



UNIVERSITAT AUTÒNOMA DE BARCELONA

Departament de Bioquímica i Biologia Molecular

**Proteases i inhibidors com a models per a
estudis de relació estructura-funció i
agregació.**

Irantzu Pallarès i Goitiz

Juliol 2006



UNIVERSITAT AUTÒNOMA DE BARCELONA

Departament de Bioquímica i Biologia Molecular

Proteases i inhibidors com a models per a estudis de relació estructura-funció i agregació.

Tesi presentada per Irantzu Pallarès i Goitiz, llicenciada en Bioquímica, per obtenir el grau de Doctora en Bioquímica.

Treball realitzat en el Departament de Bioquímica i Biologia Molecular de la Facultat de Ciències de la Universitat Autònoma de Barcelona, sota la direcció del Dr. Josep Vendrell i Roca i el Dr. Salvador Ventura i Zamora.

Irantzu Pallarès Goitiz Dr. J. Vendrell i Roca Dr. S. Ventura i Zamora

Bellaterra, 30 de Maig de 2006

Als meus pares
Al Jesús

Caminante, son tus huellas
el camino, y nada más
caminante, no hay camino,
se hace camino al andar.
Al andar se hace camino,
y al volver la vista atrás
se ve la senda que nunca
se ha de pisar.
Caminante, no hay camino,
sino estelas en la mar.

*Antonio Machado,
Proverbios y Cantares – XXIX*

Agraïments

Vull expressar el meu més sincer agraïment a tots els que, amb la vostra ajuda, heu fet possible la realització d'aquest treball.

En primer lloc al Pep i al Salva, directors d'aquesta tesi, per la seva constant labor de direcció i suport que m'han demostrat al llarg de tots aquests anys, així com la correcció i discussió d'aquesta memòria. Tanmateix, el meu agraïment al Xavier per obrir-me les portes del seu laboratori i encaminar-me en el camp de la recerca, alhora que m'ha donat la possibilitat de gaudir treballant en altres laboratoris arreu del món.

De les estades a fora en guardo un magnífic record i de bona part d'elles en vaig obtenir resultats que avui formen part de la present tesi. Començar doncs, agraint a en Lloyd Fricker, amb qui vaig iniciar aquesta tesi aventurera visitant el seu laboratori a New York a on va acollir-me com a un més i em va contagiar del seu caràcter entusiasta.

A l'Institut Max-Planck de Bioquímica de München hi vaig estar durant set mesos al grup d'en Robert Huber i Wolfram Bode als que vull agrair que em donessin aquesta oportunitat. M'agradaria reconèixer-li especialment a la Mireia no només la seva paciència infinita ensenyant-me tot el que avui sé de cristal·lografia, sinó que juntament amb l'Eva i la Sofia aconseguissin que durant tot el temps que vaig estar a allí em trobes com a casa.

A Alemanya, també vaig realitzar una estada a l'empresa Sanofi-Aventis, Frankfurt, que em va permetre viure la recerca des d'un nou enfoc. Des d'aquí les gràcies a tots els companys amb els que vaig compartir aquests quatre mesos. Aquest mateix projecte em va portar al laboratori d'en Jan Enghild a Århus, Dinamarca, a on vaig aprendre a purificar proteïnes de fonts naturals i d'on guardo un record especial pel càlid acolliment que vaig tenir de la família que formen al laboratori en Jan, la Ida i la Zuzanna.

Dels meus companys de laboratori, n'hi ha que ja no hi són com la Verònica, la Sònia, el Roman, i altres que encara donen "guerra": l'Àlex, la Núria, la Natàlia i el Daniel, a tots ells, gràcies per fer del laboratori dels PROs un lloc a on em trobo súper a gust. A l'altra banda de carrer estan els companys d'enzimo, que són una pila..., moltes gràcies també a vosaltres per aquest temps que portem junts.

A tots els companys de departament, adh's, cromatínics, ribos, edu, llevats, dna's, plantes, rmn's, els hi voldria agrair les bones estones que hem passat junts. Un raconet especial pel Salva i l'Helena per lo molt que ens mimen i el seu constant suport. I no em puc oblidar del grupet dels sopars de noies: Núria, Esther, Sílvia, Susanna, Emma i Anna per la "teràpia de grup" que fem amb l'excusa de celebrar aniversaris.

A la colla de la Seu els hi vull donar les gràcies per estar sempre, sempre al meu costat .

Finalment agrair a la meva família el seu constant recolzament sense el qual no hagués pogut arribar fins aquí. I molt especialment al meu germanet i a la Amuma que, sense saber-ho, em va donar la il·lusió que necessitava al final de la tesi.

ÍNDEX GENERAL

ÍNDEX	xi
Índex de figures i taules	xx
ABREVIATURES	xxiii
Nota prèvia sobre el format de la tesi	xxxi
I.- INTRODUCCIÓ.....	33
I.1.- LES PROTEASES.....	35
I.1.1.- Classificació de les proteases.....	35
I.1.1.1.- Classificació de les metal·loproteases	38
I.1.2.- Les metal·locarboxipeptidases.....	39
I.1.2.1.- Classificació de les metal·locarboxipeptidases.....	39
I.1.2.2.- Subfamília A/B	41
I.1.2.3.- Subfamília N/E.....	45
I.1.2.4.- Estructura de les metal·locarboxipeptidases de la subfamília A/B	49
I.1.2.5.- Activació de les metal·locarboxipeptidases	51
I.1.2.6.- Mecanisme catalític de les metal·locarboxipeptidases	52
I.1.3.- Inhibidors proteics de metal·locarboxipeptidases	55
I.1.3.1.- L'inhibidor de carboxipeptidases de patata	55
I.1.3.2.- L'inhibidor de carboxipeptidases de tomàquet	56
I.1.3.3.- L'inhibidor de carboxipeptidases d' <i>Ascaris suum</i>	56
I.1.3.4.- L'inhibidor de carboxipeptidases de sangonera	56
I.1.3.5.- L'inhibidor de carboxipeptidases de paparra	57
I.1.3.6.- L'inhibidor endogen de carboxipeptidases	58
I.1.3.7.- Mecanisme d'inhibició de les metal·locarboxipeptidases pels seus inhibidors proteics.....	59
I.2.- PLEGAMENT DE PROTEÏNES I AGREGACIÓ.....	60
I.2.1.- El plegament proteic	60
I.2.2.- Els fonaments del plegament proteic, des del postulat d'Anfinsen fins la "New View" (Nova Visió).....	60
I.2.2.1.- El postulat d'Anfinsen i la paradoxa de Levinthal	60
I.2.2.2.- Models clàssics de plegament proteic.....	61
I.2.2.3.- La nova visió: els embuts de plegament	62
I.2.3.- Agregació proteica i formació d'amiloïdes.....	64
I.2.4.- Les proteïnes amiloïdogèniques	68

M.3.7.3.- Cromatografia líquida d'alta resolució (FPLC) de filtració en gel.....	124
M.4.- CRISTAL·LITZACIÓ DE PROTEÍNES	125
M.4.1.- Cristal·lització de la CPA1 humana.....	125
M.4.2.- Cristal·lització del complex latexina/hCPA4	126
PART I.....	127
CHAPTER 1.- Three-dimensional structure of human carboxipeptidase A1	129
1.1.- SUMMARY	131
1.2.- INTRODUCTION.....	132
1.3.- MATERIALS AND METHODS	133
1.3.1.- Protein preparation and crystallization	133
1.3.2.- Data collection and processing	134
1.3.3.- Structure determination and processing	134
1.4.- RESULTS AND DISCUSSION	136
1.4.1.- Overall structure of human carboxypeptidase A1	136
1.4.2.- The active site.....	138
1.4.3.- Comparison of the overall hCPA1 structure with other structures of CPs	140
CHAPTER 2.- Structure of human carboxypeptidase A4 with its endogenous protein inhibitor, latexin	145
2.1.- SUMMARY	147
2.2.- INTRODUCTION.....	148
2.3.- MATERIALS AND METHODS	151
2.3.1.- Molecular cloning of human latexin.....	151
2.3.2.- Molecular cloning of the human hPCPA4	152
2.3.3.- Recombinant overexpression and purification of human latexin	152
2.3.4.- Recombinant overexpression and purification of human PCPA4.....	152
2.3.5.- Preparation of the hCPA4/latexin complex	153
2.3.6.- Latexin inhibitory assays	153
2.3.7.- Crystallization of the hCPA4/latexin complex, structure solution and refinement	154
2.3.8.- Miscellaneous	156

2.4.- RESULTS.....	158
2.4.1.- Inhibition profile of latexin	158
2.4.2.- The human latexin structure and comparison with the mouse orthologue.....	160
2.4.3.- Structural details of the hCPA4/latexin complex	162
2.4.4.- Differences of hCPA4 between hCPA4 and the latexin complex..	167
2.4.5.- Structural similarities of latexin	169
2.4.6.- A novel family of potential A/B-MCP inhibitors.....	170
2.5.- DISCUSSION	173
PART II.....	177
CHAPTER 3.- Human latexin forms amyloid-like oligomers by association of unfolded polypeptide segments	179
3.1.- SUMMARY	181
3.2.- INTRODUCTION.....	182
3.3.- MATERIALS AND METHODS	184
3.3.1.- Materials	184
3.3.2.- Sample preparation.....	184
3.3.3.- Prediction of sensitive regions for aggregation	184
3.3.4.- Circular dichroism	185
3.3.5.- NMR spectroscopy.....	185
3.3.6.- Dyes binding assays	185
3.3.7.- Transmission electron microscopy.....	186
3.3.8.- ATR-FTIR spectroscopy.....	186
3.3.9.- Measuring molecular dimensions	186
3.4.- RESULTS.....	187
3.4.1.- Conformation and stability of latexin	187
3.4.2.- Latexin self-association.....	189
3.4.3.- Far-UV circular dichroism studies of urea-induced latexin oligomerization	191
3.4.4.- Amyloidotic dyes binding	192
3.4.5.- Oligomerization evaluated by electron microscopy.....	194
3.4.6.- Generation of an aggregation profile and identification of “hot spots”	194
3.4.7.- Microcrystals of VLHLAWVA resemble amyloid fibrils.....	196
3.5.- DISCUSSION	198

CHAPTER 4.- Amyloid fibril formation by a partially structured intermediate state of α -chymotrypsin	201
4.1.- SUMMARY	203
4.2.- INTRODUCTION.....	204
4.3.- MATERIALS AND METHODS	206
4.3.1.- Materials	206
4.3.2.- Turbidity measurements	206
4.3.3.- Circular dichroism	206
4.3.4.- Intrinsic fluorescence	206
4.3.5.- Dye binding assays.....	207
4.3.6.- ATR-FTIR spectroscopy.....	207
4.3.7.- Aggregation kinetics.....	207
4.3.8.- Transmission electron microscopy.....	207
4.4.- RESULTS.....	208
4.4.1.- Effect of TFE on α -chymotrypsin conformation.....	208
4.4.2.- Influence of backbone conformation on α -chymotrypsin aggregation.....	209
4.4.3.- TFE-induced aggregates display amyloid-like features	210
4.4.4.- Kinetics of aggregation	214
4.4.5.- Aggregation occurs from a partially structured intermediate.....	214
4.5.- DISCUSSION	217
CHAPTER 5.- Prediction of “hot spots” of aggregation in disease-linked polypeptides.....	221
5.1.- SUMMARY	223
5.2.- INTRODUCTION.....	224
5.3.- MATERIALS AND METHODS	226
5.3.1.- Experimental determination of amino acids aggregation propensities	226
5.3.2.- Generation of aggregation profiles and identification of “hot spots”.....	226
5.3.3.- Analysis of the effect of changes in the polypeptide sequence on aggregation.....	227
5.4.- RESULTS.....	228
5.4.1.- Aggregation propensities of natural amino acids	228

5.4.2.- Generation of protein aggregation profiles and prediction of the effects of protein mutation on the aggregation propensity	228
5.4.3.- Analysis of disease-related polypeptide sequences	230
5.4.4.- Intrinsically unstructured proteins.....	230
5.4.5.- Globular proteins.....	238
5.5.- DISCUSSION	247
DISCUSSIÓ GENERAL	251
CONCLUSIONS.....	265
BIBLIOGRAFIA	279
SUMMARY	315

ÍNDEX DE FIGURES I TAULES

Figures

I.1	37
I.2	40
I.3	50
I.4	53
I.5	54
I.6	63
I.7	66
I.8	68
I.9	72
I.10	73
I.11	74
I.12	75
I.13	76
I.14	77
M1	90
M2	94
M3	96
M4	114
1.1	133
1.2	136
1.3	137
1.4	138
1.5	140
1.6	142
2.1	154
2.2	158
2.3	161
2.4	163

2.5	166
2.6	168
2.7	170
2.8	172
3.1	188
3.2	189
3.3	190
3.4	191
3.5	192
3.6	193
3.7	195
3.8	196
3.9	197
4.1	209
4.2	210
4.3	211
4.4	212
4.5	213
4.6	214
4.7	216
5.1	231
5.2	234
5.3	237
5.4	240
5.5	242
5.6	244
5.7	246

Taules

1.1	45
1.2	48
1.3	65
1.4	66
1.1	135
1.2	139
1.3	141
2.1	157
2.2	159
2.3	164
5.1	229
5.2	232
5.3	235
5.4	249

ABREVIATURES

3D	Tridimensional
Å	Angstrom
A	Adenosina
AAFP	N-(4-Metoxifenilazoformil)-L-fenilalanina
ACLP	<i>Aortic carboxypeptidase-like protein</i>
AD	<i>Activation domain</i>
ADAM	<i>A disintegrin and metalloprotease</i>
ADEPT	<i>Antibody directed enzyme prodrug therapy</i>
Ae	Activitat específica
AEBP	<i>Adipocyte enhancer-binding protein</i>
Aht	Absorbància hidròlisi total
ANS	<i>Aniline-8-naphthalene sulfonate</i>
ATP	5' trifosfat d'adenosina
ATR/FT-IR	<i>Attenuated Total Reflectance Fourier Transform Infrared</i>
BAPA	N-Benzoil-L-arginina-p-nitroanilida
BMGY	<i>Buffered glycerol-complex medium</i>
BMMY	<i>Buffered methanol-complex medium</i>
C	Citidina
CD	Dicroisme circular
cDNA	Àcid desoxirribonucleic còpia
CHC	<i>Central hydrophobic cluster</i>
CJD	<i>Cruetsfeldt-Jakob Disease</i>
cm	centímetre
CP	Carboxipeptidasa
CPA	Carboxipeptidasa A
CPA1	Carboxipeptidasa A1
CPA2	Carboxipeptidasa A2
CPA4	Carboxipeptidasa A4
CPA5	Carboxipeptidasa A5
CPA6	Carboxipeptidasa A6
CPAHa	Carboxipeptidasa A d' <i>Helicoverpa armigera</i>
CPB	Carboxipeptidasa B
CPBHz	Carboxipeptidasa B d' <i>Helicoverpa zea</i>
CPD	Carboxipeptidasa D
CPE	Carboxipeptidasa E
CPN	Carboxipeptidasa N
CPO	Carboxipeptidasa O
CPT	Carboxipeptidasa T
CPX1	Carboxipeptidasa X1
CPX2	Carboxipeptidasa X2
CPY	Carboxipeptidasa Y
CPZ	Carboxipeptidasa Z
C-terminal	Carboxiterminal
CTS	Subdomini carboxiterminal
DALI	<i>Distance matrix alignment</i>
ddNTPs	Didesoxinucleòtids trifosfat
DEAE	Dietilaminoetil
DESY	<i>Deutsches electronen synchrotron</i>
DLS	<i>Dynamic light scattering</i>
DNA	Àcid desoxirribonucleic
dNTPs	Desoxinucleòtids trifosfat
DTT	Ditiotreitol
EBI	<i>European bioinformatics institute</i>
ECI	Inhibidor endogen de carboxipeptidasa
EDTA	Tetraacetat d'etilè diamina

Abreviatures

ESRF	<i>European synchrotron radiation facility</i>
<i>et al.</i>	Et alter, i col·laboradors
FAAK	N-furilacrilòil-L-alanil-L-lisina
FAP	<i>Familial amyloidotic polyneuropathy</i>
FAPP	N-furilacrilòil-L-fenilalanil-L-fenilalanina
FPLC	<i>Fast protein liquid chromatography</i>
g	grams
G	Guanosina
GFP	<i>Green fluorescent protein</i>
GST	Glutatió-S-tranferasa
h	hora
HEPES	àcid 4-(2-hidroxietil)-1-piperazinetanesulfònic
IAPP	<i>Islet amyloid polypeptide</i>
K	Kelvin
kb	Kilobases
k_{cat}	Constant catalítica
kDa	Kilodaltons
K_i	Constant cinètica d'inhibició
K_M	Constant de Michaelis-Menten
K_{off}	Constant cinètica de dissociació
K_{on}	Constant cinètica d'associació
LB	Luria Bertrani
LCI	Inhibidor de carboxipeptidasa de sangonera
m	mil·li-
M	molar
mA	Mil·liamper
MAD	<i>Multiple wavelength anomalous dispersion</i>
MALDI-TOF	<i>Matrix-assited laser desorption ionization time-of-flight</i>
MC-CPA3	Carboxipeptidasa de mastòcit
MCP	Metal·locarboxipeptidasa
MCPI	Inhibidor de carboxipeptidasa de tomàquet
MeCPA	Carboxipeptidasa de fong
MEP	Metal·loendopeptidasa
MET	Microscopi electrònic de transmissió
min	minuts
MIR	<i>Multiple isomorphous replacement</i>
mL	mil·lilitre
MMP	Metal·loproteases de matriu extracel·lular
MPD	2-metil-2,4-pentanediol
Mr	Massa molecular relativa
mRNA	Àcid ribonucleic missatger
MS	Espectrometria de masses
n	nano-
N	Normal
NAC	<i>Non Aβ component of amyloid polypeptide</i>
NCBI	<i>National center for biotechnology information</i>
nm	Nanòmetre
N-terminal	Aminoterminal
NTS	Subdomini aminoterminal
p	pico-
p.ex.	per exemple
p/p	pes/pes
PAA	Àcid poliacrílic
PAGE	Electroforesi en gel de poliacrilamida
pb	Parells de bases

PBS	<i>Phosphate-buffered saline</i>
PCI	Inhibidor de carboxipeptidasa de patata
PCP	Procarboxipeptidasa
PCPA	Procarboxipeptidasa A
PCPB	Procarboxipeptidasa B
PCPAHa	Procarboxipeptidasa A de <i>Helicoverpa armigera</i>
PCPBHz	Procarboxipeptidasa B de <i>Helicoverpa zea</i>
PCR	Reacció en cadena de la polimerasa
PD	Pro domini
PDB	<i>Protein data bank</i>
PEG	Polietilè glicol
PPE	Proteïna E
PrP	Proteïna del prió
PrP _c	Proteïna del prió normal
PrP _{Sc}	Proteïna del prió patogènica
PVDF	Difluorur de polivinilidè
QTC	Quimotripsinogen C
RE	Reticle endoplasmàtic
RMN	Ressonància magnètica nuclear
RMSD	<i>Root-mean-square deviation</i>
RNA	Àcid ribonucleic
rpm	revolucions per minut
s	segons
SCP-B	<i>Scytilidocarboxil peptidasa B</i>
SDS	Dodecilsulfat sòdic (laurilsulfat sòdic)
SIDA	Síndrome d'Immunodeficiència Adquirida
SSA	Amiloïdosi senil sistèmica
T	Timidina
TAE	Tampó Tris-acetat-EDTA
TAFI	Inhibidor de la fibrinòlisi activable per trombina
TAFIa	TAFI activa
TBE	Tampó Tris-borat-EDTA
TCI	Inhibidor de carboxipeptidasa de paparra
TE	Tampó Tris-EDTA
TEMED	N, N, N', N'-tetrametilè diamí
TFA	Àcid trifluoroacètic
TFE	2,2,2-trifluoroetanol
Th-T	Tioflavina-T
TIMP	Inhibidors de metal-loproteases de teixit
Tm	Temperatura de fusió
Tris	Tris-hidroximetilamimetà
TTR	Transtirretina
UDO	Unitat de densitat òptica
V	volum
v/v	volum/volum
V _i	Velocitat inicial
V _{max}	Velocitat màxima
w/w	<i>weight/weight</i>
WT	Forma salvatge
Zn	Zinc
β2-m	β2-microglobulina

Abreviatures

ΔAP

λ

μ

Propensió d'agregació promig

Longitud d'ona

micro-

Taula d'abreviatures dels aminoàcids proteïnogènics

Nom Complet	Codi de tres lletres	Codi d'una lletra
Alanina	Ala	A
Arginina	Arg	R
Asparragina	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
Triptofan	Trp	W
Tirosina	Tyr	Y
Valina	Val	V

Nota prèvia sobre el format d'aquesta tesi

Aquesta Tesi es presenta, en la seva part de resultats, estructurada en dues parts de dos i tres capítols respectivament. Els cinc capítols corresponen, en el seu contingut fonamental, a sis articles, quatre d'ells ja publicats i dos més properament enviats per a publicació. Malgrat que els apartats Introducció, Materials i Mètodes, Discussió i Conclusions estan escrits només en català, els cinc capítols de l'apartat de Resultats, redactats en llengua anglesa, contenen tots la introducció i discussió corresponents, a banda de la descripció d'aquelles metodologies específiques per a cada capítol. Dues notes al final dels capítols 2 i 5 especifiquen la participació de l'autora en aquells treballs fruit de la col·laboració amb altres investigadors.

INTRODUCCIÓ

I.1.- LES PROTEASES

Les proteases són enzims que catalitzen la hidròlisi d'enllaços peptídics de proteïnes i pèptids. El conjunt de les proteases engloba un gran nombre d'enzims (~2% dels productes d'un genoma determinat) amb funcions molt diverses i que afecten a diferents aspectes del metabolisme, variant des de la funció més puntual i específica, com el processament de pro-hormones i pro-neuropèptids, fins a la funció més exhaustiva i inespecífica com és la degradació de les proteïnes provinents de la ingesta. Les proteases tenen un gran interès en medicina i biotecnologia i constitueixen dianes molt atractives per al disseny de fàrmacs.

Existeixen diversos termes redundants per referir-se als enzims proteolítics: proteasa, peptidasa i proteïnasa. Tot i que la Unió Internacional de Bioquímica i Biologia Molecular recomana l'ús del terme peptidasa, el terme proteasa és el que s'utilitza normalment.

El mecanisme catalític general pel qual una proteasa hidrolitza un enllaç peptídic consisteix en la polarització de l'enllaç peptídic mitjançant l'atac sobre el carboni del grup carbonil, el qual es pot produir directament per un grup nucleòfil o bé per acció d'una base a través d'una molècula d'aigua, i paral·lelament té lloc la donació d'un protó a l'àtom de nitrògen. Cada tipus de proteasa té uns residus específics que realitzen les funcions de nucleòfils i de donadors de protons.

I.1.1.- Classificació de les proteases

Les proteases es poden classificar en base a diferents criteris: especificitat de substrat, mecanisme catalític, relacions estructurals evolutives, localització intracel·lular, funció en l'organisme, etc.

Una de les classificacions més usades és la descrita per Rawlings i Barret (1993), basada en el mecanisme catalític de les proteases. Aquests autors han creat una base de dades de proteases anomenada MEROPS (merops.sanger.ac.uk) (Rawlings & Barret, 2000), on cada grup de proteases és dividit en famílies seguint criteris d'homologia estructural. Els set grups de proteases que s'han definit són: serina proteases, treonina proteases, cisteïna proteases, aspàrtic proteases,

Introducció

metal·loproteases, proteases de mecanisme catalític desconegut i més recentment les glutàmic proteases (Fig. I.1).

El grup de les **serina proteases** engloba un conjunt d'enzims que es caracteritzen per presentar una serina reactiva al centre actiu. El mecanisme catalític, així com l'estructura tridimensional del centre actiu, és molt semblant en totes les serina proteases estudiades. Les diferències de funció que presenten són degudes a canvis en l'especificitat de substrat. El grup de les serina proteases es divideixen en 38 famílies, cadascuna de les quals presenta un membre representatiu. La família de la quimotripsina (S1) és molt nombrosa i conté endoproteases com la tripsina, l'elastasa, l'acrosina, les calicreïnes, els granzims, factors de coagulació (VII, IX, X, XI), l'activador del plasminogen, la trombina, la plasmina, la proteïna C i factors del sistema del complement. Una altra família a destacar és la de les serina carboxipeptidases (S10), amb la carboxipeptidasa Y (CPY) de llevat com a enzim representatiu. En aquest grup trobem diverses carboxipeptidases presents en fongs, gramínies i dicotiledònies.

Les **treonina proteases** s'han descrit recentment (Seemuller *et al.*, 1995) i, de moment, és un grup de proteases poc nombrosos. Només està format per quatre famílies, entre les que cal destacar la família T1, amb la subunitat catalítica del proteasoma de *Thermoplasma acidophilum* com a membre representatiu. Dins d'aquesta família trobem subunitats catalítiques tant del proteasoma procariota com de l'eucariota. Aquests enzims es caracteritzen per presentar una treonina en posició N-terminal que té la funció de nucleòfil del centre actiu.

Les **cisteïna proteases** s'anomenen també sulfhidril o tiol proteases i formen un grup de 50 famílies, la majoria de les quals corresponen a proteases de virus, tot i que també n'hi ha de plantes. El mecanisme catalític d'aquest grup s'ha caracteritzat en la família de la papaïna (C1) i s'ha observat que hi són involucrats un residu de cisteïna i un d'histidina. Altres famílies remarcables són la de les calpaïnes (C2) i la de les caspases (C14).

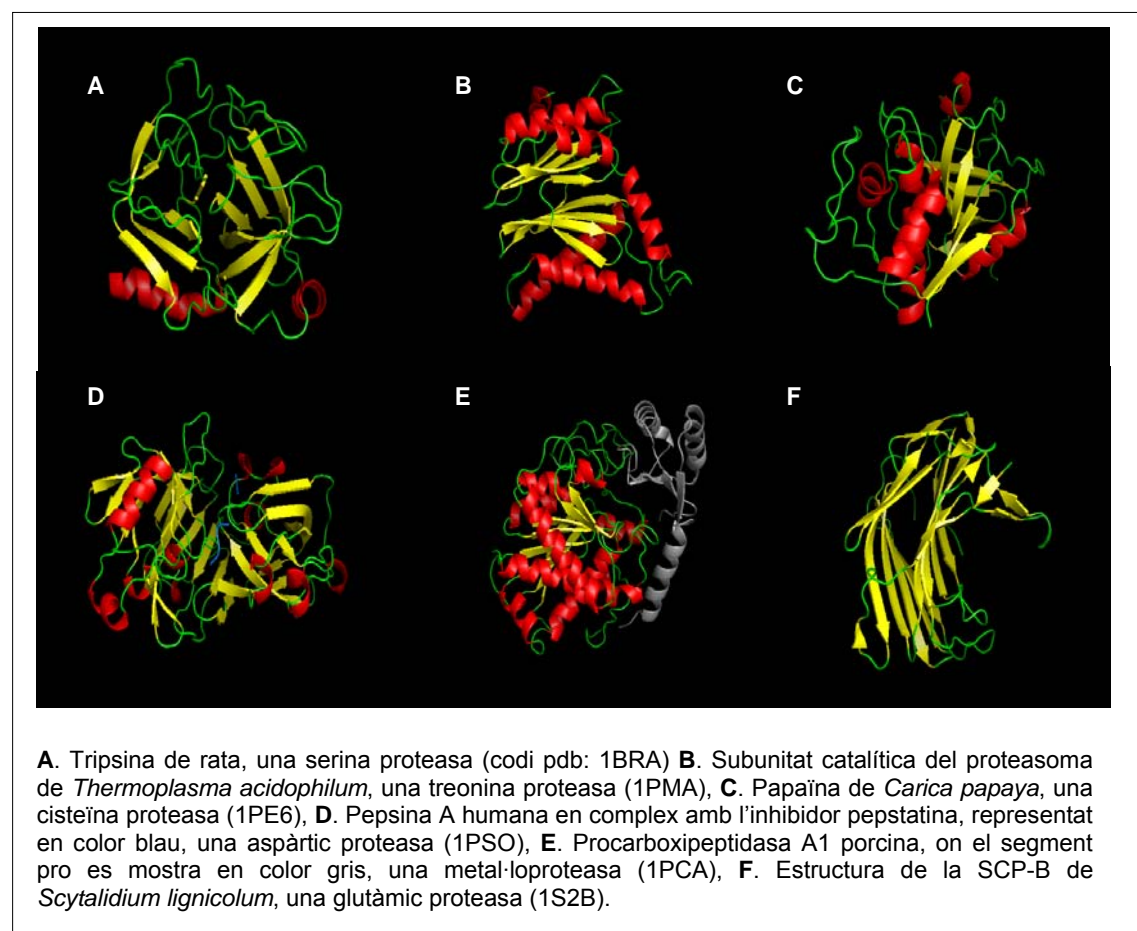
El grup de les **aspàrtic proteases** està constituït per 18 famílies. Aquests enzims també són coneguts com a proteases àcides o carboxil proteases. Són endoproteases que es caracteritzen per la presència d'un residu aspàrtic en el centre actiu, cosa que fa que el seu pH òptim sigui àcid. La família A1, amb la pepsina A humana com a membre representatiu és molt nombrosa i conté, a més de moltes

pepsines de diferents espècies, la quimosina i la renina. La família A2 conté retropepsines dels virus de la SIDA de diverses espècies i proteïnes relacionades.

Les **metal·loproteases** són un conjunt molt variat d'enzims que es caracteritzen per presentar un ió metàl·lic en el centre actiu, que normalment és zinc. A partir de l'estudi de les estructures de diferents enzims pertanyents a aquest grup, s'ha observat que l'ió metàl·lic presenta tres lligands amb la cadena polipeptídica i un quart lligand que és una molècula d'aigua. Aquesta és activada per iniciar l'atac nucleofílic sobre l'enllaç a hidrolitzar. Es coneixen 51 famílies de metal·loproteases, algunes de les quals són endopeptidases, com les famílies de la termolisina (M4) i l'astacina (M12), i famílies d'exopeptidases, com la de la carboxipeptidasa A1 (M14) i la de l'aminopeptidasa N (M1).

Les **glutàmic proteases** s'han descrit recentment (Fujinaga *et al.*, 2004) i, de moment, aquest grup només està format per la família G1 amb la scyतालidocarboxil peptidasa B (SCP-B) com a membre representatiu. La família G1 només conté proteases d'origen fúngic.

Figura I.1.- Diagrama de cintes de l'estructura d'un representant de cadascun dels sis grups



En un setè grup s'engloben aquells enzims amb mecanisme catalític desconegut. Dins d'aquest grup trobem les secretases, que són tres enzims, α , β i γ secretasa, responsables del processament de la proteïna precursora dels pèptids amiloides que es disposen extracel·lularment en el cervell, causant la malaltia de l'Alzheimer (Domínguez *et al.*, 2001).

I.1.1.1.- Classificació de les metal·loproteases

Rawlings i Barret (1995) van establir una classificació específica per a les metal·loproteases basada també en relacions estructurals evolutives. Segons aquesta classificació, les diferents famílies de metal·loproteases es divideixen en quatre grups quant a la manera de coordinar l'ió metàl·lic; dos dels quals contenen el motiu HEXXH, que, tal i com s'ha demostrat per cristal·lografia de raigs-X, forma part del lloc d'unió de l'ió metàl·lic, que generalment és zinc. Els quatre grups que s'han definit són:

1. “**HEXXH + E**”, en què els quatre lligands de l'ió zinc són una molècula d'aigua, les dues histidines del motiu HEXXH i un residu d'àcid glutàmic. Dins d'aquest grup trobem 18 famílies de les quals corresponen en gran majoria endopeptidases, com la família de la termolisina (M4). La família de l'aminopeptidasa N (M1), la de la carboxipeptidasa Taq de *Thermus aquaticus* (M32) i la de la glicil aminopeptidasa (M3) és l'única d'entre totes les famílies de metal·loproteases que conté tant endo com exopeptidases.
2. “**HEXXH + H/D**”, que presenten l'ió de zinc coordinat amb una molècula d'aigua, dos residus d'histidina del motiu HEXXH i un tercer residu d'histidina o aspàrtic. Aquest grup està constituït per 12 famílies d'endopeptidases, entre les quals hi ha la família de l'astacina (M12) i la de la col·lagenasa I (M10).
3. Aquest és un grup heterogeni que engloba diferents famílies de metal·loproteases que presenten un motiu d'unió a ió metàl·lic que és diferent a HEXXH. La majoria de les famílies que integren aquest grup presenten histidines com a lligand predominant de l'ió metàl·lic. Per altra banda, les famílies d'aminopeptidases que hi trobem es caracteritzen per

presentar dos ions metàl·lics en el centre actiu, els quals estan coordinats majoritàriament per residus àcids, com és el cas de la família leucil aminopeptidasa (M17) i de la metionil aminopeptidasa (M24). En aquest últim cas, junt amb la família M52, ens trobem amb exemples de metal·loproteases que presenten en el centre actiu ions metàl·lics diferents del zinc, com són el cobalt i el manganès per la família M24 i el níquel per a la família M52. Dins d'aquest tercer grup trobem la família M14, que conté les carboxipeptidases que són objecte d'estudi en aquest treball i presenta un motiu d'unió a zinc del tipus HXXE+H.

4. Aquest grup engloba totes les famílies (actualment un total de sis) per les quals els lligands del metall són encara desconeguts.

I.1.2.- Les metal·locarboxipeptidases

Les metal·locarboxipeptidases són enzims que catalitzen la hidròlisi d'enllaços peptídics dels residus C-terminals d'altres proteïnes o pèptids com un pas en la degradació de molècules substrat o durant la maduració d'altres i que presenten un ió de zinc directament involucrat en la catàlisi (Vendrell *et al.*, 2004). D'entre les metal·locarboxipeptidases, el grup més àmpliament estudiat ha estat el de les M14. En aquesta tesi en centrarem en la descripció d'alguns membres d'aquesta família i dels seus inhibidors proteics.

I.1.2.1.- Classificació de les metal·locarboxipeptidases M14

Les M14 es poden classificar en dues subfamílies tenint en compte criteris seqüencials i estructurals: la subfamília de la carboxipeptidasa A/B (CPA/B) i la subfamília de la carboxipeptidasa N/E (CPA N/E) (Vendrell *et al.*, 2000). El grau d'identitat seqüencial és superior al 40% dintre cada subfamília i al voltant del 15-20% entre subfamílies.

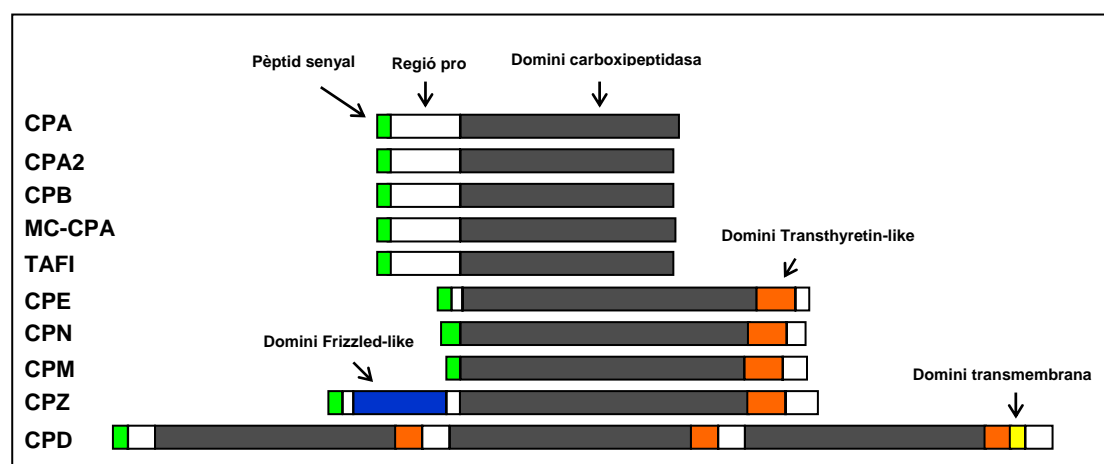
Tots els membres de la subfamília CPA/B són proteïnes d'aproximadament 34-36 kDa sintetitzades com a precursors inactius. Els membres d'aquesta subfamília que

Introducció

han estat caracteritzats tenen un pH òptim d'activitat dintre de la neutralitat, d'acord amb el pH de l'entorn a on desenvolupen la seva activitat.

A diferència dels membres de la subfamília CPA/B, els enzims N/E no es sintetitzen en forma de precursors inactius que necessitin de proteòlisi per obtenir la forma activa. En els enzims N/E és l'especificitat de substrat i la compartimentalització subcel·lular l'estratègia per a prevenir talls inapropiats que podrien causar danys cel·lulars. D'altra banda, tots els membres de la subfamília N/E presenten un domini extra al C-terminal que no es troba present en les proteïnes de l'altra subfamília. Aquest domini de 80 residus té un plegament que recorda el barril- β que podem trobar en la transtirretina i altres proteïnes. La funció d'aquest domini és desconeguda; una possibilitat seria que participés en el plegament del domini carboxipeptidasa o bé que estigués implicat en la formació d'oligòmers així com en la regulació de l'activitat enzimàtica (Reznik & Fricker, 2001) (Fig. I.2).

Figura I.2.- Comparativa dels membres de la família de les metal·locarboxipeptidases



A part d'aquesta classificació de les metal·locarboxipeptidases basada en relacions estructurals, tradicionalment se segueixen altres classificacions alternatives en base a diferents criteris. Així, les carboxipeptidases s'han classificat en formes A i B, segons l'especificitat que presenten pels residus C-terminal. Les formes A, en general, presenten preferència per residus C-terminal hidrofòbics, mentre que les formes B escindeixen residus bàsics. Es va observar que alguns organismes presentaven dues isoformes amb especificitat tipus A (Gardell *et al.*, 1988; Pascual *et al.*, 1989), que es van anomenar A1 i A2. La isoforma A1 correspon a la que anteriorment es coneixia com a forma A i la isoforma A2 és l'enzim que es va

caracteritzar de nou, el qual mostra preferència per residus C-terminal hidrofòbics del tipus aromàtic.

Una subdivisió alternativa d'acord amb la funció fisiològica que realitzen també és possible. Així, distingim entre les carboxipeptidases pancreàtiques A1, A2 i B, les quals estan involucrades en la degradació de les proteïnes durant la digestió, mentre que la resta de les carboxipeptidases de mamífer, que participen en diversos processos biològics generalment en teixits no digestius, s'han anomenat carboxipeptidases reguladores per la seva participació en reaccions més específiques, tals com el processament de pèptids. Aquest grup inclou, per una part, la TAFI (procarboxipeptidasa B de plasma), la CPA4 i la MC-CPA3 (CPA de mastòcit), que, segons la classificació en base a l'estructura, formen part de la subfamília A/B i també són anomenats enzims "pancreatic-like". Tota la resta de carboxipeptidases reguladores pertanyen a la subfamília N/E i els enzims que es troben són la CPE, CPD, CPM, CPN, CPZ, AEBP, CPX1 i CPX2.

La lleu confusió que es pot percebre en l'anterior descripció de les classificacions de les carboxipeptidases fa que sigui la classificació estructural a on tenim la subfamília A/B per una banda i la N/E per l'altra la més recomanada darrerament. Aquesta classificació estructural fa que convisquin, al mateix grup enzims amb funcions "digestives" i "reguladores" (Wei *et al.*, 2002).

En els darrers anys, s'han trobat A/B MCPs no només en mamífers sinó també en tots els regnes des dels archaea i bacteris, protozous, fongs, nematodes, insectes i altres invertebrats fins a plantes, amfibis i ocells.

I.1.2.2.- Subfamília A/B

Les secrecions pancreàtiques contenen quantitats considerables de les formes precursors de les **carboxipeptidases A1** (PCPA1), **A2** (PCPA2) i **B** (PCPB).

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). La CPB presenta afinitat per residus bàsics (Folk *et al.*, 1960). En el procés de digestió dels vertebrats, la CPA1 continua la degradació dels productes de la hidròlisi quimotríptica i la CPB dels de la hidròlisi trípica (Puigserver *et al.*, 1986). Les carboxipeptidases pancreàtiques es troben en forma soluble i no glicosilada i es

sintetitzen com a precursors inactius amb un segment d'activació d'uns 94-96 residus. Són estables en la forma de zimògens mentre estan emmagatzemats en els grànuls pancreàtics i fins el moment de la secreció. La presència de zimògens a les secrecions pancreàtiques és un fet habitual, però a diferència d'altres zimògens pancreàtics, els segments d'activació de les PCPs suposen una quarta part de la longitud total de la molècula. Aquesta característica diferencial es veu reflectida en el fet que aquests pro-segments juguen un paper més complex en la modulació del procés d'activació d'aquests proenzims. Posteriorment, mitjançant proteòlisi limitada, la tripsina genera enzims actius de 305-309 residus (Avilés *et al.*, 1993) que contribueixen a la degradació de les proteïnes alimentàries en l'intestí. Gràcies a aquesta clara compartimentalització de l'activitat, aquests proenzims poden servir de marcadors de diverses patologies. Així, per exemple, la procarboxipeptidasa B (PCPB) trobada en sèrum constitueix un marcador de pancreatitis aguda (Yamamoto *et al.*, 1992), i també de deficiència renal crònica (Chen *et al.*, 1996).

En mamífers, la PCPA1 pot trobar-se en forma monomèrica, formant complexos binaris amb la proproteïnasa E (PPE) o amb el quimotripsinogen C (QTC), o bé formant un complex ternari amb ambdues serina proteases. Les formes B, en canvi, apareixen tan sols com a monòmers. El complex ternari (PCPA1/PPE/QTC) s'ha descrit únicament en el sistema boví (Pascual *et al.*, 1990; Ventura *et al.*, 1997) i s'ha resolt la seva estructura tridimensional per difracció de raigs-X (Gomis-Rüth *et al.*, 1997a) (codi pdb: 1PYT). Existeixen diverses hipòtesis sobre quin és el paper biològic de les estructures quaternàries de les PCPA1: control i potenciació de l'activació, modulació de la activitat, protecció dels enzims front el pH àcid del suc pancreàtic i coordinació de l'aparició de les diferents activitats proteolítiques.

El sistema de les PCPs pancreàtiques humanes ha estat caracteritzat detalladament pel nostre grup (Pascual *et al.*, 1989). A partir d'extractes aquosos de polvoritzat pancreàtic humà s'han aïllat tres PCPAs (A1, A2 i A3) i dues PCPBs (B1 i B2). La PCPA1 i la PCPA2 s'obtenen en forma monomèrica, mentre que l'espècie PCPA3 és un complex binari format per la PCPA1 i la proproteïnasa E. La PCPB1 i la PCPB2 corresponen a dues formes al·lomòrfiques del mateix enzim. Més tard es van aïllar els cDNAs de les diferents procarboxipeptidases humanes i es va determinar les seves seqüències nucleotídiques i les aminoacídiques derivades (Catasús *et al.*, 1992; Catasús *et al.*, 1995; Aloy *et al.*, 1998).

La **carboxipeptidasa A4** s'ha identificat recentment en humà en cèl·lules canceroses de pròstata. El gen que es va caracteritzar expressa un mRNA de 2795

pb i codifica per un marc obert de lectura de 421 aminoàcids. La hPCPA4 presenta un 37-63% d'identitat seqüencial amb les zinc carboxipeptidases d'altres mamífers. I també comparteix un 27-43% d'identitat seqüencial amb zinc carboxipeptidases de diverses espècies de no mamífers. La similitud estructural entre la CPA4 i els seus homòlegs més propers indica que la proteïna conté un pèptid senyal de 16 residus, un segment d'activació N-terminal de 95 residus, i un domini carboxipeptidasa de 310 aminoàcids.

L'observació de la inducció de la PCPA4 per butirat en diversos càncers de pròstata van portar a estudiar la via implicada en la inducció del mRNA de la PCPA4. La Trichostatina A, un potent inhibidor específic de la histona desacetilasa, també induïx l'expressió del mRNA de la PCPA4, fet que indica que la inducció d'aquest gen ve intervinguda per la via d'hiperacetilació de les histones (Huang *et al.*, 1999). Aquest és un dels enzims objecte d'estudi en aquesta tesi.

La **carboxipeptidasa A3**, o carboxipeptidasa A de mastòcits, es troba en els grànuls de secreció dels mastòcits, on forma complexos macromoleulars amb proteoglicans. És una proteïna que hidrolitza substrats peptídics en tàndem amb altres proteases (Reynolds *et al.*, 1989). La funció precisa d'aquest enzim no es coneix, però pot estar relacionada amb els processos de resposta al·lèrgica, asmàtics o inflamatoris, en què els mastòcits estan implicats. Es sintetitza en forma de proenzim, amb un segment d'activació de 94 residus i un domini enzimàtic de 308 residus. El segment d'activació presenta dos llocs d'N-glicosilació i el tall proteolític que l'allibera del proenzim té lloc sobre un residu àcid, a diferència de les formes pancreàtiques, en què el tall activant té lloc sobre un residu bàsic. S'ha observat que el processament del zimogen a forma activa té lloc en els grànuls de secreció (Rath-Wolfson, 2001). L'enzim generat presenta activitat de tipus A, tot i que presenta un percentatge d'identitat superior respecte la forma B.

Tres nous membres de la subfamília A/B han estat identificats recentment en el genoma humà (Wei *et al.*, 2002): la **CPA5** i **CPA6** hidrolitzen preferentment residus C-terminals alifàtics/aromàtics, amb una preferència per part de la CPA5 vers cadenes laterals més petites. En canvi, el centre actiu predit per a la **CPO** el residu en posició 255, responsable de l'especificitat de substrat, és una Arg suggerint que la CPO presentaria activitat en front a aminoàcids C-terminal àcids. S'ha detectat expressió de CPA5 principalment en testicles i pàncrees. En el cas de la CPA6 s'han detectat nivells elevats de mRNA en el bulb olfatori (Fontenele-Neto *et al.*, 2005).

Introducció

Recentment s'han identificat i caracteritzat carboxipeptidases de diversos insectes. En el cas del lepidòpter *Helicoverpa armigera* (Lepidoptera: *Noctuidae*), una plaga del cotó, s'ha caracteritzat una carboxipeptidasa, **CPAHa**, amb activitat tipus A (Bown *et al.*, 1998). Aquest enzim, expressat en intestí, està involucrat en la digestió de nutrients. Es sintetitza en forma proenzimàtica amb un segment pro de 91 residus i un domini actiu de 318 residus. La seva estructura (codi pdb: 1JQG) ha estat resolta recentment (Estébanez-Perpiñá *et al.*, 2001) i s'observa que, tot i tenir un nivell d'identitat baix amb els altres membres de la seva subfamília (al voltant d' un 30%), presenten estructures tridimensionals semblants. També ha estat cristal·litzada recentment una carboxipeptidasa amb activitat de tipus B del lepidòpter *Helicoverpa zea*, **CPBHz** (Bayés *et al.*, 2005) (codi pdb: 2C1C). D'altra banda, s'ha identificat una altra carboxipeptidasa en els dípters hematòfags *Simulium vittatum* (Diptera: *Simuliidae*) (Ramos *et al.*, 1993), *Anopheles gambiae* (Diptera: *Culicidae*), mosquit de la malària africana, (Edwards *et al.*, 1997) i *Aedes aegypti* (Diptera: *Culicidae*), mosquit de la febre groga, (Edwards *et al.*, 2000) que presenta un nivell d'identitat entre aquestes tres espècies superior al 60%, pel que podria correspondre a la mateixa forma enzimàtica. En contrast, el nivell d'identitat entre aquesta forma i la CPAHa és d'un 30%. L'enzim actiu té 311 residus i és sintetitzat amb un segment pro de 95 residus. Aquest enzim presenta un 40% d'identitat amb carboxipeptidases pancreàtiques, valor superior al que presenta la CPAHa.

La **MeCPA** és una carboxipeptidasa del fong entomopatogen *Metarhizium anisopliae*. Aquest enzim és secretat en els teixits infectats del seu hoste, l'insecte *Manduca sexta*, on, juntament amb altres proteases del tipus endoproteasa, s'ocupa de procurar nutrients per al fong paràsit (Joshi & St. Leger, 1999). La MeCPA és sintetitzada en forma proenzimàtica, amb un segment pro de 76 residus i un domini enzimàtic de 319 residus amb un % d'identitat amb les formes A pancreàtiques del 40%. L'enzim actiu presenta especificitat A1+A2, és a dir, és capaç d'hidrolitzar tant residus C-terminals alifàtics com aromàtics. A diferència de les carboxipeptidases bacterianes, no presenta activitat tipus B.

La **carboxipeptidasa T** és un enzim extracel·lular que s'ha trobat en bacteris com *Thermoactinomyces vulgaris* (Smulevitch *et al.*, 1991) i *Streptomyces griseus* (Narahashi, 1990). Té la característica de processar residus C-terminals bàsics i hidrofòbics amb eficiències comparables. Es sintetitza en forma proenzimàtica, amb un segment pro de 75 residus i un domini enzimàtic de 326 residus. Constitueix un homòleg remot de les carboxipeptidases de mamífers, d'aquí que presenti un nivell

d'identitat baix en relació amb les altres carboxipeptidases de la seva subfamília (al voltant d'un 30%). Tot i aquesta baixa identitat seqüencial, s'ha observat que l'estructura tridimensional de la CPT (codi pdb: 1OBR) és molt similar a la de la CPA1 (Teplyakov *et al.*, 1993) (Taula I.1.).

Finalment convé comentar que, arrel de la seqüenciació dels genomes d'alguns organismes, s'han identificat gens que presenten homologia seqüencial amb la procarboxipeptidasa A1, però la potencial proteïna que se'n derivaria encara no ha estat identificada ni caracteritzada. Com a exemple, dins d'aquest grup trobem diverses hipotètiques procarboxipeptidases de *Drosophila melanogaster* (p.ex, número accés TrEMBL Q9VL86 i O61532) i una de *Saccharomyces cerevisiae* (número accés SwissProt P38836).

Taula I.1.- Percentatges d' identitat seqüencial entre diverses carboxipeptidases de la subfamília de la CPA1 i la CPE.

	CPA1	CPA2	CPB	TAFI	CPA3	CPA4	CPAHa	CPT	CPE
CPA1	100	66	47	42	41	59	28	27	14
CPA2	66	100	46	43	42	68	32	31	14
CPB	47	46	100	48	54	47	26	28	13
TAFI	42	43	48	100	48	42	26	28	17
CPA3	41	42	54	48	100	43	22	28	6
CPA4	59	68	47	42	43	100	30	27	16
CPAHa	28	32	26	26	22	30	100	24	13
CPT	27	31	28	28	28	27	24	100	19
CPE	14	14	13	17	6	16	13	19	100

En tots els casos s'han considerat les seqüències humanes, excepte per a la CPAHa, que és un enzim d' *Helicoverpa armigera*, i per a la CPT, que és l'enzim de *Thermoactinomyces vulgaris*.

I.1.2.3.- Subfamília N/E

La **carboxipeptidasa E**, també coneguda com a CPH, actua en el processament d'hormones peptídiques i neurotransmissors (Fricker *et al.*, 1989). La CPE es troba en teixits neuroendocrins, sobretot pituitària i cervell. Presenta un propèptid de tan sols 17 residus que no afecta a l' activitat de l'enzim, un domini enzimàtic tipus carboxipeptidasa d'uns 300 residus, un domini homòleg a la transtiretina d'uns 80

Introducció

residus i una hèlix amfòtera d'ancoratge a membrana de 25 residus. La CPE presenta especificitat pels aminoàcids C-terminalis bàsics lisina, arginina i histidina i és regulada per pH (Greene *et al.*, 1992), de manera que és inactiva a pH neutre (reticle endoplasmàtic i aparell de Golgi) i té una màxima activitat a pH 5-5,5 (vesícules secretores madures). En els ratolins diabètics obesos amb la mutació Cpe^{fat}/Cpe^{fat} s'ha demostrat que presenten una deficiència en el processament de la proinsulina, degut a una mutació puntual Ser/Pro que es dona en la CPE, quedant la seva activitat completament suprimida (Fricker *et al.*, 1996). En aquests ratolins mutants es va observar la presència de petites quantitats de pèptids processats, cosa que indicava que una altra carboxipeptidasa contribuïa al processament de pèptids neuroendocrins. La recerca de nous enzims va conduir a la identificació de la carboxipeptidasa D.

La **carboxipeptidasa D** va ser descoberta simultàniament per tres grups. El fet que s'observessin petites quantitats de pèptids processats correctament en ratolins Cpe^{fat}/Cpe^{fat} implicava que havia d'existir una altra carboxipeptidasa actuant en la via secretora. Els estudis realitzats (Song & Fricker, 1995a) van mostrar que aquesta activitat carboxipeptidasa estava unida a membrana i posteriorment es va purificar aquest nou enzim, que presenta una grandària de 180 kDa. D'altra banda, la CPD també es va identificar en fetge d'ànec com una proteïna de 180 kDa que s'uneix al virus de l'hepatitis B (Kuroki *et al.*, 1994; Eng *et al.*, 1998) i com a responsable de la mutació *Silver* en *Drosophila melanogaster* (Settle *et al.*, 1995). L'any 1997 es va clonar la CPD humana (Tan *et al.*, 1997) i es va observar que conté tres dominis semblants a la CPE seguits d'una regió transmembranal i una cua citoplasmàtica de 58 residus. La CPD es troba en tots els teixits examinats, però en cervell i pituitària és on més abunda. El pH òptim d'activitat de la CPD és de 5,5 a 7 (Song & Fricker, 1995b). Aquest rang ampli és degut als pH òptims solapants del primer i segon domini carboxipeptidasa. El tercer domini és inactiu, ja que li manquen els residus catalítics essencials (Aloy *et al.*, 2001). Recentment s'ha resolt l'estructura del domini II de la CPD d'ànec (codi pdb: 1QMU) (Gomis-Rüth *et al.*, 1999). La funció més important de la CPD sembla ser el processament de proteïnes i pèptids que transiten per la via secretora. La seva funció també inclou el processament de pèptids neuroendocrins, però la CPD no pot compensar completament la manca de CPE en ratolins Cpe^{fat}/Cpe^{fat} .

La **carboxipeptidasa M** és una glicoproteïna que va ser purificada i caracteritzada de microvellositats placentàries (Skidgel, 1989). És present en una gran varietat de

teixits i cèl·lules, però abunda especialment en placenta i pulmons. Es troba unida a membrana gràcies a una àncora de glicosilfosfatidilinositol a la regió C-terminal, com demostra el seu alliberament de la membrana després de tractar amb una fosfolipasa específica per a fosfatidilinositol (Skidgel *et al.*, 1991). La carboxipeptidasa M és un enzim de 426 residus sintetitzat sense segment d'activació i amb especificitat front residus C-terminal bàsics. La seva localització en membrana indica que podria estar implicada en el control de l'activitat d'hormones peptídiques en la superfície cel·lular, en la modulació de receptors i en el processament o degradació de proteïnes extracel·lulars. Concretament, s'ha observat que la carboxipeptidasa M podria estar implicada en el metabolisme de factors de creixement degut a la seva capacitat de generar des-Arg53-EGF i a la coincidència de la distribució de l'enzim i el factor de creixement epidèrmic (McGuire & Skidgel, 1995).

La **carboxipeptidasa N** es sintetitza al fetge sense segment pro i s'allibera a la sang, on és present en elevades concentracions. Se la coneix també pels noms de lisina carboxipeptidasa (CPK), inactivador d'anafilotoxina i carboxipeptidasa B de sèrum. Presenta especificitat per residus C-terminal bàsics i en humans és una glicoproteïna que consta de quatre subunitats no associades covalentment: dues subunitats catalítiques de 438 residus cadascuna (50-55 kDa) i dues subunitats glicosilades de 83 kDa sense activitat (Erdös, 1979). Les subunitats no catalítiques estableixen les subunitats actives front la temperatura corporal i les manté en circulació evitant la seva eliminació de la sang per filtració glomerular. Aquesta proteasa està implicada en la protecció de l'organisme contra pèptids amb potent acció que van a parar al torrent sanguini, com les quinines i les anafilotoxines.

La **carboxipeptidasa Z** s'ha clonat i expressat en baculovirus (Song & Fricker, 1997). Tant la forma humana com la de rata contenen un domini N-terminal de 100 aminoàcids altament conservat que presenta un 20-30% d'identitat amb el domini "frizzled" que es troba en proteïnes que interaccionen amb Wnt, una proteïna involucrada en la polaritat de teixits en les primeres fases d'embriogènesi. La carboxipeptidasa Z abunda en la placenta i també es troba, encara que en quantitats menors, en cervell, pulmons, timus i ronyó.

L' **AEBP**, *adipocyte enhancer-binding protein*, és un repressor transcripcional amb activitat carboxipeptidasa, l'expressió del qual és suprimida durant la diferenciació dels adipòcits (He *et al.*, 1995). S'ha observat que l' AEBP és activat a través d'un nou mecanisme, en què la unió a DNA estimula l'activitat de l'enzim (Muisse & Ro,

Introducció

1999). Aquest enzim, a l'igual que la CPX1 i la CPX2 consta d'un domini similar a discoïdina i un domini carboxipeptidasa. S'ha descrit una isoforma de l'AEBP, anomenada ACLP (*aortic carboxypeptidase-like protein*), que presenta una extensió N-terminal de 380 aminoàcids. L'expressió d'aquesta isoforma augmenta durant la diferenciació de les cèl·lules del múscul llis de l'aorta (Layne *et al.*, 1998).

Les **carboxipeptidases X1 i X2** s'han identificat recentment en ratolí (Lei *et al.*, 1999; Xin *et al.*, 1998). Ambdues proteïnes contenen una regió N-terminal de 160 aminoàcids seqüencialment similar al domini de discoïdina d'altres proteïnes. Alguns residus importants per a l'activitat catalítica d'altres carboxipeptidases no es troben conservats en aquestes, característica que comparteixen amb el domini III de la CPD. Així, la carboxipeptidasa X1 no té activitat front substrats estàndard per a carboxipeptidases, mentre que per a la carboxipeptidasa X2 no s'ha detectat cap activitat. La carboxipeptidasa X1 s'ha trobat en molts teixits embrionaris, cosa que suggereix que la seva funció deu estar lligada al desenvolupament, possiblement mitjançant interaccions cel·lulars a través del seu domini de discoïdina. La carboxipeptidasa X2 té una àmplia distribució en el cervell, on probablement té una funció com a proteïna d'unió. (Taula I.2.).

Taula I.2.- Percentatges d' identitat seqüencial entre diverses carboxipeptidases de la subfamília de la CPE

	CPE	CPN	CPM	CPD	CPZ	AEBP1	CPX1	CPX2
CPE	100	52	42	50	50	40	48	46
CPN	52	100	42	49	50	41	47	48
CPM	42	42	100	50	39	29	35	35
CPD	50	49	50	100	41	40	44	42
CPZ	50	50	39	41	100	41	47	43
AEBP1	40	41	29	40	41	100	51	58
CPX1	48	47	35	44	47	51	100	56
CPX2	46	48	35	42	43	58	56	100

S'han trobat membres d'aquesta subfamília en organismes invertebrats, tals com el mol·lusc marí *Aplysia californica*, on s'ha detectat la presència d'enzims CPE-like (Juvvadi *et al.*, 1997) i CPD-like (Fan *et al.*, 1999) i l' insecte *Manduca sexta* (Stone

et al., 1994). En el bacteri *Porphyromonas gingivalis* s'ha caracteritzat una carboxipeptidasa que presenta homologia amb aquesta subfamília (Masuda *et al.*, 2002), però la funció que se li atribueix és la d'obtenir nutrients. D'altra banda, s'han detectat seqüències de DNA amb homologia respecte la CPE en diferents organismes, com *Caenorhabditis elegans* (número accés EMBL U80448 i Z81054) o *Pseudomonas aeruginosa* (número accés EMBL AE004730).

I.1.2.4.- Estructura de les metal·locarboxipeptidases de la subfamília A/B

La determinació de les estructures tridimensionals de les metal·locarboxipeptidases té gran interès de cara al disseny d'inhibidors o altres molècules que permetin la variació d'aspectes de la seva funcionalitat. Aquesta família de proenzims té importància dins el camp terapèutic, ja que molts d'aquests enzims participen en processos biològics específics com pot ser el processament de pèptids bioactius (CPD, CPN, CPA5) (Song & Fricker, 1995b; Wei *et al.*, 2003) o la participació en la cascada fibrinolítica (CPB de plasma o TAFI) (Wang *et al.*, 1998), i la seva activitat descontrolada pot ser causa de disfuncions en l'organisme (per exemple, nivells anòmals de TAFI a plasma s'han associat amb diverses malalties com la hemofília, entre altres) (Franco *et al.*, 2001; Mosnier *et al.*, 2001). També tenen importància per la seva aplicació en procediments terapèutics tals com l'ADEPT (*Antibody Directed Enzyme Prodrug therapy*) (Smith *et al.*, 1997).

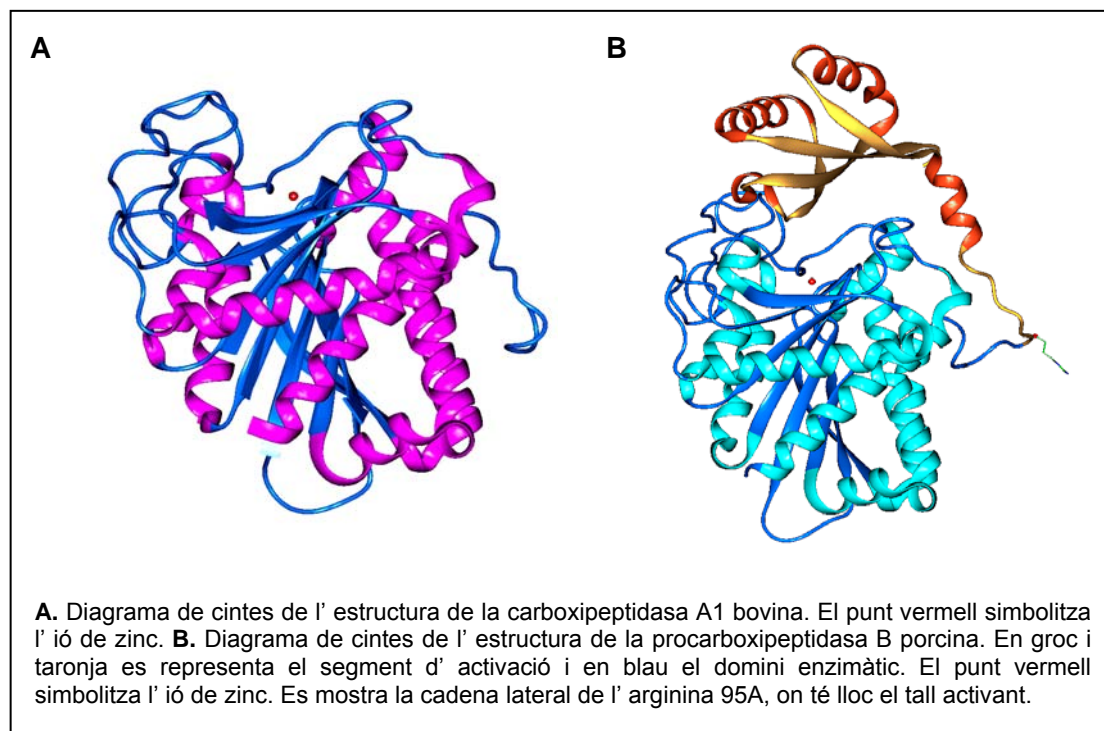
La carboxipeptidasa A bovina és un dels enzims més ben caracteritzats estructuralment d'aquesta família i n'és el model principal. S'ha determinat l'estructura cristal·logràfica de l'enzim aïllat i en complex amb inhibidors (Quioco & Lipscomb, 1971; Rees *et al.*, 1983; Christianson & Lipscomb, 1989) (codis pdb: 3CPA, 4CPA, 5CPA). Es tracta d'un enzim amb forma globular, que presenta 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 $-\alpha$ laterals (Fig. I.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.

Els primers proenzims de carboxipeptidases dels que es va resoldre l'estructura tridimensional són d'origen porcí (Coll *et al.*, 1991 i Guasch *et al.*, 1992). (Fig. I.3) i corresponen a dues isoformes, A i B. També s'ha resolt l'estructura del domini

Introducció

globular del segment d'activació, en dissolució mitjançant NMR, de la forma B porcina (Vendrell *et al.*, 1990a; 1990b; 1991).

Figura I.3.



Tot i que els segments d'activació de les diferents formes proenzimàtiques presenten un nivell d'identitat baix, la seva disposició a l'espai és molt similar. Tots 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 mirant a la carboxipeptidasa per un costat i amb dues hèlixs en el costat oposat. El segment que connecta la regió globular amb l'enzim està estructurat en una hèlix de longitud variable en funció dels diferents enzims i un llaç on, generalment, es troben els punts d'activació.

Els segments d'activació inhibeixen la carboxipeptidasa quan s'hi troben units, tot i que per substrats petits s'ha descrit un cert grau d'activitat en algunes procarboxipeptidasas del tipus A (Serra, 1992). Existeixen una sèrie de contactes específics entre el segment d'activació i l'enzim que bloquegen els subestis implicats en la unió de substrats peptídics. De tots els residus del segment d'activació, la triada Asp36-Phe37-Trp38 està molt conservada i es creu que és essencial per a la interacció.

Actualment es coneix l'estructura de varies procarboxipeptidases de la subfamília A/B: 2 pro-enzims humans (A2 i B) (García-Sáez *et al.*, 1997; Barbosa-Pereira *et al.*, 2002), així com les dues formes porcines ja esmentades (A i B), la procarboxipeptidasa A d'origen boví en complex amb dues serina proteases, les dels enzims de lepidòpter PCPAHa i la CPBHz (Estébanez-Perpiñá *et al.*, 2001; Bayés *et al.*, 2005) i, en el moment d'escriure aquesta memòria, la CPA1 i CPA4 humanes que són objecte d'aquesta tesi.

I.2.2.5.- Activació de les metal-loprocarboxipeptidases

Moltes proteases són sintetitzades com a precursors inactius anomenats proenzims o zimògens per tal de modular la seva activitat, de manera que són activades només en el lloc i en moment adequat. Els zimògens són activats per un procés de proteòlisi limitada que implica la hidròlisi específica d'un o més enllaços per part d'una proteasa.

Els segments responsables de la inactivació de les proteases solen trobar-se en posició N-terminal. Aquests segments tenen una longitud que pot variar des d'uns pocs residus, com en el cas del tripsinogen, en què s'allibera un pèptid de 6 residus, fins als 560 residus que presenta el plasminogen, grandària major que la de la proteïna activa. El fet que existeixin segments d'activació amb una longitud considerable fa pensar que puguin tenir funcions secundàries, com passa amb el pèptid d'activació del tripsinogen, que produeix una supressió hormonal de la secreció gàstrica (Bower *et al.*, 1974). D'altra banda, s'ha observat que en alguns casos els segments d'activació són essencials perquè la proteïna adopti un plegament correcte, de manera que se'ls atribueix la funció general de xaperones co-transduccionals (Sorenson *et al.*, 1993; Phillips & Rutter, 1996).

Els mecanismes d'inhibició dels segments d'activació són diversos. Poden induir al zimogen a adoptar una conformació catalíticament no competent, com en les serina proteases, o poden actuar com a elements d'impediment estèric per a l'accés al centre actiu, com passa en les procarboxipeptidases. En aquests casos el centre catalític està preformat als enzims i pot succeir que alguns proenzims presentin activitat intrínseca residual (Kassell & Kay, 1973; Guasch *et al.*, 1992).

L'activació de les procarboxipeptidases constitueix un exemple de proteòlisi N-terminal. Els dominis d'activació d'aquestes proteases són força llargs en presentar entre 94 i 95

residus, el que significa una quarta part de tota la molècula. La longitud i el grau d'estructuració dels pro-segments podria dirigir el plegament de tot el proenzim, com s'esperaria dels seu comportament com a unitat independent i eficient de plegament i de la seva influència en l'expressió *in vivo* de les PCPs (Villegas *et al.*, 1995). Per als enzims de mamífers l'enzim responsable de produir el tall que activi la carboxipeptidasa és la tripsina. S'ha descrit per algunes procarboxipeptidases del tipus A que després de ser activades els seus dominis d'activació es comporten com inhibidors força potents (K_i nM), si més no durant un cert temps després de l'activació i abans d'una més profunda degradació. Una anàlisi de l'activitat enzimàtica generada durant el procés d'activació pot determinar si el segment pro està actuant com a inhibidor. Si del seguiment de la generació de l'activitat s'obté una corba sigmoide estem en un cas de domini pro inhibidor; si, contràriament, la forma és d'una hipèrbola no existiria aquesta inhibició.

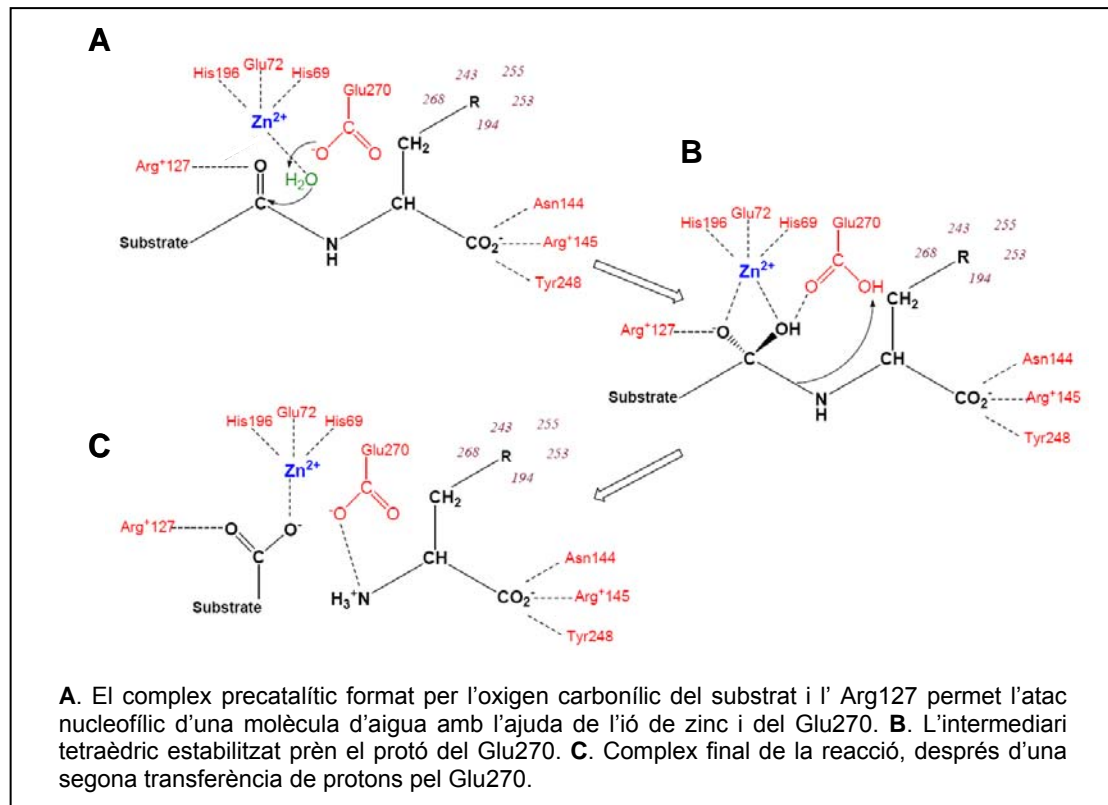
I.2.2.6.- Mecanisme catalític de les metal·locarboxipeptidases

Totes les carboxipeptidases presenten un motiu d'unió al zinc del tipus **HXXE + H**, on els quatre lligands del zinc són una histidina, un glutamat situat tres residus més enllà, una altra histidina localitzada entre 108 i 135 residus cap al C-terminal de la primera histidina i una molècula d'aigua.

Encara avui no està del tot clar el mecanisme catalític d'aquests enzims. Hi ha dues hipòtesis, la "via de l'acil" i la "via promoguda per l'aigua", de les que la segona és la més acceptada (Phillips *et al.*, 1990; Kim & Lipscomb, 1991; Álvarez-Santos *et al.*, 1994) (Fig. I.4). Segons aquest mecanisme, el zinc del centre actiu polaritza la molècula d'aigua amb la que està coordinat, i aquesta ataca l'enllaç peptídic a hidrolitzar originant un intermediari tetraèdric. El Glu270 i l'Arg127 (numeració estàndard de la CPA1 bovina) col·laboren en aquest atac mitjançant la formació d'un pont d'hidrogen amb l'aigua i l'estabilització de l'oxoni de l'intermediari tetraèdric, respectivament. La interacció del zinc amb l'aigua origina un hidroxil que actua com a potent nucleòfil i l'Arg127, junt amb l'ió metàl·lic, estabilitza les càrregues negatives. Les carboxipeptidases mostren activitat peptidàsica i esteràsica, i el mecanisme proposat funcionaria tant per a l'una com per a l'altre, amb diferents passos limitants.

Gràcies als estudis cristal·logràfics i cinètics amb substrats de diferent longitud i composició i als estudis de mutagènesi dirigida (Gardell *et al.*, 1985 i Phillips *et al.*, 1990), s'han identificat els residus del centre actiu i els responsables de la fixació del substrat.

Figura I.4.- Mecanisme catalític de la carboxipeptidasa A1 bovina



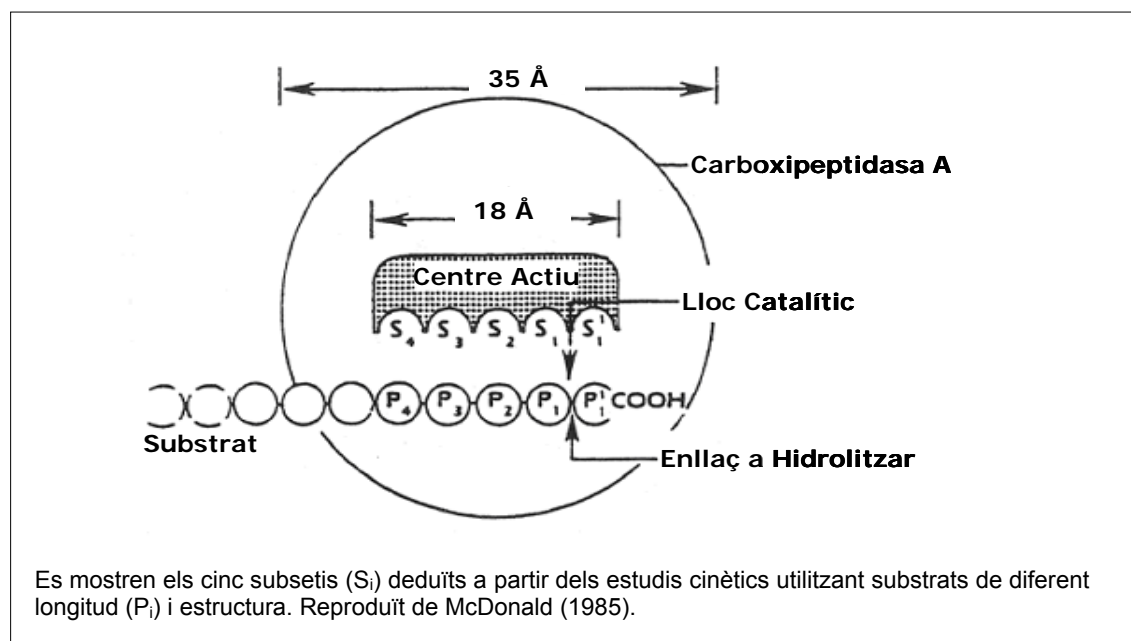
Així, s'ha observat que els residus d'unió al zinc són: His69, Glu72 i His196 i s'han definit una sèrie d'aminoàcids importants per a la catàlisi i unió a substrat, els quals són classificats en diversos subconjunts (Fig. I.5), els residus principals dels quals són els següents:

- S1': Asn144, Arg145 i Tyr248.
- S1: Arg127 i Glu270.
- S2: Arg71, Ser197, Tyr198 i Ser199.
- S3: Phe279.
- S4: Glu122, Arg124.

Introducció

El grup carboxilat terminal del substrat peptídic és fixat per l'Asn144, Arg145 i Tyr248 i el grup carbonil del enllaç peptídic a hidrolitzar està prop de Glu270, Arg127 i el zinc. Tots aquests residus estan conservats en les carboxipeptidases digestives i molts també ho estan en les reguladores. Les diferències més importants es troben en els aminoàcids que defineixen la especificitat de substrat o en els que es troben en regions superficials.

Figura I.5.- Esquema del centre actiu de la carboxipeptidasa A1



Es creu que l'especificitat de substrat està definida principalment per un únic residu, que es troba en la posició 255 de la seqüència primària. S'ha observat que els enzims amb activitat front a residus alifàtics o aromàtics presenten residus amb aquestes característiques en la posició 255, mentre que les carboxipeptidases que hidrolitzen residus C-terminals bàsics sempre presenten residus àcids en aquesta posició. Recentment, a *Helicoverpa armigera*, s'ha identificat un enzim amb un residu bàsic en la posició que determina l'especificitat de substrat i s'ha demostrat experimentalment que l'activitat proteasa que presenta és en front a residus C-terminals àcids (Bown & Gatehouse, 2004). Probablement hi hagi d'altres residus que juguin un cert paper en l'especificitat de substrat però no s'han pogut identificar.

I.1.3.- Inhibidors proteics de metal·locarboxipeptidases

Les metal·locarboxipeptidases són inhibides per una varietat d'agents quelants dels que n'és un exemple la 1,10-fenantrolina. L'àcid benzil succínic inhibeix els enzims tipus A, i compostos com els àcids i el 2-mercaptometil-3-guanidinoetilopranòic inhibeixen els enzims de tipus B, amb constants d'inhibició (K_i) dintre dels marges micromolar o nanomolar (Vendrell & Avilés, 1999). Altres inhibidors sintètics de la carboxipeptidasa A han estat emprats per poder analitzar el seu mecanisme catalític. El nombre d'inhibidors de carboxipeptidases identificats en organismes vius és força baix, sobretot si el comparem amb el nombre d'inhibidors que s'han trobat per endoproteases (Vendrell *et al.*, 2000).

S'han identificat alguns inhibidors heteròlegs de metal·locarboxipeptidases com l'inhibidor de carboxipeptidasa de patata (PCI), el seu ortòleg de tomàquet (MCPI), l'inhibidor del paràsit intestinal *Ascaris suum* i de la sangonera *Hirudo medicinalis*, i el més recent, l'inhibidor de la paparra *Rhipicephalus bursa*. Fins el moment, l'únic inhibidor endogen específic present en mamífers es la latexina, també coneguda com a *endogenous carboxypeptidase inhibitor* (ECI).

I.1.3.1.- L'inhibidor de carboxipeptidases de patata (PCI)

És una proteïna de 39 residus, i una de les proteïnes globulars més petites descrites. L'estructura tridimensional es coneix tant en dissolució (Clore *et al.*, 1987), com en complex cristal·logràfic amb la carboxipeptidasa A (Rees & Lipscomb, 1982). El nucli globular de 27 residus del PCI es troba estabilitzat per tres ponts disulfur i manca d'estructura secundària regular, exceptuant una curta hèlix interna de cinc residus. D'aquest nucli sobresurten una cua N-terminal de set residus i una C-terminal de cinc residus. Encara no s'ha pogut assignar cap funció a la cua N-terminal (residus 1-6), mentre que els cinc residus de l'extrem C-terminal juguen un paper molt important en la inhibició i es defineixen com el lloc d'interacció primària. Aquesta cua interacciona directament amb els residus del centre actiu, bloquejant-lo. A més, una segona regió del PCI es defineix com el lloc d'unió secundari. Aquesta està formada pels residus que es troben al voltant del triptòfan 28: His15, Trp22, Phe23, Asn29 i Ser30 (Rees & Lipscomb, 1982). Degut a tot això, la interacció entre PCI i les

carboxipeptidases és molt forta i la constant d'inhibició calculada està al voltant d'1nM per als enzims de mamífer.

I.1.3.2.- L'inhibidor de carboxipeptidases de tomàquet (MCPI)

Els tomàquets contenen un inhibidor de metal·locarboxipeptidases (MCPI) amb una seqüència aminoacídica altament similar a la del PCI (Hass & Hermodson, 1981). Dels 37 residus d'aquest inhibidor, 26 són idèntics, i 7 dels 11 canvis podrien haver-se donat amb el canvi d'una única base. A més a més, tots sis residus cisteïna, que donen lloc a tres ponts disulfur, es troben alineats sense la necessitat d'introduir cap espai en les seves seqüències. Els valors de K_i d'ambdós inhibidors són indistingibles per a les carboxipeptidases A i B.

I.1.3.3.- L'inhibidor de carboxipeptidasa d'*Ascaris suum*

Ascaris suum són paràsits intestinals que produeixen inhibidors contra les pepsines de l'estómac de l'hoste, i inhibidors de la quimotripsina, elastasa i tripsina del intestí de l'hoste. L'inhibidor de metal·locarboxipeptidases trobat en aquest paràsit és un polipèptid de 65 o 66 residus (Homandberg *et al.*, 1989). La seva seqüència no presenta homologia ni amb els inhibidors de quimotripsina/elastasa ni amb els inhibidors de tripsina d'*Ascaris*. S'ha observat certa homologia amb l'inhibidor de tripsina de carbassa, i amb els factors VIII i X. A més a més, presenta similitud amb els inhibidors de *Solanacea* en una curta regió interna de 9 residus i amb els 5 residus del C-terminal.

I.1.3.4.- L'inhibidor de carboxipeptidasa de sangonera (LCI)

Així com fins el moment s'han aïllat diversos inhibidors de serina proteases a partir de la sangonera *Hirudo medicinalis*, l'inhibidor de carboxipeptidases anomenat LCI és el primer inhibidor de metal·locarboxipeptidases trobat en aquesta espècie. El LCI no presenta similitud seqüencial amb cap altra proteïna, excepte en la seva cua C-terminal, la qual és molt similar a la dels inhibidors de *Solanacea* (Reverter *et al.*, 1998a). El LCI és un polipèptid de 66 residus ric en cisteïnes que es comporta com un inhibidor de metal·locarboxipeptidases competitiu i d'unió forta amb una K_i de

l'ordre nanomolar. El cDNA del LCI codifica per una seqüència de 81 aminoàcids que conté un pèptid senyal hidrofòbic de 15 residus que precedeix la Ser N-terminal de la proteïna madura. L'espectroscòpia de CD i NMR 1D indiquen que el LCI recombinant és una proteïna globular compacta, molt estable en un ampli rang de pH i de condicions desnaturalitzants. La seva estructura tridimensional es va determinar en dissolució per NMR i acomplexat amb la carboxipeptidasa humana A2 per cristal·lografia de raigs-X (Reverter *et al.*, 2000). El LCI es plega en un domini compacte format per cinc cadenes- β antiparal·leles i una α -hèlix curta, en una topologia β_3 - β_1 - β_2 - β_5 - β_4 , i l' α -hèlix curta s'empaqueta en la regió més compacta de l'estructura β , interaccionant amb el principi i el final de les cadenes β_1 i β_2 respectivament. El 45% dels residus del LCI es troben formant part d'elements d'estructura secundària regulars, un percentatge elevat quan es compara amb altres proteïnes de grandària similar. Aquest inhibidor es troba estabilitzat per la presència de 4 ponts disulfur formats entre les cisteïnes 11-34, 18-62, 19-43 i 22-58, tots ells localitzats dintre d'elements d'estructura secundària.

I.1.3.5.- L'inhibidor de carboxipeptidasa de paparra (TCI)

Recentment s'ha aïllat un nou inhibidor de metal·locarboxipeptidases, el primer trobat en la paparra *Rhipicephalus bursa*. El TCI és una proteïna de 75 residus rica en cisteïnes (12). El cDNA del TCI conté un marc obert de lectura que codifica per una proteïna de 97 residus, que conté un pèptid senyal hidrofòbic de 22 residus precedint a la proteïna madura. Estudis estructurals de CD i NMR indiquen que el rTCI és una proteïna fortament constreta pels sis ponts disulfur que el mantenen correctament plegat en un ampli rang de pH i temperatura i extraordinàriament resistent a les condicions desnaturalitzants (Arolas *et al.*, 2005a). La major part del TCI no presenta similitud seqüencial amb proteïnes conegudes; tot i això, la seva cua C-terminal recorda a la dels inhibidors de *Solanacea*, *Ascaris* i *Hirudo*, fet que porta a pensar en un mecanisme d'acció comú. El rTCI inhibeix les carboxipeptidases de la subfamília A/B amb K_i s de l'ordre nanomolar.

Les paparres secreten el TCI durant la ingesta i la digestió, probablement per poder mantenir la sang en estat líquid ja que el TCI inhibeix la TAFI, un atenuador de la fibrinòlisi. A més a més, s'ha pogut observar que, *in vitro*, el TCI estimula la fibrinòlisi (Arolas *et al.*, 2005a).

S'han resolt les estructures de dos complexos del TCI amb la hCPB i la bCPA per cristal·lografia de raigs-X i ambdues són pràcticament idèntiques. L'estructura del inhibidor conté dos dominis, tot i que els dos dominis presenten una baixa homologia seqüencial entre ells, estructuralment són molt similars. El domini N-terminal està format per una α -hèlix curta i un fulla- β antiparal·lela formada per tres cadenes adquirint una topologia $\alpha\beta_1\beta_2\beta_3$. El domini C-terminal també adopta una estructura de fulla- β de tres cadenes que es troba connectada al domini N-terminal per unllaç que estructuralment és similar a una petita α -hèlix (Arolas *et al.*, 2005b).

A diferència de la resta dels inhibidors, el TCI s'uneix a la carboxipeptidasa en forma de doble cap. Els últims tres residus C-terminals interactuen amb el centre actiu de l'enzim imitant una unió enzim/substrat i constituint el lloc primari d'unió. Per altra banda, el domini N-terminal del TCI estableix interaccions amb la superfície de l'enzim, que són diferents de les que es donen en el PCI o el LCI, i que es troben localitzades al voltant del subseti S4 de la carboxipeptidasa. Si es realitza una cerca electrònica amb el servidor DALI s'observa que l'estructura d'aquests dominis és similar al plegament de la família de les β -defensines. Tenint en compte que un gran nombre dels inhibidors de trombina aïllats d'animals xucladors de sang són inhibidors de doble cap, podríem pensar en el TCI com un possible fàrmac amb propietats trombolítiques (Arolas *et al.*, 2005b).

1.1.3.6.- L'inhibidor endogen de carboxipeptidases (ECI o Latexin)

És conegut també com a inhibidor tissular de carboxipeptidases o latexina i és l'inhibidor de metal·locarboxipeptidases més gran descrit fins el moment. Inicialment es va purificar de cervell de rata després de detectar-se activitat inhibidora de CPA. El seu cDNA codifica una proteïna de 223 aminoàcids (PM 26 KDa) (Normant *et al.*, 1995). A nivell de seqüència no presenta homologia significativa amb la resta dels inhibidors de carboxipeptidasa descrits fins el moment. La seva forma recombinant produïda a *E.coli* es comporta com un potent inhibidor de tipus no-competitiu de les carboxipeptidases A1, A2, B i MCPA.

Inicialment, aquest inhibidor es va trobar àmpliament distribuït en teixits de rata (p.ex. en cervell, pulmó i tracte digestiu), però estudis posteriors van detectar una expressió dependent del tipus cel·lular tant en el sistema nerviós central com en el perifèric. L'ECI també es troba àmpliament distribuït a macròfags de ratolí i es

sobreexpressa quan s'estimulen les cèl·lules amb factors de creixement o mitjançant estímuls pro-inflamatoris. Un homòleg de la latexina ha estat identificat en humans (Liu *et al.*, 2000), i aquesta és una de les proteïnes objecte d'estudi en aquesta tesi, pel fet de reunir les qualitats de ser el primer inhibidor proteic de CPs descrit en vertebrats i posseir un pes molecular indicatiu d'una probable nova estructura.

I.1.3.7.- Mecanisme d'inhibició de les metal·locarboxipeptidases pels seus inhibidors proteics

El mecanisme d'inhibició del PCI, LCI i TCI es basa en la interacció de la seva cua C-terminal amb la cavitat del centre actiu de la carboxipeptidasa mimetitzant les unions enzim-substrat (Vendrell *et al.*, 2000; Arolas *et al.*, 2005b). Com s'observa en les estructures cristal·logràfiques del PCI, LCI i TCI en complex amb la bCPA, hCPA2, bCPA i la hCPB respectivament (Rees & Lipscomb, 1982; Reverter *et al.*, 2000; Arolas *et al.*, 2005b), la cua C-terminal de l'inhibidor s'insereix dins del centre actiu de l'enzim, constituint la regió primària d'unió.

Allà, els subsetis S1, S2 i S3 de la carboxipeptidasa també es troben establint contactes amb l'inhibidor. Les nombroses interaccions establertes contribueixen en l'alta eficiència d'inhibició. El domini N-terminal del TCI estableix interaccions amb la superfície de l'enzim que no havien estat identificades prèviament per cap altre inhibidor i que es localitzen a prop del subseti S4 de la carboxipeptidasa (Arolas *et al.*, 2005b). En conjunt, el nombre de contactes observats en els complexos TCI/bCPA i TCI/hCPB són majors, malgrat que, les constants d'inhibició del PCI, LCI i TCI per les diferents carboxipeptidases de la subfamília A/B es troben totes dintre del rang nanomolar.

En general, els plegaments de PCI, LCI i TCI són totalment diferents i també ho són en els seus respectius complexos, excepte en la seva cua C-terminal, on una seqüència similar dona un plegament de l'esquelet pràcticament idèntic. Aquest és l'únic motiu conservat entre els inhibidors de carboxipeptidases aïllats a partir d'organismes evolutivament molt distants com són, per exemple, les plantes de *Solanacea* i els animals xucladors de sang, i representen un bon exemple d'evolució convergent dictada per l'estructura tridimensional dels seus enzims diana.

I.2.- PLEGAMENT DE PROTEÏNES I AGREGACIÓ

I.2.1.- El plegament proteic

Un organisme viu conté al voltant de 100,000 tipus diferents de proteïnes les quals intervenen i controlen tots els processos dels que depèn la vida. Proteïnes diferents difereixen en la seva seqüència polipeptídica. El plegament genera una estructura tridimensional a partir d'una informació essencialment monodimensional. Així, després de la seva biosíntesi al ribosoma, cada molècula de proteïna s'ha de plegar en la seva conformació nativa codificada específicament a la seva seqüència per poder així desenvolupar la seva funció biològica. Entendre com aquest procés té lloc és un dels reptes més fascinants de la biologia estructural moderna.

A les cèl·lules, el plegament té lloc en un entorn complex a on les proteïnes i altres macromolècules assoleixen concentracions de 300-400 mg/ml. Això fa necessària la presència d'un bon nombre de proteïnes que ajuden durant el procés de plegament. Aquestes proteïnes inclouen les xaperones moleculars, que tenen com a paper principal el d'evitar que la cadena polipeptídica que no està totalment plegada estableixi interaccions incorrectes, especialment aquelles que donen lloc a fenòmens d'agregació; i els catalitzadors del plegament, que s'encarreguen d'accelerar els passos cinèticament limitants del procés de plegament com per exemple, la formació de ponts disulfur i la isomerització dels residus de prolina. Per tal de poder assolir la seva estructura funcional, la molècula de proteïna ha de poder destriar el camí cap a l'única conformació nativa entre les incomputables alternatives.

I.2.2.- Els fonaments del plegament proteic, des del postulat d'Anfinsen fins la “New View” (Nova Visió).

I.2.2.1.- El postulat d'Anfinsen i la paradoxa de Levinthal

Anfinsen va demostrar a finals dels 60 que en desplegar l'enzim ribonucleasa A amb urea i mercaptoetanol augmentava el seu volum aparent i desapareixien les seves

propietats catalítiques. En dialitzar la proteïna tornava a plegar-se. Així va demostrar que el plegament de les proteïnes no està induït per la cèl·lula sinó que és el resultat de la interacció de la seqüència peptídica amb l'aigua i va concloure que “tota la informació necessària per assolir la conformació nativa d'una proteïna està continguda en la seva seqüència aminoacídica” (Anfinsen, 1973). Això implicava, que l'estructura nativa d'una proteïna és termodinàmicament estable, representa un mínim en l'energia lliure de Gibbs i algun dia podria ser predita a partir de la seqüència.

Aquest procés va ser discutit per Levinthal, ja que a la natura el procés de plegament de les proteïnes és molt ràpid tot i que donada la flexibilitat dels polipèptids el nombre de conformacions possible d'una proteïna es enorme. Una cerca a l'atzar (representada per la hipòtesi del control termodinàmic d'Anfinsen) de totes les conformacions accessibles no pot donar-se a la realitat, donat que requeriria un temps quasi infinit. A aquest principi se'l coneix com la paradoxa de Levinthal. D'aquí que l'autor suggereixi que les proteïnes no realitzen una cerca a l'atzar, sinó que segueixen un camí concret de plegament. Aquest camí aniria conduït per un control cinètic del plegament proteic (en contraposició al control termodinàmic postulat per Anfinsen).

A partir d'aquest punt, han sorgit diversos models per explicar quins són els mecanismes subjacents al plegament de les proteïnes.

I.2.2.2.- Models clàssics de plegament proteic

Per discernir com una cadena polipeptídica pot obviar la paradoxa de Levinthal, s'han proposat diferents models de plegament tant a partir d'observacions experimentals (Kim & Baldwin, 1990) com d'observacions teòriques (Karplus & Sali, 1995). Es van desenvolupar diferents models de plegament tot basats en la teoria de la existència de “vies de plegament”. Aquesta visió del plegament suposa que totes les molècules segueixen el mateix camí passant en el seu cas pels mateixos intermediaris i estat de transició. Podem distingir fonamentalment 2 tipus de models: els seqüencials, en els quals es proposa primer la formació de l'estructura secundària en front de la terciària p.ex., el model de *framework* on després del primer esdeveniment, la nucleació, es formen les estructures secundàries que a la seva vegada s'associen per formar les estructures supersecundàries. I models de

col·lapse, en els quals l'adquisició d'ambdues estructures es produeix simultàniament p.ex. el model del col·lapse hidrofòb on el primer esdeveniment del plegament proteic consisteix en un col·lapse a través d'interaccions hidrofòbiques de llarg abast i té lloc abans de la formació de l'estructura secundària (Dill, 1985). Una de les principals aportacions del model de col·lapse hidrofòbic va ser el concepte de *molten globule*. El primer fenomen observable en la via de plegament de moltes proteïnes es el col·lapse de la cadena polipeptídica desplegada, flexible i desordenada, en un estat globular parcialment organitzat que s'anomena *molten globule*. Es caracteritza per tenir un elevat contingut d'estructura secundària nativa però es menys compacte que l'estructura nativa. No es tracta d'una única entitat estructural sinó, de fet, d'un conjunt d'estructures relacionades que s'estan interconvertint ràpidament.

Semblava, emperò, que totes aquestes "vies de plegament" eren massa rígides conceptualment i poc generals com per a explicar globalment el procés de plegament proteic.

I.2.2.3.- La nova visió: els embuts de plegament

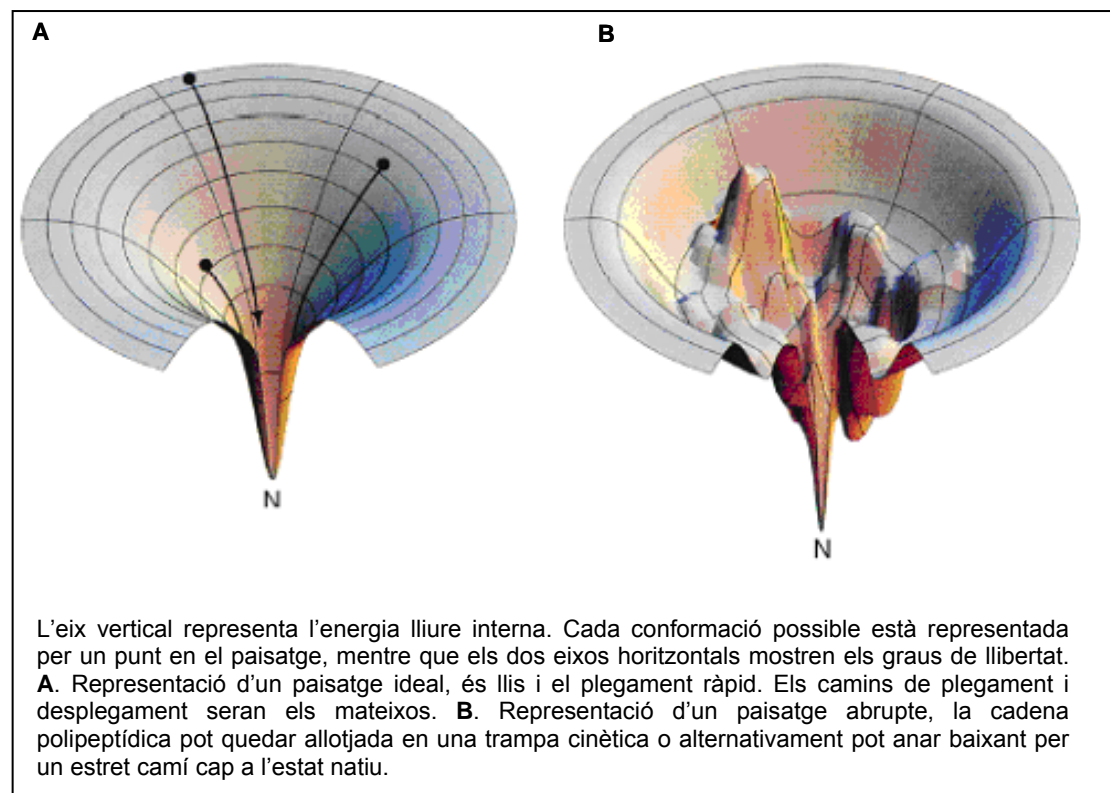
L'anomenada nova visió ha evolucionat a partir dels models clàssics seqüencials abans esmentats, durant els darrers anys gràcies a dades tant teòriques com experimentals mitjançant l'ús de models mecànics simplificats, i està il·lustrada pel model de "l'embut de plegament" (Wolynes *et al.*, 1995). En aquests models es considera que cada molècula es plega tot seguint un camí distint.

Els embuts de plegament 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 & Chan, 1997). Es basen en unes superfícies tridimensionals, on la dimensió vertical representa l'energia lliure del sistema (o també el nombre de contactes nadius presents) i les altres l'entropia i el nombre de conformacions possibles. La forma d'embut ve donada pel fet que se suposa que es parteix d'un conjunt molt ampli i heterogeni de conformacions desplegadas de màxima energia, el qual es va reduint a mida que disminueix tant l'energia com l'entropia durant el procés de plegament, fins a arribar a l'estructura nativa, on el nombre de conformacions és extremadament limitat, pràcticament únic, tot i que dinàmic (Fig. I.6)

Aquest model sembla que és genèric i a la vegada explica els plegaments característics de diferents proteïnes, de manera que el que variaria serien només les formes dels embuts, característiques de cada proteïna (de cada seqüència) i de les condicions on es porti a terme el plegament. Contempla, també, l'existència d'intermediaris, els quals de fet són mínims energètics locals, a diferència de l'estructura nativa, que és considerada un mínim energètic global.

A mesura que la mida i la complexitat de les proteïnes augmenta el procés de plegament esdevé més complex. Intermediaris que només tenen estructures parcialment formades poden començar a ser abundants i tenir temps de vida significatius. Les xaperones moleculars són capaces de mitigar en part aquest comportament i donar protecció per a les cadenes polipeptídiques parcialment plegades (Ellis & Hartl, 1999), però la idea de que les proteïnes puguin, sota certes circumstàncies, escapar aquest control desplegant-se, o plegant-se en intermediaris que poden portar a processos com l'agregació, ens fan adonar dels problemes que poden derivar-se d'un mal plegament proteic, doncs les proteïnes no tan sols no assolirien la seva funció sinó que podrien guanyar una nova funció no desitjada com és la toxicitat cel·lular.

Figura I.6.- Diferents paisatges energètics representats en forma d'embuts de plegament



Els fenòmens de plegament i desplegament proteic es troben associat a molts dels processos essencials d'un sistema biològic com són, el translocament de proteïnes a través de la membrana, el tràfic proteic, la secreció extracel·lular de proteïnes, i el control i regulació del cicle cel·lular (Radford & Dobson 1999). És per tot això, que la incapacitat de les proteïnes a plegar-se, o de mantenir-se correctament plegades en condicions fisiològiques, és la causa de mal funcionament i origina malalties.

Un bon nombre de les malalties humanes s'associen amb aberracions durant el procés de plegament. Algunes d'aquestes malalties (p.ex. la fibrosi quística) resulten del fet que si les proteïnes no es pleguen correctament no serà possible que puguin exercir les seves funcions de forma efectiva. En altres casos, les proteïnes desplegadas escapen de tots els mecanismes de protecció i formen agregats intractables dintre de les cèl·lules o a l'espai extracel·lular. Un nombre creixent de patologies, incloent les malalties de l'Alzheimer, el Parkinson i les encefalopaties espongiformes, es coneix que estan directament associades amb la deposició d'aquests agregats en els teixits. Les malalties d'aquest tipus es troben entre les més debilitants, disruptores socials i cares del món modern, i donat que són degeneratives i associades a l'edat estan esdevenint cada cop més prevalents a mesura que la nostra societat envellaix.

I.2.3.- Agregació proteica i formació d'amiloides

Cada malaltia amiloïdòtica està relacionada amb l'agregació majoritària d'una proteïna específica, tot i que un seguit d'altres proteïnes i carbohidrats són incorporats als depòsits quan aquests es formen *in vivo*. En les malalties neurodegeneratives, la quantitat dels agregats implicats pot ser tan petita que sigui indetectable, en canvi, en algunes malalties sistèmiques es poden trobar quilograms d'aquests en els òrgans (Tan & Pepys, 1994). Les característiques de les formes solubles de les 20, o més, proteïnes implicades en les diferents amiloïdosis identificades són molt variades, sense cap relació seqüencial, estructural o de mida, comprenent des de proteïnes globulars grans fins a molècules peptídiques petites altament desestructurades. Tot i això, els agregats que formen tenen moltes característiques en comú (Sunde & Blake, 1997) (Taula I.3.).

Els depòsits d'amiloides presenten un comportament òptic específic (com la birrefringència) en unir-se a certes molècules colorants com el congo red o la thioflavina-T. La morfologia comú de les fibres amiloides pot observar-se fàcilment al

microscopi electrònic; les fibres formades per diferents proteïnes comparteixen un diàmetre entre 7 i 12 nm i estan sempre formades per un conjunt de protofilaments entre 0.2 – 0.35 nm que s'associen entre ells (Jiménez *et al.*, 2002).

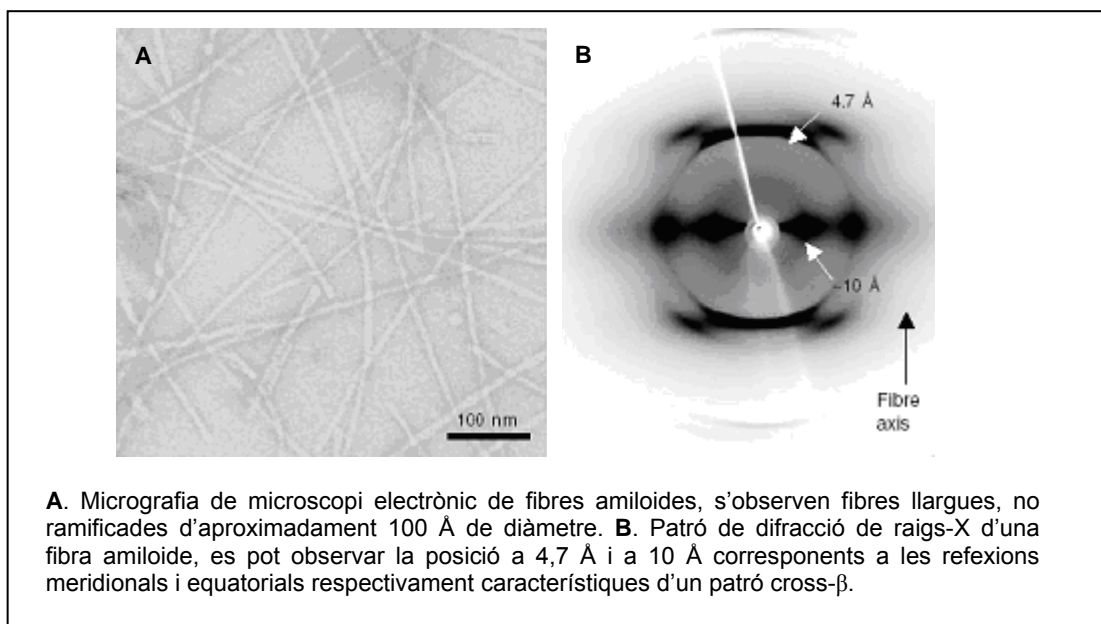
Taula I.3. Exemples de malalties conformacionals. Reproduït parcialment de Dobson, 2001.

Malaltia	Proteïna involucrada	Lloc de plegament
Hipercolesterolèmia	Receptor de la lipoproteïna de baixa densitat (LDL)	RE
Fibrosi cística	Regulador transmembrana de la fibrosi cística	RE
Fenilcetonúria	Fenilalanina hidroxilasa	Citosol
Malaltia de Huntington	Huntingtina	Citosol
Anèmia falciforme	Hemoglobina	Citosol
Malaltia de Tay-Sachs	β -hexosaminidasa	RE
Alzheimer	β -amiloide / presenilina	RE
Malaltia de Parkinson	α -sinucleïna	Citosol
Scrapie / malaltia de Creutzfeldt-Jacob (humans)	Proteïna priònica (Prp)	RE
Amiloïdosis familiars	Transtirretina / lisozim	RE
Retinosi pigmentària	Rodopsina	RE
Cataractes	Cristal·lins	Citosol
Càncer	p53	Citosol

Per difracció de raigs X totes les fibres amiloides presenten un patró típic d'estructura en fulla- β creuada, a on les diferents fulles- β formades per la proteïna agregada es troben ordenades en direcció perpendicular a l'eix de la fibra (Fig. I.7). L'elevat contingut d'estructura secundària ordenada que contenen aquests agregats, permet, així mateix, el seu estudi amb tècniques clàssiques per monitoritzar aquestes estructures proteiques com poden ser el dicroisme circular o l'espectrometria d'infraroig. Aquestes tècniques són especialment útils per estudis inicials dels processos agregatius, on es produeix indefectiblement un guany de nova estructura beta respecte a la conformació nativa de la proteïna.

Durant molts anys s'havia assumit que la habilitat de formar estructures amiloides estava restringida al petit nombre de proteïnes que estan associades a malalties i que d'alguna manera aquestes posseïen seqüències amb motius específics que codificaven pel nucli de l'estructura amiloide. Estudis recents suggereixen que la capacitat de formar aquestes estructures es pot considerar una habilitat general de totes les cadenes polipeptídiques (Dobson, 2001).

Figura I.7.- Estructura de les fibres amiloides



L'evidència més directa és que al laboratori es pot induir la formació de fibres de moltes proteïnes diferents tot i que no estan associades amb malalties, incloent proteïnes globulars riques en estructura hèlix- α com la mioglobina (Fändrich *et al.*, 2001), així mateix, també s'ha demostrat que homopolimers com politreonines o polilisines poden formar aquestes estructures en condicions adients (Fändrich & Dobson, 2002) (Taula I.4.)

Taula I.4. Proteïnes amiloïdogèniques no patogèniques. Exemples de proteïnes formadores de fibres amiloides no relacionades amb malalties conegudes. Reproduït de Dobson, 2001.

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

Tot i que la habilitat de formar fibres sembli ser general, la propensió per fer-ho varia molt entre seqüències diferents. Això implica que la concentració de cadena polipeptídica desplegada que es necessita assolir per a que es doni l'agregació serà major per uns polipèptids que per altres. A més a més, el procés d'agregació, com el de cristal·lització, necessita, en la majoria dels casos, ser nucleat i el llindar en el que s'inicia aquest procés és fortament depenent de molts factors diferents. Es coneix que el canvi d'un únic residu de la seqüència de la proteïna pot fer canviar la taxa a la que una cadena polipeptídica desestructurada agrega, en un ordre de magnitud o més. S'ha vist que és possible correlacionar canvis en la taxa d'agregació deguda a mutacions amb els canvis en les propietats físico-químiques de les respectives cadenes laterals, com la càrrega, la hidrofobicitat, els patrons polar-apolar i la propensió d'adoptar certs elements d'estructura secundària (Chiti *et al.*, 2003; de Groot *et al.*, 2006).

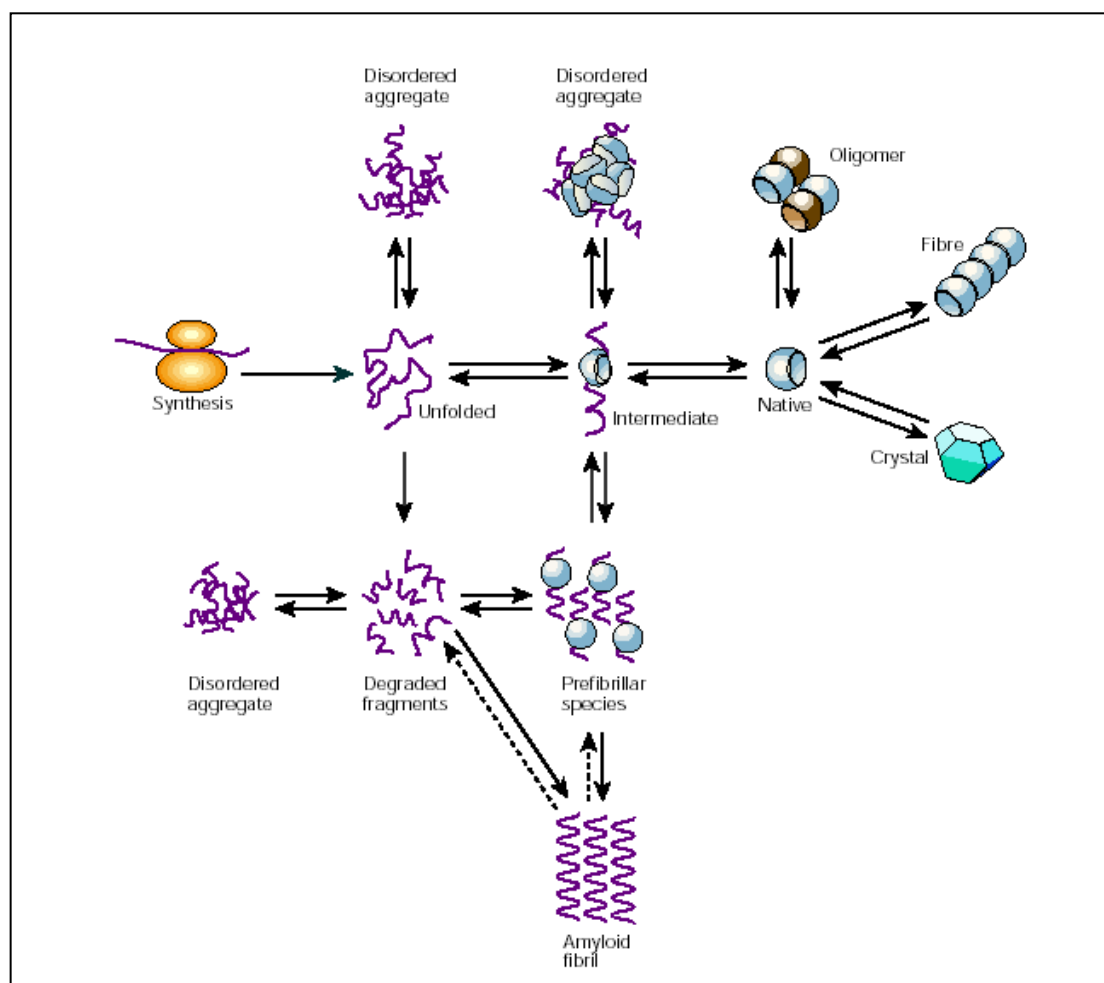
Partint de la base de que totes les fibres, independentment de l'estructura original de la proteïna amiloïdogènica, comparteixen una estructura comuna en fulla- β creuada, això implica que la formació de l'estructura fibril·lar suposa fortes reorganitzacions conformacionals de les cadenes polipeptídiques així com l'establiment d'un nou ordre d'interaccions no covalents. Aquests canvis no poden tenir lloc en la típica conformació fortament empaquetada de les proteïnes natives, degut a les constriccions conformacionals i termodinàmiques que imposa l'estructura terciària i al fet de que, com es veurà després, la majoria de regions amiloïdogèniques estan enterrades o implicades en interaccions intra-moleculares, i per tant, protegides. Això apareix com una exitosa estratègia evolutiva per evitar l'agregació, doncs molt poques proteïnes són capaces d'agregar directament des del seu estat natiu i estable. Es per tant, necessària en la majoria dels casos la formació d'una conformació no-nativa parcialment desplegada. Aquesta conformació parcialment estructurada possibilita interaccions intermoleculares específiques que són necessàries per a l'oligomerització i fibril·lació (Fig. 1.8).

La formació de fibres tipus amiloide no és l'únic distintiu de les malalties conformacionals. En diversos desordres, així com en nombrosos experiments *in vitro*, els depòsits de proteïna estan compostats d'agregats amorfs, sense un ordre local. De la mateixa manera, els oligòmers solubles representen un altre producte final alternatiu del procés d'agregació. L'elecció entre els tres camins de l'agregació, fibril·lació, formació d'agregats amorfs o l'oligomerització, sembla estar determinada tant per la seqüència aminoacídica com per l'entorn a on es troba la proteïna.

1.2.4.- Les proteïnes amiloïdogèniques

Com s'ha dit, les proteïnes implicades en aquestes malalties no es troben relacionades a nivell de seqüència ni de contingut d'estructura secundària. Des d'un punt de vista conformacional es poden distingir dues classes principals: les proteïnes globulars que tenen una única conformació en l'estat natiu i per altra banda, les proteïnes que es troben intrínsecament desestructurades (Uversky & Fink, 2004).

Figura I.8.- . Esquema de formació de fibres amiloides



Com s'ha esmentat, **les proteïnes globulars** rarament agreguen des del seu estat natiu, i es requereix de la desestabilització d'aquesta conformació. La desestabilització provoca un augment de la població de molècules total o parcialment desplegadas en un moment donat i s'ha establert com un factor desencadenant en els desordres associats amb els dipòsits de proteïnes que són

globulars en els seus estats funcionals normals, com és el cas de la β 2-microglobulina, el lisozim, la transtirretina i la proteïna del prió.

Les proteïnes intrínsecament desestructurades són proteïnes que sembla que necessitin d'un grau elevat de desordre estructural en el seu estat natiu per desenvolupar les seves funcions (Wright & Dyson, 1999; Uversky, 2002a; Tompa, 2002). La plasticitat estructural de les proteïnes no estructurades en el seu estat natiu s'ha suggerit que afavoriria les seves interaccions amb els lligands (Iakoucheva *et al.*, 2002; Uversky, 2002b). La selecció natural ha dotat a aquestes proteïnes d'un alt contingut en residus polars i carregats per evitar la seva agregació a partir de zones hidrofòbiques exposades al dissolvent. Tot i això, dintre d'aquest grup trobem proteïnes que donen lloc a malalties humanes devastadores, incloent-hi l'amilina, la proteïna β -amiloide i l' α -sinucleïna entre d'altres. El principal problema rau en que en aquests casos no es requereix d'un procés de desplegament i l'agregació té lloc per un autoacoblament directe de les cadenes polipeptídiques no estructurades a partir d'una certa concentració crítica, que malauradament s'assoleix en molts casos en condicions fisiològiques.

1.2.5.- Regions facilitadores de l'agregació

Un dels interrogants principals pel que respecta a l'agregació de proteïnes és el de l'especificitat amb la que la seqüència primària determina la propensió d'agregació des d'estats total o parcialment desplegats. Conèixer la resposta a aquesta qüestió permetria la predicció de les regions proteiques implicades en l'agregació tot obrint la possibilitat de controlar la formació de depòsits no desitjats mitjançant l'ús de molècules específiques dirigides a aquestes dianes. Un primer avenç en aquesta direcció és el recent descobriment que indica que no totes les regions d'una cadena polipeptídica són igual d'importantes a l'hora de determinar la seva tendència a agregar, tant en les proteïnes desplegades com en les proteïnes globulars. En aquesta direcció s'ha provat recentment que regions molt curtes d'aminoàcids poden actuar com a disparadors o inhibidors de la formació de fibres amiloides (Ventura *et al.*, 2004; Ivanova *et al.*, 2004). Aquestes regions són col·loquialment conegudes com a "hot spots". Aquestes seqüències facilitadores de l'agregació es troben normalment bloquejades en l'estat natiu de les proteïnes globulars ja que acostumen

Introducció

a estar amagades al interior de la proteïna o participant en la xarxa de contactes que estableixen les proteïnes. Així, es pot dir que de fet, el plegament i l'agregació són processos contraposats que estan contínuament competint dintre de la cèl·lula, evitant o promovent els contactes intermoleculars d'aquestes regions, respectivament. En canvi, en el cas de les proteïnes intrínsecament desestructurades aquestes regions facilitadores de l'agregació es troben ja accessibles per establir contactes intermoleculars que poden acabar amb la formació d'agregats. D'acord amb això, s'ha comprovat que la presència de "hot spots" és molt més freqüent en les seqüències de proteïnes globulars que no en les que codifiquen per proteïnes nativament desestructurades (Linding *et al.*, 2004). A hores d'ara s'ha predit la presència de regions facilitadores de l'agregació en la majoria dels pèptids i proteïnes implicades en desordres neurodegeneratius i en amiloïdosis sistèmiques (Fernández-Escamilla *et al.*, 2004; Pawar *et al.*, 2005).

I.3.- CRISTAL·LITZACIÓ DE PROTEÏNES I RESOLUCIÓ DE LA SEVA ESTRUCTURA

I.3.1.- La cristal·lització de proteïnes

Per tal de determinar l'estructura tridimensional d'una proteïna a nivell atòmic disposem de dues aproximacions: l'ús de la ressonància magnètica nuclear (NMR) o la utilització de la difracció de raigs-X sobre cristalls de proteïnes. Altres tècniques com ara la microscòpia electrònica o la microscòpia de força atòmica ofereixen alternatives prometedores, però encara no estan llestes com a mètodes usuals per determinar el model atòmic d'una cadena polipeptídica, subministrant, en la majoria dels casos, una imatge de baixa resolució del relleu general de la superfície de la proteïna.

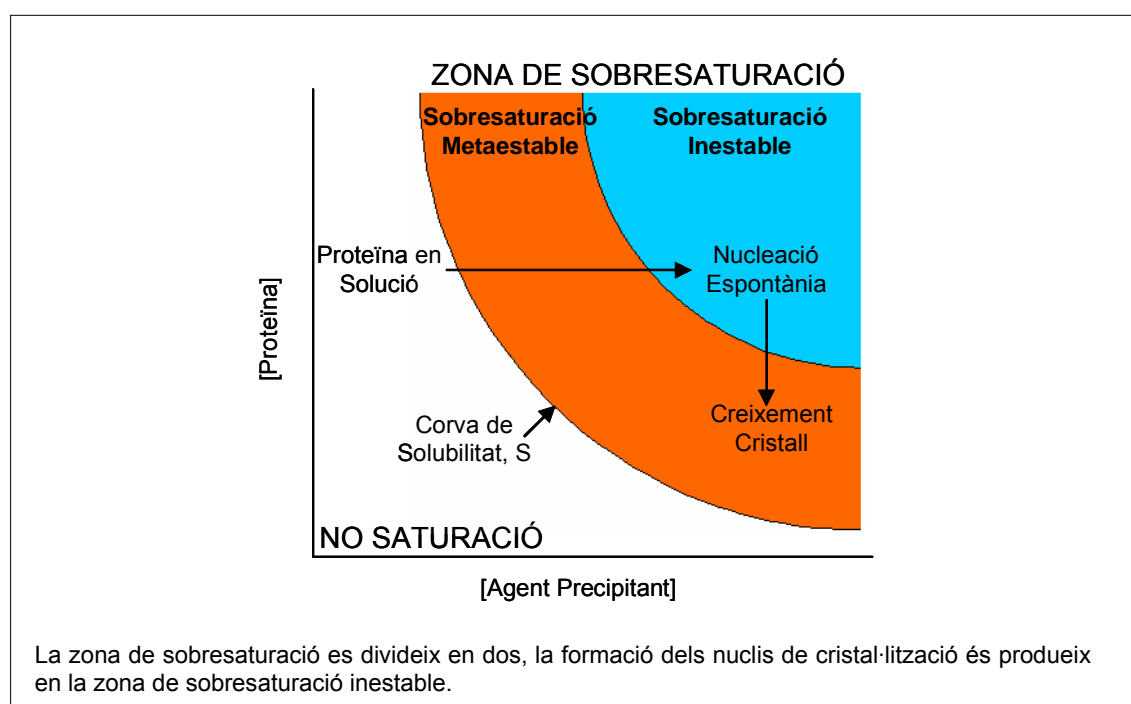
La RMN té el seu camp d'aplicació en les proteïnes de fins a 250-300 residus que puguin preparar-se pures a altes concentracions i mantenir-se solubles. Ens permet l'estudi de la dinàmica de les proteïnes així com la caracterització d'estats de la proteïna que no són accessibles a la cristal·lografia i que són importants en l'estudi de les vies de plegament. La tècnica d'elecció per a determinar l'estructura de proteïnes més grans així com complexes macromoleculars és la difracció de raigs-X. Però això, només és possible si disposem de cristalls de la proteïna.

Els cristalls de qualsevol substància s'aconsegueixen partint d'una dissolució sobresaturada, mitjançant l'addició ordenada de molècules de solut a un nucli inicial de cristal·lització (Fig. I.9). La condició de sobresaturació suposa que hi ha d'haver més proteïna dissolta que la que accepta la dissolució en el punt de saturació. Si la dissolució estigués solament saturada no es produiria el canvi de fase que volem: la saturació és un punt d'equilibri, en el que la dissolució ha acceptat tot el solut que es pot dissoldre, per això no hi ha la tendència a que el solut deixi de ser-ho i es separi com a sòlid. Per a la cristal·lització d'una proteïna, doncs, cal assolir un estat de sobresaturació; en aquesta condició, es poden produir uns quants nuclis de cristal·lització sobre els que creixin cristalls més grans i ben ordenats. Si la concentració de proteïna sobrepassa molt el punt de saturació la situació es torna inestable, donant lloc, molt sovint, a la precipitació en forma d'agregats amorfs absolutament inadequats per a la difracció. És per això que l'obtenció de cristalls requereix assolir la sobresaturació poc a poc, de forma gradual.

Introducció

Les condicions òptimes en les que s'ha de trobar una proteïna per tal d'iniciar un experiment de cristal·lografia són: elevada concentració, idealment de 10 mg/mL o més; màxima puresa i mínima heterogeneïtat de la mostra, que normalment s'associa a modificacions post-traducionals, especialment les glicosilacions. A més, és desitjable que la mostra sigui el més estable possible, doncs aquesta mena d'experiments poden arribar a durar mesos.

Figura I.9. Diagrama de solubilitat d'una proteïna en presència d'un agent precipitant

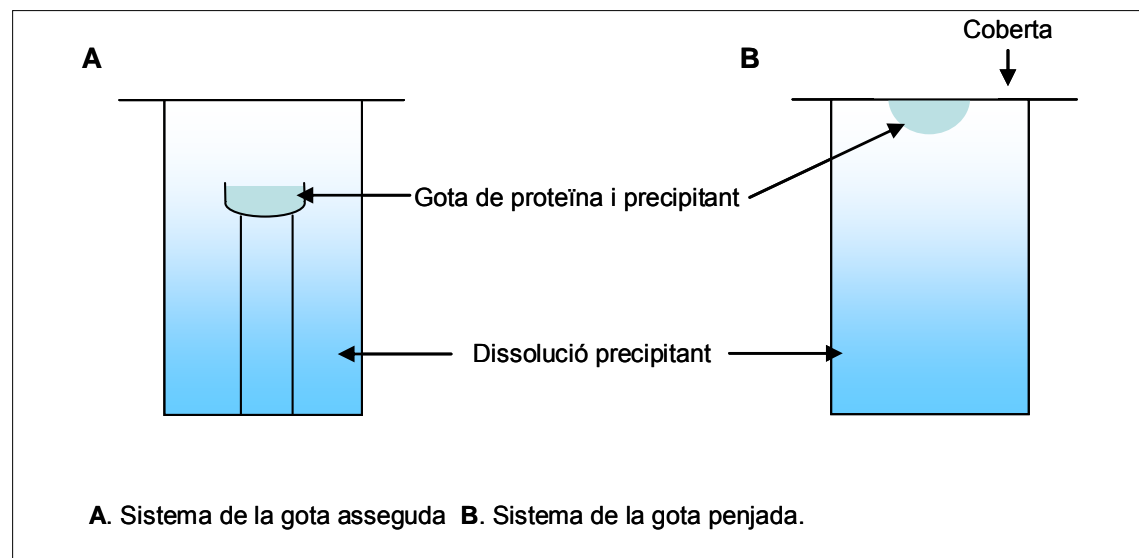


Per tal d'arribar al punt de sobresaturació hi ha dues aproximacions que es complementen: l'augment de la concentració de proteïna i la reducció de la seva solubilitat. Per augmentar la concentració de proteïna s'elimina l'aigua de la dissolució mitjançant un procés d'evaporació que es coneix com difusió de vapor. La reducció de la solubilitat de la proteïna s'assoleix manipulant el pH, la temperatura, la presència o absència de sal, l'addició de solvents orgànics o la de precipitants polimèrics.

Hi ha varies tècniques per a la cristal·lització de proteïnes però les més comunes es basen, com ja s'ha dit, en assolir la sobresaturació mitjançant la difusió de vapor. Una gota preparada barrejant a parts iguals la dissolució de proteïna i la de cristal·lització es manté dins d'una cambra hermètica formada per un dipòsit que conté un volum molt major de la dissolució de cristal·lització, sense que la gota i la

dissolució del reservori entrin en contacte. La pressió de vapor en la gota és major que en el dipòsit, i per tant, la gota perd poc a poc aigua fins que les concentracions de precipitants s'igualen. Existeixen principalment dues modalitats d'aquesta tècnica (Fig. I.10). L'anomenada de la gota penjada, en la que aquesta es diposita sobre un cobreobjectes siliconat que s'inverteix i s'utilitza per a segellar un pouet i la de la gota assegurada en la que la gota s'assenta sobre una elevació plana o còncava situada per sobre del nivell de la dissolució de cristal·lització.

Figura I.10. Esquema de cristal·lització mitjançant difusió de vapor



Per tal de trobar les condicions òptimes de cristal·lització, i ja que el nombre de factors a controlar es excessiu per realitzar cerques sistemàtiques en que s'explorin tots els factors possibles, s'han dissenyat estratègies de mostreig que utilitzen matrius que exploren tot l'espai de cristal·lització. Quan s'obtenen cristalls, si aquests són de proteïna i no de sal, cal, en la majoria dels casos, millorar-los i trobar les condicions en les que els cristalls resulten més grans i homogenis. Això es duu a terme provant diferents condicions de cristal·lització al voltant de les que han produït els primers cristalls.

1.3.2.- Estructura de proteïnes per cristal·lografia de raigs-X

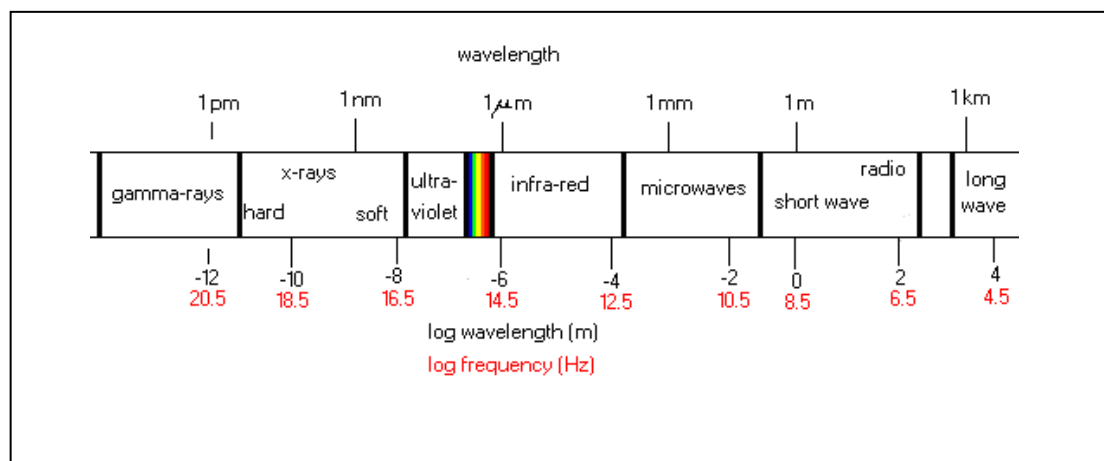
1.3.2.1.- Principis de la cristal·lografia de raigs-X

Per què raigs-X? El “límit de difracció” implica que no ens podem visualitzar coses que siguin molt més petites que la longitud d’ona que estem fent servir per il·luminar-les. La λ de la llum visible és de centenars de nm; en canvi els àtoms estan separats per distàncies de l’ordre d’1 Å. Mirant l’espectre electromagnètic, els raigs-X ens donen el rang de longitud d’ona adient. (Fig. I.11)

Per què densitats electròniques? El que observem com a resultat d’un experiment cristal·logràfic, no és una “foto” real dels àtoms sinó un mapa de la distribució dels electrons en la molècula. Però, tenint en compte que els electrons es troben majoritàriament localitzats al voltant del nucli, el mapa de densitat electrònica ens proporciona una bona imatge de la molècula.

Per què cristalls? La dispersió de raigs-X a partir d’una única molècula seria massa dèbil com per a poder ser diferenciada del soroll de fons. Un cristall ordena un elevat nombre de molècules en una mateixa orientació, de forma que les ones dispersades poden sumar-se permetent l’augment de la senyal fins a un nivell detectable.

Figura I.11. Espectre electromagnètic

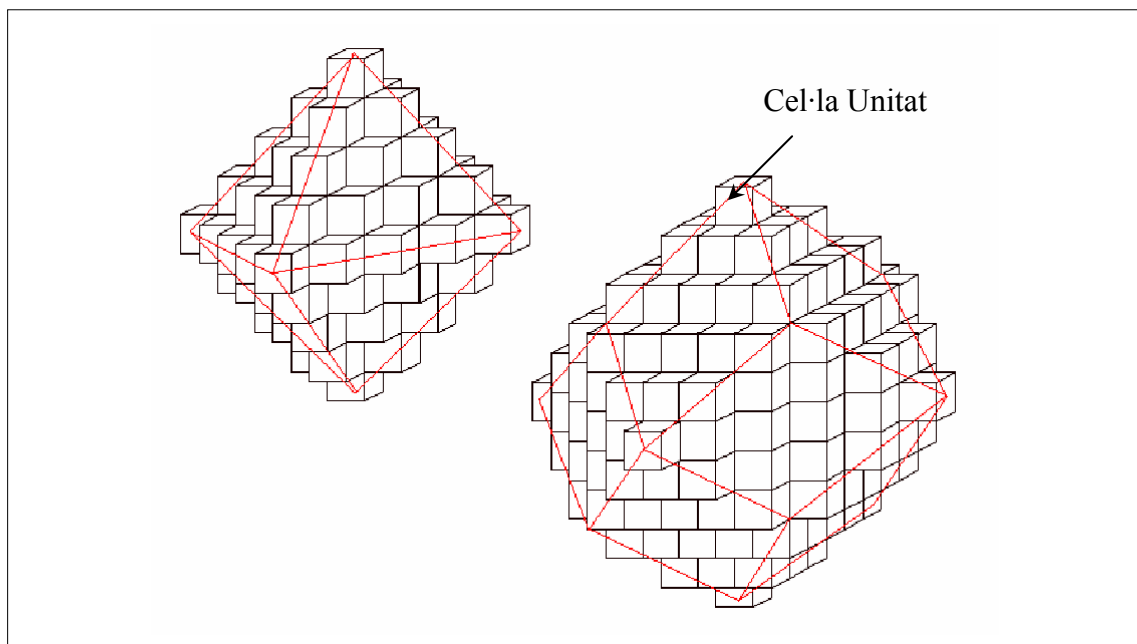


Els cristalls són agrupacions de molècules que segueixen un model de repetició periòdica tridimensional. Els components dels cristalls es repeteixen des de pocs fins a centenars d’angstroms (Å) ($1 \text{ Å} = 10^{-10} \text{ m}$) i aquesta repetitivitat, en tres dimensions, s’anomena xarxa cristal·lina (Fig. I.12). El conjunt mínim que es repeteix es coneix com la **cel·la unitat**. Dins de la cel·la unitat podem trobar una o més molècules de proteïna organitzades al voltant d’un eix de simetria. En funció del nombre i

disposició de les molècules que formen la cel·la unitat trobem diferents eixos de simetria (binaris, ternaris, quaternaris i senaris).

La difracció dels raigs-X és el fenomen físic fonamental a través del qual es manifesta la interacció dels raigs amb els cristalls. Quan un feix de raigs-X incideix sobre un cristall aquest interacciona amb els electrons dels àtoms que formen la proteïna, fent-los vibrar acoblament amb les vibracions periòdiques del seu camp elèctric. Els electrons oscil·lants es converteixen així en focus de nova radiació X que s'emet en forma esfèrica, fenomen que anomenem dispersió, normalment elàstica. Els raigs-X dispersats pels electrons interfereixen entre si, i ja que els àtoms d'un cristall estan ordenats de manera regular i periòdica també es produeix que, en determinades direccions, es combinen reforçant-se per donar lloc a dispersions cooperatives, també anomenades difraccions.

Figura I.12. Esquema de dos cristalls on es mostra la repetitivat de la cel·la unitat



La descripció del fenomen de difracció fou realitzat el 1915 per William H. Bragg i W. Lawrence Bragg a la Universitat de Cambridge, en el seu model proposen que la difracció dels raigs-X en un cristall es produeix de forma similar a les reflexions òptiques. Aquestes reflexions es produeixen sobre uns plans virtuals de la xarxa cristal·lina (Fig. I.13). La difracció es produeix per la interferència entre dos raigs paral·lels que pertanyen al mateix front d'ona que es reflexen en dos plans diferents.

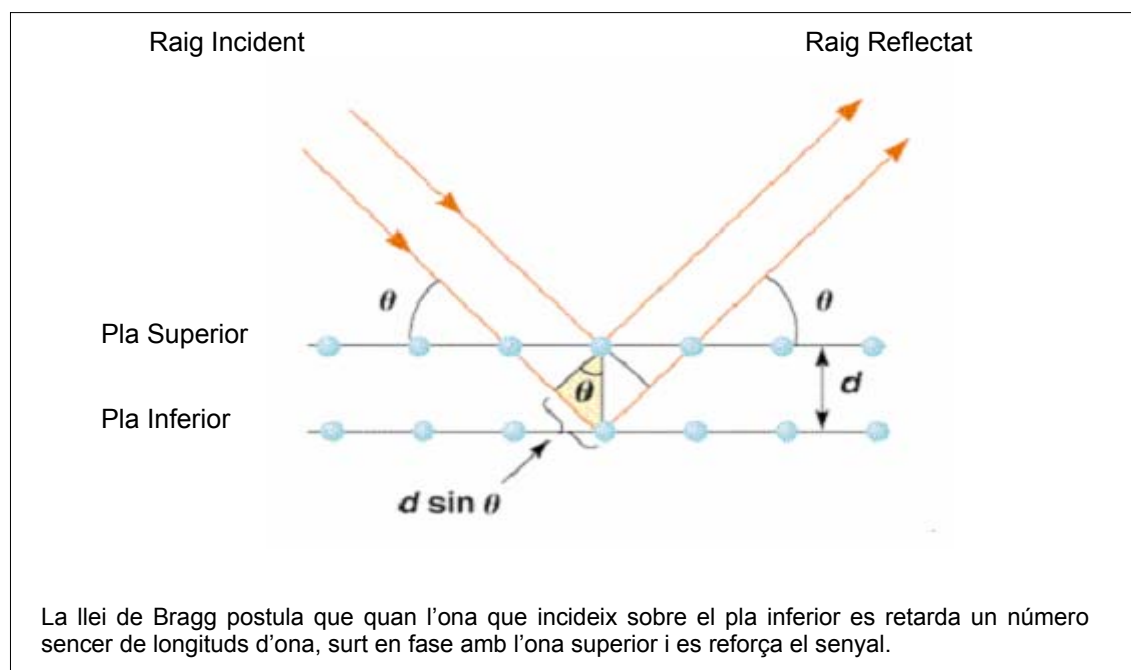
Introducció

L'anomenada condició de difracció de Bragg s'obté quan dita interferència és constructiva, i això passa quan la diferència de camins òptics recorreguts pels dos raigs és un nombre sencer de longituds d'ona. La representació matemàtica d'aquesta condició és $2d \sin\theta = n\lambda$ (Fig. I.13) on d és la distància entre dos plans, θ l'angle en que incideix el raig, n el nombre sencer de longituds d'ona i λ la longitud d'ona.

Per tal que es puguin produir tots els feixos difractats possibles, el cristall ha de girar-se de tal manera que tots els plans virtuals es col·loquin en la posició adequada per complir la llei de Bragg. Per últim els feixos difractats incideixen sobre un detector electrònic que recull la informació per poder-la processar posteriorment.

L'objectiu de l'estudi per difracció de raigs-X és la determinació de la densitat electrònica, $\rho(xyz)$, per a cada punt de la cel·la unitat. Per a obtenir-la partim de la informació de l'espai recíproc (espectre de difracció).

Figura I.13. Representació gràfica del fenomen descrit per W.H.Bragg i W.L. Bragg en que els raigs-X es difracten en plans virtuals del Cristall

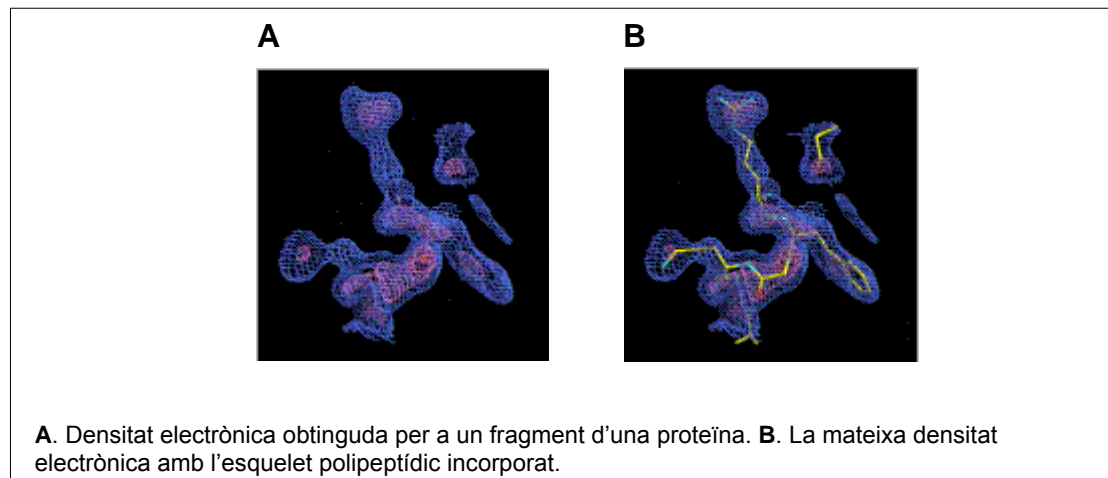


La transformació de l'espai recíproc a l'espai cristal·lí implica la resolució d'una funció anomenada transformada de Fourier. Per tal de resoldre aquesta equació, i per tant obtindre la densitat electrònica del cristall, cal conèixer l'amplitud dels raigs difractats així com la seva fase. L'amplitud es pot deduir de les intensitats dels raigs

X difractats però les fases no es poden mesurar directament, el que constitueix el que es coneix com el problema de les fases. Hi ha diferents maneres de resoldre el problema de les fases:

El mètode del *Molecular Replacement*: Aquest mètode només es pot usar si disposem d'un model estructural d'una proteïna amb una seqüència d'aminoàcids prou homòloga. L'estructura de la proteïna homòloga es considera com si fos la proteïna que va a determinar-se i serveix com un primer model que posteriorment serà refinat. Cal transferir l'estructura molecular de la proteïna coneguda, des del seu propi empaquetament cristal·lí al cristall de la proteïna desconeguda. El posicionament de la molècula coneguda en la cel·la unitat de la desconeguda requereix determinar la seva correcta orientació i la posició precisa. Les operacions de rotació i translació que cal dur a terme per a un correcte posicionament de l'estructura coneguda es calculen mitjançant unes funcions específiques. Una vegada hem traslladat adequadament la molècula coneguda es calcula un mapa de densitat electrònica utilitzant les dades obtingudes experimentalment (Fig. I.14A).

Figura I.14.



Finalment cal acoblar l'esquelet polipeptídic dins del mapa de densitat electrònica (Fig. I.14B). Això es realitza mitjançant l'ús de programes informàtics de tractament d'imatges. La qualitat dels mapes de densitat electrònica depèn de l'exactitud de les dades de difracció i per tant determinen la qualitat final de l'estructura. Per valorar la qualitat del model estructural generat també es comprova que aconsegueixi els criteris estereoquímics establerts, que hi hagi absència de xocs entre cadenes, les distàncies i els angles d'enllaç siguin adequats i els paràmetres tèrmics raonables.

Introducció

També és important contrastar que les conformacions que adopten els enllaços N-C α i C α -C de l'enllaç peptídic no donin lloc a aproximacions interatòmiques energèticament improbables.

Multiple isomorphous replacement (MIR): És el mètode clàssic per determinar fases d'estructures noves. En aquest mètode es comparen varis conjunts de dades obtinguts per difracció de raigs-X en diversos cristalls derivats de la proteïna, els quals són el resultat de l'addició d'ions d'un determinat àtom pesat. Aquests ions poden localitzar-se amb precisió i serveixen com a referència artificial. La comparació de l'efecte dels ions en els diferents conjunts de dades adquirits permet la determinació de les fases.

Multiple wavelength anomalous dispersion (MAD): És el més emprat actualment per a la determinació de fases en estructures noves. A partir d'un únic cristall derivat es va variant la longitud d'ona dels raigs-X que incideixen en l'àtom pesat i es recullen conjunts de dades per a cada una d'elles. El mateix àtom pesat actua de manera diferent en funció de la longitud d'ona i la comparació de l'efecte dels ions en els diferents conjunts de dades adquirits ens donarà les fases.

OBJECTIUS

Els objectius i treball d'aquesta tesi s'inscriuen dins l'àmbit descrit a la introducció i que constitueixen dues de les línies de recerca principals del nostre grup: la relació estructura/funció en metal·locarboxipeptidases i inhibidors i, per altra banda, l'estudi del plegament i de l'agregació de proteïnes.

Un primer bloc d'objectius comprèn l'estudi d'aspectes estructurals de carboxipeptidases humanes de la subfamília A/B i del seu inhibidor endogen humà (ECI).

Els objectius d'aquesta primera part han estat:

- Cristal·lització i resolució per difracció de raigs X de l'estructura tridimensional de la variant humana de l'enzim de referència entre les carboxipeptidases: la CPA1.
- Comparació de l'estructura tridimensional de la CPA1 humana amb la d'altres carboxipeptidases.
- Aïllament i clonació de l'inhibidor endogen de carboxipeptidases humà.
- Obtenció de l'inhibidor en forma pura mitjançant la seva expressió recombinant i posterior purificació.
- Caracterització de l'activitat de l'inhibidor envers diferents proteases.
- Determinació de la seva estructura tridimensional en complex amb la carboxipeptidasa CPA4 humana i elucidació del seu mecanisme d'inhibició.
- Caracterització de la família proteica a la que pertany la latexina en base a cerques d'homologia estructural.

La segona part d'aquesta tesi s'ha centrat en l'estudi d'aspectes de l'agregació de proteïnes emprant com a models l'inhibidor de carboxipeptidases humà i la quimotripsina. Els objectius concrets d'aquesta part del treball han estat:

- Estudi del procés de desplegament i l'estabilitat conformacional de la latexina
- Caracterització del procés d'oligomerització de la latexina i anàlisi de les propietats amiloïdogèniques dels oligòmers.
- Identificació de seqüències aminoacídiques implicades en l'oligomerització de la latexina. Anàlisi i comparació de les propietats amiloïdogèniques d'una d'aquestes regions amb les de la proteïna sencera.
- Caracterització del procés de formació d'estructures amiloïdogèniques de la primera proteïna amb plegament de barril- β , l' α -quimotripsina.

Objectius

- El·laboració i validació d'un mètode que permeti la detecció de regions que puguin dirigir el procés d'agregació d'una cadena polipeptídica i que permeti la predicció dels efectes de canvis seqüencials en la propensió a agregar dels polipèptids.

MATERIALS I MÈTODES

Pròleg

Aquest apartat general de materials i mètodes de la tesi pretén recollir exhaustivament els protocols i procediments generals seguits en les diferents parts d'aquest treball. Els capítols de resultats i discussió que es presenten més endavant inclouen apartats de materials i mètodes propis que indiquen els procediments específics seguits.

M.1.- APARELLS I PRODUCTES QUÍMICS

M.1.1.- Aparells d'ús general

Agitador rotatiu de gels Biotron (Espanya).

Agitador vòrtex Heidolph, model Reax 2000 (Alemanya).

Agitadors magnètics SBS, model A-03.

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).

Cabina de flux laminar Faster, model TWO-30 (Itàlia).

Centrifugues Beckman models J2-HS i J2-21 (EE.UU.) amb rotors JA20 i JA14.

Concentrador Speed Vac, refrigerador i bomba VP100 Savant (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 d'electroforesi per gels d'agarosa Miniphor Submarine Electrophoresis Unit LKB, model 2013 (Suècia) i cubeta Pharmacia, model GNA-100 (Suècia).

Equip de seqüenciació Beckman, model LF 3000 Protein Sequencer (EE.UU.).

Equip de seqüenciació de DNA ALF, Pharmacia (Suècia).

Espectrofotòmetre UV-Vis Cary100 Bio, Varian (EE.UU.).

Espectròmetre de masses Bruker model Ultraflex (Alemanya).

Espectropolarímetre Jasco-810

Estufa incubadora de cultius Sanyo, model Mir-152 (Japó).

Fluorímetre Perkin Elmer, model 650-40 (Japó).

Font d'alimentació per a electroforesi Hoefer Scientific Inst., model PS500XT (EE.UU.).

FPLC Amersham Pharmacia Biotech, model Äkta purifier (Suècia).

Incubador d'aire amb agitació orbital Braun, model Certomat S (Alemanya) i incubador Infors AG CH-4103 (Suïssa).

Liofilitzador Virtris

Microcentrifuga IEC, model Micro Max (EE.UU.) i microcentrifugues refrigerades Sigma, models 2MK i 2K15 (Alemanya).

Microscopi òptic Leica, model DMRB (Alemanya).

Materials i Mètodes

Microscopi electrònic de transmissió Hitachi, model H-7000 (Japó).

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

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

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

M.1.2.- Productes químics

Els productes químics d'ús general utilitzats han estat de grau analític de les firmes Merck (Alemanya) i Sigma Chemical Co (EE.UU.). La procedència dels productes d'ús més específic s'indicarà a mesura que apareguin en el text.

M.2.- MÈTODES DE BIOLOGIA MOLECULAR

M.2.1.- Material Biològic i medis de cultiu

M.2.1.1.- Soques i vectors

Soca *E.coli* MC1061 (Casadaban *et al.*, 1980)

Genotip: hsd R2, hsd M+, hsd S+, ara D139, Δ (ara-leu) 7-697, Δ (lac)X47, gal E15, gal K16, rpsL (Strr) Mcr A, Mcr B1.

S'ha utilitzat com a hoste de vectors plasmídics.

Soca *E.coli* BL21 DE3 (Grodberg & Dunn, 1988)

Genotip: F-, *ompT*, *hdsS _{β}* (*r _{β}* -*m _{β}* -), *dcm*, *gal*, (DE3) tonA

S'ha utilitzat com a soca d'elevada expressió de proteïnes ja que presenta una deleció de les proteases *ompT* i *lon*.

Vector pGEM-T (Promega):

Vector d'aproximadament 3Kb, està dissenyat per al clonatge directe de productes de PCR generats amb la polimerasa de *T. Aquaticus*. És un vector linealitzat el qual conté timines en els extrems. S'ha usat per el clonatge i posterior seqüenciació de fragments obtinguts per PCR.

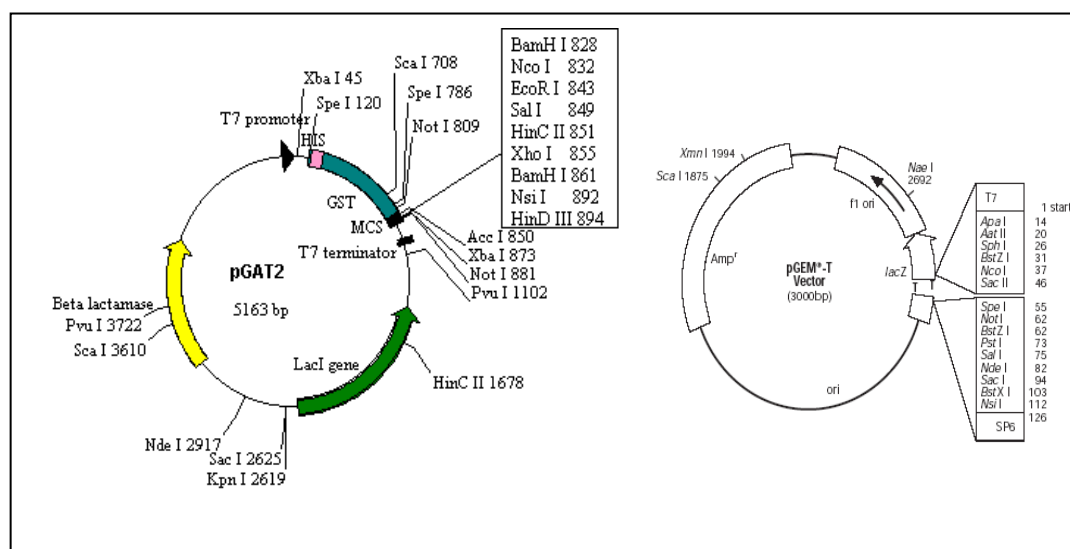
Vector pGAT2 (Peränen *et al.*, 1996):

Vector d'aproximadament 5 Kb, està dissenyat per l'expressió intracel·lular de proteïnes en *E. Coli*. És un plasmidi de baix nombre de còpies que permet clonar la proteïna d'interès sota control del promotor reconegut per la RNA polimerasa del fag T7 i de l'operador *lac* tot afegint la la proteïna expressada una cua N-terminal de 6-histidines i la proteïna de fusió Glutatió S-Transferasa (GST). La GST ajuda al

Materials i Mètodes

plegament de la proteïna recombinant que volem expressar alhora que li dóna major solubilitat. Entre la GST i l'extrem N-terminal de la proteïna que volem clonar hi ha una seqüència que codifica per la diana de reconeixement de la trombina, permetent que un cop la proteïna d'interès ha estat purificada es puguin eliminar tant la cua d'histidines com la GST (Fig. M.1).

Figura M.1. Esquema del vector pGAT2 i el pGEM-T



M.2.1.2.- Manteniment de soques bacterianes.

Per al manteniment de les soques a curt termini (un mes com a màxim) aquestes es poden plaquejar en plaques de Petri que continguin el medi adequat i guardar-les invertides a 4°C.

Per al manteniment de les soques a llarg termini es fan minicultius en glicerol que permeten conservar un cultiu durant com a mínim dos anys:

S'inocula la soca desitjada en 10 ml de medi de cultiu i s'incuba a 37°C en agitació durant 12 h. Un cop el cultiu ha crescut s'aliquota en fraccions de 850 µl en tubs estèrils i a cadascun s'hi afegeixen 150 µl de glicerol 80%, esterilitzat prèviament a l'autoclaui. Es barreja amb l'ajut del vòrtex i es congela immediatament a -80°C i a -20°C.

M.2.1.3.- Medis de cultiu

A continuació es detallen els medis de cultiu més utilitzats en aquest treball, descrits en el manual de Sambrook *et al.*, 1988.

Medi LB (Luria-Bertani)

Per 1 litre:

10 g triptona

5 g extracte de llevat

10 g NaCl

pH ajustat a 7,0 amb NaOH. Esterilització per autoclau.

Medi 2YT

Per 1 litre:

16 g triptona

10 g extracte de llevat

5 g de NaCl

pH ajustat a 7,0 amb NaOH. Esterilització per autoclau.

M.2.2.- “Polimerase chain reaction” (PCR)

La tècnica de la reacció en cadena de la polimerasa (PCR) ha estat una eina fonamental per a la biologia molecular dels darrers anys. És un sistema ràpid i senzill per amplificar DNA *in vitro*. En aquest treball hem emprat la tècnica de PCR per introduir unes dianes de restricció adequades als extrems dels cDNAs. Quan es realitza una reacció de PCR, sempre és necessari seqüenciar les construccions per comprovar que la polimerasa no ha introduït cap error al amplificar la seqüència.

En el disseny dels encebadors cal tenir en compte els següents punts:

1.- La temperatura de fusió (T_m) de l'extrem 3' de l'encebador ha de ser prou alta com per permetre una bona hibridació amb el DNA motlle. La T_m es pot calcular mitjançant la següent fórmula: $T_m = 4(G+C) + 2(A+T)$ on A, T, C i G indiquen,

respectivament, el nombre d'adenines, timines, citosines i guanines en la regió de solapament. Es considera com a prou bona una T_m més gran o igual a 50°C .

2.- S'ha de comprovar que les seqüències oligomèriques no són potencialment capaces de formar estructures secundàries ni són complementàries a altres encebadors de la mateixa reacció.

3.- La temperatura de fusió (T_m) dels dos encebadors ha de ser semblant. Una manera aproximada de calcular la temperatura d'hibridació és restar 5°C a la T_m dels encebadors.

M.2.2.1.- Amplificació del cDNA de la latexina

El cDNA de la latexina ha estat aïllat a partir d'una barreja de primeres cadenes de cDNA de cervell fetal (INVITROGEN).

Es va realitzar una estratègia de NESTED-PCR per poder obtenir el cDNA de la proteïna completa. L'objectiu és realitzar una primera reacció de PCR amb un encebador 5' que contingui la primera metionina i un encebador 3' que hibridi amb la regió del cDNA posterior a la senyal d'STOP. Si aquesta primera reacció de PCR dóna més d'un producte degut a problemes d'inespecificitat muntarem una segona reacció de PCR usant la primera reacció com a motllo a on l'encebador 5' serà el mateix que a la primera reacció però el 3' serà un nou encebador que conte la part final de la seqüència i acaba en el codó d'STOP (Fig. M.2). D'aquesta manera s'obté una major especificitat.

Els encebadors dissenyats per a pescar la latexina són:

Encebador 5': GGAATGGAAATCCCGCCGACCAACTACCC

Encebador 3' extern: GGCATCTAGCTTATGAGGCCAAATAATCCCA

Encebador 3' intern: GCCGTCTGCCAAAGGAAGTACAACTGGAATAA

La barreja de reacció que es va usar per fer la primera amplificació ha estat:

5 μl encebador 5' a concentració 10 μM

5 µl encebador 3' extern a concentració 10 µM
 8 µl DNA motlle (INVITROGEN)
 5 µl dNTPs a concentració 2 mM (Pharmacia Biotech)
 1 µl Taq polimerasa (New England Biolabs)
 2 µl MgSO₄
 5 µl tampó de la polimerasa x10
 19 µl aigua

i el programa emprat per a l'amplificació ha estat el següent:

cicle 1(X1):	1.5 min	94°C
cicle 2(X40):	1.0 min	95°C
	1.0 min	50°C
	1.0 min	72°C
cicle 3(X1):	1.0 min	95°C
	1.0 min	50°C
	10 min	72°C

Es va analitzar la PCR en un gel d'agarosa al 1.2%. No s'observava cap banda ben definida a l'alçada de 762 parells de bases, es per això és va muntar una segona reacció de PCR usant el producte de la amplificació com a motlle, l'encebador 3' més intern, una T_m més alta que en la primera PCR i es va reduir el nombre de cicles d'amplificació. La barreja de reacció que es va usar per fer la segona amplificació ha estat:

5 µl encebador 5' a concentració 10 µM
 5 µl encebador 3' intern a concentració 10 µM
 3 µl DNA motlle (producte de la reacció de PCR anterior)
 5 µl dNTPs a concentració 2 mM (Pharmacia Biotech)
 1 µl Taq polimerasa (New England Biolabs)
 2 µl MgSO₄
 5 µl tampó de la polimerasa x10
 24 µl aigua

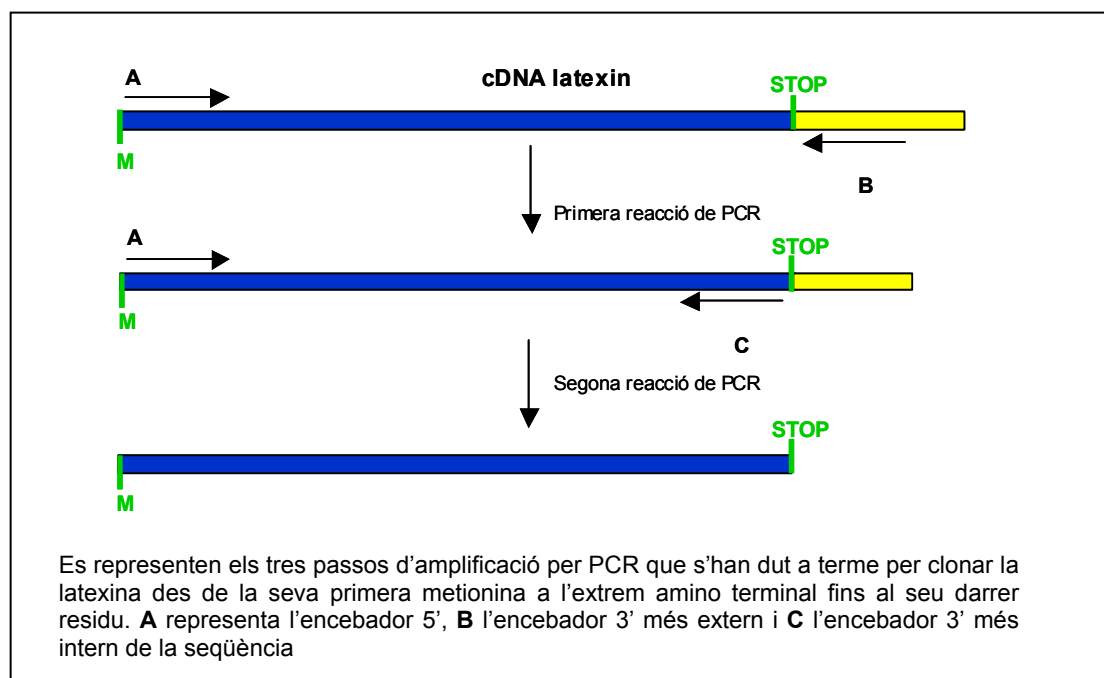
i el programa emprat per a l'amplificació ha estat el següent:

cicle 1(X1):	1.5 min	94°C
cicle 2(X30):	1.0 min	95°C
	1.0 min	60°C
	1.0 min	72°C
cicle 3(X1):	1.0 min	95°C
	1.0 min	60°C
	10 min	72°C

El producte de PCR obtingut es va clonar en el Vector pGEMT-Easy (Promega) i es va portar a seqüenciar.

Els encebadors emprats en aquest treball han estat sintetitzats per la casa Roche.

Figura M.2.- Esquema del procés de clonació de la latexina



M.2.2.2.- Introducció de dianes de restricció al cDNA de la latexina

Usant com a motlle aquest cDNA clonat en pGEMT-Easy i amb la tècnica de PCR s'han introduït als extrems de la seqüència de la latexina les dianes de tall dels enzims de restricció *NcoI* i *EcoRI* per tal de poder clonar-la en el vector d'expressió d'*Escherichia coli* pGAT2. Una part dels oligonucleòtids encebadors utilitzats en aquesta segona reacció de PCR hibriden amb el cDNA i en l'altra part es troben les dianes de restricció de *NcoI* i *EcoRI* (Fig. M.3).

La barreja de reacció que es va usar per fer les amplificacions ha estat:

- 5 µl encebador 5' a concentració 10 µM
- 5 µl encebador 3' a concentració 10 µM
- 1 µl DNA motlle
- 5 µl dNTPs a concentració 2 mM (Pharmacia Biotech)
- 1 µl High Expand polimerasa (New England Biolabs)
- 5 µl tampó de la polimerasa x10
- 28 µl aigua

i el programa emprat per a l'amplificació ha estat el següent:

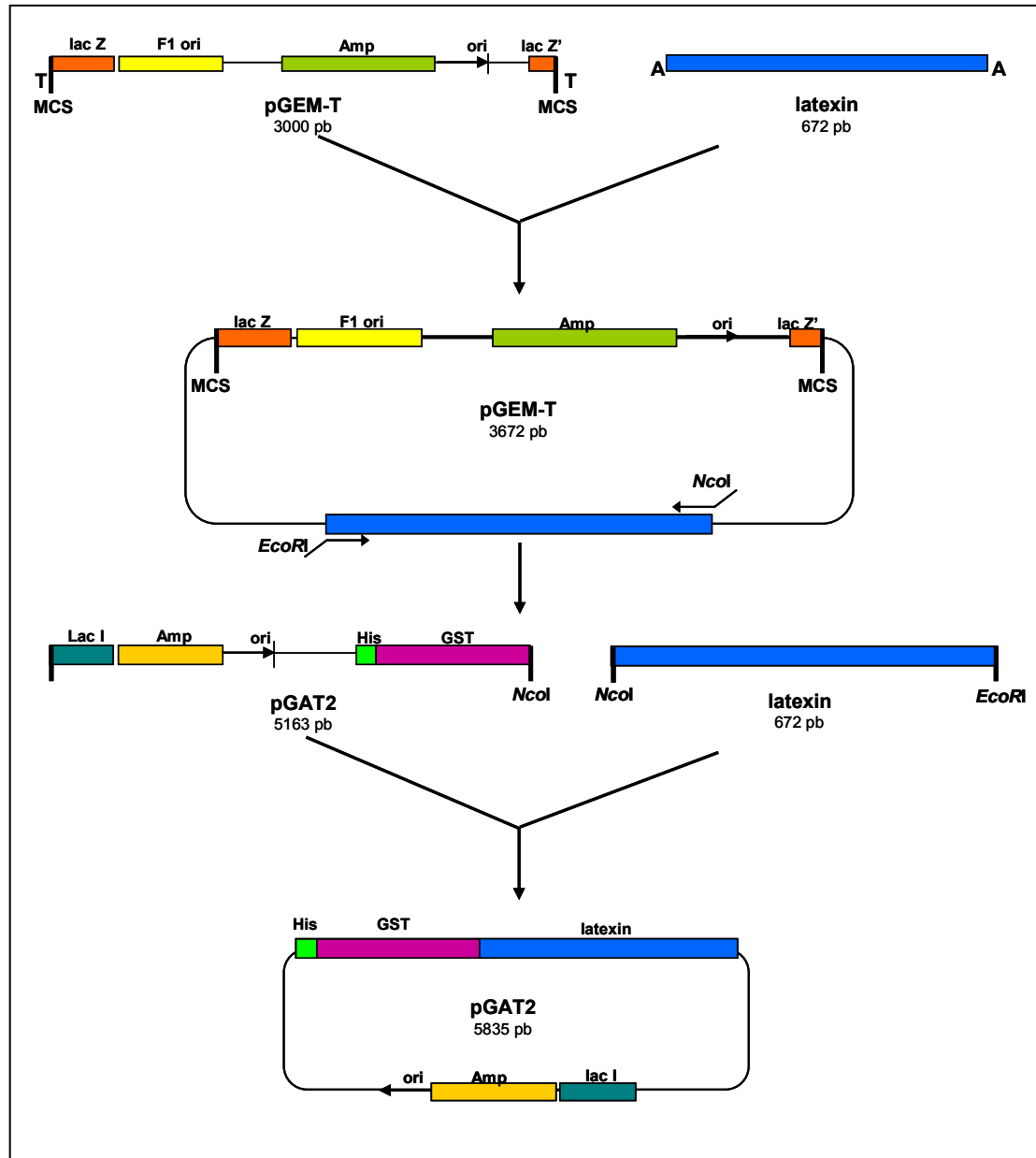
cicle 1(X1):	1.5 min	94°C
cicle 2(X30):	1.0 min	95°C
	1.0 min	55°C
	1.0 min	72°C
cicle 3(X1):	1.0 min	95°C
	1.0 min	55°C
	10 min	72°C

M.2.3.- Seqüenciació de DNA

La seqüenciació de DNA es va realitzar a Sistemas Genómicos. El mètode de seqüenciació és una variació del mètode de Sanger (Sanger *et al.*, 1977) basat en la seqüenciació cíclica. S'usen encebadors marcats amb un grup fluorescent

(ALFexpress), els quals permeten l'elongació del DNA en presència de Taq polimerasa. Es duen a terme separatament quatre reaccions amb cadascun dels quatre ddNTPs.

Figura M.3.- Esquema de la construcció dels vectors pGEM-T i pGAT2-latexina



M.2.4.- Extracció de DNA plasmídic

M.2.4.1.- Procediment de la lisi alcalina

Aquest mètode (Birnboim & Dolly 1979) degrada tot el DNA cel·lular excepte el circular. Se segueix el protocol següent:

- 1.- S'inocula una colònia de la soca desitjada en 5 mL d'LB fresc amb l'antibiòtic apropiat. S'incuba durant 12 a 15 h amb agitació a 37°C.
- 2.- Es centrifuga 1,5 mL de cultiu a 12.000 xg durant 2 min.
- 3.- S'elimina el sobrenedant i s'afegeix 1,5 mL més de cultiu i es torna a centrifugar. Opcionalment es pot tornar a repetir el mateix pas amb 1,5 mL més de cultiu. Al final s'ha de deixar el pellet de bacteris el més eixut possible.
- 4.- Es resuspèn el pellet en 100 µL de la dissolució I. Es barreja amb vòrtex. Es deixa a temperatura ambient entre 5 i 7 min.

Dissolució I de la lisi alcalina:

Glucosa 50 mM

Tris-HCl 25 mM (pH 8,0)

EDTA 10 mM (pH 8,0)

- 5.- S'afegeix 200 µL de la dissolució II acabada de preparar. Es barreja per inversió, no amb vòrtex. Es deixa en gel durant 5 min.

Dissolució II de la lisi alcalina:

NaOH 0,2 N

SDS 1%

- 6.- S'afegeix 150 µL de la dissolució III temperada prèviament en gel. Es barreja amb el vòrtex. Es deixa en gel durant 5 min.

Dissolució III de la lisi alcalina (per a 100 mL):

Acetat potàssic 60 mL

Àcid acètic glacial 11,5 mL

Materials i Mètodes

Aigua destil·lada Milli-Q 28,5 mL

Aquesta dissolució és 3 M en potassi i 5 M en l'acetat.

7.- Es centrifuga a 12.000 xg durant 5 min. Es transfereix el sobrenedant a un nou tub.

8.- Extracció amb fenol/cloroform.

9.- Precipitació del DNA plasmídic amb 2,5 volums d'etanol.

10.- Es resuspèn el precipitat de DNA en 10 a 50 µL de TE o aigua estèril.

M.2.4.2.- Extracció amb *kit* comercial

S'ha utilitzat el *kit* d'extracció plasmídica de la casa comercial Pharmacia Biotech, basat en el mètode de la lisi alcalina. Aquest *kit* proporciona un bon rendiment i deixa el DNA net i lliure d'RNA.

M.2.5.- Determinació de la concentració de DNA

M.2.5.1.- Determinació espectrofotomètrica

Consisteix en enregistrar l'espectre d'absorció 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 DNA de cadena senzilla.

La puresa del DNA es pot valorar amb el coeficient A₂₆₀/A₂₈₀, que ha de ser proper a 1,8. Valors inferiors indiquen contaminació per proteïnes, i valors superiors, contaminació per RNA.

M.2.5.2.- 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 bromur d'etidi, el qual s'intercala entre les bases nitrogenades.

Es realitza una electroforesi en agarosa i es visualitza il·luminant 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.

M.2.6.- Digestions enzimàtiques d'àcids nucleics

Les endonucleases de restricció són enzims que tallen per seqüències específiques de DNA. Són les eines bàsiques de la biologia molecular. L'activitat dels enzims de restricció depèn del pH, de la força iònica i de la temperatura de la reacció. Cada enzim té unes condicions òptimes de funcionament.

Normalment, la quantitat de DNA assajada en les digestions oscil·la entre 0,5 i 10 µg i s'afegeixen d'1 a 10 unitats d'enzim per µg de DNA. Cal tenir en compte que la quantitat d'enzim no pot superar mai el 10% del volum total de la reacció, ja que els enzims de restricció estan dissolts en glicerol al 50%, i una concentració de glicerol en el medi de reacció superior al 5% pot fer perdre l'especificitat de reconeixement. La temperatura de reacció varia per a cada enzim. El temps de reacció és de 1-2 h. Si s'utilitzen dos enzims alhora, es pot augmentar el temps fins a 3 o 4 h.

Els enzims de restricció emprats en aquest treball han estat subministrats per Roche Molecular Biochemicals o New England Biolabs.

M.2.7.- Separació del DNA mitjançant gels d'agarosa

M.2.7.1.- Electroforesi de DNA

L'electroforesi horitzontal en gels d'agarosa s'ha utilitzat com a mètode habitual d'anàlisi del DNA. L'entramat molecular del gel d'agarosa és el principal factor de la separació, d'aquesta manera les molècules de mida més petita avancen més ràpidament a través del gel.

Materials i Mètodes

Existeixen diferents tipus d'agarosa i es poden utilitzar a diferents concentracions. Per realitzar aquest treball s'ha utilitzat agarosa tipus I-A. La concentració d'agarosa utilitzada varia segons la grandària de les molècules que cal separar:

<u>% Agarosa</u>	<u>Rang de separació (Kb)</u>
0,7	0,8-10
0,9	0,5-7
1,2	0,4-6
2,0	0,1-2

Com a tampó de recorregut s'utilitza:

Dissolució TAE 10X (per a 1 litre):

Tris	48,40 g
Àcid acètic glacial	11,42 mL
EDTA 0,2 M (pH 8,0)	55,00 mL

a) Preparació del gel

- 1.- Es pesa l'agarosa i s'introdueix en un erlenmeier. Per a un minigel s'afegeix 50 mL de tampó d'electroforesi (TAE o TBE) i per a un gel gran s'afegeix 200 mL.
- 2.- Es fon l'agarosa escalfant l'erlenmeier fins a ebullició. S'agita durant el procés per a obtenir una dissolució homogènia.
- 3.- Es deixa refredar i s'afegeix bromur d'etidi fins a una concentració final de 0,5 µg/mL.
- 4.- Quan l'agarosa està a una temperatura inferior a 50°C, es vessa dins el motlle del gel que duu una pinta per formar les butxaques. Cal evitar la formació de bombolles.
- 5.- Es deixa solidificar durant 20-30 min.
- 6.- Un cop solidificat es col·loca el gel dins la cubeta d'electroforesi i es recobreix amb el tampó d'electroforesi fins 2-3 mm per sobre del gel.

b) Electroforesi:

- 1.- S'afegeix a les mostres el tampó de càrrega.

Tampó de càrrega 10X:

Glicerol 50%

Colorant Orange-G (punta d'espàtula)

2.- Es carreguen de 10 a 20 μL de mostra per butxaca. En una d'elles es carrega també 1 μg de marcadors de pes molecular. En aquest treball s'ha utilitzat com a marcador el *1 Kb DNA ladder* de Life Technologies.

3.- Es connecta l'electroforesi a 70 – 90 volts.

c) Tinció del gel:

Les bandes es visualitzen per transil·luminació dels gels amb llum ultraviolada, ja que el bromur d'etidi intercalat entre les bases del DNA emet fluorescència.

d) Imatges dels gels:

Les imatges dels gels d'agarosa s'han obtingut amb un aparell *Gel Doc 1000* de BioRad.

M.2.7.2.- Purificació de bandes de DNA de gels d'agarosa

Per a l'extracció del DNA de l'agarosa s'ha utilitzat un *kit* comercial de Pharmacia Biotech. El protocol d'utilització és el següent:

1.- Es retalla la banda del gel d'agarosa i s'afegeix 1 μL de tampó caotrópic subministrat pel *kit* per cada mil·ligram de banda.

2.- S'incuba a 60°C fins a fondre l'agarosa (5-10 min).

3.- S'incuba la mostra durant 1 minut en una columneta que conté una reïna que reté el DNA.

4.- Es centrifuga 13.000 rpm 30 s. Es llença l'eluit i es posa 500 μL de tampó de rentat a la columna i es torna a centrifugar 30 s.

5.- El DNA purificat s'elueix amb 50 μL d'aigua Milli-Q estèril. S'incuba la mostra amb l'aigua durant 1 min, i es centrifuga a 13.000 rpm 1 min.

M.2.8.- Clonatge de fragments de DNA en plàsmids

En aquest treball s'ha realitzat només clonatge d'extrems cohesius. El procediment comprèn els següents passos:

- 1.- Es restringeixen el vector i l'insert amb els enzims de restricció adequats.
- 2.- Es realitza una electroforesi en gel d'agarosa, es tallen les bandes que interessin i es purifica el DNA.
- 3.- Reacció de lligació: ens permet unir fragments de DNA amb extrems compatibles. L'enzim usat per a dur a terme aquesta reacció és la T4 DNA lligasa (Promega o Invitrogen). Les quantitats de DNA i les condicions de la lligació varien en funció de que els extrems siguin roms o cohesius. Per a la lligació d'extrems cohesius compatibles cal que el quocient entre el nombre de molècules de vector i el nombre de molècules d'insert sigui proper a 0,5, és a dir, 2 molècules de fragment per cada molècula de vector. Es barregen vector linealitzat i insert en la proporció indicada anteriorment, una unitat de lligasa, 1 µL del tampó de la lligasa 10X (composició: Tris-HCl 300 mM, MgCl₂ 100 mM, DTT 100 mM, ATP 10 mM (pH 7,8)) i aigua Milli-Q estèril fins a un volum final de 10 µL. La reacció es pot dur a terme durant tota la nit a 4°C o 3 hores a temperatura ambient.

M.2.9.- Incorporació de DNA forani en bacteris

M.2.9.1.- Preparació de cèl·lules competents d'*E. coli*

La introducció de vectors plasmídics en cèl·lules d'*E. coli* es realitza mitjançant la transformació de cèl·lules competents d'*E. coli*. Per a la preparació de cèl·lules competents s'ha usat el mètode del clorur càlcic (Mandel & Higa, 1970), que inclou els següents passos:

- 1.- S'inocula 10 mL d'LB fresc amb una colònia o 50 µL de glicerinat i s'incuba durant 12-15 h a 37°C amb agitació a 300 rpm.

2.- A partir d'aquest cultiu saturat es prepara un cultiu en creixement exponencial diluint-lo 1:100 en LB i incubat a 37°C amb agitació fins que UDO_{550} del cultiu estigui entre 0,4 i 0,5 (sol tardar 2 h).

3.- Es transfereix el cultiu a un tub de polipropilè estèril de 50 mL i es refreda el cultiu en gel durant 15 minuts com a mínim.

4.- Es centrifuga a 5000 rpm durant 10 minuts a 4°C. Es decanta el sobrenedant, i s'asseca el precipitat per inversió durant 10 minuts. Es resuspèn el *pellet* en V/2 de $CaCl_2$ 100 mM temperat prèviament en gel. Es deixa reposar en gel durant 1 hora.

5.- Es centrifuga a 5000 rpm durant 10 minuts a 4°C. Es decanta el sobrenedant i s'asseca el precipitat per inversió durant 10 minuts. Es resuspèn suaument en V/20 de $CaCl_2$ 100 mM temperat prèviament en gel. Es deixa reposar en gel durant 1 hora. Al cap d'aquest temps ja tenim preparades les cèl·lules competents, que es poden aliquotar.

El $CaCl_2$ 100 mM s'esterilitza per filtració. Les alíquotes de cèl·lules competents es poden congelar a -80°C si prèviament s'afegeix glicerol a una concentració final del 15%.

M.2.9.2.- Transformació de cèl·lules competents d'*E. coli*

Les cèl·lules competents poden ser transformades amb DNA que procedeix d'una lligació o d'una minipreparació plasmídica. El protocol que s'ha seguit és el següent:

1.- Es barreja 100 μ L de cèl·lules competents amb 50 ng de DNA de la barreja de lligació. Per comprovar l'eficiència del procés es fan dos controls. Un control negatiu on només s'incuben cèl·lules competents sense DNA, i un control positiu on es transformen cèl·lules competents amb un plasmidi circular que presenta resistència a ampicil·lina però sense cap insert.

2.- S'incuba en gel durant 30 min.

3.- Es transfereixen les mostres a un bany a 42°C durant 2 min i després 5 min en gel.

4.- Opcionalment es pot afegir 1 mL de LB i s'incuba a 37°C amb agitació suau (100 rpm) durant 1 hora.

5.- Es fa un pols de centrifuga per sedimentar les cèl·lules (30 s a 8000 rpm) s'extreu medi suficient com perquè quedin 200 μ L.

- 6.- Es plaqueja amb una nansa de Digrafsky i es deixa reposar la placa durant 10 min.
- 7.- S'inverteix la placa i s'incuba a 37°C durant 14 h.
- 8.- Selecció de transformants per resistència a ampicil·lina.

M.2.10.- Expressió de proteïna recombinant

M.2.10.1.- Producció a nivell analític

Per tal de seleccionar les millors condicions d'expressió es preparen minicultius i s'analitza la producció de proteïna mitjançant SDS-PAGE.

- 1.- Es parteix d'una colònia aïllada i s'inocula 15 mL de 2YT amb ampicil·lina en un tub de 50 mL. Es fa créixer el cultiu a 37°C amb una agitació de 250-300 rpm durant 12 h.
- 3.- Es preparen tants tubs de 50 mL com condicions diferents es volen analitzar amb 3 mL de 2YT amb ampicil·lina i s'inocula en cada un d'ells amb 6 µl del cultiu crescut.
- 4.- Un cop inoculat, les variables que podem analitzar, entre d'altres, són: la UDO_{550} a la que induïrem el cultiu, la temperatura a la que creixerem les cèl·lules o les hores que deixarem que tingui lloc la inducció.
- 5.- Un cop transcorregut el temps òptim d'inducció es pren una alíquota del cultiu i es centrifuga a 12.000 xg durant 5 min i s'elimina el sobrenedant.
- 6.- pGAT2 és un vector que té un promotor fort, i en aquests casos a vegades part de la proteïna recombinant no es troba en forma soluble sinó que la trobem en forma de cossos d'inclusió. És per això que es recomana analitzar per a cadascuna de les condicions analitzades quina part de la proteïna es troba en forma soluble i quina formant part dels cossos d'inclusió. El protocol seguit per a la separació de la fracció soluble de la insoluble és el següent:
 - a) Es resuspèn el pellet en 50 µL de tampó PBS X1.
 - b) Es congela i descongela 4 vegades.
 - c) S'afegeix lisozim a una concentració final d'1mg/mL i es deixa incubant en gel 30 min.

- d) Un cop incubat s'afegeix Tritó X100 a una concentració final d'1,1% i s' agita 30 min a 4°C.
- e) S'afegeix RNAsa a la concentració final de 20 µg/mL, DNAsa I a 10 µg/mL i es deixa en agitació durant 10 min a temperatura ambient.
- f) Es centrifuga a 17.000 rpm durant 30 min. Per separar el sobrenedant, que és la fracció de proteïnes soluble del medi intracel·lular d'*E. coli*. Li afegim tampó de càrrega per realitzar l'electroforesi en SDS.
- g) El segon pellet correspon a la fracció insoluble del medi intracel·lular d'*E. coli*, el resuspenem en el mateix volum que ens ha quedat de fracció soluble i afegim també el tampó de càrrega per realitzar l'electroforesi en SDS.

7.- La proteïna recombinant es pot visualitzar per tinció amb blau de coomassie.

M.2.10.2.- Producció preparativa de proteïna recombinant

Un cop s'ha determinat quines condicions ofereixen una major producció de proteïna recombinant soluble es procedeix a una producció preparativa.

- 1.- Es parteix d'una colònia aïllada i s'inocula 15 mL de 2YT amb ampicil·lina en un tub de 50 mL. Es fa créixer el cultiu a 37°C amb una agitació de 250-300 rpm durant 12 h.
- 2.- A partir d'aquest cultiu saturat es prepara un cultiu en creixement exponencial diluint-lo 1:500 en 1 L de 2YT i incubat a 37°C amb agitació fins que UDO_{550} del cultiu estigui entre 0,4 i 0,5. Es baixa la temperatura d'incubació fins a 18°C i es deixa amb agitació fins que la UDO_{550} del cultiu estigui entre 0,7 i 0,8.
- 3.- S'indueix el cultiu amb IPTG a una concentració final d'1 mM i es deixa amb agitació a 20°C durant 24 h.
- 4.- Un cop acabada la inducció es centrifuga el cultiu, es recullen les cèl·lules, es processa el pellet com anteriorment i es recull la fracció soluble, la qual es sotmet seguidament al procés de purificació.

M.3.- MÈTODES D'ANÀLISI DE PROTEÏNES

M.3.1.- Electroforesi discontinua en presència d' SDS

L'electroforesi en gels de poliacrilamida en presència de dodecilsulfat sòdic (SDS-PAGE) és una tècnica molt emprada per la seva gran resolució. En aquest sistema electroforètic les proteïnes migren a una velocitat que està en funció inversa del logaritme de la seva massa molecular. L' SDS és un detergent iònic que desplega les proteïnes i les cobreix amb un núvol de càrregues negatives. Això té un efecte d'apantallament de la càrrega intrínseca de les proteïnes (Shapiro *et al.*, 1967). L' SDS s'uneix a les proteïnes proporcionalment a la seva massa i per tant la seva separació en el gel de poliacrilamida es farà segons aquest criteri.

L'electroforesi discontinua descrita per Laemmli (1970) augmenta la resolució perquè els complexos SDS-proteïna corren a través de dos gels de diferent densitat: el gel superior o concentrador, que és prou lax com per permetre la formació d'un front homogeni i el gel inferior o separador, que és més dens en el qual cada proteïna té una velocitat diferent en funció de la seva massa molecular.

a) Preparació del gel:

Es fan servir les següents dissolucions:

Dissolució A: dissolució al 40% d'acrilamida/bisacrilamida, 37, 5:1

Dissolució B: Tris-HCl, 1,5 M, SDS 0,4% (pH 8,8)

Dissolució C: Tris-HCl, 0,46 M, SDS 0,4% (pH 6,8)

Composició del gel separador (12% d'acrilamida):

Dissolució A:	1,5 mL
Dissolució B:	1,2 mL
Aigua Milli-Q:	2,25 mL
Persulfat amònic 10%:	50 µL
TEMED:	10 µL

Composició del gel concentrador (2,4% d'acrilamida):

Dissolució A:	0,3 mL
Dissolució C:	0,5 mL
Aigua Milli-Q:	1,2 mL
Persulfat amònic 10%:	25 µL
TEMED:	2,5 µL

b) Electroforesi:

1.- Es prepara la cubeta d'electroforesi i s'omplen els compartiments superior i inferior amb el tampó de recorregut diluït amb aigua Milli-Q.

Tampó de recorregut 10X:

Tris 0,25 M

Glicina 1,87 M

SDS 1%

S'ajusta el pH a 8,4 amb HCl

2.- Es barregen les mostres amb el tampó de d'aplicació en una relació mostra/tampó de 3:1. El tampó de mostres es prepara de forma concentrada amb β -mercaptoetanol, un agent reductor per trencar els ponts disulfur.

Tampó d'aplicació 3X:

Tris 180 mM

Glicerol 30%

SDS 9%

β -mercaptoetanol 15%

Blau de Bromofenol 0,05%

S'ajusta el pH a 6,8 amb HCl

3.- Es posen les mostres durant 2 o 3 minuts a 100°C i s'apliquen a la butxaca corresponent. En un carril paral·lel s'aplica també una barreja de proteïnes marcadores de pes molecular. D'aquesta manera la massa molecular relativa (Mr) de les proteïnes analitzades pot ser calculada per interpolació en una corba

Materials i Mètodes

construïda amb aquesta barreja de proteïnes marcadores de pes molecular conegut. Habitualment s'ha usat una barreja comercial de la casa BioRad formada per:

	Mr (kDa)
Miosina	200,00
β -galactosidasa	116,25
Fosforilasa b	94,00
Albúmina sèrica bovina	67,00
Ovoalbúmina	45,00
Anhidrasa carbònica	31,00
STI	21,50
Lisozima	14,40
Aprotinina	6,50

4.- Es connecta l'electroforesi a un amperatge constant de 20 mA, fins que surti el colorant del tampó de càrrega.

5.- Es desmunta el *Kit* i es procedeix a la tinció.

c) Tinció amb blau de Coomassie:

Té un límit de detecció de 0,3 a 1 μ g per banda d'electroforesi. El protocol és el següent:

1.- Un cop acabada l'electroforesi, els gels es submergeixen directament dins la dissolució colorant. El temps de tinció depèn del gruix del gel, però per minigels de 0,75 mm n'hi ha prou amb 15-30 minuts en agitació suau a temperatura ambient. En aquest treball s'ha utilitzat el *Phast Gel Blue R* de Pharmacia per preparar la dissolució colorant.

Dissolució colorant:

Metanol 30%

Àcid acètic 8%

Blau de Coomassie 0,15%

2.- Passat aquest temps es destenyeix amb diversos canvis de la dissolució decolorant fins que quedi contrastat, amb el fons blanc i les bandes de proteïna en blau.

Dissolució Decolorant:

Metanol 25%

Àcid Acètic 8%

d) Imatges dels gels:

Les imatges dels gels d'electroforesi s'han obtingut per escanejat amb l'aparell *Model GS-700 Imaging Densitometer* de BioRad.

En aquest treball en la part corresponent a l'oligomerització de la latexina el tampó de càrrega que emprat no contenia ni SDS ni β -mercaptoetanol així com tampoc es van bullir les mostres.

M.3.2.- Electrotransferència

L'Electrotransferència permet transferir les proteïnes separades en un gel d'electroforesi a una membrana de material sintètic (p.ex. nitrocel·lulosa o PVDF) gràcies a l'aplicació d'un camp elèctric perpendicular al gel. El protocol que es segueix és el següent:

- 1.- Realitzar l'electroforesi segons l'apartat M.3.1.
- 2.- Submergir el gel d'electroforesi en el tampó de transferència durant 15 minuts.

Tampó de transferència:

Glicina 39 mM

Tris 48 mM

SDS 0,04%

Metanol 20%

S'ajusta a pH 9,0

Materials i Mètodes

3.- Les membranes usades per a les transferències són de PVDF, de la casa Millipore. S'han de manipular amb guants i amb molta cura per evitar que es trenquin. La membrana es talla a la mateixa mida que el gel, es tracta durant quinze segons amb metanol, es renta durant dos minuts en aigua Milli-Q i s'equilibra en tampó de transferència durant cinc minuts.

4.- Es tallen dos trossos de paper Whatman (Schleicher & Schuell) a la mida del gel.

5.- Es remullen els fragments de paper Whatman i les esponges en tampó de transferència i en munta el sistema segons el següent ordre a partir del pol negatiu (costat negre del sistema):

Esponja
Paper Whatman
Gel de poliacrilamida
Membrana
Paper Whatman
Esponja

6.- El sistema es col·loca dins una cubeta de mini-trans-blot de BioRad i s'omple amb tampó de transferència.

7.- La transferència es realitza a amperatge constant a 4°C.

8.- Es tenyeix la membrana amb blau de coomassie R-250 Brilliant Blue (BioRad).

Dissolució de Tinció:

Metanol 50%
Blau de Coomassie 0,1%

9.- La destinció es realitza amb una dissolució de metanol 50% i àcid acètic 10%.

M.3.3.- Seqüenciació N-terminal de proteïnes.

La determinació de la seqüència N-terminal de proteïnes s'ha portat a terme en el Servei de Seqüenciació de Proteïnes de l'Institut de Biotecnologia i Biomedicina (IBB) de la Universitat Autònoma de Barcelona. L'equip de seqüenciació utilitzat és un Applied Biosystems model Procise 492 Protein Peptide Sequencer (USA).

Els seqüenciadors automàtics utilitzen el mètode de la degradació d'Edman amb fenilisotiocianat (PITC). El mètode es pot resumir així: el PICT s'acoba en primer lloc al grup amino terminal de la proteïna; seguidament l'enllaç N-terminal es trenca en un medi àcid anhidre, i s'allibera una anilina tiazolina (ATZ-aminoàcid); aquesta és transformada en un medi àcid aquós en una feniltiohidantoïna (PTH-aminoàcid) estable i capaç de ser identificada espectrofotomètricament.

Quan la proteïna a seqüenciar no es troba en estat pur s'ha realitzat una electroforesi per separar la banda de proteïna d'interès i posteriorment s'ha transferit a una membrana de PVDF (Millipore) pel mètode descrit en l'apartat M.3.2. La banda d'interès s'ha retallat de la membrana i s'ha seqüenciat. Les bandes es poden conservar guardant-les a -20°C .

M.3.4.- Mètodes espectroscòpics

M.3.4.1.- Mètodes espectrofotomètrics

M.3.4.1.1.- Quantificació espectrofotomètrica de proteïnes

Per calcular la concentració de proteïna s'aplica la llei de Lambert-Beer a una longitud d'ona de 280 nm:

$$C = DO_{280} / \epsilon_{280} \cdot d$$

On C equival a la concentració de proteïna, DO_{280} és l'absorbància de la mostra problema a 280 nm, ϵ_{280} el coeficient d'extinció molar ($\text{M}^{-1} \text{cm}^{-1}$) per a la proteïna d'interès a la mateixa longitud d'ona i d l'amplada del pas de llum en cm.

M.3.4.1.2.- Determinació espectrofotomètrica d'activitats enzimàtiques

Mesura de l'activitat carboxipeptidasa A:

Substrat:

N-furilacrilòil-L-fenilalanina- L-fenilalanina (FAPP) (Peterson *et al.*, 1982)

Condicions:

Concentració de substrat: 0,2 mM

Tampó: Tris-HCl 50 mM, NaCl 0,45 M (pH 7,5)

Longitud d'ona: 330 nm (se segueix la disminució)

Substrat:

N-(4-Methoxyphenylazoformyl)-L-fenilalanina (Mock *et al.*, 1996)

Condicions:

Concentració de substrat: 0,1 mM

Tampó: Tris-HCl 50 mM, NaCl 0,1 M (pH 7,5)

Longitud d'ona: 350 nm (se segueix la disminució)

Mesura de l'activitat carboxipeptidasa B:

Substrat:

N-furilacrilòil-L-alanil-L-lisina (FAAK) (Plummer & Kimmel, 1980)

Condicions:

Concentració de substrat: 0,2 mM

Tampó: Tris-HCl 20 mM, NaCl 0,45 M (pH 7,5)

Longitud d'ona: 330 nm (se segueix la disminució)

Substrat:

N-(4-Methoxyphenylazoformyl)-L-arginina

Condicions:

Concentració de substrat: 0,1 mM

Tampó: Tris-HCl 50 mM, NaCl 0,1 M (pH 7,5)

Longitud d'ona: 350 nm (se segueix la disminució)

Mesura de l'activitat carboxipeptidasa N:

Substrat:

N-furilacrilòil-L-alanil-L-lisina (FAAK) (Plummer & Kimmel, 1980)

Condicions:

Concentració de substrat: 0,35 mM

Tampó: Tris-HCl 50 mM, NaCl 0,1 M (pH 7,5)

Longitud d'ona: 340 nm (se segueix la disminució)

Mesura de l'activitat papaïna:

Substrat:

N-Benzoïl-L-arginina-p-nitroanilida (BAPA) (Gaertner & Puigserver, 1992)

Condicions:

Concentració de substrat: 1 mM

Tampó: NaAc 300 mM, EDTA 4 mM, Brij 0,05%, DTT 1 mM (pH 5,5)

Longitud d'ona: 404 nm (se segueix l'augment)

Normalment, s'afegeixen entre 1 i 20 µL de mostra a 1 mL de tampó de substrat i les mesures es realitzen durant 2 a 10 minuts en un espectrofotòmetre Cary 100 Bio o Cary 400 Bio a una temperatura de 25°C. L'activitat enzimàtica es dedueix a partir del pendent del registre de l'absorbància de la dissolució enzim-substrat al llarg del temps, i s'expressa com la quantitat de substrat consumit (o de producte format) per unitat de temps. Per poder estandaritzar els assajos, es realitza una hidròlisi total del substrat amb excés d'enzim, per mesurar l'increment absolut en densitat òptica d'una dissolució de substrat (Aht). El valor obtingut ha de ser igual o pròxim al teòric, que per al benzoïl-glicil-L-arginina i el benzoïl-alanil-L-glutàmic és de 0,36 UDO/mM i pel FAPP és de 2 UDO/mM. La forma més adequada d'expressar l'activitat és referint-la a la quantitat d'enzim present a la mostra. Aquesta expressió es denomina activitat específica (Ae):

$$Ae = \frac{\text{UDO / min}}{\text{Aht x mg enzim/mL substrat}}$$

D'aquesta manera l'activitat específica queda expressada en:

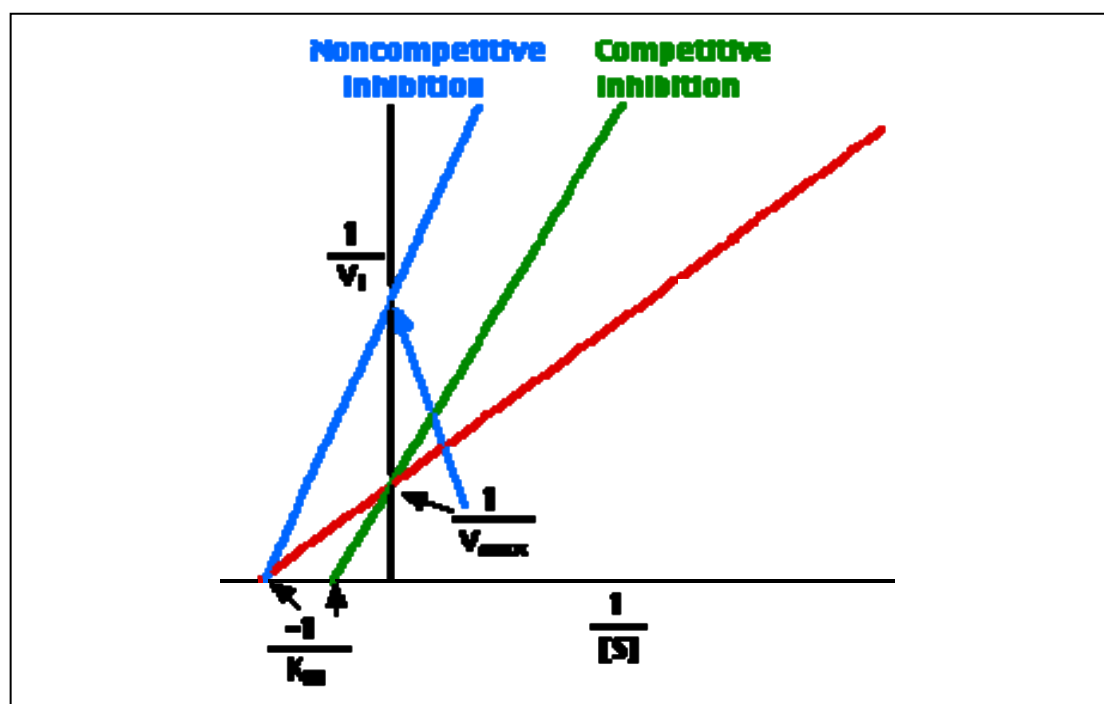
$$\frac{\text{UDO/ min}}{\text{UDO/mmol/L x mg/mL}}$$

és a dir: $\mu\text{mols de substrat/mg enzim x min}$

M.3.4.1.3.- Determinació del tipus d'inhibició

Per a determinar quin tipus d'inhibició presenta un determinat inhibidor amb els seus enzims diana es realitza un gràfic de Lineweaver-burk. La quantitat d'enzim necessària i el rang de concentracions de substrat al que es realitzen les mesures varien per a cada substrat. En aquesta treball s'han fet mesures de 5 concentracions de substrat diferent i de 6 concentracions d'inhibidor diferent per a cadascuna d'elles. Els valors de pendent obtinguts per a cada una de les corbes s'han representat en un gràfic de dobles recíprocs (Fig. M.4.).

Figura M.4. Gràfic Lineweaver-Burk



M.3.4.1.4.- Determinació de la constant de dissociació (K_i) d'inhibidors proteics per a carboxipeptidases

Per a calcular la constant de dissociació (K_i) es va usar el mètode de la cinètica de l'estat pre-estacionari ($K_i = k_{off}/k_{on}$) per a inhibidors proteics d'unió forta (tight-binding inhibitors) descrit per Morrison (1982). En aquest assaig, l'inhibidor, a una concentració prèviament establerta després de proves variant les quantitats d'enzim i inhibidor, és afegit a la cubeta mentre la corba progressiva que segueix la generació de producte d'una barreja enzim substrat és monitoritzada continuament. A partir d'aquesta corba i mitjançant un procediment matemàtic es poden derivar els valors de les constants k_{on} , k_{off} i en conseqüència la K_i . Per confirmar aquest valors de K_i obtinguts, es van realitzar assajos d'inhibició amb el mètode de l'estat estacionari per alguns dels enzims caracteritzats anteriorment amb el mètode descrit per Morrison, mesurant l'activitat a varies proporcions enzim/inhibidor i a diferents concentracions de substrat.

M.3.4.1.5.- Quantificació espectrofotomètrica de l'agregació de proteïnes

Les mesures de la turbidesa a 350 nm al llarg del temps indiquen agregació, quanta major és la dispersió més agregació presenta la mostra. En el present treball la tendència a agregar de les mostres de proteïna incubada a diferents concentracions de TFE va ser monitoritzada per la mesura de la seva turbidesa a 350 nm al llarg del temps.

M.3.4.1.6.- Quantificació espectrofotomètrica de la unió de congo red

El congo red és un colorant que s'uneix específicament a les estructures amiloides. L'anàlisi de les mostres s'ha realitzat mitjançant l'assaig espectrofotomètric del desplaçament de la banda descrit per Klunk *et al.*, 1989: s'afegeix entre 10 i 100 μ L de mostra a 1 mL de dissolució de reacció (ha de ser preparada fresca i protegida de la llum) i se n'obté l'espectre al visible.

Dissolució de Reacció:

5 mM fosfat sòdic

150 mM NaCl

5 μ M congo red
S'ajusta a pH 7,0

El congo red sol presenta el seu màxim d'absorbància a 490 nm, que augmenta i es desplaça cap al vermell quan s'uneix a estructures amiloides.

M.3.4.2.- Espectrometria de masses

L'espectrometria de masses permet determinar en pocs minuts la massa molecular relativa de qualsevol pèptid o proteïna petita amb un error menor al 0,1%. En aquest treball s'ha utilitzat un sistema MALDI-TOF (Matrix-Assisted Laser Desorption Ionization-Time of Fly) (Hillenkamp *et al.*, 1991) model Bruker Ultraflex (Alemanya). El sistema MALDI es caracteritza per la formació d'una espècie protonada gràcies a que la matriu cedeix un protó al ser impactada pel làser, generant-se un nombre d'estats de ionització menor que en altres sistemes, i per tant, espectres més fàcils d'analitzar. La mostra a analitzar (1 μ L) es barreja a parts iguals amb una matriu orgànica 1 μ L, 1 μ L d'aquesta barreja es col·loca sobre una placa de metall i es deixa assecar durant 5 minuts. La matriu que s'utilitza depèn de la naturalesa de la mostra que volem analitzar. Un cop s'ha assecat la barreja, s'introdueix la placa dins de l'aparell i s'optimitzen els paràmetres energètics de les mesures.

M.3.4.3.- Dicroisme circular

Aquest és un mètode d'anàlisi estructural de pèptids i proteïnes basat en les propietats quirals d'algunes molècules o de les seves ordenacions regulars internes per desviar de manera diferencial la llum polaritzada. Les estructures secundàries de les proteïnes posseeixen patrons de dicroisme circular característics, permetent així la seva identificació i el seguiment de processos de desnaturalització i renaturalització de proteïnes. Els aparells utilitzats van ser el JASCO J-715 del Servei d'Anàlisi Química de la Universitat Autònoma de Barcelona; l'aparell JASCO J-810 dels Serveis Científicotècnics de la Universitat de Barcelona.

M.3.4.4.- Ressonància Magnètica Nuclear Monodimensional

Aquesta tècnica experimental s'utilitza per la resolució d'estructures tridimensionals de proteïnes en dissolució. En el nostre cas sols es van fer espectres monodimensionals, que donen una primera referència sobre la dispersió de senyals de l'espectre de proteïnes i permeten fer seguiments de la seva desnaturalització tèrmica.

Els espectres de ressonància magnètica nuclear de ^1H a 500 MHz de la latexina es varen obtenir amb un espectròmetre BRUKER ARX 500 narrow-bore equipat amb un sistema de gradients de camp magnètic del Servei de Ressonància Magnètica Nuclear de la Universitat Autònoma de Barcelona.

Per al seguiment de la desnaturalització tèrmica de la latexina es van dissoldre 1.2 mg en 0.6 ml de tampó fosfat 20mM, 10% D_2O (pH 7,5). Es va ajustar l'homogeneïtat del camp magnètic i es van acumular 260 espectres a 27, 67, 62, 57, 52 i 22°C. La mostra no es va treure del magnet en tot el procés.

M.3.4.5- Fluorimetria

M.3.4.5.1.- Determinació de la fluorescència intrínseca

Les proteïnes contenen tres aminoàcids aromàtics (Trp, Tyr, Phe) els quals poden contribuir a la seva fluorescència intrínseca. La fluorescència d'una proteïna plegada és una barreja de la fluorescència de cada residu aromàtic individual. La longitud d'ona d'excitació es fixa generalment a 280 nm, i la major part de les emissions són degudes a l'excitació dels residus triptòfan. Canvis en la fluorescència intrínseca s'usen per seguir canvis en l'estructura de la proteïna. En aquest treball s'ha seguit la fluorescència intrínseca de la α -quimotripsina i de la latexina. En el primer cas s'han seguit els canvis de la fluorescència intrínseca a diferents % de TFE (trifluoroetanol), i en el cas de la latexina s'ha seguit els canvis estructurals per fluorescència a diferents concentracions d'urea. L'aparell emprat va ser un Perkin Elmer 650-40. En ambdós casos la concentració de proteïna era de 100 $\mu\text{g}/\text{mL}$ i es va usar una longitud d'ona d'excitació de 280 nm i l'emissió es va recollir entre 300 i 400 nm.

M.3.4.5.2.- Determinació de la unió de tioflavina-T (Th-T)

La Th-T és un colorant que s'usa àmpliament per la caracterització de la presència d'estructures amiloides. En presència d'amiloides la Th-T i en ser excitada a una longitud d'ona de 440 nm la Th-T emet molta fluorescència amb un màxim d'emissió a 485 nm, en absència d'estructures amiloides la Th-T té una fluorescència molt dèbil.

En el present treball els assaigs d'unió de Th-T s'han dut a terme a una concentració de proteïna entre 10 i 100 µg/mL, la mostra s'ha diluït en un tampó fosfat sòdic 10 mM, NaCl 150 mM que contenia 65 µM de Th-T fins a un volum final d' 1mL. Es va usar una longitud d'ona d'excitació de 440 nm i l'emissió es va recollir entre 460 i 560 nm.

M.3.4.5.3.- Determinació de la unió d'ANS

L'anió 1-anilino-8-naftalen sulfonat (ANS) s'uneix a les regions hidrofòbiques de la proteïna mitjançant els seu grup apolar. Aquesta unió es pot seguir pel creixement de la intensitat de la fluorescència del ANS. La unió d'ANS també ha estat àmpliament usada per l'estudi d'intermediaris *molten globule*. Nosaltres hem seguit la unió d'ANS en estats parcialment desplegats de l'α-quimotripsina amb una concentració de proteïna de 100 µg/mL i usant una longitud d'ona d'excitació de 365 nm i recollint l'emissió entre 420 i 600 nm.

M.3.5.- Mètodes de microscòpia

M.3.5.1.- Microscopi electrònic de transmissió

El microscopi electrònic de transmissió (MET) te un límit de resolució de prop de 2 nm. En el MET la mostra s'observa després d'haver estat tenyida amb ions de metalls pesats a sobre d'una reixeta de carboni. Els electrons són dispersats quan passen a través d'una fina secció de la mostra, i després són detectats i projectats

cap a una imatge sobre una pantalla fluorescent. En el present treball el protocol emprat per preparar les mostres ha estat el següent:

- 1.- Es realitzen una sèrie de dilucions de la mostra original per poder determinar quina concentració permet obtenir la millor preparació. Si hi ha massa mostra podem arribar a trencar la reixeta i si ens ha quedat massa diluïda pot succeir que les estructures amiloides es dissolguin o be que n'hi hagi molt poques en la preparació.
 - 2.- De cada dilució es posen 5 μL a sobre de la reixeta amb molta cura de no tocar-la amb la pipeta, i es deixa la gota durant aproximadament 2 minuts.
 - 3.- S'eixuga la gota amb unes tires de paper Whatman des dels laterals de la reixeta evitant el contacte del paper amb la reixeta.
 - 4.- Es deixa assecar la gota.
 - 5.- Opcionalment es pot netejar la mostra dipositada a la reixeta amb aigua Milli-Q.
 - 6.- Un cop la mostra està seca es posen 5 μL d'una preparació d'acetat d'uranil prèviament centrifugada, i es deixa incubant durant 2 min.
 - 7.- Es retira l'acetat d'uranil amb unes tires de paper Whatman des dels laterals de la reixeta evitant el contacte del paper amb la reixeta.
 - 8.- Es deixa assecar la mostra i ja està llesta per anar al MET. Opcionalment es pot realitzar un rentat de la mostra amb aigua Milli-Q per treure l'excés d'acetat d'uranil.
- En aquest treball el microscopi que hem usat és un Hitachi H-7000.

M.3.5.2.- Microscopi òptic de fluorescència

El microscopi de fluorescència es basa en el fet que hi ha certs compostos que emeten energia detectable com a llum visible quan són irradiats amb llum d'una determinada longitud d'ona. La mostra pot ser fluorescent per ella mateixa o pot haver estat tractada amb agents químics fluorescents. En el nostre treball l'hem usat per veure la fluorescència emesa per α -quimotripsina i la latexina que prèviament han estat incubades amb Th-T. El protocol seguit per la preparació de les mostres ha estat el següent:

- 1.- S'incuba la mostra amb una dissolució que contingui Th-T durant 5 min.
- 2.- Es diposita la barreja a sobre d'un portaobjectes.
- 3.- Es col·loca un cobreobjectes al damunt de la mostra i es segella.

M.3.5.3.- Microscopi òptic de contrast de fase

És un tipus de microscopi que es basa en la diferent velocitat de llum quan aquesta travessa diferents parts de la mostra i per tant hi ha un canvi en la fase. És un microscopi que augmenta els contrastos de mostres transparents ja que influeix en el camí òptic de la llum. En aquest treball s'ha emprat per veure quin aspecte tenien les mostres que s'havien d'analitzar amb el microscopi de fluorescència.

M.3.5.4.- Microscopi òptic de llum polaritzada

Quan les mostres que uneixen congo red es visualitzen en un microscopi de llum polaritzada presenten birrefringència. La mostra passa de tenir un color vermell intens a ser totalment negra i només aquelles estructures de la mostra que han unit el colorant presenten un color verd poma. En el nostre treball la mostra que es va analitzar mitjançant el microscopi de llum polaritzada no va presentar aquesta birrefringència característica.

M.3.6.- Diàlisi

La diàlisi s'ha emprat per eliminar les sals de dissolucions proteiques. Els sacs de diàlisi usats són de la casa Medicell Internacional LTD (Gran Bretanya) de mida d'exclusió de 12 kDa. Es dialitza la mostra front un volum de tampó en una relació 1:1000 (v/v) *over-night* a 4°C.

Per a volums petits s'ha utilitzat la tècnica de *drop* diàlisi descrita per Görisch (1988), en què s'utilitzen les membranes sèrie VS de 0,025 µm i 25 mm de diàmetre (Millipore).

M.3.7.- Mètodes cromatogràfics

M.3.7.1.- Cromatografia atmosfèrica d'afinitat amb histidines

La cromatografia d'afinitat permet obtenir la proteïna en un elevat grau de puresa mitjançant un únic pas de purificació. En el present treball hem utilitzat una reïna *Chelating Sepharose Fast Flow* (Amersham Biosciences).

La fase sòlida està formada per un llit de sefarosa a on hi unirem el metall, en el nostre cas l'ió Ni^{+2} . Aquesta cromatografia es basa en la interacció específica que es dona entre l'anell imidazola de la cua de 6 histidines de la proteïna recombinant amb el Ni^{+2} que hi ha immobilitzat a la matriu de sefarosa. En el nostre cas, hem utilitzat columnes de 2 x 23 cm. La cromatografia s'ha realitzat a 4°C i a un flux màxim de 0,6 mL/min. S'ha utilitzat una bomba peristàltica P-1 d'Amersham Biosciences.

La fracció soluble del medi intracel·lular d'E. coli (M.2.10.2) ha estat la mostra tractada sistemàticament mitjançant aquesta cromatografia. S'equilibra la columna amb 5 volums de columna del tampó A i posteriorment es carrega la mostra, es renta a continuació amb el mateix tampó fins que l'absorbància del rentat sigui inferior a 0,2 UDO_{280} . Es realitza un segon pas de rentat amb 10 volums de columna del tampó B. La proteïna d'interès s'elueix amb un gradient de 60 mM a 1 M d'imidazola. Finalment per regenerar la columna es passa una dissolució que conté 50 mM EDTA, 1 M NaCl.

En aquest treball la cromatografia d'afinitat s'ha utilitzat també com a segon pas de purificació. Un cop eluïda la proteïna recombinant pura del primer pas de purificació aquesta es dialitza front tampó PBS. Un cop equilibrada la mostra en PBS es procedeix al tall de la proteïna amb trombina per eliminar la proteïna de fusió, GST, i la cua de 6 histidines de l'extrem amino terminal. La relació trombina/proteïna recombinant és de 10 unitats de trombina per 1 mg de proteïna. Es deixa incubant en agitació a temperatura ambient 12 h. A continuació es torna a passar la mostra per la columna d'histidines equilibrada com s'ha indicat anteriorment en tampó A. En aquest segon pas de purificació la proteïna d'interès no s'unirà a la reïna ja que li hem eliminat la cua d'histidines i per tant ens eluïrà al principi en tampó A. Per aquest motiu és necessari que el grau de puresa del primer pas de purificació, abans de procedir al tall amb trombina, sigui molt elevat ja que, si no és així, ens eluïrà en el segon pas de purificació juntament amb les proteïnes contaminants.

Materials i Mètodes

Els tampons usats han estat:

Tampó A :

20 mM Tris-HCl
0.5 M NaCl
5 mM imidazola
S'ajusta a pH 7,9

Tampó B:

20 mM Tris-HCl
0,5 M NaCl
60 mM imidazola
S'ajusta a pH 7,9

Tampó C:

20 mM Tris-HCl
0,5 M NaCl
1 M imidazola
S'ajusta a pH 7,9

Tampó PBS X1

120 mM NaCl
2,7 mM KCl
10 mM fosfat sòdic
S'ajusta a pH 7,4

Tots els tampons utilitzats han estat filtrats a través de membranes de 0,22 μm de porus.

M.3.7.2.- Cromatografia líquida d'alta resolució (FPLC) de bescanvi aniónic

Aquesta cromatografia es basa en la interacció que es produeix entre una reïna carregada i una proteïna de càrrega oposada. En aquest treball s'han utilitzat

columnes amb reïnes carregades positivament i les proteïnes s'han dissolt en tampons a pH superior al seu punt isoelèctric, de manera que han adquirit càrrega negativa.

La utilització del FPLC en el fraccionament cromatogràfic permet la purificació en menys temps que altres mètodes, proporciona un alt rendiment i permet una recuperació òptima d'espècies pures i natives en menys etapes.

L'equip utilitzat ha estat un ÄKTA Purifier (Amersham Biosciences). Tots els disolvents usats en l'FPLC han estat preparats amb aigua destil·lada i desionitzada mitjançant un equip Milli-Q. Posteriorment han estat filtrats a través de membranes de 0,22 µm (Sartorius) i desgasificats en trompa de buit durant 10 min. La mostra dialitzada front el disolvent d'equilibrat de la columna, es filtra amb una membrana de 0,22 µm de porus de baixa adsorció de proteïnes (Millex-GV-Millipore) abans de ser injectada en la columna.

Les condicions de purificació de la latexina així com del complex latexina/hCPA4 han estat les següents:

Columna d'intercanvi aniònic: TSK DEAE-5PW de TosoHaas amb un diàmetre de partícula de 10 µm, un porus de 0,1 µm i unes dimensions de 7,5 x 75 mm, protegida amb una precolumna del mateix material.

Tampons usats:

Tampó A: Tris-HCl 20 mM pH 8,0

Tampó B: Tris-HCl 20 mM pH 8,0 Acetat amònic 0,8 M

Gradient: de 0% de tampó B al 60% en 200 minuts

de 60% de tampó B al 100% en 14 minuts

Flux: 0,5 mL/min

L'elució de les proteïnes es segueix per mesura de l'absorbància a 280nm.

M.3.7.3.- Cromatografia líquida d'alta resolució (FPLC) de filtració en gel

L'equip FPLC ha estat el mateix que en M.3.8.2.

En la cromatografia de gel filtració les molècules se separen en funció de la seva mida. La separació depèn de la diferent capacitat que tenen les molècules analitzades per entrar en la reïna usada en la cromatografia. Mentre molècules molt grosses es mouen molt ràpidament pel llit cromatogràfic, les més petites, que poden entrar pels porus del gel, es mouen lentament.

Els tampons per aquesta cromatografia han estat tractats tal i com s'indica en l'apartat anterior.

Condicions de la cromatografia:

Columna: Sephadex 75 HR 10/30

Rang de fraccionament: 3.000 – 70.000 Da

Tampó: Tris 20 mM, NaCl 0,1 M (pH 7,5)

Flux: 1 mL/min.

M.4.- CRISTAL·LITZACIÓ DE PROTEÏNES

M.4.1.- Cristal·lització de la CPA1 humana

La cristal·lització de la CPA1 humana es va aconseguir després de l'activació i posterior purificació de l'enzim actiu seguint el mètode indicat a l'apartat M.3.8.3.

La concentració i diàlisi de la mostra es va realitzar mitjançant el sistema d'Amicon Centriprep 10 Concentrator. El tampó utilitzat per la diàlisi va ser Tris-HCl 5 mM, 50 mM NaCl (pH 7,5), que permet assajar la cristal·lització de l'enzim des de condicions de força iònica baixes i varietat de pHs. La CPA1 es va concentrar fins a 6 mg/mL. El creixement dels cristalls de proteïna es va realitzar a dues temperatures, 4 i 20°C, només a 20°C van aparèixer cristalls.

Per trobar les condicions de creixement dels cristalls d'una proteïna s'analitza empíricament totes les condicions a on s'han donat més casos de cristal·lització proteica. Aquest assaig es va realitzar mitjançant els *Kits* de Hampton Crystal Screen I i II, Index I i II i Index I i II. La cristal·lització es va realitzar per la tècnica de la gota assegurada: es posa una gota de la barreja de la proteïna amb la dissolució precipitant (1 µL) sobre de cada elevació que conté cada pouet de la placa; a cada pouet hi ha una condició diferent de cristal·lització del *Kit* i el buit s'aconsegueix amb un film que queda enganxat a la superfície de la placa.

Una vegada observats cristalls en una condició (22% (w/v) PAA 5100, HEPES 0,1 M, clorur de magnesi 0,02 M (pH 7,5)), es va optimitzar el seu creixement variant el pH des de 6 a 8,5, variant la concentració de clorur de magnesi de 5 mM a 85 mM, la de PAA del 15 al 30%, augmentant la concentració de proteïna de les gotes fins a 2 i 4 µL de CPA1 per 1 o 2 µL de dissolució precipitant, així com augmentant el volum de la gota.

Per poder congelar els cristalls es va afegir un 30% de glicerol a la dissolució de cristal·lització.

La presència de la hCPA1 als cristalls es va determinar per espectrometria de masses i per seqüenciació N-terminal per veure si l'extrem N-terminal havia sofert alguna degradació al llarg del procés de cristal·lització.

M.4.2.- Cristal·lització del complex latexina/hCPA4

El complex latexina/hCPA4 complex es va concentrar fins aproximadament 7 mg/ml. Per trobar les condicions de creixement dels cristalls es va realitzar com en l'apartat anterior tot un seguit d'assajos amb *kits* comercial. En aquest cas la cristal·lització es va realitzar usant la tècnica de la gota penjant: es posa una gota de la barreja de la proteïna amb la dissolució precipitant sobre uns cobreobjectes; aquests s'inverteixen i es posicionen amb la gota penjada sobre els diferents orificis d'una placa de cultiu cel·lular; a cada orifici hi ha una condició diferent de cristal·lització del *kit* i el buit s'aconsegueix gràcies a que els cobreobjectes tenen silicona al seu voltant. El creixement dels cristalls de proteïna es va realitzar a dues temperatures, 4 i 20°C. Els cristalls que millor van difractar van créixer en una condició, 40% 2-metil-2,4-pentanediol (MPD), Bis-Tris 0,1 M (pH 6,5) i 0,2 µL d'acetona 40% després de 4 dies a 20°C. Els cristalls es van congelar directament en N₂ un cop agafats de la gota. La presència d'ambdues proteïnes als cristalls es va confirmar per espectrometria de masses.

PART I

CHAPTER 1

Three-dimensional structure of human carboxypeptidase A1

1.1- SUMMARY

Zinc metalloproteinases comprise a large family of enzymes with a wide variety of biological roles. One of the most widely studied zinc metalloenzymes is carboxypeptidase A (CPA), which catalyzes both exopeptidase and exoesterase reactions.

Although CPA has been the subject of extensive research in the last 30 years, the exact catalytic mode of action of this representative enzyme and the rest of the zinc proteinase family is still not completely clarified. Up to now only the 3D structures of porcine and bovine carboxipeptidase A and human CPA2 were available, but not that of human CPA1. Purified recombinant hCPA1 has been crystallized using a sitting-drop vapor diffusion method; the crystals obtained belong to the tetragonal space group $P4_32_12$ with unit cell parameters $a = b = 129.56 \text{ \AA}$, $c = 90.86 \text{ \AA}$, and present two carboxypeptidase A1 molecules per asymmetric unit. The crystal diffracted up to 1.6 \AA and was suitable for detailed X-ray analysis. The structure of human CPA1 at high resolution may contribute to improve our understanding of the catalytic mechanism of CPA, with possible implications for the general family of zinc proteinases. This structure provides a template for modeling the structure of physiologically relevant carboxypeptidases, like those implied in hormone and neuropeptide processing, as well as for the specific design of inhibitors.

1.2.- INTRODUCTION

Carboxypeptidases (CPs) are found throughout the animal kingdom, with at least one member in every animal species whose genome has been sequenced to date. CPs are also found in plants and some bacteria and unicellular organisms. Altogether, there are 17 members in the human genome, most of which have been expressed and characterized. Within mammals, at least one CP is present in every tissue or fluid that has been examined. However, most CPs have a fairly restricted pattern of tissue distribution, reflecting a specific physiological function.

Human pancreas secretes many of the proteolytic enzymes that participate in the digestive processes of alimentary proteins and peptides. Three forms of monomeric procarboxypeptidases, A1, A2 and B occur in human pancreas, together with a binary complex between the A1 form and proproteinase E. The mature forms of metalloprocarboxypeptidases catalyze the removal of C-terminal amino acids from proteins and/or peptides. The different members of the CP family perform different functions, typically differing in their specificity: carboxypeptidase B preferentially hydrolyses basic residues, while procarboxypeptidases A1 and A2 excise aliphatic and aromatic side-chain residues, respectively. No three-dimensional structure of human CPA1 is available until now.

Many structural studies with bovine CPA involving the use of inhibitors, substrate or transition-state analogs, as well as site-directed mutagenesis studies (Gardell *et al.*, 1985; Phillips *et al.*, 1992) have led to the identification of the residues involved in the mechanism of catalysis. Although some minor details remain unclear, the enzyme mechanism has been essentially elucidated (Christianson & Lipscomb, 1989; Kim & Lipscomb 1990) and is briefly summarized. Important elements for the catalysis are the Zn⁺² coordinated to His 69, Glu 72 and His 196. The terminal carboxylate group of the peptide substrate is fixed by Arg 145 and Tyr 248, while the carbonyl group of the scissile peptide bond becomes positioned near to Glu 270, Arg 127 and Zn. This favors a nucleophilic attack on the carbonyl carbon atom by a water molecule activated by Zn and Glu 270. Alternatively, a direct nucleophilic attack by Glu 270 to form an acid anhydride has been proposed. Arg 127 contributes to the polarization of the oxygen atom of the carbonyl group. Kinetic studies have shown that substrate binding extends over five residues, P₄ to P₁' , which interact with subsites S₄ to S₁' of the enzyme. The subsite that determines the primary specificity is S₁'.

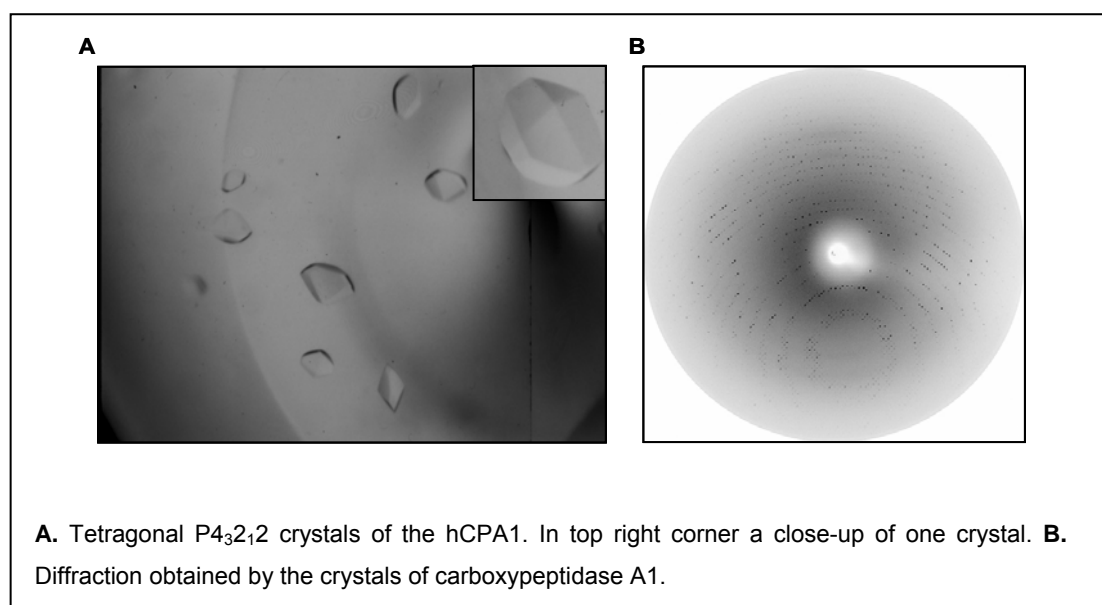
1.3.- MATERIALS AND METHODS

1.3.1.- Protein preparation and crystallization

Human PCPA1 was over-expressed in the supernatant medium of the yeast using the *Pichia pastoris* heterologous system. Purification of the proenzyme was performed with a two-step chromatographic approach, consisting of an atmospheric hydrophobic interaction chromatography on a butyl column, followed by FPLC anion-exchange chromatography on a TSK DEAE-5PW preparative column (Segura-Martin, 2002). Recombinant hPCPA1 was activated with trypsin at 40:1 ratio (w/w) at 25°C. The removal of the pro domain was performed in a gel filtration column (Sephadex 75 HR) connected to an FPLC system. The purity of the active hCPA1 form was confirmed by SDS-acrylamide gel electrophoresis.

The purified protein at 6 mg/mL dissolved in Tris 5 mM, NaCl 50 mM (pH 7.5), was mixed at 1:1 ratio with the crystallization solutions. Crystals grew in 22% (w/v) PAA 5100, HEPES 0.1 M, magnesium chloride 0.02 M (pH 7.5) crystallization buffer using a sitting drop vapor diffusion set-up at 20°C (Fig. 1.1).

Figure 1.1.- Human carboxypeptidase A1 crystals and its diffraction pattern



1.3.2.- Data collection and processing

For data collection, crystals were soaked in a cryobuffer containing 30% (v/v) glycerol before freezing in a nitrogen stream at 100K (Oxford Cryosystems Cryostream). High resolution data at 1.6 Å were collected at the DESY BW6 beamline (Deutsches Elektronen Synchrotron, Hamburg, Germany) at a wavelength of 1.050 Å. Indexing and integration of the diffraction data was carried out with DENZO (Otwinowski & Minor, 1997). The data were merged, scaled and truncated using programs supported by the Collaborative Computational Project Number 4 (The CCP4 suite, 1994).

Crystals belonged to the tetragonal space group $P4_32_12$ with unit cell parameters $a=b= 129.565$ Å, $c= 90.856$ Å and contained two molecules per asymmetric unit.

1.3.3.- Structure determination and processing

The structure of human CPA1 was solved by molecular replacement using the program suite AMoRe (Navaza, 1994) and a bovine CPA (5CPA) search model (Rees *et al.*, 1983), which consisted of the residues conserved in both enzymes, while all other residues were mutated to alanine. After rigid body fitting, the best solution reached a correlation coefficient and R factor of 36.1% and 46.1%, respectively, for 15.0–3.5 Å data. Model building was performed on a Silicon Graphics Indigo 2 workstation using the program MAIN (Turk, 1992). Positional and individual temperature factor refinement was carried out with the program CNS (Brunger *et al.*, 1998) until convergence without imposing non crystallographic symmetry constrains, and the final model was obtained yielding a final R factor of 17.8%. The two hCPA1 molecules present in the crystal asymmetric unit are structurally equivalent with an r.m.s.d. of 0.23 Å. Therefore, the following discussion will consider only molecule A.

The hCPA1 chain encompasses protein residues Ala 1 to Pro 308 and a Zn^{+2} ion which was identified and refined with full occupancy. Additional electron density near the active site was interpreted as a polyacrylic acid (PAA) molecule and three HEPES molecules were found on the surface. During the final steps of refinement, two alternative conformations were identified and built for a few side chains.

The Ramachandran plot calculated for the final CPA1 model with the program PROCHECK (Laskowski *et al.*, 1993) shows that 87.5% of the residues are in the 'most favored' regions of the ϕ , ψ plane, 12.1% of residues are in the 'additionally allowed' regions and only 1 amino acid residue appears in a 'disallowed region' (Fig. 1.3.).

The crystal data and the refinement parameters are given in Table 1.1.

Table 1.1

X-ray data and refinement statistics	
Space group	P4 ₃ 2 ₁ 2
Cell Constants	
a, b, c (Å)	129.565, 129.565, 90.856
α , β , γ	90.0, 90.0, 90.0
Molecules/asymmetric unit	2
Resolution (Å) (overall/last shell)	19.99-1.6/1.68-1.60
Reflections measured	521,057
I/ σ (I)	33.5/5.0
Unique reflections	101,271
R_{merge}^a (%) (overall)	5.4
Completeness (%) (overall/last shell)	98.8/83.1
Reflections used for refinement(total/test set)	100,653/2012
R_{factor}^b (%) / R_{free}^c (%)	17.82/19.18
r.m.s.d. bond lengths (Å)	0.0066
r.m.s.d. bond angles (°)	1.38
r.m.s.d. bonded B-factors lengths (Å ²)	0.939
Average B-factor for protein atoms (Å ²)	19.99
Average B-factor for PAA (Å ²)	22.98
Average B-factor for metallic ions (Å ²)	15.37
Ramachandran core conformations (%)	87.3
Additionally allowed conformations (%)	12.5
Disallowed conformations (%)	0.2

$$^a R_{\text{merge}} = \sum_{\text{hkl}} |<I> - I| / \sum_{\text{hkl}} I$$

$$^b R_{\text{factor}} = \sum_{\text{hkl}} ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum_{\text{hkl}} |F_{\text{obs}}|$$

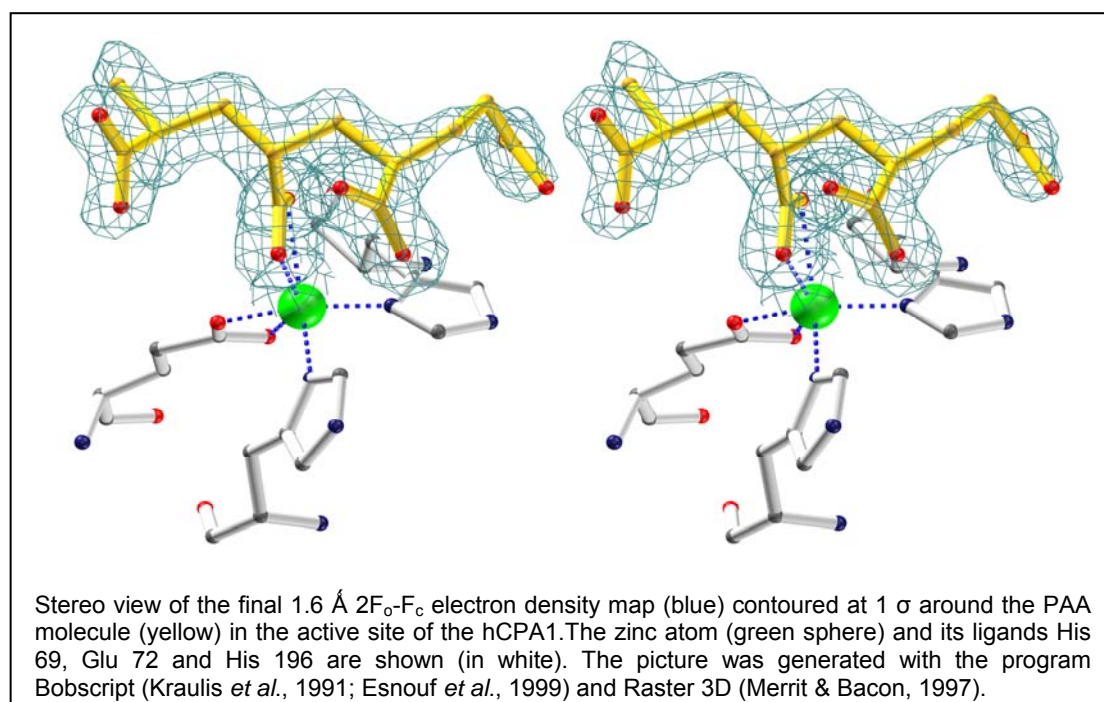
$^c R_{\text{free}}$ = is the R value calculated with 500 reflections that were not used for the refinement

1.4.- RESULTS AND DISCUSSION

1.4.1.- Overall structure of human carboxypeptidase A1

The final model of native hCPA1 at 1.6 Å resolution consists of 2509 non-H atom positions representing 308 protein residues, one Zn⁺² ion (at the active site), one PAA molecule and three HEPES molecules. The final crystallographic *R* factor and *R* free values were 17.82% and 19.18% respectively. Both of these parameters indicate a consistent and reliable refinement of the structure. The final electron-density map is of excellent clarity, with the vast majority of the protein molecule displaying well defined electron-density in the (2F_o-F_c) map (Fig. 1.2.).

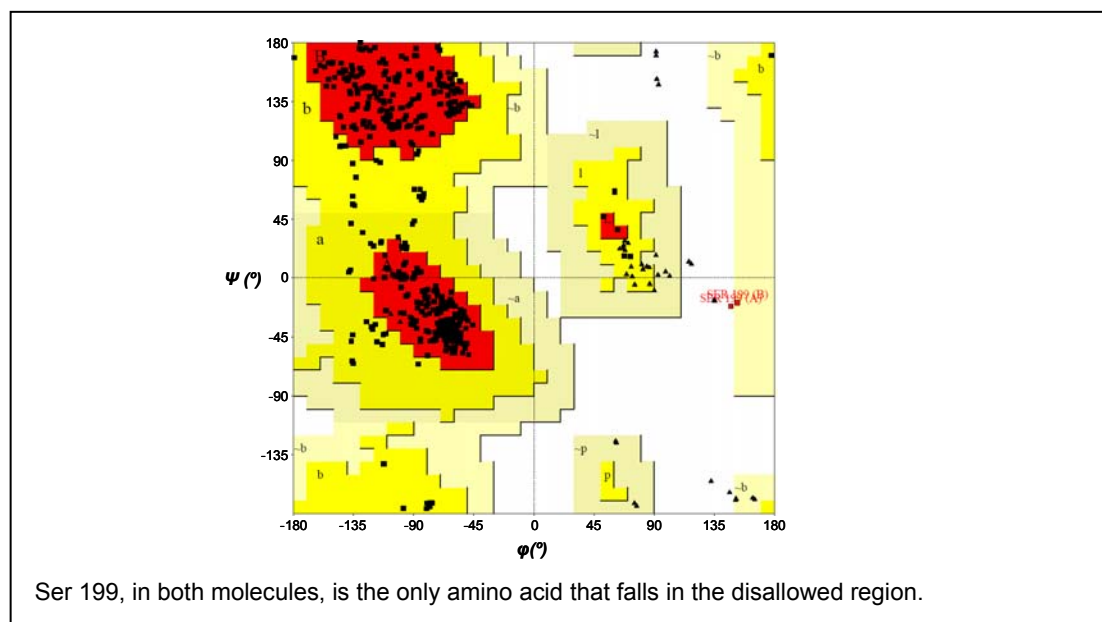
Figure 1.2.- Electron density map around the PAA molecule a in the active site of hCPA at 1.6 Å resolution



For both molecules, all residues fall in the allowed regions of the Ramachandran plot except Ser 199 (Fig 1.3), which is nevertheless unambiguously defined in the electron density maps. The abnormal conformation of Ser 199 is explained by the presence of a cis peptide bond between Ser 199 and Tyr 198 (also present in bCPA, hPCPA2 and pPCPA), an unusual cis conformation which is probably required for the

correct folding of the protein and is forced because of the well-structured arrangement around His 196 coordinating the active site Zn^{+2} .

Figure 1.3. The Ramachandran plot for the native hCPA1 model

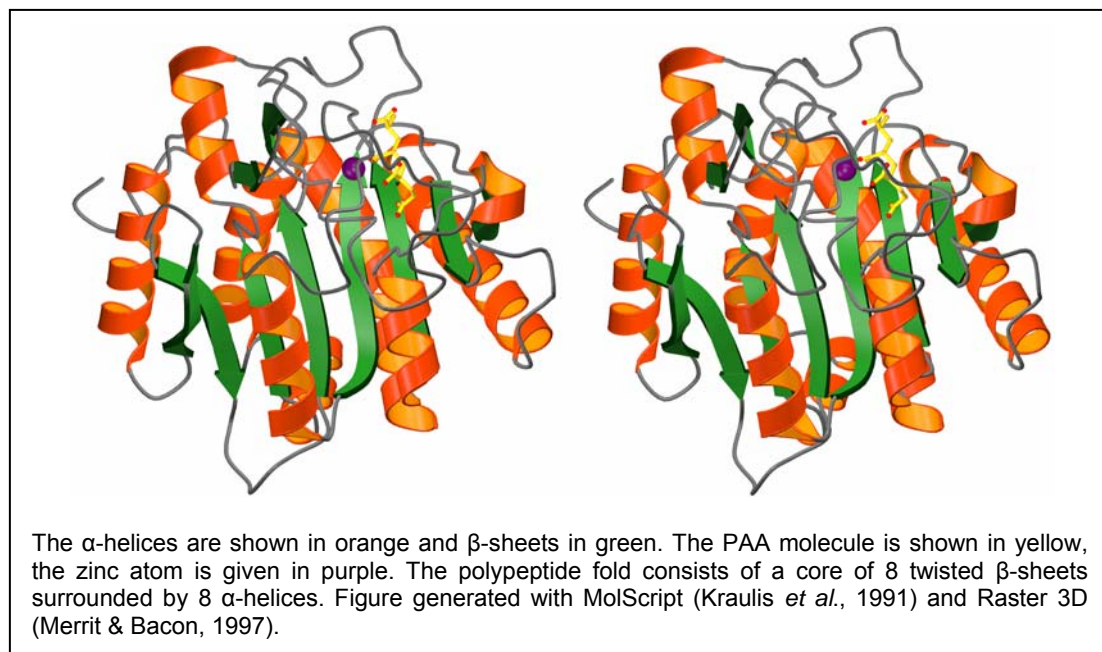


Like its other non-human homologues (Vendrell *et al.*, 2004), the tertiary fold of hCPA1 corresponds to the α/β hydrolase fold and is formed by a central mixed parallel/antiparallel 8-stranded β -sheet, with a 120° twist between the first and the last strand, over which 8 α -helices pack on both sides to form a globular molecule (Fig. 1.4). As in bovine and porcine CPAs, a disulfide bond is found in human CPA1 between Cys 138 and Cys 161. Moreover, as previously reported for other pancreatic carboxypeptidases, three peptide bonds, those between residues Ser 197-Tyr 198, Pro 205-Tyr 206 and Arg 272-Asp 273 are found in the cis conformation. The well structured arrangement of His 196 in coordination with the Zn^{+2} ion forces the formation of the Ser 197-Tyr 198 cis-peptide bond between the two following residues; similarly, the two remaining cis peptide bonds are very close to residues involved in the active site, favoring appropriate structural arrangements of residues around the active site region.

1.4.2.- The active site

The active site is located in a cavity formed by parallel strands of the internal β -sheet at the switch point between strands 3 and 5, two helices and a no-regular extension that partially covers it.

Figure 1.4. Stereo ribbon plot representation of hCPA1



It has been reported that the active site of CPA is found in two different conformations depending on the binding of substrates, substrate analogs or inhibitors (Rees & Lipscomb, 1981). In the presence of substrates, inhibitors or anions that occupy the specificity pocket, Tyr 248 rotates about its χ_1 torsion angle in a movement that spans about 12 Å and fixes it in the down position, closing the access to the active site. Tyr 248 has been found in the “down” position in the PCI-CPA complex (Rees & Lipscomb, 1982 (4CPA)), bovine PCPA (Guasch *et al.*, 1991 (1PCA)), the LCI-CPA2 complex (Reverter *et al.*, 2000 (1DTD)) and now in hCPA1, among other CPs, while it is in found the “up” position in CPA (Kilshtain-Vardi *et al.*, 2003 (1M4L)) and PCPB (Coll *et al.*, 1991 (1NSA)).

Some of the residues of hCPA1 have been found in the conformation induced by substrate binding. An unexpected strong density, corresponding to a PAA molecule, has been observed in the active site. The binding of a PAA molecule to the active site

as an inhibitor analog delineates the positions of the S_1' and S_1 subsite pockets of the enzyme (Fig. 1.5). These sites are occupied by the oxygens of groups γ and β . $O_{\gamma 1}$ forms an hydrogen bond with the OH of Tyr 248 (2.63 Å) in the “down” position and with the NH2 of Arg 145 (2.82 Å); $O_{\gamma 2}$ group of the PAA molecule also interacts with the NH1 group of Arg 145 forming a hydrogen bond of 2.80 Å; the N δ 2 atom of Asn 144, the last residue associated with the S_1' subsite, interacts with the $O_{\gamma 2}$ of the PAA (2.87 Å). The S_1 subsite is defined by Arg 127 and Glu 270, both contributing to the polarization of the carbonyl group of the scissile peptide bond and to proton exchange. The interaction of PAA with this subsite involves several hydrogen bonds. The $O_{\epsilon 1}$ of the carboxylate group of Glu 270 is within hydrogen bonding distance (2.69 Å) from the $O_{\beta 2}$ of PAA. The other $O_{\beta 1}$ oxygen from PAA also forms a hydrogen bond with the guanidinium group of Arg 127 (2.86 Å) (Table 1.2).

Similar interactions that involve the two main residues of the S_1 site, Arg 127 and Glu 270, were also found in the LCI-CPA2 complex but were not present in the PCI-CPA complex.

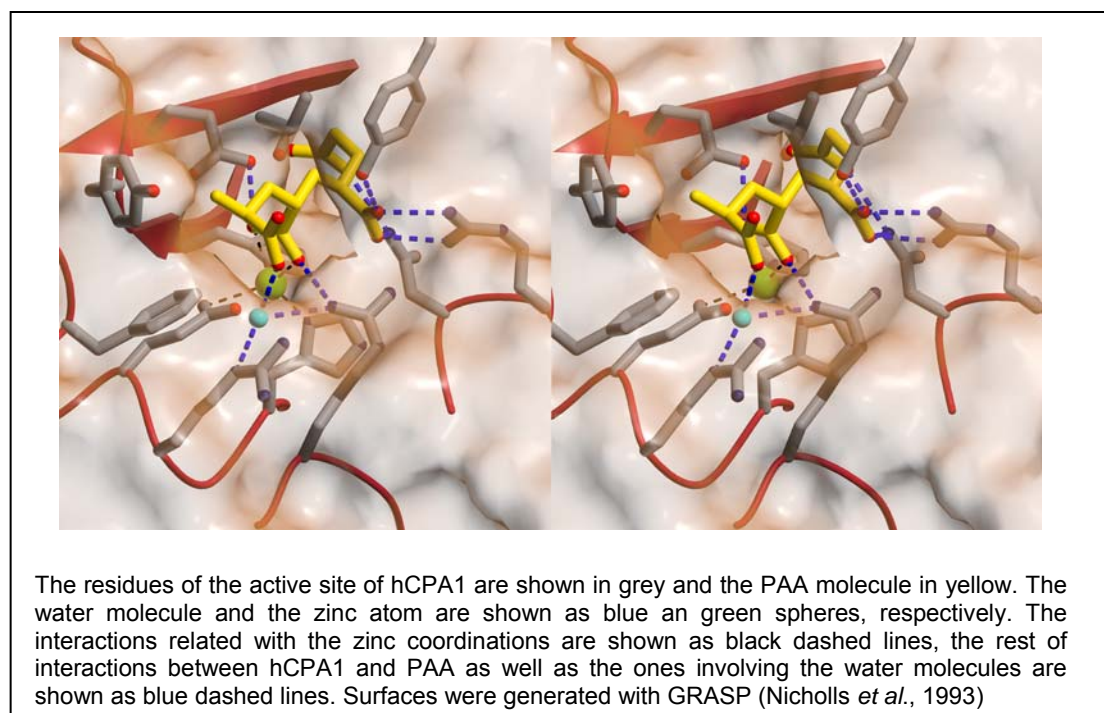
Table 1.2.

Intermolecular contacts between the PAA molecule and the enzyme moiety

	CPA	PAA	Distance (Å)	Type of interaction
S_1'	Tyr 248	$O_{\gamma 1}$	2.63	The OH of Tyr 248 with the $O_{\gamma 1}$ of PAA
	Arg 145	$O_{\gamma 1}$	2.82	The NH2 of Arg 145 with the $O_{\gamma 1}$ of PAA
	Arg 145	$O_{\gamma 2}$	2.80	The NH1 of Arg 145 with the $O_{\gamma 2}$ of PAA
	Asn 144	$O_{\gamma 2}$	2.87	The N δ 2 of Asn 144 with the $O_{\gamma 2}$ of PAA
S_1	Glu 270	$O_{\beta 2}$	2.69	The $O_{\epsilon 1}$ of Glu 270 with the $O_{\beta 2}$ of PAA
	Arg 127	$O_{\beta 1}$	2.86	The NH2 of Arg 127 with the $O_{\beta 1}$ of PAA

The coordination of Zn^{+2} in the active site is conserved compared to other metallo-carboxypeptidases. The first coordination shell of the zinc atom consists of four protein atoms, $O_{\epsilon 1}$ and $O_{\epsilon 2}$ of Glu 72 and the N δ 1 atoms of His 69 and His 196, and one water molecule. In the present structure the catalytic water molecule has been displaced by the PAA and is instead substituted by the $O_{\beta 1}$ and $O_{\beta 2}$ oxygens of PAA that interact with Zn^{+2} (Fig. 1.5).

Figure 1.5. Stereo view of the section around the S₁' pocket of hCPA1 and interactions with the PAA



The water molecule, now 3.8 Å away from the zinc, is interacting with the N δ 1 group of Arg 71 (2.92 Å), the NH₂ of Arg 127 (2.96 Å) and with the O α 2 group of the PAA molecule (2.65 Å).

1.4.3.- Comparison of the overall hCPA1 structure with other structures of CPs.

In order to evaluate the new information obtained from the current hCPA1 structure, it was compared with the most recent structures of CPs available from the Protein Data Bank (1PCA; Guasch *et al.*, 1992; 1m4l; Kilshtain-Vardi *et al.*, 2003; 1DTD; Reverter *et al.*, 2000; 1JQG; Estébanez-Perpiñá *et al.*, 2001; 1KWM; Barbosa-Pereira *et al.*, 2002; 1QMU; Gomis-Rüth *et al.*, 1999; 1OBR; Teplyakov *et al.*, 1992) Table 1.3.

Table 1.3.

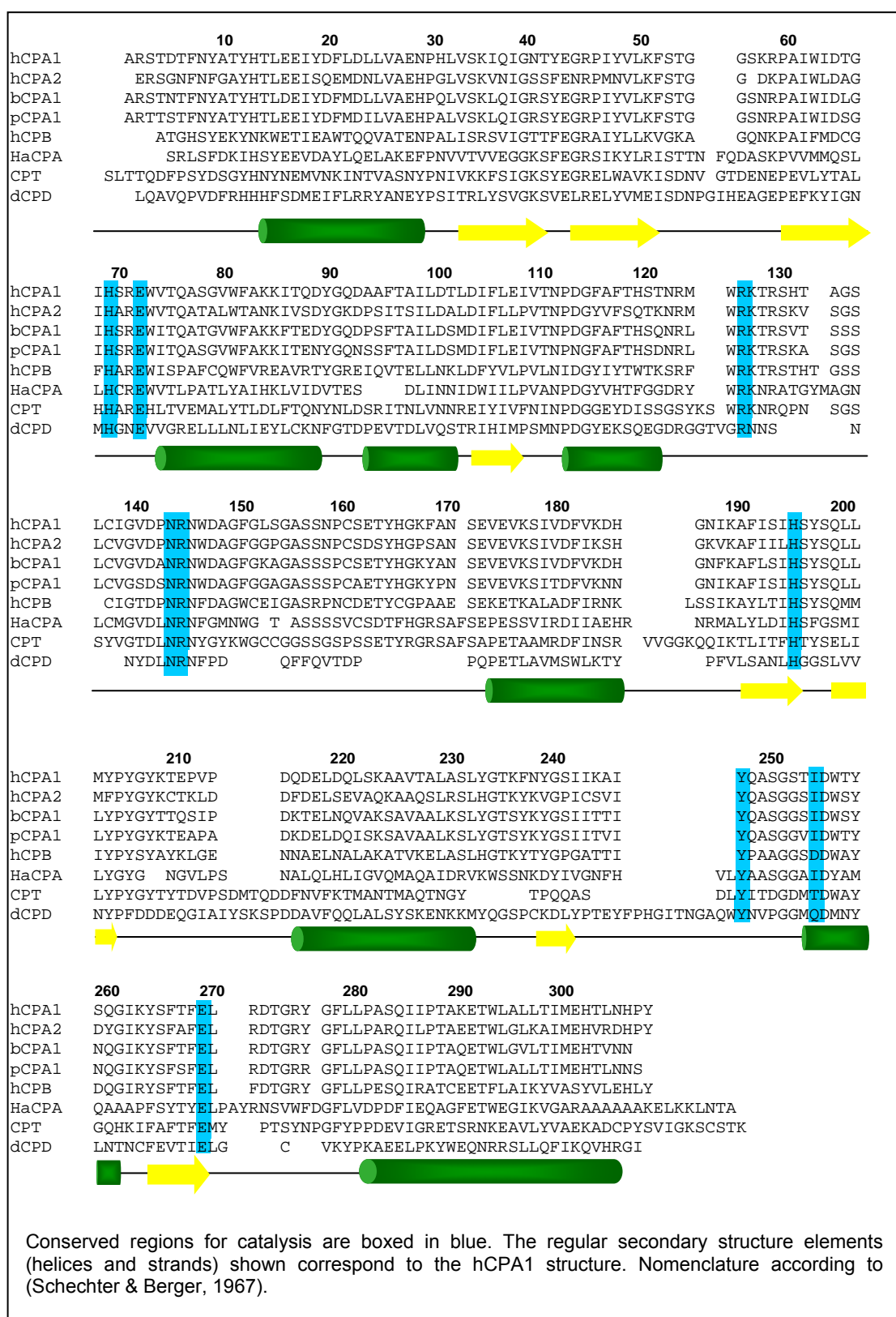
CPs	rmsd (Å)
bCPA (1PCA)	0.38
pCPA (1m4l)	0.41
hCPA2 (1DTD)	0.59
CPA Ha (1JQG)	2.48
hCPB (1KWM)	1.44
CPD (1QMU)	6.68
CPT (1OBR)	3.73

In a superposition using main-chain atoms alone, r.m.s.d. values of 0.38 Å and 0.41 Å were found for bovine and porcine CPA (bCPA and pCPA) respectively, the main differences being located in loops: 92 to 94; 132 to 135; 211 to 213; 237 to 239 and 245 to 248. Most of them can be attributed to differences in sequences in both structures (Fig. 1.6), except the latter which is attributed to a change in the position of Tyr 248 (“down” in hCPA1 and pCPA structures and “up” in bCPA).

CPA2 is the human enzyme bearing a higher degree of homology with hCPA1 and the pairwise comparison of 3D structures renders a r.m.s.d. value of 0.59 Å. The specificity pocket of both enzymes is lined by residues at equivalent positions in the 3D structure. In the A1 form, these residues are Asn 144, Ser 194, Met 203, Ile 243, Ile 247, Ala 250, Ser 253, Thr 254, Ile 255, Asp 256 and Thr 268. Ser 194 is substituted by a hydrophobic isoleucine in human CPA2. Besides, Gly 253, Ser 254 and Ala 268 in human CPA2 substitute bigger or more polar residues in the A1 form (García-Sáez *et al.*, 1997). Overall, these changes define a wider pocket for hCPA2, able to accommodate bulkier residues, and also with a stronger hydrophobic character.

At the active site of human CPB, although nearly identical to that observed in hCPA1, His 69 displays two alternate conformations, with approximately equal occupancies (Barbosa-Pereira *et al.*, 2002). This generates several changes in other close residues. The amino acid residue position 255 appears to be the main determinant of specificity for metallocarboxypeptidases (Fig. 1.6).

Figure 1.6.- Structure-based sequence alignment of human carboxypeptidase A1



As in all mammal CPAs, residue 255 is an isoleucine in hCPA1, while hCPB has an aspartic acid at this position, accounting for the specificity of B-type enzymes for C-

terminal basic residues. The surface loops of hCPB also show some differences as compared to the equivalent loops in hCPA1, which, together with the different orientation of the N-termini, gives an r.m.s.d. value of 1.44 Å.

Superposition of the 303 common C α atoms of hCPA1 and CPT gives an average deviation of 3.73 Å, arising from the different position of some secondary structure elements, from insertions in surface loops, and from the different orientations of the N- and C-termini. The same is observed when hCPA1 is superimposed to *Helicoverpa armigera* CPA, as the presence of insertions in surface loops induces several changes in close secondary structure elements, leading to an r.m.s.d value found of 2.48 Å.

Most members of the E/N subfamily of CPs function in the processing of biologically active peptides, whereas most members of the A/B subfamily act in the digestion of dietary protein. These functional differences are also reflected in the 3D structures. hCPA1 displays a close topological similarity to the CPD-2 catalytic subdomain. The main chain atoms of both structures can be superimposed fairly well over most parts of the structures as denoted by an r.m.s.d. of 6.68 Å. This similarity is nonetheless lower than that observed between hCPA1 and CPT. All regular secondary structure elements of hCPA1 lie in regions equivalent to CPD-2, except one extra helix present in CPD-2 that would be located in hCPA1 between the end of helix 6 and the beginning of strand 7 in hCPA1 (Fig. 1.6). The most important differences between the catalytic subdomains can be observed in the chain segments that form the funnel-like access to the active site cleft. The loop Ser 124 to Val 133 in CPD is folded inwards over the molecular body partially covering the access to the active site, while this loop is shorter in hCPA1 and does not disturb the entrance of the active site cleft. Overall, insertions and deletions shape an entrance to the active site of CPD-2 that is clearly distinguishable from that in hCPA1 and other A/B enzymes (Gomis-Rüth *et al.*, 1999). All these differences suggest that CPD-2 may have a distinct selectivity towards larger substrates.

CHAPTER 2

Structure of human
carboxypeptidase A4 with its
endogenous protein inhibitor,
latexin

2.1.- SUMMARY

The only endogenous protein inhibitor known for metallocarboxypeptidases (MCPs) is latexin, a 25-kDa protein discovered in the rat brain. Latexin, alias endogenous carboxypeptidase inhibitor (ECI), inhibits human CPA4 (hCPA4), whose expression is induced in prostate cancer cells after treatment with histone deacetylase inhibitors. hCPA4 is a member of the A/B subfamily of MCPs and displays the characteristic α/β -hydrolase fold.

Human latexin consists of two topologically equivalent subdomains, reminiscent of cystatins, consisting of an α -helix enveloped by a curved β -sheet. These subdomains are packed against each other through the helices and linked by a connecting segment encompassing a third α -helix. The enzyme is bound at the interface of these subdomains. The complex occludes a large contact surface but makes rather few contacts, despite a nanomolar inhibition constant.

Such an inhibitory mode is evocative of the recently reported structure of the *h*PCPA4. The main contacting structure is similar to the one employed for pro-domain (PD) inhibition. In both cases, active site blocking relies mainly on a loop provided by the central part of a β -sheet. This low specificity explains the flexibility of latexin in inhibiting all vertebrate A/B MCPs tested, even across species barriers. In contrast, modeling studies reveal why the E/N subfamily of MCPs and invertebrate A/B MCPs are not inhibited. Major differences in the loop segments shaping the border of the funnel-like access to the protease active site impede complex formation with latexin. Several sequences ascribable to diverse tissues and organs have been identified in vertebrate genomes as being highly similar to latexin.

They are proposed to constitute the latexin family of potential inhibitors. Because they are ubiquitous, latexins could represent for vertebrate A/B MCPs the counterparts of tissue inhibitors of metalloproteases for matrix metalloproteinases.

2.2. INTRODUCTION

Among the varied mechanisms of enzymatic activity control protein inhibitors are especially important in the cases where the active form of the enzyme is irreversibly generated. This is, for instance, the case of zinc-dependent metalloproteases (MCPs), for which a number of heterologous protein inhibitors have been described. To date, however, the only specific endogenous inhibitor described in mammals is latexin, also known as tissue or endogenous carboxypeptidase inhibitor (ECI). This 222-residue protein (in humans) bears no sequence relationship with any structurally characterized protein. It was discovered in rodents, in which it has a widespread tissue expression pattern, by two independent research lines. It is expressed in neural tissues, such as in the lateral neocortex of rats (Arimatsu, 1994; Aagaard *et al.*, 2005), and is considered a marker of regionality and development in rodent central and peripheral nervous systems (Arimatsu, 1994; Miyasaka *et al.*, 1999). It has also been suggested that it intervenes in the modulation of sensory perception and it is down-regulated in the brain of presenilin-1 knock-out mice, thus putatively playing a role in Alzheimer's disease (Mirnics *et al.*, 2003). Other studies identified latexin as an inhibitor of CPA, responsible for the lack of proteolytic activity despite the presence of mRNA encoding CPA in a series of non-pancreatic tissues (Normant *et al.*, 1995a). An animal model showed that latexin expression is also induced in rat acute pancreatitis, besides in inflammatory lung disease (Ji *et al.*, 2003). The tissue distribution of CPA and latexin correlate well in the rat (Normant *et al.*, 1995b). Latexin has also been shown to play a role in inflammation, as it is expressed in rat mast cells and inhibits mast cell CP. It is highly expressed upon induction in macrophages, together with other MCPs and cystatins, which are cysteine protease inhibitors (Arimatsu, 1994; Aagaard *et al.*, 2005). Latexin is also widely expressed in humans, though the tissue distribution is different: high in heart, prostate, ovary, kidney, pancreas and colon and only moderate in brain (Normant *et al.*, 1995b; Liu *et al.*, 2000; Uratani *et al.*, 2000).

MCPs can be classified into two subfamilies, the A/B (M14A according to the MEROPS database at <http://merops.sanger.ac.uk>) and the N/E forms (M14B), previously referred to as pancreatic and regulatory CPs, respectively (Liu *et al.*, 2000). A/B MCPs were among the first proteases studied as digestive enzymes synthesized in the pancreas of mammals (Waldschmidt-Leitz & Purr, 1929).

Molecular prototypes of the A/B MCPs are pancreatic bovine CPA (bCPA) and bovine CPB (bCPB) that excise C-terminal hydrophobic and basic amino acids, respectively. More recently, members of this subfamily have been found in archaea and bacteria, protozoa, fungi, nematodes, insects, and other invertebrates, plants, amphibians, birds, and mammals (Auld, 2004). In the last few years, functional and local ascription of A/B MCPs has moved away from the mere proteolysis of intake proteins in the digestive tract. In particular, they have been localized in brain, heart, stomach, colon, testis, and lung (Normant *et al.*, 1995a). They participate in peptide hormone activity and hormone-regulated tissue growth or differentiation, in fibrinolysis inhibition and bradykinin activation in blood serum, and in cellular response or complementing chymase in mast cells (Springman, 1994). One example is a gene product, human procarboxypeptidase A4 (hPCPA4), involved in prostate cancer (Huang *et al.*, 1999). It is up-regulated via the histone hyperacetylation pathway as a downstream effect during sodium butyrate treatment of prostate cancer cell lines. The *hPCPA4* gene is imprinted and may be responsible for prostate-cancer aggressiveness (Kayashima *et al.*, 2003). Expression was detected in human hormone-regulated tissues; however, levels are very low in normal human adult tissues, including prostate, ovary, testis, and pancreas (Huang *et al.*, 1999; Kayashima *et al.*, 2003).

A/B MCPs are secreted as inactive zymogens encompassing an N-terminal prodomain (PD) that blocks access to the active site cleft of the enzyme. Activation occurs through limited proteolysis in a connecting segment at the end of the PD. This reaction releases the active CP from its PD, which acts as an autologous inhibitor (Vendrell *et al.*, 2000). Heterologous MCP protein inhibitors have been reported, including the archetypal potato carboxypeptidase inhibitor (PCI), the orthologue from tomato (MCPI), inhibitors from the intestinal parasite *Ascaris suum* and the medical leech *Hirudo medicinalis* (LCI) (Vendrell *et al.*, 2000), and, most recently, the inhibitor from the tick *Rhipicephalus bursa* (Arolas *et al.*, 2004). A number of 3D structures are available for A/B MCPs, either in their active, inhibitor-complexed or zymogenic forms, and for members of the E/N subfamily (Gomis-Rüth *et al.*, 1999; Reverter *et al.*, 2004). No structure of an endogenous human inhibitor for MCPs has been reported to date. In this chapter the structure of *hCPA4* in complex with the inhibitor latexin is presented, together with a discussion on the role of latexin as a global inhibitor of vertebrate A/B MCPs. A detailed molecular comparison between the inhibition mode of A/B-type carboxypeptidases in the zymogen state and by the

Chapter 2

endogenous inhibitor latexin is also provided. As a protein-degrading enzyme potentially overexpressed in prostate cancer, the knowledge of the atomic structure of *hCPA4*, as well as the structural determinants of its inhibition within the zymogen and through an endogenous protein inhibitor, may provide another piece of the complex puzzle of the pathology.

2.3.- MATERIALS AND METHODS

2.3.1 Molecular cloning of the human latexin

The human brain latexin/ECl nucleotide sequence. Using:

sense 5' GGAATGGAAATCCCGCCGACCAACTACC 3'

antisense 5' GGCATCTAGCTTATGAGGCCAATAATCCCA 3' oligonucleotide

primers (designed based on the genomic human latexin sequence, GenBank access code n° 020169), a PCR reaction in a 50 µl reaction mixture was carried out under the action of Taq DNA polymerase, with human fetal brain first strand cDNAs (INVITROGEN) as a template. Amplification conditions were 40 cycles of 1 min at 95°C, 1 min at 50°C, and 1 min at 72°C. The final reaction mixture was applied onto a 1.2% agarose gel, separated by electrophoresis, and visualized by ethidium bromide staining. A discrete 762-base pair PCR product detected was excised from the gel, and the DNA fragment therein was isolated by spin filtration. A nested PCR reaction using the same sense oligonucleotide and an internal antisense

5' GCCGTCTGCCAAAGGAAGTACAAGTGGGAATAA 3' primer was performed using the first PCR product as a template to ensure product specificity. Amplification conditions were 30 cycles of 1 min at 95°C, 1 min at 60°C, and 1 min at 72°C. In this way, a 673-base pair product encoding all the latexin protein was amplified and isolated. The cDNA was then subcloned into a T/A cloning vector (Promega) to further manipulation. The cDNA was sequenced automatically and completely in both senses.

The cloned human brain latexin sequence was amplified by PCR and subcloned into the *NcoI* and *EcoR1* restriction sites of the prokaryotic expression vector pGAT2 (EMBL, Heidelberg, Germany). Using:

sense 5' CATGCCATGGAAATCCCGCCGACCAACTACCC 3'

antisense 5' CGTCTGCCAAAGGAAGTACAAGTGGGAATAGTAAGAATTCCGG 3'

oligonucleotide primers. Amplification conditions were 30 cycles of 1 min at 95°C, 1 min at 55°C, and 1 min at 72°C. In this vector, the target sequence is fused in frame downstream from the gene encoding glutathione-S-transferase (GST), which is preceded by a polyhistidine-tag (His-tag). Vectors containing the construct were transformed into competent BL21(DE3) *Escherichia coli* cells.

2.3.2.- Molecular cloning of the human PCPA4

The cDNA for *hPCPA4*, containing the full pro-enzyme sequence, was kindly provided by D. I. Smith and H. Huang (Mayo Clinic, Rochester, Minn.). The *hPCPA4* cDNA was cloned into vector pPIC9 of the methylotrophic yeast *Pichia pastoris*. pPIC9 provides the α -mating factor signal for secretion.

2.3.3.- Recombinant overexpression and purification of human latexin

For protein expression, bacterial cultures were grown in 2YT at 37°C to an OD of 0.8. The temperature was subsequently lowered to 18°C, and the culture induced with 1 mM isopropyl- β -D-thiogalactopyranoside (IPTG) for 24h. Cells were harvested by centrifugation, resuspended in PBS buffer, treated with Triton X-100, DNase and RNase, and sonicated. The protein was finally recovered in the supernatant fraction after centrifugation at 17,000 rpm (35,000 g) and loaded onto a Chelating Sepharose Fast Flow (Amersham Biosciences) nickel-sepharose His-tag affinity column equilibrated with Tris-HCl 20 mM, NaCl 0.5 M, imidazole 5 mM (pH 7.9). After the binding step, the column was washed with Tris-HCl 20 mM, NaCl 0.5 M, imidazole 60 mM (pH 7.9) and the His-GST-latexin construct was eluted with a linear gradient (from 0.1 to 1 M imidazole) in Tris-HCl 20 mM, NaCl 0.5 M (pH 7.9).

The eluted protein was dialyzed against PBS buffer before thrombin cleavage of the His-GST fusion through overnight incubation at room temperature (10 unit thrombin per milligram of His-GST latexin). The resulting mixture was loaded again onto the His-tag affinity column to separate the cleaved latexin from the His-GST fusion construct. Latexin was finally purified by size-exclusion chromatography in Tris-HCl 20 mM pH 8.0 and its purity was assessed by SDS-PAGE and mass spectrometry.

2.3.4.- Recombinant overexpression and purification of human *hPCPA4*

Heterologous overexpression and production of recombinant protein was performed in yeast as described previously for other PCPs (Ventura *et al.*, 1999; Reverter *et al.*, 1998b). Accordingly, the *hPCPA4* cDNA was cloned, expressed and secreted to the extracellular medium using vector pPIC9 of the methylotrophic yeast *Pichia pastoris*.

2.3.5.- Preparation of the hCPA4/latexin complex

The hCPA4/latexin complex was produced using fresh preparations of both proteins: the imidazole free His-GST-latexin product obtained after elution from the nickel affinity chromatography and subsequent dialysis, and the hCPA4 product obtained after hydrophobic interaction chromatography, dialysis, and subsequent trypsin activation (at a 200:1 w/w ratio) for 90 min at room temperature. Starting from equimolar amounts of both proteins, the formation of the complex was monitored spectrophotometrically by measuring the loss of hCPA4 activity against the FAPP substrate. The resulting complex was incubated overnight with thrombin (10 units thrombin/mg His-GST-latexin) to cleave off the His-GST tag and subsequently purified by anion exchange chromatography (TSK-DEAE 5PW) using an FPLC-Åkta system with a linear salt gradient from 0 to 60% of 0.8 M ammonium acetate in Tris-HCl 20 mM pH 8.0 (Reverter *et al.*, 2000). After its purity had been checked by SDS-PAGE, the complex was desalted and concentrated to ~6–7 mg/ml by Amicon centrifugal filter devices.

2.3.6.- Latexin inhibitory assays

The K_i values for latexin against MCPs were calculated by pre-steady-state kinetics ($K_i = k_{off}/k_{on}$), using a very sensitive substrate and a continuous photometric assay. The following chromogenic substrates were used: N-(4-methoxyphenyl-azoformyl)-L-Phe-OH for hCPA1, hCPA2, bCPA, haCPA and hCPA4; N-(4-methoxyphenyl-azoformyl)-Arg-OH for hCPB and hTAFI; and N-(4-furylacryloyl)-Ala-Lys-OH for the hCPN assay. The assays were performed in Tris-HCl 50 mM, NaCl 0.1 M pH 7.5 with a substrate concentration of 100 μ M and enzyme concentrations of 2 nM. In this assay, the inhibitor, at a concentration pre-established after different trials with varying amounts of enzyme and inhibitor, is added to a continuously monitored progress curve that follows the generation of product from an enzyme/substrate mixture. Values for k_{on} , k_{off} and subsequently K_i can be derived from mathematical processing of the data.

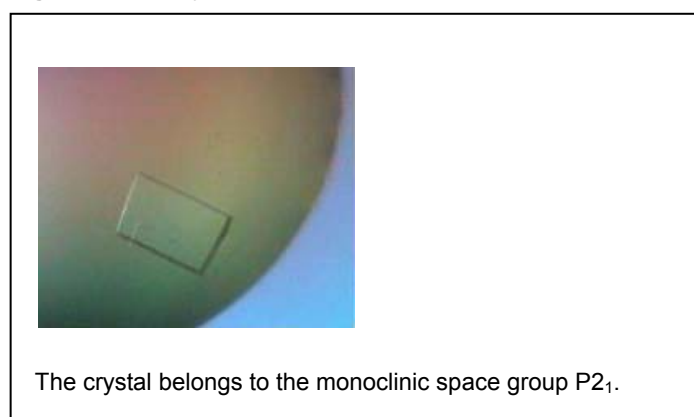
To assess the type of inhibition and to confirm the K_i values obtained previously, steady-state inhibition tests were carried out by measuring the CPA4 activity at

various enzyme/latexin ratios and at different substrate concentrations. The substrate used in this case was FAPP.

2.3.7.- Crystallization of the hCPA4/latexin complex, structure solution and refinement.

The hCPA4/latexin complex was crystallized from hanging drops containing 1 μ l of protein solution (\sim 7 mg/ml), 1 μ l of reservoir solution 2-methyl-2,4-pentanediol 40% (MPD) / Bis-Tris 0.1 M pH 6.5, and 0.2 μ l acetone 40 % after about 4 days at 20 $^{\circ}$ C (Fig. 2.1). The presence of both proteins in the crystals was assessed by mass spectrometry. The crystals were flash-cryogenized in liquid N₂ directly out of the drop.

Figure 2.1.- A crystal of the hCPA/latexin complex



The structure of the complex was solved by a combination of Patterson search and multiple-wavelength anomalous diffraction (MAD) at the zinc absorption K-edge. To this end, three diffraction datasets were collected at 100K from a single crystal on a marCCD 225 detector at beamline ID23–1 at the European Synchrotron Radiation Facility (Grenoble, France). The anomalous absorption edge was precisely determined by a fluorescence scan, with which the wavelengths corresponding to the absorption maximum (f''_{max}), the inflection point (f'_{min}) and a high-energy remote were chosen. A further high-resolution dataset was collected from the same crystal (Table 2.1), which was later subjected to mass spectrometry analysis. The result of this assay was somewhat noisy but clearly revealed the presence of two peaks at 25,312 and 34,649 Da, compatible with the presence of both the inhibitor and the mature protease within the crystal. The measured crystal belongs to the monoclinic space

group P2₁, harbors two complexes per asymmetric unit and diffracts beyond 1.6 Å resolution.

Data were processed, scaled, merged, and reduced with MOSFLM v. 6.2.2. (Leslie, 1999) and SCALA from the CCP4 suite (CCP4; 1994). To calculate initial phases, a Patterson search was performed with AMoRe (Navaza, 1994) employing the coordinates corresponding to the protein part of the active enzyme (excluding the catalytic zinc cation) of the hPCPA4 structure and diffraction data in the 12 – 4.0 Å resolution range. A single solution was encountered for two molecules in the asymmetric unit at 10.8, 90.3, 325.8, 0.0302, 0.0001, 0.0652 and 149.6, 89.2, 146.1, 0.9697, 0.0451, 0.5563, respectively. This solution gave an overall CC_F of 40.6% and a crystallographic R_{factor} of 45.9% (second-highest unrelated peak, CC_F 25.1%, R_{factor} 50.7%). The rotated and translated model was refined against the 1.6 Å resolution dataset and the position of the two catalytic zinc cations was determined by difference Fourier synthesis. These positions were then used to compute experimental phases to 2.0 Å using the three datasets of the MAD experiment and program MLPHARE within CCP4. These phases were combined (σ_A weighted) with those from the positioned and refined hCPA4 moieties (also computed up to 2.0 Å resolution) and subjected to a density modification step (with DM within CCP4) under two-fold averaging and phase extension to the full resolution of the 1.6 Å high-resolution dataset.

This produced an electron density map of excellent quality, which enabled straight forward tracing of the two complexes. Subsequently, manual model building alternated with crystallographic refinement until the final model, featuring protein residues Ser3 to Leu308 of each of the two hCPA4 molecules (chains A and C) and residues Met1 to Lys217 of the two latexin molecules (chains B and D), was obtained. The polypeptide chains were well defined by electron density, except for loops $\beta 1\beta 2$ (centered on Pro46B and Pro46D) and loops $\beta 3\beta 4$ (around Gly80B and Gly80D) of the latexin molecules, which are somewhat flexible at these points. All main-chain angles of the polypeptide chains were placed in permitted regions of the Ramachandran plot. Ser199 of both hCPA4 polypeptide chains A and C showed values close to those in the zymogen, though within a generously allowed region ($\Phi = 156$, $\Psi = -8$ in Ser199A; $155/-6$ in Ser199C). Some residues were modeled with double occupancy of their side chains. One N-acetylglucosamine module (Nag901A and Nag901C) was attached to Asn148 N $\delta 2$ of each hCPA4 moiety. A free-standing valine residue was found at each of the hCPA4 active sites (Val998A and Val998C).

Disulphide bonds and cis-peptide bonds of hCPA4 were as in the zymogen. Each latexin chain showed an additional cis-peptide bond (Ile122-Pro123). One thousand one hundred and fifty-three solvent molecules (Hoh1W–Hoh1153W), 12 MPD molecules (Mpd1L–Mpd12L) and an acetone molecule (Acn13L) were also identified.

2.3.8.- Miscellaneous

The structures of both hPCPA4 and the hCPA4/ latexin complex have two copies in their respective asymmetric units. These copies are structurally equivalent, with an r.m.s.d. of 0.29 Å for the two hPCPA4 molecules (suffixes A and B) and of 0.27 Å for the two complexes (protease moieties are A and C and the inhibitor chains are B and D). Accordingly, discussion will consider only molecule A of hPCPA4 (residues ArgA4A to GlnA99A of the PD and Glu1A to Tyr309A of the enzyme moiety; nomenclature according to Coll *et al.*, 1991) and the complex consisting of molecule A of hCPA4 (Ser3A to Leu308A) and molecule B of latexin (Met1B to Lys217B). Figures were prepared with the TURBO-Frodo, GRASP, and SETOR programs; superimpositions and searches for structural similarity were performed with SIM, TURBO-Frodo and DALI; and multiple sequence alignments, with MULTALIN and ESPRIPT 2.2. Of particular help was program DBAli2.0 (Marti-Renom *et al.*, 2001). Close contacts and interaction surfaces were calculated with CNS. The final coordinates of hPCPA4 and the hCPA4/latexin complex have been deposited with the Protein Data Bank (PDB access code 2BOA and 2BK7 respectively) at the European Bioinformatics Institute (www.ebi.ac.uk/msd), Hinxton (UK).

Table 2.1.- Crystallographic statistics on data collection and refinement

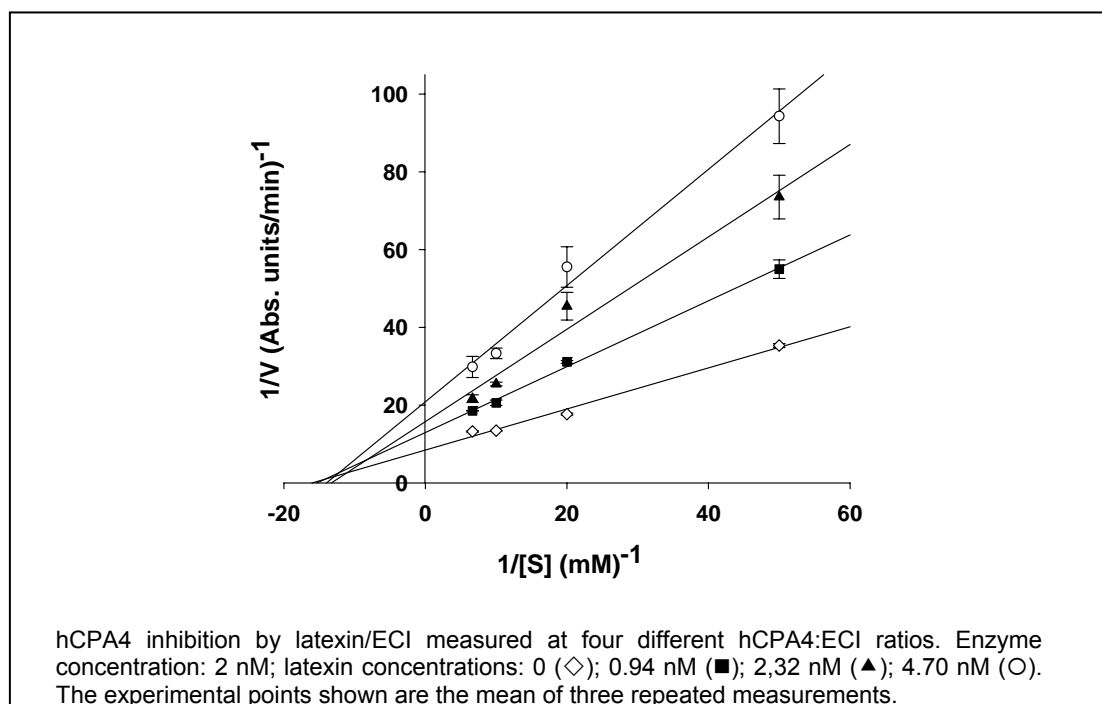
Dataset	hCPA4/latexin (high resolution)	hCPA4/latexin (f'min)	hCPA4/latexin (f''max)	hCPA4/latexin (remote)
Space group	P2 ₁		P2 ₁	
Cell constants (a, b, c; in Å; β in °)	78.9, 96.7, 92.4, 99.6		78.9, 96.7, 92.4, 99.6	
Wavelength (Å)	1.0067	1.2828	1.2818	1.2770
Number of measurements	678,157	350,563	350,066	350,334
Number of unique reflections	179,401	92,181	92,184	92,185
Resolution range(Å) (outermost shell) ^a	48.3-1.60 (1.69-1.60)		60.6-2.00 (2.11-2.00)	
Completeness (%)/anom.completeness (%)	99.8 (99.3)		99.8 (99.2)/98.6 (96.8)	
R _{merge} ^b	0.068 (0.257)	0.062 (0.135)	0.057 (0.107)	0.057 (0.083)
R _{anomalous} ^c	-	0.044 (0.101)	0.044 (0.085)	0.042 (0.063)
Average intensity (< I>/σ(< I>))	11.8 (4.8)	10.6 (5.4)	13.2 (7.9)	13.8 (10.8)
B-factor (Wilson) (Å ²)	13.5	19.6	19.5	
Average multiplicity	3.8 (3.7)		3.8 (3.8)	
Estimated values of f/f' (in electrons)	-	-9.2/2.3	-7.3/4.7	-4.6/4.9
Resolution range used for MAD phasing (Å) ^d	-		48.3-2.00	
Mean figure of merit (fom) MLPHARE/phase comb./DM ^e	-		0.25 (2.0)/0.70 (2.0)/0.83 (1.6)	
Resolution range used for refinement (Å)	48.3-1.60			
Number of reflections used (test set)	178,706 (678)			
Crystallographic R _{factor} (free R _{factor}) ^f	0.149 (0.178)			
Number of protein atoms (assymetric unit)	8,463 ^h			
Number of solvent molecules/ions/other molecules	1,153/2 (Zn ²⁺)/12 (MPD), 1 (acetone)			
R.m.s. deviation from target values				
Bonds (Å)	0.011			
Angles (°)	1.35			
Bonded B-factors (Å ²) main-chain/side chain	0.88/2.10			
Average B-factors for protein (Å ²)	16.1 ^g			

2.4.- RESULTS

2.4.1.- Inhibition profile of latexin

The results of inhibition studies of mature hCPA4, obtained after tryptic activation, with human latexin are shown in figure 2.2. The profile corresponds to non-competitive inhibition, since addition of latexin affects only V_{\max} and not K_m . However, this is valid for synthetic, small substrates like those used here and in most CP activity determinations. The presence of a huge inhibitor molecule on top of the active site might preclude not only the binding of natural substrate molecules, but even their docking. In this case, the action of latexin in a cellular context is more difficult to define and its mechanism is better considered a mixed inhibition. Since obtaining kinetic data of CPs with natural peptides has always been difficult, we suggest that, with the substrates and the data presented here, the definition of the action of latexin on CPs as tight-binding non-competitive inhibition is adequate.

Figure 2.2.- Lineweaver-Burk plot for hCPA4 inhibition by latexin



All vertebrate A/B-type MCPs tested were inhibited. The kinetic inhibition constants (K_i), calculated by pre-steady-state analysis (Table 2.2) and confirmed by parallel

steady-state measurements, are in the nanomolar range and similar to those obtained with the LCI invertebrate inhibitor from leech (Bayés *et al.*, 2003). Thus, K_i values calculated for bCPA, hCPA1, hCPA2, hCPB, hTAFI and hCPA4 ranged from 1.1 to 3.5 nM. However, human latexin fails to inhibit not only members of the E/N class like hCPN or duck (d)CPD domain 2, but also an invertebrate A/B-MCP from the cotton bollworm, *Helicoverpa armigera*, haCPA, despite 30% sequence identity of its zymogen with hPCPA4 (A. Bayés, personal communication). The invertebrate enzyme was more efficiently inhibited by PCI than by LCI, unlike other vertebrate CPs (Bayés *et al.*, 2003). This suggests further diversification inspecificity between vertebrate and invertebrate A/B MCPs. The inhibition results reported here are in accordance with those described for rat latexin against various MCPs. Those studies, however, established an IC_{50} value 100-fold larger for bCPB and 10-fold larger for mast cell CP than for bCPA (Normant *et al.*, 1994b). It was also shown that rodent latexin does not inhibit E/N CPs like mouse (m)CPH and hCPM, nor other metalloproteases like gluzincins (aminopeptidases, neprilysin, dipeptidyl peptidase leukotriene A4 hydrolase) and serine proteases (trypsin, chymotrypsin, elastase, and yeast CPY) (Normant *et al.*, 1994b).

Table 2.2.- Inhibition of MCPs by human endogenous CP inhibitor

MCP	Inhibition constant K_i, nM
bCPA	1.2±0.2
hCPA1	1.6±0.2
hCPA2	3.5±0.3
hCPB	1.1±0.1
hCPA4	3.0±0.3
hTAFI	1.8±0.3
haCP	No inhibition
hCPN	No inhibition

TAFI, Thrombin-activable fibrinolysis inhibitor. Latexin does not inhibit dCPD2 and hCPE (L.D. Fricker, personal communication).

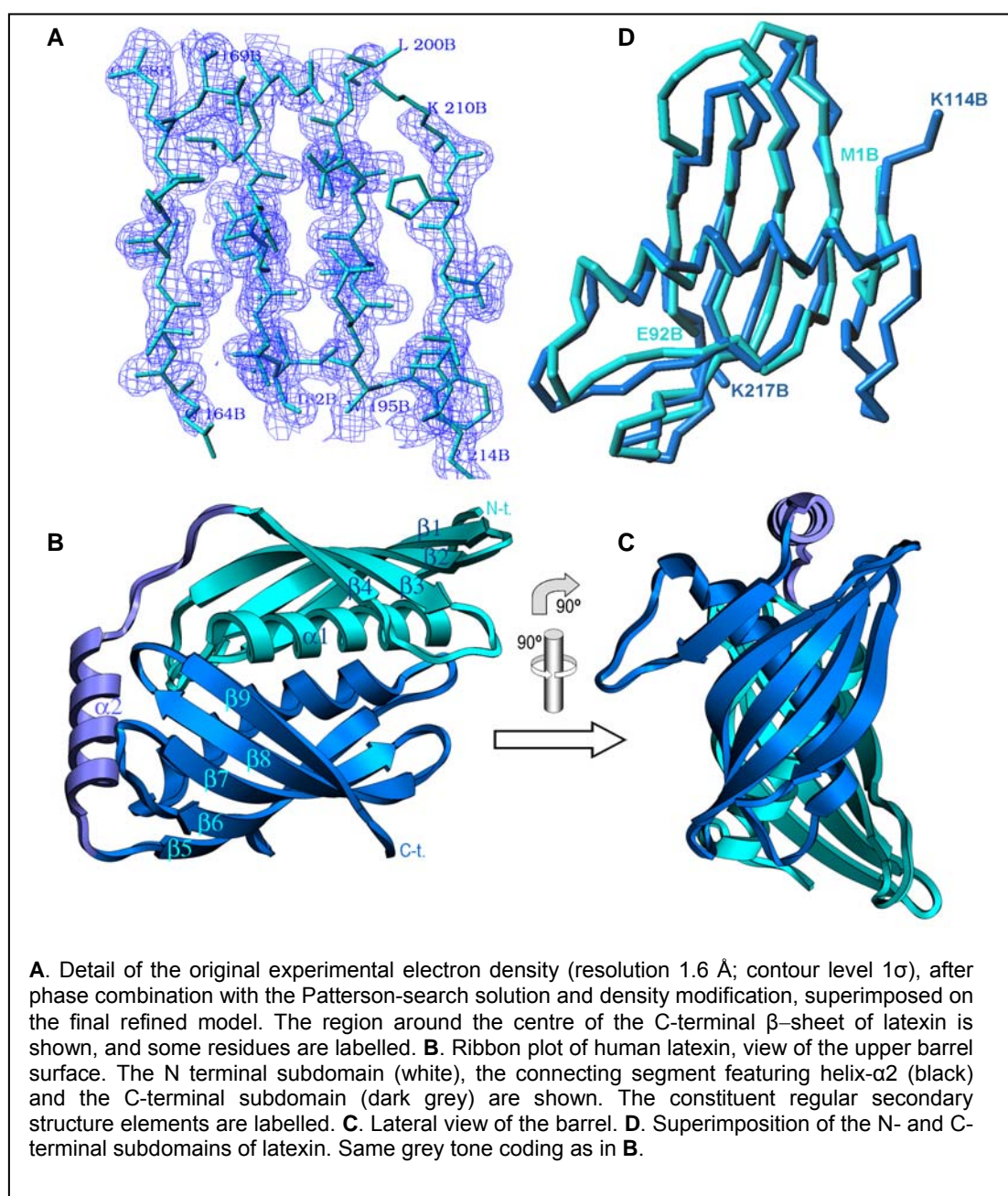
2.4.2.- The human latexin structure and comparison with the mouse orthologue

Human latexin has an elongated shape, maximum dimensions of about 65x40x40 Å (Fig. 2.3) and an α/β topology within a single 222-residue polypeptide chain. The latter five residues are not visible in the electron density maps. The protein can be split into two subdomains, an N terminal subdomain NTS; from Met1B to Glu92B; (Fig. 2.3B, D) and a C-terminal one (CTS; Lys114B–Lys217B). The sub-domains share an equivalent fold despite minute sequence similarity (14% identity), such that 87 of their C_{α} atoms can be superimposed with an r.m.s.d. of 2.1 Å (Fig. 2.3D). The subdomains comprise a long α -helix (α_1 in NTS, α_3 in CTS) and a strongly twisted four-stranded antiparallel β -sheet of simple up-and-down connectivity (β_1 – β_4 in NTS, β_6 – β_9 in CTS), which accommodates the helix on its concave side (see Fig. 2.3B). The topology of this ensemble of regular secondary-structure elements recalls a left hand, with the helix mimicking the thumb and the β strands the four fingers (Fig. 2.3D). The β -sheet of CTS contains an additional strand, β_5 , which runs antiparallel to β_6 and is sequentially N terminal to the cognate α -helix. The side chains of both helices, α_1 and α_3 , are aliphatic, aromatic or uncharged, except for one residue (Arg12B and Glu132B, respectively). This causes these structural elements to interact with their cognate β -sheets mainly through hydrophobic contacts. The subdomains are bridged by a connecting segment running along the molecule surface (Gly93B–Met113B). It is mainly constituted by an α -helix, α_2 , which is predominantly charged, with 6 out of 13 residues with formal charges (Fig. 2.8). The quaternary structure of latexin results from packing the two subdomains against each other through the α -helices. These run antiparallel to each other, with their axes ~ 7.5 Å away and rotated $\sim 50^\circ$ relative to each other around the point at which Ile21B of α_1 and Gly145B of α_3 meet. This architecture entails that both curved β -sheets are placed on the molecular surface, together forming a flat, incomplete β barrel with two main surfaces, an 'upper' and a 'lower' one. Each surface is shaped by the central parts of helices α_1 and α_3 on opposite faces and the beginning and ends of the β strands (Fig. 2.3B).

The structure of unbound mouse latexin has been reported recently (Aagaard *et al.*, 2005). The rodent orthologue displays the same number of residues as the human form and shares 85% sequence identity (Fig. 2.8). The C_{α} backbones of the two proteins fit well and may be superimposed with an r.m.s.d. of 0.63 Å over 213 residues out of the 217 amino acid positions defined by electron density in the two

structures. The mouse structure reported has three additional residues preceding the N terminal methionine due to cloning strategies. The high similarity of the unbound and the bound forms suggests that there is no induced fit of the inhibitor upon complex formation. This is akin to the structure of the target protease moieties, already preformed in the zymogens. The inhibitory loop of latexin and, in particular, Gln190B, is in the same conformation in both structures, as are all the major interaction points with the protease.

Figure 2.3.- Structure of human latexin



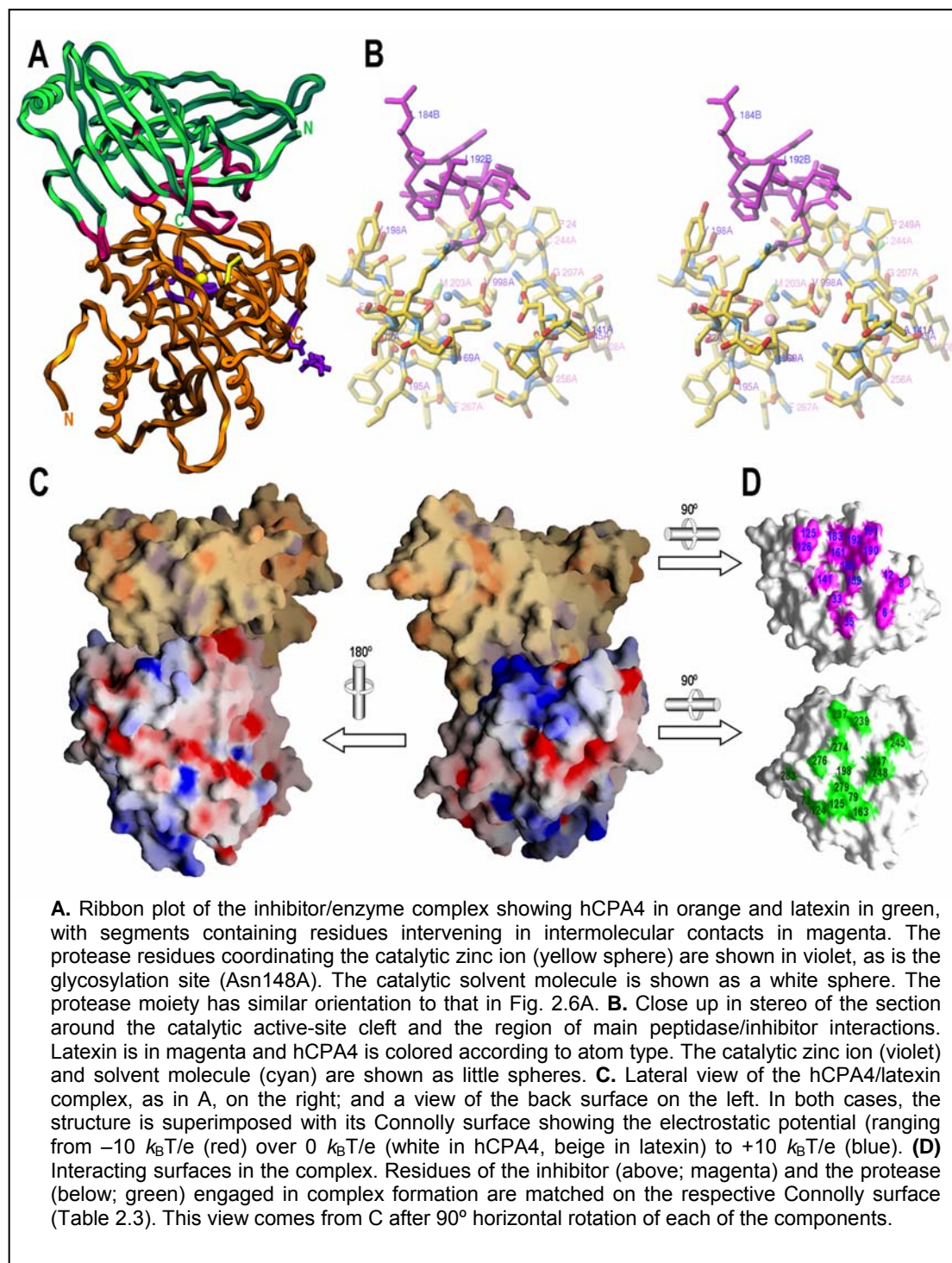
Chapter 2

Minor details worth mentioning that could be attributable to complex formation include position 125B, whose side chain rotates about 260° around its χ_1 angle in the transition from the unbound mouse to the complexed human structure to prevent collision with the side chains of Arg124A and Trp73A of the protease. Glu191B rotates 90° around χ_1 , which brings it in contact with the protease main chain at Glu163A N. The remainder is very similar in the two structures and significant deviations can only be found at the tip of loop $\beta_3 \beta_4$ (2.4 Å at Gly80B C α) and at the beginning of CTS (1.5 Å at Lys114B C α), possibly due to crystal contacts.

2.4.3.- Structural details of the hCPA4/latexin complex

In the human complex structure, the inhibitor is placed on top of the funnel bordering the active site and traps loop $\beta_{12} \alpha_{12}$ (encompassing residues Asp273A–Pro282A) of the protease through its lower barrel surface at the subdomain interface (Fig. 2.4). This lower surface of the inhibitor has a greater indentation than the upper, shallower surface. This results in a cavity of 235 Å³, once the complex with hCPA4 is formed, which is partially occupied by an MPD molecule. The complex buries a total surface of 2,340 Å² at the protein interface with a shape complementarity of 71%, a value which indicates a good fit between interacting surfaces (Lawrence & Colman, 1993). The surface is rather large if compared with typical protease-inhibitor complexes, which span between 1250 and 1750 Å² (Janin & Chothia, 1990), and results in several solvent and MPD molecules being enclosed between the protein molecules. However, complex formation entails rather few direct interactions, as only 48 intermolecular contacts below 4 Å are observed, among them 13 hydrogen bonds and 7 hydrophobic interactions (Table 2.3). The number of hydrogen bonds is within the range of the numbers of previously described protease-inhibitor complexes, which, however, are joined through roughly half the surface area. The interaction involves 15 residues of both hCPA4 and latexin, which again is unlike other protease-inhibitor complexes, in which, characteristically, a similar number of inhibitor residues (10 – 15) interact with twice as many protease residues (Janin & Chothia, 1990). These peculiarities can be attributed to the rather shallow interacting surface of latexin and to the protease mainly interacting through the border of the funnel that gives access to the active site.

Figure 2.4.- The hCPA4/latexin complex



The most interesting feature of the interaction is that none of the latexin termini is involved in protease inhibition. This is in contrast to what has been described for other complexes of MCPs with exogenous inhibitors like PCI and LCI, in which a conserved seven-residue C-terminal segment contributes to inhibition in a substrate-like binding manner. Substrate-like binding occurs in other complexes through the N-

Chapter 2

terminal tails of inhibitors, as observed in metalloprotease/inhibitor complexes of matrix metalloproteinases (MMPs) with tissue inhibitors of metalloproteinases (TIMPs) and of serralyins with their cognate inhibitors (Reverter *et al.*, 2000; Rees & Lipscomb, 1980; Baumann *et al.*, 1995; Gomis-Rüth *et al.*, 1997b). Rather, latexin inhibition is mainly caused by an inhibitory loop provided by the central part of the β -sheet of the CTS, reminiscent of the inhibition through PD in the zymogen. This loop is shaped by the end of strand $\beta 7$, the beginning of $\beta 8$ and the connecting loop $\beta 7 \beta 8$, which penetrates the protease moiety. This behavior is also found in cystatins, inhibitors of cysteine proteases which also employ a β ribbon structure for inhibition, as seen in the steffin/papain complex (Stubbs *et al.*, 1990) (see below). Gln190B N ϵ 2, at the tip of loop $\beta 7 \beta 8$, most closely approaches the active site (5.8 Å away from the catalytic zinc ion).

Table 2.3.- Interaction scheme between hCPA4 and latexin

hCPA4 latexin	Arg71 N η 2	Trp73 C ζ 3	Arg124 N ϵ	Leu125 C δ 2	Glu163 N	Tyr198 O η	Glu237 O ϵ 2	Gln239 C β	Thr245 O	Val247 O	Tyr248 O η	Thr274 O γ 1	Thr274 O	Thr276 O γ 1	Thr276 C γ 2	Phe279 C ζ	Ala283 N
Thr6 C γ 2								3.79									
Tyr8 N									2.90								
Arg12 N η 1										2.94							
Glu33 O ϵ 2												3.17		2.79			
Gln35 N ϵ 2							2.81										
Asn125 O			2.94														
Phe126 O																	3.67
Phe126 C ϵ 2		3.77															
Trp141 C η 2																	
Lys159 N ζ													3.04				
Val161 C γ 1				3.90													
Leu183 C δ 1				3.68													
His185 N δ 1						2.68											
His185 C ϵ 1																	3.37
Gln190 O	2.83																
Gln190 N ϵ 2											2.91						
Gln190 O ϵ 1	2.89																
Glu191 O ϵ 1					2.91												
Ile192 C γ 2				3.95													

The distance (in Å) is shown at the intersection between interacting atoms. There are a total of 47 contacts below 4 Å, with 13 hydrogen bonds and seven hydrophobic interactions.

In the latter, a free valine (Val998A), probably left behind after a proteolytic event during purification, is encountered in the specificity pocket (Fig. 2.4B). This finding was described previously for pPCPA1 (Guasch *et al.*, 1992). Strikingly important for complex stability is the interaction of Gln190B with Arg71A (Table 2.3). The latter is a basic residue present throughout A/B-MCPs and is required for maintenance of the crucial salt bridge with AspA36 (AsnA36A in PCPA4) of the PDs of PCPs (see

above). This basic residue is absent in N/E forms, which are not secreted as pro-enzymes. Gln190B, furthermore, interacts with Tyr248A in 'down' conformation (see below). This tyrosine is a residue involved in hydrogen bonding of the P₁ amide nitrogen of substrates through its O_η group. The position of Gln190B within the inhibitory loop is maintained by an intramolecular interaction with the Nε2 atom of the preceding His185B (2.8 Å). This histidine further interacts via an edge-to-face van-der-Waals contact with the side chains of Phe279A and Tyr198A of the protease moiety. Among the structurally important interactions of the inhibitory loop are also the one established by Glu191B, approaching Glu163A, and by both Ile192B and Leu183B, contacting Leu125A. The position of the inhibitory loop is further fixed through intramolecular main-chain interactions with the neighboring strands β6 and β9, as well as with loop α3β6. Here, Val161B makes a hydrophobic interaction with Leu125A, and Lys159B a hydrogen bond with Thr274A. Further complex stabilization is provided by latexin loop β5α3 of the CTS, which comes close to the funnel border of the protease. In particular, the tip of the loop (around Phe126B) contacts Arg124A, Trp73A and, weakly, Ala283A (Table 2.3). Two other regions, belonging to the NTS of latexin, participate in contacts, the beginning of helix-α1 and the preceding residue, and the adjacent central part of strand β1. Here we observe six interactions (Tyr8B N-Thr245A O; Arg12B N_η1-Val247A O; Thr6BC_γ2-Gln239A C_β; Glu33B O_ε2-Thr274A O_γ1 and -Thr276A O_γ1; Gln35B N_ε2-Glu237A O_ε2). Last but not least, Trp141B makes an additional isolated hydrophobic contact with the methyl group of Thr276A.

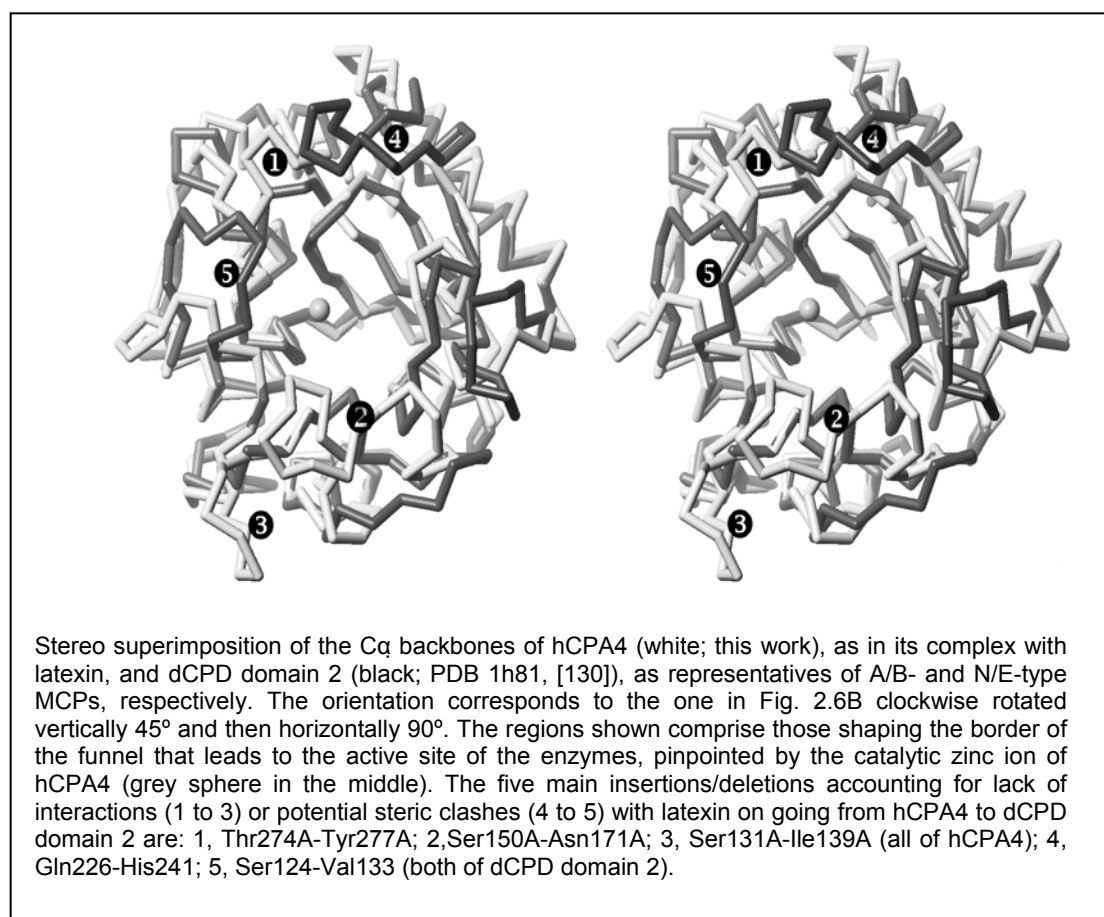
The mode of latexin inhibition described may explain why those CPs with the characteristic PD, the A/B-MCPs, are inhibited by latexin, whereas CPs without PD, e.g. N/E-MCPs like hCPM and dCPD domain 2, are not inhibited. In general, five loops in the regions furnishing the funnel border can be distinguished for either the A/B or the N/E-MCPs, and potential steric clashes and lack of interactions may justify the selective inhibitory profile of latexin.

The absence of a long insertion in hCPA4, Ser150A–Asn171A, reduced to Ala142–Ser149 in hCPM and Gln150–Pro157 in dCPD domain 2 (Fig. 2.8) (Gomis-Rüth *et al.*, 1999; Reverter *et al.*, 2004), and the adjacent Ser131A–Ile139A, assisting in hCPA4 to strut the previous loop and short-circuited to a single residue in hCPM (Asn131) and dCPD domain 2 (Asn139), dramatically diminish the interaction possibilities. The same occurs in the loop region Thr274A–Tyr277A in hCPA4, absent in hCPM and dCPD domain 2.

Chapter 2

However, the most important features are two characteristic loop insertions of E/N-MCPs, Lys221–Asn233 (hCPM) or Gln226–His241 (dCPD domain2), as well as Val116–Ser124 (hCPM) or Ser124–Val133 (dCPD domain 2), which would sterically clash with the inhibitor (Figs. 2.4, 2.5). In the case of haCPA, shown not to be inhibited by latexin despite being an A/B MCP (see above), the only structural difference observed with hCPA4 is an insertion in the invertebrate protein between the positions equivalent to Leu271A and Gly278A of the human counterpart. This results in a loop that is four residues longer and that would collide with latexin helix- α 3.

Figure 2.5.- The funnel rims in MCPs

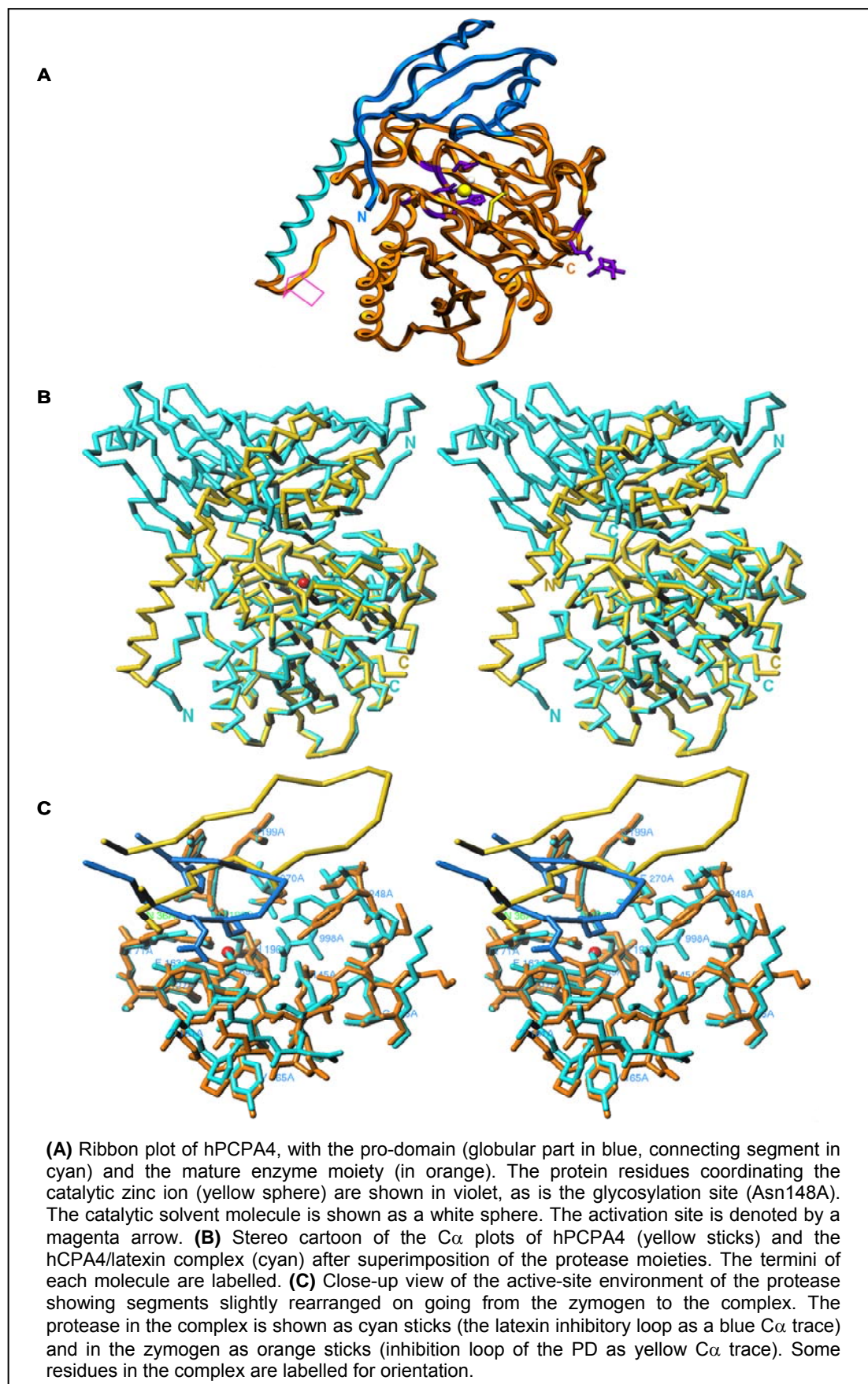


2.4.4.- Differences of hCPA4 between hPCPA4 and the latexin complex

The overall structure of hCPA4 in the zymogen and in the inhibitor complex is, in general, equivalent (303 topologically equivalent C_α atoms from Phe7A to Leu308A; r.m.s.d. 0.40 Å; Fig. 2.6B, C). However, closer inspection revealed some noteworthy structural differences. First, the N-terminal four-residue segment of the mature enzyme is rotated (mainly around Asn6A C-C_α) away from its position in the zymogen such that the new N terminus, Ser3A N, is anchored to the molecular surface by a triple interaction, with Arg28A Oδ1 (3.2 Å) and Oδ2 (2.8 Å) and with Asn24A Oδ1 (2.9 Å). Other differences arise, on the one hand, from the different interactions made with the PD and latexin and, on the other, from the left-behind free valine (Val998A) mentioned above. This valine, present in the complex but not in the zymogen, causes a rearrangement of the side chain of Tyr248A from an 'up' or 'open' conformation (Coll *et al.*, 1991) to a 'closed' or 'down' conformation, as observed in (P)CPs with an 'occupied' pocket (Gomis-Rüth *et al.*, 1999; Guasch *et al.*, 1992; García-Sáez *et al.*, 1997). This puts Tyr248A Oη in a position to interact with both a carboxylate oxygen group (2.5 Å) and the free amino group of valine (3.0 Å), besides the mentioned interaction with Gln190B of latexin (see above). This rearrangement of Tyr248A entails also a displacement of the whole polypeptide chain from Val247A to Ser251A (maximal at Pro249A C_α with 1.0 Å; Fig. 2.6C). In a concerted movement, the adjacent segment Gly153A–Val164A is also shifted (maximal at Gly155A C_α with 1.6 Å). However, this second rearrangement may also be influenced by a contact with the N terminus of a crystallographically related protease moiety.

Other minor changes include the side chain of Arg71A, slightly reallocated from the position adopted to exchange a hydrogen bond with AsnA36A in the zymogen to interact with Gln190B of the inhibitor (see above). Also, Arg127A replaces its hydrogen bond with the main chain carbonyl of Glu163A in the zymogen by a salt bridge with the carboxylate group of the free valine. Finally, the side chain of Arg145A is also rearranged to contact this carboxylate group (Fig. 2.6C).

Figure 2.6.- hPCPA4 and comparison with the hCPA4/latexin complex



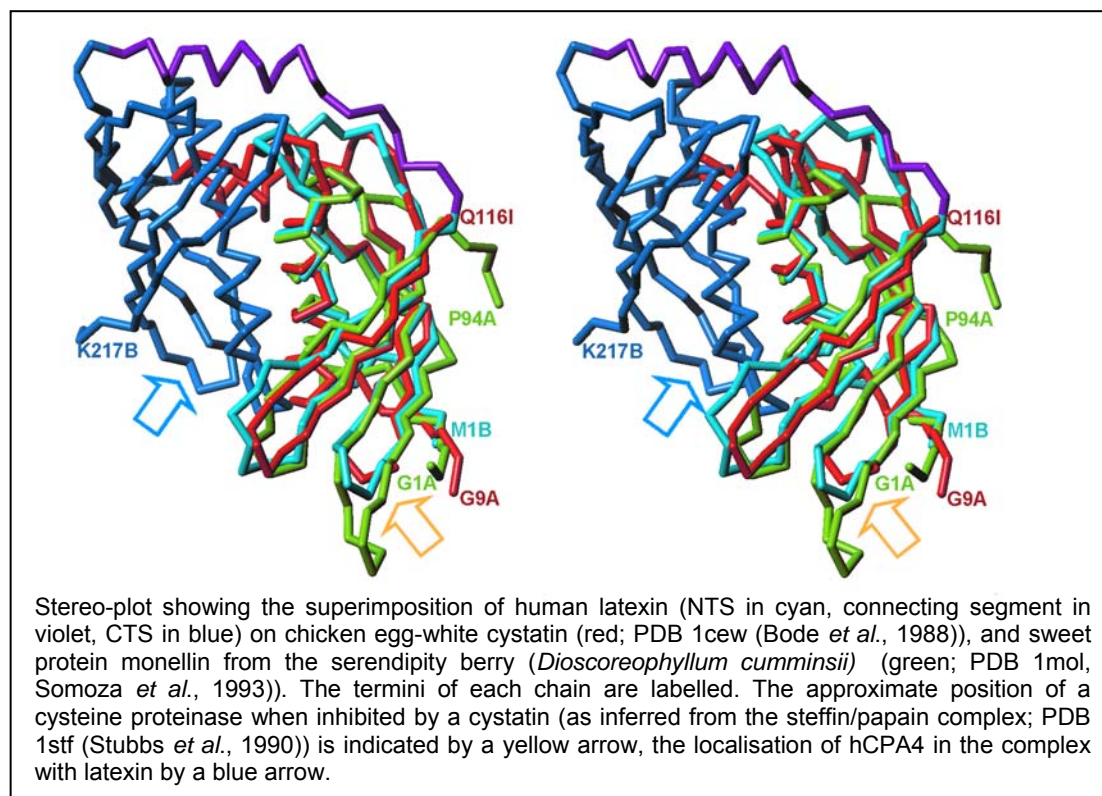
2.4.5.- Structural similarities of latexin

Interestingly, the overall topology of latexin resembles the structure of serralyisin inhibitors (Baumann *et al.*, 1995; Hege *et al.*, 2001), which target the serralyisin family of bacterial metalloendopeptidases (MEPs), part of the metzincin clan of MEPs (Gomis-Rüth, 2003). In these cases, the endogenous inhibitors are also β barrels of simple antiparallel up-and-down connectivity, but they are shaped by eight strands and closed, with a helix inserted between the third and fourth strands. All in all, these bacterial inhibitors are about half the size of latexin and their inhibition mode is different. Latexin also has some similarities with cystatins in the way it functions and bioinformatic searches have de facto identified chicken egg-white cystatin PDB 1cew (Bode *et al.*, 1988) and monellin PDB 1 mol (Somoza *et al.*, 1993) as the closest structural homologues to latexin (Fig. 2.7). Monellin is the main agent responsible for the sweetness of plant berries and is among the sweetest substances that have been identified so far. It bears close structural similarity to cystatins. The structural similarity of latexin with the latter is, however, limited to less than half of the complete molecule, as it encompasses only one of the two subdomains, and entail slow sequential similarity. In particular, 87 C_{α} atoms of cystatin and latexin can be aligned with an r.m.s.d. of 2.2 Å (16% sequence identity), as can 81 C_{α} atoms of monellin and latexin (r.m.s.d. 2.4 Å; sequence identity 14%). Closer inspection reveals that the structural unit consisting of the α -helix and the surrounding β -sheet can be superimposed.

One main difference is that cystatin displays an additional-helix between the strands equivalent to $\beta 2$ and $\beta 3$ of latexin, while monellin shows shorter strands equivalent to $\beta 1$, $\beta 2$ and $\beta 3$. Inhibition of cysteine proteinases through cystatins, as observed in the steffin/papain complex (Stubbs *et al.*, 1990), occurs through the tips of the fingers following the left-hand similitude in the fold, through the loop connecting the strands equivalent to $\beta 1$ $\beta 2$ and $\beta 3$ $\beta 4$ of latexin NTS (or $\beta 6$ $\beta 7$ and $\beta 8$ $\beta 9$ considering CTS), and through the N terminus. In contrast, latexin inhibits by elements participating in the interface between the two subdomains located on the opposite face of the inhibitor (Fig. 2.7).

Furthermore, slight conformational differences, mainly due to deletions/insertions, of both previously mentioned loop pairs of NTS and CTS suggest that these should not inhibit cysteine proteinases, a finding experimentally proved in the case of papain [our unpublished results].

Figure 2.7.- Structural similarities to latexin



Interestingly, two tandem copies of cystatin-like repeats have been predicted for inhibitors of metzincins belonging to the astacin and adamalysin/ADAM families. In particular, the endogenous protein inhibitor of carpnephrosin, an astacin, and two proteins isolated from snake venoms, *Trimeresurus flavoviridis* Habu serum factor and *Bothrops jararaca* BJ46a, inhibitors of snake venom adamalysins/ADAMs, have been described (Tsai *et al.*, 2004). They are all members of the fetuin family of proteins and, as in the case of latexins, are unable to inhibit cysteine proteases despite structural similarity with cystatins. However, no structural information on their mode of inhibition is currently available.

2.4.6.- A novel family of potential A/B-MCP inhibitors

Bioinformatic retrieval within complete and unfinished genome sequence projects has revealed a number of putative proteins which have significant sequence similarity to human latexin (Fig. 2.8). These sequences go beyond mammals to include other vertebrates like frogs, birds and fish. The aligned sequences were compiled from EST libraries from diverse tissues and organs, including kidney, spleen, liver, brain,

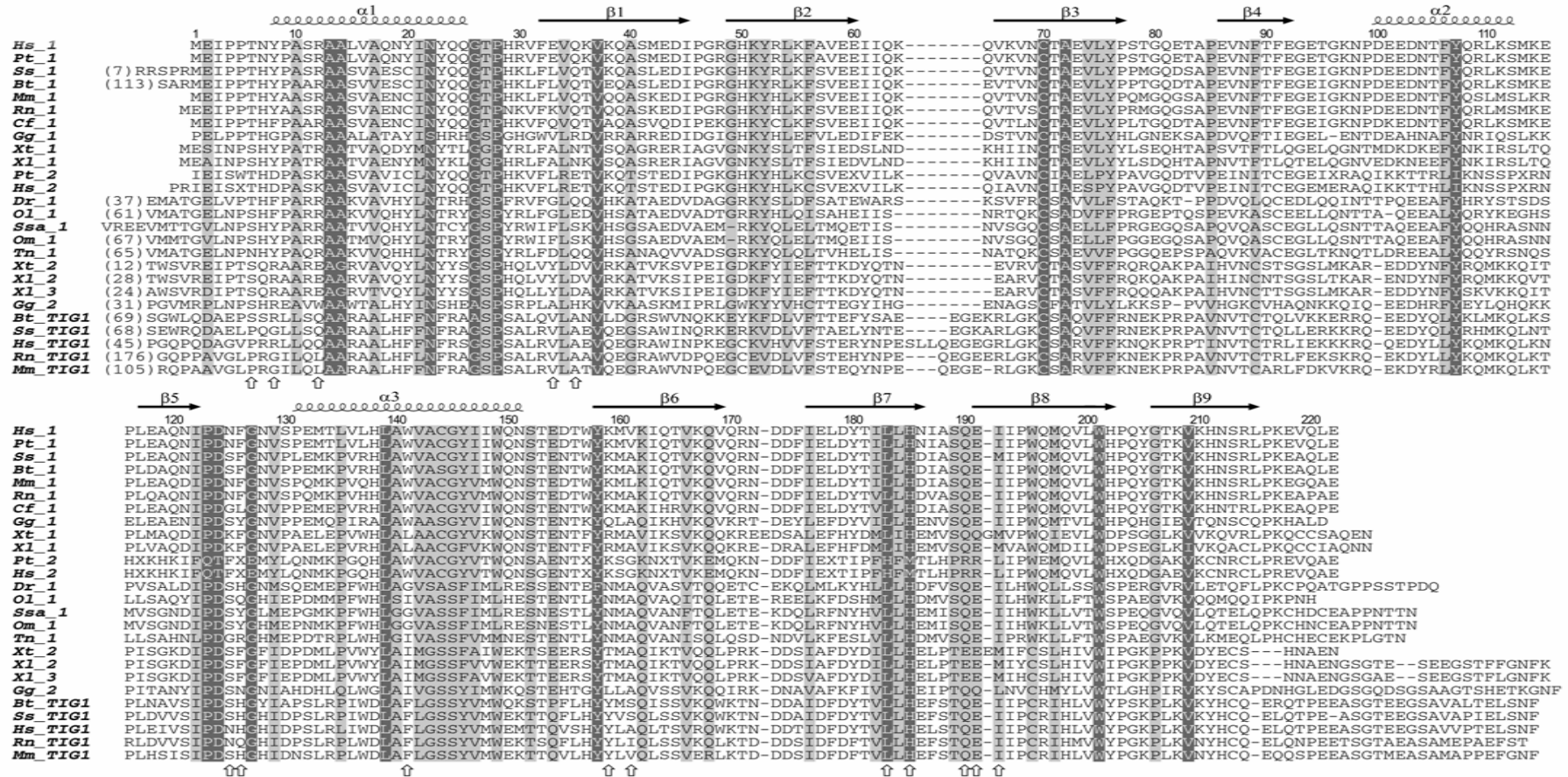
lung, ovary, testis, aorta, cartilage, placenta, thymus, gut and small intestine. Accordingly, the broad tissue distribution strongly suggests temporal and spatial ubiquity in the transcription of these genes. Inspection of the residues mainly engaged in inhibition suggests that they are mostly highly conserved or conservatively substituted. Therefore, it has recently been proposed that these sequences be grouped in the latexin family, which comprises biochemically characterized inhibitors (from human, rat and mouse, grouped into MEROPS inhibitor family I47 (Rawlings *et al.*, 2004) as well as potential MCP inhibitors, and may also harbor structurally related proteins with other functions.

Apart from the closer rodent orthologues, a sequence relative that has been studied at the biochemical level is ovocalyxin-32, a chicken eggshell protein highly expressed and secreted by surface epithelial cells of the isthmus and the uterus during eggshell formation. Its molecular interacting partners have not been discovered, but it has been ascribed to the matrix (Gautron *et al.*, 2001). Further similarity can be found with a protein whose expression is induced by retinoids in skin raft cultures, retinoic acid receptor responder protein 1 (RARRES1 alias tazarotene induced gene 1 protein, TIG1), a putative tumor suppressor gene with various splice variants (Youssef *et al.*, 2004). Retinoids are known to down-regulate the expression of metalloproteases, cytokines and other genes involved in cell proliferation and inflammation through ligand-dependent transcription factors that are retinoic acid receptors (Nagpal *et al.*, 1996). Thus, RARRES1/TIG1 could be a response to hyperproliferative and inflammatory processes like skin diseases, putatively acting against MCPs engaged in inflammation. RARRES1/TIG1 orthologues were found in cattle, pig, rat and mouse, with sequence identities ranging from 28 to 31% of human latexin (Fig. 2.8) in a common stretch of ~220 residues. They show additional N terminal residues that may encode membrane anchors (Aagaard *et al.*, 2005; Nagpal *et al.*, 1996).

Interestingly, multiple sequences very similar to latexin have been found in several organisms (Fig. 2.8). In the African clawed frog there are even three potential paralogues, with 49, 34 and 33% sequence identity with human latexin. These sequences could encode putative latexin isoforms. This is reminiscent of TIMPs, for which four different variants have been reported (TIMP-1 to TIMP-4), with multiple presence in several organisms (Woessner & Nagase, 2000).

Chapter 2

Figure 2.8.- The latexin family. Sequence alignment of members of the latexin family of putative MCP inhibitors.



Sequences were retrieved from SwissProt/TrEMBL (www.expasy.ch), GenBank (www.ncbi.nlm.nih.gov), TIGR (tigblast.tigr.org/tgi), and ENSEMBL (www.ensembl.org). Hs, *Homo sapiens* (human); Pt, *Pan troglodytes* (chimpanzee); Ss, *Sus scrofa* (pig); Bt, *Bos taurus* (cattle); Mm, *Mus musculus* (mouse); Rn, *Rattus norvegicus* (rat); Cf, *Canis familiaris* (dog); Gg, *Gallus gallus* (chicken); Xt, *Xenopus tropicalis* (pipid frog); Xl, *Xenopus laevis* (African clawed frog); Dr, *Danio rerio* (zebrafish); Ol, *Oryzias latipes* (Japanese medaka fish); Ssa, *Salmo salar* (salmon); Om, *Oncorhynchus mykiss* (rainbow trout); Tn, *Tetraodon nigroviridis* (pufferfish). Some sequences are from unfinished genome projects. In these cases, they may still bear errors and include signal and/or pro-peptides; X stands for unidentified amino acids. Hs_2 and Pt_2 are probably processed pseudogenes. Similar though incomplete sequences could be also identified in opossum, killifish and Japanese pufferfish. Positions conserved in more than 90% of the sequences are shown in white over dark grey; those with a maximum of three different amino acid types or with conservative substitutions (all apolar; i.e. aliphatic or aromatic) appear with a light grey background. The number of additional N-terminal protein residues is indicated in parenthesis if they occur. Amino-acid numbering and regular secondary structure elements correspond to the human latexin structure. Cringles stand for α -helices (labelled $\alpha 1$ to $\alpha 3$) and arrows for β -strands ($\beta 1$ to $\beta 9$). Small vertical arrows denote residues engaged in interactions with hCPA4 (see table 2.3).

2.5.-DISCUSSION

There are some similarities in the way latexin and the PD block the hCPA4 enzyme moiety, basically a comparably large interface area and relatively few contacts. The main interaction area includes the lower barrel surface of latexin around the CTS β -sheet. This arrangement is somewhat reminiscent of the outer face of the four-stranded β -sheet of the globular part employed by PD for inhibition. In both cases, active-site blocking occurs with participation of an inhibitory loop provided by the central part of the mentioned β -sheets. However, while latexin completely blocks access to hCPA4, thus inhibiting in a non-competitive manner, the activation segment in hCPA4 does not shield wholly the active-site cleft, which smaller peptide substrates can reach, as occurs in other type A PCPs, though not type B ones (García-Sáez *et al.*, 1997).

The striking structural similarity of latexin subdomains and cystatins suggests that the latter can be seen as evolutionary ancestors. Genetic duplication of such a cystatin module, including the incorporation of a 41-residue linker in between and subsequent sequential divergence may have given rise to the current structure of latexin. The latter, however, employs different regions to perform the common task of inhibiting proteases.

Latexins seem to be restricted to vertebrates and have been found in organisms ranging from fish and frogs to mammals and other higher vertebrates. No sequences could be retrieved from invertebrates or other less complex organisms. Invertebrate A/B MCPs and E/N subfamily MCPs are not inhibited, as happens with several endopeptidases studied. Our molecular studies show that structural determinants simultaneously confer vertebrate linked specificity to latexins and sufficient variability to strongly inhibit all vertebrate forms tested, even across the species barrier.

There is a parallel evolution in the studies of A/B MCPs and MMPs. These MEPs were found because of their non-specific degradation of extracellular matrix components in vertebrates, as occurs during tissue turnover and developmental and metamorphic processes, but also during tissue destruction in pathologies such as cancer and arthritis. They were located in the extracellular space where these events take place. However, decades of work led to a re-examination of the roles of MMPs as they were observed in a specific context to be engaged in limited proteolysis events of single or few peptide bonds. Furthermore, they were expressed in several

Chapter 2

organs and tissues and also in non-vertebrates, even possibly in bacteria and viruses (Gomis-Rüth, 2003). Examples of these new selective functions are the ectodomain shedding of non-matrix substrates like growth factors and growth-factor-binding proteins to liberate membrane-anchored substrates and thus facilitate autocrine and paracrine interactions. Other substrates include pro-proteinases and clotting factors that must be activated, proteins regulating apoptosis, and others engaged in cell migration and in intercellular communication (Woessner & Nagase, 2000). Likewise, the perceived importance of A/B MCPs has evolved from indiscriminate protein crunchers to sophisticated regulators of hormone homeostasis or fibrinolysis, as well as of other important functions associated with selective limited proteolysis. MMPs are inhibited by their endogenous tissue inhibitors, the TIMPs, which show complementary spatial and temporal expression patterns. The putative latexin member ovocalixin-32 was identified as a matrix protein, like TIMPs and MMPs. All these findings suggest that latexins are the endogenous inhibitors of A/B MCPs in vertebrates, as TIMPs are of MMPs.

Annex to chapter 2

The work presented in this chapter has been done in collaboration with the group of Dr. Xavier Gomis-Rüth from the Unitat de Cristal·lografia de l'Institut de Bioquímica i Biologia Molecular del CSIC a Barcelona. Specifically, my contribution has been the production of latexin, as well as its further biochemical characterization. The obtention of the protein was preceded by the isolation of its cDNA from a gene pool first strand cDNAs of human brain using PCR, tested in several expression systems for the recombinant production of latexin. Further optimization of the best expression system would be required for the production of large quantities of the protein. It is worth to mention that the purification of latexin has been a hard work due to its tendency to aggregate during the purification process.

PART II

CHAPTER 3

Human Latexin forms amyloid-like
oligomers by association of
unfolded polypeptide segments

3.1.- SUMMARY

In amyloid diseases, it is not evident which protein aggregates induce cell death via specific molecular mechanisms and which cause damage because of their mass accumulation and mechanical properties. There is an increasing interest in understanding the characteristics of metastable polypeptide associations preceding the aggregated states of proteins. Here, we show that human latexin, which shares the same fold with Cystatin C, assembles into soluble and stable amyloid-like oligomers at 6M Urea. The ensemble of oligomers binds thioflavin-T and congo red similarly to common amyloid structures. The self-association process resulting in oligomer formation has been studied using CD, electron microscopy, DLS and electrophoretic methods. The native α/β structure is lost at 6M urea and the protein becomes mostly unfolded. A strong “hot spot” of aggregation is predicted in α -helix 3 of latexin. Analysis of an internal, 8-residue synthetic peptide, suggests the involvement of this region in the self-assembly of the polypeptide. The results support the hypothesis that short specific amino acid stretches can act as mediators or facilitators in the incorporation of globular proteins into amyloid structures and the notion of the protective role of the native structure against protein polymerization.

3.2.- INTRODUCTION

Amyloidosis are a group of protein misfolding diseases that are characterized by the polymerization of normally innocuous and soluble proteins or peptides into insoluble proteinaceous deposits. No sequence or structural similarities are apparent between any of the proteins that display the ability to form amyloid aggregates in diseases like Alzheimer's disease, type II diabetes, systemic amyloidosis or the transmissible spongiform encephalopathies (Tan & Pepys, 1994; Sunde & Blake, 1998).

There is a striking difference between the amounts of amyloid depositions in various types of amyloid disorders. In systemic lysozyme amyloidosis, for example, the deposits can grow to kilogram quantities in the liver (Pepys *et al.*, 1993; Harrison *et al.*, 1996). In neurodegenerative diseases, by contrast, there is no clear correlation between the amount of amyloid deposition and the clinical severity of the disease. Significant cognitive impairment of Alzheimer's patients was observed without noticeable amyloid deposits in the brain, although the level of readily soluble amyloid oligomeric assemblies in the Alzheimer's brain was found to be greatly elevated (Lee *et al.*, 2004; Gong *et al.*, 2003). Evidence is accumulating that prefibrillar aggregates are cytotoxic both *in vivo* and *in vitro* (Andersson *et al.*, 2002; Gong *et al.*, 2003; Chromy *et al.*, 2003), although this question is still open to debate. Moreover, aggregates of proteins which are not related to clinical amyloidoses such as human α -lactalbumin (Svensson *et al.*, 2000), SH3 domain, or HypF-N protein, have also been found to be cytotoxic, which implies a common mechanism for cytotoxicity of misfolded proteins (Bucciantini *et al.*, 2002). By contrast, it has been shown that the mature amyloid fibrils of amyloid β peptide and transthyretin are not toxic (Andersson *et al.*, 2002; Hoshi *et al.*, 2003). It still is not evident, however, which particular amyloid structures induce cell death by specific molecular mechanisms and which play the role of "inert" material and cause disease due to their quantity or mechanical properties. In spite of extensive studies of the fibrillar state, scarce structural information is available, and there is an increasing interest in the understanding of the biochemical and biophysical properties of metastable polypeptide associations preceding the aggregated states of proteins.

Here, we show that latexin, a carboxypeptidase inhibitor with an α/β fold, forms metastable oligomers with amyloidotic properties from a mainly unfolded state in 6-

7M urea. Latexin adds up to other proteins unrelated to any known human disease that display the characteristic of amyloid structures (Dobson, 2001)

It has been shown that not all regions of a polypeptide are equally important for determining its aggregation tendency, both in natively unfolded and globular proteins. Single amino acid mutations in these regions can change the aggregation rates dramatically, while similar changes in other regions may have relatively little effect (Chiti *et al.*, 2002; Williams *et al.*, 2004). In addition, it has been shown that it is possible to describe with considerable accuracy the *in vitro* amyloid aggregation propensities of polypeptides using algorithms that take into account the physico-chemical properties of their sequences and of their environment (Chiti *et al.*, 2003; Tartaglia *et al.*, 2004). The presence of aggregation-prone regions has been described in most of the peptides and proteins underlying neurodegenerative and systemic amyloidogenic disorders (Fernández-Escamilla *et al.*, 2004) To determine which part of the unfolded sequence of latexin could be responsible for the oligomerization we used an automated prediction method developed in our lab. (Sánchez de Groot *et al.*, 2005). A peptide containing the predicted region located in α -helix 3 was synthesized and analyzed. Our results indicate that, in physiological conditions, this peptide is able to form microcrystals that display features similar to those of amyloidotic materials, as the binding of Th-T and a distinctive FTIR spectrum, suggesting its involvement in the polymerization of latexin.

3.3.- MATERIALS AND METHODS

3.3.1.- Materials

Recombinant latexin was produced in *E. coli* and purified as previously described (Pallarès et al., 2005). The peptide VLHLAWVA was synthesized by the American Peptide Company, Sunnyvale, CA. Urea, thioflavin-T (Th-T) and congo red were purchased from Sigma. Unless otherwise mentioned, all solutions used for latexin analysis were made in phosphate buffer 50 mM (pH 7.5) (buffer N), and phosphate buffer 5 mM (pH 7.5) was used for peptide preparations (buffer P). All the experiments were performed at 25 °C.

3.3.2- Sample preparation

Oligomeric samples were obtained by incubating protein at 1mg/ml in buffer N in the presence of different concentrations of urea at 25 °C from several hours to several days. Peptide VLHLAWVA was incubated at 500 µM final concentration in buffer P at 25 °C for 24 h.

3.3.3.- Prediction of sensitive regions for aggregation

To determine which regions of latexin sequence could be involved in aggregation an in-house developed method was used (http://bioinf.uab.es/aap/aap_ov.html) and the predictions cross-checked against a second method, TANGO. Both algorithms detected the sequence stretch corresponding to residues 132-150 as the main region which could be potentially involved in the aggregation of latexin from a mainly unfolded state (Sánchez de Groot *et al.*, 2005; Fernández-Escamilla *et al.*, 2004). A peptide covering residues V136-A143 (VLHLAWVA) has been synthesized in order to check experimentally its relevance for latexin aggregation.

3.3.4.- Circular dichroism

Circular dichroism (CD) spectra in the far-UV region was obtained by using a Jasco-810 spectropolarimeter at 25 °C. Spectra were recorded for native proteins and for proteins after incubation in proper urea concentrations (ranging from 0 to 9 M) at 100 µg/ml protein concentration. Twenty accumulations were averaged to obtain each spectrum. The mean residue ellipticity at 222 and 217 nm was plotted versus urea concentration to obtain denaturation curves. For heat-denaturation assays, a 100 µg/ml protein solution in buffer N, was placed in the CD cell holder and the temperature increased slowly at a rate of 0.5 °C/min. A thermo-couple was used to monitor the temperature inside the cuvette and the CD signal at 217 nm was acquired from 4 °C to 96 °C.

3.3.5.- NMR Spectroscopy

NMR samples were prepared by dissolving the proteins in buffer N to a final 1 mM concentration.¹H-NMR spectra were acquired on a Bruker AMX 500 MHz spectrometer at different temperatures.

3.3.6.- Dyes binding assays

Thioflavin-T binding assays were carried out using aliquots of 100 µl drawn from 1 mg/ml protein samples in buffer N after 24 h incubation. These aliquots were diluted into buffer (10 mM sodium phosphate, 150 mM NaCl) containing 65 µM thioflavin-T, and adjusted to a final volume of 1 ml. Fluorescence data were collected after five minutes to ensure that thermal equilibrium had been achieved. Fluorescence emission spectra were recorded using an excitation wavelength fixed at 440 nm. The Th-t stained samples were also analyzed under UV light using a Leica DMBR microscope. Samples were tested for amyloid-specific congo red binding by the spectroscopic band-shift assay as described by Klunk (Klunk *et al.*, 1989). Briefly, aliquots of 1 mg/ml protein were diluted in reaction solution (sodium phosphate 5 mM/NaCl 150 mM, pH 7.0) containing 5 µM congo red and absorption spectra were collected together with negative control solutions of dye in absence of protein and of

protein samples in the absence of dye, subtracting the absorption of the dye and the scattering contribution from the samples spectra.

3.3.7.- Transmission electron microscopy

Samples containing 1 mg/ml of latexin or 500 μ M of VLHLAWVA peptide were incubated as above. A 5 μ l aliquot was then placed on carbon-coated copper grids, and allowed to stand for two minutes. The grids were then washed with distilled water and stained with 2% (w/v) uranyl acetate for another two minutes prior to analysis using a Hitachi H-7000 transmission electron microscope operating at an accelerating voltage of 75 kV.

3.3.8.- ATR-FTIR spectroscopy

The VLHLAWVA peptide was incubated in buffer P for 24 h at room temperature and dried before analysis to reduce water interference in the infrared spectra. The structure of the peptide aggregates was then analyzed in a Bruker Tensor FT-IR spectrometer. For each spectrum, 20 interferograms were collected and averaged. All processing procedures were carried out so as to optimize the quality of the spectrum in the amide I region, between 1550 cm^{-1} and 1700 cm^{-1} .

3.3.9.- Measuring molecular dimensions

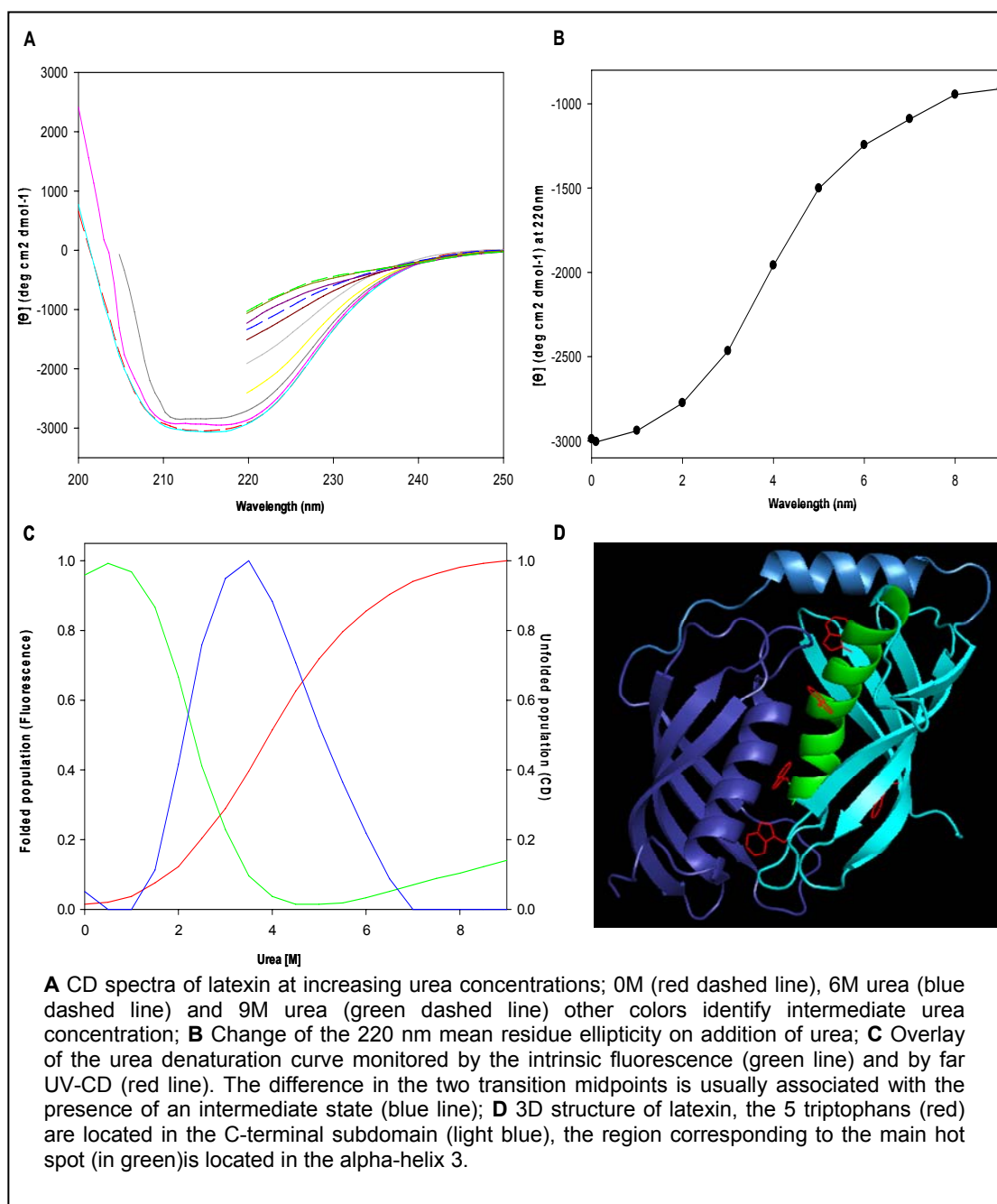
The dimensions of protein species were measured in a DynaPro V6 Molecular sizing dynamic light scattering apparatus. Native samples and protein in 6M urea were incubated at 1mg/ml and room temperature for several hours to several days. Samples were analyzed at 8°C and data was acquired every 5 seconds during three minutes.

3.4.- RESULTS

3.4.1.- Conformation and stability of latexin

Figure 3.1A shows the far-UV circular dichroism (CD) spectra of latexin at different conditions. The spectrum of the protein acquired under native conditions is consistent with the recently determined 3D crystal structure and corresponds to a well-structured protein with an α/β topology. This globular structure is progressively lost, however, in the presence of increasing concentrations of urea (Fig. 3.1A). Figure 3.1B shows the dependence of the secondary structure, measured as the mean residue ellipticity at 220 nm, on urea concentration. The denaturation curve presents a single sigmoidal transition corresponding to a two-state unfolding reaction. The urea denaturation of latexin was also followed by monitoring changes in the intrinsic fluorescence of its tryptophan residues (a better reporter of tertiary structural changes), resulting also in a sigmoidal curve whose transition midpoint was, however, much lower (2.3M) than the value obtained by far-UV CD (4M). The difference in the two transition midpoints is usually associated to the presence of an intermediate state (Fig. 3.1C). From the normalized unfolding curves, the population in the native, intermediate and unfolded states can easily be calculated. The intermediate state occurs between 1.5 and 6.5M urea concentration with a maximum population at 3.5M urea. Nevertheless, in the case of latexin it is important to stress the striking asymmetry of tryptophan distribution between its two subdomains. All five tryptophan residues are located in the second subdomain of latexin, four of them shielded at the interface between both subdomains; the remaining tryptophan is exposed to solvent, and presumably contributes little to variations in fluorescence upon unfolding (Fig. 3.1D). The observed changes in fluorescence may therefore reflect only a local structural rearrangement, such as a loss of contacts between subdomains, rather than a global unfolding of the protein or of one of its subdomains. This is also likely from the study of binding to TNS. The TNS hydrophobic probe is becoming the method of choice to determine partially folded states of a protein. TNS emission spectra of latexin were recorded as a function of urea concentration and no significant binding could be detected either in the folded, completely unfolded state or at 3-4M urea where the intermediate maximally populates.

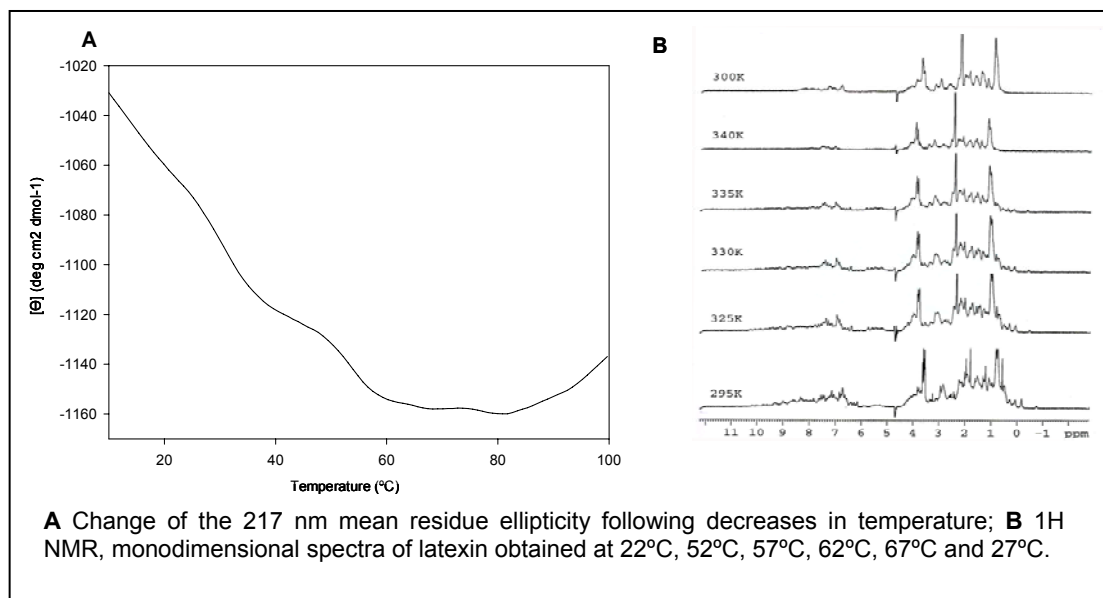
Figure 3.1.- Conformational properties of latexin



We also monitored the thermal unfolding of latexin by far-UV CD at 217 nm (Fig. 3.2A). The resulting unfolding curve is rather complex, with several transitions. Moreover, the heat-denatured latexin was not able to recover its initial native-like ellipticity upon cooling, and the presence of small aggregates was already detected in the protein solution, clearly indicating that the thermal unfolding of latexin is not reversible. Latexin appears as a rather unstable protein in front of the temperature, losing completely its secondary structure approximately above 60 °C and displaying

at 30°C only about 50% of the secondary structure content exhibited at 10°C. This behaviour, as well as the irreversibility of the process, was confirmed by analyzing the solution ¹H-NMR spectra of latexin at different temperatures (Fig. 3.2B).

Figure 3.2.- Latexin thermal desnaturalization and ¹H-NMR spectra at different temperatures

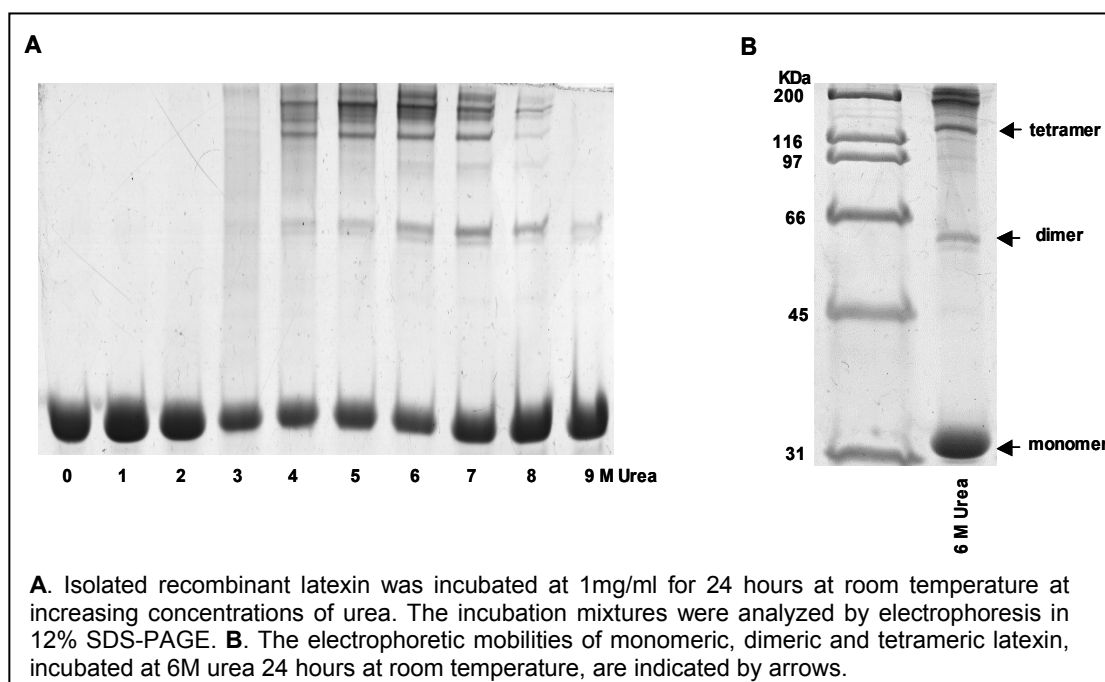


3.4.2.- Latexin self-association

In a number of cases, in particular for cystatins (Janowski *et al.*, 2004), the appearance of folding intermediates has been linked to the self-association of polypeptides into oligomers, probably by docking of these monomeric precursors, to yield insoluble amyloid-like structures. To assess whether the accumulation of an intermediate during the unfolding of latexin could result in the formation of stable oligomeric species, we used SDS-PAGE electrophoresis to analyze samples of latexin incubated for 24h at different urea concentrations. This method is useful for simultaneously detecting the formation of oligomers with different sizes. The results shown in Figure 3.3 demonstrate that latexin self-associates into stable oligomers in an urea dependent manner. SDS resistant self-associated latexin oligomers start to appear at 3-4M urea and accumulate maximally at 6-7M urea. Although a decrease is observed above 7M, the dimeric species are the most stable being detected, and are still present even at 9M urea. The two cysteine residues in latexin are not involved in a disulfide bond in the native state are neither responsible for the

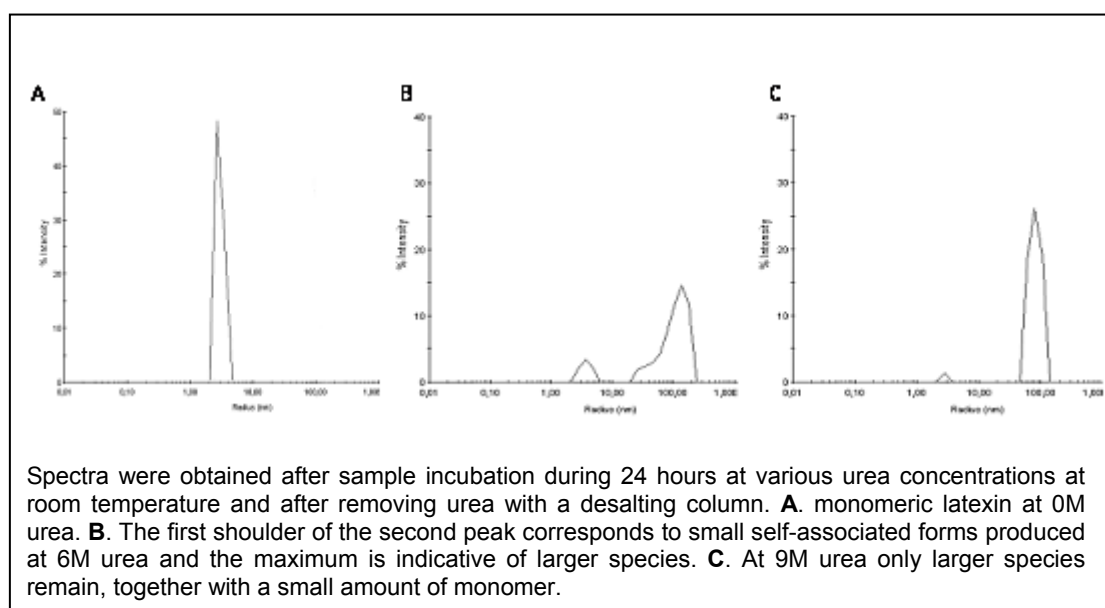
association of latexin in the presence of urea since the distribution of oligomers does not change in the presence of β 2-mercaptoethanol. The observed data indicate that self-association starts at urea concentrations where latexin is partially unfolded and that at 6-7M urea, where the maximum self-association is achieved, latexin is completely unfolded (Fig. 3.1A,B).

Figure 3.3.- SDS-PAGE electrophoresis of latexin in various concentrations of urea



Dynamic light scattering (DLS) experiments at selected urea concentrations confirmed the assumptions drawn from the electrophoretic analysis (Fig. 3.4). In the absence of urea the sample showed a single population of $R_s=2.7$ nm, corresponding to the latexin monomer (Fig. 3.4A). The population of monomeric species sharply decreases at 6M urea and a new component appears in addition to the monomer peak (Fig. 3.4B). The broadness of the peak appearing at 6M urea indicates the presence of a heterogeneous mixture of oligomers and two components can be distinguished within it: a first one of $R_s=3.8$ nm, expected to contain small self-associated forms like dimers or tetramers, and a second one, with a higher atom radius distribution of $R_s=116.2$ nm, corresponding to much larger species. For samples incubated at the highest urea concentration, the smallest self-associated forms disappear and only the larger ones remain in this condition, together with small amounts of monomer (Fig. 3.4C).

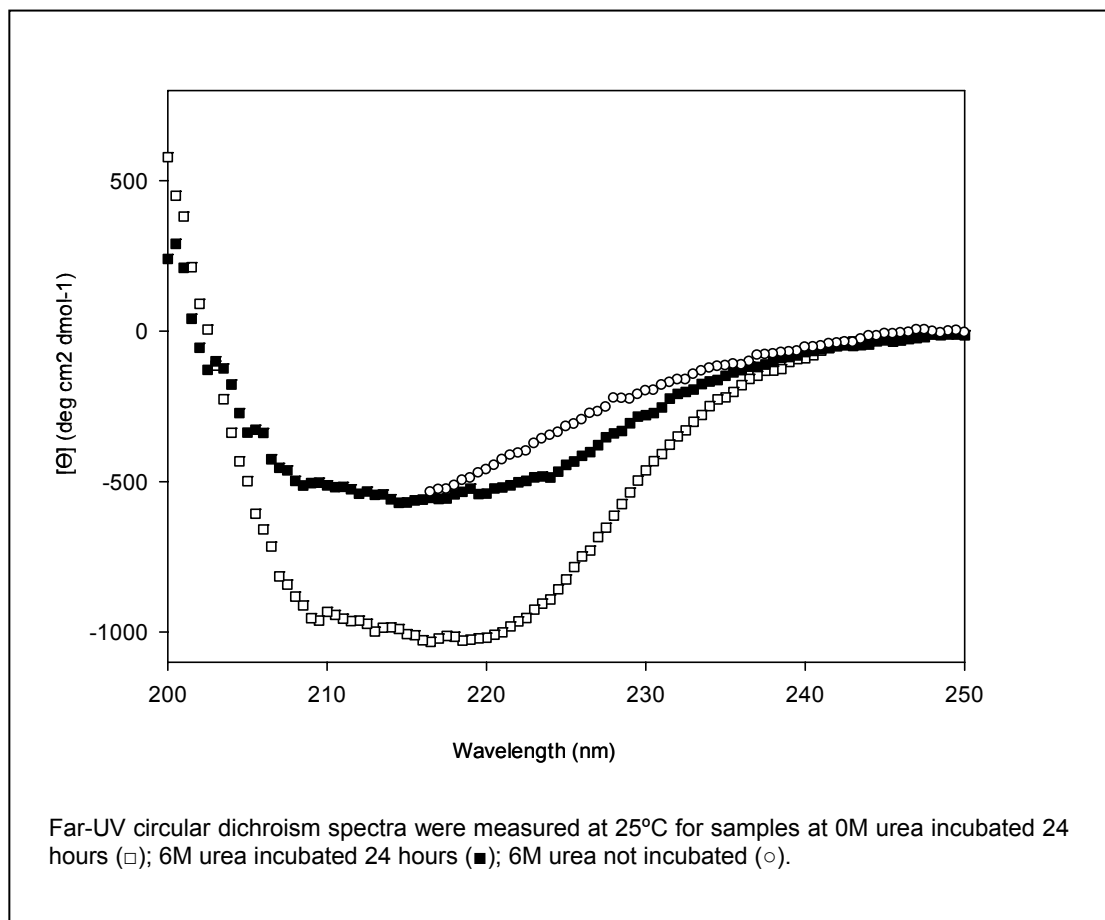
Figure 3.4.- DLS for latexin self-associated forms



3.4.3- Far-UV circular dichroism studies of urea-induced latexin oligomerization

To understand the conformational changes occurring during oligomerization, samples incubated for 24 hours at 0M urea and 6M urea were analyzed by far-UV circular dichroism and, after urea elimination with a desalting column, compared with fresh samples containing the same amount of urea (Fig. 3.5). The CD spectra of 24 hour incubated sample is typical of a largely extended β -sheet conformation as revealed by the single negative band at 217-220 nm. This spectrum differs in shape from the native conformation of latexin, which is also rich in beta sheet structure, in shape. The inhibitory capability of the incubated sample was measured to discard that the presence of properly folded latexin could contribute to its spectrum. No activity was detected, indicating the absence of molecules in native conformation. In addition, the fresh 6M urea sample displays a CD spectrum corresponding to a much more unfolded protein, suggesting that the self-associated forms of latexin could result from interactions between unfolded molecular segments that originate aggregates with a β -sheet conformation.

Figure 3.5.- Latexin secondary structural changes after 24 hour incubation in urea monitored by CD.

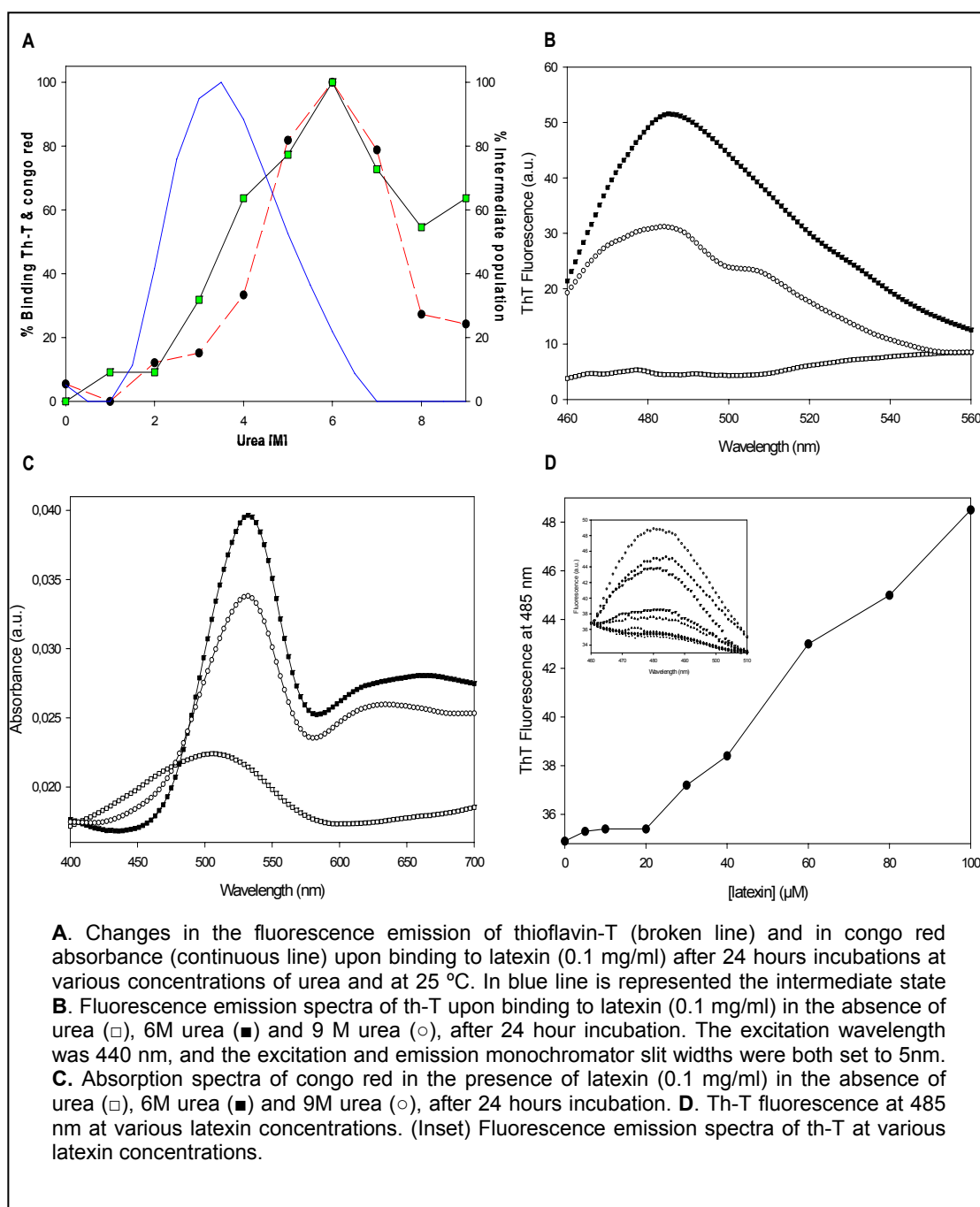


3.4.4.- Amyloidotic dyes binding

The oligomers of latexin formed in the presence of urea were further analyzed with two dyes specific for beta-sheet conformation in amyloid fibrils. In a binding assay of thioflavin-T (Th-T) to latexin previously incubated for 24 hours in the presence of different urea concentrations, changes in the fluorescence of the dye are only detected above 2M urea, and the maximum increase in the fluorescence intensity at 485 nm is observed at 6M urea as a consequence of dye binding to the protein (Fig. 3.6A,B). Beyond 6M urea the intensity of the dye fluorescence at 485 nm decreases. A second Th-T binding assay using different latexin concentrations incubated one day at 6M urea (Fig. 3.6D), showed that, above 20 μ M latexin, a linear correlation exists between latexin concentration and fluorescence emission, indicating the specificity of the assay.

The second amyloid dye used, congo red, has an absorbance maximum at 490 nm that increases and shifts to red upon binding to amyloid material. The changes can be quantified by obtaining the difference spectra between the dye alone and when incubated with the protein. After correction, latexin incubated at 6M urea showed a maximum change at 533 nm, similar to A β amyloid peptide (Klunk *et al.*, 1989), the variations being larger for samples incubated with urea concentrations higher than

Figure 3.6.- Binding of amyloid dyes by latexin



2M with a maximum at 6M urea (Fig. 3.6A,C). Though latexin oligomers bind congo red, the characteristic green birefringence under polarized light microscope, a classic feature of amyloid, was not observed. This behavior has also been observed for well-known amyloid-forming proteins like yeast Sup 35 or β 2-microglobulin (Balbirnie *et al.*, 2001; McParland *et al.*, 2000).

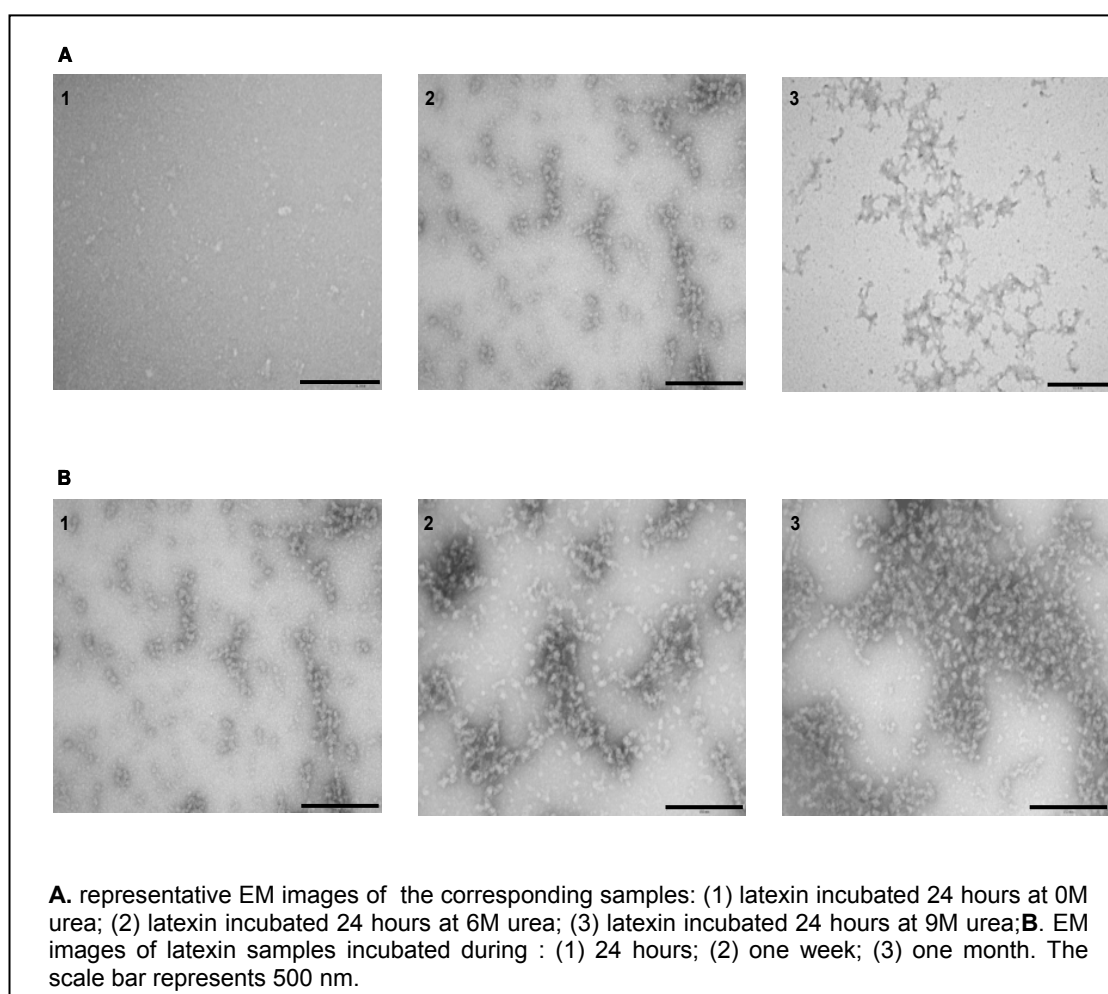
3.4.5.- Oligomerization evaluated by electron microscopy

Solutions of protein incubated for 24 hours at various urea concentrations were analyzed by EM. Results are shown in figure 3.7. No deposition or particles were detected for latexin incubated at 0M urea in accordance with the above mentioned results. Latexin incubated 24 hours at 6M urea displays particles corresponding to the formation of round-shaped oligomers whereas samples incubated one day at 9M urea present amorphous aggregates. To determine if the oligomers formed at 6M change their morphology when they are incubated for longer periods, EM analyses were also performed for latexin incubated at 6M urea for one week and one month. As can be seen in figure 3.7B, after one week of incubation, the round-shaped oligomers tend to grow in size and self associate, this observation is even more clear after one month incubation, although no evolution to fibrillar structures was observed in the preparation.

3.4.6.- Generation of an aggregation profile and identification of “hot spots”

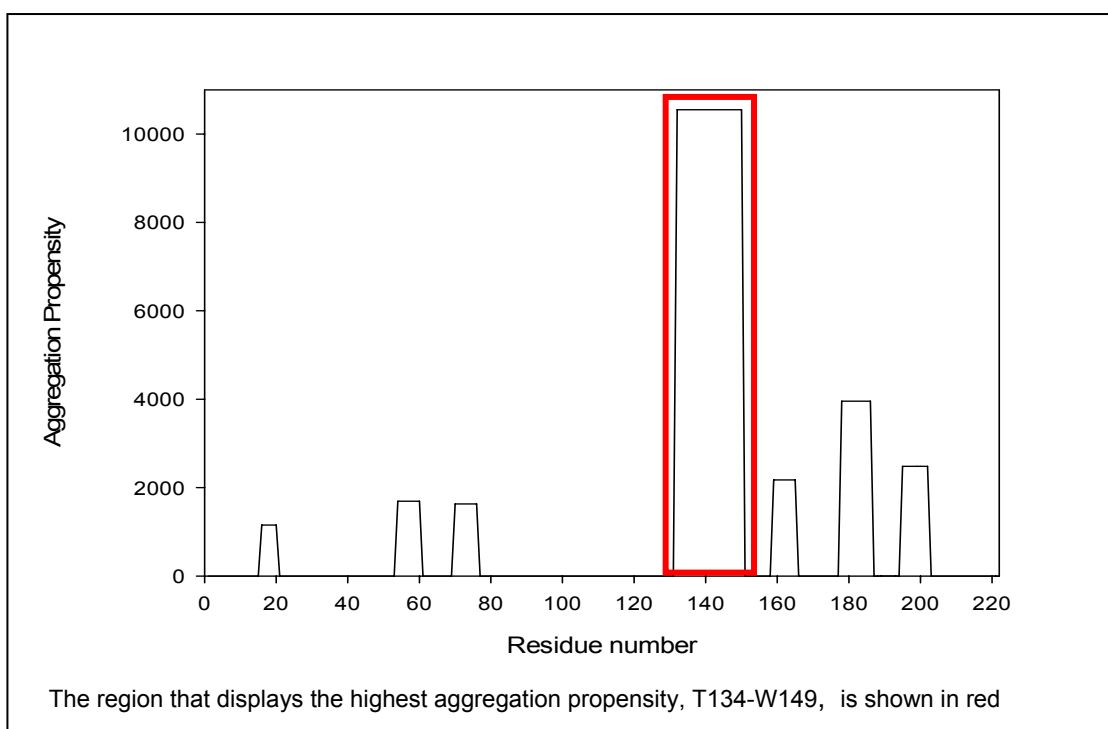
One of the most intriguing recent observations in studies of the kinetics of amyloid formation is that polypeptide sequences appear to contain local regions that are “sensitive” for aggregation. It has been shown that it is possible to describe with considerable accuracy the *in vitro* amyloid aggregation propensities of polypeptides using algorithms that take into account the physico-chemical properties of their sequences (Sánchez de Groot *et al.*, 2005; Fernández-Escamilla *et al.*, 2004; Pawar *et al.*, 2005). This leads directly to the definition of “hot spot” of aggregation as a certain region that displays higher aggregation propensity than the rest of the sequence.

Figure 3.7.- Representative images of latexin protein incubated at various urea concentrations for more than one day.



Latexin has been shown to form the oligomeric species from its unfolded state. Under these conditions aggregation-prone regions, if present, may promote and drive the aggregation event. According to the analysis of the aggregation profile, shown in figure 3.8, this protein displays one main “hot spot” encompassing residues 127-156. To find out whether this region could play a critical role in the aggregation of latexin, we decided to synthesize a peptide containing its sequence and that is located at the α -helix 3. We designed a first peptide covering this region T134-W149, but its high aggregation propensity made it impossible to be synthesized in a soluble form. We subsequently focused on the analysis of a shorter 8-residue corresponding to the central part of the “hot spot” V136-A143 (VLHLAWVA).

Figure 3.8.- Aggregation profile of human latexin



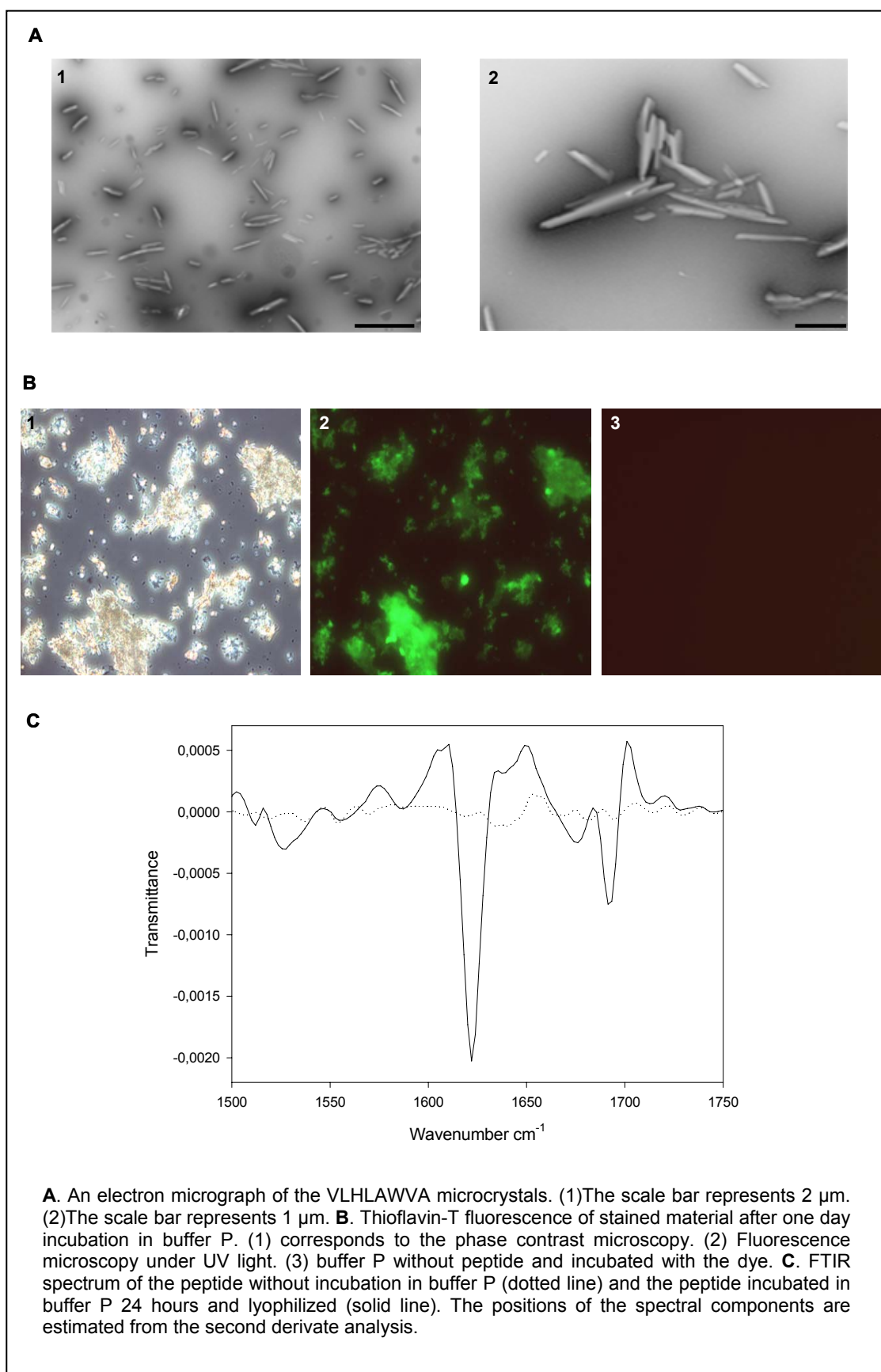
3.4.7.- Microcrystals of VLHLAWVA resemble amyloid fibrils

The peptide VLHLAWVA, when incubated at 500 μ M in sodium phosphate 5 mM pH 7.5 over-night, forms well-ordered microcrystals with amyloid properties, including the following:

Th-T binding. The binding of Th-T to the VLHLAWVA microcrystals was visualized by fluorescence microscopy (Fig. 3.9B). At pH 7.5 and after 24 hour incubation, areas rich in amyloidotic material appear stained with Th-T, giving a bright green-yellow fluorescence against a dark background. No fluorescence was detected for the buffer without sample.

Structural transition to β -sheet. The structural analysis of VLHLAWVA microcrystals by Fourier transform IR shows that they possess an internal structure that is highly similar to amyloid fibrils with a strong peak at 1620 cm^{-1} dominating the amide I region. An additional sharp peak at 1690 accounts for a highly ordered intermolecular antiparallel β -sheet structure (Fig. 3.9C). The peak at 1670 cm^{-1} can be attributed to residual TFA in the solution.

Figure 3.9.- Microcrystals of VLHLAWVA



3.5.- DISCUSSION

It has been proposed that almost every protein, when incubated under the right conditions, can aggregate (Dobson, 2001). Globular proteins rarely aggregate from their native states and destabilization, resulting in an increased population of unfolded molecules, is well established as a triggering factor in disorders associated with the deposition of proteins that are globular in their normal functional states (Kelly, 1998a), like of β 2-microglobulin, lysozyme, transthyretin and the prion protein. Interestingly enough, many proteins involved in depositional disorders are mostly unstructured within the cell (Uversky & Fink, 2004). These include amylin, amyloid- β -protein, and α -synuclein, among others. In these cases, protein polymerization and deposition does not require unfolding and can occur by direct self-assembly of the unstructured polypeptide chains.

In the last few years, proteins unrelated to any known human disease have been found to convert *in vitro* into higher order structures that also present a cross- β conformation and fulfill all characteristics of amyloid fibrils (Dobson, 2001; Guijarro *et al.*, 1998; Mayer, 2001). These results have supported the hypothesis that, in theory, all proteins under appropriated conditions can form amyloid structures. Latexin is a vertebrate carboxypeptidase inhibitor with an α/β fold, closely resembling that of cystatins. Here we show that this fold is destabilized at low urea concentrations and that the protein is completely unfolded at 6-7 M urea. The oligomerization of Latexin correlates with conditions in which the protein is mostly unfolded, rather than with environments that generate partially folded intermediates. The main intermediate detected in the unfolding reaction corresponds most likely to an open form in which the two domains of latexin have lost their contacts but conserve their domain architecture. The intermediate seems to have no relevance to aggregation since its peak formation conditions do not overlap with those of maximum oligomerization. These data argue that in this case, like in the case of intrinsically unstructured proteins, the structural precursors in the first states of polymerization are polypeptide chain segments that lack stable elements of secondary structure. This case differs, from protein models in which polymerization depends on the population of partially folded states rich in β -sheet structure.

Protein aggregation generates a very broad range of structures including oligomers, protofibrils, and fibrils. Determining the role of these so-called “prefibrillar species” in assembly is of paramount importance, since such complexes have been proposed for many proteins to be responsible for the cellular toxicity associated with amyloid disease (Klein *et al.*, 2004; Caughey & Lansbury, 2003). In this work dimeric, tetrameric and higher order latexin oligomers were visualized and preliminary characterized by EM, DLS and SDS-PAGE gel. From this study, it appears that the latexin system could be an excellent model for the isolation and deep structural and biochemical characterization of stable low-molecular weight protofibrillar species, since they appear to be the final association state of latexin under strongly denaturing conditions. No conversion into fibrils could be observed, probably because only strong intermolecular contacts are maintained in these conditions and the network of hydrogen bonds that are needed to keep the well defined orientation, characteristic of amyloid fibrils, cannot be established. This inability to form fibrils would allow the study of the, otherwise, usually elusive and always transient small molecular forms in protein polymerization processes.

Detecting the region of latexin that allows the maintenance of intermolecular contacts in strongly denaturing conditions is therefore of interest to understand the oligomerization process. Provided that this occurs in conditions where the polypeptide chain is mainly unfolded, it is suitable to use algorithms like the one developed in our lab which predicts regions with the highest propensity to aggregate merely from the protein sequence. The central part of this region corresponds in latexin to the peptide VLHLAWVA, which forms well-ordered microcrystals with amyloid properties as analyzed by EM, Th-T binding and FTIR. Consistently with its putative role in latexin oligomerization, the formation of peptide microcrystals in physiological conditions is usually associated to the rapid formation of strong and specific intermolecular interactions. These microcrystals are of interest since they closely resemble those formed by a seven-residue peptide of the yeast Sup35 prion from which the group of Eisenberg (Nelson *et al.*, 2005) have determined the only high resolution atomic structure of the cross-beta spine. Accordingly, experiments are underway in our lab to solve the intimate structure of this peptide assembly.

Aggregation-prone regions are likely to be blocked in the native state of globular proteins because their side chains are usually hidden in the inner hydrophobic core or already involved in the network of contacts that stabilizes a protein. The

Chapter 3

VLHLAWVA region of latexin is located in α -helix 3, where its anomalous interactions are being blocked not only because it belongs to a secondary structure element, but also because its side chains are involved in the interface contacts between the two domains of latexin. From the data shown, it appears that the opening of the interface is not enough to promote oligomerization and desestructuration of the alpha-helical structure, but it is probably a prerequisite for this event to occur. Once in a non structured context and exposed to solvent by denaturation, the VLHLAWA region is ready for the establishment of intermolecular contacts as proven by the analysis of the eight-residue peptide. Thus, latexin constitutes yet another example of the protective role played by the protein native structure against aggregation.

CHAPTER 4

Amyloid fibril formation by a partially structured intermediate state of alpha-chymotrypsin

4.1.- SUMMARY

In this report the effects of 2,2,2-trifluoroethanol (TFE) on the structure of α -chymotrypsin are investigated. The protein aggregates maximally in 35% (v/v) TFE. Congo red and thioflavin-T binding experiments suggest that the aggregates induced by TFE have amyloid-like properties, and transmission electron microscopy data show that these aggregates have a fibrillar morphology. Fluorescence, circular dichroism, anilino-8-naphthalene sulfonate binding, and Fourier-transformed infrared spectroscopy data suggest that formation of a partially structured intermediate state precedes the onset of the aggregation process. The native β -barrel structure of α -chymotrypsin appears to be disrupted in the partially structured intermediate state in favor of a non-native extended β -sheet conformation with exposed hydrophobic surfaces. The protein becomes “sticky” under these conditions and aggregates into amyloid-like structures.

The data support the hypothesis that amyloid formation involves the ordered self-assembly of partially folded species that are critical soluble precursors of fibrillar aggregates.

4.2.- INTRODUCTION

The deposition of amyloid fibrils has been linked to a variety of slow-onset degenerative diseases, such as Alzheimer's disease, Parkinson's disease, type II diabetes mellitus, and spongiform encephalopathies (Tan & Pepys, 1994; Dobson 2001; Ellis & Pinheiro, 2002). Identification of novel therapeutic avenues to prevent or cure diseases associated with amyloid fibril deposition in tissue and aggregates toxicity requires a rational understanding of the determinants driving protein aggregation and organized assembly of proteins into fibrils rich in β -structure. Recent clues to the molecular basis of amyloid diseases and biological significance of protein aggregation have been provided by observations that a range of proteins unrelated to any known human disease have been found to convert *in vitro* into fibrils that fulfill all characteristics of those found in pathological conditions (Ventura *et al.*, 2002; Villegas *et al.*, 2000; Fandrich *et al.*, 2001; Bucciantini *et al.*, 2002; Gujjarro *et al.*, 1998; Litvinovich *et al.*, 1998; Chiti *et al.*, 1999). In fact, several lines of evidence suggest that amyloid represents a generic form of polypeptide conformation, and most peptides/proteins have the potential to form amyloid-like structures under appropriate conditions (Dobson 2001; Bucciantini *et al.*, 2002; Chiti *et al.*, 1999).

In disease-related models (Kelly, 1998a; Lai *et al.*, 1996; Booth *et al.*, 1997; Uversky *et al.*, 2001; McParland *et al.*, 2000; Khurana *et al.*, 2001) amyloid aggregates do not develop directly from the native conformations of proteins. Amyloid formation appears to be preceded by conformational changes in the native protein leading to the formation of sticky partially structured intermediates. These partially folded species seem to interact intermolecularly to render protein oligomers that further rearrange into amyloid fibrils (Kelly, 1998a; Harper & Lansbury, 1997). Similarly, formation of amyloid by natural proteins not related to disease is promoted *in vitro* by partially denaturing conditions such as low pH (Ventura *et al.*, 2002; Gujjarro *et al.*, 1998), high temperature (Litvinovich *et al.*, 1998) or the presence of moderate concentrations of alcohols (Chiti *et al.*, 1999). However, very little is known about the conformational features of the sticky partially structured intermediate state(s). In this context, the study of conditions which populate these partially folded intermediates in different protein models should provide insights into the onset of the aggregation process.

α -Chymotrypsin is an all β -protein, which is folded in two antiparallel β -barrel domains, each containing six β -strands with the same topology (Wright, 1973). Both six-stranded antiparallel β -barrels consist of a Greek key motif followed by an antiparallel hairpin motif. Only two short lengths of α -helix are present in α -chymotrypsin, one in each domain. Recently, the induction of amyloid properties by partially denaturing conditions in the acidic fibroblast growth factor from newt, a protein whose fold resembles that of α -chymotrypsin, has been reported (Srisailam *et al.*, 2003). Thus, an investigation on the potential of α -chymotrypsin to form amyloid-like structures under partially denaturing conditions in which it still retains a measurable amount of residual structure was undertaken. TFE, a solvent known to stabilize partially folded proteins and to promote amyloid formation in several models (Srisailam *et al.*, 2003; Chiti *et al.*, 2000) was chosen for this purpose. TFE is observed to induce amyloid-like features in α -chymotrypsin and, resembling what happens to proteins involved in conformational diseases, its aggregation into amyloid-like structures is triggered by the accumulation of a sticky and partially folded intermediate

4.3.- MATERIALS AND METHODS

4.3.1.- Materials

α -Chymotrypsin, trifluoroethanol, 1-anilino-8-naphthalene sulfonate (magnesium salt), thioflavin-T (Th-T), and congo red were purchased from Sigma. Unless otherwise mentioned, all solutions were made in 50 mM phosphate buffer (pH 7.0). All experiments were performed at 25°C.

4.3.2.- Turbidity measurements

Turbidity measurements were performed on a Cary-100 Varian spectrophotometer. All measurements were made after 12 hours of incubation (at 25°C) of the protein in appropriate concentrations of TFE. The concentration of the protein used in the turbidity experiments was 100 μ g/ml. The turbidity of the solutions was measured by absorbance at 350 nm. The path length of the sample cell used was 10 mm.

4.3.3.- Circular dichroism

Circular dichroism (CD) spectra in the far and near-UV region were obtained by using a Jasco 710 spectropolarimeter at 25°C. Protein was assayed at 100 μ g/ml in proper TFE concentrations. Twenty accumulations were averaged to obtain each spectrum.

4.3.4.- Intrinsic fluorescence

Fluorescence experiments were performed on a Perkin–Elmer 650-40 fluorescence spectrophotometer. Intrinsic fluorescence measurements at a protein concentration of 100 μ g/ml were made using an excitation wavelength of 280 nm and emission was collected between 300 nm and 440 nm. The fluorescence of ANS was excited at 365 nm and emission collected between 420 nm and 600 nm.

4.3.5.- Dye binding assays

Thioflavin-T binding assays were carried out using aliquots of 100 μ l drawn from 100 μ g/ml protein samples incubated as indicated above. These aliquots were diluted into buffer (sodium phosphate 10 mM, NaCl 150 mM) containing thioflavin-T 65 μ M, and adjusted to a final volume of 1 ml. Fluorescence data were collected after five minutes to ensure that thermal equilibrium had been achieved. Fluorescence emission spectra were recorded using an excitation wavelength fixed at 440 nm. Congo red binding assays were performed as reported (Ventura *et al.*, 2002).

4.3.6.- ATR-FTIR spectroscopy

Protein aggregates incubated one day in 35% TFE were centrifuged at 12,000 rpm for ten minutes, and the supernatant removed. The precipitant was dried overnight in a Speed-Vac system. The dry aggregates were analyzed in a Bruker Tensor FT-IR spectrometer. FT-IR spectra of the native protein were acquired after air-drying the protein solution. For each spectrum, 20 interferograms were collected and averaged. All processing procedures were carried out so as to optimize the quality of the spectrum in the amide I region, between 1550 cm^{-1} and 1700 cm^{-1} .

4.3.7.- Aggregation kinetics

The assay was performed on a Cary-100 Varian spectrophotometer with samples containing 100 μ g/ml of α -chymotrypsin in 35% TFE by monitoring the absorbance at 350 nm during 200 minutes. Changes in the 217 nm CD signal were also followed in a time-dependent manner using a Jasco 710 spectropolarimeter at 25 $^{\circ}$ C.

4.3.8.- Transmission electron microscopy

Samples containing μ g/ml of α -chymotrypsin were incubated as indicated above. A 5 μ l aliquot was then placed on carbon-coated copper grids, and allowed to stand for two minutes. The grids were then washed with distilled water and stained with 2% (w/v) uranyl acetate for another two minutes prior to analysis using a Hitachi H-7000 transmission electron microscope operating at accelerating voltages of 75 kV.

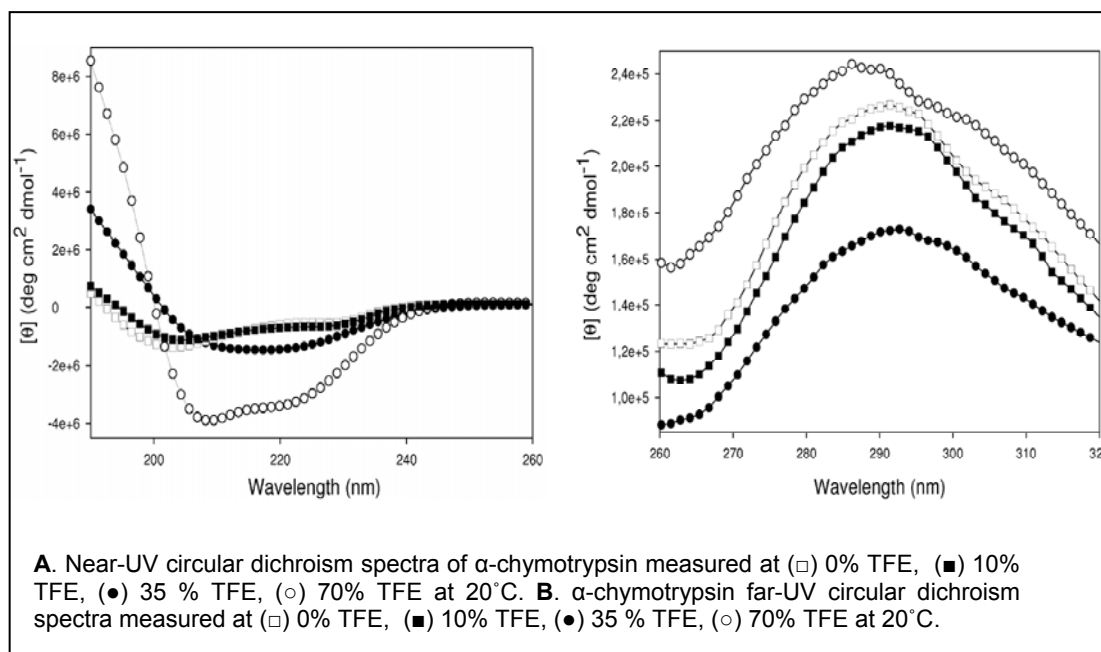
4.4.- RESULTS

4.4.1.- Effect of TFE on α -chymotrypsin conformation

The secondary structural changes in α -chymotrypsin were monitored at various concentrations of TFE using both near and far-UV circular dichroism (CD) spectroscopy. The CD spectrum of α -chymotrypsin shows the characteristic negative band around 230 nm, which is related to the extent of native α -helix in the secondary structure, and a global minimum in ellipticity at around 205 nm, typical of β -strands in β -barrel architectures (Fig. 4.1A). No significant changes occur in the CD spectrum at TFE concentrations up to 10% (v/v) (Fig. 4.1A), whereas the far-UV CD spectra of the protein show dramatic changes above this TFE concentration. At concentrations ranging from 20% to 35% the CD spectra are typical of a largely extended β -sheet conformation as revealed by the single negative band at 217–220 nm (Fig. 4.1A). At higher concentrations of TFE (> 40%), the far-UV CD spectra of the protein show two negative ellipticity bands at 208 nm and 222 nm, indicating the induction of extensive new non-native helical conformation (Figure 4.1A).

Tertiary structure sensitive near-UV CD spectra of native α -chymotrypsin show a positive band at 290 nm, which is ascribed to tryptophan and tyrosine residues (Fig. 4.1B). Changes in the intensity of this band in the presence of different concentrations of TFE correlate well with those observed in the far-UV part of the spectra. Little changes are detected in the CD spectrum at a TFE concentration up to 10. Above this TFE concentration, the intensity of the band suffers a decrease with a minimum at 35% (Fig. 4.1B), followed by a progressive recovery of band intensity until a maximum is reached at 70% TFE, where the protein exists in non-native helical conformation (Fig. 4.1B). Thus, the tertiary structure of α -chymotrypsin in the presence of 35% TFE appears to be different both from the native and non-native conformations of the protein.

Figure 4.1.- α -chymotrypsin secondary structural changes at various percentages of TFE monitored by CD

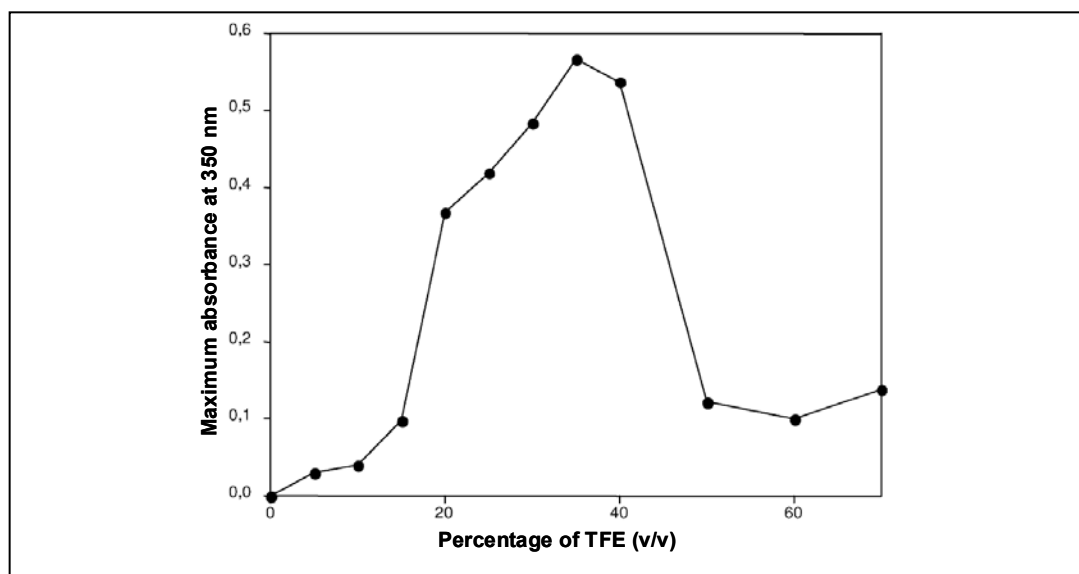


4.4.2.- Influence of backbone conformation on α -chymotrypsin aggregation

The protein samples treated with TFE in the concentration range of 15% to 70% and incubated at 25°C for more than three hours developed a slight turbidity indicating aggregation, which was systematically monitored by measuring light scattering changes at 350 nm at various concentrations of TFE. The aggregation profile shows that the protein tends to aggregate maximally between 20% and 40% TFE, with a maximum detected at a TFE concentration of 35% (Fig. 4.2), wherein the protein is observed to exist in an extended β -sheet conformation, as assessed by CD. Above 40% TFE, turbidity decreases significantly and the protein solution turns clearer, with a minimum value of 350 nm absorbance at 60% TFE (Fig. 4.2).

To rule out that aggregation could be due to structural alterations resulting from α -chymotrypsin self-processing in the presence of TFE, the integrity of the enzyme during the aggregation process at 35% TFE was confirmed by SDS-PAGE analysis (results not shown).

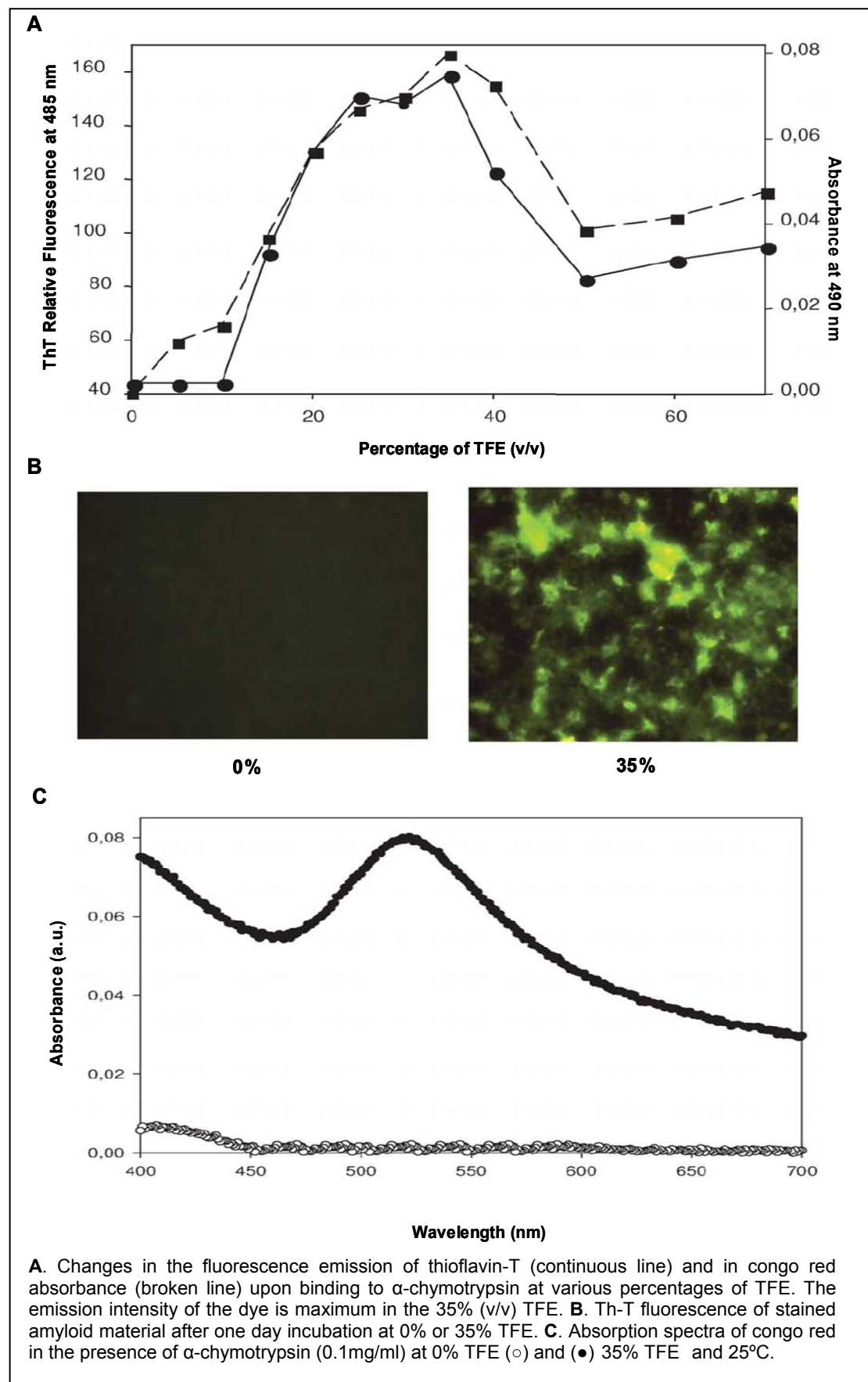
Figure 4.2.- Turbidity changes of α -chymotrypsin at various percentages of TFE, at 0.1mg/ml and 25°C in the presence of various concentrations of TFE, monitored by measuring absorbance at 350nm.



4.4.3.- TFE-induced aggregates display amyloid-like features

The possibility that the aggregates of α -chymotrypsin formed in 35% TFE possess amyloid-like properties was examined both by analyzing the ability of the TFE-induced aggregates to bind amyloid dyes and by inspection of their ultrastructural characteristics through electron microscopy and attenuated total reflection Fourier transform infrared spectroscopy (ATR-FTIR).

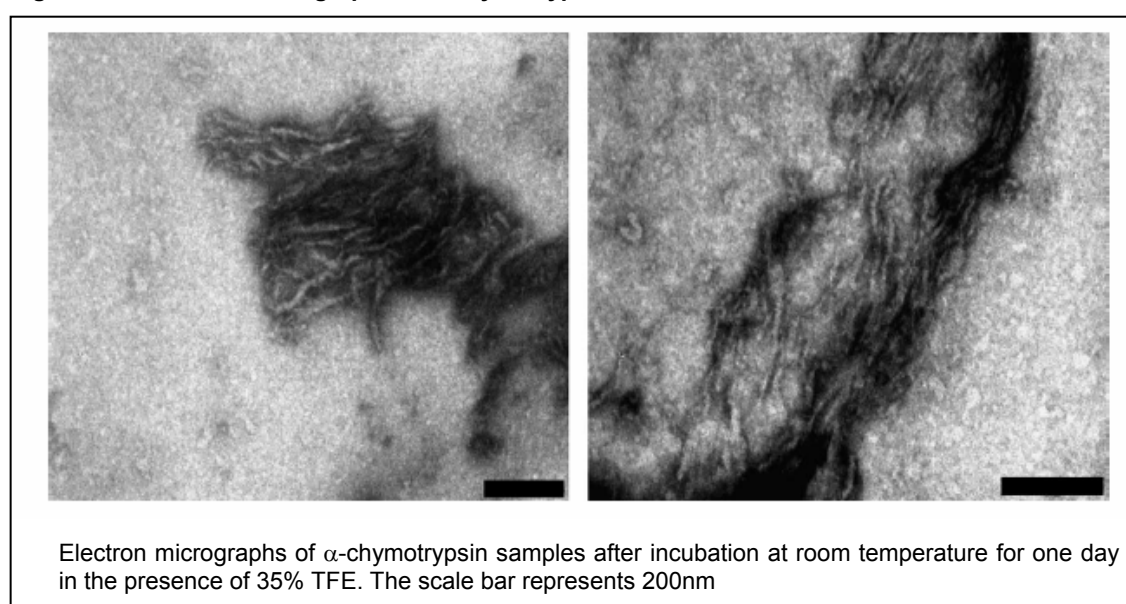
Thioflavin-T (Th-T) is an amyloid azo-free diagnostic dye that by a so far unknown mechanism specifically interacts with the crossed- β -pleated sheet structure common to a variety of amyloid fibrils. This property allows, in principle, the Th-T assay to distinguish highly ordered amyloid structures from amorphous precipitates. The binding of this dye to ordered crossed β -sheet aggregates results in a huge increase in fluorescence emission (Le Vine, 1993; Le Vine, 1997). In a binding assay with α -chymotrypsin, no appreciable changes in the fluorescence of the dye at TFE concentrations lower than 10% are detected, but a four fold increase in the fluorescence intensity at 485 nm is observed as a consequence of dye binding to the protein at 35% TFE (Fig. 4.3A). Beyond 35% TFE the intensity of the dye fluorescence at 485 nm decreases.

Figure 4.3.- Binding of amyloid dyes by α -chymotrypsin

The binding of Th-T to aggregates was visualized in parallel by fluorescence microscopy (Fig. 4.3B). At 35% TFE areas rich in protein fibrous material appear stained with Th-T, giving a bright green–yellow fluorescence against a dark background. No fluorescence was detected for the protein in its native state (Fig. 4.3B). The second amyloid dye assayed, congo red, has an absorbance maximum at 490 nm that increases and shifts to red upon binding to ordered repetitive β -sheet structures in the amyloid fibrils (Roterman *et al.*, 2001; Piekarska *et al.*, 1999). Congo red binding displays a maximum at 35% TFE in a fashion very much similar to what is observed with Th-T (Fig. 4.3A). The dye binding is detected both by a significant increase in the absorbance intensity and by a prominent red shift in the absorbance maximum (490 nm to 530 nm). Thus, the changes can be better quantified by obtaining the difference spectra between the dye free or incubated with the protein. After correction for the presence of 35% TFE, a maximum difference is apparent at 530 nm, similarly to that described for the A β amyloid peptide (Klunk *et al.*, 1999) and other amyloid proteins, while no such effect was observed for the native protein (Fig. 4.3C).

Micrographs obtained by transmission electron microscopy of the aggregates formed in 35% TFE reveal that in addition to some amorphous aggregates, fibrillar material is clearly present (Fig. 4.4). The fibrils are long and unbranched with diameters of \approx 8–15 nm, similar to the width of the fibrillar structures formed from other amyloidogenic proteins (Sunde & Blake, 1997).

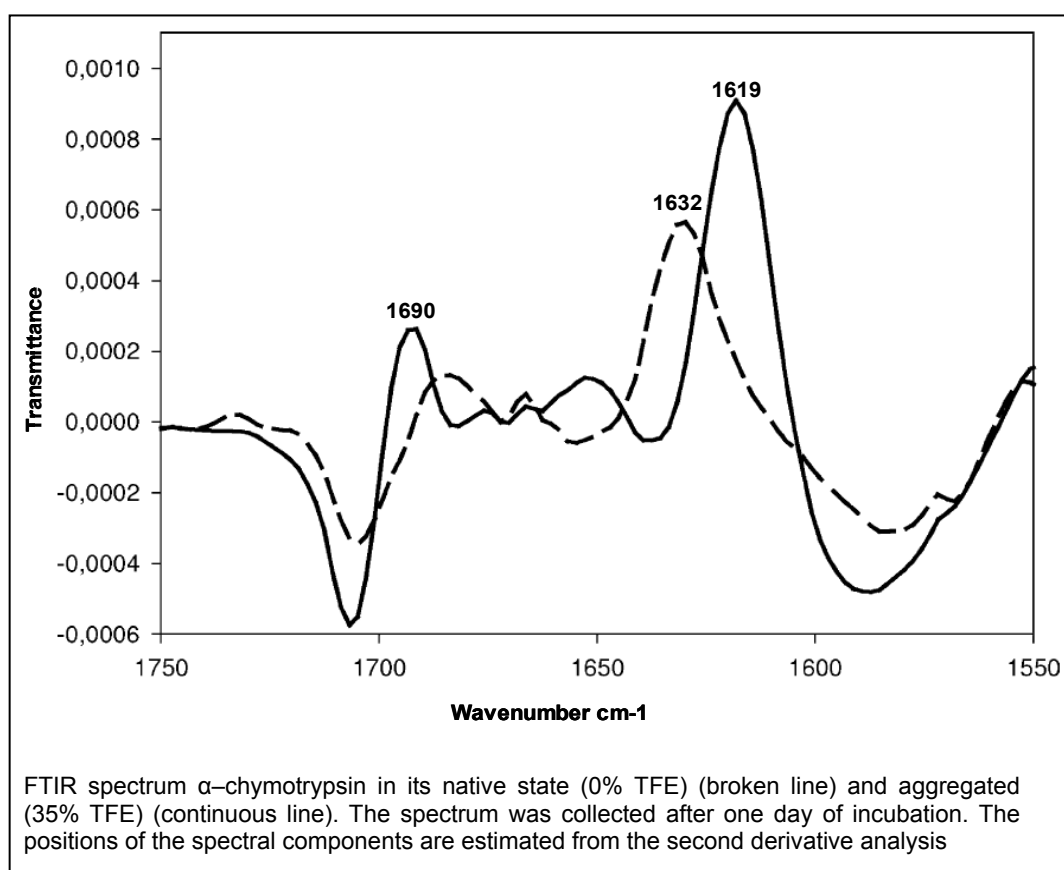
Figure 4.4.- Electron micrographs of α -chymotrypsin



These fibrils were examined by ATR-FTIR spectroscopy in order to assess whether they show the characteristic polypeptide chain arrangement of aggregated β -sheet structures or not. The FTIR spectrum of native α -chymotrypsin in the amide I region is dominated by a peak at 1632 cm^{-1} attributable to the presence of β -sheets in a β -barrel architecture (Fig. 4.5).

By contrast, aggregated material displays a shifted absorption peak centered at 1619 cm^{-1} , corresponding to an extensively extended β -sheet. In addition, the appearance of a band at 1690 cm^{-1} strongly suggests the formation of intermolecular β -sheets, typical in protein amyloid filaments (Jaikaran *et al.*, 2001; Zurdo *et al.*, 2001).

Figure 4.5.- FTIR of α -chymotrypsin

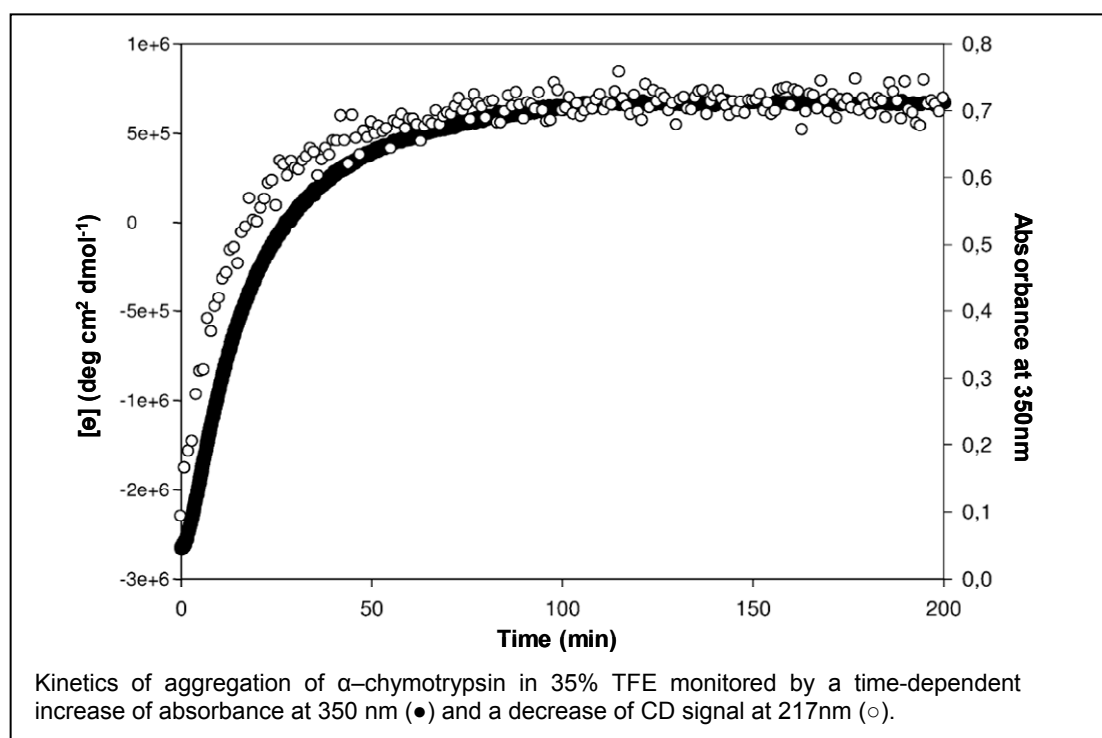


4.4.4.- Kinetics of aggregation

The kinetics of aggregation of α -chymotrypsin in 35% TFE was investigated by monitoring the time-dependent changes(s) in the 350 nm absorbance. At 0.1 mg/ml protein concentration and after a very short lag/nucleation stage, α -chymotrypsin aggregates following hyperbolic kinetics and reaching maximal aggregation after three hours of incubation at 25 °C (Fig. 4.6).

A parallel CD evaluation of the aggregation process by measuring the intensity of the 217 nm band showed a time-dependent decrease of this band, which occurred following the same kinetics obtained from absorbance monitoring (Fig. 4.6).

Figure 4.6.- Kinetics of aggregation of α -chymotrypsin in 35% TFE



4.4.5.- Aggregation occurs from a partially structured intermediate

The experimental results obtained in different proteins models, strongly suggest that amyloid fibril formation involves the ordered self-assembly of partially folded soluble precursors (Booth *et al.*, 1997; MacPhee & Dobson, 2000; Chiti *et al.*, 2002; Glenner

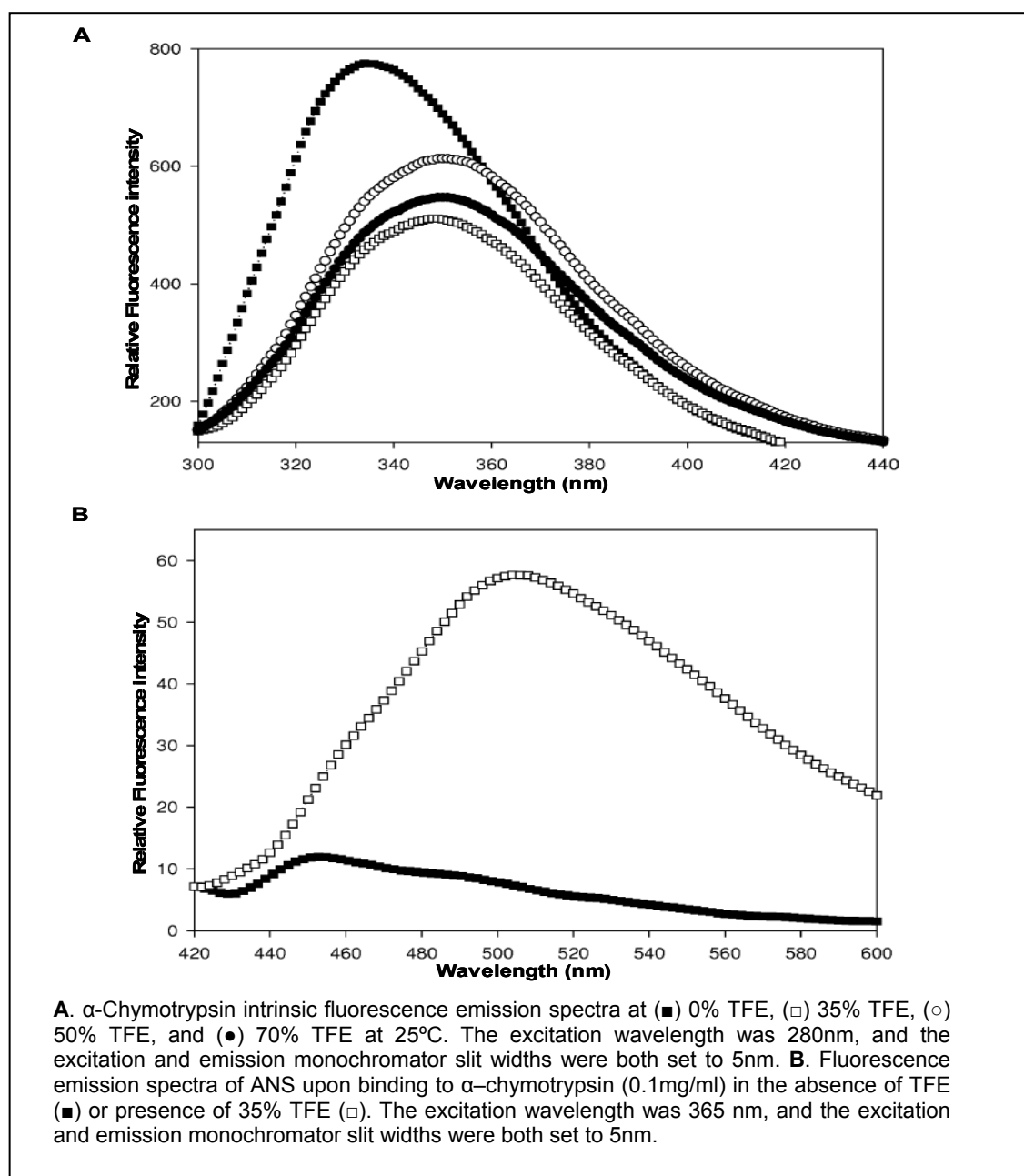
1980). With the aim of characterizing the conformational changes in α -chymotrypsin that precede its assembly into amyloid-like structures, we investigated the effect of increasing TFE concentrations on the tertiary structure by measuring changes in protein fluorescence emission. The fluorescence spectrum is determined chiefly by the polarity of the environment of the tryptophan and tyrosine residues and by their specific interactions and provides a sensitive means of characterizing proteins and their conformation, since the fluorescence emission maximum suffers a red shift when chromophores become more exposed to solvent, and the quantum yield of fluorescence decreases when the chromophores interact with quenching agents either in a solvent or in the protein itself.

The fluorescence spectrum of α -chymotrypsin shows an emission maximum at around 334 nm and does not change appreciably at TFE concentrations lower than 10%, but it red-shifts to 350 nm in the presence of 35% TFE, in addition to an intensity decrease to half of the original value, indicating an increased exposure of the tryptophan residues of the protein to the solvent (Fig. 4.7A).

After reaching a minimum at 35% TFE, the fluorescence emission of α -chymotrypsin increases substantially between 40% and 50%, suggesting a different burring of the aromatic residues upon induction of non-native helical conformation in the protein, to steadily decrease above 50% TFE (Fig. 4.7A). These data, in conjunction with those obtained using far-UV CD (indicating β -sheet segments in the protein) suggest that α -chymotrypsin at 35% TFE exists in a partially structured state(s), with extended β -sheets and loosely packed side-chains, which is different both from the native state and from the conformation achieved by the protein at higher concentrations of TFE.

The polarity-sensitive fluorescent probe aniline-8-naphthalene sulfonate (ANS) was also used in order to investigate the nature of the conformational states promoted by TFE. ANS binds to exposed hydrophobic surfaces in partially folded intermediates with much higher affinity than to native or completely unfolded proteins (Jones *et al.*, 1994; Safar *et al.*, 1994), resulting in a increase in fluorescence emission compared with the emission of free ANS in aqueous solution.

Figure 4.7. Conformational changes in α -chymotrypsin monitored by intrinsic and ANS fluorescence emission at various concentrations of TFE



No ANS binding to native α -chymotrypsin was detected, while the presence of 35% TFE in the protein solution led to an increase in fluorescence emission, indicating the exposure of hydrophobic clusters previously hidden in the native protein (Fig. 4.7B). Interestingly, high concentrations of TFE (>50%) also promote some ANS binding suggesting a certain exposure of non-polar surfaces in the non-native helical conformation of α -chymotrypsin (data not shown).

4.5.- DISCUSSION

α -Chymotrypsin folds into the canonical conformation of the serine proteases, one of the largest and most comprehensively studied enzyme families. Their common fold consists of two β -barrels, with the catalytic site found at the interface of the two domains (Perona & Craik, 1997). The structural determinants controlling specificity reside on surface loops, and this allows for the possibility of rapid and varied evolutionary divergence with conservation of the overall tertiary fold. Here, we show that this architecture can be forced to form amyloid-like structures. Our results are in good agreement with those obtained in another β -barrel protein, the acidic fibroblast growth factor from newt (Srisailam *et al.*, 2003; Srisailam *et al.*, 2002), and support the idea that amyloid fibril formation is an intrinsic property of many polypeptide chains (Dobson, 1999).

A large number of diseases are now recognized as being caused by the presence of a non-native protein conformation (Dobson, 1999; Rochet & Lansbury, 2000; Kopito & Ron, 2000). It is also generally accepted that pathological protein aggregation and fibril formation arise from a partially folded conformation of the aggregating protein (Fink, 1998). Proteins in their native state locate most of their non-polar side-chains and main-chain hydrogen bond formers inaccessible to solvent in order to reduce no desired intermolecular bonding and thus aggregation. On the other hand, strong denaturing conditions usually result in soluble and unfolded protein conformations because the non-covalent inter- and intramolecular interactions that a protein may form are avoided. Hence, conditions in which partially folded conformations are populated appear as more favorable to aggregation than those that populate the native or unfolded conformations. TFE is a solvent that diminishes native tertiary interactions and induces partly folded states of polypeptides by disrupting hydrophobic contacts without preventing the establishment of intermolecular interactions (Gast *et al.*, 1999). In this respect, a TFE stabilized partly folded state and, an aggregation prone conformational state of a pathological protein are similar.

Our results indicate that the conformational transitions induced by TFE in α -chymotrypsin occur in two stages. In the first stage, the hydrophobic contacts stabilizing the native β -barrel structure are disorganized, resulting in the formation of an extended β -sheet conformation. In the second stage, portions of the backbone of the protein form non-native helix conformation. Formation of non-native extended

β -sheet conformation appears to promote aggregation whereas the acquisition of non-native helix conformation reduces this propensity. This suggests that the aggregation of α -chymotrypsin is related to the nature of non-native secondary structural elements induced in the protein. Our results are in agreement with those obtained from the exhaustive analysis of point mutations in the muscle acylphosphatase where it was found that protein aggregation correlated significantly with the β -sheet propensity of the point mutated protein regions (Chiti *et al.*, 2002). In addition, both in muscle acylphosphatase and in the activation domain of procarboxypeptidase A2, mutations that stabilize helix conformation have been shown to decrease the aggregation process, whereas those that destabilized the helices increased the aggregation rate (Villegas *et al.*, 2000; Chiti *et al.*, 2002).

The partially structured intermediate state of α -chymotrypsin that is populated at moderate fluoro-alcohol concentration has a sticky nature, which promotes its gradual ensemble into higher order amyloid-like structures (Fig. 4.4). The presence of solvent-exposed, hydrophobic surfaces in the partially folded state is evident from its ability to bind ANS (Fig. 4.7C). Similarly, the disorganization of the native β -barrel architecture into an extended β -sheet conformation in the intermediate state results in a higher exposure of aromatic residues to solvent as seen by the consequent decrease of intrinsic fluorescence emission (Fig. 4.7A). Aggregation of the protein in TFE concentrations ranging of 20% to 40% is probably triggered by intermolecular hydrophobic contacts between the sticky-extended β -sheet regions in the partially folded intermediate. Rearrangement and annealing of the extended β -sheet elements results in the formation of amyloid-like structures in which the different polypeptidic chains are condensed in an extended β -sheet manner, as deduced from the ATR-FTIR spectrum of the TFE-induced aggregated material (Fig. 4.5). These structures are probably stabilized by hydrogen bonds between different polypeptidic chains resulting in the formation of intermolecular β -sheets. The reduced aggregation propensity observed at higher TFE concentrations (i.e., 50% TFE) is possibly related to partial shielding of the sticky hydrophobic clusters present in the extended β -sheet conformation inside a non-native helical structure, as can be deduced from the increase of intrinsic fluorescence observed under these conditions (Fig. 4.7A). In addition, the low polarity of TFE will weaken the hydrophobic interactions that stabilize the intermolecular β -sheet structures in the aggregates, thus favoring the formation of internal hydrogen atoms bonds over intermolecular bonding.

From our results it appears that protein aggregation depends both on the population of a partially structured state and on the backbone conformation of this intermediate. Formation of non-native β -sheet conformation promotes aggregation and induction of non-native helical segments in the backbone appears to reduce aggregation. Amyloid formation seems to depend on the presence of mild denaturing conditions that result in the formation of partially structured states of proteins with β -sheet structure, such as those observed here and reported for several disease-related proteins under conditions where fibrils develop readily (Kelly, 1998b; MacParland *et al.*, 2002). These partially folded states are likely to possess a high aggregation propensity as both hydrophobic side chains and main-chain hydrogen bond donors and acceptors are at least partially exposed. Myoglobin constitutes a remarkable exception to this rule, since its fibrillization correlates with conditions in which the polypeptide chain is unfolded, rather than with those that promote only partial unfolding (Fändrich *et al.*, 2003). This apparent discrepancy can be understood if one devises amyloid fibril formation as a selective process which occurs under conditions selectively unfavorable to native protein folding but that still permit the formation of non-covalent intermolecular contacts. While in the case of myoglobin these conditions occur only in residual or transient structured states, they seem to be particularly favorable to the population of partially structured intermediates. In this context, the presence of partially folded states with β -sheet conformation, such as the one described here for α -chymotrypsin, are expected to be found coincident with amyloid formation in many cases, specially for β -sheet proteins. Overall, it appears that the presence of a high extent of native structure impedes the ensemble of proteins in ordered amyloid fibrils. Thus, therapeutic strategies designed to inhibit amyloid fibril formation through stabilization of the native state (Hammarstrom *et al.*, 2001; Hammarstrom *et al.*, 2003) appear as particularly attractive, since they will prevent the formation of partially folded or unfolded precursors which may act as triggers of the disease.

CHAPTER 5

Prediction of "hot spots" of
aggregation in disease-linked
polypeptides

5.1.- SUMMARY

The polypeptides involved in amyloidogenesis may be globular proteins with a defined 3D-structure or natively unfolded proteins. The first class includes polypeptides such as β 2-microglobulin, lysozyme, transthyretin or the prion protein, whereas β -amyloid peptide, amylin or α -synuclein all belong to the second class. Recent studies suggest that specific regions in the proteins act as “hot spots” driving aggregation. This should be especially relevant for natively unfolded proteins or unfolded states of globular proteins as they lack significant secondary and tertiary structure and specific intra-chain interactions that can mask these aggregation-prone regions. Prediction of such sequence stretches is important since they are potential therapeutic targets.

In this study we exploited the experimental data obtained in an *in vivo* system using β -amyloid peptide as a model to derive the individual aggregation propensities of natural amino acids. These data are used to generate aggregation profiles for different disease related polypeptides. The approach detects the presence of “hot spots” which have been already validated experimentally in the literature and provides insights into the effect of disease-linked mutations in these polypeptides.

The proposed method might become a useful tool for the future development of sequence-targeted anti-aggregation pharmaceuticals.

5.2.- INTRODUCTION

In the last decade, protein aggregation has moved beyond being a mostly ignored area of protein chemistry to become a key topic in medical sciences (Smith, 2003), mainly because the presence of insoluble deposits in human tissues correlates with the development of many debilitating human disorders including the amyloidoses and several neurodegenerative diseases (Rochet & Lansbury, 2000). The proteins involved in these diseases are not related in terms of sequence or secondary structure content. From the conformational point of view, two major classes can be distinguished: globular proteins with a stable unique conformation in the native state and intrinsically unstructured proteins (Uversky & Fink, 2004). Globular proteins rarely aggregate from their native states and destabilization, resulting in an increased population of unfolded molecules, is well established as a triggering factor in disorders associated with the deposition of proteins that are globular in their normal functional states (Kelly, 1998b), as in the cases of β 2-microglobulin, lysozyme, transthyretin and the prion protein. Interestingly enough, many proteins involved in depositional disorders are mostly unstructured within the cell (Uversky & Fink, 2004). These include amylin, amyloid- β -protein, and α -synuclein, among others. In these cases, protein deposition does not require unfolding and can occur by direct self-assembly of the unstructured polypeptide chains.

One of the major unanswered questions of protein aggregation is the specificity with which the primary sequence determines the aggregation propensity from totally or partially unfolded states. Deciphering the answer to this question will give us a chance to control the unwanted protein deposition events through specific sequence-targeted therapeutics. A first advance in this direction is the recent discovery that not all regions of a polypeptide are equally important for determining its aggregation tendency, both in natively unfolded and globular proteins. In this way, some authors, including ourselves, have proved recently that very short specific amino acid stretches can act as facilitators or inhibitors of amyloid fibril formation (Ventura *et al.*, 2004; Ivanova *et al.*, 2004). These relevant regions are usually known as aggregation "hot spots". Aggregation-prone regions are likely to be blocked in the native state of globular proteins because their side chains are usually hidden in the inner hydrophobic core or already involved in the network of contacts that stabilizes a protein. This accounts for the protective role of the native structure against

aggregation (Dobson, 1999). In contrast, aggregation-prone regions are already exposed to solvent in natively unfolded proteins, available for the establishment of inter-molecular contacts that may finally lead to the formation of aggregates. Accordingly, the presence of putative “hot spots” of aggregation is much more frequent in the sequences of globular proteins than in those coding for natively unfolded proteins (Linding *et al.*, 2004). The presence of aggregation-prone regions has been described in most of the peptides and proteins underlying neurodegenerative and systemic amyloidogenic disorders (Fernández-Escamilla *et al.*, 2004).

We have used a simple *in vivo* system to study the aggregation effects of a complete set of mutations in one of the best characterized “hot spots” in a disease-linked protein: the central hydrophobic cluster (CHC) of the Amyloid- β -protein (A β) (De Groot *et al.*, 2005; Bitan *et al.*, 2003). The results in this and other studies on protein models not related to disease (Chiti *et al.*, 2003), suggested that common and simple principles underlie protein aggregation, at least from totally or partially unfolded states, and that the propensities of proteins backbones to aggregate are sharply modulated by the sequences that dress them. Based on these assumptions, we have developed a simple approach that identifies the presence of “hot spots” of aggregation in globular and unstructured disease-linked polypeptides and predicts the aggregation effects of mutations in their sequences.

5.3.- MATERIALS AND METHODS

5.3.1.- Experimental determination of amino acids aggregation propensities

The CHC of A β 42 peptide was chosen as a paradigmatic aggregation-prone region for the calculation of the individual effect of each natural amino acid on protein aggregation. The specific effect on A β 42's deposition promoted by the 20 different natural amino acids when located in the central position of this model "hot spot" were evaluated. Briefly, the wild type A β 42 gene and its 19 mutants were inserted as a fusion protein upstream of the green fluorescence protein (GFP) and expressed individually in bacteria. In this system, the levels of GFP fluorescence in the cells depend exclusively on the *in vivo* aggregation propensity of the A β 42 variant (de Groot *et al.*, 2005; Wurth *et al.*, 2002), in such a way that changes in aggregation propensities promoted by the different mutations can be easily monitored by measuring the fluorescence emission of the cells expressing each particular variant and normalizing it relative to that emitted by the cells bearing the wild type sequence. Three independent clones were analyzed for each mutation and each clone was analyzed at least by triplicate to generate consistent data. To obtain the individual aggregation propensities in Table 5.1, the change promoted by each amino acid was normalized relative to the average change of the pool of 20 amino acids.

5.3.2.- Generation of aggregation profiles and identification of "hot spots"

Different experimental data suggest that the aggregation of A β 42 occurs from a mostly unfolded conformation in which the CHC is exposed to solvent (Santini *et al.*, 2004). Assuming that the individual intrinsic aggregation propensities obtained analyzing this particular protein region will probably apply for any unfolded sequence; an aggregation profile was generated for every protein in this study through a simple assignment of the values in Table 5.1 to each individual residue in the corresponding sequences. Since "hot spots" are clusters of consecutive residues, the sequence was scanned by using a five residues sliding window. "Hot spots" in the sequence were

identified as those protein regions at least five residues in length (the minimal size shown to date to be required for a peptide to form amyloid fibrils similar to those formed by whole polypeptides (Haspel *et al.*, 2005), in which the aggregation propensity is above the average aggregation propensity of the complete sequence. The average propensity of the polypeptide was calculated as the sum of the aggregation propensities of its individual amino acids divided by the number of residues.

5.3.3.- Analysis of the effect of changes in the polypeptide sequence on aggregation

The concept of “hot spot” of aggregation implies that the contribution of a particular residue in a protein sequence on protein aggregation is somehow modulated by its immediate neighbors. According to this, the effects of mutation on protein aggregation can not be properly calculated by a simple subtraction of the intrinsic aggregation propensities of the wild type and mutant residues. Instead, to provide a more general description of the effect of the change on the overall aggregation propensity, the individual aggregation profiles for the wild type protein and the different mutants are obtained and the differences between the areas below the corresponding profiles are calculated. The area between each profile was always normalized by the number of residues in the considered species to compare between the aggregation propensities of the complete protein and fragments coming from proteolysis, chemical synthesis or other processes.

The difference between normalized areas, multiplied by a 100 factor, was designed as the change in average aggregation propensity (ΔAP). ΔAP will be positive if the mutation is predicted to increase the aggregation propensity of the polypeptide chain and negative if it is predicted to increase solubility.

5.4.- RESULTS

5.4.1.- Aggregation propensities of natural amino acids

The rationale behind our study is based on two recent observations in the field. First, not all the polypeptide sequence is relevant for the aggregation of a given protein, but rather there exist specific regions that drive the process (Ventura *et al.*, 2004; Ivanona *et al.*, 2004) and second, similar simple rules appear to underlie the aggregation propensities of unrelated proteins from unfolded states (Chiti *et al.*, 2003). According to these two assumptions one may expect that the conclusions obtained from the study of a relevant “hot spot” of aggregation in a specific protein could apply to other unrelated proteins involved in disease. As commented upon previously, we have exploited an *in vivo* reporter method to calculate the relative aggregation propensities of each individual natural amino acid when placed in the central position of the CHC of A β (see Material and Methods). The highest aggregation propensities correspond to isoleucine, phenylalanine, valine, and leucine, whereas aspartic, glutamic, asparagine, and arginine exhibit the lowest (Table 5.1). In general, hydrophobic residues tend to induce aggregation whereas polar ones promote solubility, matching the general assumption that hydrophobic interactions are supposed to play an important role in protein aggregation (Fink *et al.*, 1998).

5.4.2.- Generation of protein aggregation profiles and prediction of the effects of protein mutation on the aggregation propensity

Provided that a given polypeptide aggregates from an at least partially unstructured state, the experimental intrinsic aggregation propensities shown in Table 5.1 should apply independently of the protein context. Thus, a profile can be theoretically generated for any protein sequence to detect those regions with aggregation propensities above the average value of the whole sequence. This leads directly to the definition of “hot spot” of aggregation as a certain region that displays higher aggregation propensity than the rest of the sequence. Interestingly, a related approach has been reported very recently for the analysis of unstructured proteins

associated with neurodegenerative diseases (Pawar *et al.*, 2005). A good number of natural occurring mutations have been reported in proteins associated to depositional diseases. In many cases they result in changes in the global protein aggregation propensity and sometimes in the appearance of premature or acute pathological symptoms. The change in average aggregation propensity (ΔAP) between the wild type and the different mutants should predict the effect of sequence variations on the aggregation propensities, provided that they rely on changes in the intrinsic polypeptide properties.

Table 5.1.- Relative experimental aggregation propensities of the 20 natural amino acids derived from the analysis of mutants in the central position of the CHC in amyloid- β -protein

Amino acid	Aggregation propensity
I	1.822
F	1.754
V	1.594
L	1.380
Y	1.159
W	1.037
M	0.910
C	0.604
A	-0.036
T	-0.159
S	-0.294
P	-0.334
G	-0.535
K	-0.931
H	-1.033
Q	-1.231
R	-1.240
N	-1.302
E	-1.412
D	-1.836

5.4.3.- Analysis of disease-related polypeptide sequences

In this section the above described analysis is applied to a set of proteins linked to depositional diseases and the obtained results are compared with the available experimental data.

5.4.4.- Intrinsically unstructured proteins

Amyloid- β -protein

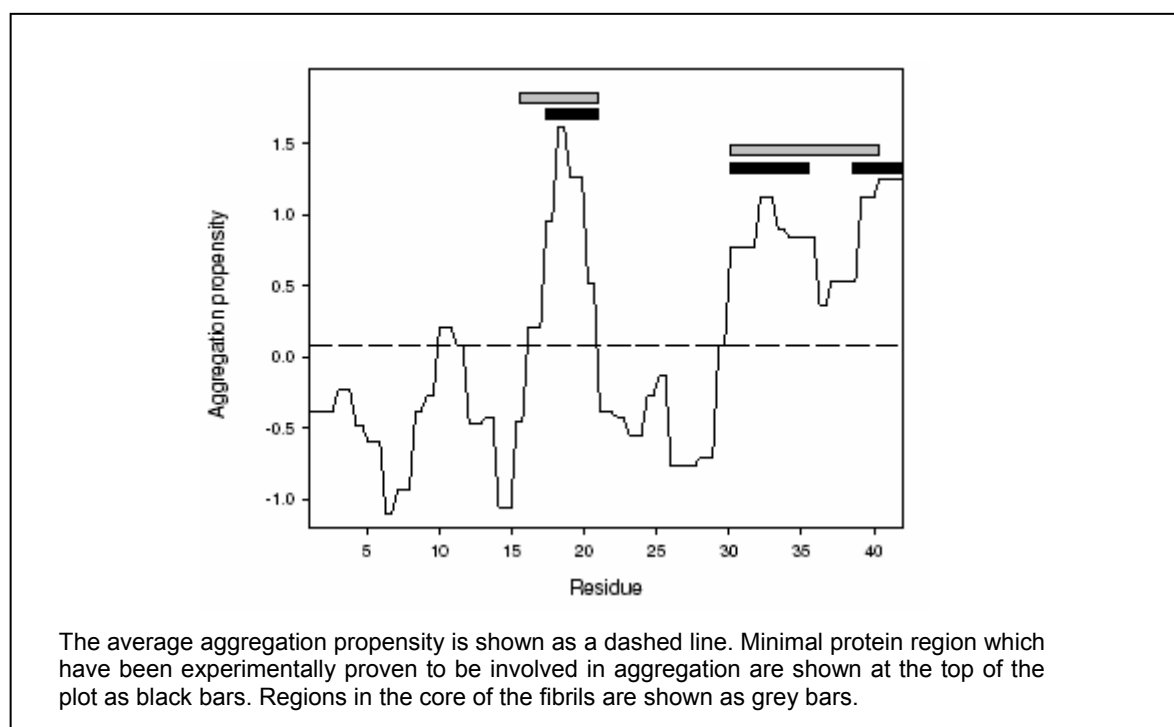
As a proof of principle, our approach was first tested in the molecule from which the experimental amino acid aggregation propensities were derived. Alzheimer's disease (AD) is a progressive neurodegenerative disorder characterized by the patient's memory loss and impairment of cognitive abilities. The extracellular amyloid is found in the brain and is widely believed to be involved in the progression of the disease (Selkoe, 2001). The principal component of the lesions is the hydrophobic polypeptide A β . The most abundant forms found in amyloid plaques are a 40-mer (A β 40) and a 42-mer (A β 42). Although less abundant, A β 42 is more amyloidogenic than A β 40 and is the major component of neuritic plaques (Nagele *et al.*, 2005).

Two main regions with high aggregation propensity can be distinguished in the aggregation profile for this polypeptide (Fig. 5.1). The second region arises from the contribution of two sequence stretches comprising residues 30-36 and 38-42, respectively. The predicted aggregation-prone regions are in excellent agreement with the experimental data in the literature. Residues 16-21 overlap with the CHC sequence comprising residues 17-21, a particular region recognized to play a key role in A β aggregation and that is defined as specially relevant for the amyloidogenesis of the A β 40 and A β 42 peptides by two recent proline-scanning-mutagenesis studies (Williams *et al.*, 2004; Morimoto *et al.*, 2004). In addition, structural studies using solid state-NMR (Petkova *et al.*, 2002) and site-directed spin labeling (Torok *et al.*, 2002) have revealed that residues 16-21 are located in the core of A β fibrils. Accordingly, a short 7 residues fragment comprising residues 16-22 is able to form ordered amyloid fibrils (Balbach *et al.*, 2000) and, more interestingly, 16-LVAFF-20 and derived peptides have been shown to bind to A β 42 and act as potent inhibitors of amyloid formation (Findeis *et al.*, 1999). The region 30-42, including both 30-36 and 38-42 stretches, has been also implicated in A β aggregation. Proline-scanning-mutagenesis revealed that the region 31-36 is sensitive to proline

replacement and likely to include a β -sheet portion of the A β fibrils (Williams *et al.*, 2004; Morimoto *et al.*, 2004). The contribution of the C-terminal region 38-42 to A β amyloidogenesis becomes clear from the observation that, although A β 40 is produced in greater abundance in vivo, the prevalence of the full-length 42-mer in plaques is much higher (Nagele *et al.*, 2005).

Experiments with truncated synthetic A β peptides have confirmed that A β 39 and A β 40 are kinetically soluble for several days, whereas A β 42 immediately aggregates into amyloid fibrils (Jarrett *et al.*, 1993). The relevance of the predicted 30-42 region is confirmed by structural studies that demonstrate that residues 30-40 are located in the core of the A β fibrils (Torok *et al.*, 2002).

Figure 5.1.- Aggregation profile of amyloid- β -protein



A set of mutations in the CHC and adjacent positions of A β 42 is intimately associated to early-onset familial Alzheimer diseases (FAD). The substitutions include A21G (Flemish), E22Q (Dutch) and E22G (Arctic) (Yamamoto *et al.*, 2004a). A β 42 congeners bearing these mutations display distinct aggregation kinetics. The rate of fibril formation by the Flemish mutant is decreased relative to WT A β 42, whereas the Dutch mutant peptide aggregates substantially faster. The Arctic peptide does not shows an overall change in the rate of fibrillogenesis relative to WT A β 42, but rather

accelerated protofibril formation. To assess whether the effect of such mutations could be predicted by the present approach we calculated ΔAP for the different sequences.

The results obtained describe accurately the effects documented in the literature (Table 5.2). Adding to the mutations present in the population, a large set of mutations has been artificially introduced on A β that result in changes in its aggregation propensity. ΔAP values were also calculated for several of them and the results compared with the experimental data (Table 5.2).

Table 5.2.- Comparison of predicted and experimental changes in aggregation for A β variants

Mutation	ΔAP^*	Observed aggregation \ddagger
A21G	-1.22	-
E22G	+2.14	+
E22Q	+0.44	+
F19P	-5.09	-
F19T	-4.73	-
I31L	-1.07	-
I32L	-1.07	-
I41G	-5.76	-
I41A	-4.52	-
I41L	-1.075	-
A42G	-1.21	-
A42V	+3.95	+
$\Delta 1-4$	+4.82	+
$\Delta 1-9$	+21.86	+
$\Delta 40-42$	-8.34	-
$\Delta 41-42$	-4.26	-
V12E+V18E+M35T+I41N	-18.96	-
F19S+L34P	-9.18	-

* Change in average aggregation propensity

\ddagger Changes in aggregation determined experimentally

The calculated changes in aggregation propensity are in excellent agreement with the trends reported in the literature. Briefly, we predict the changes in aggregation of F19 mutants, those of I31 and I32 in the 30-36 region and those of I41 and A42 in the C-terminal region, as well as the effects of deletions both in the N and the C ends.

Finally, we also predict the high solubility of A β versions generated by random mutagenesis (Wurth *et al.*, 2002).

Islet amyloid polypeptide

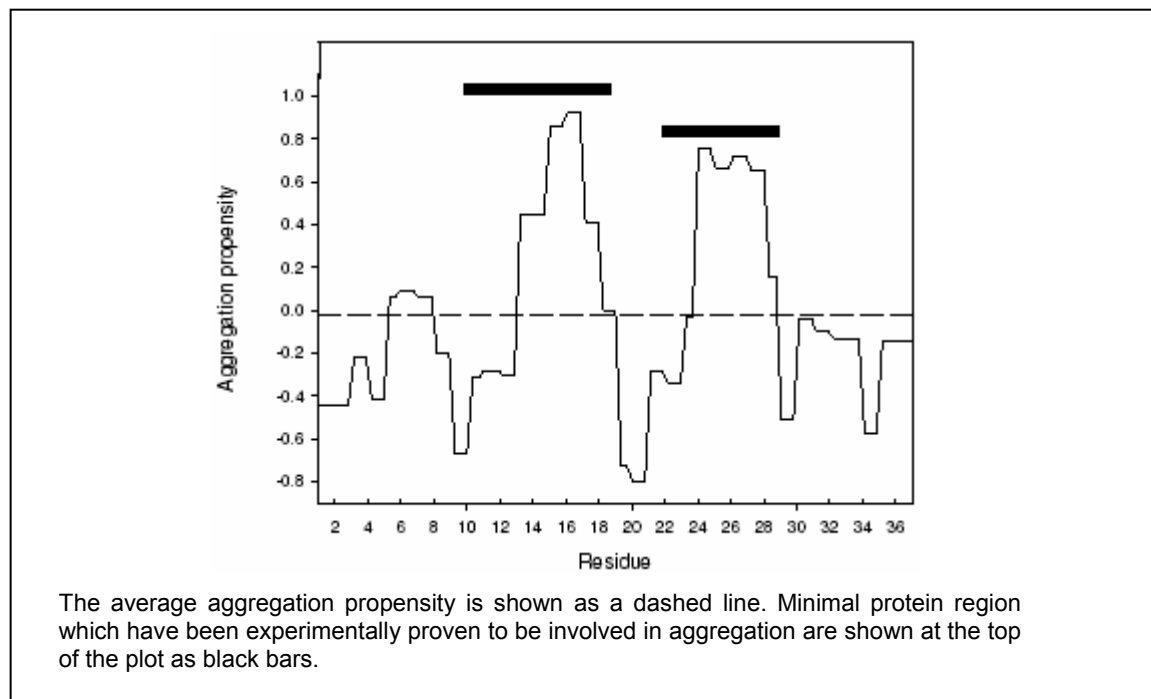
Type II diabetes is associated with progressive beta-cell failure manifested as a decline in insulin secretion and increasing hyperglycemia. A growing body of evidence suggests that beta-cell failure in type II diabetes correlates with the formation of pancreatic islet amyloid. Islet amyloid polypeptide (IAPP, amylin), the major component of islet amyloid, is co-secreted with insulin from beta-cells. In type II diabetes, this peptide aggregates to form amyloid fibrils that are toxic to beta-cells (Clark *et al.*, 1987). IAPP is an unstructured peptide hormone of 37 amino acid residues.

Two “hot spots” of aggregation comprising residues 12-18 and 22-28 are detected for this peptide (Fig. 5.2). Interestingly enough, a 8-37 IAPP-fragment including both “hot spots”, has been shown to form amyloid fibrils under physiological conditions (Goldsbury *et al.*, 2000). The two aggregation prone regions sharply coincide with those protected in the core of the fibrils in a recently described structural model of IAPP aggregates (Kajava *et al.*, 2005). In this study, residues 12-17 and 22-27 are proposed to form the inner β -sheets in the fibril protofilament structure. According to this hypothesis, peptides corresponding to residues 8-20, 10-19, 20-29 of human IAPP, which include one of the “hot spots” described here, all form amyloid (Scrocchi *et al.*, 2003; Tracz *et al.*, 2004; Moriarty & Raleigh, 1999). Smaller peptides derived from these regions have also been shown to form amyloid, and a recent investigation suggests that the minimal amyloid forming fragment of IAPP consists of residues 22-27. This hexapeptide fragment, NFGAIL, forms β -sheet-containing fibrils that coil around each other in typical amyloid fibril morphology (Azriel & Gazit, 2001).

The analysis also explains the available mutational data on IAPP. Diabetes associated IAPP amyloid occurs in primates and cats but not in rodents (Green *et al.*, 2003). Consistently, the sequences of peptides 20-29 of rodents display reduced average aggregation propensity relative to that of cat and human (Table 5.3). We also predict the slightly increased aggregation propensities of single or multiple mutations of rat IAPP to the corresponding residues of human IAPP (Green *et al.*, 2003): R18H, L23F or V26I, as well as the results from alanine-scanning-

mutagenesis in a peptide encompassing residues 22-27 (Azriel & Gazit, 2001) (Table 5.3).

Figure 5.2.- Aggregation profile of islet amyloid polypeptide



It has been found that a substitution at position 20 (S20G) in the IAPP molecule in a reduced subpopulation of Japanese people with type II diabetes is associated with an earlier onset and more severe form of disease (Sakagashira *et al.*, 1996). In this case, our approach does not predict an increased but a slightly reduced aggregation propensity in the mutant, suggesting that the pathological symptoms in this variant may arise from non-intrinsic factors. In fact, it has been suggested that the accelerated aggregation of the S20G variant could be related to structural reasons, resulting from a better packing of the turns connecting the β -sheets in the final protofilament structure (Kajava *et al.*, 2005) that cannot be predicted by the present approach. Several mechanisms have been proposed for IAPP fibril formation in type II diabetes. One widely accepted mechanism is that in type II diabetes, increased production and secretion of IAPP associated with increased demand for insulin might result in accumulation and aggregation of IAPP (Porte & Kahn, 1989). A second view considers that impaired processing of the IAPP precursor molecule, proIAPP, by islet beta-cells may lead to hypersecretion of unprocessed or partially processed forms of proIAPP that may have a higher tendency for aggregation compared to mature IAPP

(Porte & Kahn, 1989). Our calculated average aggregation propensities for proIAPP and processed IAPP support this view (Table 5.3).

Table 5.3.- Comparison of predicted and experimental changes in aggregation for IAPP variants, relative to the corresponding human IAPP sequence

Variant	ΔAP^*	Observed aggregation ‡
(20-29) Cat	-5.12	=
(20-29) Rat	-16.46	-
(20-29) Hamster	-32.73	-
R18H	+0.94	+
L23F	+1.70	+
V26I	+0.42	+
R18H+L23F+V26I	+3.06	+
(22-27) N22A	+31.53	+
(22-27) F23A	-42.96	-
(22-27) G24A	+11.96	-
(22-27) I26A	+44.5872	-
(22-27) L27A	-33.99	+
S20G	-1.09	+?
ProIAPP	+31.40	+

* Change in average aggregation propensity

‡ Changes in aggregation determined experimentally

? Not yet proved experimentally

α -Synuclein

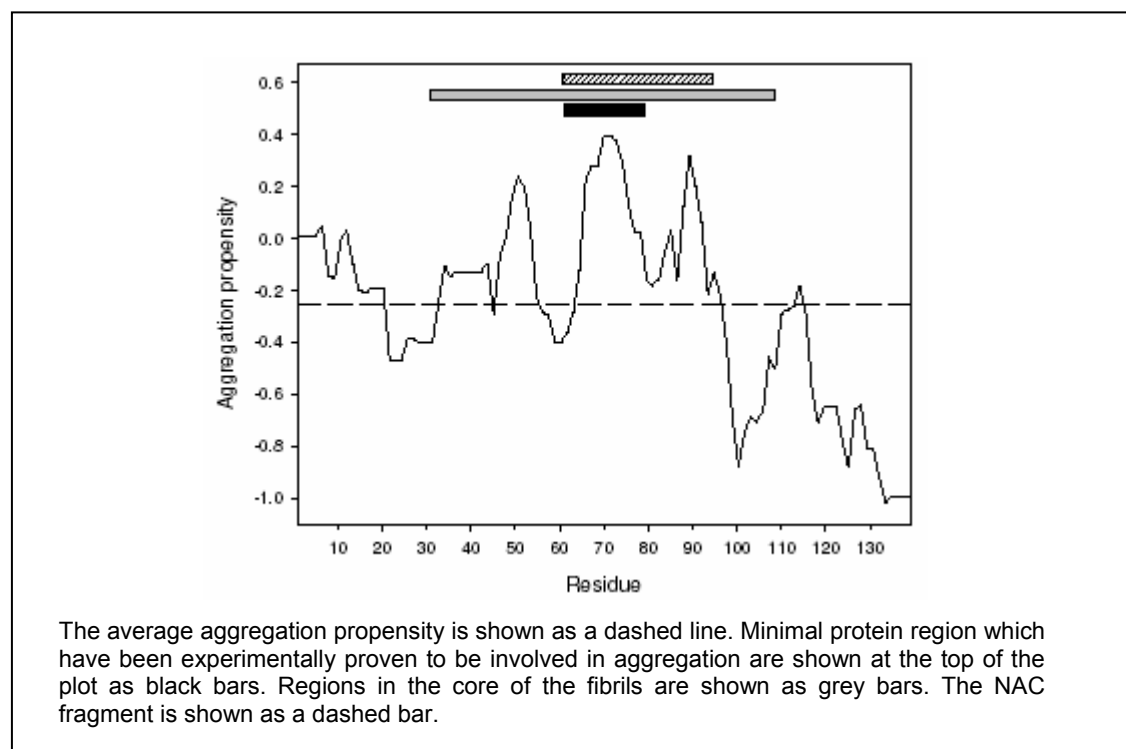
Parkinson disease is the most common neurodegenerative movement disorder and is pathologically characterized by the presence of neuronal intracytoplasmatic deposits of aggregated protein called Lewy bodies (Spillantini, 1997). Lewy bodies also occur in other cognitive disorders, globally known as α -synucleinopathies. α -synuclein is the major component of the fibrils that form the Lewy bodies (Spillantini, 1997). It is a small (137 residues), natively unfolded, soluble, presynaptic and highly conserved protein without a well-defined function.

The aggregation profile for this polypeptide is shown in Fig. 3. Several large aggregation-prone stretches were predicted for the α -synuclein sequence: region 1-18, region 27-56 and specially region 61-94. Again, our predictions are in complete

agreement with the experimental data in the literature, as many studies suggest that the central region of the protein, known as the non-A β component of amyloid plaques (NAC, amino acids 61-95), is the responsible for its aggregation process (Goedert *et al.*, 2001). A peptide comprising residues 68–78 of α -synuclein has been shown to be the minimum fragment that, like α -synuclein itself, forms amyloid fibrils and exhibits toxicity towards cells in culture (Bodles *et al.*, 2001). This fragment is included in the region 62-80 which we predict as the sequence stretch with the highest aggregation propensity. All the α -synucleinopathies are characterized by the accumulation of the 35 residues NAC fragment in the insoluble deposits (Goedert *et al.*, 2001). Accordingly, this central region is predicted to have a much higher average aggregation propensity than its soluble precursor (Δ AP = +38.47). The importance of this hydrophobic stretch is further supported by its absence in β -synuclein, a homologue of α -synuclein, with strongly reduced propensity for fibril formation. It has been shown that the deletion of amino acids 71-82 within the hydrophobic region abrogated the ability of human α -synuclein to polymerize into fibrils (Giasson *et al.*, 2001).

Protease digestion studies suggest that the core region of α -synuclein in the fibrils could be longer, since a 7-kDa fragment (comprising residues 31–109) was shown to be protected from proteinase K digestion (Miake *et al.*, 2002). This region contains the putative 12-residue core domain, as well as the NAC region and includes the second and third “hot spots” in our profile. A structural study on the organization of α -synuclein in the fibrillar state using site-directed spin labeling confirms that the 34-101 residues region constitutes the core of the fibrils forming a parallel in-register β -sheet structure whereas the N terminus is structurally more heterogeneous and the C terminus (40 amino acids) is completely unfolded (Der-Sarkissian *et al.*, 2003).

Several α -synuclein mutations appear associated with familial early-onset Parkinson Disease: A30P, A53T and E46K. All they map into our predicted second “hot spot”. The rates of fibril assembly of the E46K and A53T mutants have been shown to be greater than those of the wild type and A30P proteins (Choi *et al.*, 2004). We predict a similar average aggregation propensity for the wild-type and the A30P mutant and a slightly increased aggregation propensity for the E46K mutant, but fail to foresee the effect of the A53T mutation in promoting the formation of protofibrils.

Figure 5.3.- Aggregation profile of α -synuclein

Obviously, other functional factors apart from the intrinsic aggregation propensities can strongly influence the aggregation tendency of unfolded polypeptide chains within the cell. In fact the effects of α -synuclein mutations have been associated either to an impaired degradation inside lysosomes or to a reduced axonal transport of the variants (Cuervo *et al.*, 2004; Saha *et al.*, 2004). Both situations may result in increased concentrations of the protein in certain regions of the neuron that may favor the nucleation step of amyloid formation. According to this, α -synuclein gene triplication identified in two independent families (Huang *et al.*, 2004) has been shown to accelerate the development of Parkinson disease. Thus, an increase in the amount of cellular α -synuclein appears to be important for the pathogenesis of Parkinson disease, suggesting that the effects of the different α -synuclein mutations on protein aggregation could be quantitative, in terms of local concentration, rather than qualitative. Thus, experimental deviations from the theoretical predictions in natively unfolded proteins, in addition to reflect limitations of the approach, might also contain relevant information, prompting to find alternative structural, as in the case of amylin, or functional, as in the case of α -synuclein, explanations for the observed behavior.

5.4.5.- Globular proteins

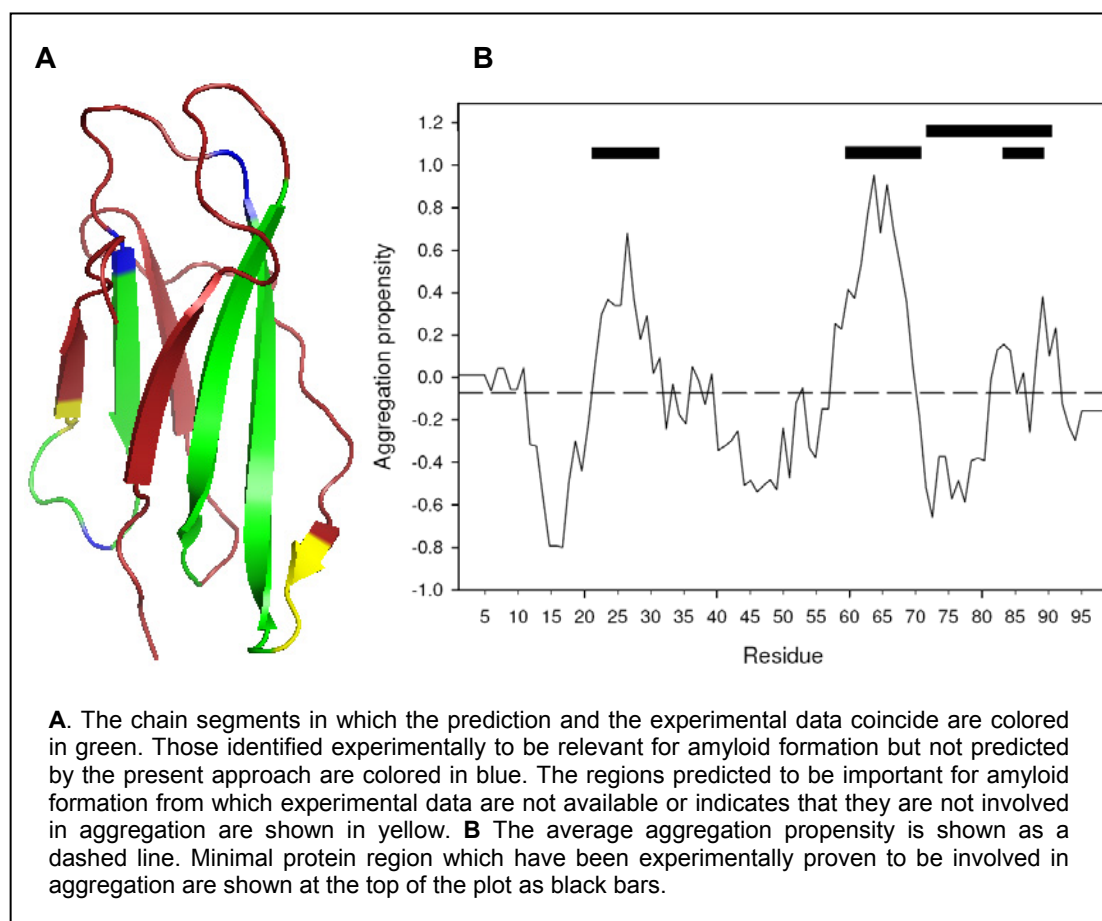
β 2-Microglobulin

β 2-Microglobulin-related amyloidosis is a common and serious complication in patients on long term hemodialysis (Koch *et al.*, 1992). Intact β 2-microglobulin is a major structural component of the amyloid fibrils. β 2-Microglobulin (β 2-m) is a small (99 residues) non glycosylated protein with an immunoglobulin-like fold consisting in two antiparallel pleated β -sheets linked by a disulfide bond (Fig. 5.4A). β 2-m has been shown to form amyloid fibrils *in vitro* under different conditions, but in all cases β 2-m populates unfolded non-native states as precursors to fibril assembly (McParland *et al.*, 2000). Under these conditions aggregation-prone regions, if present, may promote and drive the aggregation event. According to the analysis of the aggregation profile, shown in Fig. 5.4B, this protein displays four “hot spots” encompassing residues 21-31, 56-69 and 79-85, and 87-91. These regions sharply coincide with four different secondary structure elements in β 2-m: β -strand 2, formed by residues 21-31; β -strand 6, formed by residues 61-71; β -strand 7; formed by residues 77-85 and β -strand 8, formed by residues 86-95 (Fig. 5.4B). In agreement with our prediction a peptide comprising residues 21-41 has been shown to form fibrils in isolation (Kozhukh *et al.*, 2002). In addition, a N-terminal fragment of this short peptide corresponding exactly to our first “hot spot” [21-31] is also able to self-assemble into fibrillar structures (Hasegawa *et al.*, 2003). Interestingly enough, the peptides 23-31 and 21-29 exhibited reduced amyloidogenesis (Hasegawa *et al.*, 2003). Thus, in this particular “hot spot” the prediction delimits not only the overall region important for aggregation but also its precise size. The amino acid stretches 59-79 and its shorter version 59-71 which overlap with the predicted second aggregation-prone region of β 2-m have been also shown to form fibrils (Jones *et al.*, 2003). The C-terminal fragment 72-99 of β 2-m has been also reported to form amyloid (Ivanova *et al.*, 2003). This 29 residues sequence includes our third and fourth “hot spots” of aggregation. The peptide 91-99 does not aggregate, indicating that the last 9 residues of β 2-m are not relevant for amyloidogenesis as predicted here (Hasegawa *et al.*, 2003). The N-terminal region, for which no aggregation propensity is predicted, is probably not involved in the aggregation process as evidenced by the fact that the fragment 6-12 does not form fibrils (Hasegawa *et al.*, 2003). This observation could be physiologically relevant since the N-terminus of β 2-m is truncated in 30% of the molecules extracted from ex vivo fibrils (Bellotti *et al.*,

1998). In contrast to the human protein, mouse β 2-m does not form fibrils even at high concentration (Ivanova *et al.*, 2004). Based on this observation a seven residues region corresponding to residues 83-89 of human β 2-m has been suggested to be particularly important for aggregation, since it corresponds to the sequence with the highest divergence between both species. This hypothesis has been tested experimentally, since a heptapeptide bearing the human sequence is able to self-assemble whereas the mouse version is not (Ivanova *et al.*, 2004). The complete mouse sequence is predicted to have a strongly reduced aggregation propensity (Δ AP = -47.86).

Overall, our predictions on the presence and location of “hot spots” in β 2-m are extremely accurate and overlap with the experimentally found relevant regions (Fig. 5.4A). The observation that short peptides including the aggregation-prone regions described here form amyloids implies that exposure of previously hidden short segments can nucleate native proteins into the amyloid state and reinforces the hypothesis that fibril formation is sequence specific. One of the most urgent issues in the study of amyloid fibrils is to reproduce the formation of fibrils under physiological conditions. Recently, it has been found that low concentrations of SDS around the critical micelle concentration induce the extensive growth of β 2-m amyloid fibrils at physiological pH, probably through the SDS-induced conformational change of β 2-m monomers (Yamamoto *et al.*, 2004b). Contrarily to what was expected, the presence of low concentration of SDS had little effect on the stability of the protein and did not promote global protein unfolding.

Our results strongly suggest that in β 2-m the parts of the molecule involved in aggregation are located in pre-formed β -strands. Therefore, it is possible that local unfolding events may allow anomalous intermolecular interaction between this preformed elements leading to the formation of an aggregated β -sheet structure. This would explain the formation of amyloid deposits in hemodialysis patients in which no major unfolding of the protein is expected to occur, as well as the effect of seeds, which may have exposed aggregation prone β -strands, in strongly accelerating the aggregation process of β 2-m under physiological conditions (Kihara *et al.*, 2005).

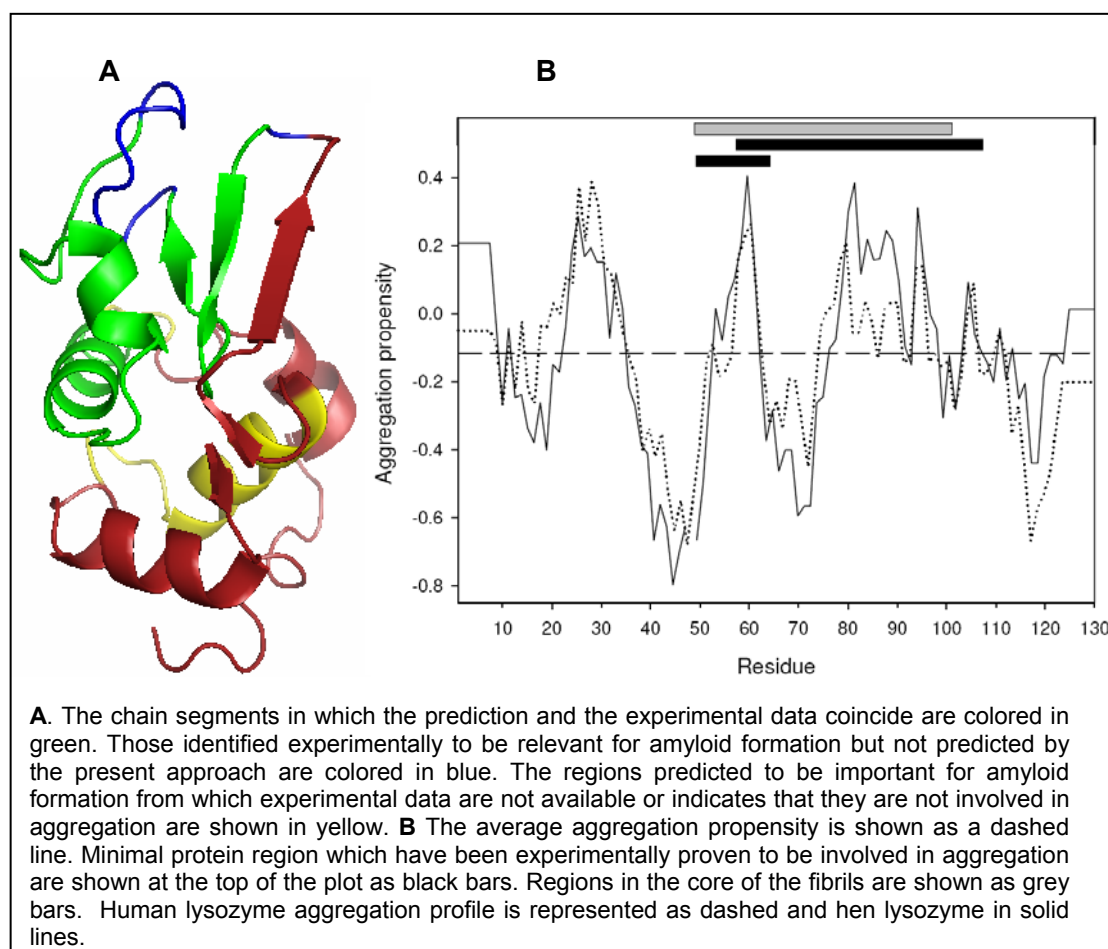
Figure 5.4.- Representation of the 3D structure and aggregation profile of β 2-Microglobulin

Lysozyme

Human lysozyme has been shown to form amyloid fibrils in individuals suffering from non neuropathic systemic amyloidosis. The disease is always associated to point mutations in the lysozyme gene and fibrils are deposited widely in tissues (Pepys *et al.*, 1993). The properties of two amyloidogenic lysozyme mutants (I56T and D67H) have been studied in detail and, when compared to those of the wild-type protein, the mutants were found to have reduced structural stability allowing unfolding to take place at least partially at physiologically relevant temperatures (Booth *et al.*, 1997; Canet *et al.*, 2002). Thus, the formation of amyloid fibrils by human lysozyme is likely to occur by the exposure of aggregation-prone region previously hidden in the native structure. The aggregation profile of lysozyme identifies three main “hot spots” corresponding to residues 20-34, 50-62 and 73-104 (Fig. 5.5B). The last large aggregation-prone region includes several local maxima. The first “hot spot” maps in

helix B, the second in a β -hairpin of the β -domain and the third includes helix C and a large flanking unstructured region at its N-terminus (Fig. 5.5A). Although there is no experimental characterization of amyloidogenic regions in human lysozyme in the literature, this information is available for the homologous hen lysozyme molecule, which displays an almost identical 3D-structure. The aggregation profile for the hen protein is very similar to that of the human one despite the fact that our input consists solely on the sequence and the identity between both molecules is only of 40 %. The equivalent “hot spots” in hen lysozyme comprise residues 24-34, 50-62 and 76-98. Experimental data suggests that the sequence of the β -domain could be of particular relevance for lysozyme aggregation since it unfolds prior to the α -domain (Canet *et al.*, 2002). Two peptides encompassing the β -domain of native lysozyme displayed very different behavior: peptide 61-82 appeared to be predominantly unstructured whereas peptide 41-60 showed a high tendency to aggregate and form extended β -sheet structures (Yang *et al.*, 1995). The first peptide coincides with a region of very low aggregation propensity in the aggregation profiles, whereas the second one covers the region with the highest aggregation propensity in the profile (residues 50-64). Interestingly enough, a peptide spanning residues 49-64 has been shown to form fibrils with the typical structure of amyloid showing that the first residues of the 41-60 peptide are not relevant for aggregation, as predicted by our approach (Krebs *et al.*, 2000). Another study has reported that the major fragment incorporated in the core of the fibril structure, as monitored using proteolysis, encompasses the chain region 49–101 (Frare *et al.*, 2004). These lysozyme fragments contain helix C and two of the three β -strands of the β -domain of the native protein structure and coincide with the limits of the second and third regions in our predictions (Fig. 5.5A and 5.5B). This observation could be biologically relevant, since the β -domain and C-helix of the human lysozyme have been shown to unfold locally in the amyloidogenic variant D67H, which is associated with the familial cases of systemic amyloidosis linked to lysozyme deposition (Canet *et al.*, 2002). The C-helix is the α -helix with the lowest helical propensity of hen lysozyme according to both theoretical and peptide based studies (Yang *et al.*, 1995).

Figure 5.5 Representation of the 3D structure and aggregation profile of lysozyme



This low propensity might be related to the ability of this region to be incorporated into the β -sheet rich fibrillar structured as have been reported for other protein systems (Kallberg *et al.*, 2001). Limited proteolysis of hen lysozyme renders fragments 57–107 and 1–38/108–129 (Frare *et al.*, 2004). In the 1–38/108–129 fragment the N-terminal and C-terminal ends of the molecule are joined by a disulfide bond. Only fragment 57–107, but not fragment 1–38/108–129, is able to generate well defined amyloid (Frare *et al.*, 2004). Whereas the behavior of the 57-107 fragment is expected from the analysis, one should also expect the fragment 1-38 to have a high tendency to aggregate.

Two explanations are possible to account for this discordance. First, it could occur that the helical structure of this region prevents its conversion to β -sheet conformation, since the A-helix displays the highest helical propensity out of all lysozyme helices (Yang *et al.*, 1995). The second possibility is that, being joined to the 108-129 region, predicted to have lower aggregation propensity, steric

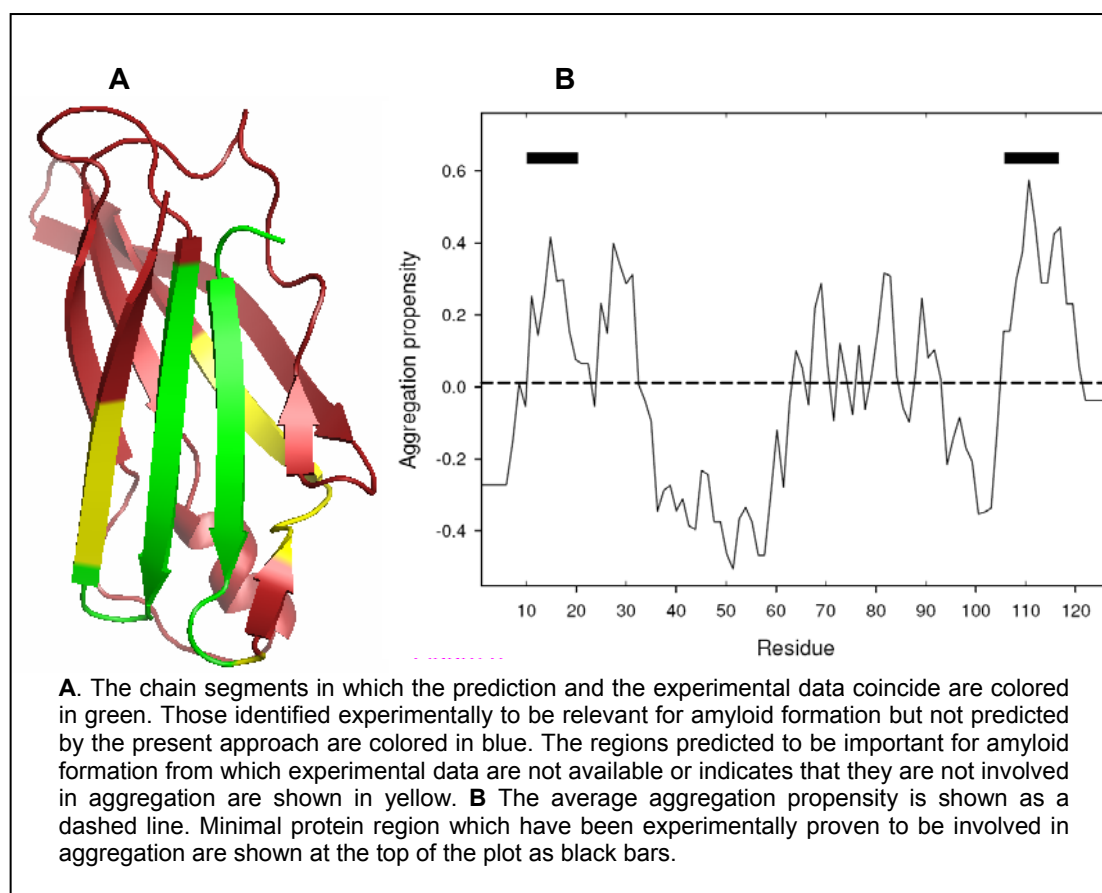
hindrances limit self-assembly or alternatively the average aggregation tendency of this peptide becomes reduced. The analysis supports this last hypothesis reporting a decrease in aggregation propensity ($\Delta AP = -5.34$) in the joined peptide respect the 1-38 peptide alone.

Transthyretin

Transthyretin (TTR) is a homotetramer of 127-amino acid subunits. TTR is found in human plasma and cerebral spinal fluid, the plasma form being the amyloidogenic precursor. TTR constitutes the fibrillar protein found in familial amyloidotic polyneuropathy (FAP) and senile systemic amyloidosis (SSA) (Saraiva *et al.*, 1984). In the case of FAP, the amyloid is associated with a point mutation in the TTR gene. To date, 100 different TTR mutations have been reported, many of which are amyloidogenic (McCutchen *et al.*, 1995). The FAP associated variants characterized thus far although tetrameric, are destabilized (Hornberg *et al.*, 2000). This destabilization allows tetramer dissociation to the amyloidogenic monomeric intermediate to occur under the influence of mild denaturing denaturation conditions. More than 10 FAP-related variants crystal structures have been solved, revealing that the tertiary and quaternary structures are essentially identical to the wild type form (Hornberg *et al.*, 2000). This observation suggests that the partial denaturation of TTR is a requirement for amyloidogenesis. In this state, the presence of “hot spots” of aggregation could play an especially important role in promoting/driving amyloid formation. According to the analysis of the aggregation profile shown in Fig. 6B, the TTR monomer displays three main “hot spots” encompassing residues 10-20, 23-33 and 105-118. Also in this case, aggregation-prone sequences appear to be located in preformed β -sheet structures: A β -strand (11-19), part of the B β -strand (28-36) and G and beginning of H β -strands (104-123) (Fig. 5.6A). Most of these secondary structure elements are involved in the formation of the tetrameric structure: H strands mediate the dimerization whereas A and G provide the contacts for the tetramerization of two preformed dimers. This explains the protective role played by the TTR quaternary structure against aggregation, since it hides or blocks most of the aggregation prone regions. Dissociation of the tetramer has been reported as a prerequisite for amyloidosis and according to our results might be associated to the exposure of previously hidden amyloidogenic sequences. We detect several short peaks exhibiting high aggregation propensities in the central region (63-94) of TTR. These result from the presence of almost regularly placed

residues with low aggregation propensity (Asp, Glu, Arg, Lys, Gly) in this rather hydrophobic sequence, which probably act as disrupters, significantly lowering the aggregation tendency of this 19-particular region, a strategy suggested to be used by nature to avoid edge-to-edge aggregation (Richardson & Richardson, 2002). To date two different fragments of TTR have been shown to form amyloid fibrils. The peptide 105-115 can be assembled into homogeneous amyloid fibrils with favorable spectroscopic properties (Jaroniec *et al.*, 2002). This has allowed to solve its fibrillar structure at high-resolution, showing that it adopts an antiparallel extended beta-strand conformation in the amyloid fibrils (Jaroniec *et al.*, 2004). This peptide coincides with the region with the highest aggregation propensity in the profile. Also in excellent agreement with the prediction, the peptide 10-20 is the only other fragment of TTR reported to form amyloid fibrils (Jarvis *et al.*, 1994). No data are available on the region 23-33 but the success of the present method in predicting relevant regions in TTR suggests that it is worth to characterize its *in vitro* aggregation capabilities.

Figure 5.6.- Representation of the 3D structure and aggregation profile of transthyretin

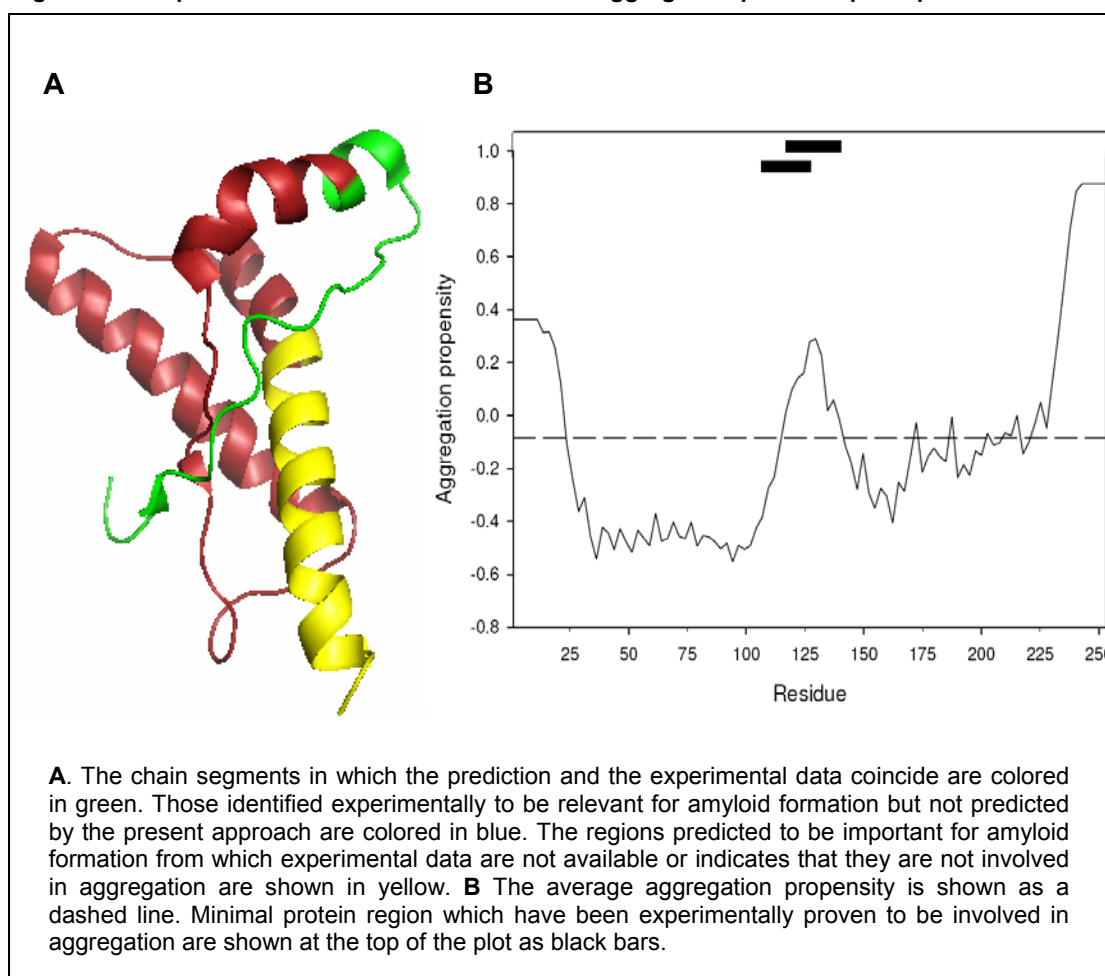


Prion protein

Misfolded isoforms of the naturally occurring prion protein (PrP) have been shown to be the causative agents in many mammalian neurodegenerative disorders, including Cruetzfeldt–Jakob disease (CJD) in human, scrapie in sheep, and bovine spongiform encephalopathy in cows. Prion infectivity is unique in that the pathogenic prion form (PrP^{Sc}) is involved in the conversion of the endogenous conformation (PrP^C) into transformed PrP^{Sc}. The "protein-only" hypothesis (Prusiner *et al.*, 1998) asserts further that no extraneous agents are necessary to explain the unusual behavior of prions. Prion diseases can have infectious, familial, and sporadic origins. The basic infectious mechanism is thought to be a conformational change of the normal prion protein (PrP^C) into the pathogenic PrP^{Sc} catalyzed by PrP^{Sc} itself. The normal prion protein (PrP^C) is a GPI-anchored glycoprotein constitutively expressed on the surface of primarily neuronal cells. It consists of two structurally different parts; a C-terminal, globular part mainly α -helical in nature (Fig. 5.7A) and an unstructured, N-terminal part (Kaneko *et al.*, 1997). Misfolding of PrP^C into PrP^{Sc} occurs post translationally and results in increased β -sheet content and gain of protease resistance. Fig. 5.7B shows the predicted "hot spots" in the aggregation profile of the full length human prion protein. They are located at the N-terminus (1-32), in the central region (105-146) and the C-terminus (208-252), respectively. The role of the detected aggregation-prone sequence at the N-terminus is uncertain since it is out of the protease resistant core of PrP^{Sc}. Little information exists about the role of this region, although it appears to be unnecessary both for prion transmission and aggregation. The predicted C-terminal "hot spot" includes almost all the C-terminal α -helix, named C, from the globular domain (Fig. 5.7A). Interestingly, some of the human mutations linked to Creutzfeldt-Jakob disease occur in this region of the prion protein and it has been related to the conversion of PrP^C into the toxic PrP^{Sc}. Moreover, some strains of PrP resistant to conversion to PrP^{Sc} have been found to bear mutations in helix C, and positions 214 and 218 have been shown to modulate PrP^{Sc} formation (Kaneko *et al.*, 1997). It is also important to note that the main structural differences between prion proteins from different species have been found at the end of helix C (Zahn *et al.*, 2000). The central region of PrP^C linking the unstructured N-terminal part with the globular C-terminal domain is believed to play a pivotal role in the PrP^C conformational changes. Extensive studies on the secondary structure and fibrillogenic properties of synthetic peptides of PrP have established that the continuous segment of the prion protein spanning residues 106–147, coincident with

the second “hot spot” predicted using our approach, is important for the fibrillogenic properties of the protein (Tagliavini *et al.*, 1993). One of the synthetic peptides, that named PrP106-126 within the central region of PrP and near the N-terminal of the protease resistant core of PrP^{Sc}, shares many properties with the infectious form as it readily forms amyloid fibrils with a high β -sheet content, shows partial proteinase K resistance and is neurotoxic *in vivo* (Gu *et al.*, 2002). The neurotoxicity of PrP106-126 depends on the expression of endogenous PrP^C which makes PrP106-126 a relevant model for PrP^{Sc} neurotoxicity (Gu *et al.*, 2002). Also another prion derived peptide - PrP118-135 - has been found to cause neuronal death via induction of apoptosis (Chabry *et al.*, 2003). The toxicity of PrP118-135 is, however, independent of endogenous PrP^C expression. Both peptides map in our predicted central aggregation-prone region of PrP^C.

Figure 5.7.- Representation of the 3D structure and aggregation profile of prion protein



5.5.- DISCUSSION

Overall, the method described here appears as a useful tool for the identification of protein regions that are especially relevant for protein aggregation and amyloidogenesis both in natively unfolded and properly folded globular proteins (Table 5.4). The results provide support to the hypothesis that short specific amino acid stretches can act as triggers for the incorporation of polypeptides into amyloid structures. It is interesting to note that in those cases in which structural information allows to delimitate the region incorporated in the core of the fibrillar structure, our predicted “hot spots” and those proved experimentally are considerably shorter than the whole region, suggesting that the role of “hot spots” is to act as specific nucleation points from which the ordered fibrillar structure is expanded. Nature has provided globular proteins with a reasonable conformational stability in the native state in which, as proved here, aggregation-prone sequences are buried or involved in intra-molecular interactions. This appears as a very successful evolutive strategy to avoid aggregation, since few proteins aggregate from their stable native conformation. Accordingly, amyloid-related mutations in globular proteins usually result in destabilization of the folded state allowing the exposure of previously hidden “hot spots”, as those reported here. This explains the scarce success in predicting the effect of mutations in the aggregation of globular proteins (data not shown), whereas the prediction of fatal sequence changes in intrinsically unstructured proteins involved in disease is generally accurate. The effects of such mutations can be explained in most cases by intrinsic factors, as they directly result in changes on the average propensity of the full polypeptide to aggregate. Besides providing important clues about the mechanism of protein aggregation, this study may be relevant for the therapeutics of amyloid disease, since the identified “hot spots” could be regarded as preferential targets to tackle the deleterious disorders linked to protein deposition. According to our results, different specific strategies should be employed when designing methods to avoid aggregation, depending on the disease being caused by natively unfolded or by globular proteins. In Alzheimer, type II diabetes and Parkinson diseases, shielding the already exposed aggregation-prone regions in the polypeptides by using small compounds or antibodies appears as a promising approach, whereas compounds that will stabilize the native conformation and avoid the exposure of the deleterious “hot spots” will be more effective in the

Chapter 5

case of globular proteins. Additionally, when gene therapy eventually comes to age, mutations that disrupt aggregation-prone regions in unstructured polypeptides or those which over-stabilize the native state of globular aggregation-prone proteins are expected to be useful approaches to avoid protein deposition and meliorate neurodegenerative and systemic amyloidogenic disorders.

Table 5.4.- List of the predicted "hot spots" in the different disease-linked polypeptides in this study and comparison with the available experimental data.

Protein	Predicted "Hot Spots"	Experimental "Hot Spots"	Regions in the fibrils
Amyloid-β-protein	16-21	+	+
	30-36	+	+
	38-42	+	+
Islet amyloid polypeptide	12-18	+	uncertain
	22-28	+	uncertain
α-Synuclein	1-18	No experimental data available	uncertain
	27-56	+	uncertain
β2-Microglobulin	61-94	+	+
	21-31	+	+
	56-69	+	+
	79-85	+	+
	87-91	+	+
	24-34	-	-
Lysozyme (hen)	60-62	+	+
	76-98	+	+
	10-20	+	+
Transthyretin	23-33	No experimental data available	uncertain
	105-118	+	+
Prion Protein	1-32	No experimental data available	uncertain
	105-146	+	+
	208-252	No experimental data available	uncertain

Experimental "hot spots" refer to those protein regions shown to be involved in the aggregation process of the corresponding polypeptide. It is also noted if the predicted "hot spot" has been described as a structural element of the amyloid fibrils formed by the different peptides and proteins in the study.

Annex to chapter 5

The work presented in this chapter has been done in tandem with Natalia Sánchez de Groot, a researcher from our group.

DISCUSSIÓ GENERAL

La present tesi és un reflex de les dues línies d'investigació principals del nostre grup de recerca. Per una banda, l'estudi de la relació estructura funció de les carboxipeptidases i els seus inhibidors, que constitueix la línia de treball en la que el grup hi porta treballant durant més anys. Aquesta tesi doctoral es va iniciar en aquest àmbit i els resultats obtinguts conformen la primera de les dues parts d'aquest treball. Per altra banda, la segona part de la tesi està centrada en l'estudi del plegament i l'agregació proteica, més concretament en l'estudi de models proteics en els que es pot induir la formació d'estructures amiloïdòtiques, i en el desenvolupament de mètodes de predicció de l'agregació de les proteïnes a partir de la seva seqüència. Aquest segon bloc correspon a la línia més recent d'investigació oberta al grup, en la que s'ha treballat durant la darrera etapa de la present tesi doctoral.

PART I

Les metal·locarboxipeptidases són un grup de proteases que requereixen d'un ió metàl·lic, normalment zinc, per a dur a terme la seva activitat enzimàtica. Aquesta consisteix en la hidròlisi del residu col·locat en posició C-terminal d'un pèptid o proteïna. Les metal·locarboxipeptidases es divideixen en dues subfamílies definides per criteris estructurals i evolutius: la subfamília A/B i la subfamília N/E. A nivell estructural, la diferència més significativa entre ambdues subfamílies és la presència de segments d'activació grans (75-76 residus per a procarboxipeptidases de bacteris o fongs i 91-96 per als altres organismes) en la subfamília A/B, mentre que els membres de la subfamília N/E presenten uns segments "pro" molt menors (5-17 residus) o no en presenten cap i, en canvi, contenen un domini homòleg a la transtirretina a l'extrem C-terminal. Malgrat que els membres de la primera subfamília se solen identificar amb els seus prototips d'origen pancreàtic i, per tant, amb un activitat digestiva de proteïnes exògenes, la seva activitat també s'estén a funcions reguladores no merament degradatives, sobretot en les formes no pancreàtiques. Tots els membres coneguts de la segona subfamília duen a terme funcions reguladores com pot ser l'activació de pro-hormones o neuropèptids, entre d'altres. Els membres de la subfamília N/E presenten especificitat envers residus C-terminals bàsics, mentre que els enzims de la subfamília A/B presenten una especificitat de substrat més variable. L'ampli ventall de funcions fisiològiques dutes a terme per les metal·locarboxipeptidases fa que tingui interès l'estudi de la regulació de la seva activitat. En els contextos cel·lular i extracel·lular, aquesta regulació és

duta a terme, principalment, per inhibidors proteics, la caracterització dels quals té interès biomèdic i biotecnològic tant pel que fa a la comprensió dels mecanismes involucrats com al disseny de molècules alternatives.

En la primera part d'aquest estudi es presenta l'estructura tridimensional de dues carboxipeptidases humanes de la subfamília A/B: en el capítol 1 la corresponent a la carboxipeptidasa A1 i en el segon capítol el complex de la CPA4 amb un inhibidor endogen, la latexina.

En els darrers anys s'ha anat ampliant la visió inicialment limitada de les metal·locarboxipeptidases A/B com uns enzims implicats exclusivament en la proteòlisis de les proteïnes de la ingesta, ja que s'han pogut localitzar enzims d'aquesta subfamília en teixits com el cervell o els pulmons i s'ha observat que participen en processos biològicament importants com la cascada de la coagulació, la regulació de l'activitat d'algunes hormones peptídiques o la resposta cel·lular, entre d'altres. La carboxipeptidasa A1 és l'enzim prototip de les carboxipeptidases de la subfamília A/B. Tot i que en treballs anteriors la CPA1 humana ha estat àmpliament estudiada en el nostre laboratori a nivell funcional i que es pot trobar treballs a la literatura descrivint la caracterització detallada de les estructures tridimensionals de les carboxipeptidases A1 bovina i porcina, no es disposa encara de cap treball estructural referent a la CPA1 humana. En aquesta tesi s'ha obtingut l'estructura tridimensional de la CPA1 humana a la resolució de 1,6 Å. El seu plegament global és molt semblant al que presenten les carboxipeptidases prèviament caracteritzades, és a dir, el plegament α/β d'hidrolases que conté una làmina β de vuit cadenes envoltada d'hèlix- α . El fet característic de presentar una molècula de PAA dintre del centre actiu permet extrapolar les interaccions que establiria un inhibidor amb els residus dels subetis S1' i S1 involucrats en la catàlisi i en la fixació del residu C-terminal del substrat, com són els residus Arg 127, Asn 144, Arg 145 i Tyr 248.

L'estructura de la CPA1 humana a alta resolució aporta detalls addicionals per a aprofundir en l'estudi del seu mecanisme catalític, extrapolable també al conjunt de les zinc metal·locarboxipeptidases i per a dissenyar inhibidors específics d'aquesta proteasa i d'interès biomèdic. A més a més, constitueix un motlle excel·lent per a modelar sobre d'ella l'estructura de carboxipeptidases de gran rellevància fisiològica com les implicades en el processament d'hormones i neurotransmissors.

Entre les carboxipeptidases A/B que tenen funcions reguladores s'hi troba la carboxipeptidasa A4. Aquesta és una forma que s'ha descrit com implicada en el càncer de pròstata, un dels càncers més diagnosticat en homes. Les carboxipeptidases d'aquesta subfamília són sintetitzades en forma de proenzim, i un cop es dona el tall activador que elimina el segment "pro" N-terminal, l'activació és irreversible. Donada aquesta irreversibilitat del procés d'activació, la regulació de la activitat carboxipeptidasa mitjançant inhibidors específics és de gran interès en general i, específicament, en el control d'enzims possiblement lligats al progrés de malalties. El nombre d'inhibidors aïllats a partir de fonts naturals és, però, reduït, i d'entre ells només els de patata, sangonera i paparra (PCI, LCI i TCI) han estat caracteritzats en detall tant a nivell estructural com funcional (Vendrell *et al.*, 2004; Arolas *et al.*, 2005a; Arolas *et al.*, 2005b). Fins ara, no existeix cap inhibidor de metal·locarboxipeptidases d'origen humà que hagi estat caracteritzat.

En el segon capítol de la primera part de la tesi es presenta l'estructura tridimensional del complex entre la latexina i la carboxipeptidasa A4 humanes, dues proteïnes anteriorment descrites en situacions fisiològiques diferents però que oferien l'oportunitat d'obtenir, d'una sola vegada, informació sobre la seva estructura tridimensional. D'ambdues es coneixia la seva seqüència i, en el cas de la latexina, tan sols s'havia demostrat la seva capacitat inhibidora amb enzims d'origen pancreàtic (Normant *et al.*, 1995b), malgrat que existien indicis de la presència de transcrits d'isoformes alternatives de carboxipeptidases A1 i A2 a cervell, el teixit d'origen de la latexina en la caracterització inicial de la proteïna (Normant *et al.*, 1995a).

El cDNA de la latexina humana va ser aïllat a partir d'una biblioteca de cDNAs i codifica per una proteïna de 222 residus que no presenta homologia seqüencial amb cap altra proteïna coneguda. Aquesta observació inclou el seu extrem C-terminal, el qual no guarda relació amb les seqüències d'altres inhibidors heteròlegs de carboxipeptidases que exerceixen la seva activitat inhibidora mitjançant la interacció d'aquesta regió de la proteïna amb el seti actiu de l'enzim. Aquesta absoluta absència d'homologia suggereix un mecanisme d'actuació diferent al de la resta d'inhibidors de carboxipeptidases. La caracterització funcional de la latexina mitjançant estudis d'inhibició permet descriure-la com un inhibidor de carboxipeptidases d'unió forta i de tipus no-competitiu. Totes les carboxipeptidases de vertebrats de tipus A/B analitzades són inhibides per la latexina i les constants d'inhibició determinades són similars a les obtingudes amb altres inhibidors com el

LCI o el TCI (Reverter *et al.*, 1998a; Arolas *et al.*, 2005a). La latexina no és, però, capaç d'inhibir les carboxipeptidases de la subfamília N/E ni les A/B d'invertebrats. En aquest sentit, manté la selectivitat per la subfamília A/B que ja havia estat demostrada pels inhibidors de la família del PCI. Aquest espectre de capacitat inhibidora, el fet de constituir l'únic inhibidor endogen descrit per a carboxipeptidases i les incògnites sobre el seu mecanisme d'actuació van fer créixer el nostre interès en la determinació de l'estructura tridimensional de la latexina.

La determinació estructural de la latexina no hauria estat possible en absència de derivats dels seus cristalls amb metalls pesants, donat que, en el moment d'iniciar la seva determinació estructural per difracció de raigs X, no es disposava de cap molècula anàloga per procedir a la determinació estructural per reemplaçament molecular. Gràcies a l'obtenció del seu complex amb la CPA4 va ser possible obtenir-ne l'estructura mitjançant una combinació de cerca de Patterson i difracció anòmala de longitud d'ona múltiple en base a l'àtom de Zn aportat per la carboxipeptidasa. L'estructura de la latexina dins el complex, posteriorment contrastada amb l'obtinguda en dissolució pel seu ortòleg de ratolí (Aagard *et al.*, 2005), mostra que l'inhibidor té un plegament de tipus α/β que presenta dos subdominis estructuralment equivalents tot i la seva baixa similitud seqüencial. Cada subdomini conté una hèlix α llarga i una fulla β de quatre cadenes antiparal·leles fortament retorçades que embolcallen l'hèlix mitjançant contactes hidrofòbics. Els dos subdominis es troben connectats per una hèlix α . Quan es compara aquesta estructura amb la de la resta dels inhibidors heteròlegs caracteritzats, s'observa que no segueix el patró de proteïna petita que, amb diferents graus d'estructura secundària, és sempre rica en ponts disulfur que l'ajuden a mantenir un estructura altament compacta i molt estable.

En el complex, l'inhibidor es situa a la part superior del canal d'entrada al centre actiu de la carboxipeptidasa. La superfície coberta per la latexina és molt més gran que la descrita en altres complexos proteasa-inhibidor i, a diferència d'altres inhibidors de metal·loproteases, en la inhibició no hi participa cap dels seu extrems, sinó que ve donada per les nombroses interaccions que s'estableixen amb la superfície de la carboxipeptidasa, principalment a càrrec d'un llaç present en el subdomini C-terminal de l'inhibidor. Aquesta interacció recorda, en la seva apariència global i en certs detalls a nivell atòmic en els que hi ha coincidència estricta, el mecanisme d'inhibició del segment pro en els zimògens de les CPs del tipus A/B, tant pel que fa a les interaccions específiques com a la conformació

d'algunes cadenes laterals de la part enzimàtica. Convé afegir que una interacció d'aquesta mena, governada per interaccions que obturen l'entrada al centre actiu i no per les equivalents a les d'un pseudo-substrat, ja havia estat avançada pels autors del primer treball on es descrivia la latexina (Normant et al., 1995b), malgrat que en aquest cas la seva especulació sobre la zona inhibidora, deduïda en base a homologies seqüencials amb els dominis d'activació de procarboxipeptidases, apuntava a una zona que no té cap intervenció en la interacció. El tipus d'inhibició de la latexina explica per què només les carboxipeptidases A/B són inhibides i no les de la subfamília N/E, en contenir les primeres un segment d'activació no present a les segones. A banda d'això, i principalment, existeixen també diferències en els llaços que envolten el centre actiu en les dues subfamílies de carboxipeptidases que, en les N/E, generen impediments estèrics o fan impossibles determinades interaccions entre la carboxipeptidasa i l'inhibidor.

L'aparent absència a la bibliografia de proteïnes amb similitud a la vegada estructural i funcional amb la latexina va conduir a la realització de cerques bioinformàtiques, tant a nivell de genomes de diferents organismes com de biblioteques de EST de diversos teixits. Aquestes cerques han permès trobar tota una sèrie de possibles proteïnes homòlogues que, en espera d'una més completa determinació d'estructures 3D, presenten una similitud seqüencial significativa amb la latexina. L'origen de les proteïnes detectades per homologia seqüencial va des dels ocells i els peixos fins a mamífers, amb una àmplia distribució tissular que indica una transcripció ubiqüa en el temps i en l'espai dels seus gens. En canvi, no s'han trobat seqüències a invertebrats o a organismes menys complexos. En base a aquests resultats, en el present treball es suggereix que les latexines constitueixen la família d'inhibidors endògens de les A/B metal·locarboxipeptidases de vertebrats, de manera equivalent al que representen els membres de la família dels inhibidors tissulars de les metal·loproteases de matriu.

Per altra banda, els resultats de les cerques d'homologia estructural en base a topologia de dominis van permetre comprovar que els subdominis de la latexina comparteixen el mateix plegament amb les cistatines. La cistatina C humana és un component del depòsit amiloides en les artèries del cervell a edats avançades. El mecanisme pel qual la cistatina C oligomeritza es basa en la formació de dímers mitjançant l'intercanvi de dominis (domain swapping). Aquesta homologia estructural entre la latexina i les cistatines, el comportament d'aquesta darrera en condicions patològiques i la nostra pròpia experiència amb la latexina com una proteïna

inestable en estat aïllat, van conduir a l'estudi de les propietats conformacionals de la latexina per tal de comprovar si era també capaç de formar estructures amb propietats amiloïdogèniques. Aquest tercer treball és el que obre la segona part de la present tesi.

PART II

Les malalties amiloïdòtiques es troben entre les més debilitants i cares en termes socials i econòmics del món modern i, donat que són degeneratives i associades a l'edat, estan esdevenint cada cop més prevalents a mesura que la nostra societat envelleix.

Tradicionalment s'han associat els efectes simptomàtics en aquestes malalties degeneratives a la creixent acumulació de dipòsits proteics insolubles dintre o fora del teixit afectat. Aquests dipòsits estan formats, sempre i de manera independent de la malaltia i de la proteïna que es tracti, d'estructures fibril·lars comunament anomenades fibres amiloides. Tot i això, darrerament està guanyant força la hipòtesi que recolza que serien, de fet, les estructures amiloïdòtiques prefibril·lars precursors dels agregats les formes amb més alta citotoxicitat en malalties neurodegeneratives com l'Alzheimer, el Parkinson o les encefalopaties espongiformes i no les fibres madures insolubles, de les que es postula que podrien arribar a tenir fins i tot un paper protector, en evitar l'acumulació de formes oligomèriques tòxiques (Lee *et al.*, 2004; Gong *et al.*, 2003). D'aquestes estructures, a diferència de les fibres, es coneix poc sobre les seves propietats bioquímiques i biofísiques degut a que es tracta sovint de formes heterogènies i transitòries molt difícils d'aïllar i caracteritzar. Per tant, el desenvolupament de sistemes proteics model no relacionats amb malalties i que a la vegada siguin fàcils de manipular al laboratori i aplicables a l'estudi d'estats prefibril·lars, és de gran interès ja que s'espera que aportin informació rellevant sobre aquestes espècies. Val a dir que aquesta aproximació s'ha demostrat molt útil en l'estudi de la formació dels estats fibril·lars, on bona part dels coneixements disponibles provenen de l'estudi de models no patogènics.

Les proteïnes implicades en malalties conformacionals es divideixen en dos grans tipus: les proteïnes globulars i les intrínscament desestructurades. Com a exemple del primer grup tenim la proteïna del prió o el lisozim, entre d'altres. Les proteïnes globulars no acostumen a agregar des del seu estat natiu sinó que és necessària

una desestabilització de la conformació nativa i la formació d'estats parcialment desplegats rics en estructura β . Aquests estats parcialment plegats tenen una elevada propensió d'agregació ja que regions que en l'estructura nativa quedaven amagades ara es troben exposades al solvent. Per altra banda, hi ha proteïnes com l'amilina o la β -amiloide que es troben pràcticament desestructurades dintre de les cèl·lules i per a les quals no és necessari un pas de desplegament previ a la formació d'estructures amiloïdòtiques.

El segon bloc d'aquesta tesi consta de tres capítols independents però tots ells encarats a aportar nova informació al coneixement dels processos que es donen lloc al voltant de les malalties deposicionals.

El tercer capítol, amb el que s'obre aquesta segona part, demostra que la latexina podria esdevenir un bon model per a l'estudi dels estats prefibril·lars formats a partir d'un estat totalment desplegat. Com s'ha esmentat en el capítol anterior, la latexina presenta un plegament de tipus α/β en dos dominis estructuralment equivalents.

El procés de desnaturalització de la latexina a concentracions creixents d'urea s'ha seguit tant per fluorescència, indicador de canvis en l'estructura terciària, com per diroisme circular. La diferència en els punts mitjos d'ambdues corbes denota la presència d'un intermediari que correspondria, probablement, a un estat on els dos dominis de la latexina perdrien els contactes que els mantenen units i es donaria una obertura de l'estructura global de l'inhibidor sense comportar la pèrdua de la seva estructura secundària.

Cal destacar que la latexina és una de les poques proteïnes amb dos dominis utilitzada com a model per al coneixement dels canvis i reorganitzacions que tenen lloc a la cadena polipeptídica durant un procés d'oligomerització, ja que, generalment, les proteïnes amb les que es treballa *in vitro* acostumen a ser proteïnes senzilles o dominis de proteïnes més grans. I, també, que la resolució de l'estructura tridimensional de la proteïna descrita al capítol anterior ens ha permès observar que tots els triptòfans es troben localitzats en el segon subdomini i més concretament encarats a la interfase entre els dos dominis, de manera que el seguiment de la fluorescència ens permet monitoritzar de manera específica els canvis conformationals que tenen lloc en aquesta regió i ens ha de facilitar en un futur la caracterització cinètica del plegament d'aquesta proteïna.

De forma sorprenent, en el procés d'oligomerització de la latexina no hi ha participació de l'intermediari parcialment plegat ja que no és fins que la proteïna ha

perdut pràcticament tota la seva estructura secundària que es desencadena el mecanisme d'autoassociació i formació d'estructures oligomèriques. S'ha de destacar que els oligomers són altament estables en front de la desnaturalització química, resistents a SDS i a agents reductors. És per això, que la latexina pot ser una bona proteïna model per conèixer millor els mecanismes primers de formació d'estructures amiloïdogèniques en malalties associades amb proteïnes intrínsecament desestructurades. Cal fer èmfasi en que aquestes espècies oligomèriques semblen ser l'estat final d'associació de la latexina ja que no s'observa una evolució en el temps cap a la formació de fibres. Aquesta consideració és molt important ja que la seva gran estabilitat permetria que aquests oligòmers poguessin ser aïllats i analitzats en detall. Aquesta fita ha estat fins ara de difícil assoliment en altres models estudiats ja que, en general, es tracta d'espècies transitòries que no s'acumulen sinó que són el pas previ a la formació de les fibres amb les quals es troben en equilibri.

Es coneix que no totes les regions de la seqüència són igual d'importantes a l'hora de determinar la tendència a agregar d'una proteïna; és més, es suggereix que regions curtes de la seqüència actuen com a disparadors del procés d'agregació. Aquestes regions facilitadores de l'agregació es troben bloquejades en l'estat natiu de la proteïna, mentre que en proteïnes que són intrínsecament desestructurades o en aquelles que es troben en un estat parcialment plegat, aquestes regions es troben exposades al solvent i poden desencadenar la formació d'agregats. Recentment s'han desenvolupat un seguit d'aproximacions teòriques que prediuen acuradament la presència d'aquestes regions en polipèptids.

En el cas de la latexina el fet que oligomeritzi des del seu estat desplegat la fa especialment adient per a ser analitzada mitjançant aquests algorismes. D'entre els diversos mètodes de predicció d'aquestes regions que es podem trobar a la literatura, en aquest treball s'ha usat el desenvolupat en el nostre grup i presentat en aquesta tesi dintre del capítol 5. El perfil d'agregació de la latexina mostra clarament una regió destacada dintre de la seqüència que presenta una elevada propensió d'agregació i que es localitza a l'hèlix $\alpha 3$. El pèptid sintètic corresponent a aquesta regió forma microcristalls que també presenten propietats amiloides i tenen l'estructura β típica de les fibres amiloides. Només s'han descrit la formació de dues formes de microcristalls semblants per a pèptids priònics. La caracterització d'aquest cristalls a nivell atòmic hauria de permetre resoldre un dels pocs exemples de l'estructura de la fulla beta amiloide.

A diferència de la latexina que pot ser emprada com a model de proteïna intrínsecament desestructurada, en el quart capítol de la tesi es presenta un treball desenvolupat amb l' α -quimotripsina. És una proteïna que en l'estat natiu té un plegament de tipus barril- β , un dels més habituals a la natura. En aquest cas la formació de fibres amiloides requereix la formació prèvia d'un intermediari parcialment plegat ric en estructura β . Es coneix molt poc sobre les propietats d'intermediaris com el que s'acaba de comentar. És per això que l'estudi de les condicions en les que s'acumulen aquestes espècies parcialment plegades en diferents proteïnes models com ara l' α -quimotripsina poden aportar elements necessaris per entendre millor el procés d'agregació *in vivo*, especialment per proteïnes amb estructura β .

El tractament de l' α -quimotripsina amb TFE indueix un canvi conformacional de l' α -quimotripsina en dos estats. En primer lloc hi ha una desorganització de les interaccions que estableixen l'estructura nativa de barril- β i la proteïna adopta una conformació en fulla β estesa no nativa. En un segon pas la cadena polipeptídica adquireix una conformació en α hèlix no nativa. L'agregació i formació d'estructures amiloïdogèniques només es desencadena des de l'intermediari parcialment estructurat, i és la presència de superfícies hidrofòbiques exposades al solvent la que promou la formació de contactes intermoleculars hidrofòbics entre les regions amb conformació de fulla- β estesa dels intermediaris parcialment plegats. En canvi, l'adquisició d'una conformació en hèlix no nativa resulta en una disminució en la propensió d'agregació que està relacionada amb el fet que les regions enganxoses que quedaven exposades al solvent en l'intermediari, ara es troben situades a dintre de l'estructura helicoidal, impedit així l'establiment d'aquests contactes intermoleculars que dirigeixen el procés de formació d'estructures amiloïdogèniques. Per tant, en aquest cas el procés d'agregació depèn tant de la presència d'un intermediari parcialment plegat així com de la conformació d'aquest intermediari.

El primer i el segon treball d'aquesta segona part són un reflex dels dos models de proteïnes que es troben implicades en el desenvolupament de les malalties amiloïdogèniques. En el cas de la latexina, la formació d'estructures amiloides requereix del desplegament total de la proteïna, pel que seria un model adequat per a l'estudi de les malalties en les que la proteïna que origina l'agregació es troba intrínsecament desestructurada. En canvi, el model de l' α -quimotripsina s'ajustaria al de les proteïnes globulars on el procés d'agregació requereix de la formació d'un

estat parcialment plegat ric en estructura β . Els resultats obtinguts en aquests dos treballs i les observacions dels darrers estudis realitzats en el camp dels amiloides, són indicatius de que la capacitat de formar aquestes estructures es pot considerar una habilitat general de totes les cadenes polipeptídiques (Dobson, 2001). Tot i això, la propensió per fer-ho varia molt entre seqüències diferents, alhora que regions molt curtes d'aminoàcids poden actuar com a disparadors o inhibidors de la formació de fibres amiloides (Ventura *et al.*, 2004; Ivanova *et al.*, 2004).

En el darrer i cinquè treball de la present tesi, s'ha desenvolupat un mètode per a predir aquestes regions facilitadores de l'agregació, les quals són potencials dianes terapèutiques. La diferència principal entre els escassos mètodes de predicció trobats a la literatura i el que es presenta en la tesi és l'origen de les dades que s'utilitzen per generar el perfil d'agregació de la proteïna; mentre que en la resta de mètodes descrits s'usen fórmules matemàtiques amb valors teòrics que són assignats a regions específiques de la seqüència, en aquest treball s'han usat les dades experimentals d'un sistema *in vivo* desenvolupat anteriorment al grup, emprant el pèptid β amiloide com a model per a calcular la propensió de cada aminoàcid natural quan es col·loca al centre del nucli hidrofòbic del pèptid A β (de Groot *et al.*, 2006). En general, els residus hidrofòbics tendeixen a induir l'agregació mentre que els polars promouen la solubilitat, dades que encaixen amb el supòsit general que indica que les interaccions hidrofòbiques juguen un paper important en l'agregació proteica.

Donat que una determinada cadena polipeptídica agrega des d'un estat total o parcialment desplegat, els valors intrínsecs d'agregació de cadascun dels aminoàcids poden ser aplicats, en teoria, a qualsevol proteïna amb independència del seu context. I, en conseqüència, es pot generar un perfil d'agregació per a qualsevol seqüència proteica i determinar quines regions presenten propensions d'agregació superiors a la mitjana global de la proteïna. En el treball s'han definit aquestes regions de la seqüència que presenten una major propensió d'agregació com a "hot spots" o regions disparadores de l'agregació.

A la natura es presenten casos en els que la formació de fibres amiloides d'una determinada proteïna ve donada per una mutació puntual en la seva cadena polipeptídica. Aquestes mutacions puntuals provoquen canvis en la propensió d'agregació. El mètode desenvolupat en el cinquè treball permet predir la diferent

propensió d'agregació de les proteïnes salvatges respecte de les mutants. Aquesta és una eina útil de cara a la predicció, abans de realitzar la seva expressió recombinant, de l'efecte que pot tenir la introducció d'una determinada mutació en la tendència a agregar d'una proteïna. La validació del mètode amb grup de proteïnes associades a malalties conformacionals, comparant les regions que experimentalment s'han demostrat implicades en l'agregació dels corresponents polipèptids ha donat un resultat excel·lent i confirma la hipòtesi que regions curtes específiques de la seqüència d'una proteïna poden actuar com a desencadenants de l'agregació, actuant probablement com a punts de nucleació a partir dels quals s'expandiria l'estructura fibril·lar ordenada. La natura, com ja em vist també en els dos capítols anteriors, ha dotat a les proteïnes globulars d'una estabilitat en el seu estat natiu que fa que els "hot spots" es trobin soterrats o bé implicats en interaccions intramoleculares. Aquesta es presenta com una estratègia evolutiva fructífera ja que molt poques proteïnes agreguen des de la seva conformació nativa. En canvi, en les proteïnes intrínsecament desestructurades les regions disparadores de l'agregació es troben directament exposades al dissolvent, i és a conseqüència d'això que la presència de "hot spots" és molt menor en aquest tipus de cadenes polipeptídiques. Aquest mètode pot ser rellevant per al tractament de les malalties amiloidogèniques, ja que les regions determinades com a "hot spots" poden ser considerades com a dianes preferents a ser abordades en el tractament d'aquestes malalties i en el disseny d'estratègies per al bloqueig del procés d'agregació.

CONCLUSIONS

PART I

Capítol 1: Estructura tridimensional de la carboxipeptidasa A1 humana

1.- La proteïna purificada cristal·litza en 22% (w/v) PAA 5100, 0,1 M HEPES pH 7,5, 0,02 M clorur de magnesi.

2.- L'estructura de l'enzim actiu es va resoldre per cristal·lografia de raigs-X i reemplaçament molecular, usant com a motlle l'estructura tridimensional de la CPA bovina, a una resolució final d'1,6 Å.

3.- L'estructura de la CPA1 humana presenta un plegament tipus α/β hidrolasa conservat en la resta de metal·locarboxipeptidases. Aquest està format per vuit fulles- β centrals rodejades de vuit hèlix- α .

4.- En el centre actiu hi ha una molècula de PAA que ocupa les posicions S1' i S1 i estableix interaccions amb residus de la proteïna similars als trobats en els complexos entre les carboxipeptidases i els seus inhibidors.

5.- Essent la CPA1 l'enzim el prototip de la subfamília A/B, aquesta estructura a alta resolució constitueix un motlle adequat per a modelar sobre ella l'estructura de carboxipeptidases de gran rellevància fisiològica i proporciona detalls addicionals per a aprofundir en l'estudi del seu mecanisme catalític i per al disseny d'inhibidors específics extrapolables també al conjunt de les zinc metal·loproteases.

Capítol 2: Estructura de la carboxipeptidasa A4 humana i del seu inhibidor endogen, la latexina.

1.- S'ha clonat la latexina a partir de primeres cadenes de cDNAs de cervell de fetus humà. Gràcies al desenvolupament d'un sistema d'expressió recombinant eficient, s'han pogut obtenir quantitats d'inhibidor suficients per a l'obtenció de complexos amb la hCPA4.

Conclusions

2.- La latexina és un inhibidor de tipus no competitiu de metal·locarboxipeptidases de tot els vertebrats de la subfamília A/B. Presenta K_i 's en el rang nM que són similars als determinats per a la resta d'inhibidors proteics descrits. La latexina no pot inhibir les carboxipeptidases de la subfamília N/E ni una MCP de la subfamília A/B d'invertebrat.

3.- El complex hCPA4/latexina purificat cristal·litza en 40% 2-metil-2,4-pentanediol (MPD), 0,1 M Bis-Tris pH 6,5, i 0,2 μ l 40% d'acetona al cap de quatre dies.

4.- L'estructura tridimensional de la latexina té una topologia de tipus α/β i es pot dividir en dos subdominis que presenten un mateix plegament. Els dos subdominis superimposats presenten un r.m.s.d. de 2,1 Å, tot i la seva baixa identitat seqüencial (14%).

5.- Cada subdomini està format per una hèlix llarga seguida d'una fulla- β formada per quatre cadenes- β antiparal·leles. Els dos subdominis estan units per un segment connector format per una hèlix- α .

6.- El domini carboxipeptidasa de la CPA4 humana presenta l'estructura observada en altres enzims de la subfamília A/B. Aquest està format per vuit fulles- β centrals envoltades per vuit hèlix- α .

7.- El mecanisme d'inhibició de la latexina difereix del de la resta d'inhibidors proteics descrits ja que cap dels seus extrems participa en la inhibició de la proteasa. L'inhibidor es situa a la superfície de la proteasa tapant l'entrada al centre actiu, i la seva acció és duta a terme principalment per un llaç del subdomini C-terminal comprès entre la β -7 i la β -8.

8.- El modus d'inhibició de la latexina presenta similituds amb la forma en la que el prodomini de les PCP bloqueja el domini catalític. Aquesta similitud podria explicar el per què aquelles CPs que tenen el característic PD, les de la subfamília A/B, són inhibides per la latexina mentre que les que manquen de prodomini, les MCPs de la subfamília N/E, no poden ser inhibides.

9.- Cadascun dels subdominis de la latexina presenta una topologia molt homòloga a la de les cistatines, tot i que no inhibeix les cisteïna proteases en ser diferents les regions implicades en la inhibició en els dos tipus d'inhibidors. La similitud estructural

dels subdominis de la latexina amb les cistatines suggereix l'existència d'un ancestre comú a partir del qual s'hauria esdevingut un procés de duplicació gènica.

10.- Les cerques bioinformàtiques han permès trobar tot un seguit de seqüències similars a la latexina en mamífers i altres vertebrats. Per altra banda, la latexina presenta una distribució ubiqua en el temps i en diferents teixits. Aquests resultats suggereixen que les latexines constitueixen una família d'inhibidors endògens de les A/B-MCPs de vertebrats.

PART II

Capítol 3: La latexina humana forma oligòmers de tipus amiloide mitjançant l'associació de segments polipeptídics que es troben desplegats

1.- La latexina s'autoassocia per formar espècies oligomèriques que tenen propietats amiloïdòtiques, com la unió dels colorants Th-t i congo red.

2.- El principal intermediari detectat a la reacció de desplegament correspon, probablement, a una forma oberta en la que els dos dominis han perdut els seus contactes intermoleculars però conserven la seva arquitectura. Aquest estat parcialment plegat no és rellevant per l'associació de la latexina.

3.- L'oligomerització de la latexina correlaciona amb les condicions en les que la proteïna es troba majoritàriament desplegada i no amb ambients que afavoreixin la població d'estats intermediaris parcialment plegats rics en estructura beta.

4.- La latexina pot ser un bon sistema per a l'aïllament i la caracterització estructural i bioquímica de les sovint elusives espècies protofibrilars de baix pes molecular, donat que aquestes constitueixen l'estat final d'associació d'aquesta proteïna sota condicions fortament desnaturalitzants.

5.- S'ha localitzat una regió de la seqüència de la latexina que podria ser la responsable de l'agregació. El pèptid sintètic corresponent a aquesta regió forma microcristalls que també presenten propietats amiloides i tenen l'estructura β típica de les fibres amiloides. Només s'ha descrit la formació de dues formes de microcristalls semblants per a pèptids priònics.

Capítol 4: Formació de fibres amiloides a partir d'un intermediari parcialment estructurat de l' α -quimotripsina.

1.- En presència de 35 % TFE l' α -quimotripsina perd parcialment la seva estructura terciària i s'indueixen elements d'estructura secundària no nadius.

2.- La desorganització del plegament natiu i l'adopció d'una estructura en fulla- β estesa de l'intermediari parcialment plegat, resulta en una major exposició dels residus aromàtics al dissolvent. Els contactes hidrofòbics intermoleculars entre les regions "enganxoses" de l'intermediari són les que promouen el seu ensamblatge i per tant, la formació d'estructures de tipus amiloide.

3.- Els experiments de tioflavina-T i congo red indiquen que els agregats formats en presència de 35% TFE tenen propietats amiloïdòtiques, i les imatges del MET mostren que aquests agregats tenen morfologia fibril·lar.

4.- A concentracions més elevades de TFE es redueix la propensió d'agregació ja que s'indueixen conformacions en hèlix on les regions hidrofòbiques que quedaven exposades en la fulla- β ja no són accessibles.

5.- L'agregació de l' α -quimotripsina depèn tant de la presència d'un intermediari parcialment estructurat com de la conformació de l'esquelet polipeptídic d'aquest.

Capítol 5: Prediccions de regions "hot spots" d'agregació en els polipèptids relacionats amb malalties

1.- L'èxit en les prediccions del mètode desenvolupat recolza la hipòtesi que regions curtes específiques de la seqüència d'una proteïna poden actuar com a desencadenants de l'agregació.

2.- En general, els residus hidrofòbics tendeixen a induir l'agregació, mentre que els polars promouen la solubilitat, confirmant-se així, que les interaccions hidrofòbiques juguen un paper important en l'agregació proteica.

Conclusions

3.- El paper de les regions “hot spots” és, probablement, el d’actuar com a punts específics de nucleació a partir dels quals l’estructura fibril·lar ordenada s’anirà expandint.

4.- Les proteïnes globulars presenten un major nombre de regions “hot spots” que aquelles proteïnes intrínsecament desestructurades; és per això que la natura les ha dotat d’un plegament estable en el qual les regions amb propensió a agregar queden soterrades a l’interior de l’estructura o es troben participant en interaccions intramoleculares.

5.- Mutacions que interrompin aquelles regions amb major propensió a agregar, en el cas de les proteïnes intrínsecament desestructurades, o que sobreestabilitzin l’estat natiu de les proteïnes globulars, esdevindrien una aproximació útil per evitar els depòsits de proteïnes i en conseqüència una millora per al tractament de les malalties amiloïdòtiques.

PART I

Chapter 1: Three dimensional structure of human carboxypeptidase A1

1.- The purified protein crystallized in 22% (w/v) PAA 5100, 0.1 M HEPES pH 7.5, 0.02 M magnesium chloride.

2.- The structure of the active form was solved by molecular replacement using bovine CPA as search model and yielding a 1.6 Å final resolution.

3.- hCPA1 exhibits the same fold as the other known carboxypeptidase structures. It consists of a central twisted eight-stranded β -sheet flanked by eight α -helices, which together form a globular α/β protein.

4.- A PAA molecule occupies positions S1' i S1 at the active site and establishes interactions with the protein residues similar to those involved in other carboxipeptidase-inhibitor complexes.

5.- This is the structure of the human form of carboxypeptidase A1, the prototype of the A/B subfamily. Its high resolution provides an excellent template for modeling the structure of physiologically relevant carboxypeptidases. Moreover, it gives additional details to the study of its catalytical mechanism and provides a more accurate picture of the interactions between the active-site zinc and its lligand for the design of specific inhibitors that can be extended to the rest of the zinc metalloproteases

Chapter 2: Structure of human carboxypeptidase A4 with its endogenous protein inhibitor, latexin.

1.- Latexin has been cloned from a gene pool cDNA of human fetal brain. The development of an efficient recombinant expression system has made possible to obtain amounts of the inhibitor sufficient to form a complex with hCPA4.

Concluding remarks

2.- Latexin is a non-competitive inhibitor of vertebrate A/B metalloproteases. The kinetic inhibition constants are in the nanomolar range and are similar to those obtained for the other protein inhibitors of metalloproteases. Latexin does not inhibit members of the N/E class or an invertebrate A/B MCP from insects.

3.- The hCPA4-latexin complex was crystallized in 40% 2-methyl-2,4-pentanediol (MPD), 0.1 M Bis-Tris pH 6.5, and 0.2 μ l of 40% acetone.

4.- Human latexin is an elongated protein, with an α/β topology. It can be divided into two subdomains of the same fold and both subdomains can be superimposed with an r.m.s.d. of 2.1 Å, despite negligible sequence identity (14%).

5.- Each subdomain contains an extended α -helix followed by a strongly twisted four-stranded antiparallel β -sheet. The two subdomains are linked by a connecting segment which contains α 2-helix.

6.- The carboxypeptidase moiety has the classical α/β hydrolase fold of A/B-MCPs, formed by a core of eight twisted β -sheets flanked on both sides by several α -helices.

7.- The structural determinants for latexin inhibition differ with the rest of inhibitors described so far because any of its termini is involved. The inhibitor sits on top of the funnel rim occluding the entrance to the active site and inhibition is mainly caused by an inhibitory loop provided by the C-terminal subdomain.

8.- The way how latexin inhibits MCPs presents some similarity with the prodomain inhibition in the zymogen. This similarities may explain why those CPs with the characteristic pro-domain, the A/B-MCPs, are inhibited by latexin whereas CPs without it, the N/E-MCPs, are not inhibited.

9.- Each of the latexin subdomains shows strong similarity with cystatins. However, different regions are involved in target protein inhibition in both cases. The striking structural similarity suggests that latexin and cystatins have a common ancestry. Genetic duplication of a cystatin module may have given rise to the structure of latexin.

10.- Bioinformatic searches have found a number of putative proteins similar to human latexin in mammals and other vertebrates. On the other hand, latexin presents a

temporal and spatial ubiquity. All these results suggest that latexins could represent the A/B MCPs endogenous inhibitors in vertebrates.

PART II

Chapter 3: Human latexin forms amyloid-like oligomers by association of unfolded polypeptide segments.

1.- Human latexin assembles into oligomeric species with amyloid-like properties, as the binding to the specific dyes Th-T and congo red.

2.- The main intermediate detected in the unfolding reaction corresponds most likely to an open form in which the contacts between the subdomains are lost rather than to a globally unfolded form. The intermediate is not involved in the aggregation.

3.- Latexin oligomerization correlates with conditions where the protein is mostly unfolded, rather than with environments that result in the formation of partially structured states rich in β -sheet structure.

4.- Latexin could be an excellent model for the isolation and a deep structural and biochemical characterization of stable low-molecular protofibrillar species, since they appear to be the final association state of latexin under strongly denaturing conditions.

5.- A region in the latexin polypeptide sequence that may promote and drive aggregation has been detected. A peptide corresponding to that region forms well-ordered microcrystals with amyloid properties and the typical β -sheet structure of amyloid fibrils.

Chapter 4: Amyloid fibril formation by a partially structured intermediate state of alpha-chymotrypsin.

1.- In the presence of 35% TFE, α -chymotrypsin partially loses its native tertiary structure and non-native secondary structure elements are induced.

2.- The disorganization of the native state and the formation of an extended β -sheet conformation in the intermediate state result in a higher exposure of aromatic residues to the solvent. The intermolecular hydrophobic contacts between the sticky-extended β -sheet elements promote aggregation and the formation of amyloid-like structures.

3.- The Th-T and congo red binding experiments indicate that the aggregates formed in 35% TFE possess amyloid-like properties. Moreover, micrographs obtained by TEM reveal that the aggregates have fibrillar morphology.

4.- The aggregation propensity is reduced at higher TFE concentration, probably due to the induction of a non-native helical conformation where the sticky hydrophobic clusters present in the extended β -sheet are now partially shielded inside the structure.

5.- α -chymotrypsin aggregation depends both on the population of a partially structured state and on the backbone conformation of this intermediate.

Chapter 5: Prediction of “hot spots” of aggregation in disease-linked polypeptides.

1.- The success of the predictions of the developed method, supports the hypothesis that short specific amino acid stretches can act as triggers of the aggregation process.

2.- Overall, hydrophobic residues tend to induce aggregation whereas polar ones promote solubility, in accordance with the general assumption that hydrophobic interactions are supposed to play an important role in protein aggregation.

3.- The role of “hot spots” is, probably, to act as specific nucleation points from which the ordered fibrillar structure is expanded.

4.- Globular proteins present a higher number of “hot spots” in comparison with natively unfolded proteins. That is the reason why nature has provided globular proteins with stable native states in which the aggregation-prone sequences are buried or involved in intra-molecular interactions.

5.- Mutations that disrupt aggregation-prone regions in unstructured polypeptides or those which over-stabilize the native state of globular proteins, are expected to be useful approaches to avoid protein deposition.

BIBLIOGRAFIA

Aloy, P., Catusas, L., Villegas, V., Reverter, D., Vendrell, J. & Avilés, F. X. (1998). Comparative analysis of the sequences and three-dimensional models of human procarboxypeptidases A1, A2 and B. *Biol Chem* **379**(2), 149-55.

Aloy, P., Companys, V., Vendrell, J., Avilés, F. X., Fricker, L. D., Coll, M. & Gomis-Ruth, F. X. (2001). The crystal structure of the inhibitor-complexed carboxypeptidase D domain II and the modeling of regulatory carboxypeptidases. *J Biol Chem* **276**(19), 16177-84.

Alvarez-Santos, S., Gonzalez-Lafont, A., Lluch, J.M., Oliva, O. & Avilés, F.X. (1994). On water-promoted mechanism of peptide cleavage by carboxypeptidase A. A theoretical study. *Canadian J Chemistry* **72**, 2077-083.

Andersson, K., Olofsson, A., Nielsen, E. H., Svehag, S. E. & Lundgren, E. (2002). Only amyloidogenic intermediates of transthyretin induce apoptosis. *Biochem Biophys Res Commun* **294**(2), 309-14.

Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science* **181**(96), 223-30.

Anson, M.L. (1935). Crystalline carboxypeptidase. *Science* **21**, 467-8.

Arimatsu, Y. (1994). Latexin: a molecular marker for regional specification in the neocortex. *Neurosci Res* **20**(2), 131-5.

Arolas, J. L., Lorenzo, J., Rovira, A., Castella, J., Avilés, F. X. & Sommerhoff, C. P. (2005a). A carboxypeptidase inhibitor from the tick *Rhipicephalus bursa*: isolation, cDNA cloning, recombinant expression, and characterization. *J Biol Chem* **280**(5), 3441-8.

Arolas, J. L., Popowicz, G. M., Lorenzo, J., Sommerhoff, C. P., Huber, R., Avilés, F. X. & Holak, T. A. (2005b). The three-dimensional structures of tick carboxypeptidase inhibitor in complex with A/B carboxypeptidases reveal a novel double-headed binding mode. *J Mol Biol* **350**(3), 489-98.

Bibliografia

Auld D.S. (2004). Carboxypeptidase A. In: Handbook of proteolytic Enzymes, Barrett, A.J., Rawlings N.D. & Woessner, J. F. Jr (eds), Elsevier, London, vol 1, 812-21.

Avilés, F. X., 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**(3), 381-9.

Azriel, R. & Gazit, E. (2001). Analysis of the minimal amyloid-forming fragment of the islet amyloid polypeptide. An experimental support for the key role of the phenylalanine residue in amyloid formation. *J Biol Chem* **276**(36), 34156-61.

Balbach, J. J., Ishii, Y., Antzutkin, O. N., Leapman, R. D., Rizzo, N. W., Dyda, F., Reed, J. & Tycko, R. (2000). Amyloid fibril formation by A beta 16-22, a seven-residue fragment of the Alzheimer's beta-amyloid peptide, and structural characterization by solid state NMR. *Biochemistry* **39**(45), 13748-59.

Balbirnie, M., Grothe, R. & Eisenberg, D. S. (2001). An amyloid-forming peptide from the yeast prion Sup35 reveals a dehydrated beta-sheet structure for amyloid. *Proc Natl Acad Sci U S A* **98**(5), 2375-80.

Barbosa Pereira, P. J., Segura-Martin, S., Oliva, B., Ferrer-Orta, C., Avilés, F. X., Coll, M., Gomis-Ruth, F. X. & Vendrell, J. (2002). Human procarboxypeptidase B: three-dimensional structure and implications for thrombin-activatable fibrinolysis inhibitor (TAFI). *J Mol Biol* **321**(3), 537-47.

Baumann, U., Bauer, M., Letoffe, S., Delepelaire, P. & Wandersman, C. (1995). Crystal structure of a complex between *Serratia marcescens* metallo-protease and an inhibitor from *Erwinia chrysanthemi*. *J Mol Biol* **248**(3), 653-61.

Bayés, A., Sonnenschein, A., Daura, X., Vendrell, J. & Avilés, F. X. (2003). Procarboxypeptidase A from the insect pest *Helicoverpa armigera* and its derived enzyme. Two forms with new functional properties. *Eur J Biochem* **270**(14), 3026-35.

Bayés, A., Comellas-Bigler, M., Rodríguez de la Vega, M., Maskos, K., Bode, W., Avilés, F. X., Jongsma, M. A., 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**(46), 16602-7.

Bellotti, V., Stoppini, M., Mangione, P., Sunde, M., Robinson, C., Asti, L., Brancaccio, D. & Ferri, G. (1998). Beta2-microglobulin can be refolded into a native state from ex vivo amyloid fibrils. *Eur J Biochem* **258**(1), 61-7.

Birnboim, H. C. & Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res* **7**(6), 1513-23.

Bitan, G., Vollers, S. S. & Teplow, D. B. (2003). Elucidation of primary structure elements controlling early amyloid beta-protein oligomerization. *J Biol Chem* **278**(37), 34882-9.

Bode, W., Engh, R., Musil, D., Thiele, U., Huber, R., Karshikov, A., Brzin, J., Kos, J. & Turk, V. (1988). The 2.0 Å X-ray crystal structure of chicken egg white cystatin and its possible mode of interaction with cysteine proteinases. *Embo J* **7**(8), 2593-9.

Bodles, A. M., Guthrie, D. J., Greer, B. & Irvine, G. B. (2001). Identification of the region of non-Aβ component (NAC) of Alzheimer's disease amyloid responsible for its aggregation and toxicity. *J Neurochem* **78**(2), 384-95.

Booth, D. R., Sunde, M., Bellotti, V., Robinson, C. V., Hutchinson, W. L., Fraser, P. E., Hawkins, P. N., Dobson, C. M., Radford, S. E., Blake, C. C. & Pepys, M. B. (1997). Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis. *Nature* **385**(6619), 787-93.

Bower, J. M., Hayward, C. F., Morley, I. S. & Morley, J. S. (1974). Trypsinogen activation peptide and related peptides as inhibitors of gastric secretion. *Biochem Biophys Res Commun* **60**(2), 820-4.

Bown, D. P., Wilkinson, H. S. & Gatehouse, J. A. (1998). Midgut carboxypeptidase from *Helicoverpa armigera* (Lepidoptera: Noctuidae) larvae: enzyme characterisation, cDNA cloning and expression. *Insect Biochem Mol Biol* **28**(10), 739-49.

Bibliografia

- Bown, D. P. & Gatehouse, J. A. (2004). Characterization of a digestive carboxypeptidase from the insect pest corn earworm (*Helicoverpa armigera*) with novel specificity towards C-terminal glutamate residues. *Eur J Biochem* **271**(10), 2000-11.
- Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54 (Pt 5)**, 905-21.
- Bucciantini, M., Giannoni, E., Chiti, F., Baroni, F., Formigli, L., Zurdo, J., Taddei, N., Ramponi, G., Dobson, C. M. & Stefani, M. (2002). Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. *Nature* **416**(6880), 507-11.
- Canet, D., Last, A. M., Tito, P., Sunde, M., Spencer, A., Archer, D. B., Redfield, C., Robinson, C. V. & Dobson, C. M. (2002). Local cooperativity in the unfolding of an amyloidogenic variant of human lysozyme. *Nat Struct Biol* **9**(4), 308-15.
- Casadaban, M. J. & Cohen, S. N. (1980). Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. *J Mol Biol* **138**(2), 179-207.
- Catusus, L., Villegas, V., Pascual, R., Avilés, F. X., Wicker-Planquart, C. & Puigserver, A. (1992). cDNA cloning and sequence analysis of human pancreatic procarboxypeptidase A1. *Biochem J* **287 (Pt 1)**, 299-303.
- Catusus, L., Vendrell, J., Avilés, F. X., 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**(12), 6651-7.
- Caughey, B. & Lansbury, P. T. (2003). Protofibrils, pores, fibrils, and neurodegeneration: separating the responsible protein aggregates from the innocent bystanders. *Annu Rev Neurosci* **26**, 267-98.

- Clark, A., Cooper, G. J., Lewis, C. E., Morris, J. F., Willis, A. C., Reid, K. B. & Turner, R. C. (1987). Islet amyloid formed from diabetes-associated peptide may be pathogenic in type-2 diabetes. *Lancet* **2**(8553), 231-4.
- Clore, G. M., Gronenborn, A. M., Nilges, M. & Ryan, C. A. (1987). Three-dimensional structure of potato carboxypeptidase inhibitor in solution. A study using nuclear magnetic resonance, distance geometry, and restrained molecular dynamics. *Biochemistry* **26**(24), 8012-23.
- Chabry, J., Ratsimanohatra, C., Sponne, I., Elena, P. P., Vincent, J. P. & Pillot, T. (2003). In vivo and in vitro neurotoxicity of the human prion protein (PrP) fragment P118-135 independently of PrP expression. *J Neurosci* **23**(2), 462-9.
- Chen, C. C., Wang, S. S., Chen, T. W., Jap, T. S., Chen, S. J., Jeng, F. S. & Lee, S. D. (1996). Serum procarboxypeptidase B, amylase and lipase in chronic renal failure. *J Gastroenterol Hepatol* **11**(5), 496-9.
- Chiti, F., Webster, P., Taddei, N., Clark, A., Stefani, M., Ramponi, G. & Dobson, C. M. (1999). Designing conditions for in vitro formation of amyloid protofilaments and fibrils. *Proc Natl Acad Sci U S A* **96**(7), 3590-4.
- Chiti, F., Taddei, N., Baroni, F., Capanni, C., Stefani, M., Ramponi, G. & Dobson, C. M. (2002). Kinetic partitioning of protein folding and aggregation. *Nat Struct Biol* **9**(2), 137-43.
- Chiti, F., Taddei, N., Bucciantini, M., White, P., Ramponi, G. & Dobson, C. M. (2000). Mutational analysis of the propensity for amyloid formation by a globular protein. *Embo J* **19**(7), 1441-9.
- Chiti, F., Stefani, M., Taddei, N., Ramponi, G. & Dobson, C. M. (2003). Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature* **424**(6950), 805-8.
- Choi, W., Zibae, S., Jakes, R., Serpell, L. C., Davletov, B., Crowther, R. A. & Goedert, M. (2004). Mutation E46K increases phospholipid binding and assembly into filaments of human alpha-synuclein. *FEBS Lett* **576**(3), 363-8.

Bibliografia

Christianson, D. W., Mangani, S., Shoham, G. & Lipscomb, W. N. (1989). Binding of D-phenylalanine and D-tyrosine to carboxypeptidase A. *J Biol Chem* **264**(22), 12849-53.

Chromy, B. A., Nowak, R. J., Lambert, M. P., Viola, K. L., Chang, L., Velasco, P. T., Jones, B. W., Fernandez, S. J., Lacor, P. N., Horowitz, P., Finch, C. E., Krafft, G. A. & Klein, W. L. (2003). Self-assembly of Abeta(1-42) into globular neurotoxins. *Biochemistry* **42**(44), 12749-60.

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

Collaborative Computational Project No. 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog sect D* **50**, 760-3.

Cuervo, A. M., Stefanis, L., Fredenburg, R., Lansbury, P. T. & Sulzer, D. (2004). Impaired degradation of mutant alpha-synuclein by chaperone-mediated autophagy. *Science* **305**(5688), 1292-5.

de Groot, N. S., Avilés, F. X., Vendrell, J. & Ventura, S. (2006). Mutagenesis of the central hydrophobic cluster in Abeta42 Alzheimer's peptide. *Febs J* **273**(3), 658-668.

Der-Sarkissian, A., Jao, C. C., Chen, J. & Langen, R. (2003). Structural organization of alpha-synuclein fibrils studied by site-directed spin labeling. *J Biol Chem* **278**(39), 37530-5.

Dill, K. A. (1985). Theory for the folding and stability of globular proteins. *Biochemistry* **24**(6), 1501-9.

Dill, K. A. & Chan, H. S. (1997). From Levinthal to pathways to funnels. *Nat Struct Biol* **4**(1), 10-9.

Dobson, C. M. (1999). Protein misfolding, evolution and disease. *Trends Biochem Sci* **24**(9), 329-32.

Dobson, C. M. (2001). The structural basis of protein folding and its links with human disease. *Philos Trans R Soc Lond B Biol Sci* **356**(1406), 133-45.

Dominguez, D. I., De Strooper, B. & Annaert, W. (2001). Secretases as therapeutic targets for the treatment of Alzheimer's disease. *Amyloid* **8**(2), 124-42.

Edwards, M. J., Lemos, F. J., 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**(12), 1063-72.

Edwards, M. J., Moskalyk, L. A., Donnelly-Doman, M., Vlaskova, M., Noriega, F. G., Walker, V. K. & Jacobs-Lorena, M. (2000). Characterization of a carboxypeptidase A gene from the mosquito, *Aedes aegypti*. *Insect Mol Biol* **9**(1), 33-8.

Ellis, R. J. & Hartl, F. U. (1999). Principles of protein folding in the cellular environment. *Curr Opin Struct Biol* **9**(1), 102-10.

Ellis, R. J. & Pinheiro, T. J. (2002). Medicine: danger--misfolding proteins. *Nature* **416**(6880), 483-4.

Eng, F. J., Novikova, E. G., Kuroki, K., Ganem, D. & Fricker, L. D. (1998). gp180, a protein that binds duck hepatitis B virus particles, has metallo-carboxypeptidase D-like enzymatic activity. *J Biol Chem* **273**(14), 8382-8.

Erdös, E.G. (1979). A: Handobook of experimental pharmacology Erdös EG (Ed) Springer Verlag, Heidelberg, 427-448.

Estebanez-Perpina, E., Bayés, A., Vendrell, J., Jongsma, M. A., Bown, D. P., Gatehouse, J. A., Huber, R., Bode, W., Avilés, F. X. & Reverter, D. (2001). Crystal structure of a novel mid-gut procarboxypeptidase from the cotton pest *Helicoverpa armigera*. *J Mol Biol* **313**(3), 629-38.

Fan, X., Qian, Y., Fricker, L. D., Akalal, D. B. & Nagle, G. T. (1999). Cloning and expression of *Aplysia* carboxypeptidase D, a candidate prohormone-processing enzyme. *DNA Cell Biol* **18**(2), 121-32.

Bibliografia

Fandrich, M., Fletcher, M. A. & Dobson, C. M. (2001). Amyloid fibrils from muscle myoglobin. *Nature* **410**(6825), 165-6.

Fandrich, M. & Dobson, C. M. (2002). The behaviour of polyamino acids reveals an inverse side chain effect in amyloid structure formation. *Embo J* **21**(21), 5682-90.

Fandrich, M., Forge, V., Buder, K., Kittler, M., Dobson, C. M. & Diekmann, S. (2003). Myoglobin forms amyloid fibrils by association of unfolded polypeptide segments. *Proc Natl Acad Sci U S A* **100**(26), 15463-8.

Fernandez-Escamilla, A. M., Rousseau, F., Schymkowitz, J. & Serrano, L. (2004). Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat Biotechnol* **22**(10), 1302-6.

Findeis, M. A., Musso, G. M., Arico-Muendel, C. C., Benjamin, H. W., Hundal, A. M., Lee, J. J., Chin, J., Kelley, M., Wakefield, J., Hayward, N. J. & Molineaux, S. M. (1999). Modified-peptide inhibitors of amyloid beta-peptide polymerization. *Biochemistry* **38**(21), 6791-800.

Fink, A. L. (1998). Protein aggregation: folding aggregates, inclusion bodies and amyloid. *Fold Des* **3**(1), R9-23.

Floege, J., Schaffer, J., Koch, K. M. & Shaldon, S. (1992). Dialysis related amyloidosis: a disease of chronic retention and inflammation? *Kidney Int Suppl* **38**, S78-85.

Folk, J.E., Piez, K.A., Carroll, W.R. & Gladner, J.A. (1960). Carboxypeptidase B. 4. Purification and characterization of the porcine enzyme. *J Biol Chem* **235**, 2272-7.

Fontenele-Neto, J.D., Kalinina, E., Feng, Y. & Fricker, L.D. (2005). Identification and distribution of mouse carboxypeptidase A-6. *Brain Res Mol Brain Res* **137**(1-2), 132-42.

Franco, R. F., Fagundes, M. G., Meijers, J. C., Reitsma, P. H., Lourenco, D., Morelli, V., Maffei, F. H., Ferrari, I. C., Piccinato, C. E., Silva, W. A., Jr. & Zago, M. A. (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**(5), 510-7.

Frare, E., Polverino De Laureto, P., Zurdo, J., Dobson, C. M. & Fontana, A. (2004). A highly amyloidogenic region of hen lysozyme. *J Mol Biol* **340**(5), 1153-65.

Fricker, L. D., Adelman, J. P., Douglass, J., Thompson, R. C., von Strandmann, R. P. & 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**(4), 666-73.

Fricker, L. D., Berman, Y. L., Leiter, E. H. & Devi, L. A. (1996). Carboxypeptidase E activity is deficient in mice with the fat mutation. Effect on peptide processing. *J Biol Chem* **271**(48), 30619-24.

Fujinaga, M., Cherney, M. M., Oyama, H., Oda, K. & James, M. N. (2004). The molecular structure and catalytic mechanism of a novel carboxyl peptidase from *Scytalidium lignicolum*. *Proc Natl Acad Sci U S A* **101**(10), 3364-9.

Gaertner, H. F. & Puigserver, A. J. (1992). Increased activity and stability of poly(ethylene glycol)-modified trypsin. *Enzyme Microb Technol* **14**(2), 150-5.

García-Castellanos, R., Bonet-Figueredo, R., Pallares, I., Ventura, S., Avilés, F. X., Vendrell, J. & Gomis-Rutha, F. X. (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**(17), 1996-2014.

García-Sáez, I., Reverter, D., Vendrell, J., Avilés, F. X. & 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**(23), 6906-13.

Gardell, S. J., Craik, C. S., Hilvert, D., Urdea, M. S. & Rutter, W. J. (1985). Site-directed mutagenesis shows that tyrosine 248 of carboxypeptidase A does not play a crucial role in catalysis. *Nature* **317**(6037), 551-5.

Bibliografia

Gardell, S. J., Craik, C. S., Clauser, E., Goldsmith, E. J., Stewart, C. B., Graf, M. & Rutter, W. J. (1988). A novel rat carboxypeptidase, CPA2: characterization, molecular cloning, and evolutionary implications on substrate specificity in the carboxypeptidase gene family. *J Biol Chem* **263**(33), 17828-36.

Gast, K., Zirwer, D., Muller-Frohne, M. & Damaschun, G. (1999). Trifluoroethanol-induced conformational transitions of proteins: insights gained from the differences between alpha-lactalbumin and ribonuclease A. *Protein Sci* **8**(3), 625-34.

Gautron, J., Hincke, M. T., Mann, K., Panheleux, M., Bain, M., McKee, M. D., Solomon, S. E. & Nys, Y. (2001). Ovocalyxin-32, a novel chicken eggshell matrix protein. isolation, amino acid sequencing, cloning, and immunocytochemical localization. *J Biol Chem* **276**(42), 39243-52.

Giasson, B. I., Murray, I. V., Trojanowski, J. Q. & Lee, V. M. (2001). A hydrophobic stretch of 12 amino acid residues in the middle of alpha-synuclein is essential for filament assembly. *J Biol Chem* **276**(4), 2380-6.

Glenner, G. G. (1980). Amyloid deposits and amyloidosis: the beta-fibrilloses (second of two parts). *N Engl J Med* **302**(24), 1333-43.

Goedert, M. (2001). Alpha-synuclein and neurodegenerative diseases. *Nat Rev Neurosci* **2**(7), 492-501.

Goldsbury, C., Goldie, K., Pellaud, J., Seelig, J., Frey, P., Muller, S. A., Kistler, J., Cooper, G. J. & Aebi, U. (2000). Amyloid fibril formation from full-length and fragments of amylin. *J Struct Biol* **130**(2-3), 352-62.

Gomis-Rüth, F. X., Gomez, M., Bode, W., Huber, R. & Avilés, F. X. (1995). The three-dimensional structure of the native ternary complex of bovine pancreatic procarboxypeptidase A with proproteinase E and chymotrypsinogen C. *Embo J* **14**(18), 4387-94.

Gomis-Rüth, F. X., Gomez-Ortiz, M., Vendrell, J., Ventura, S., Bode, W., Huber, R. & Avilés, F. X. (1997a). 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**(5), 861-80.

Gomis-Rüth, F. X., Maskos, K., Betz, M., Bergner, A., Huber, R., Suzuki, K., Yoshida, N., Nagase, H., Brew, K., Bourenkov, G. P., Bartunik, H. & Bode, W. (1997b). Mechanism of inhibition of the human matrix metalloproteinase stromelysin-1 by TIMP-1. *Nature* **389**(6646), 77-81.

Gomis-Rüth, F. X., Companys, V., Qian, Y., Fricker, L. D., Vendrell, J., Avilés, F. X. & Coll, M. (1999). Crystal structure of avian carboxypeptidase D domain II: a prototype for the regulatory metallo-carboxypeptidase subfamily. *Embo J* **18**(21), 5817-26.

Gomis-Rüth, F. X. (2003). Structural aspects of the metzincin clan of metalloendopeptidases. *Mol Biotechnol* **24**(2), 157-202.

Gong, Y., Chang, L., Viola, K. L., Lacor, P. N., Lambert, M. P., Finch, C. E., Krafft, G. A. & Klein, W. L. (2003). Alzheimer's disease-affected brain: presence of oligomeric A beta ligands (ADDLs) suggests a molecular basis for reversible memory loss. *Proc Natl Acad Sci U S A* **100**(18), 10417-22.

Gorisch, H. (1988). Drop dialysis: time course of salt and protein exchange. *Anal Biochem* **173**(2), 393-8.

Greene, D., Das, B. & Fricker, L. D. (1992). Regulation of carboxypeptidase E. Effect of pH, temperature and Co²⁺ on kinetic parameters of substrate hydrolysis. *Biochem J* **285** (Pt 2), 613-8.

Green, J., Goldsbury, C., Mini, T., Sunderji, S., Frey, P., Kistler, J., Cooper, G. & Aebi, U. (2003). Full-length rat amylin forms fibrils following substitution of single residues from human amylin. *J Mol Biol* **326**(4), 1147-56.

Grodberg, J. & Dunn, J. J. (1988). ompT encodes the Escherichia coli outer membrane protease that cleaves T7 RNA polymerase during purification. *J Bacteriol* **170**(3), 1245-53.

Bibliografia

Gu, Y., Fujioka, H., Mishra, R. S., Li, R. & Singh, N. (2002). Prion peptide 106-126 modulates the aggregation of cellular prion protein and induces the synthesis of potentially neurotoxic transmembrane PrP. *J Biol Chem* **277**(3), 2275-86.

Guasch, A., Coll, M., Avilés, F. X. & 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**(1), 141-57.

Guijarro, J. I., Sunde, M., Jones, J. A., Campbell, I. D. & Dobson, C. M. (1998). Amyloid fibril formation by an SH3 domain. *Proc Natl Acad Sci U S A* **95**(8), 4224-8.

Hammarstrom, P., Schneider, F. & Kelly, J. W. (2001). Trans-suppression of misfolding in an amyloid disease. *Science* **293**(5539), 2459-62.

Hammarstrom, P., Wiseman, R. L., Powers, E. T. & Kelly, J. W. (2003). Prevention of transthyretin amyloid disease by changing protein misfolding energetics. *Science* **299**(5607), 713-6.

Harper, J. D. & Lansbury, P. T., 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.

Harrison, R. F., Hawkins, P. N., Roche, W. R., MacMahon, R. F., Hubscher, S. G. & Buckels, J. A. (1996). 'Fragile' liver and massive hepatic haemorrhage due to hereditary amyloidosis. *Gut* **38**(1), 151-2.

Hasegawa, K., Ohhashi, Y., Yamaguchi, I., Takahashi, N., Tsutsumi, S., Goto, Y., Gejyo, F. & Naiki, H. (2003). Amyloidogenic synthetic peptides of beta2-microglobulin--a role of the disulfide bond. *Biochem Biophys Res Commun* **304**(1), 101-6.

Haspel, N., Zanuy, D., Ma, B., Wolfson, H. & Nussinov, R. (2005). A comparative study of amyloid fibril formation by residues 15-19 of the human calcitonin hormone: a single beta-sheet model with a small hydrophobic core. *J Mol Biol* **345**(5), 1213-27.

Hass, G. M. & Hermodson, M. A. (1981). Amino acid sequence of a carboxypeptidase inhibitor from tomato fruit. *Biochemistry* **20**(8), 2256-60.

He, G. P., Muise, A., Li, A. W. & Ro, H. S. (1995). A eukaryotic transcriptional repressor with carboxypeptidase activity. *Nature* **378**(6552), 92-6.

Hege, T., Feltzer, R. E., Gray, R. D. & Baumann, U. (2001). Crystal structure of a complex between *Pseudomonas aeruginosa* alkaline protease and its cognate inhibitor: inhibition by a zinc-NH₂ coordinative bond. *J Biol Chem* **276**(37), 35087-92.

Homandberg, G. A., Litwiller, R. D. & Peanasky, R. J. (1989). Carboxypeptidase inhibitors from *Ascaris suum*: the primary structure. *Arch Biochem Biophys* **270**(1), 153-61.

Hornberg, A., Eneqvist, T., Olofsson, A., Lundgren, E. & Sauer-Eriksson, A. E. (2000). A comparative analysis of 23 structures of the amyloidogenic protein transthyretin. *J Mol Biol* **302**(3), 649-69.

Hoshi, M., Sato, M., Matsumoto, S., Noguchi, A., Yasutake, K., Yoshida, N. & Sato, K. (2003). Spherical aggregates of beta-amyloid (amylospheroid) show high neurotoxicity and activate tau protein kinase I/glycogen synthase kinase-3beta. *Proc Natl Acad Sci U S A* **100**(11), 6370-5.

Huang, H., Reed, C. P., Zhang, J. S., Shridhar, V., Wang, L. & Smith, D. I. (1999). Carboxypeptidase A3 (CPA3): a novel gene highly induced by histone deacetylase inhibitors during differentiation of prostate epithelial cancer cells. *Cancer Res* **59**(12), 2981-8.

Huang, Y., Cheung, L., Rowe, D. & Halliday, G. (2004). Genetic contributions to Parkinson's disease. *Brain Res Brain Res Rev* **46**(1), 44-70.

Iakoucheva, L. M., Brown, C. J., Lawson, J. D., Obradovic, Z. & Dunker, A. K. (2002). Intrinsic disorder in cell-signaling and cancer-associated proteins. *J Mol Biol* **323**(3), 573-84.

Ivanova, M. I., Gingery, M., Whitson, L. J. & Eisenberg, D. (2003). Role of the C-terminal 28 residues of beta2-microglobulin in amyloid fibril formation. *Biochemistry* **42**(46), 13536-40.

Bibliografia

- Ivanova, M. I., Sawaya, M. R., Gingery, M., Attinger, A. & Eisenberg, D. (2004). An amyloid-forming segment of beta2-microglobulin suggests a molecular model for the fibril. *Proc Natl Acad Sci U S A* **101**(29), 10584-9.
- Jaikaran, E. T., Higham, C. E., Serpell, L. C., Zurdo, J., Gross, M., Clark, A. & Fraser, P. E. (2001). Identification of a novel human islet amyloid polypeptide beta-sheet domain and factors influencing fibrillogenesis. *J Mol Biol* **308**(3), 515-25.
- Janin, J. & Chothia, C. (1990). The structure of protein-protein recognition sites. *J Biol Chem* **265**(27), 16027-30.
- Jaroniec, C. P., MacPhee, C. E., Astrof, N. S., Dobson, C. M. & Griffin, R. G. (2002). Molecular conformation of a peptide fragment of transthyretin in an amyloid fibril. *Proc Natl Acad Sci U S A* **99**(26), 16748-53.
- Jaroniec, C. P., MacPhee, C. E., Bajaj, V. S., McMahon, M. T., Dobson, C. M. & Griffin, R. G. (2004). High-resolution molecular structure of a peptide in an amyloid fibril determined by magic angle spinning NMR spectroscopy. *Proc Natl Acad Sci U S A* **101**(3), 711-6.
- Jarrett, J. T., Berger, E. P. & Lansbury, P. T., Jr. (1993). The carboxy terminus of the beta amyloid protein is critical for the seeding of amyloid formation: implications for the pathogenesis of Alzheimer's disease. *Biochemistry* **32**(18), 4693-7.
- Jarvis, J. A., Kirkpatrick, A. & Craik, D. J. (1994). ¹H NMR analysis of fibril-forming peptide fragments of transthyretin. *Int J Pept Protein Res* **44**(4), 388-98.
- Ji, B., Chen, X. Q., Misek, D. E., Kuick, R., Hanash, S., Ernst, S., Najarian, R. & Logsdon, C. D. (2003). Pancreatic gene expression during the initiation of acute pancreatitis: identification of EGR-1 as a key regulator. *Physiol Genomics* **14**(1), 59-72.
- Jimenez, J. L., Nettleton, E. J., Bouchard, M., Robinson, C. V., Dobson, C. M. & Saibil, H. R. (2002). The protofilament structure of insulin amyloid fibrils. *Proc Natl Acad Sci U S A* **99**(14), 9196-201.

Jones, B. N. & Reifler, B. V. (1994). Depression coexisting with dementia. Evaluation and treatment. *Med Clin North Am* **78**(4), 823-40.

Jones, S., Manning, J., Kad, N. M. & Radford, S. E. (2003). Amyloid-forming peptides from beta2-microglobulin-Insights into the mechanism of fibril formation in vitro. *J Mol Biol* **325**(2), 249-57.

Joshi, L. & St Leger, R. J. (1999). Cloning, expression, and substrate specificity of MeCPA, a zinc carboxypeptidase that is secreted into infected tissues by the fungal entomopathogen *Metarhizium anisopliae*. *J Biol Chem* **274**(14), 9803-11.

Juvvadi, S., Fan, X., Nagle, G. T. & Fricker, L. D. (1997). Characterization of Aplysia carboxypeptidase E. *FEBS Lett* **408**(2), 195-200.

Kajava, A. V., Aebi, U. & Steven, A. C. (2005). The parallel superpleated beta-structure as a model for amyloid fibrils of human amylin. *J Mol Biol* **348**(2), 247-52.

Kallberg, Y., Gustafsson, M., Persson, B., Thyberg, J. & Johansson, J. (2001). Prediction of amyloid fibril-forming proteins. *J Biol Chem* **276**(16), 12945-50.

Kaneko, K., Zulianello, L., Scott, M., Cooper, C. M., Wallace, A. C., James, T. L., Cohen, F. E. & Prusiner, S. B. (1997). Evidence for protein X binding to a discontinuous epitope on the cellular prion protein during scrapie prion propagation. *Proc Natl Acad Sci U S A* **94**(19), 10069-74.

Karplus, M. & Sali, A. (1995). Theoretical studies of protein folding and unfolding. *Curr Opin Struct Biol* **5**(1), 58-73.

Kassell, B. & Kay, J. (1973). Zymogens of proteolytic enzymes. *Science* **180**(90), 1022-7.

Kayashima, T., Yamasaki, K., Yamada, T., Sakai, H., Miwa, N., Ohta, T., Yoshiura, K., Matsumoto, N., Nakane, Y., Kanetake, H., Ishino, F., Niikawa, N. & Kishino, T. (2003). The novel imprinted carboxypeptidase A4 gene (CPA4) in the 7q32 imprinting domain. *Hum Genet* **112**(3), 220-6.

Bibliografia

Kelly, J. W. (1998a). The environmental dependency of protein folding best explains prion and amyloid diseases. *Proc Natl Acad Sci U S A* **95**(3), 930-2.

Kelly, J. W. (1998b). The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways. *Curr Opin Struct Biol* **8**, 101-6.

Khurana, R., Gillespie, J. R., Talapatra, A., Minert, L. J., Ionescu-Zanetti, C., Millett, I. & Fink, A. L. (2001). Partially folded intermediates as critical precursors of light chain amyloid fibrils and amorphous aggregates. *Biochemistry* **40**(12), 3525-35.

Kihara, M., Chatani, E., Sakai, M., Hasegawa, K., Naiki, H. & Goto, Y. (2005). Seeding-dependent maturation of beta2-microglobulin amyloid fibrils at neutral pH. *J Biol Chem* **280**(12), 12012-8.

Kilshtain-Vardi, A., Glick, M., Greenblatt, H. M., Goldblum, A. & Shoham, G. (2003). Refined structure of bovine carboxypeptidase A at 1.25 Å resolution. *Acta Crystallogr D Biol Crystallogr* **59**(Pt 2), 323-33.

Kim, P. S. & Baldwin, R. L. (1990). Intermediates in the folding reactions of small proteins. *Annu Rev Biochem* **59**, 631-60.

Kim, H. & Lipscomb, W. N. (1991). Comparison of the structures of three carboxypeptidase A-phosphonate complexes determined by X-ray crystallography. *Biochemistry* **30**(33), 8171-80.

Klein, W. L., Stine, W. B., Jr. & Teplow, D. B. (2004). Small assemblies of unmodified amyloid beta-protein are the proximate neurotoxin in Alzheimer's disease. *Neurobiol Aging* **25**(5), 569-80.

Klunk, W. E., Pettegrew, J. W. & Abraham, D. J. (1989). Quantitative evaluation of congo red binding to amyloid-like proteins with a beta-pleated sheet conformation. *J Histochem Cytochem* **37**(8), 1273-81.

Kopito, R. R. & Ron, D. (2000). Conformational disease. *Nat Cell Biol* **2**(11), E207-9.

- Kozhukh, G. V., Hagihara, Y., Kawakami, T., Hasegawa, K., Naiki, H. & Goto, Y. (2002). Investigation of a peptide responsible for amyloid fibril formation of beta 2-microglobulin by achromobacter protease I. *J Biol Chem* **277**(2), 1310-5.
- Krebs, M. R., Wilkins, D. K., Chung, E. W., Pitkeathly, M. C., Chamberlain, A. K., Zurdo, J., Robinson, C. V. & Dobson, C. M. (2000). Formation and seeding of amyloid fibrils from wild-type hen lysozyme and a peptide fragment from the beta-domain. *J Mol Biol* **300**(3), 541-9.
- Kuroki, K., Cheung, R., Marion, P. L. & Ganem, D. (1994). A cell surface protein that binds avian hepatitis B virus particles. *J Virol* **68**(4), 2091-6.
- Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**(5259), 680-5.
- Lai, Z., Colon, W. & Kelly, J. W. (1996). The acid-mediated denaturation pathway of transthyretin yields a conformational intermediate that can self-assemble into amyloid. *Biochemistry* **35**(20), 6470-82.
- Laskowski, R., MacArthur, M., Hutchinson, E. & Thornton, J. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallog* **26**, 283-91.
- Lawrence, M. C. & Colman, P. M. (1993). Shape complementarity at protein/protein interfaces. *J Mol Biol* **234**(4), 946-50.
- Layne, M. D., Endege, W. O., Jain, M. K., Yet, S. F., Hsieh, C. M., Chin, M. T., Perrella, M. A., Blonar, M. A., Haber, E. & Lee, M. E. (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**(25), 15654-60.
- Lee, K. W., Lee, S. H., Kim, H., Song, J. S., Yang, S. D., Paik, S. G. & Han, P. L. (2004). Progressive cognitive impairment and anxiety induction in the absence of plaque deposition in C57BL/6 inbred mice expressing transgenic amyloid precursor protein. *J Neurosci Res* **76**(4), 572-80.

Bibliografia

Lei, Y., Xin, X., Morgan, D., Pintar, J. E. & Fricker, L. D. (1999). Identification of mouse CPX-1, a novel member of the metallocarboxypeptidase gene family with highest similarity to CPX-2. *DNA Cell Biol* **18**(2), 175-85.

Leslie, A. G. (1999). Integration of macromolecular diffraction data. *Acta Crystallogr D Biol Crystallogr* **55 (Pt 10)**, 1696-702.

LeVine, H., 3rd. (1993). Thioflavine T interaction with synthetic Alzheimer's disease beta-amyloid peptides: detection of amyloid aggregation in solution. *Protein Sci* **2**(3), 404-10.

LeVine, H., 3rd. (1997). Stopped-flow kinetics reveal multiple phases of thioflavin T binding to Alzheimer beta (1-40) amyloid fibrils. *Arch Biochem Biophys* **342**(2), 306-16.

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**(1), 345-53.

Litvinovich, S. V., Brew, S. A., Aota, S., Akiyama, S. K., Haudenschild, C. & Ingham, K. C. (1998). Formation of amyloid-like fibrils by self-association of a partially unfolded fibronectin type III module. *J Mol Biol* **280**(2), 245-58.

Liu, Q., Yu, L., Gao, J., Fu, Q., Zhang, J., Zhang, P., Chen, J. & Zhao, S. (2000). Cloning, tissue expression pattern and genomic organization of latexin, a human homologue of rat carboxypeptidase A inhibitor. *Mol Biol Rep* **27**(4), 241-6.

Lu, L., Gunja-Smith, Z., Woessner, J. F., Ursell, P. C., Nissen, T., Galardy, R. E., Xu, Y., Zhu, P. & Schwartz, G. G. (2000). Matrix metalloproteinases and collagen ultrastructure in moderate myocardial ischemia and reperfusion in vivo. *Am J Physiol Heart Circ Physiol* **279**(2), H601-9.

MacPhee, C. E. & Dobson, C. M. (2000). Chemical dissection and reassembly of amyloid fibrils formed by a peptide fragment of transthyretin. *J Mol Biol* **297**(5), 1203-15.

- Mandel, M. & Higa, A. (1970). Calcium-dependent bacteriophage DNA infection. *J Mol Biol* **53**(1), 159-62.
- Marti-Renom, M. A., Ilyin, V. A. & Sali, A. (2001). DBAli: a database of protein structure alignments. *Bioinformatics* **17**(8), 746-7.
- Masuda, K., Yoshioka, M., Hinode, D. & Nakamura, R. (2002). Purification and characterization of arginine carboxypeptidase produced by *Porphyromonas gingivalis*. *Infect Immun* **70**(4), 1807-15.
- Mayer, B. J. (2001). SH3 domains: complexity in moderation. *J Cell Sci* **114**(Pt 7), 1253-63.
- McCutchen, S. L., Lai, Z., Miroy, G. J., Kelly, J. W. & Colon, W. (1995). Comparison of lethal and nonlethal transthyretin variants and their relationship to amyloid disease. *Biochemistry* **34**(41), 13527-36.
- McGwire, G. B. & Skidgel, R. A. (1995). Extracellular conversion of epidermal growth factor (EGF) to des-Arg53-EGF by carboxypeptidase M. *J Biol Chem* **270**(29), 17154-8.
- McParland, V. J., Kad, N. M., Kalverda, A. P., Brown, A., Kirwin-Jones, P., Hunter, M. G., Sunde, M. & Radford, S. E. (2000). Partially unfolded states of beta(2)-microglobulin and amyloid formation in vitro. *Biochemistry* **39**(30), 8735-46.
- McParland, V. J., Kalverda, A. P., Homans, S. W. & Radford, S. E. (2002). Structural properties of an amyloid precursor of beta(2)-microglobulin. *Nat Struct Biol* **9**(5), 326-31.
- Miake, H., Mizusawa, H., Iwatsubo, T. & Hasegawa, M. (2002). Biochemical characterization of the core structure of alpha-synuclein filaments. *J Biol Chem* **277**(21), 19213-9.
- Mirnic, Z. K., Mirnic, K., Terrano, D., Lewis, D. A., Sisodia, S. S. & Schor, N. F. (2003). DNA microarray profiling of developing PS1-deficient mouse brain reveals complex and coregulated expression changes. *Mol Psychiatry* **8**(10), 863-78.

Bibliografia

Miyasaka, N., Hatanaka, Y., Jin, M. & Arimatsu, Y. (1999). Genomic organization and regulatory elements of the rat latexin gene, which is expressed in a cell type-specific manner in both central and peripheral nervous systems. *Brain Res Mol Brain Res* **69**(1), 62-72.

Mock, W. L., Liu, Y. & Stanford, D. J. (1996). Arazoformyl peptide surrogates as spectrophotometric kinetic assay substrates for carboxypeptidase A. *Anal Biochem* **239**(2), 218-22.

Moriarty, D. F. & Raleigh, D. P. (1999). Effects of sequential proline substitutions on amyloid formation by human amylin₂₀₋₂₉. *Biochemistry* **38**(6), 1811-8.

Morimoto, A., Irie, K., Murakami, K., Masuda, Y., Ohigashi, H., Nagao, M., Fukuda, H., Shimizu, T. & Shirasawa, T. (2004). Analysis of the secondary structure of beta-amyloid (A_β42) fibrils by systematic proline replacement. *J Biol Chem* **279**(50), 52781-8.

Mosnier, L. O., Lisman, T., van den Berg, H. M., Nieuwenhuis, H. K., Meijers, J. C. & Bouma, B. N. (2001). The defective down regulation of fibrinolysis in haemophilia A can be restored by increasing the TAFI plasma concentration. *Thromb Haemost* **86**(4), 1035-9.

Muise, A. M. & Ro, H. S. (1999). Enzymic characterization of a novel member of the regulatory B-like carboxypeptidase with transcriptional repression function: stimulation of enzymic activity by its target DNA. *Biochem J* **343 Pt 2**, 341-5.

Nagele, R. G., Wegiel, J., Venkataraman, V., Imaki, H. & Wang, K. C. (2004). Contribution of glial cells to the development of amyloid plaques in Alzheimer's disease. *Neurobiol Aging* **25**(5), 663-74.

Nagpal, S., Patel, S., Asano, A. T., Johnson, A. T., Duvic, M. & Chandraratna, R. A. (1996). Tazarotene-induced gene 1 (TIG1), a novel retinoic acid receptor-responsive gene in skin. *J Invest Dermatol* **106**(2), 269-74.

Narahashi, Y. (1990). The amino acid sequence of zinc-carboxypeptidase from *Streptomyces griseus*. *J Biochem (Tokyo)* **107**(6), 879-86.

Navaza, J. (1994). An automated package for molecular replacement. *Acta Crystallog sect A* **50**, 157-63.

Nelson, R., Sawaya, M.R., Balbirnie, M., Madsen, A.O., Riekel, C., Grothe, R., & Eisenberg, D. (2005). Structure of the cross-beta spine of amyloid-like fibrils. *Nature* **435**(7043), 773-8.

Normant, E., Gros, C. & Schwartz, J. C. (1995a). Carboxypeptidase A isoforms produced by distinct genes or alternative splicing in brain and other extrapancreatic tissues. *J Biol Chem* **270**(35), 20543-9.

Normant, E., Martres, M. P., Schwartz, J. C. & Gros, C. (1995b). Purification, cDNA cloning, functional expression, and characterization of a 26-kDa endogenous mammalian carboxypeptidase inhibitor. *Proc Natl Acad Sci U S A* **92**(26), 12225-9.

Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzym* **276**, 307-26.

Pallares, I., Bonet, R., García-Castellanos, R., Ventura, S., Avilés, F. X., Vendrell, J. & Gomis-Ruth, F. X. (2005). Structure of human carboxypeptidase A4 with its endogenous protein inhibitor, latexin. *Proc Natl Acad Sci U S A* **102**(11), 3978-83.

Pascual, R., Burgos, F. J., Salva, M., Soriano, F., Mendez, E. & Avilés, F. X. (1989). Purification and properties of five different forms of human procarboxypeptidases. *Eur J Biochem* **179**(3), 609-16.

Pascual, R., Vendrell, J., Avilés, F. X., 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**(1-2), 37-41.

Pawar, A. P., Dubay, K. F., Zurdo, J., Chiti, F., Vendruscolo, M. & Dobson, C. M. (2005). Prediction of "aggregation-prone" and "aggregation-susceptible" regions in proteins associated with neurodegenerative diseases. *J Mol Biol* **350**(2), 379-92.

Pepys, M. B., Hawkins, P. N., Booth, D. R., Vigushin, D. M., Tennent, G. A., Soutar, A. K., Totty, N., Nguyen, O., Blake, C. C., Terry, C. J. & et al. (1993). Human

Bibliografia

lysozyme gene mutations cause hereditary systemic amyloidosis. *Nature* **362**(6420), 553-7.

Peranen, J., Rikkonen, M., Hyvonen, M. & Kaariainen, L. (1996). T7 vectors with modified T7lac promoter for expression of proteins in *Escherichia coli*. *Anal Biochem* **236**(2), 371-3.

Perona, J. J. & Craik, C. S. (1997). Evolutionary divergence of substrate specificity within the chymotrypsin-like serine protease fold. *J Biol Chem* **272**(48), 29987-90.

Peterson, L. M., Holmquist, B. & Bethune, J. L. (1982). A unique activity assay for carboxypeptidase A in human serum. *Anal Biochem* **125**(2), 420-6.

Petkova, A. T., Ishii, Y., Balbach, J. J., Antzutkin, O. N., Leapman, R. D., Delaglio, F. & Tycko, R. (2002). A structural model for Alzheimer's beta -amyloid fibrils based on experimental constraints from solid state NMR. *Proc Natl Acad Sci U S A* **99**(26), 16742-7.

Phillips, M. A., Fletterick, R. & Rutter, W. J. (1990). Arginine 127 stabilizes the transition state in carboxypeptidase. *J Biol Chem* **265**(33), 20692-8.

Phillips, M. A., Kaplan, A. P., Rutter, W. J. & Bartlett, P. A. (1992). Transition-state characterization: a new approach combining inhibitor analogues and variation in enzyme structure. *Biochemistry* **31**(4), 959-63.

Phillips, M. A. & Rutter, W. J. (1996). Role of the prodomain in folding and secretion of rat pancreatic carboxypeptidase A1. *Biochemistry* **35**(21), 6771-6.

Piekarska, B., Rybarska, J., Stopa, B., Zemanek, G., Krol, M., Roterman, I. & Konieczny, L. (1999). Supramolecularity creates nonstandard protein ligands. *Acta Biochim Pol* **46**(4), 841-51.

Plummer, T. H., Jr. & Kimmel, M. T. (1980). An improved spectrophotometric assay for human plasma carboxypeptidase N1. *Anal Biochem* **108**(2), 348-53.

Porte, D., Jr. & Kahn, S. E. (1989). Hyperproinsulinemia and amyloid in NIDDM. Clues to etiology of islet beta-cell dysfunction? *Diabetes* **38**(11), 1333-6.

Prusiner, S. B., Scott, M. R., DeArmond, S. J. & Cohen, F. E. (1998). Prion protein biology. *Cell* **93**(3), 337-48.

Puigserver, A., Chapus, C. & Kerfelec, B. (1986). Pancreatic exopeptidases. In: Molecular and cellular basis of digestion. Eds. Desnuelle P. *et al.* Elsevier Publ., London, 235-247.

Quioco, F. A. & Lipscomb, W. N. (1971). Carboxypeptidase A: a protein and an enzyme. *Adv Protein Chem* **25**, 1-78.

Radford, S. E. & Dobson, C. M. (1999). From computer simulations to human disease: emerging themes in protein folding. *Cell* **97**(3), 291-8.

Ramos, A., Mahowald, A. & Jacobs-Lorena, M. (1993). Gut-specific genes from the black fly *Simulium vittatum* encoding trypsin-like and carboxypeptidase-like proteins. *Insect Mol Biol* **1**(3), 149-63.

Rath-Wolfson, L. (2001). An immunocytochemical approach to the demonstration of intracellular processing of mast cell carboxypeptidase. *Appl Immunohistochem Mol Morphol* **9**(1), 81-5.

Rawlings, N. D. & Barrett, A. J. (1993). Evolutionary families of peptidases. *Biochem J* **290** (Pt 1), 205-18.

Rawlings, N. D. & Barrett, A. J. (1995). Evolutionary families of metallopeptidases. *Methods Enzymol* **248**, 183-228.

Rawlings, N. D. & Barrett, A. J. (2000). MEROPS: the peptidase database. *Nucleic Acids Res* **28**(1), 323-5.

Rawlings, N. D., Tolle, D. P. & Barrett, A. J. (2004). Evolutionary families of peptidase inhibitors. *Biochem J* **378**(Pt 3), 705-16.

Bibliografia

Rees, D. C. & Lipscomb, W. N. (1980). Structure of the potato inhibitor complex of carboxypeptidase A at 2.5-Å resolution. *Proc Natl Acad Sci U S A* **77**(8), 4633-7.

Rees, D. C. & Lipscomb, W. N. (1981). Binding of ligands to the active site of carboxypeptidase A. *Proc Natl Acad Sci U S A* **78**(9), 5455-9.

Rees, D. C. & Lipscomb, W. N. (1982). Refined crystal structure of the potato inhibitor complex of carboxypeptidase A at 2.5 Å resolution. *J Mol Biol* **160**(3), 475-98.

Rees, D. C., Lewis, M. & Lipscomb, W. N. (1983). Refined crystal structure of carboxypeptidase A at 1.54 Å resolution. *J Mol Biol* **168**(2), 367-87.

Reverter, D., Vendrell, J., Canals, F., Horstmann, J., Avilés, F. X., Fritz, H. & Sommerhoff, C. P. (1998a). A carboxypeptidase inhibitor from the medical leech *Hirudo medicinalis*. Isolation, sequence analysis, cDNA cloning, recombinant expression, and characterization. *J Biol Chem* **273**(49), 32927-33.

Reverter, D., Ventura, S., Villegas, V., Vendrell, J. & Avilés, F. X. (1998b). Overexpression of human procarboxypeptidase A2 in *Pichia pastoris* and detailed characterization of its activation pathway. *J Biol Chem* **273**(6), 3535-41.

Reverter, D., Fernandez-Catalan, C., Baumgartner, R., Pfander, R., Huber, R., Bode, W., Vendrell, J., Holak, T. A. & Avilés, F. X. (2000). Structure of a novel leech carboxypeptidase inhibitor determined free in solution and in complex with human carboxypeptidase A2. *Nat Struct Biol* **7**(4), 322-8.

Reverter, D., Maskos, K., Tan, F., Skidgel, R. A. & Bode, W. (2004). Crystal structure of human carboxypeptidase M, a membrane-bound enzyme that regulates peptide hormone activity. *J Mol Biol* **338**(2), 257-69.

Reynolds, D. S., Stevens, R. L., Gurley, D. S., Lane, W. S., Austen, K. F. & Serafin, W. E. (1989). Isolation and molecular cloning of mast cell carboxypeptidase A. A novel member of the carboxypeptidase gene family. *J Biol Chem* **264**(33), 20094-9.

Reznik, S. E. & Fricker, L. D. (2001). Carboxypeptidases from A to z: implications in embryonic development and Wnt binding. *Cell Mol Life Sci* **58**(12-13), 1790-804.

- Richardson, J. S. & Richardson, D. C. (2002). Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation. *Proc Natl Acad Sci U S A* **99**(5), 2754-9.
- Rochet, J. C. & Lansbury, P. T., Jr. (2000). Amyloid fibrillogenesis: themes and variations. *Curr Opin Struct Biol* **10**(1), 60-8.
- Roterman, I., KrUl, M., Nowak, M., Konieczny, L., Rybarska, J., Stopa, B., Piekarska, B. & Zemanek, G. (2001). Why Congo red binding is specific for amyloid proteins - model studies and a computer analysis approach. *Med Sci Monit* **7**(4), 771-84.
- Safar, J., Roller, P. P., Gajdusek, D. C. & Gibbs, C. J., Jr. (1994). Scrapie amyloid (prion) protein has the conformational characteristics of an aggregated molten globule folding intermediate. *Biochemistry* **33**(27), 8375-83.
- Saha, A. R., Hill, J., Utton, M. A., Asuni, A. A., Ackerley, S., Grierson, A. J., Miller, C. C., Davies, A. M., Buchman, V. L., Anderton, B. H. & Hanger, D. P. (2004). Parkinson's disease alpha-synuclein mutations exhibit defective axonal transport in cultured neurons. *J Cell Sci* **117**(Pt 7), 1017-24.
- Sakagashira, S., Sanke, T., Hanabusa, T., Shimomura, H., Ohagi, S., Kumagaye, K. Y., Nakajima, K. & Nanjo, K. (1996). Missense mutation of amylin gene (S20G) in Japanese NIDDM patients. *Diabetes* **45**(9), 1279-81.
- Sambrook, J., Fritsch, E.F., & Maniatis, T. (1989). "Molecular cloning: a laboratory manual" 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- Sanchez de Groot, N., Pallares, I., Avilés, F. X., 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, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **74**(12), 5463-7.
- Santini, S., Wei, G., Mousseau, N. & Derreumaux, P. (2004). Pathway complexity of Alzheimer's beta-amyloid Abeta16-22 peptide assembly. *Structure* **12**(7), 1245-55.

Bibliografia

Saraiva, M. J., Birken, S., Costa, P. P. & Goodman, D. S. (1984). Amyloid fibril protein in familial amyloidotic polyneuropathy, Portuguese type. Definition of molecular abnormality in transthyretin (prealbumin). *J Clin Invest* **74**(1), 104-19.

Schechter, I., & Berger, A. (1967). On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun* **27**, 157-162.

Scrocchi, L. A., Ha, K., Chen, Y., Wu, L., Wang, F. & Fraser, P. E. (2003). Identification of minimal peptide sequences in the (8-20) domain of human islet amyloid polypeptide involved in fibrillogenesis. *J Struct Biol* **141**(3), 218-27.

Seemuller, E., Lupas, A., Zuhl, F., Zwickl, P. & Baumeister, W. (1995). The proteasome from *Thermoplasma acidophilum* is neither a cysteine nor a serine protease. *FEBS Lett* **359**(2-3), 173-8.

Segura-Martin, S. (2002). Procarboxipeptidases humanes: Estudis funcionals i estructurals de formes pancreàtiques i de noves formes reguladores. *Ph D Thesis*, Universitat Autònoma de Barcelona.

Selkoe, D. J. (2001). Alzheimer's disease results from the cerebral accumulation and cytotoxicity of amyloid beta-protein. *J Alzheimers Dis* **3**(1), 75-80.

Serra, M.A. (1992). Estudi de la cinètica d'hidròlisi de pèptids per la carboxipeptidasa A pancreàtica. *Ph D Thesis*, Universitat Autònoma de Barcelona.

Settle, S. H., Jr., Green, M. M. & Burtis, K. C. (1995). The silver gene of *Drosophila melanogaster* encodes multiple carboxypeptidases similar to mammalian prohormone-processing enzymes. *Proc Natl Acad Sci U S A* **92**(21), 9470-4.

Shapiro, A. L., Vinuela, E. & Maizel, J. V., Jr. (1967). Molecular weight estimation of polypeptide chains by electrophoresis in SDS-polyacrylamide gels. *Biochem Biophys Res Commun* **28**(5), 815-20.

Skidgel, R. A., Weerasinghe, D. K. & Erdos, E. G. (1989). Structure of human carboxypeptidase N (kininase I). *Adv Exp Med Biol* **247A**, 325-9.

Skidgel, R. A., Tan, F. L., Deddish, P. A. & Li, X. Y. (1991). Structure, function and membrane anchoring of carboxypeptidase M. *Biomed Biochim Acta* **50**(4-6), 815-20.

Smith, A. (2003). Protein misfolding. *Nature* **426**, 883.

Smith, G. K., Banks, S., Blumenkopf, T. A., Cory, M., Humphreys, J., Laethem, R. M., Miller, J., Moxham, C. P., Mullin, R., Ray, P. H., Walton, L. M. & Wolfe, L. A., 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**(25), 15804-16.

Smulevitch, S. V., Osterman, A. L., Galperina, O. V., Matz, M. V., Zagnitko, O. P., Kadyrov, R. M., Tsaplina, I. A., Grishin, N. V., Chestukhina, G. G. & Stepanov, V. M. (1991). Molecular cloning and primary structure of *Thermoactinomyces vulgaris* carboxypeptidase T. A metalloenzyme endowed with dual substrate specificity. *FEBS Lett* **291**(1), 75-8.

Somoza, J. R., Jiang, F., Tong, L., Kang, C. H., Cho, J. M. & Kim, S. H. (1993). Two crystal structures of a potently sweet protein. Natural monellin at 2.75 Å resolution and single-chain monellin at 1.7 Å resolution. *J Mol Biol* **234**(2), 390-404.

Song, L. & Fricker, L. D. (1995a). Calcium- and pH-dependent aggregation of carboxypeptidase E. *J Biol Chem* **270**(14), 7963-7.

Song, L. & Fricker, L. (1995b). Processing of procarboxypeptidase E into carboxypeptidase E occurs in secretory vesicles. *J Neurochem* **65**(1), 444-53.

Song, L. & Fricker, L. D. (1997). Cloning and expression of human carboxypeptidase Z, a novel metallo-carboxypeptidase. *J Biol Chem* **272**(16), 10543-50.

Sorenson, P., Winther, J. R., Kaarsholm, N. C. & Poulsen, F. M. (1993). The pro region required for folding of carboxypeptidase Y is a partially folded domain with little regular structural core. *Biochemistry* **32**(45), 12160-6.

Spillantini, M. G., Schmidt, M. L., Lee, V. M., Trojanowski, J. Q., Jakes, R. & Goedert, M. (1997). Alpha-synuclein in Lewy bodies. *Nature* **388**(6645), 839-40.

Bibliografia

Springman, E.B. (2004). In: Handbook of proteolytic Enzymes, Barrett, A.J., Rawlings N.D. & woessner, J. F. Jr (eds), Elsevier, London, vol 1, 828-30.

Srisailam, S., Wang, H. M., Kumar, T. K., Rajalingam, D., Sivaraja, V., Sheu, H. S., Chang, Y. C. & Yu, C. (2002). Amyloid-like fibril formation in an all beta-barrel protein involves the formation of partially structured intermediate(s). *J Biol Chem* **277**(21), 19027-36.

Srisailam, S., Kumar, T. K., Rajalingam, D., Kathir, K. M., Sheu, H. S., Jan, F. J., Chao, P. C. & Yu, C. (2003). Amyloid-like fibril formation in an all beta-barrel protein. Partially structured intermediate state(s) is a precursor for fibril formation. *J Biol Chem* **278**(20), 17701-9.

Stone, T. E., Li, J. P. & Bernasconi, P. (1994). Purification and characterization of the *Manduca sexta* neuropeptide processing enzyme carboxypeptidase E. *Arch Insect Biochem Physiol* **27**(3), 193-203.

Stubbs, M. T., Laber, B., Bode, W., Huber, R., Jerala, R., Lenarcic, B. & Turk, V. (1990). The refined 2.4 Å X-ray crystal structure of recombinant human stefin B in complex with the cysteine proteinase papain: a novel type of proteinase inhibitor interaction. *Embo J* **9**(6), 1939-47.

Sunde, M. & Blake, C. (1997). The structure of amyloid fibrils by electron microscopy and X-ray diffraction. *Adv Protein Chem* **50**, 123-59.

Sunde, M. & Blake, C. C. (1998). From the globular to the fibrous state: protein structure and structural conversion in amyloid formation. *Q Rev Biophys* **31**(1), 1-39.

Svensson, M., Hakansson, A., Mossberg, A. K., Linse, S. & Svanborg, C. (2000). Conversion of alpha-lactalbumin to a protein inducing apoptosis. *Proc Natl Acad Sci U S A* **97**(8), 4221-6.

Tagliavini, F., Prelli, F., Verga, L., Giaccone, G., Sarma, R., Gorevic, P., Ghetti, B., Passerini, F., Ghibaudi, E., Forloni, G. & et al. (1993). Synthetic peptides

homologous to prion protein residues 106-147 form amyloid-like fibrils in vitro. *Proc Natl Acad Sci U S A* **90**(20), 9678-82.

Tan, S. Y. & Pepys, M. B. (1994). Amyloidosis. *Histopathology* **25**(5), 403-14.

Tan, F., Rehli, M., Krause, S. W. & Skidgel, R. A. (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), 81-7.

Tartaglia, G. G., Cavalli, A., Pellarin, R. & Caflisch, A. (2004). The role of aromaticity, exposed surface, and dipole moment in determining protein aggregation rates. *Protein Sci* **13**(7), 1939-41.

Teplyakov, A., Polyakov, K., Obmolova, G., Strokopytov, B., Kuranova, I., Osterman, A., Grishin, N., Smulevitch, S., Zagnitko, O., Galperina, O. & et al. (1992). Crystal structure of carboxypeptidase T from *Thermoactinomyces vulgaris*. *Eur J Biochem* **208**(2), 281-8.

Tompa, P. (2002). Intrinsically unstructured proteins. *Trends Biochem Sci* **27**(10), 527-33.

Torok, M., Milton, S., Kaye, R., Wu, P., McIntire, T., Glabe, C. G. & Langen, R. (2002). Structural and dynamic features of Alzheimer's A β peptide in amyloid fibrils studied by site-directed spin labeling. *J Biol Chem* **277**(43), 40810-5.

Tracz, S. M., Abedini, A., Driscoll, M. & Raleigh, D. P. (2004). Role of aromatic interactions in amyloid formation by peptides derived from human Amylin. *Biochemistry* **43**(50), 15901-8.

Tsai, P. L., Chen, C. H., Huang, C. J., Chou, C. M. & Chang, G. D. (2004). Purification and cloning of an endogenous protein inhibitor of carp nephrosin, an astacin metalloproteinase. *J Biol Chem* **279**(12), 11146-55.

Turk, D. (1992). Weitereentwicklung eines Programms für Molekulargraphik und Elektronendichte-Manipulation und seine Anwendung auf verschiedene Protein-Strukturaufklärungen (Munich: Technische Universität München).

Bibliografia

Uratani, Y., Takiguchi-Hayashi, K., Miyasaka, N., Sato, M., Jin, M. & Arimatsu, Y. (2000). Latexin, a carboxypeptidase A inhibitor, is expressed in rat peritoneal mast cells and is associated with granular structures distinct from secretory granules and lysosomes. *Biochem J* **346 Pt 3**, 817-26.

Uversky, V. N., Li, J. & Fink, A. L. (2001). Evidence for a partially folded intermediate in alpha-synuclein fibril formation. *J Biol Chem* **276**(14), 10737-44.

Uversky, V. N. (2002a). What does it mean to be natively unfolded? *Eur J Biochem* **269**(1), 2-12.

Uversky, V. N. (2002b). Natively unfolded proteins: a point where biology waits for physics. *Protein Sci* **11**(4), 739-56.

Uversky, V. N. & Fink, A. L. (2004). Conformational constraints for amyloid fibrillation: the importance of being unfolded. *Biochim Biophys Acta* **1698**(2), 131-53.

Vendrell, J., Avilés, F. X., Vilanova, M., Turner, C. H. & Crane-Robinson, C. (1990a). 1H-n.m.r. studies of the isolated activation segment from pig procarboxypeptidase A. *Biochem J* **267**(1), 213-20.

Vendrell, J., Wider, G., Avilés, F. X. & 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**(32), 7515-22.

Vendrell, J., Billeter, M., Wider, G., Avilés, F. X. & Wuthrich, K. (1991). The NMR structure of the activation domain isolated from porcine procarboxypeptidase B. *Embo J* **10**(1), 11-5.

Vendrell, J. & Avilés, F. X. (1999). Carboxypeptidases. In: *Proteases: New Perspectives*. Eds. Turk V., Birkhäuser Verlag Basel, Switzerland, 1-12.

Vendrell, J., Querol, E. & Avilés, F. X. (2000). Metalloproteases and their protein inhibitors. Structure, function and biomedical properties. *Biochim Biophys Acta* **1477**(1-2), 284-98.

Vendrell, J., Avilés, F. X. & Fricker, L.D. (2004). Metalloproteases. In *Handbook of Metalloproteins*. Messerschmidt A., Bode W., Cygler, M. Wiley & Sons, Ltd, Chichester, 176-189.

Ventura, S., Gomis-Ruth, F. X., Puigserver, A., Avilés, F. X. & Vendrell, J. (1997). Pancreatic procarboxypeptidases: oligomeric structures and activation processes revisited. *Biol Chem* **378**(3-4), 161-5.

Ventura, S., Villegas, V., Sterner, J., Larson, J., Vendrell, J., Hershberger, C. L. & Avilés, F. X. (1999). Mapping the pro-region of carboxypeptidase B by protein engineering. Cloning, overexpression, and mutagenesis of the porcine proenzyme. *J Biol Chem* **274**(28), 19925-33.

Ventura, S., Lacroix, E. & Serrano, L. (2002). Insights into the origin of the tendency of the PI3-SH3 domain to form amyloid fibrils. *J Mol Biol* **322**(5), 1147-58.

Ventura, S., Zurdo, J., Narayanan, S., Parreno, M., Manges, R., Reif, B., Chiti, F., Giannoni, E., Dobson, C. M., Avilés, F. X. & 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**(19), 7258-63.

Villegas, V., Vendrell, J. & Avilés, X. (1995). The activation pathway of procarboxypeptidase B from porcine pancreas: participation of the active enzyme in the proteolytic processing. *Protein Sci* **4**(9), 1792-800.

Villegas, V., Zurdo, J., Filimonov, V. V., Avilés, F. X., Dobson, C. M. & Serrano, L. (2000). Protein engineering as a strategy to avoid formation of amyloid fibrils. *Protein Sci* **9**(9), 1700-8.

Wang, W., Boffa, M. B., Bajzar, L., Walker, J. B. & Nesheim, M. E. (1998). A study of the mechanism of inhibition of fibrinolysis by activated thrombin-activable fibrinolysis inhibitor. *J Biol Chem* **273**(42), 27176-81.

Bibliografia

Wei, S., Segura, S., Vendrell, J., Avilés, F. X., Lanoue, E., Day, R., Feng, Y. & Fricker, L. D. (2002). Identification and characterization of three members of the human metallocarboxypeptidase gene family. *J Biol Chem* **277**(17), 14954-64.

Wei, S., Feng, Y., Kalinina, E. & Fricker, L. D. (2003). Neuropeptide-processing carboxypeptidases. *Life Sci* **73**(6), 655-62.

Williams, A. D., Portelius, E., Kheterpal, I., Guo, J. T., Cook, K. D., Xu, Y. & Wetzel, R. (2004). Mapping abeta amyloid fibril secondary structure using scanning proline mutagenesis. *J Mol Biol* **335**(3), 833-42.

Wolynes, P. G., Onuchic, J. N. & Thirumalai, D. (1995). Navigating the folding routes. *Science* **267**(5204), 1619-20.

Wright, H. T. (1973). Comparison of the crystal structures of chymotrypsinogen-A and alpha-chymotrypsin. *J Mol Biol* **79**(1), 1-11.

Wright, P. E. & Dyson, H. J. (1999). Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol* **293**(2), 321-31.

Wurth, C., Guimard, N. K. & Hecht, M. H. (2002). Mutations that reduce aggregation of the Alzheimer's Abeta42 peptide: an unbiased search for the sequence determinants of Abeta amyloidogenesis. *J Mol Biol* **319**(5), 1279-90.

Xin, X., Day, R., Dong, W., Lei, Y. & Fricker, L. D. (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**(10), 897-909.

Yamamoto, K. K., Pousette, A., Chow, P., Wilson, H., el Shami, S. & French, C. K. (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**(4), 2575-81.

Yamamoto, N., Hasegawa, K., Matsuzaki, K., Naiki, H. & Yanagisawa, K. (2004). Environment- and mutation-dependent aggregation behavior of Alzheimer amyloid beta-protein. *J Neurochem* **90**(1), 62-9.

Yang, J. J., Buck, M., Pitkeathly, M., Kotik, M., Haynie, D. T., Dobson, C. M. & Radford, S. E. (1995). Conformational properties of four peptides spanning the sequence of hen lysozyme. *J Mol Biol* **252**(4), 483-91.

Youssef, E. M., Chen, X. Q., Higuchi, E., Kondo, Y., García-Manero, G., Lotan, R. & Issa, J. P. (2004). Hypermethylation and silencing of the putative tumor suppressor Tazarotene-induced gene 1 in human cancers. *Cancer Res* **64**(7), 2411-7.

Zahn, R., Liu, A., Luhrs, T., Riek, R., von Schroetter, C., Lopez García, F., Billeter, M., Calzolari, L., Wider, G. & Wuthrich, K. (2000). NMR solution structure of the human prion protein. *Proc Natl Acad Sci U S A* **97**(1), 145-50.

Zurdo, J., Guijarro, J. I., Jimenez, J. L., Saibil, H. R. & Dobson, C. M. (2001). Dependence on solution conditions of aggregation and amyloid formation by an SH3 domain. *J Mol Biol* **311**(2), 325-40.

SUMMARY

SUMMARY

The present thesis started with the study of metallo-carboxypeptidases and their inhibitors, a field in which our group has been working for many years, and that constitutes the first part of the thesis. It contains two chapters, both of them discussing upon the structure - function relationship in metallo-carboxypeptidases and their inhibitors. Two crystal structures are presented in the first and second works: that of human CPA1 and the structure of latexin, the only known endogenous inhibitor of carboxypeptidases, in complex with human carboxypeptidase A4, in the second. Subsequent conformational studies of human latexin showed that it displays the capability to associate into stable oligomeric structures from its unfolded state. Thus, aggregation studies of latexin open the second part of the thesis which consists of three independent research works that correspond to the field of protein aggregation, a line of research recently developed in our group.

Chapter 1. Three dimensional structure of human carboxypeptidase A1

Human PCPA1 was over-expressed in the supernatant medium of the yeast *Pichia pastoris* heterologous system. Purified protein at 6 mg/mL dissolved in Tris 5 mM, NaCl 50 mM (pH 7.5) was mixed at 1:1 ratio with the crystallization solutions. Crystals grew in 22% (w/v) PAA 5100, HEPES 0.1 M, magnesium chloride 0.02 M (pH 7.5) crystallization buffer using a sitting drop vapor diffusion set-up at 20°C. High resolution data at 1.6 Å were collected at the DESY BW6 beam line. Crystals belonged to the tetragonal space group $P4_32_12$ with unit cell parameters $a=b=129.565$ Å, $c=90.856$ Å. The final crystallographic R factor and R free were 17.82% and 19.18% respectively. The hCPA1 chain encompasses protein residues Ala 1 to Pro 308 and a Zn^{+2} ion. Like its other non-human homologues the tertiary fold of hCPA1 corresponds to the α/β hydrolase fold and is formed by a central mixed parallel/antiparallel 8-stranded β -sheet over which 8 α -helices pack on both sides to form a globular molecule. Additional electron density near the active site was interpreted as a polyacrylic acid (PAA) molecule. The binding of the PAA molecule to the active site as an inhibitor analog delineates the positions of the S_1' and S_1 subsite pockets of the enzyme. Similar interactions, that involve the two main residues of the S_1 site, Arg 127 and Glu 270, are also found in the LCI-CPA2 complex but are not present in the PCI-CPA complex. This structure provides a template for modeling the

structure of physiologically relevant carboxypeptidases, as the ones implied in hormone and neuropeptide processing, as well as for the specific design of inhibitors.

Chapter 2 Structure of human carboxypeptidase A4 with its endogenous protein inhibitor, latexin

Latexin is the only endogenous protein inhibitor known for metalloproteases. Human latexin was isolated from human a fetal brain gene pool cDNA using a two step PCR strategy. Latexin cDNA contains an open reading frame that encodes for a 222 residue protein and does not share sequence similarity with any other structurally characterized protein. The three dimensional structure of human latexin in complex with human CPA4 was been solved by a combination of Patterson search and multiple-wavelength anomalous diffraction (MAD). The measured crystal belongs to the monoclinic space group $P2_1$, harbors two complexes per asymmetric unit and diffracts beyond 1.6 Å resolution. The human latexin structure is elongated, with an α/β topology. It can be divided into two subdomains that share the same fold, each one shaped by an extended α -helix followed by a strongly twisted four-stranded antiparallel β -sheet. The two subdomains are linked by a connecting helical segment that runs along the surface. In the latexin/hCPA4 complex, the inhibitor sits on top of the funnel rim covering an area of 2,340 Å² at the protein interface, a value higher than in other typical protease-inhibitor interfaces. In contrast to other protein inhibitors of metalloproteases, inhibition by latexin does not involve any of its termini, since it is mainly caused by the inhibitory loop $\beta 7\beta 8$. There are some similarities in the way latexin and the prodomains of carboxypeptidases block the hCPA4 enzyme moiety, consisting of a comparably large interface area and relatively few contacts. Latexin-like proteins seem to be restricted to vertebrates since they have been found in organisms ranging from fish and frogs to mammals and other higher vertebrates, but no sequences could be retrieved from invertebrates or other less complex organisms. Invertebrate A/B MCPs and those from the E/N subfamily are not inhibited, neither are several endopeptidases studied. Our molecular studies show that structural determinants simultaneously confer vertebrate-linked specificity to latexins and sufficient variability to strongly inhibit all vertebrate forms tested. Latexins could represent the family of endogenous MCP inhibitors in vertebrates.

Chapter 3. Human Latexin forms amyloid-like oligomers by association of unfolded polypeptide segments

Human latexin unfolds via the population of an intermediate at low urea concentrations, being completely unfolded at 6-7 M urea. Latexin self-associates into stable oligomers in a urea dependent manner. SDS and DTT resistant self-associated latexin oligomers start to appear at 3-4M urea, and accumulate maximally at 6-7M urea, decreasing above 8M urea. At 6 M urea, the latexin oligomers exhibit maximal binding to two dyes specific for beta-sheet conformation in amyloid fibrils. SDS-PAGE and DLS analysis showed that dimeric, tetrameric and higher order latexin oligomers are populated under strongly denaturing conditions. These species appear to be the final association state of latexin and no conversion into fibrils could be observed by EM. This inability to form fibrils would allow the study of the otherwise usually elusive and always transient small molecular forms in protein polymerization processes. Latexin is thus shown to form the oligomeric species from its unfolded state. Under these conditions aggregation-prone regions, may promote and drive the oligomerization event. Detecting the region of latexin allowing to maintain intermolecular contacts in strongly denaturing conditions is therefore of interest to understand the oligomerization process. The analysis of the aggregation profile of this protein displays one main “hot spot” encompassing residues 127-156. A peptide corresponding to the 8 central residues of the “hot spot” (VLHLAWVA) forms well-ordered microcrystals that bind thioflavin-T and whose internal structure is highly similar to amyloid fibrils as assessed by FTIR.

Chapter 4. Amyloid fibril formation by a partially structured intermediate state of alpha-chymotry

α -Chymotrypsin is an all β -protein, that folds in two antiparallel β -barrel domains, each containing six β -strands with the same topology. The induction of amyloid properties in α -chymotrypsin by TFE is analyzed. The protein aggregates maximally in 35% TFE and the amyloidogenic properties of the induced structures were confirmed by congo red and thioflavin-T binding experiments. Transmission electron microscopy shows that these aggregates have a fibrillar morphology. The conformational transition induced by TFE in α -chymotrypsin occurs in two stages. In

Summary

the first stage, the hydrophobic contacts stabilizing the native β -barrel structure are disorganized, resulting in the formation of an extended β -sheet conformation. In the second stage, portions of the backbone of the protein form non-native helix conformation. Fluorescence, circular dichroism and Fourier-transformed infrared spectroscopy data suggest that formation of a partially structured intermediate state promotes the aggregation process. α -chymotrypsin at 35% TFE exists in a partially structured state(s), with extended β -sheets and loosely packed side-chains, which differs both from the native and unfolded state. The presence of solvent-exposed hydrophobic surfaces in the partially folded state is evident from its ability to bind ANS. Formation of non-native extended β -sheet conformation appears to promote aggregation whereas the acquisition of non-native helix conformation reduces this propensity, possibly due to the partial shielding of the sticky hydrophobic clusters present in the extended β -sheet conformation inside the helical structure. From our results, α -chymotrypsin aggregation depends both on the population of a partially structured state and on the backbone conformation of this intermediate.

Chapter 5 Prediction of "hot spots" of aggregation in disease-linked polypeptides

In this study the experimental data obtained in an *in vivo* system using β -amyloid peptide has been used as a model to derive the relative aggregation propensities of each individual natural amino acid when placed in the central position of the CHC of A β . Provided that a given polypeptide aggregates from an at least partially unstructured state, the experimental intrinsic aggregation propensities obtained analyzing this particular protein region will probably apply for any unfolded sequence. A profile can be theoretically generated for any protein sequence to detect those regions with aggregation propensities above the average value of the whole sequence. This leads directly to the definition of "hot spot" of aggregation as a certain region that displays higher aggregation propensity than the rest of the sequence. Based on these assumptions, a simple approach that identifies the presence of "hot spots" of aggregation in globular and unstructured disease-linked polypeptides and predicts the aggregation effects of mutations in their sequences has been developed.

The approach detects the presence of “hot spots” which have been already validated experimentally in the literature and provides insights into the effect of disease-linked mutations in these polypeptides.