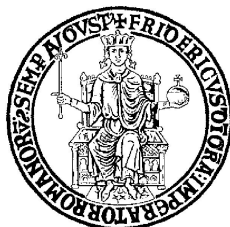


**UNIVERSITA' DEGLI STUDI DI NAPOLI
FEDERICO II**



TESI DI DOTTORATO IN SCIENZE CHIMICHE
XXVI CICLO

**STRUCTURAL STUDY OF PROTEIN-PROTEIN
AND NUCLEIC ACID-PROTEIN COMPLEXES:
STABILITY AND RECOGNITION SPECIFICITY**

ANDREA PICA

TUTORE
Prof.ssa Filomena Sica

RELATORE
Prof.ssa Flavia Nastri

COORDINATORE DEL XXVI CICLO
Prof. Luigi Paduano

CONTENTS

Contents	3
Preface	7
PART I: THROMBIN – APTAMER RECOGNITION	11
State of the art	13
<i>Nucleic Acid Aptamers</i>	13
<i>In vitro</i> selection	13
Structure	14
Advantages of aptamers	18
<i>Aptamer as anticoagulants</i>	18
Human alpha thrombin	19
Chapter I – HD1: you never forget the first love	21
Introduction	23
Results	25
<i>Overall survey of the two complexes</i>	25
<i>Focusing on thrombin</i>	26
<i>Analyzing the aptamer moieties</i>	27
<i>Differences between the complexes</i>	28
Discussion	32
<i>G-quadruplex scaffold and thrombin exosite I</i>	32
Conclusions and future perspectives	34
<i>Aptamers targeting exosite I</i>	34
Improving aptamers by chemical modification	35
Materials and methods	36
<i>Sample preparation</i>	36
<i>Crystallization and data collection</i>	36

Structure determination and refinement	37
Chapter II – HD22: when a G-quadruplex is not enough	41
Introduction	43
Results	45
<i>Crystal structure of thrombin-HD22[27] complex</i>	45
<i>A novel duplex-quadruplex motif</i>	46
<i>Interaction with thrombin</i>	50
<i>Circular dichroism study</i>	53
Discussion	55
<i>Duplex-quadruplex motif and thrombin exosite II</i>	55
Conclusions and future perspectives	57
<i>Aptamers targeting exosite II</i>	57
Materials and methods	58
<i>Sample preparation</i>	58
<i>Crystallization and data collection</i>	59
<i>Structure determination and refinement</i>	60
<i>Circular Dichroism</i>	62
References	63
Part II – DOMAIN SWAPPING AND RIBONUCLEASES	73
State of the art	75
Chapter III – Domain swapping: giving something to take “something else” back	79
Introduction	81
Results	82
<i>Solution studies</i>	82

<i>Crystal growth</i>	86
<i>Crystallographic analysis</i>	86
Discussion	91
Materials and Methods	93
<i>Protein production and purification</i>	93
<i>Dissociation studies</i>	93
<i>Crystallization and data collection</i>	94
<i>Structure determination and refinement</i>	94
Chapter IV – Too much is a bad thing	97
Introduction	99
Results	101
<i>desHP overall structure</i>	101
<i>Supramolecular assembly</i>	105
<i>Sulfate anions</i>	106
<i>Solution study</i>	107
Discussion	111
Conclusion	114
Materials and Methods	114
<i>Crystallization and data collection</i>	114
<i>Structure determination and refinement</i>	115
<i>Atomic Force Microscopy imaging</i>	115
<i>Fluorescence experiments</i>	116
Chapter V – Two is better than one	117
Introduction	119
Results and discussion	121
<i>Overall crystal structure</i>	121

<i>Oligomerization studies</i>	123
Conclusions and future perspectives	125
Materials and Methods	126
<i>Protein production and purification</i>	126
<i>Cathodic gel electrophoresis</i>	126
<i>Circular Dichroism</i>	127
<i>Crystallization and data collection</i>	127
References	129
Candidate's Publication List	137
<i>Structures deposited in the Protein Data Bank</i>	146

PREFACE

Molecular recognition is the central element from which an entire biochemical universe is built. The function of proteins, the dominant biomolecular machines, is mediated by their interaction with partners that can vary from small molecules to very large molecules, such as other proteins and nucleic acids. Protein-protein and protein-nucleic acid interactions are ubiquitous and essential to all known cellular and physiological processes. Macromolecular recognition is the origin of biological specificity. It takes place in a crowded noisy biochemical environment and requires the recognition of a specific target within a background of various often similar competing molecules. In such biological environment molecular selectivity and affinity provide the ground for any recognition phenomenon. The conformational flexibility of recognizers plays a key role in the optimization of the intermolecular interactions.

One limitation that delays the understanding of molecular recognition is the lack of high-resolution three-dimensional structures of biomacromolecules in assemblies. The challenge of understanding the molecular driving forces in the recognition process has indeed stimulated over the last fifty years the determination of the atomic structure of a great number of protein-protein and protein-DNA complexes. In this field a key role has been played by X-ray crystallography. In particular, structural studies suggest that high affinity interactions arise from the sum of many close range non-covalent contacts that represent the driving force of molecular recognition. The crystallographic approach has provided more insights on biological properties of complexes between macromolecules, allowing a deeper knowledge of physiological processes and opening new perspectives to drug design.

The central objective of this PhD project has been the structural characterization of recognition processes related to human diseases. The communication between the interacting molecules has been analyzed using X-ray crystallog-

raphy as the main technique, in combination with other biophysical techniques, such as CD, fluorescence spectroscopy and AFM.

This thesis is divided in two parts. In the first part, modulation of human α -thrombin function with nucleic acid aptamers is investigated from a structural point of view. The molecular determinants in the recognition between thrombin and the best known 15-mer aptamer HD1 are deeply reviewed and discussed on the basis of the structure of two complexes between the protein and two deletion mutants of HD1. Moreover, the structure of the more powerful aptamer HD22, targeting a different thrombin exosite, is finally unraveled and discussed. On the basis of this structure, a bunch of literature data, till now been considered unclear and ambiguous, could be explained and rationalized.

The second part of the thesis is focused on protein self-recognition and oligomerization through 3D domain swapping. It has been shown how this mechanism can lead to the formation of dimers, oligomers, and, as an extreme case, fibrils. The work on simple model proteins, usually non-directly associated with a disease, has helped the understanding of the more complex behavior of disease-associated proteins. In this context a peculiar role has been played by the mammalian pancreatic-type ribonuclease family, which has been used as model system in various studies on the protein oligomerization process. 3D domain swapping has been found to endow some of this proteins with special functions, besides their normal enzymatic activity, such as cytotoxic activity. In particular, the oligomerization has been used to convert the non-cytotoxic bovine pancreatic ribonuclease (RNase A) in multimeric protein with medical relevance as anti-tumor drugs. The final aim is the production of new ribonuclease oligomers with improved antitumor activity for the treatment of human cancer and able to be well-tolerated by patients. The main idea developed in this thesis has been the use of protein engineering to force monomeric proteins to form high stable domain-swapped dimers or oligomers that could be used as a new generation of antitumor agents.

Finally, a new very interesting evidence that domain swapping is involved in the assembly of fibrillar aggregates has been obtained. These results support the hypothesis that the subunits of the fibril retain much of the native structure of the protein.

**PART I: THROMBIN – APTAMER
RECOGNITION**

STATE OF THE ART

Nucleic Acid Aptamers

Since their discovery in the second half of 19th century (Dahm, 2008), nucleic acids have been thought exclusively as genetic information carriers, containing the encoded information for the translation to proteins (Watson, *et al.*, 1953; Crick, 1970). However, in recent years, new activities for nucleic acids have been discovered since they have been found to be involved in a series of other less known processes (Bloomfield, *et al.*, 2000). The discovery of nucleic acid sequences with catalytic and regulatory activities over the past decades illuminated a whole new array of functions that can be conducted by these macromolecules. In particular, at the beginning of the 1990s, an interesting property of ribonucleic acid was disclosed. It was observed that short RNA molecules could fold into three-dimensional structures and specifically bind different non-nucleic acid targets (Ellington, *et al.*, 1990). The conformation of nucleic acids, encoded in their primary sequence, is the element that drives the specific and selective interaction with a target. Despite their rather high flexibility, these DNA or RNA molecules, referred to as nucleic acid aptamers, can undergo adaptive binding upon target recognition, by switching from a loose conformation to a distinct more ordered structure (Hermann, *et al.*, 2000; Tan, *et al.*, 2011).

In vitro selection

Aptamers are single-stranded DNA or RNA sequences, generated through *in vitro* selection, which adopt stable three-dimensional structures having a good shape complementarity with their target. A procedure to derive sequences recognizing specific targets was first described by Tuerk and Gold under the name of Systematic Evolution of Ligands by EXponential Enrichment (SELEX) (Ellington, *et al.*, 1990; Tuerk, *et al.*, 1990). It is a combinatorial chemistry procedure that allows rapid selection of a DNA or RNA aptamers that have appropriate binding affinity to a given molecular target, starting from a large initial

library of oligonucleotides (10^8 random sequences are usually screened). Up to date a large number of different DNA and RNA aptamers were isolated against a wide range of targets (Zhu, *et al.*, 2012; Radom, *et al.*, 2013): from small molecules (Hermann, *et al.*, 2000; Miyachi, *et al.*, 2009) to proteins (Bock, *et al.*, 1992; Chi-hong, *et al.*, 2003; Daniels, *et al.*, 2003; Khati, *et al.*, 2003; Parekh, *et al.*, 2010), virus-infected cells (Tang, *et al.*, 2009; Keefe, *et al.*, 2010), stem cells (Guo, *et al.*, 2006), and cancer cells (Chen, *et al.*, 2008).

Structure

Aptamer can fold in unique and well-defined three-dimensional structures. Sometimes they assume intricate and unpredictable conformations, in order to interact closely with or adhere to the surface of their target. The multitude of folding motifs they can assume allows them to bind to almost any small molecule, nucleic acid or protein target with high specificity and affinity.

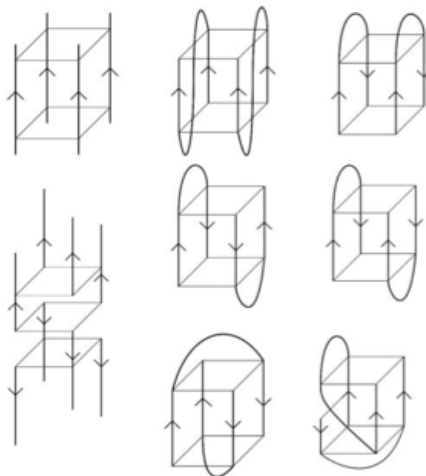
Several motifs of aptamers have been described, the most famous being the hairpin structure (Fan, *et al.*, 1996), the *pseudo-knot* (Lorsch, *et al.*, 1994) and the G-quadruplex (Macaya, *et al.*, 1993). Furthermore, it is well-known and documented that their conformation is strictly dependent on the concentration and kind of ions in solution (Draper, *et al.*, 2005; Girardot, *et al.*, 2010).

Among the large number of structural arrangements adopted by aptamers, the four-stranded topology called G-quadruplex has become of great interest, since it has been found to be implicated in relevant biological processes (Mergny, *et al.*, 1998; Hurley, 2002; Neidle, *et al.*, 2002; Kumari, *et al.*, 2007; Patel, *et al.*, 2007; Fernando, *et al.*, 2009; Lipps, *et al.*, 2009). Moreover it represents an attractive molecular scaffold to be used for drug design (Balasubramanian, *et al.*, 2009).

Quadruplex nucleic acids may be defined as higher order structures formed by DNA or RNA sequences containing at least one tract of contiguous guanine nucleotides (Neidle, 2009). The main component of the G-quadruplex is the G-tetrad or G-quartet, a roughly planar arrangement of four guanine bases associ-

ated through a cyclic array of Hoogsteen-like hydrogen bonds, in which each guanine base both accepts and donates two hydrogen bonds. This planar arrangement of guanines has a high propensity to self-stack. G-quadruplexes contain a core of at least two stacked G-quartets, stabilized by interaction of cations with the O6 atoms of guanines belonging to the cyclic array (Williamson, 1994). Indeed, the overall G-quadruplex arrangement is critically dependent on the nature of counter ions (Engelhart, *et al.*, 2009). Quadruplex arrangements can be classified on the basis of the molecularity of the quadruplex formation process. Thus monomolecular (intramolecular), bimolecular, and tetramolecular quadruplexes do exist. Within the rather rigid structure of the G-quadruplex core, a great topological variation is allowed. The sequences linking the G-tracts in a bimolecular or monomolecular G-quadruplex are called loops and the number of ways they can connect the guanines constituting the G-core generates several topological variations of the G-quadruplex itself. Some G-quadruplex topologies and loops types are described in Figure 1 and Figure 2.

(a)



(b)

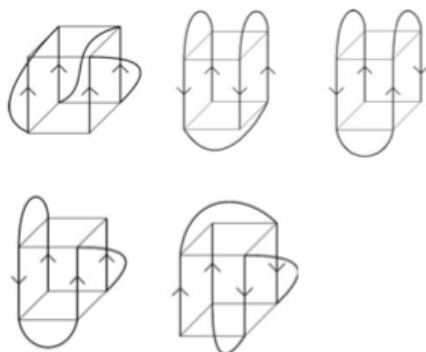


Figure 1. (a) Some possible topologies for simple tetramolecular (on the left-hand side) and bimolecular quadruplexes. Strand polarities are shown by arrows. (b) Some possible topologies for simple unimolecular quadruplexes. Image taken from (Burge, *et al.*, 2006).

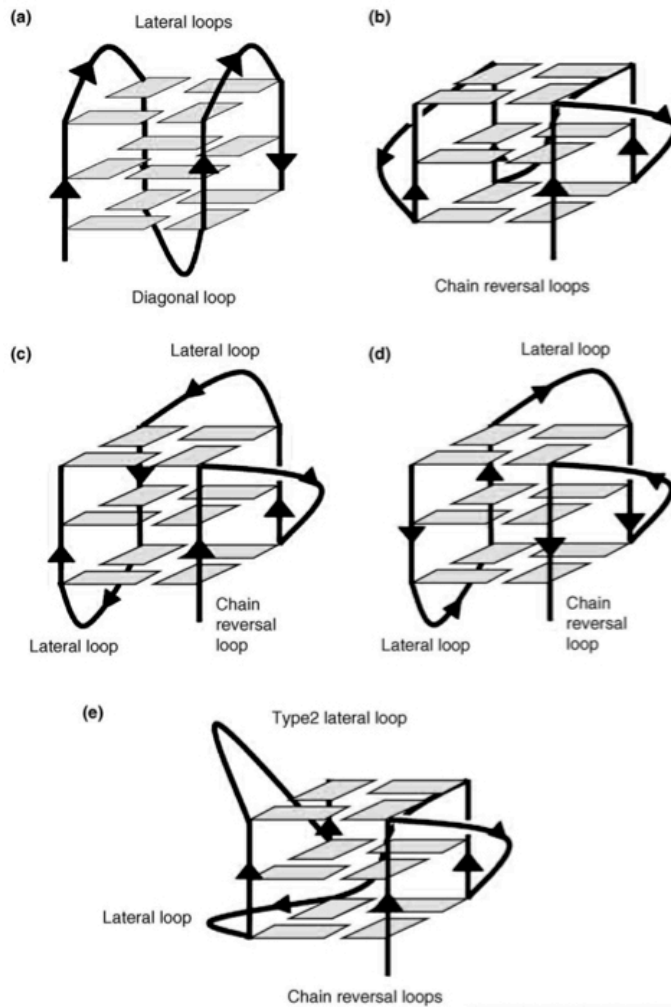


Figure 2. Schematic views of various intramolecular quadruplex topologies. (a) Antiparallel topology as found in the Na^+ form of a human telomeric quadruplex (Wang, *et al.*, 1993). (b) Parallel fold as found in the crystal structure of a K^+ form of a human telomeric quadruplex (Parkinson, *et al.*, 2002). (c) (3 + 1) hybrid topology (form 1) as found in the NMR structures of the K^+ form of human telomeric quadruplex sequences (Ambrus, *et al.*, 2006; Luu, *et al.*, 2006). (d) (3 + 1) hybrid topology (form 2) as found in the NMR structures of the K^+ form of human telomeric quadruplex sequences (Phan, *et al.*, 2006; Dai, *et al.*, 2007). Note that the order of loops is now reversed compared with that in form 1. (e) Topology of the c-kit promoter quadruplex (Phan, *et al.*, 2007). Image taken from (Neidle, 2009).

Advantages of aptamers

Aptamers are considered a sort of nucleic acid version of antibodies, holding many advantages over them: they have higher thermal stability, tolerance to wide ranges of pH and salt concentration, and they can be much more easily synthesized (Jayasena, 1999). Nucleic acid aptamers can be developed against a wide range of targets, including those toxic to organisms and therefore beyond the reach of antibody development. Moreover, aptamer identification through *in vitro* SELEX is usually more efficient and cost-effective than antibody development. The advancement of automated nucleic acid synthesis enables easy, cost-effective chemical synthesis and modification of functional moieties, as well as large-scale commercial production. Other advantages include rapid tissue penetration based on the relatively small molecular weights, low immunogenicity, and ease of antidote development (Rusconi, *et al.*, 2004; Oney, *et al.*, 2009). These advantages makes aptamers more than attractive for clinical or biomedical applications, such as disease diagnosis and therapy.

Aptamer as anticoagulants

In addition to specific recognition abilities, some aptamers can bind to and further modulate the biological activities of the molecular targets which are involved in pathogenesis or in key physiological processes. As a consequence, aptamers, as therapeutics, can readily regulate biological pathways and interfere with disease development (Nimjee, *et al.*, 2005b). In particular, aptamer technology is extensively used to modulate the function of most of the serine proteases and the cofactors involved in the coagulation pathway.

The high affinity of the selected aptamers allows for therapeutic dosing at submicromolar levels, which should reduce potential non-specific effects. Moreover, they are non-immunogenic and show tunable and predictable pharmacokinetic properties. In fact, aptamer *in vivo* half-life, which is usually very short and in the order of few minutes, can be prolonged (up to few hours) by chemical modifications (e.g. high molecular weight polyethylene glycol can be

conjugated to the aptamer). Furthermore and most importantly, antidotes can be rationally designed to control their pharmacologic activity. This property is unique to aptamers and makes this class of molecules very interesting for anti-coagulant drugs (Nimjee, *et al.*, 2005a).

Human alpha thrombin

Thrombin activity is central to hemostasis, the balance between thrombosis and bleeding (Stubbs, *et al.*, 1995), and consequently this enzyme is an important target of anticoagulant therapies.

Human α -thrombin (thrombin) is a trypsin-like serine protease that has been the focus of intense study since its discovery and continues to draw attention and research funds in the biomedical field. This is in part due to the position of thrombin at the end of the blood clotting cascade, and to its unique ability to cleave fibrinogen to the polymerogenic fibrin. Thrombin plays a pivotal role in the feedback activation of coagulation factors: it participates in the cleavage of fibrinogen to form blood clots, but also initiates anticoagulation by activating protein C. The delicate equilibrium between the coagulant and anticoagulant activation by thrombin is essential in the normal physiological state, preventing clot formation in undamaged vessels and triggering the coagulation cascade in the damaged ones. Indeed, thrombin generation is closely regulated to locally achieve rapid hemostasis after injury, without causing uncontrolled systemic thrombosis. In the absence of efficient and timely thrombin generation, stable blood clots cannot form, resulting in hemorrhage. On the other hand, an excessive coagulation function results in dissemination of the clot in undamaged tissues, causing thrombosis (Di Cera, *et al.*, 1997; Coughlin, 1999; Huntington, 2005; Crawley, *et al.*, 2007; Di Cera, 2007).

Thrombin capability to perform such different functions rely on the ability to recognize a large variety of substrates, inhibitors and cofactors (Bock, *et al.*, 2007). This ability is finely regulated by two distinct regions on the surface of thrombin, known as exosites I and II, which allosterically modulate its function,

thus providing specificity to the proteolytic activity of the protein (Di Cera, *et al.*, 2007). In particular, exosite I is also known as the fibrinogen recognition site (FRE) since it is the binding site of thrombin physiological substrate fibrinogen, whereas exosite II, which is positively charged, is the binding site of thrombin inhibitor heparin (Bode, 2006). Given the ability to modulate the recognition with a large number of molecules, thrombin is usually considered a suitable target for anticoagulant and antithrombotic agents. It is thus clear that the ability of inhibiting and regulating thrombin activity *in vivo* by synthetic compounds is an important goal in the prevention and treatment of clotting abnormalities.

Thrombin inhibition can be achieved by different mechanisms (Lombardi, *et al.*, 1999). An option is blocking its active site, through the binding of substrate analogues that form covalent bonds with thrombin catalytic residues: in this way the hydrolysis of both macromolecular and small substrates is prevented. Another mechanism is blocking the binding of macromolecular substrates at exosite I or even favoring the binding of thrombin to SERPINs (SERine Protease INhibitors), which irreversibly suppress thrombin activity. The last mechanism is that by which heparin acts: it binds thrombin exosite II and drives the formation of the complex between thrombin and a SERPIN, antithrombin (Rau, *et al.*, 2007; Nutescu, *et al.*, 2008).

The X-ray crystal structure of thrombin was solved for the first time in 1992 by Bode *et al.* (Bode, *et al.*, 1992) and determined in its complex with the specific thrombin inhibitor D-Phe-Pro-Arg chloromethylketone (PPACK). PPACK binds covalently two of the three residues composing the catalytic triad (the characteristic serine-histidine-aspartate catalytic triad), thus preventing thrombin autolysis.

**CHAPTER I – HD1: YOU NEVER FORGET THE FIRST
LOVE**

INTRODUCTION

The most extensively studied thrombin binding aptamer is a 15-mer DNA oligonucleotide with sequence 5'-GGTTGGTGTGGTTGG-3', named HD1. It has been discovered in 1992 by *in vitro* selection (Bock, *et al.*, 1992) and it is known to bind at exosite I (Wu, *et al.*, 1992; Paborsky, *et al.*, 1993; Tsiang, *et al.*, 1995), inhibiting thrombin-catalyzed fibrin clot formation (Griffin, *et al.*, 1993; Li, *et al.*, 1994). HD1 adopts a monomolecular antiparallel G-quadruplex structure with two G-tetrads, built around a centrally located monovalent cation (usually K⁺), and three edge-wise loops: a TGT loop protruding in the solvent from one side of the G-quadruplex and two TT loops from the opposite side (Macaya, *et al.*, 1993). The π - π stacking of aromatic bases of the guanines plays an important role in the stability of the G-core, in which the two G-tetrads, assembled via the Hoogsteen-type hydrogen bond pattern, stack on top of each other to form a quadruple helix segment (Figure 3) (Burge, *et al.*, 2006).

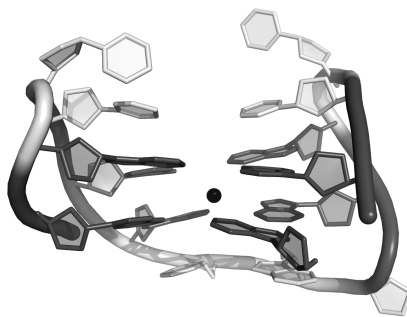


Figure 3. Cartoon representation of the G-quadruplex structure of HD1 (PDB entry 4dih). The eight guanines involved in the two G-tetrads are colored in dark grey, and the loop residues in light grey. The black sphere represents the cation stabilizing the G-core.

Several studies aimed at modulating thrombin function through modification of the aptamer TT loops have also been reported (Heckel, *et al.*, 2005; Bonifacio, *et al.*, 2008; Nagatoishi, *et al.*, 2011). Up to now, only very few modifications have led to an improvement of the anticoagulant activity

(Musumeci, *et al.*, 2012). Recently, it has been shown that mutation of TT loops by the replacement of a thymine residue with adenine or with a dSpacer decreases the thermal stability and the binding affinity of HD1 for bovine thrombin (Nagatoishi, *et al.*, 2012). This result was partly ascribed to a modification near the deletion site of the hydration pattern that mediates the recognition between the modified aptamer and the protein (Nagatoishi, *et al.*, 2012).

Intramolecular G-quadruplex scaffolds can be easily modified to increase their stability and nuclease resistance and/or to improve the binding with a specific target molecule (Borbone, *et al.*, 2012; Sagi, 2013). These features have been extensively exploited to enhance antithrombotic activity, to tailor the pharmacokinetic profile, to modulate renal clearance and the rate of tissue distribution (Bunka, *et al.*, 2006). The effects of the explored modifications can be more easily and clearly interpreted on the basis of well-defined structural models of the enzyme-aptamer interaction site. Recently, the structural details of thrombin complexed with HD1 and mTBA, an analogue that contains a 5'-5' polarity inversion site between T3 and T4, have been unambiguously established at high resolution (Russo Krauss, *et al.*, 2011; Russo Krauss, *et al.*, 2012). These crystallographic models have clearly highlighted the role of the aptamer TT loops performed in the recognition process. In particular, the two loops act as a pincer-like system that embraces the protruding region of thrombin exosite I.

In order to better clarify the contribution of each TT loop to the binding, in my PhD I have focused my work on two complexes between thrombin and two HD1 mutants. In these mutants, the nucleobase of either thymine 3 (T3) or thymine 12 (T12) is substituted with an abasic furan (dSpacer), from now on referred to as dS. Both variants, hereafter denoted as HD1 Δ T3 and HD1 Δ T12, show a decreased affinity toward the target protein with respect to the unmodified HD1 (Nagatoishi, *et al.*, 2012).

RESULTS

Overall survey of the two complexes

Structural features of the complex between thrombin and HD1 Δ T12, and HD1 Δ T3, have been investigated by X-ray diffraction (Figure 4).

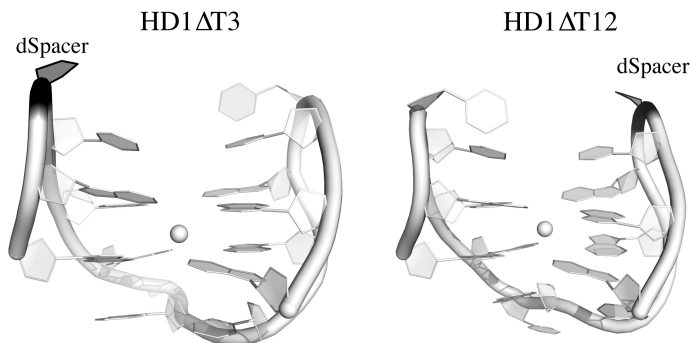


Figure 4. Deletion mutants of HD1 represented as cartoon. The dSpacer is colored in black in both oligonucleotides.

Well-ordered crystals, diffracting to 1.65 Å, were grown for the complex with HD1 Δ T12. Crystals are triclinic with one complex in the unit cell. The refined atomic model converged to R/Rfree values of 0.155/0.190.

The complex with HD1 Δ T3 produces less ordered crystals, whose diffraction pattern fades out at 2.55 Å. The asymmetric unit of the monoclinic P2₁ unit cell contains two molecules of the complex, related by a *pseudo*-two-fold symmetry axis parallel to **a**. At convergence the R/Rfree values were 0.168/0.227, respectively. The two structural models are shown in Figure 5.

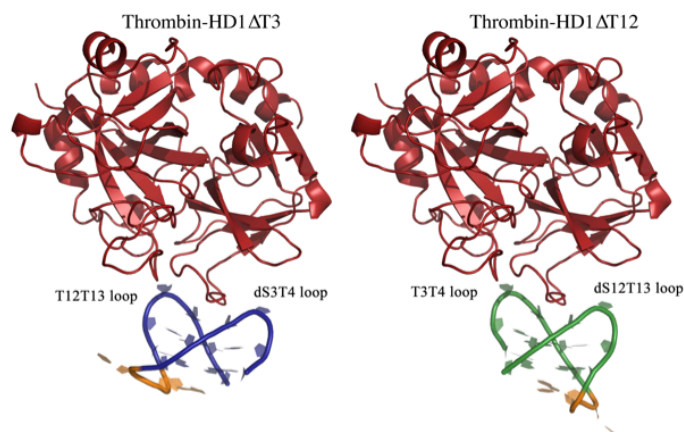


Figure 5. Overall structures of thrombin–HD1 Δ T3 and thrombin–HD1 Δ T12 complexes (PDB entry 4lz1 and 4lz4).

Focusing on thrombin

The aptamer binding does not substantially modify the thrombin structure. No significant variations are found in the protein structure when the structural models of thrombin-HD1 Δ T12 and thrombin-HD1 Δ T3 complexes are compared (RMSD of 0.90 Å, calculated on all atoms). Similar results are obtained when the models of thrombin-HD1 Δ T12 and thrombin-HD1 Δ T3 are compared with that of thrombin-HD1 (PDB entry 4dih) complex (RMSD of 0.70 Å and 0.96 Å for thrombin-HD1 Δ T12 and thrombin-HD1 Δ T3 complexes, respectively).

Furthermore, while for the structure of thrombin-HD1 Δ T12 no electron density is observed in the region of thrombin γ -autolysis loop (residues 147–151 of the heavy chain), as usually found in thrombin crystal structure, this region is well-defined in thrombin-HD1 Δ T3 complex. In the latter crystal organization, the loop conformation is stabilized by the interactions between the two molecules in the asymmetric unit. After superposition, the root mean square deviation (RMSD) between all the atoms of the two thrombin-HD1 Δ T3 complexes in the asymmetric unit was 0.62 Å (0.37 Å when only the protein atoms were considered).

Analyzing the aptamer moieties

Regarding the oligonucleotide units, only small changes in HD1 Δ T12 and HD1 Δ T3 overall fold are observed. Both the aptamers retain the expected quadruplex chair-like structure stabilized by a potassium ion localized at the center of the two G-quartets. The cation guarantees the formation of coordination bonds with the O6 atoms of the eight-purine according to a distorted anti-prism geometry (Figure 6). Thus, it can be stated that the nucleobase deletion, either of T12 or T3, does not alter the overall G-quadruplex architecture, which is assured by the π -stacking of the two bases T4 and T13 with the guanine tetrads.

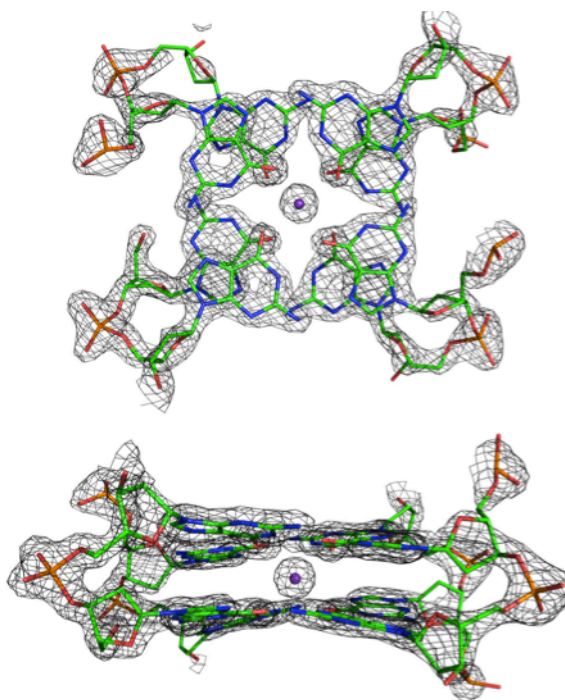


Figure 6. 2Fo-Fc electron density maps at 2.0 σ level of the G-tetrads (thrombin-HD1 Δ T12 complex) in two different orientations with the potassium ion at the center. Carbon atoms are green, nitrogen blue, oxygen red, phosphorus orange and potassium purple.

Differences between the complexes

The main structural differences between the two complexes arise from the orientation of the bound aptamer with respect to the protein (Figure 5). Indeed, after superposition of thrombin molecules, the two aptamers are related by a 180° rotation around the axis normal to the quadruplex moiety and passing through the cation (quadruplex axis). The RMSD is of about 0.4 Å, calculated excluding the TGT loop. In particular, in the complex of HD1ΔT12, the T3T4 loop interacts with residues Arg75, Glu77, Arg77A, Asn78 and Ile79 of thrombin, hereafter referred to as A-region, and the dS12T13 loop makes contacts with Arg75 and Tyr76 (B-region) (Figure 7). On the contrary, in the complex of HD1ΔT3, the T12T13 loop interacts with the A-region and the dS3T4 loop with the B-region. Therefore, in both cases, the unmodified TT loop interacts with the A-region while the base-deficient dST loop interacts with the B-region. Arg75 represents the boundary residue between the two zones; it interacts with both TT loops through contacts with the two stacked bases T4 and T13 (Figure 7).

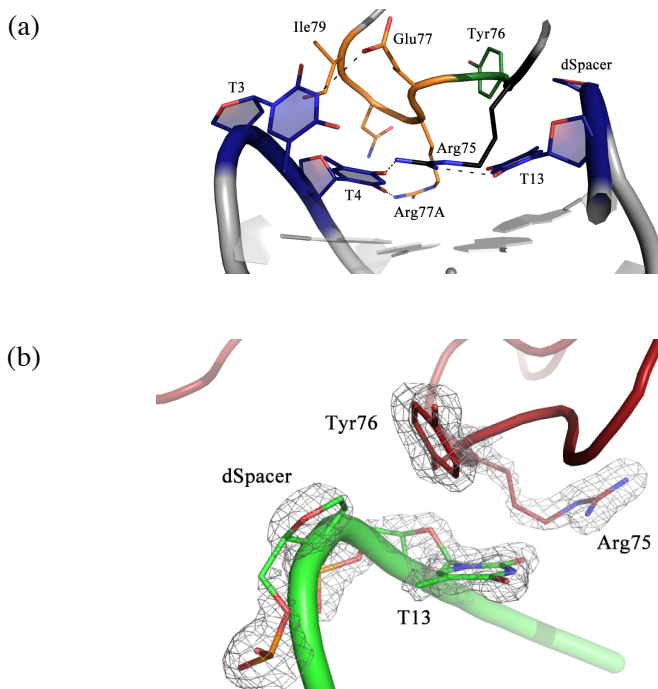


Figure 7. Thrombin–aptamer recognition site in thrombin–HD1 Δ T12. (A) A-region of exosite I interacts with the TT loop (on the left) and is colored orange, while B-region interacts with the modified dST loop and is colored green. Arg75, drawn in black, interacts with both loops and marks the transition from the A- to the B-region. (B) Electron density of the dST loop and Tyr76 and Arg75 of the B-region contoured at 2.0 σ level. Oxygen atoms are red and nitrogen blue. Carbon atoms are colored according to the structural element they belong to (dark blue for the aptamer loops, orange for thrombin region-A, black for Arg75 and green for region-B).

The interacting surface area (thrombin/aptamer) for all the known thrombin-TBAs complexes is reported in Table 1.

Table 1. Interface area (\AA^2) between thrombin and the aptamer in several complexes reported in the PDB.

	HD1 Δ T3 (4lz4)	HD1 Δ T12 (4lz1)	HD1 in K ⁺ (4dii)
Space group	P2 ₁	P1	P1
Whole interface	524	529	540
Thrombin A-region contribution	294 (T12T13)	295 (T3T4)	282 (T3T4)
Thrombin B-region contribution	157 (dS3T4)	163 (dS12T13)	184 (T12T13)

	HD1 in Na ⁺ (4dih)	mTBA (3qlp)	HD1 (1hao)
Space group	P1	I222	P2 ₁ 2 ₁ 2 ₁
Whole interface	563	663	544
Thrombin A-region contribution	295 (T3T4)	306 (T12T13)	281 (T12T13)
Thrombin B-region contribution	195 (T12T13)	280 (T35'-5'T4)	183 (T3T4)

The relative contribution of the A- and B-region is highlighted. The data indicate that the binding of a TT loop at the A-region of the protein generally engages a higher contact area compared to the B-region. The binding to the A-region is mostly contributed by a crevice on the protein surface lined by the hydrophobic side chains of Ile24, Ile79 and the phenyl ring of Tyr117, where the sugar moiety of T3 (in HD1 Δ T12) or T12 (in HD1 Δ T3) is located (Figure 8). In the more external part of the cavity the nucleobase of the residue is tightly packed among the side chains of His71, Ser72, Arg75, and Glu77, where the base forms hydrogen bonds with OE1 of Glu77 and NH1 of Arg75 (Figure 8).

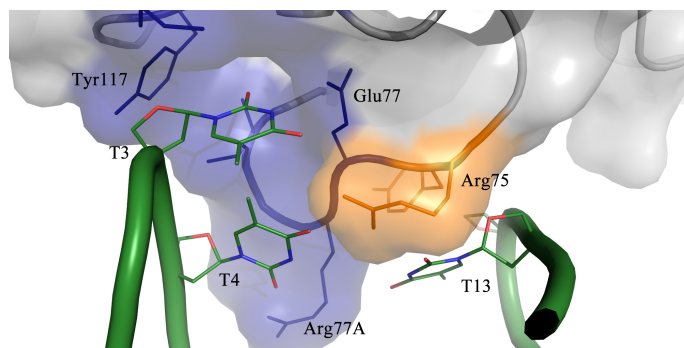


Figure 8. Thrombin crevice hosting the TT loop in thrombin–HD1 Δ T12. Residues lining the crevice are colored blue. Arg75, which is external to the crevice, is colored orange. For the residues T3, T4 and T12, carbon atoms are green, nitrogen blue and oxygen red.

The cavity is capped by the backbone of the aptamer, which forms an additional hydrogen bond with the hydroxyl group of Tyr117. These features make the crevice a highly stable and specific anchoring region for a thymine residue on the protein surface. The B-region is considerably less specific for the second TT loop of the aptamer. In particular, in the structure of the thrombin-HD1 complex (Russo Krauss, *et al.*, 2012), the interactions involve the side chains of Arg75 (see above) and Tyr76 that makes hydrogen bonds with T13 and hydrophobic contacts with the T12 nucleobase. No structural modification in the protein is observed as a consequence of the loss of the interaction between Tyr76 and the missing nucleobase (T3 or T12 for HD1 Δ T3 and HD1 Δ T12, respectively). However, the aptamer undergoes to a local conformational modification involving the sugar ring of the dS residue (in both HD1 Δ T3 and HD1 Δ T12) that gets slightly closer to Tyr76 with respect to the corresponding T12 in the thrombin-HD1 structure. This rearrangement gives rise to hydrophobic interactions between the ribose and Tyr76, which partly compensate the interaction lost with the nucleobase deletion.

As far as the TGT loop is concerned, in neither complexes the electron density is well-defined. Moreover, its conformation depends on packing interac-

tions and, thus, from the crystal space group, which is, by the way, different in the two cases.

The high quality of the diffraction data measured for thrombin-HD1 Δ T12 complex has provided a detailed picture of the bound water molecules at the interface between the TT loops and thrombin; the resulting hydration pattern does not show any major difference with that found in the high resolution structures of thrombin-HD1 complexes (PDB entries 4dih and 4dii). Unfortunately, the lower resolution of the thrombin-HD1 Δ T3 diffraction data leads to less clear-cut description of the solvent molecules, making ambiguous a detailed comparison of the hydration pattern at protein-aptamer recognition site.

DISCUSSION

G-quadruplex scaffold and thrombin exosite I

The versatility of G-quadruplex aptamers as ligands is currently used in several topic research areas (Davis, 2004). In order to improve their performance, it is critical to understand the relative contributions of the quadruplex plasticity and the specificity of the protein-ligand interactions in the recognition process.

This work has dissected the contribution of each TT loop in the recognition process between thrombin and HD1 by a crystallographic analysis of the interaction between two aptamer variants lacking a thymine nucleobase and the human enzyme. The structures of thrombin-HD1 Δ T12 and thrombin-HD1 Δ T3 provide new information about the specific contribution to the binding of the two regions in which exosite I, involved in the interaction with the TT loops, is split. It should be stressed that, with the only exception of the TGT loop (that does not interact with thrombin), HD1 possesses an almost exact two-fold axis coincident with the quadruplex axis. Therefore, in the unmodified HD1 two equivalent modes of binding, which differ only in the orientation of the TGT loop, are allowed. These two modes of binding have been found in two different

crystal structures of thrombin-HD1 complex (PDB entries 1hao and 4dih). In particular, the A-region of exosite I interacts with T12T13 in 1hao and with T3T4 in 4dih, and *vice versa* for the B-region (Figure 9).



Figure 9. Interaction between thrombin and HD1. The two structures of thrombin-HD1 complex deposited in the PDB as 1hao and 4dih are shown as cartoon, after superposition of the thrombin molecule. The two oligonucleotides (light blue for 1hao and dark blue for 4dih) differ in the position of their TGT loops (in orange).

Any chemical modification of one TT loop removes the degeneracy of the two binding modes, as it happens for example in the case of mTBA (Russo Krauss, *et al.*, 2011), a variant of HD1 in which a 5'-5' inversion of polarity site has been introduced between T3 and T4 (Martino, *et al.*, 2006). Interestingly, in the structures of thrombin-HD1 Δ T12 and thrombin-HD1 Δ T3 the aptamer binds at exosite I in such a way to preserve the interactions of the unmodified TT loop with the A-region, whereas the modified dST loop interacts with the B-region. Therefore, the overall interactions of the two modified aptamers with thrombin are practically the same, in good agreement with the values of the binding constants recently reported in the literature (Nagatoishi, *et al.*, 2012).

CONCLUSIONS AND FUTURE PERSPECTIVES

Aptamers targeting exosite I

HD1 targets thrombin acting as a pincer-like system with the two TT loops embracing the protruding region of exosite I. The identity of the two TT loops gives rise to two practically equivalent binding modes to thrombin, which differ only for the dissymmetry element TGT, exposed to the solvent and located far away from the protein. In solution the two different complexes are likely to be equally populated and their selection in the solid state is likely determined by the steric hindrance of the TGT loop and, therefore, by the space group of the crystal (Padmanabhan, *et al.*, 1996; Russo Krauss, *et al.*, 2012). Interestingly, modifications that lower the symmetry of the two TT loops (mTBA, HD1 Δ T3, HD1 Δ T12) differentiates the energetics of the two modes of binding, thus presumably leading to a different population of the two species in solution. In all these cases, in the solid state the unmodified TT loop (T3T4 in HD1 Δ T12; T12T13 in mTBA and in HD1 Δ T3) targets the same thrombin region, referred to as A-region. This observation indicates that the difference in the binding affinity between the A- and B-regions towards the intact TT loop is more important than the packing interactions in which the TGT loop is involved. Therefore, it is reasonable to conclude that the binding mode found in the crystal structure of the complexes between thrombin and the aptamer mutants is representative of that adopted by the most abundant species in solution. From these data it emerges the existence of a sort of hierarchy among the interactions that govern the recognition process between thrombin and HD1 and that the binding could be improved by increasing the contacts of the nucleotide with the B-region. In this perspective, the presence of two lysine side chains (Lys81 and Lys110) at a distance of about 10 Å from the TT loop interacting with the B-region could be used to create new anchorage points of the aptamer on protein

surface. With this aim, proper modifications of T3 (or equivalently of T12) are under design.

Improving aptamers by chemical modification

Among the few chemical modifications that have led to an improvement of the binding properties of HD1, it is particularly interesting the replacement of nucleotide G8 with the modified nucleotide (R)-1-(4-(1-pyrenylethynyl) phenylmethyl) glycerol (a.k.a. TINA), called HD1-r8p (Figure 10).

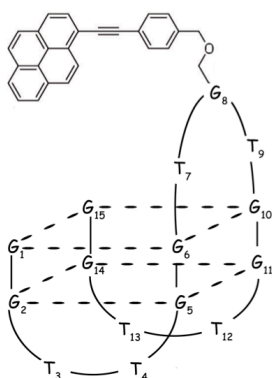


Figure 10. Schematic representation of the modified HD1, HD1-r8p.

The modification strongly improves the anticoagulant activity, affinity and thermal stability of HD1 (Rohrbach, *et al.*, 2012). In the last months, I have been trying to determine the structure of HD1-r8p in complex with thrombin in order to establish the cause of such higher affinity with respect to its parent aptamer. HD1-r8p has been provided by Prof. Gunter Mayer (University of Bonn, Germany) and Prof. Erik Pedersen (University of Southern Denmark, Denmark), with whom I started a collaboration.

MATERIALS AND METHODS

Sample preparation

Human D-Phe-Pro-Arg-chloromethylketone (PPACK) inhibited α -thrombin was purchased from Haemtech and the initial buffer was changed to 0.75 M NaCl using Centricon mini-concentrator (Vivaspin 500) and a refrigerated centrifuge (Z216MK, Hermle Labortechnik).

The two aptamers, lacking T3 (HD1 Δ T3) and T12 (HD1 Δ T12) nucleobase respectively, were synthesized by Nagatoishi *et al.* as previously described (Nagatoishi, *et al.*, 2012). Each aptamer was dissolved in 10 mM potassium phosphate buffer pH 7.1 up to 2 mM concentration. Aptamer solutions were heated for 10 minutes at 85°C, slowly cooled down, and stored at 4°C overnight. This annealing procedure is essential to induce the correct fold of the aptamer.

A standard protocol has been followed for the preparation of the complex between thrombin and each aptamer (Russo Krauss, *et al.*, 2012). A two-fold molar excess of the annealed aptamer solution was laid on a frozen sample of thrombin and the sample left for three hours at 4°C. After complete melting, the solution was diluted and the buffer changed to 25 mM potassium phosphate buffer pH 7.1, 100 mM KCl. The solution was then extensively washed, in order to take off the excess of aptamer and finally concentrated to about 0.2 mM using Centricon mini-concentrator and a refrigerated centrifuge.

Crystallization and data collection

Initial crystallization trials, involving screening around the conditions previously used for crystal growth of thrombin complexes with HD1 (Russo Krauss, *et al.*, 2012) or mTBA (Russo Krauss, *et al.*, 2011), were unsuccessful. Subsequent sitting-drop crystallization experiments were set up, for all the three complexes prepared, at 20°C in 96-well plates (Greiner Bio-One, USA) using an

Automated Protein Crystallization Workstation (Hamilton Robotics) and precipitant solutions of commercially available crystallization screens (Hampton Research Crystal Screen 1, Crystal Screen 2 and Index). Optimization of promising conditions by fine-tuning pH, precipitant and additive concentration (Russo Krauss, *et al.*, 2010) resulted in the growth of good diffraction-quality crystals of the complexes.

In particular, thrombin-HD1 Δ T3 crystals grew from 60% (v/v) Tacsimate pH 7, 1-3% (w/v) MPEG 5K or PEG 8000, while those of thrombin-HD1 Δ T12 from 25% (w/v) PEG 3350, 0.1 M Bis/Tris pH 5.8, 0.2 M ammonium acetate. In both cases concentration of the complex was about 0.2 mM. Optimization of crystallization conditions produced well-diffracting crystals using the hanging drop vapor diffusion method, at a protein-to-reservoir solution ratio of 1:1 in 1 μ L drops and using a protein concentration of about 0.2 mM.

After the addition of 35% glycerol to the harvesting solution, crystals were flash-cooled at 100 K in super-cooled N₂ gas produced by an Oxford Cryosystem and maintained at 100 K during data collection. Diffraction data were collected by using synchrotron light ($\lambda = 1.2651$ Å) and a Dectris Pilatus 2M detector on the XRD1 beamline 5.2 at Elettra Synchrotron Radiation Facility (Trieste, Italy). Data were indexed, processed and scaled with HKL2000 (Otwinowski, *et al.*, 1997).

Thrombin-HD1 Δ T3 crystallizes in the monoclinic space group P2₁, with two molecules of the complex in the asymmetric unit, whereas thrombin-HD1 Δ T12 crystals belong to the triclinic space group. Detailed data collection statistics are reported in Table 2.

Structure determination and refinement

Initial phases were determined by molecular replacement method using Phaser (McCoy, *et al.*, 2007), using the whole thrombin-HD1 complex (PDB entry 4DII) as search model. Crystallographic refinement was performed using CNS (Brunger, *et al.*, 1998) and REFMAC5 (Vagin, *et al.*, 2004), while manual

model building using Coot (Emsley, *et al.*, 2010). After the inclusion of low-resolution data and bulk solvent correction, the final R/Rfree were 0.155/0.190, for thrombin–HD1 Δ T12, and 0.168/0.227, for thrombin–HD1 Δ T3. For the latter structure, in the last steps of refinement the restraints arising from the non-crystallographic symmetry were added. The protein geometry of the refined structures was monitored using PROCHECK (Laskowski, *et al.*, 1993). Statistics and parameters of the refinement are given in Table 2. The drawings were prepared with Pymol (Schrodinger, 2010). The coordinates of the structures were deposited in the Protein Data Bank with codes 4lz1 and 4lz4.

Table 2. Data collection and refinement statistics. Values in parentheses refer to the last resolution shell.

Data collection	Thrombin– TBAΔT12	Thrombin– TBAΔT3
Space group	P1	P2 ₁
Unit-cell parameters (Å, °)	<i>a</i> = 42.9 <i>b</i> = 44.5 <i>c</i> = 52.6 α = 65.2 β = 82.5 γ = 63.7	<i>a</i> = 61.7 <i>b</i> = 120.7 <i>c</i> = 67.2 α = 90.0 β = 94.2 γ = 90.0
Resolution range (Å)	30.0–1.65 (1.68–1.65)	50.0–2.55 (2.63–2.55)
Measured reflections	125952	52191
Unique reflections	36813 (1727)	29590 (2446)
<i>R</i> _{merge} (%)	5.3 (48.0)	13.2 (39.7)
Mean <i>I</i> / σ (<i>I</i>)	25.5 (2.4)	7.3 (2.3)
Completeness (%)	96.7 (93.9)	94.4 (95.4)
Multiplicity	3.4 (2.8)	1.8 (1.8)
Complexes per asymmetric unit	1	2
Refinement statistics		
Resolution limits (Å)	25.2–1.65 (1.69–1.65)	29.6–2.56 (2.63–2.56)
<i>R</i> factor (%)	15.5 (21.8)	16.8 (23.9)
<i>R</i> _{free} (%)	19.0 (24.2)	22.7 (26.3)
Number of reflections used in the refinement	34943 (2265)	27048 (1664)
Number of reflections used for <i>R</i> _{free} calculation	1820 (141)	1430 (73)
No. of protein atoms	2294	4713
No. of aptamer atoms	306	612
No. of ions	2	4
No. of water molecules	229	219
Average <i>B</i>	27.0	37.2
Protein, overall (Å ²)	24.4	35.7
Aptamer atoms (Å ²)	40.3	50.4
Solvent atoms (Å ²)	34.6	32.2
rmsd bond lengths (Å)	0.022	0.014
rmsd bond angles (°)	2.1	1.7
Ramachandran statistics		
Favored region (%)	97.4	97.2
Allowed region (%)	2.6	2.8
Outlier region (%)	0.0	0.0

**CHAPTER II – HD22: WHEN A G-QUADRUPLEX IS
NOT ENOUGH**

INTRODUCTION

Polynucleotide backbone is characterized by an intrinsic large flexibility that allows the chain to reach a variety of three-dimensional shapes. In addition to the nucleobase sequence, the selection of a specific form is also dependent on the environmental parameters in which the nucleotide is embedded, such as pH, ionic strength, types of ions, temperature, etc. For this characteristics, the nucleotide scaffold can be easily designed to specifically interact with a variety of binding pockets on large receptors or even to bind and transport small metabolites. Cooperative effect resulting from internucleobase interactions, however, are particularly important in determining the stabilization of a sequence of specific conformations along the chain that produces a particular motif such as the duplex motif, or the variety of quadruplex motifs, or even a more complex combination of these motifs. While detailed studies of RNA and DNA aptamers with a single structural motif have been performed, up to now only very little is known about the structure of aptamers in which more than one motif can be found.

Several examples of aptamers that are supposed to adopt a mixed duplex/quadruplex structure have been reported in the literature (Tasset, *et al.*, 1997; Chinnapen, *et al.*, 2002; Pileur, *et al.*, 2003; Mori, *et al.*, 2004; Chou, *et al.*, 2005; Gatto, *et al.*, 2009; Li, *et al.*, 2009). In all the above mentioned cases the presence of both structural motifs is required for strong and specific binding of the target. Both duplex and quadruplex are characterized by a large difference in shape and electrostatic potential. For example, with respect to a double helix, a G-quadruplex exhibits a substantially higher negative charge density, that have implications on the specificity in target recognition. To date, no crystallographic data exist on bimodular aptamers (i.e. aptamers that have a fold in which more than a motif coexist) or on their complexes with targets. Their interest lies in the fact that they exhibit efficient recognition for a variety of targets related to severe pathologies or disorders, such as AIDS (Pileur, *et al.*,

2003), rheumatoid arthritis (Mori, *et al.*, 2004), hepatitis (Jones, *et al.*, 2006), and thrombosis (Tasset, *et al.*, 1997; Russo Krauss, *et al.*, 2013), as well as to endogenous small molecules such as nicotinamide (Lauhon, *et al.*, 1995), thyroxin (Levesque, *et al.*, 2007), and hemin (Chinnapen, *et al.*, 2002). Only recently the structure of an RNA molecule with a mixed duplex/quadruplex fold in complex with a peptide from the human fragile X mental retardation protein has been solved by NMR (Phan, *et al.*, 2011).

One of the most interesting results of SELEX protocols directed towards human α -thrombin (thrombin) is the identification of potent second-generation aptamers that are supposed to adopt a mixed duplex/quadruplex conformation (Tasset, *et al.*, 1997). In particular this family of aptamers includes a 29-mer, named HD22[29] (5'-AGTCCGTGGTAGGGCAGGTTGGGGTGACT-3'), and a 27-mer (HD22[27]), which lacks the first and the last residue with respect to the 29-mer. A schematic model of HD22[27] is shown in Figure 11.

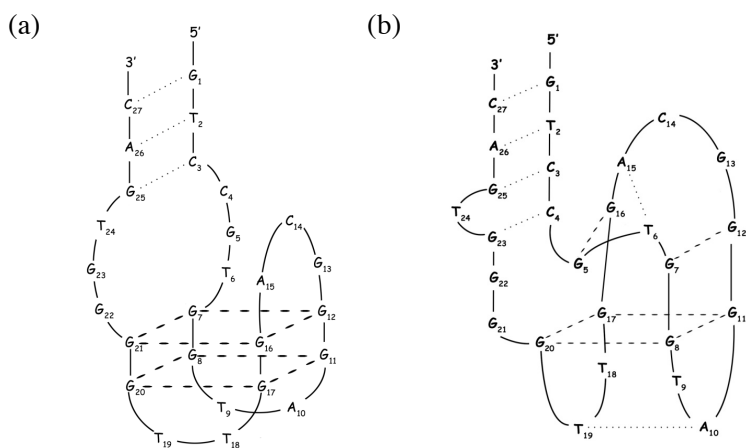


Figure 11. Schematic representations of the HD22[27] organization (a) as proposed by (Tasset, *et al.*, 1997) and (b) as found in the crystal structure of the complex with thrombin. Hydrogen bonds between bases are represented by dotted lines (dashed lines represent G-G interactions).

Both aptamers embody a 15-residue sequence very similar to that of the best characterized thrombin aptamer HD1 (5'-GGTTGGTGTGGTTGG-3') and bind

thrombin with higher affinity ($K_d \approx 0.7$ nM) (Tasset, *et al.*, 1997) compared to HD1 ($K_d \approx 100$ nM) (Macaya, *et al.*, 1993). Several evidences indicate that HD22 aptamers recognize exosite II rather than exosite I (Tasset, *et al.*, 1997). However, high resolution structural data on these aptamers and their complexes with thrombin are still lacking.

Throughout this PhD project only the HD22[27] has been used. In particular, this work has been focused on the structural analysis of the thrombin-HD22[27] complex both in solution and in the crystalline state. Furthermore, a solution study of the free aptamer has been performed by spectroscopic techniques. The X-ray model of the thrombin-HD22[27] complex reveals interesting new features regarding the overall fold of the aptamer and the way it interacts with its target.

RESULTS

Crystal structure of thrombin-HD22[27] complex

Best crystals of the complex between thrombin and HD22[27] diffract X-rays to 2.4 Å resolution and belong to monoclinic space group $P2_1$. Electron density maps calculated from initial phases obtained by molecular replacement, using thrombin as search model, allowed unambiguous building of the whole aptamer. For what concerns thrombin, the heavy chain (residues 16-245) and the light chain (residues 1B-14K) were well-defined in the electron density map, with the exception of the γ -autolysis loop (residues 146-150) of the heavy chain. Crystallographic analysis shows that HD22[27] aptamer binds the thrombin molecule in the exosite II region forming a 1:1 complex (Figure 12).

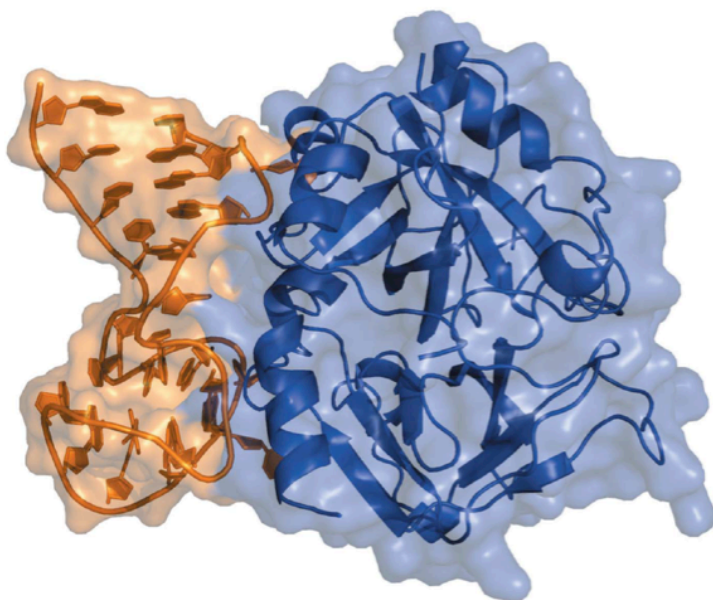


Figure 12. Overall structure of the complex between thrombin and HD22[27]. Surface representation of the complex, with thrombin colored blue and HD22[27] colored orange.

Interestingly, the ligand does not adopt a simple duplex or quadruplex fold, but it shows a mixed structure (Figure 11b). The crystal structure was refined to R/Rfree values of 0.181/0.249.

A novel duplex-quadruplex motif

The X-ray analysis of HD22[27] in complex with thrombin reveals that the oligonucleotide fold (Figure 1b and Figure 3a) is considerably different from the proposed and widely accepted model (Tasset, *et al.*, 1997; Spiridonova, *et al.*, 2003; Hasegawa, *et al.*, 2008; Müller, *et al.*, 2008; Mayer, *et al.*, 2009; Zhou, *et al.*, 2010; Marson, *et al.*, 2012) reported in Figure 1a.

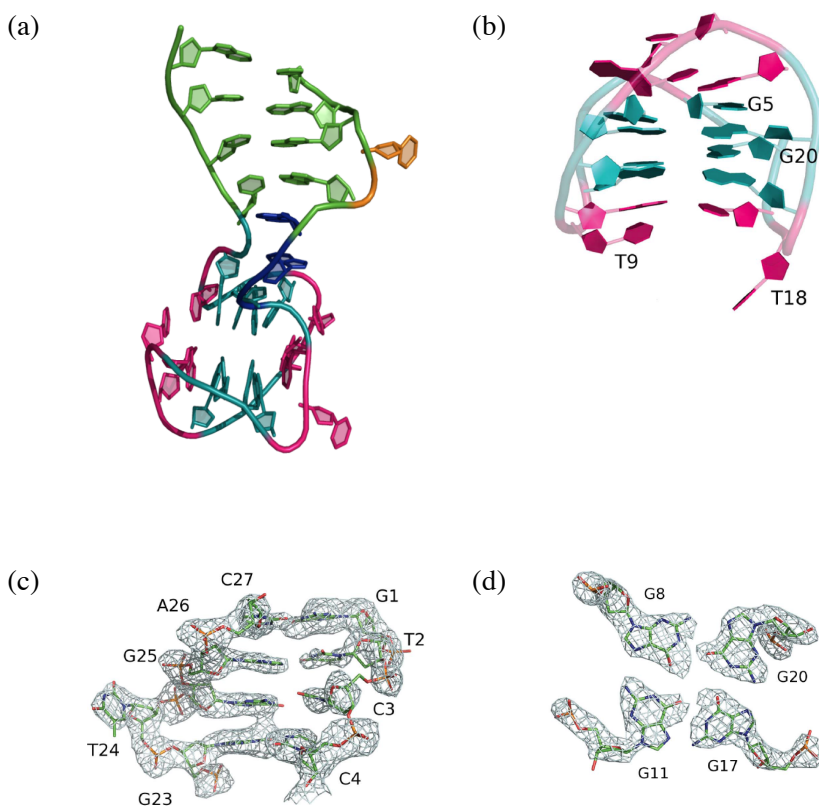


Figure 13. Crystal structure of the HD22[27] aptamer. (a) Cartoon representation: the duplex region is colored green, the G-core cyan, the loops pink, the connecting residues blue and the bulged-out thymine orange. (b) A different view of the *pseudo*-G-quadruplex architecture. Fo-Fc unbiased OMIT simulated-annealing maps contoured at the 3.0σ level are shown for the duplex motif (c), and for the G-tetrad (d).

As expected, residues 1-3 and 25-27 form the duplex motif; however, with respect to the proposed model, the crystal structure contains the additional C4-G23 Watson-Crick base pair (Figure 13a). This is favored by the bulging out of T24, which is inserted into a pocket on the thrombin surface (see below). The duplex structure is followed by sixteen residues (sequence number 5-20) organized in a G-tetrad core capped by the two-residue loops T9-A10 and T18-T19 on one side, and the three-residue loop G13-C14-A15 on the opposite side (Figure 13b). Electron density maps corresponding to the two structural motifs,

duplex and quadruplex, are shown in panels c, d, and e of Figure 13. It has to be noted that on the basis of the sequence alignment between HD22[27] and HD1 the G-quadruplex arrangement was expected to involve the fifteen residues 7-21. The shape of the HD22[27] G-tetrad core is roughly similar to that of HD1, with the extra residue T6 forming an additional one-residue loop adjacent to the three-residue loop (Figure 11b and Figure 13b). However, the chain topology is markedly different. In the core region the four guanines G8, G11, G17, G20 form a canonical G-tetrad (Figure 13d and Figure 14a): it presents the classical alternation of *syn* and *anti* conformations of the guanine residues (with G8 and G17 adopting the *anti* conformation and G11 and G20 adopting the *syn* conformation) and the nucleobases are only slightly tilted out from the average plane of the tetrad. The remaining tetrad (Figure 14b) begins with G5, two residue upstream from the corresponding tetrad in the Tasset model (Figure 11). is located two residues upstream with respect to the putative HD1 sequence. This tetrad does not possess the canonical *anti-syn-anti-syn* but an *anti-syn-anti-anti* pattern of glycosidic angles (only G12 adopts a *syn* conformation). Moreover, the presence of a short one-residue loop slightly pushes apart the backbone atoms of G5 and G7 causing the break of the Hoogsteen hydrogen bond cyclic pattern. Thus, the four guanine residues G5, G7, G12 and G16 can be better described as forming two pairs: in one pair G5 and G16 are linked through a reversed Hoogsteen N2-N7 hydrogen bond, whereas in the second pair only a loose polar contact between O6 (G12) and N7 (G7) is present (Figure 14b). Despite these modifications, an almost perfect stacking of G5 and G20 (Figure 14c) and of G16 and G17 (Figure 14d) is observed. The latter pair also makes a good stacking interaction with Ade15 belonging to the three-residue loop. This unusual four guanine set has been named *pseudo-G-tetrad* and the overall organization of residues 5-20 *pseudo-G-quadruplex*. Interestingly, base pairing is also observed for residues of the loops. In particular, T6-A15 form a *trans* Watson-Crick/Hoogsteen pair (Figure 14e), and A10-T19 a *trans* Watson-

Crick/Watson-Crick pair (Figure 14f). The sharp transition from the duplex motif to the *pseudo*-G-quadruplex one produces a somewhat strained kink of the molecule (Figure 3a). Indeed, G5, which is the first residue of the *pseudo*-G-quadruplex segment, is also linked to C4, the last residue of the duplex. Thus, G5 can be seen as the pivot residue that mediates the conversion from a duplex to a quadruplex arrangement. In the facing strand, the hinge segment between the two motifs includes two residues: G22, which does not interact with other residues of the aptamer, and G21, which forms a G-fork with G5 (Figure 14g). As result of the different number of residues involved in the loops connecting the duplex-quadruplex structures the helical axes of the two motifs form an angle of approximately 90°. Interestingly, no electron density assignable to a cation is observed in this *pseudo*-G-quadruplex region.

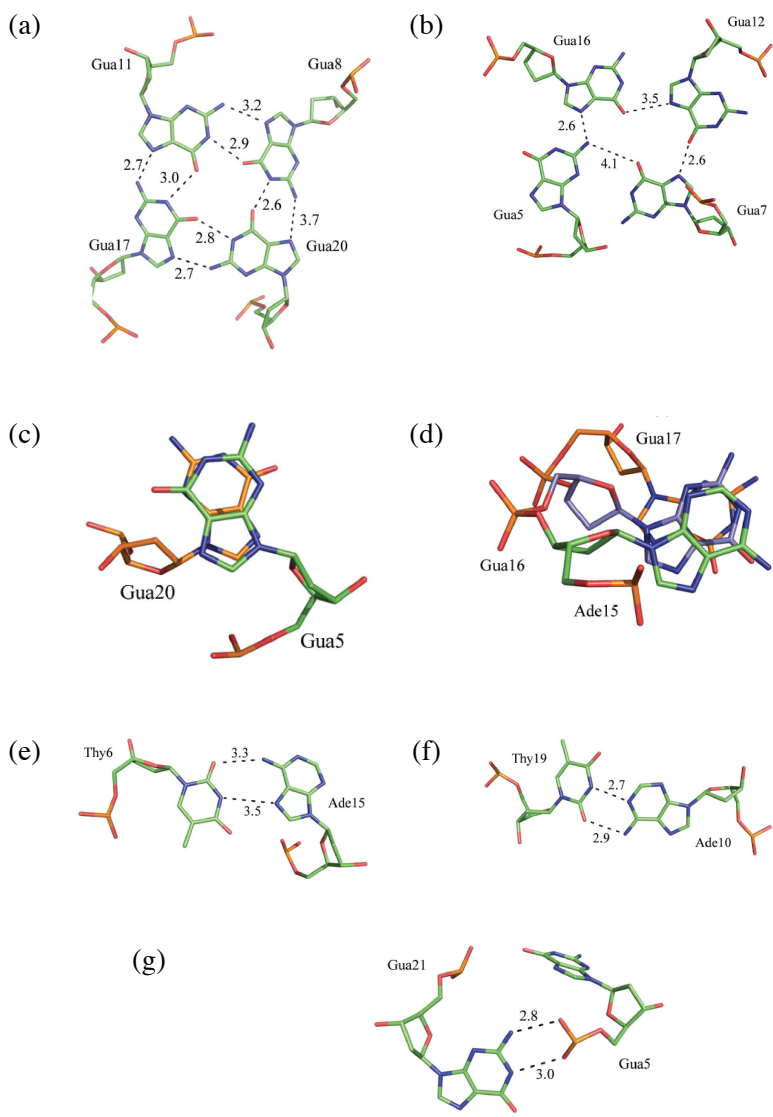


Figure 14. Snapshots of aptamer structural motifs. (a) G-tetrad, (b) *pseudo*-G-tetrad, (c) and (d) stacked bases, (e) trans Watson–Crick/Hoogsteen pair, (f) trans Watson–Crick/Watson–Crick pair and (g) G-fork. All the distances are reported in Å.

Interaction with thrombin

The unusual mixed structure of HD22[27] extensively adheres to thrombin surface (Figure 12). In particular, the aptamer binds at exosite II burying a total

of 1118 Å² of the protein accessible surface. Interaction between HD22[27] and thrombin involves numerous residues of both molecules: T9, T18, T19, G20, G23, T24, C27 and marginally G21, G22 and A26 of the aptamer, and segments 89-101, 230-245 of thrombin, with a further contribution from residues Leu130, Arg165 and Lys169. A few examples of the contacts between thrombin and HD22[27] is are presented in Figure 15.

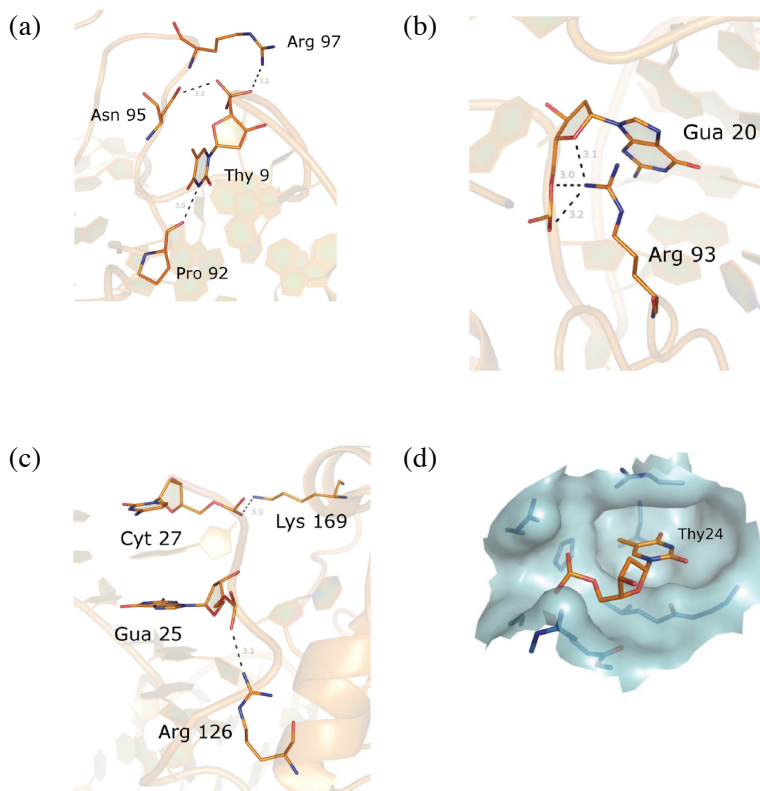


Figure 15. Some examples of the interactions between thrombin and HD22[27] focusing on T9 (a), G20 (b), G25 and C27 (c). (d) Insertion of HD22[27] T24 into a pocket on the thrombin surface. The aptamer residue is represented as sticks and the thrombin molecule as a surface. All the distances are reported in Å.

As expected on the basis of the strong electropositive potential of exosite II, many polar contacts are observed at the interface. In particular, ion pairs are formed between the phosphate backbone of the nucleotide and positive charged residues of thrombin. Hydrophobic contacts, mainly involving loop residues T9, T18 and T19 and A26, also contribute to the stability of the complex. A further anchorage is represented by T24, which bulges out from the duplex region of the nucleotide into a protein pocket (Figure 15) where it is mainly involved in polar contacts.

The thrombin-HD22[27] complex has been compared with two other thrombin-aptamer complexes: that with HD1 (PDB entry 4dih), which binds at exosite I, and that with Toggle-25t (PDB entry 3dd2), an RNA aptamer that embodies two short duplex regions (Long, *et al.*, 2008) and binds at exosite II. In particular, using Cocomaps tool (Vangone, *et al.*, 2011), the protein-aptamer interfaces have been compared (Table 3).

Table 3. Thrombin-aptamer interface features.

Structure	HD22-27mer	TBA	Toggle-25
PDB code	4i7y	4dih	3dd2
Interface area (\AA^2)	1118.5	563.2	754.5
Polar interface area (\AA^2)	515.4	287.6	279.7
Nonpolar interface area (\AA^2)	603.1	275.6	474.8
Residues at the interface			
Complex	55	22	32
Aptamer	15	8	10
Thrombin	40	14	22
Surface-complementarity index [†]	0.64	0.76	0.72

[†] See Russo Krauss *et al.* (2011).

The much higher affinity of HD22[27] (Tasset, *et al.*, 1997) with respect to HD1 (Macaya, *et al.*, 1995; Nagatoishi, *et al.*, 2011), which binds at exosite I, correlates very well with the almost doubled value of the interface area and of the number of interacting residues displayed by the HD22[27] complex. A similar good correlation is also found when the comparison is performed with Toggle-25t that recognizes exosite II with a K_d of about 3 nM (Long, *et al.*, 2008).

Circular dichroism study

Conformation and thermal stability of HD22[27] were investigated by means of circular dichroism in different experimental conditions. In particular, the role of monovalent cations on the folding of the aptamer was monitored. In the presence of sodium ions, the CD spectrum shows (Figure 16) two overlapping positive signals at ~ 288 nm and ~ 260 nm and a negative band at ~ 240 nm,

which do not correspond to a canonical G-quadruplex arrangement. A different behavior is observed for solutions containing potassium ions: the spectrum shows an intense positive peak at 292 nm, a less intense positive peak at about 250 nm, and two deep wells at 265 and 235 nm, respectively.

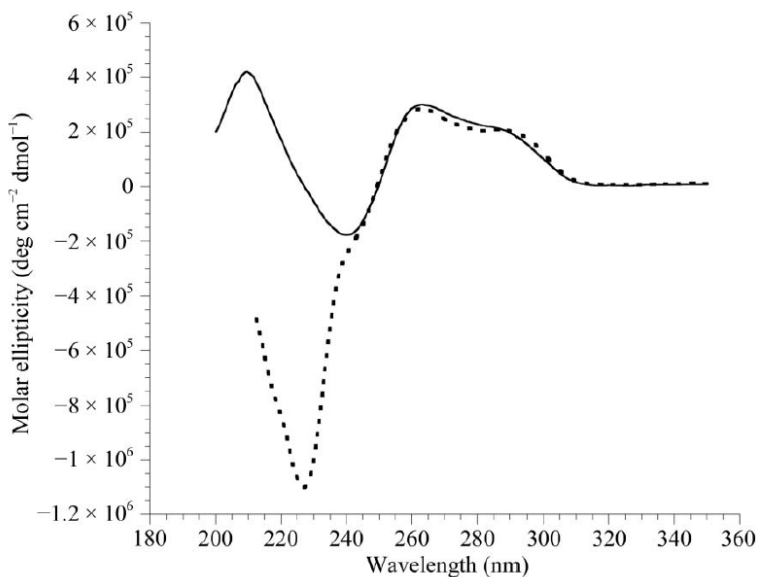


Figure 16. CD spectra of HD22[27] in the presence of sodium before (bold line) and after (dashed line) the addition of thrombin.

These spectral features are typical of antiparallel G-quadruplexes. These data suggest that the aptamer can adopt two different conformations, denoted as M (sodium containing samples) and T (potassium containing samples). The thermal stability of the two nucleotide folds has also been investigated in the 10-90°C range. The results indicate that the transition temperature (T_m) is significantly lower for the M conformer ($T_m = 36^\circ\text{C}$) with respect to the T conformer ($T_m = 47^\circ\text{C}$). The influence of thrombin binding on the aptamer structure and stability has also been analyzed. CD measurements were performed after addition of an equimolar amount of protein to the solutions of the nucleotide containing either sodium or potassium ions. The thrombin addition induces only

small variations to the spectral features of the oligonucleotide (Figure 16), but significantly increases the stability of the aptamer both in the M and in the T conformation, as indicated by their T_m values (48.0 and 57.0°C, respectively). A similar thrombin-induced stabilization effect was previously observed in the case of HD1 (Nagatoishi, *et al.*, 2011; Russo Krauss, *et al.*, 2012).

It should be noted that all the crystallization trials carried out using the aptamer in the T conformation failed.

DISCUSSION

Duplex-quadruplex motif and thrombin exosite II

The crystallographic analysis of the complex with thrombin shows that HD22[27] adopts a novel, sharply kinked, conformation, in which the helical axis of the regular duplex segment and that of the *pseudo*-G-quadruplex motif are approximately at right angle. The resulting overall shape of the molecule allows both motifs to interact with the protein (see below). The *pseudo*-G-quadruplex segment lacks the stabilizing effects produced by cyclic Hoogsteen hydrogen bonds in one of the two tetrads and by the cation binding (Williamson, 1994), but preserves the base stacking among the guanines of the core (panels c and d of Figure 14). Indeed, although the pivot residue G5 is displaced from the position required for the formation of a canonical G-tetrad, the *anti* conformation of this residue efficiently places the base in a position to build up the close packing of the eight guanine bases of the core. Moreover, hydrogen bonds between bases of loops (panels e and f of Figure 14) also contribute to the intramolecular stabilization of this fold. Thus, it may be surmised that the aptamer conformation found in the crystal structure of the complex with thrombin represents one of the low-energy conformations adopted by the free nucleotide.

Protein-aptamer interaction shows interesting new features. In the known crystallographic structures of HD1 complexes (Russo Krauss, *et al.*, 2011;

Russo Krauss, *et al.*, 2012) the formation of the complex is essentially driven by the TT loops of the quadruplex. By contrast, in the present structure the protein recognition involves an extended region of the aptamer that includes the duplex segment (G23, A26, C27), the bulged out residue T24, the *pseudo*-G-quadruplex core (G20), as well as the quadruplex connecting loops (T9, T18, T19). Despite the lack of a well-defined binding structural motif (such as the TT loops of HD1), this large contact area (1118 Å²) displays high complementarity with the thrombin surface. Indeed, the surface complementarity index (Lawrence, *et al.*, 1993) is 0.65, a value only slightly lower than the ones in the range 0.72-0.76 calculated for the complexes of HD1 family, in which the contact area is much smaller (560 Å²).

Circular dichroism studies indicated that oligonucleotides embodying duplex-quadruplex domains display a large conformational variability (Zhou, *et al.*, 2010; Lin, *et al.*, 2011; Zavyalova, *et al.*, 2011; Dolinnaya, *et al.*, 2012; Marson, *et al.*, 2012) depending on the experimental conditions, such as the nature and concentration of salts, temperature, pH, etc. In order to establish a correlation between solution and solid state structure, the crystallographic analysis has been complemented with a CD study. Comparison between CD spectra of HD22[27] free and in complex with thrombin (see methods for details) indicates that the binding does not produce significant modifications on aptamer spectral features, but enhances its thermal stability. Similar results were obtained in the case of HD1 (Nagatoishi, *et al.*, 2011; Russo Krauss, *et al.*, 2012). This finding supports the hypothesis that the crystal structure is representative of the aptamer conformation in solution.

In particular, the aptamer shows a spatial organization that differs from the proposed model and widely accepted (Tasset, *et al.*, 1997; Spiridonova, *et al.*, 2003; Hasegawa, *et al.*, 2008; Müller, *et al.*, 2008; Mayer, *et al.*, 2009; Zhou, *et al.*, 2010; Marson, *et al.*, 2012), such as the one sketched in Figure 11a that is characterized by a definite separation between the duplex and the quadruplex

regions. On the contrary, in the X-ray structure the duplex is directly enchainned to a *pseudo*-G-quadruplex topology never observed before. This structural arrangement optimizes the adhesion of HD22[27] on thrombin exosite II. Circular dichroism (CD) measurements also suggest that the crystal structure is representative of the solution structure of the aptamer in the presence of sodium.

In the absence of crystallographic data on the T conformation of the aptamer, the CD features have been interpreted on the basis of the literature data on bimodular nucleotides. The CD spectrum of the T conformation is similar to those of HD1-like aptamers, with peaks characteristic of an antiparallel G-quadruplex fold. In agreement with the current literature models, it may be speculated that this form includes a duplex segment separated by a two- three-residue linker from a canonical G-quadruplex. The latter is indeed particularly stabilized by K^+ , in line with the features of the CD spectra recorded for HD22[27] in the presence of potassium.

The existence of two HD22[27] conformers would explain also the results of the crystallization experiments. The failure of the trials carried out using HD22[27] in the T conformation may indicate that the spatial separation of the two folding motifs does not allow an efficient binding at exosite II. Moreover, the presence of a canonical G-quadruplex with exposed TT and TA loops may favor the binding of HD22[27] at exosite I, similar to that observed for HD1.

CONCLUSIONS AND FUTURE PERSPECTIVES

Aptamers targeting exosite II

In the last years several attempts to improve the biological activity of thrombin binding aptamers via chemical and structural modifications have been performed (Müller, *et al.*, 2007; Müller, *et al.*, 2008; Pasternak, *et al.*, 2011; Musumeci, *et al.*, 2012; Rangnekar, *et al.*, 2012). The present contribution based on a combined crystallographic and spectroscopic study provides the scientific

community with an enhanced structural and conformational knowledge of bimodular aptamers that will serve as a platform for a rational design of modified molecules for use in anticoagulant therapies and diagnostic applications. In particular, the structural results emphasize the role of residues controlling the spatial transition from the duplex to the quadruplex architecture and the recruitment of both motifs in thrombin binding.

On more general grounds, the behavior of HD22[27] indicates that bimodular oligonucleotides should not be thought as fixed, structured building blocks linked together. Their intrinsic structural plasticity makes them multifaceted ligands able to assume different folds depending on even small variations of environmental parameters.

Finally, the turn of the oligonucleotide chain that originate the sharp transition from the duplex to the *pseudo*-quadruplex motif may represent an intriguing example of the way in which a double helix can evolve in a duplex-quadruplex structure, such as that found in the biologically relevant regions of nucleic acids.

MATERIALS AND METHODS

Sample preparation

Human D-Phe-Pro-Arg-chloromethylketone (PPACK) inhibited thrombin was purchased from Haemtech and the initial buffer was changed to 0.75 M NaCl using Centricon mini-concentrator (Vivaspin 500) and a refrigerated centrifuge (Z216MK, Hermle Labortechnik).

The aptamer HD22[27] was purchased from Sigma-Aldrich. Stock solutions were prepared by dissolving the lyophilized oligonucleotide at a concentration of 4 mM in 10 mM sodium phosphate buffer at pH 7.1. Aptamer solutions were heated for 10 minutes at 85°C, slowly cooled down, and stored at 4°C over-

night. This annealing procedure is essential to induce the correct fold of the aptamer.

A standard protocol has been followed for the preparation of the complex between thrombin and each aptamer (Russo Krauss, *et al.*, 2012). A two-fold molar excess of the annealed aptamer solution was laid on a frozen sample of thrombin and the sample was left for three hours at 4°C. After complete melting, the solution is diluted and the buffer changed to 25 mM sodium phosphate buffer pH 7.1, 0.1 M NaCl. The solution was then extensively washed, in order to take off the excess of aptamer and finally concentrated to about 0.2 mM using Centricon mini-concentrator and a refrigerated centrifuge.

Crystallization and data collection

Initial crystallization trials, involving screening around the conditions previously used for crystal growth of thrombin complexes with HD1 or mTBA, were unsuccessful. Subsequent sitting-drop crystallization experiments were set up, for all the three complexes prepared, at 20°C in 96-well plates (Greiner Bio-One, USA) using an Automated Protein Crystallization Workstation (Hamilton Robotics) and precipitant solutions of commercially available crystallization screens (Hampton Research Crystal Screen 1, Crystal Screen 2 and Index). Optimization of promising conditions by fine-tuning pH, precipitant and additive concentration (Russo Krauss, *et al.*, 2010) resulted in the growth of good diffraction-quality crystals.

Best crystals of thrombin-HD22[27] complex grew from solutions containing 14-16% (w/v) of PEG 3350 and 0.2 M sodium citrate. Optimization of crystallization conditions produced well-diffracting crystals using the hanging drop vapor diffusion method, at a protein-to-reservoir solution ratio of 1:1 in 1 μ L drops and using a protein concentration of about 0.2 mM.

After the addition of 35% glycerol to the harvesting solution, crystals were flash-cooled at 100 K in super-cooled N₂ gas produced by an Oxford Cryosystem and maintained at 100 K during data collection. Diffraction data were col-

lected by using synchrotron light ($\lambda = 1.2651 \text{ \AA}$) and a Dectris Pilatus 2M detector on the XRD1 beamline 5.2 at Elettra Synchrotron Radiation Facility (Trieste, Italy). Data were indexed, processed and scaled with HKL2000 (Otwinowski, *et al.*, 1997).

Thrombin-HD22[27] crystals belong to the monoclinic space group $P2_1$ and diffract up to 2.4 \AA resolution. In this case, the Matthews coefficient calculations suggested the presence of a 1:1 complex in the asymmetric unit. Detailed data collection statistics are reported in Table 4.

Table 4. Data collection statistics. Values in parentheses refer the highest resolution shell.

Space group	$P2_1$
Unit-cell parameters (\AA , $^\circ$)	$a = 48.14$, $b = 81.90$, $c = 53.72$, $\beta = 99.7$
Resolution (\AA)	50.0–2.40 (2.49–2.40)
R_{merge} (%) [†]	9.9 (57.8)
$\langle I/\sigma(I) \rangle$	13.8 (2.0)
No. of observations	77242
No. of unique reflections	16175
Completeness (%)	99.8 (98.9)
Average multiplicity	4.8 (4.1)
V_M ($\text{\AA}^3 \text{ Da}^{-1}$)	2.4
Solvent content (%)	53.4

Structure determination and refinement

The structure of the complex between HD22[27] and thrombin was solved by molecular replacement method using the program Phaser (McCoy, *et al.*, 2007) with the coordinates of both inhibited thrombin (derived from the structure of its complex with HD1; PDB entry 4dih) and a DNA duplex (derived from the structure of a DNA dodecamer duplex; PDB entry 1lu5) as search models. To avoid bias, PPACK, ions and water molecules were removed from the starting model. A clear solution was obtained with a log-likelihood gain (LLG) of 1511. The model was subjected to few cycles of rigid body refinement followed by several cycles of coordinate minimization and B-factor refinement using CNS (Brunger, *et al.*, 1998) and REFMAC5 (Vagin, *et al.*, 2004). Each

run was alternated with manual model building using Coot (Emsley, *et al.*, 2010). Fourier difference maps, calculated with (Fo–Fc) and (2Fo–Fc) coefficients, showed continuous electron density in proximity of the DNA duplex segment. The analysis of these maps, calculated at various stages of refinement, allowed the building of the whole aptamer molecule, as well as the fitting of PPACK in the active site, the identification of the protein glycosylation site, and the positioning of several water molecules. After inclusion of low resolution data and bulk solvent correction, the crystallographic R/Rfree for the final model of the complex between thrombin and HD22[27], in the resolution range 50.0–2.40 Å, were 0.181/0.249. At the end of refinement the geometry of the protein structure was monitored using PROCHECK (Laskowski, *et al.*, 1993) and WHATCHECK (Hoof, *et al.*, 1996). A full list of refinement statistics is reported in Table 5. The drawings were prepared with Pymol (Schrodinger, 2010). The coordinates of the structure have been deposited in the Protein Data Bank with code 4i7y.

Table 5. Summary of refinement statistics.

Resolution (Å)	50.00–2.40 (2.49–2.40)
No. of reflections	14526
$R_{\text{work}}/R_{\text{free}}$ (%)	18.1/24.9
No. of atoms	
Total	2951
Protein	2288
Aptamer	564
Ions	1
Water	98
Average B factor (Å ²)	
Overall	50.1
Protein	47.2
Aptamer	62.2
Ion	46.8
Water	48.3
R.m.s. deviations	
Bond lengths (Å)	0.011
Bond angles (°)	1.687
Ramachandran plot, residues in (%)	
Most favoured region	91.9
Additionally allowed region	8.1
Generously allowed region	0

Circular Dichroism

Circular dichroism (CD) spectra were recorded at 10°C using a Jasco J-710 spectropolarimeter equipped with a Peltier thermostatic cell holder (Model PTC-348WI). CD measurements were carried out in the 200–350 nm range using a 0.1 cm path length cell and 40 μM solutions of the free aptamer. Before measurements, the instrument was calibrated with an aqueous solution of D(+)-10-camphorsulfonic acid at 290 nm and the samples were pre-equilibrated at 10°C for 5 minutes.

Thermal unfolding curves were recorded in the 10–90°C range, in 1°C steps with 0.5 minute equilibration time between readings, at heating rate 1°C/min. Transition temperatures were calculated from the second derivative of the ellipticity change vs temperature.

REFERENCES

- Ambrus A, Chen D, Dai J, Bialis T, Jones RA, Yang D. (2006) Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic Acids Res*, **34**:2723-2735.
- Balasubramanian S, Neidle S. (2009) G-quadruplex nucleic acids as therapeutic targets. *Curr Opin Chem Biol*, **13**:345-353.
- Bloomfield VA, Crothers DM, Tinoco I: (2000) Nucleic acids: structures, properties, and functions. Sausalito, CA: University science books
- Bock LC, Griffin LC, Latham JA, Vermaas EH, Toole JJ. (1992) Selection of single-stranded DNA molecules that bind and inhibit human thrombin. *Nature*, **355**:564-566.
- Bock P, Panizzi P, Verhamme I. (2007) Exosites in the substrate specificity of blood coagulation reactions. *J Thromb Haemost*, **5**:81-94.
- Bode W, Turk D, Karshikov A. (1992) The refined 1.9 - Å X - ray crystal structure of d - Phe - Pro - Arg chloromethylketone - inhibited human α - thrombin: Structure analysis, overall structure, electrostatic properties, detailed active - site geometry, and structure - function relationships. *Protein Sci*, **1**:426-471.
- Bode W. (2006) Structure and interaction modes of thrombin. *Blood Cells Mol Dis*, **36**:122-130.
- Bonifacio L, Church FC, Jarstfer MB. (2008) Effect of locked-nucleic acid on a biologically active g-quadruplex. A structure-activity relationship of the thrombin aptamer. *Int J Mol Sci*, **9**:422-433.
- Borbone N, Bucci M, Oliviero G, Morelli E, Amato J, D'Atri V, D'Errico S, Vellecco V, Cirino G, Piccialli G *et al.* (2012) Investigating the role of T7 and T12 residues on the biological properties of thrombin-binding aptamer: enhancement of anticoagulant activity by a single nucleobase modification. *J Med Chem*, **55**:10716-10728.
- Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS *et al.* (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Cryst D*, **54**:905-921.
- Bunka DH, Stockley PG. (2006) Aptamers come of age - at last. *Nat Rev Microbiol*, **4**:588-596.

- Burge S, Parkinson GN, Hazel P, Todd AK, Neidle S. (2006) Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res*, **34**:5402-5415.
- Chen HW, Medley CD, Sefah K, Shangguan D, Tang Z, Meng L, Smith JE, Tan W. (2008) Molecular Recognition of Small - Cell Lung Cancer Cells Using Aptamers. *ChemMedChem*, **3**:991-1001.
- Chi-hong BC, Chernis GA, Hoang VQ, Landgraf R. (2003) Inhibition of heregulin signaling by an aptamer that preferentially binds to the oligomeric form of human epidermal growth factor receptor-3. *Proc Natl Acad Sci USA*, **100**:9226-9231.
- Chinnapen DJ, Sen D. (2002) Hemin-stimulated docking of cytochrome c to a hemin-DNA aptamer complex. *Biochemistry*, **41**:5202-5212.
- Chou SH, Chin KH, Wang AH. (2005) DNA aptamers as potential anti-HIV agents. *Trends Biochem Sci*, **30**:231-234.
- Coughlin SR. (1999) How the protease thrombin talks to cells. *Proc Natl Acad Sci USA*, **96**:11023-11027.
- Crawley J, Zanardelli S, Chion C, Lane D. (2007) The central role of thrombin in hemostasis. *J Thromb Haemost*, **5**:95-101.
- Crick F. (1970) Central dogma of molecular biology. *Nature*, **227**:561-563.
- Dahm R. (2008) Discovering DNA: Friedrich Miescher and the early years of nucleic acid research. *Hum Genet*, **122**:565-581.
- Dai J, Carver M, PUNCHIHEWA C, Jones RA, Yang D. (2007) Structure of the Hybrid-2 type intramolecular human telomeric G-quadruplex in K⁺ solution: insights into structure polymorphism of the human telomeric sequence. *Nucleic Acids Res*, **35**:4927-4940.
- Daniels DA, Chen H, Hicke BJ, Swiderek KM, Gold L. (2003) A tenascin-C aptamer identified by tumor cell SELEX: systematic evolution of ligands by exponential enrichment. *Proc Natl Acad Sci USA*, **100**:15416-15421.
- Davis JT. (2004) G-quartets 40 years later: from 5'-GMP to molecular biology and supramolecular chemistry. *Angew Chem, Int Ed Engl*, **43**:668-698.
- Di Cera E, Dang Q, Ayala Y. (1997) Molecular mechanisms of thrombin function. *Cell Mol Life Sci*, **53**:701-730.
- Di Cera E. (2007) Thrombin as procoagulant and anticoagulant. *J Thromb Haemost*, **5**:196-202.

- Di Cera E, Page MJ, Bah A, Bush-Pelc LA, Garvey LC. (2007) Thrombin allostery. *Phys Chem Chem Phys*, **9**:1291-1306.
- Dolinnaya NG, Yuminova AV, Spiridonova VA, Arutyunyan AM, Kopylov AM. (2012) Coexistence of G-quadruplex and duplex domains within the secondary structure of 31-mer DNA thrombin-binding aptamer. *J Biomol Struct Dyn*, **30**:524-531.
- Draper DE, Grilley D, Soto AM. (2005) Ions and RNA folding. *Annu Rev Biophys Biomol Struct*, **34**:221-243.
- Ellington AD, Szostak JW. (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature*, **346**:818-822.
- Emsley P, Lohkamp B, Scott WG, Cowtan K. (2010) Features and development of Coot. *Acta Cryst D*, **66**:486-501.
- Engelhart AE, Plavec J, Persil O, Hud NV: (2009) Metal Ion Interactions with G-Quadruplex Structures. In: *Nucleic Acid-Metal Ion Interactions*. 118-147, edn. Edited by Hud NV. Cambridge, UK: RCS Publishing;
- Fan P, Suri AK, Fiala R, Live D, Patel DJ. (1996) Molecular recognition in the FMN-RNA aptamer complex. *J Mol Biol*, **258**:480-500.
- Fernando H, Sewitz S, Darot J, Tavaré S, Huppert JL, Balasubramanian S. (2009) Genome-wide analysis of a G-quadruplex-specific single-chain antibody that regulates gene expression. *Nucleic Acids Res*, **37**:6716-6722.
- Gatto B, Palumbo M, Sissi C. (2009) Nucleic acid aptamers based on the G-quadruplex structure: therapeutic and diagnostic potential. *Curr Med Chem*, **16**:1248-1265.
- Girardot M, Gareil P, Varenne A. (2010) Interaction study of a lysozyme - binding aptamer with mono - and divalent cations by ACE. *Electrophoresis*, **31**:546-555.
- Griffin LC, Tidmarsh GF, Bock LC, Toole JJ, Leung L. (1993) In vivo anticoagulant properties of a novel nucleotide-based thrombin inhibitor and demonstration of regional anticoagulation in extracorporeal circuits. *Blood*, **81**:3271-3276.
- Guo Z, Eisenberg D. (2006) Runaway domain swapping in amyloid-like fibrils of T7 endonuclease I. *Proc Natl Acad Sci USA*, **103**:8042-8047.
- Hasegawa H, Taira K-i, Sode K, Ikebukuro K. (2008) Improvement of aptamer affinity by dimerization. *Sensors*, **8**:1090-1098.
- Heckel A, Mayer G. (2005) Light regulation of aptamer activity: an anti-thrombin aptamer with caged thymidine nucleobases. *J Am Chem Soc*, **127**:822-823.

Hermann T, Patel DJ. (2000) Adaptive recognition by nucleic acid aptamers. *Science*, **287**:820-825.

Hooft R, Vriend G, Sander C, Abola EE. (1996) Errors in protein structures. *Nature*, **381**:272-272.

Huntington JA. (2005) Molecular recognition mechanisms of thrombin. *J Thromb Haemost*, **3**:1861-1872.

Hurley LH. (2002) DNA and its associated processes as targets for cancer therapy. *Nat Rev Cancer*, **2**:188-200.

Jayasena SD. (1999) Aptamers: an emerging class of molecules that rival antibodies in diagnostics. *Clin Chem*, **45**:1628-1650.

Jones LA, Clancy LE, Rawlinson WD, White PA. (2006) High-affinity aptamers to subtype 3a hepatitis C virus polymerase display genotypic specificity. *Antimicrob Agents Chemother*, **50**:3019-3027.

Keefe AD, Pai S, Ellington A. (2010) Aptamers as therapeutics. *Nat Rev Drug Discov*, **9**:537-550.

Khati M, Schüman M, Ibrahim J, Sattentau Q, Gordon S, James W. (2003) Neutralization of infectivity of diverse R5 clinical isolates of human immunodeficiency virus type 1 by gp120-binding 2' F-RNA aptamers. *J Virol*, **77**:12692-12698.

Kumari S, Bugaut A, Huppert JL, Balasubramanian S. (2007) An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation. *Nat Chem Biol*, **3**:218-221.

Laskowski RA, MacArthur MW, Moss DS, Thornton JM. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr*, **26**:283-291.

Lauhon CT, Szostak JW. (1995) RNA aptamers that bind flavin and nicotinamide redox cofactors. *J Am Chem Soc*, **117**:1246-1257.

Lawrence MC, Colman PM. (1993) Shape Complementarity at Protein-Protein Interfaces. *J Mol Biol*, **234**:946-950.

Levesque D, Beaudoin J, Roy S, Perreault J. (2007) In vitro selection and characterization of RNA aptamers binding thyroxine hormone. *Biochem J*, **403**:129-138.

Li T, Wang E, Dong S. (2009) A grafting strategy for the design of improved G-quadruplex aptamers and high-activity DNazymes. *PLoS One*, **4**:e5126.

Li W, Kaplan A, Grant G, Toole J, Leung L. (1994) A novel nucleotide-based thrombin inhibitor inhibits clot-bound thrombin and reduces arterial platelet thrombus formation. *Blood*, **83**:677-682.

Lin PH, Chen RH, Lee CH, Chang Y, Chen CS, Chen WY. (2011) Studies of the binding mechanism between aptamers and thrombin by circular dichroism, surface plasmon resonance and isothermal titration calorimetry. *Colloid Surface B*, **88**:552-558.

Lipps HJ, Rhodes D. (2009) G-quadruplex structures: *in vivo* evidence and function. *Trends Cell Biol*, **19**:414-422.

Lombardi A, De Simone G, Galdiero S, Staiano N, Natri F, Pavone V. (1999) From natural to synthetic multisite thrombin inhibitors. *Biopolymers*, **51**:19-39.

Long SB, Long MB, White RR, Sullenger BA. (2008) Crystal structure of an RNA aptamer bound to thrombin. *RNA*, **14**:2504-2512.

Lorsch JR, Szostak JW. (1994) In vitro selection of RNA aptamers specific for cyanocobalamin. *Biochemistry*, **33**:973-982.

Luu KN, Phan AT, Kuryavyi V, Lacroix L, Patel DJ. (2006) Structure of the human telomere in K⁺ solution: an intramolecular (3 + 1) G-quadruplex scaffold. *J Am Chem Soc*, **128**:9963-9970.

Macaya RF, Schultze P, Smith FW, Roe JA, Feigon J. (1993) Thrombin-Binding DNA Aptamer Forms a Unimolecular Quadruplex Structure in Solution. *Proc Natl Acad Sci USA*, **90**:3745-3749.

Macaya RF, Waldron JA, Beutel BA, Gao HT, Joesten ME, Yang MH, Patel R, Bertelsen AH, Cook AF. (1995) Structural and Functional-Characterization of Potent Antithrombotic Oligonucleotides Possessing Both Quadruplex and Duplex Motifs. *Biochemistry*, **34**:4478-4492.

Marson G, Palumbo M, Sissi C. (2012) Folding versus charge: understanding selective target recognition by the thrombin aptamers. *Curr Pharm Des*, **18**:2027-2035.

Martino L, Virno A, Randazzo A, Virgilio A, Esposito V, Giancola C, Bucci M, Cirino G, Mayol L. (2006) A new modified thrombin binding aptamer containing a 5' -5' inversion of polarity site. *Nucleic Acids Res*, **34**:6653-6662.

Mayer G, Müller J, Mack T, Freitag DF, Höver T, Pöttsch B, Heckel A. (2009) Differential regulation of protein subdomain activity with caged bivalent ligands. *ChemBioChem*, **10**:654-657.

- McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. (2007) Phaser crystallographic software. *J Appl Crystallogr*, **40**:658-674.
- Mergny J-L, Hélène C. (1998) G-quadruplex DNA: a target for drug design. *Nat Med*, **4**:1366-1367.
- Miyachi Y, Shimizu N, Ogino C, Fukuda H, Kondo A. (2009) Selection of a DNA aptamer that binds 8-OHdG using GMP-agarose. *Bioorg Med Chem Lett*, **19**:3619-3622.
- Mori T, Oguro A, Ohtsu T, Nakamura Y. (2004) RNA aptamers selected against the receptor activator of NF-kappaB acquire general affinity to proteins of the tumor necrosis factor receptor family. *Nucleic Acids Res*, **32**:6120-6128.
- Müller J, Wulffen B, Pöttsch B, Mayer G. (2007) Multidomain Targeting Generates a High - Affinity Thrombin - Inhibiting Bivalent Aptamer. *ChemBioChem*, **8**:2223-2226.
- Müller J, Freitag D, Mayer G, Pöttsch B. (2008) Anticoagulant characteristics of HD1 - 22, a bivalent aptamer that specifically inhibits thrombin and prothrombinase. *J Thromb Haemost*, **6**:2105-2112.
- Musumeci D, Montesarchio D. (2012) Polyvalent nucleic acid aptamers and modulation of their activity: a focus on the thrombin binding aptamer. *Pharmacol Ther*, **136**:202-215.
- Nagatoishi S, Isono N, Tsumoto K, Sugimoto N. (2011) Loop residues of thrombin-binding DNA aptamer impact G-quadruplex stability and thrombin binding. *Biochimie*, **93**:1231-1238.
- Nagatoishi S, Sugimoto N. (2012) Interaction of water with the G-quadruplex loop contributes to the binding energy of G-quadruplex to protein. *Mol Biosyst*, **8**:2766-2770.
- Neidle S, Parkinson G. (2002) Telomere maintenance as a target for anticancer drug discovery. *Nat Rev Drug Discovery*, **1**:383-393.
- Neidle S. (2009) The structures of quadruplex nucleic acids and their drug complexes. *Curr Opin Struct Biol*, **19**:239-250.
- Nimjee SM, Rusconi CP, Harrington RA, Sullenger BA. (2005a) The potential of aptamers as anticoagulants. *Trends Cardiovasc Med*, **15**:41-45.
- Nimjee SM, Rusconi CP, Sullenger BA. (2005b) Aptamers: an emerging class of therapeutics. *Annu Rev Med*, **56**:555-583.

- Nutescu EA, Shapiro NL, Chevalier A. (2008) New anticoagulant agents: direct thrombin inhibitors. *Cardiol Clin*, **26**:169-187.
- Oney S, Lam RT, Bompiani KM, Blake CM, Quick G, Heidel JD, Liu JY-C, Mack BC, Davis ME, Leong KW. (2009) Development of universal antidotes to control aptamer activity. *Nat Med*, **15**:1224-1228.
- Otwinowski Z, Minor W: (1997) Processing of X-ray diffraction data collected in oscillation mode. In: *Methods Enzymol. Volume 276*: 307-326, edn. Edited by Charles W. Carter J: Academic Press;
- Paborsky L, McCurdy SN, Griffin LC, Toole JJ, Leung L. (1993) The single-stranded DNA aptamer-binding site of human thrombin. *J Biol Chem*, **268**:20808-20811.
- Padmanabhan K, Tulinsky A. (1996) An ambiguous structure of a DNA 15-mer thrombin complex. *Acta Cryst D*, **52**:272-282.
- Parekh P, Tang Z, Turner PC, Moyer RW, Tan W. (2010) Aptamers recognizing glycosylated hemagglutinin expressed on the surface of vaccinia virus-infected cells. *Anal Chem*, **82**:8642-8649.
- Parkinson GN, Lee MP, Neidle S. (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*, **417**:876-880.
- Pasternak A, Hernandez FJ, Rasmussen LM, Vester B, Wengel J. (2011) Improved thrombin binding aptamer by incorporation of a single unlocked nucleic acid monomer. *Nucleic Acids Res*, **39**:1155-1164.
- Patel DJ, Phan AT, Kuryavyi V. (2007) Human telomere, oncogenic promoter and 5' -UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics. *Nucleic Acids Res*, **35**:7429-7455.
- Phan AT, Luu KN, Patel DJ. (2006) Different loop arrangements of intramolecular human telomeric (3+1) G-quadruplexes in K⁺ solution. *Nucleic Acids Res*, **34**:5715-5719.
- Phan AT, Kuryavyi V, Burge S, Neidle S, Patel DJ. (2007) Structure of an unprecedented G-quadruplex scaffold in the human c-kit promoter. *J Am Chem Soc*, **129**:4386-4392.
- Phan AT, Kuryavyi V, Darnell JC, Serganov A, Majumdar A, Ilin S, Raslin T, Polonskaia A, Chen C, Clain D. (2011) Structure-function studies of FMRP RGG peptide recognition of an RNA duplex-quadruplex junction. *Nat Struct Mol Biol*, **18**:796-804.

Pileur F, Andreola ML, Dausse E, Michel J, Moreau S, Yamada H, Gaidamakov SA, Crouch RJ, Toulmé JJ, Cazenave C. (2003) Selective inhibitory DNA aptamers of the human RNase H1. *Nucleic Acids Res*, **31**:5776-5788.

Radom F, Jurek P, Mazurek M, Otlewski J, Jeleń F. (2013) Aptamers: Molecules of great potential. *Biotechnol Adv*, **31**:1260–1274.

Rangnekar A, Zhang AM, Li SS, Bompiani KM, Hansen MN, Gothelf KV, Sullenger BA, LaBean TH. (2012) Increased anticoagulant activity of thrombin-binding DNA aptamers by nanoscale organization on DNA nanostructures. *Nanomed-Nanotechnol*, **8**:673-681.

Rau J, Beaulieu L, Huntington J, Church F. (2007) Serpins in thrombosis, hemostasis and fibrinolysis. *J Thromb Haemost*, **5**:102-115.

Rohrbach F, Fatthalla MI, Kupper T, Potzsch B, Muller J, Petersen M, Pedersen EB, Mayer G. (2012) Chemical maturation of a bivalent aptamer by single domain variation. *Chembiochem*, **13**:631-634.

Rusconi CP, Roberts JD, Pitoc GA, Nimjee SM, White RR, Quick G, Scardino E, Fay WP, Sullenger BA. (2004) Antidote-mediated control of an anticoagulant aptamer in vivo. *Nat Biotech*, **22**:1423-1428.

Russo Krauss I, Merlino A, Randazzo A, Mazzarella L, Sica F. (2010) Crystallization and preliminary X-ray analysis of the complex of human α -thrombin with a modified thrombin-binding aptamer. *Acta Cryst F*, **66**:961-963.

Russo Krauss I, Merlino A, Giancola C, Randazzo A, Mazzarella L, Sica F. (2011) Thrombin-aptamer recognition: a revealed ambiguity. *Nucleic Acids Res*, **39**:7858-7867.

Russo Krauss I, Merlino A, Randazzo A, Novellino E, Mazzarella L, Sica F. (2012) High-resolution structures of two complexes between thrombin and thrombin-binding aptamer shed light on the role of cations in the aptamer inhibitory activity. *Nucleic Acids Res*, **40**:8119-8128.

Russo Krauss I, Pica A, Merlino A, Mazzarella L, Sica F. (2013) Duplex–quadruplex motifs in a peculiar structural organization cooperatively contribute to thrombin binding of a DNA aptamer. *Acta Cryst D*, **69**:2403-2411.

Sagi J. (2013) G-quadruplexes incorporating modified constituents: a review. *J Biomol Struct Dyn*, **32**:477-511.

Schrodinger L. (2010) *The PyMOL Molecular Graphics System, Version 13*, Schrödinger, LLC.

- Spiridonova V, Rog E, Dugina T, Strukova S, Kopylov A. (2003) Aptamer DNA: a new type of thrombin inhibitors. *Russ J Bioorg Chem*, **29**:450-453.
- Stubbs MT, Bode W. (1995) The clot thickens: clues provided by thrombin structure. *Trends Biochem Sci*, **20**:23-28.
- Tan W, Wang H, Chen Y, Zhang X, Zhu H, Yang C, Yang R, Liu C. (2011) Molecular aptamers for drug delivery. *Trends Biotechnol*, **29**:634-640.
- Tang Z, Parekh P, Turner P, Moyer RW, Tan W. (2009) Generating aptamers for recognition of virus-infected cells. *Clin Chem*, **55**:813-822.
- Tasset DM, Kubik MF, Steiner W. (1997) Oligonucleotide inhibitors of human thrombin that bind distinct epitopes. *J Mol Biol*, **272**:688-698.
- Tsiang M, Jain AK, Dunn KE, Rojas ME, Leung LL, Gibbs CS. (1995) Functional mapping of the surface residues of human thrombin. *J Biol Chem*, **270**:16854-16863.
- Tuerk C, Gold L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, **249**:505-510.
- Vagin AA, Steiner RA, Lebedev AA, Potterton L, McNicholas S, Long F, Murshudov GN. (2004) REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Cryst D*, **60**:2184-2195.
- Vangone A, Spinelli R, Scarano V, Cavallo L, Oliva R. (2011) COCOMAPS: a web application to analyze and visualize contacts at the interface of biomolecular complexes. *Bioinformatics*, **27**:2915-2916.
- Wang Y, Patel DJ. (1993) Solution structure of the human telomeric repeat d[AG3(T2AG3)3] G-tetraplex. *Structure*, **1**:263-282.
- Watson JD, Crick FH. (1953) Molecular structure of nucleic acids. *Nature*, **171**:737-738.
- Williamson JR. (1994) G-quartet structures in telomeric DNA. *Annu Rev Biophys Biomol Struct*, **23**:703-730.
- Wu Q, Tsiang M, Sadler JE. (1992) Localization of the single-stranded DNA binding site in the thrombin anion-binding exosite. *J Biol Chem*, **267**:24408-24412.
- Zavyalova E, Golovin A, Reshetnikov R, Mudrik N, Panteleyev D, Pavlova G, Kopylov A. (2011) Novel Modular DNA Aptamer for Human Thrombin with High Anticoagulant Activity. *Curr Med Chem*, **18**:3343-3350.

Zhou GP, Huang XR, Qu YB. (2010) The binding effect of aptamers on thrombin. *Biochem Eng J*, **52**:117-122.

Zhu G, Ye M, Donovan MJ, Song E, Zhao Z, Tan W. (2012) Nucleic acid aptamers: an emerging frontier in cancer therapy. *Chem Comm*, **48**:10472-10480.

**PART II – DOMAIN SWAPPING AND
RIBONUCLEASES**

STATE OF THE ART

Protein-protein recognition is a critical event in a large number of cell processes. In fact, in biological systems proteins rarely act in isolation but they bind other biomolecules to elicit specific cellular responses (Jones, *et al.*, 1995; Marianayagam, *et al.*, 2004). In particular, the self-association of proteins to form dimers and higher-order oligomers is a very common phenomenon in nature. Protein-protein interactions form the basis of the quaternary structure of multimeric proteins, and represent one of the most complex levels of structural organization in biological molecules. Multimeric proteins represent just one system in which protein-protein interactions are important and interactions have been studied in many protein-inhibitor, protein-ligand and antibody-antigen complexes (Janin, *et al.*, 1990). However, it has been shown that the interactions between subunits in dimeric proteins are amongst the strongest and most extensive (*accessible surface area* up to about 5000 Å² per subunit).

Among the variety of mechanisms by which protein oligomerization has evolved, an interesting role is played by 3D domain swapping. It is a mechanism by which two monomer exchange (swap) an identical structural domain to form dimers or higher-order oligomers. In a 3D domain-swapped dimer, a domain of one subunit replaces the identical domain of the other subunit, and *vice versa*, restoring all the interaction already present in the monomers. According to Eisenberg's definitions (Bennett, *et al.*, 1995; Schlunegger, *et al.*, 1997), the interface between domains that exists in both the monomer and the dimer is termed closed interface (C-interface) and the interface that is present only in the dimer is termed open interface (O-interface). However, the segments that links the two domains (hinge loops) present a different conformation for the monomer and the dimer (Figure 17).

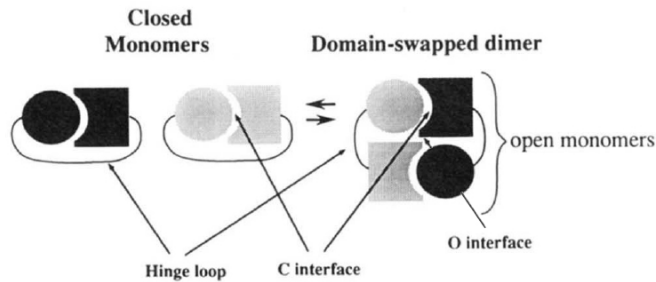


Figure 17. Closed monomers are comprised of tertiary or secondary structural domains (represented by circle and square) linked by polypeptide linkers (hinge loops). The interface between domains in the closed monomer is referred to as the C- (closed) interface. Closed monomers may be opened by mildly denaturing conditions or by mutations that destabilize the closed monomer. Open monomers may dimerize by domain swapping. The domain-swapped dimer has two C-interfaces identical to those in the closed monomer, however, each is formed between a domain from one subunit (black) and a domain from the other subunit (gray). The only residues whose conformations significantly differ between the closed and open monomers are in the hinge loop. The interface that does not exist in the closed monomers is referred to as O- (open) interface.

To date, despite numerous efforts, no unifying molecular mechanism of domain swapping has emerged. What seems clear is the existence of distinct intermediates, in which some hydrophobic part of the monomeric protein becomes exposed and, thereby, is available for interaction with a ‘like’ molecule (Gronenborn, 2009).

It has been speculated about the possibility that 3D domain swapping could be a dimerization mechanism for all proteins with free N- and C- termini (Bennett, *et al.*, 1995). This elegant mechanism of protein oligomerization has an increasing frequency in the literature, but its biological significance remains still unclear. For example, enzymatic proteins could use 3D domain swapping for functional regulation, in particular in allostery, which involves multiple active sites. It could also rescue protein function after deleterious mutation (Chirgadze, *et al.*, 2004). So far, however, it appears that 3D domain swapping is predominantly connected with molecular pathology, through its involvement in conformational disorders, of which amyloid aggregation is perhaps the most important (Janowski, *et al.*, 2001; Knaus, *et al.*, 2001; Bennett, *et al.*, 2006; Liu, *et al.*, 2011).

Over the years many proteins have been found to be domain-swapped. However, bovine pancreatic ribonuclease (RNase A) has represented and it still represents a suitable model system for the study of such mechanism. It is well-known that the first hypothesis of exchange of N-terminal regions was based on biochemical experiments performed on this protein (Crestfield, *et al.*, 1963). RNase A was also the first protein showing the possibility to exchange different regions of the polypeptide chain, giving rise to a number of experimental and modeling studies to characterize multimeric assemblies (Liu, *et al.*, 2002a; Liu, *et al.*, 2002b). Upon lyophilization from acetic acid, RNase A forms two different swapping dimers: in one dimer (C-Dimer) the two C-terminal β -strands are swapped (Liu, *et al.*, 2001), whereas in the other (N-Dimer) the N-terminal α -helices (Liu, *et al.*, 1998) are swapped. The two dimers have different quaternary structures: the N-Dimer presents a very compact structure, whereas the C-Dimer shows a very loose O-interface. Moreover, with the same procedure, other high-order RNase A oligomers are generated (trimers, tetramers, etc.). Each oligomeric species exists in the form of at least two conformational isomers (Gotte, *et al.*, 1999). 3D domain swapping has been found to endow oligomers with additional properties besides the basal enzymatic activity (Liu, *et al.*, 1998). Indeed, oligomeric forms of RNase A have generated great interest as engineered therapeutics, as they, in contrast with monomeric RNase A, show antitumor activity (even if mild). Different members of the pancreatic-type ribonuclease family are able to acquire oligomeric swapped forms. Gaining a better understanding of the mechanism by which these aggregates form is important not only for understanding the process of domain swapping but also for improving methods by which to engineer stable oligomeric forms. Furthermore these studies of oligomers of well-characterized RNases may illuminate the formation of amyloid and other protein aggregates.

**CHAPTER III – DOMAIN SWAPPING: GIVING
SOMETHING TO TAKE “SOMETHING ELSE” BACK**

INTRODUCTION

Bovine seminal ribonuclease (BS-RNase) holds a unique position in the RNase A family since it is the only covalent dimeric member. Native BS-RNase is an equilibrium mixture of swapped (MxM) and unswapped (M=M) isoforms (Piccoli, *et al.*, 1992), whose subunits are linked by two intersubunit disulphide bridges formed by Cys31 of one subunit and Cys32 of the other, and *vice versa*. The MxM form, which is further stabilized by additional non-covalent interactions between the two subunits through the swapping of the N-terminal α -helices, displays a potentially therapeutic cytotoxic activity for tumor cells (Matousek, 1973; Kim, *et al.*, 1995). This peculiar property has been related to the shape (Sica, *et al.*, 2004) and stability (Giancola, *et al.*, 2011) of its non-covalent swapped dimeric derivative (NCD-BS), which can be obtained upon selective reduction of the two MxM interchain disulphide bridges (Piccoli, *et al.*, 1992). Indeed, it has been shown that in reduction conditions, as those found in the cytosol, the MxM isoform is converted in the NCD form whose quaternary structure is very similar to that of MxM (Sica, *et al.*, 2004) and prevents the binding with the Ribonuclease Inhibitor (RI), a protein very abundant in the cytosol of mammalian cells. It is worth to note that RI binds with very high affinity RNase A (Lee, *et al.*, 1989) as well as other monomeric, pancreatic-like, RNases (Kobe, *et al.*, 1993; Kobe, *et al.*, 1996).

Studies on dimeric variants of RNase A, obtained by introducing in its sequence residues playing a key role in the acquisition of the BS-RNase 3D domain-swapped structure, have confirmed the strict relation between antitumor activity and NCD-BS quaternary architecture. Indeed, none of these dimeric variants retains in its non-covalent form a compact quaternary structure able to evade RI (Merlino, *et al.*, 2009b). This result well correlates with the much lower cytotoxic activity of these RNase A variants as compared to that of BS-RNase (Ercole, *et al.*, 2009). BS-RNase and RNase A share more than 80% of the primary structure and a strictly conserved three-dimensional fold

(Spadaccini, *et al.*, 2012), but they have a completely different behavior with respect to the swapping process. In particular, up to now there has been no experimental evidence of C-swapped BS-RNase dimers. For this reason it was believed that, in contrast to RNase A, the C-swapping was precluded for the BS-RNase sequence, due to the presence of a lysine residue replacing asparagine 113 in the C-terminal hinge region (residues 111-116). However, very recently, experimental evidences of the formation of BS-RNase derivatives stabilized by the C-terminal swapping have been shown (Gotte, *et al.*, 2012).

In this PhD project, the first clear experimental evidences on the formation of a C-terminal swapped dimer of the monomeric BS-RNase derivatives, after incubation in acetic acid and lyophilization, have been produced. The crystallographic model of one of these dimers has revealed a quaternary structure different with respect to that adopted by the CD-RNase A that well correlates with the data on the stability against dissociation of this dimer.

RESULTS

Solution studies

Monomeric samples of BS-RNase with the cysteine residues 31 and 32 alkylated with iodoacetamide have been incubated in 40% acetic acid for one hour at room temperature and lyophilized. In a typical experiment, 10 mg of protein are dissolved in 2.0 mL of a solution of 40% acetic acid, at a final concentration of 5 mg/mL. After overnight lyophilization, the protein is dissolved either in 50 mM Tris/HCl buffer pH 7.3, 0.13 M NaCl or in 0.2 M phosphate buffer pH 6.7 and loaded on a G-75 column (equilibrated with the same buffer) and eluted at 4°C. The chromatographic pattern is reported in Figure 18 (dashed line). Peak integration revealed that more than 60% of the protein is eluted as monomer, whilst the remaining part corresponds to aggregated forms of different size, alt-

though the relative amount of each aggregate varied slightly with the elution buffer.

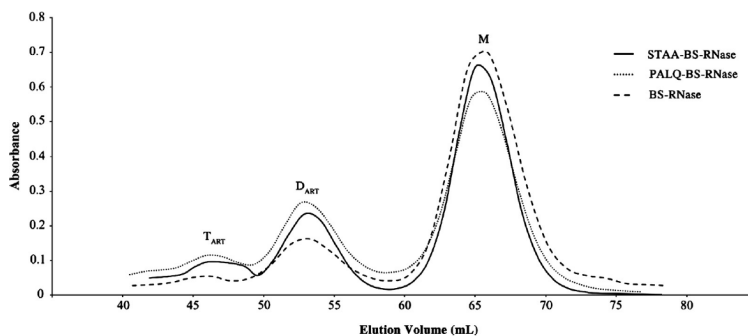


Figure 18. Chromatographic pattern of a gel-filtration on Sephadex G-75 upon acetic acid incubation and lyophilization of the monomeric derivatives of STAA-BS-RNase, PALQ-BS-RNase, and BS-RNase.

From the elution volume it is possible to calculate that the first peak corresponds to the size of a tetramer, the second (indicated as T_{ART}) to the size of a trimer, and the third (indicated as D_{ART}) to the size of a dimer. The remaining part, equivalent roughly to the 10% of the total, corresponds to higher-size aggregates. The dimer accounts for about 20-25% of the total protein amount.

The same experiment has been performed for the STAA-mBS mutant, obtaining the chromatographic profile also reported in Figure 18 (solid line), which closely resembles the one obtained with the parent (wild-type) protein, except for the slightly higher amount of aggregates. The fractions corresponding to the artificial dimers of both BS-RNase and STAA-BS-RNase have been pooled and a small amount of these protein samples has been loaded on a polyacrylamide gel run under native conditions (pH 4.5). Aliquots of BS-RNase and of NCD-BS together with the monomeric forms of BS-RNase and STAA-BS-RNase (mBS and STAA-mBS) were also loaded on the gel (Figure 19).

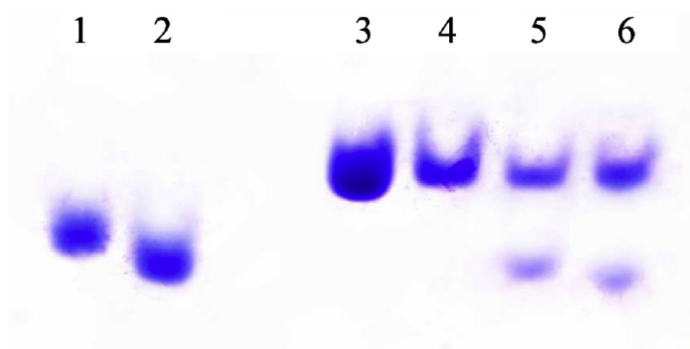


Figure 19. Polyacrylamide gel-electrophoresis under native conditions of the dimeric forms obtained by incubation in acetic acid followed by lyophilization of the monomeric derivatives of BS-RNase and STAA-BS-RNase (mBS and STAA-mBS) (lanes 5 and 6). The corresponding monomeric forms (lanes 1 and 2), together with BS-RNase (lane 3) and NCD-BS, obtained by selective reduction of the swapped dimer (lane 4), are also reported for comparison.

The analysis of the gel indicated clearly that BS-RNase and NCD-BS (lanes 3 and 4) are characterized by a single band. On the contrary, both the artificial dimers (lanes 5 and 6) showed the presence of a significant amount of the corresponding monomeric forms (lanes 1 and 2), which necessarily had to be produced upon dissociation of the dimeric species after the gel-filtration and, in part, even while running the gel. Interestingly, the comparison of lanes 3 and 4 with 5 and 6 indicates that the artificial dimers migrate on the gel with a mobility very close to that of NCD-BS, thus suggesting the presence of a single dimeric form even in the samples prepared by acetic acid lyophilization, in contrast with the results obtained for RNase A. It is indeed well-known that the two dimers formed when RNase A is subjected to the same treatment can be separated by cathodic gel-electrophoresis (Gotte, *et al.*, 1999).

The fast dissociation of the artificial dimers indicated by the gel-electrophoresis prompted a further characterization of their stability in solution. Therefore, a comparison of the dissociation processes of the NCD forms of BS-RNase and STAA-BS-RNase with the corresponding artificial dimers, *i.e.* D_{ART-BS} and $D_{ART-STAA-BS}$ has been performed. The proteins have been incubated

at 37°C and, at regular time intervals, aliquots corresponding to 0.2 mg of each sample were loaded on Superdex 75 to evaluate the relative amount of dimer and monomer. Surprisingly, the kinetic stability of the two artificial dimers was found significantly lower than that of the corresponding NCD forms: the $t_{1/2}$ of the artificial dimers were 1.5 hours for D_{ART}-BS and 0.6 hours for D_{ART}-STAA-BS, while the values of the partner NCDs, measured in the same experimental conditions, were 6.1 hours and 2.2 hours, respectively (Table 6).

Table 6. Half-life times $t_{1/2}$ of dimers from dissociation experiments.

	$t_{1/2}$ (h)	E_a (kcal/mol)
NCD-BS	6.1	–
D _{ART} -BS	1.5	–
NCD-STAA-BS	2.2	92.2
D _{ART} -STAA-BS	0.6	56.8

These results suggest that the two dimeric forms obtained in different experimental conditions, *i.e.* the NCDs and the artificial dimers, have different quaternary structures. Interestingly, the two variants conserved the same relative stability, with the STAA mutant displaying the faster dissociation rate.

A deeper comparison of the behavior of NCD and D_{ART} dimers has been performed for the STAA-BS-RNase variant, by measuring the dissociation rate at different temperatures. Straight lines were obtained at all the temperature explored (in the range 15-37°C), confirming the homogeneity of all the protein samples. Furthermore, the Arrhenius plot of the kinetic constants allowed the calculation of the activation energy of the dissociation process, reported in Table 6, which were 92.2 Kcal/mol and 56.8 Kcal/mol for NCD and D_{ART}, respectively.

These results clearly indicated that NCD and D_{ART} dimers have different quaternary structures, and consequently different stabilities, in solution. However, the fast dissociation of the artificial dimers prevented any further characterization in solution.

Crystal growth

The extensive search for crystallization conditions has produced different results in the case of the native enzyme and of its STAA variant. Crystals of D_{ART}-BS grew in a very short time (about 12 hours), showed an even shorter lifetime (about 10 hours), were quite disordered and their diffraction power decayed after few minutes of X-ray exposure. As a consequence, only few diffraction images were recorded. In the case of D_{ART}-STAA-BS, the higher stability of the crystals allowed a complete X-ray diffraction data collection at 2.48 Å resolution. Unfortunately, in addition to their low diffraction power, essentially due to their high solvent content (about 60%), crystals are twinned by *pseudo*-merohedry (see section Materials and Methods of this chapter).

Crystallographic analysis

The features of D_{ART}-STAA-BS crystals (low diffraction power and twinning) made both solution and refinement processes not trivial. The structure was solved by molecular replacement using the BS-RNase monomer as a search model (PDB entry 1n1x). The asymmetric unit contains two protein chains, whose position and orientation are compatible with the swapping of their C-terminal arms. The structure was refined at 2.48 Å resolution to R/Rfree values of 0.170/0.186 and the final model includes 1889 protein atoms and 35 water molecules. The average B-factor computed for all protein non-hydrogen atoms is 16.4 Å². A full list of refinement statistics is reported in Table 7, and the crystallographic structure of D_{ART}-STAA-BS is reported in Figure 20.

Table 7. Data collection and refinement statistics. Values in parentheses refer to the highest resolution shell.

<i>Data collection</i>	
Space group	$P2_1$
Unit-cell parameters (Å, °)	$a = 28.80$ $b = 79.04$ $c = 75.38$ $\beta = 101.05$
Resolution range (Å)	20.0–2.48 (2.57–2.48)
Measured reflections	31 997
Unique reflections	9497 (467)
Rmerge (%)	6.3 (19.1)
Mean $I/\sigma(I)$	15.9 (4.5)
Completeness (%)	78.9 (39.0)
Multiplicity	3.4 (2.0)
Twin law	$h, -k, -h-l$
Twin fraction	0.49
<i>Refinement statistics</i>	
Resolution limits (Å)	20.0–2.48
R factor (%)	17.0
R_{free} (%)	18.6
Number of reflections used in the refinement	8385
Number of reflections used for R_{free} calculation	420
No. of protein atoms	1889
No. of ions	2
No. of water molecules	35
<i>Average B</i>	
Protein, overall (Å ²)	16.4
Solvent atoms (Å ²)	14.3
Rmsd bond lengths (Å)	0.009
Rmsd bond angles (°)	1.289
<i>Ramachandran statistics</i>	
Favoured region (%)	93.6
Allowed region (%)	5.20
Outliers (%)	1.20

The quality of the electron density maps allowed a detailed description of nearly the whole molecule. The two chains share an almost identical structure, as judged from their very low root mean square deviation (RMSD) of 0.4 Å (RMSD calculated for main chain atoms excluding residues 111-116). No major topological variation has been found in the common ribonuclease fold adopted by each subunit (defined as the structural unit formed by residue 1-110 of one chain, the core of the protein, and residues 117-124 of the other chain, the swapped domain). Indeed the RMSD between one of the two subunits and the monomeric enzyme (PDB entry 1n1x) is ~0.7 Å (RMSD calculated for main

chain atoms of residues 1-110 and 117-124). A phosphate ion is bound to each composite catalytic site stabilizing the dimer by mimicking the enzyme substrate, as typically observed in several other ribonuclease structures (Liu, *et al.*, 2001; Merlino, *et al.*, 2005b; Merlino, *et al.*, 2008; Merlino, *et al.*, 2009a; Merlino, *et al.*, 2012).

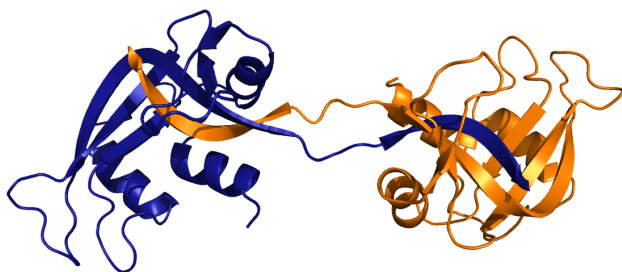


Figure 20. Crystallographic structure of D_{ART}-STAA-BS represented as cartoon. The two chains are colored differently to highlight the C-terminal swapping.

The electron density associated with the hinge region (residues 111–116), which connects the swapped domain to the rest of the protein, is continuous for the A subunit (Figure 21a) but less well-defined in the case of the B subunit. The conformation of segment 111A-116A is stabilized by a strong intra-chain H-bond between the side chains of K113 and S115 (Figure 21b).

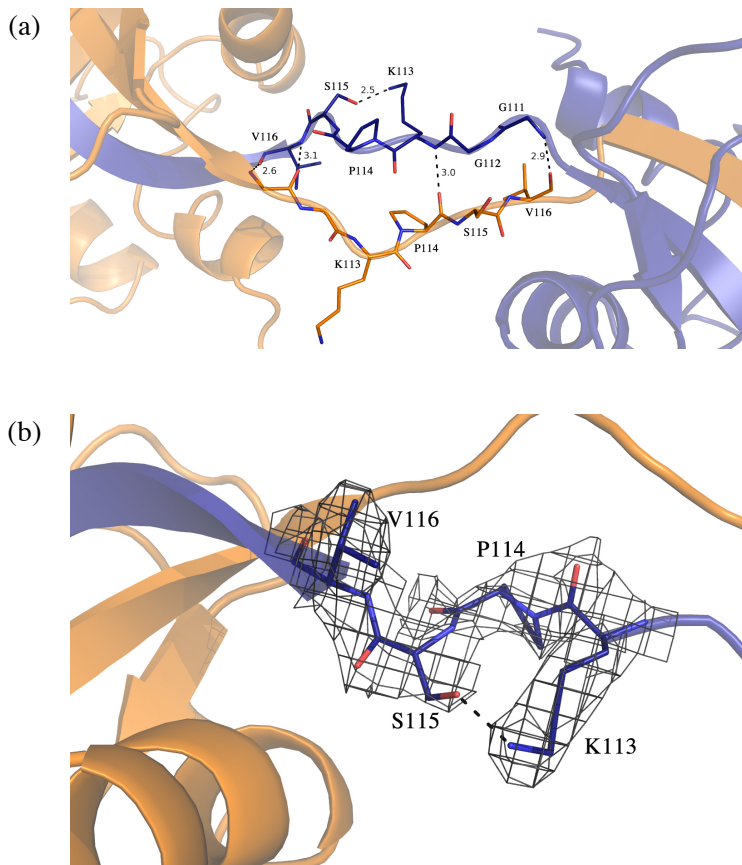


Figure 21. (a) Structure of the hinge peptides of D_{ART} -STAA-BS. Distances are reported in Angstrom. Chain A is colored blue while chain B is colored orange. (b) $2F_o - F_c$ electron density map of the hinge peptide of chain A contoured at 1.0σ .

None of the two hinge peptides adopts the β -sheet conformation shown by the same regions in the CD-RNase A structure. In the latter, the N113 enforces the hydrogen bonding network in the β -sheet alignment of the two hinge peptides. In the seminal enzyme the presence of a lysine in that position replaces the stabilizing H-bond with a positive charge repulsion. Moreover, the mutations of glutamic acid 111 with glycine and tyrosine 115 with serine confer a much higher flexibility to this region in the structure of the C-dimer of BS-

RNase when compared to that of CD-RNase A, as indicated by the low definition of the electron density of the B subunit hinge peptide, and thus to the whole swapped dimer. As a consequence, in the case of $D_{\text{ART}}\text{-STAA-BS}$ there is a decrease in the number of the inter-hinge contacts. It has to be noticed that in the case of the C-swapped dimers of both RNases the only additional interactions formed upon domain swapping (open interface) involve residues of the hinge peptides. In particular, the swapped dimer is stabilized by the burying at the interface between the two hinge peptides of an area of about 390 and 520 Å² for $D_{\text{ART}}\text{-STAA-BS}$ and CD-RNase A, respectively (Figure 21b). The peculiar structural features of the hinge segments produce a quaternary organization of $D_{\text{ART}}\text{-STAA-BS}$ very different from that of CD-RNase A (Liu, *et al.*, 2001). Indeed, after superposition of one subunit of each dimer, a further rotation of about 150° has to be applied in order to best overlap the second subunits (Figure 22).

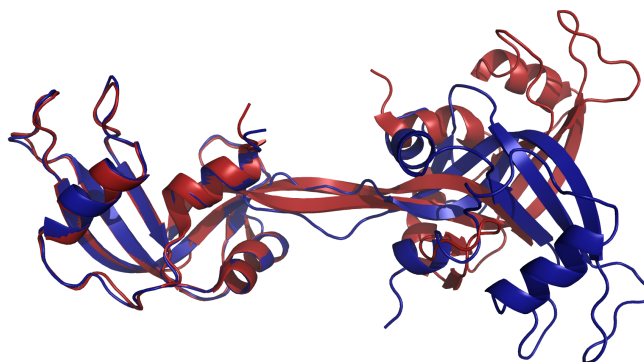


Figure 22. Comparison of the overall structures of $D_{\text{ART}}\text{-STAA-BS}$ (in blue) and CD-RNase A (in red). Only one subunit (composed by residues 1-110 of chain A and residues 117-124 of chain B) of each dimer has been superposed; a further 150° rotation is needed to best superpose the second structural units (residues 1-110 of chain B and residues 117-124 of chain A) of the two dimers.

DISCUSSION

The comparison of the swapping properties of RNase A and BS-RNase can provide interesting clues to investigate the 3D domain swapping mechanism and its functional implications. Despite the high similarity of their amino acid sequence (more than 80% identity), the two proteins have a completely different behavior with respect to the swapping (Sica, *et al.*, 2004; Picone, *et al.*, 2005; Merlino, *et al.*, 2008; Merlino, *et al.*, 2009b; Merlino, *et al.*, 2012). This has been confirmed by the crystallographic model of D_{ART}-STAA-BS. In BS-RNase the exchange of the C-terminal arms between the two subunits produces a quaternary organization very different with respect to that observed in the case of CD-RNase A (the RNase A dimer obtained by the swapping of the same structural elements). This variation is mainly due to the change in the structure adopted by the hinge peptides (111-116) of the two enzymes. In particular, a key role is played by the lysine residue at position 113. On this respect, it is worth to mention here that the lyophilization from acetic acid of the monomeric variant of BS-RNase where lysine 113 was replaced with asparagine, i.e. the residue present at the same position in RNase A, produces a higher amount of the C-terminal swapped dimer and other multimers in comparison with the wild-type enzyme (Gotte, *et al.*, 2012). The substitution also results in an increased stability of the oligomeric forms (Gotte, *et al.*, 2012). The other two substitutions present in the hinge peptide of the seminal enzyme, E111G and Y115S, confer a high conformational flexibility to this region, facilitating the dissociation of the dimer in solution.

Interestingly, no traces of a N-terminal swapped (artificial) dimer of BS-RNase and STAA-BS-RNase have been found in our experimental conditions. This unexpected result could be due to a quick dissociation of the dimer or to a very high energy barrier separating the monomer from the NCD form. The experimental conditions, in turn, could push the refolding pathway towards the formation of the C-swapping dimer, which is formed more easily despite its low

stability, and represents the highest quantity (about 16%) among the aggregated forms recovered from the size exclusion chromatography.

As far as the biological implications are concerned, unfortunately the fast dissociation of the artificial dimers in solution hampered a functional characterization, although it is possible to foresee that they should be devoid of any significant antitumor activity. Indeed, modeling studies indicate that the X-ray structure of D_{ART} -STAA-BS fits nicely in the RI cavity, as shown in Figure 23, thus reducing the high potential associated with 3D domain swapping.

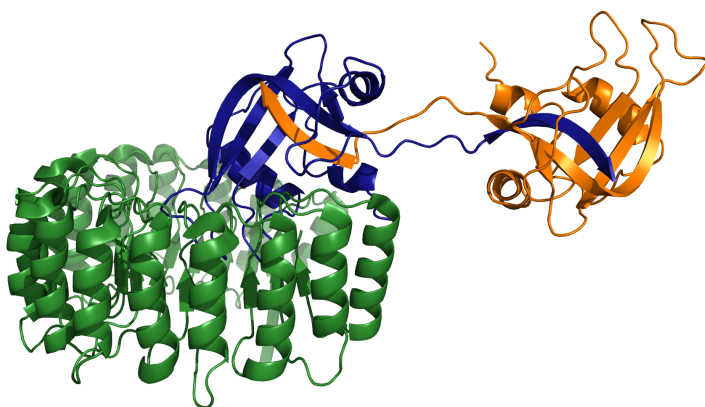


Figure 23. Putative model of the complex between the ribonuclease inhibitor (PDB entry 1z7x) and D_{ART} -STAA-BS in two different orientations.

Based on these considerations, it can be inferred that the C-swapping of BS-RNase represents an unfavorable event with respect to the N-swapping, therefore the substitutions of E111, N113, and Y115 of RNase A with G111, K113 and S115 in BS-RNase should be considered a favorable event in the evolution, since they reduce the effects of C-terminal swapping. Finally, these data enforce the role of the disulphide bridges linking the two subunits of BS-RNase, which prepare the N-terminal swapping (Mazzarella, *et al.*, 1995) and facilitate its occurrence under mild physiological conditions, giving the right orientation to the subunits.

MATERIALS AND METHODS

Protein production and purification

Monomeric and dimeric forms of BS-RNase, as well as its mutant G16SN17TP19AS20A (STAA-BS-RNase), have been kindly provided by the group of Prof. Picone, Department of Chemical Sciences, University of Naples “Federico II”, prepared and purified as described in (Ercole, *et al.*, 2003; Picone, *et al.*, 2005; Ercole, *et al.*, 2007).

The non-covalent dimeric forms of BS-RNase and of its mutant STAA-BS-RNase have been obtained by following two different protocols. The non-covalent dimer obtained by selective reduction of the swapped isoform (MxM) of the covalent dimer are referred to as NCD-BS-RNase and to NCD-STAA-BS-RNase, respectively. Non-covalent dimers obtained by lyophilization upon 40% acetic acid incubation of the monomeric derivatives (mBS-RNase and STAA-mBS-RNase) are referred to as artificial dimers (D_{ART}). The monomeric forms are both obtained by iodoalkylation of cysteine residues 31 and 32 by iodoacetamide (IAA), as reported elsewhere (Merlino, *et al.*, 2012). In both cases the dimers were purified at 4°C by Size Exclusion Chromatography (SEC) of Sephadex G-75 (1.5 x 72 cm).

Cathodic non-denaturing polyacrylamide gel-electrophoresis was performed at pH 4.5, following the protocol described in (Gotte, *et al.*, 2012).

Dissociation studies

The dissociation kinetics of the dimers has been analyzed following the amount of dimeric and monomeric proteins as a function of the incubation time at different temperatures, as previously reported (Merlino, *et al.*, 2012). At regular time intervals, aliquots of the protein samples were loaded on a Superdex 75 HR 10/30 column on FLPC system (Pharmacia, Uppsala, Sweden) equilibrated and eluted in 0.1 M Tris/HCl pH 7.3, 0.13 M NaCl.

Crystallization and data collection

Solutions of purified D_{ART} obtained from mBS (D_{ART} -BS) and STAA-mBS (D_{ART} -STAA-BS) at concentration of about 20 mg/mL in 25 mM sodium phosphate pH 6.7 have been used to search for the crystallization conditions. Initial screening has been carried out using Hampton Research Crystal Screen 1, Crystal Screen 2, Index, SaltRx and PEG/Ion at 20°C by the hanging-drop and sitting-drop vapor-diffusion methods. Drops consisting of 0.5 μ L protein mixed with an equal volume of precipitating solution were equilibrated against 500 μ L of reservoir solution. Best crystals of D_{ART} -BS have been grown by using a crystallization solution containing 36-40% w/v PEG4K, 0.1 M lithium chloride, while crystals of D_{ART} -STAA-BS have been obtained in 0.1 M sodium cacodylate buffer pH 6.2 or 16-20% w/v PEG35K, 0.2 M lithium chloride, 0.1 M sodium cacodylate buffer pH 6.0 and 3% v/v acetonitrile.

X-ray diffraction data have been collected at 100 K at the ELETTRA Synchrotron (Trieste, Italy) using a 165 mm CCD detector from MAR-Research. Diffraction data have been processed using the program suite HKL2000 (Otwinowski, *et al.*, 1997). A summary of the indicators commonly used to estimate the quality of datasets is given in Table 7. *Pseudo*-merohedral twinning (twin law h,-k,-h-l and twin fraction 0.49) have been detected using phenix.xtriage from the Phenix suite (Adams, *et al.*, 2010).

Structure determination and refinement

The structure of D_{ART} -STAA-BS has been solved by molecular replacement using the structure of BS-RNase monomeric derivative (PDB entry 1n1x) as a search model. The refinement has been carried out with REFMAC5 (Vagin, *et al.*, 2004) and Phenix (Adams, *et al.*, 2010). Several alternating cycles of positional refinement, energy minimization, individual temperature factor refinement and manual model building have been performed. Model rebuilding has been carried out using Coot (Emsley, *et al.*, 2010). The program PROCHECK

(Laskowski, *et al.*, 1993) was used to analyze the quality of the final structures. The refinement statistics are presented in Table 7. Figures have been drawn using Pymol (Schrodinger, 2010).

CHAPTER IV – TOO MUCH IS A BAD THING

INTRODUCTION

A striking case of intermolecular interaction is given by protein molecules that can produce large aggregates through a self-association process, that eventually become insoluble and precipitate. The formation of aggregates by numerous proteins and peptides unrelated in sequence, structure, and function suggests that the propensity to aggregate is an innate property of polypeptides (Dobson, 2004; Rousseau, *et al.*, 2006). Proteins occur *in vivo* in many cases at concentrations above their critical concentration for aggregation (Baldwin, *et al.*, 2011), but energy barriers against aggregation keep them in a metastable soluble state (Buell, *et al.*, 2012). Understanding the molecular determinants behind biological aggregation has been one of the major goals of structural biologists for the last thirty years. Protein deposits are the hallmark of several severe human diseases, including Alzheimer's, Parkinson's, and Huntington's diseases or type 2 diabetes (Dobson, 2001). Amyloidoses, for example, are a range of clinical disorders caused by extracellular deposition of β -sheet-rich elongated, unbranched fibrils, known as amyloids, derived from aggregation of misfolded or partially unstructured peptides and proteins (Pepys, 2006). Sometimes amyloidoses are referred to as "conformational diseases" in order to emphasize that the aggregation pathway occurs *via* a conformational change, which induces the formation of non-native interactions between adjacent protein molecules (Carrell, *et al.*, 1997). However, not all reported aggregates and fibrils are enriched in β -sheet-structure (Dykes, *et al.*, 1978; Lomas, *et al.*, 2002) or they are associated with deposition diseases (Holmes, *et al.*, 1990; Downing, *et al.*, 1999).

Three main mechanisms exist by which proteins can specifically self-associate to form fibrils: cross β -spine (Nelson, *et al.*, 2005), end-to-end stacking (Harrington, *et al.*, 1997; Elam, *et al.*, 2003), and 3D domain swapping (Bennett, *et al.*, 2006; Guo, *et al.*, 2006). Although cross β -spine aggregation has drawn most of the attention of the scientific community, end-to-end stacking is the mechanism by which notable proteins do aggregate. The two cases of

sickle cell hemoglobin (Harrington, *et al.*, 1997) and SOD1 (Elam, *et al.*, 2003) are the best-known.

As already described above, 3D domain swapping gives rise to dimer having two structural units that are very similar to the monomeric species and formed by residues from the two protein chains. Such a dimer is referred to as closed-ended as there are no unsatisfied (exposed) domains. The open interface is the additional interface not present in the monomer and unique in the dimeric species. Open-ended runaway swapped oligomers can also form by consecutive exchange of the swapped domain of three or more monomers. The hypothesis that open-ended runaway swapping is implicated in the formation of protein fibrils has been supported by the observation that human prion protein (Knaus, *et al.*, 2001), cystatin C (Janowski, *et al.*, 2001), and β 2-microglobulin (Liu, *et al.*, 2011), which form fibrils related to diseases, are domain-swapped.

Even closed-ended swapped dimers do have structural features that facilitate aggregation. Indeed, the stability of these dimers is mostly guaranteed by the closed interface that pre-exists in the monomeric form. Moreover some extra-stability is gained upon generation of the open interface, which is usually less important, as the two structural units are often linked by two flexible hinge peptides. Therefore, in contrast with common protein dimers that have acquired the quaternary organization through the development of an extensive open interface, the domain-swapped dimers are often endowed with a larger freedom in their quaternary assembly (Merlino, *et al.*, 2005a). This feature may greatly facilitate self-recognition processes and make these dimers more prone to association.

Part of this PhD project has been focused on the characterization of a variant of human pancreatic ribonuclease (desHP), forced to swap by deleting five residues (16-20) in the loop linking the N-terminal segment (residues 1-15) to the core of the protein (Russo, *et al.*, 2000). The segment is partially folded as an α -helix and embodies some residues important for the constitution of the active site (His12 and Asn13). In the variant this helix cannot take over the position

occupied in the wild type enzyme and it remains exposed to the solvent, eventually promoting the formation of a stable closed-ended swapped dimer, whose specific activity is virtually identical to that of the monomeric wild-type enzyme (Russo, *et al.*, 2000). The structural characterization by X-ray crystallography reveals that desHP molecules give rise to a densely packed rod-like structure, which extends infinitely along one crystallographic axis. On the basis of this arrangement the formation of fibrils in solution has been predicted and indeed observed.

RESULTS

desHP overall structure

desHP was crystallized and diffraction data were collected up to a resolution of 2.70 Å. The crystal belongs to space group $P2_12_12_1$. The molecular replacement approach produced a plausible solution with 4 copies of the monomeric search model in the asymmetric unit and a very high solvent content (about 60%) that could explain, at least in part, the poor diffracting power of the crystal.

The final model, which includes 3836 protein atoms, 211 water molecules and 13 sulfate ions, has been refined to R/Rfree values of 0.199/0.248. The statistics of refinement are listed in Table 8.

Table 8. Data collection and refinement statistics. Values in parentheses refer to the highest resolution shell.

Data collection	
Space group	$P2_12_12_1$
Unit-cell parameters (Å)	$a = 71.51, b = 74.32,$ $c = 128.53$
Resolution range (Å)	20.0–2.70 (2.80–2.70)
Measured reflections	119522
Unique reflections	19442 (1909)
R_{merge} (%)	5.7 (16.1)
Mean $I/\sigma(I)$	28.6 (9.9)
Completeness (%)	100 (100)
Multiplicity	6.1 (6.3)
Dimers per asymmetric unit	2
Refinement statistics	
Resolution limits (Å)	20.0–2.70 (2.80–2.70)
R factor (%)	19.8 (28.0)
R_{free} (%)	24.6 (36.0)
No. of reflections used in refinement	18448 (1806)
No. of reflections used for R_{free} calculation	992 (102)
No. of protein atoms	3836
No. of ions	17
No. of water molecules	211
Average B factors (Å ²)	
Protein, overall	26.7
Solvent atoms	24.3
R.m.s.d., bond lengths (Å)	0.014
R.m.s.d., bond angles (°)	1.7
Ramachandran statistics (%)	
Favoured region	95.8
Allowed region	3.8
Outliers	0.4

In order to facilitate reading, residue numbering of the parent protein (wild-type human pancreatic ribonuclease) has been adopted; therefore, in the present case, Ser15 is covalently linked to Ser21. The monomers are associated as N-terminal swapped dimers with approximate two-fold symmetry. In each dimer, residues 1-14 of one chain and residues 24-125 of the partner chain form a structural unit (SU) that closely resemble the native enzyme (PDB entry 1e21); with respect to the latter the RMSD based on C α of the four SU is in the range 0.30-0.35 Å.

Each active site is assembled with residues of the two chains and is practically indistinguishable from that of the wild-type enzyme (Figure 24).

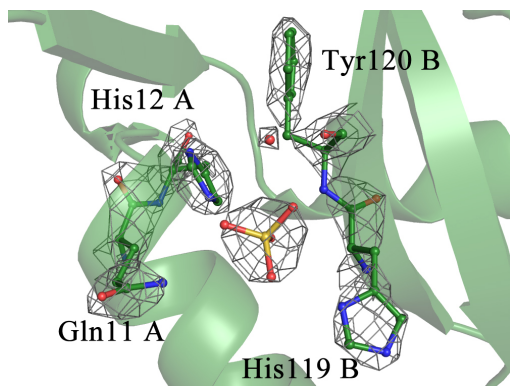


Figure 24. A composite active site generated by residues belonging to subunits A and B of one of the two dimers in the asymmetric unit. The electron density map is calculated with $2F_o - F_c$ Fourier coefficients and is contoured at the 2.2σ level.

The swapping is associated with a four residues long hinge peptide, whose conformational features control the quaternary assembly of the dimer by determining the relative orientation of the N-terminus with respect to the body of the chain. The two dimers have slightly different quaternary structures; this is shown in Figure 25a, where the two dimers are drawn after superposition of one SU. The rotation that has to be applied to best superpose the second SU of the two dimers amount to 23° and is achieved with a modest energy penalty through very small variations of main chain torsion angles of the four hinge residues.

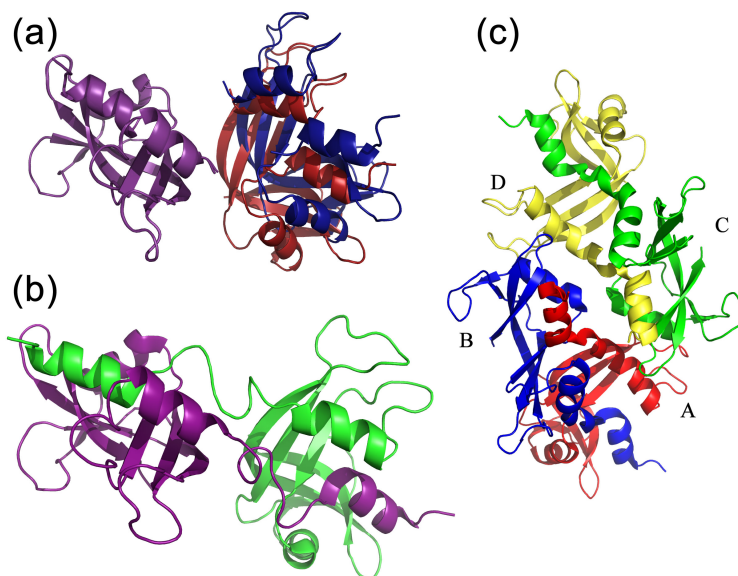


Figure 25. (a) View of the two dimers in the asymmetric unit with monomers A and C (purple) superposed. A further 23° rotation is needed to best superpose the second monomers of the two dimers: B (red) and D (blue). For clarity, the hinge peptides have not been drawn. (b) The structure of a domain-swapped dimeric mutant of human pancreatic ribonuclease (PDB entry 1h8x; Canals et al., 2001) is shown as a cartoon for comparison with the desHP dimers. (c) The two dimers in the independent unit are shown as cartoons: chains A (red) and B (blue) form the first dimer and chains C (green) and D (yellow) form the second dimer.

In both dimers the position of Asp14 is mostly unperturbed by the swapping: the carboxylate group and the carbonyl oxygen are hydrogen bonded to the side chain of Tyr25 and His48, respectively, of the partner chain; these interactions are typical features of pancreatic-like monomeric ribonucleases. The carboxylate group also makes a further hydrogen bond with O γ of Ser21. The polyserine stretch (Ser15, Ser21, Ser22, Ser23), which constitutes the hinge loop, forms a rather regular three-fold helix stabilized by internal hydrogen bonds. In comparison, the swapped dimer (PM8) of a mutated form of the human pancreatic ribonuclease (Canals, *et al.*, 2001), shown in Figure 25b, has a more opened structure and the two SU are further apart with respect to both desHP dimers due to the greater length of the hinge peptide. The association of the two dimers (formed by chains A-B and C-D) in the asymmetric unit is also

shown in Figure 25c. Interestingly, the extension of the open surface (about 390 Å² and 255 Å² for the two dimers, respectively) is much smaller than that between dimers (see below). It should be noted that both dimers display a *pseudo*-two-fold symmetry, whose axis is approximately orthogonal and almost intersects the crystallographic screw axis parallel to **c**.

Supramolecular assembly

The molecular packing displays particularly interesting features. The two dimers are related by an approximate rotation of 90° about an axis that is practically coincident with the crystallographic two-fold screw axis parallel to the **c** axis, combined with a *c*/4 screw translation. These symmetry operators, together with the two-fold symmetry of the dimers, build up rods parallel to the **c** axis with an approximate 4₃22 symmetry. The interdimer association is very strong, burying a surface of about 900 Å² (930 Å² and 820 Å² for the two crystallographically independent dimer-dimer interfaces) that propagates along the **c** axis producing tightly packed rods. By contrast, contacts between rods are weak (Figure 26).

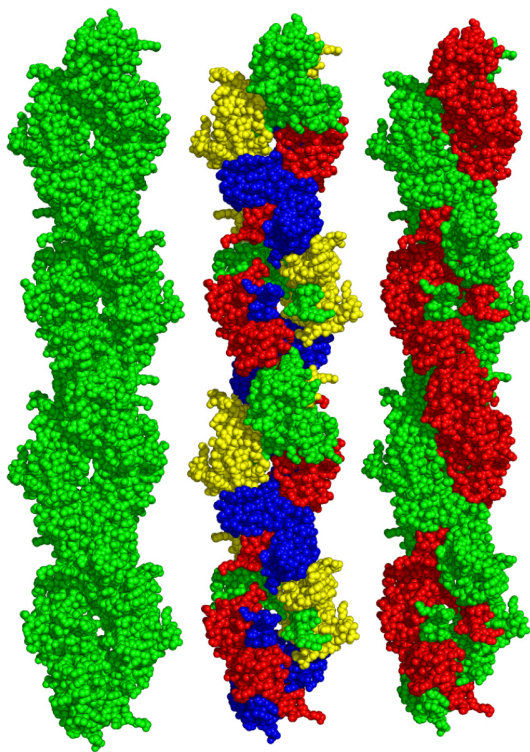


Figure 26. View of the rod structure along the **a** axis. The rod on the left is colored green. In the central rod, each chain of the two dimers in the asymmetric unit has a different color according to the color code used in Figure 25c. In the third rod, the chains are colored to highlight the antiparallel double-helical arrangement of the monomers.

Sulfate anions

Thirteen sulfate anions have been identified in the electron density map of the asymmetric unit of the crystal, confirming that this ion plays a fundamental role in the crystallization process. With the exception of the anions positioned at the four active sites, as already observed for many others structures of RNases (Vitagliano, *et al.*, 1999; Vitagliano, *et al.*, 2000), all the remaining ones are located on positive patches of the rod surface. The electrostatic features of the surface calculated with the software APBS (Baker, *et al.*, 2001) and the bound anions is shown in Figure 27.

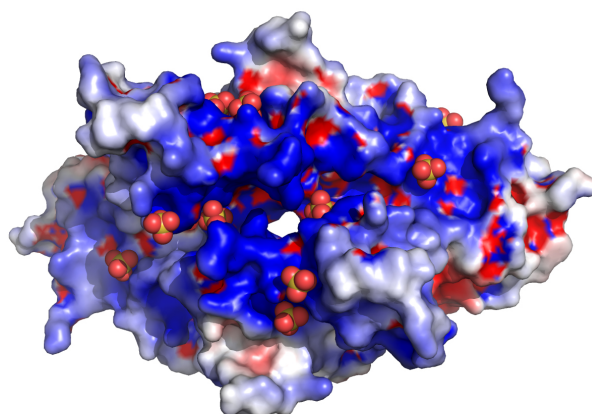


Figure 27. Electrostatic surface of the two dimers in the asymmetric unit. Sulfate ions are shown in ball-and-stick representation.

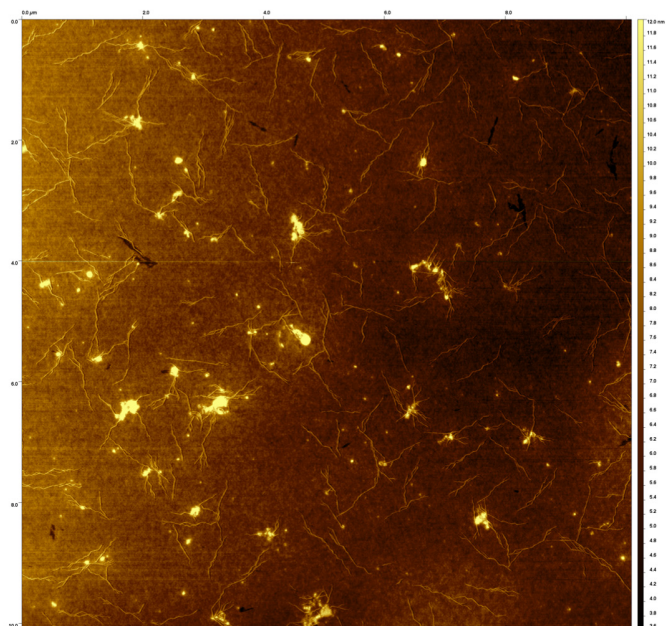
In particular, four sulfates are bound at equivalent position of the four monomers and interact with Arg32 and Arg33. Two more sulfate ions are located at nearly equivalent positions mediating the interaction between rods. The remaining two ions are located at the border of the interdimer surfaces. Thus, the presence of sulfate ions may be important for the stabilization of such a positively charged rod-like structure.

Solution study

In solution desHP displays a strong tendency to aggregate, making difficult the growth of crystals suitable for X-ray diffraction. Indeed, at low ionic strength, the protein solution is opalescent even at very low concentration (1 mg/mL). Increasing ionic strength (0.3 M in NaCl), it was possible to prepare a solution more concentrated in protein (up to 2 mg/mL) from which crystals were obtained. Moreover, a tendency to form fibrils is also suggested by the features of the observed crystal packing. Eventually, fibrils were grown by incubating a solution of the protein at 0.6 mg/mL concentration in 0.1 M Tris-acetate pH 7.1, 0.2 M NaCl, both in the presence and in the absence of 0.1 M sodium sulfate, at room temperature for two weeks. Samples of the solution were deposited on a

mica surface, dried and washed with water (see section Materials and Methods of this chapter). The substrate was then imaged with an atomic force microscope, using tapping mode in air. Several AFM images (Figure 28) clearly showed the presence of long fibrils with a highly homogeneous cross sectional diameter of about $18 \pm 4 \text{ \AA}$.

(a)



(b)

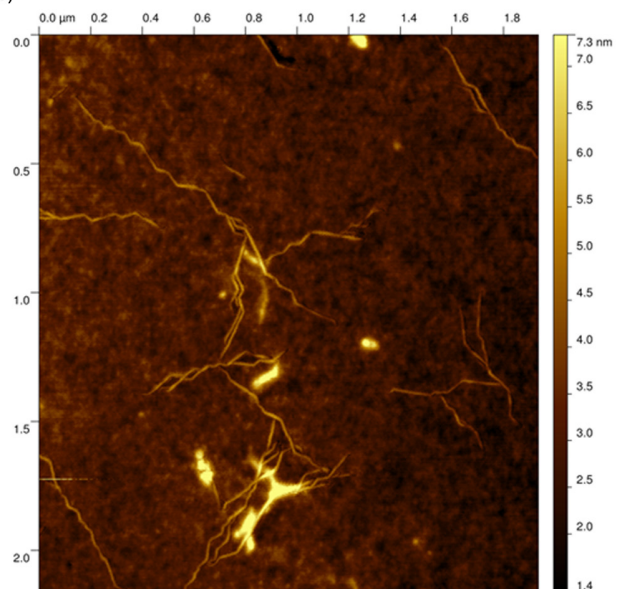


Figure 28. AFM images of desHP fibrils at low (a) and high (b) magnification.

In Figure 29, a histogram of the heights derived from the images is reported. Moreover, fibrils appear not to be twisted, in contrast to most amyloid fibrils described to-date (Figure 28b).

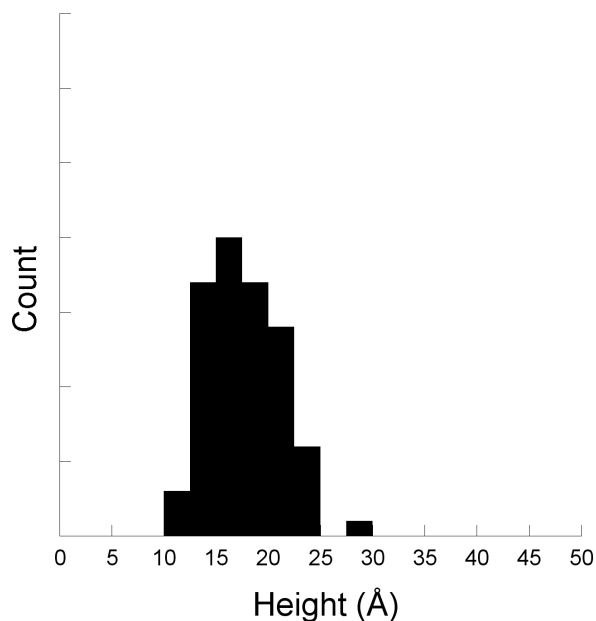


Figure 29. Histogram of desHP fibril heights derived from AFM images.

The ability of these fibrils formed in solution to bind ThT and to lead to the enhanced fluorescence emission at 480 nm upon excitation at 440 nm has also been tested. No difference in fluorescence signal was observed between a fresh protein solution and the fibril-containing solution (not shown).

DISCUSSION

Although proteins can be robust against point mutations, enduring significant numbers of amino acid substitutions with little changes in protein structure and function (Taverna, *et al.*, 2002), it is not uncommon to find cases in which adding/deleting short fragments to/from their sequences partially destabilizes the native fold and favors the formation of dimers or higher-order structures. Eisenberg and co-workers converted RNase A in a protein able to form fibrils by simply expanding the hinge peptide connecting the core domain of the protein with the exchanging C-terminal β -strand, with an insertion of ten glutamine residues (Sambashivan, *et al.*, 2005). The designed amyloid-like fibrils of RNase A contain 3D domain-swapped molecules endowed with enzymatic activity. The β -sheet arrangement arises from stacking of the expanded hinge peptides.

Here, it is shown that the deletion of five residues in the hinge loop of human pancreatic ribonuclease induces the formation of a domain-swapped dimer that leads to the generation of linear aggregates of desHP molecules, as revealed by the crystal packing, in which domain swapping and flexibility of the dimeric assembly are intimately involved. Domain swapping provides the cohesive interaction energy for the formation of dimers through an extensive interchain closed interface, whereas the size of the open interface is often small; this feature confers flexibility to the quaternary structure, particularly when the hinge peptide is long, that facilitates self-recognition of the dimers and favors further aggregation. In this respect, the behavior of domain-swapped dimers is quite different from that of dimers whose quaternary structure is held together only by the interactions across an open interface; in this case the stability of the dimer relies on the extension of the interface and this necessarily produces a rather rigid quaternary assembly. In the present case, two dimers with slightly different quaternary structures are present in the crystal: their close interfaces are nearly the same and amount to about 1500 \AA^2 , whereas the open interface is small and only slightly different (about 390 \AA^2 and 250 \AA^2 , respectively). These two di-

mers alternate on top of each other forming tightly packed rods, which sit on the crystallographic two-fold screw axis parallel to c . In the rod, the molecules form a left-handed helix with approximate four-fold symmetry. The small difference in the quaternary assembly of the dimers efficiently increases the complementarity of the two crystallographically independent dimer-dimer contact surfaces that are 930 and 820 Å², respectively. These tightly packed rods run side by side through the crystal with weak lateral contacts. The present case closely resembles the one described for SOD1 (Elam, *et al.*, 2003), in which the interfaces within dimers and between dimers along the rod axis are approximately similar and amount to about 650 Å². These values are much higher in our case because the interface between dimers is about 50% higher and, most remarkably, the one stabilizing the dimeric form through domain swapping amounts to about 2000 Å². These data strongly underline the cooperative role between domain swapping and stacking interactions in building up supramolecular structures. It should be also noted that desHP rods, as in the case of SOD1, are not formed by crystallographic operators only, but its generation involves a non-crystallographic symmetry operation. This suggests that the formation of filaments is not simply a necessary consequence of crystallization (Elam, *et al.*, 2003).

Interestingly, each rod of desHP can also be alternatively described as formed by two intertwined antiparallel four-fold helices: one helix is formed by one of the monomers, monomer A, C, A*... of the dimers successively stacked along the rod, and the second helix is built up by monomers B, D, B*... of the same dimers, where the star indicates screw symmetry related units. Domain swapping between A and B, C and D..., cross-links monomers belonging to different helices and strongly stabilizes the whole structure.

The crystal structure of desHP described in this work suggests that the protein could form fibrils in solution, and this prediction was confirmed by the observation of long, unbranched thin fibrils by AFM. Filaments of several mi-

micrometers in length were observed in different samples, and the thickness was measured (see histogram of fibril heights in Figure 29).

Figure 30 shows a crystal plane perpendicular to the c axis; from the picture it is clear that fibrils are almost cylindrical in shape. A rough estimation of fibril width derived from the crystallographic model is about 25-30 Å. The diameter measured by AFM (≈ 20 Å) of these fibrils is smaller than the crystallographic dimensions suggest, a feature that can be explained by the fact that the AFM images of fibrils have been acquired in a dry state. The macromolecular crystal, on the other hand, is highly hydrated (about 60% of total volume) and more than 200 water molecules have been found to interact directly with the protein.

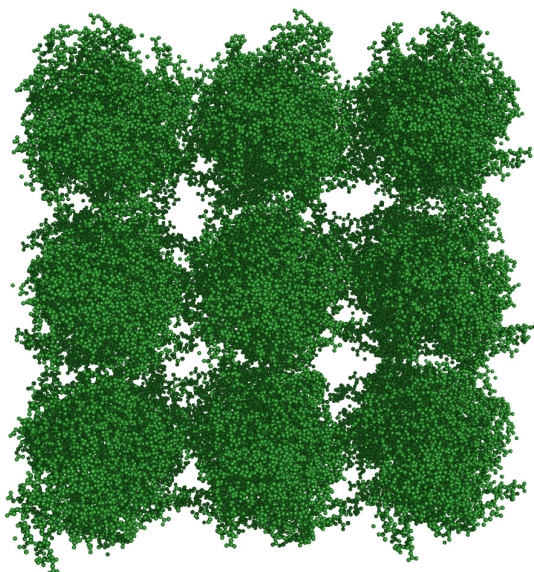


Figure 30. View of the structure along the crystallographic c axis.

Based on the AFM images, which show substantial amounts of amorphous protein, it appears that only a small fraction of the protein sample is converted into fibrils, rendering further biophysical examination, for example by CD spectroscopy, difficult. However, the fact that the formation of long, unbranched fi-

brils have been observed under quiescent conditions and at low concentrations suggests that these fibrils are indeed the same entities found in the crystal state. Moreover, the absence of enhanced ThT fluorescence strongly supports the hypothesis that the fibrils observed are not cross- β rich amyloid fibrils but rather fibrils derived from the fibrillar assembly observed in the crystal.

This result represents the first example of a structure-based prediction of protein fibril formation in solution, and in turn suggests that these fibrils can be modeled at atomic detail on the basis of the crystal structure.

CONCLUSION

3D domain swapping is a mechanism for forming elaborate, large assemblies that retain a protein's original fold, referred to as native-like aggregation. In particular, the interplay between the interchain closed interface and the quaternary structure flexibility, which are characteristic features of a swapping dimer, can play a crucial role in determining self-aggregation processes through a precise hierarchical organization of the native proteins. In this thesis, it has been shown how, starting from the observation of the peculiar supra-molecular assembly found in the crystal, the formation of long and unbranched protein fibrils has been predicted and indeed observed. These results highlight the importance of domain swapping in inducing fibril formation by a mechanism alternative to the most common runaway swapping.

MATERIALS AND METHODS

Crystallization and data collection

Protein production and purification were performed as described previously (Russo, *et al.*, 2000). Crystals of desHP were obtained after one week at 4°C by the hanging drop vapor diffusion method, using 22% (w/v) PEG 8000 and 0.1

M ammonium sulfate as precipitants, protein concentration 1.6 mg/ml, in a 0.1 M sodium citrate buffer pH 6.5.

X-ray diffraction data were collected at 100 K at the ELETTRA Synchrotron (Trieste, Italy) using a 165 mm CCD detector from MAR Research. Diffraction data were processed using the program suite HKL2000 (Otwinowski, *et al.*, 1997). A summary of the indicators commonly used to estimate the quality of datasets is given in Table 8.

Structure determination and refinement

The initial phase determination was carried out by molecular replacement method using the program PHASER (McCoy, *et al.*, 2007). The refined coordinates of des(1-7)HP-RNase (PDB entry 1e21) were used as a search model (Pous, *et al.*, 2001).

The refinement was carried out with the programs CNS v1.3 (Brunger, *et al.*, 1998) and REFMAC5 (Vagin, *et al.*, 2004). Several alternating cycles of positional refinement, energy minimization, individual temperature factor refinement and manual model building were performed. Model rebuilding was carried out using Coot (Emsley, *et al.*, 2010). Water molecules and sulfate ions were added manually into the model using the same software. The program PROCHECK (Laskowski, *et al.*, 1993) was used to analyze the quality of the final structures. The refinement statistics are presented in Table 8. Pymol (Schrodinger, 2010) was used to draw figures. The final model and structure factors have been deposited in the Protein Data Bank (PDB entry 4kxh).

Atomic Force Microscopy imaging

A 10 μ L drop of the protein solution (0.6 mg/mL) was deposited on a freshly cleaved mica, and let stand for about 30 minutes in air, then rinsed with MilliQ water to remove unbound molecules and salt dissolved in the buffer. The remaining water was blown away with nitrogen. Atomic force microscopy was performed under ambient conditions in intermittent contact mode using a Nan-

owizard II (JPK Instruments). Mikromasch NSC36 cantilevers with a resonant frequency in the range of 75-150 kHz and nominal spring constants of approx. 1 N/m were used. Height images were recorded at 2048x2048 pixels resolution. The scan rate varied between 0.2 and 1.0 Hz. All image data sets were analyzed using Gwyddion AFM software.

Fluorescence experiments

The fluorescence spectra of solutions of 1.2 μM Thioflavin T (ThT) were measured with a UV/Vis fluorimeter (constant excitation wavelength 440 nm, emission scanned from 450 to 600 nm). Then, the fluorescence of a solution of 0.6 mg/mL desHP fibrils in 0.3 M NaCl at pH 7.1 and 1.2 μM ThT was measured.

CHAPTER V – TWO IS BETTER THAN ONE

INTRODUCTION

Ribonucleases (RNases) are potentially cytotoxic agents thanks to their ability to degrade RNA, and therefore to inhibit protein biosynthesis at both transcriptional and translational levels (Ilinskaya, *et al.*, 2005). The cytotoxic activity of RNases is critically dependent on the ribonucleolytic activity and on the ability of these molecules to reach the cytosol and to degrade RNA by evading the action of the ribonuclease inhibitor (RI) (Youle, *et al.*, 1997; Leland, *et al.*, 2001a; Matousek, 2001; Makarov, *et al.*, 2003; Dickson, *et al.*, 2005). This intriguing feature has stimulated since many years numerous studies on the anti-tumor effects of RNases, leading to a slow but continuous progress toward the development of ribonuclease-based cancer therapeutics. Engineering of these enzymes is actively pursued in order to increase their cytotoxicity and selectivity towards malignant cells (Leland, *et al.*, 1998; Bretscher, *et al.*, 2000; Leland, *et al.*, 2001b; Dickson, *et al.*, 2003). The great interest in this class of proteins is confirmed by the high number of patents and papers recently published in this field and by ongoing clinical trials (Raines, 2013).

In the last decade, it has been shown that oligomerization is essential for providing special properties to RNases (Libonati, *et al.*, 2004). For example, the human pancreatic ribonuclease (HP-RNase), a monomeric enzyme very sensitive to RI inactivation, has been transformed into dimeric RNases, called HHP-RNase and HHP2-RNase, selectively cytotoxic for malignant cell lines (Di Gaetano, *et al.*, 2001; Merlino, *et al.*, 2009a). Dimers of bovine pancreatic ribonuclease (RNase A) formed by cross-linking with amine-reactive imido esters, inhibit tumor cell proliferation *in vitro* (Bartholeyns, *et al.*, 1976) and *in vivo* (Tarnowski, *et al.*, 1976; Bartholeyns, *et al.*, 1979). Cross-linked trimers of RNase A generated with a dimethyl suberimidate linker displayed higher toxicity towards a cervical carcinoma cell line with respect to the cross-linked dimers (Gotte, *et al.*, 1997). RI-evasive trimeric conjugates of G88C RNase A and

G89C HP-RNase, prepared by using a thiol-reactive homo-bifunctional linker efficiently inhibit tumor growth in tumor-bearing mice (Rutkoski, *et al.*, 2010).

Domain-swapped ribonucleases represent a class of RNase oligomers deserving a special interest. In particular, domain-swapped oligomers of RNase A were shown to be selectively cytotoxic toward tumor cells. Purified RNase A domain-swapped dimers, trimers, and tetramers (Gotte, *et al.*, 1999), obtained by lyophilizing the enzyme protein from 40% acetic acid solutions (Crestfield, *et al.*, 1963), display antitumor activities that increase remarkably as a function of the oligomer mass and, at the same time, show a complete lack of embryotoxicity (Matousek, *et al.*, 2003). Unfortunately, the short half-life time of domain-swapped oligomers hampers their use as cytotoxic agents. A peculiar example is given by BS-RNase (see Chapter III), that selectively kills human multidrug-resistant neuroblastoma cells *via* induction of apoptosis and severely reduces the appearance of lung metastases generated from Lewis lung carcinoma cells *in vivo*. The cytotoxic action of this enzyme is related to the survival of a non-covalent swapped dimer in the cytosol (Piccoli, *et al.*, 1992). This activity significantly increases when the dimeric enzyme is transformed in tetramers by double domain swapping (Gotte, *et al.*, 2012).

In this PhD project, it has been tried to use the 3D domain swapping mechanism to modulate RNase A self-assembly. The idea is to force the reciprocal exchange of polypeptide segments between two RNase A monomers by shortening the hinge peptide (also called hinge loop) connecting the N-terminal arm to the core of the protein. As illustrated in Chapter IV, this strategy has already been successfully applied to HP-RNase (Russo, *et al.*, 2000). The aim is the generation of a stable dimer to be used as a building block to create higher-order oligomers, with a longer half-life time and of defined subunit composition, stoichiometry, and quaternary structure. In particular, as in the case of desHP, a variant of RNase A (desBP) was forced to form a domain-swapped dimer by deleting five residues (16-20) in the hinge loop. This line of research is still under-

way and, at the moment, only a preliminary crystal structure of desBP has been obtained. The study of the oligomerization tendency of this dimer has revealed that under mildly denaturing conditions tetramers, and higher-order oligomers are formed. The final aim of the research is the structural characterization of the dimeric and tetrameric forms whose cytotoxic activity will also be analyzed.

RESULTS AND DISCUSSION

Overall crystal structure

desBP was crystallized and diffraction data were collected up to a resolution of 1.78 Å. The crystal belongs to space group P2₁2₁2. The asymmetric unit contains two half-dimers, each located on a crystallographic two-fold axis. The statistics of data collection are listed in Table 9.

Table 9. Data collection statistics. Values in parentheses refer to the highest resolution shell.

Space group	P2 ₁ 2 ₁ 2
Unit cell parameters (Å)	a = 66.5 b = 72.4 c = 52.8
Resolution range (Å)	50.0 - 1.78 (1.81 - 1.78)
Measured reflections	179381
Unique reflections	25149 (1235)
Rmerge (%)	8.4 (63.1)
Mean I/σ(I)	22.2 (2.9)
Completeness (%)	99.8 (100)
Multiplicity	7.1 (7.2)
Chains per asymmetric unit	2

Refinement of the structure is still underway. One of the two reconstructed dimers in the asymmetric unit is shown in Figure 31.

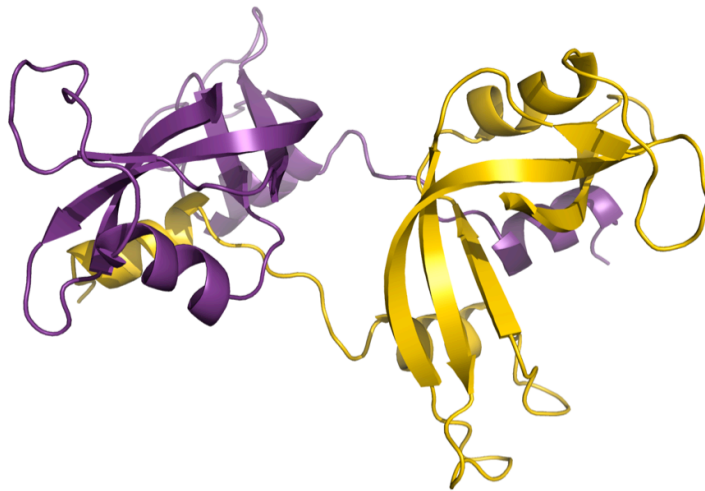


Figure 31. A desBP dimer represented in a cartoon representation.

As in the case of desHP (Pica, *et al.*, 2013), in order to facilitate reading, residue numbering scheme of the wild-type RNase A has been kept; therefore, in the present case, Ser15 is covalently linked to Ser21. The unit composed by the core of one chain (residues from 24 to 124) and the swapped N-terminal helix of the other chain (residues from 1 to 14) is defined as structural unit (SU). Thus, the four SU in the unit cell have the same folding of the wild-type parent protein, RNase A (PDB entry 1kf4).

The two dimers of desBP (each generated by a binary symmetry operation) have approximately the same quaternary organization: if any two SU belonging to each dimer is superposed, then a rotation of about 2° (χ angle) would overlap the second SU. χ is a suitable parameter for comparing the quaternary structures between pairs of dimers.

The quaternary organization of desBP has been compared to that of its human counterpart, desHP (PDB entry 4kxh), as well as to that of the N-terminal swapped dimer of wild-type RNase A (ND-RNase A), obtained upon lyophilization from acetic acid (PDB entry 1a2w) and the non-covalent swapped dimer of

BS-RNase (PDB entry 1tq9). The value of χ for each pair of dimers analyzed is reported in Table 10.

Table 10. χ angle for pair of N-terminal domain-swapped dimers of several pancreatic-like ribonucleases. The two crystallographically different dimers of desBP and desHP are arbitrarily denoted as desBP₁, desBP₂, desHP₁, and desHP₂.

	desBP ₁	desBP ₂	desHP ₁	desHP ₂	ND-RNase A
desBP ₁	-				
desBP ₂	1.7°	-			
desHP ₁	8.2°	9.1°	-		
desHP ₂	18.3°	19.2°	22.6°	-	
ND-RNase A	119.1°	119.8°	111.5°	126.3°	-
NCD-BS	95.0°	95.1°	101.6°	80.2°	160.2°

From the comparison of the χ values reported in Table 10, it appears clear that the quaternary organization of desBP is much more similar to that of desHP rather than ND-RNase A. The shortening of five residues in the sequence of RNase A, as well as in that of HP-RNase, leads to the lost of degrees of freedom in the respective swapped dimers, which is thus characterized by a lower conformational flexibility with respect to the swapped dimers formed by the wild-type proteins. The lower conformational flexibility constrains, in some extent, the two dimers towards a common quaternary organization.

Despite this similarity in the overall organization desBP shows a completely different crystal packing with respect to desHP, as well as a different solubility in solution. As already stated in Chapter IV, desHP has a great tendency to aggregate and, in fact, it forms fibrils in solution and it is organized as rod-like structures in the crystalline state. desBP is much more soluble than its human counterpart.

Oligomerization studies

desBP has been lyophilized from 40% acetic acid in order to promote the swapping of its C-terminal tail, thus leading to the formation of oligomers in

which the central subunits swap both their N- and C-terminal tails while the lateral subunits swap only their N-terminus. Preliminary results showed that tetramers, hexamers, and higher-order oligomers migrate differently on a cathodic gel electrophoresis under non-denaturing conditions (Figure 32).

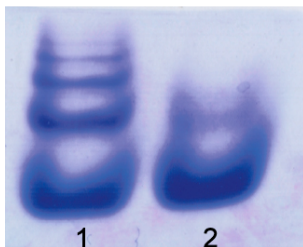


Figure 32. Cathodic gel electrophoresis under non-denaturing conditions. Lane 1: Upon lyophilization from 40% acetic acid, desBP gives rise to a several multimers that migrate differently on the gel. Lane 2: untreated sample of desBP is shown for comparison.

The tetramer can be easily separated from the oligomeric mixture by size-exclusion chromatography (SEC) with a Sephadex G-75 column (Figure 33).

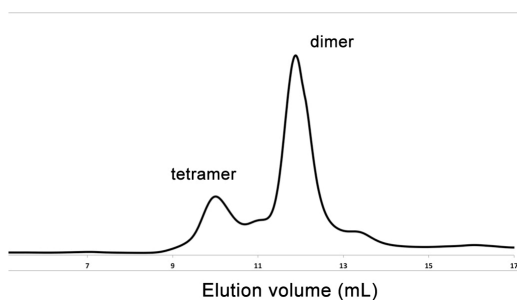


Figure 33. Size exclusion chromatogram of desBP after lyophilization from 40% acetic acid. Dimer and tetramer species can be easily separated.

The tetramer should have a quaternary organization similar to that of the model represented in Figure 34.

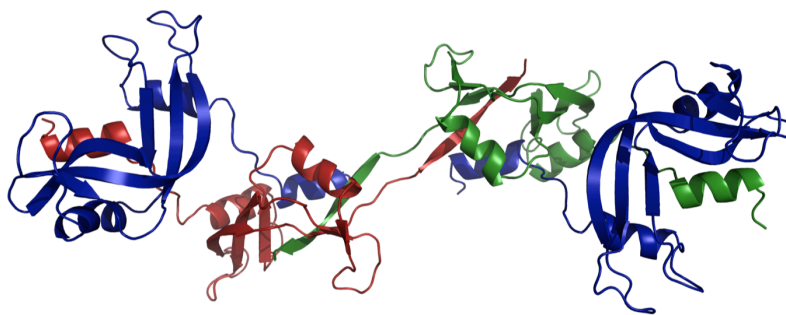


Figure 34. Model of the tetramer of RNase A obtained by C-terminal swapping of two N-terminal swapped dimers, represented as cartoon.

CONCLUSIONS AND FUTURE PERSPECTIVES

As it is well known, under specific conditions bovine pancreatic ribonuclease easily generates two domain-swapped dimers: one dimer (N-dimer) is obtained through swapping of the N-terminal helix (residues 1-15), the second (C-dimer) is formed through swapping of the C-terminal strand (residues 115-124), the latter being promoted by the transition of the *cis* to *trans* conformation of proline 114. Thus, RNase A can be classified as a bi-functional swapping protein. In principle, through a sequential alternate swapping of the N- and C-segment, a linear chain of swapped monomer can be built. However, in the preparative assays that promote both N- and C-swapping many different species are observed, such as the two dimers (N- and C-swapped), several trimers and many other higher-order oligomers. This variety of different oligomers makes the isolation of a single species not an easy task, and, therefore, the chemical and biological properties of these oligomers difficult to study. From this point of view, desBP is particularly interesting. As shown in the present study, the deletion of the residues from 16 to 20 does indeed force the monomer to dimerize *via* the exchange of the N-terminal helix (residues 1-14) between the two subunits, leading to a high stability of the dimer. This property suggests that the dimer represents an intriguing molecule that can act as a building-block to perform oligomerization experiments. Indeed, the dimer can still be considered a bi-

functional swapping unit, having the possibility to exchange two C-terminal strands each belonging to the two subunits composing the dimer. The C-terminal swapping of desBP may thus yield only one tetramer (dimer of dimer) and/or a more limited number of higher-order oligomers, thus making the isolation of a single species simpler to achieve. Moreover, the oligomerization experiments may be conducted in the best conditions that promote the C-terminal swapping. Preliminary, positive results have already been obtained and the work along this line is in progress.

MATERIALS AND METHODS

Protein production and purification

The deletion mutant des(16-20)-RNase A (desBP) has been kindly provided by the group of Prof. Picone, Department of Chemical Sciences, University of Naples “Federico II”, and prepared as described in (Picone, *et al.*, 2005).

Oligomers have been obtained by lyophilization upon 40% acetic acid incubation of the dimeric derivative. Purification and analyses of desBP oligomers have been performed through size-exclusion chromatography using a Superdex 75 HR 10/30 column, flow rate 0.5 mL/min, attached to an AKTA FPLC system (GE-Healthcare), at room temperature. The column has been equilibrated with 10 mM sodium phosphate pH 7.1, 0.1 M NaCl buffer. Elution was performed with the same buffer. The purified oligomers have been kept at 4°C until use.

Cathodic gel electrophoresis

Cathodic gel electrophoresis under non-denaturing conditions has been performed according to the procedure described by Goldenberg (Goldenberg, 1989) with slight modifications, using a β -alanine/acetic acid buffer, pH 4.0. Gels (10% polyacrylamide) have been run at 20 mA for 4 hours at 4 VC. Fixing and

staining were performed with 12.5% trichloroacetic acid and 0.1% Coomassie brilliant blue (Gotte, *et al.*, 1999).

Circular Dichroism

Circular dichroism (CD) spectra were recorded at 20°C using a Jasco J-710 spectropolarimeter equipped with a Peltier thermostatic cell holder (Model PTC-348WI). CD measurements were carried out in the 190-250 nm range using a 0.1 cm path length cell and 0.2 mg/mL solutions of the protein, in 10 mM sodium phosphate pH 7.1, 100 mM NaCl. Before measurements, the instrument was calibrated with an aqueous solution of D(+)-10-camphorsulfonic acid at 290 nm and the samples were pre-equilibrated at 20°C for 5 minutes.

Thermal unfolding curves were recorded in the 20–80°C range, in 0.5°C steps at heating rate 30°C/hour. Transition temperatures were calculated from the second derivative of the ellipticity change *vs* temperature.

Crystallization and data collection

A purified solution of desBP at a concentration of 20 mg/mL in 25 mM sodium phosphate pH 6.7 has been used to search for the crystallization conditions. Crystals of have been grown from a 1:1 (v/v) mixture of a desBP protein solution and a reservoir solution (0.1 M sodium citrate pH 5.6, 16% (w/v) polyethylene glycol 4000, 8-12% (v/v) isopropanol) using the hanging-drop vapor diffusion method at 4°C. Drops consisting of 0.5 µL protein mixed with an equal volume of precipitating solution have been equilibrated against 500 µL of reservoir solution. Prior to the collection of the X-ray data, the crystals have been soaked in a reservoir solution containing 20-25% (w/v) glycerol as a cryo-protectant and flash-frozen in a stream of liquid nitrogen.

X-ray diffraction data have been collected at 100 K at the ELETTRA Synchrotron (Trieste, Italy) using a 165 mm CCD detector from MAR-Research. Diffraction data have been processed using the program suite HKL2000

(Otwinowski, *et al.*, 1997). A summary of the indicators commonly used to estimate the quality of datasets is given in Table 9.

The structure of desBP has been solved by molecular replacement using the structure of wild-type RNase A (PDB entry 2e3w) as a search model. Refinement of the structure is still in progress. Figures have been drawn using Pymol (Schrodinger, 2010).

REFERENCES

- Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst D*, **66**:213-221.
- Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci USA*, **98**:10037-10041.
- Baldwin AJ, Knowles TP, Tartaglia GG, Fitzpatrick AW, Devlin GL, Shammass SL, Waudby CA, Mossuto MF, Meehan S, Gras SL *et al.* (2011) Metastability of native proteins and the phenomenon of amyloid formation. *J Am Chem Soc*, **133**:14160-14163.
- Bartholeyns J, Baudhuin P. (1976) Inhibition of tumor cell proliferation by dimerized ribonuclease. *Proc Natl Acad Sci USA*, **73**:573-576.
- Bartholeyns J, Zenebergh A. (1979) *In vitro* and *in vivo* antitumor effect of dimerized ribonuclease A. *Eur J Cancer*, **15**:85-91.
- Bennett MJ, Schlunegger MP, Eisenberg D. (1995) 3D domain swapping: a mechanism for oligomer assembly. *Protein Sci*, **4**:2455-2468.
- Bennett MJ, Sawaya MR, Eisenberg D. (2006) Deposition diseases and 3D domain swapping. *Structure*, **14**:811-824.
- Bretscher LE, Abel RL, Raines RT. (2000) A ribonuclease A variant with low catalytic activity but high cytotoxicity. *J Biol Chem*, **275**:9893-9896.
- Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS *et al.* (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Cryst D*, **54**:905-921.
- Buell AK, Dhulesia A, White DA, Knowles TP, Dobson CM, Welland ME. (2012) Detailed analysis of the energy barriers for amyloid fibril growth. *Angew Chem, Int Ed Engl*, **51**:5247-5251.
- Canals A, Pous J, Guasch A, Benito A, Ribo M, Vilanova M, Coll M. (2001) The structure of an engineered domain-swapped ribonuclease dimer and its implications for the evolution of proteins toward oligomerization. *Structure*, **9**:967-976.
- Carrell RW, Lomas DA. (1997) Conformational disease. *Lancet*, **350**:134-138.

- Chirgadze DY, Demydchuk M, Becker M, Moran S, Paoli M. (2004) Snapshot of protein structure evolution reveals conservation of functional dimerization through intertwined folding. *Structure*, **12**:1489-1494.
- Crestfield AM, Stein WH, Moore S. (1963) On the preparation of bovine pancreatic ribonuclease A. *J Biol Chem*, **238**:618-621.
- Di Gaetano S, D'Alessio G, Piccoli R. (2001) Second generation antitumour human RNase: significance of its structural and functional features for the mechanism of antitumour action. *Biochem J*, **358**:241-247.
- Dickson KA, Dahlberg CL, Raines RT. (2003) Compensating effects on the cytotoxicity of ribonuclease A variants. *Arch Biochem Biophys*, **415**:172-177.
- Dickson KA, Haigis MC, Raines RT. (2005) Ribonuclease inhibitor: structure and function. *Prog Nucleic Acid Res Mol Biol*, **80**:349-374.
- Dobson CM. (2001) The structural basis of protein folding and its links with human disease. *Philos T Roy Soc B*, **356**:133-145.
- Dobson CM. (2004) Principles of protein folding, misfolding and aggregation. *Semin Cell Dev Biol*, **15**:3-16.
- Downing KH, Nogales E. (1999) Crystallographic structure of tubulin: implications for dynamics and drug binding. *Cell Struct Funct*, **24**:269-275.
- Dykes G, Crepeau RH, Edelstein SJ. (1978) Three-dimensional reconstruction of the fibres of sickle cell haemoglobin. *Nature*, **272**:506-510.
- Elam JS, Taylor AB, Strange R, Antonyuk S, Doucette PA, Rodriguez JA, Hasnain SS, Hayward LJ, Valentine JS, Yeates TO *et al.* (2003) Amyloid-like filaments and water-filled nanotubes formed by SOD1 mutant proteins linked to familial ALS. *Nat Struct Biol*, **10**:461-467.
- Emsley P, Lohkamp B, Scott WG, Cowtan K. (2010) Features and development of Coot. *Acta Cryst D*, **66**:486-501.
- Ercole C, Avitabile F, Del Vecchio P, Crescenzi O, Tancredi T, Picone D. (2003) Role of the hinge peptide and the intersubunit interface in the swapping of N-termini in dimeric bovine seminal RNase. *Eur J Biochem*, **270**:4729-4735.
- Ercole C, Spadaccini R, Alfano C, Tancredi T, Picone D. (2007) A new mutant of bovine seminal ribonuclease with a reversed swapping propensity. *Biochemistry*, **46**:2227-2232.

- Ercole C, Colamarino RA, Pizzo E, Fogolari F, Spadaccini R, Picone D. (2009) Comparison of the structural and functional properties of RNase A and BS-RNase: a stepwise mutagenesis approach. *Biopolymers*, **91**:1009-1017.
- Giancola C, Ercole C, Fotticchia I, Spadaccini R, Pizzo E, D'Alessio G, Picone D. (2011) Structure-cytotoxicity relationships in bovine seminal ribonuclease: new insights from heat and chemical denaturation studies on variants. *FEBS J*, **278**:111-122.
- Goldenberg DP: (1989) Analysis of protein conformation by gel electrophoresis. In: *Protein Structure: A practical approach*. edn. Edited by Creighton TE. Oxford, UK: IRL Press;
- Gotte G, Testolin L, Costanzo C, Sorrentino S, Armato U, Libonati M. (1997) Cross-linked trimers of bovine ribonuclease A: activity on double-stranded RNA and antitumor action. *FEBS Lett*, **415**:308-312.
- Gotte G, Bertoldi M, Libonati M. (1999) Structural versatility of bovine ribonuclease A. Distinct conformers of trimeric and tetrameric aggregates of the enzyme. *Eur J Biochem*, **265**:680-687.
- Gotte G, Mahmoud Helmy A, Ercole C, Spadaccini R, Laurents DV, Donadelli M, Picone D. (2012) Double domain swapping in bovine seminal RNase: formation of distinct N- and C-swapped tetramers and multimers with increasing biological activities. *PLoS One*, **7**:e46804.
- Gronenborn AM. (2009) Protein acrobatics in pairs—dimerization via domain swapping. *Curr Opin Struct Biol*, **19**:39-49.
- Guo Z, Eisenberg D. (2006) Runaway domain swapping in amyloid-like fibrils of T7 endonuclease I. *Proc Natl Acad Sci USA*, **103**:8042-8047.
- Harrington DJ, Adachi K, Royer WE, Jr. (1997) The high resolution crystal structure of deoxyhemoglobin S. *J Mol Biol*, **272**:398-407.
- Holmes KC, Popp D, Gebhard W, Kabsch W. (1990) Atomic model of the actin filament. *Nature*, **347**:44-49.
- Ilinskaya O, Makarov A. (2005) Why ribonucleases induce tumor cell death. *Mol Biol*, **39**:1-10.
- Janin J, Chothia C. (1990) The structure of protein-protein recognition sites. *J Biol Chem*, **265**:16027-16030.
- Janowski R, Kozak M, Jankowska E, Grzonka Z, Grubb A, Abrahamson M, Jaskolski M. (2001) Human cystatin C, an amyloidogenic protein, dimerizes through three-dimensional domain swapping. *Nat Struct Biol*, **8**:316-320.

- Jones S, Thornton JM. (1995) Protein-protein interactions: a review of protein dimer structures. *Prog Biophys Mol Biol*, **63**:31-65.
- Kim JS, Soucek J, Matousek J, Raines RT. (1995) Structural basis for the biological activities of bovine seminal ribonuclease. *J Biol Chem*, **270**:10525-10530.
- Knaus KJ, Morillas M, Swietnicki W, Malone M, Surewicz WK, Yee VC. (2001) Crystal structure of the human prion protein reveals a mechanism for oligomerization. *Nat Struct Biol*, **8**:770-774.
- Kobe B, Deisenhofer J. (1993) Crystal structure of porcine ribonuclease inhibitor, a protein with leucine-rich repeats. *Nature*, **366**:751-756.
- Kobe B, Deisenhofer J. (1996) Mechanism of ribonuclease inhibition by ribonuclease inhibitor protein based on the crystal structure of its complex with ribonuclease A. *J Mol Biol*, **264**:1028-1043.
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr*, **26**:283-291.
- Lee FS, Shapiro R, Vallee BL. (1989) Tight-binding inhibition of angiogenin and ribonuclease A by placental ribonuclease inhibitor. *Biochemistry*, **28**:225-230.
- Leland PA, Schultz LW, Kim BM, Raines RT. (1998) Ribonuclease A variants with potent cytotoxic activity. *Proc Natl Acad Sci USA*, **95**:10407-10412.
- Leland PA, Raines RT. (2001a) Cancer chemotherapy-ribonucleases to the rescue. *Chem Biol*, **8**:405-413.
- Leland PA, Staniszewski KE, Kim BM, Raines RT. (2001b) Endowing human pancreatic ribonuclease with toxicity for cancer cells. *J Biol Chem*, **276**:43095-43102.
- Libonati M, Gotte G. (2004) Oligomerization of bovine ribonuclease A: structural and functional features of its multimers. *Biochem J*, **380**:311-327.
- Liu C, Sawaya MR, Eisenberg D. (2011) beta(2)-microglobulin forms three-dimensional domain-swapped amyloid fibrils with disulfide linkages. *Nat Struct Mol Biol*, **18**:49-55.
- Liu Y, Hart PJ, Schlunegger MP, Eisenberg D. (1998) The crystal structure of a 3D domain-swapped dimer of RNase A at a 2.1 Å resolution. *Proc Natl Acad Sci USA*, **95**:3437-3442.
- Liu Y, Gotte G, Libonati M, Eisenberg D. (2001) A domain-swapped RNase A dimer with implications for amyloid formation. *Nat Struct Biol*, **8**:211-214.

- Liu Y, Eisenberg D. (2002a) 3D domain swapping: as domains continue to swap. *Protein Sci*, **11**:1285-1299.
- Liu Y, Gotte G, Libonati M, Eisenberg D. (2002b) Structures of the two 3D domain-swapped RNase A trimers. *Protein Sci*, **11**:371-380.
- Lomas DA, Carrell RW. (2002) Serpinopathies and the conformational dementias. *Nat Rev Genet*, **3**:759-768.
- Makarov AA, Ilinskaya ON. (2003) Cytotoxic ribonucleases: molecular weapons and their targets. *FEBS Lett*, **540**:15-20.
- Marianayagam NJ, Sunde M, Matthews JM. (2004) The power of two: protein dimerization in biology. *Trends Biochem Sci*, **29**:618-625.
- Matousek J. (1973) The effect of bovine seminal ribonuclease (AS RNase) on cells of Crocker tumour in mice. *Experientia*, **29**:858-859.
- Matousek J. (2001) Ribonucleases and their antitumor activity. *Comp Biochem Physiol C Toxicol Pharmacol*, **129**:175-191.
- Matousek J, Gotte G, Pouckova P, Soucek J, Slavik T, Vottariello F, Libonati M. (2003) Antitumor activity and other biological actions of oligomers of ribonuclease A. *J Biol Chem*, **278**:23817-23822.
- Mazzarella L, Vitagliano L, Zagari A. (1995) Swapping structural determinants of ribonucleases: an energetic analysis of the hinge peptide 16-22. *Proc Natl Acad Sci USA*, **92**:3799-3803.
- McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. (2007) Phaser crystallographic software. *J Appl Crystallogr*, **40**:658-674.
- Merlino A, Ceruso MA, Vitagliano L, Mazzarella L. (2005a) Open interface and large quaternary structure movements in 3D domain-swapped proteins: insights from molecular dynamics simulations of the C-terminal swapped dimer of ribonuclease A. *Biophys J*, **88**:2003-2012.
- Merlino A, Mazzarella L, Carannante A, Di Fiore A, Di Donato A, Notomista E, Sica F. (2005b) The importance of dynamic effects on the enzyme activity: X-ray structure and molecular dynamics of onconase mutants. *J Biol Chem*, **280**:17953-17960.
- Merlino A, Ercole C, Picone D, Pizzo E, Mazzarella L, Sica F. (2008) The buried diversity of bovine seminal ribonuclease: shape and cytotoxicity of the swapped non-covalent form of the enzyme. *J Mol Biol*, **376**:427-437.

Merlino A, Avella G, Di Gaetano S, Arciello A, Piccoli R, Mazzarella L, Sica F. (2009a) Structural features for the mechanism of antitumor action of a dimeric human pancreatic ribonuclease variant. *Protein Sci*, **18**:50-57.

Merlino A, Russo Krauss I, Perillo M, Mattia CA, Ercole C, Picone D, Vergara A, Sica F. (2009b) Toward an antitumor form of bovine pancreatic ribonuclease: the crystal structure of three non-covalent dimeric mutants. *Biopolymers*, **91**:1029-1037.

Merlino A, Picone D, Ercole C, Balsamo A, Sica F. (2012) Chain termini cross-talk in the swapping process of bovine pancreatic ribonuclease. *Biochimie*, **94**:1108-1118.

Nelson R, Sawaya MR, Balbirnie M, Madsen AO, Riekkel C, Grothe R, Eisenberg D. (2005) Structure of the cross-beta spine of amyloid-like fibrils. *Nature*, **435**:773-778.

Otwinowski Z, Minor W: (1997) Processing of X-ray diffraction data collected in oscillation mode. In: *Methods Enzymol. Volume 276*: 307-326, edn. Edited by Charles W. Carter J: Academic Press;

Pepys MB. (2006) Amyloidosis. *Annu Rev Med*, **57**:223-241.

Pica A, Merlino A, Buell AK, Knowles TPJ, Pizzo E, D'Alessio G, Sica F, Mazzarella L. (2013) Three-dimensional domain swapping and supramolecular protein assembly: insights from the X-ray structure of a dimeric swapped variant of human pancreatic RNase. *Acta Cryst D*, **69**:2116-2123.

Piccoli R, Tamburrini M, Piccialli G, Di Donato A, Parente A, D'Alessio G. (1992) The dual-mode quaternary structure of seminal RNase. *Proc Natl Acad Sci USA*, **89**:1870-1874.

Picone D, Di Fiore A, Ercole C, Franzese M, Sica F, Tomaselli S, Mazzarella L. (2005) The role of the hinge loop in domain swapping. The special case of bovine seminal ribonuclease. *J Biol Chem*, **280**:13771-13778.

Pous J, Mallorqui-Fernandez G, Peracaula R, Terzyan SS, Futami J, Tada H, Yamada H, Seno M, de Llorens R, Gomis-Ruth FX *et al.* (2001) Three-dimensional structure of human RNase 1 delta N7 at 1.9 Å resolution. *Acta Cryst D*, **57**:498-505.

Raines RT: (2013) Enzymes as Chemotherapeutic Agents. In: *Chembiomolecular Science*. 281-291, edn. Edited by Shibasaki M, Iino M, Osada H: Springer Japan;

Rousseau F, Schymkowitz J, Serrano L. (2006) Protein aggregation and amyloidosis: confusion of the kinds? *Curr Opin Struct Biol*, **16**:118-126.

Russo N, Antignani A, D'Alessio G. (2000) *In vitro* evolution of a dimeric variant of human pancreatic ribonuclease. *Biochemistry*, **39**:3585-3591.

- Rutkoski TJ, Kink JA, Strong LE, Schilling CI, Raines RT. (2010) Antitumor activity of ribonuclease multimers created by site-specific covalent tethering. *Bioconjug Chem*, **21**:1691-1702.
- Sambashivan S, Liu Y, Sawaya MR, Gingery M, Eisenberg D. (2005) Amyloid-like fibrils of ribonuclease A with three-dimensional domain-swapped and native-like structure. *Nature*, **437**:266-269.
- Schlunegger MP, Bennett MJ, Eisenberg D. (1997) Oligomer formation by 3D domain swapping: a model for protein assembly and misassembly. *Adv Prot Chem*, **50**:61-122.
- Schrodinger L. (2010) *The PyMOL Molecular Graphics System, Version 13*, Schrödinger, LLC.
- Sica F, Di Fiore A, Merlino A, Mazzarella L. (2004) Structure and stability of the non-covalent swapped dimer of bovine seminal ribonuclease: an enzyme tailored to evade ribonuclease protein inhibitor. *J Biol Chem*, **279**:36753-36760.
- Spadaccini R, Ercole C, Gentile MA, Sanfelice D, Boelens R, Wechselberger R, Batta G, Bernini A, Niccolai N, Picone D. (2012) NMR studies on structure and dynamics of the monomeric derivative of BS-RNase: new insights for 3D domain swapping. *PLoS One*, **7**:e29076.
- Tarnowski GS, Kassel RL, Mountain IM, Blackburn P, Wilson G, Wang D. (1976) Comparison of antitumor activities of pancreatic ribonuclease and its cross-linked dimer. *Cancer Res*, **36**:4074-4078.
- Taverna DM, Goldstein RA. (2002) Why are proteins so robust to site mutations? *J Mol Biol*, **315**:479-484.
- Vagin AA, Steiner RA, Lebedev AA, Potterton L, McNicholas S, Long F, Murshudov GN. (2004) REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Cryst D*, **60**:2184-2195.
- Vitagliano L, Adinolfi S, Sica F, Merlino A, Zagari A, Mazzarella L. (1999) A potential allosteric subsite generated by domain swapping in bovine seminal ribonuclease. *J Mol Biol*, **293**:569-577.
- Vitagliano L, Merlino A, Zagari A, Mazzarella L. (2000) Productive and nonproductive binding to ribonuclease A: X-ray structure of two complexes with uridylyl(2',5')guanosine. *Protein Sci*, **9**:1217-1225.
- Youle RJ, D'Alessio G: (1997) Antitumor RNases. In: *Ribonucleases: Structures and Functions*. edn. Edited by Giuseppe DA, James FR. San Diego: Academic Press;

CANDIDATE'S PUBLICATION LIST

- Merlino A, Russo Krauss I, Albino A, Pica A, Vergara A, Masullo M, De Vendittis E, Sica F. (2011) Improving Protein Crystal Quality by the Without-Oil Microbatch Method: Crystallization and Preliminary X-ray Diffraction Analysis of Glutathione Synthetase from *Pseudoalteromonas haloplanktis*. *Int J Mol Sci*, **12**:6312-6319.

Abstract:

Glutathione synthetases catalyze the ATP-dependent synthesis of glutathione from L- γ -glutamyl-L-cysteine and glycine. Although these enzymes have been sequenced and characterized from a variety of biological sources, their exact catalytic mechanism is not fully understood and nothing is known about their adaptation at extremophilic environments. Glutathione synthetase from the Antarctic eubacterium *Pseudoalteromonas haloplanktis* (PhGshB) has been expressed, purified and successfully crystallized. An overall improvement of the crystal quality has been obtained by adapting the crystal growth conditions found with vapor diffusion experiments to the without-oil microbatch method. The best crystals of PhGshB diffract to 2.34 Å resolution and belong to space group $P2_12_12_1$, with unit-cell parameters $a = 83.28 \text{ \AA}$, $b = 119.88 \text{ \AA}$, $c = 159.82 \text{ \AA}$. Refinement of the model, obtained using phases derived from the structure of the same enzyme from *Escherichia coli* by molecular replacement, is in progress. The structural determination will provide the first structural characterization of a psychrophilic glutathione synthetase reported to date.

- Pica A, Russo Krauss I, Castellano I, Rossi M, La Cara F, Graziano G, Sica F, Merlino A. (2012) Exploring the unfolding mechanism of gamma-glutamyltranspeptidases: the case of the thermophilic enzyme from *Geobacillus thermodenitrificans*. *Biochim Biophys Acta*, **1824**:571-577.

Abstract:

γ -glutamyltranspeptidases (γ -GTs) are ubiquitous enzymes that catalyze the hydrolysis of γ -glutamyl bonds in glutathione and glutamine and the transfer of the released γ -glutamyl group to amino acids or short peptides. These enzymes are generally synthesized as precursor proteins, which undergo an intra-molecular autocatalytic cleavage yielding a large and a small subunit. In this study, circular dichroism and intrinsic fluorescence measurements have been used to investigate the structural features and the temperature- and guanidinium hydrochloride (GdnHCl)-induced unfolding of the mature form of the γ -GT from *Geobacillus thermodenitrificans* (GthGT) and that of its T353A mutant, which represents a mimic of the precursor protein. Data indicate that a) the mutant and the mature GthGT have a different secondary structure content and a slightly different exposure of hydrophobic regions, b) the thermal unfolding processes of both GthGT forms occur through a three-state model, characterized by a stable intermediate species, whereas chemical denaturations proceed through a single transition, c) both GthGT forms exhibit remarkable stability against temperature, but they do not display a strong resistance to the denaturing action of GdnHCl. These findings suggest that electrostatic interactions significantly contribute to the protein stability and that both the precursor and the mature form of GthGT assume compact and stable conformations to resist to the extreme temperatures where *G. thermodenitrificans* lives. Owing to its thermostability and unique catalytic properties, GthGT is an excellent candidate to be used as a glutaminase in food industry.

- Merlino A, Fuchs MR, Pica A, Balsamo A, Dworkowski FS, Pompidor G, Mazzarella L, Vergara A. (2013) Selective X-ray-induced NO photodissociation in haemoglobin crystals: evidence from a Raman-assisted crystallographic study. *Acta Cryst D*, **69**:137-140.

Abstract:

Despite their high physiological relevance, haemoglobin crystal structures with NO bound to haem constitute less than 1% of the total ligated haemoglobins (Hbs) deposited in the Protein Data Bank. The major difficulty in obtaining NO-ligated Hbs is most likely to be related to the oxidative denitrosylation caused by the high reactivity of the nitrosylated species with O₂. Here, using Raman-assisted X-ray crystallography, it is shown that under X-ray exposure (at four different radiation doses) crystals of nitrosylated haemoglobin from *Trematomus bernacchii* undergo a transition, mainly in the β -chains, that generates a pentacoordinate species owing to photodissociation of the Fe-NO bond. These data provide a physical explanation for the low number of nitrosylated Hb structures available in the literature.

- Pica A, Merlino A, Buell AK, Knowles TPJ, Pizzo E, D'Alessio G, Sica F, Mazzarella L. (2013a) Three-dimensional domain swapping and supramolecular protein assembly: insights from the X-ray structure of a dimeric swapped variant of human pancreatic RNase. *Acta Cryst D*, **69**:2116-2123.

Abstract:

The deletion of five residues in the loop connecting the N-terminal helix to the core of monomeric human pancreatic ribonuclease leads to the formation of an enzymatically active domain-swapped dimer (desHP). The crystal structure of desHP reveals the generation of an intriguing fibril-like aggregate of desHP molecules that extends along the **c** crystallographic axis. Dimers are formed by three-dimensional domain swapping. Tetramers are formed by the aggregation of swapped dimers with slightly different quaternary structures. The tetramers interact in such a way as to form an infinite rod-like structure that propagates throughout the crystal. The observed supramolecular assembly captured in the crystal predicts that desHP fibrils could form in solution; this has been confirmed by atomic force microscopy. These results provide new evidence that three-dimensional domain swapping can be a mechanism for the formation of elaborate large assemblies in which the protein, apart from the swapping, retains its original fold.

- Pica A, Russo Krauss I, Castellano I, La Cara F, Graziano G, Sica F, Merlino A. (2013b) Effect of NaCl on the conformational stability of the thermophilic gamma-glutamyltranspeptidase from *Geobacillus thermodenitrificans*: Implication for globular protein halotolerance. *Biochim Biophys Acta*, **1834**:149-157.

Abstract:

The transpeptidation activity of γ -glutamyltranspeptidase from *Geobacillus thermodenitrificans* (GthGT) is negligible and the enzyme is highly thermostable. Here we have examined the effect of concentrated NaCl solutions on structure, stability, dynamics and enzymatic activity of GthGT. The protein exhibited hydrolytic activity over a broad range of NaCl concentrations. Even at 4.0 M NaCl, GthGT retained more than 90% of the initial activity and showed unaltered fluorescence emission, secondary structure and acrylamide quenching on tryptophan fluorescence. Furthermore, at 2.8 M and 4.0 M NaCl the temperature-induced unfolding profiles are dramatically changed with large ($> 20^{\circ}\text{C}$) positive shifts in the denaturation temperature. These features make GthGT an ideal system to be used in industrial processes that require high temperatures and high-salt environments. A general explanation of the NaCl effect by means of a statistical thermodynamic model is also provided, together with an analysis of residue distribution between protein surface and interior in 15 non-redundant families of halophilic and non-halophilic proteins. The results are in line with a comparative sequence and structural analysis between halophilic and non-halophilic γ -glutamyltranspeptidases which revealed that a major role in halotolerance should be played by solvent exposed negatively charged residues.

- Pica A, Russo Krauss I, Merlino A, Nagatoishi S, Sugimoto N, Sica F. (2013c) Dissecting the contribution of thrombin exosite I in the recognition of thrombin binding aptamer. *FEBS J*, **280**:6581-6588.

Abstract:

Thrombin plays a pivotal role in the coagulation cascade; therefore, it represents a primary target in the treatment of several blood diseases. The 15-mer DNA oligonucleotide 5'-GGTTGGTGTGGTTGG-3', known as thrombin binding aptamer (TBA), is a highly potent inhibitor of the enzyme. TBA folds as an antiparallel chair-like G-quadruplex structure, with two G-tetrads surrounded by two TT loops on one side and a TGT loop on the opposite side. Previous crystallographic studies have shown that TBA binds thrombin exosite I by its TT loops, T3T4 and T12T13. In order to get a better understanding of the thrombin-TBA interaction, we have undertaken a crystallographic characterization of the complexes between thrombin and two TBA mutants, TBA Δ T3 and TBA Δ T12, which lack the nucleobase of T3 and T12, respectively. The structural details of the two complexes show that exosite I is actually split into two regions, which contribute differently to TBA recognition. These results provide the basis for a more rational design of new aptamers with improved therapeutic action.

- Russo Krauss I, Pica A, Merlino A, Mazzarella L, Sica F. (2013) Duplex-quadruplex motifs in a peculiar structural organization cooperatively contribute to thrombin binding of a DNA aptamer. *Acta Cryst D*, 69:2403-2411.

Abstract:

Potent second-generation thrombin aptamers adopt a duplex-quadruplex bimodular folding and recognize thrombin exosite II with very high affinity and specificity. A sound model of these oligonucleotides, either free or in complex with thrombin, is not yet available. Here, a structural study of one of these aptamers, HD22-27mer, is presented. The crystal structure of this aptamer in complex with thrombin displays a novel architecture in which the helical stem is enchainned to a *pseudo*-G-quadruplex. The results also underline the role of the residues that join the duplex and quadruplex motifs and control their recruitment in thrombin binding.

- Sica F, Pica A, Merlino A, Russo Krauss I, Ercole C, Picone D. (2013) The multiple forms of bovine seminal ribonuclease: Structure and stability of a C-terminal swapped dimer. *FEBS Lett*, **587**:3755-3762.

Abstract:

Bovine seminal ribonuclease (BS-RNase) acquires an interesting anti-tumor activity associated with the swapping on the N-terminal. The first direct experimental evidence on the formation of a C-terminal swapped dimer (C-dimer) obtained from the monomeric derivative of BS-RNase, although under non-native conditions, is here reported. The X-ray model of this dimer reveals a quaternary structure different from that of the C-dimer of RNase A, due to the presence of three mutations in the hinge peptide 111–116. The mutations increase the hinge peptide flexibility and decrease the stability of the C-dimer against dissociation. The biological implications of the structural data are also discussed.

- Stelitano V, Brandt A, Fernicola S, Franceschini S, Giardina G, Pica A, Rinaldo S, Sica F, Cutruzzola F. (2013) Probing the activity of diguanylate cyclases and c-di-GMP phosphodiesterases in real-time by CD spectroscopy. *Nucleic Acids Res*, **41**:e79.

Abstract:

Bacteria react to adverse environmental stimuli by clustering into organized communities called biofilms. A remarkably sophisticated control system based on the dinucleotide 3'-5' cyclic diguanylic acid (c-di-GMP) is involved in deciding whether to form or abandon biofilms. The ability of c-di-GMP to form self-intercalated dimers is also thought to play a role in this complex regulation. A great advantage in the quest of elucidating the catalytic properties of the enzymes involved in c-di-GMP turnover (diguanylate cyclases and phosphodiesterases) would come from the availability of an experimental approach for in vitro quantification of c-di-GMP in real-time. Here, we show that c-di-GMP can be detected and quantified by circular dichroism (CD) spectroscopy in the low micromolar range. The method is based on the selective ability of manganese ions to induce formation of the intercalated dimer of the c-di-GMP dinucleotide in solution, which displays an intense sigmoidal CD spectrum in the near-ultraviolet region. This characteristic spectrum originates from the stacking interaction of the four mutually intercalated guanines, as it is absent in the other cyclic dinucleotide 3'-5' cyclic adenilic acid (c-di-AMP). Thus, near-ultraviolet CD can be used to effectively quantify in real-time the activity of diguanylate cyclases and phosphodiesterases in solution.

Structures deposited in the Protein Data Bank

PDB entry

- 4g51 Crystallographic analysis of the interaction of nitric oxide with hemoglobin from *Trematomus bernacchii* in the T quaternary structure (fully ligated state)
Merlino, A, Balsamo, A, Pica, A, Mazzarella, L, Vergara, A
- 4kxh The X-ray crystal structure of a dimeric variant of human pancreatic ribonuclease
Pica, A, Merlino, A, Mazzarella, L, Sica, F
- 4i7y Crystal Structure of Human Alpha Thrombin in Complex with a 27-mer Aptamer Bound to Exosite II
Pica, A, Russo Krauss, I, Merlino, A, Mazzarella, L, Sica, F
- 4n4c Crystal structure of the C-terminal swapped dimer of a Bovine seminal ribonuclease mutant
Pica, A, Russo Krauss, I, Merlino, A, Sica, F
- 4lz4 X-ray structure of the complex between human thrombin and the TBA deletion mutant lacking thymine 3 nucleobase
Pica, A, Russo Krauss, I, Merlino, A, Sica, F
- 4lz1 X-ray structure of the complex between human thrombin and the TBA deletion mutant lacking thymine 12 nucleobase
Pica, A, Russo Krauss, I, Merlino, A, Sica, F