UNIVERSITÀ DEGLI STUDI DI MILANO

PhD Course in Molecular and Cellular Biology

XXX Cycle

DISSECTING STRUCTURAL ASPECTS OF PROTEIN STABILITY

BENEDETTA MARIA SALA

PhD Thesis

Scientific tutor: Prof. Stefano Ricagno

Academic year: 2016-2017

SSD: Bio/10

This thesis was performed at Department of Biosciences, University of Milan.

Contents

Amino acids in symbol

Ala	A	Alanine
Arg	R	Arginine
Asn	N	Asparagine
Asp	D	Aspartic Acid
Cys	C	Cysteine
Glu	E	Glutamic Acid
Gln	Q	Glutamine
Gly	G	Glycine
His	H	Histidine
Ile	I	Isoleucine
Leu	L	Leucine
Lys	K	Lysine
Met	M	Methionine
Phe	F	Phenylalanine
Pro	P	Proline
Ser	S	Serine
Thr	T	Threonine
Trp	W	Tryptophan
Tyr	Y	Tyrosine
Val	V	Valine

Summary

The contribution of a single residue in the structural stability of a whole protein can seem small if its position in the structure is not considered. Indeed, the stability of a protein is determined by the contribution of many forces among amino acids: hydrophobicity, hydrogen-bonding network and van der Waals interactions, salt bridges. Three independent projects were carried out during my PhD studies addressing mutations that affect protein fold stability.

I

β2-Microglobulin (β2m), the light chain of the Major Histocompatibility Class I complex (MHC-I), can assemble into amyloid fibrils and deposit in joints and bones of patients receiving long term haemodialysis, leading to dialysis-related amyloidosis (DRA).
β2m has been object of many mutational studies to investigate the molecular bases of its aggregation propensity. Among other residues, these analyses have pointed out that Trp60 plays a crucial role for β2m stability, in particular, W60G β2m, has an increased conformational stability and a reduced amyloidogenicity compared to the wild type protein.
To identify the key-residues that prompt the amyloidogenicity of β2m, the native state dynamics of WT β2m and W60G β2m have been compared, combining solution NMR spectroscopy and molecular dynamics (MD) simulations. As a result, three β2m variants were designed and characterized: W60G-Y63W, W60G-N83V and V85E. All mutants have displayed the expected structure, stability and aggregation propensity, giving insights into structural determinants of aggregation propensity of WT β2m.

II

To date, only a single natural β2m variant at residue Asp76, D76N β2m, has been reported. This mutation destabilizes the protein making it more prone to aggregation and has been related to a systemic and hereditary amyloid disorder.
To study the fundamental properties of protein dynamics and stability in solution and in crystal, D76N, WT and W60G β2m variants were chosen. These proteins display significant different stabilities and aggregation propensities, although conserving well comparable structures, with the same crystal symmetry and number of molecules in the asymmetric unit. These properties make them suitable as model system to obtain results free of spurious effects, reflecting the changes on protein dynamics related to the mutation.

Thus, these three variants were used to investigate the correlation between protein stability in solution and in crystalline form by Fourier transform infrared spectroscopy (FT-IR).

FT-IR spectra indicate that each β2m variant displays an increased structural order in the crystals compared to the protein in solution; temperature ramps show that increased melting temperatures for crystals compared to protein in solution. Nevertheless, the differences in protein stability between the three variants are well detectable in crystalline form and they well correlate with the trend observed in solution.

III

MHC-I complexes are hetero-trimers composed by the polymorphic heavy chain, the light chain β2m and an antigenic peptide.

The overall structural stability of the complex together with peptide-complex affinity, is fundamental for an effective presentation of the antigen: a strong peptide/MHC-I (pMHC) interaction and a time-extended peptide presentation trigger the best immunogenic CD8+ T-cell response.

Accordingly, aim of the work is the exploration of the impact on protein fold stability of complex-stabilizing point mutations on the tumour antigenic peptide NY–$ESO_{157-165}$ (SLLMWITQC). This peptide is expressed by a broad range of tumour types, but not in healthy adult somatic tissue, making it an ideal cancer vaccine candidate.

The *in vitro* stability of eight complexes has been monitored following the thermal unfolding. All complexes bearing the mutated peptides displayed an increase in stability compared to that of the complex with the NY–$ESO_{157-165}$ WT.

The structural contributions to the stability of the complexes have been investigated through X-ray crystallography. Thanks to the structural studies it has been possible to deepen the molecular characteristics determining peptide binding to MHC- I molecules and rationalize the optimization of the MHC- I anchor residues in tumour epitopes to enhance binding of the peptide to the complex.

Part I

1. State of the art

1.1. Protein folding, stability and aggregation

1.1.1. Protein fold, stability and dynamics

The amino acid sequence of a protein contains the information that guides its folding and, therefore, its three-dimensional native structure (*1*). Typically, this native conformation corresponds to a local energy minimum reached with the establishment of energetically favourable intra-molecular interactions (*2, 3*).

The protein fold stability is determined by the combination of many forces that constrain the polypeptide chain in its native configuration (*4, 5*). Most of these are weak non-covalent interactions that can act synergically to hold different regions of a polypeptide chain closely together involving backbone atoms as well as atoms of the side chains (*6*).

Hydrogen bonds stabilize the secondary structural elements of a folded protein connecting the carbonyl and amide groups along the polypeptide backbone. In addition, several amino acid side chains are also capable of establishing inter- and intra-molecular hydrogen bonds (*7*).

The chemistry of amino acid side chains, and therefore distribution of polar and non-polar amino acids, is determinant to the protein structure. The non-polar (hydrophobic) side chains interact via weak van der Waals interactions. Thus, they tend to cluster in the core of the molecule, avoiding contacts with the surrounding solvent. In contrast, polar (hydrophilic) amino acids are usually arranged in the external regions of the molecules where they can form hydrogen bonds with other polar molecules or with water (*8, 9*). Furthermore, ionic bonds can occur among charged amino acid.

Cysteine residues, when oxidized, are the only amino acids capable of forming covalent bonds other than peptide bonds, known as disulfide bridges. These covalent bonds can link together two segments of a polypeptide chain, further stabilizing the folded structure (*10, 11*).

The relationship between structural conformation of proteins and their stability has been widely studied. Thermodynamic studies are a suitable tool to predict how proteins changes their conformation under different environmental conditions giving information about the stability of proteins.

Moreover, different strategies for improving protein stability have been set up (*12*). These include increase in the hydrophobic interactions, stabilization through the insertion of disulphide bonds and prevention of irreversible processes like aggregation and incorrect folding.

Once folded, proteins populate an ensemble of conformations at equilibrium. This allows the proteins to maintain a high degree of flexibility, often crucial for their function. The biological function of a

protein, hence, may often depend not only on its three-dimensional structure, but also on its dynamical properties. The dynamics in a protein allow its conformational changes to respond to the presence of other molecules and/or to variations in the environment. Biological and biochemical processes such as signal transduction, antigen recognition, protein transport and enzyme catalysis rely on this ability to change conformation or to adapt to change.

Quantitative characterization of protein dynamics can be carried out by Nuclear Magnetic Resonance (NMR) spectroscopy, allowing its precise assessment under equilibrium conditions over different time scales, from ps-ns to s-min.

1.1.2. Amyloidosis

The native protein dynamics can be perturbed by several factors leading to misfolding and aggregation. Main factors affecting the level of aggregation prone conformers are changes in the intracellular environment such as pH, temperature, oxidative stress and metal ions, the protein concentration as well as somatic or genetic mutations (*3*).

Protein conformational disorders have been linked to various human diseases, defined as diseases by misfolding and protein aggregation (*13*). Among these are amyloidosis, a group of pathologies where specific proteins, in parallel or as an alternative to the physiological fold, generate highly structured fibrillar aggregates that deposit in tissues.

About 20 different non-related proteins can undergo into amyloid aggregation *in vivo* (*14*), forming fibrils with cross-β-sheet quaternary structure (*Fig.1*): the polypeptide chains assume a cross-β arrangement along the growth axis, with their β filaments perpendicular to it (*15*). The super helical structure of the protofilament allows the extension of hydrogen bonds among β filaments along the total length of amyloid fibrils, defining their distinctive rigidity and stability (*16*). Protofilaments wrap around each other forming amyloid fibres with diameter ranging between 50 and 150 Å (*17*).

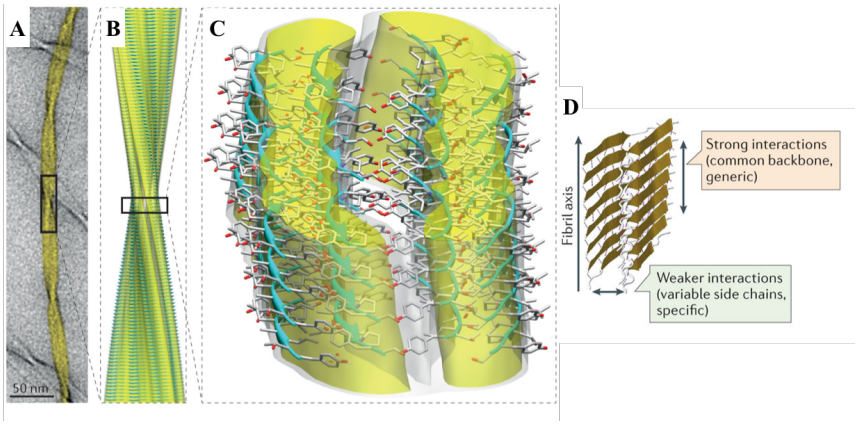


Figure 1. Amyloid structures possess common features. **(A)** Cryo-electron microscopy imaging of amyloid fibrils. **(B, C)** Atomic-level structure determined by NMR analysis. **(D)** Schematic representation of the 'cross-β' structure common to amyloid fibrils. Adapted from T. P. Knowles, M. Vendruscolo, C. M. Dobson, The amyloid state and its association with protein misfolding diseases. *Nat Rev Mol Cell Biol* 15, 384-396 (2014) (*17*).

Several are the causes that trigger proteins to amyloid aggregation (*14*). Amyloidosis can result from altered conformations assumed by a natively folded protein (*18*), a drastic increase in protein concentration (*19*), or proteolysis that generate amyloidogenic peptides (*20*). Moreover, also a denaturing stress and/or a mutation can lead to conformational changes affording an intermediate that self-assembles into amyloid (*21-24*).

The kinetics of amyloid formation can be described with a sigmoidal curve where first a lag phase is followed by a quick growth phase and a final plateau regime. Protein monomers undergo conformational changes and associate to form oligomeric nuclei. Elongation of fibrils occurs with addition of monomers to the ends of existing aggregates leading to form mature fibrils (*14, 25*). Fibrils present a catalytic surface, allowing the formations of new aggregates through the so called "secondary nucleation". Moreover, fragmentation of fibrils may also provide new ends for elongation. All these processes, (primary and secondary nucleation, elongation and fragmentation), occur during all phases of the aggregation process with different rates depending by the concentration of reacting species (*Fig.2*) (*26*).

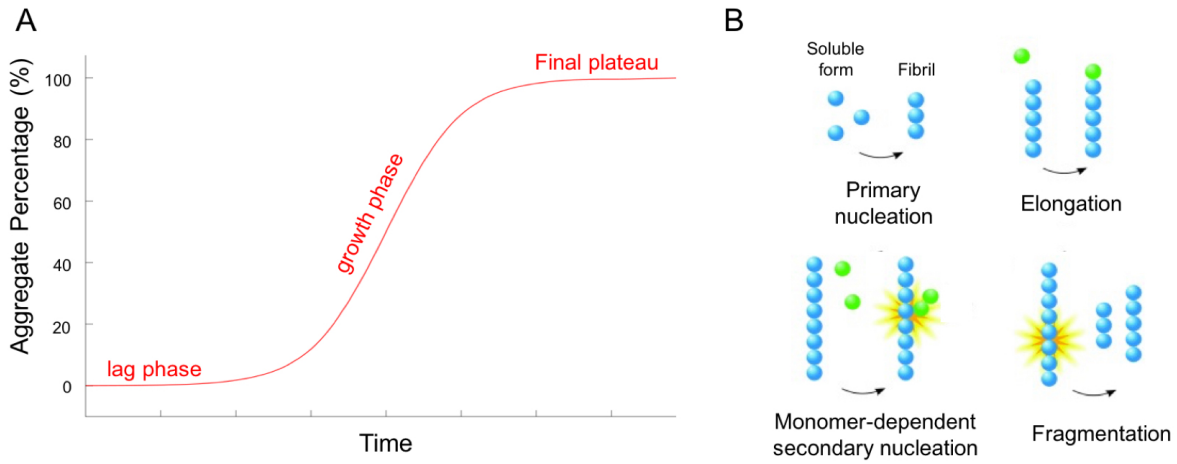


Figure 2. Amyloid aggregation. **(A)** Characteristic aggregation curve for amyloid fibril formation. The curve is typically divided into a lag phase, a growth phase and a final plateau. **(B)** Processes occurring during amyloid formation: primary nucleation from monomers in solution, elongation (growth) by monomers addition to existing aggregates, surface catalysed secondary nucleation from monomers on fibril surface and fragmentation. A from P. Arosio, T. P. Knowles, S. Linse, On the lag phase in amyloid fibril formation. *Phys Chem Chem Phys* 17, 7606-7618 (2015) (*26*).

Diagnosis and detection of amyloid fibrils are usually obtained by using amyloid binding molecules, such as Congo Red staining and Thioflavin-T (ThT) fluorescence (*27*). Congo Red staining can detect amyloid deposits thanks to their characteristic apple green birefringence under polarized light conditions. ThT assays measure the changes of fluorescence intensity of ThT upon binding to amyloid fibrils both *in vivo* and *in vitro (28, 29)*. The enhanced fluorescence can be observed by fluorescence microscopy or by fluorescent spectroscopy.

It is widely accepted that the amyloid state is accessible to various amino acid sequences under proper conditions and can therefore be considered an inherent property of polypeptide chains (*3, 30, 31*). Therefore, it is necessary to define the specific mechanisms by which proteins assume pathogenic behaviours that lead to amyloid aggregation and how this is generally avoided in living systems.

1.1.3. β2-microglobulin

1.1.3.1. Structure, function and dialysis-related amyloidosis

Among amyloidosis-related proteins, β2-Microglobulin (β2m) is a well studied protein model. Wild-type β2m (WT β2m) is composed by 99 amino arranged in seven anti-parallel β-strands organized in two facing β-sheets (ABED and CFG). It thus displays the typical immunoglobulin fold, with the β-sandwich stabilized by an inter-strand disulphide bond between cysteine 25 and 80 (*Fig. 3B, C*) (*32-35*). It has also been demonstrated that this bond is necessary for β2m folding in the endoplasmic reticulum, and it is not broken during fibrillogenesis *in vivo* (*34*).

β2m is the non-covalently bound light chain of the Major Histocompatibility Complex class I (MHC-I, see also paragraph 1.2.1.1.), where, it plays a crucial role in chaperoning assembly of the complex for antigen presentation on the cell surface (*Fig. 3A*) (*36, 37*).

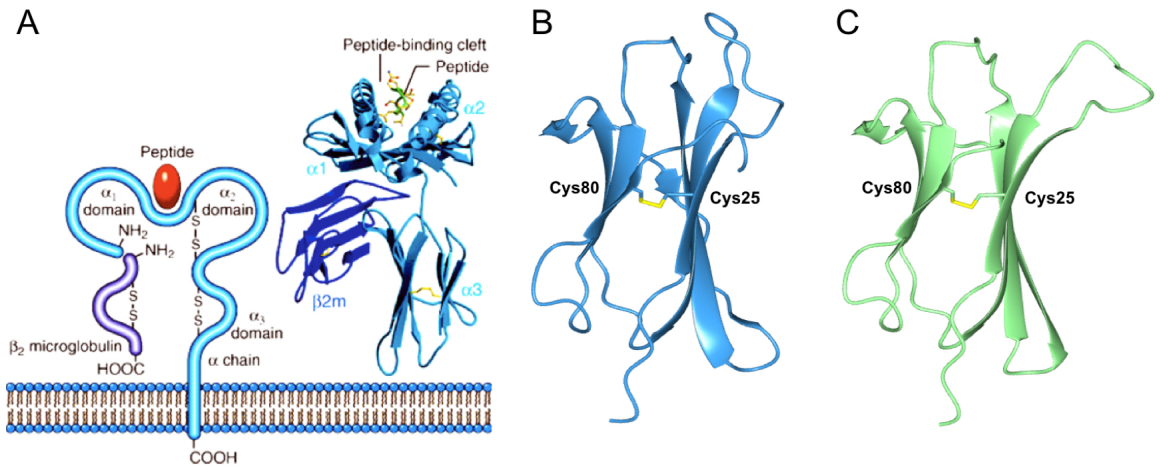


Figure 3. X-ray crystal structure of MHC-I and β2m. **(A)** Schematic representation and 3D model (pdb code: 1S9W) showing the β2m non-covalently bounded to the heavy chain (α1, α2, α3) carrying the antigenic peptide in its cleft. **(B)** 3D model (pdb code: 2BSS) of the structure assumed by β2m in the MHC-I complex, AB loop has a "close" conformation. **(C)** 3D model (pdb code: 1LDS) of the monomeric β2m, AB loop has an "open" conformation. Cysteine residues 25 and 80 are showed as sticks in the ribbon structures of β2m, linked by the disulphide bridge.

The region of the protein involved in the interaction with the heavy chain of the complex are the AB-loop (residues 12-21) that assumes a "close conformation" (Fig. 3B), the A and D strands, and the DE-loop (pdb code: 1S9W) (*38*).

β2m is highly stabilized thanks to the interactions within the MHC- I (*39*) and in physiological conditions its catabolism occurs through two steps: β2m dissociates as a monomer from the MHC-I heavy chain, then it is released into the serum (with a typical concentration of 0.1 μM), and finally degraded proximal tubules in the kidney (*40*).

In patients suffering from renal dysfunction undergoing to regular haemodialysis treatments, β2m concentration starts gradually to concentrate into their serum, reaching up to 60-time that of healthy individuals (*41*).

This causes a condition known as dialysis-related amyloidosis (DRA), where the β2m deposits as amyloid fibrils in joints and bones of the patients, leading to pathological bone fragility and articular pain (*42, 43*).

1.1.3.2. *In vitro* studies

DRA was firstly discovered by Assenat *et al.* in 1980 (*44*), but the mechanisms of β2m aggregation are still unclear.

In vitro studies have demonstrated that increased β2m concentration is not sufficient on its own to trigger amyloid aggregation. In fact, incubating β2m for several months at physiological conditions (neutral pH, 37°C), the protein remains mostly monomeric even if at concentrations more than 20-fold higher than those found in DRA patients (*45, 46*). Thus, several factors have been shown to trigger β2m aggregation both *in vitro* and *in vivo (47)*: presence of a divalent metal ions such as Cu^{2+} (*48, 49*), glycosaminoglycans (*50*), collagen (*51*) and others (*52-55*).

Among vertebrates β2m possess a highly conserved sequence. The study of artificial mutants of WT β2m has being been a key tool to investigate the molecular process of fibrillogenesis. To this aim several mutants have been characterized *in vitro* from a biochemical and structural point of view in relation to their amyloidogenic properties.

Chiti *et al.* (*56*) through a series of spectroscopic analyses demonstrated that WT β2m folds via two structurally distinct intermediates, I_1 and I_2, in three sequential phases folding process. The first intermediate, the faster one, is characterized by species showing partially folded elements and hydrophobic regions exposed to solvent. The second one, indeed, displays native-like secondary structure and can accumulate, reaching concentration well correlating with its involvement in amyloid elongation (*57*).

Mutagenesis studies were performed to better investigate this mechanism, demonstrating that the transition between intermediate I_2 and the native structure is rate limited by *trans* to *cis* isomerization of the His31-Pro32 peptidyl–prolyl bond (*58*). Furthermore, the His31 residue was related to the copper ion-dependent fibrillogenesis: this residue mask the ion, which promotes the isomerization of

the Pro32 (*59*). Moreover, Pro32 residue mutations suggested that *cis/trans* isomerization of this residual constitutes a kinetically limiting stage in the formation speed of amyloid fibres (*60*).

The correct structural fold and function of β2m is correlated with two evolutionary conserved Tryptophan (Trp) residues, with differential and complementary impact in the structure of β2m (*61, 62*). Trp95, located at the end of β-strand G, is partially buried. When mutated, the fold stability of the protein is decreased and the protein unfold in solution, yielding non fibrillar aggregates.
Conversely, Trp60 is exposed to the solvent at the apex of DE-loop (residues 57–60) that is involved in the interaction of β2m with the heavy chain of MHC-I. This region becomes totally exposed to solvent in monomeric β2m with an enhanced backbone dynamics (*63*) and variety in conformations in different crystal structures (*Fig. 4*). Consequently, due to its high frequency of intermolecular contacts, Trp60 is involved in the early steps of fibrillogenesis, as suggested by molecular dynamics simulations (*64*). Moreover, NMR H/D exchange experiments pointed out that, once the protein has aggregated, Trp60 becomes solvent accessible in the fibrils (*65*).

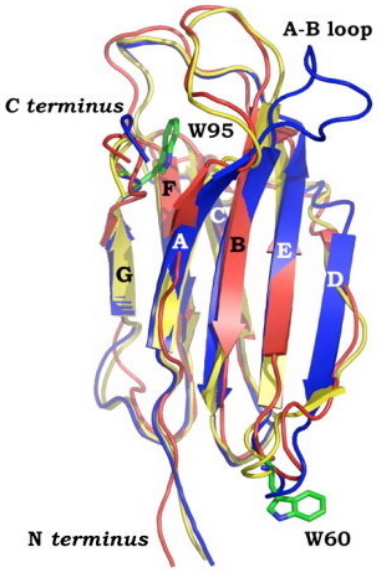


Figure 4. Cartoon representation of the W60G crystal (blue) and NMR average (red) structures superposed on WT β2m (yellow) as crystallized in the MHC-I complex (PDB code: 2BSS). Strands are labelled according to the convention adopted for β2m; residues Trp60 and Trp95 from crystal structures are shown as stick models. Figure acquired from G. Esposito *et al.*, The controlling roles of Trp60 and Trp95 in beta2-microglobulin function, folding and amyloid aggregation properties. *J Mol Biol* 378, 887-897 (2008) (*61*).

Rational mutation studies of the DE-loop have shown that alterations in this region can affect overall protein structure, fold stability, folding kinetics and amyloid aggregation properties. These properties have been related to the strained conformation in the WT β2m backbone structure. Indeed, residues in the DE-loop populate unfavorable regions of the Ramachandran plot in all the published structures of both monomeric β2m and in complex with the MHC-I heavy chain.

The replacement of Trp60 with a Gly residue, generates the W60G β2m variant and results in a structural relaxation of the DE-loop compared to the WT protein. This increases the W60G β2m stability, relative to the WT protein. Thermal protein unfolding monitored by near- and far-UV CD shows that the W60G mutation substantially increases thermal stability of the β2m fold, shifting the T_m of about +8°C (*61*).

Additionally, the removal of the bulky hydrophobic Trp side chain also increase the solubility of the monomeric β2m and make it aggregation-resistant under mild conditions (*61*).

As resulted from Nano-ESI-MS experiments, both WT β2m and W60G mutant assume a partially folded form at pH 2.5. However, despite having comparable structural compactness, the W60G mutant at pH 2.5 displays a slower kinetics of disulphide reduction. This results can be related either to its slightly more structured conformation, and reduced flexibility, that affect both native state and partially unfolded intermediate (*66*).

Given these observations, other DE-loop mutants were designed. A Proline residue at position 59 (Asp59→Pro) has been introduced to further increase the rigidity of the DE loop of the WT protein. The D59P crystal structure displays backbone conformational strains in the mutated region and thermal unfolding carried out on the mutant shows a decrease in stability relative to WT β2m and higher propensity to aggregation, confirming that the improvement of the backbone conformational strain can trigger long-range effects (*62*).

The Trp60→Cys substitution, indeed, induces the same regular β-turn conformation as observed in the W60G β2m mutant. However, contrary to W60G, W60C β2m displays a measurable – though reduced - aggregation propensity compared to WT. This can suggest that the relaxation of the DE-loop geometry, alone, is not fully accounting for the lack of amyloidogenicity observed for W60G (*Fig. 5*) (*62*).

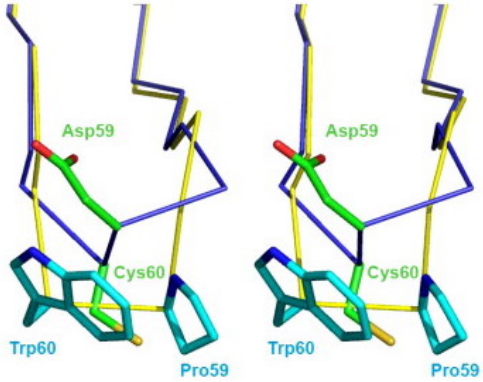


Figure 5. Mutation of residues Asp58 and Trp60 in the DE-loop of β2m. Superposition of W60C (blue) and D59P (yellow) DE-loops (stereo view). Stick representation of residues 59 and 60 is shown in green for W60C and cyan for D59P. Figure acquired from S. Ricagno *et al.*, DE loop mutations affect beta2-microglobulin stability and amyloid aggregation. *Biochem Biophys Res Commun* 377, 146-150 (2008) (*62*).

As W60C, also W60V mutant showed a thermal stability close to that of WT β2m, the same low efficiency in beta-aggregation at neutral pH and the common DE-loop conformation of W60G variant. The reduced tendency to form fibrils at mild conditions of the W60V and W60C β2m mutants, compared to those of WT protein, highlights the key role of the aromatic residues in the DE-loop in promoting β2m fibrillogenesis (*66, 67*).

Colombo *et al.* introduced a Lys58→Pro mutation to the W60G β2m variant to stabilize the overall structure through an ideal type-I β-turn, mimicking the effect of a 'native Pro' residue (*68*).
Compared to the WT protein, the K58P-W60G mutant shows enhanced stability and faster folding kinetic without passing through the folding slow phase (*56*).
K58P-W60G β2m displays also peculiar feature: while all β2m variants show a completely oxidized disulphide bond, this mutant is purified as a mixture of oxidized (~70%) and reduced (~30%) molecules.
This proved that in the K58P-W60G mutant the disulphide bond is not essential for its folding process, in apparent divergence with what proven for WT β2m (*34*), and demonstrate that DE-loop properties affect not only fold stability but also the folding pathway.

In this context *Paper 1*, aims to investigate whether W60G mutation has other effects on β2m beyond its stabilization. The work thus considers if the aggregation propensity is the result of an altered protein dynamics.
The native state dynamics of both proteins were compared combining NMR spectroscopy and molecular dynamics simulations underlining an increased average protection of the aggregation-prone residues at W60G β2m surface. These findings allowed the rational design of three new β2m variants displaying at the same time an altered aggregation propensity and an un-altered protein stability.

1.1.3.3. Hereditary systemic amyloidosis

In 2012 Valleix *et al.* reported the first natural variant of human β2m discovered in members of a French family. It consists in a point mutation of residue 76 from Asp to Asn (D76N β2m) (*22*). This results in a new hereditary systemic amyloidosis with large widespread amyloid deposits found in visceral organs (liver, spleen, kidney, and heart) and whose main symptoms are chronic diarrhoea, loss of weight, sicca syndrome and polyneuropathy.
Unlike DRA, the plasma concentration of monomeric β2m ranges in normal values and amyloid deposits display the presence of only full length mutated protein without any truncated forms or WT (*22*).

From the biophysical point of view, Asp76→Asn strongly destabilizes the protein and makes it much more prone to amyloidogenic aggregation than WT β2m even if its crystal structure shows only subtle conformational changes (*Fig. 6*). Mangione P. *et al.* also reported the significant increase of the partially structured I_2 folding intermediate in D76N β2m variant compared to the WT protein (*56, 69*). These findings confirm the correlation between the amyloidogenic propensity of β2m species with the existence of a concentration of this partially structured intermediate (*70*).

The molecular bases of D76N β2m instability and aggregation propensity have been further investigated characterizing different single point mutation variants, where all Asp residues have been individually substituted to Asn (D34N, D38N, D53N, D59N, D96N and D98N) (*71*). All new variants displayed thermal stability and aggregation propensity close to those of WT β2m suggesting that the amyloidogenic properties of the D76N mutant is not simply due to the loss of a negative charge (*Fig. 6*).

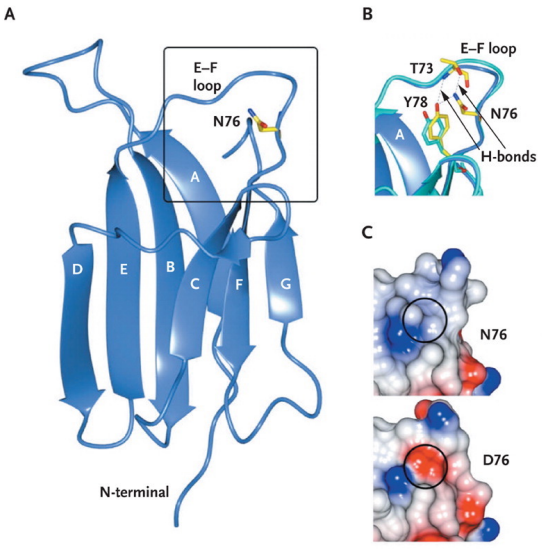


Figure 6. Crystal structure of D76N variant. **(A)** Ribbon representation of the D76N mutant. **(B)** is a close-up view of the E–F loop (residues 70 to 80) shown in (A), superimposed on the wild-type β_2-microglobulin structure (blue–green). Residues belonging to the D76N variant are yellow, and Tyr78 belonging to wild-type β_2-microglobulin is blue–green; hydrogen (H) bonds are shown as dashed lines. **(C)** shows the surface electrostatic potential of the E–F loop region in the D76N variant (top) and in wild-type β2-microglobulin (bottom); circles indicate the mutated residue. Blue represents positively charged regions, and red negatively charged regions. Figure acquired from S. Valleix et al., Hereditary systemic amyloidosis due to Asp76Asn variant beta2-microglobulin. N Engl J Med 366, 2276-2283 (2012) (*22*).

Given D76N β2m instability, Halabelian L. *et al.* investigated the mechanisms that allow the D76N variant to escape the unfolded protein response system that targets misfolded/aggregated polypeptides (*72*). They, thus, performed structural and biophysical study on a MHC-I bearing the D76N variant, pointing out the high stability of the complex, matching with that of MCH-I bearing WT β2m (*39*).

Therefore, MHC-I has a role in the stabilization of D76N variant, acting as chaperone and protecting this variant from intracellular degradation.

The set of these three β2m variants, WT, W60G and D76N is well suited for a comparative analysis of protein stability in solution and in crystal, having different thermal stability and aggregation properties, but a well comparable crystal packing and crystal structures.
Manuscript 1, "New insights in protein stability: a crystalline point of view", deals with a Fourier Transform Infrared spectroscopy (FT-IR) study on protein stability of these β2m variants. The stability of the proteins in solution and in crystal has been compared to evaluate whether the proteins maintain a dynamic also in crystallized conditions. The existence of these correlation would open protein dynamics exploration not only to solution state NMR but also to crystalline solid state NMR.

1.2. Protein stability in immune response

Several efforts have been focused on the understanding of the molecular basis of protein thermostability. Thermostability seems to be linked to the modulation of the canonical forces occurring in all proteins such as electrostatic (hydrogen bonds and ion-pairs (*73*)) and hydrophobic interactions together with many other stabilizing features (greater hydrophobicity, optimized folding, reduction of loops, increased occurrence of proline residues, increased helical content, increased polar surface area (*74*)).
With this perspective, even the contribution of a single residue amino acid, can become crucial in determining the structural stability of a protein or a protein complex.

1.2.1.1. MHC-I structure and stability

The Major Histocompatibility Class I (also known as Human Leucocyte Antigen, HLA) is a heterotrimeric complex composed by a highly polymorphic heavy chain (α) with a transmembrane domain, a monomorphic light chain (β2m) and an antigenic peptide. It is located on the cell surface of almost all nucleated cells, exposing peptides derived from intracellular proteins, eliciting specific immune responses mediated by cytotoxic T lymphocytes (CTLs) (*75, 76*).
The structure of MHC-I can be divided in two globular units. The region close to the membrane consists of the β2m and α3 heavy chain domains that non-covalently associate through the four-strand β-sheet. The second region, consists of α1 and α2 domains, composed of four antiparallel β-strands with a solvent exposed helical region, that associate forming a peptide-binding groove (*77*).
The two domains are hydrogen bonded in a way that the eight β-strands form a flat surface on top of the β2m and α3 domains. The two α-helices form a cleft which is the antigen binding site.

The peptide, composed of 8-11 amino acids, is bound into the cleft and the most of residues side chains pointing outward for interaction with the TCR (*Fig. 7*). The side chains of residues in position 1, 2 and at the C-*terminus* are, instead, involved in the binding with the heavy chain, acting as anchors (Fig. 8) (*78, 79*).

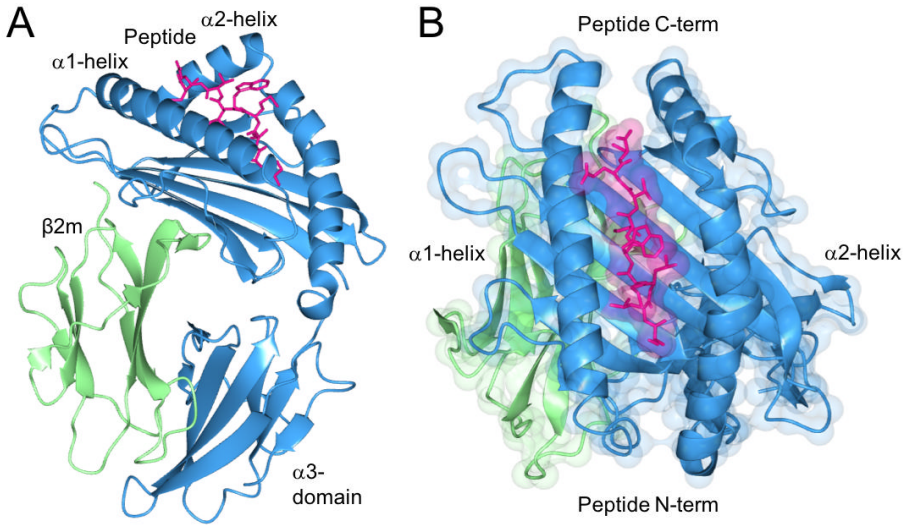


Figure 7. Crystal structure of MHC-I complex (PDB code: 1S9W). **(A)** Complete view of MHC-I X-ray structure with heavy chain (blue), β2m (green) and peptide (magenta). **(B)** Zoomed view of the peptide-binding cleft.

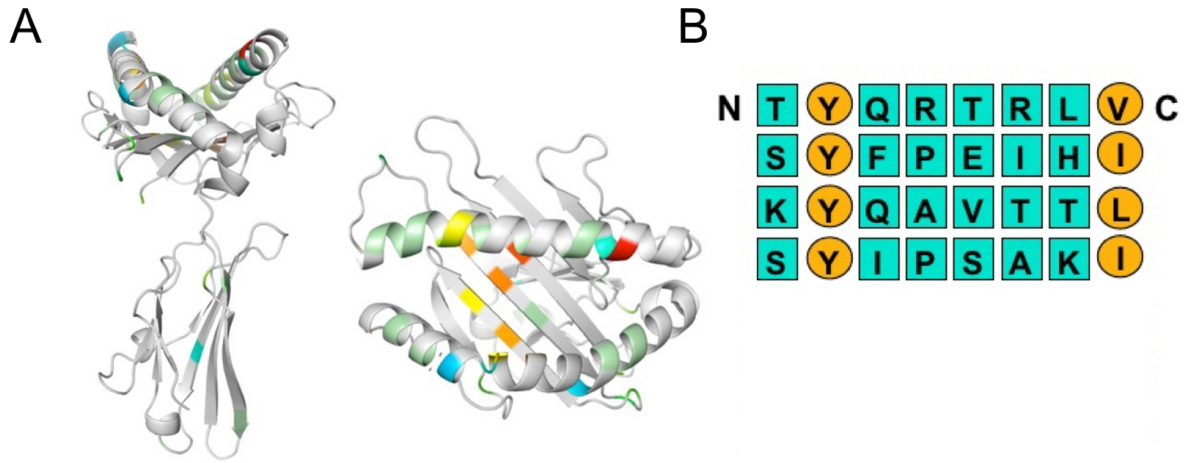


Figure 8. Allelic variations and anchor residues. **(A)** Mapping the sequence variability onto the structure of the HLA-A protein illustrates that the distribution of amino acid substitutions among HLA-A alleles in the population is not random. Large regions of the protein are unaffected by genetic polymorphism whereas the beta-sheet and the two alpha-helices defining the peptide antigen-binding site, are highly polymorphic (color code for variability, highest to lowest: red, orange, yellow, green and light blue). **(B)** Examples of peptides are reported. The most common anchor residues are highlighted in orange

Thanks to the polymorphic nature of the heavy chain *(Fig. 8 B)*, the antigenic surface formed by the bound peptide and the neighbouring α-helices is unique to each pMHC structure and is determinant for recruiting of antigen-specific T-Cell Receptors (TCRs) (*80*).

Despite the high amount of antigenic peptides and, obviously, the high polymorphic variability of the α-chain, the overall structures of the MHC-I are virtually identical, as determined by X-ray crystallography (*81*).

On the other hand, depending on the antigenic peptide, the stability of MHC-I measured *in vitro* can considerably change ranging from 40°C to nearly 70°C (Fig 9.) (*82-84*). Experimental evidences have demonstrate the impact of MHC-I fold stability on downstream events determining the T cells activation (*85, 86*). The general idea is that the higher is the observed *in vitro* stability of MHC-I, the longer is the cell surface presentation, and the more effective is T-cell responses (*87*). Indeed, it has been claimed that stability, rather than affinity, of pMHC-I complexes better correlates with immunogenicity and immune-dominance (*88*).

Peptide name	Sequence	T_m by DSF (°C)	T_m by CD (°C)
Class I peptide/MHC complexes			
MART-1_{27-35}	AAGIGILTV	40.8 (±0.1)	37
MART-$1_{27(2L)-35}$	ALGIGILTV	59.2 (±0.2)	51
MART-1_{26-35}	EAAGIGILTV	45.7 (±0.3)	39
MART-$1_{26(2L)-35}$	ELAGIGILTV	59.4 (±0.3)	55
TAX	LLFGYPVYV	66.3 (±0.1)	65
LLE	LLEEMFLTV	54.0 (±0.6)	52
$gp100_{209(2M)-217}$	IMDQVPFSV	62.2 (±0.5)	60
HUD	LGYGFVNYI	44.8 (±0.3)	40
$HUD_{87(2L,9V)-95}$	LLYGFVNYV	65.1 (±0.5)	63
$NS3_{1406-1415}$	KLVALGINAV	64.0 (±0.6)	64
$NS3_{1406(4T)-1415}$	KLVTLGINAV	63.0 (±0.1)	64
$NS3_{1406(7N)-1415}$	KLVALGNNAV	55.8 (±0.2)	59
$NS3_{1406(3S,4G,7L)-1415}$	KLSGLGLNAV	61.1 (±0.1)	61
WT1	RMFPNAPYL	56.3 (±0.3)	54
TYR	YMDGTMSQV	67.3 (±0.1)	65
TYR (T5F)	YMDGFMSQV	61.4 (±0.3)	63

Figure 9. Comparison of peptide/MHC apparent T_m values determined by differential scanning fluorimetry and circular dichroism. Notice how different peptides bound to the same MHC allomorph can affect complex stability. Figure acquired from L. M. Hellman *et al.*, Differential scanning fluorimetry based assessments of the thermal and kinetic stability of peptide-MHC complexes. *J Immunol Methods* 432, 95-101 (2016) (*82*).

1.2.1.2. Cancer vaccines: antigen stabilization

MHC-I has an essential role in autoimmune diseases and cancer immune surveillance (*89, 90*). Eradication of tumours is associated with a strong cytotoxic T cell response to antigens expressed by the tumour (tumour-associated antigens (TAA)). Many TAA peptides are self-proteins or closely

related to self-proteins, and interact weakly with MHC-I heavy chain, thus being poorly immunogenic (*91*). For this reason, many efforts have been done to improve the efficacy of the immunogen, engineering synthetic peptides (*92*) to dispense to patients through peptide-base vaccines.

Peptide-based vaccination strategies must consider the intrinsic instability of native peptides, since the extracellular environment is rich in proteases and, thus, peptides are rapidly cleared from the body.

Therefore, peptide-based vaccines need to be engineered to improve formulation and storage properties. Furthermore, incorporation of non-natural amino acids with modified side chains or modified peptide bonds there have been engineered to improve peptide binding stability and maintain T cell cross-reactivity (*93-95*).

Currently another approach being adopted to overcome peptide stability include the optimization of the binding to MHC-I by substituting key residues with more appropriate anchor residues.

The ideal TAA for vaccination and immunotherapy should be expressed exclusively in cancerous tissue and are essential for its malignant phenotype. These features are rare, indeed, most TAAs are either self-proteins overexpressed in tumours or mutated self-proteins not easily detectible by the immune system.

Several tumor-specific antigens have been discovered through the application of serological analysis of recombinant cDNA libraries from human tumours(*96*).

Among these, NY-ESO-1 seems to be one of the most promising tumour-specific antigens (*97*): indeed about 20% to 40% of tumours express NY-ESO-1 mRNA (*98-102*), including melanoma, prostate, breast, lung and cervical carcinoma (Fig. 10).

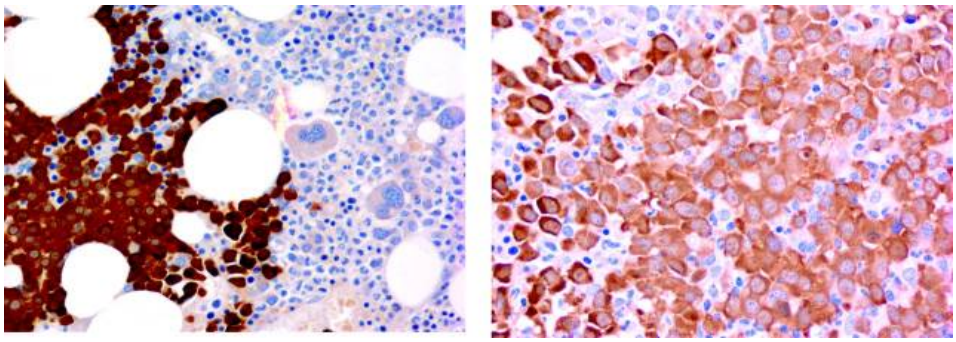


Figure 10. Bone marrow biopsy stained with NY-ESO-1–specific murine IgG1 monoclonal Ab shows strong expression of NY-ESO-1 protein in myeloma plasma cells. Images (left panel, original magnification × 10, numerical aperture 0.3; right panel, original magnification × 20, numerical aperture 0.46) were visualized with an Olympus microscope, type BH2-RFCA. Images were acquired with a SPOT camera (Diagnostics Instruments, Sterling Heights, MI) and SPOT Advanced version 3.5.6 software without image manipulation. Figure acquired from F. van Rhee *et al.*, NY-ESO-1 is highly expressed in poor-prognosis multiple myeloma and induces spontaneous humoral and cellular immune responses. *Blood* 105, 3939-3944 (2005) (*102*)

Thanks to its features, NY-ESO-1 thus offers the possibility to develop a broad-spectrum tumor-specific cancer vaccine. However, several studies have highlighted problems of stability and bioavailability associated with NY-ESO-1 derived-peptide immunization and failure in eliciting robust T-Cells response to kill tumours (*103, 104*).

Jager *et al.* reported three NY-ESO-1 derived peptides (residues 155–163, QLSLLMWIT; residues 157–165, SLLMWITQC; residues 157–167, SLLMWITQCFL) as HLA-A*0201-restricted T-CD8+ cells response from a melanoma patient (*97*).

In a trial vaccination, a mixture of these peptides was administered to patients with NY-ESO-1 expressing tumours, to evaluate their immunogenicity underlining a strong T-CD8+ reactivity and response against NY-ESO-$1_{157–165}$ (*105*).

The presence of a Cys residue at the C-*terminus* of the peptide NY-ESO-$1_{157–165}$ affects vaccines formulation reducing their immunogenic efficacy, the residue has been replaced with more conventional anchor residues generating peptide analogues (*106, 107*). These analogues have shown an increased binding affinity to the complex, but not an enhancement in the immunogenic reaction.

It has been recently proposed a novel method for the design of altered 'super-peptide' ligands (APLs): non-anchoring peptide amino acids are mutate to obtain peptides that bind with high affinity to MHC-I (*108, 109*).

Accordingly, recent outcomes provide insights into the mechanisms controlling the enhanced immunogenicity of super agonist peptides by demonstrating a correlation between the stability of the peptide–MHC–TCR complex, its ability to stimulate a faster formation of conjugates, resulting in improved immunogenic response.

In *Manuscript 2*, "Novel MHC-I complex-stabilizing "superantigens" derived from the cancer target NY-ESO-1", eight MHC-I bearing independent variants of the tumour antigenic peptide NY–$ESO_{157-165}$ have purified. The effects of such variants on MHC-I fold stability has been evaluated. Moreover, the crystal structure of one of the most stable complexes has been determined and its structural analysis helps to rationalise the roles of peptide residues in determining the overall MHC-I stability.

2. Aims of the thesis

The variation of the canonical forces that constrain a polypeptide chain, such as electrostatic forces and hydrophobic interactions, can modulate the stability of a protein and hence its physiological and pathological properties.

The main project is aimed at dissecting the crucial role that a definite amino acid can play on the protein stability and/or on its pathogenicity through two systems and two different biological contexts: β2m and the MHC-I.

I

β2m, responsible for Dialysis Related Amyloidosis in its wild type form, has been subjected to several mutational studies to investigate the molecular bases of its amyloid propensity. These analyses have revealed that among other mutants the W60G variant has an increased conformational stability and a reduced aggregation propensity if compared to WT protein.

The work combines solution NMR spectroscopy and molecular dynamics simulations on WT and W60G β2m variants to investigate whether aggregation properties are exclusively related to the stability of the protein or is the result of an altered protein dynamics.

Thus, to shed light on the molecular determinants that trigger β2m aggregation, three new β2m variants were designed and characterized both biochemically and structurally. The inserted mutations on WT and W60G β2m variants, should alter their aggregation prone exposed surface, modifying their aggregation propensity without altering their stability.

II

To date, it is not fully understood whether protein dynamics and stability are conserved in protein crystals, in other words, if such biophysical properties assessed *in crystallo* are representative of protein behaviour in solution.

In order to start to address such questions, the present work takes as model system the above described three β2m variants WT, W60G and D76N, to carry out a comparative study of protein stability in solution and in the crystalline form.

Thermal unfolding analyses of the proteins in solutions and in crystal have been performed following change in protein secondary structure by FT-IR spectroscopy to assess whether the differences in stability of these variants in solution correlates with the stability in the protein crystal.

III

The heterotrimeric complex MHC-I triggers an immune response by displaying antigenic peptides on the cell surface for recognition by appropriate T cells. It has been previously demonstrated how the stability of the MHC-I complex *in vitro* is correlated to the bioavailability of the complex on the cell surface and to the affinity toward T cell receptors.

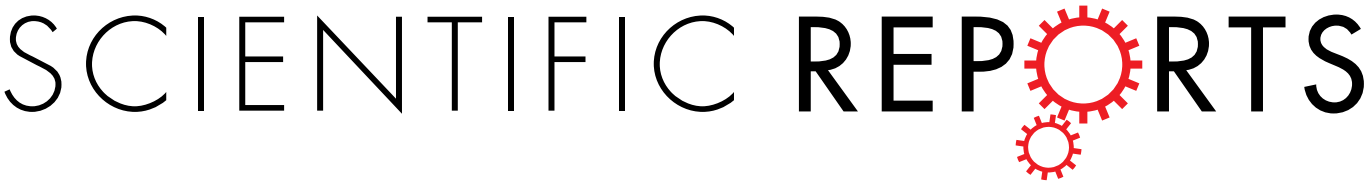
Aim of the work is the exploration of the impact on protein fold stability of complex-stabilizing point mutations on the tumour antigenic peptide NY–ESO-$1_{157\text{-}165}$, valid candidate for cancer vaccines. In this context, eight MHC-I bearing eight peptide variants have been biophysically characterized. Moreover, a structural characterization of the complexes has been performed to investigate the structural features that may contribute to MHC-I stabilization.

3. Results

3.1. Published Paper I

Rational design of mutations that change the aggregation rate of a protein while maintaining its native structure and stability

Camilloni, C., Sala, B.M., Sormanni, P., Porcari, R., Corazza, A., De Rosa, M., Zanini, S., Barbiroli, A., Esposito, G., Bolognesi, M., Bellotti, V., Vendruscolo, M., Ricagno, S. (2016).



OPEN

Rational design of mutations that change the aggregation rate of a protein while maintaining its native structure and stability

Received: 01 March 2016
Accepted: 20 April 2016
Published: 06 May 2016

Carlo Camilloni[1,2], Benedetta Maria Sala[3], Pietro Sormanni[1], Riccardo Porcari[4], Alessandra Corazza[5], Matteo De Rosa[3], Stefano Zanini[5], Alberto Barbiroli[6], Gennaro Esposito[5,7], Martino Bolognesi[3,8], Vittorio Bellotti[4], Michele Vendruscolo[1] & Stefano Ricagno[3]

A wide range of human diseases is associated with mutations that, destabilizing proteins native state, promote their aggregation. However, the mechanisms leading from folded to aggregated states are still incompletely understood. To investigate these mechanisms, we used a combination of NMR spectroscopy and molecular dynamics simulations to compare the native state dynamics of Beta-2 microglobulin (β2m), whose aggregation is associated with dialysis-related amyloidosis, and its aggregation-resistant mutant W60G. Our results indicate that W60G low aggregation propensity can be explained, beyond its higher stability, by an increased average protection of the aggregation-prone residues at its surface. To validate these findings, we designed β2m variants that alter the aggregation-prone exposed surface of wild-type and W60G β2m modifying their aggregation propensity. These results allowed us to pinpoint the role of dynamics in β2m aggregation and to provide a new strategy to tune protein aggregation by modulating the exposure of aggregation-prone residues.

Protein misfolding and aggregation is a widespread phenomenon associated with the intrinsic properties of polypeptide chains[1,2], indeed, it has been observed that the cross-β structure characteristic of amyloid aggregates often represents the most thermodynamically stable state of polypeptide chains[3,4]. Thus, at least at typical cellular concentrations, proteins can be thermodynamically metastable in their native states and do not aggregate only because of the presence of high kinetic barriers[3,4]. It is therefore crucial to uncover the strategies that proteins adopt to remain soluble and escape aggregation *in vivo*, and shedding light on the design principles against aggregation would allow a better control of this process[5,6].

Protein aggregation mechanisms can be identified from the analysis of the many amyloidogenic mutants that have been characterized in recent years, including those of transthyretin, lysozyme, β2μ, and gelsolin[5,6]. Known mutations in these proteins destabilize their native states, often by increasing their dynamics and flexibility[7–11]. When destabilized, native states can undergo larger structural fluctuations, or even unfolding, thus favoring the formation of non-native interactions and, ultimately, the deposition into amyloid aggregates[11].

To determine the molecular origins of these processes, we compare here wild type β2m and its W60G mutant, to investigate the correlation of thermodynamic stability and conformational flexibility *versus* aggregation propensity. β2m is a single-domain protein, characterized by a seven-stranded β-sandwich fold typical of the immunoglobulin domain family; β-strands within the protein are named from A to G (Fig. 1A)[12]. β2m is the light

[1]Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, UK. [2]Department of Chemistry and Institute for Advanced Study, Technische Universität München, Lichtenbergstraße 4, D-85748 Garching, Germany. [3]Dipartimento di Bioscienze, Università degli Studi di Milano, 20133 Milano, Italy. [4]Wolfson Drug Discovery Unit, Centre for Amyloidosis and Acute Phase Proteins, University College London, London NW3 2PF, UK. [5]Dipartimento di Scienze Mediche e Biologiche, Università di Udine, 33100 Udine, Italy. [6]Dipartimento di Scienze per gli Alimenti, la Nutrizione e l'Ambiente, Università degli Studi di Milano, 20133 Milano, Italy. [7]Science and Math Division, New York University Abu Dhabi, Saadiyat Island, Abu Dhabi, UAE. [8]CIMAINA and CNR Istituto di Biofisica, c/o Dipartimento di Bioscienze, Università degli Studi di Milano, 20133 Milano, Italy. Correspondence and requests for materials should be addressed to C.C. (email: carlo.camilloni@ch.tum.de) or S.R. (email: stefano.ricagno@unimi.it)

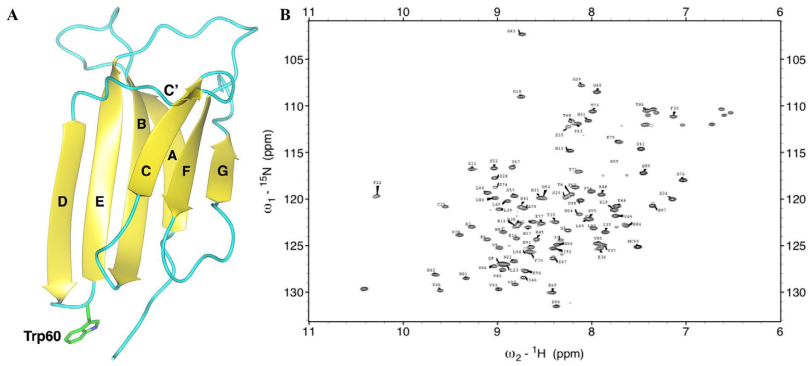


Figure 1. NMR assigment of β2m. (**A**) Ribbon representation of the crystal structure of human β2m (pdb code 2YXF) with β-strands labeled according to standard nomenclature. W60 is shown as sticks. (**B**) HSQC [^{1}H, ^{15}N] spectrum recorded at 11.7 T (500.13 MHz for 1H), 310K, of [U-^{13}C, U-^{15}N W60G β2m] 0.5 mM dissolved in 70 mM phosphate buffer at pH = 6.6 and 100 mM NaCl.

chain of the major histocompatibility complex (MHC) class I[12–14]. While β2m is highly stabilized by the interactions within the MHC class I[15], when β2m is released as a monomer, it may turn into the etiological agent of dialysis-related amyloidosis, a condition triggered by renal impairment and the subsequent chronic accumulation of abnormally high concentrations of β2m in body fluids. β2m deposits in joints, bones and muscles, lead to movement impairment, bones fragility and articular pain[16,17].

Recent efforts in elucidating the molecular determinants of β2m aggregation have highlighted the complex role played by W60, an evolutionary-conserved, solvent-exposed residue. *In silico* simulations of wild-type β2m suggested that W60 is heavily involved in intermolecular interactions[18], therefore the non-natural mutation to Gly (W60G) was prepared, yielding to a variant with unexpected properties[19]. In particular, the (unfolding) thermodynamic stability of this mutant is higher than that of the wild-type β2m, with $\Delta G°(H_2O)$ 22.2 ± 2.0 and 27.6 ± 3.3 kJ mol^{-1}, for wild-type and W60G variants, respectively[19]. While both wild-type and W60G aggregate under strongly denaturing conditions (pH 2.5), W60G shows low aggregation propensity using the standard protocol at pH 7.4 and 20% TFE, conditions under which wild-type β2m aggregates abundantly[19,20]. A wealth of data indicates that the DE loop, where W60 is located, is under backbone geometrical strain in wild-type β2m[18,21–23]. The introduction in the DE loop of a residue endowed with higher conformational freedom, such as Gly, relieves the local conformational strain and results in a β2m variant thermodynamically more stable in solution, with a decreased aggregation propensity[19,24]. The above observations prompted us to select wild-type β2m and its W60G variant as a well-suited system to investigate, besides protein thermodynamic stabilities, other mechanisms that may tune protein aggregation propensity.

To gain further understanding into the molecular bases of the different solution properties displayed by wild-type and W60G forms, we characterized the equilibrium distribution of conformations for the two variants. We employed NMR solution chemical shifts measured under native conditions in combination with replica-averaged metadynamics (RAM) simulations[25–27]. RAM simulations allow integrating the physico-chemical knowledge of a system (*i.e.* the interactions among its atoms as defined in molecular force fields) with the knowledge derived by experimental measurements (*i.e.* chemical shifts), to provide an atomistic description of the Boltzmann distribution of structures that satisfy the experimental data and the maximum entropy principle[28]. Comparison of the resulting ensembles shows how the W60G mutation decreases the overall aggregation propensity by increasing the average β-structure content and by perturbing the overall dynamics of surface residues. To test whether a better average protection of aggregation-prone surface residues can actually be key in regulating protein aggregation, we designed surface mutations aiming to either decrease the aggregation propensity of wild-type β2m, or increase that of the W60G variant. We found that the designed mutants do not cause any relevant structural rearrangement in the protein core, nor they alter β2m thermodynamic stability, but display striking different aggregation propensities. Thus, the approach presented here helps to decouple different molecular determinants contributing together to determine a high aggregation propensity, this strategy may be used to understand and modulate the aggregation properties of other folded proteins without affecting their structures and thermal stabilities.

Results

NMR chemical shift assignment of the W60G mutant. Wild-type β2m was already characterized in terms of ^{1}H, ^{15}N and ^{13}C backbone and side chain assignment[29], whereas for the W60G only ^{1}H resonances and amide nitrogens were previously attributed[19]. Here we extended the NMR analysis to the full backbone, as well as to Cβ and Hα, and in favorable cases to Hβ. The ^{15}N-^{1}H HSQC spectrum at 310 K (Fig. 1B), shows 108 (96 belonging to HNs) of the 110 peaks expected. The missing amide peaks belong to K58 and S88, which were not observed also in the wild-type type form due to unfavorable local exchange broadening[30]. In fact only at very low protein concentration (less than 50 μM), where the presence of dynamic dimers is scarce, these two peaks are visible. Moreover A15 and Y66 NHs overlap in the HSQC spectrum acquired at 500 MHz. The sequential backbone assignment was obtained using heteronuclear 3D triple-resonance experiments acquired on ^{13}C, ^{15}N

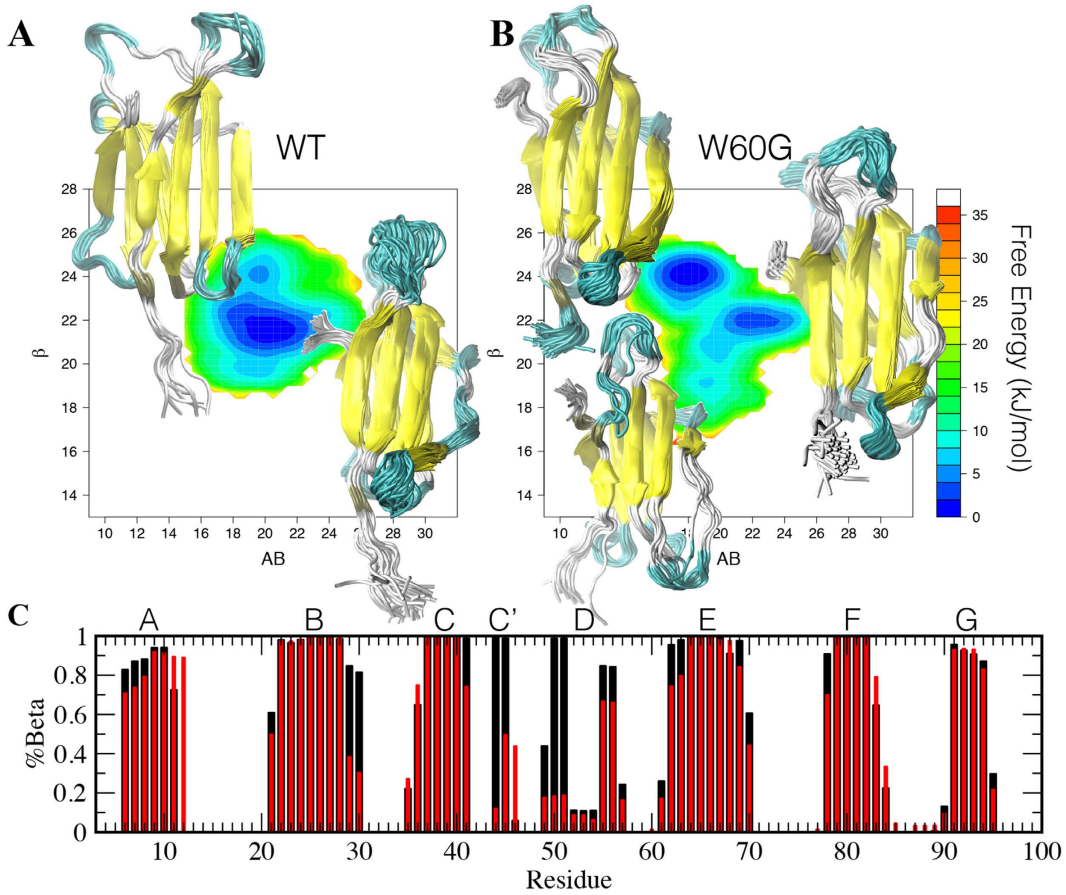


Figure 2. β2m RAM Ensembles. Free energy surfaces (in kJ/mol) for wild-type (**A**) and W60G β2m (**B**) as a function of the side chains rotameric state, AB, and the antiparallel β-structure content (β). (**C**) β-structure populations for wild-type (red bars) and W60G (black bars) β2m. The seven β-strands building the protein fold are identified by A through G labels.

uniformly labeled sample. The assignment percentage for C', Cα, Cβ were 93%, 97%, 86%, respectively. 95% of Hα and only 23% of Hβ were identified through (H)CCH-TOCSY spectra.

Native state dynamics of β2m. We determined the conformational properties of wild-type and W60G β2m using NMR chemical shifts and molecular simulations in the RAM framework[25–27]. In RAM the sampling of the conformational space is enhanced by employing bias-exchange metadynamics[31], while the quality of the force field is improved by including the information content of equilibrium experiments, *i.e.* NMR chemical shifts[25,26], within the framework of the maximum entropy principle[28].

Using the RAM approach we obtained a converged sampling, resulting in a free energy landscape within a statistical uncertainty of less than 2 kJ/mol for free energies up to 20 kJ/mol (Figs S1 and S2). The free energy surfaces for the wild-type and W60G β2m variants, as a function of the side-chain rotamer distribution (AlphaBeta collective variable, AB, see Methods) and the anti-parallel β-structure content (AntiBetaRMSD collective variable, β, see Methods), are shown in Fig. 2A,B. The free energy surfaces indicate that W60G hosts on average a higher content of β-structure than the wild-type protein. By closer inspection, the main differences in the β-sheet populations (β) are localized in the B, C' and D strands (Fig. 2C). These differences shows how the local relaxation of the strain of the backbone in position 60 propagates along the D strand up to C' due to its reduced freedom on one side and through space to the C-terminal of the B strand on the other. Wild-type β2m shows a single well-defined minimum at β = 22 and AB = 20, and a low-populated minimum at higher β-sheet content (β = 24), compatible with a fully structured D strand. Conversely, W60G shows two minima of comparable energy, one at β = 24 and AB = 18, and the other at β = 22 and AB = 22; both minima are close to the two minima of the wild-type protein (Fig. 2). The effect of the W60G mutation on the dynamics of β2m is also reflected by the reduced structural fluctuations in the 25–31 and 40–60 regions (B,C,C' and D strands) as shown in Fig. S3.

Analysis of sequence and structural properties leading to protein aggregation. Wild-type and W60G are characterized by significantly different aggregation properties. While the wild-type protein can form fibrils under both neutral (pH 7.4) and acidic (pH 2.5) conditions, the mutant yields fibrils only under denaturing (pH 2.5) conditions[19,20]. Such difference was explained in terms of the different thermodynamic stability of the two variants. NMR relaxation studies under conditions far from those favoring aggregation confirm that the

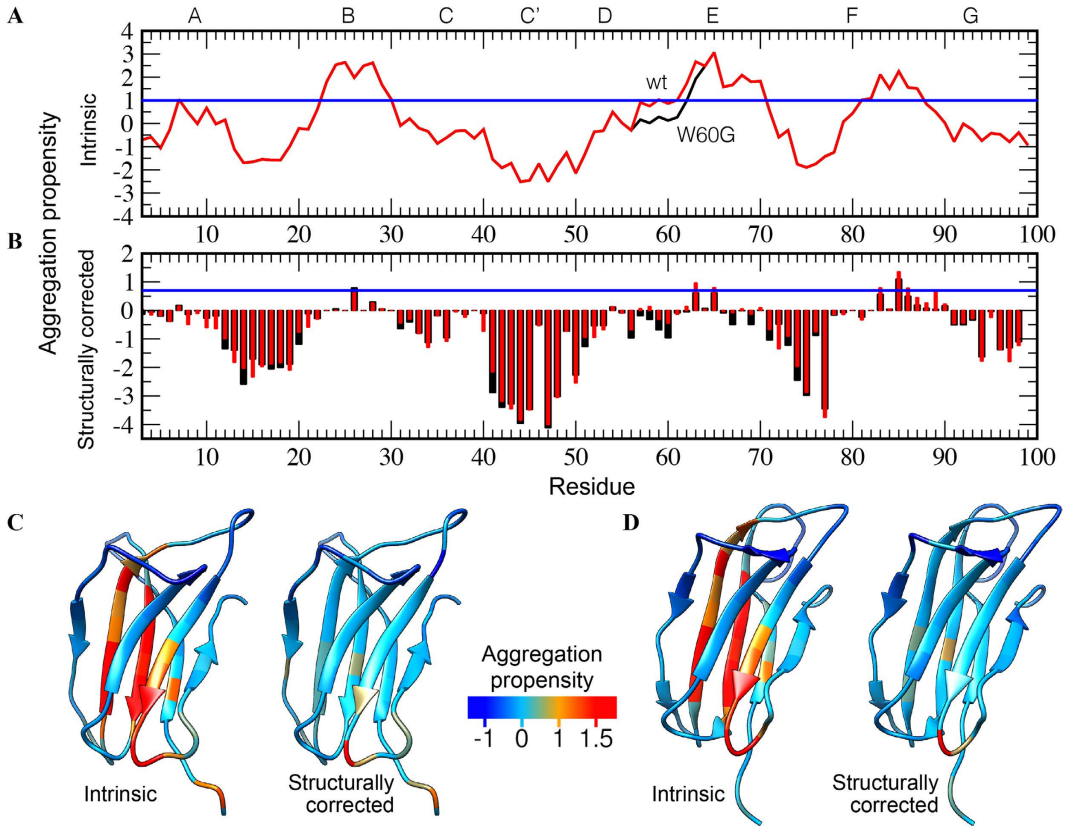


Figure 3. Predicted aggregation properties of β2m wild-type and W60G variants. (**A**) Intrinsic aggregation profiles of the wild-type (red) and W60G (black) variants. (**B**) Structurally corrected profiles calculated as ensemble averages for β2m wild-type (red bars) and W60G (black bars). In (**A**,**B**) values larger than 1 denote aggregation-prone regions. (**C**) Projection on the structure of wild-type β2m of the intrinsic profile (left) and the structurally corrected profile (right). The color code is blue for aggregation-resistant residues and red for aggregation-prone residues. (**D**) Same as (**C**) for the structure of W60G β2m.

wild-type species exhibits a much more pronounced association propensity than the W60G mutant[30]. However, additional insight can be obtained by analyzing the amyloidogenic properties of β2m in terms of sequence- and structure-based properties. The intrinsic and structurally corrected aggregation propensities were assessed using the CamSol method (Fig. 3A,B)[32], which is highly accurate in predicting mutation-induced changes in protein solubility, a property often strongly correlated with aggregation[33]. In the present work we used aggregation propensity scores that are the inverse of the CamSol solubility scores (so that higher scores represent higher aggregation propensity) and, in order to analyze the RAM ensemble, we have defined a 'total structurally-corrected score' (see Methods). The latter can be used to predict the tendency to aggregate of individual conformations of the ensemble; it is calculated from the structurally corrected CamSol profile in the same way the CamSol intrinsic score is calculated from the intrinsic profile[32]. The ensemble average of the total structurally corrected scores for the two β2m variants is -0.59 ± 0.01 and -0.66 ± 0.01, for the wild-type and W60G, respectively. Such scores are in keeping with the experimental aggregation data, *i.e.* wild-type β2m is more aggregation prone than W60G β2m under native conditions.

The intrinsic and the structurally corrected profiles are represented in Fig. 3C,D, color-coded on the structures of wild-type and W60G. Aggregation-prone residues are mostly located in the B, E strands, and at the C-terminal of the F strand (Fig. 3A). These β-strands are the central strands in the two β2m β-sheets; the B and E strands are flanked by the A and D strands (in the 4-stranded β-sheet), while the F strand is surrounded by the C and G strands (in the 3-stranded β-sheet) (Fig. 1A). It is noteworthy that all these flanking β-strands consistently prove less aggregation-prone than the central ones, according to both the intrinsic and the structurally corrected CamSol profiles (Fig. 3B), an observation that may be interpreted as a strategy to protect the protein from unwanted aggregation. Furthermore, at least half of the aggregation-prone residues point towards the protein core, thus being excluded from the chance of unwanted interactions with other proteins under native conditions (Fig. 3B). The analysis recognizes the A strand and the CD and EF loops as the protein protective regions. Interestingly, the strong aggregation propensity of the naturally occurring truncated β2m variant, devoid of the first six residues (ΔN6)[34], can be rationalized in the above context; in fact, by shortening the protective A strand, the C-terminal residues of the B strand, which are the most aggregation-prone, become solvent exposed. By the same token, one may also explain the behavior of D76N-β2m, the naturally occurring mutant responsible for a severe systemic amyloidosis[10]. In fact, the 'fatal' mutation leading to the most amyloidogenic isoform of β2m so far described occurs in the protective EF loop.

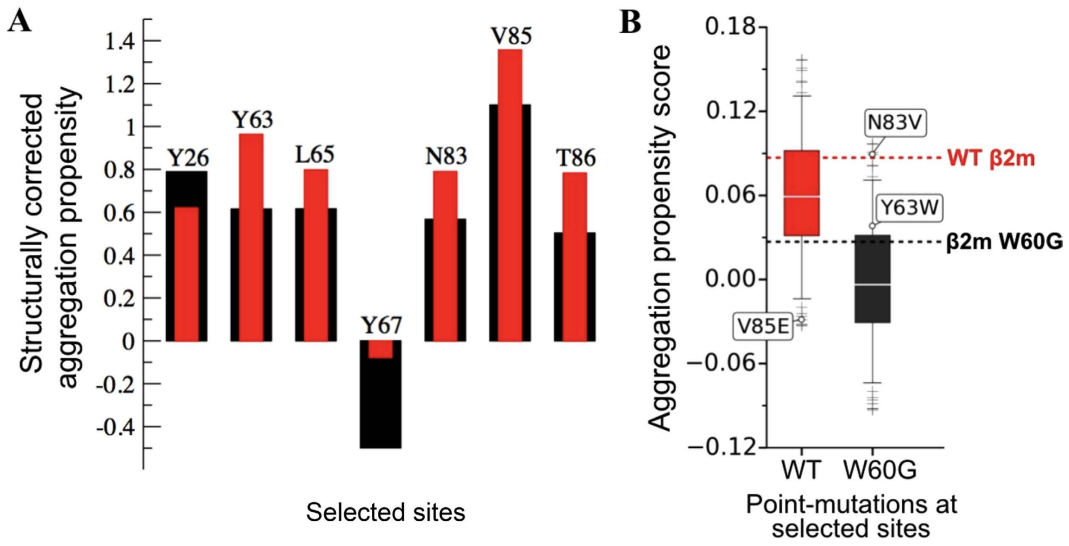


Figure 4. Rational design of β2m mutants. (**A**) Aggregation-prone residues (intrinsic score > 1) for which the protective effect of the W60G mutation is the highest, defined as those residues with the largest change in the structurally-corrected score between the wild-type (red) and W60G (black) β2m. (**B**) Predicted aggregation propensity scores calculated using CamSol for all possible amino acid substitutions at the sites shown in panel A in the wild type (left box-plot in red) and in the W60G variant (right box-plot in black). The red and black dashed lines correspond to the aggregation propensity score of the WT and the W60G variant, respectively. Aggregation-protective and aggregation-promoting mutations selected for experimental validation are flagged. V85E and N83V mutations have been chosen as those with the largest predicted effect, while Y63W was chosen as a weakly aggregation promoting mutation. In each box the horizontal line denotes the median of the distribution; whiskers extend from the 5th to the 95th percentile of the distribution and boxes from the lower to the upper quartile.

The structurally corrected aggregation propensity profile for W60G shows an increase in the average protection of the aggregation-prone residues (Figs 3B,D and 4). Such an effect can be linked to the reduced strain at site 60 that propagates thorough the D strand in the 4-stranded β-sheet to the opposite β-sheet (Fig. 3D); indeed W60G shows on average a higher β-structure content distributed along all its β-strands (Fig. 2A–C) together with reduced fluctuations (Fig. S3).

Rational design of mutants with different aggregation propensities. The above analysis suggests then that the role of the W60G mutation in reducing aggregation is not exclusively due to stabilization of the protein. In particular, by optimizing the geometry of the D strand through the W60G mutation, all the aggregation-prone residues of the central β-strands, in both β-sheets, become on average more protected (Fig. 3). These observations suggest that by specifically modulating the average surface properties arising from the native state dynamics, it may be possible to tune the aggregation properties of β2m.

To test the role of dynamics on β2m native state aggregation we searched for mutations able to decrease the aggregation propensity of wild-type β2m, on one hand, or increasing the aggregation propensity of W60G, on the other, without affecting their thermodynamic stabilities and structures. In Fig. 4A the residues with an intrinsic score higher than 1, for which the W60G mutation results in a large variation (>0.2) of the structurally corrected score with respect to the wild-type are highlighted. These seven residues are all located in the central strands of both β-sheets. Table S1 reports the amino acid substitutions, known within the β2m protein family, that occur at such selected residue sites[35]; the sequence data highlight that none of these sites show strong residue conservation. Mutations of residues 65 and 67, in the E strand, have already been studied[36], even though under different experimental conditions, showing that residues in this β-strand are critical for aggregation. Mutations in the F strand have not yet been investigated. By calculating the change in the protein intrinsic score upon all the possible single mutations at these seven sites (Fig. 4B), we suggest that the V85E β2m variant should reduce the wild-type aggregation propensity, while two alternative specific mutations in the W60G β2m variant (W60G-Y63W and W60G-N83V) are expected to weakly (the former) or largely (the latter) increase the W60G aggregation propensity. Indeed, the score of V85E β2m is lower than that of W60G, the one of W60G-N83V is comparable to that of the wild-type while W60G-Y63W is slightly larger than that of W60G (Fig. 4B). All three mutations are located on the protein surface, and two of them are also naturally occurring in vertebrates other than human; hence they should cause little perturbation to the protein structure and stability. It is noteworthy that the effects of surface mutations introducing or removing surface charges, do not correlate simply with aggregation propensity: the V84D and Y86E mutations in acylphosphatase from *S. solfataricus* do not protect the protein from unfolding and aggregation[37]. Furthermore, a D to N systematic scanning in β2m indicated that the removal of a negative charge triggers measurable effects in aggregation propensity only when the D to N mutation occurs at a specific site[38].

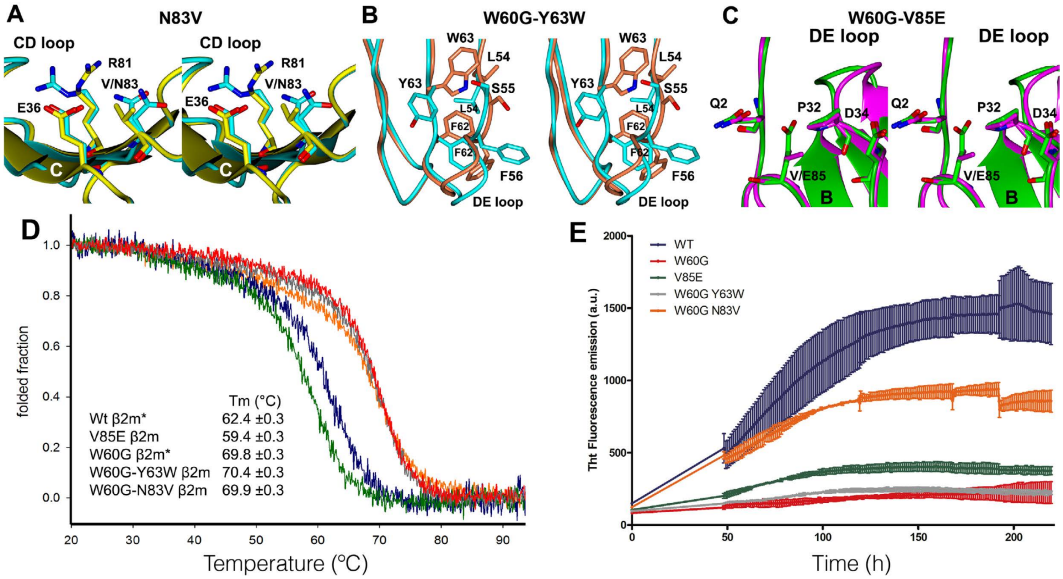


Figure 5. Structure, stability and aggregation of the β2m variants. (**A**) The superimposed structures of the W60G and W60G-V83N mutants are shown in cyan and yellow, respectively. (**B**) Zoom of the (**D** and **E**) strands: the superposed structures of the W60G and W60G-Y63W are shown in cyan and coral, respectively. (**C**) Superimposed structures of wild-type β2m and the V85E variant, at the mutation site, shown in magenta and green, respectively. (**D**) Thermal unfolding of wild-type β2m, of the W60G variant and of the three surface mutants, monitored by far-UV CD at 202 nm. In the table the measured Tm are reported (*the Tm of wild-type and W60G variants have been previously reported[23]). (**E**) Comparison of the kinetics of fibril formation of the various mutational variants of β2m analyzed in this work monitored by fluorescence in a thioflavin T assay.

Structural characterization of three rationally-designed β2m surface mutants. The three β2m surface mutants that we rationally designed (V85E, W60G-N83V, W60G-Y63W) were produced as recombinant proteins, purified to homogeneity, crystallized, and their structures determined at high resolution (1.75 Å, 1.49 Å, 1.70 Å, for V85E, W60G-Y63W, and W60G-N83V, respectively). Although the three crystals belong to the same space group, unit cell edges differ and crystal packing is distinct for each of the three mutants (see methods). The three structures display electron density of excellent quality along the polypeptide chains. Data collection and refinement statistics are reported in Table S2.

Inspection of the refined X-ray crystal structures shows that all three surface mutations are well tolerated in the β2m fold, and backbone conformational changes are relegated to loops (Table S3). Indeed, the largest structural deviations in the three mutants occur far from the mutation sites, at the AB loop (residues 12–21) that is typically in an open conformation in the crystal structures of isolated β2m, while it is closed over the rest of the β-sandwich when β2m is part of the MHC class I and in few structures of monomeric β2m[39–41]. Notably, the AB loop in a closed conformation is consistently observed in solution NMR structures[42]. In particular, for the crystal structures here reported of the W60G-Y63W and V85E mutants a closed AB loop conformation was observed; by contrast, the wild-type β2m, W60G and W60G-N83V 3D structures exhibit an open AB loop conformation. All these data suggest that the AB loop can easily adjust to different conformations, which, in the case of the present mutants, is likely the result of convenient molecular packing in the crystals.

Detailed inspection shows that the structure of the W60G-N83V mutant is perfectly superimposable onto that of W60G β2m. In fact, in W60G the side chain of N83, which is located on the F-strand, although surrounded by two charged residues (E36 and R81), does not establish any hydrogen bond with the neighboring residues; thus, the N83V substitution leaves both the conformation of residue 83 and of the surrounding side chains unaffected (Fig. 5A).

Residue 63 lies in the E strand, in close vicinity of the DE loop (residues 57–60). The Y63W mutation triggers just a minor reorientation of the DE loop backbone, but is responsible of extensive, though local, side chain reorganization. In the structure of the W60G-Y63W mutant the bulkier W63 side chain flips away from the DE loop and towards the D-strand, compared to what is observed in W60G. Such movement has two consequences: (1) the shift of S55, F56, L54 and F62 side chains away from W63, and (2) the DE loop moves about 2 Å towards residue 63 (Fig. 5B).

The comparison of the wild-type β2m and the V85E mutant structures reveals that the mutation does not elicit local rearrangements in the FG loop where E85 is located, nor in the surrounding residues. The 3D-structure of V85E mutant displays modified conformations in the DE and C'D loops; however, given that both regions are far from the mutation site, these differences may be due to the specific crystal packing (Fig. 5C).

Stability of the rationally-designed β2m surface mutants. The stability of all three mutants described above was evaluated by monitoring their thermal denaturation through far-UV CD measurements.

In keeping with the goals set (*i.e.* designing mutants that maintained the same thermodynamic stability either of wild-type β2m or of the W60G mutant) and with the outcomes of the crystallographic analyses, none of the three mutations was found to alter dramatically the protein stability. Upon V85E mutation, the wild-type β2m stability was slightly impaired, with a T_m value decreased by about 3 °C relative to wild-type β2m. The negative charge of E85 lies 4.5 Å away from the carboxylate of D34, the electrostatic repulsion between these two residues may contribute to the slightly lower stability observed for the V85E mutant compared to the wild-type protein (Fig. 5C). Conversely, introduction of the Y63W and N83V mutations (added to the W60G mutation) did not vary significantly the T_m value recorded for the W60G mutant (Fig. 5D).

Aggregation measurements of the rationally-designed β2m surface mutants. To further test the effects of the three designed mutations, the aggregation propensity of the V85E, W60G-Y63W, and W60G-N83V mutants was tested at pH 7.4 and in 20% TFE, conditions under which wild-type β2m aggregates readily while the W60G mutant yields only small amounts of aggregates[19,43]. Figure 5E shows that the V85E mutation, despite being slightly less stable than the wild-type protein, almost totally abolished wild-type β2m aggregation-propensity, and that the W60G-N83V variant presented a markedly increased aggregation trend with respect to W60G β2m. The W60G-Y63W variant displayed the same very low aggregation propensity as the W60G mutant, suggesting that mutations should be chosen in order to maximize the difference of the total average CamSol score (Fig. 4B).

Discussion

As protein aggregates are often thermodynamically stable in living organisms[1], the avoidance of the aggregation process has been optimized through evolution by means of multiple strategies[44]. While efficient protection mechanisms work at the cellular level by regulating protein homeostasis[45,46], many strategies to prevent aggregation concern the properties of the native states of proteins, including a high thermodynamic and kinetic stability, limited conformational fluctuations, and the presence of structural motifs that are incompatible with the cross-β assembly[1,4,47]. As folded proteins are metastable with respect to aggregation[3], mutation that destabilize the folded states can dramatically increase the rates of aggregation. Furthermore, mutations that change the accessibility to a high free energy state, without changing the overall stability of the protein native state, are enough to enhance aggregation[11,48]. Conversely, mutations that introduce structural protections on edge strands play a protective role independently of the overall protein stability[37].

Here we investigated the different roles played by stability and dynamics in protein aggregation using human β2m, a protein that, when released as a monomer and accumulated in patients' sera, may become the etiological agent of dialysis-related amyloidosis, forming amyloid deposits in joints, bones and muscles[17]. Esposito *et al.* reported previously an engineered β2m variant bearing the W60G mutation, which was found to be thermodynamically very stable (Cm increased by 0.7 M GdHCl) and resistant to aggregation relative to the wild type β2m[19,20]. To elucidate the molecular determinants underlying such distinct properties, we characterized the native state fluctuations of the wild-type protein and of the stable W60G mutant using NMR chemical shifts and RAM simulations, and analyzed the aggregation properties of the resulting conformational ensembles.

Our simulation results show that the W60G mutation greatly reduces the flexibility of the protein native state (Fig. S3), and increases the overall content of residues adopting β-structure (Fig. 2), in keeping with previous experimental data[19,20,30]. Moreover, our analysis shows that β2m hosts three aggregation-prone regions located essentially in three β-strands that are central in the β2m two β-sheets, *i.e.* strands B, E and F (Fig. 1A), and that in these regions those residues that are exposed on the surface are better protected on average in the W60G variant than in the wild-type (Fig. 3). The differences in the native state dynamics together with the similarity of their sequence-based intrinsic scores, explains the different behaviors observed under native and denaturing conditions, *i.e.* while both protein variants aggregate under denaturing conditions (pH 2.5), W60G shows little aggregation propensity under the standard aggregation protocol, at pH 7.4 and 20% TFE[19,20].

To probe the role played by the residues in the three central β-strands in determining β2m aggregation propensity, and to further investigate the protective effect of the W60G mutation, we designed three mutations aimed to modulate β2m aggregation trends under native conditions, while leaving β2m stability and structure unaltered. To this aim, we focused on mutations of protein surface residues that turned out to be better protected in the W60G ensemble compared to the wild-type protein. The V85E single site β2m mutant, together with the W60G-Y63W and the W60G-N83V double mutants were accordingly designed and produced in recombinant form. Consistent with such rational approach, we found that all three surface mutants maintain unaltered the β2m core structure (Fig. 5 and Table S2) and display fold stabilities very similar to wild-type β2m (for V85E), or to W60G (for the two double mutants), respectively. Conversely, analysis of their aggregation properties indicates that surface mutations can indeed modulate β2m aggregation propensity. The V85E mutation, on the F strand, abrogates completely wild-type β2m amyloidogenicity, under the tested conditions, despite being slightly less stable than the wild-type; the W60G-N83V mutant (residue 83 falls on the F strand) gains a non-negligible aggregation propensity relative to W60G without any difference in protein stability. We cannot rule out that the fibrillar aggregate may form following a different aggregation pathway compared to wild-type β2m as it was shown in details for mutants of acylphosphatase from *S. Solfataricus*[49]. However in such case this phenomenon was coupled to very different thermodynamic stability of the native protein[49]. The W60G-Y63W double mutant, on the other hand, displays an aggregation trend comparable to W60G and in agreement with the small change in the aggregation score (Fig. 4B), indicating that this conservative mutation cannot encode noticeable changes of the aggregation propensity.

β2m hosts extensive aggregation regions on protein surface (Fig. 3), this being rooted in its sequence and structure that are specifically optimized to be part of the MHC class I. In particular, the aggregation-prone residues W60 and Y63 are buried at the interface between β2m and the MHC class I heavy chain[19], consistent with the observation that protein interface residues are typically aggregation prone[50]. On the contrary, the F

strand, which has not been previously studied, plays a pivotal role in the aggregation propensity of β2m, and it is solvent-exposed in MHC class I. In particular, site 85 in human β2m bears a dangerously aggregation-prone Val residue. Such apparent design error, however, could be reconciled with a recent observation[15] whereby the MHC class I heavy chain stabilizes markedly the β2m fold.

In summary our results provide an explanation for the resistance to aggregation associated with the W60G mutation of β2m. This mutation not only brings a significant increase in the thermodynamic stability of the protein, but also generates a more regular and less flexible overall structure in which the surface aggregation-prone regions are better protected. We suggest that such second feature is primarily responsible for the reduced aggregation propensity of this variant. Based on these observations, we showed that the aggregation propensity of β2m can be tuned by mutating surface residues without affecting its structure and stability, thus highlighting the crucial role played by the dynamics of surface residues in protein aggregation. Our results elucidate a mechanism whereby the aggregation propensity of a protein can be modulated by mutations that change the average protection of its aggregation-prone surface residues with minor effects on the structure and stability of the protein itself.

Materials and Methods

Multidimensional NMR spectroscopy. All data were recorded on a U-^{13}C, U-^{15}N labeled W60G β2m sample 0.5 mM in phosphate buffer 70 mM, NaCl 100 mM, pH 6.6, at 310K with a Bruker Avance 500 and with a Varian INOVA spectrometers operating at proton frequency of 500 and 800 MHz, respectively. Proton chemical shifts were referenced to 2,2-Dimethyl-2-silapentane-5-sulfonate sodium salt, DSS, whose resonance was set to 0.0 ppm. ^{13}C and ^{15}N chemical shifts were referenced indirectly to DSS, using absolute frequency ratio[51]. Sequence specific assignment were achieved using [1H, 15N]-HSQC and [1H, 13C]-HSQC together with 3D triple-resonance experiments, in particular by using the [15N, 1H]- HNCA/[15N, 1H]- HN(CO)CA for the identification of patterns of sequentially linked spin systems and 3D [15N, 1H]- HNCACB/[15N, 1H]-HN(CO)CACB pair of experiments, which allowed the identification of the amino acid types. 3D (H)CCH-TOCSY was used for the assignment of Hα and Hβ resonances.

The spectra were processed with Topspin 2.0 (Bruker Biospin) and analysed into the Sparky (T. D. Goddard and D. G. Kneller, SPARKY3, University of California, San Francisco) framework.

Molecular dynamics simulations. All the simulations in the present work were performed using GROMACS compiled with PLUMED and ALMOST. The system was simulated using the Amber03W force field in explicit TIP4P05 water. A time step of 2 fs was used together with LINCS constraints. van der Waals and short-range electrostatic interactions were cut-off at 0.9 nm while long range electrostatic was treated with the Particle Mesh Ewald method and a mesh size of 0.12 nm. The canonical ensemble was enforced by keeping the volume fixed and by thermosetting the system with the Nosé-Hoover thermostat at 310K. The starting conformation for the wild-type was taken from the 1BMG X-Ray structure while for W60G was taken from the NMR structure 2VB5. The structures were protonated and solvated with 11500 water molecules in a dodecahedron box of 360 nm^3 of volume.

Replica-averaged Metadynamics. RAM simulations were performed using chemical shifts as replica-averaged restraints and bias-exchange metadynamics. The bias-exchange metadynamics approach combines replica exchange[52] with metadynamics, in which several metadynamics simulations are performed in parallel on different replicas of the system, each replica biasing a different collective variable (CV). Exchanges between the replicas are attempted periodically according to a replica-exchange scheme. Four replicas of the system were simulated in parallel at 310 K with a restraint applied on the average value of the CamShift back-calculated NMR chemical shifts where the force constant is set to 24 kJ/(mol ppm^2).

Each of the four replica is bias along one of the following four collective variables, CVs: the anti-parallel β content (the 'β' CV), the parallel β-sheet content (the 'Pβ' CV), the AlphaBeta collective variable defined over all the chi1 angles for the hydrophobic side-chains (the 'AB' CV) and the AlphaBeta collective variable defined over all the phi and psi backbone dihedral angles of the protein (the 'bbAB' CV). The choice of the secondary structures CVs was guided by the fact that the protein is a full beta protein. The to AlphaBeta CVs are designed to sped up the fluctuations of the backbone and the side-chains irrelevant of their specific conformation. Gaussians deposition was performed with an initial rate of 0.125 kJ/mol/ps, a bias-factor of 8 and with σ values set to 0.1, 0.04, 0.16 and 0.25, for β, Pβ, AB and bbAB, respectively. Furthermore, in order to limit the extent of accessible space along each collective variable and correctly treat the problem of the borders, we set the bias as constant outside a defined interval for each CV; intervals were set to 19–29, 3–7, 14–30 and 47–71 for the four CVs, respectively. Each replica have been evolved for 200 ns, with exchange trials every 50 ps.

The sampling of the four replicas was used to generate a four-dimensional free energy landscape, as a function of the before mentioned CVs, where a set of microstates is identified by dividing the four-dimensional CV-space into a homogeneous grid of small dimensional hypercubes and their free energy is obtained using a standard weighted histogram analysis. Bi-dimensional FESes are obtained integrating out the variables not showed.

Cloning, mutagenesis, expression and purification. Recombinant vectors pEX containing genes encoding for W60G-Y63W, W60G-N83V variants, flanking NdeI/XhoI sites were purchased from Eurofins MWG Operon©. An empty vector pET-29b and the vectors carrying the β2m variants were digested with NdeI and XhoI enzymes (New England Biolabs (UK) Ltd©) to obtain a linearized pET29 vector and the isolated genes. Wild-type β2m, W60G-Y63W and W60G-N83V genes were inserted in pET-29b vector with a ligation reaction (Thermo Scientific Rapid DNA Ligation Kit, Life Technologies©). pET-29b-WTβ2m plasmid was used as template DNA for Site-Directed Mutagenesis to obtain the V85E mutant. The following primers were used: 5′-GCC TGC

CGT GTG AAC CAT **GAG** ACT TTG TCA CAG CCC-3′, 3′-GG GCT GTG ACA AAG **TCT** CAT GGT TCA CAC GGC AGG C-5′. All mutants were expressed and purified as previously reported[34].

Crystallization and X-ray structure determination. All mutants were crystallized with sitting drops technique. Lyophilized W60G-Y63W, W60G-N83V and V85E β2m mutants were solubilized in ddH_2O at a concentration of 10 mg/mL, 8.5 mg/mL and 8 mg/mL respectively at 20 °C. β2mW60G-Y63W crystals were obtained in 0.1 M HEPES pH 7.5, 20% (w/v) PEG 10K buffer (Crystal Screen HT™, Hampton). The crystals were cryoprotected in 0.08 M HEPES pH 7.5, 18% PEG 10K, and 20% (v/v) glycerol and flash-frozen in liquid nitrogen. W60G-N83V crystallized in 0.1 M MES pH 6, 21% PEG 4K, 15% glycerol, 0.2 M ammonium acetate, while V85E mutant was crystallized in 0.1 M ammonium acetate pH 5.5, 30% PEG 4K, 15% glycerol and 0.2 M ammonium acetate. The crystals were flash-frozen using mother liquor as cryoprotectant.

X-ray data were collected at ESRF (European Synchrotron Radiation Facility of Grenoble–France) at the ID23-2 beam line. The diffraction data were analyzed and processed using MOSFLM[53], the crystal symmetry was then verified by POINTLESS[54] and the intensities were merged with SCALA[55]. The crystal structure was determined by molecular replacement using MOLREP[56] using the structure of W60V β2m mutant (pdb code: 2Z9T) as searching model[22]. The model molecules placed in the asymmetric unit were subjected firstly to a rigid-body refinement and then to a restrained refinement using Phenix Refine in the Phenix program suite[57]. Manual model building, addition of water molecules and ligands were then performed using the molecular graphic software Coot[58].

All three mutants crystallized in C2 space group however the underlying crystal packing is different: the two double mutants (W60G-Y63W, W60G-N83V) have one molecule in the asymmetric unit (AU); however, the different conformation of the AB loop determines distinct intermolecular interactions in the crystals. Unusually, the AU of the V85E structure contains two molecules. These different crystal packings determine some changes in the protein regions involved in the intermolecular interactions, however, these differences are not deemed relevant for the discussion on the properties of the three mutants.

Circular dichroism spectroscopy. Thermal stability experiments, performed in the far-UV region, were carried out on a J-810 spectropolarimeter (JASCO Corp., Tokyo, Japan) equipped with a Peltier system for temperature control. The protein concentration was 0.1 mg/mL in 50 mM sodium phosphate pH 7.4. The temperature ramp measurements were recorded from 20 to 95 °C (temperature slope 50 °C/hour) in a 0.1 cm path length cuvette and monitored at 202 nm wavelength. T_m was calculated as the first-derivative minimum of the traces. Spectra before and after unfolding ramp were recorded (260–190 nm).

Aggregation assays. Samples (100 μL) of recombinant variants W60G, V85E, W60G-Y63W, W60G-N83V, and wild-type β2m at 100 μM in 50 mM phosphate buffer (pH 7.4), 100 mM NaCl, 20% trifluoroethanol and containing 10 μM Thioflavin T (ThT) (SIGMA), were incubated at 37 °C in Costar 96-well black- wall plates sealed with clear sealing film (4TITUDE) and were subjected to 900 rpm double-orbital shaking. In each well fragmented fibrils of wild type β2m were added as seeds. Bottom fluorescence was recorded at 15-min intervals (BMG LABTECH FLUOstar Omega). Fluorescence was monitored in three or more replicate tests.

NMR assignment deposition. 1H, 15N and 13C assignments of W60G β2m have been deposited to Biological Magnetic Resonance Bank with accession code 25809.

Structure deposition. Atomic coordinates and structure factors for the β2m mutants W60G-Y63W, W60G-N83V, V85E have been deposited with the Protein Data Bank, with accession codes 5CFH, 5CKA and 5CKG, respectively.

References

1. Dobson, C. M. Protein folding and misfolding. *Nature* **426,** 884–890, doi: 10.1038/nature02261 (2003).
2. Knowles, T. P., Vendruscolo, M. & Dobson, C. M. The amyloid state and its association with protein misfolding diseases. *Nat Rev Mol Cell Biol* **15,** 384–396, doi: 10.1038/nrm3810 (2014).
3. Baldwin, A. J. *et al.* Metastability of native proteins and the phenomenon of amyloid formation. *J Am. Chem. Soc.* **133,** 14160–14163, doi: 10.1021/ja2017703 (2011).
4. Tartaglia, G. G., Pechmann, S., Dobson, C. M. & Vendruscolo, M. Life on the edge: a link between gene expression levels and aggregation rates of human proteins. *Trends Biochem. Sci.* **32,** 204–206, doi: 10.1016/j.tibs.2007.03.005 (2007).
5. Chiti, F., Stefani, M., Taddei, N., Ramponi, G. & Dobson, C. M. Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature* **424,** 805–808, doi: 10.1038/nature01891 (2003).
6. Tartaglia, G. G. & Vendruscolo, M. The Zyggregator method for predicting protein aggregation propensities. *Chem. Soc. Rev.* **37,** 1395–1401, doi: 10.1039/b706784b (2008).
7. Chen, C. D. *et al.* Furin initiates gelsolin familial amyloidosis in the Golgi through a defect in Ca(2+) stabilization. *The EMBO J* **20,** 6277–6287, doi: 10.1093/emboj/20.22.6277 (2001).
8. Hammarstrom, P., Wiseman, R. L., Powers, E. T. & Kelly, J. W. Prevention of transthyretin amyloid disease by changing protein misfolding energetics. *Science* **299,** 713–716, doi: 10.1126/science.1079589 (2003).
9. Merlini, G. & Bellotti, V. Molecular mechanisms of amyloidosis. *N Engl J Med* **349,** 583–596 (2003).
10. Valleix, S. *et al.* Hereditary systemic amyloidosis due to Asp76Asn variant beta2-microglobulin. *N Engl J Med* **366,** 2276–2283, doi: 10.1056/NEJMoa1201356 (2012).
11. De Simone, A. *et al.* Experimental free energy surfaces reveal the mechanisms of maintenance of protein solubility. *Proc. Natl Acad. Sci. USA* **108,** 21057–21062, doi: 10.1073/pnas.1112197108 (2011).
12. Bjorkman, P. J. *et al.* Structure of the human class I histocompatibility antigen, HLA-A2. *Nature* **329,** 506–512 (1987).
13. Achour, A. *et al.* Structural basis of the differential stability and receptor specificity of H-2Db in complex with murine versus human beta2-microglobulin. *J Mol. Biol.* **356,** 382–396, doi: 10.1016/j.jmb.2005.11.068 (2006).
14. Porcelli, S. A. & Modlin, R. L. The CD1 system: antigen-presenting molecules for T cell recognition of lipids and glycolipids. *Annu Rev Immunol* **17,** 297–329 (1999).

15. Halabelian, L. *et al.* Class I Major Histocompatibility Complex, the Trojan Horse for Secretion of Amyloidogenic beta2-Microglobulin. *J Biol. Chem.* **289,** 3318–3327, doi: 10.1074/jbc.M113.524157 (2014).
16. Floege, J. & Ketteler, M. beta2-microglobulin-derived amyloidosis: an update. *Kidney Int Suppl* **78,** S164–171, doi: kid7823 (2001).
17. Gejyo, F. *et al.* A new form of amyloid protein associated with chronic hemodialysis was identified as beta 2-microglobulin. *Bioch. Biophys. Res. Comm.* **129,** 701–706 (1985).
18. Fogolari, F. *et al.* Molecular Dynamics Simulation Suggests Possible Interaction Patterns at Early Steps of {beta}2-Microglobulin Aggregation. *Biophys J* **92,** 1673–1681 (2007).
19. Esposito, G. *et al.* The controlling roles of Trp60 and Trp95 in beta2-microglobulin function, folding and amyloid aggregation properties. *J Mol. Biol.* **378,** 885–895 (2008).
20. Rennella, E. *et al.* Folding and fibrillogenesis: clues from beta2-microglobulin. *J Mol. Biol.* **401,** 286–297 (2010).
21. Gianni, S. *et al.* Understanding the frustration arising from the competition between function, misfolding, and aggregation in a globular protein. *Proc. Natl Acad. Sci. USA* **111,** 14141–14146, doi: 10.1073/pnas.1405233111 (2014).
22. Ricagno, S., Raimondi, S., Giorgetti, S., Bellotti, V. & Bolognesi, M. Human beta-2 microglobulin W60V mutant structure: Implications for stability and amyloid aggregation. *Bioch. Biophys. Res. Comm.* **380,** 543–547 (2009).
23. Santambrogio, C. *et al.* DE-loop mutations affect beta2 microglobulin stability, oligomerization, and the low-pH unfolded form. *Protein Sci* **19,** 1386–1394 (2010).
24. Ami, D. *et al.* Structure, stability, and aggregation of beta-2 microglobulin mutants: insights from a Fourier transform infrared study in solution and in the crystalline state. *Biophys J* **102,** 1676–1684, doi: 10.1016/j.bpj.2012.02.045 (2012).
25. Camilloni, C., Cavalli, A. & Vendruscolo, M. Replica-Averaged Metadynamics. *J. Chem. Theory Comput.* **9,** 5610–5617 (2013).
26. Camilloni, C., Robustelli, P., De Simone, A., Cavalli, A. & Vendruscolo, M. Characterization of the conformational equilibrium between the two major substates of RNase A using NMR chemical shifts. *J Am. Chem. Soc.* **134,** 3968–3971, doi: 10.1021/ja210951z (2012).
27. Camilloni, C. & Vendruscolo, M. Statistical mechanics of the denatured state of a protein using replica-averaged metadynamics. *J Am. Chem. Soc.* **136,** 8982–8991, doi: 10.1021/ja5027584 (2014).
28. Cavalli, A., Camilloni, C. & Vendruscolo, M. Molecular dynamics simulations with replica-averaged structural restraints generate structural ensembles according to the maximum entropy principle. *J Chem. Phys.* **138,** 094112 doi: 10.1063/1.4793625 (2013).
29. Katou, H. *et al.* The role of disulfide bond in the amyloidogenic state of beta(2)-microglobulin studied by heteronuclear NMR. *Protein Sci.* **11,** 2218–2229, doi: 10.1110/ps.0213202 (2002).
30. Gumral, D. *et al.* Reduction of conformational mobility and aggregation in W60G beta2-microglobulin: assessment by 15N NMR relaxation. *Magn. Res. Chem.: MRC* **51,** 795–807, doi: 10.1002/mrc.4018 (2013).
31. Piana, S. & Laio, A. A bias-exchange approach to protein folding. *J Phys. Chem. B* **111,** 4553–4559, doi: 10.1021/jp067873l (2007).
32. Sormanni, P., Aprile, F. A. & Vendruscolo, M. The CamSol method of rational design of protein mutants with enhanced solubility. *J Mol. Biol.* **427,** 478–490, doi: 10.1016/j.jmb.2014.09.026 (2015).
33. Agostini, F., Vendruscolo, M. & Tartaglia, G. G. Sequence-based prediction of protein solubility. *J Mol. Biol.* **421,** 237–241, doi: 10.1016/j.jmb.2011.12.005 (2012).
34. Esposito, G. *et al.* Removal of the N-terminal hexapeptide from human beta2-microglobulin facilitates protein aggregation and fibril formation. *Protein Sci* **9,** 831–845 (2000).
35. Raimondi, S. *et al.* The two tryptophans of beta2-microglobulin have distinct roles in function and folding and might represent two independent responses to evolutionary pressure. *BMC Evol Biol* **11,** 159, doi: 10.1186/1471-2148-11-159 (2011).
36. Platt, G. W., Routledge, K. E., Homans, S. W. & Radford, S. E. Fibril growth kinetics reveal a region of beta2-microglobulin important for nucleation and elongation of aggregation. *J Mol. Biol.* **378,** 251–263 (2008).
37. de Rosa, M. *et al.* Edge strand engineering prevents native-like aggregation in Sulfolobus solfataricus acylphosphatase. *FEBS J*, doi: 10.1111/febs.12861 (2014).
38. de Rosa, M. *et al.* Decoding the Structural Bases of D76N ss2-Microglobulin High Amyloidogenicity through Crystallography and Asn-Scan Mutagenesis. *Plos One* **10,** e0144061, doi: 10.1371/journal.pone.0144061 (2015).
39. Azinas, S. *et al.* D-strand perturbation and amyloid propensity in beta-2 microglobulin. *FEBS J* **278,** 2349–2358, doi: 10.1111/j.1742-4658.2011.08157.x (2011).
40. Colombo, M., de Rosa, M., Bellotti, V., Ricagno, S. & Bolognesi, M. A recurrent D-strand association interface is observed in beta-2 microglobulin oligomers. *FEBS J* **279,** 1131–1143, doi: 10.1111/j.1742-4658.2012.08510.x (2012).
41. Eakin, C. M., Berman, A. J. & Miranker, A. D. A native to amyloidogenic transition regulated by a backbone trigger. *Nat Struct Mol Biol* **13,** 202–208 (2006).
42. Verdone, G. *et al.* The solution structure of human beta2-microglobulin reveals the prodromes of its amyloid transition. *Protein Sci* **11,** 487–499 (2002).
43. Natalello, A. *et al.* Wild type beta-2 microglobulin and DE loop mutants display a common fibrillar architecture. *Plos One* **10,** e0122449, doi: 10.1371/journal.pone.0122449 (2015).
44. Dobson, C. M. Protein misfolding, evolution and disease. *Trends Bioch. Sci.* **24,** 329–332 (1999).
45. Tyedmers, J., Mogk, A. & Bukau, B. Cellular strategies for controlling protein aggregation. *Nat Rev Mol Cell Biol* **11,** 777–788, doi: 10.1038/nrm2993 (2010).
46. Vendruscolo, M. Proteome folding and aggregation. *Curr Opin Struct Biol* **22,** 138–143, doi: 10.1016/j.sbi.2012.01.005 (2012).
47. Richardson, J. S. & Richardson, D. C. Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation. *Proc. Natl Acad. Sci. USA* **99,** 2754–2759 (2002).
48. Neudecker, P. *et al.* Structure of an intermediate state in protein folding and aggregation. *Science* **336,** 362–366, doi: 10.1126/science.1214203 (2012).
49. Soldi, G., Bemporad, F. & Chiti, F. The degree of structural protection at the edge beta-strands determines the pathway of amyloid formation in globular proteins. *J Am. Chem. Soc.* **130,** 4295–4302, doi: 10.1021/ja076628s (2008).
50. Pechmann, S., Levy, E. D., Tartaglia, G. G. & Vendruscolo, M. Physicochemical principles that regulate the competition between functional and dysfunctional association of proteins. *Proc. Natl Acad. Sci. USA* **106,** 10159–10164, doi: 10.1073/pnas.0812414106 (2009).
51. Wishart, D. S. *et al.* 1H, 13C and 15N chemical shift referencing in biomolecular NMR. *J Biomol. NMR* **6,** 135–140 (1995).
52. Sugita, Y. & Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **314,** 141–151 (1999).
53. Leslie, A. G. W. Recent changes to the MOSFLM package for processing film and image plate data. Joint CCP4+ ESF-EACMB Newsletter on Protein Crystallography (1992).
54. Evans, P. Scaling and assessment of data quality. *Acta Cryst.* **62,** 72–82, doi: 10.1107/S0907444905036693 (2006).
55. CCP4. The CCP4 suite: programs for protein crystallography. *Acta Cryst.* **50,** 760–763 (1994).
56. Vagin, A. A. & Teplyakov, A. MOLREP: an automated program for molecular replacement. *J App. Crystallogr.* **30,** 1022–1025 (1997).
57. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst.* **66,** 213–221 doi: 10.1107/S0907444909052925 (2010).
58. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Cryst.* **60,** 2126–2132 (2004).

Acknowledgements
ESRF is kindly acknowledged for beamtime. This work was supported by the Italian Ministry of University and Research Project FIRB RBFR109EOS.

Author Contributions
C.C., B.M.S., P.S., R.P., A.C., M.d.R., S.Z. and A.B. performed the experiments; C.C., A.C., M.B., M.V. and S.R. designed the study; C.C., G.E., M.B., V.B., M.V. and S.R. wrote the paper.

Additional Information
Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Camilloni, C. *et al.* Rational design of mutations that change the aggregation rate of a protein while maintaining its native structure and stability. *Sci. Rep.* **6**, 25559; doi: 10.1038/srep25559 (2016).

3.2. Manuscript I

New insights in protein stability: a crystalline point of view

Sala, B.M., Natalello, A., Ricagno, S. (2017).

New insights in protein stability: a crystalline point of view

Benedetta Maria Sala[1], Antonino Natalello[2], Stefano Ricagno[1]*

[1] *Department of Bioscience, Università degli Studi di Milano, Milano 20133, Italy;*

[2] *Department of Biotechnology and Biosciences, University of Milano-Bicocca, 20126 Milan, Italy;*

*Corresponding author: Stefano Ricagno, stefano.ricagno@unimi.it.

Abstract

β2-Microglobulin (β2m) is the 99 residues light chain of Major Histocompatibility Class I complex (MHC-I). Wild type (WT) β2m is known to aggregate as amyloid fibrils Dialysis Related Amyloidosis (DRA) while the D76N destabilizing mutation is linked to a genetic systemic amyloidosis.

β2m has been object of many mutational studies to investigate the molecular bases of its aggregation propensity. These analyses have revealed that Trp60, among other residues, plays a crucial role for β2m stability. In particular, W60G β2m mutant, has an increased conformational stability and a reduced amyloidogenicity compared to WT β2m.

In this work, we seek a better understanding of the relationship between protein stability (usually measured in solution) and the interactions present in a protein crystal. To such aim, the set of these three β2m variants, WT, W60G and D76N, is an ideal system given their differences in stability and aggregation properties but their identical crystal packing.

The changes in secondary structure content have been monitored at increasing temperature on protein crystals compared to the protein in solution by Fourier transform infrared spectroscopy (FT-IR). Denaturation of proteins in the crystalline state is observed at much higher temperatures compared to solution; however, the differences in protein stability between the three variants are evident in crystalline form and they agree with the trend observed in solution.

This work provides further insights in the biophysical properties of protein molecules in solid crystalline form.

Introduction

Human β2-Microglobulin (β2m), the light chain of Major Histocompatibility Class I complex (MHC-I), is a 99 residues protein with an immunoglobulin-like fold. It is composed by seven β-strands that arrange in two β-sheets organized in the classic β-sandwich structure stabilized by an intramolecular disulphide bond (*1*).

WT β2m is the causing agent of an amyloidosis affecting patients undergoing to long-term haemodialysis, condition known as dialysis-related amyloidosis (DRA) (*2*). Moreover, its only known genetic variant D76N is responsible for a familial systemic amyloidosis (*3*). These two kinds of amyloidosis are completely unrelated and differ for deposits tissue localization (*4*). In DRA, WT β2m aggregates selectively in bones and ligaments due to the high concentration of protein in the plasma (*5*). The genetic-type of amyloidosis, instead, is not triggered by any increase in β2m concentration; the amyloid fibres deposit in visceral organs, particularly in liver, kidney, spleen, and heart (*3*).

WT and D76N β2m displays different properties *in vitro*: the D76N mutant presents a strongly decreased protein stability, higher propensity to aggregation, besides forming fibrils even under native conditions (*6, 7*).

β2m has been object of many mutational studies to investigate the molecular bases of its aggregation propensity. Particular attention has been focused on D-E loop (*8, 9*). This loop is involved in the binding with the heavy chain for MHC-I formation, and becomes solvent exposed once β2m is released from the complex (*10*). The dynamic properties of this region have been associated to the strained conformation in the WT β2m backbone structure. The replacement of Trp60 with a Gly residues (W60G) is associated to a structural relaxation of the DE-loop, resulting in an increased protein stability and a decreased amyloidogenicity of W60G variant (*9, 11, 12*).

Despite the β2m variants above described (WT, D76N and W60G) display quite different stabilities and aggregation propensities, crystallographic analyses of such mutants showed strong overall conservation of the native WT β2m structure (*Figure S1*) (*3, 13*).

Indeed, the structure determined by X-ray diffraction on protein crystals is an average of contributions from all the different conformational states of the molecule during the data collection time. Thus, X-ray diffraction experiments provide only limited information on protein dynamics (*14*) (*i.e.* poor defined electron density in flexible and/or disordered regions (*15*)), and may be affected by crystal packing interactions.

Consequently, protein crystal structures are regarded as rigid and molecules are caged in the crystal lattice, leading to the general view that little can be learnt about stability and protein dynamics from proteins in the crystalline form. Nevertheless, it should be possible to derive biophysical properties from protein crystals, given the fact that proteins retains their native structure and thus their specific intra-molecular interactions (*16, 17*).

Protein crystallization is determined by specific chemical attraction between macromolecules that result in the onset of intermolecular (ionic) bonds within the crystal. Bond strengths may be estimated by assigning binding energy to hydrogen bond or other ionic contacts, water-mediated hydrogen bonds, and other van der Waals interactions (*18*).

Recent advances in high-resolution solid-state NMR (ssNMR) illustrate the potential for detailed, and site-specific dynamic studies of microcrystalline proteins under time scales not observable by other techniques. SsNMR also opens to the possibility of detecting structural heterogeneity in sample as the presence of slightly different conformations. Interestingly, this structural polymorphism observed in ssNMR spectra matches with regions that showed dynamics in solution NMR experiments on different timescales (*19-21*). To date, concerns remain on whether protein stability and dynamics assessed in crystals may be representative of what is observed in solution and ultimately *in vivo*.

The present work aims to address this issue using the above-described WT, D76N and W60G β2m variants. Such proteins offer crucial advantages: their biophysical properties have been thoroughly characterised; furthermore these three β2m variants crystallise under the same conditions and have an identical crystal packing (*3, 22*). Thus, we expect that the results we obtained should be free of spurious crystallisation artefacts, while reflecting changes on protein stability due to the different mutations.

Therefore, we followed through Fourier transform infrared (FT-IR) spectroscopy the changes in secondary structure of these proteins in solution and as crystals. FT-IR spectroscopy has been widely adopted for the analysis of protein samples in different conditions, resulting well suitable to compare the secondary structures of proteins both in solution and in crystals (*23-26*). Moreover, a recently published work by our group, gave new insights into the correlation between the structural changes induced by single mutations and β2m conformational stability and aggregation propensity through FT-IR analysis and protocols how to acquire high quality spectra in solution and in single crystal were reported (*27*).

In the present work FT-IR spectra of the three β2m variants were collected in solution and in crystalline form. Temperature ramps were carried out to monitor fold stability. Expectedly our data indicate an increased structural order in the crystals compared to the protein in solution; while temperature ramps show that the crystalline form exerts a marked stabilising effect on protein molecules compared to protein in solution. Nevertheless, the differences in protein stability between the three variants are well detectable in crystalline form and they well correlate with the trend observed in solution.

Methods

β2-Microglobulin expression and purification.

All β2m variants were expressed in BL21 (DE3) pLysS *E. coli* strain as inclusion bodies. The proteins were extracted and purified following previously published protocols (*28*).

Crystallization and sample preparation for Fourier transform infrared (FT-IR) spectroscopy.

Lyophilized WT, W60G and D76N β2m variants were solubilized in ddH2O at a concentration of 8.5 mg/ml. All mutants were crystallized with sitting drops technique at 20°C. 160 µl of protein solution was mixed in a bridge with the same amount of 0.1 M MES pH 6, 27% PEG 4K, 15% glycerol, and placed against 0.1 M MES pH 6, 30% PEG 4K, 15% glycerol as reservoir solution. In few days, needle-like crystals were grown. Before FT-IR measurement the crystallization drop was collected and centrifuged at 2000 rcf for two minutes. Once removed the supernatant, the crystal pellet was washed and centrifuged three times with 0.1 M MES pH 6, 27% PEG 4K, 15% glycerol, all component solubilized in deuterated water to achieve the full D to H exchange of all exchangeable H. Crystals were then incubated overnight in 350 µl of the same deuterated buffer and washed once again before measurement.

FT-IR of β2m crystals.

For the IR analysis, crystals were resuspended in 20 µl of buffer (0.1 M MES pH 6, 27% PEG 4K, 15% glycerol) and then transferred in a temperature-controlled transmission cell with two BaF_2 windows separated by a 150 µm Teflon spacer.

FTIR spectra were collected at 25°C, in transmission mode, using a Varian 670-IR spectrometer (Varian Australia, Mulgrave, Australia), equipped with a mercury cadmium telluride detector nitrogen-cooled, under accurate dry air purging. Spectra were collected under the following conditions: 2 cm^{-1} resolution, 25 kHz scan speed, 1000 scan coaddition, and triangular apodization.

Thermal stability experiments, were carried out heating the sample from 25 °C to 100 °C at a rate of 0.4 °C/min and 1 °C/min, collecting each transmission spectrum every 3.45 °C and 4.35 °C respectively. The FT-IR spectra were obtained after subtraction of solvent absorption, collected under the same conditions. Second derivative analysis (*29*) was performed by the Resolutions-Pro software (Varian Australia, Mulgrave, Australia) after Savitzky Golay smoothing of the measured spectra.

FT-IR of β2m in solution.

WT, W60G and D76N β2m variant were dissolved in deuterated buffer 50 mM Sodium Phosphate pH 7.4, 100 mM NaCl, at a final concentration of 2.5 mg/ml. 20 µl of the sample were transferred in a

temperature-controlled transmission cell with two BaF_2 windows separated by a 150 μm Teflon spacer. Spectra and thermal stability experiment were collected at a rate of 0.4 °C/min following the conditions described above for the crystalline pellet and previously reported for β2m variants in solution (*27*).

Results

FT-IR spectra of WT β2m and mutated variants in solution and in crystalline state.

The secondary structure of crystalline WT β2m was analysed by FT-IR spectroscopy in its crystallisation condition and the resulted spectrum was compared to the spectrum recorded in phosphate buffer solution.

All in solution FT-IR experiments here presented have been firstly performed on WT β2m and on the other variants (see below) in the crystallisation solution to better compare the resulting data with the ones obtained from the crystalline state. However, under such conditions proteins are unstable and tend to precipitate leading to low quality non-reproducible measurements (see *Supporting material* and *Figures S3*). Thus, from now, experiments carried out on proteins in solution will refer to deuterated phosphate buffer solution as described in Methods.

Figure 1 A shows the second derivative of the FT-IR absorption spectra of the samples. The second derivative of the spectrum of crystalline WT β2m display sharper peaks suggesting more rigid protein molecules compared to what observed in the solution spectrum below; this is in agreement with previous FT-IR studies on single WT β2m crystals (*27*). The peak related to the native antiparallel β-sheets, as expected, is downshifted from ~1636 cm^{-1} in solution to ~1628 cm^{-1}, reflecting a stronger hydrogen bonding between the β-sheets of the protein related to the crystal packing (*26, 27*).

Due to the increased structural order of the crystalline state, new and better-resolved bands are detectable in the crystalline sample. In fact, the turn absorption band at ~1670 cm^{-1} (*30*) is narrowed in WT β2m crystalline sample compared to solution. Moreover, two closed components at ~1689 and the new at ~1677 cm^{-1} have been assigned to intramolecular β-sheet structures (*30*) and the new band at ~1649 cm^{-1} has been related to loop regions.

Spectra of the D76N and W60G variants were also performed both in solution and in crystalline state and compared to the analyses carried out on WT β2m, as shown respectively in *Figure* 1B and 1C. As mentioned above for the WT protein, the same relevant differences are notable between the spectra measured in solution and in crystal. Better resolved components are well detectable together with the downshift of the main native β-sheets band.

Importantly, neither the comparison of spectra taken from crystalline samples nor the one from samples in solution detect significant differences between the three variants.

Temperature ramps of the WT, D76N and W60G variants: protein unfolding.

Temperature ramp of WT, W60G and D76N β2m variants in solution has been carried out at a heating rate of 0.4 °C/min. The loss of native β-sheet structure is followed by protein aggregation, as

demonstrated from the appearance of new bands at ~1619 and 1684 cm^{-1}, related to the formation of intermolecular β-sheet structures (*Figure 4A and S2 A*).

The change in native secondary structure of WT, D76N and W60G crystallized variants were analysed increasing the temperature at a rate of 0.4 °C/min and 1 °C/min. FT-IR spectra on crystals were recorded in the 25-100 °C temperature range (*Figure 2 B, C* and *Figure S2 B*, *C)*.

In all crystalline samples the progressive protein aggregation following the unfolding process is reduced compared to the protein in solution. Thus, the bands at ~1619 and 1684 cm^{-1} have a minor intensity at higher temperature in their spectra while the random coil component (peak around 1648 cm^{-1}) is stronger, suggesting an increased unfolded population in the sample. Moreover, this trend is better noticeable increasing the temperature of the sample at 1 °C/min rate (Figure 4B and S2B).

In addition, WT and D76N crystals display differences compared to the W60G variant: in *Figure* S2B, in WT and D76N, the magenta spectra have a higher intensity than in red (at 100 °C) indicating a partial melting of the formed aggregates at higher temperatures.

Different rates result in non-negligible differences in the measured thermal stabilities (*Figure 3A* and *Table 1*). Thus, different heating rates result in an apparent change of the system stability, in other words the stability observed for all crystalline samples seems to have a relevant kinetic component. Crucially, at both heating rates the trend in fold stability of the three variants is in keeping with the one previously reported by experiments in solution (W60G > WT > D76N) (*3, 11*) (*Table 1*).

We then compare the results recorded on crystalline material with those on solubilized proteins taking in consideration the different components described above in the previous paragraph. The temperature dependence of the native β-sheet peak intensity of the three variants in crystal and in solution are shown in *Figure 3*, while *Table 1* reports the calculated melting temperature (T_m) extrapolated from the unfolding curves.

It is noticeable an increasing in stability of the crystallized samples respect to the solubilized ones. This behaviour is in accordance with the downshift of the intramolecular β-structure component of the protein in the crystalline state thus linkable to the more order and tight structure of the protein under these conditions. Interestingly, the differences in stability among the three variants previously determined by CD experiments (*3, 11, 13, 27*) are visible in all conditions monitored by FT-IR: W60G β2m is the most stable variant followed in order by WT and D76N.

Temperature ramps of the WT, D76N and W60G variants: crystal melting

Further comparative analysis can be performed on the different events occurring during protein unfolding in the solution and crystalline state by monitoring the tyrosine band at ~1514 cm^{-1} (*Figure 4*). This component reflects the environment of Tyr residues and displays minor changes in case of protein unfolding in solution (*Figure 4 A*). Conversely, in the crystalline samples the peak position corresponding to the maximum absorption by Tyr residues is shifting to higher wavenumbers. Thus, the

temperature-dependence position of this peak can be plotted, resulting in an unfolding curve (*Figure 4 B*) well reflecting the thermal stability of the variants. The change in Tyr absorption can be related to the crystal melting: the Tyr residues embedded into the crystal "are forced" to interact with the surrounding residues while, when crystals are disintegrating releasing protein molecules, the environment around the Tyr residues undergoes a drastic change.

Another effect of the temperature, remarkable only on the crystalline samples, is the variation of the area of the absorption spectrum in the amide I region (data not shown). This area does not change significantly in the spectra recorded on the proteins in solution. Conversely, in the crystalline samples, as the temperature starts destabilizing the crystal, this band becomes more intense. This effect may provide a measure of how restrained the protein molecules are: indeed, they experience a highly constrained environment in the crystal lattice while once the crystal has melted unfolded protein molecules are completely free in solution.

Discussion

The present work aims to understand whether biophysical properties assessed *in crystallo* are representative of protein behaviour in solution. We thus take as model system three β2m variants WT, W60G and D76N, to carry out a comparative study of protein stability in solution and in the crystalline form given their equal crystal properties and their differences in stability and aggregation propensity. Thermal unfolding analyses of the proteins in solutions and in crystal have been performed following change in protein secondary structure by FT-IR spectroscopy.

As FT-IR spectra on single crystals (*27*), spectra recorded on the crystalline samples report more information compare with the spectra recorded in solution, displaying better resolved components, suggesting that crystals are ideal samples to acquire FT-IR spectra of excellent quality.

The thermal stabilities, here reported for WT, D76N and W60G β2m variants, are generally in good agreement with previous data (*3, 22*) confirming the hypothesis that information about fold stability can be assessed also from proteins in crystalline form.

Additionally, in the spectra recorded on crystalline samples at increasing temperature (figure S2) is evident the reduction of the fraction of protein that aggregates. This event can be explained by the fact that, if soluble, the protein that unfolds at lower temperature starts to aggregate progressively. Conversely, in the crystalline samples, the late unfolding of the protein delays the aggregation process, resulting in a bigger population of unfolded protein at the end of the ramp.

Moreover, FT-IR spectra collected at increasing temperature on crystalline samples bear information beyond protein unfolding: the temperature-dependent variation in position of the Tyr peak allows monitoring crystal melting.

Therefore, the FT-IR experiments here reported point at the conservation of fold stability of proteins in crystalline state setting the base for future experiments comparing molecular dynamics of proteins in solution and in crystalline state. Such data together will be required for an exhaustive characterisation of the biophysical properties of proteins in the crystalline form.

References

1. P. J. Bjorkman *et al.*, Structure of the human class I histocompatibility antigen, HLA-A2. *Nature* **329**, 506-512 (1987).
2. F. Gejyo *et al.*, A new form of amyloid protein associated with chronic hemodialysis was identified as beta 2-microglobulin. *Biochem Biophys Res Commun* **129**, 701-706 (1985).
3. S. Valleix *et al.*, Hereditary systemic amyloidosis due to Asp76Asn variant beta2-microglobulin. *N Engl J Med* **366**, 2276-2283 (2012).
4. M. Stoppini, V. Bellotti, Systemic amyloidosis: lessons from beta2-microglobulin. *J Biol Chem* **290**, 9951-9958 (2015).
5. F. Danesh, L. T. Ho, Dialysis-related amyloidosis: history and clinical manifestations. *Semin Dial* **14**, 80-85 (2001).
6. P. P. Mangione *et al.*, Structure, folding dynamics, and amyloidogenesis of D76N beta2-microglobulin: roles of shear flow, hydrophobic surfaces, and alpha-crystallin. *J Biol Chem* **288**, 30917-30930 (2013).
7. M. de Rosa *et al.*, Decoding the Structural Bases of D76N ss2-Microglobulin High Amyloidogenicity through Crystallography and Asn-Scan Mutagenesis. *PLoS One* **10**, e0144061 (2015).
8. M. Colombo *et al.*, The effects of an ideal beta-turn on beta-2 microglobulin fold stability. *J Biochem* **150**, 39-47 (2011).
9. S. Ricagno *et al.*, DE loop mutations affect beta2-microglobulin stability and amyloid aggregation. *Biochem Biophys Res Commun* **377**, 146-150 (2008).
10. J. P. Hodkinson, T. R. Jahn, S. E. Radford, A. E. Ashcroft, HDX-ESI-MS reveals enhanced conformational dynamics of the amyloidogenic protein beta(2)-microglobulin upon release from the MHC-1. *J Am Soc Mass Spectrom* **20**, 278-286 (2009).
11. C. Camilloni *et al.*, Rational design of mutations that change the aggregation rate of a protein while maintaining its native structure and stability. *Sci Rep* **6**, 25559 (2016).
12. D. Gumral *et al.*, Reduction of conformational mobility and aggregation in W60G beta2-microglobulin: assessment by 15N NMR relaxation. *Magn Reson Chem* **51**, 795-807 (2013).
13. C. Santambrogio *et al.*, DE-loop mutations affect beta2 microglobulin stability, oligomerization, and the low-pH unfolded form. *Protein Sci* **19**, 1386-1394 (2010).
14. D. Ringe, G. A. Petsko, Study of protein dynamics by X-ray diffraction. *Methods Enzymol* **131**, 389-433 (1986).
15. K. Djinovic-Carugo, O. Carugo, Missing strings of residues in protein crystal structures. *Intrinsically Disord Proteins* **3**, e1095697 (2015).
16. A. A. Chernov, Protein crystals and their growth. *J Struct Biol* **142**, 3-21 (2003).
17. C. M. Vander Zanden, M. Carter, P. S. Ho, Determining thermodynamic properties of molecular interactions from single crystal studies. *Methods* **64**, 12-18 (2013).
18. Y. Matsuura, A. A. Chernov, Morphology and the strength of intermolecular contacts in protein crystals. *Acta Crystallogr D Biol Crystallogr* **59**, 1347-1356 (2003).
19. S. G. Zech, A. J. Wand, A. E. McDermott, Protein structure determination by high-resolution solid-state NMR spectroscopy: application to microcrystalline ubiquitin. *J Am Chem Soc* **127**, 8618-8626 (2005).
20. J. L. Lorieau, A. E. McDermott, Conformational flexibility of a microcrystalline globular protein: order parameters by solid-state NMR spectroscopy. *J Am Chem Soc* **128**, 11505-11512 (2006).
21. H. K. Fasshuber *et al.*, Structural heterogeneity in microcrystalline ubiquitin studied by solid-state NMR. *Protein Sci* **24**, 592-598 (2015).
22. G. Esposito *et al.*, The controlling roles of Trp60 and Trp95 in beta2-microglobulin function, folding and amyloid aggregation properties. *J Mol Biol* **378**, 887-897 (2008).
23. J. M. Hadden, D. Chapman, D. C. Lee, A comparison of infrared spectra of proteins in solution and crystalline forms. *Biochim Biophys Acta* **1248**, 115-122 (1995).
24. J. T. Sage *et al.*, Infrared protein crystallography. *Biochim Biophys Acta* **1814**, 760-777 (2011).

25. K. L. Chan, L. Govada, R. M. Bill, N. E. Chayen, S. G. Kazarian, Attenuated total reflection-FT-IR spectroscopic imaging of protein crystallization. *Anal Chem* **81**, 3769-3775 (2009).
26. A. Barth, C. Zscherp, What vibrations tell us about proteins. *Q Rev Biophys* **35**, 369-430 (2002).
27. D. Ami *et al.*, Structure, stability, and aggregation of beta-2 microglobulin mutants: insights from a Fourier transform infrared study in solution and in the crystalline state. *Biophys J* **102**, 1676-1684 (2012).
28. G. Esposito *et al.*, Removal of the N-terminal hexapeptide from human beta2-microglobulin facilitates protein aggregation and fibril formation. *Protein Sci* **9**, 831-845 (2000).
29. H. Susi, D. M. Byler, Resolution-enhanced Fourier transform infrared spectroscopy of enzymes. *Methods Enzymol* **130**, 290-311 (1986).
30. J. L. Arrondo, F. M. Goni, Structure and dynamics of membrane proteins as studied by infrared spectroscopy. *Prog Biophys Mol Biol* **72**, 367-405 (1999).

Figures

Figure 1. **FT-IR spectroscopy of WT, D76N and W60G β2m variants. (A)** The second derivative of the absorption spectrum of WT β2m in solution compared with that of WT protein in the crystalline state. **(B)** Comparison of the second derivatives of the absorption spectra of WT, D76N and W60G β2m in phosphate solution. **(C)** Comparison of the second derivatives of the absorption spectra of crystallized WT, D76N and W60G β2m. The peak position of the main Amide I components is reported.

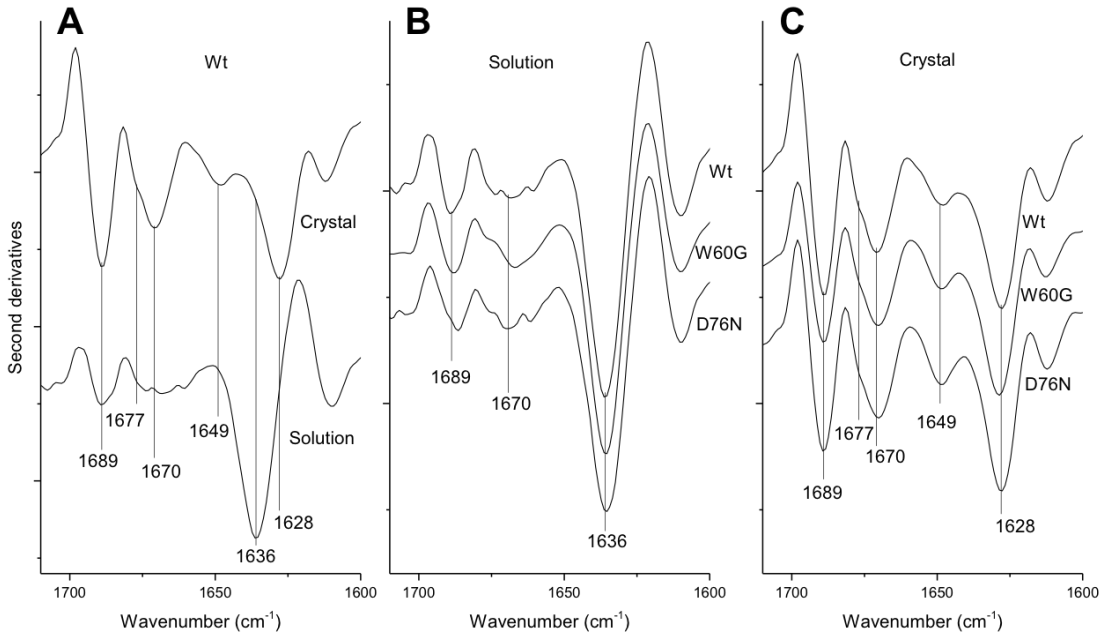


Figure 2. **Second derivatives of the absorption spectra in the amide I region of WT, D76N and W60G β2m variants. (A)** The second derivatives of the absorption spectra of WT, D76N and W60G β2m variants in phosphate buffer, reported in the 1700 – 1590 cm^{-1} region. **(B)** The second derivatives of the absorption spectra of WT, D76N and W60G β2m crystallized variants heated at a rate of 1 °C/min, reported in the 1700 – 1590 cm^{-1} region. **(C)** The second derivative of the absorption spectra of WT, D76N and W60G β2m crystallized variants heated at a rate of 0.4 °C/min, reported in the 1700 – 1590 cm^{-1} region.

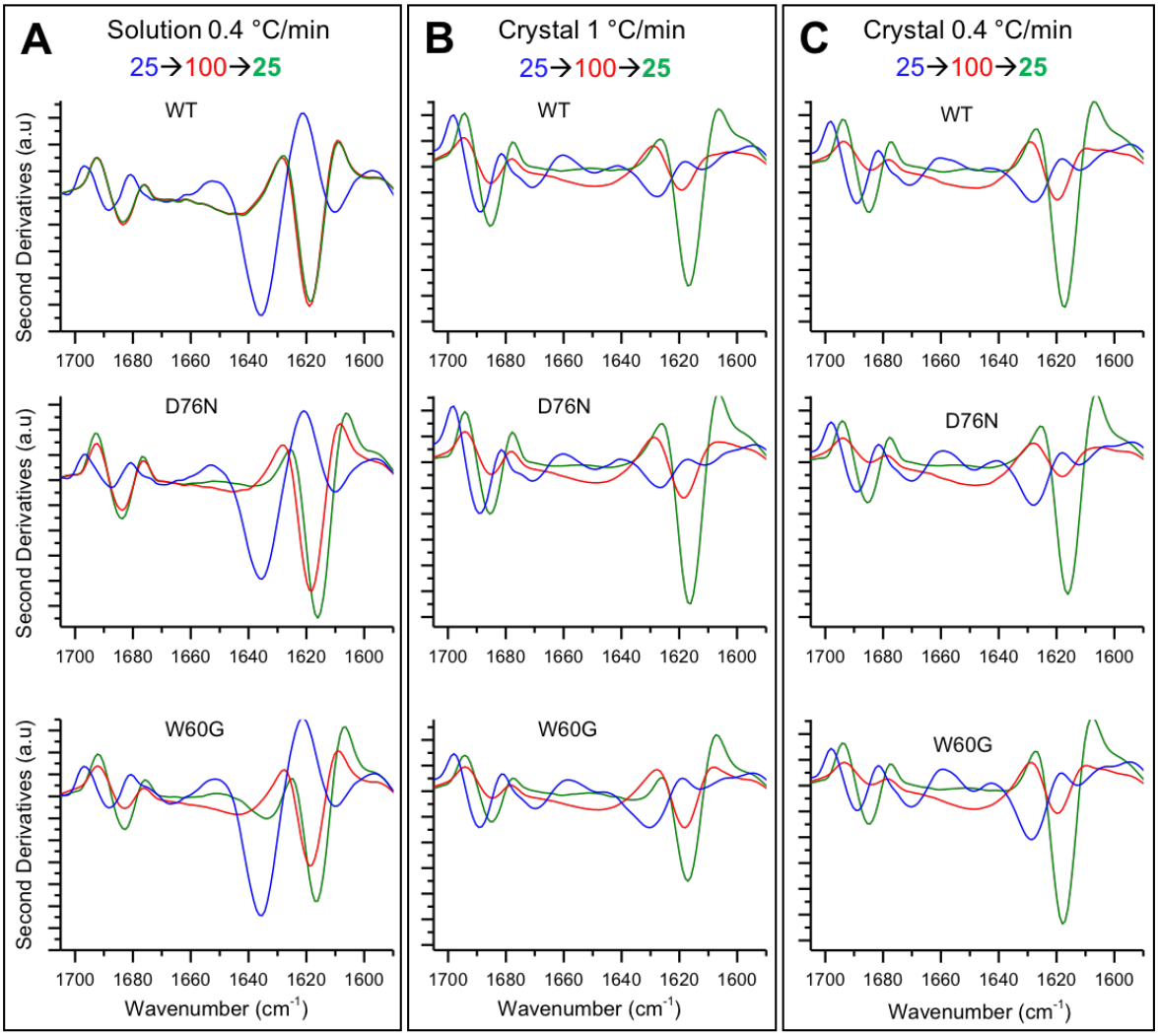


Figure 3. **Thermal stability of β2m variants assessed by FT-IR spectroscopy. (A)** The temperature dependence of the native β-sheet peak intensity at ~1628 cm^{-1}, taken from second derivative spectra of β2m analysed crystallized variants heated at a rate of 0.4 °C/min and 1 °C/min. **(B)** The temperature dependence of the native β-sheet peak intensity at ~1636 cm^{-1}, taken from second derivative spectra of β2m analysed variants in phosphate solution at a rate of 0.4 °C/min.

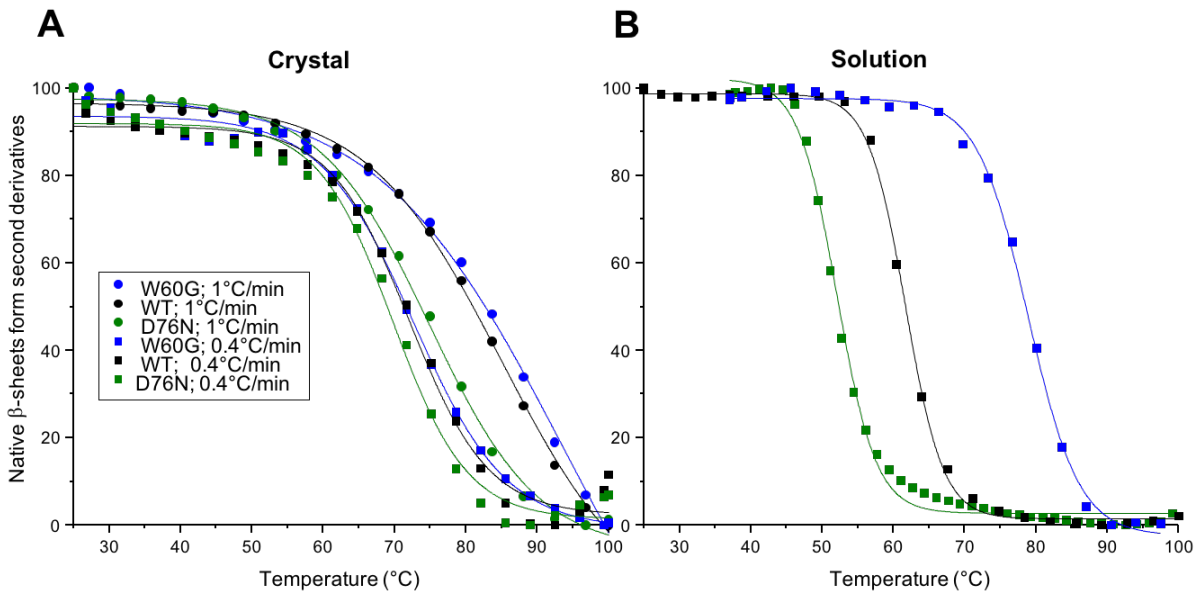
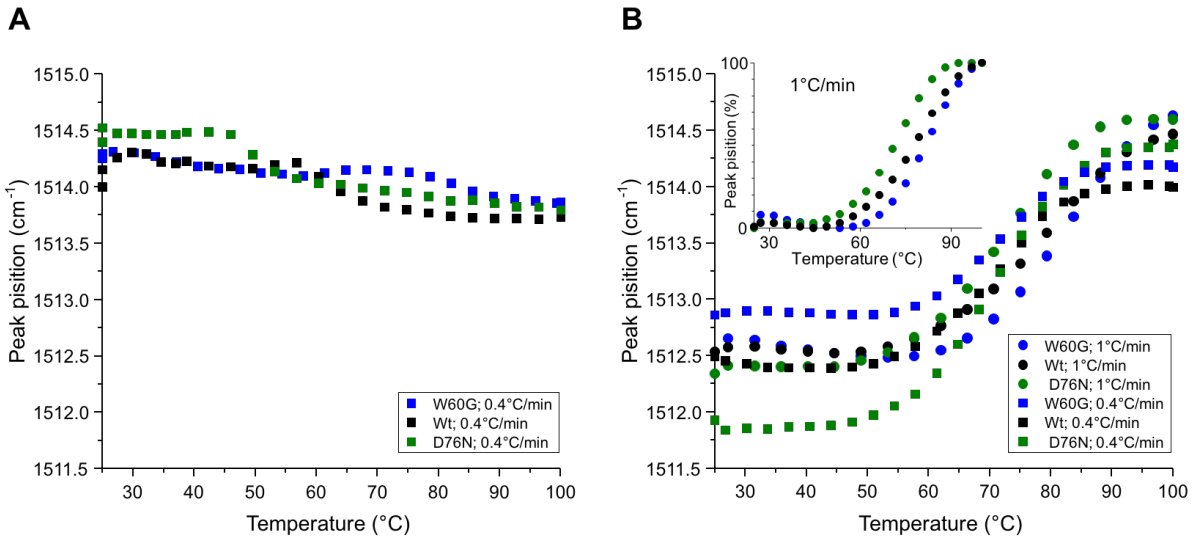


Figure 4. **Thermal stability of β2m variants assessed from the temperature-dependent position of the tyrosine peak. (A)** The temperature dependence of the tyrosine peak position taken from the second derivative spectra of β2m analysed variants in phosphate solution. **(B)** The temperature dependence of the tyrosine peak position taken from the second derivative spectra of β2m analysed crystallized variants heated at a rate of 0.4 °C/min and 1 °C/min. Inset: Curves recorded at a rate of 1 °C/min have been normalized to highlight the trends.



Tables

Table 1. **Calculated melting temperature for the complexes.** T_m values are assessed on native β-sheet peak intensity measured on the protein in phosphate solution, in crystalline state at a heating rate of 0.4 °C/min and 1 °C/min, and on far UV CD curves (*3, 11*). Values are calculated fitting the data with the Boltzann function.

Protein	**Tm / °C (solution FT-IR), PBS**	**Tm / °C (crystal FT-IR, 0.4 °C/min)**	**Tm / °C (crystal FT-IR, 1 °C/min)**	**Tm / °C (CD)**
WT	61.6	71.3	85.0	62.4
D76N	52.1	70.2	74.8	53.2
W60G	78.9	81.1	93.6	69.8

Supporting Material

New insights in protein stability: a crystalline point of view

Benedetta Maria Sala[1], Antonino Natalello[2], Stefano Ricagno[1]*

[1] *Department of Bioscience, Università degli Studi di Milano, Milano 20133, Italy;*

[2] *Department of Biotechnology and Biosciences, University of Milano-Bicocca, 20126 Milan, Italy*

*Corresponding author: Stefano Ricagno, stefano.ricagno@unimi.it.

Contents

FT-IR experiments with WT, D76N and W60G β2m variants solubilized in the crystallization buffer. 1 mg of lyophilized protein for each β2m variant was dissolved in 2 ml of deuterated buffer 0.1 M MES pH 6, 27% PEG 4K, 15% glycerol and let overnight at 4 °C. The samples, containing a high amount of pellet (precipitated protein), was centrifuged for 20 minutes at 10000 rpm. 20 µl of the sample were then transferred in a temperature-controlled transmission cell with two BaF_2 windows separated by a 150 µm Teflon spacer. Spectra and thermal stability experiment were collected following the conditions described in Materials and Methods paragraph for the crystalline pellet.

The thermal stability of β2m variants assessed by FT-IR spectroscopy on protein solubilized in crystallization buffer are shown in *Figure S3*.

Figure S1. **Crystal structure superposition for WT, D76N and W60G β2m variants.** Structure superposition of WT (green), D76N (blue) and W60G (yellow) β2m variants (pdb codes 1LDS, 4FXL, 2Z9T respectively). The main β-sheet strands are labelled according to the accepted β2m nomenclature. Asp76 and Trp60 are shown in sticks in WT β2m structure.

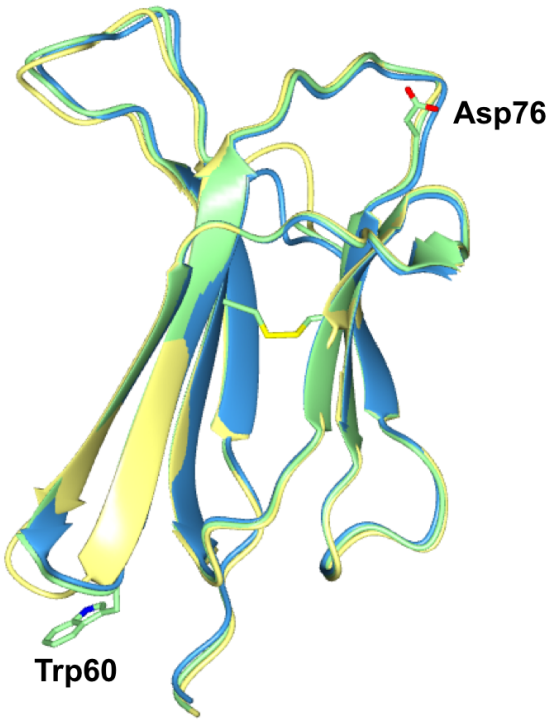


Figure S2. **Thermal stability of β2m variants assessed by FT-IR spectroscopy. (A)** The second derivative of the absorption spectra of WT, D76N and W60G β2m variants in phosphate buffer recorded at increasing temperature. **(B)** The second derivative of the absorption spectra of WT, D76N and W60G β2m crystallized variants recorded at increasing temperature (1 °C/min). **(C)** The second derivative of the absorption spectrum of WT, D76N and W60G β2m crystallized variants recorded at increasing temperature (0.4 °C/min).

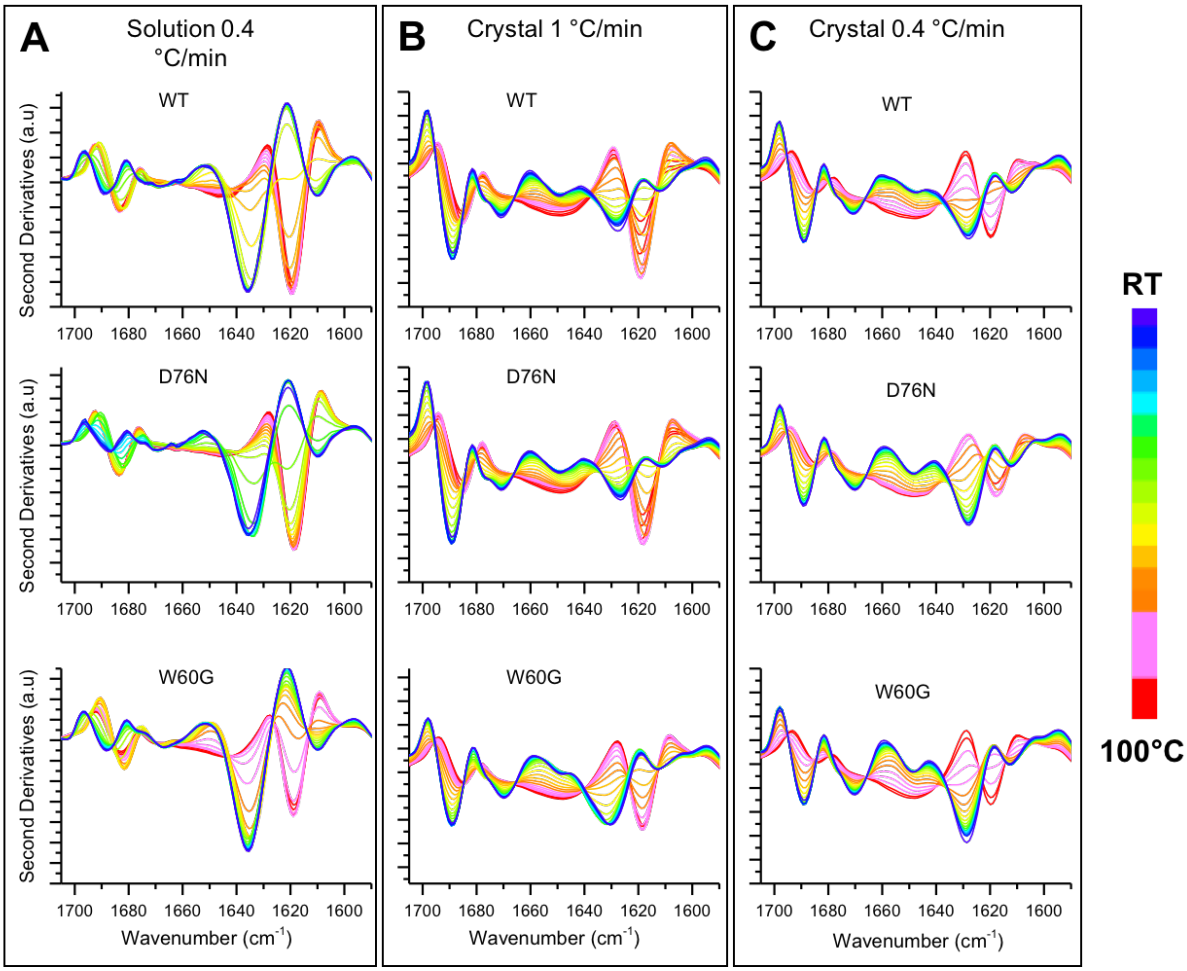
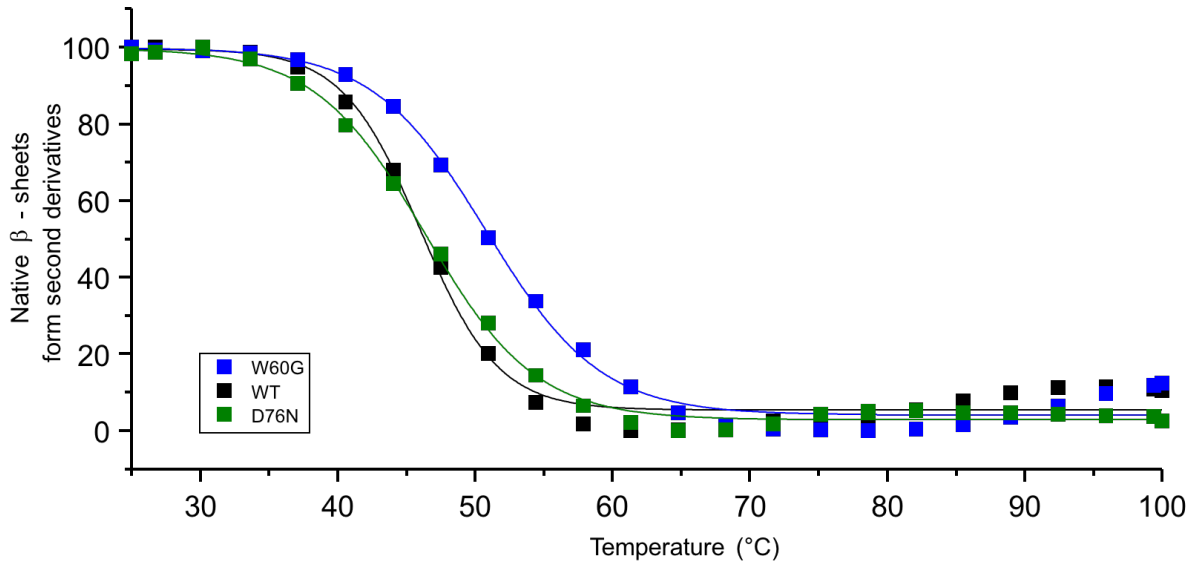


Figure S3. **Thermal stability of β2m variants assessed by FT-IR spectroscopy on protein solubilized in crystallization buffer.** The temperature dependence of the native β-sheet peak intensity at ~1636 cm^{-1}, taken from the second derivative spectra of β2m analysed variants solubilized in the crystallization solution.



3.3. Manuscript II

Novel MHC-I complex-stabilizing "superantigens" derived from the cancer target NY-ESO-1

Sala, B.M., Broggini, L., Pellegrino, S., Barbiroli, A., Achour, A., Ricagno, S. (2017).

Novel MHC-I complex-stabilizing "superantigens" derived from the cancer target NY-ESO-1

Benedetta Maria Sala[1], Luca Broggini[1], Sara Pellegrino[2], Alberto Barbiroli[3], Adnane Achour[4]*, Stefano Ricagno[1]*

[1] *Department of Bioscience, Università degli Studi di Milano, Milano 20133, Italy;*

[2] *University of Milano, Department of Pharmaceutical Sciences, Milano 20133, Italy;*

[3] *DeFENS, Dep. of Food Environmental and Nutritional sciences, University of Milan, Milano 20133, Italy;*

[4] *Science for Life Laboratory, Department of Medicine, Karolinska Institute, Stockholm, Sweden.*

*Corresponding authors: Adnane Achour, adnane.achour@ki.se; Stefano Ricagno, stefano.ricagno@unimi.it.

Abstract

Major Histocompatibility Class I complexes (MHC-I) are hetero-trimers composed by the polymorphic heavy chain, the light chain β2-microglobulin (β2m) and an antigenic peptide. It has an essential role in immune response against external agents and in cancer immune surveillance, exposing antigenic peptides on the cell surface and triggering T cells activation. It has been assessed that stability, rather than affinity, of the peptide with the complexes better correlates with immunogenicity. A promising strategy for tumor eradication may consists in inducing a strong T-cell response against tumor-associated antigens (TAA).

The antigen NY-ESO-1 is expressed in many cancer types, thus it is a valid candidate for cancer vaccine design. The studies on vaccine preparation with the antigenic peptide NY-ESO-$1_{157-165}$ have revealed problems of stability and bioavailability associated with peptide immunization and the frequent failure to activate cytotoxic T lymphocytes. In this context, here we adopt a novel generation of engineered "super-peptides" ligands (APLs), to stabilize MHC-I and increase the affinity toward T cell receptors.

Eight MHC-I bearing different NY-ESO-$1_{157-165}$-derived antigenic peptides have been produced and purified; their stability has been monitored *in vitro*; though to different degrees, all considered peptide variants increase the fold stability of MHC-I.

The X-ray crystal structure of one of the most stable MHC-I has been determined and its structure compared to those already published. Together, our results can contribute to the rationalization of the MHC-I anchor residues in tumour epitopes to enhance binding of the peptide to the complex, improving cancer vaccine design.

Introduction

The Major Histocompatibility Class I complex (MHC-I) consists in a highly polymorphic heavy chain (α) with a transmembrane domain, a monomorphic light chain (β2m) and an antigenic peptide (*1*). Presentation of peptides by MHC-I on cell surfaces results in the activation of immune responses mediated by CD8+ T lymphocytes (TCD8), including cytotoxic T lymphocytes (CTLs), and the subsequent lysis of target cells (*2, 3*).

Cancer cells express a wide range of antigens that, once proteolized and exposed by MHC-I, are recognized by CTLs. Several clinical trials have highlighted that peptides coming from tumor-associated antigens (TAA) interact weakly with MHC-I heavy chain (*4*). Thus, many efforts have been done to improve the efficacy of the peptides processed from TAA, engineering synthetic peptides to dispense to patients through peptide-base vaccines. (*5, 6*).

NY-ESO-1, a 180-amino acid long protein with unknown function, is considered one of the most promising tumor-specific antigens (*7*). It is expressed by many cancer-type cells (such as breast, prostate, ovarian cancer, as well as melanoma) and not by healthy adult somatic tissues (*8-11*). Thanks to its high immunogenicity, NY-ESO-1 is a valid candidate for design specific cancer vaccines (*12, 13*). Nevertheless, several studies have pointed out problems of stability and bioavailability associated with immunization using NY-ESO-1 derived-peptides (*14, 15*).

So far, the most studied vaccine preparation for NY-ESO-1 makes use of wild type peptides administered with various adjutants for immunization of cancer populations expressing the antigen (*16-18*).

Trial vaccinations with a mixture of NY-ESO-1 derived peptides, revealed a strong T-CD8+ reactivity and response against NY-ESO-$1_{157–165}$ bound to human MHC-I (*7, 19*). This peptide presents at the *C-terminus* a cysteine (Cys) residue that, due to cysteinylation and dimerization, affect vaccines formulation reducing their immunogenic efficacy (*20*). In fact, under physiological conditions, cysteine is the most chemically reactive amino acid due to its free sulphydryl group.

It has been demonstrated either *in vitro* and *in vivo*, that reducing agents significantly enhanced the immunogenicity of cysteine-containing peptides (*20*): MHC-I molecules bearing reduced NY-ESO-$1_{157-165}$ are, indeed, recognized ten times more efficiently (*12*).

Similar improved effects were obtained by replacing the Cys 165 with more common anchor residues generating new peptide variants (*12, 21*).

In this context, the properties of MHC-I/NY-ESO-$1_{157-165}$WT (from now MHC-I/WT) have been compared by Webb and colleagues to that of the complexes bearing the peptide with mutations C9A and C9S (from now MHC-I/9A and MHC-I/9S) (*21*). Substitution of the C-terminal Cys with an Ala do not affect complex thermal stability and CTL recognition. The mutation to Ser, instead, resulted in

a lower cytotoxic response, although keeping peptide conformation in the binding groove of the heavy chain.

Variants of NY-ESO-$1_{157\text{-}165}$ where the C-terminal Cys had been replaced by valine, isoleucine, or leucine, have been generated by Chen and colleagues (*12*). These peptides bind more efficiently to MHC-I and are recognized at least 100 times more efficiently than the wild-type peptide by specific CTL. Among these new peptides analogues, it has been demonstrated that NY-ESO-$1_{157\text{-}165}$9V (from now MHC-I/9V) can improve T-Cell Receptors (TCR) affinity compared to the others. The introduction of a Val residue, optimize the antigenic peptide binding without altering the orientation of the key residue for TCR recognition (Met4 and Trp5) (*22*).

Given the background, van Stipdonk has recently reported a novel procedure for the design of altered "super-peptides" ligands (APLs) to improve binding affinity to MHC-I (*23*). These modified peptides hold a structural conformation similar to the WT infection-derived or non-immunogenic tumor-associated peptides and act as highly immunogenic mimotopes of tumor associated antigens (TAAs) in several different cancer mouse models (*23*).

This work aims to design a novel generation of APLs on the human cancer antigen NY–ESO-1 to identify peptides that strongly stabilize MHC-I. Seven variants of NY-ESO-$1_{157\text{-}165}$ have been designed following van Stipdonk's procedure, based on the previous observations that peptides, stabilizing MHC-I *in vitro*, also increase its half-life in cell yielding to an enhanced immune response (*12*). Chen *et al.* showed that positions 1, 3 and 9 are crucial for the interaction of the peptide with the heavy chain and for the interaction between MHC-I and TCRs. Mutations on N- and C-terminal residues are suggested to be directly involved in stabilizing the complex while mutations on the third residues should affect the conformation of the peptide within the groove, which entails a different spatial organization of the complex (*12*).

Hence, in the present work one, two or three of the mutations - S1Y, L3Y, L3P, C9V - have been introduced into new NY-ESO-$1_{157\text{-}165}$ peptide variants. Eight MHC-I bearing NY-ESO-$1_{157\text{-}165}$ peptide variants (1Y3P9V, 3Y9V, 1Y9V, 1Y, 1Y3P, 9V, 3P9V and WT) have been purified. The effects caused on the complex thermal stability by a given peptide have been evaluated for all designed variants bound to MHC-I and compared to MHC-I/WT (*21*). Furthermore, the structural features, responsible for differences in stability have been examined through the X-ray structure of one of the most stable MHC-I.

Methods

Peptide Synthesis.

The eight NY-ESO-$1_{157-165}$ variant peptides (WT: SLLMWITQC, 1Y3P9V: **Y**L**P**MWITQ**V**, 3Y9V: SL**Y**MWITQ**V**, 1Y9V: **Y**LLMWITQ**V**, 1Y: **Y**LLMWITQC, 1Y3P: **Y**L**P**MWITQC, 9V: SLLMWITQ**V**, 3P9V: SL**P**MWITQ**V**) were prepared by microwave-assisted solid phase synthesis (*24*) based on Fmoc chemistry on preloaded pre-loaded Wang (0.6 eq/g loading) using a 5-fold molar excess of 0.2 M Fmoc-protected amino acids dissolved in dimethylformamide and using N-hydroxybenzotriazole/O-(benzotriazol-1-yl)-N ,N, N′, N′-tetramethyluronium hexafluorophosphate/diiso-propylethylamine (5: 5: 10) as activators. Coupling reactions were performed for 5 min at 40 watts with a maximum temperature of 75 °C. Deprotection was performed in two stages using 20% piperidine in dimethylformamide (5 and 10 min each). Cleavage was performed using 10 ml of reagent K (TFA/phenol/water/thioanisole/ethanedithiol; 82.5/5/5/5/2.5) for 180 min. Following cleavage, peptides were precipitated out and washed using ice-cold anhydrous ethyl ether. All peptides were purified by reverse phase HPLC using a gradient elution of 5–70% solvent B (solvent A: water/acetonitrile/TFA 95/5/0.1; solvent B: water/acetonitrile/TFA 5/95/0.1) over 20 min at a flow rate of 10 ml/min. The purified peptides were freeze-dried and stored at 0 °C.

MHC-I Purification.

The HLA-A*0201 (HLA-A2) heavy chain and β2-microglobulin (β2m) proteins were expressed individually as inclusion bodies using the BL21 (DE3) *E.coli* strain, following previous published protocols (*25, 26*).

The refolding of MHC-I was carry out by dilution. The peptide and β2m were added firstly to the refolding buffer (100 mM Tris pH 8, 450 mM L-Arginine, 5 mM L-Glutathione reduced, 0.5 mM L-Glutathione oxidized, 2 mM EDTA, 0.5 mM AEBSF) and the solution was left at 4°C under stirring for half an hour. The unfolded HLA-A2, solubilized in 6 M GdmHCl, 20 mM TrisHCl pH 8 was added. The final concentration of the components within the refolding buffer was 1 μm HLA-A2, 2 μm WT β2m, 10 μm peptide.

The refolding was completed after 72 hours at 4°C under stirring. The solution was concentrated trough Stirred Ultrafiltration Cell (Millipore) and Amicon Ultra-15 Centrifugal Filters (EMD Millipore) up to approximately 3 mL. The sample was then purified by size exclusion chromatography using a Superdex 75 column equilibrated with 20 mM TrisHCl pH 8, 150 mM NaCl.

The eluted protein was analyzed by SDS-PAGE, frozen in liquid nitrogen and stored at -20°C.

Circular Dichroism.

Thermal unfolding experiments were performed by circular dichroism (CD) in the far-UV region on a J-810 spectropolarimeter (JASCO Corp., Tokyo, Japan) equipped with a Peltier system for temperature control. The protein concentration was 0.2 mg/mL in 20 mM Tris-HCl, pH 8.0, 150 mM Sodium Chloride. The temperature ramp measurements were recorded from 20 to 95°C (temperature slope 60° C/hour) in a 0.1 cm path length cuvette and monitored at 218 nm wavelength. Tm was calculated as the first-derivative minimum of the traces. Far-UV CD spectra were recorded before and after unfolding ramp (260-190 nm).

Crystallization.

MHC-I/3Y9V has been concentrated up to 6 mg/ml. Crystals were grown at 20°C with sitting drop techniques by mixing equal amounts of protein solution, and the reservoir solution in two conditions: (A) 0.2 M $MgCl_2$, 0.1 M Tris-HCl pH 8.1, 31% PEG 3350 and (B) 28 % w/v Polyethylene glycol 6000, 500 mM LiCl, 100 mM Tris-HCl pH 8.5. Crystals from condition (B) were cryoprotected with with 20 % w/v Polyethylene glycol 4000 and crystals from both conditions were then flash-frozen in liquid nitrogen. X-ray diffraction data were collected at the beam line ID-30a at ESRF (European Synchrotron Radiation Facility of Grenoble, France).

Structure Determination.

The diffraction data were processed using XDS (*27*) and the intensities were merged and scaled with AIMLESS (*28*). The structure was solved by molecular replacement using PHASER (*29*) using the structure of MHC-I bearing the WT peptide as search model (PDB code: 1S9W (*21*)). The model molecules were subjected firstly to a rigid-body refinement and then to a restrained refinement using REFMAC (*30*) program suite. Manual model building, was carried out using the COOT (Crystallographic Object-Oriented Toolkit) (*31*).

Results

MHC-I refolding and purification

To evaluate effects on MHC-I stability by NY-ESO-$1_{157-165}$ variants, HLA-A0201 (HLA-A2) heavy chain has been refolded in presence of β2m and eight different peptide variants. Following conventional protocols (*26*), the refolding process resulted in efficient production of most of the MHC-I molecules, except for MHC-I bearing WT and 1Y variant peptides. Purifications yields of all MHC-I are reported in *Table1* (gel filtration chromatograms are shown in *Figure S1*).

The lower MHC-I/WT and MHC-I/1Y purification yield may be explained by an altered refolding outcome. It is notably from the far UV spectra monitored by circular dichroism (*Figure S2*), that the profile of MHC-I/WT and MHC-I/1Y have a shifted intersection of the X-axis compared to all other complexes, suggesting a slightly different secondary content.

MHC-I Stability measured by CD

Unfolding of MHC-I bearing NY-ESO-$1_{157-165}$ variants was monitored at increasing temperature by far-UV CD to investigate specific effects on complex stability due to the designed peptide mutations.

As shown in *Figure 1* the unfolding curves display a cooperative behaviour, thus confirming the presence of secondary structured proteins. Unfolding of MHC-I/WT, instead, is poorly cooperative and displays two inflection points and altered behaviour in different purification batches (WT and WT b in *Figure 1*). Due to MHC-I/WT instability, the heavy chain probably experiences denaturation prior to β2m. This could explain the presence of a first initial denaturation (the heavy chain unfolding) followed by a plateau (heavy chain aggregation) and then a second inflection point (β2m collapsing together with the formed aggregation). Likewise, also previous papers reporting thermodynamic stability measurement on MHC-I/WT revealed a very poor cooperativity of complex unfolding (*21*).

Table 2 reports the calculated melting temperature (T_m) for all complexes. The MHC-I/WT T_m has been not extrapolated from the unfolding curves, but it was considered that calculated by Webb *et al.* (*21*). T_m values of all other complexes bearing NY-ESO-$1_{157-165}$ mutated peptides, revealed an increase in stability compared to the MHC-I/WT. In particular, MHC-I/1Y, MHC-I/9V, MHC-I/1Y9V and MHC-I/3Y9V are the most promising complexes since their peptides have the greatest effects on the stability of MHC-I.

Structures of the MHC/ 3Y9V

The X-ray crystal structure of MHC-I/3Y9V, one of the most stable MHC-I, has been determined at high resolution and its structure compared to those already published MHC-I/WT and MHC-I/1Y9V. This structure allows to inspect the structural contributions to the improved stability determined by the mutation at the third peptide residue.

The complex crystallized in two different conditions showing different space groups and crystal packing: the crystal grown in condition A (MHC-I/3Y9Va) has P 21 21 21 space group and two molecules in the asymmetric unit (AU); while, the crystal grown in condition B (MHC-I/3Y9Vb) has P 21 space group and one molecule in the AU. Data collection and refinement statistics are reported in *Table 3*.

In both crystal structures the electron density map is of excellent quality for HLA-A2, β2m and the peptide, allowing the model building of the whole three chains (*Figure 2 A*). L3Y, C9V peptide mutations are clearly detectable (*Figure 2 B* and *C*). The three structures comprise residues 1–274 of the HLA-A2 heavy chain, residues 1–99 of β2m, and the nine residues of the bound peptide and a variable number of water molecules.

Overall, the quaternary structure of MHC-I/3Y9V seems not to be affected by the mutation of the peptide. In fact, it results very similar and perfectly superimposable to the previously reported MHC-I/WT and MHC-I/1Y9V structures (*21, 32*). Therefore, our analysis will focus on the conformations that the peptides assume in cleft of HLA-A2.

Peptide conformations of the MHC/ 3Y9V

The NY-ESO-$1_{157–165}$ variant peptide is bound in an extended conformation in both structures (*Figure 2 B, C* and *3A*).

The residues of their the N- and C- *termini* are well superposable (*Figure 3 A*) while differences are located in the central region of the peptide. In particular, in MHC-I/3Y9Va, Trp5 is flipped and it heads towards the heavy chain and not to the solvent (*Figure 4 A* and *B*), as it does in the MHC-I/3Y9Vb (*Figure 4 C* and *D*). As consequence, in MHC-I/3Y9Va, Trp5 establishes an interaction with Arg97 of the heavy chain, and the Ile6 is displaced and points to the solvent.

Interestingly, despite different peptide conformations, no big side chains rearrangements are detectable in the heavy chain. This observation suggests that the presence of two alternative conformations should be due to the flexibility of the peptide bound into the heavy chain cleft.

MHC-I/3Y9Va and MHC-I/3Y9Vb peptides have been superposed to those of the previously reported MHC-I/WT and MHC-I/1Y9V structures (*Figure 3 B* and *C*) (*21, 32*). MHC-I/3Y9Vb peptide conformation results generally comparable to that of MHC-I/WT and MHC-I/1Y9V peptides. As reported above, in the MHC-I/3Y9Vb structure, Met4 and Trp5 residues form a bulky protrusion from the surface of the pMHC, known to play a crucial role for TCR recognition (*Figure 3* and *4 C* and *D*) (*22*).

Such structural comparison reveals that L3Y and C9V substitution causes only very subtle difference in the general orientation of the side chain within the pocket of the HLA-A2 peptide-binding groove (Figure 3). The increase in stability of MHC-I/3Y9V, thus, cannot be associated to a stabilizing local rearrangement in the cleft but rather, it can be associated to the highly optimised local interactions

between the peptide and the heavy chain. In both MHC-I/3Y9V structure, the position of the mutated residue Tyr3 is conserved and induces the displacement of Gln155 side chain in the binding cleft (*Figure* 5). Moreover, Tyr3 establishes stacking interactions with the adjacent Trp5, Gln155, Leu156, and Tyr159. Finally, this residue is further stabilized by an interaction with a water molecule, not present in both MHC-I/WT and 1Y9V structures. As also previously described by Chen *et al*. (*22*), the Val side chain in position 9 improved shape complementarity together with the buried surface within the binding pocket compared to a Cys, enhancing the van der Waal's contacts between the mutated peptide and heavy chain residues Asp77, Leu81, Tyr116 and Trp147. All these interactions established by the 3Y9V peptide are likely to account for the increased stability observed by thermal unfolding.

The two alternative conformations of the 3Y9V peptide provide information into the role of peptide positions in determining the complex stability. Residues in position 4, 5 and 6 assume different conformations and establish different interactions in the two solved structures of MHC-I/3Y9V while the positions of residues 1, 3, 9 in NY-ESO-$1_{157\text{-}165}$ is highly conserved. It is noteworthy that a single unfolding step is observed in the temperature ramp (*Figure 1*) indicating that the MHC-I/3Y9V complexes may have different peptide binding mode but they all have the same stability and just the different crystallization conditions may have favoured one of the two conformations. Crucially this observation directly rules out an important role for residues 4, 5, and 6 in determining MHC-I stability and further stress the relevance of the anchoring residues such as 1 3 and 9.

Discussion

In this work, eight MHC-I bearing peptide variants of the tumour antigenic peptide NY–ESO-$1_{157\text{-}165}$ have been purified and their fold stability has been characterized to assess the impact on MHC-I fold stability of designed mutations at positions 1, 3 and 9 of the bound peptide.

As reported by previous studies, peptide residues 1 and 3 play a crucial role for the specificity of peptide binding together with the two "anchor" residues at position 2 and 9 (*33, 34*).

Indeed, all MHC-I variants displayed a higher stability compared to MHC-I/WT in thermal unfolding experiments. Moreover, the structure of one of the most stable complexes, MHC-I/3Y9V, has been determined at high resolution in two different space groups, allowing the characterization of the structural contributions to the improved stability. As reported by Chen *et al.* for MHC-I/9V (*22*), the structural arrangement of the complex is conserved, and only subtle differences in peptide conformation can be observed despite the significant boost in stability compared to MHC-I/WT (*21*).

In keeping with previous data (*34*), we demonstrate that the heavy chain binding pocket of residue 9 ideally accommodate small size non-polar residues such as Val, indeed, Val has been shown to be the most dominant COOH-terminal peptide anchor residue bound to HLA-A2 haplotype (*35-37*). In addition, an aromatic residue such as Tyr is favoured at the peptide third position, as previously reported (*34, 35*).

Thus, in line with reported HLA-A2 peptide preference, the mutations introduced to NY–ESO-$1_{157\text{-}165}$ seems to be enough to improve the van der Waal's contacts between the altered ligand peptide and HLA-A2, conferring enhanced stability to the complex.

Our results provide new evidence into the mechanisms controlling the higher stability induced by APLs binding by confirming a correlation between the stability of the MHC-I complex and a rational peptide mutagenesis. In particular, our experiments have successfully identified several NY-ESO-$1_{157\text{-}165}$ variants, which display an excellent performance *in vitro*. In future experiments, the relative immunogenicity of each analogue compared to the WT peptide shall be tested for TCR recognition and for their effects on MHC-I half-live in cells.

References

1. M. A. Batalia, E. J. Collins, Peptide binding by class I and class II MHC molecules. *Biopolymers* **43**, 281-302 (1997).
2. A. Townsend, H. Bodmer, Antigen recognition by class I-restricted T lymphocytes. *Annu Rev Immunol* **7**, 601-624 (1989).
3. I. A. York, K. L. Rock, Antigen processing and presentation by the class I major histocompatibility complex. *Annu Rev Immunol* **14**, 369-396 (1996).
4. C. D. Platsoucas *et al.*, Immune responses to human tumors: development of tumor vaccines. *Anticancer Res* **23**, 1969-1996 (2003).
5. M. Vergati, C. Intrivici, N. Y. Huen, J. Schlom, K. Y. Tsang, Strategies for cancer vaccine development. *J Biomed Biotechnol* **2010**, (2010).
6. A. W. Purcell, J. McCluskey, J. Rossjohn, More than one reason to rethink the use of peptides in vaccine design. *Nat Rev Drug Discov* **6**, 404-414 (2007).
7. E. Jager *et al.*, Simultaneous humoral and cellular immune response against cancer-testis antigen NY-ESO-1: definition of human histocompatibility leukocyte antigen (HLA)-A2-binding peptide epitopes. *J Exp Med* **187**, 265-270 (1998).
8. A. J. Simpson, O. L. Caballero, A. Jungbluth, Y. T. Chen, L. J. Old, Cancer/testis antigens, gametogenesis and cancer. *Nat Rev Cancer* **5**, 615-625 (2005).
9. Y. T. Chen *et al.*, A testicular antigen aberrantly expressed in human cancers detected by autologous antibody screening. *Proc Natl Acad Sci U S A* **94**, 1914-1918 (1997).
10. A. A. Jungbluth *et al.*, Immunohistochemical analysis of NY-ESO-1 antigen expression in normal and malignant human tissues. *Int J Cancer* **92**, 856-860 (2001).
11. Y. Sugita *et al.*, NY-ESO-1 expression and immunogenicity in malignant and benign breast tumors. *Cancer Res* **64**, 2199-2204 (2004).
12. J. L. Chen *et al.*, Identification of NY-ESO-1 peptide analogues capable of improved stimulation of tumor-reactive CTL. *J Immunol* **165**, 948-955 (2000).
13. T. Nicholaou *et al.*, Directions in the immune targeting of cancer: lessons learned from the cancer-testis Ag NY-ESO-1. *Immunol Cell Biol* **84**, 303-317 (2006).
14. D. Valmori *et al.*, Naturally occurring human lymphocyte antigen-A2 restricted CD8+ T-cell response to the cancer testis antigen NY-ESO-1 in melanoma patients. *Cancer Res* **60**, 4499-4506 (2000).
15. E. Jager *et al.*, Monitoring CD8 T cell responses to NY-ESO-1: correlation of humoral and cellular immune responses. *Proc Natl Acad Sci U S A* **97**, 4760-4765 (2000).
16. S. Adams *et al.*, Immunization of malignant melanoma patients with full-length NY-ESO-1 protein using TLR7 agonist imiquimod as vaccine adjuvant. *J Immunol* **181**, 776-784 (2008).
17. Y. Tian *et al.*, The novel complex combination of alum, CpG ODN and HH2 as adjuvant in cancer vaccine effectively suppresses tumor growth in vivo. *Oncotarget* **8**, 45951-45964 (2017).
18. S. G. Reed, M. T. Orr, C. B. Fox, Key roles of adjuvants in modern vaccines. *Nat Med* **19**, 1597-1608 (2013).
19. E. Jager *et al.*, Induction of primary NY-ESO-1 immunity: CD8+ T lymphocyte and antibody responses in peptide-vaccinated patients with NY-ESO-1+ cancers. *Proc Natl Acad Sci U S A* **97**, 12198-12203 (2000).
20. W. Chen, J. W. Yewdell, R. L. Levine, J. R. Bennink, Modification of cysteine residues in vitro and in vivo affects the immunogenicity and antigenicity of major histocompatibility complex class I-restricted viral determinants. *J Exp Med* **189**, 1757-1764 (1999).
21. A. I. Webb *et al.*, Functional and structural characteristics of NY-ESO-1-related HLA A2-restricted epitopes and the design of a novel immunogenic analogue. *J Biol Chem* **279**, 23438-23446 (2004).
22. J. L. Chen *et al.*, Structural and kinetic basis for heightened immunogenicity of T cell vaccines. *J Exp Med* **201**, 1243-1255 (2005).

23. M. J. van Stipdonk *et al.*, Design of agonistic altered peptides for the robust induction of CTL directed towards H-2Db in complex with the melanoma-associated epitope gp100. *Cancer Res* **69**, 7784-7792 (2009).
24. S. Pellegrino, C. Annoni, A. Contini, F. Clerici, M. L. Gelmi, Expedient chemical synthesis of 75mer DNA binding domain of MafA: an insight on its binding to insulin enhancer. *Amino Acids* **43**, 1995-2003 (2012).
25. G. Esposito *et al.*, Removal of the N-terminal hexapeptide from human beta2-microglobulin facilitates protein aggregation and fibril formation. *Protein Sci* **9**, 831-845 (2000).
26. A. Achour *et al.*, Murine class I major histocompatibility complex H-2Dd: expression, refolding and crystallization. *Acta Crystallogr D Biol Crystallogr* **55**, 260-262 (1999).
27. W. Kabsch, Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-132 (2010).
28. P. Evans, Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* **62**, 72-82 (2006).
29. A. J. McCoy *et al.*, Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674 (2007).
30. G. N. Murshudov, A. A. Vagin, E. J. Dodson, Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**, 240-255 (1997).
31. P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics. *Acta crystallographica* **60**, 2126-2132 (2004).
32. L. Halabelian *et al.*, Class I major histocompatibility complex, the trojan horse for secretion of amyloidogenic beta2-microglobulin. *J Biol Chem* **289**, 3318-3327 (2014).
33. P. Cano, B. Fan, A geometric and algebraic view of MHC-peptide complexes and their binding properties. *BMC Struct Biol* **1**, 2 (2001).
34. K. C. Parker *et al.*, Sequence motifs important for peptide binding to the human MHC class I molecule, HLA-A2. *J Immunol* **149**, 3580-3587 (1992).
35. P. Guan, I. A. Doytchinova, V. A. Walshe, P. Borrow, D. R. Flower, Analysis of peptide-protein binding using amino acid descriptors: prediction and experimental verification for human histocompatibility complex HLA-A0201. *J Med Chem* **48**, 7418-7425 (2005).
36. K. Falk, O. Rotzschke, S. Stevanovic, G. Jung, H. G. Rammensee, Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature* **351**, 290-296 (1991).
37. D. F. Hunt *et al.*, Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry. *Science* **255**, 1261-1263 (1992).

Figures

Figure 1. MHC-I stability measured by CD. Variation of Far-UV CD signal as a function of temperature. Thermal unfolding was monitored at 218 nm of MHC-I bearing NY-ESO-$1_{157\text{-}165}$ variants. Melting temperatures assessed using CD curves are reported in *Table 2*.

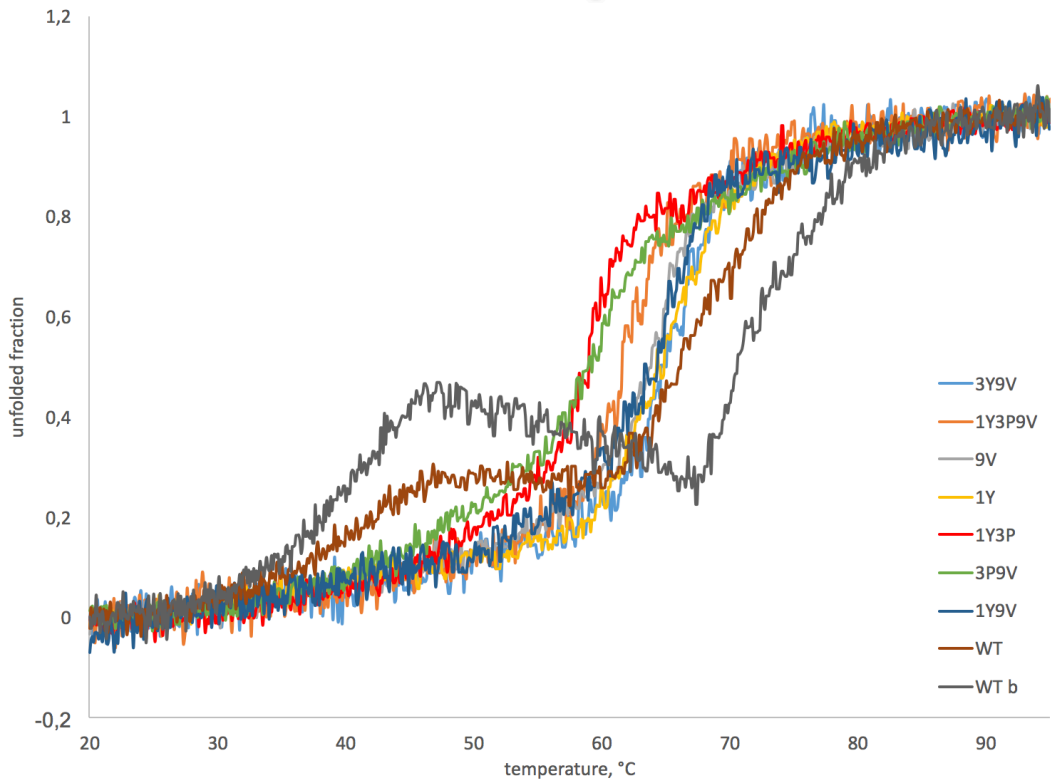


Figure 2. X-ray Crystal structure of MHC-I/3Y9Va and MHC-I/3Y9Vb. **(A)** Superposition of the structures of MHC-I/3Y9Va (cyan) and MHC-I/3Y9Vb (pink) (RMSD: 0.91 382/382 Cα) **(B)** 3Y9V peptide in the binding pocket of MHC-I/3Y9Va structure. Electron density map is clipped around the peptide. **(C)** 3Y9V peptide in the binding pocket of MHC-I/3Y9Vb structure. Electron density map is clipped around the peptide.

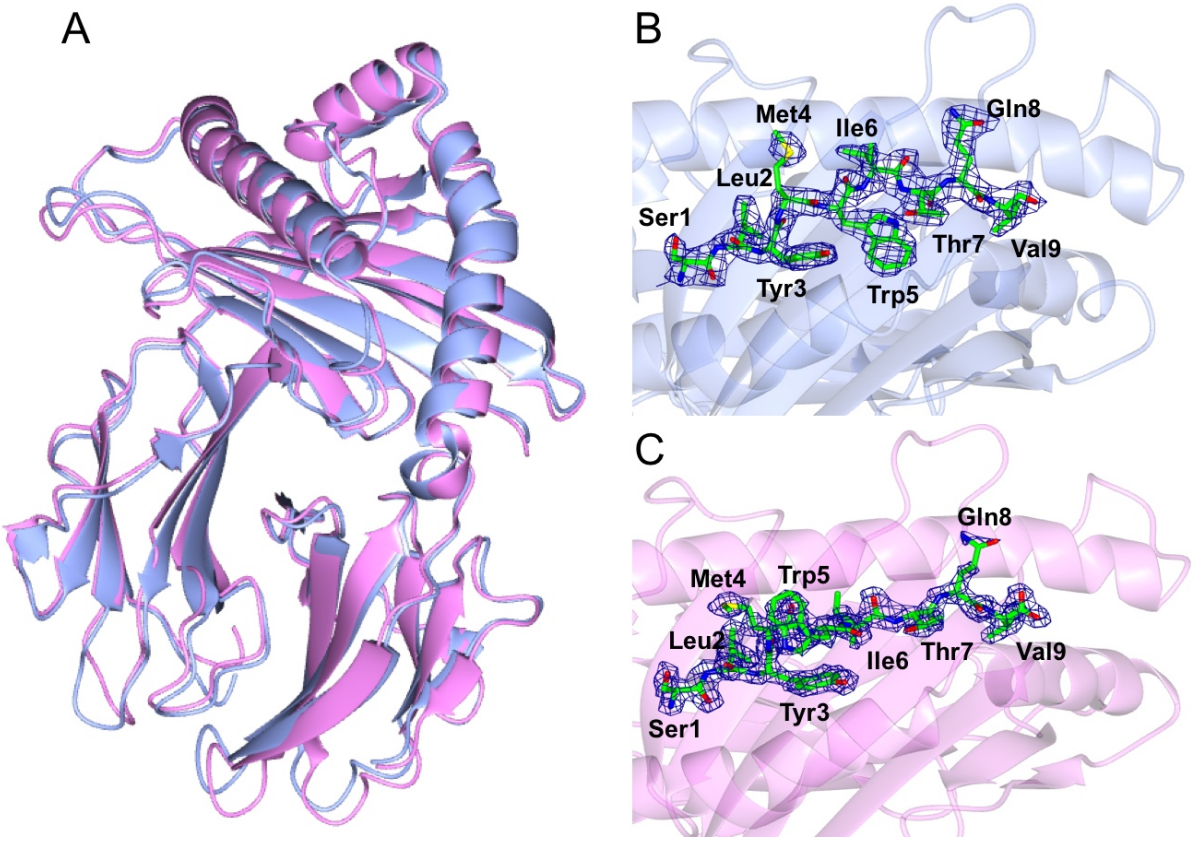


Figure 3. 3Y9V, WT and 1Y9V peptides superimposition. **(A)** Superposition of the peptides of MHC-I/3Y9Va (cyan) and MHC-I/3Y9Vb (pink) structures. **(B)** Superposition of the peptides of MHC-I/3Y9Va (cyan) and MHC-I/WT (yellow, pdb 1S9W) and MHC-I/1Y9V (green, pdb 4L29) structures. **(C)** Superposition of the peptides of MHC-I/3Y9Vb (pink) and MHC-I/WT (yellow, pdb 1S9W) and MHC-I/1Y9V (green, pdb 4L29) structures.

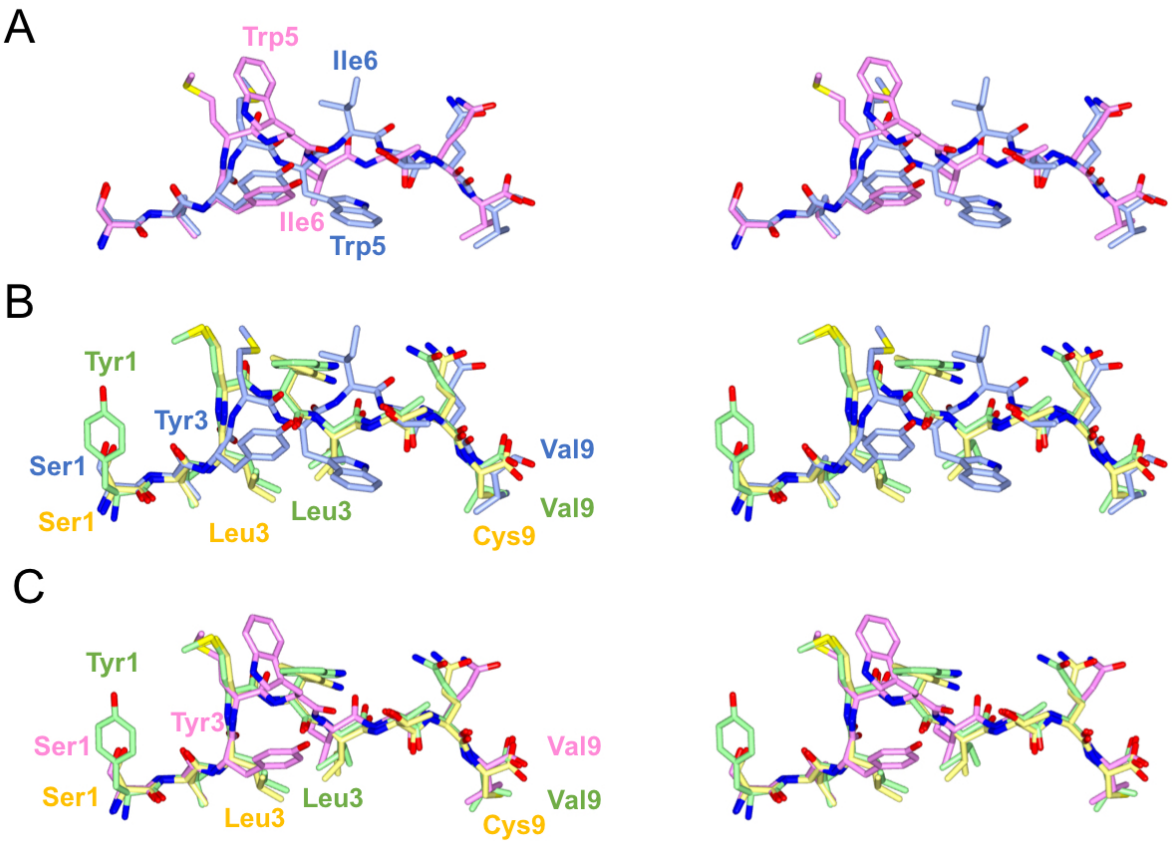


Figure 4. Relative position of Trp5 and Arg97 in the crystal structure of MHC-I/3Y9Va and MHC-I/3Y9Vb. **A** and **B** show the site on MHC-I/3Y9Va from two different perspectives. **C** and **D** show the site on MHC-I/3Y9Vb.

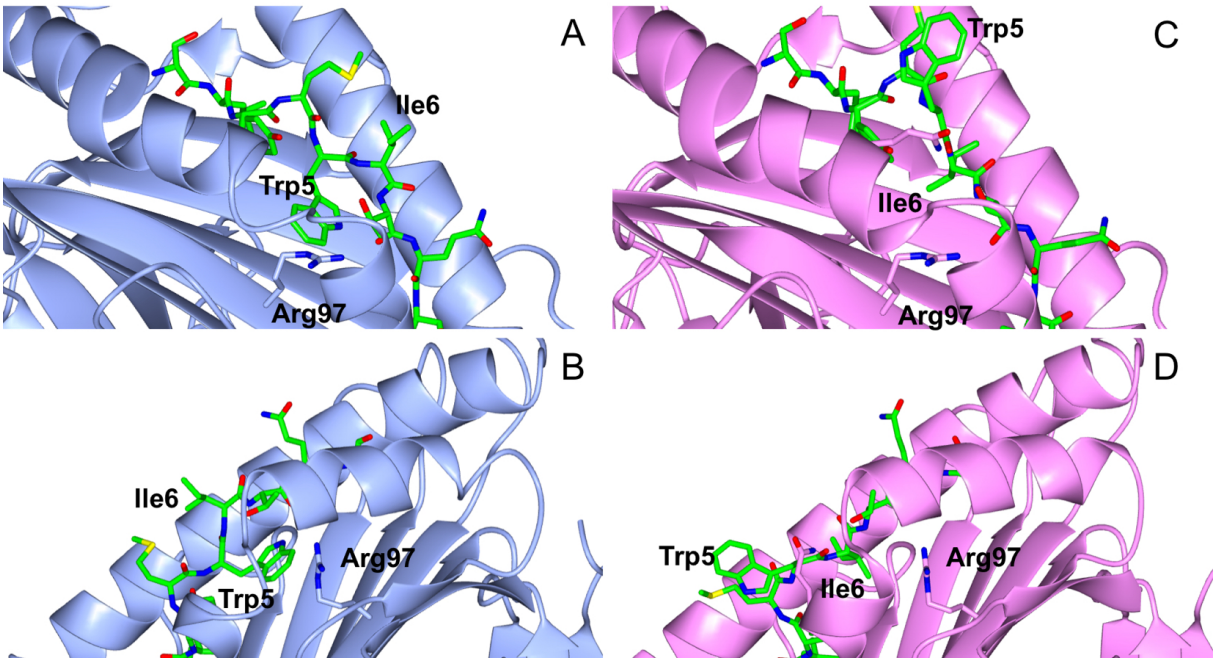
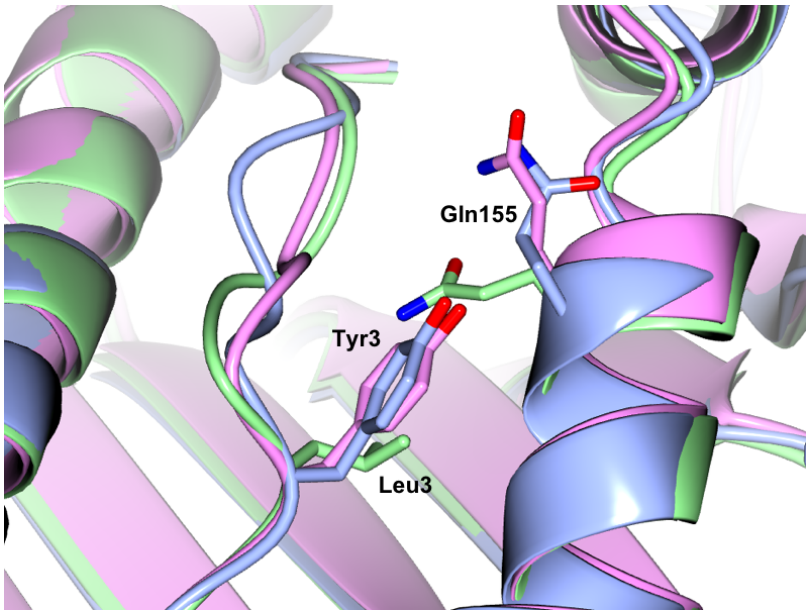


Figure 5. Relative position Gln155 in the crystal structure of MHC-I/3Y9Va (cyan) and MHC-I/3Y9Vb (pink) and MHC-I/WT (green, pdb code 1S9W).



Tables

Table 1. Purification yields for all complexes are reported. Notably, complexes bearing the WT peptide and the 1Y variant, has the lowest yield.

Peptide	Complex	Yield %
SLLMWITQC	WT	4.2
YL**P**MWITQ**V**	1Y3P9V	15.9
SL**Y**MWITQ**V**	3Y9V	14.8
YLLMWITQ**V**	1Y9V	16.8
YLLMWITQC	1Y	2.6
YL**P**MWITQC	1Y3P	8.8
SLLMWITQ**V**	9V	7.9
SL**P**MWITQ**V**	3P9V	11.9

Table 2. Calculated melting temperature for the complexes. T_m values are assessed on Far-UV CD curves.

Complex	Tm / °C
WT	57*
1Y3P9V	61.7
3Y9V	65.5
1Y9V	64.9
1Y	64.5
1Y3P	59.0
9V	64.1
3P9V	59.5

* as reported by Webb *et al.* (*21*).

Table 3. Data collection and refinement statistics for MHC-I/3Y9V. Values in parenthesis are for the highest resolution shell.

Structure	**Crystallization condition (A)**	**Crystallization condition (B)**
Beam Line	ID30a (ESRF)	ID30a (ESRF)
Space group	P 21 21 21	P 1 21 1
Unit cell constants (Å, °)	a = 63.94, b = 86.40, c = 153.83; α = 89.95, β = 89.80, γ = 89.94	a = 41.04, b = 53.35, c = 85.16; α = 90.13, β = 94.11, γ = 90.05
Resolution (Å)	49.17 - 2.50 (2.60 - 2.50)	45.22 – 1.65 (1.68 – 1.65)
R_{merge} (%)[a]	16.5 (69.7)	7.6 (35.4)
I/σI	8.6 (2.9)	8.1 (2.2)
Completeness (%)	100.0 (100.0)	99.8 (99.1)
Multiplicity	3.5 (3.6)	3.3 (3.1)
Unique reflections	30297 (3353)	44267 (2175)
Refinement		
R_{work} (%)[b]	19.13	19.4
R_{free} (%)	25.49	23.1
Number of atoms	6707	3581
Protein	6306	3147
Water	401	434
Heteroatoms	-	-
Ramachandran plot, *n* (%)		
Most favoured region	722 (96.52)	352 (96.70)
Allowed region	22 (2.94)	10 (2.75)
Outliers	4 (0.53)	2 (0.55)

[a] $R_{merge} = \Sigma_{hkl} \Sigma_j I_{hkl,j} - \langle I_{hkl} \rangle / \Sigma_{hkl} \Sigma_j I_{hkl,j}$ where I is the observed intensity and <I> is the average intensity.

[b] $R_{work} = \Sigma_{hkl} F_o - F_c / \Sigma_{hkl} F_o$ for all data except 5–10%, which were used for the R_{free} calculation.

Supporting Tables and Figures

Figure S1. Gel filtration chromatograms of MHC-I/WT and MHC-I/3P9V. The rectangle highlight the elution peak of the refolded complexes. MHC-I/WT and MHC-I/1Y display similar elution profiles and lower yields compared to that of MHC-I/3P9V (as example of all the other MHC-I/NY-ESO-$1_{155-167}$ variant peptides).

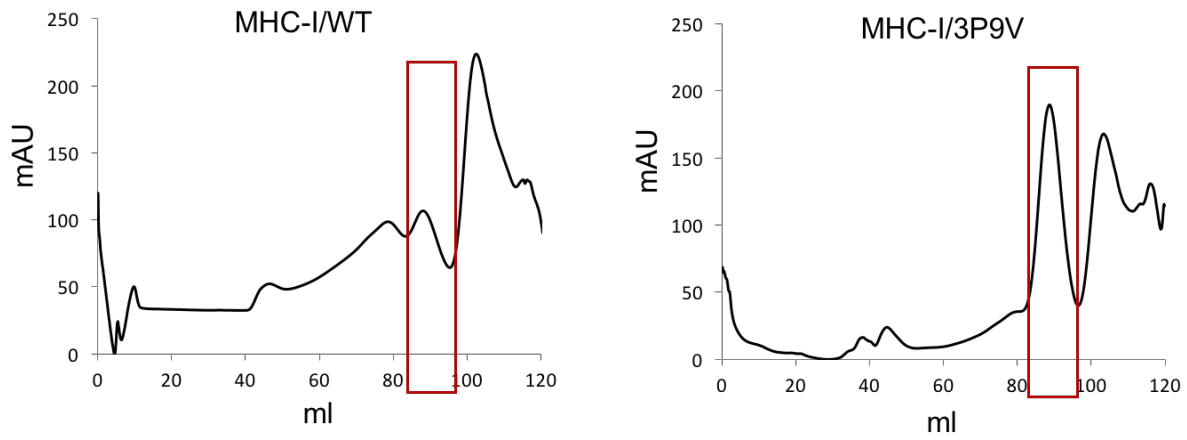
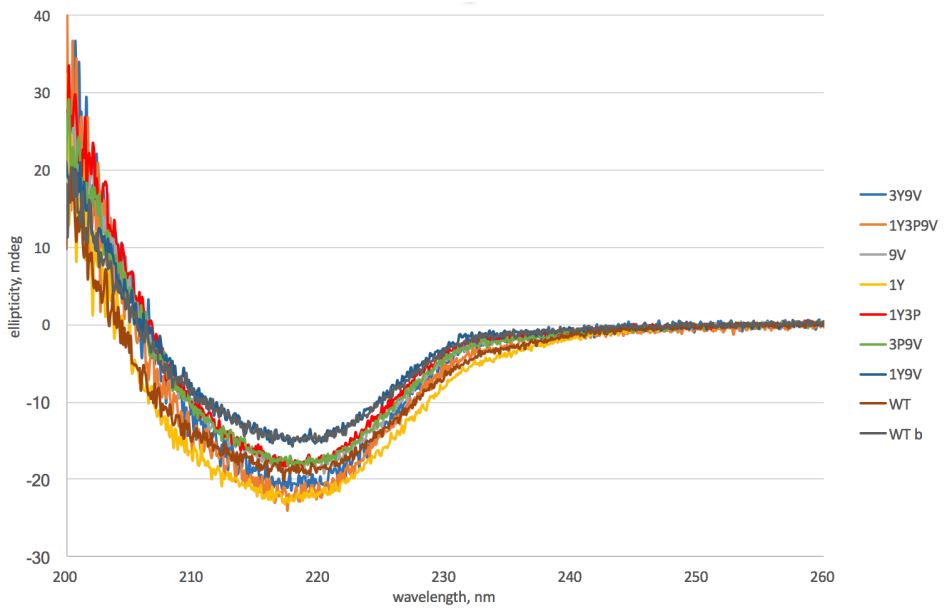


Figure S2. Far-UV spectra of all complexes monitored by circular dichroism.



4. Conclusions

My thesis aims to provide further understanding on the effects that a single amino acid mutation can play on protein stability and protein physiologic and pathologic properties using two systems and two different biological contexts: β2m and the MHC-I.

I

To identify key resides that trigger β2m amyloid aggregation, the native state dynamics of WT β2m and W60G β2m have been compared, combining solution NMR spectroscopy and molecular dynamics (MD) simulations.

The work provides insights into structural determinants of the lack of aggregation propensity of W60G β2m. The W60G mutation not only increases the protein fold stability but also generates a more regular and more rigid overall structure in which the superficial aggregation-prone regions are better protected, drastically reducing the aggregation properties typical of WT β2m.

Thus, to validate the role of protein dynamics in determining aggregation propensity, three new β2m variants (V85E, W60G-N83V and W60G-Y63W) were designed and characterized both biochemically and structurally. Mutations on WT and W60G β2m were inserted to alter their aggregation prone exposed surface, modifying their aggregation propensity without altering their stability. All newly designed variants have displayed the expected structure, stability and aggregation propensity.

These results demonstrated that the aggregation propensity of β2m can be modulated by mutating surface residues that regulate the protection of aggregation-prone surfaces, with negligible effects on protein structure and stability.

II

To date, it is not fully understood whether biophysical properties assessed in crystallized proteins are representative of protein behaviour in solution.

The project deals with a comparative study of protein stability in solution and in the crystalline form on three β2m variants: WT, W60G and D76N. These variants, besides their differences in stability and aggregation propensity, display equal crystal properties, thus resulting in an ideal system for this kind of experiments.

Thermal unfolding analyses of the proteins in solutions and in crystal were carried out. The change in protein secondary structure was monitored at increasing temperature by Fourier transform infrared (FT-IR) spectroscopy.

Crystals results to be valid alternative samples for acquiring FT-IR spectra. Indeed, the increased structural order of the crystalline state contributes to provide better resolved spectral components compared to those collected in solution and crucially, the crystalline samples displayed thermal stabilities in good agreement with trend observed in solution (W60G > WT > D76N). Besides being a reliable tool to evaluate protein stability, FT-IR on crystalline samples allowed us to obtain also information about crystal melting through the temperature-dependent variation in maximum of the Tyr peak detectable only in crystals spectra.

Therefore, the FT-IR experiments carried out on β2m variants suggest that protein stability can be studied in crystals point at the possibility to study other biophysical properties such as protein dynamics exploration not only in solution also in crystalline samples.

III

The heterotrimeric complex MHC-I triggers an immune response by displaying antigenic peptides on the cell surface for recognition by appropriate T cells. It has been widely demonstrated that the overall MHC-I stability together with peptide-complex affinity is fundamental for successful antigen presentation.

In this context, this work dissects the impact on protein fold stability of designed complex-stabilizing point mutations on the tumour antigenic peptide NY–ESO-$1_{157\text{-}165}$ to produce more effective cancer vaccines.

Eight MHC-I bearing peptide variants were purified and characterized. Thermal stability experiments revealed a higher stability in all MHC-I variants compared to MHC-I/WT validating the strategy of "super-peptides" engineering. Moreover, the structural characterization of one of the most stable complexes allowed to deeply inspect the contributions to the improved stability obtained by rational peptide mutagenesis. Finally, our experiments have successfully identified several NY-ESO-$1_{157\text{-}165}$ variants, which display an excellent performance *in vitro* and shall be tested for TCR recognition and for their effects on MHC-I half-live in cells.

5. Bibliography

1. C. B. Anfinsen, Principles that govern the folding of protein chains. *Science* **181**, 223-230 (1973).
2. I. Guzman, M. Gruebele, Protein folding dynamics in the cell. *J Phys Chem B* **118**, 8459-8470 (2014).
3. C. M. Dobson, Protein folding and misfolding. *Nature* **426**, 884-890 (2003).
4. C. B. Anfinsen, The formation and stabilization of protein structure. *Biochem J* **128**, 737-749 (1972).
5. R. L. Baldwin, Energetics of protein folding. *J Mol Biol* **371**, 283-301 (2007).
6. C. Nick Pace, J. M. Scholtz, G. R. Grimsley, Forces stabilizing proteins. *FEBS Lett* **588**, 2177-2184 (2014).
7. M. M. Gromiha, S. Selvaraj, Inter-residue interactions in protein folding and stability. *Prog Biophys Mol Biol* **86**, 235-277 (2004).
8. L. Lins, R. Brasseur, The hydrophobic effect in protein folding. *FASEB J* **9**, 535-540 (1995).
9. T. Ooi, Thermodynamics of protein folding: effects of hydration and electrostatic interactions. *Adv Biophys* **30**, 105-154 (1994).
10. D. Fass, Disulfide bonding in protein biophysics. *Annu Rev Biophys* **41**, 63-79 (2012).
11. M. V. Trivedi, J. S. Laurence, T. J. Siahaan, The role of thiols and disulfides on protein stability. *Curr Protein Pept Sci* **10**, 614-625 (2009).
12. R. D. Socha, N. Tokuriki, Modulating protein stability - directed evolution strategies for improved protein function. *FEBS J* **280**, 5582-5595 (2013).
13. A. Horwich, Protein aggregation in disease: a role for folding intermediates forming specific multimeric interactions. *J Clin Invest* **110**, 1221-1232 (2002).
14. G. Merlini, V. Bellotti, Molecular mechanisms of amyloidosis. *N Engl J Med* **349**, 583-596 (2003).
15. M. Sunde *et al.*, Common core structure of amyloid fibrils by synchrotron X-ray diffraction. *J Mol Biol* **273**, 729-739 (1997).
16. A. W. Fitzpatrick, S. T. Park, A. H. Zewail, Exceptional rigidity and biomechanics of amyloid revealed by 4D electron microscopy. *Proc Natl Acad Sci U S A* **110**, 10976-10981 (2013).
17. T. P. Knowles, M. Vendruscolo, C. M. Dobson, The amyloid state and its association with protein misfolding diseases. *Nat Rev Mol Cell Biol* **15**, 384-396 (2014).
18. P. Westermark, J. Bergstrom, A. Solomon, C. Murphy, K. Sletten, Transthyretin-derived senile systemic amyloidosis: clinicopathologic and structural considerations. *Amyloid* **10 Suppl 1**, 48-54 (2003).
19. F. Danesh, L. T. Ho, Dialysis-related amyloidosis: history and clinical manifestations. *Semin Dial* **14**, 80-85 (2001).
20. D. Allsop, J. Mayes, Amyloid beta-peptide and Alzheimer's disease. *Essays Biochem* **56**, 99-110 (2014).
21. J. W. Kelly, Alternative conformations of amyloidogenic proteins govern their behavior. *Curr Opin Struct Biol* **6**, 11-17 (1996).
22. S. Valleix *et al.*, Hereditary systemic amyloidosis due to Asp76Asn variant beta2-microglobulin. *N Engl J Med* **366**, 2276-2283 (2012).
23. M. D. Benson, J. C. Kincaid, The molecular biology and clinical features of amyloid neuropathy. *Muscle Nerve* **36**, 411-423 (2007).
24. T. L. Poshusta *et al.*, Mutations in specific structural regions of immunoglobulin light chains are associated with free light chain levels in patients with AL amyloidosis. *PLoS One* **4**, e5169 (2009).
25. A. Natalello *et al.*, Co-fibrillogenesis of Wild-type and D76N beta2-Microglobulin: THE CRUCIAL ROLE OF FIBRILLAR SEEDS. *J Biol Chem* **291**, 9678-9689 (2016).
26. P. Arosio, T. P. Knowles, S. Linse, On the lag phase in amyloid fibril formation. *Phys Chem Chem Phys* **17**, 7606-7618 (2015).

27. M. Girych *et al.*, Combined thioflavin T-Congo red fluorescence assay for amyloid fibril detection. *Methods Appl Fluoresc* **4**, 034010 (2016).
28. H. Naiki, K. Higuchi, M. Hosokawa, T. Takeda, Fluorometric determination of amyloid fibrils in vitro using the fluorescent dye, thioflavin T1. *Anal Biochem* **177**, 244-249 (1989).
29. H. Naiki *et al.*, Fluorometric examination of tissue amyloid fibrils in murine senile amyloidosis: use of the fluorescent indicator, thioflavine T. *Lab Invest* **62**, 768-773 (1990).
30. C. M. Dobson, Protein misfolding, evolution and disease. *Trends Biochem Sci* **24**, 329-332 (1999).
31. M. Fandrich, C. M. Dobson, The behaviour of polyamino acids reveals an inverse side chain effect in amyloid structure formation. *EMBO J* **21**, 5682-5690 (2002).
32. G. Verdone *et al.*, The solution structure of human beta2-microglobulin reveals the prodromes of its amyloid transition. *Protein Sci* **11**, 487-499 (2002).
33. J. W. Becker, G. N. Reeke, Jr., Three-dimensional structure of beta 2-microglobulin. *Proc Natl Acad Sci U S A* **82**, 4225-4229 (1985).
34. V. Bellotti *et al.*, Beta2-microglobulin can be refolded into a native state from ex vivo amyloid fibrils. *Eur J Biochem* **258**, 61-67 (1998).
35. D. P. Hong, M. Gozu, K. Hasegawa, H. Naiki, Y. Goto, Conformation of beta 2-microglobulin amyloid fibrils analyzed by reduction of the disulfide bond. *J Biol Chem* **277**, 21554-21560 (2002).
36. E. A. Hughes, C. Hammond, P. Cresswell, Misfolded major histocompatibility complex class I heavy chains are translocated into the cytoplasm and degraded by the proteasome. *Proc Natl Acad Sci U S A* **94**, 1896-1901 (1997).
37. K. M. Paulsson *et al.*, Distinct differences in association of MHC class I with endoplasmic reticulum proteins in wild-type, and beta 2-microglobulin- and TAP-deficient cell lines. *Int Immunol* **13**, 1063-1073 (2001).
38. P. J. Bjorkman *et al.*, Structure of the human class I histocompatibility antigen, HLA-A2. *Nature* **329**, 506-512 (1987).
39. L. Halabelian *et al.*, Class I major histocompatibility complex, the trojan horse for secretion of amyloidogenic beta2-microglobulin. *J Biol Chem* **289**, 3318-3327 (2014).
40. J. Floege *et al.*, Clearance and synthesis rates of beta 2-microglobulin in patients undergoing hemodialysis and in normal subjects. *J Lab Clin Med* **118**, 153-165 (1991).
41. J. Floege, G. Ehlerding, Beta-2-microglobulin-associated amyloidosis. *Nephron* **72**, 9-26 (1996).
42. J. Floege, M. Ketteler, beta2-microglobulin-derived amyloidosis: an update. *Kidney Int Suppl* **78**, S164-171 (2001).
43. F. Gejyo *et al.*, A new form of amyloid protein associated with chronic hemodialysis was identified as beta 2-microglobulin. *Biochem Biophys Res Commun* **129**, 701-706 (1985).
44. H. Assenat *et al.*, [Hemodialysis: carpal tunnel syndrome and amyloid substance]. *Nouv Presse Med* **9**, 1715 (1980).
45. T. Eichner, S. E. Radford, A generic mechanism of beta2-microglobulin amyloid assembly at neutral pH involving a specific proline switch. *J Mol Biol* **386**, 1312-1326 (2009).
46. S. L. Myers *et al.*, A systematic study of the effect of physiological factors on beta2-microglobulin amyloid formation at neutral pH. *Biochemistry* **45**, 2311-2321 (2006).
47. C. M. Eakin, A. D. Miranker, From chance to frequent encounters: origins of beta2-microglobulin fibrillogenesis. *Biochim Biophys Acta* **1753**, 92-99 (2005).
48. R. Srikanth, V. L. Mendoza, J. D. Bridgewater, G. Zhang, R. W. Vachet, Copper binding to beta-2-microglobulin and its pre-amyloid oligomers. *Biochemistry* **48**, 9871-9881 (2009).
49. N. B. Borotto, Z. Zhang, J. Dong, B. Burant, R. W. Vachet, Increased beta-Sheet Dynamics and D-E Loop Repositioning Are Necessary for Cu(II)-Induced Amyloid Formation by beta-2-Microglobulin. *Biochemistry* **56**, 1095-1104 (2017).
50. S. Yamamoto *et al.*, Glycosaminoglycans enhance the trifluoroethanol-induced extension of beta 2-microglobulin-related amyloid fibrils at a neutral pH. *J Am Soc Nephrol* **15**, 126-133 (2004).
51. A. Relini *et al.*, Heparin strongly enhances the formation of beta2-microglobulin amyloid fibrils in the presence of type I collagen. *J Biol Chem* **283**, 4912-4920 (2008).

52. H. Pal-Gabor *et al.*, Mechanism of lysophosphatidic acid-induced amyloid fibril formation of beta(2)-microglobulin in vitro under physiological conditions. *Biochemistry* **48**, 5689-5699 (2009).
53. G. Esposito *et al.*, Removal of the N-terminal hexapeptide from human beta2-microglobulin facilitates protein aggregation and fibril formation. *Protein Sci* **9**, 831-845 (2000).
54. K. Hasegawa *et al.*, Growth of beta(2)-microglobulin-related amyloid fibrils by non-esterified fatty acids at a neutral pH. *Biochem J* **416**, 307-315 (2008).
55. T. Niwa, Dialysis-related amyloidosis: pathogenesis focusing on AGE modification. *Semin Dial* **14**, 123-126 (2001).
56. F. Chiti *et al.*, Detection of two partially structured species in the folding process of the amyloidogenic protein beta 2-microglobulin. *J Mol Biol* **307**, 379-391 (2001).
57. F. Chiti *et al.*, A partially structured species of beta 2-microglobulin is significantly populated under physiological conditions and involved in fibrillogenesis. *J Biol Chem* **276**, 46714-46721 (2001).
58. M. Sakata *et al.*, Kinetic coupling of folding and prolyl isomerization of beta2-microglobulin studied by mutational analysis. *J Mol Biol* **382**, 1242-1255 (2008).
59. C. Rosano *et al.*, beta2-microglobulin H31Y variant 3D structure highlights the protein natural propensity towards intermolecular aggregation. *J Mol Biol* **335**, 1051-1064 (2004).
60. C. M. Eakin, A. J. Berman, A. D. Miranker, A native to amyloidogenic transition regulated by a backbone trigger. *Nat Struct Mol Biol* **13**, 202-208 (2006).
61. G. Esposito *et al.*, The controlling roles of Trp60 and Trp95 in beta2-microglobulin function, folding and amyloid aggregation properties. *J Mol Biol* **378**, 887-897 (2008).
62. S. Ricagno *et al.*, DE loop mutations affect beta2-microglobulin stability and amyloid aggregation. *Biochem Biophys Res Commun* **377**, 146-150 (2008).
63. J. P. Hodkinson, T. R. Jahn, S. E. Radford, A. E. Ashcroft, HDX-ESI-MS reveals enhanced conformational dynamics of the amyloidogenic protein beta(2)-microglobulin upon release from the MHC-1. *J Am Soc Mass Spectrom* **20**, 278-286 (2009).
64. F. Fogolari *et al.*, Molecular dynamics simulation suggests possible interaction patterns at early steps of beta2-microglobulin aggregation. *Biophys J* **92**, 1673-1681 (2007).
65. M. Hoshino *et al.*, Mapping the core of the beta(2)-microglobulin amyloid fibril by H/D exchange. *Nat Struct Biol* **9**, 332-336 (2002).
66. C. Santambrogio *et al.*, DE-loop mutations affect beta2 microglobulin stability, oligomerization, and the low-pH unfolded form. *Protein Sci* **19**, 1386-1394 (2010).
67. G. W. Platt, K. E. Routledge, S. W. Homans, S. E. Radford, Fibril growth kinetics reveal a region of beta2-microglobulin important for nucleation and elongation of aggregation. *J Mol Biol* **378**, 251-263 (2008).
68. M. Colombo *et al.*, The effects of an ideal beta-turn on beta-2 microglobulin fold stability. *J Biochem* **150**, 39-47 (2011).
69. P. P. Mangione *et al.*, Structure, folding dynamics, and amyloidogenesis of D76N beta2-microglobulin: roles of shear flow, hydrophobic surfaces, and alpha-crystallin. *J Biol Chem* **288**, 30917-30930 (2013).
70. M. Stoppini, V. Bellotti, Systemic amyloidosis: lessons from beta2-microglobulin. *J Biol Chem* **290**, 9951-9958 (2015).
71. M. de Rosa *et al.*, Decoding the Structural Bases of D76N ss2-Microglobulin High Amyloidogenicity through Crystallography and Asn-Scan Mutagenesis. *PLoS One* **10**, e0144061 (2015).
72. D. Ron, P. Walter, Signal integration in the endoplasmic reticulum unfolded protein response. *Nat Rev Mol Cell Biol* **8**, 519-529 (2007).
73. G. Vogt, S. Woell, P. Argos, Protein thermal stability, hydrogen bonds, and ion pairs. *J Mol Biol* **269**, 631-643 (1997).
74. S. Kumar, C. J. Tsai, R. Nussinov, Factors enhancing protein thermostability. *Protein Eng* **13**, 179-191 (2000).
75. J. D. Comber, R. Philip, MHC class I antigen presentation and implications for developing a new generation of therapeutic vaccines. *Ther Adv Vaccines* **2**, 77-89 (2014).
76. I. A. York, K. L. Rock, Antigen processing and presentation by the class I major histocompatibility complex. *Annu Rev Immunol* **14**, 369-396 (1996).

77. M. A. Batalia, E. J. Collins, Peptide binding by class I and class II MHC molecules. *Biopolymers* **43**, 281-302 (1997).
78. K. Falk, O. Rotzschke, S. Stevanovic, G. Jung, H. G. Rammensee, Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature* **351**, 290-296 (1991).
79. L. J. Stern, D. C. Wiley, Antigenic peptide binding by class I and class II histocompatibility proteins. *Behring Inst Mitt*, 1-10 (1994).
80. F. Latron *et al.*, A critical role for conserved residues in the cleft of HLA-A2 in presentation of a nonapeptide to T cells. *Science* **257**, 964-967 (1992).
81. K. J. Smith *et al.*, An altered position of the alpha 2 helix of MHC class I is revealed by the crystal structure of HLA-B*3501. *Immunity* **4**, 203-213 (1996).
82. L. M. Hellman *et al.*, Differential scanning fluorimetry based assessments of the thermal and kinetic stability of peptide-MHC complexes. *J Immunol Methods* **432**, 95-101 (2016).
83. R. C. Hillig *et al.*, Thermodynamic and structural analysis of peptide- and allele-dependent properties of two HLA-B27 subtypes exhibiting differential disease association. *J Biol Chem* **279**, 652-663 (2004).
84. N. G. Walpole *et al.*, The structure and stability of the monomorphic HLA-G are influenced by the nature of the bound peptide. *J Mol Biol* **397**, 467-480 (2010).
85. C. Motozono, S. Yanaka, K. Tsumoto, M. Takiguchi, T. Ueno, Impact of intrinsic cooperative thermodynamics of peptide-MHC complexes on antiviral activity of HIV-specific CTL. *J Immunol* **182**, 5528-5536 (2009).
86. E. Riquelme, L. J. Carreno, P. A. Gonzalez, A. M. Kalergis, The duration of TCR/pMHC interactions regulates CTL effector function and tumor-killing capacity. *Eur J Immunol* **39**, 2259-2269 (2009).
87. N. A. Bowerman *et al.*, Engineering the binding properties of the T cell receptor:peptide:MHC ternary complex that governs T cell activity. *Mol Immunol* **46**, 3000-3008 (2009).
88. S. H. van der Burg, M. J. Visseren, R. M. Brandt, W. M. Kast, C. J. Melief, Immunogenicity of peptides bound to MHC class I molecules depends on the MHC-peptide complex stability. *J Immunol* **156**, 3308-3314 (1996).
89. A. Garcia-Lora, I. Algarra, F. Garrido, MHC class I antigens, immune surveillance, and tumor immune escape. *J Cell Physiol* **195**, 346-355 (2003).
90. D. S. Singer, E. Mozes, S. Kirshner, L. D. Kohn, Role of MHC class I molecules in autoimmune disease. *Crit Rev Immunol* **17**, 463-468 (1997).
91. C. D. Platsoucas *et al.*, Immune responses to human tumors: development of tumor vaccines. *Anticancer Res* **23**, 1969-1996 (2003).
92. A. W. Purcell, J. McCluskey, J. Rossjohn, More than one reason to rethink the use of peptides in vaccine design. *Nat Rev Drug Discov* **6**, 404-414 (2007).
93. S. Calbo *et al.*, Antitumor vaccination using a major histocompatibility complex (MHC) class I-restricted pseudopeptide with reduced peptide bond. *J Immunother* **23**, 125-130 (2000).
94. G. Guichard *et al.*, Partially modified retro-inverso pseudopeptides as non-natural ligands for the human class I histocompatibility molecule HLA-A2. *J Med Chem* **39**, 2030-2039 (1996).
95. G. Guichard *et al.*, Melanoma peptide MART-1(27-35) analogues with enhanced binding capacity to the human class I histocompatibility molecule HLA-A2 by introduction of a beta-amino acid residue: implications for recognition by tumor-infiltrating lymphocytes. *J Med Chem* **43**, 3803-3808 (2000).
96. T. Boon, P. van der Bruggen, Human tumor antigens recognized by T lymphocytes. *J Exp Med* **183**, 725-729 (1996).
97. E. Jager *et al.*, Simultaneous humoral and cellular immune response against cancer-testis antigen NY-ESO-1: definition of human histocompatibility leukocyte antigen (HLA)-A2-binding peptide epitopes. *J Exp Med* **187**, 265-270 (1998).
98. A. A. Jungbluth *et al.*, Immunohistochemical analysis of NY-ESO-1 antigen expression in normal and malignant human tissues. *Int J Cancer* **92**, 856-860 (2001).
99. Y. Sugita *et al.*, NY-ESO-1 expression and immunogenicity in malignant and benign breast tumors. *Cancer Res* **64**, 2199-2204 (2004).

100. Y. Li, R. Song, X. Li, F. Xu, Expression and immunogenicity of NY-ESO-1 in colorectal cancer. *Exp Ther Med* **13**, 3581-3585 (2017).
101. J. A. Veit *et al.*, Expression and clinical significance of MAGE and NY-ESO-1 cancer-testis antigens in adenoid cystic carcinoma of the head and neck. *Head Neck* **38**, 1008-1016 (2016).
102. F. van Rhee *et al.*, NY-ESO-1 is highly expressed in poor-prognosis multiple myeloma and induces spontaneous humoral and cellular immune responses. *Blood* **105**, 3939-3944 (2005).
103. D. Valmori *et al.*, Naturally occurring human lymphocyte antigen-A2 restricted CD8+ T-cell response to the cancer testis antigen NY-ESO-1 in melanoma patients. *Cancer Res* **60**, 4499-4506 (2000).
104. E. Jager *et al.*, Monitoring CD8 T cell responses to NY-ESO-1: correlation of humoral and cellular immune responses. *Proc Natl Acad Sci U S A* **97**, 4760-4765 (2000).
105. E. Jager *et al.*, Induction of primary NY-ESO-1 immunity: CD8+ T lymphocyte and antibody responses in peptide-vaccinated patients with NY-ESO-1+ cancers. *Proc Natl Acad Sci U S A* **97**, 12198-12203 (2000).
106. J. L. Chen *et al.*, Identification of NY-ESO-1 peptide analogues capable of improved stimulation of tumor-reactive CTL. *J Immunol* **165**, 948-955 (2000).
107. A. I. Webb *et al.*, Functional and structural characteristics of NY-ESO-1-related HLA A2-restricted epitopes and the design of a novel immunogenic analogue. *J Biol Chem* **279**, 23438-23446 (2004).
108. E. M. Doorduijn *et al.*, TAP-independent self-peptides enhance T cell recognition of immune-escaped tumors. *J Clin Invest* **126**, 784-794 (2016).
109. M. J. van Stipdonk *et al.*, Design of agonistic altered peptides for the robust induction of CTL directed towards H-2Db in complex with the melanoma-associated epitope gp100. *Cancer Res* **69**, 7784-7792 (2009).

Part II

Side researches not included in the main frame of the thesis

Side Project I - Published Paper II

FRET studies of various conformational states adopted by transthyretin

Ghadami S.A., Bemporad F., Sala B.M., Tiana G., Ricagno S., Chiti F.

The paper deals with the investigation of the conformational changes occurring to TTR during the amyloid aggregation and provides a deep insight into the processes of folding, unfolding and aggregation of TTR.

The laboratory of Prof. Chiti at the Section of Biochemistry of the University of Florence has exploited fluorescence resonance energy transfer (FRET) experiments in TTR, labelled with a coumarin derivative, N-(7-dimethylamino-4-methylcoumarin-3-yl) maleimide, (DACM).

My task aimed to structurally characterize the labelled WT and M-TTR variants providing the exact position of DACM to better interpret the FRET data and assess if the DACM labelling interferes with the native TTR fold. Both TTR variants had been crystallized, DACM-WT-TTR and DACM-M-TTR are perfectly superimposable onto the non-labelled ones, demonstrating that the bound DACM does not perturb the native fold of the protein, thus validating FRET results.

Cell. Mol. Life Sci. (2017) 74:3577–3598
DOI 10.1007/s00018-017-2533-x

Cellular and Molecular Life Sciences

ORIGINAL ARTICLE

FRET studies of various conformational states adopted by transthyretin

Seyyed Abolghasem Ghadami[1] · Francesco Bemporad[1] · Benedetta Maria Sala[2] · Guido Tiana[3] · Stefano Ricagno[2] · Fabrizio Chiti[1]

Received: 7 December 2016 / Revised: 15 April 2017 / Accepted: 2 May 2017 / Published online: 6 May 2017

Abstract Transthyretin (TTR) is an extracellular protein able to deposit into well-defined protein aggregates called amyloid, in pathological conditions known as senile systemic amyloidosis, familial amyloid polyneuropathy, familial amyloid cardiomyopathy and leptomeningeal amyloidosis. At least three distinct partially folded states have been described for TTR, including the widely studied amyloidogenic state at mildly acidic pH. Here, we have used fluorescence resonance energy transfer (FRET) experiments in a monomeric variant of TTR (M-TTR) and in its W41F and W79F mutants, taking advantage of the presence of a unique, solvent-exposed, cysteine residue at position 10, that we have labelled with a coumarin derivative (DACM, acceptor), and of the two natural tryptophan residues at positions 41 and 79 (donors). Trp41 is located in an ideal position as it is one of the residues of β-strand C, whose degree of unfolding is debated. We found that the amyloidogenic state at low pH has the same FRET efficiency as the folded state at neutral pH in both M-TTR and W79F-M-TTR, indicating an unmodified Cys10–Trp41 distance. The partially folded state populated at low denaturant concentrations also has a similar FRET efficiency, but other spectroscopic probes indicate that it is distinct from the amyloidogenic state at acidic pH. By contrast, the off-pathway state accumulating transiently during refolding has a higher FRET efficiency, indicating non-native interactions that reduce the Cys10–Trp41 spatial distance, revealing a third distinct conformational state. Overall, our results clarify a negligible degree of unfolding of β-strand C in the formation of the amyloidogenic state and establish the concept that TTR is a highly plastic protein able to populate at least three distinct conformational states.

Keywords Protein aggregation · Protein misfolding · Protein folding · Folding intermediate · SSA · FAP · FAC

Abbreviations

FRET	Fluorescence resonance energy transfer
TTR	Transthyretin
M-TTR	Monomeric variant of TTR
RBP	Retinol binding protein
CSF	Cerebrospinal fluid
SSA	Senile systemic amyloidosis
FAP	Familial amyloid polyneuropathy
FAC	Familial amyloid cardiomyopathy
DMSO	Dimethyl sulfoxide
GSH	Glutathione
TCEP	Tris(2-carboxyethyl)phosphine hydrochloride
TFA	Trifluoroacetic acid
DACM	*N*-(7-Dimethylamino-4-methylcoumarin-3-yl)maleimide
DLS	Dynamic light scattering
MD	Molecular dynamics
AU	Asymmetric unit
CD	Circular dichroism

✉ Fabrizio Chiti
fabrizio.chiti@unifi.it

[1] Dipartimento di Scienze Biomediche Sperimentali e Cliniche "Mario Serio", Sezione di Scienze Biochimiche, Università degli Studi di Firenze, Viale Morgagni 50, 50134 Florence, Italy

[2] Dipartimento di Bioscienze, Università degli Studi di Milano, 20133 Milan, Italy

[3] Center for Complexity and Biosystems, Department of Physics, Università degli Studi di Milano and INFN, via Celoria 16, 20133 Milan, Italy

Introduction

Transthyretin (TTR) is a homotetrameric protein with a total molecular mass of 55 kDa that is mainly synthesized in the liver, choroid plexus of the brain, and retina [1, 2], albeit further studies have demonstrated significant synthesis in the pancreas, kidneys, Schwann cells and neurons [3–7]. In the plasma TTR transports thyroxine (T4) and the retinol binding protein (RBP), whereas in the cerebrospinal fluid (CSF) TTR is the primary transporter of T4 [8, 9].

TTR is also one of the few proteins associated with systemic amyloidoses, a group of disorders resulting from the extracellular deposition of well-defined protein aggregates characterized by a fibrillar morphology, cross-β structure and peculiar tinctorial properties in the presence of specific dyes such as Congo red or thioflavin S [10, 11]. Amyloid deposition of wild-type TTR occurs in the heart of 10–25% of humans older than 80 years, resulting in senile systemic amyloidosis (SSA), often leading to congestive heart failure [12, 13]. Amyloid deposition by TTR is accelerated by the presence of any of the approximately 100 different amyloidogenic mutations of TTR responsible for early-onset TTR amyloidosis with autosomal dominant inheritance, such as familial amyloid polyneuropathy (FAP), familial amyloid cardiomyopathy (FAC) and leptomeningeal amyloidosis [14, 15]. A list of all TTR mutations associated with such disorders is reported in a database named "Mutations of Hereditary Amyloidosis" (http://www.amyloidosismutations.com/mut-attr.php).

While the main component of amyloid fibrils in patients suffering from sporadic SSA is wild-type TTR and its fragments, the main component of amyloid fibrils in TTR familial amyloidoses is mutant TTR, with the wild-type counterpart only scarcely represented [10, 16–18]. The single-site mutations that cause such diseases do not significantly change the tertiary or quaternary structure of tetrameric TTR; instead they appear to be responsible for the reduced thermodynamic and/or kinetic stability of native tetrameric TTR in favor of the monomeric, partially unfolded amyloidogenic intermediate, which self-assembles into amyloid fibrils under mildly denaturing conditions, at pH 3.9–5.0 [19–21].

Considerable effort has been expended to determine the structure of the monomeric amyloidogenic state of TTR populated at weakly acidic pH values. Biophysical and limited proteolysis experiments have led to the conclusion that upon acidification, tetrameric TTR dissociates into monomers with the subsequent unfolding of the β-strand C, the β-strand D and the connecting CD loop [22, 23]. In another report it was proposed, on the basis of hydrogen/deuterium exchange experiments coupled to NMR, that the β-sheet CBEF is the most destabilized upon acidification of TTR at pH 4.5 [24]. Significant advances have been reached with a double mutant of TTR carrying two single mutations (F87M/L110M) designed to destabilise the molecular interface between the subunits that compose the tetramer [25]. This double mutant, generally referred to as M-TTR, was shown to be stable as a monomer at neutral pH and has proved a unique tool for characterizing, at the molecular level, the transition from the folded non-amyloidogenic monomer populated at neutral pH to the amyloidogenic state populated at low pH [25]. Solution NMR studies have shown that M-TTR is folded at neutral pH, with many of the resonances of the β-sheet DAGH exhibiting line broadening, particularly the β-strand H, indicative of structural fluctuations [26]. By contrast, the β-sheet CBEF is not just folded, but also substantially packed. Upon acidification to pH values promoting amyloid fibril formation, line broadening extends to the β-strand D, the D–E loop, the E–F helix, and the residues of the A–B loop forming interactions with the helix, whereas most residues of the β-sheet CBEF maintain sharp, non-broadened resonances [26]. These observations have been confirmed later on tetrameric TTR at mildly low pH [27]. Hence, it is not yet clear if the partial unfolding of the monomeric folded unit of TTR to produce the amyloidogenic state involves the β-strands C and D and the interconnecting loop [22, 23], the CBEF β-sheet [24] or only the peripheral portion of the DAGH β-sheet [26].

In an independent study, the folding process of M-TTR was characterized kinetically and at equilibrium using a number of biophysical probes, leading to the identification of a molten globule state populated at equilibrium at low urea concentrations and to an off-pathway partially folded state populated transiently during the folding process of M-TTR [28]. The correspondence between any of such conformational states and the amyloidogenic state of M-TTR populated at weakly acidic pH has not yet been established.

To clarify these unresolved issues, we have carried out fluorescence resonance energy transfer (FRET) experiments, taking advantage of the presence of a unique, solvent-exposed, cysteine residue at position 10 and of the two naturally occurring tryptophan residues at positions 41 and 79 in M-TTR. We have labelled the Cys10 residue of M-TTR with a coumarin derivative that acts as a FRET acceptor of the intrinsic fluorescence emitted by the Trp41 and Trp79 residues that thus act as FRET donors. Trp41 is located in an ideal position as it is one of the residues of β-strand C which has been reported to be either unfolded or fully folded in the various structural characterizations described above. Cys10 is solvent-exposed and its labelling is expected to have only minor influence on the structure of M-TTR. Hence, labelled M-TTR is expected to contain a

new solvent-exposed environmental probe at position 10, thus permitting the measurement of the spatial distance between the probe linked to Cys10 and Trp41/Trp79. We have also produced and labeled with coumarin two single mutants of M-TTR, namely W41F and W79F, to monitor the spatial distance between Cys10 and the two tryptophan residues separately. The utility of FRET results from the strong dependence of the efficiency of energy transfer on the sixth power of the spatial distance between donor and acceptor, which has led to the often quoted reference to FRET as a 'spectroscopic ruler' [29].

We will report the FRET efficiency values of the various labelled mutants under various experimental conditions to shed light on the structure of the various conformational states of M-TTR so far characterized, including the fully folded and fully unfolded states at neutral pH, the amyloidogenic state at low pH, the equilibrium partially folded state at low urea concentrations and the kinetically trapped state transiently populated during folding.

Materials and methods

Materials

Dimethyl sulfoxide (DMSO), glutathione (GSH), Tris(2-carboxyethyl)phosphine hydrochloride (TCEP), trifluoroacetic acid (TFA) and urea were purchased from Sigma-Aldrich (St Louis, MO, USA). *N*-(7-Dimethylamino-4-methylcoumarin-3-yl)maleimide (DACM) was purchased from Thermo Fisher Scientific (Waltham, MA, USA).

Protein expression, purification and mutagenesis

Production of M-TTR and its mutants was carried out as previously described [9], except for the bacterial growth temperature of the W41F-M-TTR which was an overnight incubation at 25 °C. Purified proteins were stored at −20 °C in 20 mM phosphate buffer, pH 7.4. Mutations in the gene coding for M-TTR were generated using the QuickChange site-directed mutagenesis kit (Agilent Technologies, Santa Clara, CA, USA). The DNA sequences of the wild-type and mutated genes were checked with DNA sequencing. The molecular masses of the purified variants were checked with MALDI mass spectrometry. Protein purity was found by SDS-PAGE to be >95% in all cases.

Labeling with DACM

Each protein variant was diluted to 0.2 mM in 20 mM phosphate buffer at pH 7.4, 25 °C. Aliquots of DACM dissolved in pure DMSO were added to a tenfold molar excess of dye. The sample was wrapped with aluminium foil and incubated under shaking for 1 h at 37 °C. The reaction was quenched with 5 µl of TFA. The unbound dye was removed by extensive dialysis, using membranes with a 3.0 kDa molecular weight cut off, and the sample was then centrifuged to remove any precipitate. DACM concentration of the resulting labelled protein sample was determined using $\varepsilon_{381} = 27{,}000\ M^{-1}\ cm^{-1}$. Protein concentration was measured at 280 nm using $\varepsilon_{280} = 18{,}450\ M^{-1}\ cm^{-1}$ for labelled M-TTR, $\varepsilon_{280} = 12{,}950\ M^{-1}\ cm^{-1}$ for labelled W79F and W41F-M-TTR, and after subtraction of the contribution of an equimolar concentration of DACM-GSH. An $\varepsilon_{280} = 77{,}600\ M^{-1}\ cm^{-1}$ for labelled WT-TTR was used as previously reported [19], To this purpose, the optical absorption spectra of DACM-GSH and DACM-M-TTR (or its mutants) were normalised to match identical maximum coumarin absorbance and have fully comparable spectra.

Dynamic light scattering (DLS) measurements

Samples were prepared at a final protein concentration of 15 µM in 20 mM phosphate buffer, pH 7.4, 25 °C or 20 mM acetate buffer, pH 4.4, 25 °C. Before the measurements, the protein samples were filtered with Anotop filters having a cutoff of 20 nm (Whatman, Little Chalfont, UK). DLS measurements were performed using a Zetasizer Nano S device from Malvern Instruments (Malvern, Worcestershire, UK) thermostated with a Peltier system. Low-volume 10 × 4 mm disposable cells were used. The values of refractive index and viscosity set on the instrument were determined using the software provided with the instrument, based on the information of buffer and temperature provided by the user. The presented size distributions were the average of three consecutive measurements.

X-ray crystallography

DACM-WT-TTR and DACM-M-TTR were concentrated and buffer-exchanged at 4 °C with Amicon Ultra-4 Centrifugal Filter Unit with Ultracel-3 membrane, Ultracel®-3K (Merck Millipore, Billerica, MA, USA) down to a final concentration of 12 mg ml^{-1}, in 100 mM KCl, 20 mM Sodium Phosphate, pH 7.4. Both were crystallized by sitting-drops technique at 20 °C: DACM-WT-TTR crystallized in 0.2 M of $CaCl_2$, 0.1 M of HEPES sodium, pH 7.5, 28% v/v polyethylene glycol 400; DACM-M-TTR crystallized in 0.2 M of $CaCl_2$, 0.1 M of sodium acetate, pH 4.6, 20% v/v 2-propanol. Both were cryoprotected adding glycerol to the crystallization solution (final concentration 30% v/v) and flash-frozen in liquid nitrogen. Diffraction data for DACM-WT-TTR and DACM-M-TTR were collected at the European Synchrotron Radiation

Facility (ESRF, Grenoble, France) at the ID23-1 and at the ID23-2 beam line to the resolution of 1.42 and 1.7 Å, respectively.

To assess if DACM was damaged during the X-ray data-collection, crystal absorption spectra were collected before and after the diffraction experiment at ID29S (ESRF): both spectra display the maximum absorption at 384 nm typical of DACM and their intensities are comparable ruling out major radiation damage (data not shown).

The diffraction data were integrated and processed using MOSFLM [30], POINTLESS and SCALA [31–33]. The crystal structure of DACM-WT-TTR was determined using BALBES [34] while that of DACM-M-TTR was determined by MOLREP [35] using the structure of M-TTR (pdb code: 1GKO [25]) as a search model. Refinement was performed using Refmac5 and Phenix Refine [36, 37]. Manual model building, visual inspection, addition of water molecules were carried out with Coot [38]. Images of all TTR structures were generated by CCP4mg [33].

Molecular dynamics (MD) simulations

M-TTR was modelled with the Amber99sb force field [39], while DACM was modelled with the GAFF force field [40]. Starting from the X-ray crystallographic structure of DACM-M-TTR, the system was solvated with 5171 TIP3P water molecules in a volume of 1750 $Å^3$. A MD simulation was carried out with Gromacs 4.5.5 for 20 ns at the temperature of 300 K, maintained by a Nosè–Hoover thermostat. Atomic coordinates and structure factors for the TTR variants DACM-WT-TTR and DACM-M-TTR have been deposited with the Protein Data Bank, with accession codes 5LLL and 5LLV, respectively.

Fluorescence spectroscopy

Fluorescence emission spectra (excitation 290 nm) were recorded using a PerkinElmer LS 55 spectrofluorimeter (Waltham, MA, USA) equipped with a thermostated cell holder attached to a Haake F8 water bath (Karlsruhe, Germany). For typical FRET measurements protein samples were diluted to a concentration of 3 μM in 0.5 ml, in 20 mM phosphate buffer, 0.0–7.8 M urea, pH 7.4 25 °C, unless stated otherwise. Pipetting of the protein mother solution for the various protein samples was carried out using a Handy Step from Brand (Wertheim, Germany) to secure accuracy. Each sample was analysed in triplicate (three samples) and three spectra were recorded in each case, implying that each spectrum is the average of nine spectra. For FRET measurements during aggregation protein samples were diluted to a concentration of 1, 3, 7, 14, 28 μM in 0.7 ml under the conditions reported in the "Turbidimetry" section. A 2 × 10 mm and a 4 × 10 mm quartz cuvette were used for the two sets of experiments, respectively.

Equilibrium urea unfolding

28–34 samples were prepared containing 3 μM protein in 20 mM phosphate buffer, pH 7.4, with urea concentrations ranging from 0 to 7.8 M and were incubated for 1 h at 25 °C. Pipetting of the protein mother solution for the various protein samples was carried out using a Handy Step from Brand (Wertheim, Germany) to secure accuracy. Each sample was analysed in triplicate (three samples) and three spectra were recorded in each case, implying that each spectrum is the average of nine spectra. Fluorescence spectra were acquired at 25 °C from 300 to 550 nm (excitation 290 nm) using the spectrofluorimeter, water bath and cuvette described above. Plots of fluorescence at a given wavelength (either 348, 362 or 462 nm) versus urea concentration were analyzed with the two-state unfolding model provided by Santoro and Bolen [25, 41] to obtain quantitative measurements of the free energy change upon denaturation in the absence of denaturant ($\Delta G^{U-F}_{H_2O}$), the dependence of the free energy change upon denaturation on urea concentration (m value) and the concentration of middle denaturation (C_m).

Stopped-flow measurements

Unfolding and refolding reactions were followed using a Bio-Logic (Claix, France) SFM-3 stopped-flow device equipped with an FC-08 cuvette, coupled to a fluorescence detection system and thermostated with a Haake F8 water bath (Karlsruhe, Germany). Excitation wavelength was 290 nm. Band-pass filters cutting emission below 385 and 320 nm were used to monitor DACM and tryptophan fluorescence, respectively. All the experiments were performed in 20 mM phosphate buffer at pH 7.4 and 25 °C, at final protein concentrations of 1.5–2.9 μM. For the unfolding experiments, native proteins in 0.5 M urea were diluted into solutions containing urea at final concentrations ranging from 3.0 to 6.5 M. Refolding reactions were initiated by 10- to 20-fold dilutions of the protein denatured with 5 M urea into solutions containing low urea at final concentrations ranging from 0.25 to 3.5 M. The dead time was generally 10.4 ms. Folding/unfolding kinetic traces were analysed as previously described [28]. In another set of experiments, the equilibrium signal measured at the end of unfolding kinetics was plotted versus urea concentration to linearly extrapolate the fluorescence of the labelled or unlabelled unfolded protein under native conditions (0.5 and 2.1 M urea).

Turbidimetry

15 μM M-TTR and 15 μM DACM-M-TTR were incubated in 20 mM acetate buffer, pH 4.4, 37 °C, in the presence of 0, 30, 60, 90 or 137 mM NaCl. 15 μM W79F-M-TTR and 15 μM DACM-W79F-M-TTR were incubated in 20 mM acetate buffer, pH 5.6, 37 °C, in the presence of 0, 30, 60, 90 or 137 mM NaCl. The time dependent development of turbidity at 450 nm was followed at 37 °C using a Jasco V-630 spectrophotometer (Tokyo, Japan), thermostated within ±0.1 °C by a Haake F8 water bath (Karlsruhe, Germany) and using a 10 mm path length cell. All turbidity values were blank subtracted.

Far-UV CD spectroscopy

Far-UV CD spectra of samples containing 16 μM non-mutated, W41F and W79F-M-TTR in 20 mM sodium phosphate buffer, pH 7.4, 25 °C were collected from 185 to 250 nm using a J-810 Spectropolarimeter from Jasco (Tokyo, Japan) equipped with a thermostated cell holder attached to a Thermo Haake C25P water bath (Karlsruhe, Germany). A 0.1 mm path-length cell (Hellma, Müllheim, Germany) was used. All spectra were blank subtracted and converted to mean residue ellipticity per residue [Θ].

Results

Purification and labeling of M-TTR

We first purified M-TTR and labelled it with *N*-(7-dimethylamino-4-methylcoumarin-3-yl)maleimide (DACM), as described in the "Materials and methods" section. The samples containing the purified unlabelled and labelled proteins were analysed with MALDI mass spectrometry (Fig. 1a). A single peak at 13,895 ± 10 and 14,193 ± 10 Da were detected for unlabelled and labelled M-TTR, respectively, in agreement with the expected molecular weights of the two protein samples, i.e. 13,894.7 and 14,193.0 Da, respectively. In the mass spectrum of DACM-M-TTR we could not detect peaks corresponding to the molecular weights of M-TTR with double or triple DACM labelling, excluding the presence of M-TTR labelled with two or more DACM moieties. We could not detect any peak corresponding to unlabelled M-TTR, indicating that no detectable unlabelled protein remained at the end of the labeling reaction.

As an evidence that unlabelled and labelled M-TTR are monomeric, we assessed the size distributions of M-TTR, DACM-M-TTR and wild-type tetrameric TTR (WT-TTR) by means of dynamic light scattering (DLS) at pH 7.4, 25 °C (Fig. 1b). An apparent hydrodynamic diameter of 4.20 ± 0.11 nm was observed for M-TTR, which is consistent with the hydrodynamic diameter of monomeric folded M-TTR, as reported previously [25, 28, 42]. An apparent hydrodynamic diameter of 5.03 ± 0.10 nm was observed for DACM-M-TTR. This diameter is higher than that of unlabelled M-TTR, possibly due to the presence of the covalently attached and solvent-exposed DACM moiety. A diameter of 5.03 ± 0.10 nm is, however, significantly lower than that measured for wild-type tetrameric TTR (6.17 ± 0.02 nm) and of that expected for labelled dimeric M-TTR (>5.2 nm) and labelled tetrameric M-TTR (>6.5 nm) based on purely geometric considerations and of the diffusion of dimeric and tetrameric species in solution in the three dimensions. Overall, the DLS results indicate that both M-TTR and DACM-M-TTR are monomeric.

Next, to gauge the labelling degree of DACM-M-TTR, we acquired optical absorption spectra of DACM-M-TTR and of DACM covalently attached to glutathione (DACM-GSH) at pH 7.4, 25 °C and at the same probe concentration (Fig. 1c). The concentration of DACM bound to M-TTR was determined by measuring the optical absorption at 381 nm (A_{381}), using a molar extinction coefficient (ε_{381}) of 27,000 cm^{-1} M^{-1} (Fig. 1c, blue spectrum). Protein optical absorption in the DACM-M-TTR sample was then reconstructed by subtracting the spectrum of DACM-GSH (Fig. 1c, green spectrum) from that of DACM-M-TTR sample (Fig. 1c, blue spectrum). In the resulting difference spectrum (Fig. 1c, red spectrum), protein concentration was measured at 280 nm (A_{280}), using a molar extinction coefficient (ε_{280}) of 18,450 cm^{-1} M^{-1} (an ε_{280} value of 12,950 cm^{-1} M^{-1} was used for the W79F mutant described below). Labelling degree d was then calculated using:

$$d(\%) = 100 \times (A_{381}/\varepsilon_{381})/(A_{280}/\varepsilon_{280}). \quad (1)$$

The spectroscopic investigation described below was carried out with samples having $d \geq 95\%$.

The fluorescence spectrum of M-TTR at pH 7.4, 25 °C (excitation at 290 nm) is dominated by a major large band at 330–350 nm (Fig. 1d, red spectrum), resulting largely from Trp41 as the fluorescence of Trp79 is largely quenched [23]. The fluorescence spectrum of an equimolar quantity of DACM-GSH, acquired under the same conditions (excitation at 290 nm), is very low in intensity and negligible (Fig. 1d, green spectrum). By contrast, the fluorescence spectrum of an equimolar quantity of DACM-M-TTR under the same conditions (excitation at 290 nm) is characterized by a significant decrease of tryptophan fluorescence intensity, relative to the spectrum of M-TTR, and a substantial enhancement of the DACM emission intensity, relative to the spectrum of DACM-GSH, indicating that FRET is occurring. FRET is largely due to the energy transfer from Trp41 to DACM, both because the

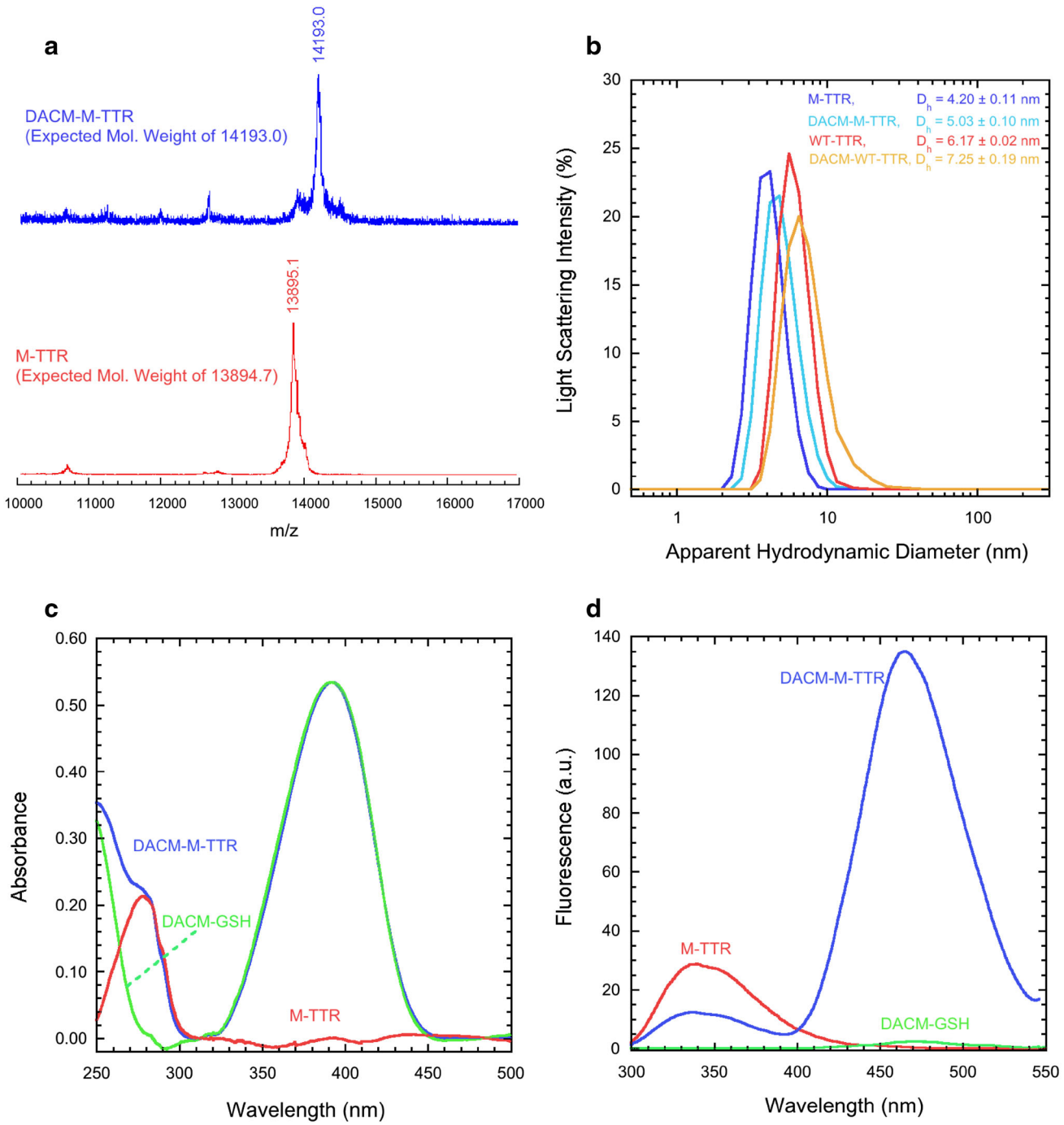


Fig. 1 Labelling of M-TTR. **a** MALDI mass spectrometry analysis of M-TTR (*red*) and DACM-M-TTR (*blue*). The expected molecular weights for human M-TTR and DACM-M-TTR are 13,894.7 and 14,193.0, respectively. **b** Size distributions of M-TTR (*blue*), DACM-M-TTR (*pale blue*), WT-TTR (*red*) and DACM-WT-TTR (*orange*) samples obtained with DLS at pH 7.4, 25 °C. **c** Optical absorption spectra of DACM-M-TTR (*blue*) and DACM-GSH (*green*) at pH 7.4, 25 °C and at the same probe concentration. The difference spectrum obtained by subtracting the latter from the former is also shown (*red*). **d** Fluorescence spectra of M-TTR (*red*), DACM-M-TTR (*blue*) and DACM-GSH (*green*) at pH 7.4, 25 °C (excitation 290 nm)

fluorescence of Trp79 is quenched and because Trp79 is more distant from DACM compared to Trp41. Importantly, the fluorescence intensity of the tryptophan band is not completely cancelled, leading to a FRET efficiency value (E) intermediate between 0 (no FRET) and 1 (full FRET), making it possible to monitor changes of the distance between donor (Trp41) and acceptor (DACM) by FRET measurements.

Overall, these data indicate that DACM-M-TTR is monomeric, singly labelled with an efficiency close to 100% and with donor/acceptor distance ideal for FRET measurements.

X-ray crystal structures of DACM-M-TTR and DACM-WT-TTR

To determine if DACM labeling on Cys10 affects M-TTR structure, the X-ray crystal structures of DACM-WT-TTR and DACM-M-TTR were determined at high resolution and compared with that of WT-TTR and M-TTR previously deposited in the Protein Data Bank [25, 43]. Data collection and refinement statistics are reported in Table 1. In both structures, from Cys10 to Asn124, all polypeptide chains are well traceable into the electron density map and, in the case of DACM-M-TTR, the F87M/L110M mutations are clearly detectable. As observed in most of the previous TTR structures, the first nine residues are flexible and not present in the electron density. In both DACM-WT-TTR and DACM-M-TTR, the presence of some extra density around Cys10 suggests the presence of DACM. The quality of such density is too poor to allow the building of DACM in the model, suggesting a high degree of conformational flexibility of the DACM moiety.

DACM-WT-TTR and DACM-M-TTR crystallised in $P2_122_1$ and $P2_12_12_1$ space group, respectively: DACM-WT-TTR has two molecules in the asymmetric unit (AU), whereas the AU of the DACM-M-TTR structure contains four molecules, as already observed in the M-TTR structure [25]. Superimposition of the DACM-labelled variants with the corresponding non-labeled variants shows that the structures of the DACM-WT-TTR and DACM-M-TTR are perfectly superimposable onto those of WT-TTR and M-TTR, respectively, indicating that DACM labeling does not interfere with the TTR native fold in the crystal (Fig. 2a, b; Table 1).

As mentioned above, DACM was not traceable in the electron density, due to its high mobility. Thus, MD simulations of the solvated DACM-M-TTR were performed, as described in the Materials and Methods, to provide the correct distribution of the DACM conformers, weighted according to Boltzmann statistics. The DACM moiety was not found to be randomly distributed in space during the simulation, with certain conformations more populated than others (Fig. 2c). The mean Trp41–DACM and Trp79–DACM spatial distances, calculated over the centers of mass, were found to be 23.2 ± 2.0 and 26.2 ± 2.0 Å, respectively (Fig. 2d).

Table 1 Data collection and refinement statistics for DACM-WT-TTR and DACM-M-TTR

Crystal[a]	DACM-WT-TTR	DACM-M-TTR
Space group	P 2_122_1	P $2_12_12_1$
Unit cell constants (Å)	$a = 43.68$, $b = 64.49$, $c = 85.27$	$a = 64.44$, $b = 83.76$, $c = 86.53$
	$\alpha, \beta, \gamma = 90$ °C	$\alpha, \beta, \gamma = 90$ °C
Resolution (Å)	22.35–1.42 (1.50–1.42)	22.04–1.70 (1.79–1.70)
R_{merge} (%)[b]	9.7 (73.0)	11.1 (77.2)
$I/\sigma I$	8.1 (2.1)	8.1 (1.9)
Completeness (%)	99.1 (99.2)	99.5 (99.8)
Multiplicity	5.0 (5.0)	4.1 (4.3)
Molecules per asymmetric unit	2	4
Refinement		
R_{work} (%)[c]	16.0	15.0
R_{free} (%)	21.7	22.3
Ramachandran plot, *n* (residues)		
Most favored region	200	421
Allowed region	3	8
Outliers	0	2
RMSD[d]	0.31 Å 114/114 Cα	0.60 Å 115/115 Cα

[a] Values given in parenthesis refer to the high-resolution shell

[b] $R_{merge} = \Sigma hkl\ \Sigma jIhkl,j - \langle Ihkl\rangle / \Sigma hkl \Sigma jIhkl,j$ where I is the observed intensity and $\langle I\rangle$ is the average intensity

[c] $R_{work} = \Sigma hklFo - Fc/\Sigma hklFo$ for all data except 5–10%, which were used for the R_{free} calculation

[d] RMSD values calculated from the structural superposition of the DACM-WT-TTR and DACM-M-TTR structures with the non-labeled WT-TTR (pdb code: 1BMZ) and M-TTR (pdb code: 1GKO)

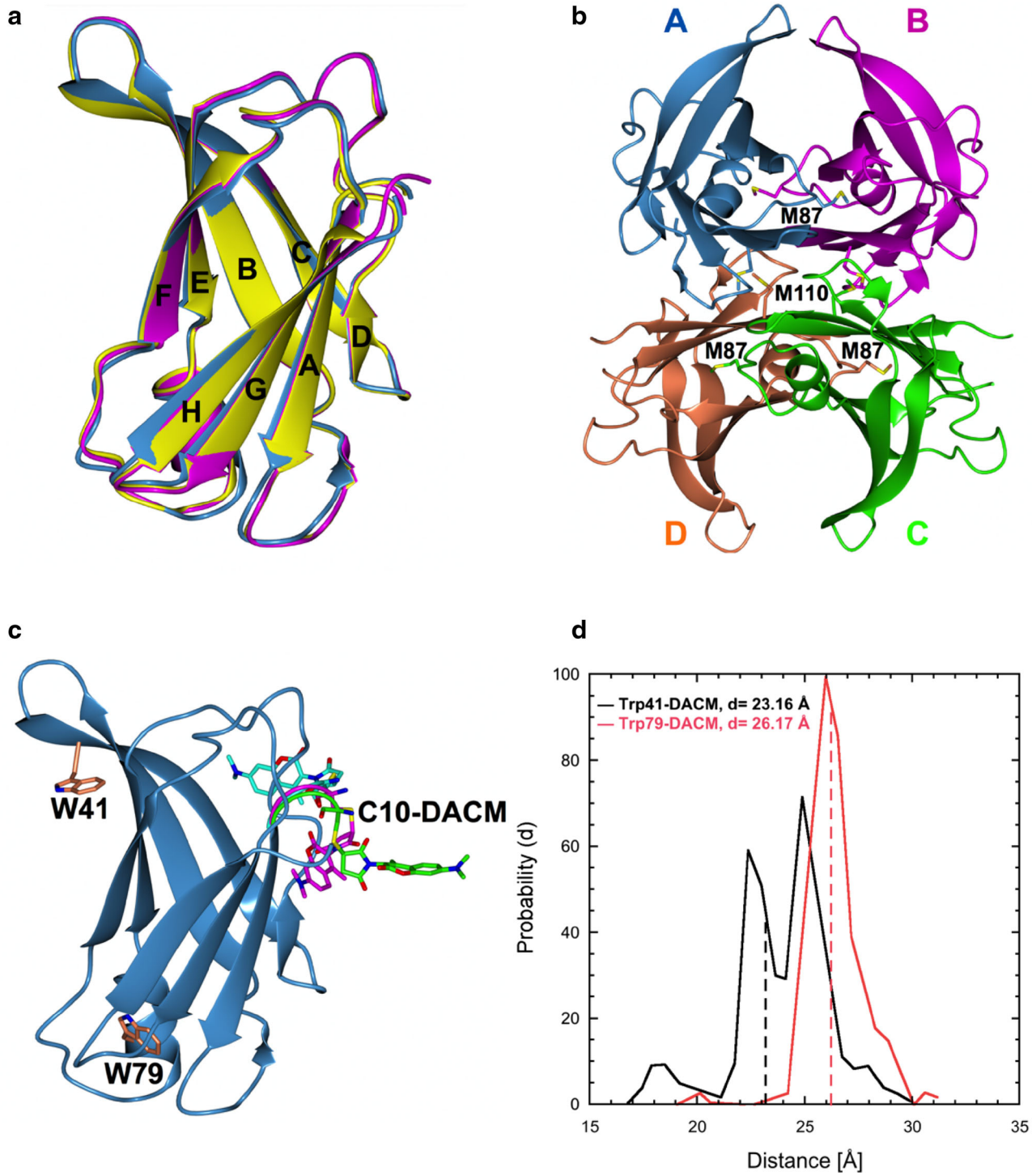


Fig. 2 X-ray crystal structures of DACM-M-TTR and DACM-WT-TTR. **a** Superposition of the structures of DACM-M-TTR (*blue*), DACM-WT-TTR (*magenta*) and WT-TTR (*yellow*, pdb code: 1BMZ [43]). **b** Tetrameric structure of DACM-M-TTR. The residues Met87 and Met110, that destabilize the tetrameric structure in solution, are shown as *sticks*. **c** MD simulations of DACM-M-TTR: Trp41, Trp79 and the most populated conformation of Cys10-DACM are shown as *sticks*. **d** Distributions of the spatial distances between the centers of mass of Trp41 and DACM (*black*) and those of Trp79 and DACM (*red*), calculated using MD simulations starting from the crystal structure of DACM-M-TTR. The *dashed lines* indicate the mean spatial distances assuming a single average conformer for the DACM moiety

FRET efficiency of native M-TTR and Förster distance of the Trp/DACM pair

We first determined the FRET efficiency of DACM-M-TTR under conditions in which the protein is folded, at pH 7.4, 25 °C. To this aim, we mixed different volumes of samples containing M-TTR and DACM-M-TTR, each at the same concentration of 3 μM, to achieve different percentages of labelled M-TTR, from 0 to 100%. Fluorescence emission spectra were obtained for all the resulting samples using an excitation wavelength of 290 nm (Fig. 3a). The emission spectrum of unlabeled M-TTR (0% DACM-M-TTR) showed an intense peak at 330–350 nm, resulting from tryptophan fluorescence emission, and no emission at 465 nm due to the absence of the DACM moiety (Fig. 3a, red spectrum). By contrast, the emission spectrum of labeled M-TTR (100% DACM-M-TTR) shows a low intensity peak at 330–350 nm and an intense peak at 465 nm, due to the energy transfer from the excited tryptophan residues to the DACM group, which fluoresces at 465 nm (Fig. 3a, blue spectrum). The samples presenting intermediate degrees of labelling have an intermediate behaviour, with the donor and acceptor peaks decreasing and increasing in intensity with labelling degree, respectively (Fig. 3a, red to blue spectra).

The FRET efficiency can be calculated using [44]:

$$E = 1 - \frac{F_{DA}}{F_D}, \tag{2}$$

where E is the FRET efficiency, F_{DA} and F_D are the fluorescence emission values of the donor in the presence and absence of acceptor, respectively, i.e. the tryptophan emission values at 348 nm measured for the samples containing 100 and 0% DACM-M-TTR, respectively. This led to $E = 0.70 \pm 0.02$. To achieve more accurate measurements of F_{DA}, F_D and E we plotted the tryptophan fluorescence emission at 348 nm versus the percentage of DACM-M-TTR and analyzed the resulting plot with a procedure of best fitting using a linear equation (Fig. 3b). This provided more accurate values of F_{DA} and F_D and a value of $E = 0.72 \pm 0.02$.

We then measured the fluorescence spectra of M-TTR and DACM-M-TTR at various protein concentrations ranging from 1 to 10 μM (Fig. 3c) and measured the E values at each of these concentrations (Fig. 3d). E was found to be independent of protein concentration, indicating that M-TTR and DACM-M-TTR remain largely monomeric under this range of protein concentration. In addition, the value of E remains lower than that measured for tetrameric DACM-WT-TTR (data not shown), corroborating the absence of protein tetramerisation under the conditions studied here.

To determine the distance between Trp41 (donor) and DACM (acceptor) in the folded monomer using the E value, we first determined experimentally the Förster distance using:

$$R_0^6 = 8.79 \times 10^{-5}\left(\kappa^2 n^{-4} Q_D J(\lambda)\right), \tag{3}$$

where R_0 is the Förster distance (in Å), κ^2 is the orientation factor describing the relative orientation of the transition dipole moments of the donor and the acceptor, n is the refractive index of the aqueous solution, Q_D is the quantum yield of an isolated donor, and $J(\lambda)$ is the integral expressing the degree of spectral overlap between donor emission and acceptor absorption (units of $nm^4\ M^{-1}\ cm^{-1}$) [44, 45]. A k^2 value of 0.87 was used as a good approximation of a donor–acceptor orientation factor of a folded protein [46], whereas n, Q_D and $J(\lambda)$ were determined experimentally (Table 2). This allowed the R_0 parameter to be determined for folded M-TTR as 24.9 ± 1.4 Å (Table 2). We then used these R_0 and E values determined experimentally for folded DACM-M-TTR to determine the donor–acceptor distance (R) using the equation [44]:

$$R^6 = R_0^6(1/E - 1), \tag{4}$$

where R is the distance between the centers of mass of Trp41 and DACM (in Å). This led to a value of R of 21.3 ± 1.6 Å, in good agreement with that determined from X-ray crystallography (R = 23.2 ± 2.0 Å).

FRET efficiency of urea-induced unfolded and molten globule states of M-TTR

Next, we acquired fluorescence spectra of M-TTR and DACM-M-TTR (excitation 290 nm) at urea concentrations ranging from 0 to 7 M at pH 7.4, 25 °C (Fig. 4a, b). The corresponding urea denaturation curves (spectroscopic signal versus urea concentration at equilibrium) were obtained using tryptophan fluorescence at 362 nm as a spectroscopic probe for both the labelled and unlabeled proteins (Fig. 4c). The analysis yielded values of conformational stability in the absence of denaturant ($\Delta G_{H_2O}^{U-F}$) of 20.4 ± 1.5 and 11.1 ± 1.5 kJ mol^{-1}, dependencies of $\Delta G_{H_2O}^{U-F}$ on urea concentration (m value) of 6.9 ± 0.5 and 4.3 ± 0.5 kJ $mol^{-1}\ M^{-1}$ and midpoint of denaturation (C_m) of 3.0 ± 0.1 and 2.6 ± 0.1 M for M-TTR and DACM-M-TTR, respectively. The values obtained for M-TTR are in reasonable agreement with those obtained previously under slightly different conditions [28]. The urea denaturation curve of DACM-M-TTR was also obtained using DACM fluorescence at 462 nm as a spectroscopic probe (Fig. 4d). This led to values of 12.4 ± 1.3 kJ mol^{-1}, 4.9 ± 0.4 kJ $mol^{-1}\ M^{-1}$ and

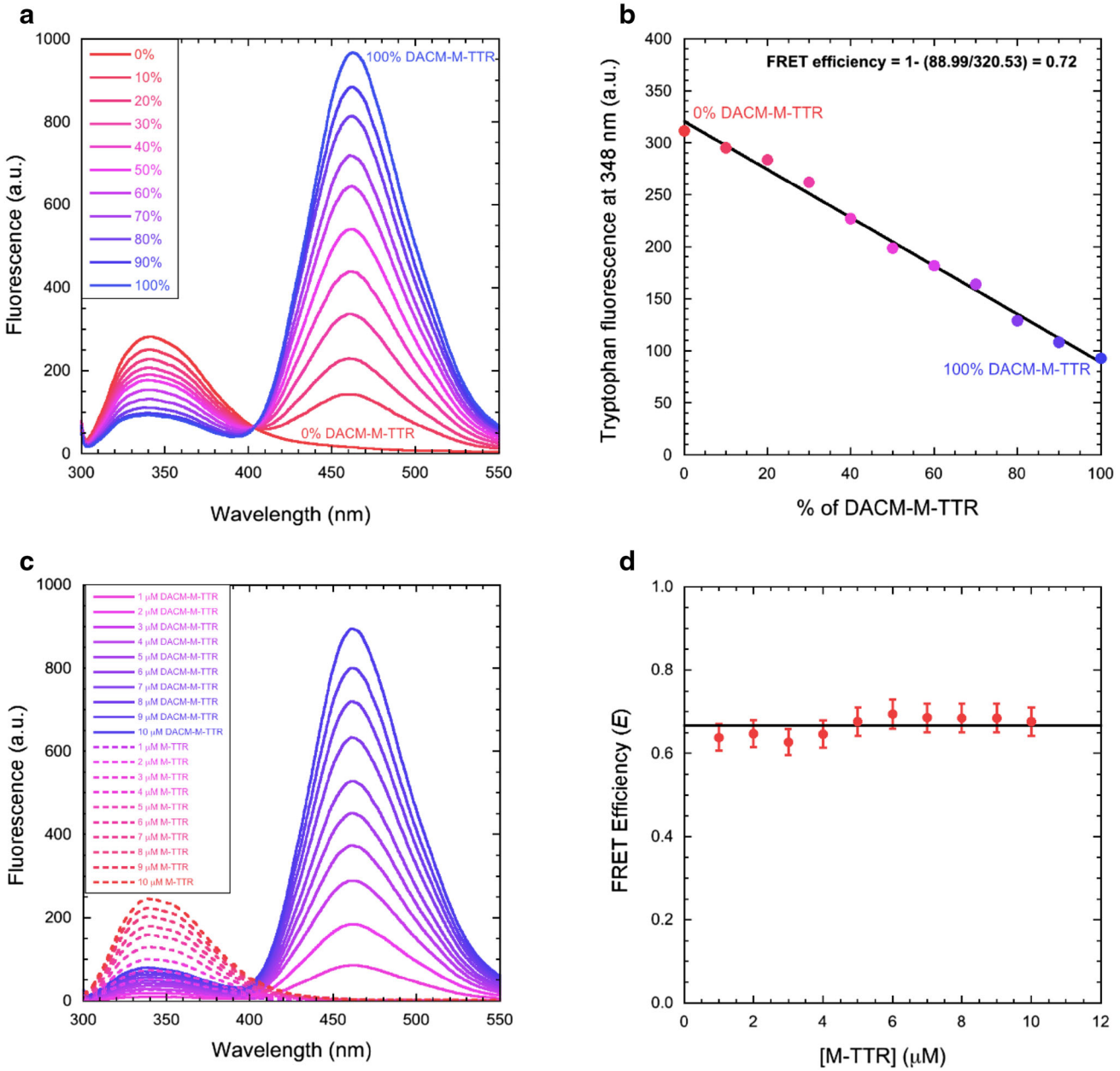


Fig. 3 FRET of native M-TTR. **a** Fluorescence spectra of mixtures of M-TTR and DACM-M-TTR at the indicated percentages of the latter, at 3 μM total protein concentration, pH 7.4, 25 °C. **b** Tryptophan fluorescence emission at 348 nm versus the percentage of DACM-M-TTR. The *straight line* represents the best fit of the data points to a linear function. The equation indicates how E was determined. **c** Fluorescence spectra of M-TTR (*dashed lines*) and DACM-M-TTR (*continuous lines*) at various protein concentrations ranging from 1 to 10 μM, pH 7.4, 25 °C (excitation 290 nm). **d** Plot of E versus M-TTR concentration. The *straight line* represents the average value

2.5 ± 0.1 M for $\Delta G^{U-F}_{H_2O}$, m and C_m, respectively, in agreement with those obtained using tryptophan fluorescence as a spectroscopic probe. This analysis indicates that the conformational stability of DACM-M-TTR is significantly lower than that of M-TTR, indicating that the labeling has an effect on protein conformational stability, as it is often observed for mutations.

In principle, one can obtain a urea denaturation curve by plotting E versus urea concentration. However, such a curve suffers from different conformational stabilities of labelled and unlabelled M-TTR which makes it difficult to compare fluorescence spectra for the two proteins at the same urea concentration and determine E at each urea concentration. In spite of these difficulties, it is possible to determine the E values at low (<1.2 M) and high (>4.8 M) urea concentrations when the major conformational transition has not yet and has completely occurred, respectively, for both labelled and unlabeled proteins. The

Table 2 FRET parameters of M-TTR adopting different conformational states

Conformational state	R (Å)	E	R_0 (Å)	Q_D	n	$J(\lambda)$ ($nm^4\ M^{-1}\ cm^{-1}$)	κ^2
Folded (0 M urea)	21.3 (±1.6)	0.72 (±0.02)	24.9 (±1.4)	0.050 (±0.008)	1.332 (±0.001)	1.96×10^{14} ($\pm 0.20 \times 10^{14}$)	0.87 (±0.2)
Unfolded (6 M urea)	30.8 (±1.7)	0.29 (±0.01)	26.5 (±1.4)	0.070 (±0.004)	1.384 (±0.002)	3.09×10^{14} ($\pm 0.15 \times 10^{14}$)	0.67 (±0.2)
Equilibrium partially folded (1 M urea)	24.0 (±1.6)	0.63 (±0.02)	26.2 (±1.3)	0.054 (±0.003)	1.358 (±0.001)	2.66×10^{14} ($\pm 0.13 \times 10^{14}$)	0.87 (±0.2)
Amyloidogenic partially folded (0 M urea, pH 4.5)	21.1 (±1.2)	0.73 (±0.02)	24.9 (±1.4)	0.049 (±0.003)	1.332 (±0.001)	2.03×10^{14} ($\pm 0.10 \times 10^{14}$)	0.87 (±0.2)

E values were determined experimentally from the fluorescence spectra of DACM-M-TTR and M-TTR, respectively. Q_D was determined experimentally using free tryptophan as a reference. n was determined experimentally using a 2WAJ ABBE bench refractometer from Optika Microscopes (Bergamo, Italy). $J(\lambda)$ was determined experimentally as the integral expressing the degree of spectral overlap between donor emission and acceptor absorption [44, 45]. κ^2 was assumed to be 0.87 ± 0.2 for folded proteins [46] and 2/3 (0.67 ± 0.2) for unfolded proteins [44]. R_0 values were determined using Eq. 3. R values were determined from E and R_0 values using Eq. 4

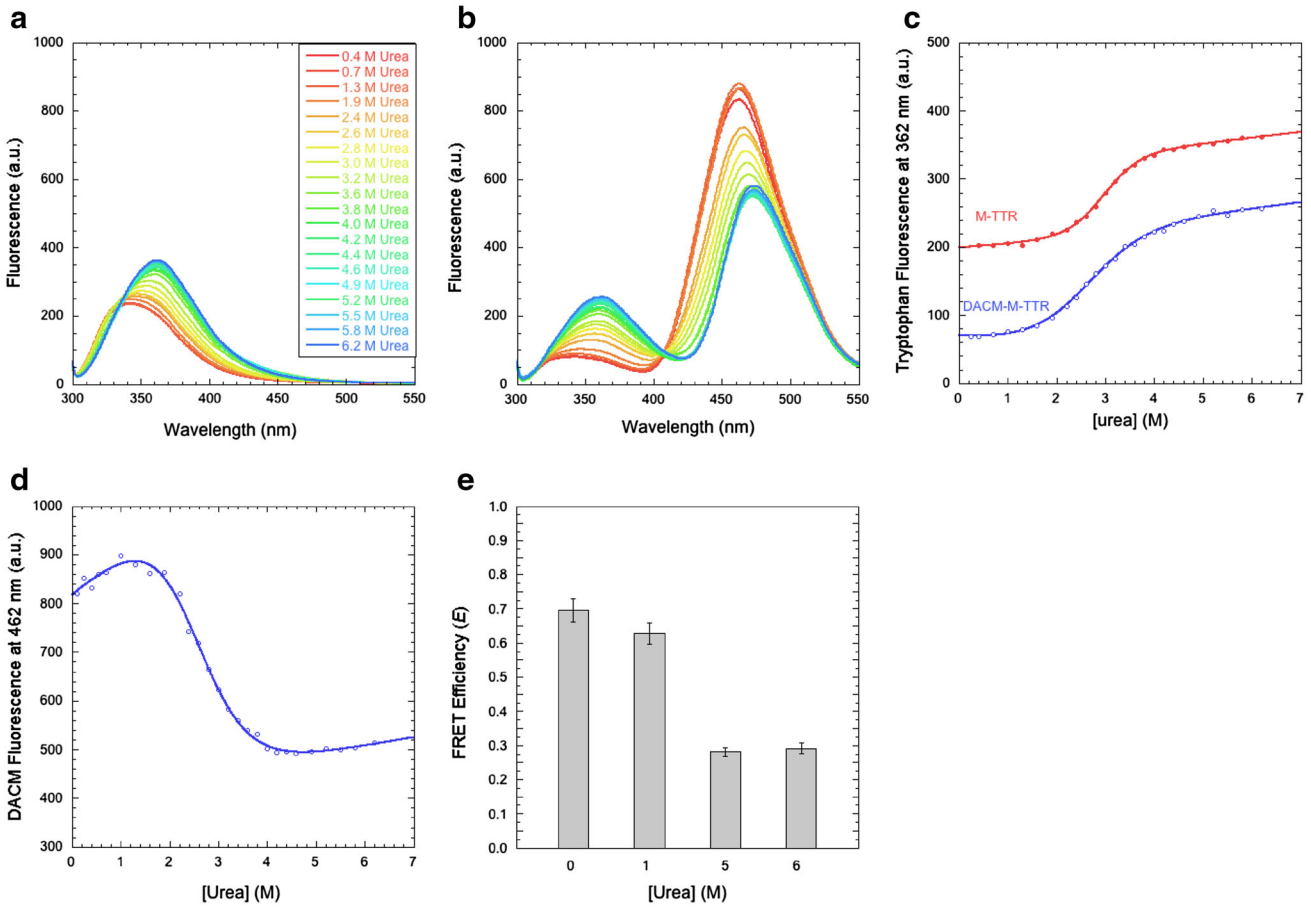


Fig. 4 FRET of urea-unfolded and molten globule states of M-TTR. **a**, **b** Fluorescence spectra of M-TTR (**a**) and DACM-M-TTR (**b**) at urea concentrations ranging from 0 to 6.2 M at pH 7.4, 25 °C (excitation 290 nm). **c** Urea denaturation curves (spectroscopic signal versus urea concentration) using tryptophan fluorescence at 362 nm as a spectroscopic probe for both M-TTR and DACM-M-TTR. **d** Urea denaturation curve of DACM-M-TTR, using DACM fluorescence at 462 nm as a spectroscopic probe. **e** FRET E values at the indicated urea concentrations, pH 7.4, 25 °C

FRET E values at 5.0 and 6.0 M urea, when DACM-M-TTR is unfolded, are remarkably lower than those determined at 0.0 and 1.0 M urea, indicating a significantly higher distance between the DACM moiety attached to Cys10 and the two tryptophan residues (Fig. 4e). The R_0 and R values, determined as described above using Eqs. 3

and 4, confirm a larger mean distance between donor and acceptor in the unfolded state (Table 2). The large decrease of the FRET E value upon unfolding indicates that such distance can be monitored with high sensitivity. The DACM fluorescence at 462 nm also decreases from 0.0–1.0 M to 5.0–6.0 M urea (Fig. 4d), confirming that the spatial distance between DACM and the two tryptophan residues increases upon unfolding and gives us the opportunity to observe the movement of DACM from Trp41 in the transition region in which the FRET E value could not be calculated.

Importantly, the FRET E value at 1.0 M urea is similar, within experimental error, to that determined at 0.0 M urea (Fig. 4e). This also results into a similar R value, determined as described above using Eqs. 3 and 4 (Table 2). The DACM fluorescence also does not change significantly (Fig. 4d). This indicates that the equilibrium molten globule state previously detected at low urea concentrations and having a transition at 0.0–2.4 M urea [28] has an unmodified distance between the DACM moiety and Trp41.

FRET efficiency of the transiently populated partially folded state of M-TTR

We then monitored the folding and unfolding processes of M-TTR and DACM-M-TTR in real time by diluting the urea-unfolded proteins into solutions containing low urea concentrations (folding) and by diluting the folded protein into solutions containing high urea concentrations (unfolding) using a stopped-flow device coupled to a fluorescence detection system. Representative folding traces recorded in 0.5 M urea, pH 7.4, 25 °C and monitored with intrinsic tryptophan fluorescence at 320–385 nm are shown for M-TTR and DACM-M-TTR (Fig. 5a, b). In addition to the major folding phase, two more low-amplitude exponential phases were observed for M-TTR, which were complete on a longer time scale, as observed previously [28]. Moreover, in addition to the major unfolding phase, one additional low-amplitude unfolding phase was apparent on a long time scale, again as observed previously [28]. This behaviour was also observed for the labelled protein.

The analysis of the resulting kinetic traces of folding and unfolding with a procedure of best fit using exponential functions allowed the rate constants for the major folding/unfolding transition (k) to be determined at the various urea concentrations (see "Materials and methods" for further details). These values were used to build the plot of ln(k) versus urea concentration, generally referred to as Chevron plot, for both the labelled and unlabelled proteins (Fig. 5c). The C_m values obtained from the analysis of these plots, calculated as the urea concentration at which extrapolated refolding and unfolding rate constants are equal (3.2 ± 0.2 and 2.2 ± 0.2 M for M-TTR and DACM-M-TTR, respectively), were in good agreement with those determined from equilibrium experiments (3.0 ± 0.1 and 2.5 ± 0.1 M, respectively). The downward curvature in the folding branch of the M-TTR plot results from a rapidly formed, off-pathway, partially folded species that forms rapidly and accumulates during the folding process [28]. Such a curvature was also observed for DACM-M-TTR (Fig. 5c). The refolding rate constant did not vary as the concentration of DACM-M-TTR ranged from 0.005 to 0.080 mg ml^{-1}, ruling out that the downward curvature arises from DACM-M-TTR aggregation during folding (data not shown). The extrapolated unfolding rate constants in the absence of denaturant were 1.82 ± 0.18 s^{-1} [28] and 6.73 ± 0.70 s^{-1} for M-TTR and DACM-M-TTR, respectively. Given the conformational stabilities obtained by equilibrium experiments, the folding rate constants under the same conditions would be $\sim$7100 s^{-1} [28] and $\sim$890 s^{-1}, respectively, on the assumption they were two-state folders. By contrast, the refolding rate constants extrapolated by experimental data were orders of magnitude lower than the calculated values, i.e. 81 ± 8 s^{-1} [28] and 60 ± 5 s^{-1} for M-TTR and DACM-M-TTR, respectively. This corroborates the hypothesis that a partially folded species accumulates for both M-TTR and DACM-M-TTR.

The spectroscopic signals at 320–385 nm recorded for M-TTR and DACM-M-TTR at the beginning and at the end of the kinetic traces in 0.5 M urea, pH 7.4, 25 °C (Fig. 5a, b) were thus used to calculate transient and equilibrium FRET E values of the folded and partially folded states under these conditions. These were found to be 0.66 ± 0.03 and 0.72 ± 0.03, indicating a slightly higher FRET E value for the partially folded state relative to the folded state. We also monitored the DACM fluorescence emission of such a partially folded state by inspecting the kinetic trace for folding in 0.5 M urea using a band-pass filter that cuts the signal below 385 nm, to monitor the fluorescence of the DACM moiety and exclude that of the tryptophan indole groups (Fig. 5d). Such a kinetic trace involved a decrease in DACM fluorescence, complete in $\sim$50 ms in 0.5 M urea, as DACM-M-TTR underwent refolding (Fig. 5d). The signal of the folded protein and that of the unfolded protein extrapolated down to 0.5 M urea by linear extrapolation from measurements at high urea concentrations, where the protein is 100% unfolded, were remarkably lower (Fig. 5d). This analysis suggests that the transient partially folded state of M-TTR has a DACM fluorescence higher not just than the unfolded state, but also than the fully folded state. The observation of an increase of FRET efficiency by monitoring both tryptophan and DACM fluorescence and using the same stopped-flow apparatus and experimental conditions

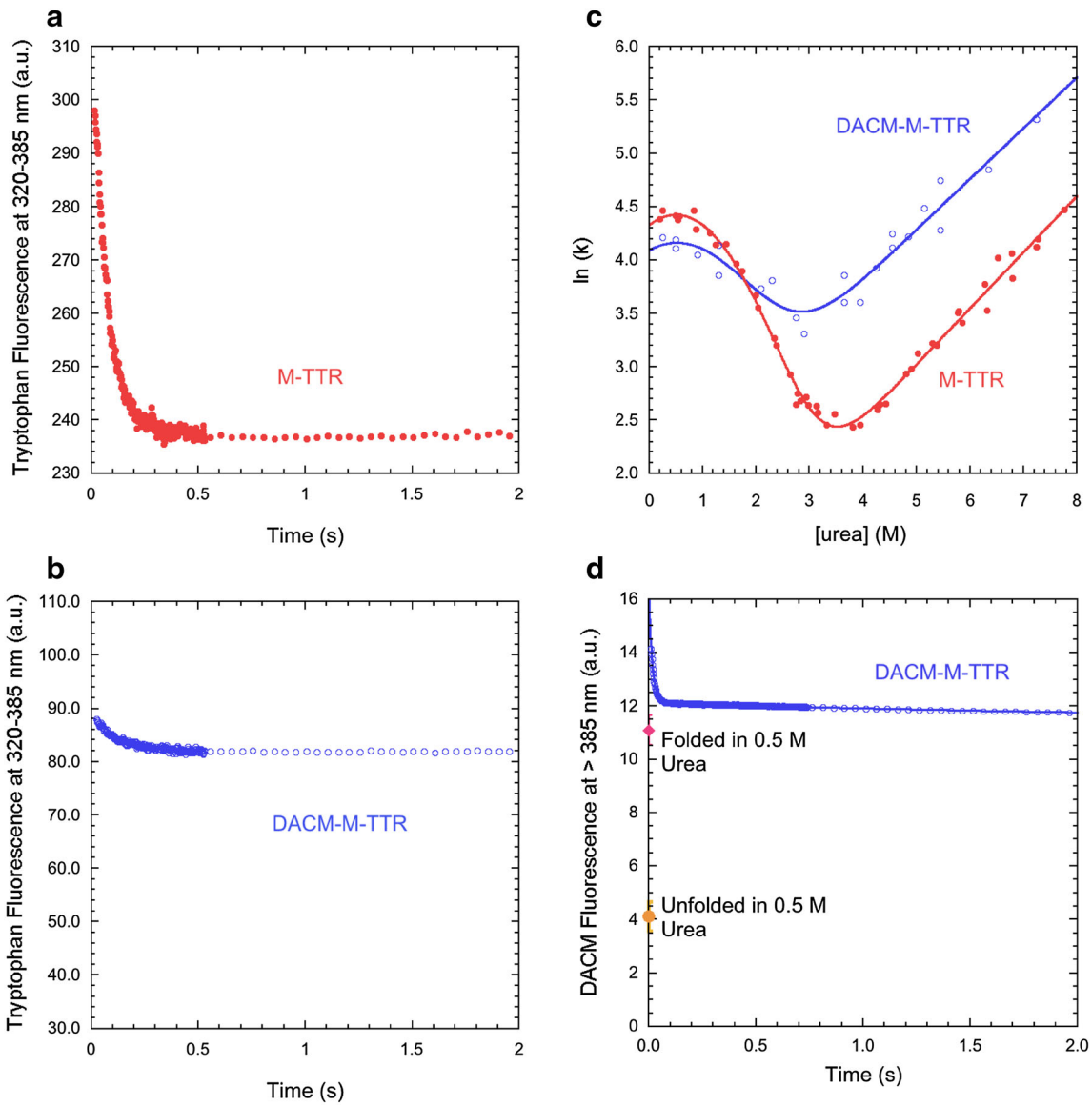


Fig. 5 FRET during M-TTR refolding. **a**, **b** Refolding time course of M-TTR (*red filled circles*) and DACM-M-TTR (*blue open circles*) monitored by tryptophan fluorescence (excitation 290 nm, emission 320–385 nm) in 0.5 M urea at pH 7.4, 25 °C. **c** Natural logarithm of the observed folding/unfolding microscopic rate constants for DACM-M-TTR (*blue open circles*) compared to data previously published for M-TTR (*red filled circles*) [28], plotted as a function of urea concentration (chevron plot), at pH 7.4, 25 °C. **d** Refolding time course of DACM-M-TTR (*blue open circles*) monitored by DACM fluorescence (excitation 290 nm, emission >385 nm) in 0.5 M urea at pH 7.4, 25 °C. The signals of folded and unfolded DACM-M-TTR are also indicated

indicates that the increase is significant. As a control, the kinetic trace for folding in the presence of 2.1 M urea, where the partially folded state cannot be detected, did not show such a burst-phase increase of DACM fluorescence (data not shown). Phenomena different from FRET may concur to the hyperfluorescence exhibited by DACM in the partially folded state, including a different local environment experienced by the coumarin moiety.

DACM fluorescence emission was also followed during unfolding. A decrease in DACM emission was observed, in agreement with an increase in distance between donor and acceptor during the process. As previously observed using intrinsic fluorescence as an optical probe [28], such decrease involved two phases, lasting ca. 0.1 and 30 s, respectively, in the trace recorded in 4.5 M urea taken here as an example (data not shown).

FRET efficiency of the amyloidogenic state of M-TTR

The FRET analysis was also used to investigate the structure of M-TTR under conditions in which the protein

is initially monomeric but slowly converts into amyloid fibrils, i.e. at pH 4.4, 37 °C [23, 25]. To check that these conditions of pH and temperature also promoted aggregation of DACM-M-TTR and identify conditions suitable for slow aggregation (so that the initial monomeric DACM-M-TTR can be monitored), DACM-M-TTR aggregation was studied at a protein concentration of 15 μM, in 20 mM acetate buffer, pH 4.4, 37 °C, under the effect of different ionic strengths, that is in the presence of 0, 30, 60, 90 and 137 mM NaCl. The time courses of turbidity at 450 nm show a gradual increase of the rate of turbidity development with the increase of ionic strength (Fig. 6a). The sigmoidal pattern of the intermediate time courses indicates that formation of amyloid fibrils involves the typical three phases of the amyloid fibril formation process: an initial lag phase, a subsequent growth phase and a final saturation phase (plateau). At 30 mM NaCl the time length of the first phase was relatively large but aggregation could still be observed, so that we selected this condition of ionic strength to study the aggregation of the protein.

DLS measurements confirmed that DACM-M-TTR was monomeric immediately after incubation at pH 4.4 (Fig. 6b). Indeed, the size distributions of M-TTR and DACM-M-TTR at pH 7.4, used here as references of monomeric conformations, had major peaks of hydrodynamic diameters at 4.20 ± 0.11 and 5.03 ± 0.10 nm, respectively, with the slightly higher increase of the latter due to the presence of the DACM moiety (see above). The size distributions acquired immediately after incubation at pH 4.4 were found to have major peaks at 4.65 ± 0.26 and 5.60 ± 0.30 nm, respectively (Fig. 6b). These were both slightly higher than those of the corresponding proteins at pH 7.4, most likely as a consequence of the partial unfolding of the proteins at pH 4.4 [23], but were still lower than those of partially unfolded dimers or higher order oligomers.

Hence, both M-TTR and DACM-M-TTR were incubated at 1, 3, 7, 14, 28 μM in 20 mM acetate buffer, 30 mM NaCl, pH 4.4, 37 °C, and fluorescence spectra were collected at different time points, to measure the FRET efficiency (*E*) values as aggregation proceeded (Fig. 6c–e). Since M-TTR and DACM-M-TTR aggregate with different rates under the selected conditions, the FRET *E* values are meaningful only immediately after incubation (time 0 s) and at the plateau of the aggregation process (time 3000 s) when both proteins are completely monomeric and aggregated, respectively. The FRET *E* value of the amyloidogenic state at 0 s from the five different kinetic traces and the resulting *R* value determined with Eqs. 3 and 4 were found to be 0.73 ± 0.02 and 21.1 ± 1.2 Å, respectively, which are very similar to those measured for the folded proteins at pH 7.4, 25 °C, indicating an unchanged distance between Trp41 and Cys10 upon partial unfolding at pH 4.4, 37 °C (Fig. 6f). The FRET *E* value at 3000 s was found to be 0.82 ± 0.01. This value is, however, difficult to interpret because both tryptophan and coumarin fluorescences are quenched in the aggregated state, the Förster distance can change following aggregation and is difficult to determine and a population of different aggregates may exist in the sample.

Fig. 6 FRET of the amyloidogenic and aggregated states of M-TTR. ▶ **a** Effect of different NaCl concentrations on the kinetics of 15 μM DACM-M-TTR aggregation in 20 mM acetate buffer, pH 4.4, 37 °C, monitored with turbidimetry at 450 nm. **b** DLS size distributions of M-TTR and DACM-M-TTR at both pH 7.4 and 4.4. **c**, **d** Fluorescence spectra at different time points for M-TTR (**c**) and DACM-M-TTR (**d**) in 20 mM acetate buffer, 30 mM NaCl, pH 4.4, 37 °C. **e** Change of *E* during aggregation. Conditions as in **c**, **d**. **f** FRET *E* values for native M-TTR at pH 7.4, 25 °C (*first bar*), amyloidogenic monomeric M-TTR at pH 4.4, 37 °C, 0 s (*second bar*) and aggregated M-TTR at pH 4.4, 37 °C, 3000 s (*third bar*)

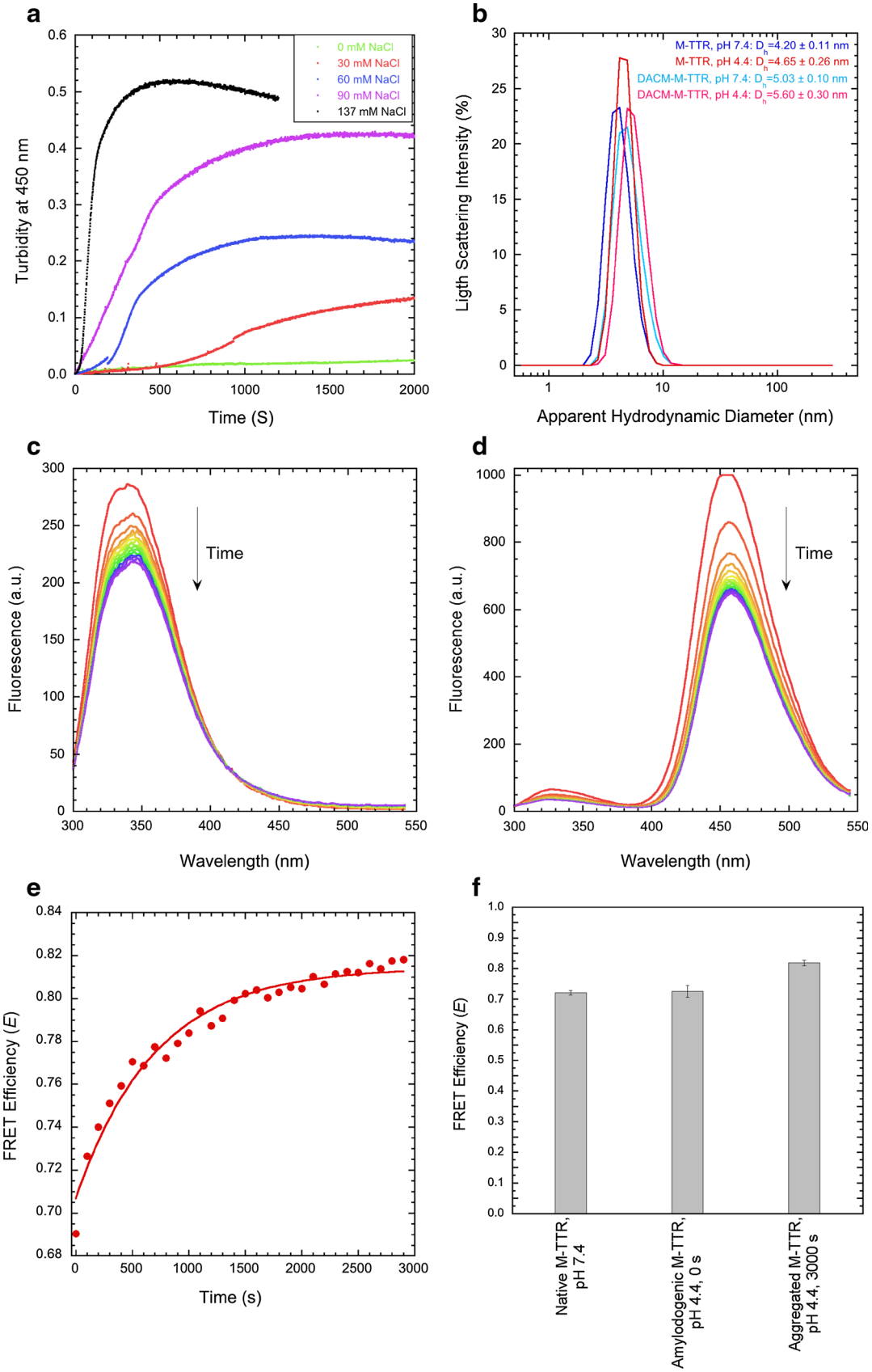
Purification and analysis of W41F and W79F mutants of M-TTR

Although Trp79 is quenched and more distant than Trp41 to Cys10 in the folded state, doubts remain in partially or fully unfolded states, where Trp79 may not be quenched and may be placed closer to Cys10 than Trp41. We therefore, inserted single point mutations in the gene encoding M-TTR and we purified and labelled two single mutants of M-TTR, replacing either Trp41 or Trp79 by a phenylalanine residue (W79F and W41F) and repeated the FRET experiments to assess the involvement of individual inter-residue distances in the overall measured FRET.

The W79F mutant has a far-UV circular dichroism (CD) spectrum very similar to that of non-mutated M-TTR in 20 mM phosphate buffer, pH 7.4, 25 °C, with a negative peak around 215 nm and a positive peak at 196 nm, typical of β-sheet structure (Fig. 7a). By contrast, the W41F mutant is remarkably different, with mean residue ellipticity peaks characterized by a low intensity, that is reminiscent of protein aggregation (Fig. 7a). The DLS analysis at pH 7.4, 25 °C confirmed that both W79F-M-TTR and DACM-W79F-M-TTR are monomeric, whereas both W41F-M-TTR and DACM-W41F-M-TTR are aggregated (Fig. 7b). As a further evidence that W41F-M-TTR aggregates rapidly under native conditions, the time courses of turbidity at 15 μM protein concentration, 20 mM phosphate buffer, 0–137 mM NaCl, pH 7.4, 37 °C indicate that W41F-M-TTR has a high turbidity at time 0 s in all kinetic traces under native conditions, including at 0 mM NaCl (Fig. 7c).

Comparison of the FRET efficiency values of mutant and non-mutated M-TTR

Since W79F-M-TTR, unlike the W41F variant, appears to be soluble and monomeric at pH 7.4, 25 °C, we repeated


a
Turbidity at 450 nm
Time (S)
0 mM NaCl
30 mM NaCl
60 mM NaCl
90 mM NaCl
137 mM NaCl
b
Ligth Scattering Intensity (%)
Apparent Hydrodynamic Diameter (nm)
M-TTR, pH 7.4: D_h=4.20 ± 0.11 nm
M-TTR, pH 4.4: D_h=4.65 ± 0.26 nm
DACM-M-TTR, pH 7.4: D_h=5.03 ± 0.10 nm
DACM-M-TTR, pH 4.4: D_h=5.60 ± 0.30 nm
c
Fluorescence (a.u.)
Wavelength (nm)
Time
d
Fluorescence (a.u.)
Wavelength (nm)
Time
e
FRET Efficiency (E)
Time (s)
f
FRET Efficiency (E)
Native M-TTR, pH 7.4
Amylodogenic M-TTR, pH 4.4, 0 s
Aggregated M-TTR, pH 4.4, 3000 s

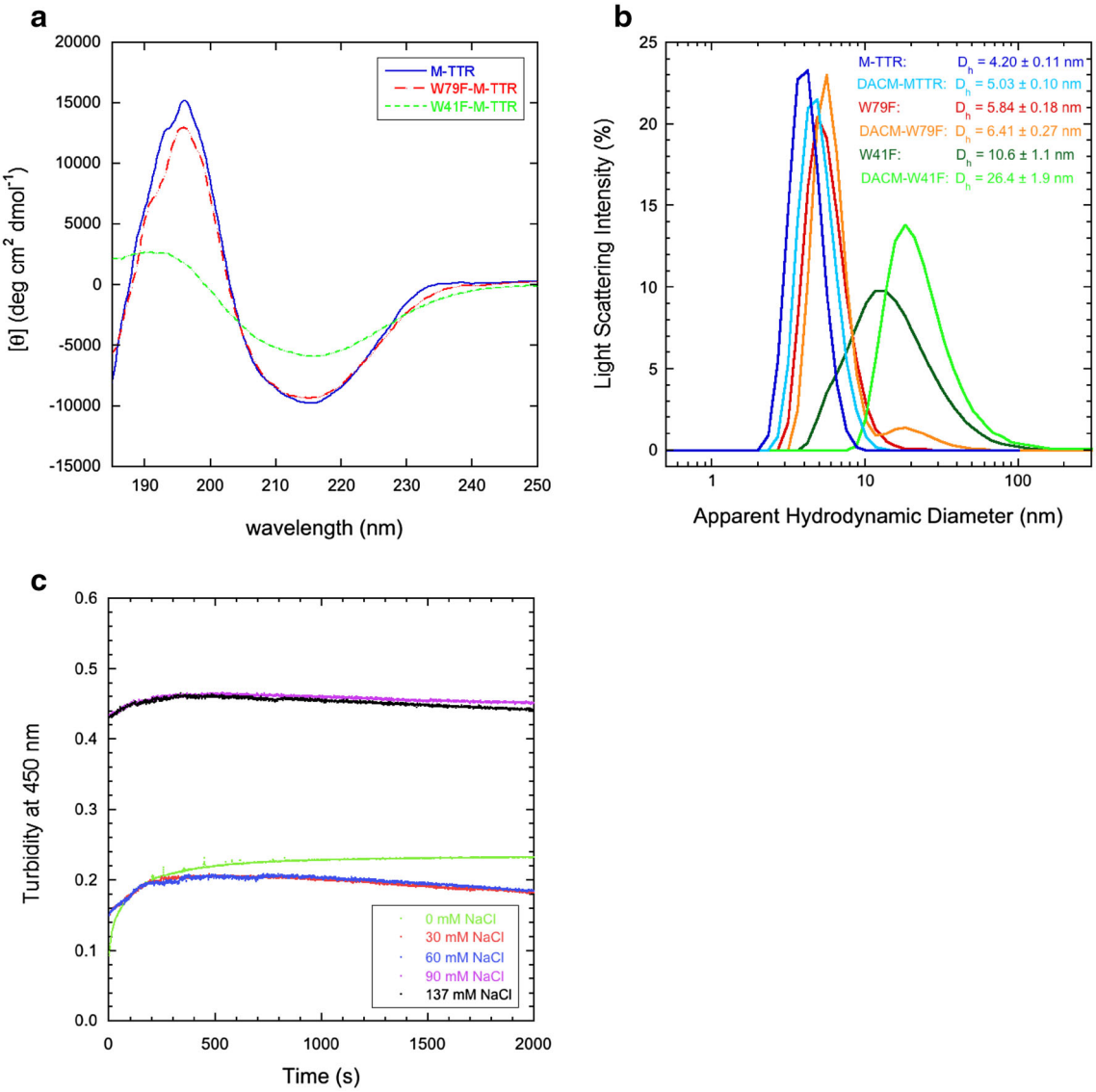


Fig. 7 Structural and aggregation properties of W41F and W79F-M-TTR mutants. **a** Far-UV CD spectra of the non-mutated and mutant forms of M-TTR obtained in 20 mM phosphate buffer, pH 7.4, 25 °C. **b** Size distributions by DLS of unlabelled and labelled non-mutated, W79F- and W41F-M-TTR at pH 7.4, 25 °C. **c** Aggregation time courses of W41F-M-TTR at pH 7.4, 37 °C in the presence of 0–137 mM NaCl

the same experiments carried out with non-mutated M-TTR, as the mutant provides the opportunity to measure the distance between DACM and Trp41 in the absence of possible artefacts (Fig. 8). The FRET E value of the folded state of W79F-M-TTR at pH 7.4, 25 °C was found to be 0.76 ± 0.02 (Fig. 8a), which is close to the value of non-mutated M-TTR ($E = 0.72 \pm 0.02$). This indicates that the mutation does not alter significantly the inter-residue distance between Cys10 and Trp41 in native M-TTR and reinforces the view that the FRET E value determined for wild-type M-TTR mainly reports on Trp41 and its distance from DACM.

The analysis of the urea denaturation curves of the unlabelled and labelled W79F mutant, monitored with tryptophan fluorescence, yielded values of $\Delta G^{U-F}_{H_2O}$ of 13.0 ± 1.4 and 8.6 ± 0.9 kJ mol^{-1}, respectively, m values of 4.3 ± 0.4 and 4.2 ± 0.4 kJ mol^{-1} M^{-1}, respectively, C_m values of 3.0 ± 0.1 and 2.0 ± 0.2 M, respectively (Fig. 8b). The $\Delta G^{U-F}_{H_2O}$ values of W79F-M-TTR and DACM-W79F-M-TTR are significantly lower than those of the corresponding wild-type proteins, indicating that the mutation lowers the conformational stability of the folded state. Moreover, the $\Delta G^{U-F}_{H_2O}$ value of DACM-W79F-M-TTR is significantly lower than that of W79F-M-TTR, indicating that the DACM-labelling lowers the conformational stability of the folded state, as observed for non-mutated M-TTR. The FRET E values measured after denaturation in 5 and 6 M urea ($E = 0.30 \pm 0.05$ and

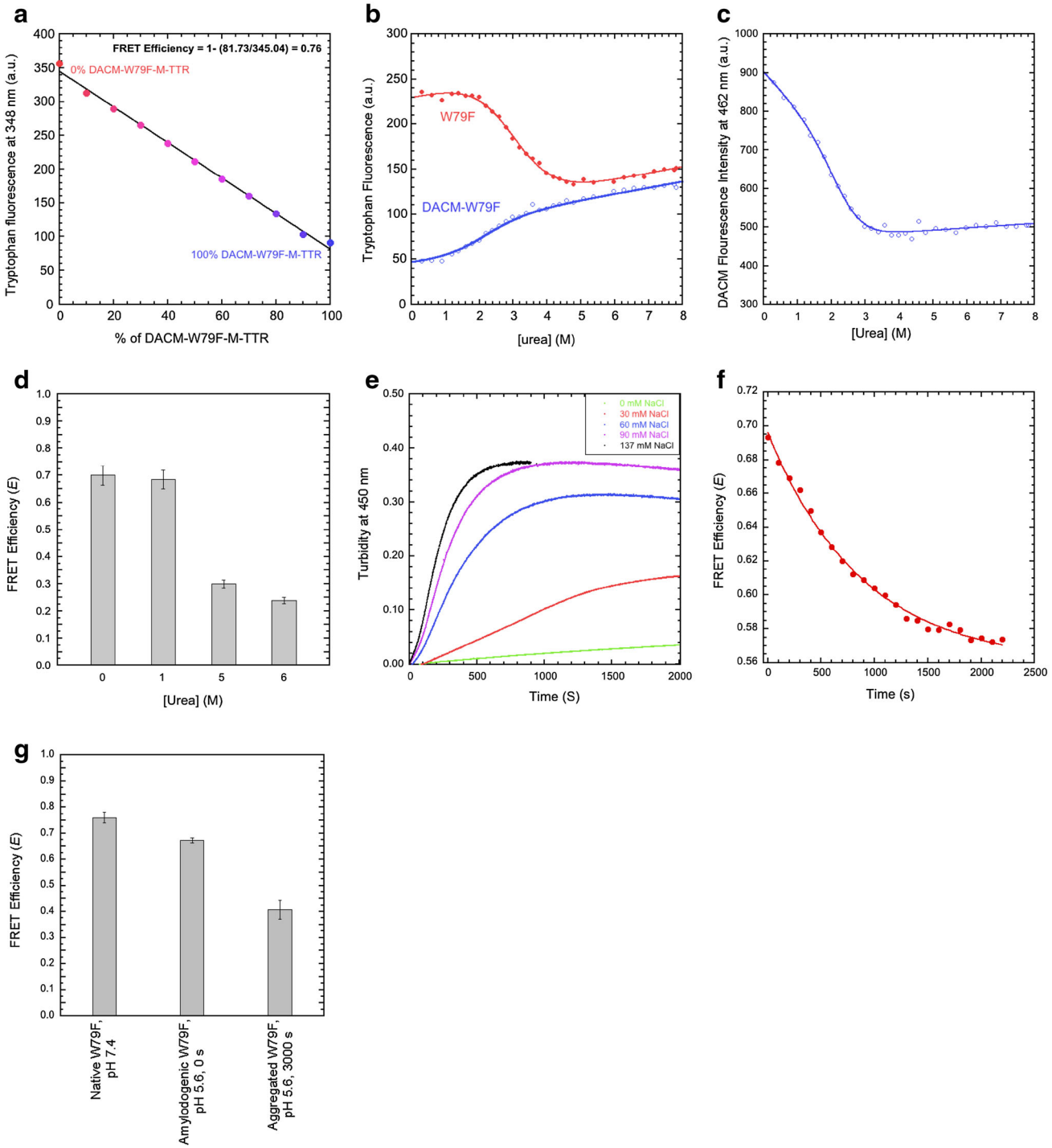


Fig. 8 FRET of W79F-M-TTR under different conditions. **a** Tryptophan fluorescence emission at 348 nm versus the percentage of DACM-W79F-M-TTR, at pH 7.4, 25 °C. The *straight line* represents the best fit of the data points to a linear function. **b** Urea denaturation curves (spectroscopic signal versus urea concentration at equilibrium) using tryptophan fluorescence at 330 and 365 nm as a spectroscopic probe for W79F-M-TTR and DACM-W79F-M-TTR, respectively, pH 7.4, 25 °C. **c** Urea denaturation curve of DACM-W79F-M-TTR, using DACM fluorescence at 462 nm as a spectroscopic probe, pH 7.4, 25 °C. **d** FRET *E* values at the indicated urea concentrations. **e** Effect of different NaCl concentrations on the kinetics of 15 μM DACM-W79F-M-TTR aggregation in 20 mM acetate buffer, pH 5.6, 37 °C, monitored with turbidimetry at 450 nm. **f** FRET *E* during aggregation. Conditions were 20 mM acetate buffer, 30 mM NaCl, pH 5.6, 37 °C. **g** FRET *E* values for native W79F-M-TTR at pH 7.4 (*first bar*), amyloidogenic monomeric W79F-M-TTR at pH 5.6, 0 s (*second bar*) and aggregated W79F-M-TTR at pH 5.6, 3000 s (*third bar*)

$E = 0.24 \pm 0.02$, respectively) are lower than those in 0 and 1 M urea ($E = 0.70 \pm 0.04$ and $E = 0.68 \pm 0.03$, respectively), indicating a higher distance between the DACM group and the Trp41 residue (Fig. 8d). The FRET E value in 1.0 M urea is similar or just slightly lower than that determined in 0.0 M urea, confirming that the equilibrium molten globule state detectable at low urea concentrations has an unmodified distance between the DACM moiety and Trp41. The plot of DACM fluorescence versus urea concentration confirms that the distance between the DACM moiety and the Trp41 residue increases upon unfolding (Fig. 8c).

Folding of W79F-M-TTR and DACM-W79F-M-TTR was monitored in real time using the stopped-flow apparatus, similarly to the non-mutated protein (Fig. 9a–c). The spectroscopic signals analysed by imposing an off-pathway model [28] were thus used to calculate transient and equilibrium FRET E values of the folded (F), unfolded (U), and partially folded (PF) states (Fig. 9d). Interestingly, the folded and partially folded states of W79F-M-TTR were found to have similar FRET E values, unlike non-mutated M-TTR (Fig. 9d). In addition, the transiently populated partially folded state for W79F-M-TTR has a DACM fluorescence significantly higher than those of both the folded and unfolded states (Fig. 9c), but this effect is less dramatic than that observed for the non-mutated protein (Fig. 5d). Kinetic data analysis and calculation of the FRET E values highlighted several interesting features (Fig. 9a–d). Apparent refolding rates in 0.5 M urea were found equal to 16.3 ± 0.2 and 172 ± 20 s^{-1} for M-TTR and W79F-M-TTR, respectively, suggesting a strongly destabilized partially folded state following the W79F mutation. However, when the DACM label is attached to M-TTR, refolding rates in 0.5 M urea were found to be 20.2 ± 0.2 for both DACM-M-TTR and DACM-W79F-M-TTR, suggesting a higher energy barrier for refolding when the label is attached to the W79F mutant. The FRET E value (Fig. 9d) of the W79F-M-TTR partially folded state in 0.5 M urea is slightly lower than that of M-TTR (0.68 ± 0.03, as opposed to 0.72 ± 0.03), an observation that, together with that obtained with the refolding rates, is reminiscent of a slightly less compact partially folded state in the mutant.

We then attempted to study the aggregation process of the labelled and unlabeled W79F mutant at pH 4.4, 37 °C, as we did for the non-mutated M-TTR. However, the aggregation process was too fast and DLS measurements showed that the protein was aggregated immediately after incubation under these conditions. We therefore, performed a preliminary study with DLS to identify a pH value and ionic strength ideal for slow aggregation of the labelled and unlabeled W79F mutant at a concentration of 15 μM and we found that at pH 5.6, both W79F-M-TTR and DACM-W79F-M-TTR aggregated sufficiently slowly to allow the detection and analysis of both the monomeric amylodogenic state and the aggregated state (data not shown). We therefore, acquired time courses of aggregation for 15 μM W79F-M-TTR and DACM-W79F-M-TTR in 20 mM acetate buffer, pH 5.6, 37 °C in the presence of 0, 30, 60, 90 and 137 mM NaCl, using turbidimetry at 450 nm as a probe for aggregation (Fig. 8e). We selected a NaCl concentration of 30 mM for FRET measurements and performed the experiments as described above for the wild-type protein (Fig. 8f) determining a FRET E value of 0.67 ± 0.01 for the amyloidogenic state (Fig. 8g). A FRET E value of 0.41 ± 0.04 was obtained for the aggregated state (Fig. 8g), although this value is difficult to interpret for the reasons explained above. Overall, the FRET E values measured for the native states at pH 7.4 and for the amyloidogenic states at mildly acidic pH for M-TTR and W79F-M-TTR are all similar, within experimental error, indicating that the distance between the DACM moiety attached to Cys10 and Trp41 does not change significantly in the transition from the native to the amyloidogenic state.

Discussion

The production of a mutant of human wild-type TTR that is stable as a monomer, generally referred to as M-TTR [25], has greatly facilitated the study of the folding process of this protein, of the structural perturbations occurring in the conversion of the fully folded monomer to the amyloidogenic state at weakly acidic pH values, of the aggregation process from such a state and of the mechanism of action of TTR as a toxin detoxifier [25, 26, 28, 47–49]. Following the observation that M-TTR contains only a solvent-exposed cysteine residue at position 10 that can be potentially labelled with a coumarin derivative able to act as an acceptor of the fluorescence of the naturally present tryptophan residues at positions 41 and 79, we generated a coumarin-labelled M-TTR molecule (DACM-M-TTR) to investigate the spatial distance between the coumarin moiety at position 10 and Trp41/Trp79, particularly Trp41, which is non-quenched and spatially closer to Cys10.

Five conformational states of M-TTR have been investigated, namely (1) the fully folded state at pH 7.4, (2) the urea-unfolded state at pH 7.4, (3) the molten globule state populated at equilibrium at low urea concentrations, pH 7.4, (4) the partially folded state accumulating transiently during folding, pH 7.4 and (5) the amyloidogenic state at pH 4.4.

The first study performed at pH 7.4, 25 °C, no added denaturants, indicated that DACM-M-TTR is folded, monomeric in solution, singly and uniformly labelled, and maintains a structure superimposable to that of unlabeled M-TTR. The labelled protein is less stable than the

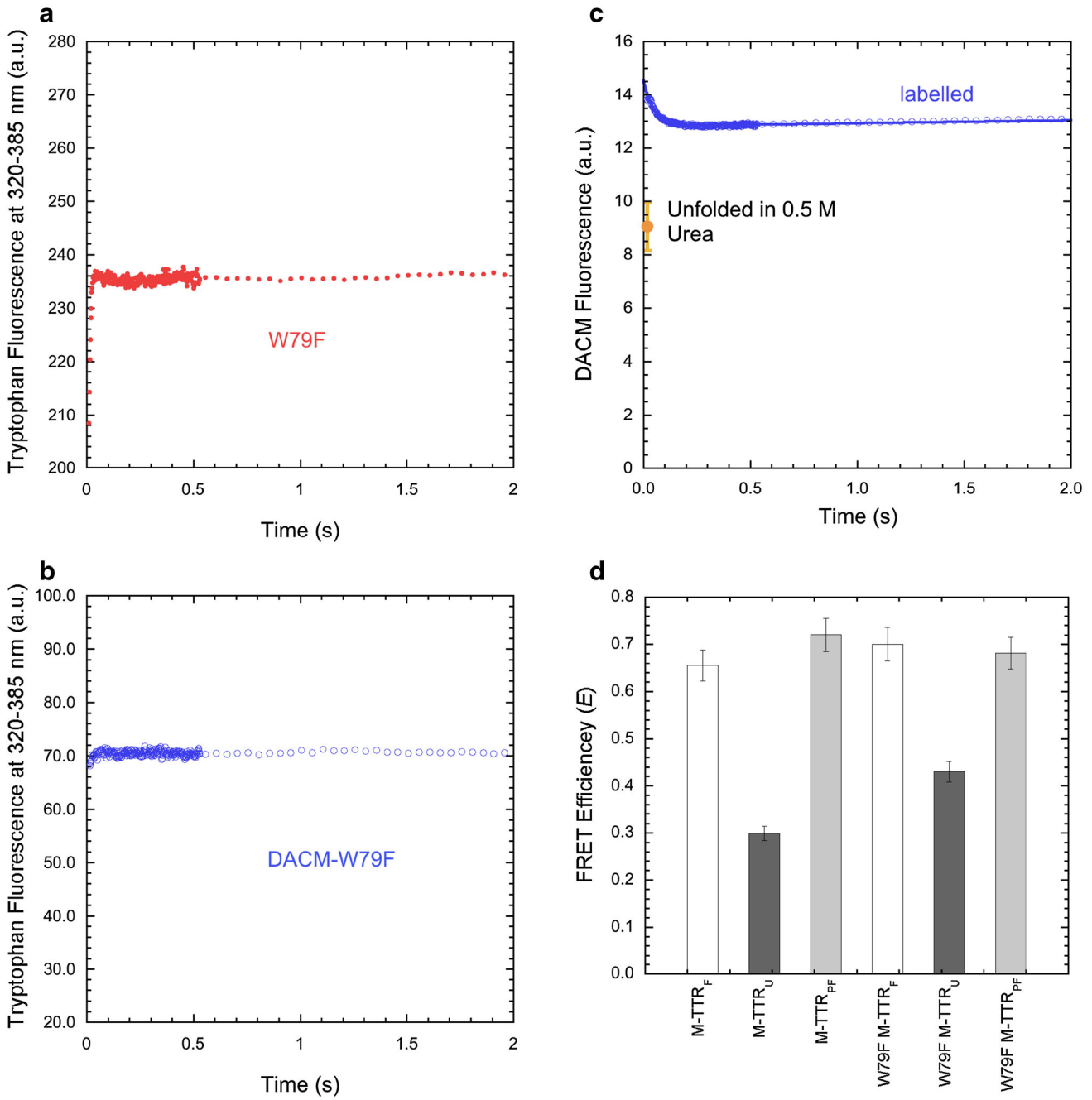


Fig. 9 RET of W79F-M-TTR during refolding. **a**, **b** Refolding time course of W79F-M-TTR (*red filled circles*) and DACM-W79F-M-TTR (*blue open circles*) monitored by tryptophan fluorescence (excitation 290 nm, emission 320–385 nm) in 0.5 M urea at pH 7.4, 25 °C. **c** Refolding time course of DACM-W79F-M-TTR (*blue open circles*) monitored by DACM fluorescence (excitation 290 nm, emission >385 nm) in 0.5 M urea at pH 7.4, 25 °C. **d** Comparison between the FRET *E* values calculated, from the spectroscopic data, for the folded (F), unfolded (U) and partially folded (PF) states of M-TTR and W79F-M-TTR in 0.5 M urea at pH 7.4, 25 °C

unlabeled one, as indicated by the lower free energy change of unfolding ($\Delta G^{U-F}_{H_2O}$), but this is attributable to the high sensitivity of the $\Delta G^{U-F}_{H_2O}$ value to covalent modifications, as indicated by the many observations accumulated so far that most mutants of TTR are indeed less stable than the wild-type protein, regardless of the position and type of the mutation [20]. A FRET *E* value of 0.72 ± 0.02 was obtained for DACM-M-TTR at pH 7.4. Such a value is largely due to the energy transfer from Trp41 to DACM, both because the fluorescence of Trp79 is quenched [23] and because Trp79 is more distant than Trp41 from Cys10 [25, 43]. The observation that the W79F mutant of DACM-M-TTR features a similar FRET *E* value under the same conditions, supports this hypothesis. A FRET *E* value of

0.72 ± 0.02 results into a spatial distance of 21.3 ± 1.6 nm, between Trp41 and DACM.

The conversion of the non-mutated and W79F mutant forms of M-TTR from their folded state to the amyloidogenic state at pH 4.4 (or 5.6 in the case of the W79F-M-TTR) does not change significantly the FRET *E* value and the related distance between DACM and Trp41, indicating that β-strand C, where Trp41 is located, does not increase its distance from Cys10 and does not unfold upon such conversion. This observation is particularly important as previous structural investigations on WT-TTR, carried out with spectroscopic methods, limited proteolysis and hydrogen/deuterium exchange monitored with NMR, attributed such a structural conversion mainly to the unfolding of β-strands C/D and the interconnecting loop [23] or to the entire β-sheet CBEF [24]. If this was the case, a decrease of the FRET *E* value should have been observed upon acidification, given the high sensitivity of FRET to even minor structural perturbations or increased dynamics. Our observation is indeed in agreement with more recent observations obtained with solution NMR that the whole CBEF β-sheet, including β-strand C, does not unfold, nor does it increase its structural dynamics, upon conversion of native M-TTR and WT-TTR into the amyloidogenic state at mildly acidic pH [26, 27, 50]. A unifying picture of the amyloidogenic state of TTR at mildly low pH (pH 3.9–5.5) that is consistent with all recent data obtained so far with different techniques is that of a largely native-like state with the entire DAGH β-sheet, the D–E loop, the E–F α-helix and its associated residues from the A–B loop, exhibiting structural fluctuations, in the presence of a substantially folded and packed CBEF β-sheet.

The equilibrium molten globule state of M-TTR, or W79F-M-TTR, populated at low urea concentrations, pH 7.4, has a FRET *E* value and a resulting *R* value similar to those of the folded and amyloidogenic states of the same protein isoforms, suggesting that the Cys10–Trp41 distance is not significantly altered in such a conformational state. It is interesting to compare the structural information collected so far on such molten globule state of M-TTR [28] with that obtained for the amyloidogenic state of M-TTR and WT-TTR at pH 3.9-5.0 [23, 25–27, 50], to assess whether the two conformational states are essentially similar albeit populated under different conditions. The two conformational states share similar FRET *E* and *R* values and similar near-UV CD spectra, but different intrinsic fluorescence and far-UV CD spectra. The molten globule state of M-TTR in 2.0 M urea, pH 7.4, has a far-UV CD spectrum different from that of the folded/amyloidogenic states [27], which appear to be similar between them [23, 28]. Indeed, the molten globule state indicates a substantial unfolding of some portions of the protein [28]. The molten globule state in 2.0 M urea, pH 7.4, unlike the amyloidogenic state at pH 4.4, has an intrinsic fluorescence spectrum similar to that of the folded state at pH 7.4 [23, 28]. Hence, it appears that the two conformational states are structurally distinct. It will be important to establish the amyloidogenic nature of the molten globule state, which may be as important as the amyloidogenic state populated at low pH.

As far as the off-pathway, partially folded state formed transiently during folding at pH 7.4 is concerned, it has a FRET *E* value significantly higher than that of the folded and amyloidogenic states, suggesting that Trp41 is closer to Cys10 than it is in the folded/amyloidogenic states. This result reinforces the view that such a conformational state is an off-pathway species with some non-native contacts that need to be unfolded before folding can proceed [28]. The kinetic and amyloidogenic conformational states share similar values of mean residue ellipticity, at least at 219 nm, and both induce an increase and blue-shift of the ANS dye [23, 25, 28]. However, in addition to showing different FRET *E* values, they also differ for the intrinsic fluorescence spectra, exhibiting a fluorescence intensity higher and lower than the folded state at pH 7.4, respectively [23, 28]. Again, it appears that the two conformational states are structurally distinct.

Another interesting observation emerging from these results is the high aggregation propensity exhibited by the W41F mutant, which was indeed found to be insoluble even under native conditions, limiting our FRET analysis. Among the many mutants of TTR associated with TTR diseases, one has been reported to involve Trp41 (W41L) and to cause vitreous opacity with amyloid-material positive for TTR in the vitreous fluid of the index patient [51, 52]. W41F and W41L TTRs share similarly high aggregation propensities, observed in vivo and in vitro, respectively, suggesting that Trp41 is a very important amino acid residue for maintaining the stability and solubility of native TTR.

Conclusions

In conclusion, the FRET study described here provides insight into the structural changes occurring during the folding, unfolding and aggregation of TTR, a protein whose aggregation has been linked to systemic amyloidoses. The most important achievements of our results is the clarification of the degree of unfolding and dynamics of β-strand C in the formation of the amyloidogenic state and the establishment of the concept that both the molten globule state and kinetically trapped species accumulating during folding are structurally distinct from such a state, indicating that TTR is a highly plastic protein able to populate a number of structurally discrete conformational states.

Acknowledgements We thank the Iranian Ministry of Science Research and Technology for providing the studentship for S.A.G. and the University of Florence for Fondi di Ateneo. F.B.'s research is funded by the Italian MIUR (Programma per Giovani Ricercatori Rita Levi Montalcini 2010). We also thank Martino Bolognesi for insightful discussion. We thank Joel Buxbaum and Xinyi Li for providing the gene coding M-TTR.

References

1. Soprano DR, Herbert J, Soprano KJ, Schon EA, Goodman DS (1985) Demonstration of transthyretin mRNA in the brain and other extrahepatic tissues in the rat. J Biol Chem 260(21):11793–11798
2. Stauder AJ, Dickson PW, Aldred AR, Schreiber G, Mendelsohn FA, Hudson P (1986) Synthesis of transthyretin (pre-albumin) mRNA in choroid plexus epithelial cells, localized by in situ hybridization in rat brain. J Histochem Cytochem 34(7):949–952
3. Herbert J, Wilcox JN, Pham KT, Fremeau RT Jr, Zeviani M, Dwork A, Soprano DR, Makover A, Goodman DS, Zimmerman EA et al (1986) Transthyretin: a choroid plexus-specific transport protein in human brain. The 1986 S. Weir Mitchell award. Neurology 36(7):900–911
4. Jacobsson B (1989) Localization of transthyretin-mRNA and of immunoreactive transthyretin in the human fetus. Virchows Arch A Pathol Anat Histopathol 415(3):259–263
5. Jacobsson B, Collins VP, Grimelius L, Pettersson T, Sandstedt B, Carlstrom A (1989) Transthyretin immunoreactivity in human and porcine liver, choroid plexus, and pancreatic islets. J Histochem Cytochem 37(1):31–37
6. Murakami T, Ohsawa Y, Zhenghua L, Yamamura K, Sunada Y (2010) The transthyretin gene is expressed in Schwann cells of peripheral nerves. Brain Res 1348:222–225. doi:10.1016/j.brainres.2010.06.017
7. Wakasugi S, Maeda S, Shimada K (1986) Structure and expression of the mouse prealbumin gene. J Biochem 100(1):49–58
8. Buxbaum JN, Reixach N (2009) Transthyretin: the servant of many masters. Cell Mol Life Sci 66(19):3095–3101. doi:10.1007/s00018-009-0109-0
9. Reixach N, Foss TR, Santelli E, Pascual J, Kelly JW, Buxbaum JN (2008) Human-murine transthyretin heterotetramers are kinetically stable and non-amyloidogenic. A lesson in the generation of transgenic models of diseases involving oligomeric proteins. J Biol Chem 283(4):2098–2107. doi:10.1074/jbc.M708028200
10. Sekijima Y (2015) Transthyretin (ATTR) amyloidosis: clinical spectrum, molecular pathogenesis and disease-modifying treatments. J Neurol Neurosurg Psychiatry 86(9):1036–1043. doi:10.1136/jnnp-2014-308724
11. Sipe JD, Benson MD, Buxbaum JN, Ikeda SI, Merlini G, Saraiva MJ, Westermark P (2016) Amyloid fibril proteins and amyloidosis: chemical identification and clinical classification International Society of Amyloidosis 2016 Nomenclature Guidelines. Amyloid 23(4):209–213. doi:10.1080/13506129.2016.1257986
12. Lie JT, Hammond PI (1988) Pathology of the senescent heart: anatomic observations on 237 autopsy studies of patients 90 to 105 years old. Mayo Clin Proc 63(6):552–564
13. Tanskanen M, Peuralinna T, Polvikoski T, Notkola IL, Sulkava R, Hardy J, Singleton A, Kiuru-Enari S, Paetau A, Tienari PJ, Myllykangas L (2008) Senile systemic amyloidosis affects 25% of the very aged and associates with genetic variation in alpha2-macroglobulin and tau: a population-based autopsy study. Ann Med 40(3):232–239. doi:10.1080/07853890701842988
14. Connors LH, Lim A, Prokaeva T, Roskens VA, Costello CE (2003) Tabulation of human transthyretin (TTR) variants, 2003. Amyloid 10(3):160–184. doi:10.3109/13506120308998998
15. Garzuly F, Vidal R, Wisniewski T, Brittig F, Budka H (1996) Familial meningocerebrovascular amyloidosis, Hungarian type, with mutant transthyretin (TTR Asp18Gly). Neurology 47(6):1562–1567
16. Cornwell GG 3rd, Sletten K, Johansson B, Westermark P (1988) Evidence that the amyloid fibril protein in senile systemic amyloidosis is derived from normal prealbumin. Biochem Biophys Res Commun 154(2):648–653
17. Saraiva MJ, Costa PP, Goodman DS (1988) Transthyretin (prealbumin) in familial amyloidotic polyneuropathy: genetic and functional aspects. Adv Neurol 48:189–200
18. Westermark P, Sletten K, Johansson B, Cornwell GG 3rd (1990) Fibril in senile systemic amyloidosis is derived from normal transthyretin. Proc Natl Acad Sci USA 87(7):2843–2845
19. Mccutchen SL, Colon W, Kelly JW (1993) Transthyretin mutation Leu-55-Pro significantly alters tetramer stability and increases amyloidogenicity. Biochemistry 32(45):12119–12127. doi:10.1021/bi00096a024
20. Sekijima Y, Wiseman RL, Matteson J, Hammarstrom P, Miller SR, Sawkar AR, Balch WE, Kelly JW (2005) The biological and chemical basis for tissue-selective amyloid disease. Cell 121(1):73–85. doi:10.1016/j.cell.2005.01.018
21. Steinrauf LK, Hamilton JA, Braden BC, Murrell JR, Benson MD (1993) X-ray crystal structure of the Ala-109 → Thr variant of human transthyretin which produces euthyroid hyperthyroxinemia. J Biol Chem 268(4):2425–2430
22. Kelly JW, Colon W, Lai ZH, Lashuel HA, McCulloch J, McCutchen SL, Miroy GJ, Peterson SA (1997) Transthyretin quaternary and tertiary structural changes facilitate misassembly into amyloid. Adv Protein Chem 50:161–181. doi:10.1016/S0065-3233(08)60321-6
23. Lai Z, Colon W, Kelly JW (1996) The acid-mediated denaturation pathway of transthyretin yields a conformational intermediate that can self-assemble into amyloid. Biochemistry 35(20):6470–6482. doi:10.1021/bi952501g
24. Liu K, Cho HS, Hoyt DW, Nguyen TN, Olds P, Kelly JW, Wemmer DE (2000) Deuterium–proton exchange on the native wild-type transthyretin tetramer identifies the stable core of the individual subunits and indicates mobility at the subunit interface. J Mol Biol 303(4):555–565. doi:10.1006/jmbi.2000.4164
25. Jiang X, Smith CS, Petrassi HM, Hammarstrom P, White JT, Sacchettini JC, Kelly JW (2001) An engineered transthyretin monomer that is nonamyloidogenic, unless it is partially denatured. Biochemistry 40(38):11442–11452
26. Lim KH, Dyson HJ, Kelly JW, Wright PE (2013) Localized structural fluctuations promote amyloidogenic conformations in transthyretin. J Mol Biol 425(6):977–988. doi:10.1016/j.jmb.2013.01.008
27. Lim KH, Dasari AK, Hung I, Gan Z, Kelly JW, Wemmer DE (2016) Structural changes associated with transthyretin misfolding and amyloid formation revealed by solution and solid-state NMR. Biochemistry 55(13):1941–1944. doi:10.1021/acs.biochem.6b00164
28. Conti S, Li X, Gianni S, Ghadami SA, Buxbaum J, Cecchi C, Chiti F, Bemporad F (2014) A complex equilibrium among partially unfolded conformations in monomeric transthyretin. Biochemistry 53(27):4381–4392. doi:10.1021/bi500430w
29. Clegg RM (1995) Fluorescence resonance energy transfer. Curr Opin Biotechnol 6(1):103–110
30. Leslie AGW (1992) Joint CCP4 and ESF-EACMB newsletter on protein crystallography. SERC Daresbury Laboratory, Warrington

31. Collaborative Computational Project N (1994) The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr 50(Pt 5):760–763. doi:10.1107/S0907444994003112
32. Evans P (2006) Scaling and assessment of data quality. Acta Crystallogr D Biol Crystallogr 62(Pt 1):72–82. doi:10.1107/S0907444905036693
33. McNicholas S, Potterton E, Wilson KS, Noble ME (2011) Presenting your structures: the CCP4mg molecular-graphics software. Acta Crystallogr D Biol Crystallogr 67(Pt 4):386–394. doi:10.1107/S0907444911007281
34. Long F, Vagin AA, Young P, Murshudov GN (2008) BALBES: a molecular-replacement pipeline. Acta Crystallogr D Biol Crystallogr 64(Pt 1):125–132. doi:10.1107/S0907444907050172
35. Vagin A, Teplyakov A (1997) MOLREP: an automated program for molecular replacement. J Appl Crystallogr 30:1022–1025. doi:10.1107/S0021889897006766
36. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66(Pt 2):213–221. doi:10.1107/S0907444909052925
37. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr 53(Pt 3):240–255. doi:10.1107/S0907444996012255
38. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 60(Pt 12 Pt 1):2126–2132. doi:10.1107/S0907444904019158
39. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins 65(3):712–725. doi:10.1002/prot.21123
40. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general amber force field. J Comput Chem 25(9):1157–1174. doi:10.1002/jcc.20035
41. Santoro MM, Bolen DW (1988) Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. Biochemistry 27(21):8063–8068
42. Pires RH, Karsai A, Saraiva MJ, Damas AM, Kellermayer MS (2012) Distinct annular oligomers captured along the assembly and disassembly pathways of transthyretin amyloid protofibrils. PLoS One 7(9):e44992. doi:10.1371/journal.pone.0044992
43. Peterson SA, Klabunde T, Lashuel HA, Purkey H, Sacchettini JC, Kelly JW (1998) Inhibiting transthyretin conformational changes that lead to amyloid fibril formation. Proc Natl Acad Sci USA 95(22):12956–12960
44. Lakowicz JR (2006) Principles of fluorescence spectroscopy, 3rd edn. Springer, New York
45. Nazarov PV, Koehorst RB, Vos WL, Apanasovich VV, Hemminga MA (2006) FRET study of membrane proteins: simulation-based fitting for analysis of membrane protein embedment and association. Biophys J 91(2):454–466. doi:10.1529/biophysj.106.082867
46. Visser NV, Westphal AH, van Hoek A, van Mierlo CP, Visser AJ, van Amerongen H (2008) Tryptophan-tryptophan energy migration as a tool to follow apoflavodoxin folding. Biophys J 95(5):2462–2469. doi:10.1529/biophysj.108.132001
47. Hurshman AR, White JT, Powers ET, Kelly JW (2004) Transthyretin aggregation under partially denaturing conditions is a downhill polymerization. Biochemistry 43(23):7365–7381. doi:10.1021/bi0496211
48. Cascella R, Conti S, Mannini B, Li X, Buxbaum JN, Tiribilli B, Chiti F, Cecchi C (2013) Transthyretin suppresses the toxicity of oligomers formed by misfolded proteins in vitro. Biochim Biophys Acta 1832 12:2302–2314. doi:10.1016/j.bbadis.2013.09.011
49. Cappelli S, Penco A, Mannini B, Cascella R, Wilson MR, Ecroyd H, Li X, Buxbaum JN, Dobson CM, Cecchi C, Relini A, Chiti F (2016) Effect of molecular chaperones on aberrant protein oligomers in vitro: super-versus sub-stoichiometric chaperone concentrations. Biol Chem 397(5):401–415. doi:10.1515/hsz-2015-0250
50. Das JK, Mall SS, Bej A, Mukherjee S (2014) Conformational flexibility tunes the propensity of transthyretin to form fibrils through non-native intermediate states. Angew Chem Int Ed Engl 53(47):12781–12784. doi:10.1002/anie.201407323
51. Yazaki M, Connors LH, Eagle RC Jr, Leff SR, Skinner M, Benson MD (2002) Transthyretin amyloidosis associated with a novel variant (Trp41Leu) presenting with vitreous opacities. Amyloid 9(4):263–267
52. Yazaki M, Varga J, Dyck PJ, Benson MD (2002) A new transthyretin variant Leu55Gln in a patient with systemic amyloidosis. Amyloid 9(4):268–271

Side Project II

Optimization of a protocol for expression and purification of a glycosylated protein through *Leishmania tarentolae* expression system.

E. coli can express only the 15% of eukaryotic proteins in their active form. This prokaryotic expression system, indeed, lacks the posttranslational modifications often essential for the activity of eukaryotic proteins. To face this problem alternative many expression systems have been developed. Among others, *Leishmania tarentolae* expression system (LEXSY), combines eukaryotic protein expression with highly complex and efficient folding and post-translational machineries and with easy handling broadly comparable to the one of bacterial expression systems.
L. tarentolae is a trypanosomatid protozoan host not pathogenic to mammals, first isolated from the Moorish gecko *Tarentola mauritanica*. *Trypanosomatidae* are rich in glycoproteins, which can account for more than 10% of their proteome. Probably due to their parasitic lifestyle, the oligosaccharide structures of their glycoproteins are often chemically similar to those of mammalians. Therefore, LEXSY is a very promising expression system to heterologously express glycosylated proteins.
In this context, I spent a three months period in Prof. Achour laboratory the SciLifeLab (www.scilifelab.se, Karolinska Institute, Stockholm, Sweden) to optimize a protocol to express and purify a Neuroserpin, human glycosylated protein expressed in the central nervous system. Prof. A. Achour, is a long-term collaborator and in his lab all the necessary expertise for LEXSY was available.

A synthetic gene was cloned into pLexsy-Sat2 plasmid for a secreted constitutive protein expression in LEXSY with an 6xHis-Tag to facilitate the purification process. The sequence of the protein was optimized based on the *L. tarentolae* codon usage, and fused with a Leishmania secretory signal peptide for secreted expression.
After sequencing, the plasmid was amplified and linearized. LEXSY host P10 cells (Jena Bioscience) were, thus, transfected with the linearized construct. The transfected cells were selected by plating on agar plates. LEXSY colonies were then selected and transferred into culture plates.
The constitutive secretory expression of the protein was tested through dot blot and the culture with highest expression was scaled up to 4 L culture to purify the glycosylated protein from the medium (*Figure 1*).
The glycosylated protein was isolated first through a nickel affinity chromatography followed by a gel filtrations. The SDS-PAGE of the eluted fractions showed that the protein was pure.

Therefore, during this stage, I could learn all the crucial steps of heterologous protein expression in LEXSY and applied all relevant protocols for the constitutive expression of human Neuroserpin, which I successfully purified at the end of the stage. In the months following the stage, I collaborated with other lab members to install this technology in Prof. Ricagno's lab (Dept. of Bioscience, University of Milan).

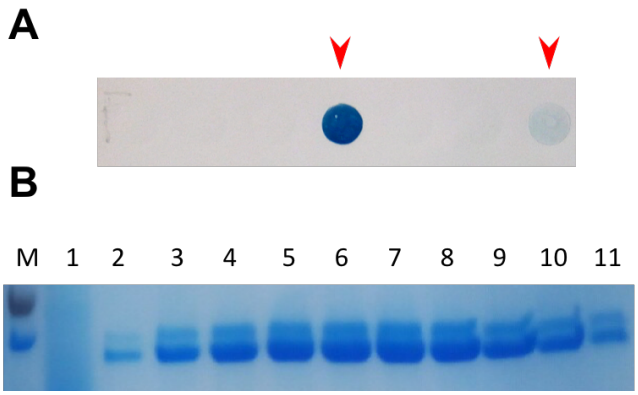


Figure 1: (A) Dot Blot of the medium of transfected cells. Only two clones, highlighted with the red arrow, show expression of NS. **(B)** SDS-PAGE of the eluted fractions of the affinity chromatography. Two distinct bands around NS molecular weight indicate that spontaneous proteolysis has occurred.