Universidade do Algarve

Faculdade de Ciências e Tecnologia

STRUCTURAL CHARACTERIZATION AND COMPARATIVE ANALYSIS OF HUMAN AND PISCINE CARTILAGE ACIDIC PROTEIN (CRTAC1/CRTAC2)

Marta Lúcia Amaro Guerreiro

Dissertação de Mestrado em Biotecnologia

2014

Universidade do Algarve

Faculdade de Ciências e Tecnologia

STRUCTURAL CHARACTERIZATION AND COMPARATIVE ANALYSIS OF HUMAN AND PISCINE CARTILAGE ACIDIC PROTEIN (CRTAC1/CRTAC2)

Marta Lúcia Amaro Guerreiro

Dissertação de Mestrado em Biotecnologia

Trabalho efectuado sob orientação de :
Professora Drª Deborah Mary Power
Drª Liliana Anjos
Drª Isabel Morgado

2014

Declaração de autoria

Declaro ser a autora desta tese e todo o seu contéudo é da minha exclusiva responsabilidade. Os dados bibliográficos consultados estão devidamente citados e constam da listagem de referências.

Copyright

Acknowledgements

My special thanks go to my thesis supervisors Drª Liliana Anjos and Drª Isabel Morgado for the fondness and teaching given, for their dedication and time spent to help me throughout my academic journey and the support given in the laboratory experience acquisition.

Very thankful to Prof. Drª Deborah Mary Power for the opportunity to work and learn in their team and share their knowledge and teaching me.

I'm also thankful to all members of CCMAR especially for the knowlodge transmitted.

I wish to express my heartfelt gratitude to my family, friends and master colleagues.

Abstract

CRTAC (Cartilage Acidic Protein) firstly identified as a chondrocyte marker in humans and implicated in a number of diseases. This ancient protein is present from prokaryotes to vertebrates and the teleost are the only group that contain duplicates (CRTAC1/CRTAC2). The structure of CRTACs is poorly characterized and was the starting point of the present study.

To establish the molecular and structural characterization of CRTAC, three recombinant proteins [human (h) CRTAC1 and sea bass (*Dicentrarchus labrax,* dl) CRTAC1 and CRTAC2] were over-expressed in *E.coli* as inclusion bodies and their identity was confirmed by mass spectrometry. The resulting refolded recombinant proteins were obtained with a productivity of 11,51, 8,4 and 11,9 mg of protein per gram of biomass for hCRTAC1, dlCRTAC1 and dlCRTAC2 respectively and approximately 23,48%, 9,09% and 33,46% of, hCRTAC1, dlCRTAC1 and dlCRTAC2 were lost as insoluble aggregates. Size exclusion chromatography revealed the presence of mostly soluble aggregated forms of piscine CRTACs and a mixture of aggregates and monomeric form for hCRTAC1. Spectroscopic studies of human CRTAC1 and sea bass CRTAC1/CRTAC2 showed that all proteins possess secondary and tertiary structure and are particularly rich in β- pleated sheet structure (≈40%), have ≈10,3% alpha-helix and the remainder is disordered. The thermal stability of CRTAC´s structure was assessed considering heating (from 20-90ºC) and freezing (-80ºC). CRTACs retain their native secondary and tertiary structures upon heating, however a slight loss of structure was observed for hCRTAC1 at 60ºC. Freezing induced loss of secondary structure and conformational changes more pronounced for hCRTAC1 and third cycle of dlCRTAC2 and less perceptible for dlCRTAC1. Amyloid formation by CRTACs was assessed in Thioflavin-T assays and a decrease in fluorescence was observed after incubation. In summary, CRTAC´s have propensity to form soluble aggregates that are highly thermostable and these structural properties are conserved from teleosts to humans.

Keywords: CRTAC; β-sheet; thermostability; diseases; amyloid; structure

Resumo

A CRTAC (proteína acídica da cartilagem) é uma proteína da matriz extracelular, inicialmente identificada em humanos como marcador de condrócitos e que está também associada a várias doenças, como a esclerose múltipla. Esta proteína ancestral está presente desde os procariotas até aos vertebrados e os teleósteos são o único grupo que contém o gene duplicado (CRTAC1/CRTAC2). Ambas as proteínas apresentam estrutura similar e possuem na região N-terminal um domínio semelhante ao das cadeias α presente nas integrinas, mas diferem na região C-terminal pelo facto da CRTAC1 ter um domínio adicional descrito como um factor de crescimento epidérmico com ligação a iões Ca^{2+}. A estrutura conservada das CRTAC indica que estas proteínas podem ter uma função importante que ainda não foi identificada e o facto da sua estrutura estar pouco caracterizada foi o objecto deste estudo.

Para o estudo de caracterização molecular e estrutural das CRTAC, três proteínas recombinantes [hCRTAC1 humana e CRTAC1 e 2 de robalo (*Dicentrarchus labrax* ,dl)] foram sobrexpressas como corpos de inclusão em *E.coli* e a sua identidade foi confirmada por espectroscopia de massa. Após a solubilização, a purificação e o "refolding", a produtividade das proteínas recombinantes resultantes foi de, 11,89 para a hCRTAC1, 8,4 para a dlCRTAC1 e 11,9 para a dlCRTAC2, expressa em mg de proteína obtida por g de biomassa. Durante este processo, foram registadas perdas proteicas na forma de agregados insolúveis de aproximadamente 23,48% de hCRTAC1, 9,09% de dlCRTAC1 e 33,46% de dlCRTAC2. Análises de cromatografia de exclusão molecular, das proteínas CRTAC solubilizadas, revelaram a presença de agregados de elevado peso molecular da dlCRTAC1 e 2 de peixes e uma mistura de agregados e monómero da CRTAC1 humana. Os estudos espectroscópicos, demonstraram que, as proteínas recombinantes CRTAC têm estrutura secundária e terciária e que a sua estrutura secundária é particularmente rica em folha-β (≈40%), tendo apenas 10,3% de hélice-α e a restante estrutura é desordenada.

Neste estudo foi avaliada a estabilidade térmica da estrutura das CRTAC, considerando o efeito do aquecimento (de 20º a 90ºC) e do congelamento em vários ciclos (-80ºC). Os resultados indicaram que, as CRTAC mantêm a sua estrutura secundária e terciária após o aquecimento, no entanto, no caso da CRTAC1 humana foi observada uma ligeira perda de estrutura a partir dos 60ºC. O efeito do congelamento induziu a uma perda de

estrutura secundária e também a alterações conformacionais mais evidentes na CRTAC1 humana relativamente á dlCRTAC1 e dlCRTAC2 de peixes.

A capacidade de formação de fibrilas amilóides pelas CRTAC, sob certas condições de incubação, foi avaliada utilizando ensaios com Tioflavina-T. Após a incubação foi observada uma diminuição da fluorescência e estes resultados sugerem que, as CRTAC´s podem facilmente formar agregados de grandes dimensões com propriedades não-amilóides que, podem ser degradados ou sair de solução após longos periodos de incubação. Admite-se a hipótese de que tais agregados possam ser estruturas funcionais.

Resumindo, as proteínas CRTAC tendem a formar agregados solúveis que são altamente termoestáveis e estas propriedades estruturais são conservadas desde os teleósteos até aos humanos.

Abbreviations

CRTAC- Cartilage acidic protein

CD- Circular dichroism

dlCRTAC1/2- Sea bass (*Dicentrarchus labrax)* Cartilage Acidic Protein ½

ECM- Extracellular matrix

hCRTAC1 – human Cartilage acidic protein 1

MW-Molecular Weight

MRW- Mean residue weight

SDS-PAGE- Sodium Dodecyl Sulphate-PolyAcrylamide Gel Electrophoresis

Table of Contents

ACKNOWLEDGEMENTS 4

ABSTRACT 5

RESUMO 6

ABBREVIATIONS 8

CHAPTER I 10

GENERAL INTRODUCTION 10

1.1 Framework on Biotechnology 11
1.2 Protein structures and functions 11
1.2.1 Importance of the study of protein structures 11
1.2.2 Protein structural organization 11
1.2.3 Structures and functions relationship 12
1.3 Proteins and conformational changes 13
1.4 Methods to analyse protein structures and protein interactions 14
1.5 The Cartilage Acidic Protein (CRTAC) 15
1.5.1 Human CRTAC1 15
1.5.2 Piscine CRTAC1/CRTAC2 16
1.5.3 CRTAC1/CRTAC2 protein structures 16
1.5.4 CRTAC1/2 biochemichal/biophysical characterization and related functions 18
1.6 Objectives 19

CHAPTER II 20

STRUCTURAL CHARACTERIZATION AND COMPARATIVE ANALYSIS OF HUMAN AND PISCINE CARTILAGE ACIDID PROTEIN (CRTAC1/CRTAC2) 20

1. Introduction 22
2. Material and Methods 24
2.2. Recombinant CRTACs expression 24
2.2.2 CRTAC's purification and refolding 25
2.3 Molecular Characterization of CRTAC's 26
2.3.2 Western blot analysis 26
2.4 Structural Characterization of CRTAC'S 28
3. Results 30
3.1 CRTAC's recombinant proteins expression 30
3.2 Molecular characterization 33
3.2.2 CRTAC's immunoreactivity 33
3.3 Structural characterization 35
4. Discussion and conclusions 40

CHAPTER IV 44

FUTURE PERSPECTIVES 44

Future perspectives 45

REFERENCES 46

CHAPTER I

General Introduction

1.1 Framework on Biotechnology

The production of a recombinant protein is a biotechnological process because it results in a engineered form to produce a protein. In this process the production of heterologous protein by a host organism generate large quantities of proteins with biotechnological potencial interest.

The biological diversity of the marine environments is a powerful resource for the discovery of new bioactive compounds (proteins, drugs, biomaterials etc.) with high potential for exploitation by biotechnology. In the present study novel marine derived biomolecules (proteins), with potential biotechnological interest will be explored.

1.2 Protein structures and functions

1.2.1 Importance of the study of protein structures

Proteins are the most abundant macromolecules in biological systems and they are involved in practically all structural and functional properties of life.

Between various functions played by proteins, several can be highlight: enzymatic catalyses, hormone regulation, support, movement and protection. Proteins are constituted by a linear association of amino acids which ultimately determine the protein's organization in different structural levels: the secondary, tertiary and quaternary structure. The wide range of functions that proteins have the ability to play is a consequence of their elaborate three-dimensional structures, which allow them to convert forms of energy or generate the essential structural parts of an organism [1]. The three dimensional structure of proteins is a source of important information because it gives the various responses, such as shape and domain structure, protein classification, prediction of function for uncharacterized proteins, interaction with other macromolecules, interactions with small ligands, evidence for enzyme mechanism, structure-based drug development, post-translational modifications and experimental evidence for trans-membrane domains.

1.2.2 Protein structural organization

The structural organization of proteins is divided in 4 main levels: primary, secondary, tertiary and quaternary. The primary structure is simpler and results from a polymer of amino acid linked by a peptide bond and the protein acquire a linear conformation. The secondary structure is associated with the structural arrangements that result from the convolution process of the polypeptide chain. The tertiary structure is the three-dimensional visualization that depended of secondary structure. The Quaternary structure refers to the spatial arrangement of subunits and the type of their interactions.

The secondary structure of proteins is classified in two main structural arrangements: α-helix and β-sheets. This arrangements sometimes are accomplished by structural elements called the turn and loops.

The α-helix is a coiled structure stabilized by intra-chain hydrogen bonds established between the hydrogen of N-H on the bottom coil and oxygen of C=O on top coil of the main chain. Two or more helixes can form highly stable coiled coils commonly found in sequence motif structures, for example in fibrous proteins, globular proteins, growth hormone, myoglobin and others.

The β-sheets are stabilized by hydrogen bonding between polypeptide strands. A polypeptide chain, called a β strand, in a β-sheet is almost fully extended rather than being tightly coiled as in the α-helix. The β-sheet has formed by linking two or more β strands by hydrogen bonds. Adjacent chains in a β-sheet that run in opposite directions are called antiparallel β-sheets and in the same direction are called parallel β-sheets. The β-sheet is an important structural element in many proteins, such as, for example, fatty acid binding proteins (FABP), important for lipid metabolism. The biological membranes has formed by lipids, this are a barrier that define the inside and the outside of the cell. Our rich β-sheet structure promotes a permeability, diffusion and channel formation.

Turn and loops participate in the interactions between proteins and other molecules. (Voet, D. 2004)

1.2.3 Structures and functions relationship

Proteins are involved in several cellular functions. The elucidation of their structure and function not only provides a strong basis for the understanding of these processes, but it also allows the identification and/or improvement of therapeutic approaches to fight against human, animal or plant diseases. Functional proteins can be divided into

different groups: 1) enzymes that catalyse chemical reactions of crucial importance for living organisms, 2) proteins that regulate the biosynthesis and metabolism of RNA/DNA, 3) cellular membrane proteins, membrane-bound proteins are called 'transporters', 4) proteins involved in the signal transductions of hormones and neurotransmitters and finally 5) proteins involved in the regulation of 'life and death' of bacterial and eukaryotic cells.

It is generally known that, in proteins, helices and strands are connected to each other and combined in many different ways. The most common motifs present in proteins are: β-hairpin (two antiparallel β-strands connected by a tight turn of a few amino acids between them), greek key (4 β-strands folded over into a sandwich shape), helix-loop-helix as example the "EF-Hand" motif (α-helix bound by a looping stretch of amino acids which structure is stabilized by binding calcium ion) and zinc finger (two β-strands with an α-helix end folded over to bind a zinc ion, important in DNA binding proteins).

Protein domains consist in a conserved part of the protein sequence that can be associated with functions and exist independently of the rest of protein chain. Domains are distinct functional and/or structural units in proteins. Usually they are responsible for a particular function and interaction, contributing to the overall role of a protein [1]..In this thesis the proteins under study had 3 characteristic domains: ASPIC/UnbV, N-terminal integrin α-chain like domain and EGF-Calcium binding domain. The ASPIC/UnbV domains are a conserved sequence found associated with Tetratricopeptide (TPR) in several paralogous proteins in marine bacteria Pirellula (*Rhodopirellula baltica)* [2]. These domains were found associated with FG-GAP repeat in several eukaryotic integrin-like proteins and in several other bacterial proteins. The N-terminal integrin α-chain like domains are involved in inside-out cellular signalling mechanism[3]. The EGF-Calcium binding domains are crucial for protein-protein interactions [4].

1.3 Proteins and conformational changes

Proteins are usually flexible and dynamic and can change its shape in response to environmental changes or other factors. The form is then called conformation and the transition between two forms is called conformational changes. Conformational changes

can be induced by various factors such as variation in temperature, pH, voltage, ionic concentration, phosphorylation and binding to a ligand [5].

In the presence of factors that induce conformational changes, the protein will acquire structures other than the native one that can compromise its functionality. In certain conditions, proteins may have a tendency to misfold or the inability to sustain their folded state and this can lead to the phenomenon of protein aggregation. Protein aggregates are classified on the basis of *in vivo* and *in vitro*, and ordered and disordered. Amyloid fibrils are the example of ordered aggregates, both *in vivo* and *in vitro*. Inclusion bodies are the example of disordered aggregates *in vivo*, these are formed in the refolding of denaturant-unfolded protein or under mild conditions at native high protein concentration [6]

When producing recombinant proteins the major problem encountered is the formation of inclusion bodies. These aggregates of insoluble recombinant proteins are normally found in bacteria that are forced to produce heterologous protein [7]

1.4 Methods to analyse protein structures and protein interactions

The development of structural biology and recombinant DNA technologies made possible the production of recombinant proteins. It required the development and implementation of more accurate structural studies of the possible conditions in which proteins operate, and can then understand their structural changes and their biological function [8].

One of the most used methods to analyze the secondary structure of proteins is circular dichroism (CD). Measurement, in solution, of the far-UV CD spectra allows estimate the percentage of secondary structure elements of a protein by differential absorption of enantiomers in polarized light. The wavelengths used depend on prior knowledge of protein structure and the abundance of certain amino acids [9].

The native tertiary structure of proteins can be determined by x-ray crystallography among other methods (e.g. NMR, Cryo electron microscopy). Fluorescence emission spectroscopy can also be used to assess tertiary structure conformational changes or the formation of amyloid aggregates (when using the dye thioflavin-T).

The basis of X-ray crystallography is to determine the atomic coordinates of a crystallized protein molecule using x-ray diffraction and therefore characterize the three-dimensional structure of that protein . The biggest challenge of crystallography is

to crystallize a highly concentrated sample of purified protein as this process is dependent on multiple parameters, such as pH, temperature, protein concentration, solvent, precipitating ions and other ligands [10].

Fluorescence spectroscopy can reveal conformational changes in the tertiary structure of a protein. Proteins have intrinsic fluorescence due to the presence of aromatic amino acids (tryptophan, tyrosine, phenylalanine). Proteins can be excited at 280nm and usually have a maximum emission between 300-350 nm. A protein folded in its native state has characteristic emission spectra. Structural changes (e.g. unfolding) caused by different factors (e.g. temperature, chemicals, pH, ligands, etc) can be detected by a shift in the emission spectra, for example due to exposure of the aromatic amino acids buried in the native conformation.

Fluorescence can also be used to identify protein amyloid aggregates as the ones present in diseases like Alzheimer's disease, Parkinson's disease or type II diabetes. In this case a dye, Thioflavin-T is added to the protein sample. When bound to highly β-sheet rich structures, as the ones of amyloid aggregates, Thioflavin-T emits enhanced fluorescence and a shift of its emission maximum. [11].

1.5 The Cartilage Acidic Protein (CRTAC)

1.5.1 Human CRTAC1

The human cartilage acidic protein 1 (CRTAC1) also known as chondrocyte expressed protein 68kDa (CEP-68), was first described as a chondrocyte biomarker [12] with ability to distinguish human chondrocytes from osteoblasts and mesenchymal stem cells, in culture [13]. Recently CRTAC1 has also been described as a potential therapeutic target for detection/monitoring damages in cartilage and associated to diseases of the human cardiovascular, hematological, neurological, respiratory and urinary systems [14], although its function remains unknown. In studies of patients with multiples sclerosis was found the CRTAC-1B transcript in cerebrospinal fluid [15] and expression in several brain regions of embryonic mouse, including olfactory bulb, hippocampus, cerebral cortex, thalamus and spinal cord [16]. CRTAC1 was found associated to glomus tumors in neurofibromatosis type 1 disease [17]. Recently in studies of human keratoconus was found CRTAC1 expressed in the cornea. [18]

CRTAC1 gene encodes a glycosylated extracellular matrix protein that is found in articular cartilage. This acidic protein is composed to 661 amino acids, a N-terminal 27 amino acids long signal peptide and an EGF-Like Ca^{2+} binding motif near the C-terminal. The presence of seven phenylalanine-glycin (FG)-glycin-alanin-prolin (GAP) domains and Arginine-glycine-aspartate (RGD) integrin-like binding motifs suggests that CRTAC1 may be involved in cell-cell interactions or cell-matrix interactions [14].

The human CRTAC1 show two transcript variants (CRTAC1-A and CRTAC1-B) that result from alternative splicing. The CRTAC1-A transcript were found expressed in human brain (low levels), lung, blood, chondrocytes, and transcript CRTAC1-B in brain (high levels), eye and pineal gland. The ortologs of CRTAC1 were identified in chimpanzee house mouse, Norway rat, domestic cattle, pig, dog chicken, lizard, zebrafish, torafugu, green pufferfish and japonese medaka [19].

1.5.2 Piscine CRTAC1/CRTAC2

Analysis of sea bream (sparus auratus) pituitary gland expressed sequence tags (EST) reveal a transcript homologue to human CRTAC1 and a CRTAC isoform that lacks the C-terminal calcium binding EGF domain. This smaller gene was called CRTAC2. The two isoforms of CRTAC (CRTAC1 and 2) have emerged in teleost fish as a result of the teleost specific whole genome duplication [20].

The CRTAC1 and 2 are highly expressed in the liver and kidney followed by nervous tissues such as brain, pituitary and spinal cord in teleost fishes[19] [21]. CRTAC's form aggregates under non reducing conditions and different size distributions are found in tissues (CRTAC1 and 2 monomer with 69 kDa and 56,9 kDa respectively and a dimer of 120kDa)[21]

1.5.3 CRTAC1/CRTAC2 protein structures

CRTAC1 as CRTAC2 both share the same domains, except the EGF-like calcium binding domains, present in CRTAC1 (Fig 1). In silico analysis predict one N-terminal integrin α-chain-like region combined with ASPIC/UnbV domain suggesting the possibility of alternative functions for this region.

The ASPIC and UnbV domain is a β-strand-rich region with unknown functions that is to our knowledge exclusively found in CRTAC1/CRTAC2 protein family[19]. The N-

terminal integrin α chain-like domain is a receptor domain of integrins involved in inside-out signalling mechanisms through the cell membrane.

Fig. 1 Structural domains of CRTAC family members. SP- Signal peptide, integrin alpha chain-like- responsible for cellular signalling mechanism cell-cell/cell-matrix interactions, ASPIC/UnbV- Unknown ancestral domain, EGF- Epidermal growth factor domain.

The structure of this receptor domain is composed of a seven-bladed β-propeller and each blade is composed of four antiparallel β-sheets [19]. Homology model built for sea bream CRTAC2 (sbCRTAC2) N-terminal region also predicts a seven-bladed β-propeller structure for this region (Fig 2A). Mapping of sbCRTAC2 β-propeller structure revealed five Calcium binding β hairpin loops located in the lower face of CRTAC2 propeller assumed as essential for molecular recognition mechanisms (fig 2). These Calcium binding motifs, composed of up to 12 residues, with acid residues in position 1,3,5 and 9 that are responsible for the coordination with calcium ion (Fig 2B) [19]. The predicted secondary structures of CRTAC2 showed that 9% of the molecule is α-helix, 30% is extended strand and 60% is random coil.

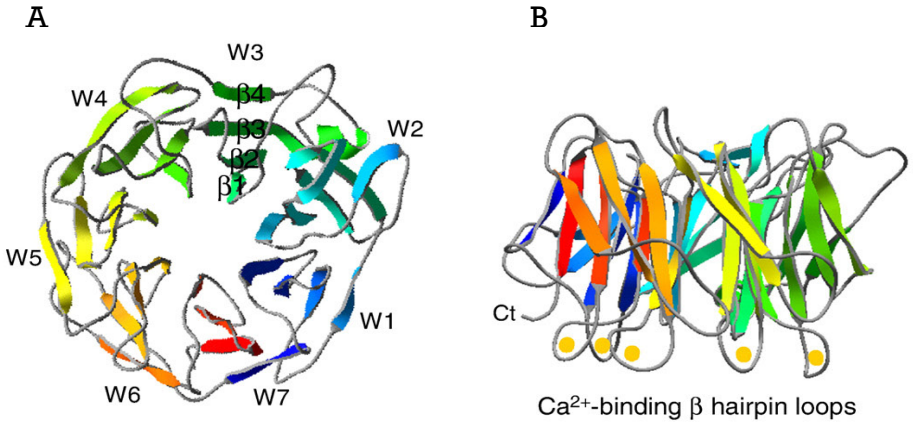

Fig 2 Structure of sea bream CRTAC2 modelling (adapted from [19]). Upper and lateral views of CRTAC2 β –propeller and in different colours the four antiparallel strands of each blade. (A) W1 to W7 represent the seven FG-GAP repeats of the N-terminal integrin alfa chain-like domain; β1 to β4 represent the β-hairpin loops when hypothetical calcium ions bind (B) Hypothetical calcium binding ions (yellow balls) located in the lower face of the propeller.

1.5.4 CRTAC1/2 biochemichal/biophysical characterization and related functions

Biochemical and biophysical studies recently performed on sea bream recombinant CRTAC2, demonstrated that there is a tendency for aggregate formation in tissues, which is majorly insoluble [21]. The formation of these aggregates and their subsequent characterization may be the key to understanding its function in the degenerative diseases implications. Studies conducted have shown the presence of CRTAC1 in patients with multiple sclerosis, being the main cause of this disease the amyloid plates formation.

Previous studies in predicted structural modelling showed the existence of five binding sites of the calcium ions, but more recent studies have shown that there are six binding sites and calcium binding does not induce secondary or tertiary structure modifications, suggesting a conformational stability [21].

Anjos et al. have demonstrated that CRTAC2 probably aggregate before the loss of the secondary and tertiary structure until temperatures of 90 °C. This ability for aggregation and your thermostability may suggest important function still unknown.

1.6 Objectives

The general aim of the present thesis is to characterize the structure, stability and aggregation properties of human CRTAC1 and piscine CRTAC1 and CRTAC2 by comparative analysis. This involves the construction of a pipeline to produce recombinant CRTAC proteins and a number of biophysical studies. Three main tasks are proposed:

1) Large scale production and purification of the 3 recombinant CRTAC proteins (hCRTAC1, dlCRTAC1 and dlCRTAC2). This task involves the application of several techniques such as protein expression, solubilisation, purification, refolding and dialysis. The recombinant proteins production will be monitored and the proteins analysed by SDS-PAGE. The identity of the CRTACs will be confirmed by MALDI-TOF and western blot.

2) Structural characterisation of native CRTACs and evaluation of their thermal stability using circular dichroism and intrinsic fluorescence emission spectroscopy. In this study it will be explored the effect of heating and freezing on the secondary and tertiary structure of the proteins.

3) Analysis of CRTAC's ability to aggregate and form amyloid structures. In this task the proteins will be incubated and the formation of aggregates will be assessed. The putative formation of amyloid-like aggregates will be tested using the dye Thioflavin-T.

The scientific contribution of the thesis was to increase the knowledge about the structure of the duplicates CRTAC1 and 2 that could be of relevance to the biotechnological and pharmaceutical industries.

CHAPTER II

STRUCTURAL CHARACTERIZATION AND COMPARATIVE ANALYSIS OF HUMAN AND PISCINE CARTILAGE ACIDID PROTEIN (CRTAC1/CRTAC2)

STRUCTURAL CHARACTERIZATION AND COMPARATIVE ANALYSIS OF HUMAN AND PISCINE CARTILAGE ACIDIC PROTEIN (CRTAC1/CRTAC2)

Marta Guerreiro, Liliana Anjos, Isabel Morgado, DM Power

Centre of Marine Science (CCMAR), University of Algarve, Campus of Gambelas, Faro, 8005-139, Portugal

Correspond to:

Running title: Structural analysis of CRTACs

Keywords: CRTAC; β-sheet; thermostability; degenerative diseases; fibrill; amyloid

1. Introduction

Cartilage acidic protein (CRTAC) is an extracellular matrix (ECM) protein of unknown function firstly identified in humans as a chondrocytes marker with ability to distinguish human chondrocytes from osteoblasts and mesenchymal cells in culture[12]. It is an evolutionary ancient protein that is found from prokaryotes to vertebrates and its conservation suggests an important, although as yet unknown, function. Two isoforms of CRTAC (CRTAC1 and 2) were found in teleost fish as a result of the teleost specific whole genome duplication [20] and are homologues of the recently identified CRTAC1 in humans (hCRTAC1). Both proteins have a similar structure and possess a conserved N-terminal seven-bladed β-propeller structure containing an integrin α chain-like Ca^{2+} -binding motif. However, CRTAC1 has acquired a C-terminal (EGF)-Calcium binding domain [19], crucial for protein–protein interactions [22] which is absent in CRTAC2. In silico analysis predict also the presence of an ASPIC/UnbV domain combined with a N-terminal integrin α chain-like domain which is involved in inside-out cellular signaling mechanisms through the cell membrane [23].The ASPIC/Unbv domain is a β- strand-rich region, exclusively found in bacterial proteins and in eukaryotic integrin-like proteins (ex: CRTAC's family) [24]. Structural prediction analysis identified seven FG-GAP amino acid repeats present in the N-terminal region of the CRTAC1/CRTAC2 family members. This region forms a conserved seven-bladed β-propeller structure containing Ca-binding β-hairpin loops and each blade of the propeller is composed of four antiparallel β-sheets [19].

In addition to the chondrogenic tissue, hCRTAC1 was also found in lung, brain, eye, pineal and parathyroid gland [14]. A CRTAC1B splice variant is the dominant transcript in brain [12] and is an antagonist of Nogo receptor-1 playing a key role in olfactory tract formation [16]. The distribution of CRTAC2 in teleosts is similar to human CRTAC1 and their broad tissue distribution suggests that it may not only function as an ECM protein [19]. Interest in this protein has been linked to its localization in articular cartilage and related diseases and it has also been associated with multiple sclerosis, bone fracture, neurofibromatosis type 1-associated glomus tumors [17]. However CRTAC's relation to pathogenesis is unclear. Recent studies showed the increase of CRTAC1 expression in patients with multiple sclerosis [15] and this disease is strongly related to formation of amyloid plates.

Previous biochemical and biophysical studies revealed that sea bream CRTAC2 is a highly thermostable β-sheet ordered protein that binds Calcium with high affinity and tends to form soluble aggregates *in vivo* [21]. Aggregate formation by sbCRTAC2 native protein was also observed in tissues like pituitary and liver [21]. In fact proteins with a β-propeller fold like CRTAC frequently play key biological roles and have been directly or indirectly associated with human diseases [25] such as human retinal degeneration, Kallmann syndrome or amyloid diseases, such as Alzheimers or atherosclerosis [26]. The formation of aggregates by CRTAC and their subsequent characterization may be the key to understanding its putative role in degenerative diseases.

Three recombinant CRTAC proteins were produced, hCRTAC1 and piscine CRTAC1 andCRTAC2, and comparative biophysical analysis were performed aiming to characterize CRTAC structure and assess its stability and propensity for fibrils formation.

2. Material and Methods

2.2. Recombinant CRTACs expression

2.2.1 Large scale production of human CRTAC1 and piscine CRTAC1/CRTAC2 recombinant proteins

Glycerol stocks of *E. Coli* transformed with the recombinant expression constructs pET11a/CRTACs had been previously prepared and were used as starting material in this project. Briefly, the piscine forms of CRTAC1 and 2 (dlCRTAC1 and 2) were isolated from cDNA of *sea bass* (*Dicentrarchus labrax,* dl) larvae, cloned into pGEMT-easy vector (Novagen) and subcloned into pET11a expression vector (Novagen). Human CRTAC1 (hCRTAC1) is described in the literature and was purchased as an ORFEXPRESS™ Gateway PLUS shuttle clone from GeneCopoeia and subcloned into pET11a expression vector. Vector inserts were analysed by PCR followed by restriction digestion and sequencing to confirm the authenticity and the correct reading frame of the inserts. Competent *E. Coli* (BL21 (DE3) and Origami^{TM}2 DE3, Novagen) were transfected with the constructs (pET11a/hCRTAC1; pET11a/dlCRTAC1; pET11a/dlCRTAC2) and tested for high expression level of the mature CRTACs forms. Single clones of BL21 (DE3)- hCRTAC1 and Origami^{TM}2(DE3)- dlCRTAC1 and dlCRTAC2 were independently inoculated in LB media containing Ampicillin (100 µg/ml) and grown overnight at 37 ºC and 250 rpm. This culture was further used to inoculate large scale cultures (2 or 1,5 L of LB enriched with 35 mM K_2HPO_4, 4 mM KH_2PO^4, 4 mM glucose and Ampicillin (100 µg/ml)) that were grown at 37 ºC and 250 rpm until they attained an OD_{600}nm=0.6-0.8. Expression was induced by adding 0.8 mM IPTG and was allowed to proceed for 6 h at 30 ºC, in all CRTAC'S and additional optimization for 12h at 30ºC was carried out for hCRTAC1. Bacterial pellets were harvested by centrifugation for 10 min at 7500 rpm, 4°C, and lysed in lysis buffer [200 mM Tris–HCl pH 8, 500 mM NaCl, 0.1 mg/ml lysozyme (Sigma, UK), 0.025 mg/ml DNase I (Sigma, USA), 4.2 mM MgCl2, 1 mM phenylmethanesulfonyl fluoride (Sigma, US)] at a ratio of 6 ml of buffer per gram of *E. coli.* The mixture was incubated for1 h at room temperature with agitation followed by 3 freeze/thaw cycles and sonication (2 cycles at 30 sec and 1 cycle at 1 min) to ensure complete lysis. After centrifugation for 10 min. at 15000 rpm, 4 ºC, the insoluble protein fraction was solubilized in denaturing buffer (8 M Urea, 50 mM Tris pH 8, 100 mM NaCl, 10 mM EDTA) at a ratio of 3 ml of buffer per gram of pellet for 16 h at 4 ºC with gentle stirring. Soluble and insoluble

protein fractions were analyzed by SDS-PAGE using a 10% polyacrylamide gel to confirm expression.

2.2.2 CRTAC's purification and refolding

Purification and analytical separation of the recombinant dlCRTAC1/ dlCRTAC2 and hCRTAC1 was carried out by preparative SDS-PAGE with continuous elution electrophoresis using a Model 491 Prep Cell (Bio-Rad, Portugal). In this method, the sample moves through a cylindrical gel where proteins separate into ring-shaped bands. Individual proteins bands migrate according to size to the bottom of the gel into a collection chamber from where they are pumped into an external fraction collector. The resolving gel (75x37mm) was composed of 10% polyacrylamide in 0.375 M Tris-HCl, pH 8.8 and the polymerization occurred into a cylindrical cast for 4 hours. The stacking gel (25x37mm) was of 4 % polyacrylamide in 0,125 M Tris-HCl, pH 6.8 and was polymerized for 2 hours. The resolving gel after polymerization had approximately between 7.5-8.4 cm and stacking gel had 1.2-3.5 cm. Protein sample was mixed (6 or 8ml) with 2x sample buffer (0.0625 M Tris–HCl pH 6.8; 0.025% bromophenol blue; 10% glycerol, 200 mM DTT) respectively and heated at 100 ºC for 5 min. Electrophoresis was carried out using 0,025 M Tris-HCl; 0,19 M glycine; 0,1% SDS pH 8.3 as running buffer and run at a constant power of 12W/80 mA approximately 8-10h, maintaining a constant temperature (18ºC) until the ion front (bromophenol blue) has migrated out the gel. Protein fractionated was collected into the elution chamber and collected in 2,5ml/fraction at 8ºC in 1x Elution buffer (0,192 M Glycine; 25 mM Tris-base). The eluted recombinant proteins were detected by analytical SDS-PAGE gel 10% and coomassie blue or silver staining (Biorad Silver Kit, Portugal). Pure fractions were pooled and concentrated >10x using Ultrafree-15 centrifugal filter devices with a molecular cut-off of 5 kDa (Millipore, Bedford, UK). The refolding process of the three CRTAC proteins was achieved by successive dialysis against different buffers (1^{st} - 0.5 M Urea, 100 mM Glycine, 100 mM Tris-HClpH 8, 0.4 M L-arginine, 0.5 mM Glutathione oxidized, 5 mM Glutathione reduced; 2^{nd} - 100 mM Glycine, 100 mM Tris-HCl pH 8, 0.2 M L-arginine, 0.5 mM Glutathione oxidized, 5 mM Glutathione reduced; 3^{rd} - 100 mM Tris-HCl pH 8, 100 mM Glycine, 0.1 M L-arginine, 0.25 mM Glutathione oxidized, 2.5 mM Glutathione reduced; 4^{th} - 100 mM Tris-HCl pH 8,100 mM Glycine , 5^{th} - 100 mM Tris-HCl pH 8, 250 mM NaCl). Each dialysis was performed at 4 ºC

during 16 h with gentle stirring and to eliminate protein precipitates the samples were centrifuged for 10 min at 12000 rpm, 4ºC in the end of the dialysis process. The final concentration of each protein was determined using the method described by Gill and Hippel [27] based on UV absorption at 280 nm and on the molar extinction coefficient estimated according to the following equation $\varepsilon M_{,Gdn.HCl} = a\ \varepsilon_{M,,Tyr} + b\ \varepsilon_{M,,Trp} + c\ \varepsilon_{M,,Cys.}$
The process of expression, purification and refolding of the three CRTACs was repeated several times whenever necessary to carry out the biophysical experiments experiments and the yield of the protein production process was monitored in each step of the process (from purification until the final product).

2.3 Molecular Characterization of CRTAC's

2.3.1 MALDI-TOF analysis

Purified recombinant CRTAC´s were resolved on SDS-PAGE (10% polyacrylamide gel) and the bands were excised and placed in MilliQ water for protein identification analysis conducted by Centro de Genómica y Proteómica (Unidad de Proteómica, Faculdad de Farmacia, UCM, Spain). An additional protein band observed during dlCRTAC2 expression but with a small molecular weight than the expected for CRTAC2 was also identified. The CRTAC's proteins were subjected in-gel to trypsin digestion and the molecular weights of the peptides and amino acid sequences were analyzed by MALDI-TOF (Matrix Assisted Laser Desorption Ionization (MALDI) tandem Time-of-Flight (TOF) mass spectrophotometer). The peptides fragmentation were identified by MS/MS and peptide mass fingerprinting. The results were used for database searches using MASCOT software (http://www.matrixscience.com).

2.3.2 Western blot analysis

Western blotting, also known as immunoblotting or protein blotting is used to detect the presence of a specific protein in a complex mixture. The western blotting procedure relies upon three key steps: the separation of the mixture by size using gel electrophoresis, the efficient transference of separated proteins to a solid support and the specific detection of the target protein by appropriately matched antibodies.
The identity and authenticity of CRTAC´s proteins was confirmed by western blot using the following specific polyclonal antibodies: commercial goat anti-hCRTAC1

(SantaCruz Biotechnology, USA); rabbit anti-dlCRTAC2 serum (produced against a specific dlCRTAC2 peptide); rabbit anti-sbCRTAC2 serum (produced against the sbCRTAC2 recombinant protein), being the last two available in house. All the antibodies in use were previously validated and the titres optimized before the beginning of this study.

The purified CRTAC proteins (1µg per well in duplicates) were separated by SDS-PAGE (10% polyacrylamide gel) and transferred onto a nitrocellulose membrane 0.2 µm (Biorad, Portugal) for 2:15 h at 300 mA in a vertical tank transfer system (TE 22 Mighthy Small Tank Tranfer, Hoeffer, Amersham GE HealthCare. USA). Nitrocellulose membrane was stained with Ponceau S dye (Sigma Aldrich, USA) for reversible detection of protein bands on membrane. After protein bands detection the membrane was blocked with 3% BSA, overnight at 4 ºC for dlCRTAC1 and dlCRTAC2 and 1h for hCRTAC1 at room temperature. CRTAC proteins immobilized on membranes were firstly incubated with primary antibodies (goat anti-hCRTAC1 1/2,500 for hCRTAC1 and rabbit, anti-sbCRTAC2 (1/300,000) or anti-dlCRTAC2 (1/100,000) for dlCRTAC1 and dlCRTAC2 respectively) during 1 h at room temperature or overnight at 4 ºC in the case of hCRTAC1. As reaction controls, membrane duplicates were incubated with rabbit pre-immune sera using the same dilutions, except for hCRTAC1 in which the control was established using pre-absorbed (with 10,4 µM of purified hCRTAC1) goat-anti hCRTAC1 1/2,500. Detection was carried out by incubating the membranes during 1 h at room temperature with the secondary antibodies (1/35,000, biotinylated anti-rabbit IgG, Amersham GE Healthcare, UK or 1/80,000 HRP-anti-goat IgG, (Amersham GE Healthcare, UK) followed by streptavidin–horseradish peroxidase conjugate (1/50,000, Amersham GE Healthcare, UK) for the biotinylated secondary antibody. Antibody working dilutions were prepared in 1XTBS (50 mM Tris, 150 mM NaCl, pH 7.6). Between each incubation step membranes were washed for 3 times during 10 min in 1xTBST. Membranes were developed by chemiluminescence using ECL™ Prime Western Blotting Detection kit (Amersham GE Healthcare, UK) following the manufacturer´s instructions. Images were captured by automatic exposure using the ImageQuant LAS 500 (GE healthcare, UK) system.

2.3.4 Size exclusion Chromatography

Size exclusion chromatography was used to evaluate the formation of aggregates by the recombinant CRTAC proteins after purification, concentration and refolding. Separation was carried in a Superdex 200 increase 10/300 GL column (GE healthcare, UK) connected to an Äkta Purifier (GE healthcare, UK) following the manufacturer's instruction and guidelines. The column was pre-equilibrated with 5 column volumes (cv) of water followed by 5 cycles of chromatography buffer (100 mM Tris-HCl pH8, 250 mM NaCl buffer). Pure CRTACs were dialyzed against the chromatography buffer and concentrated to 33 µM, 61 µM and 77 µM for hCRTAC1, dlCRTAC1 and dlCRTAC2, respectively. A sample of 1-2 ml size was injected into the column using a 2 ml loop. The elution of protein peaks was monitored by UV280 nm using the UNICORN software (GE healthcare, UK) and the protein factions recovered using a fraction collector.

2.4 Structural Characterization of CRTAC'S

2.4.1 Circular Dichroism (CD) spectroscopy

CD was used to analyze and compare the secondary structure and stability of the human as piscine CRTACs. Pure CRTACs were dialyzed against water and used at a final concentration of 5 µM. Measurements were performed in a Jasco J-810 CD Spectropolarimeter using a 1 mm path lenght quartz cell (Hellma). Far-UV spectra was recorded between 180-250 nm. Thermal scans were performed to evaluate and compare CRTAC's thermal stability in water between 20 ºC and 90 ºC at 210 nm using a peltier unit. The temperature measurements were obtained at rate 1ºC/min. The effect of low temperature (freezing) was also analyzed by measuring a scan between 180-250 nm before and after freezing each CRTAC at -80ºC. and measurements normalized to maximum emission wavelength. A buffer/water baseline was measured accordingly and subtracted from all spectra. A minimum of 3 accumulations were obtained and averaged for each measurement.

Results were expressed in mean residue weight ellipticity ([θ]mrw) in deg.cm^2/dmol using the equation:$[\theta]mrw,\lambda = MRW(\theta\lambda)/10lc$ in which MRW is the mean residue weight (molecular weight/$N-1$ where N is the number of amino acids), θ is the observed

ellipticity (in degrees), *l* is the path length (in cm) and *c* the protein concentration (in g/ml).

2.4.2 Fluorescence Emission Spectroscopy

Fluorescence spectroscopy was used to analyse the tertiary structure and thermal stability of the CRTACs. Measurements were performed in a Fluoromax3 Horiba Scientific using a rectangular quartz cell with a 10 mm path length and excitation at 280 nm. Emission spectrum was recorded between 290 and 420 nm using a 1 nm slit width. Protein samples were prepared in water to a final concentration of 1μM. Thermal stability was assessed qualitatively by taking scans at 20ºC and after heating the samples to 90ºC. Similarly scans were also taken before and after freezing the samples at -80ºC. Scans were analyzed with the program Origin and normalized to the maximum emission wavelength for comparison.

2.4.3 Fibrill formation analysis

Thioflavin T (ThT) assay was used to evaluate amyloid fiber formation by CRTACs [11]. Human and piscine CRTAC protein samples were incubated at 37ºC in Tris buffer 100 mM, 500 mM NaCl, 1mM CaCl2 at pH8 for 20 days to allow potential fibril formation. Samples was taken and mixed with ThT to a final concentration of 33 uM, 131 uM and 153 uM hCRTAC1, dlCRTAC1 and dlCRTAC2 protein, respectively and 20 μM ThT. Fluorescence was measured in using a 96 multiwell plate in a Fluorescence plate reader (BIO synergy plate reader, BioTek). Excitation and emission were at 450 and 490 nm respectively. ThT fluorescence was measured at incubation time T=0 for control.

3. Results

3.1 CRTAC's recombinant proteins expression

The expressed CRTAC's recombinant proteins were predominantly insoluble. The SDS-PAGE analysis of the large-scale production of recombinant CRTAC´s proteins revealed that all CRTAC´s were overexpressed in the insoluble cellular fraction as inclusion bodies (Fig 1A, B). Inducer concentration and the post-induction growth temperature or induction time were optimized in small-scale production however the insolubility of the CRTAC´s could not be reverted (results previously obtained by Anjos L. and not shown). Theoretical values for the isoelectric point (pI) and molecular weight (MW) of the mature CRTAC´s forms preceded by a Met (methionine) were estimated by ProtParam Expasy software (http://web.expasy.org/protparam/) and were of pI= 4,97 and MW= 68,56 kDa for hCRTAC1 and of pI=5,2 and 4,9 and MW= 68,02 and 56,87 kDa for dlCRTAC1 and 2 respectively. All CRTAC´s proteins were successfully produced with the expected molecular weight as predicted and strong bands were observed for hCRTAC1 (Fig 1A), dlCRTAC1 and dlCRTAC2 (Fig 1B) in the insoluble bacterial extract fraction after expression induction with 0.8 Mm IPTG. hCRTAC1 was expressed in *E.coli* BL21(DE3) strain (no results were obtained for *E.coli* Origami2(DE3)) and two induction times (6 and 12 h) with 0,8 mM IPTG were tested in the large-scale production. The results demonstrate an increase in hCRTAC1 protein expression with the induction time duplication (Fig 1A). dlCRTAC1 and 2 were expressed in both strains (*E.coli* BL21(DE3) and *E.coli* Origami2(DE3)), however the expression was higher in *E.coli* Origami2(DE3) and the best induction time was of 6 h after small scale optimization (results previously obtained by Anjos L. and not shown).

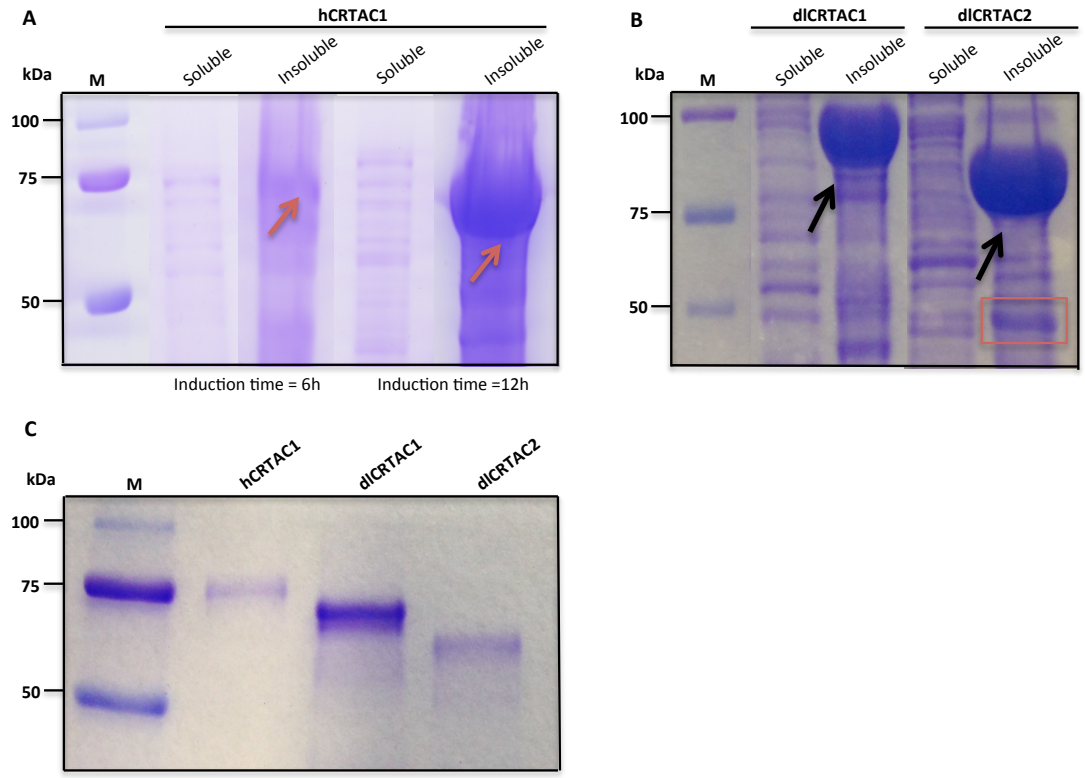

Fig. 1. Scale-up production analysis of the recombinant CRTAC´s by Coomassie blue stained SDS-PAGE (10%). A) Bacterial lysates from *E.coli* BL21(DE3) transformed with pET11a/hCRTAC1 construct, note the hCRTAC1 expression at different induction times (6 and 12 h) with 0,8 mM IPTG present in the insoluble bacterial fraction (red arrow). B) Bacterial lysates from *E.coli* Origami2(DE3) transformed with pET11a/dlCRTAC1 and pET11a/dlCRTAC2 constructs, note the dlCRTAC1 and dlCRTAC2 expressed after 6 h of induction with 0,8mM IPTG and presents in the insoluble bacterial fraction (black arrows). C) Purified CRTAC´s proteins. Orange square- small molecular weight band identified as dlCRTAC2 by MALDI-TOF and described in point 3.2.1; M - molecular weight markers (kDa).

The production line of the recombinant CRTAC´s proteins was preliminary monitored (Table 1) by protein quantification in each step of the process (purification/ concentration/ refolding) through the UV–vis absorption at 280 nm and molar extinction coefficient (Ɛ) [27], that was estimated to be of 52600 M^{-1} cm^{-1} for hCRTAC1, 58050 M^{-1} cm^{-1} for dlCRTAC1 and 38270 M^{-1} cm^{-1} dlCRTAC2.

The yield of the process from protein purification until refolding was estimated and piscine CRTAC1 and human CRTAC1 (12 h of induction) were obtained with higher yield (90.91 % and 76.52 %) and the protein loss was observed only during protein concentration step. Significant loss was observed for dlCRTAC2 (≈ 36.46%) mainly during protein concentration step after purification and this loss even with less impact extends up to the refolding process unlike which was observed for CRTAC1 forms. In this case and for all CRTAC1 forms some protein was recovered during the refolding process. Protein aggregates and precipitates were visually observed during concentration/refolding process and its appearance resembles floating wires.

Table1. Preliminary monitoring of the recombinant CRTACs proteins production process. Each step of the process from purification, concentration until refolding was monitored by estimation of the total protein concentration (mg/ml) and the yield, loss and productivity of process are also presented.

Process steps/Parameters	hCRTAC1 (6h)	hCRTAC1 (12h)	dlCRTAC1 (6h)	dlCRTAC2 (6h)
Culture volume (L)	1,5	2	2	2
Biomass (g)	5,23	13,37	9	10
Inclusion bodys solubilization buffer (ml) *	12,55	40,11	21,6	24
Sample volume load in PREPCELL for purification (ml)	6	8	6	3
Protein after Prep Cell purification [mg/ml] (Total volume/Total concentration)	0,084 (125ml/10,5mg)	0,23 (175ml/40,25mg)	0,063 (368ml/23,1mg)	0,110 (212ml/23,45mg)
Protein after Ultrafree-15 filter concentration [mg/ml] (Total volume/Total concentration)	1,68 (7ml/8,3mg)	2,41 (9,7ml/23,37mg)	1 (20ml/20mg)	0,84 (20ml/16,8mg)
Protein after Refolding [mg/ml] (Total volume/Total concentration)	1,08 (9ml/9,8mg)	2,46 (12,5ml/30,8mg)	0,7 (30ml/21mg)	0,88 (17ml/14,9mg)
Total yield (%) - from purification until refolding ❖	93,33%	76,52%	90,91%	63,54%
Total loss (%) - from purification until refolding ❖	6,67%	23,48%	9,09%	36,46%
Productivity (mg of purified protein/g of biomass)	3,92	11,55	8,4	11,9

❖ *Total yield = (Total protein concentration after refolding/ Total protein concentration after Prep Cell purification) x 100%.*
❖*Total loss was estimated by calculating the difference between 100% and total yield (%).*
Productivity was estimated according to the following equation ((volume of Inclusion Bodies solubilization buffer mass of refolded protein)/sample volume. load in Prep Cell)/Total biomass)*
All values in the table are considered preliminary since they refer to a single production process for each protein and not include replicates of the process.

The final productivity of purified, refolded solubilized recombinant hCRTAC1 (12h), dlCRTAC1 and dlCRTAC2 was respectively, 11,55, 8,4 and 11,9 mg of protein per gram of bacteria. Highest productivity was observed for hCRTAC1 (12h) and dlCRTAC2, even though the latter have gone significant losses. Remarkable was the hCRTAC1 productivity difference (from 3,92 to 11,55 mg of protein/g of biomass) obtained with the induction time duplication (from 6 to 12 h) and this results are in agreement with which was observed in gel in Fig 1A. The results obtained suggest that different *E. Coli* strains could influence CRTAC´s expression and more induction time (12h) will be needed for hCRTAC1 expression in *E.coli* BL21(DE3), comparatively to dlCRTAC1 and dlCRTAC2 expression (6h) in *E.coli* Origami2(DE3) strain.

All the recombinant CRTAC´s were obtained with a good purity as observed under denaturing conditions in Fig 1C.

3.2 Molecular characterization

3.2.1 Identification of recombinant CRTAC´s

The identity of the purified recombinant CRTAC´s was confirmed by MALDI-TOF analysis with a significant score and $p<0.05$. The MS fingerprint spectra of hCRTAC1, dlCRTAC1 and 2 were composed of 21, 20 and 24 tryptic peptides. MASCOT blast search revealed the amino acid sequence for hCRTAC1, dlCRTAC1 and 2. hCRTAC1 and dlCRTAC1 had the best match with the sequence predicted for Cartilage Acidic protein 1 isoform X1 (*Homo sapiens*, gi:530393911) and Cartilage Acidic Protein 1 (*Tetraodon nigroviridis*, gi: 85838736) respectively. The sequence coverage, the nominal mass (Mr) and pI was of 39%, 68,6 kDa and 5.05 for hCRTAC1 and of 29%, 69,6 kDa and 5.42 for dlCRTAC1. dlCRTAC2 was not possible identify by MASCOT blast search in public databases and their identity was confirmed via MASCOT search after tryptic digestion but against its own amino acid sequence (translated from the amplified full lenght cDNA of *Dicentrarchus labrax*) provided by us. Both dlCRTAC2 protein bands, the largest and the smallest indicated in Fig 1B were analyzed by MALDI-TOF and the sequence coverage was of 60% for the largest form and 30% for the smallest form, the nominal mass (M_r) and pI was of 59.2 and 5.03. MALDI-TOF analysis results are shown in appendix 1.

3.2.2 CRTAC's immunoreactivity

The immunoreactivity of the purified/refolded recombinant CRTAC's were confirmed by Western blot analysis using specific CRTAC´s polyclonal antibodies (Fig 2). Single immunoractive bands were observed for hCRTAC1, dlCRTAC1 and dlCRTAC2 (Fig 3) with the expected molecular weights of 68,6 kDa for hCRTAC1, 69,6 kDa for dlCRTAC1 and 56,7 kDa for dlCRTAC2, under denaturing conditions and no protein degradation was observed. No reaction in Western blot was obtained when anti-CRTAC´s antibody or antiserum was substituted by pre-absorbed or pre-immune serum (Fig 3). Results of antibodies validation assays previously carried out demonstrate that human anti-CRTAC1 (not published) and sea bream anti-sbCRTAC2[21] cross react with piscine CRTAC 1 and 2. The antiserum generated against sea bass CRTAC2 specific peptide show to be specific for CRTAC2 only (not published). The identity of the

CRTAC´s was confirmed and despite the cross-reacting antibodies, discrimination between proteins was also possible due to the MW differences between them.

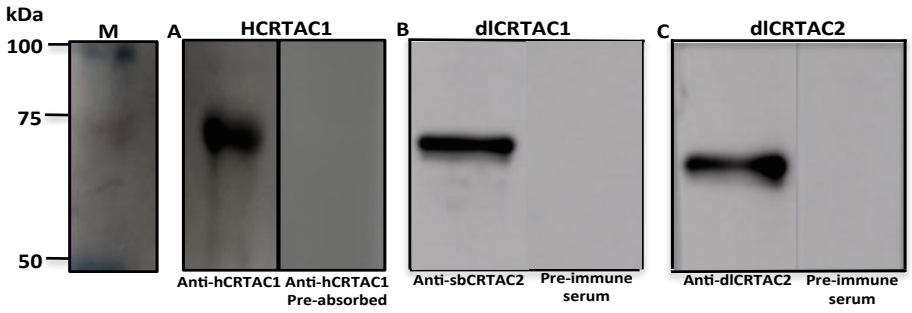

Fig. 3. CRTAC´s immunoreactivity analysis by western blot using polyconal antibodies (anti-hCRTAC1/anti-sbCRTAC2/anti-dlCRTAC2). Pure CRTAC's proteins (1 µg) was fractionated by SDS-PAGE (10%), blotted onto nitrocellulose membrane and imunostained with: (A)-goat anti-hCRTAC1 or anti-hCRTAC1 pre-absorbed (1/2,500); (B)-rabbit anti-sbCRTAC2 serum or pre-immune serum (1/300,000); (C) rabbit anti-dlCRTAC2 serum or pre-immune serum (1/100,000). The secondary antibodys were anti-goat HRP (1/80,000) or biotinylated anti-rabbit (1/35,000) and streptavidin–HRP (1/50,000). M — molecular weight markers (kDa).

3.2.3 CRTAC monomers and aggregates

Analysis and separation of different molecular weight species formed by the human and piscine CRTAC´s were carried out by size exclusion chromatography. hCRTAC1, dlCRTAC1 and dlCRTAC2 samples at 33uM, 61uM and 77uM respectively were separated in a semi-preparative gel filtration column to obtain final homogenous samples in a high concentration suitable for protein crystallography trials

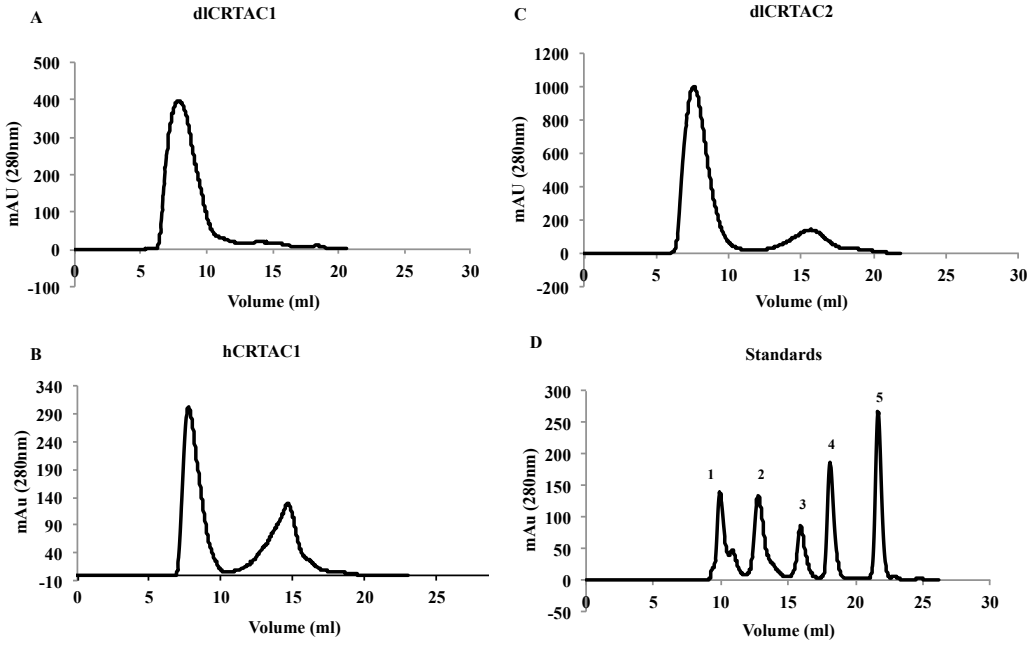

Fig. 3. Size exclusion chromatographs of dlCRTAC1 (A), dlCRTAC2 (B) and hCRTAC1 (C) and standard chromatograph with proteins of known molecular weight used for qualitative molecular weight estimation. Peaks correspond to retention volumes of each protein species: dlCRTAC1 (7.9 and 14.7 ml), dlCRTAC2 (7.6 and 15.2 ml) and hCRTAC1 (7.75 and 14.6 ml) or standard proteins (D) as indicated (1- Thyroglobulin, 2- gama-globulin, 3- Ovalbulmin, 4- Myoglobin, 5- Vitamin B12)

A chromatograph with proteins of known molecular weight was obtained in order to estimate the molecular weight of the CRTAC peaks by comparison. Large peaks with lower retention volumes (approximately 7.5 ml) times obtained for dlCRTAC1, dlCRTAC2 and hCRTAC1 (Fig 3) correspond to large molecular weights above 670 kDa as suggested by the standard graph (Fig 3D). Smaller peaks with higher retention volumes (approximately 15ml) were obtained for each protein and likely correspond to the monomeric form of each CRTAC. Peaks were separately collected and further concentrated to use in crystallization trials (future studies). Peaks corresponding to dlCRTAC1 and dlCRTAC2 monomer were very small suggesting that the protein is mostly aggregated and less stable or less soluble than hCRTAC1.

3.3 Structural characterization

3.3.1 Freezing and heating does not induce conformational changes

Pure CRTAC proteins dialysed against water were analyzed by Far-UV CD spectra and found to to possess secondary structure. Percentage of secondary structure elements was estimated using a algorithm CONTIN. The hCRTAC1 showed a negative peak (-1,0853x10^3 deg cm^2/dmol) at 211,2 nm and 10,3% α-helix, 38% β-sheet and 51,6% disordered. dlCRTAC1 showed a negative peak (-9,5017x10^3)at 211nm and 10,3% α-helix, 38% β-sheet and 51,6% disordered. dlCRTAC2 showed a negative peak (-1,00619x10^3) at 207,6 nm and 10,4% α-helix, 38,1% β-sheet and 51,5% was disordered. All CRTACs showed a high β-sheet content and CD spectra suggest high similarity (Fig 4).

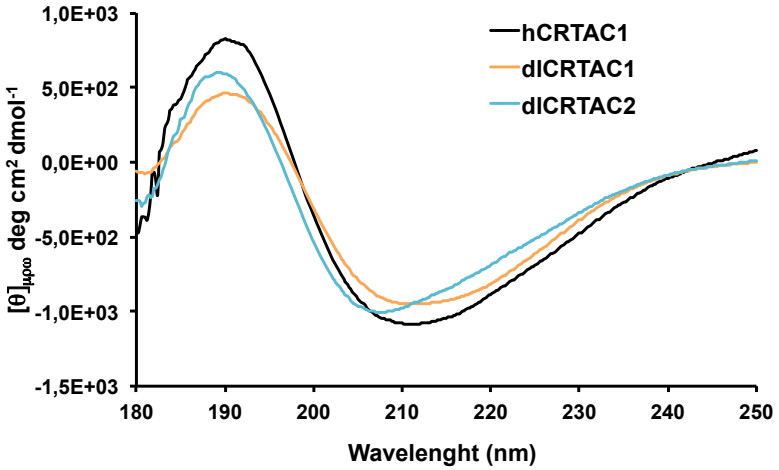

Fig 4. Far-UV spectra of purified recombinant CRTAC'S (5uM) in water at room temperature (25ºC), scanned from 250 to 180nm.

The thermostability of CRTACs was analyzed by following changes in the spectra minimum ellepticity (210 nm) with increasing temperature from 20ºC to 90ºC (Fig. 5). No sign of thermal unfolding (thermal state transition) was observed for all the CRTACs suggesting a high thermostability There was a slight loss of secondary structure after 60ºC which was much more pronounced for hCRTAC1 (to 39.6% from at 48,4% at 20ºC) suggesting that human CRTAC might be less thermostable than the piscine CRTAC form. CRTAC thermostability had been already described for the sea bream CRTAC2[21].

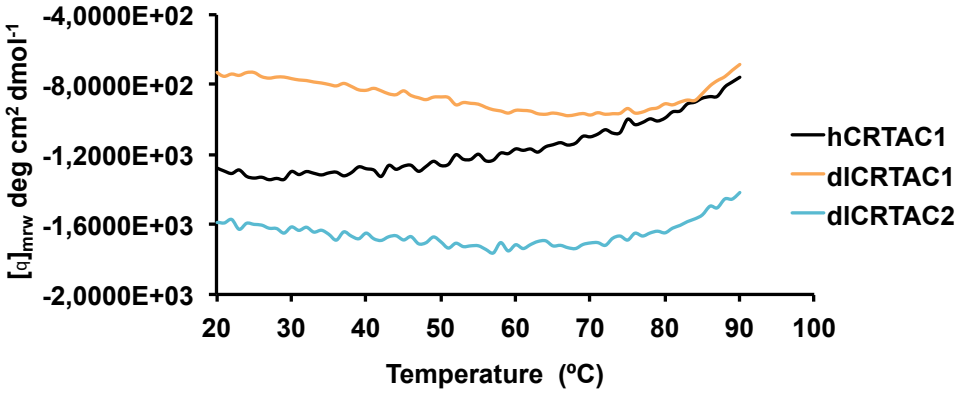

Fig. 5. Thermal scan at 210nm of purified recombinant CRTAC's in water, from 20ºC at 90ºC with 1ºC temperature measurement step.

In order to analyze the effect of low temperatures and freezing CRTAC's proteins were subjected to 4 cycles of freezing at -80ºC. Far-UV CD spectra after one freezing cycle

revealed a loss in secondary structure more pronounced for hCRTAC1 and only in third freeze of dlCRTAC2. After the third freezing cycle a shift in the minimum to lower wavelength is observed only in hCRTAC1 suggesting an decrease in ordered structure (Fig. 6).

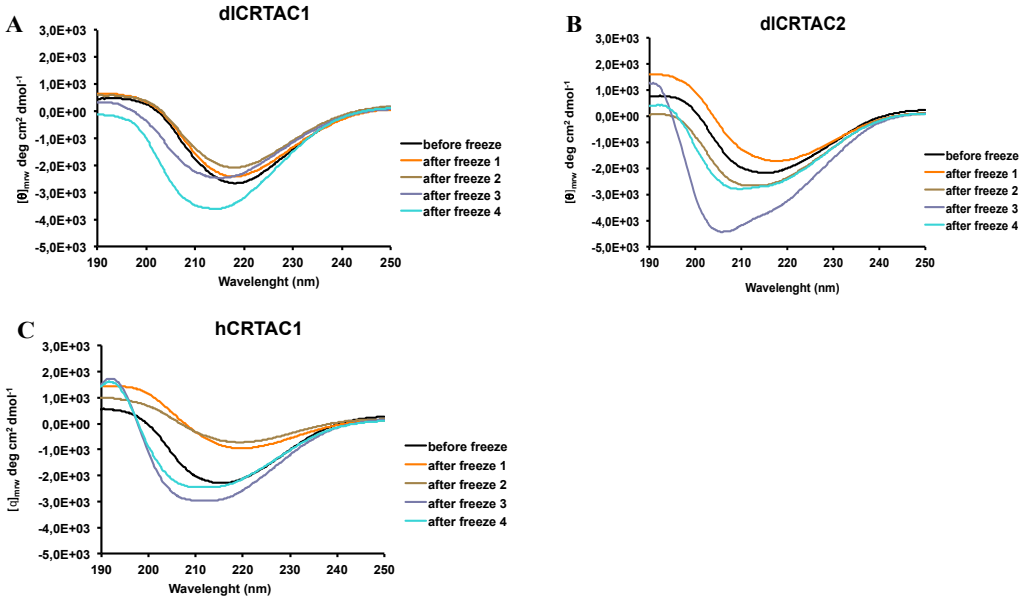

Fig. 6. Far-UV CD spectra of purified recombinant CRTAC's proteins in water, after 4 freezing cycles. A- dlCRTAC1 meausements of 4 cycles of freezing at -80ºC, B-dlCRTAC2 meausements of 4 cycles of freezing at -80ºC, C- hCRTAC1 meausements of 4 cycles of freezing at -80ºC.

3.3.2 Thermoestability of CRTAC's

Intrinsic fluorescence emission was used to analyze the tertiary structure and thermal stability of recombinant CRTAC's (1 µM). Proteins contain three amino acid residues, tryptophan, tyrosine and phenylalanine, which may contribute to their intrinsic fluorescence. CRTAC's proteins were excited at 280 nm to measure tryptophan emission spectrum between 290 and 420 nm using a 1 nm slit width. Proteins were solubilized in water in both experimental procedures. All CRTACs show similar emission spectra with maximum emission at 338, 339 and 338 for hCRTAC1, dlCRTAC1 and dlCRTAC2 respectively. After heating to 90 ºC there was no appreciable shift in the emission maximum suggesting that there was no tryptophan exposure due to unfolding (Fig 7). These results are in agreement with the CD measurements showing that CRTAC´s are thermostable and retain tertiary structure up to 90ºC. These results could be justified by your characteristic of insolubility and

ulterior aggregations. Probably both CRTAC's proteins aggregate before disintegrate the secondary and tertiary structure, as was already described [21]. One could hypothesize that CRTAC´s aggregate forming thermal resistant multimeric structures that retain 3 dimensional structure.

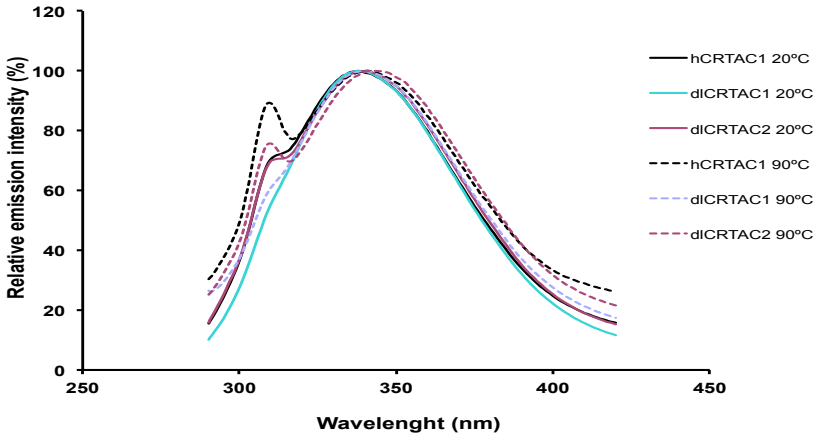

Fig. 7. Fluorescence emission spectra (λexc=280nm) of native CRTAC's (1uM) in water at 20 ºC (full line) and at 90 ºC (dotted line).

CRTAC's proteins were submitted to 4 freezing cycles. Spectra obtained do not show an appreciable deviation of emission maximum indicative of unfolding. The results obtained were consistent with those obtained in the CD measurements suggesting that all proteins maintain their native structure. However for hCRTAC1 a shift in emission maximum was observed after the first freezing cycle suggesting lower stability (Fig 8).

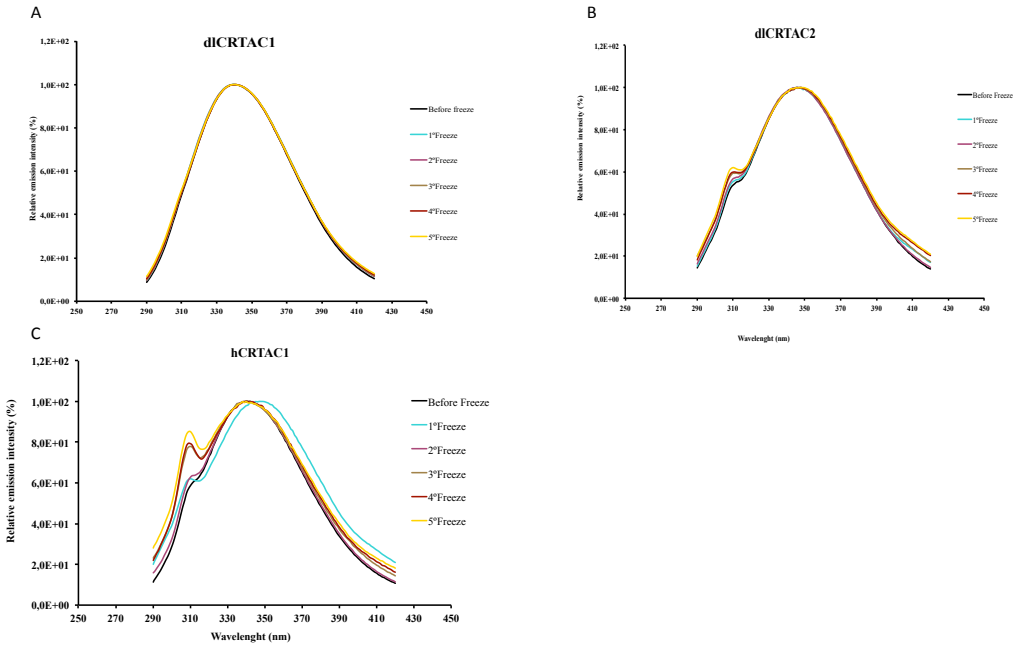

Fig. 8. Fluorescence emission spectra (λexc=280nm) of CRTAC proteins after freezing at -80 ºC. A- dlCRTAC1 fluorescence emission measurements of 5 cycles of freezing at -80 ºC, B- dlCRTAC2 fluorescence emission measurements of 5 cycles of freezing at -80 ºC, C- hCRTAC1 fluorescence emission measurements of 5 cycles of freezing at -80 ºC.

3.3.3 Aggregation Assay

Thioflavin-T assay was used to investigate putative formation of amyloid aggregates by CRTACs. When bound to highly beta-sheet rich structures, such as amyloid aggregates, ThT emits enhanced florescence 482 nm. CRTAC's proteins were incubated for 20 days at 37ºC and their fluorescence measured in the presence of ThT. Results showed a decrease of fluorescence emission relatively to CRTAC's that were not incubated (Fig. 9). These results may be due to protein degradation or formation of non-amyloid aggregates. The fact that CRTACs are β-sheet rich structures accounts for the hight ThT fluorescence in the native state.

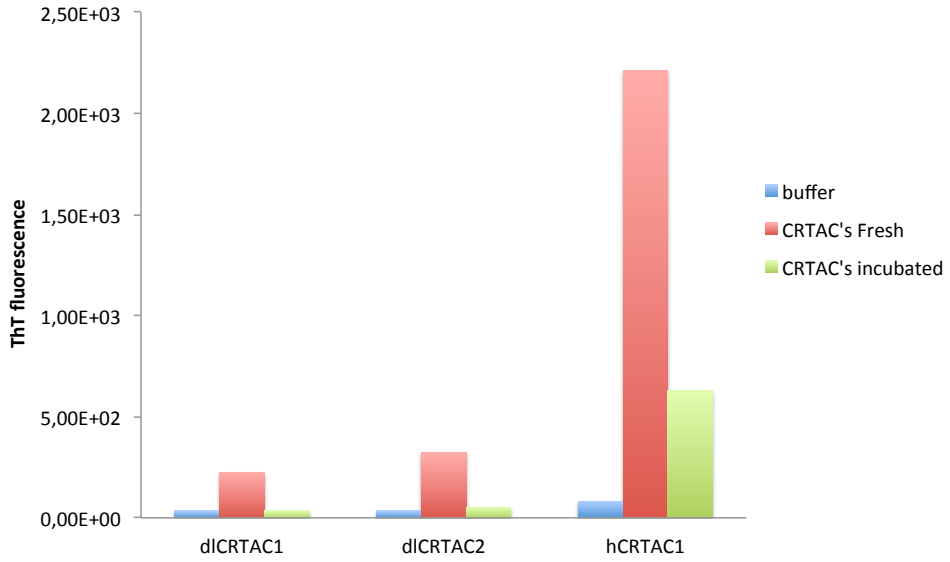

Fig. 9. Thioflavin-T assay. Blue bars correspond to the negative control comprising buffer (100 mM Tris-HCl, 500 mM NaCl, 1 mM CaCl2 at pH 8) without proteins. Red bars correspond to fluorescence measurements of dlCRTAC1, dlCRTAC2 and hCRTAC1 respectively, that were not incubated (T=0h). Green bars correspond to dlCRTAC1, dlCRTAC2 and hCRTAC1, mixed with Thioflavin-T after 20 days of incubation at 37ºC.

4. Discussion and conclusions

Human CRTAC1 and piscine dlCRTAC1 and dlCRTAC2 recombinant proteins were overexpressed in *E.coli* BL21(DE3) and Origami 2 (DE3), in the insoluble cellular fraction as inclusion bodies. The introduction of foreign genes for heterologous protein production in recombinant expression systems, such as *E.coli*, often leads to the production of expressed proteins as inclusion bodies [28].

The small-scale optimization results (not shown) suggest that different *E.coli* strains could influence the expression of the recombinant CRTAC´s and *E.coli* Origami 2 (DE3) shows to be most efficient for piscine CRTAC´s expression while *E.coli* BL21(DE3) was the most adequated for hCRTAC1 expression. However, the double of induction time (12 h) was required for the production of recombinant hCRTAC1 in *E.coli* BL21(DE3) reach a similar level to that of piscine CRTAC's productivity. In some cases, cells transfected with recombinant expression vectors modify the intrinsic cellular protein levels which adversely affect cell growth [29] and hCRTAC1 required more induction time to reach similar expression levels to those of piscine CRTAC´s.

Contrary to what was observed for human and piscine CRTAC1 form which were obtained in a high yield (estimated in 76,52 - 90,91%) after solubilisation, purification and refolding process, significant losses (36,46%) of dlCRTAC2 protein occurred and the critical steps were the concentration and refolding process. It is known that the concentration of some proteins until certain limits induce protein aggregation [30] and the refolding process by dialysis promote the proteins to return the right folding, in particular proteins rich in disulphide bonds, however under certain concentration of denaturant the proteins may form aggregates that resulting from folding intermediate species [31]. Similar results were also observed for the recombinant sea bream CRTAC2 [21] that shows a high tendency to form aggregates. CRTAC´s proteins tend to precipitate and form aggregates during the recombinant production and this property seems to be more pronounced in the teleost CRTAC2 form in comparison with human and teleost CRTAC1.

Solubilisation and refolding process are two essential steps to obtain native protein conformation [32]. The resulting refolded and solubilized CRTAC´s were used for molecular and structural characterization using a comparative approach.

All CRTAC´s were obtained with high purity and their identity was assessed by mass spectrometry that generated a confirmatory peptide mass fingerprint. Beyond the

expected MW forms of human and piscine CRTAC´s, an additional smallest protein co-expressed with dlCRTAC2 was observed and their identity was confirmed as dlCRTAC2 that suggest an effect of protein processing by *E.coli* giving rise to alternative CRTAC2 forms. In fact, CRTAC is an ancestral protein present in several prokaryotes that possesses domains such as ASPIC/UnbV and several VCBS, which are highly conserved and strongly associated with bacteria.

Purified and refolded CRTAC´s in denaturing conditions shown immunoreactivity with specific polyclonal antibodies. Single protein bands with the expected MW were observed for each CRTAC in Western blot and this confirm the identity, integrity and quality of the resulting recombinant proteins. However, the availability of antibodies that allows discriminating between CRTAC1 and 2 is a real challenge due to the high degree of similarity (72%) between them. We know in advance that the antiserum generated against sea bass CRTAC2 specific peptide is the only specific for CRTAC2 (results not published) and although the antibodies used to recognize CRTAC1 (human or sea bass) cross-react with the two forms (Anjos et al., 2013), the descrimination between proteins was also possible due to the MW differences between them (67,9 kDa for dlCRTAC1, 68,6 kDa for hCRTAC1 and 56,7k Da for dlCRTAC2).

Size exclusion chromatography revealed that piscine dlCRTAC1 and dlCRTAC2 form mostly aggregates and only very little amounts of monomer are revealed by the small peaks. hCRTAC1, however, exhibited appreciable amounts of both the aggregate and monomeric form. These findings suggest that piscine CRTACs are less stable or less soluble than the human counterpart.

In previous studies recombinant piscine CRTAC2 revealed its tendency to form aggregates under non denaturing conditions, either in vivo and in vitro [21]. The aggregate propensity of hCRTAC1 has not been previously explored although such tendency is suggested by its tissue distribution [19], overexpression in neurodegenerative diseases [15] and biomarker characteristics [12] suggest .

Circular Dichroism analysis demonstrated that the native structure of piscine and human CRTAC's is rich in beta-sheet (≈40%), contains approximately 10,3% alpha-helix and the remaining is disordered. The high Beta-sheet content of proteins, as CRTAC's coupled together with their high percentage of non-polar hydrophobic amino acid residues [19], highlight their tendency to form aggregates. CD analysis showed that dlCRTAC1, dlCRTAC2 and hCRTAC1 are highly thermostable as they did not show sign of thermal unfolding. However hCRTAC1 showed slightly higher loss of

secondary structure comparatively to the piscine forms suggesting that human CRTAC1 might be less thermostable than piscine CRTAC´s.

Intrinsic fluorescence emission demonstrated that after heating all CRTAC's to 90ºC there was no appreciable shift in the emission maximum suggesting that there was no tryptophane exposure due to unfolding. These results are in agreement with the CD measurements showing that CRTAC's are thermostable and retain tertiary structure up to 90ºC, as previously described for sea bream CRTAC2 [21]. These results could be justified by CRTAC´s insolubility and aggregation propensity: One could hypothesize that the proteins aggregate before disintegrating their secondary and tertiary structure [21] and render a highly thermal resistant multimeric structure that retain its 3 dimensional structural motifs.

Fluorescence emission studies after submitting the proteins to several freezing cycles also suggest that hCRTAC1 has lower stability than piscine CRTACs: after only one freezing cycle hCRTAC1 has a much more pronounced loss of ordered structure revealed by a loss of signal in the 215-220 nm range. Curiously after three freezing cycles the CD signal becomes larger and more negative with minimum mrw shift to lower wavelengths (approximately 208 nm) and more positive at approximately 195 nm. This may suggest a shift in secondary structure to an increase in alpha-helical content. This behaviour is more pronounced for hCRTAC1 and dlCRTAC2 while for dlCRTAC1 it is less pronounced and happens only after 4 cycles of freezing. These results suggest that dlCRTAC1 is more stable and perhaps dlCRTAC2 is more flexible and prone to conformational changes. hCRTAC1 seems to be less stable in general.

Thioflavin T is a dye which bind β-sheet structure and its used for identification of amyloid fiber formation due to the highly packed β-sheet structure of amyloids [33]. We used this dye to test if CRTACs form amyloid-like structures after incubation at 37ºC. Thioflavin-T assays showed a decrease in fluorescence emission for CRTACs incubated at 37ºC comparatively to non-incubated CRTAC. The fact that ThioflavinT fluorescence decreased after incubation may suggest protein degradation or structure loss due to the long period of incubation or formations of non-amyloid aggregates. Nevertheless the formation of amyloid structures by CRTAC´s should not be discarded. The fact that the protein forms very easily large aggregates that seem to retain tertiary structure suggests that CRTAC aggregates could have a functional role. Weather CRTAC aggregation can lead to amyloid formation remains to be established. The fact that fresh CRTACs exhibit high ThioflavinT fluorescence could be explained by their

high intrinsic β-sheet content or perhaps a "native" functional amyloid-like conformation that can provide binding sites for Thioflavin-T. Such idea is supported by the evidence that CRTAC´s form large aggregates in vivo [21]. Functional amyloids have previously been reported. For example *E.coli* amyloid curli, form extracellular fibers highly stable, insoluble and high molecular weight protein complexes rich in beta sheet secondary structure [34]. The existence of functional amyloids could explanation the tendency of certain proteins to aggregate [35].

In conclusion, structural studies demonstrated that CRTAC's were thermostable to heating and freezing effect and probably has the ability to retain your secondary and tertiary structure. The beta-sheet conformation and hydrophobic composition indicate the proteins most probably aggregate. The presence of functional amyloid and crystallography knowledge coud explain a tendency of certain proteins aggregate.

Chapter IV

Future perspectives

Future perspectives

In the present study the recombinant hCRTAC1 and piscine CRTAC1 and 2 proteins were generated in large-scale production and this are valuable tools that will permit in the future the development of structural and functional studies to establish the actions of this enigmatic protein.

The fact that CRTAC´s tends to form very easily large aggregates that retain their tertiary structure and that are highly thermostable suggest that such aggregates could have a functional role. This idea is supported by the evidence that CRTAC´s form large aggregates *in vivo*[21]. Clarify the type and mechanisms and dynamics of aggregates formed by CRTAC´s will be important in the future in order to establish if they are functional structures (non-pathological proteins) or are involved in pathologies and this will be of considerable interest for medicine and pharmaceutical industries. Your tendency to aggregate should be assessed in future studies *in vivo* and *in vitro* and the results may be the answer in diseases where this proteins are involved. Moreover, the CRTAC´s ubiquitous tissue distribution [21] [19] and association with a number of pathologies (US20120231477 A1) suggest an important role for this protein in health and disease.

The resolution of CRTAC´s crystal structures (3D structures) will be also of relevant interest, as future perspective, to give insights about protein function, evolution and the resulting information could contribute to explain the tendency of certain proteins to aggregate.

Comparative analysis between human and piscine CRTAC´s may contribute to understand protein folding, stability and give insights about the functional evolution and this information will certainly be of importance for medical application and pharmaceutical and biotechnological industries.

References

1. Berg, J. M., Tymoczko, J. L. & Stryer, L. *Biochemistry*. *Biochemistry textbook* 1120 (2006).

2. Blatch, G. L. & Lässle, M. The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *Bioessays* **21,** 932–939 (1999).

3. Campbell, I. D. & Humphries, M. J. Integrin structure, activation, and interactions. *Cold Spring Harb. Perspect. Biol.* **3,** 1–14 (2011).

4. Handford, P. A. *et al.* Key residues involved in calcium-binding motifs in EGF-like domains. *Nature* **351,** 164–167 (1991).

5. Bondos, S. E. & Bicknell, A. Detection and prevention of protein aggregation before, during, and after purification. *Anal. Biochem.* **316,** 223–231 (2003).

6. Iram, A. & Naeem, A. Protein Folding, Misfolding, Aggregation and Their Implications in Human Diseases: Discovering Therapeutic Ways to Amyloid-Associated Diseases. *Cell Biochem. Biophys.* (2014). doi:10.1007/s12013-014-9904-9

7. García-Fruitós, E., Sabate, R., de Groot, N. S., Villaverde, A. & Ventura, S. Biological role of bacterial inclusion bodies: a model for amyloid aggregation. *FEBS J.* **278,** 2419–27 (2011).

8. Kelly, S. M., Jess, T. J. & Price, N. C. How to study proteins by circular dichroism. *Biochim. Biophys. Acta* **1751,** 119–139 (2005).

9. Sreerama, N., Venyaminov, S. Y. & Woody, R. W. Estimation of protein secondary structure from circular dichroism spectra: inclusion of denatured proteins with native proteins in the analysis. *Anal. Biochem.* **287,** 243–51 (2000).

10. Smyth, M. S. & Martin, J. H. J. Review x Ray crystallography. 8–14 (2000).

11. Khurana, R. *et al.* Mechanism of thioflavin T binding to amyloid fibrils. *J. Struct. Biol.* **151,** 229–238 (2005).

12. Steck, E. *et al.* Chondrocyte expressed protein-68 (CEP-68), a novel human marker gene for cultured chondrocytes. *Biochem. J.* **353,** 169–74 (2001).

13. Rocha, B. *et al.* Secretome analysis of human mesenchymal stem cells undergoing chondrogenic differentiation. *J. Proteome Res.* **13,** 1045–54 (2014).

14. Steck, E. *et al.* Chondrocyte secreted CRTAC1: a glycosylated extracellular matrix molecule of human articular cartilage. *Matrix Biol.* **26,** 30–41 (2007).

15. Balood, M. *et al.* Elevated serum levels of lysophosphatidic acid in patients with multiple sclerosis. *Hum. Immunol.* **75,** 411–3 (2014).

16. Schwab, M. E. & Strittmatter, S. M. Nogo limits neural plasticity and recovery from injury. *Curr. Opin. Neurobiol.* **27C,** 53–60 (2014).

17. Brems, H. *et al.* Glomus tumors in neurofibromatosis type 1: genetic, functional, and clinical evidence of a novel association. *Cancer Res.* **69,** 7393–401 (2009).

18. Rabinowitz, Y. S., Dong, L. & Wistow, G. Gene expression profile studies of human keratoconus cornea for NEIBank: a novel cornea-expressed gene and the absence of transcripts for aquaporin 5. *Invest. Ophthalmol. Vis. Sci.* **46,** 1239–46 (2005).

19. Redruello, B. *et al.* CRTAC1 homolog proteins are conserved from cyanobacteria to man and secreted by the teleost fish pituitary gland. *Gene* **456,** 1–14 (2010).

20. Kassahn, K. S., Dang, V. T., Wilkins, S. J., Perkins, A. C. & Ragan, M. A. Evolution of gene function and regulatory control after whole-genome duplication : Comparative analyses in vertebrates. 1404–1418 (2009). doi:10.1101/gr.086827.108.evolution

21. Anjos, L., Gomes, A. & Melo, E. Cartilage Acidic Protein 2 a hyperthermostable, high affinity calcium-binding protein. *Acta (BBA)-Proteins* (2013).

22. Tomas, A., Futter, C. E. & Eden, E. R. EGF receptor trafficking: consequences for signaling and cancer. *Trends Cell Biol.* **24,** 26–34 (2014).

23. Hynes, R. O. Integrins: bidirectional, allosteric signaling machines. *Cell* **110,** 673–87 (2002).

24. Marchler-Bauer, A. *et al.* CDD: conserved domains and protein three-dimensional structure. *Nucleic Acids Res.* **41,** D348–52 (2013).

25. Chen, C. K.-M., Chan, N.-L. & Wang, A. H.-J. The many blades of the β-propeller proteins: conserved but versatile. *Trends Biochem. Sci.* **36,** 553–61 (2011).

26. Shin, T. M. *et al.* Formation of soluble amyloid oligomers and amyloid fibrils by the multifunctional protein vitronectin. *Mol. Neurodegener.* **3,** 16 (2008).

27. Gill, S. C. & von Hippel, P. H. Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **182,** 319–326 (1989).

28. Philo, J. S. & Arakawa, T. Mechanisms of protein aggregation. *Curr. Pharm. Biotechnol.* **10,** 348–51 (2009).

29. Han, M., Jeong, K. J., Yoo, J., Yup, S. & Lee, S. Y. Engineering Escherichia coli for Increased Productivity of Serine-Rich Proteins Based on Proteome Profiling Engineering Escherichia coli for Increased Productivity of Serine-Rich Proteins Based on Proteome Profiling. (2003). doi:10.1128/AEM.69.10.5772

30. Yamaguchi, H. & Miyazaki, M. Refolding Techniques for Recovering Biologically Active Recombinant Proteins from Inclusion Bodies. (2014).

31. Tsumoto, K., Ejima, D., Kumagai, I. & Arakawa, T. Practical considerations in refolding proteins from inclusion bodies. *Protein Expr. Purif.* **28,** 1–8 (2003).

32. Fink, a L. Protein aggregation: folding aggregates, inclusion bodies and amyloid. *Fold. Des.* **3,** R9–23 (1998).

33. Wolfe, L. S. *et al.* Protein-induced photophysical changes to the amyloid indicator dye thioflavin T. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 16863–8 (2010).

34. Hung, C., Marschall, J., Burnham, C.-A. D., Byun, A. S. & Henderson, J. P. The bacterial amyloid curli is associated with urinary source bloodstream infection. *PLoS One* **9,** e86009 (2014).

35. Otzen, D. Functional amyloid: Turning swords into plowshares. *Prion* **4,** 256–264 (2010).

Appendix 1

Protein ID MASCOT
cartilage acidic protein 1 isoform X1 [Homo sapiens] gi|530393911
Mr(Da)= 68600 pI= 5,05
1 MAPSADPGMS RMLPFLLLLW FLPITEGSQR AEPMFTAVTN SVLPPDYDSN 51 PTQLNYGVAV TDVDHDGDFE IVVAGYNGPN LVLKYDRAQK RLVNIAVDER 101 SSPYYALRDR QGNAIGVTAC DIDGDGREEI YFLNTNNAFS GVATYTDKLF 151 KFRNNRWEDI LSDEVNVARG VASLFAGRSV ACVDRKGSGR YSIYIANYAY 201 GNVGPDALIE MDPEASDLSR GILALRDVAA EAGVSKYTGG RGVSVGPILS 251 SSASDIFCDN ENGPNFLFHN RGDGTFVDAA ASAGVDDPHQ HGRGVALADF 301 NRDGKVDIVY GNWNGPHRLY LQMSTHGKVR FRDIASPKFS MPSPVRTVIT 351 ADFDNDQELE IFFNNIAYRS SSANRLFRVI RREHGDPLIE ELNPGDALEP 401 EGRGTGGVVT DFDGDGMLDL ILSHGESMAQ PLSVFRGNQG FNNNWLRVVP 451 RTRFGAFARG AKVVLYTKKS GAHLRIIDGG SGYLCEMEPV AHFGLGKDEA 501 SSVEVTWPDG KMVSRNVASG EMNSVLEILY PRDEDTLQDP APLECGQGFS 551 QQENGHCMDT NECIQFPFVC PRDKPVCVNT YGSYRCRTNK KCSRGYEPNE 601 DGTACVERTL LLGLCNLLGK

Protein ID MASCOT
Cartilage acidic protein 1 [Tetraodon nigroviridis] gi|85838736
Mr(Da)= 69600 pI= 5,42
1 MLAWVLLFLP PVSSAQRSEP VFSSITKSIL PPNYENNPTQ LNYGVAVTDV 51 DGDGDLEVFV AGYNGPNLVL KYIQDQKRLV NIAVDNRSSP FYALRDRQGN 101 AIGVTACDID GDGREEIYVL NTNNAFSGRA TYSDKLFKFR NGRFEDLLND 151 DINEHRDVAN RMAGRSVACV DRKGTGRYAI YIANYASGNV GPHALIEMDE 201 LASDLSQGVI ALSNVAEEAG VNKFTGGRGV VVGPILNQIL PDVFCDNEYG 251 PNFLFRNNGD GTFTDVAQQA GVEDPMQHGR GVALADFNRD GRTDIVYGNW 301 NGPHRLFMQL NNRRQKFKDI ASQKFSMPSP VRTVIAADFD NDNELEVFFN 351 NIAYRGPSAN RLFRVSRREH GDPQIEELNV GEASEPEGRG TGAVATDFDG 401 DGRLELLVSH GESAAQPLSV YKVLQGTSNS WLRVIPRTKF GAFARGAKVV 451 VYTKKSGTHT RIIDGGSGYL CEMEPVAHFG LGKDVATGVE VYWPDGRSVV 501 RLLEPSDLNS VLEIQYPRDV EVTPTAQTEC GHGFALNEKG RCTDEDECTR 551 YPFACPLDRP VCVNTYGSYR CRAKRRCNQG FEPSDDGSAC VGQVAYFGGT 601 RSFADRKWSG LCFWLLLPHV LRPGLL

Protein ID MASCOT
dlCRTAC2 against own sequence
Mr(Da)= 59200 pI= 5,03
1 MWASGLLLFL VGLWHQSRAQ GSEPMLRVVT ETMLPPDSLH NPTQLNYGMA 51 VTDVDGDGDL EVVVAGYNGP NLVLKYDRTQ NRLVNIAIDD SNSPYYALRD 101 RAGNAIGVTA CDVDGDGREE IYFLNTNNAY SGRATYSDKL FKFRNGRFED 151 LLSDELNVRR GVANRMAGRS VACIDRKGTG RYSVYVANYA SGNVGPHALL 201 EMDETASDVA KGIIALSDVA AVAGVNKFTG GRGVVVGPIL SQSRSDVFCD 251 NENGPNFLFK NNGDGTFVDM ARQAGVEDRY QHGRGVALAD FNGDGKTDII 301 YGNWNGPHRL FLQGSDSTFR NIATGGFAAP SPIRTVIAAD LDNDKELDVF 351 FNNIAYRGNA PNRLFRVSRR ANADPLIQEL NVGDAAEPEG RGTGGTVTDF 401 DGDGQLDLLL AHGESARQPI SVFKVTQGSS NNWLRVIPRT QFGSFARGAK 451 VTAFTSQSGA HTRIIDGGSG YLCEMEPVAH FGLGNDEVTV LEVSWPDGSS 501 ITRTLQSGEM NSVVEVAYPK EGETFLLAND TQCGNGFTVK NGHCAGL

Protein ID MASCOT
dlCRTAC2 against own sequence (small band)
Mr(Da)= 37000
1 MWASGLLLFL VGLWHQSRAQ GSEPMLRVVT ETMLPPDSLH NPTQLNYGMA 51 VTDVDGDGDL EVVVAGYNGP NLVLKYDRTQ NRLVNIAIDD SNSPYYALRD 101 RAGNAIGVTA CDVDGDGREE IYFLNTNNAY SGRATYSDKL FKFRNGRFED 151 LLSDELNVRR GVANRMAGRS VACIDRKGTG RYSVYVANYA SGNVGPHALL 201 EMDETASDVA KGIIALSDVA AVAGVNKFTG GRGVVVGPIL SQSRSDVFCD 251 NENGPNFLFK NNGDGTFVDM ARQAGVEDRY QHGRGVALAD FNGDGKTDII 301 YGNWNGPHRL FLQGSDSTFR NIATGGFAAP SPIRTVIAAD LDNDKELDVF 351 FNNIAYRGNA PNRLFRVSRR ANADPLIQEL NVGDAAEPEG RGTGGTVTDF 401 DGDGQLDLLL AHGESARQPI SVFKVTQGSS NNWLRVIPRT QFGSFARGAK 451 VTAFTSQSGA HTRIIDGGSG YLCEMEPVAH FGLGNDEVTV LEVSWPDGSS 501 ITRTLQSGEM NSVVEVAYPK EGETFLLAND TQCGNGFTVK NGHCAGL

I- Identification of CRTACs proteins by searching with tryptic peptide sequence against primary sequence databases, using the MASCOT search. dlCRTAC2 protein and dlCRTAC2(small band) identification was occur by searching with tryptic peptide sequence against own sequence.
The dlCRTAC2(small band) was visualized in Fig, 1B (Pag 31) (Orange square).