cm^{-1} amb els agregats en forma de cuc (WL) (Jahn *et al.*, 2008). Aquesta banda evoluciona al llarg de la formació de la fibra amiloide cap a 1616 cm^{-1} , reflectint la compactació de l'estructura- β de la fibra, de manera similar a l'observada anteriorment per a la proteïna transtirretina (Zandomenighi *et al.*, 2004). Les cinètiques van permetre observar una cinètica d'evolució de l'àrea de la banda de 1619 cm^{-1} , corresponent a l'estructura β -agregada, amb 3 fases clarament diferenciades. La primera etapa reflectiria la dissolució dels agregats WL referits anteriorment, la segona etapa correspondria al procés de nucleació de la fibra amiloide i l'etapa final la reorganització d'aquests nuclis que comportaran la posterior elongació de la fibra. Estudis realitzats pel grup de la Dra. Sheena E. Radford van concloure que els agregats tipus WL i les fibres eren els estadis finals de 2 camins d'agregació diferents, que es trobaven en competència entre ells, i alhora, competien amb el camí de plegament (Gosal *et al.*, 2005; Jahn i Radford, 2005). En el nostre cas, l'augment de temperatura produïa la dissolució d'aquests agregats primerencs i el desplaçament de la competència cap a l'afavoriment, en aquestes noves condicions, de la formació de fibres amiloides. En el tractament de les dades cinètiques es va emprar també l'anàlisi de correlació bidimensional, determinant els mapes sincrònics per al procés de fibril·logènesi (Iloro *et al.*, 2004; Iloro *et al.*, 2005;

Noda, 2007). Aquests mapes correlacionen les intensitats dels diferents espectres obtinguts amb la tècnica de FTIR al llarg del temps, i ens permeten observar les relacions entre les bandes que es produeixen sincrònicament, és a dir, en fase. En el cas d'ADA2h s'observava que la principal correlació es donava entre la banda corresponent a l'estructura β -agregada (1619 cm^{-1}) i la banda de les hèlixs- α parcialment desplegadas (1647 cm^{-1}), essent aquesta correlació de signe negatiu. La naturalesa negativa de la correlació indicava que la formació d'aquesta estructura β -agregada provenia, en gran part, de la desaparició de les hèlixs- α d'ADA2h. D'altra banda, s'observen correlacions entre les diverses bandes característiques de les fulles i els girs- β .

Els assaigs de les cinètiques d'agregació estudiades per CD i FTIR van ser complementats amb l'elaboració i anàlisi de variants d'ADA2h, cadascuna amb una única mutació aminoacídica. Part dels mutants havien estat prèviament dissenyats per a l'estudi de l'estat de transició de la via de plegament d'ADA2h, de manera que a cada un es trencava una interacció present a l'estat natiu (Villegas *et al.*, 1998). Això significava que tots ells es trobaven desestabilitzats respecte la variant salvatge (*wild-type*, WT). Estudis d'agregació a l'equilibri amb aquests mutants van demostrar que alguns d'ells eren incapaços de formar l'estructura β típica de les fibres amiloides a iguals condicions de medi i concentració proteica ($20\ \mu\text{M}$) que la variant salvatge. No obstant, un augment de la concentració proteica fins a 10 vegades ($200\ \mu\text{M}$) permetia l'aparició de l'estructura β -amiloide. Aquest fet demostrava que l'estabilitat, tot i ser important, no és l'únic factor que influeix a l'agregació proteica, com en un principi semblaven suggerir els experiments realitzats amb els mutants amb hèlixs- α estabilitzades d'ADA2h (veure més amunt i Villegas *et al.*, 2000) i algunes malalties causades per mutacions desestabilitzants a proteïnes amiloidogèniques (Hurle *et al.*, 1994; Booth *et al.*, 1997; Chiti *et al.*, 2000; Smith *et al.*, 2003). La relació directa entre estabilitat i agregació va ser descartada per Chiti i col·laboradors per a la proteïna acilfosfatasa durant la realització d'aquesta tesi doctoral (Chiti *et al.*, 2003). L'observació de diferents comportaments en els mutants puntuals va originar l'estudi de les contribucions a l'agregació de cada regió d'ADA2h mitjançant diferents pèptids que comprenien parts de la seva seqüència. Aquests pèptids van ser dissenyats segons el perfil d'agregació obtingut pel programa TANGO desenvolupat pel grup del Dr. Luis Serrano al Laboratori Europeu de Biologia Molecular (Fernández-Escamilla *et al.*, 2004), amb el qual es va col·laborar. TANGO és un algorisme computacional dirigit a la predicció de regions nucleadores de l'agregació a cadenes polipeptídiques desplegadas. TANGO predeia principalment les regions N i C-terminal d'ADA2h com

les més propenses a agregar en conformació- β , especialment la regió C-terminal. Els resultats tant de CD com de TEM dels pèptids van assenyalar la regió que comprèn la cadena β -4 d'ADA2h, situada a l'extrem C-terminal, com la responsable de la formació de fibres amiloides. Es demostrava, doncs, que la regió implicada en la nucleació de l'agregació era diferent de la zona implicada en la nucleació del procés de plegament (cadena β -2 i β -3, i part de l'hèlix α -2, principalment, Villegas *et al.*, 1998), i que les condicions del medi i concentració proteica afavoririen diferencialment la importància relativa de cada una. Aquesta observació coincidia amb la presència de també 2 zones diferenciades a la proteïna acilfosfatasa, una que dirigia el procés de plegament, i l'altra el procés d'agregació (Chiti *et al.*, 2002).

Amb la informació obtinguda amb l'algorisme TANGO, es van crear 11 mutants puntuals nous que, units als creats anteriorment, formaven una bateria que proporcionava una visió global de quasi tota la seqüència de la proteïna. Només el mutant K63A no va poder ésser obtingut en quantitats acceptables, la qual cosa s'adiu amb la predicció de l'algorisme TANGO, el qual atorga a aquest mutant la major propensió d'agregació. La majoria de les cinètiques d'agregació seguides per CD dels mutants presenten les mateixes dues etapes ben diferenciades observades per al WT: el procés de nucleació i el de reorganització. No obstant, alguns dels mutants no mostraven un augment d'estructura- β , de manera similar al que s'havia observat en els estudis a l'equilibri a dues concentracions de proteïna (veure paràgraf anterior). Els resultats cinètics de CD obtinguts amb els mutants van reafirmar la importància de la cadena β -4 en el procés d'agregació, ja que algunes mutacions en aquesta regió eren capaces o bé d'augmentar significativament la velocitat del procés agregatiu, o bé d'abolir-la totalment, almenys a l'escala de temps assajada. Finalment, les diferents velocitats de nucleació (k_1) van ésser comparades amb les prediccions de diferents programes i fórmules. Els resultats més satisfactoris es van obtenir amb l'algorisme TANGO (Fernández-Escamilla *et al.*, 2004), ja que aquest té en compte el context de seqüència en el càlcul de la propensió d'agregació en estructura- β . No obstant, TANGO només era efectiu quan atorgava a mutants prediccions de propensions d'agregació substancialment diferents a la variant salvatge, en canvi, predeïa propensions d'agregació similars al WT a una sèrie de mutants que experimentalment presentaven velocitats d'adquisició de conformació- β diferents a la variant salvatge. Aquesta observació segurament era deguda a una falta de reconeixement d'altres regions de la proteïna de menor importància en el procés agregatiu. Cal destacar que les correlacions obtingudes amb el paràmetre Z_{agg} del programa ZyggregatorTM van ésser també destacables, però la significació es veia dràsticament disminuïda degut a

la predicció notablement errònia d'un dels mutants. Un altre problema de la comparació realitzada és la presència probable d'estructura residual a ADA2h, tot i les condicions extensivament desnaturalitzants emprades, de la mateixa manera que ha estat demostrat en altres proteïnes en condicions similars (García *et al.*, 2001). L'existència d'estructura residual no permetria, llavors, una comparació directa exempta d'un cert error de les dades experimentals amb mètodes predictius basats simplement en la seqüència polipeptídica. TANGO i ZyggregatorTM van essent implementats periòdicament, tant amb les dades experimentals que es van publicant, a TANGO, com amb la introducció de nous paràmetres que complementen i augmenten l'eficàcia de la predicció, a ZYGGREGATORTM (TANGO es troba disponible actualment a <http://tango.crg.org> i ZYGGREGATORTM a <http://www-vendruscolo.ch.cam.ac.uk/zyggregator.php>). Amb TANGO també es van calcular les regions més implicades en l'agregació d'altres proteïnes de topologia similar a ADA2h, observant-se que la contribució en l'agregació de les regions N i C-terminal no és un tret comú de totes elles.

Per als estudis d'FTIR només van ésser estudiats 3 mutants, els quals foren escollits segons les velocitats d'agregació observades en els assaigs de CD: un mutant amb velocitat d'agregació superior a la del WT (N58A), un amb velocitat inferior (I23V), i finalment, un dels que no presentava adquisició d'estructura- β (F65A). A part de la ja esmentada presència d'agregats inicials WL, l'augment important de la concentració proteica (544 μ M) produïa l'agregació del mutant F65A, el qual no havia presentat un increment d'estructura- β en les cinètiques mesurades per CD (a concentració 40 μ M). Aquesta observació ja semblava ser intuïda pels estudis de CD a l'equilibri i a dues concentracions realitzats anteriorment, on la presència d'estructura- β a certs mutants només podia ésser visualitzada quan les condicions eren forçades (a 200 μ M de proteïna). La inducció de l'agregació a aquest mutant a 544 μ M, a l'igual que als altres dos, porta a la formació de fibres amiloides visibles per TEM. De la mateixa manera que el WT, tots els mutants presentaven ja la banda corresponent a estructura β -agregada (encara que desplaçada) a temps zero, a més de fibres WL visibles per TEM en dos d'ells. Aquest grau d'agregació inicial és diferent entre totes les variants assajades, i correlaciona força bé amb l'estabilitat proteica. Els espectres corresponents als mutants presenten bandes generalment més amples, segurament degut a la menor estabilitat i, per tant, a una estructura menys definida d'aquestes variants. Aquesta observació es feia més aparent als mapes bidimensionals, on les correlacions entre bandes són similars a les trobades per al WT, encara que més extenses. Els estudis cinètics, per contra, van mostrar que tot i les diferents

propensions d'agregació de les variants, els números d'ona i els percentatges d'àrea semblaven dirigir-se cap a una convergència de valors, la qual cosa es veia reflectida en la formació de fibres amiloides molt similars per part de totes elles, sobretot a temps més avançats.

És interessant destacar que les fibres amiloides d'ADA2h s'han obtingut en condicions inicials molt diferents: a partir de l'estat desnaturalitzat, a baixa concentració proteica i pH neutre, i a partir d'un estat plegat a pH àcid, en el qual es visualitzen agregats WL al TEM, degut especialment a l'alta concentració de proteïna. Per altra banda, encara que la via d'agregació fos inicialment diferent per als mutants d'ADA2h assajats, es va arribar en tots els casos a la formació de la mateixa fibra amiloide. A més a més, en totes les condicions estudiades s'observa una reorganització conformacional posterior a la nucleació que inicia el procés d'agregació. Aquesta reorganització conformacional és la que permet l'elongació del nucli d'agregació per tal d'obtenir la fibra amiloide final.

La utilització d'eines bioinformàtiques va significar un recurs molt valuós en l'estudi funcional d'ADA4h, ja que va ajudar a centrar la cerca de funcions del domini cap a la unió de RNA. Per a aquesta part de la tesi es va triar el domini d'activació de la procarboxipeptidasa A4 humana, degut a que les primeres observacions van mostrar-ne una singularitat respecte els altres dominis d'activació coneguts: és la proteïna amb un major punt isoelèctric (9,3), fins a 2 punts superior a qualsevol d'ells. L'objectiu d'aquesta part era estudiar la possibilitat d'una funció alternativa per a ADA4h, ja que cada cop són descobertes noves funcions alternatives per a enzims profundament caracteritzats, com les d'unió a RNA per part d'enzims metabòlics (Ciesla, 2006).

Les primeres aproximacions bioinformàtiques, en col·laboració amb el grup del Dr. Baldo Oliva de l'Institut Municipal d'Investigació Mèdica, es van fer a nivell de seqüència, per tal d'identificar relacions amb altres proteïnes amb funció diferent a carboxipeptidasa. Les cerques es van dur a terme mitjançant PSI-BLAST (Altschul *et al.*, 1997) sobre la base de dades Uniprot (Bairoch *et al.*, 2005). Les cerques no van resultar satisfactòries fins que el llindar de significació (*E-value*) de la primera iteració va ser augmentat prou com per a que la primera proteïna amb funció diferent a carboxipeptidasa pogués introduir-se a la matriu de similitud. A les següents iteracions, el llindar de significació va tornar a ésser restablert als seus valors per defecte fins que

l'algorisme va assolir la convergència. Els resultats obtinguts van assenyalar una semblança seqüencial remota d'ADA4h amb un domini d'una proteïna de llevat implicada en l'*splicing*. Com que no es disposava de l'estructura d'aquesta proteïna de llevat, aquesta última es va modelar utilitzant el programa MODELLER (Sali i Blundell, 1993) sobre l'estructura de la proteïna d'*splicing* U1A (Oubridge *et al.*, 1994), un domini d'unió a RNA (*RNA Recognition Motif*, RRM). El model va ésser avaluat amb el programa ProSall (Sippl, 1993) i va poder ser catalogat com a energèticament favorable, de manera que la proteïna d'*splicing* U1A i aquest domini de la proteïna de llevat compartien la mateixa estructura i, per tant, la mateixa família a la base de dades SCOP (*Structural Classification Of Proteins*, Murzin *et al.*, 1995). A partir d'aquest moment es va emprar l'estructura de la proteïna U1A, resolta mitjançant difracció de rajos-X en complex amb un RNA específic, per tal d'avaluar la seva similitud estructural amb ADA4h emprant el programa STAMP (*Structural Time series Analyser Modeller and Predictor*, Russell i Barton, 1992). La sobreimposició va permetre observar diferències entre les seqüències consens d'unió al RNA presents a U1A (RNP1 i RNP2) i les regions equivalents a ADA4h, als llaços 1 i 3, predient-se una unió més dèbil de l'àcid nucleic per part del domini d'activació. Amb aquesta informació es va crear una nova variant del patró d'unió a RNA i es va assajar mitjançant una cerca a la base de dades Uniprot. La cerca amb el nou motiu va identificar un conjunt de seqüències que encaixaven dins d'aquest motiu i no en el motiu clàssic d'unió al RNA. Més del 50% d'aquestes seqüències trobades estaven relacionades amb unió a RNA. Els resultats obtinguts bioinformàticament es van voler verificar experimentalment i amb aquesta finalitat es va realitzar un assaig d'unió a RNA d'ADA4h utilitzant com a RNA el que apareixia a l'estructura tridimensional resolta d'U1A. Es va dissenyar un RNA que contenia 2 còpies de la seqüència diana separades per un segment poli-A, de manera que fossin capaces de formar la mateixa estructura secundària que la que pot visualitzar-se al complex amb U1A. Als gels de retardament, (*Electroforetic Mobility Shift Assays*, EMSA) es podia observar una banda de mobilitat inferior, la qual augmentava d'intensitat a mida que s'incrementava la concentració d'ADA4h. Aquesta banda es corresponia a la sonda de RNA marcada radioactivament unida a ADA4h. No obstant, l'addició de tRNA de llevat com a competidor era capaç d'abolir quasi completament l'aparició de la banda de retard, suggerint que la unió d'aquest RNA per part d'ADA4h no seria del tot específica. Algunes proteïnes de demostrada capacitat d'unió a RNA presenten diferències respecte els motius RNP1 i RNP2 de Pfam (Maris *et al.*, 2005), com és el cas d'U1A, i en alguns casos, s'observen proteïnes que tenen patrons RNP1 o RNP2-like, els quals han evolucionat cap a la interacció amb altres proteïnes més que no pas amb àcids

nucleics (Kielkopf *et al.*, 2004). ADA4h presenta diverses diferències respecte els patrons esmentats o altres residus d'U1A, les quals han pogut ésser avaluades gràcies a la gran informació estructural disponible d'aquesta última en complex amb RNA (Oubridge *et al.*, 1994; Allain *et al.*, 1996; Allain *et al.*, 1997). La diferència probablement més notable entre les seqüències consens i les presents a ADA4h resideix en l'absència de l'aminoàcid glicina en el RNP1 d'aquesta última, el qual ha demostrat ser un residu rellevant en alguns RRM (DeAngelo *et al.*, 1995; Kim *et al.*, 1997). Aquest aminoàcid no es troba present en interaccions amb altres residus, per tant, es creu que la seva importància seria més aviat dirigida cap a atorgar flexibilitat o la longitud necessària a la seqüència per a que es donin les interaccions amb el RNA pertinents i, per tant, facilitant-ne la unió. Si relacionem la funció d'inhibició realitzada per ADA4h dins el proenzim amb la hipotetitzada mitjançant bioinformàtica, podem veure que les regions predites com a llocs rellevants en la interacció amb el RNA comprenen residus que es troben contactant amb el domini catalític del proenzim. D'aquesta manera, la hipòtesi més plausible seria que els dominis d'activació de les carboxipeptidases haurien pogut provenir d'un domini d'unió a RNA ancestral que hauria evolucionat cap a maximitzar la inhibició de l'enzim carboxipeptidasa, degut a la perillositat d'aquestes sense una regulació estricta. Aquesta evolució hauria comportat com a conseqüència la pèrdua de part de la seva capacitat d'unió a l'àcid nucleic. Aquesta pèrdua és evident als dominis pacreàtics, que mostren un grau d'identitat amb U1A molt més baix que ADA4h.

El fet de trobar homologies estructurals remotes i poder validar la funció experimentalment, encara que amb un RNA que no és probablement la diana fisiològica, obre les portes cap a la cerca de RNAs que fisiològicament tinguin la possibilitat d'interaccionar amb ADA4h. A més, l'increment de l'expressió d'ADA4h en cèl·lules canceroses que han estat tractades per a modificar el grau d'acetilació de la cromatina, és a dir, sota canvis d'expressió gènica, crea un context fisiològic molt interessant degut a la possibilitat de que aquesta proteïna realitzi una funció rellevant.

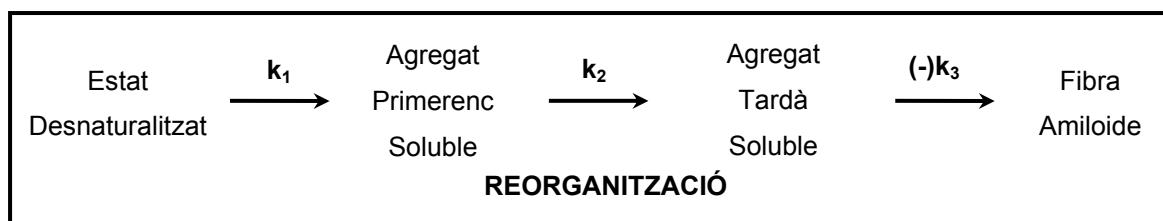
CONCLUSIONS

Capítol 1: La cinètica primerenca de formació de fibres amiloides revela la presència d'una reorganització conformacional dels agregats inicials

1. El programa de predicció de la tendència d'agregació TANGO assenjala la seqüència corresponent a la cadena β -4 del domini ADA2h com la regió més important en la direcció del procés agregatiu.
2. L'estudi de dicroisme circular (CD) i de microscòpia electrònica de transmissió (TEM) d'un pèptid representatiu de la regió N-terminal d'ADA2h i de varis representatius de la regió C-terminal han permès confirmar la funció clau de la seqüència aminoacídica corresponent a la cadena β -4 en la formació de fibres amiloides.
3. 11 nous mutants puntuals d'ADA2h han estat construïts amb l'ajuda del programa TANGO, ampliant i complementant la bateria construïda prèviament per a l'estudi de l'estat de transició del plegament d'ADA2h. La finalitat de la construcció d'aquest gran nombre de mutants ha estat l'avaluació dels efectes sobre la velocitat d'agregació produïts per mutacions puntuals a diferents regions del domini.
4. S'han expressat i purificat en quantitats suficients tant els nous mutants puntuals com els construïts prèviament, amb excepció de K63A, el qual no va poder ésser obtingut en forma soluble.
5. La disminució de l'estabilitat proteica d'ADA2h no es troba directament relacionada amb un augment de la velocitat d'agregació, ja que alguns mutants desestabilitzats són incapaços d'adquirir l'estructura agregada en β típica de les fibres amiloides en les mateixes condicions i concentració proteica que la variant salvatge (WT).
6. S'han posat a punt les condicions inicials de l'experiment, emprant urea 4 M i una temperatura de 60 °C, que han assegurat que tant la variant salvatge d'ADA2h com els seus mutants tinguessin el mateix grau de desplegament.
7. S'ha desenvolupat un mètode per a estudiar les cinètiques d'agregació d'ADA2h mitjançant CD, el qual permet l'observació de l'adquisició de senyal d'estructura secundària- β al llarg del temps. La inducció de l'agregació es va dur a terme per

salt de pH, i es va haver de realitzar la mescla directament a la cel·la òptica amb un *stopped-flow* degut a la naturalesa ràpida del procés.

8. Els resultats cinètics han estat ajustats a exponencials decreixents dobles, derivant-se, per tant, l'existència de 2 etapes diferents governades per 2 velocitats diferents.
9. L'anàlisi cinètica de l'agregació dels mutants ha confirmat també la importància de la cadena β -4 en el procés d'agregació d'ADA2h.
10. L'estudi de les cinètiques d'agregació a 2 concentracions proteiques per a 2 mutants d'ADA2h, un amb major velocitat d'agregació respecte la variant salvatge (N58A) i un altre amb menor velocitat (I23V), han fet pal·lesa la influència important de la concentració sobre k_1 i pràcticament inexistent sobre k_2 , suggerint que la segona velocitat correspondria a un procés de reorganització del monòmer dins dels nuclis d'agregació.
11. Segons aquests resultats, l'agregació d'ADA2h seguiria un esquema del tipus:



12. Les prediccions de la propensió de l'agregació d'ADA2h semblen ser més acurades amb el programa TANGO, tot i que algunes mutacions que produeixen canvis importants en l'agregació no són detectades per aquest algorisme, presumiblement perquè es troben en regions de la seqüència polipeptídica que el programa no reconeix com a rellevants en el procés agregatiu.
13. La microscòpia electrònica de transmissió ha permès observar que l'adquisició d'estructura secundària- β durant les cinètiques condueix a la formació d'agregats ordenats tipus fibra amiloide semblants als presents a les amiloïdosis.
14. Tot i que ADA2h i l'acilfosfatasa, proteïnes amb un plegament pràcticament idèntic, presenten el C-terminal com a zona més implicada en l'agregació, altres proteïnes amb topologia similar a les anteriors contindrien regions diferents que dirigirien el procés d'agregació, segons la predicció de TANGO.

Capítol 2: La caracterització de diferents processos d'amiloidogènesi per 2D-FTIR mostra convergència cap a la mateixa fibra final

1. S'han posat a punt les condicions més adequades per a l'estudi, mitjançant espectroscòpia d'infraroig amb transformada de Fourier (FTIR), de la formació de fibres amiloides d'ADA2h a partir de l'estat natiu, a pD àcid i amb inducció per temperatura.
2. Han estat escollits per a aquest estudi mutants amb diferents velocitats d'agregació respecte el WT, segons els resultats obtinguts al capítol 1 d'aquesta tesi. S'ha assajat un mutant amb velocitat d'agregació superior a la variant salvatge (N58A), un mutant amb velocitat d'agregació inferior (I23V) i un mutant que no produeix agregació en l'escala de temps dels experiments cinètics realitzats al primer capítol (F65A).
3. Les principals bandes observades a l'espectre inicial han estat adjudicades a les diferents estructures segons segueix: 1619 cm^{-1} , estructura β -agregada; 1632 cm^{-1} , component de baixa freqüència de la fulla- β antiparal·lela nativa; 1647 cm^{-1} , hèlixs- α parcialment desplegadas; 1660 cm^{-1} , hèlixs 3_{10} ; 1671 cm^{-1} , girs- β ; i 1682 cm^{-1} , component d'elevada freqüència de la fulla- β antiparal·lela.
4. La banda corresponent a l'estructura β -agregada ($1616\text{-}1619\text{ cm}^{-1}$) s'allunya força de la corresponent a l'estructura β -nativa (1632 cm^{-1}), mostrant que l'estructura β de la fibra amiloide d'ADA2h és força diferent a l'estructura β -nativa.
5. Durant el procés de fibril·logènesi la quantitat de bandes no canvia però varia la seva posició i el seu percentatge d'àrea, indicant canvis en l'entorn conformacional i interconversió d'estructures.
6. La posició i el percentatge d'àrea de la banda de 1619 cm^{-1} quasi no varien durant els primers 4 minuts de la cinètica, indicant que ens trobem en un procés inicial de desagregació on es produeixen pocs canvis conformacionals. Aquesta hipòtesi es veu recolzada per l'observació de fibres curtes en forma de cuc (*worm-like*, WL) mitjançant microscòpia electrònica de transmissió.
7. L'estudi del percentatge d'àrea de la banda de 1619 cm^{-1} per a la variant salvatge d'ADA2h indica una transició en tres estats en la formació de la fibra amiloide que poden ésser resumits segons:



8. Els percentatges d'àrea i les posicions de les bandes corresponents a l'estructura nativa d'ADA2h reflecteixen el desplegament d'aquestes estructures per tal de donar lloc -i posteriorment acomodar-se- a la fibra amiloide.
9. A les concentracions de proteïna assajades en el FTIR, tots els mutants, inclòs F65A, són capaços de formar fibres amiloides. Aquestes estructures poden ser visualitzades mitjançant TEM, i la conversió d'estructura nativa a la típica β -amiloide pot ésser observada per FTIR i CD.
10. Els mutants presenten diferent temporalitat respecte l'etapa de desagregació, però les posicions i les àrees durant el procés de reorganització pràcticament no varien, indicant que les mutacions afecten a la propensió d'agregació, però no l'estructura fibril·lar.
11. S'ha observat que el percentatge d'àrea inicial de la banda de 1619 cm^{-1} de les diferents variants correlaciona inversament amb l'estabilitat d'aquestes, exceptuant N58A, el mutant amb una major propensió d'agregació segons les prediccions dels algorismes esmentats al capítol 1. No obstant, totes les variants pràcticament convergeixen als mateixos número d'ona i percentatge d'àrea al final del procés.
12. L'elaboració del mapa bidimensional sincrònic de la variant WT ha permès detectar la relació en fase però oposada entre les intensitats de les bandes de 1616 cm^{-1} i 1647 cm^{-1} , indicant que les hèlixs- α parcialment desestructurades originarien l'estructura β -agregada. També s'observen correlacions entre les diferents bandes característiques de les fulles i girs- β , indicant l'acomodament d'aquestes estructures dins la fibra amiloide.
13. Els mapes bidimensionals sincrònics dels diferents mutants d'ADA2h són molt similars al de la variant salvatge, tot i que les correlacions apareixen més amples, reflectint la seva menor estructuració en les condicions assajades.

14. La microscòpia electrònica de transmissió ha permès la visualització de fibres amiloides al final del procés estudiat quasi idèntiques per a totes les variants.

Capítol 3: Predicció d'una funció alternativa per a ADA4h i validació experimental

1. S'ha trobat certa similitud entre la seqüència aminoacídica d'ADA4h i una proteïna de llevat associada a l'*splicing*, emprant l'eina bioinformàtica PSI-BLAST i la base de dades Uniprot.
2. L'estructura d'aquesta proteïna de llevat ha estat modelada amb èxit emprant el programa MODELLER i l'estructura d'un dels membres de la família de ribonucleoproteïnes petites U1, U1A. El model ha estat avaluat energèticament amb el programa Prosall.
3. La superposició de l'estructura d'U1A, un motiu de reconeixement de RNA (RRM), amb ADA4h mitjançant el programa STAMP mostra la similitud dels patrons d'unió de RNA RNP1 i RNP2 entre elles, tot i que les seqüències d'ADA4h es troben més allunyades de les seqüències consens. Aquesta diferència seqüencial semblava indicar que ADA4h produiria unions més dèbils amb l'àcid nucleic degut a les seves diferents propietats físico-químiques.
4. S'ha definit un nou patró de RRM segons les variacions en les seqüències de les regions RNP1 i RNP2, provant-se la seva capacitat per a identificar proteïnes implicades en la unió d'RNA.
5. S'ha construït, expressat heteròlogament i purificat ADA4h satisfactòriament, tot i que en quantitats limitades, juntament amb cua d'histidines en posició N-terminal per a facilitar-ne el seu aïllament.
6. La unió inespecífica de RNA per part d'ADA4h ha estat demostrada experimentalment mitjançant mètodes electroforètics de retardament, emprant com a àcid nucleic la seqüència present a l'estructura del complex d'U1A amb codi PDB 1m5kc.
7. La unió experimental de RNA ha resultat molt més dèbil per part d'ADA4h que no pas pel control positiu, la qual cosa indicaria que ADA4h hauria evolucionat cap a la optimització de la inhibició del domini catalític carboxipeptidasa, d'aquesta manera perdent residus importants per a la unió d'aquest àcid nucleic. No és descartable, però, l'unió específica en condicions fisiològiques a un RNA concret diferent de l'assajat.

BIBLIOGRAFIA

- Adam SA, Nakagawa T, Swanson MS, Woodruff TK, & Dreyfuss G (1986) mRNA polyadenylate-binding protein: gene isolation and sequencing and identification of a ribonucleoprotein consensus sequence. *Mol Cell Biol* **6**: 2932-2943
- Allain FH, Gubser CC, Howe PW, Nagai K, Neuhaus D, & Varani G (1996) Specificity of ribonucleoprotein interaction determined by RNA folding during complex formulation. *Nature* **380**: 646-650
- Allain FH, Howe PW, Neuhaus D, & Varani G (1997) Structural basis of the RNA-binding specificity of human U1A protein. *EMBO J* **16**: 5764-5772
- Aloy P, Catasús L, Villegas V, Reverter D, Vendrell J, & Avilés FX (1998) Comparative analysis of the sequences and three-dimensional models of human procarboxypeptidases A1, A2 and B. *Biol Chem* **379**: 149-155
- Aloy P, Companys V, Vendrell J, Avilés FX, Fricker LD, Coll M, & Gomis-Rüth FX (2001) The crystal structure of the inhibitor-complexed carboxypeptidase D domain II and the modeling of regulatory carboxypeptidases. *J Biol Chem* **276**: 16177-16184
- Aloy P, Oliva B, Querol E, Avilés FX, & Russell RB (2002) Structural similarity to link sequence space: new potential superfamilies and implications for structural genomics. *Protein Sci* **11**: 1101-1116
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, & Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389-3402
- Anand K, Pallarès I, Valnickova Z, Christensen T, Vendrell J, Wendt KU, Schreuder HA, Enghild JJ, & Avilés FX (2008) The crystal structure of TAFI provides the structural basis for its intrinsic activity and the short half-life of TAFIa. *J Biol Chem* (en publicació, manuscrit M804003200)
- Anfinsen CB, Haber E, Sela M, & White FH, Jr (1961) The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc Natl Acad Sci U S A* **47**: 1309-1314
- Anson ML (1935) Crystalline carboxypeptidase. *Science* **21**: 467-468

Argarana CE, Barra HS, & Caputto R (1978) Release of [14C]tyrosine from tubulinyl-[14C]tyrosine by brain extract. Separation of a carboxypeptidase from tubulin-tyrosine ligase. *Mol Cell Biochem* **19**: 17-21

Arolas JL, Popowicz GM, Bronsoms S, Avilés FX, Huber R, Holak TA, & Ventura S (2005) Study of a major intermediate in the oxidative folding of leech carboxypeptidase inhibitor: contribution of the fourth disulfide bond. *J Mol Biol* **352**: 961-975

Arrondo JL, Muga A, Castresana J, & Goñi FM (1993) Quantitative studies of the structure of proteins in solution by Fourier-transform infrared spectroscopy. *Prog Biophys Mol Biol* **59**: 23-56

Arrondo JL & Goñi FM (1999) Structure and dynamics of membrane proteins as studied by infrared spectroscopy. *Prog Biophys Mol Biol* **72**: 367-405

Arrondo JL, Echabe I, Iloro I, Hernando MA, de la Cruz F, & Goñi FM (2003) A bacterial TrwC relaxase domain contains a thermally stable alpha-helical core. *J Bacteriol* **185**: 4226-4232

Arrondo JLR, Iloro I, Aguirre J, & Goñi FM (2004) A two-dimensional IR spectroscopic (2D-IR) simulation of protein conformational changes. *Spectroscopy* **18**: 49-58

Aslanidis C & de Jong PJ (1990) Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res* **18**: 6069-6074

Avilés FX, Vendrell J, Guasch A, Coll M, & Huber R (1993) Advances in metallo-procarboxypeptidases. Emerging details on the inhibition mechanism and on the activation process. *Eur J Biochem* **211**: 381-389

Bairoch A, Apweiler R, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, López R, Magrane M, Martin MJ, Natale DA, O'Donovan C, Redaschi N, & Yeh LS (2005) The Universal Protein Resource (UniProt). *Nucleic Acids Res* **33**: D154-159

Barbosa Pereira PJ, Segura-Martín S, Oliva B, Ferrer-Orta C, Avilés FX, Coll M, Gomis-Rüth FX, & Vendrell J (2002) Human procarboxypeptidase B: three-dimensional structure and implications for thrombin-activatable fibrinolysis inhibitor (TAFI). *J Mol Biol* **321**: 537-547

Barra HS, Arce CA, & Argarana CE (1988) Posttranslational tyrosination/detyrosination of tubulin. *Mol Neurobiol* **2**: 133-153

Basharov MA (2003) Protein folding. *J Cell Mol Med* **7**: 223-237

Bayés A, Comellas-Bigler M, Rodríguez de la Vega, M., Maskos K, Bode W, Avilés FX, Jongsma MA, Beekwilder J, & Vendrell J (2005) Structural basis of the resistance of an insect carboxypeptidase to plant protease inhibitors. *Proc Natl Acad Sci U S A* **102**: 16602-16607

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, & Bourne PE (2000) The Protein Data Bank. *Nucleic Acids Res* **28**: 235-242

Birnboim HC & Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res* **7**: 1513-1523

Booth DR, Sunde M, Bellotti V, Robinson CV, Hutchinson WL, Fraser PE, Hawkins PN, Dobson CM, Radford SE, Blake CC, & Pepys MB (1997) Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis. *Nature* **385**: 787-793

Bouchard M, Zurdo J, Nettleton EJ, Dobson CM, & Robinson CV (2000) Formation of insulin amyloid fibrils followed by FTIR simultaneously with CD and electron microscopy. *Protein Sci* **9**: 1960-1967

Bower JM, Hayward CF, Morley IS, & Morley JS (1974) Trypsinogen activation peptide and related peptides as inhibitors of gastric secretion. *Biochem Biophys Res Commun* **60**: 820-824

Bown DP, Wilkinson HS, & Gatehouse JA (1998) Midgut carboxypeptidase from *Helicoverpa armigera* (Lepidoptera: Noctuidae) larvae: enzyme characterisation, cDNA cloning and expression. *Insect Biochem Mol Biol* **28**: 739-749

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**: 248-254

Branden C & Tooze J (1991) *Introduction to Protein Structure*. Garland Publishing: New York and London

Bruix M, Pascual J, Santoro J, Prieto J, Serrano L, & Rico M (1993) 1H- and 15N-NMR assignment and solution structure of the chemotactic *Escherichia coli* Che Y protein. *Eur J Biochem* **215**: 573-585

Bryngelson JD, Onuchic JN, Socci ND, & Wolynes PG (1995) Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* **21**: 167-195

Bucciantini M, Giannoni E, Chiti F, Baroni F, Formigli L, Zurdo J, Taddei N, Ramponi G, Dobson CM, & Stefani M (2002) Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. *Nature* **416**: 507-511

Bukau B, Weissman J, & Horwich A (2006) Molecular chaperones and protein quality control. *Cell* **125**: 443-451

Bullock WO, Fernández JM, & Short JM (1987) XL1-BLUE: A high efficiency plasmid transforming *recA Escherichia coli* strain with beta-galactosidase selection. *Biotechniques* **5**: 376-378

Burd CG & Dreyfuss G (1994) Conserved structures and diversity of functions of RNA-binding proteins. *Science* **265**: 615-621

Byler DM & Susi H (1986) Examination of the secondary structure of proteins by deconvolved FTIR spectra. *Biopolymers* **25**: 469-487

Calamai M, Canale C, Relini A, Stefani M, Chiti F, & Dobson CM (2005) Reversal of protein aggregation provides evidence for multiple aggregated States. *J Mol Biol* **346**: 603-616

Calloni G, Taddei N, Plaxco KW, Ramponi G, Stefani M, & Chiti F (2003) Comparison of the folding processes of distantly related proteins. Importance of hydrophobic content in folding. *J Mol Biol* **330**: 577-591

Campbell W & Okada H (1989) An arginine specific carboxypeptidase generated in blood during coagulation or inflammation which is unrelated to carboxypeptidase N or its subunits. *Biochem Biophys Res Commun* **162**: 933-939

Campbell W, Okada N, & Okada H (2001) Carboxypeptidase R is an inactivator of complement-derived inflammatory peptides and an inhibitor of fibrinolysis. *Immunol Rev* **180**: 162-167

- Campbell WD, Lazoura E, Okada N, & Okada H (2002) Inactivation of C3a and C5a octapeptides by carboxypeptidase R and carboxypeptidase N. *Microbiol Immunol* **46**: 131-134
- Catasús L, Villegas V, Pascual R, Avilés FX, Wicker-Planquart C, & Puigserver A (1992) cDNA cloning and sequence analysis of human pancreatic procarboxypeptidase A1. *Biochem J* **287 (Pt 1)**: 299-303
- Catasús L, Vendrell J, Avilés FX, Carreira S, Puigserver A, & Billeter M (1995) The sequence and conformation of human pancreatic procarboxypeptidase A2. cDNA cloning, sequence analysis, and three-dimensional model. *J Biol Chem* **270**: 6651-6657
- Caughey B & Lansbury PT (2003) Protofibrils, pores, fibrils, and neurodegeneration: separating the responsible protein aggregates from the innocent bystanders. *Annu Rev Neurosci* **26**: 267-298
- Cerdà-Costa N, Esteras-Chopo A, Avilés FX, Serrano L, & Villegas V (2007) Early kinetics of amyloid fibril formation reveals conformational reorganisation of initial aggregates. *J Mol Biol* **366**: 1351-1363
- Chehín R, Iloro I, Marcos MJ, Villar E, Shnyrov VL, & Arrondo JL (1999) Thermal and pH-induced conformational changes of a beta-sheet protein monitored by infrared spectroscopy. *Biochemistry* **38**: 1525-1530
- Chen CC, Wang SS, Chen TW, Jap TS, Chen SJ, Jeng FS, & Lee SD (1996) Serum procarboxypeptidase B, amylase and lipase in chronic renal failure. *J Gastroenterol Hepatol* **11**: 496-499
- Chenna R, Sugawara H, Koike T, López R, Gibson TJ, Higgins DG, & Thompson JD (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* **31**: 3497-3500
- Cheon M, Chang I, Mohanty S, Luheshi LM, Dobson CM, Vendruscolo M, & Favrin G (2007) Structural reorganisation and potential toxicity of oligomeric species formed during the assembly of amyloid fibrils. *PLoS Comput Biol* **3**: 1727-1738
- Chien P, Weissman JS, & DePace AH (2004) Emerging principles of conformation-based prion inheritance. *Annu Rev Biochem* **73**: 617-656

Chiti F, Taddei N, White PM, Bucciantini M, Magherini F, Stefani M, & Dobson CM (1999) Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding. *Nat Struct Biol* **6**: 1005-1009

Chiti F, Taddei N, Bucciantini M, White P, Ramponi G, & Dobson CM (2000) Mutational analysis of the propensity for amyloid formation by a globular protein. *EMBO J* **19**: 1441-1449

Chiti F, Taddei N, Baroni F, Capanni C, Stefani M, Ramponi G, & Dobson CM (2002) Kinetic partitioning of protein folding and aggregation. *Nat Struct Biol* **9**: 137-143

Chiti F, Stefani M, Taddei N, Ramponi G, & Dobson CM (2003) Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature* **424**: 805-808

Christianson DW & Lipscomb WN (1986) X-ray crystallographic investigation of substrate binding to carboxypeptidase A at subzero temperature. *Proc Natl Acad Sci U S A* **83**: 7568-7572

Christianson DW, Mangani S, Shoham G, & Lipscomb WN (1989) Binding of D-phenylalanine and D-tyrosine to carboxypeptidase A. *J Biol Chem* **264**: 12849-12853

Ciesla J (2006) Metabolic enzymes that bind RNA: yet another level of cellular regulatory network? *Acta Biochim Pol* **53**: 11-32

Cohen E, Keller PJ, & Neurath H (1956) Purification and properties of procarboxypeptidase. *J Biol Chem* **223**: 457-467

Coll M, Guasch A, Avilés FX, & Huber R (1991) Three-dimensional structure of porcine procarboxypeptidase B: a structural basis of its inactivity. *EMBO J* **10**: 1-9

Conchillo-Solé O, de Groot NS, Avilés FX, Vendrell J, Daura X, & Ventura S (2007) AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. *BMC Bioinformatics* **8**: 65

Copley SD (2003) Enzymes with extra talents: moonlighting functions and catalytic promiscuity. *Curr Opin Chem Biol* **7**: 265-272

Cordeiro Y, Kraineva J, Suarez MC, Tempesta AG, Kelly JW, Silva JL, Winter R, & Foguel D (2006) Fourier transform infrared spectroscopy provides a fingerprint for the tetramer and for the aggregates of transthyretin. *Biophys J* **91**: 957-967

- Csorba TR (1991) Proinsulin: biosynthesis, conversion, assay methods and clinical studies. *Clin Biochem* **24**: 447-454
- de Castro E, Sigrist CJ, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, Bairoch A, & Hulo N (2006) ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res* **34**: W362-365
- de Groot NS, Avilés FX, Vendrell J, & Ventura S (2006) Mutagenesis of the central hydrophobic cluster in Abeta42 Alzheimer's peptide. Side-chain properties correlate with aggregation propensities. *FEBS J* **273**: 658-668
- DeAngelo DJ, DeFalco J, Rybacki L, & Childs G (1995) The embryonic enhancer-binding protein SSAP contains a novel DNA-binding domain which has homology to several RNA-binding proteins. *Mol Cell Biol* **15**: 1254-1264
- Dill KA & Chan HS (1997) From Levinthal to pathways to funnels. *Nat Struct Biol* **4**: 10-19
- Dinner AR, Sali A, Smith LJ, Dobson CM, & Karplus M (2000) Understanding protein folding via free-energy surfaces from theory and experiment. *Trends Biochem Sci* **25**: 331-339
- Dobson CM (1999) Protein misfolding, evolution and disease. *Trends Biochem Sci* **24**: 329-332
- Dobson CM (2001) The structural basis of protein folding and its links with human disease. *Philos Trans R Soc Lond B Biol Sci* **356**: 133-145
- Dobson CM (2003) Protein folding and misfolding. *Nature* **426**: 884-890
- Dong X, Stothard P, Forsythe IJ, & Wishart DS (2004) PlasMapper: a web server for drawing and auto-annotating plasmid maps. *Nucleic Acids Res* **32**: W660-664
- Dreyfuss G, Swanson MS, & Pinol-Roma S (1988) Heterogeneous nuclear ribonucleoprotein particles and the pathway of mRNA formation. *Trends Biochem Sci* **13**: 86-91

DuBay KF, Pawar AP, Chiti F, Zurdo J, Dobson CM, & Vendruscolo M (2004) Prediction of the absolute aggregation rates of amyloidogenic polypeptide chains. *J Mol Biol* **341**: 1317-1326

Edwards MJ, Lemos FJ, Donnelly-Doman M, & Jacobs-Lorena M (1997) Rapid induction by a blood meal of a carboxypeptidase gene in the gut of the mosquito *Anopheles gambiae*. *Insect Biochem Mol Biol* **27**: 1063-1072

Erdös EG (1979) In *Handbook of experimental pharmacology*, Erdös EG (ed) pp 427-448. Springer Verlag: Heidelberg

Espadaler J, Querol E, Avilés FX, & Oliva B (2006) Identification of function-associated loop motifs and application to protein function prediction. *Bioinformatics* **22**: 2237-2243

Estébanez-Perpiñá E, Bayés A, Vendrell J, Jongsma MA, Bown DP, Gatehouse JA, Huber R, Bode W, Avilés FX, & Reverter D (2001) Crystal structure of a novel mid-gut procarboxypeptidase from the cotton pest *Helicoverpa armigera*. *J Mol Biol* **313**: 629-638

Esteras-Chopo A, Serrano L, & López de la Paz M (2005) The amyloid stretch hypothesis: recruiting proteins toward the dark side. *Proc Natl Acad Sci U S A* **102**: 16672-16677

Fan X, Qian Y, Fricker LD, Akalal DB, & Nagle GT (1999) Cloning and expression of *Aplysia* carboxypeptidase D, a candidate prohormone-processing enzyme. *DNA Cell Biol* **18**: 121-132

Fändrich M, Fletcher MA, & Dobson CM (2001) Amyloid fibrils from muscle myoglobin. *Nature* **410**: 165-166

Ferbitz L, Maier T, Patzelt H, Bukau B, Deuerling E, & Ban N (2004) Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. *Nature* **431**: 590-596

Fernández AM, Villegas V, Martínez JC, Van Nuland NA, Conejero-Lara F, Avilés FX, Serrano L, Filimonov VV, & Mateo PL (2000) Thermodynamic analysis of helix-engineered forms of the activation domain of human procarboxypeptidase A2. *Eur J Biochem* **267**: 5891-5899

- Fernández MR, Porté S, Crosas E, Barberà N, Farrés J, Biosca JA, & Parés X (2007) Human and yeast zeta-crystallins bind AU-rich elements in RNA. *Cell Mol Life Sci* **64**: 1419-1427
- Fernández-Escamilla AM, Rousseau F, Schymkowitz J, & Serrano L (2004) Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat Biotechnol* **22**: 1302-1306
- Fersht AR (1995) Characterizing transition states in protein folding: an essential step in the puzzle. *Curr Opin Struct Biol* **5**: 79-84
- Fersht AR (2000) Transition-state structure as a unifying basis in protein-folding mechanisms: contact order, chain topology, stability, and the extended nucleus mechanism. *Proc Natl Acad Sci U S A* **97**: 1525-1529
- Fersht AR & Sato S (2004) Phi-value analysis and the nature of protein-folding transition states. *Proc Natl Acad Sci U S A* **101**: 7976-7981
- Fink AL (1998) Protein aggregation: folding aggregates, inclusion bodies and amyloid. *Fold Des* **3**: R9-23
- Finn RD, Mistry J, Schuster-Bockler B, Griffiths-Jones S, Hollich V, Lassmann T, Moxon S, Marshall M, Khanna A, Durbin R, Eddy SR, Sonnhammer EL, & Bateman A (2006) Pfam: clans, web tools and services. *Nucleic Acids Res* **34**: D247-251
- Floege J & Ketteler M (2001) beta2-microglobulin-derived amyloidosis: an update. *Kidney Int Suppl* **78**: S164-171
- Folk JE, Piez KA, Carroll WR, & Gladner JA (1960) Carboxypeptidase B. IV. Purification and characterization of the porcine enzyme. *J Biol Chem* **235**: 2272-2277
- Fontenele-Neto JD, Kalinina E, Feng Y, & Fricker LD (2005) Identification and distribution of mouse carboxypeptidase A6. *Brain Res Mol Brain Res* **137**: 132-142
- Fowler DM, Koulov AV, Alory-Jost C, Marks MS, Balch WE, & Kelly JW (2006) Functional amyloid formation within mammalian tissue. *PLoS Biol* **4**: e6
- Fowler DM, Koulov AV, Balch WE, & Kelly JW (2007) Functional amyloid - from bacteria to humans. *Trends Biochem Sci* **32**: 217-224

Franco RF, Fagundes MG, Meijers JC, Reitsma PH, Lourenco D, Morelli V, Maffei FH, Ferrari IC, Piccinato CE, Silva WA, Jr, & Zago MA (2001) Identification of polymorphisms in the 5'-untranslated region of the TAFI gene: relationship with plasma TAFI levels and risk of venous thrombosis. *Haematologica* **86**: 510-517

Fricker LD, Adelman JP, Douglass J, Thompson RC, von Strandmann RP, & Hutton J (1989) Isolation and sequence analysis of cDNA for rat carboxypeptidase E [EC 3.4.17.10], a neuropeptide processing enzyme. *Mol Endocrinol* **3**: 666-673

Fricker LD, Berman YL, Leiter EH, & Devi LA (1996) Carboxypeptidase E activity is deficient in mice with the fat mutation. Effect on peptide processing. *J Biol Chem* **271**: 30619-30624

Fujinaga M, Cherney MM, Oyama H, Oda K, & James MN (2004) The molecular structure and catalytic mechanism of a novel carboxyl peptidase from *Scytalidium lignicolum*. *Proc Natl Acad Sci U S A* **101**: 3364-3369

Gallego-Nicasio J, López-Rodríguez G, Martínez R, Tarancón MJ, Fraile MV, & Carmona P (2003) Structural changes of low density lipoproteins with Cu²⁺ and glucose induced oxidation. *Biopolymers* **72**: 514-520

García P, Serrano L, Durand D, Rico M, & Bruix M (2001) NMR and SAXS characterization of the denatured state of the chemotactic protein CheY: implications for protein folding initiation. *Protein Sci* **10**: 1100-1112

García-Castellanos R, Bonet-Figueredo R, Pallarès I, Ventura S, Avilés FX, Vendrell J, & Gomis-Rüth FX (2005) Detailed molecular comparison between the inhibition mode of A/B-type carboxypeptidases in the zymogen state and by the endogenous inhibitor latexin. *Cell Mol Life Sci* **62**: 1996-2014

García-Sáez I, Reverter D, Vendrell J, Avilés FX, & Coll M (1997) The three-dimensional structure of human procarboxypeptidase A2. Deciphering the basis of the inhibition, activation and intrinsic activity of the zymogen. *EMBO J* **16**: 6906-6913

Gardell SJ, Craik CS, Clauser E, Goldsmith EJ, Stewart CB, Graf M, & Rutter WJ (1988) A novel rat carboxypeptidase, CPA2: characterization, molecular cloning, and evolutionary implications on substrate specificity in the carboxypeptidase gene family. *J Biol Chem* **263**: 17828-17836

Gargallo R, Hunenberger PH, Avilés FX, & Oliva B (2003) Molecular dynamics simulation of highly charged proteins: comparison of the particle-particle particle-mesh and reaction field methods for the calculation of electrostatic interactions. *Protein Sci* **12**: 2161-2172

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A (2005) Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook*, Walker JM (ed) pp 571-607. Humana Press: Totowa, NJ

Gill SC & von Hippel PH (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem* **182**: 319-326

Goldberg MS & Lansbury PT, Jr (2000) Is there a cause-and-effect relationship between alpha-synuclein fibrillization and Parkinson's disease? *Nat Cell Biol* **2**: E115-119

Gomis-Rüth FX, Gomez-Ortiz M, Vendrell J, Ventura S, Bode W, Huber R, & Avilés FX (1997) Crystal structure of an oligomer of proteolytic zymogens: detailed conformational analysis of the bovine ternary complex and implications for their activation. *J Mol Biol* **269**: 861-880

Gomis-Rüth FX, Companys V, Qian Y, Fricker LD, Vendrell J, Avilés FX, & Coll M (1999) Crystal structure of avian carboxypeptidase D domain II: a prototype for the regulatory metallocarboxypeptidase subfamily. *EMBO J* **18**: 5817-5826

González C, Neira JL, Ventura S, Bronsoms S, Rico M, & Avilés FX (2003) Structure and dynamics of the potato carboxypeptidase inhibitor by 1H and 15N NMR. *Proteins* **50**: 410-422

Gosal WS, Morten IJ, Hewitt EW, Smith DA, Thomson NH, & Radford SE (2005) Competing pathways determine fibril morphology in the self-assembly of beta2-microglobulin into amyloid. *J Mol Biol* **351**: 850-864

Gothel SF & Marahiel MA (1999) Peptidyl-prolyl cis-trans isomerases, a superfamily of ubiquitous folding catalysts. *Cell Mol Life Sci* **55**: 423-436

Govaerts C, Wille H, Prusiner SB, & Cohen FE (2004) Evidence for assembly of prions with left-handed beta-helices into trimers. *Proc Natl Acad Sci U S A* **101**: 8342-8347

Greene D, Das B, & Fricker LD (1992) Regulation of carboxypeptidase E. Effect of pH, temperature and Co²⁺ on kinetic parameters of substrate hydrolysis. *Biochem J* **285** (Pt 2): 613-618

Gruber CW, Cemazar M, Heras B, Martin JL, & Craik DJ (2006) Protein disulfide isomerase: the structure of oxidative folding. *Trends Biochem Sci* **31**: 455-464

Guasch A, Coll M, Avilés FX, & Huber R (1992) Three-dimensional structure of porcine pancreatic procarboxypeptidase A. A comparison of the A and B zymogens and their determinants for inhibition and activation. *J Mol Biol* **224**: 141-157

Guijarro JI, Sunde M, Jones JA, Campbell ID, & Dobson CM (1998) Amyloid fibril formation by an SH3 domain. *Proc Natl Acad Sci U S A* **95**: 4224-4228

Harper JD & Lansbury PT, Jr (1997) Models of amyloid seeding in Alzheimer's disease and scrapie: mechanistic truths and physiological consequences of the time-dependent solubility of amyloid proteins. *Annu Rev Biochem* **66**: 385-407

Harrington PB, Urbas A, & Tandler PJ (2000) Two-dimensional correlation analysis. *Chemom Intell Lab Syst* **50**: 149-174

He GP, Muise A, Li AW, & Ro HS (1995) A eukaryotic transcriptional repressor with carboxypeptidase activity. *Nature* **378**: 92-96

Hendriks DF (1998) Carboxypeptidase U. In *Handbook of Proteolytic Enzymes*, Barrett AJ, Rawlings ND, Woessner JF (eds) pp 1328-1330. Academic Press: San Diego

Holloway PW & Mantsch HH (1989) Structure of cytochrome b5 in solution by Fourier-transform infrared spectroscopy. *Biochemistry* **28**: 931-935

Horwich A (2002) Protein aggregation in disease: a role for folding intermediates forming specific multimeric interactions. *J Clin Invest* **110**: 1221-1232

Houry WA (2001) Chaperone-assisted protein folding in the cell cytoplasm. *Curr Protein Pept Sci* **2**: 227-244

Hu X, Kaplan D, & Cebe P (2006) Determining Beta-Sheet Crystallinity in Fibrous Proteins by Thermal Analysis and Infrared Spectroscopy. *Macromolecules* **39**: 6161-6170

- Huang H, Reed CP, Zhang JS, Shridhar V, Wang L, & Smith DI (1999) Carboxypeptidase A3 (CPA3): a novel gene highly induced by histone deacetylase inhibitors during differentiation of prostate epithelial cancer cells. *Cancer Res* **59**: 2981-2988
- Hulo N, Bairoch A, Bulliard V, Cerutti L, De Castro E, Langendijk-Genevaux PS, Pagni M, & Sigrist CJ (2006) The PROSITE database. *Nucleic Acids Res* **34**: D227-230
- Hurle MR, Helms LR, Li L, Chan W, & Wetzel R (1994) A role for destabilizing amino acid replacements in light-chain amyloidosis. *Proc Natl Acad Sci U S A* **91**: 5446-5450
- Iloro I, Chehín R, Goñi FM, Pajares MA, & Arrondo JL (2004) Methionine adenosyltransferase alpha-helix structure unfolds at lower temperatures than beta-sheet: a 2D-IR study. *Biophys J* **86**: 3951-3958
- Iloro I, Goñi FM, & Arrondo JL (2005) A 2D-IR study of heat- and [(13)C]urea-induced denaturation of sarcoplasmic reticulum Ca(2+)-ATPase. *Acta Biochim Pol* **52**: 477-483
- Jackson SE & Fersht AR (1991) Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry* **30**: 10428-10435
- Jahn TR & Radford SE (2005) The Yin and Yang of protein folding. *FEBS J* **272**: 5962-5970
- Jahn TR, Tennent GA, & Radford SE (2008) A common beta-sheet architecture underlies *in vitro* and *in vivo* beta2-microglobulin amyloid fibrils. *J Biol Chem* **283**: 17279-17286
- James MNG (2004) Catalytic pathway of aspartic peptidases. In *Handbook of proteolytic enzymes*, Barrett AJ, Rawlings ND, Woessner JF (eds) pp 12-19. Elsevier: London
- Jeffery CJ (1999) Moonlighting proteins. *Trends Biochem Sci* **24**: 8-11
- Jessen TH, Oubridge C, Teo CH, Pritchard C, & Nagai K (1991) Identification of molecular contacts between the U1 A small nuclear ribonucleoprotein and U1 RNA. *EMBO J* **10**: 3447-3456

Jiménez JL, Nettleton EJ, Bouchard M, Robinson CV, Dobson CM, & Saibil HR (2002) The protofilament structure of insulin amyloid fibrils. *Proc Natl Acad Sci U S A* **99**: 9196-9201

Jiménez MA, Villegas V, Santoro J, Serrano L, Vendrell J, Avilés FX, & Rico M (2003) NMR solution structure of the activation domain of human procarboxypeptidase A2. *Protein Sci* **12**: 296-305

Juvvadi S, Fan X, Nagle GT, & Fricker LD (1997) Characterization of *Aplysia* carboxypeptidase E. *FEBS Lett* **408**: 195-200

Kaiser CM, Chang HC, Agashe VR, Lakshmipathy SK, Etchells SA, Hayer-Hartl M, Hartl FU, & Barral JM (2006) Real-time observation of trigger factor function on translating ribosomes. *Nature* **444**: 455-460

Kajava AV, Baxa U, Wickner RB, & Steven AC (2004) A model for Ure2p prion filaments and other amyloids: the parallel superpleated beta-structure. *Proc Natl Acad Sci U S A* **101**: 7885-7890

Kalinina E, Biswas R, Berezniuk I, Hermoso A, Avilés FX, & Fricker LD (2007) A novel subfamily of mouse cytosolic carboxypeptidases. *FASEB J* **21**: 836-850

Kallberg Y, Gustafsson M, Persson B, Thyberg J, & Johansson J (2001) Prediction of amyloid fibril-forming proteins. *J Biol Chem* **276**: 12945-12950

Kielkopf CL, Lucke S, & Green MR (2004) U2AF homology motifs: protein recognition in the RRM world. *Genes Dev* **18**: 1513-1526

Kim S, Merrill BM, Rajpurohit R, Kumar A, Stone KL, Papov VV, Schneiders JM, Szer W, Wilson SH, Paik WK, & Williams KR (1997) Identification of N(G)-methylarginine residues in human heterogeneous RNP protein A1: Phe/Gly-Gly-Gly-Arg-Gly-Gly-Gly/Phe is a preferred recognition motif. *Biochemistry* **36**: 5185-5192

Konno T, Murata K, & Nagayama K (1999) Amyloid-like aggregates of a plant protein: a case of a sweet-tasting protein, monellin. *FEBS Lett* **454**: 122-126

Kubelka J & Keiderling TA (2001) The anomalous infrared amide I intensity distribution in (¹³C isotopically labeled peptide beta-sheets comes from extended, multiple-stranded structures: an *ab initio* study. *J Am Chem Soc* **123**: 6142-6150

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680-685

Layne MD, Endege WO, Jain MK, Yet SF, Hsieh CM, Chin MT, Perrella MA, Blonar MA, Haber E, & Lee ME (1998) Aortic carboxypeptidase-like protein, a novel protein with discoidin and carboxypeptidase-like domains, is up-regulated during vascular smooth muscle cell differentiation. *J Biol Chem* **273**: 15654-15660

Lees JG, Miles AJ, Wien F, & Wallace BA (2006) A reference database for circular dichroism spectroscopy covering fold and secondary structure space. *Bioinformatics* **22**: 1955-1962

Lei Y, Xin X, Morgan D, Pintar JE, & Fricker LD (1999) Identification of mouse CPX-1, a novel member of the metallo-carboxypeptidase gene family with highest similarity to CPX-2. *DNA Cell Biol* **18**: 175-185

Levin Y, Skidgel RA, & Erdős EG (1982) Isolation and characterization of the subunits of human plasma carboxypeptidase N (kininase i). *Proc Natl Acad Sci U S A* **79**: 4618-4622

LeVine H, 3rd (1993) Thioflavine T interaction with synthetic Alzheimer's disease beta-amyloid peptides: detection of amyloid aggregation in solution. *Protein Sci* **2**: 404-410

Levinthal C (1968) Are there pathways in protein folding? *J Chim Phys* **65**: 44-45

Linding R, Schymkowitz J, Rousseau F, Diella F, & Serrano L (2004) A comparative study of the relationship between protein structure and beta-aggregation in globular and intrinsically disordered proteins. *J Mol Biol* **342**: 345-353

Lindorff-Larsen K, Rogen P, Paci E, Vendruscolo M, & Dobson CM (2005) Protein folding and the organization of the protein topology universe. *Trends Biochem Sci* **30**: 13-19

Lobley A, Whitmore L, & Wallace BA (2002) DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra. *Bioinformatics* **18**: 211-212

López De La Paz M, Goldie K, Zurdo J, Lacroix E, Dobson CM, Hoenger A, & Serrano L (2002) *De novo* designed peptide-based amyloid fibrils. *Proc Natl Acad Sci U S A* **99**: 16052-16057

López de la Paz M, & Serrano L (2004) Sequence determinants of amyloid fibril formation. *Proc Natl Acad Sci U S A* **101**: 87-92

López de la Paz M, de Mori GM, Serrano L, & Colombo G (2005) Sequence dependence of amyloid fibril formation: insights from molecular dynamics simulations. *J Mol Biol* **349**: 583-596

Lyons PJ, Callaway MB, & Fricker LD (2008) Characterization of carboxypeptidase A6, an extracellular matrix peptidase. *J Biol Chem* **283**: 7054-7063

Manas ES, Getahun Z, Wright WW, DeGrado WF, & Vanderkooi JM (2000) Infrared Spectra of Amide Groups in Alpha-Helical Proteins: Evidence for Hydrogen Bonding between Helices and Water. *J Am Chem Soc* **122**: 9883-9890

Mandel M & Higa A (1970) Calcium-dependent bacteriophage DNA infection. *J Mol Biol* **53**: 159-162

Maris C, Domínguez C, & Allain FH (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J* **272**: 2118-2131

Martin J (2000) Group II chaperonins as mediators of cytosolic protein folding. *Curr Protein Pept Sci* **1**: 309-324

Marx PF, Brondijk TH, Plug T, Romijn RA, Hemrika W, Meijers JC, & Huizinga EG (2008) Crystal structures of TAFI elucidate the inactivation mechanism of activated TAFI; A novel mechanism for enzyme auto-regulation. *Blood* **112**: 2803-2809

Masuda K, Yoshioka M, Hinode D, & Nakamura R (2002) Purification and characterization of arginine carboxypeptidase produced by *Porphyromonas gingivalis*. *Infect Immun* **70**: 1807-1815

McCutchen SL, Lai Z, Miroy GJ, Kelly JW, & Colon W (1995) Comparison of lethal and nonlethal transthyretin variants and their relationship to amyloid disease. *Biochemistry* **34**: 13527-13536

- McGwire GB & Skidgel RA (1995) Extracellular conversion of epidermal growth factor (EGF) to des-Arg53-EGF by carboxypeptidase M. *J Biol Chem* **270**: 17154-17158
- McMurry JE & Fay RC (2004) *Chemistry*. Prentice Hall: New Jersey
- Mead DA, Szczesna-Skorupa E, & Kemper B (1986) Single-stranded DNA 'blue' T7 promoter plasmids: a versatile tandem promoter system for cloning and protein engineering. *Protein Eng* **1**: 67-74
- Militello V, Casarino C, Emanuele A, Giostra A, Pullara F, & Leone M (2004) Aggregation kinetics of bovine serum albumin studied by FTIR spectroscopy and light scattering. *Biophys Chem* **107**: 175-187
- Mittermaier A, Varani L, Muhandiram DR, Kay LE, & Varani G (1999) Changes in side-chain and backbone dynamics identify determinants of specificity in RNA recognition by human U1A protein. *J Mol Biol* **294**: 967-979
- Mosnier LO, Lisman T, van den Berg HM, Nieuwenhuis HK, Meijers JC, & Bouma BN (2001) The defective down regulation of fibrinolysis in haemophilia A can be restored by increasing the TAFI plasma concentration. *Thromb Haemost* **86**: 1035-1039
- Mullis KB & Faloona FA (1987) Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction. *Methods Enzymol* **155**: 335-350
- Muñoz V & Serrano L (1997) Development of the multiple sequence approximation within the AGADIR model of alpha-helix formation: comparison with Zimm-Bragg and Lifson-Roig formalisms. *Biopolymers* **41**: 495-509
- Murzin AG, Brenner SE, Hubbard T, & Chothia C (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol* **247**: 536-540
- Nagai K, Oubridge C, Ito N, Avis J, & Evans P (1995) The RNP domain: a sequence-specific RNA-binding domain involved in processing and transport of RNA. *Trends Biochem Sci* **20**: 235-240
- Narahashi Y (1990) The amino acid sequence of zinc-carboxypeptidase from *Streptomyces griseus*. *J Biochem (Tokyo)* **107**: 879-886

Nelson R, Sawaya MR, Balbirnie M, Madsen AO, Riek C, Grothe R, & Eisenberg D (2005) Structure of the cross-beta spine of amyloid-like fibrils. *Nature* **435**: 773-778

Nelson R & Eisenberg D (2006) Recent atomic models of amyloid fibril structure. *Curr Opin Struct Biol* **16**: 260-265

Neri D, Billeter M, Wider G, & Wuthrich K (1992) NMR determination of residual structure in a urea-denatured protein, the 434-repressor. *Science* **257**: 1559-1563

Noda I (2007) Two-dimensional correlation analysis useful for spectroscopy, chromatography, and other analytical measurements. *Anal Sci* **23**: 139-146

Northrop JH (1930) Crystalline pepsin. I. Isolation and tests of purity. *J Gen Physiol* **13**: 739-766

Olinger JM, Hill DM, Jakobsen RJ, & Brody RS (1986) Fourier transform infrared studies of ribonuclease in H₂O and ²H₂O solutions. *Biochim Biophys Acta* **869**: 89-98

Oubridge C, Ito N, Evans PR, Teo CH, & Nagai K (1994) Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* **372**: 432-438

Pallarès I, Bonet R, García-Castellanos R, Ventura S, Avilés FX, Vendrell J, & Gomis-Rüth FX (2005) Structure of human carboxypeptidase A4 with its endogenous protein inhibitor, latexin. *Proc Natl Acad Sci U S A* **102**: 3978-3983

Pallarès I, Fernández D, Comellas-Bigler M, Fernández-Recio J, Ventura S, Avilés FX, Bode W, & Vendrell J (2008) Direct interaction between a human digestive protease and the mucoadhesive poly(acrylic acid). *Acta Crystallogr D Biol Crystallogr* **64**: 784-791

Pascual R, Burgos FJ, Salva M, Soriano F, Méndez E, & Avilés FX (1989) Purification and properties of five different forms of human procarboxypeptidases. *Eur J Biochem* **179**: 609-616

Pascual R, Vendrell J, Avilés FX, Bonicel J, Wicker C, & Puigserver A (1990) Autolysis of proproteinase E in bovine procarboxypeptidase A ternary complex gives rise to subunit III. *FEBS Lett* **277**: 37-41

- Pastor MT, Esteras-Chopo A, & López de la Paz M (2005) Design of model systems for amyloid formation: lessons for prediction and inhibition. *Curr Opin Struct Biol* **15**: 57-63
- Pastor MT, Kümmerer N, Schubert V, Esteras-Chopo A, Dotti CG, López de la Paz M, & Serrano L (2008) Amyloid toxicity is independent of polypeptide sequence, length and chirality. *J Mol Biol* **375**: 695-707
- Pawar AP, Dubay KF, Zurdo J, Chiti F, Vendruscolo M, & Dobson CM (2005) Prediction of "aggregation-prone" and "aggregation-susceptible" regions in proteins associated with neurodegenerative diseases. *J Mol Biol* **350**: 379-392
- Phillips MA & Rutter WJ (1996) Role of the prodomain in folding and secretion of rat pancreatic carboxypeptidase A1. *Biochemistry* **35**: 6771-6776
- Plakoutsi G, Taddei N, Stefani M, & Chiti F (2004) Aggregation of the Acylphosphatase from *Sulfolobus solfataricus*: the folded and partially unfolded states can both be precursors for amyloid formation. *J Biol Chem* **279**: 14111-14119
- Plakoutsi G, Bemporad F, Calamai M, Taddei N, Dobson CM, & Chiti F (2005) Evidence for a mechanism of amyloid formation involving molecular reorganisation within native-like precursor aggregates. *J Mol Biol* **351**: 910-922
- Plaxco KW, Simons KT, & Baker D (1998) Contact order, transition state placement and the refolding rates of single domain proteins. *J Mol Biol* **277**: 985-994
- Prestrelski SJ, Byler DM, & Thompson MP (1991) Infrared spectroscopic discrimination between alpha- and 3(10)-helices in globular proteins. Reexamination of Amide I infrared bands of alpha-lactalbumin and their assignment to secondary structures. *Int J Pept Protein Res* **37**: 508-512
- Prusiner SB (1982) Novel proteinaceous infectious particles cause scrapie. *Science* **216**: 136-144
- Puigserver A, Chapus C, Kerfelec B (1986) Pancreatic exopeptidases. In *Molecular and cellular basis of digestion*, Desnuelle P, Sjöström H, Norén O (eds) pp 235-247. Elsevier
- Quiocho FA & Lipscomb WN (1971) Carboxypeptidase A: a protein and an enzyme. *Adv Protein Chem* **25**: 1-78

Ramírez-Alvarado M, Merkel JS, & Regan L (2000) A systematic exploration of the influence of the protein stability on amyloid fibril formation *in vitro*. *Proc Natl Acad Sci U S A* **97**: 8979-8984

Ranson NA, Dunster NJ, Burston SG, & Clarke AR (1995) Chaperonins can catalyse the reversal of early aggregation steps when a protein misfolds. *J Mol Biol* **250**: 581-586

Rath-Wolfson L (2001) An immunocytochemical approach to the demonstration of intracellular processing of mast cell carboxypeptidase. *Appl Immunohistochem Mol Morphol* **9**: 81-85

Rawlings ND & Barrett AJ (1993) Evolutionary families of peptidases. *Biochem J* **290** (Pt 1): 205-218

Rawlings ND & Barrett AJ (1995) Evolutionary families of metallopeptidases. *Methods Enzymol* **248**: 183-228

Rawlings ND, Tolle DP, & Barrett AJ (2004) Evolutionary families of peptidase inhibitors. *Biochem J* **378**: 705-716

Rawlings ND, Morton FR, & Barrett AJ (2006) MEROPS: the peptidase database. *Nucleic Acids Res* **34**: D270-272

Rees DC & Lipscomb WN (1982) Refined crystal structure of the potato inhibitor complex of carboxypeptidase A at 2.5 Å resolution. *J Mol Biol* **160**: 475-498

Rees DC, Lewis M, & Lipscomb WN (1983) Refined crystal structure of carboxypeptidase A at 1.54 Å resolution. *J Mol Biol* **168**: 367-387

Reisdorf WC, Jr & Krimm S (1996) Infrared amide I' band of the coiled coil. *Biochemistry* **35**: 1383-1386

Reverter D, Ventura S, Villegas V, Vendrell J, & Avilés FX (1998) Overexpression of human procarboxypeptidase A2 in *Pichia pastoris* and detailed characterization of its activation pathway. *J Biol Chem* **273**: 3535-3541

Reynolds DS, Gurley DS, Stevens RL, Sugarbaker DJ, Austen KF, & Serafin WE (1989) Cloning of cDNAs that encode human mast cell carboxypeptidase A, and

comparison of the protein with mouse mast cell carboxypeptidase A and rat pancreatic carboxypeptidases. *Proc Natl Acad Sci U S A* **86**: 9480-9484

Reznik SE & Fricker LD (2001) Carboxypeptidases from A to Z: implications in embryonic development and Wnt binding. *Cell Mol Life Sci* **58**: 1790-1804

Rochet JC & Lansbury PT, Jr (2000) Amyloid fibrillogenesis: themes and variations. *Curr Opin Struct Biol* **10**: 60-68

Rodríguez de la Vega M, Sevilla RG, Hermoso A, Lorenzo J, Tanco S, Díez A, Fricker LD, Bautista JM, & Avilés FX (2007) Nna1-like proteins are active metallocarboxypeptidases of a new and diverse M14 subfamily. *FASEB J* **21**: 851-865

Rousseau F, Schymkowitz J, & Serrano L (2006a) Protein aggregation and amyloidosis: confusion of the kinds? *Curr Opin Struct Biol* **16**: 118-126

Rousseau F, Serrano L, & Schymkowitz JW (2006b) How evolutionary pressure against protein aggregation shaped chaperone specificity. *J Mol Biol* **355**: 1037-1047

Rupert PB & Ferré-D'Amare AR (2001) Crystal structure of a hairpin ribozyme-inhibitor complex with implications for catalysis. *Nature* **410**: 780-786

Russell RB & Barton GJ (1992) Multiple protein sequence alignment from tertiary structure comparison: assignment of global and residue confidence levels. *Proteins* **14**: 309-323

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, & Erlich HA (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**: 487-491

Sali A & Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* **234**: 779-815

Sambashivan S, Liu Y, Sawaya MR, Gingery M, & Eisenberg D (2005) Amyloid-like fibrils of ribonuclease A with three-dimensional domain-swapped and native-like structure. *Nature* **437**: 266-269

Sambrook J, Fritsch EF, & Maniatis T (1989) *Molecular Cloning: a laboratory Manual*. Cold Spring Harbour Lab. Press: New York

Sánchez de Groot N, Pallarès I, Avilés FX, Vendrell J, & Ventura S (2005) Prediction of "hot spots" of aggregation in disease-linked polypeptides. *BMC Struct Biol* **5**: 18

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

Sawaya MR, Sambashivan S, Nelson R, Ivanova MI, Sievers SA, Apóstol MI, Thompson MJ, Balbirnie M, Wiltzius JJ, McFarlane HT, Madsen AO, Riekkel C, & Eisenberg D (2007) Atomic structures of amyloid cross-beta spines reveal varied steric zippers. *Nature* **447**: 453-457

Schwalbe H, Fiebig KM, Buck M, Jones JA, Grimshaw SB, Spencer A, Glaser SJ, Smith LJ, & Dobson CM (1997) Structural and dynamical properties of a denatured protein. Heteronuclear 3D NMR experiments and theoretical simulations of lysozyme in 8 M urea. *Biochemistry* **36**: 8977-8991

Seemuller E, Lupas A, Stock D, Lowe J, Huber R, & Baumeister W (1995) Proteasome from *Thermoplasma acidophilum*: a threonine protease. *Science* **268**: 579-582

Segundo BS, Martínez MC, Vilanova M, Cuchillo CM, & Avilés FX (1982) The severed activation segment of porcine pancreatic procarboxypeptidase A is a powerful inhibitor of the active enzyme. Isolation and characterisation of the activation peptide. *Biochim Biophys Acta* **707**: 74-80

Selkoe DJ (2003) Folding proteins in fatal ways. *Nature* **426**: 900-904

Serio TR, Cashikar AG, Kowal AS, Sawicki GJ, Moslehi JJ, Serpell L, Arnsdorf MF, & Lindquist SL (2000) Nucleated conformational conversion and the replication of conformational information by a prion determinant. *Science* **289**: 1317-1321

Serpell LC, Fraser PE, & Sunde M (1999) X-ray fiber diffraction of amyloid fibrils. *Methods Enzymol* **309**: 526-536

Serpell LC, Sunde M, Benson MD, Tennent GA, Pepys MB, & Fraser PE (2000) The protofilament substructure of amyloid fibrils. *J Mol Biol* **300**: 1033-1039

Shapiro AL, Viñuela E, & Maizel JV, Jr (1967) Molecular weight estimation of polypeptide chains by electrophoresis in SDS-polyacrylamide gels. *Biochem Biophys Res Commun* **28**: 815-820

Siepen JA & Westhead DR (2002) The fibril_one on-line database: mutations, experimental conditions, and trends associated with amyloid fibril formation. *Protein Sci* **11**: 1862-1866

Silva RA, Nguyen JY, & Decatur SM (2002) Probing the effect of side chains on the conformation and stability of helical peptides via isotope-edited infrared spectroscopy. *Biochemistry* **41**: 15296-15303

Sipl MJ (1993) Recognition of errors in three-dimensional structures of proteins. *Proteins* **17**: 355-362

Skidgel RA, Tan FL, Deddish PA, & Li XY (1991) Structure, function and membrane anchoring of carboxypeptidase M. *Biomed Biochim Acta* **50**: 815-820

Skidgel RA, McGwire GB, & Li XY (1996) Membrane anchoring and release of carboxypeptidase M: implications for extracellular hydrolysis of peptide hormones. *Immunopharmacology* **32**: 48-52

Smith DP, Jones S, Serpell LC, Sunde M, & Radford SE (2003) A systematic investigation into the effect of protein destabilisation on beta 2-microglobulin amyloid formation. *J Mol Biol* **330**: 943-954

Smith DP, Giles K, Bateman RH, Radford SE, & Ashcroft AE (2007) Monitoring copopulated conformational states during protein folding events using electrospray ionization-ion mobility spectrometry-mass spectrometry. *J Am Soc Mass Spectrom* **18**: 2180-2190

Smith GK, Banks S, Blumenkopf TA, Cory M, Humphreys J, Laethem RM, Miller J, Moxham CP, Mullin R, Ray PH, Walton LM, & Wolfe LA, 3rd (1997) Toward antibody-directed enzyme prodrug therapy with the T268G mutant of human carboxypeptidase A1 and novel in vivo stable prodrugs of methotrexate. *J Biol Chem* **272**: 15804-15816

Smulevitch SV, Osterman AL, Galperina OV, Matz MV, Zagnitko OP, Kadyrov RM, Tsaplina IA, Grishin NV, Chestukhina GG, & Stepanov VM (1991) Molecular cloning and primary structure of *Thermoactinomyces vulgaris* carboxypeptidase T. A metalloenzyme endowed with dual substrate specificity. *FEBS Lett* **291**: 75-78

Song L & Fricker LD (1995) Calcium- and pH-dependent aggregation of carboxypeptidase E. *J Biol Chem* **270**: 7963-7967

Song L & Fricker LD (1997) Cloning and expression of human carboxypeptidase Z, a novel metallo-carboxypeptidase. *J Biol Chem* **272**: 10543-10550

Sørensen P, Winther JR, Kaarsholm NC, & Poulsen FM (1993) The pro region required for folding of carboxypeptidase Y is a partially folded domain with little regular structural core. *Biochemistry* **32**: 12160-12166

Springman EB, Dikov MM, & Serafin WE (1995) Mast cell procarboxypeptidase A. Molecular modeling and biochemical characterization of its processing within secretory granules. *J Biol Chem* **270**: 1300-1307

Stefani M (2003) What the use of disease-unrelated model proteins can tell us about the molecular basis of amyloid aggregation and toxicity. *Ital J Biochem* **52**: 162-176

Stefani M & Dobson CM (2003) Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution. *J Mol Med* **81**: 678-699

Stone TE, Li JP, & Bernasconi P (1994) Purification and characterization of the *Manduca sexta* neuropeptide processing enzyme carboxypeptidase E. *Arch Insect Biochem Physiol* **27**: 193-203

Studier FW & Moffatt BA (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* **189**: 113-130

Sunde M & Blake C (1997) The structure of amyloid fibrils by electron microscopy and X-ray diffraction. *Adv Protein Chem* **50**: 123-159

Susi H (1969) Infrared Spectra of Biological Macromolecules and Related Systems. In *Structure and Stability of Biological Macromolecules*, pp 575-663. Dekker: New York

Swanson MS, Nakagawa TY, LeVan K, & Dreyfuss G (1987) Primary structure of human nuclear ribonucleoprotein particle C proteins: conservation of sequence and domain structures in heterogeneous nuclear RNA, mRNA, and pre-rRNA-binding proteins. *Mol Cell Biol* **7**: 1731-1739

Tan F, Rehli M, Krause SW, & Skidgel RA (1997) Sequence of human carboxypeptidase D reveals it to be a member of the regulatory carboxypeptidase family with three tandem active site domains. *Biochem J* **327 (Pt 1)**: 81-87

- Tang A & Curthoys NP (2001) Identification of zeta-crystallin/NADPH:quinone reductase as a renal glutaminase mRNA pH response element-binding protein. *J Biol Chem* **276**: 21375-21380
- Tartaglia GG, Pawar AP, Campioni S, Dobson CM, Chiti F, & Vendruscolo M (2008) Prediction of aggregation-prone regions in structured proteins. *J Mol Biol* **380**: 425-436
- Tartaglia GG & Vendruscolo M (2008) The Zyggregator method for predicting protein aggregation propensities. *Chem Soc Rev* **37**: 1395-1401
- Thompson MJ, Sievers SA, Karanicolas J, Ivanova MI, Baker D, & Eisenberg D (2006) The 3D profile method for identifying fibril-forming segments of proteins. *Proc Natl Acad Sci U S A* **103**: 4074-4078
- Vendrell J, Persson B, Avilés FX, & Jornvall H (1989) Procarboxypeptidase A activation segment compared to structures of other proteins. *Protein Seq Data Anal* **2**: 461-462
- Vendrell J, Avilés FX, Vilanova M, Turner CH, & Crane-Robinson C (1990a) 1H-n.m.r. studies of the isolated activation segment from pig procarboxypeptidase A. *Biochem J* **267**: 213-220
- Vendrell J, Wider G, Avilés FX, & Wuthrich K (1990b) Sequence-specific 1H NMR assignments and determination of the secondary structure for the activation domain isolated from pancreatic procarboxypeptidase B. *Biochemistry* **29**: 7515-7522
- Vendrell J, Billeter M, Wider G, Avilés FX, & Wuthrich K (1991) The NMR structure of the activation domain isolated from porcine procarboxypeptidase B. *EMBO J* **10**: 11-15
- Vendrell J, Querol E, & Avilés FX (2000) Metalloproteases and their protein inhibitors. Structure, function and biomedical properties. *Biochim Biophys Acta* **1477**: 284-298
- Vendrell J, Avilés FX, Fricker LD (2004) Metalloproteases. In *Handbook of Metalloproteins*, Messerschmidt A, Bode W, Cygler M (eds) pp 176-189. John Wiley & Sons, Ltd: Chichester
- Ventura S, Gomis-Rüth FX, Puigserver A, Avilés FX, & Vendrell J (1997) Pancreatic procarboxypeptidases: oligomeric structures and activation processes revisited. *J. Biol Chem* **378**: 161-165

Ventura S, Zurdo J, Narayanan S, Parreño M, Mangues R, Reif B, Chiti F, Giannoni E, Dobson CM, Avilés FX, & Serrano L (2004) Short amino acid stretches can mediate amyloid formation in globular proteins: the Src homology 3 (SH3) case. *Proc Natl Acad Sci U S A* **101**: 7258-7263

Viguera AR, Villegas V, Avilés FX, & Serrano L (1997) Favourable native-like helical local interactions can accelerate protein folding. *Fold Des* **2**: 23-33

Vilanova M, Vendrell J, López MT, Cuchillo CM, & Avilés FX (1985) Preparative isolation of the two forms of pig pancreatic pro-(carboxypeptidase A) and their monomeric carboxypeptidases A. *Biochem J* **229**: 605-609

Vilanova M, Avilés FX, Vendrell J, & Méndez E (1987) Conformational predictive studies on the activation segment of pancreatic procarboxypeptidases. *Biochem Biophys Res Commun* **149**: 729-734

Vilanova M, Vendrell J, Cuchillo CM, & Avilés FX (1988) Analysis of the conformation and ligand-binding properties of the activation segment of pig procarboxypeptidase A. *Biochem J* **251**: 901-905

Villanueva J, Canals F, Villegas V, Querol E, & Avilés FX (2000) Hydrogen exchange monitored by MALDI-TOF mass spectrometry for rapid characterization of the stability and conformation of proteins. *FEBS Lett* **472**: 27-33

Villanueva J, Villegas V, Querol E, Avilés FX, & Serrano L (2002) Protein secondary structure and stability determined by combining exoproteolysis and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *J Mass Spectrom* **37**: 974-984

Villanueva J, Villegas V, Querol E, Avilés FX, & Serrano L (2003) Monitoring disappearance of monomers and generation of resistance to proteolysis during the formation of the activation domain of human procarboxypeptidase A2 (ADA2h) amyloid fibrils by matrix-assisted laser-desorption ionization-time-of-flight-MS. *Biochem J* **374**: 489-495

Villegas V, Azuaga A, Catasús L, Reverter D, Mateo PL, Avilés FX, & Serrano L (1995) Evidence for a two-state transition in the folding process of the activation domain of human procarboxypeptidase A2. *Biochemistry* **34**: 15105-15110

- Villegas V, Viguera AR, Avilés FX, & Serrano L (1996) Stabilization of proteins by rational design of alpha-helix stability using helix/coil transition theory. *Fold Des* **1**: 29-34
- Villegas V, Martínez JC, Avilés FX, & Serrano L (1998) Structure of the transition state in the folding process of human procarboxypeptidase A2 activation domain. *J Mol Biol* **283**: 1027-1036
- Villegas V, Zurdo J, Filimonov VV, Avilés FX, Dobson CM, & Serrano L (2000) Protein engineering as a strategy to avoid formation of amyloid fibrils. *Protein Sci* **9**: 1700-1708
- Wacker JL, Zareie MH, Fong H, Sarikaya M, & Muchowski PJ (2004) Hsp70 and Hsp40 attenuate formation of spherical and annular polyglutamine oligomers by partitioning monomer. *Nat Struct Mol Biol* **11**: 1215-1222
- Walsh DM, Klyubin I, Fadeeva JV, Cullen WK, Anwyl R, Wolfe MS, Rowan MJ, & Selkoe DJ (2002) Naturally secreted oligomers of amyloid beta protein potently inhibit hippocampal long-term potentiation in vivo. *Nature* **416**: 535-539
- Wang W, Boffa MB, Bajzar L, Walker JB, & Nesheim ME (1998) A study of the mechanism of inhibition of fibrinolysis by activated thrombin-activable fibrinolysis inhibitor. *J Biol Chem* **273**: 27176-27181
- Webster DR (2004) Tubuliny-Tyr carboxypeptidase. In *Handbook of Proteolytic Enzymes*, Barrett AJ, Rawlings ND, Woessner JF (eds) pp 2111-2113. Academic: San Diego
- Wei S, Segura S, Vendrell J, Avilés FX, Lanoue E, Day R, Feng Y, & Fricker LD (2002) Identification and characterization of three members of the human metallo-carboxypeptidase gene family. *J Biol Chem* **277**: 14954-14964
- Weiner MP, Costa GL, Schoettlin W, Cline J, Mathur E, & Bauer JC (1994) Site-directed mutagenesis of double-stranded DNA by the polymerase chain reaction. *Gene* **151**: 119-123
- Westermarck P, Benson MD, Buxbaum JN, Cohen AS, Frangione B, Ikeda S, Masters CL, Merlini G, Saraiva MJ, Sipe JD, & Nomenclature Committee of the International Society of Amyloidosis (2005) Amyloid: toward terminology clarification. Report from

the Nomenclature Committee of the International Society of Amyloidosis. *Amyloid* **12**: 1-4

Whitmore L & Wallace BA (2004) DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res* **32**: W668-673

Xin X, Day R, Dong W, Lei Y, & Fricker LD (1998) Identification of mouse CPX-2, a novel member of the metallocarboxypeptidase gene family: cDNA cloning, mRNA distribution, and protein expression and characterization. *DNA Cell Biol* **17**: 897-909

Xue WF, Homans SW, & Radford SE (2008) Systematic analysis of nucleation-dependent polymerization reveals new insights into the mechanism of amyloid self-assembly. *Proc Natl Acad Sci U S A* **105**: 8926-8931

Yamamoto KK, Pousette A, Chow P, Wilson H, el Shami S, & French CK (1992) Isolation of a cDNA encoding a human serum marker for acute pancreatitis. Identification of pancreas-specific protein as pancreatic procarboxypeptidase B. *J Biol Chem* **267**: 2575-2581

Zandomeneghi G, Krebs MR, McCammon MG, & Fändrich M (2004) FTIR reveals structural differences between native beta-sheet proteins and amyloid fibrils. *Protein Sci* **13**: 3314-3321

RESUM

Els dominis d'activació són presents en la porció N-terminal de la forma inactiva de les procarboxipeptidases de la subfamília de proteases M14A, i les seves funcions descrites comprenen tant el manteniment d'aquest estat inactiu com l'assistència en el plegament de l'enzim complet. La limitada mida i la notable estabilitat derivada de la gran quantitat d'estructura secundària han fet que un d'aquests dominis, el de la procarboxipeptidasa A2 humana (ADA2h) hagi estat un bon model per a l'estudi del plegament proteic, en treballs anteriors, i de l'agregació en forma de fibres amiloides en la present memòria. Els estudis cinètics de dicromisme circular de la variant salvatge i una extensa bateria de mutants puntuals revelen la seqüència corresponent a la fulla β -4 com la responsable de la direcció del procés d'agregació ordenada en conformació β . Mutacions en aquesta zona tenen la capacitat tant d'accelerar dramàticament la velocitat d'agregació, com d'abolir-la totalment. Les correlacions amb algorismes predictius de l'agregació basats en les propietats físico-químiques de la seqüència polipeptídica posen de manifest la rellevància principal de l'estructura primària en el govern del procés. La cinètica d'agregació per dicromisme circular mostra una naturalesa dual, amb una segona etapa pràcticament independent de concentració proteica, indicant una etapa tardana de reorganització conformacional. Estudis cinètics complementaris per espectroscòpia d'infraroig mostren igualment una reorganització molecular i assenyalen que la conformació agregada en β present a la fibra amiloide sembla originar-se primerament a partir de les hèlixs- α parcialment desplegadas. L'estudi de mutants amb diferent propensió a l'agregació i estabilitat ha mostrat que les mutacions comporten canvis en la via d'agregació, però que finalment arriben al mateix punt de fibril·lació.

Una possible funció alternativa per a un d'aquests dominis, els quals tenen una extraordinària complexitat en comparació amb altres seqüències inhibidores en *cis*, ha estat cercada mitjançant eines bioinformàtiques. Certa similitud seqüencial i estructural sembla ser trobada entre el domini d'activació de la procarboxipeptidasa A4 humana (ADA4h) i un domini d'unió a RNA (RRM), suggerint una possible funció ancestral per al primer, bé que no tots els residus claus per a la interacció són presents a ADA4h. La hipòtesi ha estat comprovada experimentalment mitjançant un gel de retardament, observant-se una unió a RNA dèbil i inespecífica per a la sonda assajada, d'acord amb el que s'havia predit bioinformàticament. Aquest fet no exclou, emperò, una unió fisiològica i específica amb un altre RNA diferent a l'assajat.

APÈNDIXS

Apèndix A. Publicacions

- **Primer capítol:**

Cerdà-Costa N, Esteras-Chopo A, Avilés FX, Serrano L, & Villegas V (2007) Early kinetics of amyloid fibril formation reveals conformational reorganisation of initial aggregates. *J Mol Biol* **366**: 1351-1363

- **Segon capítol:**

El manuscrit corresponent al segon capítol ha estat enviat a la revista *Proteins: Structure, Function, and Bioinformatics* amb referència Prot-00584-2008.

- **Tercer capítol:**

El manuscrit corresponent al segon capítol ha estat enviat a la revista *Journal of Molecular Biology* amb referència JMB-D-08-01422.

Apèndix B. Adreces electròniques actualitzades dels programes i bases de dades disponibles en xarxa

| Programa | Adreça electrònica |
|-----------------|---|
| AGGRESKAN | http://bioinf.uab.es/aggrescan/ |
| CLUSTALW | http://www.ebi.ac.uk/Tools/clustalw2/index.html |
| Expasy | http://www.expasy.ch/ |
| MEROPS | http://merops.sanger.ac.uk/ |
| Oligoanalyzer | http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/ |
| PDB | http://www.rcsb.org/pdb/home/home.do |
| Pfam | http://www.sanger.ac.uk/Software/Pfam/ |
| PlasMapper | http://wishart.biology.ualberta.ca/PlasMapper/ |
| PROSITE | http://us.expasy.org/prosite/ |
| ProtParam | http://www.expasy.ch/tools/protparam.html |
| PSI-BLAST | www.ncbi.nlm.nih.gov/BLAST/ |
| ScanProsite | http://www.expasy.ch/tools/scanprosite/ |
| SCOP | http://scop.mrc-lmb.cam.ac.uk/scop/ |
| TANGO | http://tango.crg.es/ |
| UniProt | www.uniprot.org |
| ZYGGREGATOR | http://www.vendruscolo.ch.cam.ac.uk/zyggregator.php |

Apèndix C. Codis PDB de les estructures esmentades en aquesta memòria

| Nom | Codi PDB | Tècnica |
|------------------------------------|------------|---------|
| acilfosfatasa de cavall | 1aps | RMN |
| ADA2 humà | 1o6x | RMN |
| ADB porcí | 1pba | RMN |
| complex CPA bovina + PCI de patata | 4CPA | raigs-X |
| complex CPA1 + LCI de sangonera | 2abz | raigs-X |
| complex CPA4 humana + ECI humà | 2bo9 | raigs-X |
| complex U1A humà + RNA | 1m5k | raigs-X |
| CPA bovina | 3cpa, 5cpa | raigs-X |
| CPA1 humana | 2v77 | raigs-X |
| CPB de lepidòpter (CPBHz) | 2c1c | raigs-X |
| CPD domini II d'ànec | 1qmu | raigs-X |
| CPD domini II d'ànec + GEMSA | 1h8l | raigs-X |
| PCI | 1h20 | RMN |
| PCPA de lepidòpter (PCPAHa) | 1jqg | raigs-X |
| PCPA1 porcina | 1pca | raigs-X |
| PCPA2 humana | 1aye | raigs-X |
| PCPA4 humana | 2boa | raigs-X |
| PCPB1 humana | 1kwm | raigs-X |
| PCPB1 porcina | 1nsa | raigs-X |
| PCPB2 bovina | 3dgv | raigs-X |
| PCPB2 humana | 3d66 | raigs-X |
| PCPB2 humana + GEMSA | 3d67 | raigs-X |
| PCPB2 humana mutant quàdruple | 3d68 | raigs-X |
| pepsinogen | 3psg | raigs-X |
| pèptid GNNQQNY de Sup35p de llevat | 1yjp | raigs-X |