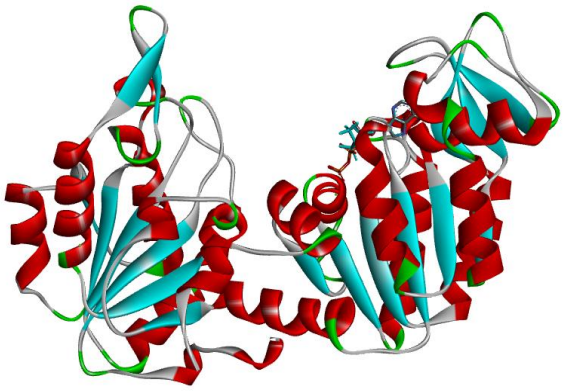
SAPIENZA
UNIVERSITÀ DI ROMA

PhD in Biochemistry
XXXI Cycle (2015-2018)

Impact of non synonymous single nucleotide variants on protein fitness: experimental analysis for a comparative study



Candidate

Maria Petrosino

Supervisor

Prof.ssa Roberta Chiaraluce

Coordinators

Prof. Stefano Gianni
Prof. Francesco Malatesta

SAPIENZA
UNIVERSITÀ DI ROMA

PhD in Biochemistry
XXXI Cycle (2015-2018)

Impact of non synonymous single nucleotide variants on protein fitness: experimental analysis for a comparative study

Candidate

Maria Petrosino

Supervisor

Prof.ssa Roberta Chiaraluce

Coordinators

Prof. Stefano Gianni
Prof. Francesco Malatesta

A tutti coloro che hanno creduto in me.
It always seems impossibile until it’s done.
(Nelson Mandela)

Acknowledgements

Though the following PhD thesis is an individual work, I could never have reached it without the help, support, guidance and efforts of a lot of people.

Firstly, I would like to express my sincere gratitude to my supervisors Prof.ssa Roberta Chiaraluce and Prof. Valerio Consalvi for believing in me and for instilling in me the qualities of being a good scientist. Your help and guidance over the years has been unmeasurable and without it I would not be where I am today. It has been a pleasure for me work with you.

I am deeply grateful to Dr.ssa Alessandra Pasquo for the knowledge you have passed on. You are my primary re source for getting my science questions answered and I will always be grateful for having the opportunity to learn under your guidance. You have made me grow so much both from a personal and professional point of view. I will treasure your advices forever.

Thanks also go out to Prof. Francesco Malatesta for taking care about my professional growth, for his encouragement and helpful advice.

Thank to Dr.ssa Francoise Dantzer (CNRS, University of Strasbourg) for the opportunity to work in her respectable laboratory. You made me feel as a component of your group from the first day, making me feel at home and very welcome. Thanks to her research group, in particular I would like to thank Jean-Christophe for all the days that we spent working and smiling in front of the column to manage the purification, Josè for his patient and precious help and above all Leonel and Kathline. These two friends formed the core of my research time in the Dantzer's group. You are wonderful and generous friends who I admire a lot for your ability to smile despite the situation. I'll never forget the help you gave me in research activities and the wonderful lunches and fun activities we've done together during my period in Strasbourg.

A very special thank you to my friends Leonore, Laura, Giusy, Marialaura, Giammarco, Ilaria, Elena, Claudia, Daniele, Flavio, Regine, Jo-Ann, Lais and to my friends and collegues Flavia, Francesca, Serena. A good support

system is important to surviving and staying sane during PhD. I was lucky to have all of you. You contribute to the following thesis in different ways: giving advices, working hard for helping me, making me smiling.

Thanks to my new collegues for helping and supporting me in this caotic period. A special thank to Giovanni for believing in me and for give me the opportunità to work in his research group and thanks to Silvia e Annalisa, too.

A special thank to my family for all their love and encouragement. Thanks becouse you supported me in all my choises being always with me and giving me the possibility to reach what I have now. For my parents that teach me what it means believe in something and work hard to reach your goal, and for my sister and her unmeasurable patient and love.

Index

Chapter 1 1
1.1 General introduction 1
1.2 Single nucleotide variants 2
1.3 Overview of possible effects induced by single amino acid substitution 9
1.4 How can you study the effect of point mutations on protein? 9
1.5 nsSNVs and cancer 14
1.5.1 nsSNVs and cancer: the experimental system 14
1.5.1i Phosphoglycerate kinase 1 18
1.5.1ii Human Frataxin 23
1.5.1iii Poly (ADP-ribose) polymerase 3 (PARP3) 29
Chapter 2 35
Aim of the study 35
Chapter 3 39
Methodology 39
3.1 Protein expression and purification 41
3.2 Spectroscopic measurements 43
3.3 Thermal denaturation experiments 43
3.4 Urea-induced equilibrium unfolding 44
3.5 PGK1 Enzyme activity and kinetic studies 45
3.6 PGK1 Protein crystallization, data collection, structures solution 46
3.7 hFXN Stopped-flow measurements 46
3.8 Data analysis 47
3.8.1 Quantitative analysis of equilibrium transition 47
3.9 Cell culture 50
3.10 Western blot analysis 50
3.11 Dose response 51
3.12 Clonogenic assay 51
3.13 PARP3 transfection in human breast cancer cell lines 52

Chapter 4 ..... .53
Results ..... .53
4.1 Experimental analysis of PGK1 nsSNVs.....56
4.1.1 Spectroscopic characterization of PGK1 wild-type and variants.....57
4.1.2 Kinetic analysis.....59
4.1.3 Thermal and thermodynamic analysis.....62
4.1.4 Structural analysis.....66
4.2 Experimental analysis of hFXN nsSNVs.....72
4.2.1 Spectroscopic characterization of hFXN wild-type and variants.....74
4.2.2 Thermal and thermodynamic analysis.....76
4.2.3 Analysis of the folding pathway of hFXN wild-type and its variants 80
4.3 Biochemical characterization of PARP3 and analysis of PARP3 nsSNVs.....83
4.3.1 PARP3 in the tumorigenicity of glioblastoma 85
4.3.2 PARP3 and nsSNVs …….....88
4.4 Analysis of the effects induced by single amino acid substitution.....92
4.4.1 Effect on protein conformation in solution.....93
4.4.2 Effect on protein structure.....95
4.4.3 Effect on protein function.....96
4.4.4 Effect on protein binding.....98
4.4.5 Effect on protein stability.....100
4.4.6 Effect of single amino acid substitution on protein production.....102
Chapter 5.....103
Discussion 103
Chapter 6.....109
Attachments 109
Common acronyms and abbreviations.....211
References.....215

Chapter 1

1.1 General introduction

Proteins are large biological molecules that control most vital cellular functions. They consist of one or more chains of amino acids in an order determined by the base sequence of nucleotides in the DNA coding for the protein. Thanks to the information from the genetic code and according to the energy landscape, proteins fold into their correct three-dimensional structures and exert their specific function. The correct fold of large portion of the structure is generally related to specific protein functions and when any even small alterations occur, it is possible to observe a decrease, an increase or a drastic change in the protein function. In several cases alterations at the amino acid level can influence the conformational rearrangement, the function or the binding properties of a given protein. On this premise, knowledge on protein structure-function relationships can be crucial in finding the molecular basis for hereditary diseases and in predicting protein function from structure and vice versa. Therefore, the study of structure-function relationships is really important nowadays to better understand several diseases at their molecular level. In particular, this kind of approach seems to be relevant in cancer research considering that several somatic variants resulting from alterations at the amino acid level have been detected in cancer genome for several proteins. The analysis of this kind of alterations is key to understand the genetic bases of disease progression, patient survival

and also response to therapy. Since knowledge of protein function in health and disease is essential to identify new and more specific cures for different diseases and to design pharmacologically active and more selective drugs, the information resulting from the analysis of somatic mutations found in cancer tissues can improve the available therapies and create new and more specific ones suggesting that precision and personalized medicine is not anymore a daydream.

1.2 Single nucleotide variants

The human genome, which consists of over 3 billion base pairs, has remained well conserved throughout evolution, in fact it is at least 99.5 % identical between any two humans on the planet [Levy S et al, 2007]. Modern genomic sequence analysis have revealed that it is more complex, diverse, and dynamic than previously thought, with possible genetic variations in the range from 0.1 % to 0.4 % [Jorde LB et al, 2004]. Sequence variations at the amino acid level may influence the conformation, function or binding properties of a given protein [Bhattacharya R et al, 2017]. When sequence changes occur, even in non-protein coding regions of the DNA, the consequent alteration in the human genome may cause or just be linked to a specific disease.

In the last decades, several studies have focused their attention on protein variants and their relationship with rare, mendelian genetic diseases and some authors have recently hypothesized that the common variants may contribute significantly to genetic risk for common disease [Lander ES et al, 1996; Risch N et al, 1996; Collins FS et al, 1997].

The most common genetic differences in the human genome are polymorphisms, generally defined as single base pair variations in DNA sequence that occur in at least 1% of the population. These differences are called single nucleotide polymorphisms (SNPs) and they are considered the most common genetic variations observed in humans [Collins FS, 1998] (The International HapMap Consortium 2003) [Shastry, 2002], since they occur, on the average, once every 300–400 base pairs [Kruglyak and Nickerson, 2001] in contrast with other types of polymorphisms (for example, differences in copy number, insertions, deletions, duplications, and rearrangements) that also occur, but much less frequently. Some of them are located within the coding or regulatory regions of genes but, considering that only about 5% of the human genome codes for the production of proteins (The International Human Genome Sequencing Consortium 2001), most single nucleotide mutations may fall in the non-coding regions of genes or in the intergenic regions. Point mutations within a coding sequence are of particular interest because they can cause qualitative and quantitative changes in gene expression, RNA splicing, protein translation, or gene function. Coding SNPs, especially nonsynonymous coding SNPs (nsSNPs), induce changes at the amino acid level of the polypeptide sequence of the protein generating protein variants with single amino acid substitution (nsSNVs), also referred to as missense mutations, that may present structural and/or functional alterations [Zhou et al, 2010].

Only in the last years, millions of human SNVs have been identified, and these variants could be strongly correlated with phenotypic variations of traits/diseases [Cao et al, 2017]. It has been reported that nearly 30% of the nsSNVs are predicted to affect the protein function [Chasman and Adams, 2001; Ng and Henikoff, 2002; Ramensky et al, 2002]. Several nsSNVs have

been found neutral or with no effect on human health, but many nsSNVs have been linked to many diseases or may be responsible for individual susceptibility to contract diseases.

In the last years, several sequencing initiatives [Hudson et al, 2012; Stratton et al, 2012] have deeply contributed to the implementation of information about nsSNVs detected in humans and many databases are nowaday available, listing human SNVs and their effects, such as dbSNP (Single Nucleotide Polymorphism database) [Sherry et al, 2001], SAAPdb [Hurst et al, 2009] or SNPdbe [Schaefer et al, 2012]. Since genetic variation in the human genome is an emerging resource for studying cancer and other diseases [Mueller et al, 2015; Kunz et al, 2016, Zang et al, 2016, Didonna et al, 2015], in the last years databases like Online Mendelian Inheritance in Man (OMIM) [Hamosh a et al, 2005] and COSMIC [Forbes et al, 2010], that contribute to catalogue nsSNVs found in pathological tissues, have been created. Even if there is a huge amount of SNVs deposited in public platform, only a small proportion of them are functional polymorphisms that contribute to disease phenotypes [Tabor et al, 2002; Yuan et al, 2006]. Recently a new important database, named Pan-Cancer Atlas (https://www.cell.com/pb-assets/consortium/pancanceratlas/pancani3/index.html), is available and catalogues the nsSNVs found in cancer tissues in the Genomic Data Commons (https://portal.gdc.cancer.gov/).

Given the high potentiality of knowledge about the relationship between structure-function changes observed in nsSNVs and diseases, it is noteworthy that this field is extensively expanding in cancer research since several somatic mutations that have been identified in cancer tissues are nsSNVs. Until now a huge dataset has been generated and made available. A stumbling-block is represented by the fact that no tumor genome has been

completely characterized but the advances in high throughput technologies and the declining costs of high-throughput sequencing are transforming our understanding of cancer and will allow us to generate a comprehensive map of polymorphisms distributed over the entire genome that will be helpful to identify new and more appropriate therapies and to make effort in drug discovery.

1.3 Overview of possible effects induced by single amino acid substitution

Single amino acid substitutions can be beneficial, adding new functionality and increasing the fitness of the cells or deleterious, damaging the protein, causing destabilization of the protein structure or reducing its functional activity because of the disruption of a site that is directly involved in the function of the protein and thereby contributing to disease phenotypes [Studer et al, 2013]. Usually, change in amino acids with similar size and physico-chemical properties (e.g. substitution from leucine to valine) has a mild effect. Similarly, if the point mutation disrupts secondary structure elements (e.g. substitution to proline in alpha helix region) such mutation usually may affect whole protein structure and function. Thus nsSNVs can potentially affect the function of the cell in a variety of ways.

The most important consequence of point mutations in different types of functional sites can be either the increase or the decrease of protein activity. The most dramatic effects have been detected when point mutations occur in the active sites of enzymes or in the binding pockets of receptors

[Stevanin et al, 2004; Yamada et al, 2006]. Considering that enzymes catalyze biochemical reactions not only thanks to catalytic residues, but also with several surrounding residues important for ensuring proper attachment of the substrates and cofactors to the active site cavity (binding sites residues), mutations that occur on the residues located in the neighborhood of the active site can influence the activity [Zhang et al, 2001; Takamiya et al, 2002; Zhang et al, 2010]. In the human population, 25% of the known nsSNVs significantly affect protein function in vivo [Yue and Moult, 2006]. It has been reported that in most cases nsSNVs lead to loss-of-function (LOF) generally due to perturbation of the active site of the enzyme or of its global structure [Kucukkal et al, 2015] or to a reduction of the thermodynamic stability of the protein leading to a shift of the folding equilibrium toward the nonfunctional unfolded state, possibly coupled to irreversible aggregation (thermal instability) and/or degradation by cellular quality control [Yue et al, 2005; Casadio et al, 2011; Shi et al, 2011; Stefl et al, 2013; Petukh et al, 2015]. Sometimes, but not commonly, nsSNVs can result in a protein that still functions, even if in a "non-canonical" way, as evident in the so-called gain-of-function (GOF) mutants that generally include oncogenic mutations in proteins such as the Ras GTPase and the epidermal growth factor receptor family of tyrosine kinases, which often drive cancer tumor cell development and proliferation [Schubbert et al, 2007; Arteaga et al, 2014].

The occurrence of a mutational event can generate a protein variant that perturbs conformational constraints of the native protein (e.g., substituting a small side chain residue to a large one and vice versa, resulting in backbone strain or overpacking) or have physicochemical effects (substitutions between hydrophilic residues and hydrophobic residues, burial of charged residues, the disruption of hydrogen bonds, loss of hydrogen

bonds, of S–S bonds) [Shirley et al, 1992], leading to considerable alterations of the stability of the native protein [de Cristofaro et al, 2006; Koukouritaki et al, 2007; Ode et al, 2007]. Protein stability is a key characteristic of a functional protein [Wang and Moult 2001; Zhang et al, 2001; Ramensky et al, 2002; Wang and Moult 2003; Capriotti et al, 2005; Karchin et al, 2005; Ye et al, 2006; Zhang et al, 2010]. It is known that a single base DNA substitution may result in alteration of stability of the corresponding native protein [Wang and Moult, 2001; Yue et al, 2005] and that about 80% of missense mutations associated with disease affects the stability of proteins by several kcal·mol^{-1} [Wang and Moult, 2001]. Generally, a nsSNV can have destabilizing or stabilizing effects: most frequently, missense mutations lead to a destabilization of the protein stability by making it susceptible to proteolysis or by changing the thermal inactivation temperature. The impact of protein destabilization can be observed in several proteins involved in many neurodegenerative diseases, such as Parkinson's disease [Lin et al, 2008; Morais et al, 2009; Nuytemans et al, 2010] or in cancer, where destabilizing mutations in the core of the protein lead to an inactivation of many tumor suppressors [Stehr et al, 2011]. An amino acid substitution can occur at a position critical for the folding of the protein destabilizing it and/or stabilizing a misfolded state [Valastyan and Lindquist, 2014]. The importance of the correct folding of a protein is clear if we consider the hundreds of diseases associated with many misfolded proteins [Dobson, 2003; Thusberg and Vihinen, 2009; Groenendyk et al, 2010; Valastyan and Lindquist, 2014].

Several investigations have been addressed to better elucidate how missense variants can induce minimal or significant alterations on the protein flexibility [Young et al, 2001; Zhang et al, 2010; Karplus and Kuriyan, 2005]

considering that the conformational flexibility of a protein is crucial for its correct folding, for the binding with partners and for its function. The presence of a point mutation, by altering the protein flexibility, can affect protein aggregation propensity. This is considered a hallmark of some neurodegenerative diseases, like Alzheimer's disease (AD), Parkinson's disease (PD), Huntington's disease (HD), amyotrophic lateral sclerosis (ALS) and prion diseases [Guijarro et al, 1998; Chiti et al, 1999; Fandrich et al, 2001; Chiti et al, 2003; Bucciantini et al, 2004; Harris and True 2006; Keage et al, 2009; Khemtemourian et al, 2008; Robinson, 2008; Yankner and Lu 2009] and for this reason research on protein misfolding and aggregation is progressively gaining increasing attention. The general mechanism by which nsSNVs promote these diseases is the promotion of toxic protein aggregation [Calamini and Morimoto, 2012; Knowles et al, 2014; Valastyan and Lindquist, 2014] triggered by the destabilization and unfolding of the native protein structure, which exposes aggregation-prone regions previously buried inside the structure.

Approximately 60% of disease-associated nsSNVs induce a perturbation in the binding site of the protein, expecially if the point mutation affects residue located on protein-protein interface [Schuster-Bockler and Batemen, 2008; Teng et al, 2009; David et al, 2012], and may lead to changes in the interactions with partners (activators, repressors or substrates) and in the binding specificity [Kohler et al, 2008; Wu et al, 2008; Bauer-Mehren et al, 2009; Vanunu et al, 2010; Barabàsi et al, 2011]. Many of these mutations induce significant perturbations that result in complete loss of interactions or complete abolishment of the function, generally leading to diseases [Wang et al, 2012; Sahni et al, 2015].

The impact of single amino acid substitutions on protein stability has been investigated by theoretical and experimental approaches [Bross et al, 1999; Wang and Moult, 2001; Ferrer-Costa et al, 2002; Steward et al, 2003; Yue et al, 2005; Scheraga et al, 2007; Dill et al, 2008; Fersht, 2008]. In general, the experimental studies are relatively limited due to the cost and time needed for the entire process which often includes mutagenesis, protein expression and purification, followed by thermal or chemical unfolding.

1.4 How can you study the impact of single point mutations on the protein?

Single point mutations can occur between different individuals of the same species. These mutations can lead to a protein without any changes in its amino acid sequence (synonymous SNVs (sSNVs)), insert a stop codon or generate a protein variant with an amino acid change (nsSNVs). Even if most SNVs are neutral or have no effect on human health and embryonic development [Kimura M, 1983; Wang Z, Moult J, 2001], a large number of nsSNVs have been associated with diseases [Wang Z et al, 2001]. Due to the complexity of human biology, the association of a nsSNV with a complex phenotype is still a problem and methods that target understanding the effects of missense mutations on various sequence, structural and functional features are progressively implemented in the hope of deciphering the phenotype–genotype relations [Stanley et al, 2013; Yates and Sternberg 2013; Kucukkal et al, 2014].

Nowadays bioinformatics tools can be really helpful to filter large datasets of genetic variation and in the last decades good computational

methods have been developed to predict the effect of nsSNVs and their pathogenicity in order to predict the experimentally observed effects for disease-related variants [Hecht et al, 2013]. These methods are based on implemented computational algorithms that allow the evaluation and analysis of mutations as a function of their pathological consequence [Balu and Purohit, 2013; Kamaraj and Purohit, 2013; Kamaraj et al, 2013; Kumar et al, 2013] giving us the possibility to distinguish the non-significant SNVs from the ones that might produce major disease associated consequences. Computational methods such as SIFT [Sim et al, 2012], Polyphen-2 [Adzhubei I. et al, 2013], StSNP [Uzun et al, 2007], Bongo [Cheng et al, 2008], Condel [Ng and Henikoff, 2006; Gonzalez-Perez and Lopez-Bigas, 2011] and MAPP [Stone E.A. et al, 2005] classify SNVs according to neutral, negative or positive effects on the structure or function of a protein, while different computational methods, such as FoldX [Schymkowitz et al, 2005], PoPMuSiC [Dehouck et al, 2011], CUPSAT [Parthiban et al, 2006], Rosetta [Leaver-Fay et al, 2011] I-Mutant [Capriotti et al, 2005] or INPS (a predictor of the Impact of Non-synonymous-variations on Protein Stability) have also been developed to classify a single amino acid substitution according to its effect on the protein stability by comparing wild-type with mutated proteins, and predicting the variation of protein stability ($\Delta\Delta G$) of the mutation. However these methods are generally based on a single folded structure, but the native state of a protein is not unique, thus it is important to consider the entire ensemble of the protein conformers in dynamic equilibrium.

A deeper understanding of the effect of a point mutation on the protein structural, physical and chemical properties requires the utilization of several techniques: only a combination of predictive analysis and experimental procedures can give us a whole picture regarding the consequences of

missense mutations on protein. Since in the last years the accepted paradigm that proteins can tolerate nearly any amino acid substitution has been replaced by the awareness of deleterious effects of mutations, especially on the thermodynamic and kinetic stability of protein, several methods have been developed to catalogue the nsSNVs that affect the stability and function of the protein [Ng and Henikoff, 2006; Yue and Moult, 2006]. In order to combine the results of the various tools, consensus predictors have been developed to allow comparison between methods that use different analytical approaches [Vendruscolo et al, 2003; Bendl et al, 2014].

One useful method largely used to provide information about the conformational changes that a mutated protein may undergo under various conditions is the molecular dynamics simulation (MD) leading to a systematic evaluation of molecular properties in dynamic molecular systems. Using specific algorithms it is possible to assess the potential impact of variants within the protein revealing the possible various mechanisms by which variants may lead to functional alteration. Some of them can be revealed energetically, others structurally or dynamically analyzing the dynamic behaviour of proteins [Zimmermann MT et al, 2017].

The importance of a combination of experimental studies with molecular dynamics simulations allow us to highlight several experimental procedures that can be used to reveal the structural and functional effects of single amino acid changes in proteins owing nsSNVs. Of particular interest are experimental methods that provide 3D structures of genetic variant proteins revealing atomic-level structural details with which it's possible to analyze the structural differences caused by the variation at the DNA level. X-ray Diffraction, solution Nuclear Magnetic Resonance or Electron Microscopy contribute to implement the PDB archive [Rose PW et al, 2015]. Even if the

availability of the 3D structure of a gene product represents the most informative source of data that can explain what is able to cause a particular phenotype, most studies that analyze the relationship between point mutations and experimentally observed 3D protein structure published to date have been restricted to individual proteins or single diseases. There is a paucity of quantitative analyses of the consequences of SNVs on 3D protein structure going beyond the realm of prediction [Arodź and Płonka, 2012], even if tools that incorporate structural information when making predictions have been implemented [Brown DK Bishop ÖT, 2017]: an excellent communication between bioinformatics and biomolecular scientists is requested to increase the amount of data about genetic variants and phenotype.

Since several nsSNVs associated with different diseases can affect the stability of the native protein [Wang and Moult, 2001], protein folding analysis and stability measurements may be useful for the study of disease-associated processes [Adhikari et al, 2015]. With this purpose mutagenesis can be used to obtain mutated proteins and followed by protein expression and purification steps in order to obtain a pure protein product that can be analyzed by thermal or chemical unfolding to determine the difference of Gibbs free energy value (ΔG) between the wild-type and the mutated proteins. Several forces contribute to the stabilization of the folding state of a protein, expecially hydrogen bonding, hydrophobic contacts between the buried side chains in the interior of the molecule out of contact with water, and the formation of intramolecular contacts between side chains [Eswar and Ramakrishnan 2000]. In their stable state, proteins are marginally stable, between −3 and −10 kcal·mol^{-1} [Taverna and Goldstein, 2002]. They can tolerate some destabilization within a narrow range [Bershtein et al, 2006;

Goldstein, 2011] but when this range is overshot a pathological condition can occur. Point mutations may add energy (i.e. more than +2 kcal·mol^{-1}) to the folded state and destabilize the equilibrium between the native state and the unfolded one, making the protein more likely to aggregate in its unfolded form: this is a crucial event in some diseases characterized by unfolded aggregates of proteins [Yue et al, 2005]. However, in some cases a mutation removes energy from the folded state (i.e. less than +2 kcal·mol^{-1}) and stabilizes it making the protein too rigid and so non-functional. Most of the disease-causing mutations were found to be destabilizing [Pasquo et al, 2012; Grothe et al, 2013; Khan et al, 2013] and, in particular, the degree of destabilization was found to be elevated for mutations that introduce drastic changes such as charged to neutral, relatively rigid to relatively flexible, or aromatic to aliphatic mutation types.

Considering the plausible alterations in the folding process caused by point mutations, it is not surprising that experimental studies about protein folding and stability are needed.

Equilibrium studies allow to understand the cooperativity of the folding process and can be used to evaluate the conformational stability. Usually denaturation is studied monitoring changes in secondary or tertiary structural elements in protein, by means of circular dichroism or fluorescence spectroscopy, when different concentrations of a chemical denaturant agent are used to induce the denaturation process. Urea and guanidine hydrochloride are the most common chemical denaturant agents used, because it is well known that the interaction of urea or guanidine hydrochloride with the constituent groups of proteins is more favorable than the interaction of those groups with water [Tanford C, 1970]. Denaturants alter the equilibrium between the native (N) and the denatured (D) states and

the free energy of unfolding reaction is given by: $\Delta G = -RT \ln([D]/[N])$, where [D] and [N] are the concentrations of protein in the denatured and native state, respectively.

Thermal denaturation experiments can also be performed due to the large amount of information that they can give about the denaturation process of a given protein. In particular, calculation of the T_m plotting the first derivative of the molar ellipticity values at around 222 nm, where helices are known to give their spectral signal, as a function of temperature, it's possible to compare the transition curves of the wild-type and of the mutated proteins looking at the midpoint of the denaturation process.

Furthermore, the folding process of a given protein can be analyzed by following the kinetics of unfolding that allow to understand the protein dynamics and provide information about the folding pathway. Kinetic studies are performed by altering the equilibrium of the system introducing a perturbation, such as mixing two solutions with different urea concentrations or increasing rapidly the temperature of the sample, and evaluating how the system relaxes to a new equilibrium. The relaxation process can be estimated by following the variation of an optical probe with time, such as the fluorescence of the aromatic residue.

1.5 nsSNVs and cancer

Cancer is an important health problem worldwide, considered as one of the most frequent causes of death. It is a genetic disease caused by changes of genes that control the way in which cells function allowing them to grow out of control and become invasive. These kind of abnormalities have been

detected studying cancer genomes, however large-scale research studies are still needed to characterize some tumour types that have not been deeply characterized. Furthermore it is important to take into consideration that each cancer type has a unique combination of genetic changes in different patients that can be inherited from the parents or can be related to the adoption of unhealthy lifestyle, including smoking, drinking, physical inactivity and ‘‘westernized’’ diets [Hua et al, 2014]. In fact, several types of cancer can be caused by environmental factor exposures that include substances, such as the chemicals in tobacco smoke, and radiation, such as ultraviolet rays from the sun, which can lead to cancer by inducing DNA damage [Levi et al, 1999]. In addition, in cancer cells genetic variation occurs with an increased frequency due to the uncontrolled proliferation and the compromised DNA integrity.

Since the mechanisms of cancer cells at their molecular level have not been identified, new technologies and the knowledge gained from previous genomic studies could be used to define the full set of alterations to DNA and RNA in many cancers. It can be really useful to set studies that compare genomic information from tumours and normal tissue from the same patient in order to allow researchers to discover genomic changes that may drive cancer. Another possibility is to expand the current use of genomic methods to investigate the molecular basis of clinical phenotypes in order to identify genetic changes that may distinguish aggressive cancers from indolent ones or to study the molecular basis of response to a given therapy, as well as mechanisms of resistance to treatment.

Nowadays the National Cancer Institute (NCI) is working to analyze the DNA and RNA of cancer cells using advanced technologies such as next-generation DNA sequencing. Using multiple genomic techniques the

landscape of the cancer genome is mapped in order to discover new changes linked to disease. In particular, a collaboration between NCI and the National Human Genome Research Institute (NHGRI), and Therapeutically Applicable Research to Generate Effective Treatments (TARGET), named The Cancer Genome Atlas (TCGA), have characterized thousands of genomes and matched normal samples making available a collection of cross-cancer analyses delving into overarching themes on cancer, including cell-of-origin patterns, oncogenic processes and signaling pathways.

It is clear that in cancer research genetic susceptibility plays an important role: genome-wide association studies (GWAS) have been conducted in large cohorts of patients allowing us to say that cancer can be considered a genetic disease sometimes caused by a single or few catastrophic somatic mutations that are responsible for cellular transformation [Zhang CZ et al, 2015], and sometimes caused by a combination of several genomic alterations that work together to promote cancer. Many of the identified causal genetic changes are related to the SNV that derives from a substitution of one base pair in DNA and that can initiate the cascades of downstream signaling and eventually transforms the cellular phenotype from normal to malignant [Hyeonju S et al, 2017]. These mutations represent novel targets for therapeutic intervention in this otherwise incurable disease.

1.5.1 nsSNVs and cancer: the experimental system

Although nsSNVs are theoretically expected in tumours, as it has been shown in the previous section, they can occur in mostly random manner across the genome or they can be significantly biased toward a specific codon transition [Hyeonju S et al, 2017]. Furthermore, it is not easy to characterize all the nsSNVs found in cancer tissues by an experimental point of view, due to the large amount of variants identified in several proteins involved in different pathways and to the cost and time required for the experiments. However, since structural and functional analysis of nsSNVs in the DNA sequences of humans may help to predict an individual's response to certain drugs, susceptibility to environmental factors, and risk of developing particular diseases [Johnson et al, 2008; Zhou et al, 2010], during the last years we focus our attention on the characterization of nsSNVs of different proteins involved in the regulation of cell metabolism and in the response of DNA damages.

In particular, during my PhD I carried on a biochemical characterization of nsSNVs that have been detected in different types of cancer of three proteins involved in the regulation of cell metabolism, such as the Phosphoglycerate kinase 1 (PGK1) and the human Frataxin (hFXN), or in the response of DNA damages, such as the Poly (ADP-ribose) polymerase 3 (PARP3). Using structural and functional analysis our aim was to investigate any possible correlations between alteration on protein structure, function or stability caused by single amino acid substitutions and cancer. Furthermore in order to present a detailed discussion about this issue, in this thesis I'm going to

combine the results obtained during my PhD with those obtained for variants of other proteins studied by our group in the last years: Bromodomains (BRDs), Peroxisome Proliferator-Activated Receptor γ (PPARγ), Serine/threonine-protein kinase pim-1 (Pim-1) and Protein tyrosine phosphatase ρ (PTPρ) in order to present a systematic and comparative analysis of properties of protein variants caused by single amino acid substitutions.

1.5.1i Phosphoglycerate kinase 1

Tumors reprogram pathways of nutrient acquisition and metabolism to meet the bioenergetic, biosynthetic and redox demands of malignant cells: altered metabolic activity supports anabolic growth during nutrient replete conditions, catabolism increases to support cell survival during nutrient limitation, and fortification of redox homeostatic systems to counteract the metabolic effects of oncogene activation, tumor suppressor loss, and other stresses [Boroughs L.K. et al, 2015].

The greater example of a reprogrammed metabolic pathway in cancer is the Warburg effect or aerobic glycolysis [Lunt S.Y. et al, 2011]. Glycolysis is a physiological response to hypoxia in normal tissues, and Otto Warburg in the 1920s observed that cancer cells are able to adapt to survive in difficult conditions, like for example in O_2 deficiency, through a reprogramming of their metabolism according to their requirements [Bertout J.A. et al, 2008]. Phosphoglycerate kinase (PGK) (EC 2.7.2.3) is an essential glycolytic enzyme for all living organisms. It catalyzes the reversible phosphotransfer reaction from 1,3-bisphosphoglycerate (1,3-BPG) to MgADP to produce 3-

phosphoglycerate (3-PG) and MgATP, an important ATP-generating step in glycolysis [Beutler E., 2007]. By controlling ATP and 3-PG levels, PGK plays an important role in coordinating energy production with biosynthesis and redox balance. PGK is a typical hinge-bending monomeric enzyme of 417 amino acids with an apparent molecular mass of approximately 45 kDa. It is characterized of two domains of equal size, the N- and C-terminal domains. The N-terminal domain binds 3-PG or 1,3-BPG, whereas the C-terminal domain binds MgADP or MgATP. The two domains are separated by a deep cleft and linked by two alpha-helices (α-helix 7 and α-helix 14) [Vas et al, 2010; Palmai et al, 2009]. During the catalytic cycle this flexible hinge region allows the two domains to approach involving a large number of conformational rearrangements: in the open form the enzyme binds the substrates while in the closed form the enzyme performs the transfer of the phosphoryl group and it exerts its catalytic activity [Vas et al, 2010].

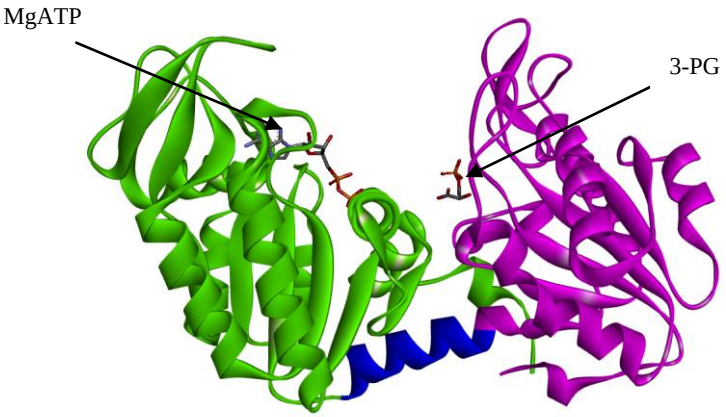


Fig. 1.1 Phosphoglycerate kinase 1 (PDB code: 2XE7, open conformation) PGK1 is composed by two domains of equal size, the N- and C-terminal domains (violet and green,

respectively), connected by a hinge region (blue). The N-terminal domain binds 3-PG or 1,3-BPG, whereas the C-terminal domain binds MgADP or MgATP.

Two human phosphoglycerate kinase isoenzymes have been identified: PGK1 and PGK2. They are characterized by distinctive tissue localization and encoded by two distinct genes [Willard et al, 1985; McCarrey et al, 1987]. PGK1 is ubiquitously expressed in all somatic cells while PGK2, also known as testis form, is unique to meiotic/postmeiotic spermatogenic cells. Interestingly, similar to many other glycolytic enzymes, human PGK1 was found implicated not only in the glycolytic pathway infact, in addition to its metabolic function, it may play different roles. In mammal cell nuclei, for example, this enzyme participates in the DNA replication and repair [Jindal et al, 1990]; it can also phosphorylate L-nucleoside analogues used in antiviral and anticancer therapies [Krishnan et al, 2003; Gallois-Montbrun et al, 2004; Gondeau et al, 2008].

Furthermore, this enzyme may be surprisingly secreted by tumors in the extracellular environment where it exhibits thiol reductase activity on plasmin that permits the cleavage of plasminogen in order to generate the vascular inhibitor angiostatin [Lay et al, 2000]. Thus the enzyme has been suggested to be an important negative regulator of the angiogenic process that is essential for tumour and metastatic growth [Shichijo et al, 2004; Wang et al, 2007]. It has been found that PGK1 regulates angiogenesis by generating angiostatin and by reducing the secretion of the proangiogenic cytokines VEGF and IL-8. CXCR4 signaling regulates PGK1 expression. At sites of high CXCL12 production, PGK1 secretion is inhibited by the CXCL12/CXCR4 axis. Thus, CXCL12 signaling through CXCR4 generates an ‘angiogenic switch’ necessary for metastatic growth. Together, these data

further show that CXCL12/CXCR4 chemokine axis and PGK1 represent at least one of the critical events necessary for metastasis of prostate cancer as well as a mechanism for a proangiogenic switch to promote tumor growth [Wang et al, 2007].

In addition, under hypoxic conditions, PGK1 may translocate from cytoplasm to mitochondrion where it may act as a protein kinase and phosphorylate different protein substrates [Li et al., 2016]. In particular, under hypoxic stress, ERK activation-dependent mitochondrial translocation of a small portion of cytosolic PGK1 have been observed [Li X. et al, 2016]. Once activated, ERK1/2 phosphorylates PGK1 at Ser203 allowing the recruitment of the PIN1 prolyl isomerase and leading to isomerisation of the Ser203. Pro204 bond and the subsequent exposed PGK1 pre-sequence (38-QRIKAA-43) on its surface is recognized by the mitochondrial translocase of the outer membrane complex, leading to the translocation of PGK1 into the mitochondria. Here, PGK1 acts as a protein kinase, interacts with and directly phosphorylates pyruvate dehydrogenase kinase isozyme 1 (PDHK1) at Thr338 using ATP as a phosphate donor. This phosphorylation activates PDHK1 and enhances PDHK1- mediated pyruvate dehydrogenase E1 a phosphorylation at Ser293, which inactivates the pyruvate dehydrogenase complex preventing the conversion of pyruvate and coenzymeA (CoA) to acetyl-CoA and CO_2 in the mitochondria. This phosphorylation inhibits mitochondrial pyruvate metabolism and ROS production and enhances lactate production, thereby promoting tumour development [Xinjian et al, 2016]. Recently, the protein kinase activity of PGK1 has been related to initiation of autophagy [Qian et al., 2017].

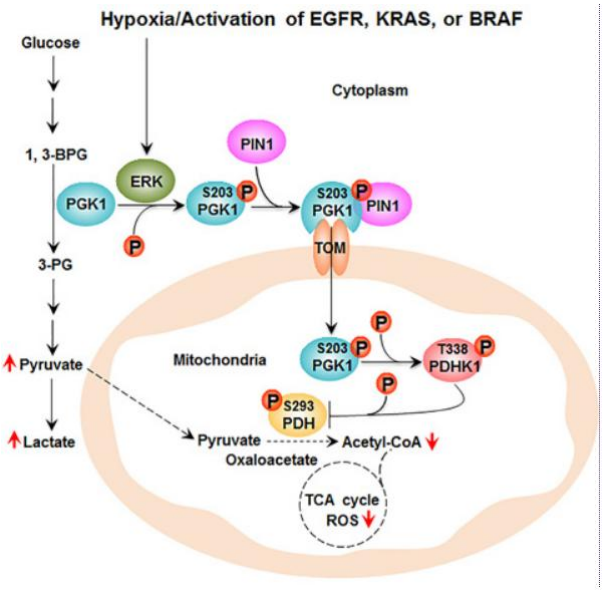


Fig. 1.2 Mechanism of Mitochondrial PGK1-Coordinated Glycolysis and TCA Cycle in Tumorigenesis [Xinjian Li et al, 2006; doi:10.1016/j.molcel.2016.02.009].

PGK1 is regulated by hypoxia-induced factor-1α (HIF-1α), the most important factor involved in the cellular response to hypoxia (Wang et al, 2007). Several solid tumors exhibit an increased expression of glycolytic enzymes such as PGK1 to generate ATP in hypoxic conditions. Elevated levels of PGK1 protein have been detected in the serum of patients with pancreatic cancer (Hwang et al, 2006) and in breast cancer tissues compared with normal tissues (Kabbage et al, 2008) and this suggests a plausible activity of the enzyme as a biomarker for cancer.

A great number of mutations of the PGK1 gene have been so far identified: most of them are related to the X-linked human PGK1 Deficiency (OMIM entry 311800) [Chiarelli et al, 2012]. PGK1 Deficiency is a genetic

disease that has been found in ~40 patients, and ~20 different mutations have been shown to be disease-causing mutations. Approximately 80% of the mutations cause single amino acid substitutions or small deletions [Fermo et al, 2012]. PGK1 deficient patients show heterogeneous phenotypes including chronic nonspherocytic hemolytic anemia, neurological dysfunctions, and myopathies, even though the presence of all three types of clinical signs in the same patient is rare.

In addition, several somatic mutations of PGK1 have been exclusively identified in different cancer types, as reported in COSMIC (Catalogue of Somatic Mutations in Cancer) (http://cancer.sanger.ac.uk/cosmic) [Forbes et al., 2011]. The role of PGK1 variants in cancer is not still clear, so this study would give a biochemical characterization of some PGK1 variants (R38M, R65W, G166D, M189I, A199V, V216F and F241S) found in cancer tissues and annotated in the COSMIC database [Forbes et al, 2010].

1.5.1ii Human Frataxin (hFXN)

Several studies in the literature about cancer highlight the relationship between iron metabolism and carcinogenesis [Richardson D.R. et al. 2009; Torti S.V. 2011, 2013]. Iron is an inorganic element which is critical for cell proliferation and growth by incorporating into iron- or heme-containing enzymes. Since these enzymes are involved in respiratory complexes, DNA synthesis, cell cycle and detoxification processes, iron is essential in terms of cell replication, cellular metabolism and growth. Iron is not only fundamental for cell survival but it can be related to carcinogenesis because, beside the effects mentioned above, iron can create reactive oxygen species (ROS) by

participating in Fenton reaction where hydroxyl radical is produced: ROS can damage DNA and be mutagenic.

Frataxin (FXN) is a small highly conserved mitochondrial protein ubiquitously expressed in prokaryotes and eukaryotes [Adinolfi et al, 2002; Gibbson et al, 1996] that plays an important role in Fe-S cluster biosynthesis and iron and heme metabolism [Muhlenhoff et al, 2002; Che net al, 2002; Busi et al, 2006]. Sequence alignment of the frataxin family shows two distinct regions: an N-terminal region of 70-90 residues completely absent in prokaryotes and poorly conserved in eukaryotes with typical unfolded features [Huynen MA et al, 2001] and a highly conserved C-terminal region of 100-120 residues with a sequence identity of about 25% and similarity around 40-70% indicating that this part of the protein is functionally important [Pastore A et al, 2013]. Frataxin is nuclear encoded, expressed in the cytoplasm and imported in the mitochondrion through an import signal contained in the N-terminal region of the protein. In fact, human frataxin (hFXN) is synthesized as a precursor of 210 amino acids and imported into the mitochondrion where the mitochondrial processing peptidase operates several cleavage processes that remove the N-terminal region and allow the formation of the mature form of the protein [Koutnikova H et al, 1998; Condò I et al, 2007; Schmucker S et al, 2008].

The mature protein consists of amino acids 81–210 and has an apparent molecular mass of about 14 kDa [Campuzano et al, 1996]. The human isoform shows an α- β sandwich structure composed of two α-helices, which form one plane over five antiparallel β-strands, which form the second plane. Furthermore, two others β-strands intersect the two planes in order to achieve the overall structure [Bencze et al, 2006].

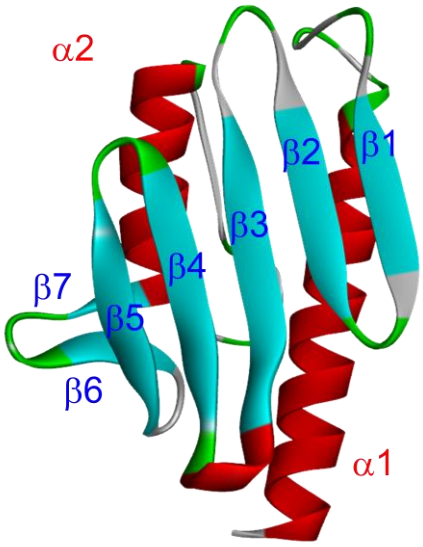


Fig. 1.3 Human Frataxin (PDB code: 1EKG) Globular structure of hFXN with two terminal α-helices which pack against a β-sheet with 7 strands, giving a planar α- β sandwich structure motif.

Several conserved acidic residues (Asp, Glu) are located between the first helix and the edge of the β 1-sheet gene rating a semi-conserved acidic ridge that generates a negatively charged surface (Fig. 1.4) that seems to be involved in iron binding and in binding with partners [Dhe-Paganon et al, 2000]. Three tryptophans are also conserved even if the occurrence of this kind of residue is relatively low in proteins, suggesting that they could have an important structural and/or functional role.

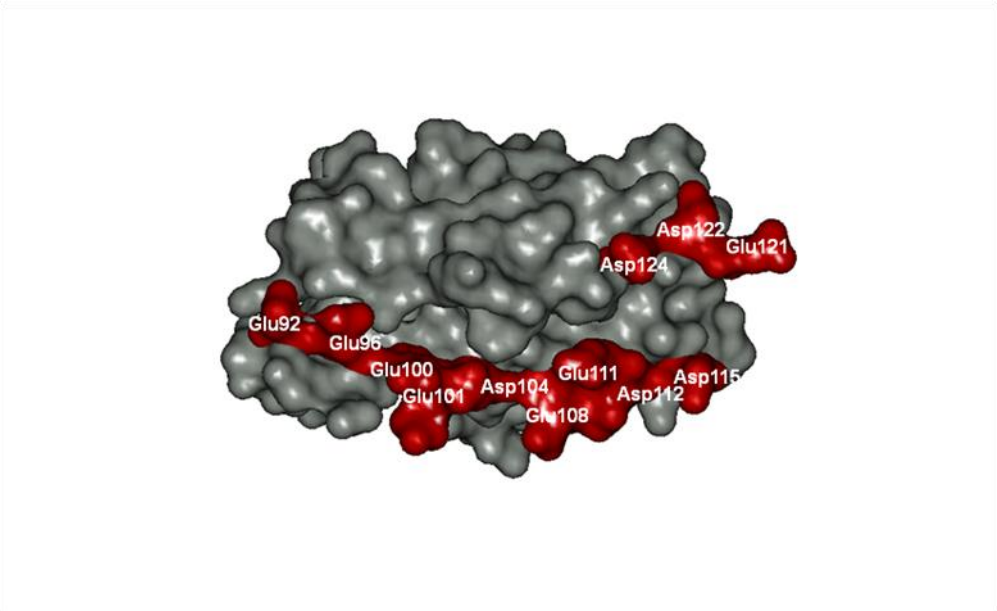


Fig. 1.4 Molecular representation of the negatively charged surface of hFXN. This acidic region is located on the interface between α1 and the first β1 strand and seems to be involved in iron binding and in binding with partners.

In mammals, frataxin is mainly espressed in tissues with a high metabolic rate, like neurons, heart, kidney and liver and normally activates mitochondrial iron-sulfur-cluster protein assembly complex, which is composed of the Nfs1 enzyme and the scaffold protein Isu [Gerber J et al, 2003; Ramazzotti A et al, 2004]. These two proteins are, respectively, the cysteine desulfurase that converts cysteine to alanine and the transient scaffold protein on which the cluster assembles [Bandyopadhyay S et al, 2008; Py B et al, 2011].

The direct frataxin partner have been extensively studied: early genetic and biochemical studies were careful to generically indicate interaction between frataxin and the Nfs1/Isu complex [Gerber J et al. 2003; Ramazzotti A et al. 2004], however later it was reported that human frataxin interact with Isu with high affinity and in an iron-dependent way [Cook JD et al. 2006; Bencze

KZ et al. 2007]; finally, independent works confirmed that hFXN interacts with a preformed complex composed of Nfs1, Isu and Isd11 in an iron-independent fashion [Tsai and Barondeau 2010; and Schmucker S et al. 2011].

The notoriety of this protein is predominantly related to Friedreich ataxia (FRDA), an autosomal-recessively inherited disorder described for the first time in 1863 by Friedreich and with an incidence of 1:50000 in the European population. 95% of FRDA patients are homozygous for unstable guanine-adenine-adenine (GAA) expansions in the first intron of the frataxin gene (*fxn*) on the positive strand of chromosome 9q21.11 [Campuzano et al, 1996], however another 5% of FA patients are compound heterozygotes with an expansion on one allele and conventional mutations on the other. At least 15 missense point mutations have been reported [Cossée M et al, 1999; Labuda M et al, 1999; Musco G et al, 2000; Leidgens S et al, 2010; Schmucker S et al, 2011]. In vitro studies demonstrated that these variants (W155R, I154F, D122Y, G130V, N146K) retain a native fold under physiological conditions but have reduced thermodynamic stabilities and binding property of frataxin for its protein partners [Musco et al, 2000; Correira A et al, 2006, 2008]. The core syndrome is an early onset, slowly progressive ataxia associated with areflexia that appears during adolescence. Ataxia arises from combined afferent (peripheral sensory neuropathy plus spinal degeneration), cerebellar and sometimes also vestibular dysfunction. In addition to ataxia of stance and gait, patients develop appendicular and truncal ataxia. Dysarthria is another cerebellar feature present in 70% with abnormal pitch variation, loudness maintenance, breath support for speech, hypernasality and consonant imprecision due to laryngeal or velopharyngeal dysfunction [Folker et al, 2010]. Cardiac involvement is also possible, with

left ventricular hypertrophy, paroxysmal or permanent supraventricular tachycardias [Bourke et al, 2011]. In 8 to 49% of the cases it's possible to see diabetes [Dürr et al, 1996; Finocchiaro et al, 1988] with an usually late manifestation in the course of FRDA (mean 15 years after onset) [De Michele et al, 1996]. The onset is often acute, sometimes with ketoacidosis [Bird et al, 1978]. The main event in the etiology of diabetes is the loss of pancreatic islet ß cells with concomitant decline of insulin secretion. The process is further aggravated by increased insulin requirements due to insulin resistance. Scoliosis is considered typical for FRDA. Its prevalence varies between 33 and 100% depending on the individual study [Harding et al, 1981; Labelle et al, 1986]. Similarly, foot deformities (pes cavus, club foot, pes planus) may significantly interfere with mobility in 55 to 90% of patients. Severity does not depend on expansion size or age of onset, but is related to disease duration and age.

Despite several studies have been performed until now on frataxin, the exact function of this protein is still unclear: in addition to its certain implication in iron homeostasis [Richardson et al, 2010], iron–sulfur cluster (ISC) biosynthesis [Stemmler ey al, 2010] and protection from oxidative stress and apoptosis [Condo et al, 2006], it seems that frataxin can be involved in several other functions.

Frataxin is a protein required for cell survival since its complete knockout is lethal. This protein protects tumour cells against oxidative stress and apoptosis but also acts as a tumour suppressor [Guccini I et al, 2011] and, until now, the molecular bases of this apparent paradox are missing. Furthermore, conflicting results have been reported on the role of frataxin in cellular growth: both frataxin knockdown and frataxin overexpression were shown to impair cell growth [Stehling O et al, 2004; Thierbach R et al, 2005;

Schulz TJ et al, 2006]. Very little is known regarding the molecular regulation of frataxin expression: histone deacetylase inhibitors, erythropoietin, cisplatin, 3-nitropropionic acid and hemin increase levels of frataxin in vitro (Gottesfeld et al, 2007). Moreover, transcription factors peroxisome proliferator-activated receptor gamma (Marmolino et al, 2009) and hypoxia-inducible factor-2 alpha (HIF-2a) (Oktay et al, 2007) positively regulate frataxin expression. It has been reported that frataxin overexpression can protect tumour cells from apoptosis but can also act as a tumour suppressor participating, by regulating p53 activation, in tumour adaptation to hypoxia, a critical feature associated with tumour growth and progression, thus suggesting that frataxin levels can influence tumour cell fate [Guccini I et al, 2011].

Furthermore, in several databases such as COSMIC database or ATLAS, several FXN variants have been reported suggesting that abnormalities in this highly conserved essential protein can be implicated in cancer onset and/or progression. This study would give a biochemical characterization of some FXN variants (D104G, A107V, F109L, Y123S, S161I, W173C, S181F and S202F) found in cancer tissues and annotated in the COSMIC database [Forbes et al, 2010].

1.5.1iii Poly (ADP-ribose) polymerase 3 (PARP3)

In addition to the accepted role for cell metabolism in cancer, it is well established that DNA-repair/DNA-damage pathways are important in cancer progression because dysregulation leads to higher levels of genomic instability, increased mutation rate, and enhanced intra-tumor heterogeneity [Burrel RA et al, 2013; Chae YKet al, 2016; Gavande NS et al, 2016].

Mutations in DNA-repair genes can confer growth and survival advantages of cancer cells, therefore it can be interesting to improve knowledge about this important aspect of cancer biology because this might lead to the development of new targeted therapies in DNA-repair deficient cancers or even can improve the efficacy of existing therapies, such as PARP inhibitors.

The biological roles of Poly (ADP-ribose) polymerase 3 (PARP3) are currently under investigation; however, several key reports indicate the integral roles of PARP3 in DNA damage repair responses [D'Amours et al, 1999; Ame et al, 2004], and thus it has been investigated as a novel target in oncology.

PARP3 is an important member of the PARP family, also referred to as diphtheria toxin-like ADP-ribosyltransferases (ARTD), composed of 17 members which all share a conserved ADP-ribosyl transferase (ART) fold and involved in a number of crucial cellular processes that are linked to genomic DNA integrity such as DNA repair, genome stability, and cellular stress responses [D'Amours et al, 1999; Ame et al, 2004; Hottiger et al, 2010]. In the PARP family, only a subset of members is predicted to have the ability to produce Poly-ADP-ribose (PARP-1 to PARP-5a and PARP-5b) while two are inactive enzymes (PARP-9 and PARP-13) and the remaining members are able to produce a mono-ADP-ribose modification [Kleine et al, 2008]. The 17 members of the PARP family show very different structures and cellular functions but are all related by the presence of the PARP signature, a conserved PARP catalytic domain [Otto H et al, 2005].

The founding and most studied member PARP1 as well as PARP2 and PARP3 play important roles in the repair of DNA strand breaks and are known to be catalytically activated through interaction with DNA damaged and catalyze auto-ADP-ribosylation and ADP-ribosylation of other nuclear

acceptor proteins [Pascal JM et al, 2015; Martin-Hernandez K et al, 2016; Wei H et al, 2016]. PARP1, PARP2 and PARP3 show a characteristic fold that is characterized by conserved tryptophan-glycine-arginine (WGR) and C-terminal catalytic (CAT) domains and of an N-terminal region (NTR) that is really different between the three proteins [Oliver AW et al, 2004]. In fact, in contrast to PARP1, the NTRs of PARP-2 and PARP-3 are much shorter (78 and 40 residues, respectively), and they do not contain the N-terminal zinc finger domains essential for initial DNA binding of PARP1 [Langlier MF et al, 2012]. Short NTRs of PARP2 and PARP3 are not strictly required for DNA-dependent activation but rather contribute to the overall binding affinity and specificity for SSB due to the presence of basic residues in this region, as shown for PARP2 [Ame et al, 1999], and contribute to the allosteric regulatory mechanism of DNA-dependent catalytic activation via local destabilization of CAT through local rearrangements of the WGR domain [Dawicki-McKenna JM et al, 2015; Obaji E et al, 2016].

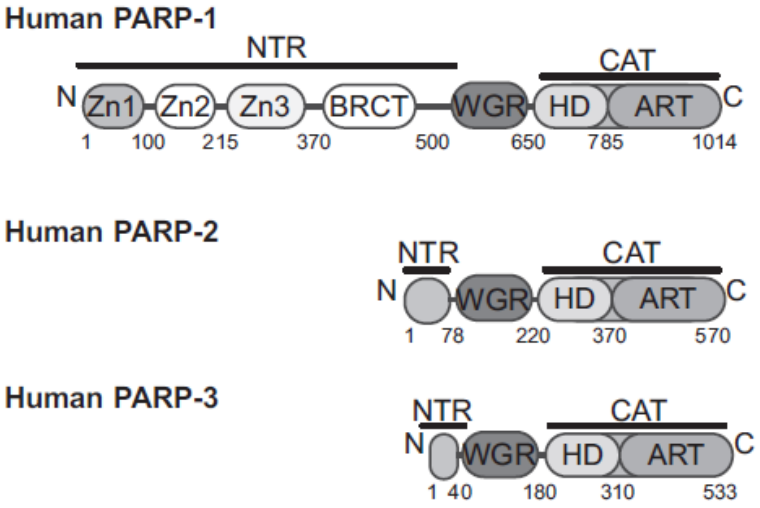


Fig. 1.5 Domain architecture of PARP-1, PARP-2 and PARP-3 The WGR and CAT domains are conserved, while the N-terminal regions (NTRs) are distinct [John M. Pascal Nucleic Acids Research, 2014, Vol. 42, No. 12]

PARP3 has been found to associate with Polycomb group proteins involved in transcriptional silencing and with DNA repair networks, including base excision repair/single-strand break repair (BER/SSBR) and nonhomologous end-joining (NHEJ), suggesting an active role for PARP3 in the maintenance of genomic integrity [Rouleau M et al, 2007]. Furthermore, PARP3 has been described as a critical player in the stabilization of the mitotic spindle and in telomere integrity notably by associating and regulating the mitotic components NuMA and Tankyrase 1 suggesting PARP3 as a positive regulator of the mitotic network containing Tankyrase 1 and NuMA with important implications in spindle dynamics and telomere integrity during mitosis [Boehler C et al, 2011].

PARP3 serves TGFβ-dependent epithelial-to-mesenchymal transition (EMT), a transdifferentiation programme that is important for organogenesis in the developing embryo, tissue injury repair and cancer progression, by stimulating a TG2-Snail-E-cadherin axis in a manner involving its response to TGFβ-induced ROS [Karicheva O et al, 2016]. The fact that PARP3 is upregulated and responds to ROS produced from both exogenous (genotoxic agents) and endogenous (TGFβ) sources makes it a major driver of the ROS response that is likely to function in various other physiological and pathophysiological events. One interesting example is suggested by the prominent role of PARP3 in neural crest cell differentiation in the zebrafish [Rouleau M et al, 2011], a process in which the generation of ROS by the

NADPH oxidase NOX4 was documented to be required for efficient differentiation [Lee JE et al, 2014].

Since point mutations in DNA repair pathways can alter repair function and contribute to cancer risk [García-Closas et al. 2006a; Goode et al. 2002], in this study we selected some nsSNVs found in cancer tissues and annotated in the COSMIC database [Forbes et al, 2010] (F125L, P147T, G148D, L233F and F440L) in order to investigate their biophysical and biochemical properties respect to its wild-type counterpart.

The data presented in this thesis will provide a comprehensive analysis of three-dimensional structure and biophysics of wild-type and nsSNVs with the purpose to give a contribution to fill the gap between the collection of thermodynamic data and disease-related information on protein variants.

Chapter 2

Aim of the study

nsSNVs are caused by missense mutations, changes in one DNA base pair that result in the substitution of an amino acid in the protein encoded by a gene, and contribute to ~80% of the total number of variants annotated as pathogenic. They are the most common type of variation among humans and pathogenic nsSNVs, or missense mutations, account for approximately half of the allelic variants causative of hereditary disease. During the last decade several databases were created collecting data associated to protein variations, above all focusing their attention on the impact of protein variants at the structural, functional and phenotypic levels. Some amino acid substitutions are functionally neutral, however most of them may have different effects on protein structure and function and can be related to several diseases. It was demonstrated that mutations can frequently affect several biophysical characteristics simultaneously and may or may not cause diseases [Gong et Blundell, 2010]. Even if the available data on single amino acid substitutions are expanding rapidly, the knowledge of the possible disease association of nsSNVs and the molecular mechanisms of genetic disease is lagging due to the limited integration between experimental data and predictions because of the laborious and time consuming nature of experimental studies. There is no clear threshold of how large the change of the wild-type characteristics should be in order to alter protein function and

result in disease. Predictions of damaging effects of mutations are further complicated by the observations that enhanced protein activity [Zhang et al, 2013], higher stability or binding affinity [Witham et al, 2011; Takano et al, 2012] are not necessarily advantageous for the cell and can be disease causing.

Nowadays several computational methods are available for predicting the possible pathogenicity of nsSNVs; they are predominantly based on evolutionary information and/or varying structural descriptors of the protein in question. These methods aim at automating the annotation process of nsSNV effects and therefore would be very useful for the mutation research community. In silico study can be useful for obtaining a statistical picture of the extent to which these SNVs influence phenotype and for prioritizing SNVs for experimental study [Hallali-Assani et al, 2009].

Computational analysis has predicted that around 30% of protein variants resulting from non synonymous single nucleotide substitutions are less stable than the wild-type. Experimental studies have been carried out on very few proteins due to the cost and time needed for the entire process that requires mutagenesis, protein expression and purification followed by thermal and chemical unfolding [Pasquo A et al, 2012; Lori C et al, 2013; Lori L et al, 2016; Gao et al, 2015; Petrosino M et al, 2017].

In this study I'm going to present a systematic and qualitative analysis of structural and functional properties of protein variants caused by single amino acid substitutions leading to non synonymous variants found in cancer tissues, including changes in enzyme activity, aggregation propensity, structural stability and binding ability. The aim of this study is the investigation of plausible correlations, if any, between alteration on protein

structure, function or stability caused by single amino acid substitutions and cancer.

During my PhD I focused my attention on three case-reports analyzing the expression, the structural characterization and stability and binding properties of some variants of three different proteins related to several cancer types:

- Phosphoglycerate kinase 1 (PGK1) is an essential enzyme that catalyzes the reversible phosphotransfer reaction from 1,3-bisphosphoglycerate (1,3-BPG) to MgADP to produce 3-phosphoglycerate (3-PG) and MgATP, an important ATP-generating step in glycolysis. Several solid tumors exhibit an increased expression of glycolytic enzymes such as PGK1 to generate ATP in hypoxic conditions [Hwang et al., 2006; Sun et al., 2015] suggesting a plausible role of PGK1 as a cancer biomarker.

- Human Frataxin (hFXN) is a nuclear-encoded protein highly conserved implicated in Fe-S cluster biosynthesis and in iron and heme metabolism. Iron is an inorganic element which is critical for cell proliferation and growth and alterations in proteins involved in the regulation of iron levels in cells can be dangerous for living organisms and can be related to carcinogenesis.

- Poly (ADP-ribose) polymerase 3 (PARP3) is a member of the PARP family, a group of enzymes that synthesize poly (ADP-ribose) on themselves or other acceptor proteins, that catalyses a post-translational modification of proteins involved in

biological processes, such as transcriptional regulation, energy metabolism and cell death [Hakmé A et al, 2008]. Mutations in DNA-repair genes can confer growth and survival advantages of cancer cells leading to DNA-repair deficient cancers.

Site-directed mutagenesis has been used to produce recombinant mutant proteins in *E. coli* cells. In this study we characterized the wild-type and the variants carrying missense mutations (nsSNVs) and identified in cancer tissues in human in order to investigate the impact of these amino acid substitutions on protein structure, conformational stability and interactions with ligands. We analyzed the effects induced by single amino acid substitution on protein structure and function, by means of structural stability, binding and activity studies. In particular we selected variants from available database such as COSMIC [Forbes et al, 2010] for somatic mutations in cancer that stores information about genetic variations on the genomic scale and aim to collect mutations in all genes.

The results obtained during my PhD will be combined with those obtained for variants of other proteins studied by our group in the last years: Bromodomains (BRDs), Peroxisome Proliferator-Activated Receptor γ (PPARγ), Serine/threonine-protein kinase pim-1 (Pim-1) and Protein tyrosine phosphatase ρ (PTPρ). . This comparative and comprehensive study aims to the evaluation of the possible effects that single amino acid substitutions can have on protein structure, function or stability. This will be useful to understand how missense variants can influence protein properties and which can be the correlation, if any, between this type of alterations and cancer.

Chapter 3

Methodology

In this study we aimed to investigate the effect or consequences that a nsSNV can induce at the level of the protein and for this reason we set an adequate experimental procedure.

We identified from available databases, such as COSMIC [Forbes et al, 2010], OMIM [Hamosh a et al, 2005], Pan-Cancer Atlas (https://www.cell.com/pb-assets/consortium/pancanceratlas/pancani3/index.html) and SNP database, nsSNVs of several proteins related to different type of cancers. We chose variants of proteins with available PDB in order to map them on the crystal structure of the protein. We categorized the nsSNVs into two main groups according to their localization on the 3D structure of the protein and considering its biological assembly: nsSNVs on surface exposed or buried residues, observing that most of the nsSNVs identified are on the protein surface and only 31% are buried into the internal part of the protein (Fig. 3.1).

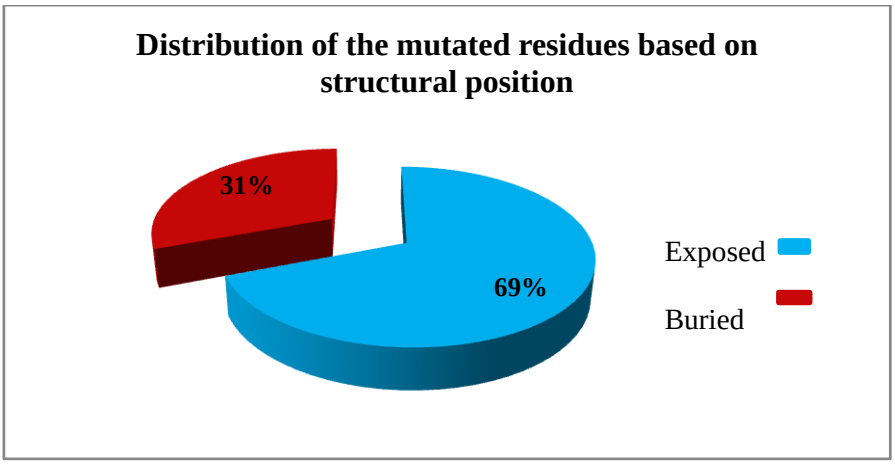


Fig. 3.1 Distribution of nsSNVs based on the structural position of the mutated residues.

We designed specific oligonucleotides in order to obtain the selected variants and we performed site-directed mutagenesis using Quick Change Site-Directed Mutagenesis Kit (Stratagene) to introduce the point mutations on wild-type plasmid. The presence of the desired mutations and the absence of unwanted mutations were confirmed by sequence analysis. Once obtained, we expressed the wild-type and the variants as soluble form in *E. coli* and then we carried out a structural characterization by means of circular dichroism and fluorescence spectroscopy. Functional analysis were performed and we used Molecular Dynamics simulations to obtain information about the dynamic properties of macromolecules and to analyze conformational ensembles and X-ray crystallography to solve the crystal structure of some of the variants obtaining three dimensional molecular features on the possible local structural alterations induced by a point mutation. Finally, we performed binding assay using BLI or ITC experiments to better understand the binding properties of the selected variants.

3.1 Protein expression and purification

The wild-type proteins and their variants were expressed in *E. coli* strain (BL21(DE3) or Rosetta cells) and cultures were grown in LB medium containing the optimal antibiotic at 37°C until OD_{600} arrived to 0.6 and then protein expression was induced with 0.5mM or 1mM isopropyl-β-d-thiogalactoside (Sigma-Aldrich). After induction, cells were grown at the optimal temperature for a period of time well set to obtain a significant amount of protein in the soluble fraction; then collected by centrifugation.

Every liter of Cell cultures were resuspended in 50 ml of appropriate Binding buffer containing 0.5-1 mM tris(2-carboxyethyl)phosphine in the presence of protease inhibitors (Complete, Roche) and disrupted by sonication. The lysate was cleared by centrifugation and the supernatant was purified using affinity chromatography with specific arrangements according to the characteristics of each protein of interest. Affinity chromatography is one of the most diverse and powerful chromatographic methods for purification of a specific molecule or a group of molecules from complex mixture, based on highly specific and reversible interactions between two molecules. Affinity chromatographic purification is frequently of great importance in the case of recombinant proteins since they are often produced in a way that they contain a fused “label” at their N- or C-terminus, resulting from genetic engineering. Thanks to that, the recombinant protein can be simply “fished out” of the cell extract via affinity chromatography. One of the most widely used of such labels fused to protein termini is the oligo-histidine tag (His-tag), which binds reversibly to metal chelates (Ni chelate immobilized on the stationary phase). With the premise that we used

plasmids which structures contain a hexahistidine tag cleavable with tobacco etch virus protease (Pro-TEV), the supernatant was loaded on a Ni-NTA (Ni2+ - nitriltriacetate) affinity resin (GE Healthcare) pre-equilibrated with Binding buffer. The column was washed with Binding buffer to elute weakly bound contaminants. The recombinant protein was eluted by passing over the column binding buffer solutions containing increasing imidazole concentrations (50 mM, 100 mM, 150 mM and 250 mM, respectively). The collected eluates were supplemented with a final concentration of 10 mM dithiothreitol (DTT) and tested for purity on SDS gel using precasted gel system (Invitrogen). The pure fractions were incubated overnight at 4°C with tobacco etch virus protease (Pro-TEV), to remove the hexahistidine tag.

For some of the obtained proteins, such as PGK1 and PARP3, it was necessary to adjust the protocol for purification by introducing an upstream step: the supernatant was loaded on a DE52 column (GE Healthcare), previously equilibrated with Binding buffer and 0.5 mM tris(2-carboxyethyl)-phosphine, to remove nucleic acids through ion exchange chromatography that separates molecules on the basis of differences in their net surface charge.

After digestion, the mixture containing TEV protease, the His-tag and the cleaved protein was applied to a 5 ml pre-packed His Trap column (GE Healthcare) previously equilibrated in Binding buffer and the flow through containing the protein without His-tag was collected and concentrated to 2 ml using Millipore concentrators. To achieve a high degree of purity, a further purification step was performed according to the characteristics of each protein of interest. In the case of PGK1, the collected 2ml were loaded onto a Superdex 200 300/10 gel filtration column previously equilibrated with 50 mM Tris/HCl, 0.25 M NaCl, 10 mM DTT, pH 7.5 at a flow rate of 1.0

ml/min; 2 ml fractions were collected and the pure protein was identified by SDS PAGE. In the case of hFXN and PARP3, the collected 2 ml were applied to a PD-10 pre-packed column (GE Healthcare). Protein concentration was determined spectrophotometrically.

3.2 Spectroscopic measurements

Intrinsic fluorescence emission measurements were carried out with a LS50B spectrofluorimeter (Perkin-Elmer) using a 1.0 cm path length quartz cuvette. Fluorescence emission spectra were recorded from 300–450 nm (1 nm sampling interval), with the excitation wavelength set at 295 nm. Far-UV (190–250 nm) CD spectra were recorded either in a 0.1 cm quartz cuvette or in a 0.5 cm quartz cuvette; near-UV (240–420 nm) CD spectra were recorded at protein concentration ranging from 1.0 to 2.4 mg/ml in a 1.0 cm quartz cuvette. The results obtained from the near-Uv CD and far-UV CD spectra were expressed as the mean residue ellipticity [θ], assuming a mean residue molecular mass of 110 per amino acid residue. All spectroscopic measurements were carried out at 10°C or 20°C.

3.3 Thermal denaturation experiments

Thermal denaturation experiments were carried out on a JASCO circular dichroism (CD) spectropolarimeter. Wild-type and variants (0.20 mg/ml) were generally heated from 20°C to 80°C in a 0.1 cm quartz cuvette with a heating rate of 1.0 and 0.5 degreexmin^{-1} controlled by a Jasco programmable Peltier element. The dichroic activity at 209 nm orat 222 nm

and the PMTV were continuously monitored in parallel every 0.5°C. All the thermal scans were corrected for the solvent contribution at the different temperatures. Melting temperature (T_m) values were calculated by taking the first derivative of the ellipticity at 208 nm or 222 nm with respect to temperature.

3.4 Urea-induced equilibrium unfolding

For equilibrium transition studies, wild-type and variants were incubated at 10°C or at 20°C at increasing concentrations of urea (0–9 M). When equilibrium was reached, and intrinsic fluorescence emission and far-UV CD spectra (0.5 cm quartz cuvette) were recorded in parallel at 10°C or at 20°C. To test the reversibility of the unfolding, wild-type and variants were unfolded at the same temperature in the optimal urea concentration for complete unfolding and, after 10 min, refolding was started by 10-fold dilution of the unfolding mixture into solutions of the same buffer used for unfolding containing decreasing urea concentrations. After incubation for a specific time, required to reach equilibrium from 2h to 24h, intrinsic fluorescence emission and far-UV CD spectra were recorded at the chosen temperature.

3.5 PGK1 enzyme activity and kinetic studies

Enzyme activity was determined at 20°C using a standard reaction mixture in a final volume of 0.5 ml. At least 10 different concentrations of substrates under identical conditions were tested. The reaction was started by adding different amounts of purified enzyme ranging from 4 ng to 20 μg. All measurements were performed in triplicate in a Lambda 16 computerized spectrophotometer (Perkin-Elmer). Kinetic data were analysed according to [Chiarelli LR et al, 2012], using GraphPadPrism 5.04.

The activity assay mixture was also incubated at increasing temperature in a thermostated cuvette. Pure enzymes, at 10°C, were added to 0.5 ml of the assay mixture equilibrated at the desired temperature to start the reaction. The solution was mixed in the thermostated cuvette and the absorbance at 340 nm was continuously monitored for 10 min. The changes of enzyme activity as a function of temperature was fitted nonlinearly to the Arrhenius equation using GraphPadPrism 5.04 to obtain the activation energies ($E_a^{\ddagger}$) for the catalytic reaction

$$k=Ae^{-Ea/RT} \qquad (1)$$

where k (s^{-1}) is the rate constant at temperature T (K), A is a reaction specific quantity, R the gas constant (1.987 cal x mol^{-1} x K^{-1}) and E_a the activation energy of the reaction, as described in [Lori C et al, 2013].

3.6 PGK1 protein crystallization, data collection, structures solution

The protein samples used for crystallization trials contained protein variants at a concentration of 12–18 mg/ml in Tris-HCl 20 mM pH 7.0, in the presence of 10 mM ADP, 25mM $MgCl_2$, 50 mM 3-PG and 10 mM DTT. Crystals were obtained at 298 K by the hanging-drop vapour diffusion method, sealing symmetric drops (1.0+1.0 μL) over 500 μL of reservoir solution. A single-wavelength data set was collected at 100 K from each crystal and processed with XDS [Kabsch W, 2010] and Aimless [Evans PR et al, 2013]. The structures were determined by molecular replacement with the program MOLREP (CCP4 suite) [Vagin A et al, 1997] using the PDB entries 2WZB and 2ZGV as search models for closed and partially closed conformations respectively. Refinement has been performed using the maximum-likelihood method with the program REFMAC [Murshudov GN et al, 1997] and model building has been done by using Coot [Emsley P et al, 2004].

3.7 hFXN stopped-flow measurements

Kinetic (un)folding experiments were carried out on a single-mixing SX-18 stopped-flow instrument (Applied Photophysics, Leatherhead, UK) the reaction being followed by tryptophan emission. The excitation wavelength was 280 nm and emission was collected using a 320 nm cut-off glass filter.

Final protein concentration was typically 2μM. The experiments were performed at 37°C with 8 M urea.

3.8 Data analysis

3.8.1 Quantitative analysis of equilibrium transition

The changes in intrinsic fluorescence emission spectra at increasing urea concentrations were quantified as the intensity-averaged emission wavelength, $\bar{\lambda}$, [Royer CA et al, 1993] calculated according to

$$\bar{\lambda} = \Sigma (I_i \lambda_i)/\Sigma(I_i) \quad (2)$$

where λ_i and I_i are the emission wavelength and its corresponding fluorescence intensity at that wavelength, respectively. This quantity is an integral measurement, negligibly influenced by the noise, which reflects changes in the shape and position of the emission spectrum. Urea-induced equilibrium unfolding transitions monitored by far-UV CD ellipticities changes was analysed by fitting baseline and transition region data to a two-state linear extrapolation model [Santoro MM et al, 1998] according to

$$\Delta G_{\text{unfolding}} = \Delta G^{H_2O} + m\,[\text{urea}] = -RT \ln K_{\text{unfolding}} \quad (3)$$

where $\Delta G_{\text{unfolding}}$ is the free energy change for unfolding for a given denaturant concentration, ΔG^{H_2O} the free energy change for unfolding in the absence of denaturant and *m* a slope term which quantifies the change in $\Delta G_{\text{unfolding}}$ per unit concentration of denaturant, *R* the gas constant, *T* the temperature and *K*unfolding the equilibrium constant for unfolding. The model expresses the signal as a function of denaturant concentration:

$$y_i = \frac{y_N + s_N[X]_i + (y_U + s_U[X]_i) * \exp\left[\left(-\Delta G^{H_2O} - m[X]_i\right) / RT\right]}{1 + \exp\left[\frac{\left(-\Delta G^{H_2O} - m[X]_i\right)}{RT}\right]} \quad (4)$$

where y_i is the observed signal, y_U and y_N are the baseline intercepts for unfolded and native protein, s_U and s_N are the baseline slopes for the unfolded and native protein, $[X]_i$ the denaturant concentration after the addition, ΔG^{H_2O} the extrapolated free energy of unfolding in the absence of denaturant, *m* the slope of a $\Delta G_{\text{unfolding}}$ versus [X] plot. Data were globally fitted with the *m* values shared between the data sets; all other parameters were not constrained. The denaturant concentration at the midpoint of the transition, $[\text{Urea}]_{0.5}$, according to equation 2, is calculated as:

$$[\text{urea}]_{0.5} = \Delta G^{H_2O}/m \quad (5)$$

All unfolding transition data were fitted by using Graphpad Prism 5.04.

Far-UV CD spectra recorded as a function of urea concentration were analyzed by a singular value decomposition algorithm (SVD) using the software MATLAB (Math-Works, South Natick, MA) to remove the high frequency noise and the low frequency random errors and to determine the number of independent components in any given set of spectra. CD spectra in the 213–250 nm or in the 250-320 nm region were placed in a rectangular matrix *A* of *n* columns, one column for each spectrum collected at each time. The *A* matrix is decomposed by SVD into the product of three matrices: $A = U*S*V^T$, where *U* and *V* are orthogonal matrices and *S* is a diagonal matrix. The *U* matrix columns contain the basis spectra and the *V* matrix columns contain the urea dependence of each basis spectrum. Both *U* and *V* columns are arranged in terms of decreasing order of the relative weight of information, as indicated by the magnitude of the singular values in *S*. The diagonal *S* matrix contains the singular values that quantify the relative importance of each vector in *U* and *V*. The signal-to-noise ratio is very high in the earliest columns of *U* and *V* while the random noise is mainly accumulated in the latest *U* and *V* columns. The wavelength averaged spectral changes induced by increasing denaturant concentrations are represented by the columns of matrix *V*; hence, the plot of the columns of *V* versus the denaturant concentrations provides information about the observed transition.

3.9 Cell culture

Human gliomablastoma cell line LN229 were maintained in Minimum Essential Media (MEM) with 10% fetal bovine serum (FBS) and 1% gentamycin (Invitrogen) in a humidified incubator with 37°C and 5% CO_2. MDA MB231 $PARP3^{-/-}$ cells and MDA MB436 $PARP3^{-/-}$ breast cancer cells were maintained in RPMI 1640 medium with 10% fetal bovine serum (FBS) and 1% gentamycin (Invitrogen) in a humidified incubator with 37°C and 5% CO_2.

3.10 Western blot analysis

For PARP3 expression analysis in LN229 cell lines, cells were collected and pellet was resuspended in hypotonic buffer (10mM Tris pH 7.3, 10mM KCl, 1.5mM $MgCl_2$, 10mM β-mercaptoethanol, 0.2mM PMSF). After centrifugation at 4000 rpm at 4°C for 5 min, the supernatant was removed and the pellet was resuspended using extraction buffer (15mM Tris pH 7.3, 0.4 mM NaCl, 1mM EDTA, 1mM $MgCl_2$, 10% glycerol, 10mM β-mercaptoethanol, 0.2mM PMSF). After incubation for 30 min. in ice, the samples were centrifuged at maximum speed at 4°C for 30 min. The cleared suspension was quantified by Bradford protein assay. Equivalent amounts of proteins were next analysed by 10% SDS-PAGE and Western blot analysis were performer using the appropriate antibodies. Bands were visualized using ECL-PLUS detection system (Amersham Biosciences) and the images were

captured using the Image Quant LAS 4000 imaging system (GE Healthcare Life Science).

3.11 Dose response

In order to induce DNA damage and thus to observe the involvment of PARP3 in DNA repair, we treated our cell lines with chemotherapeutic drugs, such as Temozolomide (TMZ) and Phleomycin (Phleo). We evaluate the IC_{50} in order to decide the concentration to use for the kinetic assay.

3.12 Clonogenic assay

Clonogenic assay, or colony formation assay, is an in vitro cell survival assay based on the ability of a single cell to grow into a colony. We analyzed the colony formation propensity of each cell line. After 10 days we fixed and stained the colonies:

1. Gently remove the media from each of the plates by aspiration.
2. Wash each plate with 5 ml 0.9% saline.
3. Fix the colonies with 5 ml 10% neutral buffered formalin solution for 15-30 minutes.
4. Stain with 5 ml 0.01% (w/v) crystal violet in dH_2O for 30-60 minutes.
5. Wash excess crystal violet with dH_2O and allow dishes to dry.

Finally digital images of the colonies were obtained using a camera or scanning device and colonies were counted using imaging analysis software.

3.13 PARP3 transfection in human breast cancer cell lines

MDA MB231 PARP3$^{-/-}$ cells and MDA MB436 PARP3$^{-/-}$ breast cancer cells were transfected with an empty vector (pEGFP), with the PARP3 wild-type (pEGFPhP3) and with two variants of the PARP3 WGR domain (pEGFPhP3-P147T, pEGFPhP3-G148D) and with one variant of the PARP3 CAT domain (pEGFPhP3-F440L).

300000 cells were plated in a P60 plate and transfection was induced after 24h, when a confluence of 50-80% was reached. In an eppendorf tube the necessary volumes of OptiMEM medium and Fugene 6 were mixed and then the plasmids with DNA (5µg) were added. After incubation for 15 min at room temperature, the mixtures were added using a pipette while trying to distribute the drops on the entire surface of the P60 and at the end, and the plates were moved by circular motions for a better distribution of the solution. The transfected cells were mainteined at 37°C in a humidified incubator with 37°C and 5% CO_2. The cell culture medium was changed 48 hours after transfection and the stable transfectants were selected with neomycin at 500 µg/ml and 250 µg/ml in MDA MB231 cells and in MDA MB436 cells respectively. Breast cancer cell lines were maintained in RPMI 1640 medium with 10% fetal bovine serum (FBS) and 1% gentamycin (Invitrogen) in a humidified incubator with 37°C and 5% CO_2.

Chapter 4

Results

Point mutations can induce multiple effects on a given protein: nsSNVs can be beneficial, adding new functionality and increasing the fitness of the cell or deleterious, damaging the protein, causing destabilization of the protein structure or reducing its functional activity because of the disruption of a site that is directly involved in the function of the protein and thereby contributing to disease phenotypes [Studer et al, 2013]. Change in amino acids with similar size and physico-chemical properties (e.g. substitution from leucine to valine) has a mild effect. Similarly, if the point mutation disrupts secondary structure elements (e.g. substitution to proline in alpha helix region) such mutation usually may affect protein structure, function and stability. Interestingly, several nsSNVs have been found in different types of cancer leading researchers to some questions: "how missense variants can influence protein properties?

In order to give an overview of the possible effect or consequences that a single amino acid substitution can have on a given protein, we analyzed 45 nsSNVs of different proteins found in cancer tissues. We chose proteins with available crystal structures in order to locate the position of the amino acid change present in nsSNV within them. In the last years we published some works regarding this topic in which we analyzed the biochemical properties of several nsSNVs of some proteins that seems to be widely mutated in different cancer types: the Bromodomains (BRDs) [Lori L et al,

2016], the Peroxisome Proliferator-Activated Receptor γ (PPARγ) [Petrosino M et al, 2017], the Serine/threonine-protein kinase pim-1 (Pim-1) [Lori C et al, 2013] and the Protein tyrosine phosphatase ρ (PTPρ) [Pasquo A et al, 2012]. I will combine the results obtained for these proteins with that obtained for the variants of PGK1, hFXN and PARP3, that I analyzed during my PhD, in order to set a comparative work with the aim to fill the gap between the collection of biochemical data and cancer-related information on protein variants.

For each protein we chose variants with an high pathogenic score (Table 4.1), a significant value generally used to give reason of the pathogenic rate. It is derived from the new FATHMM-MKL (http://fathmm.biocompute.org.uk/) algorithm that predicts the functional, molecular and phenotypic consequences of protein missense variants using hidden Markov models. The functional scores for individual mutations from FATHMM-MKL are in the form of a single p-value, ranging from 0 to 1. Scores above 0.5 are deleterious, but in order to highlight the most significant data in COSMIC, only scores ≥ 0.7 are classified as 'Pathogenic'. Mutations are classed as 'Neutral' if the score is ≤ 0.5. [Shihab Hashem A et al, 2015].

Protein	Variant	Pathogenic score	Primary Tissue
BRD4(1)	A89V	0.95	Large intestine
BRD2(1)	R100L	0.93	Lung
BRD2(1)	E140K	0.99	Lung
BRD2(1)	Y153H	0.99	Endometrium
BRD2(1)	D160N	0.99	Ovary
BRD2(1)	D160Y	0.99	Ovary
BRD2(2)	G355D	0.97	Large intestine
BRD3(2)	H395R	0.96	Lung

BRD2(2)	R419W	0.89	Skin
BRD2(2)	P430S	0.96	Large intestine
BRD2(2)	Q443H	0.99	Large intestine
BRD4(2)	A420D	0.99	Lung
hFXN	D104G	0.96	Liver
hFXN	A107V	0.85	Large intestine
hFXN	F109L	0.72	Endometrium
hFXN	Y123S	0.96	Upper aerodigestive tract
hFXN	S161I	0.99	Endometrium
hFXN	W173C	0.99	Oesophagus
hFXN	S181F	0.95	Skin
hFXN	S202F	0.50	Skin
PARP3	F125L	0.97	Large intestine
PARP3	P147T	0.72	Breast
PARP3	G148D	0.97	Liver
PARP3	L233F	0.93	Oesophagus
PARP3	F440L	0.97	Lung
PGK1	R39M	0.99	Lung
PGK1	R66W	0.94	Liver
PGK1	G167D	0.98	Endometrium
PGK1	M190I	0.99	Breast
PGK1	A200V	0.98	Breast
PGK1	V217F	0.98	Breast
PGK1	F242S	0.91	Breast
PIM1	E124K	0.99	Urinary tract
PIM1	V126A	0.91	Urinary tract
PIM1	E135K	0.97	Haematopoietic and lymphoid
PIM1	E142D	0.88	Haematopoietic and lymphoid
PPARγ	Q286P	0.90	Colon
PPARγ	R288H	0.95	Colon

PPARγ	E324K	0.99	Endometrium; Skin
PPARγ	R357C	0.88	Lung
PPARγ	E460K	0.73	Lung
PTPρ	D927G	0.91	Colon
PTPρ	Q987K	0.91	Colon
PTPρ	A1118P	0.89	Colon
PTPρ	N1128I	0.91	Rectum

Table 4.1 Forty-five analyzed nsSNVs. List of the analyzed nsSNVs with their pathogenic score and the primary tissue in which they have been found.

By determining the location of the amino acid change on the protein structure, we manually categorized the nsSNVs into two groups: 69% of the nsSNVs involved solvent exposed residues and 31% were buried into the protein (Fig. 3.1).

4.1 Experimental analysis of PGK1 nsSNVs

In order to understand the structural and functional consequences of non synonymous changes, in this study we selected seven PGK1 variants (R38M, R65W, G166D, M189I, A199V, V216F and F241S) found in cancer tissues and annotated in the COSMIC database and we report the effect of these variants analyzing the structure, the conformational stability, the protein folding and the kinetic behaviour compared with its wild-type counterpart. The selected variants involved residues located in the 3-PG binding domain, in the nucleotide binding domain and in the hinge region that connects the two domains (Fig. 4.1).

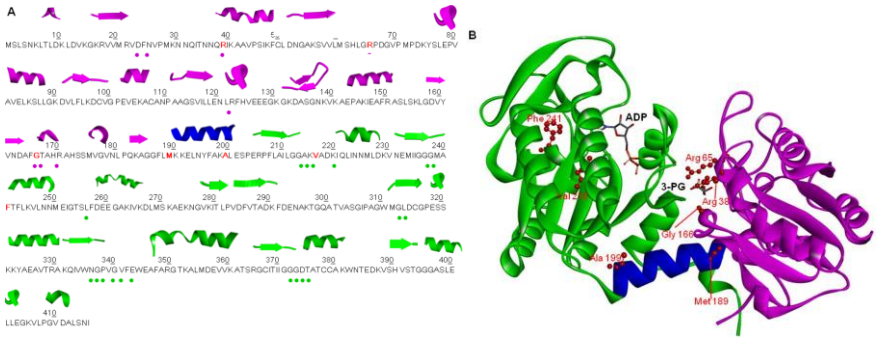


Fig. 4.1 Amino acid sequence and structure of PGK1. N-terminal domain (violet), C-terminal domain (green), hinge region (blue) (PDB: 2XE7, open conformation). (A) Secondary structural elements are shown at the top of the amino acid sequence. Mutated residues are depicted in bold red. The dots under the sequence represent the residues involved in 3-PG or 1,3-BPG binding (pink dot) and in ATP or ADP binding (green dot). (B) Location of the mutations on PGK1 structure (open conformation). Mutated residues are depicted in scaled ball and stick and 3-PG and ADP in stick. https://doi.org/10.1371/journal.pone.0199191.g001

4.1.1 Spectroscopic characterization of PGK1 wild-type and variants

Using CD and fluorescence spectroscopy we performed a detailed analysis of the conformation in solution of all PGK1 variants. The near-UV CD spectrum of PGK1 wild-type represent the spectral contributions of all aromatic residues and is characterized by a strong positive peak centred at around 292 nm flanked by a shoulder at 286 nm and by negative contributions in the region between 255 and 270 nm, with two peaks centered at 262 and 268 nm (Fig. 4.2A). We observed only small conformational

changes: R38M, G166D, A199V and V216F display near-UV CD spectra closely similar to that of the wild-type, except for a slight difference at 277 nm observed for G166D, however significant differences are observed in the near UV CD spectra of M189I, F241S and R65W. In the case of M189I and F241S, the contributions at 262 nm and 268 nm are positive; in addition, the spectrum of M189I loses fine structure features in the region between 270 and 280 nm. The variant R65W shows a near-UV CD spectrum that significantly differs from that of the wild-type in intensity, consistent with the substitution of an arginine residue with a tryptophan. The fluorescence spectra of wild-type and variants are centred at the same maximum emission wavelength at around 350 nm, characteristic of tryptophan contribution, with differences in the relative fluorescence emission intensity which is significantly decreased for A199V and increased for R65W, M189I and V216F (Fig. 4.2B).

The far-UV CD spectra of PGK1 wild-type and variants show the typical local minimum contributions of alpha helical proteins at around 208 nm and at 222 nm (Fig. 4.8C). The molar ellipticity ratio at 222 and at 208 nm ($[\Theta]_{222}/[\Theta]_{208}$) is indicative of interhelical contacts present in helix bundle and coiled coil structures and it is generally used to distinguish between coiled coil helices ($\geq$1.0) and non-interacting helices (0.8-0.9) [Shermann MA et al, 1992; Choy N et al, 2003]. The $[\Theta]_{222}/[\Theta]_{208}$ is 1.2 for the wild-type, it is decreased to 1.1 only for the variants R38M and M189I and it is increased to 1.3 for V216F. Notably, in the case of M189I the far-UV dichroic activity is decreased with respect to that of the wild-type (Fig. 4.2C). These differences in secondary structure may suggest differences in interhelical interactions and dynamic fluctuation in solution for some of PGK1 variants.

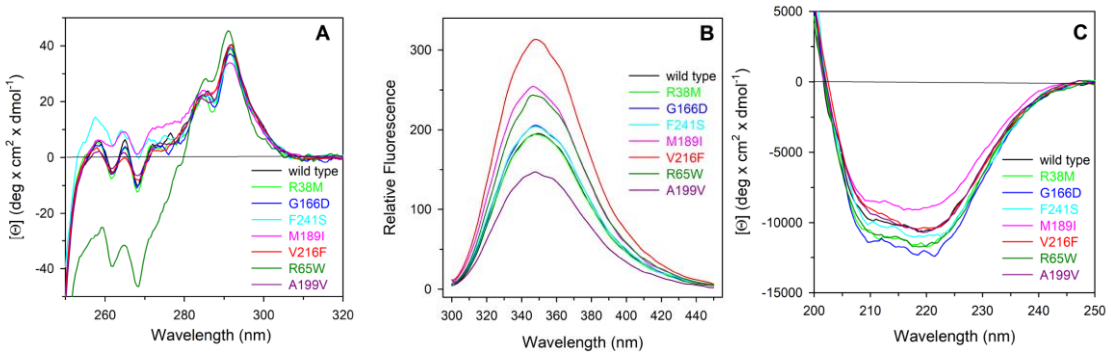


Fig. 4.2 Spectroscopic characterization of PGK1 wild-type and variants. (A) Near-UV CD spectra were recorded in a 1.0-cm quartz cuvette at 1.5-1.7 mg/ml protein concentration in 20 mM Tris-HCl pH 8.0 containing 2.0 mM DTT, 1 mM EDTA and 200 mM NaCl. (B) Intrinsic fluorescence emission were recorded at 110 μg/ml for R65W and 130 μg/ml for wild-type and the other variants spectra (0.08 AU280nm, 295 nm excitation wavelength), in 20 mM Tris-HCl, pH 8.0 containing 0.2 M NaCl and 0.2 mM DTT. (C) Far-UV CD spectra were recorded in a 0.1-cm quartz cuvette at 130-170 μg/ml in 20 mM Tris-HCl, pH 7.5 containing 0.2 M NaCl and 0.2 mM DTT. All spectra were recorded at 20°C.

4.1.2 Kinetic analysis

The effect of non synonymous single nucleotide substitutions on the kinetic properties of PGK1 enzyme was investigated performing the reverse reaction at 20°C at a fixed concentration of free Mg^{2+} and using 3-PG as variable substrate. All the variants displayed a reduced catalytic efficiency analyzed using 3PG as variable substrate over the concentration range 0.05-15.0 mM, with the exception of F241S that displays a two fold increase of activity (Table 4.2). The decrease of the enzyme activity was particularly

evident for the variant R38M, that showed a decrease of eight order of magnitude, due to a dramatic decrease of the turnover number and to an increase of K_M. R38, located in the 3-PG binding site, is critical for the positioning of the substrate in the catalytic site and for the stabilization of the transition state because it interacts electrostatically with hydrogen bound with the carboxyl group of 3-PG.

The temperature dependence of the kinase activity of PGK1 wild-type and its variants using a low concentration of 3- PG substrate (0.02 mM, well below the K_M value) was analysed over the temperature range of 10–42°C (Fig. 4.3) to study the enzymatic release of substrate and product that may be related to its flexibility. The optimal temperatures for catalysis in these conditions were estimated to be 37°C for the wild-type and most of the variants, around 35°C for F241S, and around 40°C for R38M and V216F (Table 4.3 and Fig. 4.3). As shown in Table 4.3, the lowest value of the activation energy ($E_a^{\ddagger}$) is that of the F241S variant (6.39 kcal·mol^{-1}) which is also the variant displaying a k_{cat} higher than that of the wild-type enzyme (213.0 vs 89.8 s^{-1}). As mentioned before, the activation energies have been measured at a 3-PG concentration of 0.02 mM, well below the K_M value. At this substrate concentration, the enzyme is not saturated, the complex enzyme-3-PG-ATP may dissociate to form again the enzyme-ATP complex and free 3-PG or may catalyse the formation of the two products ADP and 1,3-BPG. The activation energy measured in these conditions reflects the enzymatic release of both the substrate and the product and it may be related to the enzyme flexibility.

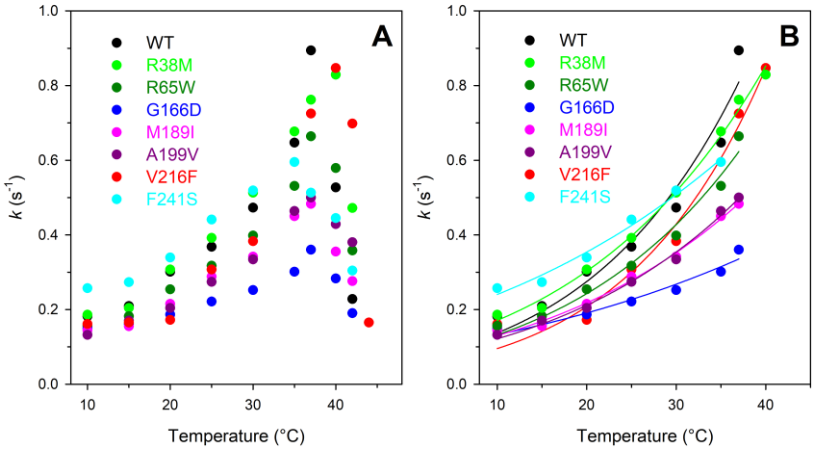


Fig. 4.3 Effect of temperature on kinase activity of PGK1 wild-type and variants. (A) Temperature dependence of kinase activity of PGK1 wild-type and variants. (B) Non-linear fit of the temperature dependence of PGK1 activity to the Arrhenius equation Eq (1). Assays were performed under the conditions described in Materials and Methods, using 0.18-12.0 nM enzyme. https://doi.org/10.1371/journal.pone.0199191.g002

	[PGK1] (nM)	[3PG] (mM)	K_M (mM)	*k*cat (s-1)	*k*cat/ K_M (s-1 mM-1)	Vmax (μM/min)	Specific activity (μM/min•μg)
Wild type	0.18	0.05–15.0	0.40 ± 0.03	89.8 ± 1.5	224.5	0.97	248.7
R38M	9.04 x10^{5}	3.00–15.0	3.15 ± 0.40	(7.2 ± 0.5) x10^{-6}	2.28x10^{-6}	0.39	0.02
R65W	0.35	1.00–15.0	1.61 ± 0.13	23.7 ± 0.5	14.7	0.50	63.8
G166D	0.70	0.90–15.0	1.54 ± 0.15	5.9 ± 0.2	3.8	0.25	15.9
M189I	0.18	0.20–15.0	0.58 ± 0.50	30.5 ± 1.4	52.7	0.33	84.6
A199V	0.18	0.80–15.0	1.95 ± 0.15	76.6 ± 1.7	39.3	0.83	212.0
V216F	7.00	0.10–15.0	0.78 ± 0.05	1.9 ± 0.03	2.4	0.93	5.5
F241S	0.18	0.20–15.0	5.87 ± 0.7	213.0 ± 14.0	31.5	2.30	515.4

Table 4.2 Kinetic parameters of PGK1 wild-type and variants. PGK1 activity was determined at 20°C, with 3-PG and 5 mM MgATP (fixed substrate concentration), as described in Materials and Methods. Kinetic parameters for PGK1 activity were determined at 20°C by using at least at 10 different concentrations of 3-PG. Data are reported as the mean ± SE of the fit. https://doi.org/10.1371/journal.pone.0199191.t001

	T_{max} (°C)	E_a (kcal/mol)
Wild type	37	11.47 ± 0.7
R38M	40	9.44 ± 0.34
R65W	37	10.03 ± 0.60
G166D	37	6.00 ± 0.30
M189I	37	8.50 ± 0.35
A199V	37	9.11 ± 0.39
V216F	40	12.82 ± 0.70
F241S	35	6.39 ± 0.41

Table 4.3 Effect of temperature on kinase activity of PGK1 wild-type and variants $E_a^{a\ddagger}$ was determined by the Arrhenius equation Eq (1) in the temperature range between 10 °C and the optimal temperature of each protein. https://doi.org/10.1371/journal.pone.0199191.t002

4.1.3 Thermal and thermodynamic analysis

The thermal stability of PGK1 wild-type and variants was investigated in the temperature range between 20°C and 80°C by continuously monitoring the ellipticity changes at 222 nm, where the main amplitude was observed (Fig. 4.4A,B). The thermal denaturation process is irreversible and occur in an apparent cooperative transition. The parameter chosen to compare the transition curves of the proteins is the melting temperature (T_m) defined as the midpoint of the denaturation process and calculated by plotting the first derivative of the molar ellipticity values as a function of temperature (Fig. 4.4A, insert). T_m values range from 49.0°C to 53.5°C (Table 4.4). The variant R65W shows the same T_m value of the wild-type, a modest increase in T_m values is observed for the variants G166D and A199V; all the other variants show T_m values lower than that of the wild-type, with F241S showing a T_m value three degrees below that of the wild-type (Table 4.4). Notably, the

differences in the amplitude observed for the thermal transitions of most of the variants (4.4B) may be referred to the difference in the dichroic activity at 222 nm of their corresponding native states, as also indicated in far-UV CD spectra reported in Fig. 4.2C.

The thermodynamic unfolding parameters of PGK1 wild-type and variants have also been detected, since they reversibly unfold in urea at 20°C. The effect of increasing urea concentrations (0–8 M) on the protein structure was analyzed by far-UV CD and fluorescence spectroscopy (Fig. 4.4C,D). The fluorescence changes at increasing urea concentration were measured by calculating the intensity averaged emission wavelength $\bar{\lambda}$, an integral measurement that depends both on the position and the shape of the spectrum. The same samples used to monitor the fluorescence changes during the unfolding transitions were used to monitor the far-UV CD ellipticty changes, to allow a direct comparison of the data. The urea-induced changes in 222 nm ellipticity and in the emission fluorescence show a sigmoidal dependence on urea concentration and follow an apparent two-state transition without any detectable intermediate (Fig. 4.4C,D). The unfolding process is fully reversible upon dilution of the denaturant either for the wild-type and variants. The thermodynamic parameters ΔG^{H_2O}, the free energy change for unfolding in the absence of denaturant, and *m*, a parameter that refers to the amount of protein surface area exposed to the solvent during unfolding, have been obtained using a two-state model (Table 4.4). ΔG^{H_2O} values of PGK1 variants are similar to those of the wild-type with the exception of G166D, M189I and F241S that show a significant decrease in the thermodynamic parameters (Table 4.4), mainly due to a decrease in *m* values which suggests either a destabilization of the native state or the presence of unfolding

intermediates not detectable at the equilibrium. The occurrence of undetectable unfolding intermediates may be the case of R38M and G166D as suggested by the non coincidence of the $[Urea]_{0.5}$ values obtained by far-UV CD and fluorescence changes, mainly due to the differences in m value (Table 4.4).

	T_m (°C)	ΔG^{H_2O} (kcal·mol^{-1})		*m* (kcal·mol^{-1}·M^{-1})		$[Urea]_{0.5}$ (M)	
		CD ($[\Theta]_{222}$)	Fluorescence ($\bar{\lambda}$)	CD ($[\Theta]_{222}$)	Fluorescence ($\bar{\lambda}$)	CD ($[\Theta]_{222}$)	Fluorescence ($\bar{\lambda}$)
Wild-type	52.5	8.29 ± 0.57	8.06 ± 0.46	3.54 ± 0.25	3.10 ± 0.17	2.34	2.60
R38M	51.5	6.91 ± 0.73	5.54 ± 0.37	3.54 ± 0.37	1.87 ± 0.12	1.95	2.96
R65W	52.5	7.04 ± 0.32	7.20 ± 0.66	3.07 ± 0.14	2.63 ± 0.24	2.29	2.73
G166D	53.0	4.19 ± 0.31	4.94 ± 0.29	2.33 ± 0.16	1.51 ± 0.09	1.75	3.27
M189I	50.5	4.87 ± 0.31	5.01 ± 0.48	2.36 ± 0.15	1.82 ± 0.17	2.06	2.75
A199V	53.5	7.14 ± 0.57	6.82 ± 0.72	3.41 ± 0.14	2.71 ± 0.22	2.23	2.63
V216F	52.0	6.45 ± 0.34	5.76 ± 0.54	2.99 ± 0.16	2.14 ± 0.20	2.16	2.69
F241S	49.0	3.02 ± 0.17	3.80 ± 0.26	2.06 ± 0.10	1.67 ± 0.11	1.47	2.27

Table 4.4 Melting temperatures and thermodynamic parameters for urea-induced unfolding equilibrium of PGK1 wild-type and mutants measured by far-UV CD and fluorescence spectroscopy. The temperature-induced changes were followed by monitoring the ellipticity at 222 nm. The T_m values were calculated by taking the first derivative of the ellipticity at 222 nm with respect to temperature. Urea-induced unfolding equilibrium data were measured at 20°C by monitoring ellipticity at 222 nm [Θ_{222}] and fluorescence intensity averaged emission wavelength $\bar{\lambda}$. ΔG^{H_2O} and *m* values were obtained from Eq 3; $[Urea]_{0.5}$ was calculated from Eq 5. Data are reported as the mean ± SE of the fit.

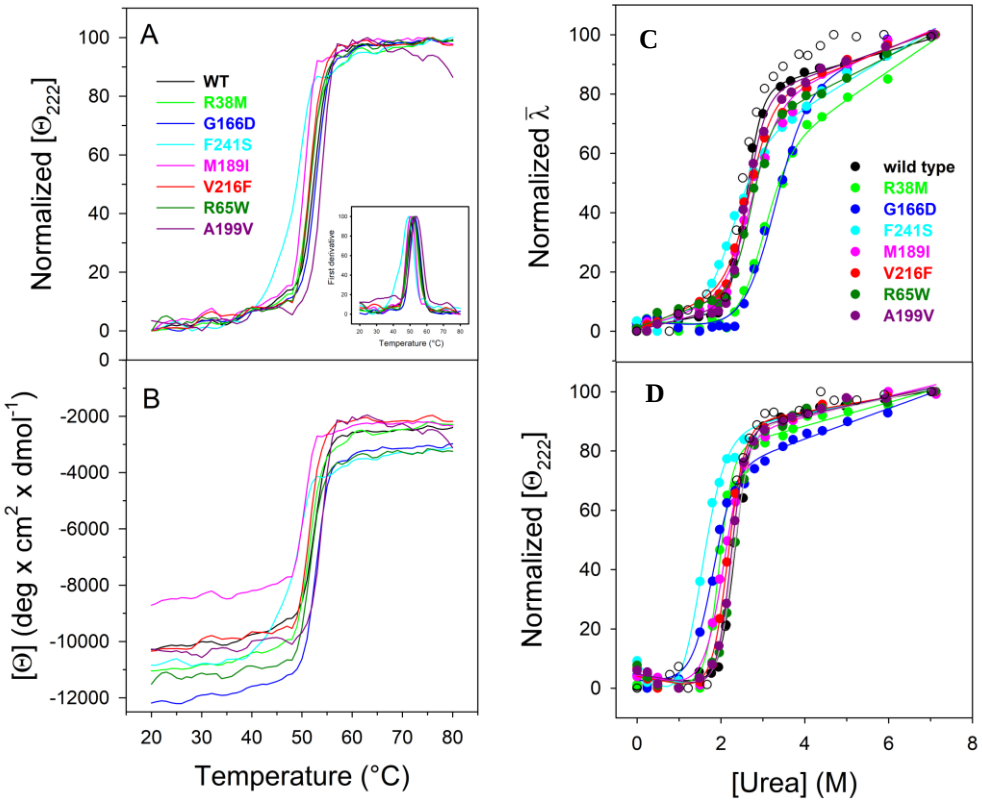


Fig. 4.4 Thermal unfolding and urea-induced equilibrium unfolding of PGK1 wild-type and variants. PGK1 variants and wild-type (130 - 170 µg/ml) were heated from 20°C to 80°C in 20 mM Tris-HCl, pH 8.0 containing 0.2 M NaCl and 0.2 mM DTT . The molar ellipticity at 222 nm ($[\Theta_{222}]$) was monitored continuously every 0.5 °C. (A) Normalized $[\Theta_{222}]$; the inset shows the first derivative of the same data as in (A). (B) $[\Theta_{222}]$ before normalization. of PGK1 wild-type and variants. (C) Normalized intensity-averaged emission wavelength $\bar{\lambda}$). (D) Normalized molar ellipticity at 222 nm ($[\Theta_{222}]$) reported after of the high-frequency noise and the low-frequency random error by SVD. The continuous lines represent the nonlinear fitting of the normalized $\bar{\lambda}$ and $[\Theta_{222}]$ data to Eq 4. The reversibility points (empty circles) are shown, for clarity, only for the wild-type and were not included in the nonlinear regression analysis. All spectra were recorded at 20 °C.

4.1.4 Structural analysis

We investigated the structure of PGK1 variants in the presence of Mg-ADP and 3-PG through X-ray crystallography in order to reveal possible local and global variations with respect to the wild-type protein. We succeed in solving the X-ray crystal structures of the variants with mutation in the N-terminal region as R38M and G166D, in the C-terminal region as V216F and in the hinge region as M189I. We solved the structures of all these mutants in the partially closed conformation, and we succeed to solve the structure of M189I also in the totally closed conformation. The overall structure is conserved in all the mutants compared to the wild-type (Table 4.5). The details of local variation are discussed case-by-case.

Mutant	Domain	Role/position of mutated residue	Conformation	Ligands	RMSD[#] (Å^2)
R38M	3-PG-binding	Binding of 3PG, charge balancing in the transition state	Partially closed	Mg-ADP	vs 2ZGV*: 0.23 vs 2XE7*: 0.96
G166D	3-PG-binding	Close to 3PG COO^-	Partially closed	Mg-ADP	vs 2ZGV*: 0.62 vs 2XE7*: 1.17
V216F	ADP-binding	Part of the catalytic loop 211-219 that favors closed conformation	Partially closed	Mg-ADP, 3-PG	vs 2ZGV*: 0.60 vs 2XE7*: 1.22
M189I	Hinge	N-terminal portion of the hinge	Partially closed	Mg-ADP, 3-PG	vs 2ZGV*: 0.60 vs 2XE7*: 1.18
			Closed	Mg-ADP, 3-PG	vs 2WZB*: 0.25

Table 4.5 Superimposition between the PGK1 wild type and the crystallized variants. RMSD[#]: root mean square deviation obtained from the superimposition of each variant with respect to wild-type PGK1. *2ZGV: wild-type binding Mg-ADP; 2XE7: wild-type binding Mg-ADP and 3-PG; 2WZB: wild-type binding Mg-ADP, 3-PG and MgF_3^-

The residue R38 is placed in the N-terminal 3-PG binding domain and is a residue important for the substrate binding and its correct positioning to react with ADP (Fig 4.5). Moreover, together with residues K215 and K219, R38 is critical for charge balancing of the transition state, directly interacting with the transferring phosphate group in the closed conformation of PGK1.

The crystal structure of R38M does not contain 3-PG (Fig. 4.5A), in line with the role of R38. This finding could account even for the reluctance to crystallize in the closed form since the binding of both substrates is required for the stabilization of this conformation. Taken together these observations justify the dramatic effect of R38M mutation on kinetic parameters (see Table 4.2). Indeed, upon R38M mutation the K_M increases from 0.40 to 3.15 mM and the turnover number strongly decreases from 89.8 to $7.2x10^{-6}$ s^{-1}.

Apart for the mutation, the position of the residues in the 3-PG binding pocket is very similar to that of the wild-type without 3-PG (PDB code: 2ZGV, Fig. 4.5B). The only differences between the two structures concern the α-helix 374-382, visible in the R38M variant but not in the wild-type 2ZGV structure, and the position of the β-phosphate group, which in R38M points towards the helix (Fig. 4.5A). However, both the helix and the phosphate present some variability upon different PGK1 structures indicating that it could be dependent on crystallization conditions.

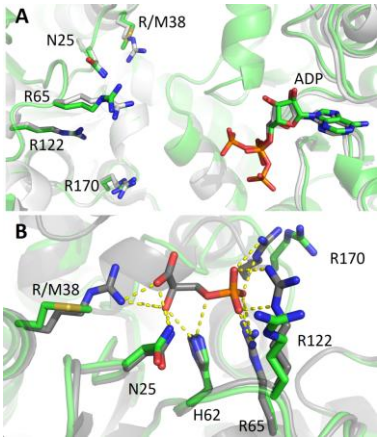


Fig. 4.5 Detail of the 3-PG binding site in the variant R38M in comparison with PGK1 wild-type. Overlap of R38M variant (green) and PGK1 wild-type crystallized (A) in the absence of 3-PG (PDB 2ZGV) (light grey) or (B) in the presence of 3-PG (PDB: 2XE7) (grey). https://doi.org/10.1371/journal.pone.0199191.g003

G166 is one of the residues lining the 3-PG binding site and its substitution for the bulky and negatively charged aspartate modifies the shape and the charge of the substrate binding pocket. The structure of G166D (Fig 4.6) reveals that the aspartate engages R38 in a salt bridge (OD-(D166)-NH1(Arg38)=3.3Å) altering its capability of interaction. These modifications of the pocket prevent the binding of 3-PG, that is not present in the structure, and the stabilization of the transition state through R38, thus affecting both k_{cat} (5.9 s-1) with a 15-fold decrease, and K_M (1.54 mM) increased 4- fold.

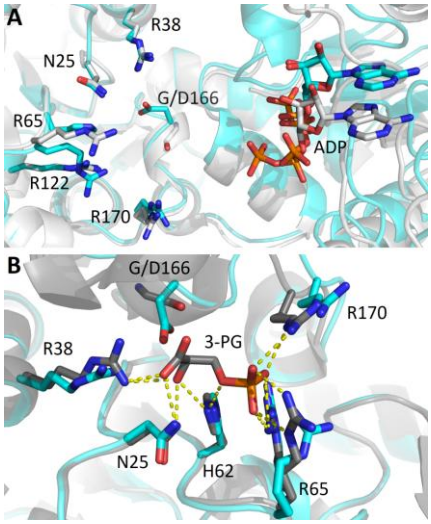


Fig. 4.6 Detail of the 3-PG binding site in the variant G166D in comparison with PGK1 wild-type. Overlap of G166D variant (cyan) and PGK1 wild-type crystallized (A) in the absence of 3-PG (PDB 2ZGV) (light grey) or (B) in the presence of 3-PG (PDB: 2XE7) (grey). https://doi.org/10.1371/journal.pone.0199191.g004

V216 belongs to the loop 214-219, close to the ADP-binding site. The rearrangement of this loop has been suggested to be important for the preparation of closed conformation by binding of ligands [Zerrad L et al, 2011]. Indeed, upon ADP-binding to wild-type the loop assumes two conformations, promoting the reorientation of K215 and D218, now ready to interact with γ-phosphate of 3-PG and D65, in the closed conformation. In the variant V216F, the loop adopts the same conformation as in apo-PGK1. As shown in Fig. 4.7B the substitution of the valine with a phenylalanine keeps the loop in this conformation since the aromatic residue tends to stay buried. The lower mobility of the loop hampers the transition from the open to the closed conformation affecting the enzyme catalytic activity (the k_{cat} of the

mutant is 1.9 s^{-1}, 45-fold lower than that of the wild-type). Conversely, no variation is found in ADP-binding. In fact, as shown in Fig. 4.7, the residues directly involved in the binding (E343, G312, A214) and the Mg(II) ion bound trough the β and γ-phosphate and D374 conserve the same positions as in the wild-type structure.

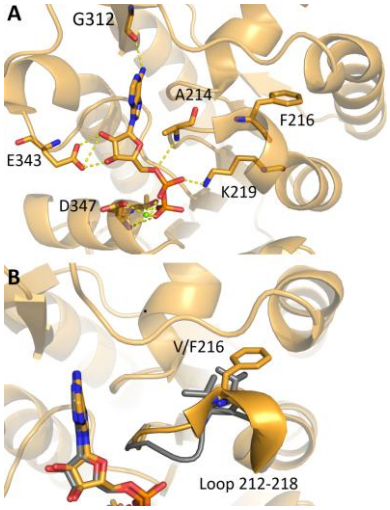


Fig. 4.7 ADP binding site in the variant V216F in comparison with PGK1 wild-type. (A) Detail of the binding of ADP to V216F variant (orange). (B) Superimpostition with wild-type PGK1 crystallized with ADP and 3-PG (PDB: 2XE7) (grey), highlighting how the mutation V216F impairs the movement of the loop 212-218. The residue V/F216 and ADP are shown as sticks. https://doi.org/10.1371/journal.pone.0199191.g005

Residue 189 is part of helix 189-201, the hinge connecting the N-terminal 3-PG binding domain and the C-terminal ADP binding domain. The structure of M189I has been solved both in open and closed conformation. As for the other variants, M189I partially open structure is coincident to that of the wild-type but in this case both ADP and 3-PG are clearly visible. It should

be noticed that in the M189I the hydroxyl carboxylate moiety of the 3-PG is 90°-rotated with respect to the position adopted in the wild-type structure. This could be due to the presence, in the active site, of a phosphate ion which may induce a change in the mode of 3-PG binding (Fig. 4.8A). Since M189I mutation causes neither overall nor local conformational change, the K_M of the mutant is very similar to that of the wild-type protein (0.58 mM *vs* 0.40 mM) and the k_{cat} is only three times lower than that of wild-type protein (30.5 *vs* 89.8 s^{-1}). The structure in closed conformation was obtained in the same condition of the wild-type protein (PDB 2WZB). As shown in Fig. 4.8B, the structure appear to be identical to the wild-type ones; also the position of the two ligands, ADP and 3-PG, is conserved. Interestingly, in the M189I structure in closed conformation MgF_3^- is not present and therefore the transition complex does not form in the M189I variant. Nevertheless, all the residues stabilizing the transition state conserve the same positions.

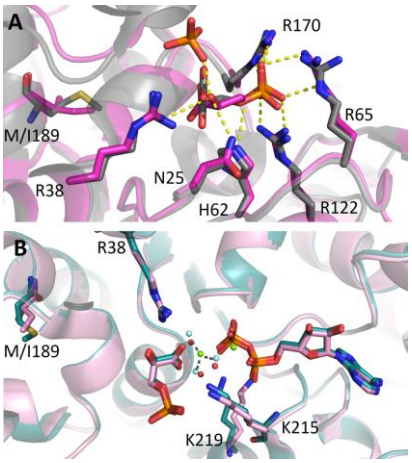


Fig. 4.8 M189I variant in comparison with PGK1 wild-type. (A) Detail of the binding of 3-PG to M189I variant in the partially-closed conformation (magenta) (B) Superimpostition of M189I variant in the closed conformation (pink) with wild-type PGK1 crystallized with

ADP, 3-PG and MgF_3^- (PDB: 2WBZ) (deep teal). MgF_3^- is shown as green and cyan spheres. https://doi.org/10.1371/journal.pone.0199191.g006

4.2 Experimental analysis of hFXN nsSNVs

The eight human Frataxin (hFXN) variants, associated to human carcinoma reported in this study were mined from the COSMIC database (http://cancer.sanger.ac.uk/cosmic) [Forbes SA et al, 2017]. The variant D104G is reported to be present in liver carcinoma, A107V, Y123S and W173C are reported in carcinoma from diverse locations of the digestive tract, F109L and S161I from endometrium carcinoma and S181F and S202F from skin malignant melanoma and carcinoma, respectively. In Figure 4.9 is reported the location of the selected mutants mapped onto the hFXN structure. It is noteworthy that all the residues are solvent exposed. Residues D104, A107 and F109 are located in the α1 helix, which belongs to the hFXN acidic region [Correira et al., 2008, Ramazzotti et al., 2004] formed at the interface with strand β1, a region placed at the protein surface, and presumably involved in iron binding [Ramazzotti et al., 2004]. The residues Y123 and S161, located in the coils between helix α1 and strand β1, and β5 and β6 respectively, are both the last coil residues prior to the next strend. Residues W173 and S181 belong to strand β6 and β7, respectively and S202 to the turn at the C-terminus. Notably the residues Y123, W173 and S202 are mutated in Friedreich's ataxia (FRDA) and reported in COSMIC. For each of the identified mutant we generated recombinant protein using site-directed mutagenesis. With the exception of W173C, all mutations resulted in soluble recombinant proteins and allowed us to investigate the results of single amino

acid substitution on the hFXN thermal and thermodynamic stability and the biological activity.

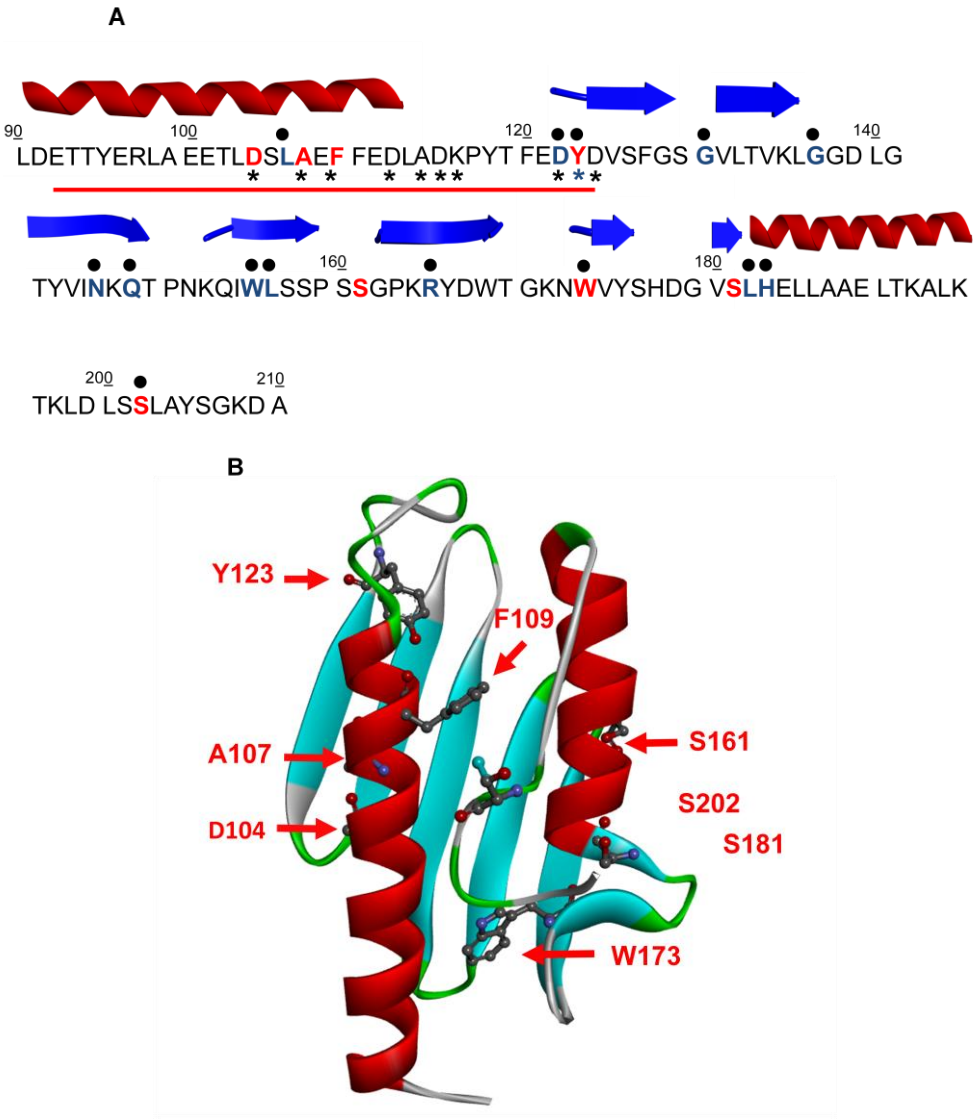


Fig. 4.9 Amino acid sequence and structure of human Frataxin. A) The mutations reported in COSMIC are depicted in bold red, the residues mutated in Friedreich's ataxia (FRDA) are in bold black and further identified by black dots. Y123, W173 and S202 are mutated in FRDA and reported in COSMIC. The black stars represent the residues involved in iron binding. The red line highlights the acidic binding region. The secondary structure

elements are as reposted in 1EKG [Dhe-Paganon et al, 2000] B) Location of the mutations on hFXN structure. Mutated residues are depicted in scaled ball and stick.

4.2.1 Spectroscopic characterization of hFXN wild-type and variants

To evaluate the effect of the hFXN variants found in cancer tissues on the protein conformation in solution, spectroscopic analysis were performed using CD and fluorescence spectroscopy. The spectral contributions of all aromatic residues present in hFXN wild-type are evidenced in the near-UV CD spectrum that is characterized by an intense negative peak between 260 and 270 nm and by two other negative contributions at around 285 and 295 nm (Fig. 4.10A). Most of the variants display near-UV CD spectra almost identical to that of the wild-type with slightly differences in the intensity of the peaks suggesting that their tertiary structure arrangements are similar to that of the corresponding wild-type protein. The most consistent alteration in the intensity of the negative contributions are observed in the near-UV CD spectrum of Y123S variant that shows a significant decrease of the dichroic activity expecially in the region between 265-285 nm, a region where the typical spectral contribution of tyrosine is expected. In line with the near-UV CD results, the fluorescence emission spectra of variants are similar to that of wild-type protein. They show the same maximum emission wavelength around 343 nm, with a decreased emission fluorescence intensity (Fig. 4.10B) with respect to that of the wild-type.

Far-UV CD spectra of hFXN wild-type and variants show local minima at around 208 and 220 nm and a zero intercept at around 200 nm indicating the

major contribution of alpha helical secondary structural elements, slightly influenced by the contribution of β sheet (Fig. 4.10C). As judged by the shape and the intensities of the signals, the secondary structure of the hFXN variants is native-like (Fig. 4.10C) suggesting that the effect of the point mutations is mainly directed and localized to the mutated residue with minor modification of tertiary and secondary arrangements. The 222/208 ellipticity ratio gives information about interhelical contacts and is generally used to distinguish between coiled coil helices and non-interacting helices (<0.9) [Choy N et al, 2003; Kiss RS et al, 2003]. The ratio of the molar ellipticity at 222 and at 208 nm ($[\Theta]_{222}/[\Theta]_{208}$) observed for the wild-type is 0.70 and is quite similar for the variants ranging from 0.69 for D104G, to 0.75 for Y123S and suggesting that the single amino acid substitutions do not alter interhelical interactions in solutions.

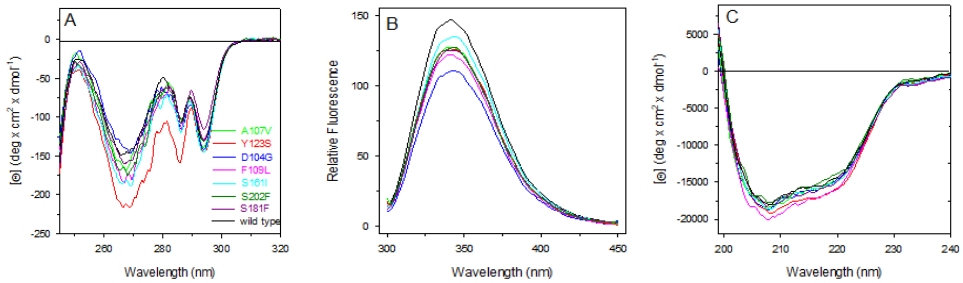


Fig. 4.10 Spectroscopic characterization of hFXN wild-type and variants. (A) Near-UV CD spectra were recorded in a 1.0-cm quartz cuvette at 1.1-1.3 mg/ml protein concentration in 20 mM Tris-HCl pH 8.0 containing 1.0 mM DTT, 1 mM EDTA and 100 mM NaCl. (B) Intrinsic fluorescence emission were recorded at 60-80 μg/ml for wild-type and the other variants spectra (0.09 AU280nm, 295 nm excitation wavelength), in 20 mM Tris-HCl, pH 8.0 containing 0.1 M NaCl, 0.1mM EDTA and 0.2 mM DTT. (C) Far-UV CD spectra were recorded in a 0.1-cm quartz cuvette at 103-130 μg/ml in 20 mM Tris-HCl, pH 8 containing 0.2 M NaCl and 0.2 mM DTT. All spectra were recorded at 20°C.

4.2.2 Thermal and thermodynamic analysis

The thermal stability of hFXN wild-type and variants was investigated by continuously monitoring the ellipticity changes at 222 nm between 20 and 95°C in comparison with that of wild-type (Fig. 4.11). By monitoring the temperature-induced ellipticity changes at 222 nm, where the main amplitude was observed, we observed an apparent cooperative transition for hFXN wild-type and all the variants, with apparent T_m values ranging from 56.0 to 73.0°C. The thermal unfolding transitions were reversible as judged by the signals upon cooling protein solutions to the starting temperature. As reported in Table 4.5, a modest increase in T_m value was observed for D104G and S202F variants, whereas all the other variants show a significant reduction of the T_m values ranging from 3 degrees below that of the wild-type, as in the case of A107V, to 14 degrees below that of the wild-type. as in the case of Y123S.

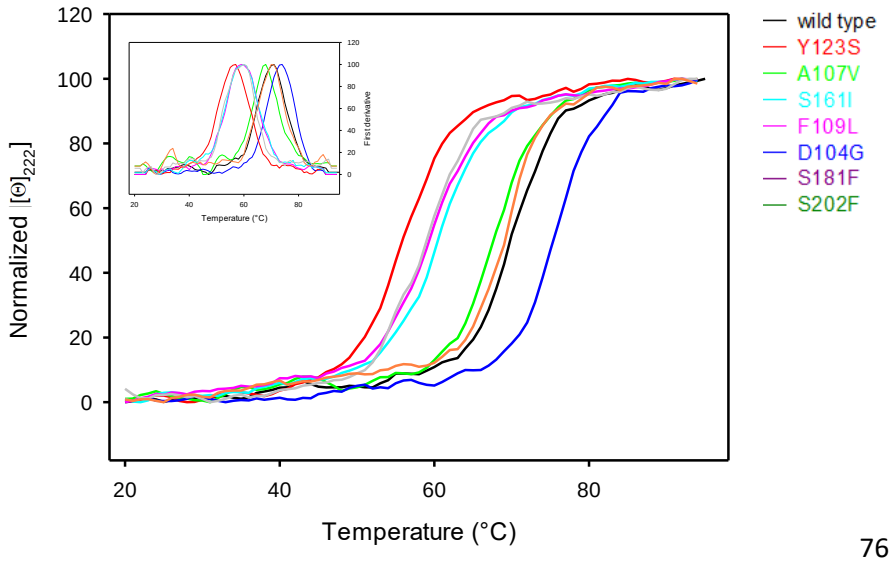


Fig. 4.11 Thermal unfolding of hFXN wild-type and variants. Normalized [Θ_{222}]; the inset shows the first derivative. hFXN wild-type and variants (103 μg/ml - 130 μg/ml) were heated from 20°C to 95°C in 20 mM Tris-HCl, pH 8.0 containing 0.2 M NaCl and 0.2 mM DTT. The molar ellipticity at 222 nm ([Θ_{222}]) was monitored continuously every 0.5 °C.

The impact of point mutations on the thermodynamic stability of hFXN was investigated at 20°C in the presence of urea as a chemical denaturant agent. hFXN wild-type and variants reversibly unfold in urea in 20 mM TrisHCl, pH 8.0, containing 0.1 M NaCl and 200 μM DTT with cooperative sigmoidal unfolding transitions that follow an apparent two-state model. The effect of increasing urea concentrations (0–9 M) on the structure of hFXN variants was analyzed by far-UV CD and intrinsic fluorescence emission spectroscopy and compared to the wild-type (Fig. 4.12).

Incubation of hFXN wild-type and variants at increasing urea concentrations at 20°C resulted in a progressive change of the intrinsic fluorescence emission intensity and a red-shift of the maximal emission wavelength. At the end of the transition, above 9 M urea, the intrinsic fluorescence emission intensity is increased about 1.5 fold and the maximal fluorescence emission wavelength shifts from around 343 nm to around 357 nm either for the wild-type and all the variants. The intrinsic fluorescence emission changes were expressed by calculating the intensity averaged emission wavelength $\bar{\lambda}$ at increasing urea concentration (Fig. 12B). This parameter is an integral measurement, negligibly influenced by the noise, and reflects changes in both the shape and the position of the emission spectrum.

The urea induced unfolding transitions monitored by far-UV CD are coincident with those obtained from $\bar{\lambda}$ change (Fig.4.12A). The

thermodynamic parameters obtained from the analysis of the far-UV CD and fluorescence changes transitions are reported in Table 4.5. D104G and S202F show conformational stability very similar to that of the wild-type, whereas the difference between the free energy of urea-induced unfolding, ΔG^{H_2O}, of F109L, Y123S, S161I and S181F variants and that of the wild-type indicates a significant decrease in thermodynamic stability of 2–4 kcal·mol^{-1} for these variants and an increase of about 1 kcal·mol^{-1} for A107V variant. The changes in *m* values may indicate differences in the solvent exposed surface area upon unfolding between the variants and the wild-type: decrease in *m* values is usually referred to a decrease in the solvent-exposed surface area upon unfolding. This is frequently ascribed to an increase in the compactness of the residual structure in the non-native state ensemble, rather than to an increase of the accessible surface area of the native state [Wrabl J et al, 1999; Pradeep L et al, 2004].

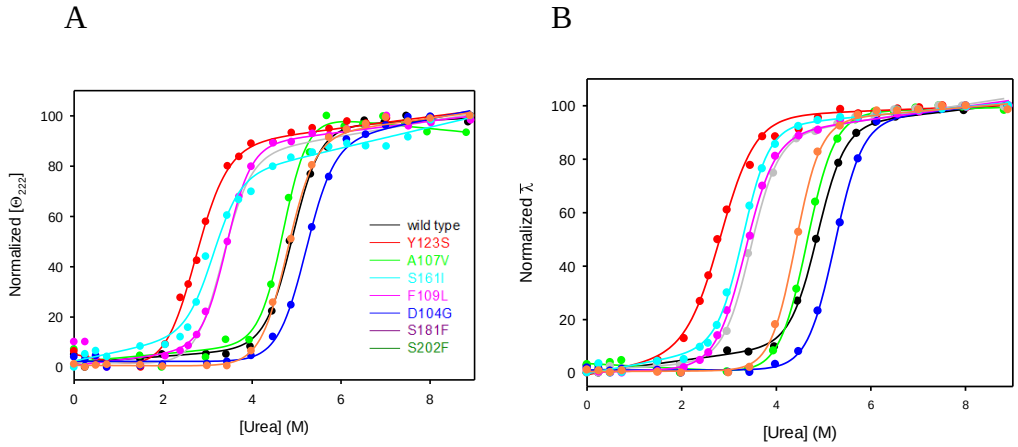


Fig. 4.12 Urea-induced equilibrium unfolding of hFXN wild-type and variants. (A) Normalized intensity-averaged emission wavelength $\bar{\lambda}$). The continuous lines represent the nonlinear fitting of the normalized $\bar{\lambda}$ and [Θ_{222}] data to Eq 4. (B) Normalized molar

ellipticity at 222 nm ([Θ_{222}]) reported after removal of the high-frequency noise and the low-frequency random error by SVD. All spectra were recorded at 20 °C.

		ΔG^{H_2O} (kcal·mol^{-1})		*m* (kcal·mol^{-1}·M^{-1})		$[Urea]_{0.5}$ (M)	
	T_m	CD ($[\Theta]_{222}$)	Fluorescence (Lambda Averaged)	CD ($[\Theta]_{222}$)	Fluorescence (Lambda Averaged)	CD ($[\Theta]_{222}$)	Fluorescence (Lambda Averaged)
WT	71.0	9.23 ± 0.47	9.77 ± 0.40	1.88 ± 0.10	2.00 ± 0.08	4.91	4.87
D104G	74.0	9.44 ± 0.46	10.39 ± 0.31	1.82 ± 0.09	1.99 ± 0.06	5.20	5.22
A107V	68.0	10.03 ± 0.91	9.06 ± 0.22	2.13 ± 0.19	1.95 ± 0.05	4.71	4.64
F109L	59.6	7.14 ± 0.34	6.18 ± 0.16	2.09 ± 0.10	1.85 ± 0.05	3.42	3.34
Y123S	56.6	4.31 ± 0.18	4.48 ± 0.18	1.60 ± 0.06	1.59 ± 0.06	2.69	2.84
S161I	60.0	5.88 ± 0.58	6.95 ± 0.24	1.88 ± 0.19	2.11 ± 0.07	3.12	3.30
S181F	59.9	6.12 ± 0.27	6.86 ± 0.33	1.84 ± 0.08	1.99 ± 0.09	3.32	3.45
S202F	70.7	9.07 ±0.32	9.60 ±0.33	1.89 ± 0.07	2.18 ± 0.07	4.78	4.40

Table 4.5 Melting temperatures and thermodynamic parameters for urea-induced unfolding equilibrium of hFXN wild-type and variants measured by far-UV CD and fluorescence spectroscopy. The temperature-induced changes were followed by monitoring the ellipticity at 222 nm. The T_m values were calculated by taking the first derivative of the ellipticity at 222 nm with respect to temperature. Urea-induced unfolding equilibrium data were measured at 20°C by monitoring ellipticity at 222 nm [Θ_{222}] and fluorescence intensity averaged emission wavelength $\bar{\lambda}$. ΔG^{H_2O} and *m* values were obtained from Eq 3; $[Urea]_{0.5}$ was calculated from Eq 5. Data are reported as the mean ± SE of the fit.

4.2.3 Analysis of the folding pathway of hFXN wild-type and its variants

In order to better understand the folding of human FXN in its wild-type form and its variants we resorted to perform kinetic experiments. In an effort to fully characterize the folding properties of hFXN and its variants we performed stopped-flow kinetic unfolding and refolding experiments, at different concentrations of urea ranging from 0M to 8M, in buffer TrisHCl 50mM, NaCl 100mM, DTT 1mM, pH 8.5, at 37°C. In every experiment conducted, the kinetics of folding and unfolding were perfectly described by a single exponential behaviour, an aspect that is usually interpreted with the absence of stable intermediates accumulating in the *ms* to *s* time range. The dependences of the logarithm of the observed rate constants versus the concentration of urea (chevron plot) for WT is reported in Fig. 4.13. In a simple two-state folding mechanism, the resulting chevron plot presents a typical V-shape with the logarithm of the observed rate constants of the unfolding and refolding arms increasing and decreasing linearly as the concentration of urea changes. However, as it might be noted from inspection of Fig. 4.13, the arms of the chevron plot of wt hFXN deviate from linearity, which suggests the presence of a more complex scenario. In a recent work [Bonetti et al., 2014] the folding kinetics of yeast frataxin was extensively characterized. The chevron plots obtained for yeast frataxin presented a curvature in both the unfolding and refolding arms and were fitted with an equation that describes a folding pathway characterized by the presence of a broad energy barrier in the transition state [Oliveberg et al., 1998]

$$k_{obs} = k_f^0 \cdot e^{\left(-m_f[UREA]+m'[UREA]^2\right)} + k_u^0 \cdot e^{\left(m_u[UREA]+m'[UREA]^2\right)}$$

Thus, we resorted to compare folding kinetics data of the WT human FXN with the ones obtained from the yeast FXN, which returned an excellent fit (Fig. 4.13). This result suggests that the folding of the human variant of hFXN can be described with a comparable mechanism.

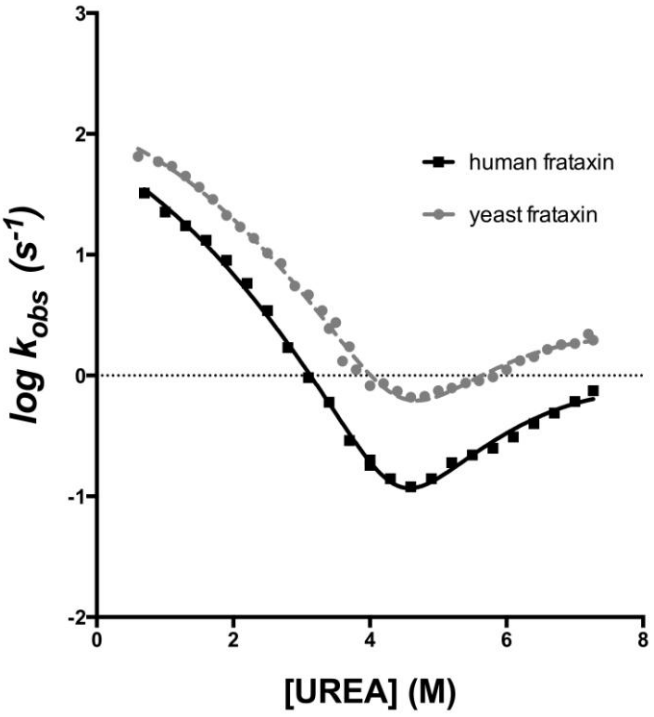


Fig. 4.13 Chevron plots of human (black squares) and yeast frataxin (data from Bonetti et al 2014) (gray circles). Data are fitted with an equation that takes into account a broad energy barrier in the transition state.

Then we performed kinetic folding experiments on the variants of hFXN. The resulting chevron plots are reported in Fig. 4.14. A global fit was performed

to analyse data, sharing the m_u and m_f values for all the data sets, which, as Fig. 4.14 reports, returned an excellent fit for every variant analysed, suggesting that mutations inferred to the protein does not have any effect on the folding mechanism of hFXN. Kinetic data obtained from the fit are shown in Table 4.6. On the other hand however, some mutations clearly affect the stability of hFXN, with the ΔG_{eq} values obtained from kinetics being in good agreement with the ones obtained from equilibrium fluorescence experiments.

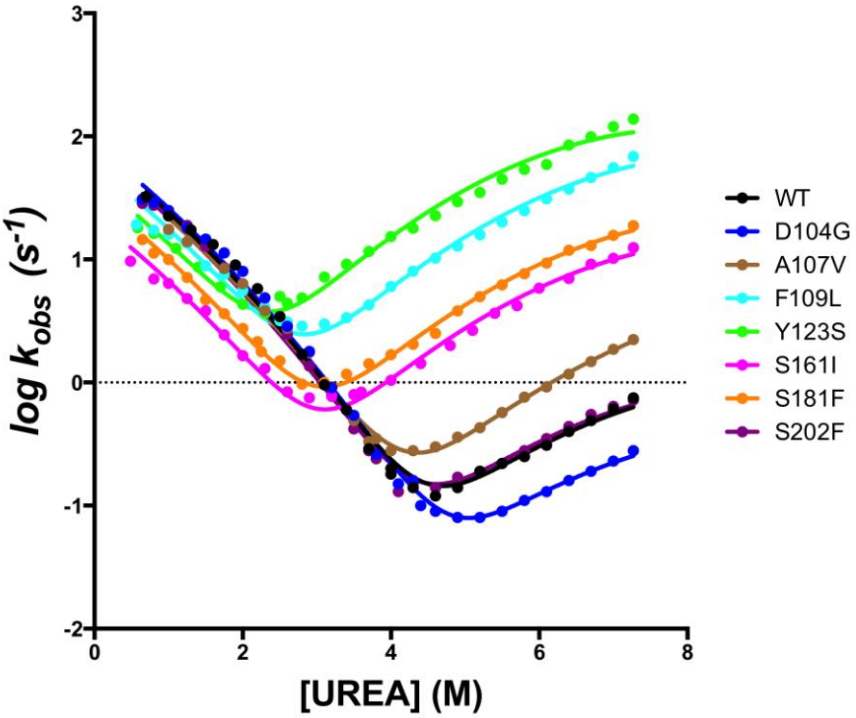


Fig. 4.14 Chevron plots of the wt hFTX and D104G, A107V, F109L, Y123S, S161I, S181F, S202F variants in buffer TrisHCl 50mM, NaCl 100mM, DTT 1mM, pH8.5, at 37°C. A

global fit analysis was performed, sharing the m_f and m_u values for all the data sets (see details in the text).

hFXN variant	Kf	Mf	Ku	Mu	m’
WT	78.5 ± 5.8	1.07 ± 0.05	$7.8 \cdot 10^{-5} \pm 0.8 \cdot 10^{-5}$	2.03 ± 0.09	0.11± 0.01
D104G	84.5 ± 5.9	1.07 ± 0.05	$3.2 \cdot 10^{-5} \pm 0.3 \cdot 10^{-5}$	2.03 ± 0.09	0.11± 0.01
A107V	70.0 ± 5.4	1.07 ± 0.05	$1.6 \cdot 10^{-5} \pm 0.2 \cdot 10^{-5}$	2.03 ± 0.09	0.11± 0.01
F109L	57.7 ± 3.6	1.07 ± 0.05	$1.1 \cdot 10^{-5} \pm 0.1 \cdot 10^{-5}$	2.03 ± 0.09	0.11± 0.01
Y123S	43.8 ± 2.9	1.07 ± 0.05	$3.4 \cdot 10^{-5} \pm 0.3 \cdot 10^{-5}$	2.03 ± 0.09	0.11± 0.01
S161F	21.7 ± 1.4	1.07 ± 0.05	$1.9 \cdot 10^{-5} \pm 0.2 \cdot 10^{-5}$	2.03 ± 0.09	0.11± 0.01
S181F	32.7 ± 2.1	1.07 ± 0.05	$3.0 \cdot 10^{-5} \pm 0.3 \cdot 10^{-5}$	2.03 ± 0.09	0.11± 0.01
S202F	71.4 ± 4.9	1.07 ± 0.05	$9.4 \cdot 10^{-5} \pm 0.9 \cdot 10^{-5}$	2.03 ± 0.09	0.11± 0.01

Table 4.6 Kinetic data for hFXN WT and variants calculated from chevron plot fitting.

4.3 Biochemical characterization of PARP3 and analysis of PARP3 nsSNVs

PARP3 plays an essential role in the maintenance of genome integrity. It is known that other proteins of the PARP family, such as PARP1 and PARP2, are involved in cancer development and tumorigenesis and, since they have been described as targets in cancer therapy, a great amount of inhibitors have been developed even if they are still not very specific. PARP1 and PARP2 have been well characterized in the last decades, while the role of

PARP3 in tumorigenesis remains largely unknown. Studies in the Françoise Dantzer's laboratory (CNRS, University of Strasbourg, Strasbourg, France) revealed an important role of PARP3 in the control of theTGFβ-induced epithelial-to-mesenchymal transition and the acquisition of stem like features in breast cancer cells lines [Karicheva O et al, 2016]. These results encouraged us to characterize PARP3 in another solid tumor, the glioblastoma, which is the most common and aggressive primary brain tumor and is resistant to current therapies. It has been described that the protein levels of PARP3 are upregulated in primary glioblastoma tissues but silencing the expression of PARP3 resulted in a synergistic radiosensitizing effect when combined with radiotherapy in glioblastoma cell lines [Quan JJ et al, 2015]. However, several questions remain unanswered. The transcriptional activity, the catalytic activity, the potential interactors of PARP3 have to be analyzed and only a more detailed characterization of the biochemical properties of PARP3 could allow us to define this protein as a therapeutic target to overcome radioresistance in glioblastoma. In this project we use the glioblastoma LN229 cell lines in order to elucidate the molecular pathways in which PARP3 could be involved. LN229 PARP3$^{-/-}$ cells have been generated using the CRISPR/Cas9 technology. To further verify the role of PARP3 activity versus the protein, the LN229 PARP3$^{-/-}$ cells were restored for the expression of either an empty vector (Flag), the PARP3 wild-type (Flag-PARP3) and a catalytically inactive mutant of PARP3 (Flag-PARP3-HE). Furthermore, the impact of single amino acid substitution on PARP3 on key cellular pathways, such as tumorigenicity and DNA repair, have been analyzed and compared with the effect of the mutations on the protein structure and stability.

4.3.1 PARP3 in the tumorigenicity of glioblastoma

In the laboratory of Francoise Dantzer, where I spent a period of my PhD, stable PARP3$^{-/-}$ LN229 cell lines were generated using the CRISPR/Cas9 method. Next these cells were restored for the expression of the PARP3 wild-type (Flag-PARP3-WT), the catalytically inactive PARP3 (Flag-PARP3-HE) and an empty vector with a Flag tag. The Flag-PARP3-HE mutant is a catalytically inactive mutant in which the histidine residue in position 384 is mutated in alanine and the glutamate residue in position 514 is mutated in alanine.

Human glioblastoma cell lines LN229 (Fig. 4.15A) were maintained in Minimum Essential Media (MEM) with 10% fetal bovine serum (FBS) and 1% gentamycin (Invitrogen) in a humidified incubator with 37°C and 5% CO_2. Western blot analysis were performed to verify the expression level of the protein after the transfection, antibiotic selection and sorting of the cells. We detected high expression of Flag-PARP3 compared to the endogenous expression level of the parental PARP3$^{+/+}$ cells (WT), while we observed more or less the same expression level of Flag-PARP3-HE respect to that observed for PARP3$^{+/+}$. The complementation with Flag showed no expression of PARP3 as expected (Fig. 4.15B).

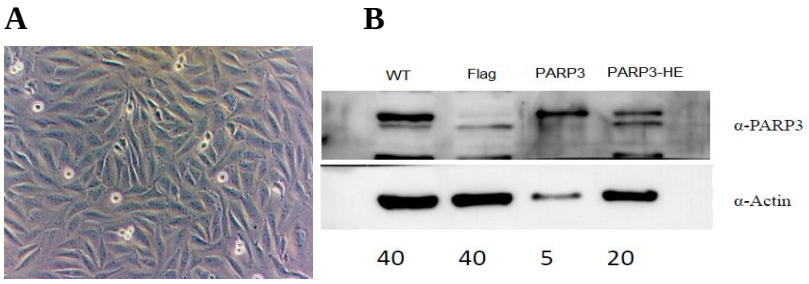


Fig. 4.15 A) Representation of adherent LN229 cell line at optical microscope. B) The protein Expression levels of PARP3 in the different complemented LN229 PARP3-/- cell lines were examined on nuclear protein extracts prepared as described in Material and method and Western blotting using the appropriate antibodies. For a better illustration different amount of protein have been loaded. 40 µg of nuclear protein extract have been loaded for the WT and the cells expressing the empty plasmid (Flag). For the Cells expressing PARP3 WT and PARP3 HE , respectively 5 µg and 20 µg have been loaded.

Using clonogenic assay we looked at the capacity of the different cells lines to proliferate and survive. After performing the experiments in triplicate, we could conclude that we didn't observe significant differences between the wild-type and the other cell lines suggesting that the absence of PARP3, the complementation with PARP3 and the complementation with the inactive PARP3 mutant do not affect the capacity of LN229 cells to form colonies or to proliferate in normal conditions. Knowing the role of PARP3 in DNA repair, we next aimed to compare the sensitivity of these different cell lines to various chemotherapeutics agents (Temozolomide (TMZ), Phleomycin (Phleo)) by clonogenic assays. For this, the four LN229 cell lines were plated following appropriate dilutions in 6-well-plates and allowed to attach overnight. The next day cells were treated for 5 h using three different concentrations of the chosen chemotherapeutic agents: 1 µM - 2 µM - 4 µM for TMZ, 1 µg/ml - 2 µg/ml -5 µg/ml for Phleo. After 5h the treated cells were plated for clonogenic assay. Our pourpose was to evaluate the capacity

of the affected cells to proliferate. After 10 days in a humidified incubator with 37°C and 5% CO2 cells were fixed with 5 ml 10% neutral buffered formalin solution for 15-30 minutes and stained with 5 ml 0.01% (w/v) crystal violet in deuterated water (dH_2O) for 30-60 minutes. The crystal violet in excess was washed out with dH_2O and dishes were allowed to dry. Finally digital images of the colonies were obtained using a camera or scanning device and colonies were counted using imaging analysis software in order to analyze the capacity of the formation of the clones (amount and size). Relative to untreated cells, LN229 wild-type treated with the chemotherapeutic agents showed a significant reduction in survival and proliferation (Fig. 4.16). Indeed, TMZ and Phleo affected the propensity of the cells to form clones and above all they affected the proliferation because of the significant reduction of the size of the formed clones that is much more evident with high concentration of drug. No obvious difference in the number or in the size of clones was detected between the different cell lines. We conclude that the absence of PARP3 has no significant impact on the survival and proliferation efficiency of LN229 cells upon exposure to either TMZ or Phleomycin.

A **B**

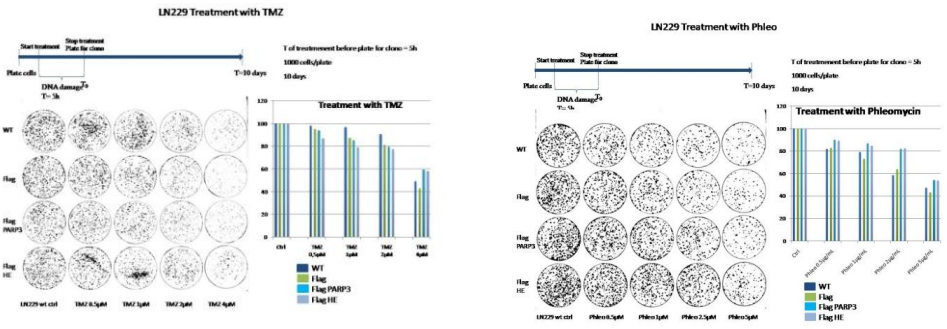


Fig. 4.16 Clonogenicity experiments of the different LN229 cells lines in presence of different concentrations of TMZ (A) and Phleo (B). The cells lines have been plated at the same dilution range for 10 days and the formed clones were stained using cristal violet. Experiments were performed three times. A representative experiment is shown.

4.3.2 PARP3 and nsSNVs

PARP3, like PARP1 and PARP2, is a DNA-dependent enzyme that is catalytically activated upon binding to DNA damage. It shows a typical fold that is characterized by a conserved C-terminal region, which contain a WGR domain and the CAT domain, and an N-terminal region not conserved between members of the PARP family. We focused our attention on the WGR and the CAT domains of the protein: WGR domain is responsible for the transmission of the information from the DNA damage recognition to the CAT site of the protein, and the CAT domain is responsible for the catalytic activity of the protein. In the COSMIC database a huge amount of PARP3 nsSNVs have been reported. We performed a screening of these variants and, according to the frequency and to the position of the variants in the 3D structure of the protein, we selected F125L, P147T and G148D, located respectively in the core and in the flexible loop of the WGR domain (Fig. 4.17A) and the variants L233F and F440L, located in the CAT domain of PARP3 (Fig. 4.17B).

A **B**

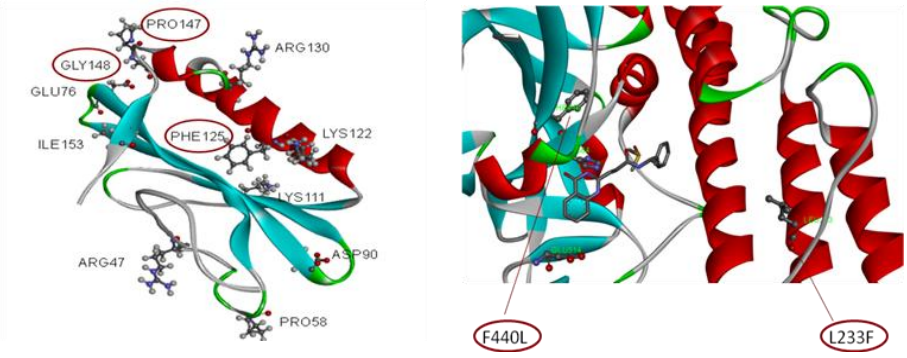


Fig. 4.17 A) Location of the selected variants on PARP3 WGR domain (Pdb code: 2EOC) B) Location of the selected variants on PARP3 CAT domain (Pdb code: 4L6Z).

We used PARP3 wild-type pET-HPARP-3-Met2 plasmid. Quick Change Site-Directed Mutagenesis Kit (Stratagene) was used to introduce the single mutations on wild-type hPARP3 plasmid. The presence of the desired mutations and the absence of unwanted additional mutations were confirmed by plasmid sequencing.

PARP3 wild-type and the variants were expressed in *E. coli* strain BL21(DE3). 15 ml of overnight culture containing 100 µg/ml ampicillin was used to inoculate 500ml LB cultures containing ampicillin as antibiotic at a final concentration of 50 µg/ml. Cultures were grown at 37 °C until optical density OD_{600} reached 0.8, the protein expression was induced using 0.5 mM isopropyl-β-D-thiogalactoside (Sigma-Aldrich) and cultures were grown at 18°C overnight with vigorous shaking. The cells were collected by centrifugation and the pellet frozen. The first attempts to purify the protein were not successful due to technical troubles.

Several attempts to purify the protein were not successful due to the low expression of the protein in soluble form. Cells were resuspended in 35 ml of

Binding buffer (20 mM Tris-HCl pH 7.5, 0.5 mM EDTA, 14 mM β-mercaptoethanol, 0.5 mM PMSF) containing a cocktail of EDTA-free protease inhibitors (Sigma-Aldrich) and disrupted by sonication in a Vibracell sonicator on ice. The lysate was cleared by centrifugation at 12000 rpm and the supernatant was applied on a DE52 column (GE Healthcare), previously equilibrated with Binding buffer and 0.5 mM tris(2-carboxyethyl)-phosphine, to remove nucleic acids The flow-through was loaded on a Ni-NTA (Ni^{2+} - nitriltriacetate) affinity column (GE Healthcare) pre-equilibrated with Binding buffer and the protein was eluted using Buffer with 250mM Imidazole. Purity and size of the protein was checked by SDS–PAGE on a 4–12% bis-Tris polyacrylamide gel and by Western blot analysis performed using the home-made SE4698 anti-hPARP3 antibody. The signal was visualized using ECL-PLUSdetection system (Amersham Biosciences) and the images were captured using the Image Quant LAS 4000 imaging system (GE Healthcare Life Science). Gels were stained with Coomassie blue R-250.

Using CD and fluorescence spectroscopy we will perform a detailed analysis of the conformation in solution of all PARP3 variants.

We also decided to analyze the impact of the PARP3 nsSNVs on the viability and proliferative capacity of human cancer cells. For this reason we used Quick Change Site-Directed Mutagenesis Kit (Stratagene) to obtain the PARP3 nsSNVs in the pEGFP vector. MDA MB231 $PARP3^{-/-}$ cells and MDA MB436 $PARP3^{-/-}$ breast cancer cells were transfected with an empty vector (pEGFP), with the PARP3 wild-type (pEGFPhP3) and with two variants of the PARP3 WGR domain (pEGFPhP3-P147T, pEGFPhP3-G148D) and with one variant of the PARP3 CAT domain (pEGFPhP3-F440L).

The cell culture medium was changed 48 hours after transfection and the stable transfectants were selected with neomycin at 500 μg/ml and 250 μg/ml in MDA MB231 cells and in MDA MB436 cells respectively. Breast cancer cell lines were maintained in RPMI 1640 medium with 10% fetal bovine serum (FBS) and 1% gentamycin (Invitrogen) in a humidified incubator with 37°C and 5% CO_2.

The expression of the variants seems to be toxic for the cells, expecially in the case of the variants G148D and F440L in both cell lines (Fig. 4.18).

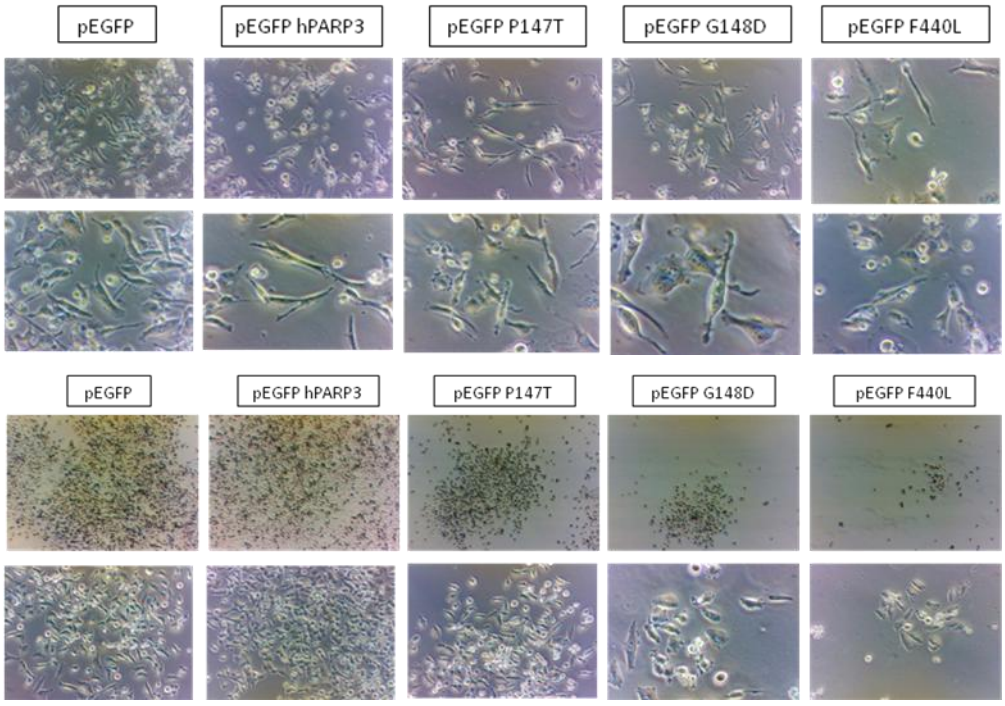


Fig. 4.18 Representation of the toxicity of the overexpression of PARP3 wild-type and variants in cancer cells A) Control (pEGFP), PARP3 WT (pEGFP hPARP3) and variants (pEGFP P147T, pEGFP G148D, pEGFP F440L) overexpressing MDA MB231 cells. B) Control (pEGFP), PARP3 WT (pEGFP hPARP3) and variants (pEGFP P147T, pEGFP G148D, pEGFP F440L) overexpressing MDA MB436 cells. In both cell lines, the overexpression of pEGFP F440L, a variant of the catalytic region of PARP3, seems to be

really toxic for the cells as evidenced by the high mortality and by the changes in the morphology of the cells.

Protein extraction and western blot analysis will be performed in order to analyze the expression levels of PARP3 wild-type and PARP3 variants. To understand the effects of the selected PARP3 variants on cell survival, we will evaluate the ability of the transfected cell lines to form colonies in clonogenic assays.

4.4 Analysis of the effects induced by single amino acid substitution

The results reported above are framed in a larger project that results from their combination with what has been obtained for variants of other proteins studied by our group in the last years: Bromodomains (BRDs), Peroxisome Proliferator-Activated Receptor γ (PPARγ), Serine/threonine-protein kinase pim-1 (Pim-1) and Protein tyrosine phosphatase ρ (PTPρ) in order to present a detailed evaluation of the possible effects that single amino acid substitutions can have on protein structure, function or stability. This will be useful to understand how missense variants can influence protein properties. We have been interested in nsSNVs that have been found in cancer tissues, as reported in COSMIC database [Forbes S.A. et al, 2011], OMIM [Hamosh a et al, 2005], Pan-Cancer Atlas (https://www.cell.com/pb-assets/consortium/pancanceratlas/pancani3/index.html) and SNP database. For each protein we chose variants with an high pathogenic score (Table 4.1)

and, after expression and characterization, we classified the selected variants according to the observed alterations with respect to the wild-type protein.

4.4.1 Effect on protein conformation in solution

39 of 45 nsSNVs analyzed caused changes in the protein conformation in solution. In most of the cases we observed only small conformational changes, as evidenced in the near-UV CD spectra of the hFXN variants (Fig. 4.10A), for example, suggesting that the tertiary structure arrangements of the nsSNVs are really similar to that of the corresponding wild-type protein. However, significant differences can also be detected, as evidenced in the near-UV CD spectrum of the PGK1 variant R65W (Fig. 4.2A) where the substitution of an arginine with a tryptophan residue caused a significant alteration in the intensity of the negative peaks in the region between 250 and 270nm.

The most significant example regarding the effect of single amino acid substitution on protein conformation in solution in our analyzed dataset of 45 nsSNVs is represented by the case of E135K variant of Pim-1 considering that this missense mutation displayed a near-UV CD spectrum which dramatically differed from the wild-type spectrum: the dichroic activity in the 290–275 nm region is markedly decreased with a positive contribution in the region below 275 nm (Fig. 4.19A). These alterations are confirmed by fluorescence emission spectroscopy, in particular the fluorescence emission spectrum of the variant is centered at the same maximum emission wavelength of wild-type at around 345 nm and shows differences in emission fluorescence intensity (Fig. 4.19B). Pim-1 is a serine/threonine protein kinase [Saris CJ et al, 1991] involved in several signalling pathways such as the

regulation of cell cycle progression and apoptosis [Anizon F et al, 2010]. It has been identified as a key cofactor regulating the expression of the oncogenic transcription factor c-Myc [Zippo A. et al, 2007] and several nsSNVs have been reported in prostate cancer and multiple myeloma [Akue´-Ge´du R. et al, 2009]. Of these variants, E135 is located in helix αD and forms a hydrogen bond with Q127 which is likely to be important in stabilizing this helix (Fig. 4.19C). Furthermore, a significant reduction in the protein activity was also observed for this variant: the mutated protein showed only about 3% of the activity of the corresponding wild-type. The decrease of the activation energy, E_a, detected suggested an increase of the flexibility respect to the wild-type counterpart [Lori L et al, PlosOne, 2013].

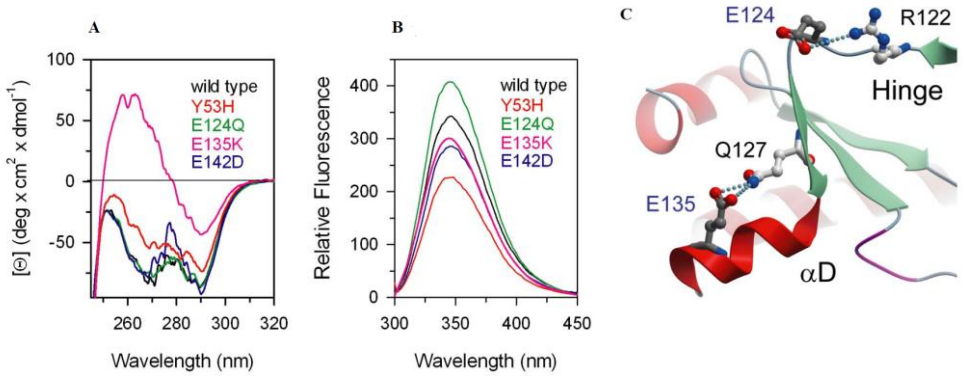


Fig. 4.19 Location and spectral properties of Pim-1 wild-type and E135K variants (A) Near-UV CD spectra The near-UV CD spectrum of wild-type Pim-1 shows the spectral contribution of all aromatic residues and is characterized by two strong negative peaks centred at 290 nm and at 269 nm accompanied by fine structure features at 275–280 nm (B) Intrinsic fluorescence emission spectra. (C) Detailed view of the local structural environment around the mutated residue E135. doi:10.1371/journal.pone.0064824.g001

4.4.2 Effect on protein structure

7 of 45 nsSNVs analyzed give rise of minimal changes in the protein structure as already put in evidence in all the crystal structures of the PGK1 variants analyzed in this thesis (Fig. 4.5-4.8).

The mutation of Arg → Leu (R100L) in the ZA loop of the Bromodomain 2 domain 1 (BRD2(1)), in close proximity of the acetyl-lysine binding site, induced loss of the hydrogen bonds with backbone residues in helix αA (Fig. 4.20B) but this did not result in significant structural changes [Lori L et al, 2016] (Fig. 4.20A). The Bromodomains (BRDs) are the only known protein recognition module that selectively targets ε-N-acetylation of lysines. Dysregulation of acetylation levels has been associated with several diseases and cancer and inhibitors that specifically target the BRDs can interfere with gene expression that mediates cellular growth and evasion of apoptosis in cancer.

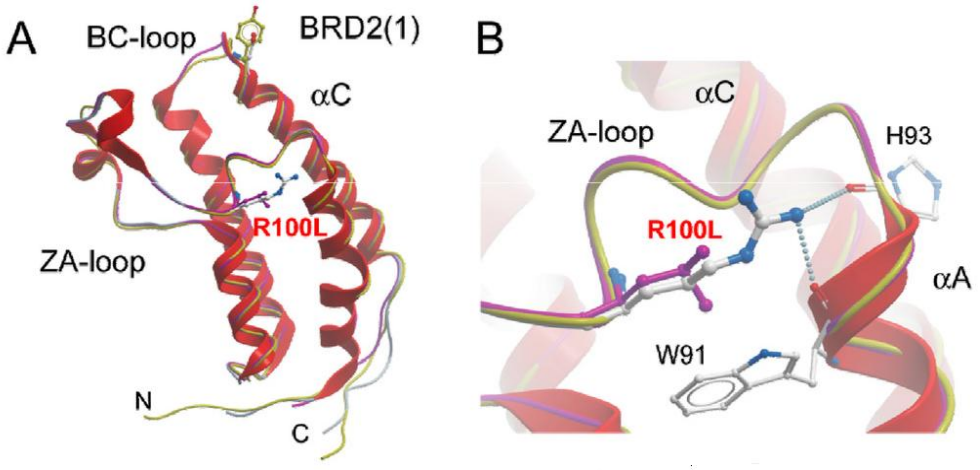


Fig. 4.20 Comparison of the crystal structures of BRD2(1) wild-type and R100L variant (A) Superimposition of wild-type BRD2(1) shown as a ribbon diagram with the mutant BRD2(1) R100L shown as protein worm in magenta. The mutated residue is shown in ball and stick representation and main structural elements are labelled. (B) Details of interactions formed by R100 in the wild-type compared to the mutated residue. doi:10.1371/journal.pone.0159180.g002

4.4.3 Effect on protein function

27 of 45 nsSNVs analyzed had effect on the protein activity, either increasing or decreasing it.

If a mutation occurs in an active site, it can affect critical components of the biological reaction, which, in turn, will alter the normal protein function [Stevanin G et al, 2004; Yamada Y et al, 2006]. However, since the biochemical reaction is very sensitive to the precise geometry of the active sites for both of the reactants and products, not only any conformational change altering the active sites can affect the biochemical reaction but also mutations that are not expected to perturb protein function by much can do it [Takamiya O et al, 2002; Zhang Z et al, 2010; Zhang Z et al, 2011].

The effect of non synonymous single nucleotide substitutions on the kinetic properties of PGK1 enzyme have already been presented in this thesis (Fig. 4.3 and Table 4.2) by putting in evidence a reduced catalytic efficiency for most of the variants with the exception of F241S that displays a two fold increase of activity.

An interesting example about changes on the protein activity due to the presence of missense mutations is represented by the case of Protein tyrosine phosphatase ρ (PTPρ) (PDB code 2OOQ). In particular, PTPρ loses

detectable enzymatic activity due to a D927G mutation that has been found in cancer tissue (Fig. 4.21A,B). PTPρ belongs to the classical receptor type IIB family of protein tyrosine phosphatase and several evidences suggest that it acts as a tumour suppressor gene [Wang Z et al, 2004]. PTPρ missense variants have been identified in human cancer tissues. The mutation D927G involves a solvent exposed residue (Asp-927) placed in a 4 residues turn between two coils and distant from the catalytic site of the enzyme. It connects different protein secondary structure regions through hydrogen bonds with three residues (Asp-947, Lys-930 and Glu-931) (Fig. 4.22C). Presumably, the change in main chain flexibility caused by the high destabilizing D927G mutation may lead to local disorder and may affect the stabilizing hydrogen bonds of residues in close proximity [Pasquo A et al, 2012].

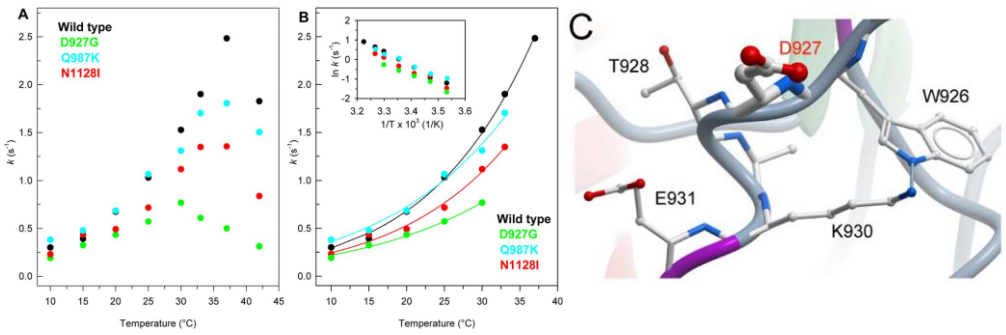


Fig. 4.21 Effect of temperature on phosphatase activity of PTPρ wild-type and mutants and location of PTPρ mutations. (A) Temperature dependence of phosphatase activity of PTPρ wild-type and mutants. (B) Non-linear fit of the temperature dependence of phosphatase activity to the Arrhenius equation; the inset shows the linear Arrhenius plot for the same data. doi:10.1371/journal.pone.0032555.g004 (C) Local environment of D927.

4.4.4 Effect on protein binding

A nsSNV can alter the binding site of the protein, can change the interactions with partners (activators, repressors or substrates) and can also affect the kinetics of interactions with partner or the binding specificity [Kohler et al, 2008; Wu et al, 2008; Bauer-Mehren et al, 2009; Vanunu et al, 2010; Barabàsi et al, 2011].

If a missense mutation occurs on a residue located in the binding interface, crucial in contributing to the interaction [Dixit A et al, 2009; Ozbabacan SEA et al, 2010], the binding affinity would be dramatically affected due to geometrical constrains and/or energetic effects [Akhavan S et al, 2005; Jones R et al, 2007]. For instance, when substituting a small side chain for a bulky side chain in a narrow binding pocket, the entrance of the partner group will be blocked and the binding process will be completely or partially prevented [Rignall TR et al, 2002; Zhang Z et al, 2011].

26 of 45 nsSNVs analyzed induced changes in the binding properties of the protein. One interesting example is represented by the case of the H395R variant of the Bromodomain 3 domain 2 (BRD3(2)). The substitution of the histidine residue in position 395 of the (BRD3(2)) with an arginine residue decreased of more than two-fold the K_d for the thienodiazepine JQ1 (K_d wild-type 23.4 ± 1.3 nM, K_d H395R 7.0 ± 0.4 nM) and increased of about nine-fold the K_d for the quinazolinone PFI-1 (K_d wild-type 161.0 ± 5.9, K_d H395R 1386.0 ± 65.0 nM), two inhibitors of pharmacological interest, as demonstrated by ITC experiments that put in evidence a significant alteration in binding enthalpy and binding affinity [Lori L et al, 2016] (Fig. 4.22C). By comparing the crystal structure of BRD3(2) wild-type with JQ1 and the

structure of BRD3(2) H395R we observed that the arginine residue introduced by this point mutation may form a more favourable hydrogen bond with the carbonyl oxygen of the ester link [Lori L et al, 2016] (Fig. 4.22A,B)

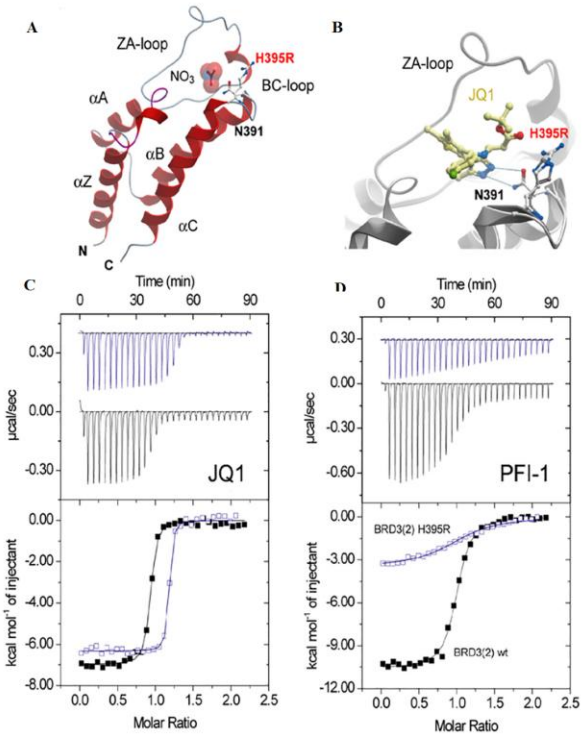


Fig. 4.22 Binding of BRD3(2) H395R variant to inhibitors (A) Structure of BRD3(2) H395R variant. The mutated residues are highlighted. A nitrate molecule, present in the crystallization solution occupied the acetyl-lysine binding site in BRD3(2) H395R. The conserved asparagine (N391) formed canonical hydrogen bonds with the nitrated ion. (B) Superimposition of the wild-type JQ1 complex with BRD3(2) H395R (C) ITC experiments measuring the interaction of the panBET inhibitor JQ1 with wild-type BRD3(2) (black curve) and BRD3(2) H395R (blue curve). Shown are raw titration heats (top panel) as well as normalized binding enthalpies as a function of ligand/protein ratio (lower panel). The best fit to a single binding site model is shown as solid lines. (D) ITC experiments showing the

binding of the pan-BET inhibitor PFI-1 with BRD3(2) and the H395R variant. doi:10.1371/journal.pone.0159180.g004.

4.4.5 Effect on protein stability

33 of 45 nsSNVs analyzed showed a reduction of the protein stability, as evident from the analysis of the thermodynamic parameters of the variants of PGK1 and hFXN reported in this thesis (Table 4.4 and 4.5) that point to a destabilization of the native state indicated by the decrease of ΔG^{H_2O} values for most of the variants.

Protein stability is also a key characteristic of a functional protein [Capriotti E et al, 2005; Ye Y et al, 2006; Zhan Z et al, 2010; Zhang Z et al, 2011].

It is known that a single base DNA variant results in alteration of stability of the corresponding native protein [Wang and Moult, 2001; Yue et al, 2005] and that about 80% of missense mutations associated with disease affects the stability of proteins by several kcal·mol^{-1} [Wang and Moult, 2001]. A nsSNV can have destabilizing or stabilizing effects: most frequently, missense mutations lead to a destabilization of the protein stability by making it susceptible to proteolysis or by changing the thermal inactivation temperature [Pasquo A et al, 2012; Lori L et al, 2013] or by affecting protein folding and producing changes in thermodynamic stability [Adhikari et al, 2015; Kroncke et al, 2015; Petrosino M et al, 2017].

The variant E140K of BRD2(1) decreases drastically the thermal and the thermodynamic stability of the protein. The thermal stability of the wild-type and of the variant was investigated by continuously monitoring the molar ellipticity changes at 222nm between 20 and 80°C and the substitution of

Glu140 with Lys induced a significant decrease in the melting temperature (T_m): the T_m determined for E140K (44°C), calculated by plotting the first derivative of the molar ellipticity at 222nm as a function of temperature, is about 1 **A** egree below the T_m of the wild- **B** e protein (54.8°C) (Fig. 4.23A) [Lori L et al, 2016].

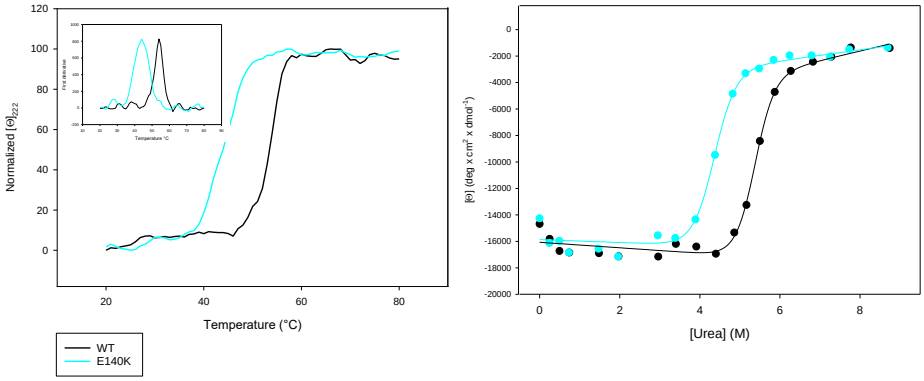


Fig. 4.23 Thermal unfolding transition and urea-induced equilibrium unfolding of BRD2(1) wild-type and E140K variant (A) Wild-type (black) and E140K variant (cyan) were heated between 20 and 80°C in a 0.1-cm quartz cuvette at 0.2mg/ml. The insert shows the first derivative of the same data. (B) Normalized molar ellipticity at 222nm reported after removal of the high-frequency noise and the low-frequency random error.

The lower T_m value and the high loss of secondary structure elements upon unfolding suggest that the point mutation induces a remarkable destabilization of the native state of BRD2(1). Studying the thermodynamic stability by urea-induced unfolding experiments, a decrease in the free-energy ΔG^{H_2O} was also detected (Fig. 4.23B). E140 is a solvent exposed residue located in the αB helix and does not participate in any other interactions with another helix but is involved in stabilizing contacts with the amino acid residues in the close coil.

4.4.6 Effect of single amino acid substitution on protein production

Several mutations can affect the production of protein, most probably reflecting their influence on folding efficiency, generally decreasing the probability of the protein reaching its active conformation.

2 of 45 nsSNVs analyzed could not be expressed in soluble fraction even when different growth and induction conditions were used. This is the case of W173C variant of hFXN (Fig. 4.9) and E324K variant of the Peroxisome Proliferator-Activated Receptor γ (PPARγ). PPARγ, a nuclear receptor that belongs to the superfamily of ligand inducible transcription factors, is involved in several biological processes and in the maintenance of cellular homeostasis [Sauer S, 2015] and several nsSNVs have been found in cancer tissues [Kersten S et al, 2000; Wang T et al, 2006; Lehrke M et al, 2005; Spiegelman BM et al, 2008]. Mapping of this mutation onto the structure of PPARγ revealed that E324 is involved (with R397) in one of the two salt bridges that play a pivotal role in the domain stabilization (Fig. 4.24) [Petrosino M et al, 2017].

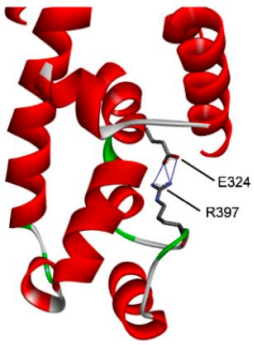


Fig. 4.24 Local environment of residue E324 involved in one salt bridge.

Chapter 5

Discussion

Single nucleotide substitution in DNA that alters an amino acid in the corresponding codon (non synonymous change) generates a mutant protein (nsSNV) with possible structural or functional alterations. A high percentage of nsSNVs have been found in different cancer types and catalogued in several databases, such as COSMIC [Forbes et al, 2010] or Pan-Cancer Atlas (https://www.cell.com/pb-assets/consortium/pancanceratlas/pancani3/index.html), suggesting that the change of an amino acid is an event that could be crucial for cellular transformation. The resulting variant carrying the missense mutation may initiate the cascade of downstream signalling and eventually transforms the cellular phenotype from normal to malignant. However, a point mutation may still affect the structure and function of the protein not causing a disease.

In this thesis a biochemical characterization of nsSNVs that have been detected in different types of cancer has been performed in order to investigate the effect of missense substitutions on the nsSNV structure, function or stability. In particular, we studied variants of three proteins involved in the regulation of cell metabolism (PGK1, hFXN) and in the response of DNA damages (PARP3). The results obtained for these variants have been combined with those obtained for variants found in cancer tissues of other proteins studied by our group in the last years (BRDs [Lori L et al, 2016], PPARγ [Petrosino M et al, 2017], Pim-1 [Lori C et al, 2013] and PTPρ

[Pasquo A et al, 2012]) in order to present a detailed evaluation of properties of protein variants caused by single amino acid substitutions. The data reported here provide an important contribution to fill the gap between the collection of thermodynamic data and disease-related information on protein variants.

Our analysis of a dataset of 45 nsSNVs (Table 4.1) underscores that the single amino acid substitutions can affect the properties of the native protein in different ways. They usually influence the protein conformation and the protein stability and in several cases it was possible to detect alteration of the protein function and of the binding properties.

Most of the variants studied show significantly alterations of their tertiary interactions, as indicated by the differences in their near-UV CD fine structure and/or in the ellipticity amplitude, when compared to the corresponding wild-type proteins (Fig. 4.2A, 4.10A, 4.19A). The changes in tertiary arrangements of most of the variants were also evident from their intrinsic fluorescence spectra that showed differences in relative intensity and/or in the maximum emission wavelength (Fig. 4.2B, 4.10B, 4.19B). Due to the usually surface exposed location of the mutated residues (69% solvent exposed, 31% buried) (Fig. 3.1), the consequences of the mutations were not obvious.

The structures of some of the variants that showed the most significant changes in their thermodynamics or functional properties were determined by X-ray crystallography. Notably, the structural comparative analysis shows that the introduced mutations do not cause large conformational changes in the protein tridimensional structure but only local changes involving the mutated residues and their chemical surrounding (Fig.

4.5-4.8, 4.20, 4.22). These local changes may be responsible for local differences in protein flexibility.

Amino acid substitutions altering the structure or the conformational rearrangement in solution of the native protein could remarkably affect the network of protein interactions, above all when the substitution involves a solvent exposed residue. These alterations may lead to change in the interactions with ligands or inhibitors as reported for the BRD3(2) H395R variant in which the significant differences observed in the near-UV CD spectrum, that indicate that its tertiary structure arrangements is different from that of its wild-type counterpart, have been associated with a significant reduction in the binding enthalpy and in the binding affinity for JQ1 and PFI-1 (Fig. 4.22) [Lori L et al, 2016].

The global effects of local alterations are evident from the analysis of the thermodynamic parameters that point to a remarkable destabilization of the native state, as suggested by the significant changes in both thermal and thermodynamic stability resulting in changes in T_m and/or in ΔG^{H_2O} values in most of the 45 analyzed variants (Fig. 4.23 and Table 4.4, 4.5). Protein stability is a key characteristic of a functional protein [Wang and Moult 2001; Zhang et al, 2001; Capriotti et al, 2005] and its evaluation could help to predict the half-life of the mutant proteins.

Independently from the position of the mutation on the protein, one of the result of the amino acid substitutions, object of this report, is a significant change in the variants catalytic properties, in fact most of the variants display an altered catalytic efficiency. This is particularly evident for the PGK1 variant R38M, that shows a decrease of eight order of magnitude of the k_{cat}/K_M value, mainly due to a dramatic decrease of the turnover number (Table 4.2) [Fiorillo A et al, 2018]. As shown in the structure (Fig. 4.3), R38

is located in the 3-PG binding site and interacts electrostatically with the carboxyl group of 3-PG, hence it is critical for the positioning of the substrate in the catalytic site and for the stabilization of the transition state. Accordingly, the substitution of the R38 implies a dramatic decrease of the k_{cat} and an increase of K_M. Given that PGK1 is a glycolytic enzyme, the reduced catalytic properties of PGK1 variants may suggest that their expression in the neoplastic cells would significantly burden the metabolic flux through the glycolytic pathway. Thus, metabolic intermediates may accumulate and be converted in other products through alternative pathway, such as the pentose shunt, that could supply for the biosynthetic need of the growing and proliferating tumour cell. On the other hand, the occurrence of a "glycolytic jam" may induce the cell to increase the uptake of glucose.

In conclusion, our data give a comprehensive understanding of three-dimensional structure, dynamics, and biophysics of wild-type and nsSNVs that can be used to develop better tools for accurate predictions regarding the consequences of single amino acid changes. Since the integration of information from multiple heterogeneous sources including sequence, structural data and cellular pathway analysis can help to understand the molecular basis of cancer, our future perspective is represented by the study of the impact of the selected nsSNVs on important pathways for the cell, such as cellular differentiation, apoptosis and propensity to tumorigenicity. This purpose is at the beginning for the PARP3 variants and we aim to extend this kind of analysis to all the 45 nsSNVs analyzed in this study from a biochemical point of view in order to test how these variants can modulate the efficiency of the cellular pathways of the protein.

Thanks to the useful information that can be provided from a detailed investigation at the molecular and cellular level of the protein variants

carrying missense mutations, the development of new therapeutic strategies could be possible, expecially in the search of small molecules able to selectively interact with the variants, which is an essential preliminary step to personalized medicine. Furthermore, since we might expect that some nsSNVs will modulate the output of the cellular pathways of the protein by making them more or less efficient, research over the next few years will likely uncover this gap in the study of nsSNVs and it will make treatment of cancer more successful.

Chapter 6

Attachments

PLOS ONE

RESEARCH ARTICLE

The phosphoglycerate kinase 1 variants found in carcinoma cells display different catalytic activity and conformational stability compared to the native enzyme

Annarita Fiorillo[1☯]*, Maria Petrosino[1☯], Andrea Ilari[2], Alessandra Pasquo[3], Alessandra Cipollone[1], Maristella Maggi[4], Roberta Chiaraluce[1]*, Valerio Consalvi[1]

1 Department of Biochemical Sciences "A. Rossi Fanelli", Sapienza University of Rome, Rome, Italy, **2** CNR-Institute of Molecular Biology and Pathology, Rome, Italy, **3** ENEA CR Frascati, Diagnostics and Metrology Laboratory, FSN-TECFIS-DIM, Frascati, Italy, **4** Department of Molecular Medicine, Unit of Immunology and General Pathology, University of Pavia, Pavia, Italy

☯ These authors contributed equally to this work.
* roberta.chiaraluce@uniroma1.it (RC); annarita.fiorillo@uniroma1.it (AF)

OPEN ACCESS

Citation: Fiorillo A, Petrosino M, Ilari A, Pasquo A, Cipollone A, Maggi M, et al. (2018) The phosphoglycerate kinase 1 variants found in carcinoma cells display different catalytic activity and conformational stability compared to the native enzyme. PLoS ONE 13(7): e0199191. https://doi.org/10.1371/journal.pone.0199191

Editor: Jose M. Sanchez-Ruiz, Universidad de Granada, SPAIN

Received: March 28, 2018

Accepted: June 2, 2018

Published: July 11, 2018

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This study was supported by Prot. FILAS-RU-2014-1020 Regione Lazio, Grant recipient: Valerio Consalvi. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Abstract

Cancer cells are able to survive in difficult conditions, reprogramming their metabolism according to their requirements. Under hypoxic conditions they shift from oxidative phosphorylation to aerobic glycolysis, a behavior known as Warburg effect. In the last years, glycolytic enzymes have been identified as potential targets for alternative anticancer therapies. Recently, phosphoglycerate kinase 1 (PGK1), an ubiquitous enzyme expressed in all somatic cells that catalyzes the seventh step of glycolysis which consists of the reversible phosphotransfer reaction from 1,3-bisphosphoglycerate to ADP, has been discovered to be overexpressed in many cancer types. Moreover, several somatic variants of PGK1 have been identified in tumors. In this study we analyzed the effect of the single nucleotide variants found in cancer tissues on the PGK1 structure and function. Our results clearly show that the variants display a decreased catalytic efficiency and/or thermodynamic stability and an altered local tertiary structure, as shown by the solved X-ray structures. The changes in the catalytic properties and in the stability of the PGK1 variants, mainly due to the local changes evidenced by the X-ray structures, suggest also changes in the functional role of PGK to support the biosynthetic need of the growing and proliferating tumour cells.

Introduction

Carcinomas are the most common type of cancer. According to the American Cancer Society, in 2016 the overall estimate is of 1685210 new cases of cancer worldwide and among them 61000 cases of female breast carcinoma in situ are expected [1]. Cancer cells are able to adapt to survive in difficult conditions, like for example in O_2 deficiency, through a reprogramming of their metabolic machinery according to their requirements [2]. Among these, the hallmark

of the metabolic reprogramming is the shift from oxidative phosphorylation to aerobic glycolysis that allows tumor cells to survive under hypoxic conditions [3], an adaptative behaviour described nearly 100 years ago and known as the Warburg effect. In the last years, glycolytic enzymes, and their role in cancer metabolism, have been the object of several studies and they have been identified as potential targets for alternative anticancer therapies [4, 5].

A central enzyme in glycolysis is the phosphoglycerate kinase 1 (PGK1). This ubiquitous glycolytic enzyme is expressed in all somatic cells, where it provides energy in form of ATP through the reversible phosphotransfer reaction from 1,3-bisphosphoglycerate (1,3-BPG) to MgADP in order to produce 3-phosphoglycerate (3-PG) and MgATP in the presence of free magnesium [6,7]. Deficiency of PGK1 has been associated with hereditary non-spherocytic haemolytic anaemia (HNSHA), a rare disorder that is caused by a variety of inherited defects in glycolysis [8].

The structure of human PGK1 was solved in the open conformation (PDB code: 2XE7) [7] in the partially closed conformation (PDB code: 2ZGV) [9] and in closed conformation bound to a transition state analogue (PDB code: 2WZB) [10] as well as in complex with inhibitors (PDB codes: 4O33, 4O3F) [11]. PGK1 is a monomeric enzyme of 417 amino acids formed by two distinct α-helical domains of equal size, the N- and C-terminal domains connected by a hinge region (Fig 1). The N-terminal domain binds 3-PG or 1,3-BPG while the C-terminal domain binds MgADP or MgATP. During the catalytic cycle four hinge points mediate inter-domain motions that result in the flexible region bending, allowing the two domains to approach each other bringing the catalytic residues in the right position. During this transition phase, a large number of conformational rearrangements, triggered by the simultaneous binding of the two substrates, lead the enzyme from its open form, which has the highest affinity for the substrates, to its closed form in which the enzyme performs the transfer of the phosphoryl group fulfilling its catalytic activity [6].

Interestingly, similarly to many other glycolytic enzymes, PGK1, in addition to its metabolic function, may acquire different functions. This enzyme may be secreted in the extracellular environment by tumour cells and act as a thiol reductase regulating angiogenesis [12]. In addition, translocation of PGK1 to the nucleus is related to binding to alpha DNA polymerase [13]. Notably, under hypoxic conditions, PGK1 may translocate from cytoplasm to mitochondrion where it may act as a protein kinase and phosphorylate different protein substrates [14]. Interestingly, the protein kinase activity of PGK1 has been related to initiation of autophagy [15, 16].

PGK1 is regulated by hypoxia-induced factor-1α (HIF-1α), the most important factor involved in the cellular response to hypoxia [12]. Several solid tumors like prostate cancer, breast cancer, pancreatic ductal adenocarcinoma, multidrug-resistant ovarian cancer and metastatic gastric cancer exhibit an increased expression of glycolytic enzymes such as PGK1 to generate ATP in hypoxic conditions [17–21]. The elevated levels of PGK1 protein, detected in the serum of patients affected by pancreatic cancer [22, 23] and in breast cancer tissues [23], suggest a plausible use of PGK1 as a cancer biomarker. Indeed, despite all the observations that report PGK1 overexpression in many cancer types, the role of this enzyme in tumorigenesis is yet unclear [14].

In different cancer types, somatic mutations of PGK1 have been identified, as reported in COSMIC (Catalogue of Somatic Mutations in Cancer) (http://cancer.sanger.ac.uk/cosmic), a database which collects somatic mutation identified in human cancers [24]. Most of these variants are missense ones, also known as nonsynonymous single nucleotide variants (nsSNVs), occurring in the coding region and leading to a polypeptide sequence with amino acid substitutions [25]. By controlling ATP and 3-PG levels, PGK1 plays an important role in coordinating energy production with biosynthesis and redox balance, so mutations of this enzyme can be responsible for alterations of metabolic profile in different cancer cell lines.

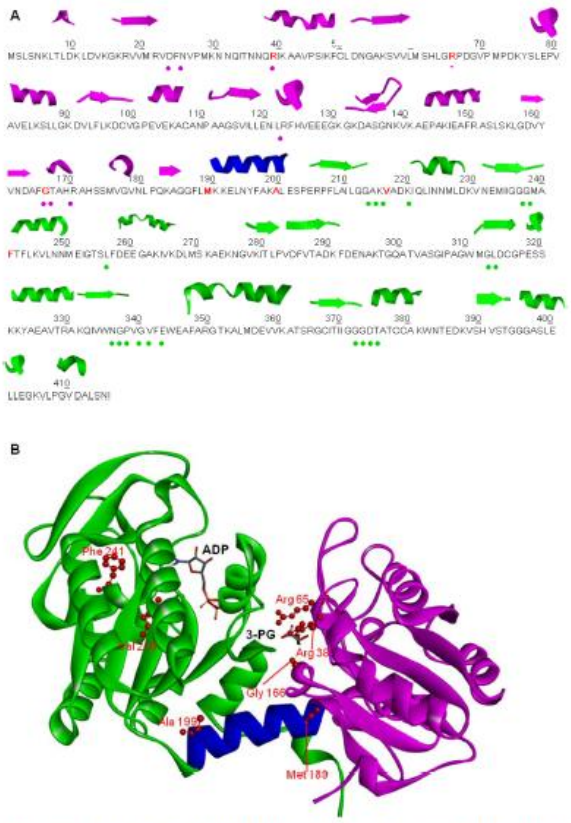


Fig 1. Amino acid sequence and structure of PGK1. N-terminal domain (violet), C-terminal domain (green), hinge region (blue) (PDB: 2XE7, open conformation). (A) Secondary structural elements are shown at the top of the amino acid sequence. Mutated residues are depicted in bold red. The dots under the sequence represent the residues involved in 3-PG or 1,3-BPG binding (pink dot) and in ATP or ADP binding (green dot). (B) Location of the mutations on PGK1 structure (open conformation). Mutated residues are depicted in scaled ball and stick and 3-PG and ADP in stick.

https://doi.org/10.1371/journal.pone.0199191.g001

The effect of nsSNVs on protein stability, protein-protein interactions and protein functions has been investigated for several other protein families [26–29]. Indeed, large-scale computational studies utilizing structural information indicate that a single amino acid

substitution will affect either protein-protein interactions and protein stability [30, 31]. Analysis of physico-chemical properties of natural variants may be helpful to reveal local structural changes that may not affect the overall folding of the structure [28]. Thus, only a detailed experimental analysis can unequivocally reveal the effect of the missense mutations on protein function and/or stability [32].

In this study we selected seven PGK1 nsSNVs found in cancer tissues and annotated in the COSMIC database. The mutated residues are located in the 3-PG binding site, in the flexible hinge region that allows the two domains of the enzyme to approach during the catalytic cycle, and in the MgADP binding site in the C-terminal domain (Fig 1). Herewith we report the effects of nsSNV on PGK1 by analysing the impact of single amino acid replacements on the variants structure, on their conformational stability, and on their catalytic activity compared to the wild type protein.

Materials and methods

Plasmids and site-directed mutagenesis

PGK1 wild type plasmid was kindly provided by Dr. Maristella Maggi, University of Pavia [PMID: 22348148]. Quick Change Site-Directed Mutagenesis Kit (Stratagene) was used to introduce single mutations in the gene encoding for the wild type PGK1 and inserted into the plasmid backbone. The mutagenic oligonucleotides used to obtain the seven selected variants (i.e, R38M, R65W, G166D, M189I, A199V, V216F, and F241S) are listed in S1 Table.

The presence of the desired mutations and the absence of unwanted additional mutations were confirmed by plasmid sequencing of the insert.

Protein expression and purification

Recombinant PGK1 protein has been expressed and purified as described in [33] with minor modifications. PGK1 wild type and mutants were expressed in *E. coli* strain BL21(DE3). 15 mL of overnight culture containing 100 μg/mL ampicillin were used to inoculate 500 mL LB cultures containing ampicillin as antibiotic at a final concentration of 50 μg/mL. Cultures were grown at 37˚C until optical density at 600 nm (OD_{600}) reached 0.6, at which point the protein expression was induced by adding 0.5 mM isopropyl-β-D-thiogalactoside (Sigma-Aldrich). Induced cultures were grown at 37˚C for 5 h in vigorous shaking. At the end of induction, cells were collected by centrifugation and the pellet frozen.

To purify the enzymes, cells were resuspended in 45 mL of Binding buffer (20 mM Tris-HCl pH 8.0, 1 mM EDTA, 1 mM tris(2-carboxyethyl)phosphine (TCEP)) containing a cocktail of EDTA-free protease inhibitors (Sigma-Aldrich) and disrupted by sonication in a Vibracell sonicator with 5 s boosts and 9 s pause, on ice. The lysate was cleared by centrifugation at 15000 rpm and the supernatant, after an additional centrifugation, was applied to a DEAE-Sepharose FF column (GE Healthcare) previously equilibrated in Binding buffer, to remove endogenous protein contaminants and nucleic acids. In the above described conditions, the recombinant PGK1, eluted in the unbound fraction, was concentrated to 2 mL using an Amicon concentrator Ultra-15 (Millipore). In order to remove protein aggregates, concentrated protein was loaded onto a Superdex 200 300/10 gel filtration column equilibrated in 20 mM Tris-HCl pH 8.0, 1 mM EDTA, 200 mM NaCl, 2 mM DTT using an AKTA FPLC system (GE Healthcare). The protein was eluted by isocratic flow at 1.0 mL/min. 2 mL fractions were collected and purity and molecular weight of the protein were checked by SDS–PAGE on a precasted NuPage 4–12% bis-Tris polyacrylamide gel (Invitrogen). Gels were stained with Coomassie blue R-250. Protein concentration was determined spectrophotometrically using a molar absorptivity coefficient (ε_{280}) corresponding to 33460 $M^{-1}cm^{-1}$ for the variant R65W

and 28335 $M^{-1}cm^{-1}$ for wild type and the other variants, based on a molecular mass of 44.615 kDa, and calculated according to Gill and Hippel [34]. Pure protein was used for all structural and stability experiments. The enzyme obtained from 1 liter culture was approximately 65 mg in the case of wild type. High level of expression, ranging from 50 mg (V216F and F241S) to 87 mg (R38M), was obtained for all the mutant enzymes too.

Spectroscopic measurements

Intrinsic fluorescence emission measurements were recorded at a protein concentration of 110 μg/mL for the R65W variant and 130 μg/mL for wild type and all the other variants (0.08 AU_{280nm}), in 20 mM Tris-HCl pH 8.0, 200 mM NaCl and 0.2 mM DTT, in a LS50B spectrofluorimeter (Perkin-Elmer) using a 1.0 cm path length quartz cuvette. Intrinsic fluorescence emission spectra were recorded from 300 to 450 nm (1 nm sampling interval), with the excitation wavelength set at 295 nm. Far-UV (190–250 nm) CD spectra were monitored at 20˚C at a protein concentration ranging over 130–170 μg/mL, using a 0.1 cm path length quartz cuvette, in 20 mM Tris-HCl pH 8.0, 200 mM NaCl and 0.2 mM DTT. Near-UV (250–320 nm) CD spectra were recorded in a 1.0 cm path length quartz cuvette at a protein concentration ranging between 1.50 and 1.70 mg/mL, in 20 mM Tris-HCl pH 8.0, 200 mM NaCl, 1.0 mM EDTA and 2.0 mM DTT. CD measurements were carried out in a JASCO-815 spectropolarimeter (Jasco, Easton, MD, USA) and the results are expressed as the mean residue ellipticity ([Θ]), assuming a mean residue molecular mass of 110 per amino acid residue. All spectroscopic measurements were carried out at 20˚C.

Enzyme activity assay and kinetic studies

PGK1 activity was determined at 20˚C, with 3-PG and MgATP as substrates, by glyceraldehyde-3-phosphate dehydrogenase (GAPDH) coupled spectrophotometric assay according to [33]. The standard reaction mixture contained 100 mM Tris-HCl pH 8.0, 0.5 mM EDTA, 2 mM $MgCl_2$, 0.24 mM NADH, 4U/0.5mL GAPDH, 15.0 mM 3-PG, and 5.0 mM MgATP, in a final volume of 0.5 mL. Kinetic parameters for PGK1 activity towards 3-PG were determined at 20˚C by using at least 10 different concentrations of 3-PG under conditions identical to those described above, at 5.0 mM MgATP fixed substrate concentration. The reaction was started by adding different amounts of PGK1 ranging from 4 ng to 20 μg. All measurements were performed in triplicate in a Lambda 16 computerized spectrophotometer (Perkin-Elmer). Kinetic data were analysed according to [33], using GraphPadPrism 5.04. Results are reported as the mean of three experiments from different protein preparations.

Temperature dependence of PGK1 activity

The activity assay mixture, containing 100 mM Tris-HCl pH 8.0, 0.5 mM EDTA, 2 mM $MgCl_2$, 0.24 mM NADH, 4U/0.5mL GAPDH, 5.0 mM MgATP and 0.2 mM 3-PG in 0.5 mL final volume, was incubated at increasing temperature in a thermostated cuvette. 4 μL of pure enzyme, at 10˚C, were added to 0.5 mL of the assay mixture equilibrated at the desired temperature to start the reaction. The final enzyme concentration was 0.18–12.0 nM. The solution was mixed in the thermostated cuvette and the absorbance at 340 nm was continuously monitored for 10 min. The changes of enzyme activity as a function of temperature was fitted nonlinearly to the Arrhenius equation using GraphPadPrism 5.04 to obtain the activation energies (E_a) for the catalytic reaction

$$k = Ae^{-Ea/RT} \quad (1)$$

where k (s^{-1}) is the rate constant at temperature T (K), A is a reaction specific quantity, R the gas constant (1.987 cal x mol^{-1} x K^{-1}) and E_a the activation energy of the reaction, as described in [27].

Urea-induced unfolding equilibrium

PGK1 wild type and variants (110–160 μg/mL final concentration) were incubated at 20˚C at increasing concentrations of urea (0–8 M) in 20 mM Tris-HCl pH 8.0, 200 mM NaCl, 5.0 mM $MgCl_2$ and 0.2 mM DTT. After 10 min, a time sufficient to reach equilibrium, intrinsic fluorescence emission and far-UV CD spectra (0.2-cm cuvette) were recorded in parallel at 20˚C. To test the reversibility of the unfolding, PGK1 wild type and variants were unfolded at 20˚C in 7.5 M urea at protein concentration ranging over 1.10–1.60 mg/mL in 20 mM Tris-HCl, pH 8.0, in the presence of 2 mM DTT, 5 mM $MgCl_2$ and 200 mM NaCl. After 10 min, refolding was started by 10-fold dilution of the unfolding mixture at 20˚C into solutions of the same buffer used for unfolding containing decreasing urea concentrations. The final protein concentration was 110–160 μg/mL. After 2 h, intrinsic fluorescence emission and far-UV CD spectra were recorded at 20˚C. All denaturation experiments were performed in triplicate.

Thermal stability

PGK1 wild type and variants (110–160 μg/mL) were heated from 20˚C to 80˚C in 20 mM Tris-HCl, pH 8.0, 200 mM NaCl, 0.2 mM DTT, in a 0.1 cm quartz cuvette with a heating rate of 1 degree x min^{-1} controlled by a Jasco programmable Peltier element as described in [27, 29]. The dichroic activity at 222 nm and the photomultiplier were continuously monitored in parallel every 0.5˚C [35]. The solvent contribution at the various temperatures was taken into consideration for all thermal scans. Melting temperature (T_m) values were calculated by taking the first derivative of the ellipticity at 222 nm with respect to temperature, as described in [29]. All denaturation experiments were performed in triplicate.

Data analysis

Far-UV CD spectra recorded as a function of urea concentration were analyzed by a singular value decomposition algorithm (SVD) using the software MATLAB (Math-Works, South Natick, MA) to remove the high frequency noise and the low frequency random errors and determine the number of independent components in any given set of spectra, as described in [27].

The changes in intrinsic fluorescence emission spectra at increasing urea concentrations were quantified as the intensity-averaged emission wavelength, ($\bar{\lambda}$), [36] calculated according to

$$\bar{\lambda} = \Sigma(I_i\lambda_i)/\Sigma(I_i) \quad (2)$$

where λ_i and I_i are the emission wavelength and its corresponding fluorescence intensity at that wavelength, respectively. This quantity is an integral measurement, negligibly influenced by the noise, which reflects changes in the shape and position of the emission spectrum.

Urea-induced equilibrium unfolding transitions monitored by far-UV CD ellipticity and intrinsic fluorescence emission changes were analysed by fitting baseline and transition region data to a two-state linear extrapolation model [37] according to

$$\Delta G_{unfolding} = \Delta G^{H_2O} + m[Urea] = -RTlnK_{unfolding} \quad (3)$$

where $\Delta G_{unfolding}$ is the free energy change for unfolding for a given denaturant concentration,

ΔG^{H_2O} the free energy change for unfolding in the absence of denaturant and m a slope term which quantifies the change in $\Delta G_{unfolding}$ per unit concentration of denaturant, R the gas constant, T the temperature and $K_{unfolding}$ the equilibrium constant for unfolding. The model expresses the signal as a function of denaturant concentration:

$$y_i = \frac{y_N + s_N[X]_i + (y_U + s_U[X]_i)^* \exp[(-\Delta G^{H_2O} - m[X]_i)/RT]}{1 + \exp\left[\frac{(-\Delta G^{H_2O} - m[X]_i)}{RT}\right]} \quad (4)$$

where y_i is the observed signal, y_U and y_N are the baseline intercepts for unfolded and native protein, s_U and s_N are the baseline slopes for the unfolded and native protein, $[X]_i$ the denaturant concentration after the ith addition, ΔG^{H_2O} the extrapolated free energy of unfolding in the absence of denaturant, m the slope of a $\Delta G_{unfolding}$ versus [X] plot. The denaturant concentration at the midpoint of the transition, $[Urea]_{0.5}$, according to Eq (3), is calculated as:

$$[Urea]_{0.5} = \Delta G^{H_2O}/m \quad (5)$$

All unfolding transition data were fitted by using Graphpad Prism 5.04.

Variants crystallization, data collection, structures solution

The protein samples used for crystallization trials contained PGK1 variants at a concentration of 12–18 mg/mL in Tris-HCl 20 mM pH 7.0, in the presence of 10 mM ADP, 25mM $MgCl_2$, 50 mM 3-PG and 10 mM DTT. Crystals were obtained at 298 K by the hanging-drop vapour diffusion method, sealing symmetric drops (1.0+1.0 μL) over 500 μL of reservoir solution.

We performed crystallization experiments with the PGK1 variants V216F, G166D, M189I, R38M in order to obtain the structure of the protein-ligands complexes in both closed and partially closed conformation. The detailed crystallization conditions for all the structures solved are indicated in S2 Table. First we tested the conditions reported in the literature for the partially closed conformation of wild type PGK1 [9] based on Na/K phosphate, obtaining diffracting crystals ($0.1 \times 0.2 \times 0.3$ mm^3) for V216F and M189I. Then a high throughput crystallization screening (HTS) approach was employed for G166D and R38M variants. The crystallization robot (Crystal Phoenix, Art Robbins Instruments) available at the bio-crystal facility at the IBPM-CNR (c/o Department of Biochemical Sciences, Sapienza University of Rome) allowed us to set up up-to crystallization trials using commercial crystallization kits (Crystal Screens I, II, Wizard I, II and Index HT, Hampton Research). The best crystals of R38M and G166D mutants have been obtained in sodium citrate. We also tried to crystallize the closed form of the mutants by adding 20 mM NH_4F in the protein solutions defined above in order to mimic the transition state of the enzyme. Again we first tested the conditions already reported for the wild type PGK1 [10], obtaining crystals only for the M189I mutant. Then we performed HTS with the other mutants without success.

A single-wavelength data set was collected at 100 K from each crystal and processed with XDS [38] and Aimless [39]. The structures were determined by molecular replacement with the program MOLREP (CCP4 suite) [40] using the PDB entries 2WZB and 2ZGV as search models for closed and partially closed conformations respectively. Refinement has been performed using the maximum-likelihood method with the program REFMAC [41] and model building has been done by using Coot [42]. Crystal parameters, beamlines specifications, data collection parameters and refinement statistics are reported in S2 Table.

Results

In this study we selected seven PGK1 variants (R38M, R65W, G166D, M189I, A199V, V216F and F241S) mined from the COSMIC database (http://cancer.sanger.ac.uk/cosmic) [24] and associated to human carcinoma. Most of these PGK1 variants, M189I, A199V, V216F and F241S (Fig 1), are reported to be found in breast carcinoma; the other variants, R38M, R65W and G166D are reported to be present in lung, liver and endometrium carcinoma. The location of the selected mutants mapped onto the PGK1 structure is shown in Fig 1. Residues R38, R65 and G166 are located in the enzyme 3-PG binding site (Fig 1) at the N-terminal domain; in particular, the residues R38 and G166 are part of two helices, whereas the residue R65 is located in a loop (Fig 1). R38 and R65 are involved in 3-PG binding; in particular, R65 NH1 and NH2 groups are involved in an electrostatic interaction with the carboxyl group of 3-PG, whereas R65 Nε and NH2 groups acquire contacts with the phosphate group of 3-PG [43]. G166 is located in the 3-PG binding pocket interacting with the substrate C2 through its main carboxylic group. M189 and A199 are located at the edges of the helix that represents the flexible hinge region that allow the two domains of the enzyme to approach each other during the catalytic cycle (Fig 1). Residue V216 is located in a turn, in proximity of the nucleotide-binding A214 residue; F241 is part of a helix, and they both belong to the MgADP binding pocket, located at the protein C-terminal domain (Fig 1). Among the mutations studied, only R65W involves a surface exposed residue, all others variants, R38M, G166D, M189I, A199V, V216F and F241S are more buried. We generated recombinant protein for each of the identified mutants using site directed mutagenesis and available bacterial expression systems. Introduction of these mutations resulted in soluble recombinant proteins and allowed us to study the consequences of the mutations on PGK1 thermal and thermodynamic stability and the kinetic activity.

Spectroscopic characterization of PGK1 wild type and variants

The conformation in solution of all PGK1 variants was studied by CD and fluorescence spectroscopy. The near-UV CD spectrum of PGK1 wild type represents the spectral contributions of all aromatic residues and is characterized by a strong positive peak centred at around 292 nm, flanked by a shoulder at 286 nm and by negative contributions in the region between 255 and 270 nm, with two peaks centred at 262 and 268 nm (S1 Fig). R38M, G166D, A199V and V216F display near-UV CD spectra closely similar to that of the wild type, except for a slight difference at 277 nm observed for G166D (S1 Fig). Significant differences are observed in the near-UV CD spectra of M189I, F241S and R65W (S1 Fig). In the case of M189I and F241S, the contributions at 262 nm and 268 nm are positive; in addition, the spectrum of M189I loses fine structure features in the region between 270 and 280 nm. The variant R65W shows a near-UV CD spectrum that significantly differs from that of the wild type in intensity, consistent with the substitution of an arginine residue with a tryptophan (S1 Fig). The fluorescence spectra of wild type and variants are centred at the same maximum emission wavelength at around 350 nm, characteristic of tryptophan contribution, with differences in the relative fluorescence emission intensity which is significantly decreased for A199V and increased for R65W, M189I and V216F (S1 Fig).

The far-UV CD spectra of PGK1 wild type and variants show the typical local minimum contributions of alpha helical proteins at around 208 nm and at 222 nm (S1 Fig). The molar ellipticity ratio at 222 and at 208 nm ($\Theta_{222}/\Theta_{208}$) is indicative of interhelical contacts present in helix bundle and coiled coil structures and it is generally used to distinguish between coiled coil helices ($\geq$1.0) and non-interacting helices (0.8–0.9) [44,45]. The $\Theta_{222}/\Theta_{208}$ is 1.2 for the wild type, it is decreased to 1.1 only for the variants R38M and M189I and it is increased to 1.3

for V216F (S1 Fig). Notably, in the case of M189I the far-UV dichroic activity is decreased with respect to that of the wild type (S1 Fig). These differences in secondary structure may suggest differences in interhelical interactions and dynamic fluctuation in solution for some of PGK1 variants.

PGK1 wild type and variants: Kinetic properties and temperature dependence of enzyme activity

Enzymatic assays of PGK1 wild type and variants. Kinetic properties of PGK1 wild type and variants were investigated by monitoring the reverse reaction at 20˚C at a fixed concentration of free Mg^{2+} and using 3-PG as variable substrate [33]. All the variants display altered kinetic properties, in comparison with the wild type. The kinetic properties of wild type, M189I and F241S, revealed a non Michaelis–Menten behaviour, as indicated by double reciprocal plots typical of substrate activation (data not shown). PGK1 specific activity, analysed using 3-PG as variable substrate over the concentration range 0.05–15.0 mM, is decreased for all the variants with the exception of F241S that displays a two fold increase of activity (Table 1). Notably, all the kinetic parameters indicate a significant decrease of the catalytic efficiency (k_{cat}/K_M), particularly evident for R38M which displays a catalytic efficiency 10 milion fold lower than that of the wild type. All the variants display an increase of 3-PG K_M values, particularly evident for F241S that shows a K_M value 14-fold higher than that of the wild type (Table 1).

Temperature dependence of enzyme activity. The kinase activity of PGK1 wild type and variants was analysed as a function of temperature (10–42˚C) using a low concentration of 3-PG substrate (0.02 mM, well below the K_M value) (Fig 2) to study the enzymatic release of substrate and product that may be related to its flexibility. The optimal temperatures for catalysis, in these conditions, were estimated to be 37˚C for the wild type and most of the variants, around 35˚C for F241S, and around 40˚C for R38M and V216F (Table 2 and Fig 2). As shown in Table 2, the lowest activation energy value (E_a) is that of the F241S variant (6.39 kcal/mol) which is also the variant displaying a k_{cat} higher than that of the wild type enzyme (213.0 vs 89.8 s^{-1}). As mentioned before, the activation energies have been measured at a 3-PG concentration of 0.02 mM, well below the K_M value. At this substrate concentration, the enzyme is not saturated, the complex enzyme-3-PG-MgATP may dissociate to form again the enzyme-MgATP complex and free 3-PG or may catalyse the formation of the two products MgADP and 1,3-BPG. The activation energy measured in these conditions reflects the enzymatic

Table 1. Kinetic parameters of PGK1 wild type and variants.

	[PGK1] (nM)	[3PG] (mM)	K_M (mM)	kcat (s-1)	kcat/ K_M (s-1 mM-1)	Vmax (μM/min)	Specific activity (μM/min•μg)
Wild type	0.18	0.05–15.0	0.40 ± 0.03	89.8 ± 1.5	224.5	0.97	248.7
R38M	9.04×10^5	3.00–15.0	3.15 ± 0.40	$(7.2 \pm 0.5) \times 10^{-6}$	2.28×10^{-6}	0.39	0.02
R65W	0.35	1.00–15.0	1.61 ± 0.13	23.7 ± 0.5	14.7	0.50	63.8
G166D	0.70	0.90–15.0	1.54 ± 0.15	5.9 ± 0.2	3.8	0.25	15.9
M189I	0.18	0.20–15.0	0.58 ± 0.50	30.5 ± 1.4	52.7	0.33	84.6
A199V	0.18	0.80–15.0	1.95 ± 0.15	76.6 ± 1.7	39.3	0.83	212.0
V216F	7.00	0.10–15.0	0.78 ± 0.05	1.9 ± 0.03	2.4	0.93	5.5
F241S	0.18	0.20–15.0	5.87 ± 0.7	213.0 ± 14.0	31.5	2.30	515.4

PGK1 activity was determined at 20˚C, with 3-PG and 5 mM MgATP (fixed substrate concentration), as described in Materials and Methods. Kinetic parameters for PGK1 activity were determined at 20˚C by using at least at 10 different concentrations of 3-PG. Data are reported as the mean ± SE of the fit.

https://doi.org/10.1371/journal.pone.0199191.t001

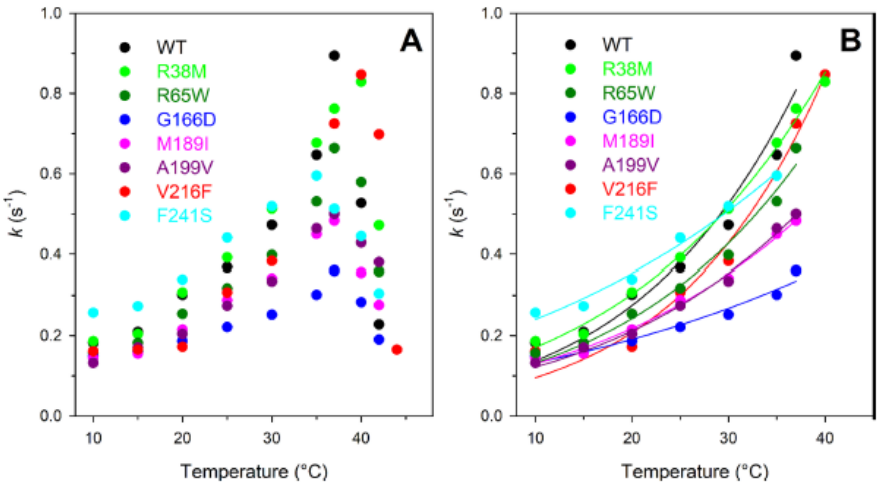


Fig 2. Effect of temperature on kinase activity of PGK1 wild type and variants. (A) Temperature dependence of kinase activity of PGK1 wild type and variants. (B) Non-linear fit of the temperature dependence of PGK1 activity to the Arrhenius equation Eq (1). Assays were performed under the conditions described in Materials and Methods, using 0.18–12.0 nM enzyme.

https://doi.org/10.1371/journal.pone.0199191.g002

release of both the 3-PG substrate and the products and it may be related to the enzyme flexibility.

Thermal stability

The thermal stability of PGK1 wild type and variants was studied between 20˚C and 80˚C by continuously monitoring the ellipticity changes at 222 nm, where the main amplitude was observed (S2 Fig). The thermal denaturation process is irreversible, as indicated by the spectra

Table 2. Effect of temperature on kinase activity of PGK1 wild type and variants.

	T_{max} (˚C)	E_a (kcal/mol)
Wild type	37	11.47 ± 0.7
R38M	40	9.44 ± 0.34
R65W	37	10.03 ± 0.60
G166D	37	6.00 ± 0.30
M189I	37	8.50 ± 0.35
A199V	37	9.11 ± 0.39
V216F	40	12.82 ± 0.70
F241S	35	6.39 ± 0.41

E_a was determined by the Arrhenius equation Eq (1) in the temperature range between 10˚C and the optimal temperature of each protein.

https://doi.org/10.1371/journal.pone.0199191.t002

Table 3. Melting temperatures and thermodynamic parameters for urea-induced unfolding equilibrium of PGK1 wild type and variants measured by far-UV CD and fluorescence spectroscopy.

	T_m (°C)	ΔG^{H_2O} (kcal/mol)		m (kcal/mol•M)		$[Urea]_{0.5}$ (M)	
		CD ($[\Theta]_{222}$)	Fluorescence ($\bar{\lambda}$)	CD ($[\Theta]_{222}$)	Fluorescence ($\bar{\lambda}$)	CD ($[\Theta]_{222}$)	Fluorescence ($\bar{\lambda}$)
Wild type	52.5	8.29 ± 0.57	8.06 ± 0.46	3.54 ± 0.25	3.10 ± 0.17	2.34	2.60
R38M	51.5	6.91 ± 0.73	5.54 ± 0.37	3.54 ± 0.37	1.87 ± 0.12	1.95	2.96
R65W	52.5	7.04 ± 0.32	7.20 ± 0.66	3.07 ± 0.14	2.63 ± 0.24	2.29	2.73
G166D	53.0	4.19 ± 0.31	4.94 ± 0.29	2.33 ± 0.16	1.51 ± 0.09	1.75	3.27
M189I	50.5	4.87 ± 0.31	5.01 ± 0.48	2.36 ± 0.15	1.82 ± 0.17	2.06	2.75
A199V	53.5	7.14 ± 0.57	6.82 ± 0.72	3.41 ± 0.14	2.71 ± 0.22	2.23	2.63
V216F	52.0	6.45 ± 0.34	5.76 ± 0.54	2.99 ± 0.16	2.14 ± 0.20	2.16	2.69
F241S	49.0	3.02 ± 0.17	3.80 ± 0.26	2.06 ± 0.10	1.67 ± 0.11	1.47	2.27

T_m values were calculated by taking the first derivative of the ellipticity at 222 nm with respect to temperature. Urea-induced unfolding equilibrium data were measured as described in Materials and Methods by monitoring the ellipticity at 222 nm ($[\Theta_{222}]$) and fluorescence intensity averaged emission wavelength ($\bar{\lambda}$). ΔG^{H_2O} and m values were obtained from Eq (3); $[Urea]_{0.5}$ was calculated from Eq (5). Data are reported as the mean ± SE of the fit.

https://doi.org/10.1371/journal.pone.0199191.t003

measured at the end of the cooling phase that differ from those of the native proteins measured at the beginning of the thermal transitions. The temperature-induced ellipticity changes at 222 nm occur in an apparent cooperative transition and overlap with the increase of the photomultiplier tube voltage above 370 V (data not shown) at increasing temperature, suggesting that the temperature-induced unfolding is accompanied by protein aggregation, as revealed by the presence of a large amount of precipitate in the cuvette at the end of the cooling phase. The parameter chosen to compare the transition curves of the proteins is the melting temperature (T_m) defined as the midpoint of the denaturation process and calculated by plotting the first derivative of the molar ellipticity values as a function of temperature (S2 Fig, inset). T_m values range from 49.0 °C to 53.5°C being 52.5°C the T_m obtained for the wild type protein (Table 3). The variant R65W shows the same T_m value of the wild type, a modest increase in T_m value is observed for the variants G166D and A199V; all the other variants show T_m values lower than that of the wild type, with F241S showing a T_m value three degrees below that of the wild type (Table 3). Notably, differences in the amplitude observed for the thermal transitions of most of the variants (S2 Fig) may be referred to the difference in the dichroic activity at 222 nm of their corresponding native states, as also indicated in far-UV CD spectra reported in S1 Fig.

Urea-induced equilibrium unfolding transitions

PGK1 wild type and variants reversibly unfold in urea at 20 °C in 20 mM Tris-HCl pH 8.0, 1 mM EDTA, 200 mM NaCl and 0. 2 mM DTT. The effect of increasing urea concentrations (0–8 M) on the protein structure was analyzed by far-UV CD and fluorescence spectroscopy (S3 Fig). The fluorescence changes at increasing urea concentration were measured by calculating the intensity averaged emission wavelength $\bar{\lambda}$, an integral measurement that depends both on the position and the shape of the spectrum. The same samples used to monitor the fluorescence changes during the unfolding transitions were used to monitor the far-UV CD ellipticity changes, to allow a direct comparison of the data. The urea-induced changes in 222 nm ellipticity and in the emission fluorescence show a sigmoidal dependence on urea concentration and follow an apparent two-state transition without any detectable intermediate (S3 Fig). The unfolding process is fully reversible upon dilution of the denaturant either for the wild type and the variant proteins. The denaturation curves relative to the apparent two-state equilibrium unfolding measured by far-UV CD and by fluorescence have been fitted to a two-state

model according to Eq (4) to obtain the thermodynamic parameters ΔG^{H_2O}, the free energy change for unfolding in the absence of denaturant, and *m*, a parameter that refers to the amount of protein surface area exposed to the solvent during unfolding (Table 3) [46]. Interestingly, the *m* values determined either for PGK1 wild type and variants are lower than that predicted for a protein of 417 amino acid residues unfolded in urea [47, 48], suggesting a more complex, presumably non-two state, unfolding process with undetectable intermediates (S3 Fig). ΔG^{H_2O} values of PGK1 variants are similar to those of the wild type, with the exception of G166D, M189I and F241S that show a significant decrease in the thermodynamic parameters (Table 3), mainly due to a decrease in *m* values. A decrease in *m* value may suggest either a destabilization of the native state or the presence of unfolding intermediates, not detectable at the equilibrium. The occurrence of undetectable unfolding intermediates may be the case of R38M and G166D, as suggested by the non coincidence of the $[Urea]_{0.5}$ values obtained by far-UV CD and fluorescence changes, mainly due to the differences in *m* value (Table 3).

Structural analysis

We investigated the structure of PGK1 variants R38M, G166D, V216F and M189I in the presence of MgADP and 3-PG through X-ray crystallography in order to reveal possible local and global variations with respect to the wild type protein. Two different crystallographic conformations of PGK1 have been already identified: the partially closed conformation, able to bind substrates, and the closed conformation, the catalytically competent state. Another form, defined as open and that is the most populated conformation in solution, has been detected through small-angle X-ray scattering [7]. It is interesting to note that all the mutants crystallized in the partially closed conformation but only M189I crystallized even in the closed form, suggesting that for the variants R38M, G166D and V216F the closed form is less favourable in the crystallization conditions. The overall structure is conserved in all the mutants compared to the wild type (Table 4, S4 Fig). The details of local variation are discussed case-by-case.

R38M variant. The residue R38 is placed in the N-terminal 3-PG binding domain and is a residue important for the substrate binding and its correct positioning towards the co-substrate MgADP (Fig 3). Moreover, together with residues K215 and K219, R38 is critical for the charge balancing of the transition state, directly interacting with the transferring phosphate group in the closed conformation of PGK1 [10].

The crystal structure of R38M contains ADP but does not contain 3-PG (Fig 3A), in line with the role of R38. The absence of the 3-PG ligand could account even for the variant

Table 4. Superimposition between the PGK1 wild type and the crystallized variants.

Mutant	Domain	Role/position of mutated residue	Conformation	Ligands	RMSD* (Å²)
R38M	3-PG-binding	Binding of 3-PG, charge balancing in the transition state	Partially closed	MgADP	vs 2ZGV*: 0.23 vs 2XE7*: 0.96
G166D	3-PG-binding	Close to 3-PG COO⁻	Partially closed	MgADP	vs 2ZGV*: 0.62 vs 2XE7*: 1.17
V216F	ADP-binding	Part of the catalytic loop 211–219 that favors closed conformation	Partially closed	MgADP, 3-PG	vs 2ZGV*: 0.60 vs 2XE7*: 1.22
M189I	Hinge	N-terminal portion of the hinge	Partially closed	MgADP, 3-PG	vs 2ZGV*: 0.60 vs 2XE7*: 1.18
			Closed	MgADP, 3-PG	vs 2WZB*: 0.25

RMSD*: root mean square deviation obtained from the superimposition of each variant with respect to wild type PGK1. 2ZGV*: wild type binding MgADP; 2XE7*: wild type binding MgADP and 3-PG; 2WZB*: wild type binding MgADP, 3-PG and MgF_3^-

https://doi.org/10.1371/journal.pone.0199191.t004

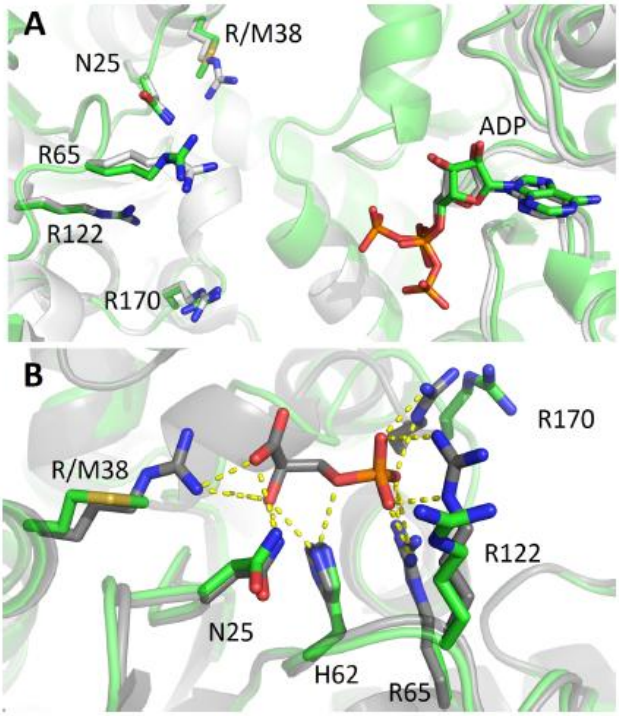


Fig 3. Detail of the 3-PG binding site in the variant R38M in comparison with PGK1 wild type. Overlap of R38M variant (green) and PGK1 wild type crystallized (A) in the absence of 3-PG (PDB 2ZGV) (light grey) or (B) in the presence of 3-PG (PDB: 2XE7) (grey). The view in B is rotated about 90° anticlockwise with respect to A. ADP, 3-PG and the residues involved in 3-PG binding are shown as sticks.

https://doi.org/10.1371/journal.pone.0199191.g003

reluctance to crystallize in the closed form since the binding of both substrates is required for the stabilization of this conformation. Taken together these observations justify the dramatic effect of R38M mutation on kinetic parameters (see Table 1). Indeed, upon R38M mutation the K_M increases from 0.40 to 3.15 mM and the turnover number strongly decreases from 89.8 to 7.2×10^{-6} s^{-1}.

Apart from the mutation, in position 38, the local geometry of residues belonging to the 3-PG binding pocket is very similar to that of the wild type protein crystallized in the absence of 3-PG (PDB code: 2ZGV, Fig 3B). The only differences between the two structures concern the α-helix 374–382, visible in the R38M variant but not in the wild type 2ZGV structure, and the position of the β-phosphate group, which in R38M points towards this helix (Fig 3A).

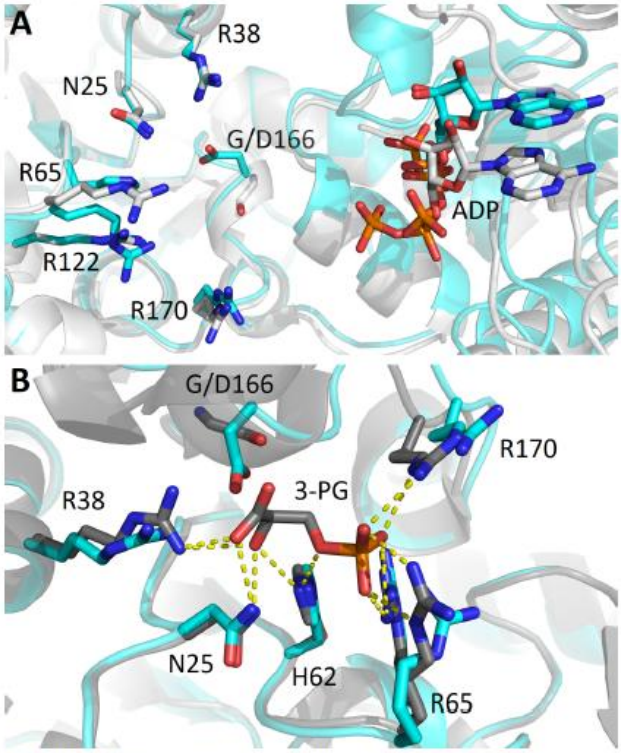


Fig 4. Detail of the 3-PG binding site in the variant G166D in comparison with PGK1 wild type. Overlap of G166D variant (cyan) and PGK1 wild type crystallized (A) in the absence of 3-PG (PDB 2ZGV) (light grey) or (B) in the presence of 3-PG (PDB: 2XE7) (grey). The view in B is rotated about 90° anticlockwise with respect to A.

https://doi.org/10.1371/journal.pone.0199191.g004

However, both the helix and the phosphate present some variability upon different PGK1 structures indicating that it could be dependent on crystallization conditions.

G166D variant. G166 is one of the residues lining the 3-PG binding site and its substitution for the bulky and negatively charged aspartate modifies the shape and the charge of the substrate binding pocket. The structure of G166D (Fig 4) reveals that the aspartate Oδ is engaged in a salt bridge with the R38 NH1 (OD-(D166)-NH1(Arg38) = 3.3 Å) altering its capability of interacting with 3-PG. These modifications of the pocket impair the binding of 3-PG, that is not present in the structure, and the stabilization of the transition state through R38, thus affecting both k_{cat} (5.9 s^{-1}) with a 15-fold decrease, and K_M (1.54 mM) increased 4- fold.

V216F variant. V216 belongs to the loop 214–219, close to the ADP-binding site. The rearrangement of this loop has been suggested to be important for the enzyme transition to the closed conformation upon binding of the two ligands [7]. Indeed, in the wild type structure in complex with the MgADP ligand the loop assumes two conformations, promoting the reorientation of K215 and D218, now ready to interact with γ-phosphate of 3-PG and D65, in the closed conformation. In the V216F variant structure in complex with MgADP, the loop adopts the same conformation as in apo-PGK1. As shown in Fig 5B, replacement of the valine with a phenylalanine keeps the loop in the apo-like conformation since the aromatic residue tends to stay buried. The lower mobility of the loop hampers the transition from the open to the closed conformation affecting the enzyme catalytic activity (the k_{cat} of the mutant is 1.9 s^{-1}, 45-fold lower than that of the wild type). Conversely, no variation is found in ADP-binding. In fact, as shown in Fig 5, the residues directly involved in the ADP binding (E343, G312, A214) and the Mg^{2+}ion bound through the β and γ-phosphate and D374 preserve the same positions as in the wild type structure.

M189I variant. Residue 189 is part of the 189–201 α-helix, the structural element that represents the hinge connecting the N-terminal 3-PG binding domain and the C-terminal ADP binding domain. The structure of M189I has been solved both in open and closed conformation. Similarly to the other variants, M189I partially open structure is coincident to that of the wild type but in this case both ADP and 3-PG are clearly visible. It should be noticed that in the M189I variant structure the hydroxyl carboxylate moiety of the 3-PG is 90°-rotated with respect to the position adopted in the wild type structure. This could be due to the presence, in the active site, of a phosphate ion which may induce a change in the mode of 3-PG binding (Fig 6A). Since M189I mutation causes neither overall nor local conformational change, the K_M of the mutant is very similar to that of the wild type protein (0.58 mM *vs* 0.40 mM) and the k_{cat} is only three times lower than that of the wild type protein (30.5 *vs* 89.8 s^{-1}). The structure in the closed conformation was obtained in the same condition of the wild type protein (PDB code: 2WZB). As shown in Fig 6B, the structure appear to be identical to the wild type ones; also the position of the two ligands, ADP and 3-PG, is conserved. Interestingly, in the M189I structure in the closed conformation MgF_3^- is not present, therefore the transition complex, evident in the wild type structure, is not formed in the M189I variant. Nevertheless, all the residues stabilizing the transition state are superposable between the wild type and mutant structure.

Discussion

Phosphoglycerate kinase 1 (PGK1) is an ubiquitous enzyme expressed in all somatic cells that plays a central role in glycolysis where it provides energy in form of ATP through the reversible phosphotransfer reaction from 1,3-bisphosphoglycerate (1,3-BPG) to MgADP in order to produce 3-phosphoglycerate (3-PG) and MgATP in the presence of free magnesium. PGK1 is a central enzyme in the cancer cells metabolism since they utilize preferentially glycolysis to produce ATP. Alteration of PGK1 levels has been reported in different cancer types and this enzyme is considered as a negative prognostic marker [23]. In the general metabolic reprogramming of cancer cells, glycolytic enzymes play a pivotal role since many intermediates of the glycolytic pathway may also be utilized for the increased biosynthetic demand of proliferating tumor cells [49]. Indeed, one of the consequence of the metabolic change in cancer cells is the request of reducing equivalent for biosynthesis that can be provided through the deviation of glycolysis intermediate in the shunt of pentose phosphate pathway [50].

In this study, we selected from the COSMIC database some somatic PGK1 single nucleotide variants found in cancer tissues. The variants involve amino acid residues located in the

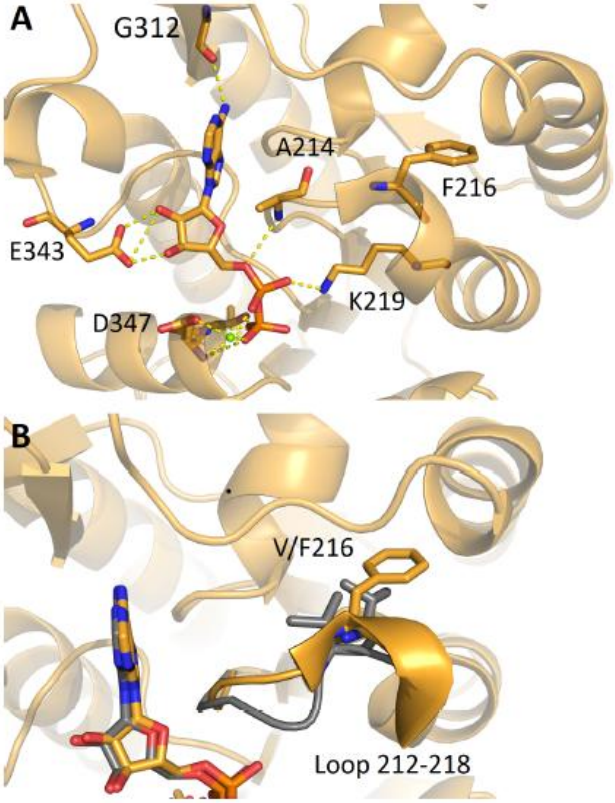


Fig 5. ADP binding site in the variant V216F in comparison with PGK1 wild type. (A) Detail of the binding of ADP to V216F variant (orange). ADP and the residues involved in the interaction are depicted as sticks, Mg^{2+} ion as a green sphere. (B) Superimpostition with wild type PGK1 crystallized with ADP and 3-PG (PDB: 2XE7) (grey), highlighting how the mutation V216F impairs the movement of the loop 212–218. The residue V/F216 and ADP are shown as sticks.

https://doi.org/10.1371/journal.pone.0199191.g005

enzyme 3-PG binding domain, in the nucleotide binding domain and in the hinge region connecting the two domains (Fig 1) and are different from those associated to PGK1 deficiency and hereditary non-spherocytic haemolytic anaemia [33].

Independently from the position of the mutation on the protein, the main result of the aminoacid substitutions, object of this report, is a significant change in the variants catalytic

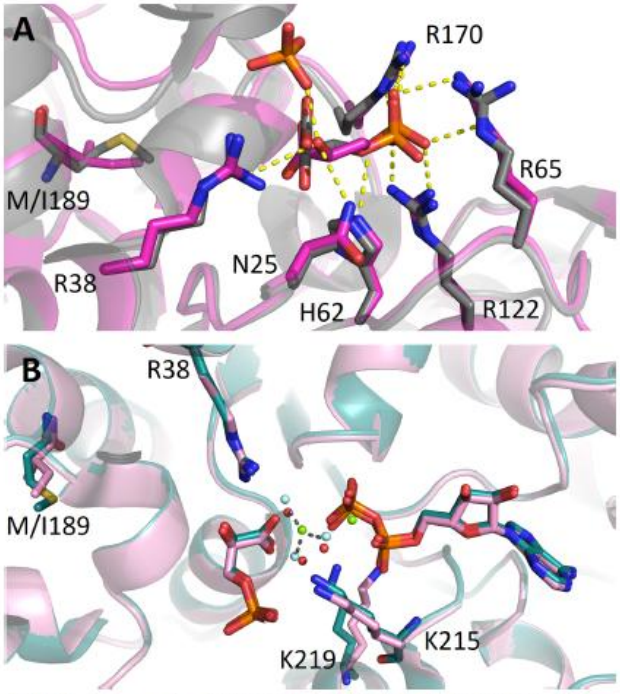


Fig 6. M189I variant in comparison with PGK1 wild type. (A) Detail of the binding of 3-PG to M189I variant in the partially-closed conformation (magenta) superimposed with partially-closed wild type PGK1 (PDB: 2XE7) (grey). 3-PG and the residues involved in the interaction are depicted as sticks. (B) Superimpostition of M189I variant in the closed conformation (pink) with wild type PGK1 crystallized with ADP, 3-PG and MgF_3^- (PDB: 2WBZ) (deep teal). MgF_3^- is shown as green and cyan spheres. Three water molecules (red spheres) bound to M189I mimic the position of F^- in MgF_3^-.

https://doi.org/10.1371/journal.pone.0199191.g006

properties. All the variants display a reduced catalytic efficiency, particularly evident for the variant R38M, that shows a decrease of eight order of magnitude of the k_{cat}/K_M value, mainly due to a dramatic decrease of the turnover number (Table 1). As shown in the structure (Fig 3), R38 is located in the 3-PG binding site and interacts electrostatically with the carboxyl group of 3-PG, hence it is critical for the positioning of the substrate in the catalytic site and for the stabilization of the transition state (Fig 6B). Accordingly, the substitution of the R38 implies a dramatic decrease of the k_{cat} and an increase of K_m (see Table 1).

Mutation of the other residues involved in 3-PG binding, R65W, a positively charged arginine with the non polar tryptophan, and G166D, a substitution of glycine with a negatively

charged residue, leads to an increase in K_m values and, in the case of G166D, also to a reduced turnover number. As shown in the structure, the introduction of a negatively charged residue in the 3-PG binding site alters the affinity of the protein for the substrate (which is not found in the structure) by occupying part of the binding site and by engaging the R38 in a new salt bridge (Fig 4). Notably, the binding affinity for 3-PG is also significantly reduced in the variant A199V, a mutation placed at the end of the hinge region. Regarding the mutants of the nucleotide binding domain, V216F shows a decrease in the turnover number of about fifty fold, while F241S a notable increase in the K_m value accompanied by a two-fold increase in the turnover number. Interestingly, the turnover numbers of M189I and A199V, both in the hinge region and that of R65W, in the 3-PG binding site, are comparable to the turnover number of the wild type, indicating that these mutations are not affecting the release of the product. The reduced catalytic efficiency of all variants is supported by the local structural changes observed in the X-ray structure of the crystallized variants that involve mutation of residues located either in the 3-PG binding domain (R38M, G166D) or in the nucleotide binding site (V216F) or in the hinge (M189I). Notably, the structural comparative analysis shows that the introduced mutations do not cause large conformational changes in the protein tridimensional structure but only local changes involving the mutated residues and their chemical surrounding. These local changes may be responsible for local differences in protein flexibility that become evident in the values of activation energy that is decreased for all the cancer associated variants, with the notable exception of V216F. In this variant, the introduction of a phenylalanine residue reinforces the hydrophobic interactions between the 211–219 loop and the 243–263 region thereby stabilizing the C-terminal domain (Fig 5). This local stabilization, which decreases the flexibility of the ATP binding site, is sufficient to determine the decrease of about 50-fold of the turnover number with respect to that of the wild type protein. The global effects of local alterations are evident from the analysis of the thermodynamic parameters that point to a destabilization of the native state, as indicated by the decrease of ΔG^{H_2O} values for all the variants, particularly evident for G166D, M189I and F241S that involve mutation of residues located in the 3-PG binding site, in the hinge region and in the nucleotide binding site, respectively (Fig 1). Indeed, the significant decrease of the *m* values observed for G166D, M189I and F241S suggests a decrease of their surface area exposed to the solvent during unfolding.

The reduced catalytic properties and the decrease of the stability of PGK1 variants may suggest that their expression in the neoplastic cells would significantly burden the metabolic flux through the glycolytic pathway. Thus, metabolic intermediates may accumulate and be converted in other products through alternative pathway, such as the pentose shunt, that could supply for the biosynthetic need of the growing and proliferating tumor cell. On the other hand, the occurrence of a "glycolytic jam" may induce the cell to increase the uptake of glucose [50].

Glycolytic enzymes may play additional roles other than sustaining the cell metabolism, and translocate from the cytoplasm into the nucleus and/or mitochondria thus displaying additional non-metabolic functions [51, 52]. Thus, mutations in a glycolytic enzyme may affect not only its metabolic activity but also its alternative non-metabolic function. PGK1, as a moonlighting enzyme, play alternative roles in different cellular compartments. In the mitochondria, PGK1 acts as a protein kinase and inhibits pyruvate metabolism promoting the Warburg effect [14]. In the nucleus, PGK1 binds to DNA polymerases α and ε and increases the synthesis of DNA [13]. In addition PGK1, when secreted in extracellular compartment by tumor cells, binds plasmin thus permitting the cleavage of plasminogen in order to generate the vascular inhibitor angiostatin [53].

It is noteworthy that some of the PGK1 variants found in cancer tissues involve residues important for mitochondrial translocation, i.e. the region between residues 38 and 43. In this

regard, R38, replaced with a methionine in the variant R38M, is placed in a critical position. Similarly, A199V is in close proximity to S203 which, prior to mitochondria translocation, is phosphorylated under hypoxia conditions [14].

In conclusion, the present paper discloses the structure and functional features of nsSNVs of PGK1 found in cancer cells: R38M, R65W, G166D, M189I, A199V, V216F and F241S. We succeed in solving the X-ray crystal structures of the variants with mutation in the N-terminal region (R38M and G166D), in the C-terminal region (V216F) and in the hinge region (M189I). We solved the structures of all these mutants in the partially closed conformation, and we succeed to solve the structure of M189I also in the totally closed conformation.

Our study reveals that most of the PGK1 nsSNVs found in cancer tissues display a decreased catalytic efficiency together with a destabilization of the native state. However, the structural and functional data show only local changes around the mutated residues, without large conformational changes in the protein structure. All these variants are less efficient with respect to the wild type protein and the loss of efficiency is related, as shown by the crystals structures, to the decreased affinity for the substrate as in the case of the R38M and G166D, or to an increased local rigidity of the substrates binding sites as in the case of V216F and M189I. A less efficient PGK1 in cancer cells may point to its moonlighting functions to satisfy cancer cell proliferation and survival requirements.

Supporting information

S1 Table. List of oligonucleotides used for site-directed mutagenesis (PDF).
(PDF)

S2 Table. Crystallization conditions, data collection parameters, refinement statistics of the PGK1 variants (PDF).
(PDF)

S1 Fig. Spectral properties of PGK1 wild type and variants (PDF).
(PDF)

S2 Fig. Thermal unfolding of PGK1 wild type and variants (PDF).
(PDF)

S3 Fig. Urea-induced equilibrium unfolding of PGK1 wild type and variants (PDF).
(PDF)

S4 Fig. Overall fold of PGK1 versions compared to the wild type (PDF).
(PDF)

Author Contributions

Conceptualization: Andrea Ilari, Alessandra Pasquo, Roberta Chiaraluce, Valerio Consalvi.

Data curation: Annarita Fiorillo, Maria Petrosino, Andrea Ilari, Alessandra Pasquo, Roberta Chiaraluce, Valerio Consalvi.

Formal analysis: Annarita Fiorillo, Maria Petrosino, Andrea Ilari, Alessandra Pasquo, Valerio Consalvi.

Funding acquisition: Roberta Chiaraluce, Valerio Consalvi.

Investigation: Annarita Fiorillo, Maria Petrosino, Andrea Ilari, Alessandra Pasquo, Alessandra Cipollone, Valerio Consalvi.

Methodology: Annarita Fiorillo, Maria Petrosino, Andrea Ilari, Alessandra Pasquo, Alessandra Cipollone, Roberta Chiaraluce.

Project administration: Roberta Chiaraluce, Valerio Consalvi.

Software: Maria Petrosino, Roberta Chiaraluce, Valerio Consalvi.

Supervision: Andrea Ilari, Alessandra Pasquo, Roberta Chiaraluce, Valerio Consalvi.

Validation: Annarita Fiorillo, Maria Petrosino, Alessandra Pasquo, Maristella Maggi, Roberta Chiaraluce, Valerio Consalvi.

Visualization: Maristella Maggi.

Writing – original draft: Maria Petrosino, Andrea Ilari, Roberta Chiaraluce, Valerio Consalvi.

Writing – review & editing: Annarita Fiorillo, Maria Petrosino, Andrea Ilari, Alessandra Pasquo, Maristella Maggi, Roberta Chiaraluce, Valerio Consalvi.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. CA Cancer J Clin. 2016; 66:7–30. https://doi.org/10.3322/caac.21332 PMID: 26742998
2. Bertout JA, Patel SA, Simon MC. The impact of O_2 availability on human cancer. Nat Rev Cancer. 2008; 8: 967–975. https://doi.org/10.1038/nrc2540 PMID: 18987634
3. Kroemer G, Pouyssegur J. Tumor cell metabolism: cancer's Achilles'heel. Cancer Cell. 2008; 13:472–482. https://doi.org/10.1016/j.ccr.2008.05.005 PMID: 18538731
4. Sheng H, Tang W. Glycolysis inhibitors for anticancer therapy: a review of recent patents. Recent Pat Anticancer Drug Discov. 2016; 11:297–308. PMID: 27087655
5. Rodríguez-Enríquez S, Gallardo-Pérez JC, Hernández-Reséndiz I, Marín-Hernández A, Pacheco-Velázquez SC, López-Ramírez SY, et al. Canonical and new generation anticancer drugs also target energy metabolism. Arch Toxicol. 2014; 88:1327–1350. https://doi.org/10.1007/s00204-014-1246-2 PMID: 24792321
6. Vas M, Varga A, Gráczer E. Insight into the mechanism of domain movements and their role in enzyme function: example of 3-phosphoglycerate kinase. Curr Protein Pept Sci. 2010; 11:118–147. PMID: 20088776
7. Zerrad L, Merli A, Schröder GF, Varga A, Gráczer É, Pernot P, et al. A spring-loaded release mechanism regulates domain movement and catalysis in phosphoglycerate kinase. J Biol Chem. 2011; 286:14040–14048. https://doi.org/10.1074/jbc.M110.206813 PMID: 21349853
8. Beutler E. PGK deficiency. Br J Haematol. 2007; 136:3–11. https://doi.org/10.1111/j.1365-2141.2006.06351.x PMID: 17222195
9. Gondeau C, Chaloin L, Lallemand P, Roy B, Périgaud C, Barman T, et al. Molecular basis for the lack of enantioselectivity of human 3-phosphoglycerate kinase. Nucleic Acids Res. 2008; 36:3620–3629. https://doi.org/10.1093/nar/gkn212 PMID: 18463139
10. Cliff MJ, Bowler MW, Varga A, Marston JP, Szabó J, Hounslow AM, et al. Transition state analogue structures of human phosphoglycerate kinase establish the importance of charge balance in catalysis. J Am Chem Soc. 2010; 132:6507–6516. https://doi.org/10.1021/ja100974t PMID: 20397725
11. Chen X, Zhao C, Li X, Wang T, Li Y, Cao C, et al., Terazosin activates Pgk1 and Hsp90 to promote stress resistance. Nat Chem Biol. 2015; 11:19–25. https://doi.org/10.1038/nchembio.1657 PMID: 25383758
12. Wang J, Wang J, Dai J, Jung Y, Wei CL, Wang Y, et al. A glycolytic mechanism regulating an angiogenic switch in prostate cancer. Cancer Res. 2007; 67:149–159. https://doi.org/10.1158/0008-5472.CAN-06-2971 PMID: 17210694
13. Boukouris AE, Zervopoulos SD, Michelakis ED. Metabolic Enzymes Moonlighting in the Nucleus: metabolic Regulation of Gene Transcription. Trends Biochem Sci. 2016; 41:712–730. https://doi.org/10.1016/j.tibs.2016.05.013 PMID: 27345518
14. Li X, Jiang Y, Meisenhelder J, Yang W, Hawke DH, Zheng Y, et al. Mitochondria-translocated pgk1 functions as a protein kinase to coordinate glycolysis and the tca cycle in tumorigenesis. Mol Cell. 2016; 61:705–719. https://doi.org/10.1016/j.molcel.2016.02.009 PMID: 26942675

15. Ariosa AR, Klionsky DJ. A novel role for a glycolytic pathway kinase in regulating autophagy has implications in cancer therapy. Autophagy. 2017; 13: 1091–1092. https://doi.org/10.1080/15548627.2017.1321723 PMID: 28537472

16. Qian X, Li X, Cai Q, Zhang C, Yu Q, Jiang Y et al. Phosphoglycerate Kinase 1 Phosphorylates Beclin1 to Induce Autophagy. Mol Cell. 2017 Mar; 65:917–931. https://doi.org/10.1016/j.molcel.2017.01.027 PMID: 28238651

17. Ahmad SS, Glatzle J, Bajaeifer K, Buhler S, Lehmann T, Konigsrainer I, et al. Phosphoglycerate kinase 1 as a promoter of metastasis in colon cancer. Int J Oncol. 2013; 43: 586–590. https://doi.org/10.3892/ijo.2013.1971 PMID: 23727790

18. Ai J, Huang H, Lv X, Tang Z, Chen M, Chen T, et al. FLNA and PGK1 are two potential markers for progression in hepatocellular carcinoma. Cell Physiol Biochem. 2011; 27: 207–216. https://doi.org/10.1159/000327946 PMID: 21471709

19. Daly EB, Wind T, Jiang XM, Sun L, Hogg, PJ. Secretion of phosphoglycerate kinase from tumour cells is controlled by oxygen-sensing hydroxylases. Biochim Biophys Acta. 2004; 1691:17–22. https://doi.org/10.1016/j.bbamcr.2003.11.004 PMID: 15053920

20. Duan Z, Lamendola DE, Yusuf RZ, Penson RT, Preffer FI, Seiden MV. Overexpression of human phosphoglycerate kinase 1 (PGK1) induces a multidrug resistance phenotype. Anticancer Res. 2002; 22:1933–1941. PMID: 12174867

21. Zieker D, Konigsrainer I, Tritschler I, Loffler M, Beckert S, Traub F, et al. Phosphoglycerate kinase 1 a promoting enzyme for peritoneal dissemination in gastric cancer. Int J Cancer. 2010; 126:1513–20. https://doi.org/10.1002/ijc.24835 PMID: 19688824

22. Hwang TL, Liang Y, Chien KY, Yu JS. Overexpression and elevated serum levels of phosphoglycerate kinase 1 in pancreatic ductal adenocarcinoma. Proteomics. 2006; 6:2259–2272. https://doi.org/10.1002/pmic.200500345 PMID: 16493704

23. Sun S, Liang X, Zhang X, Liu T, Shi Q, Song Y, et al. Phosphoglycerate kinase-1 is a predictor of poor survival and a novel prognostic biomarker of chemoresistance to paclitaxel treatment in breast cancer. J Cancer. 2015; 112:1332–1339.

24. Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, et al. COSMIC: somatic cancer genetics at high-resolution. Nucleic Acids Res. 2017; 45(D1):D777–D783. https://doi.org/10.1093/nar/gkw1121 PMID: 27899578

25. Karki R, Pandya D, Elston RC, Ferlini C. Defining "mutation" and "polymorphism" in the era of personal genomics. BMC Med Genomics. 2015; 8:37–43. https://doi.org/10.1186/s12920-015-0115-z PMID: 26173390

26. Pasquo A, Consalvi V, Knapp S, Alfano I, Ardini M, Stefanini S, et al. Structural stability of human protein tyrosine phosphatase rho catalytic domain: effect of point mutations. PLoS One. 2012; 7(2): e32555. https://doi.org/10.1371/journal.pone.0032555 PMID: 22389709

27. Lori C, Lantella A, Pasquo A, Alexander LT, Knapp S, Chiaraluce R, et al. Effect of single amino acid substitution observed in cancer on Pim-1 kinase thermodynamic stability and structure. PLoS One. 2013; 8(6):e64824. https://doi.org/10.1371/journal.pone.0064824 PMID: 23755147

28. Lori L, Pasquo A, Lori C, Petrosino M, Chiaraluce R, Tallant C, et al. Effect of BET missense mutations on bromodomain function, inhibitor binding and stability. PLoS One. 2016; 11(7):e0159180. https://doi.org/10.1371/journal.pone.0159180 eCollection 2016. PMID: 27403962

29. Petrosino M, Lori L, Pasquo A, Lori C, Consalvi V, Minicozzi V, et al., Single-Nucleotide Polymorphism of PPARγ, a Protein at the Crossroads of Physiological and Pathological Processes. Int J Mol Sci. 2017; 18(2). pii: E361. https://doi.org/10.3390/ijms18020361 PMID: 28208577

30. Casadio R, Vassura M, Tiwari S, Fariselli P, Martelli LP. Correlating disease-related mutations to their effect on protein stability: a large-scale analysis of the human proteome. Hum Mutat. 2011; 32:1161–1170. https://doi.org/10.1002/humu.21555 PMID: 21853506

31. Petukh M, Kucukkal TG, Alexov E. On human disease-causing amino acid variants: statistical study of sequence and structural patterns. Hum Mutat. 2015; 36:524–534. https://doi.org/10.1002/humu.22770 PMID: 25689729

32. Kucukkal TG, Petukh M, Li L, Alexov E. Structural and physico-chemical effects of disease and non-disease nsSNPs on proteins. Curr Opin Struct Biol. 2015; 32:18–24. https://doi.org/10.1016/j.sbi.2015.01.003 PMID: 25658850

33. Chiarelli LR, Morera SM, Bianchi P, Fermo E, Zanella A, Galizzi A, et al. Molecular insights on pathogenic effects of mutations causing phosphoglycerate kinase deficiency. PLoS One 2012; 7(2):e32065. https://doi.org/10.1371/journal.pone.0032065 PMID: 22348148

34. Gill SC, von Hippel PH. Calculation of protein extinction coefficients from amino acid sequence data. Anal Biochem. 1989; 182:319–326. PMID: 2610349

35. Benjwal S, Verma S, Rohm KH, Gursky O. Monitoring protein aggregation during thermal unfolding in circular dichroism experiments. Protein Sci. 2006; 15:635–639. https://doi.org/10.1110/ps.051917406 PMID: 16452626

36. Royer CA, Mann CJ, Matthews CR. Resolution of the fluorescence equilibrium unfolding profile of trp aporepressor using single tryptophan mutants. Protein Sci. 1993; 2:1844–1852. https://doi.org/10.1002/pro.5560021106 PMID: 8268795

37. Santoro MM, Bolen DW. Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. Biochemistry. 1988; 27:8063–8068. PMID: 3233195

38. Kabsch W. XDS. Acta Crystallogr D Biol Crystallogr. 2010; D66:125–132.

39. Evans PR and Murshudov GN. How good are my data and what is the resolution? Acta Cryst. 2013; D69, 1204–1214.

40. Vagin A, Teplyakov A. MOLREP: an automated program for molecular replacement. J. Appl. Cryst. 1997; 30: 1022–1025.

41. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr. 1997; 53(Pt 3):240–255. https://doi.org/10.1107/S0907444996012255 PMID: 15299926

42. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr. 2004; 60(Pt 12 Pt 1):2126–2132.

43. Sherman MA, Fairbrother WJ, Mas MT. Characterization of the structure and properties of the His62-fAla and Arg38fAla mutants of yeast phosphoglycerate kinase: an investigation of the catalytic and activatory sites by site-directed mutagenesis and NMR. Prot Sci. 1992; 1:752–760.

44. Choy N, Rausseens V, Narayanaswami V. Inter-molecular coiled-coil formation in human apolipoprotein E C-terminal domain. J Mol Biol. 2003; 334:527–539. PMID: 14623192

45. Kiss RS, Weers PM, Narayanaswami V, Cohen J, Kay CM, Ryan RO. Structure-guided protein engineering modulates helix bundle exchangeable apolipoprotein properties. J Biol Chem. 2003; 278:21952–21959. https://doi.org/10.1074/jbc.M302676200 PMID: 12684504

46. Myers JK, Pace CN, Scholtz JM. Denaturant m values and heat capacity changes: Relation to changes in accessible surface areas of protein unfolding. Protein Sci. 1995; 4:2138–2148. https://doi.org/10.1002/pro.5560041020 PMID: 8535251

47. Auton M, Holthauzen LM, Bolen DW. Anatomy of energetic changes accompanying urea-induced protein denaturation. Proc Natl Acad Sci USA. 2007; 104:15317–15322. https://doi.org/10.1073/pnas.0706251104 PMID: 17878304

48. Geierhaas CD, Nickson AA, Lindorff-Larsen K, Clarke J, Vendruscolo M. BPPred: A computational tool to predict biophysical quantities of proteins. Protein Sci. 2007; 16:125–134. https://doi.org/10.1110/ps.062383807 PMID: 17123959

49. Pavlova NN, Thompson CB. The emerging hallmarks of cancer metabolism. Cell Metab. 2016; 23:27–47. https://doi.org/10.1016/j.cmet.2015.12.006 PMID: 26771115

50. Lim SO, Li CW, Xia W, Lee HH, Chang SS, Shen J, et al. EGFR signaling enhances aerobic glycolysis in triple negative breast cancer cells to promote tumor growth and immune escape Cancer Res. 2016; 76:1284–1296. https://doi.org/10.1158/0008-5472.CAN-15-2478 PMID: 26759242

51. Yu X, Li S. Non-metabolic functions of glycolytic enzymes in tumorigenesis. Oncogene. 2017; 36: 2629–2636. https://doi.org/10.1038/onc.2016.410 PMID: 27797379

52. Lu Z, Hunter T. Metabolic Kinases Moonlighting as Protein Kinases. Trends Biochem Sci. 2018; pii: S0968-0004(18)30020-3. https://doi.org/10.1016/j.tibs.2018.01.006 [Epub ahead of print] PMID: 29463470

53. Butera D, Wind T, Lay AJ, Beck J, Castellino FJ, Hogg PJ. Characterization of a reduced form of plasma plasminogen as the precursor for angiostatin formation. J Biol Chem. 2014; 289:2992–3000. https://doi.org/10.1074/jbc.M113.539924 PMID: 24338014

Supporting Information

S1 Table. List of oligonucleotides used for site-directed mutagenesis

Mutant	Primer sequences (5' to 3')
R38M	**FW**ACAACCAG**ATG**ATTAAGGCTGCTGT **REV**ACAGCAGCCTTAAT**CAT**CTGGTTGT
R65W	**FW** CACCTAGGC**TGG**CCTGATGGTGT **REV** ACACCATCAGG**CCA**GCCTAGGTG
G166D	**FW** GATGCTTTT**GAC**ACTGCTCACAG **REV** CTGTGAGCAGT**GTC**AAAAGCATC
M189I	**FW** GCTGGTGGGTTTTTG**ATC**AAGAAGGAGCTG **REV** CAGCTCCTTCTT**GAT**CAAAAACCCACCAGC
A199V	**FW** C**TTT**GCAAAG**GTC**TTGGAGAGCCCA**G** **REV** CTGGGCTCTCCAA**GAC**CTTTGCAAAG
V216F	**FW** CGGAGCTAAA**TTT**GCAGACAAGATC **REV**GATCTTGTCTGC**AAA**TTTAGCTCCG
F241S	**FW** GGAATGGCT**TCT**ACCTTCCTTAAGG **REV** CCTTAAGGAAGGT**AGA**AGCCATTCC

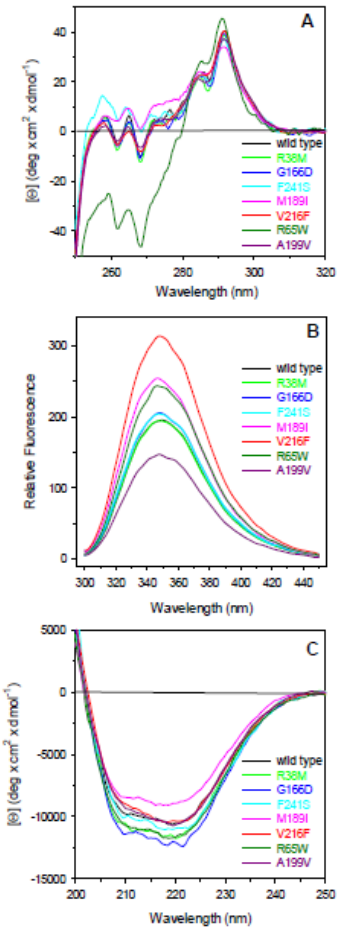
S2 Table. Crystallization conditions, data collection parameters, refinement statistics of the PGK1 variants.

	V216F	R38M	G166D	M189I (partially closed)	M189I (closed)
PDB codes	5M3U	5O7D	5M1R	5M6Z	5MXM
Resolution (Å)	1.81	1.84	1.64	1.67	2.05
Crystallization conditions	2.0 M Na/KPO_4, pH 8.4	Na citrate 1.6 M pH 6.5	Na citrate 1.6 M pH 6.5	2.5 M Na/KPO_4, pH 8.6	PEG 2K MME 30% w/v, bis-tris 0.1 M pH 7
X-ray source	ESRF-BM30-A3	ESRF ID23-2	ESRF ID23-2	ESRF ID23-2	BESSY 14.3
Wavelength (Å)	0.9677	0.8729	0.8729	0.8729	0.8943
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_12_12_1$
Unit cell parameters *a,b,c* (Å); β (°)	35.65, 106.03, 50.10; 98.71	36.05, 106.68, 50.69; 97.32	35.76, 105.57, 50.00; 97.76	35.88, 106.12, 50.47; 98.13	38.36, 91.10, 109.10; 90
No. of molecules in the asymmetric unit (modelled residue range)	1 (1-416)	1 (2-416)	1 (1-416)	1 (1-416)	1 (2-416)
Wilson B-factor ($Å^2$)	26.3	17.4	17.1	22.4	18.3
Data analysis range (highest resolution shell) (Å)	1.81-50 (1.81-1.92)	1.84–50 (1.84-1.87)	1.64-52.78 (1.64-1.74)	1.67-50 (1.67-1.77)	2.05-69.91 (2.05-2.17)
Unique reflections	32377	31685	44562	42601	24685
Completeness (%)	96 (79.1)	96.1 (99.5)	98.7 (97.3)	97.5 (94.1)	99.7 (99.3)
Redundancy	4.5 (3.8)	6.9 (6.9)	3.7 (3.7)	4.2 (4.2)	6.0 (5.9)
[a]Rmerge	0.08 (0.56)	0.17 (0.89)	0.097 (57)	0.054 (0.67)	0.14 (0.70)
CC(1/2) (%)	99.7 (75.3)	99.6 (36.7)	99.3 (68.8)	99 (59)	99.5 (88.3)
$\langle I/\sigma(I) \rangle$	11.22 (1.88)	10.5 (2.2)	7.84 (1.82)	15.08 (1.89)	13.15 (3.13)
Refinement ranges (highest resolution bin)	1.81-50 (1.81-1.85)	1.84-53.34 (1.84-1.89)	1.64-52.78 (1.64-1.66)	1.67-53.06 (1.67-1.71)	2.05-50 (2.05-2.10)
R_{cryst} (%)	19.4 (29.3)	22.8 (36.3)	16.6 (25.9)	18.5 (30.1)	17.9 (29.5)
R_{free} (%)	24.2 (38.6)	26.7 (36.6)	20.1 (30.5)	22.3 (33.2)	22.6 (39.5)
rms (angles) (°)	0.015	0.011	0.012	0.011	0.016
rms (bonds) (Å)	1.735	1.209	1.667	1.558	1.95
Residues in allowed region of Ramachandran plot / generously allowed (%)	100 / 2	100 / 3	100 / 2	100 / 1	100 / 2

Values in parentheses are for the highest-resolution shell.

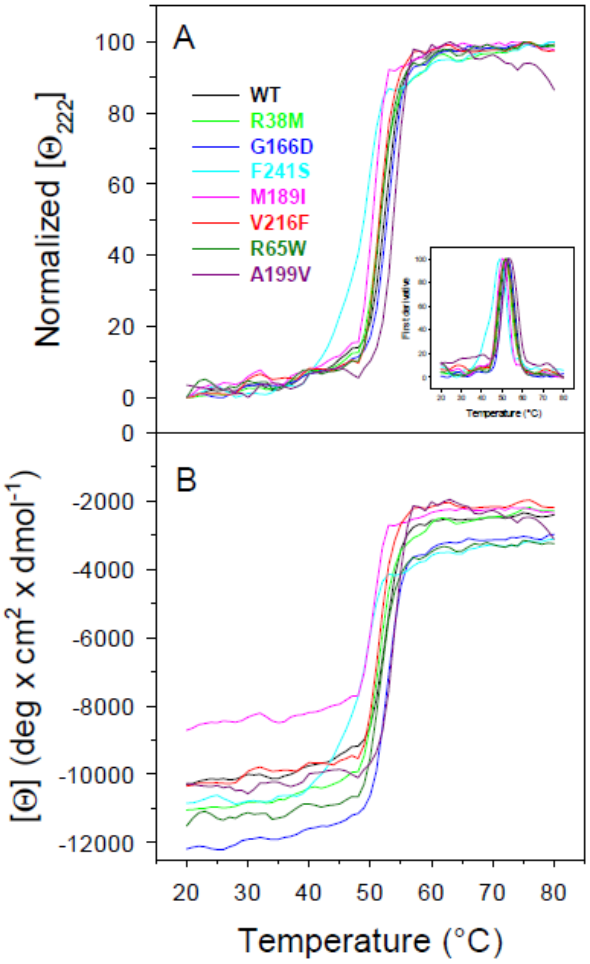
[a] $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the *i*th observation of the reflection (*hkl*) and $\langle I(hkl)\rangle$ is the mean intensity of the (*hkl*) reflection.

S1 Fig



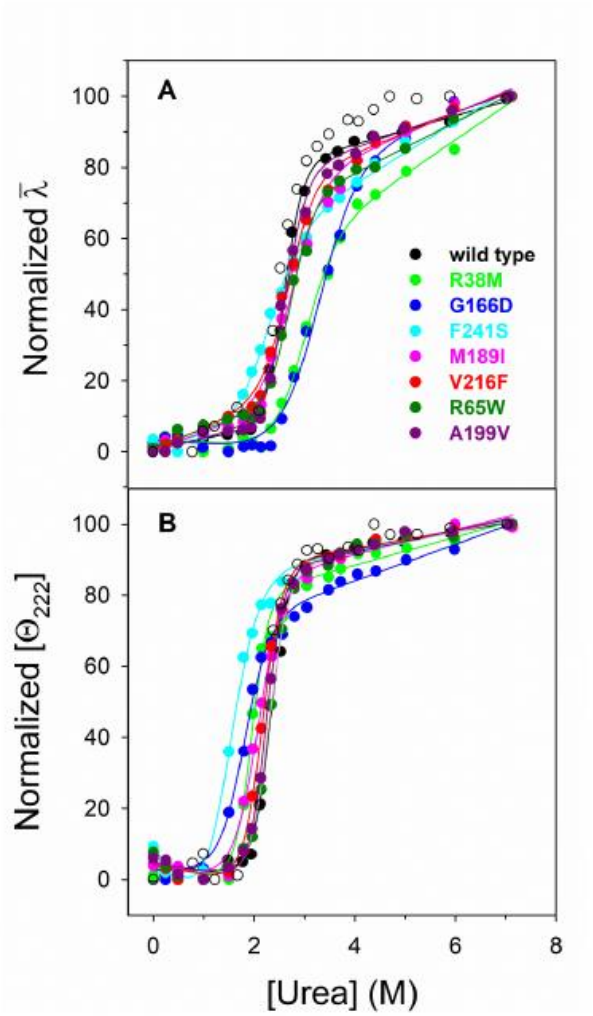
S1 Fig. Spectral properties of PGK1 wild type and variants. (A) Near-UV CD spectra were recorded in a 1.0-cm quartz cuvette at 1.5-1.7 mg/mL protein concentration in 20 mM Tris-HCl pH 8.0 containing 2.0 mM DTT, 1 mM EDTA and 200 mM NaCl. (B) Intrinsic fluorescence emission spectra were recorded at 110 μg/mL for R65W and 130 μg/mL for wild type and the other variants (0.08 AU_{280nm}, 295 nm excitation wavelength), in 20 mM Tris-HCl, pH 8.0 containing 0.2 M NaCl and 0.2 mM DTT. (C) Far-UV CD spectra were recorded in a 0.1-cm quartz cuvette at 130-170 μg/mL in 20 mM Tris-HCl, pH 7.5 containing 0.2 M NaCl and 0.2 mM DTT. All spectra were recorded at 20°C.

S2 Fig



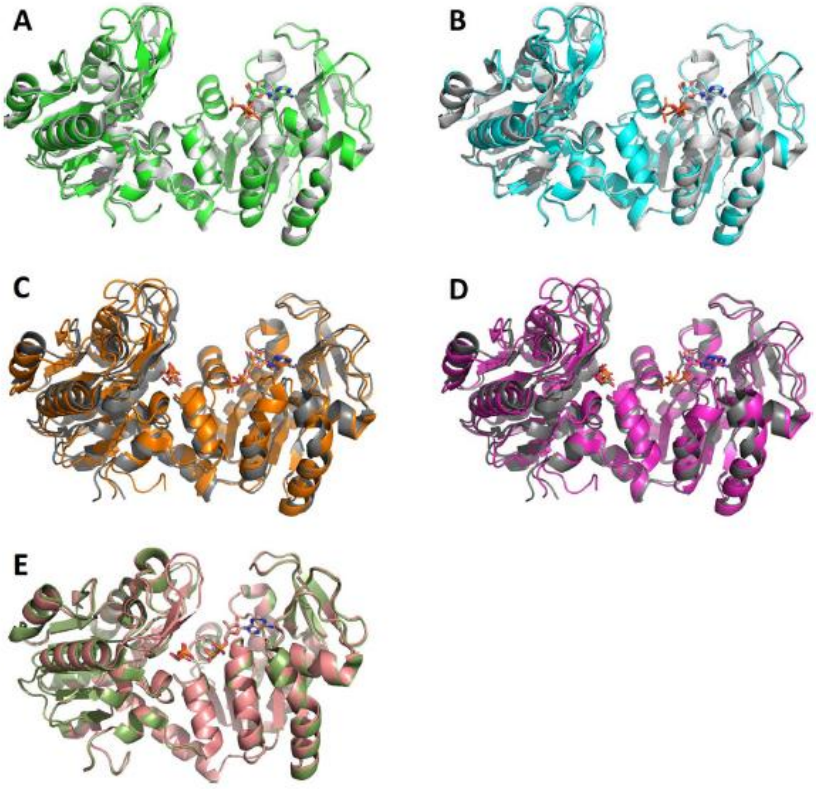
S2 Fig. Thermal unfolding of PGK1 wild type and variants. PGK1 variants and wild type (130-170 μg/mL) were heated from 20°C to 80°C in 20 mM Tris-HCl, pH 8.0 containing 0.2 M NaCl and 0.2 mM DTT . The molar ellipticity at 222 nm ([Θ_{222}]) was monitored continuously every 0.5 °C. (A) Normalized [Θ_{222}]; the inset shows the first derivative of the same data as in (A). (B) [Θ_{222}] before normalization.

S3 Fig



S3 Fig. Urea-induced equilibrium unfolding of PGK1 wild type and variants. (A) Normalized intensity-averaged emission wavelength ($\bar{\lambda}$). (B) Normalized molar ellipticity at 222 nm ([Θ_{222}]) reported after removal of the high-frequency noise and the low-frequency random error by SVD. The continuous lines represent the nonlinear fitting of the normalized $\bar{\lambda}$ and [Θ_{222}] data to Eq 4. The reversibility points (empty circles) are shown, for clarity, only for the wild type and were not included in the nonlinear regression analysis. All spectra were recorded at 20 °C, as described in Materials ad Methods.

S4 Fig



S4 Fig . Overall fold of PGK1 versions compared to the wild type. The structure of each variant has been superimposed with the most related structure of wild type, chosen according to the bound ligands and the conformation assumed. (A) R38M (green) *versus* 2ZGV (light grey). (B) G166D (cyan) versus 2ZGV (light grey). (C) V216F (orange) versus 2XE7 (dark grey). (D) M189I in partially open conformation (magenta) versus 2XE7 (dark grey). (E) M189I in closed conformation (pink) versus 2WZB (pale green). 2ZGV: wild type binding Mg-ADP; 2XE7: wild type binding Mg-ADP and 3-PG; 2WZB: wild type binding Mg-ADP, 3-PG and MgF_3^-. Mg-ADP and 3-PG are shown as sticks, $MgF_3^-$is shown as spheres.

Biochemistry and Biophysics Reports 11 (2017) 99–104

Contents lists available at ScienceDirect

Biochemistry and Biophysics Reports

journal homepage: www.elsevier.com/locate/bbrep

Unveiling the folding mechanism of the Bromodomains

Maria Petrosino[a,1], Daniela Bonetti[a,1], Alessandra Pasquo[b], Laura Lori[a], Roberta Chiaraluce[a], Valerio Consalvi[a], Carlo Travaglini-Allocatelli[a,*]

[a] Dipartimento di Scienze Biochimiche "A. Rossi Fanelli", Università di Roma "Sapienza", P.le A. Moro 5, 00185 Rome, Italy
[b] SSPT-BIOAG-BIOTEC ENEA Casaccia ENEA, Rome, Italy

ARTICLE INFO

Keywords:
Bromodomain
Protein folding
Folding intermediate

ABSTRACT

Bromodomains (BRDs) are small protein domains often present in large multidomain proteins involved in transcriptional regulation in eukaryotic cells. They currently represent valuable targets for the development of inhibitors of aberrant transcriptional processes in a variety of human diseases. Here we report urea-induced equilibrium unfolding experiments monitored by circular dichroism (CD) and fluorescence on two structurally similar BRDs: BRD2(2) and BRD4(1), showing that BRD4(1) is more stable than BRD2(2). Moreover, we report a description of their kinetic folding mechanism, as obtained by careful analysis of stopped-flow and temperature-jump data. The presence of a high energy intermediate for both proteins, suggested by the non-linear dependence of the folding rate on denaturant concentration in the millisec time regime, has been experimentally observed by temperature-jump experiments. Quantitative global analysis of all the rate constants obtained over a wide range of urea concentrations, allowed us to propose a common, three-state, folding mechanism for these two BRDs. Interestingly, the intermediate of BRD4(1) appears to be more stable and structurally native-like than that populated by BRD2(2). Our results underscore the role played by structural topology and sequence in determining and tuning the folding mechanism.

1. Introduction

Proteins involved in the regulation of histone post-translational modifications, such as those involved in acetylation, phosphorylation or methylation, play a pivotal role in the control of gene expression [1], therefore acting as "proof-editors" of the genetic code. Proteins acting on the histone acetylation processes can be grouped in at least three different subsets, on the basis of their specific function: "writer" proteins (as Histone AcetylTransferases, HATs) responsible for the addition of acetyl groups to specific lysine residues, "eraser" proteins removing specific acetyl groups (as Histone DeAcetylases, HDACs), and "reader" proteins, endowed with the ability to recognize and bind to specific histone acetylated lysines (AcK), such as the Bromodomains (BRDs) [2,3].

BRDs are conserved structural motifs of about 100 amino acids that are often present in large multidomain proteins involved in a variety of cellular processes, such as chromatin remodeling, post-translational modifications or transcriptional control [4]. Members of the BET (Bromo-Extra-Terminal domain) family (comprising human BRD2, BRD3, BRD4 and BRDT) display common modular architecture with two highly conserved amino-terminal BRDs, and a less conserved C-terminal recruitment domain. Mutations or chromosomal rearrangements affecting BRDs have been linked to various human diseases including cancer, and therefore BRDs are currently considered a promising target for the development of small-molecule inhibitors aiming at interfering with aberrant transcriptional processes in such diseases [5–7].

The structure of a variety of BRDs has been solved and shows a conserved left-handed helical bundle composed by four α-helices termed αZ, αA, αB and αC (from the N-terminal to the C-terminal helix) connected by loop regions of variable length (ZA and BC loops) (Fig. 1). From a structural point of view the BRDs therefore belong to the all-α fold class. Structural analyses of acetylated peptide-BRD complexes have shown that the AcK binding site is a hydrophobic cavity present on top of the helical bundle whose surface is shaped mainly by the ZA and BC loops. Not surprisingly, given the pivotal role played by BRDs in a variety of patho-physiological processes, a growing number of studies are currently focusing on their ligand binding affinity and specificity. However, little is known about the dynamic properties of these domains and, to our knowledge, no information is available about the mechanisms of folding of BRDs. This lack of information is somewhat surprising as it is known that for many proteins a relationship exists between

* Corresponding author.
E-mail address: carlo.travaglini@uniroma1.it (C. Travaglini-Allocatelli).
[1] M.P and D.B. contributed equally for this work.

http://dx.doi.org/10.1016/j.bbrep.2017.06.009
Received 29 March 2017; Received in revised form 23 June 2017; Accepted 28 June 2017
Available online 08 July 2017

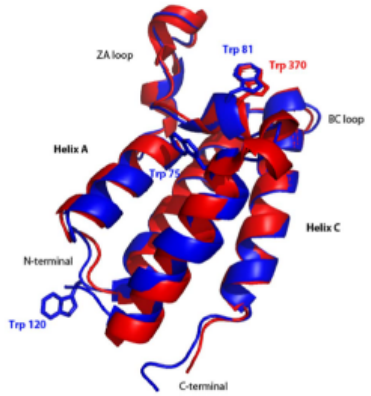


Fig. 1. Structural alignment of BRD2(2) (pdb id: 3oni), shown in red, and BRD4(1) (pdb id: 3uvx), shown in blue. The Trp residues are shown in sticks representations. BRD2(2) contains only one Trp (Trp370, structurally homologous to Trp81 of BRD4(1)); BRD4(1) presents two additional Trp residues (Trp75 and Trp120). The alignment was generated using PyMol (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC).

folding and ligand binding mechanisms, e.g. in the classical induced-fit model, in the case of other small protein domains involved in mediating protein-protein interactions [8] or in the case of intrinsically disordered proteins [9,10]. Interestingly, in the case of BRD domains, a ligand-induced conformational change has been proposed and discussed [11,12]. In light of these considerations we believe that obtaining information about the folding mechanism of the BRD domains may pave the way to a better understanding of their binding mechanism.

In this report we investigate the thermodynamic properties and the folding mechanism of two BET bromodomains: the second BRD of BRD2 (hereafter, BRD2(2)) and the first BRD of BRD4 (hereafter, BRD4(1)) by equilibrium spectroscopy and pre steady-state kinetic experiments. We decided to focus on these two BET BRDs because i) they are representative of the first and second domains generally found in the BET BRD family and ii) they represent ideal experimental system to investigate conservation (if any) of the folding mechanism among members of a fold family. Indeed these two BET BRDs display a 56% sequence similarity and, as can be seen from Fig. 1, they are structurally very similar (Cα root-mean square deviation (RMSD) is 1.2 ± 0.7 Å). It should be recalled that the folding mechanism of other all-α proteins has been studied in detail [13–16], leading to the hypothesis that formation of a folding intermediate is tuned by the specific α-helical propensities.

Quantitative analyses of stopped-flow (SF) mixing experiments and ultra-rapid temperature-jump (T-jump) data, allowed us to show that the folding mechanism of both BRDs are consistent with the presence of a folding intermediate, transiently populated in the sub-milliseconds time-regime. However, our results suggest that the two intermediate species show dissimilar thermodynamic and structural properties, highlighting different dynamic properties of these two BRDs.

2. Materials and methods

2.1. Protein expression and purification

BRD2(2) and BRD4(1) were expressed in *E.coli* and purified as previously described [17] and briefly reported in the legend to Fig. S1. [17]. Structural integrity of the purified proteins was checked by CD spectra in the far- and near-UV region (Figs. S2 and S3, respectively).

2.2. Urea-induced equilibrium unfolding

All experiments were carried out at 20 °C in 20 mM Tris/HCl, pH 7.5, 0.2 M NaCl, 200 μM DTT. Intrinsic fluorescence emission measurements were carried out with a LS50B spectrofluorimeter (Perkin-Elmer) using a 1.0 cm path length quartz cuvette. Fluorescence emission spectra were recorded from 300 to 450 nm (1 nm sampling interval), with the excitation wavelength set at 295 nm. Circular dichroism (CD) measurements were performed with a JASCO J-720 spectropolarimeter using a 0.2-cm cuvette. For urea-induced equilibrium unfolding, proteins (final concentration ranging over 50.0 −100 μg/mL) were incubated at 20 °C at increasing concentrations of urea (0 − 9.5 M). When equilibrium was reached, intrinsic fluorescence emission and far-UV CD spectra were recorded in parallel. To test the reversibility of the unfolding, BRD2(2) and BRD4(1) were denatured in 7.9 M urea at protein concentration ranging over 0.5–1.0 mg/mL. After 10 min, refolding was started by 15-fold dilution of the unfolding mixture into solutions of the same buffer used for unfolding containing decreasing urea concentrations. The final protein concentration ranged over 50.0 − 100 μg/mL. After 24 h, intrinsic fluorescence emission and far-UV CD spectra were recorded at 20 °C. All equilibrium unfolding experiments were performed in triplicate. The changes in intrinsic fluorescence emission spectra at increasing urea concentrations were quantified as the changes of the relative fluorescence intensity at 345 and at 350 nm for BRD2(2) and BRD4(1), respectively. The excitation wavelength used was 295 nm.

Urea-induced equilibrium unfolding transitions monitored by far-UV CD ellipticity and intrinsic fluorescence emission changes were analysed by fitting baseline and transition region data to a two-state linear extrapolation model [18] according to

$$\Delta G_{unf} = \Delta G^{H_2O} + m\,[\mathrm{Urea}] - RT\,\ln(K_{unf}) \quad (1)$$

where ΔG_{unf} is the free energy change for unfolding for a given denaturant concentration, ΔG^{H_2O} the free energy change for unfolding in the absence of denaturant and m a slope term which quantifies the change in ΔG_{unf} per unit concentration of denaturant, R the gas constant, T the temperature and K_{unf} the equilibrium constant for unfolding. The model expresses the signal as a function of denaturant concentration:

$$y_i = \frac{y_N + s_N[X]_i + (y_U + s_U[X]_i) * \exp\left[(-\Delta G^{H_2O} - m[X]_i)/RT\right]}{1 + \exp\left[\frac{-\Delta G^{H_2O} - m[X]_i}{RT}\right]} \quad (2)$$

where y_i is the observed signal, y_U and y_N are the baseline intercepts for unfolded and native protein, s_U and s_N are the baseline slopes for the unfolded and native protein, $[X]_i$ the denaturant concentration after the i^{th} addition, ΔG^{H_2O} the extrapolated free energy of unfolding in the absence of denaturant, m the slope in a $\Delta G_{unfolding}$ versus [X] plot.

The denaturant concentration at the midpoint of the transition, $[\mathrm{Urea}]_{0.5}$, according to Eq. (2), is calculated as:

$$[\mathrm{Urea}]_{0.5} = \Delta G^{H_2O}/m \quad (3)$$

2.3. Kinetic experiments

Stopped-flow kinetic folding experiments were carried out on a SX-17 stopped-flow instrument (Applied Photophysics, Leatherhead, UK) in Tris/HCl 50 mM buffer pH 7.5, 0.2 M NaCl, 2 mM DTT, at 20 °C; the excitation wavelength was 280 nm and the fluorescence emission was measured using a 320 nm cut-off glass filter. In all experiments, refolding and unfolding were initiated by a 11-fold dilution of the denatured or the native protein with the appropriate buffer. Usually 4–6 individual traces were accumulated and averaged. Final protein concentration was typically 5 μM.

Analysis was performed by non-linear least squares fitting of exponential phases using the fitting procedures provided in the Applied Photophysics software.

The relaxation kinetics was measured by using a Hi-Tech PTJ-64 capacitor-discharge T-jump apparatus (Hi-Tech, Salisbury, UK). Degassed and filtered samples were slowly pumped through the 0.5 × 2 mm quartz flow cell before data acquisition. Temperature was rapidly changed with a jump-size of 9 °C (from 11 °C to 20 °C). Usually 10–20 individual traces were accumulated and averaged. The excitation wavelength was 296 nm and the fluorescence emission was measured using a 320 nm cut-off glass filter. Protein concentration was typically 20 μM. The buffer used was Tris/HCl 50 mM buffer pH 7.5, 0.2 M NaCl, 200 μM DTT.

Experimental kinetic data were modeled on the basis of a three-state folding scheme, either with an on- or off-pathway intermediate, assuming that the logarithm of the microscopic rate constants linearly depends on the denaturant concentration: $\ln k_{ij} = \ln k_{ij}^{\circ} + m_{ij}(RT)^{-1}[\text{urea}]$, where k_{ij}° and m_{ij} represent the elementary rate constant in the absence of urea and the urea dependence of the rate constant (kinetic m value), 205 [19, 20].

For the two step reaction, global analysis of the two apparent rate constants λ_1 and λ_2 was performed by non-linear least-squares fitting of the kinetic data as previously described [19] using Graphpad Prism 5.04.

Kinetic ΔG values were calculated as follows: $\Delta G_{DN} = -RT \ln (k_{DI} \times k_{IN}) / (k_{ID} \times k_{NI})$; $\Delta G_{DI} = -RT \ln k_{DI} / k_{ID}$. Tanford β-values ($\beta_T$) were calculated as follows: $\beta_T(D) = 0$; $\beta_T(TS1) = -m_{DI}/m_{DN}$; $\beta_T(I) = (m_{ID} - m_{DI})/m_{DN}$; $\beta_T(TS2) = 1 - m_{NI}/m_{DN}$; $\beta_T(N) = 1$; m_{DN} is the kinetically derived 'equilibrium' m value, $m_{DN} = m_{ID} - m_{DI} + m_{NI} - m_{IN}$.

3. Results and discussion

3.1. Urea-induced equilibrium unfolding

The thermodynamic stability of BRD2(2) and BRD4(1) was determined at pH 7.5 and 20 °C by urea-induced equilibrium unfolding experiments, monitoring both the change of ellipticity at 222 nm by CD spectroscopy (Fig. 2A) and the change of intrinsic fluorescence emission (Fig. 2B). In all cases, the observed unfolding transitions were checked for reversibility. The same samples used to monitor the far-UV CD changes during the unfolding transition were used to monitor fluorescence emission changes, to allow a direct comparison of the data.

For both proteins, the denaturation curves obtained by monitoring the CD signal at 222 nm are satisfactorily fitted to a two-state model. The calculated unfolding free energy in water (ΔG^{H_2O}) for BRD2(2) is 8.83 ± 0.59 kcal mol^{-1} and the m value is 1.93 ± 0.13 kcal mol^{-1} M^{-1}, while for BRD4(1) the same thermodynamic parameters are $\Delta G^{H_2O} = 11.52 \pm 0.65$ kcal mol^{-1} and $m = 1.67 \pm 0.09$ kcal mol^{-1} M^{-1}, highlighting a larger stability for BRD4(1) compared to BRD2(2) ($\Delta\Delta G = 2.69$ kcal mol^{-1}) (Table 1).

In the case of BRD2(2), fitting the unfolding transition obtained by monitoring the changes of the fluorescence emission intensity to a two-state model allowed us to obtain thermodynamic parameters comparable to those obtained by CD ($\Delta G^{H_2O} = 9.09 \pm 0.68$ kcal mol^{-1}, $m = 1.81 \pm 0.13$ kcal mol^{-1} M^{-1}). However, in the case of BRD4(1), the unfolding transition obtained by monitoring the fluorescence changes could not be fitted to a two-state model because of a multiphasic profile (Fig. 2B). This result can be explained by the observation that BRD4(1), contrary to BRD2(2), contains more than one Trp residues, each monitoring the conformational properties of different regions of the protein (Fig. 1). Indeed, the different molecular environment of the fluorophores in the two BRDs is also mirrored by the different fluorescence emission spectra of their relative native states shown in Fig. S4.

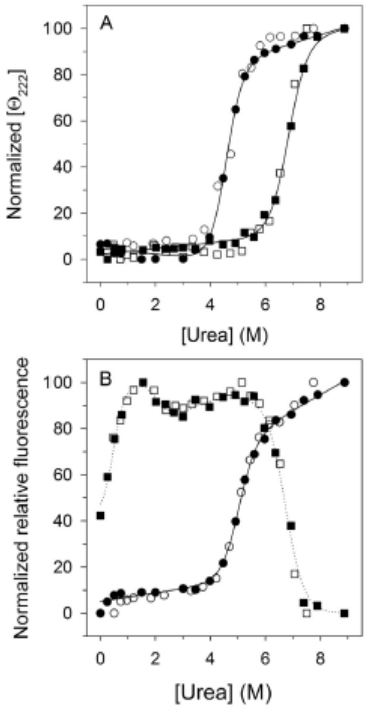


Fig. 2. Urea-induced equilibrium unfolding of BRD2(2) and BRD4(1). (A) Normalized molar ellipticity at 222 nm ($[\Theta]_{222}$) of BRD2(2) (circles) and BRD4(1) (squares) reported after removal of the high-frequency noise and the low-frequency random error by SVD. Continuous lines are the nonlinear regression to Eq. 2 of the data at different denaturant concentrations, as described in Section 2. (B) Normalized relative intrinsic fluorescence changes of BRD2(2) (circles) and BRD4(1) (squares). Continuous lines are the nonlinear regression of the data at different denaturant concentrations fitted according to Eq. 2 for BRD2(2), as described in Section 2. The dotted line interpolating the relative fluorescence intensity for BRD4(1) (squares) is drawn as a guide to the eye. The reversibility points (empty symbols) were not included in the nonlinear regression analysis.

Table 1
Thermodynamic parameters for urea-induced unfolding equilibrium of BRD2 (2) and BRD4 (1) measured by far-UV CD and fluorescence spectroscopy.

	BRD2(2)		BRD4(1)
	CD ($[\Theta]_{222}$)	Fluorescence	CD ($[\Theta]_{222}$)
$\Delta G_{H_2O}^{NU}$ (kcal/mol)	8.83 ± 0.59	9.09 ± 0.68	11.52 ± 0.65
m (kcal/mol/M)	1.93 ± 0.13	1.81 ± 0.13	1.67 ± 0.09
$[\text{Urea}]_{0.5}$ (M)	4.57	5.02	6.90

Urea-induced unfolding equilibrium data were obtained as described in Materials and methods. Data are reported as the mean ± SE of the fit. For BRD4(1) fluorescence changes could not be fitted to a two-state model because of a multiphasic profile (Fig. 2 B).

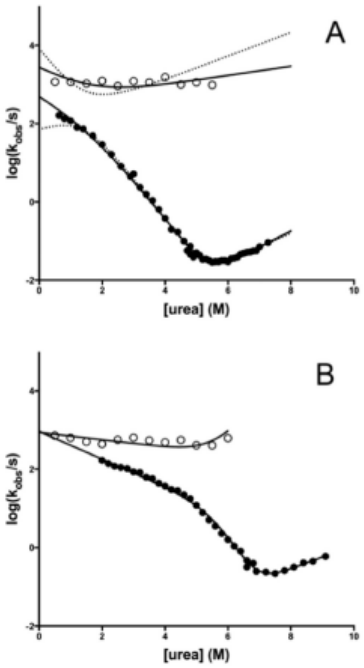


Fig. 3. Semilogarithmic plot (chevron plot) of the observed folding and unfolding rate constants as a function of urea concentration by SF (closed symbols) and T-jump (open symbols) experiments for BRD2(2) (panel A) and BRD4(1) (panel B). In panel A, the lines are the best fit to a three-state model involving on- (continuous line) or off-pathway (dashed line) intermediates [19]. BRD2(2) (Panel A): the calculated parameters for the on-pathway model are are: k_{IN} = 2357 ± 831; m_{IN} = 0.60 ± 0.33; k_{ID} = 329 ± 163; m_{ID} = 0.16 ± 0.06; k_{DN} = 570 ± 211; m_{DN} = 0.58 ± 0.34; k_{DI} = 0.00010 ± 0.00008; m_{DI} = 0.53 ± 0.07. BRD4(1) (Panel B): the calculated parameters for the on-pathway model (continuous line) are are: k_{DI} = 924 ± 130; m_{DI} = 0.47 ± 0.02; k_{ID} = 0.007 ± 0.010; m_{ID} = 1.12 ± 0.15; k_{DN} = 875 ± 115; m_{DN} = 0.12 ± 0.02; k_{NI} = 0.001 ± 0.001; m_{NI} = 0.40 ± 0.08. Rate constants are in s^{-1} and the associated m values are in kcal $mol^{-1} M^{-1}$. In the case of BRD4(1) the fit to an off-pathway model does not converge.

3.2. Kinetic folding-unfolding experiments

To gain information on the folding mechanism of these two BRDs, we carried out kinetic folding experiments at pH 7.5, 20 °C by fluorescence-monitored SF and T-jump experiments. The unfolding time courses obtained by rapid-mixing SF experiments for BRD2(2) and BRD4(1) were satisfactorily fitted to a single exponential decay at any final denaturant concentration (see Fig. S5A and S5B for representative unfolding time courses), while the refolding reaction was characterized by two processes having relaxation constants in different time regime (at [urea] around 1 M, k_1 ~100 s^{-1} and k_2 ~1 s^{-1}) (see Fig. S5C and S5D for representative refolding time courses). Since the slower refolding phase is characterized by a smaller amplitude (less than 10% of the faster phase) and is largely independent on denaturant concentration, it probably originates from *cis-trans* proline isomerization, as often observed in the folding of other proteins [21]

Fig. 3 shows the semi-logarithmic plot of the observed folding/unfolding rate constants (excluding the slow proline isomerization folding phase) *versus* denaturant concentration (chevron plot) obtained for BRD2(2) (Fig. 3A) and BRD4(1) (Fig. 3B) from SF (closed symbols) and T-jump (open symbols) experiments (see below). It should be noticed that, whereas the logarithm of the observed unfolding rate constant increases linearly with increasing denaturant concentration, both BRDs show a deviation from linearity of the observed refolding rate constants obtained by SF at low urea concentrations (roll-over effect). This non-linear dependence of the folding rate constants evidenced in chevron plots is generally interpreted as reflecting the accumulation of a folding intermediate [19,22], even if different explanations have been proposed and discussed [23].

We therefore hypothesized that an additional, and even faster folding phase undetectable by SF experiments because of the dead time of the mixing apparatus (2–3 ms), was to be observed under refolding conditions in the sub-ms time regime. Indeed, this hypothesis was experimentally confirmed by T-jump folding experiments carried out under matching temperature conditions (T = 20 °C) at different urea concentrations. The time courses obtained under folding and unfolding conditions were always satisfactorily fitted to a single exponential decay at any denaturant concentration (representative kinetic traces are reported in Fig. S6). The results of these experiments allowed us to measure not only the rate constant relative to the rapid formation of the intermediate, but also the kinetics of its unfolding at higher urea concentration. The data reported in Fig. 3 therefore represent the urea dependence of the main folding phases for BRD2(2) (panel A) and BRD4(1) (panel B): both proteins show a slower process observed by SF experiments and relative to the (un)folding of the native state, and a faster process, observed only by T-jump experiments, describing the rapid folding and unfolding of an intermediate species. However, as thoroughly discussed [19], identification of a transient folding intermediate is not sufficient to assign its kinetic role in a three state mechanism. In particular, an intermediate state may represent an obligatory species toward the formation of the native state ($D \rightleftharpoons I \rightleftharpoons N$; on-pathway intermediate), or a kinetic trap along the folding pathway ($I \rightleftharpoons D \rightleftharpoons N$; off-pathway intermediate). A clear distinction between these two alternative scenarios can be obtained if the urea dependence of all the four rate constants relative to the (un)folding of the native and intermediate species (k_{DI}, k_{ID}, k_{IN} and k_{NI}) are measured over a wide range of urea concentration. Following an approach generally used to analyze the folding mechanism of different proteins [13,19,24,25] we globally fitted the SF and T-jump data reported in the chevron plots to a three-state scheme, involving either an on- and an off-pathway intermediate (Fig. 3A/B). The four microscopic rate constants and the corresponding m-values obtained by this procedure allowed us to assign to the folding intermediate of BRD2(2) and BRD4(1) the role of an on-pathway species along the pathway of folding to the native state, since the off-pathway model fails to adequately fit the data in both cases. It is interesting to notice that an obligatory transient folding intermediate has been described also for the folding mechanism of other, non-related, four helical bundle proteins [13–15,26].

3.3. BRDs folding mechanism

Using the parameters obtained by the fit to the on-pathway model (see legend to Fig. 3), we also determined the populations of the three species at equilibrium as a function of urea concentration. Such an analysis, predicts that the intermediate species identified in the folding mechanism of both BRD domains is only negligible populated at equilibrium (< 1%) at urea concentration corresponding to the midpoint of the folding transition (e.g. [urea] = 5.2 M and 6.4 M for BRD2(2) and BRD4(1), respectively) (data not shown). The kinetic parameters obtained from the on-pathway model were also used to calculate the energy profiles, highlighting the position of the intermediate and transition states along a reaction coordinate. As shown in Fig. 4, both BRDs

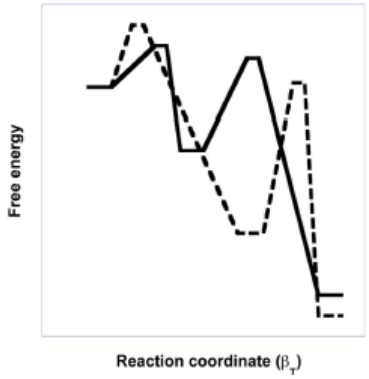


Fig. 4. Schematic energy diagrams for the folding of BRD2(2) (continuous line) and BRD4(1) (dashed line). The free energy values and the β_T values, were calculated from the kinetic parameters obtained by fitting the chevron plots of the two BRD domains to an on pathway intermediate (see the legend to Fig. 3). Free energy values were calculated with the Arrhenius equation using a pre-exponential factor of $4.8 \times 10^8\ s^{-1}$ [15]. BRD2(2): $\Delta G_{DI} = 1.14$ kcal mol^{-1}; $\Delta G_{DN} = 10.16$ kcal mol^{-1}; $\beta_T(TS1) = 0.32$; $\beta_T(I) = 0.40$; $\beta_T(TS2) = 0.71$. BRD4(1): $\Delta G_{DI} = 6.83$ kcal mol^{-1}; $\Delta G_{DN} = 14.80$ kcal mol^{-1}; $\beta_T(TS1) = 0.22$; $\beta_T(I) = 0.75$; $\beta_T(TS2) = 0.80$.

fold along a minimal three-state mechanism involving a denatured-like and a native-like transition states (TS1 and TS2, respectively), and on-pathway intermediate in between. The calculated Tanford β-values (β_T) for TS1, reflecting the buried surface area relative to the unfolded and native states, is $\beta_T = 0.32$ for BRD2(2) and $\beta_T = 0.22$ for BRD4(1), whereas the β_T for TS2 is 0.71 for BRD2(2) and 0.81 for BRD4(1). These results suggest that the two transition states of the two proteins display similar structural properties. On the contrary, the thermodynamic and conformational properties of the intermediate species appear to be largely different in the two BRDs; in fact, the intermediate populated by BRD4(1) is more stable relative to that of BRD2(2) ($\Delta\Delta G_{DI} = 5.7 \pm 0.8$ kcal mol^{-1}) and is clearly more native-like, as judged by its β_T value relative to the β_T value calculated for BRD2(2) ($\beta_T = 0.7$ and $\beta_T = 0.4$, respectively). The larger stability of the BRD4(1) intermediate relative to that of the BRD2(2) mirrors the greater stability of its native state ($\Delta\Delta G_{DN} = 4.3 \pm 1.3$ kcal mol^{-1} from kinetic data), suggesting that this intermediate is stabilized by native-like interactions. Moreover, inspection of Fig. 4 shows that the stabilities of the two transition states TS1 and TS2 and the intermediate in between in the folding pathway are correlated, as has been observed analyzing different sets of structurally unrelated proteins following a three-state folding mechanism [19,27]. Altogether, these observations suggest that formation of an intermediate species is a common property of the folding of these two BRDs; nevertheless, specific sequence features appear to be crucial to modulate its stability and conformational properties.

Together with native state topology, inherent propensity to form secondary structural elements is a key factor in modulating the folding process of proteins [28] and its contribution can be clearly evidenced studying the folding mechanism of homologous proteins sharing the same topological properties. In particular, it has been shown that, as the propensity for forming secondary structure increases, the folding mechanism of proteins of the same fold family slides from two-state to multi-state mechanisms [29,30]. In the case of BRD2(2) and BRD4(1), a comparison of helical propensities of sequence elements is meaningful, given the high structural homology between these two proteins. Fig. 5 shows that the distribution of helical propensity (as calculated by

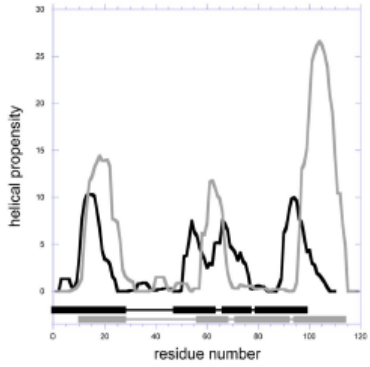


Fig. 5. Helical propensity of BRD2(2) (black line) and BRD4(1) (grey line). Helical propensity was calculated by the AGADIR program [31] (http://agadir.crg.es) Boxes represent the four helices, as determined from the X-ray structure of BRD2(2) (pdb id: 3oni; black boxes) and BRD4(1) (pdb id: 3uvz; grey boxes).

AGADIR [31],) between the two BRDs is very similar but that the overall helical propensity is higher in the case of BRD4(1), particularly at the level of the αC C-terminal helix. Phi-value analysis is in progress and it will allow us to describe the structure of the intermediate and transition states at the residue level and to check whether the C-terminal α – helix is part of the folding nucleus in these two BRDs.

4. Conclusions

A fruitful strategy for extracting general rules governing the folding of proteins is the comparison of the folding mechanism of proteins belonging to the same structural family. In this context, the bromodomains (BRDs) offer an as yet unexplored experimental system to study the folding mechanism of proteins belonging to the all-α fold class. In this paper we have reported the thermodynamic characterization at equilibrium and the folding kinetics of BRD2(2) and BRD4(1), two representative domains of domain 1 and 2 of the BET BRD family. The results reported in this work allow us to propose that formation of an obligatory intermediate may be a general feature of the folding landscape explored by these small all-α domains and add general significance to the hypothesis that members of a given fold family share a consensus folding mechanism. Our observations, however, highlight that the intermediate species populated during the folding of BRD4(1) is closer to the native state along the reaction coordinate than that populated by the less stable BRD2(2). This conclusion, which will be corroborated by structural analysis of folding intermediate(s) and transition states, underscores the role played by the sequence in modulating the stability of transiently populated states.

Acknowledgements

We thank Prof. Stefan Knapp (Institute for Pharmaceutical Chemistry and Buchmann Institute for Life Sciences (BMLS), Johann Wolfgang Goethe-University, Frankfurt am Main, Germany) for kindly providing us with the plasmids harboring the BRDs wild type genes. Work partially supported by grant C26A14AXLE – Sapienza Università di Roma to C.T.A. and by grant Regione Lazio (Prot. FILAS-RU-2014-1020) to V.C.

Appendix A. Transparency document

Transparency document associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.bbrep.2017.06.009.

Appendix B. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.bbrep.2017.06.009.

References

[1] T. Kouzarides, Cell 128 (2007) 693–705.
[2] R. Chandrasekaran, M. Thompson, Biochem. Biophys. Res. Commun. 355 (2007) 661–666.
[3] F. Gong, L.Y. Chiu, K.M. Miller, PLoS Genet. 12 (2016) e1006272.
[4] M.Y. Lubula, B.E. Eckenroth, S. Carlson, A. Poplawski, M. Chruszcz, K.C. Glass, FEBS Lett. 588 (2014) 3844–3854.
[5] P. Filippakopoulos, J. Qi, S. Picaud, Y. Shen, W.B. Smith, O. Fedorov, E.M. Morse, T. Keates, T.T. Hickman, I. Felletar, M. Philpott, S. Munro, M.R. McKeown, Y. Wang, A.L. Christie, N. West, M.J. Cameron, B. Schwartz, T.D. Heightman, N. La Thangue, C.A. French, O. Wiest, A.L. Kung, S. Knapp, J.E. Bradner, Nature 468 (2010) 1067–1073.
[6] P. Filippakopoulos, S. Picaud, M. Mangos, T. Keates, J.P. Lambert, D. Barsyte-Lovejoy, I. Felletar, R. Volkmer, S. Muller, T. Pawson, A.C. Gingras, C.H. Arrowsmith, S. Knapp, Cell 149 (2012) 214–231.
[7] L.L. Fu, M. Tian, X. Li, J.J. Li, J. Huang, L. Ouyang, Y. Zhang, B. Liu, Oncotarget 6 (2015) 5501–5516.
[8] C.N. Chi, S.R. Haq, S. Rinaldo, J. Dogan, F. Cutruzzola, A. Engstrom, S. Gianni, P. Lundstrom, P. Jemth, Biochemistry 51 (2012) 8971–8979.
[9] S. Gianni, J. Dogan, P. Jemth, Biophys. Chem. 189 (2014) 33–39.
[10] S.L. Shammas, M.D. Crabtree, L. Dahal, B.I. Wicky, J. Clarke, J. Biol. Chem. 291 (2016) 6689–6695.
[11] C. Kupitz, R. Chandrasekaran, M. Thompson, Biophys. Chem. 136 (2008) 7–12.
[12] S. Picaud, M. Strocchia, S. Terracciano, G. Lauro, J. Mender, D.L. Daniels, R. Riccio, G. Bifulco, I. Bruno, P. Filippakopoulos, J. Med. Chem. 58 (2015) 2718–2736.
[13] A. Borgia, D. Bonivento, C. Travaglini-Allocatelli, A. Di Matteo, M. Brunori, J. Biol. Chem. 281 (2006) 9331–9336.
[14] A.P. Capaldi, M.C.R. Shastry, C. Kleanthous, H. Roder, S.E. Radford, Nat. Struct. Biol. 8 (2001) 68–72.
[15] K. Tellum, K. Maki, B.B. Kragelund, F.M. Poulsen, H. Roder, Proc. Natl. Acad. Sci. USA 99 (2002) 9807–9812.
[16] Z. Zhou, Y. Huang, Y. Bai, J. Mol. Biol. 352 (2005) 757–764.
[17] L. Lori, A. Pasquo, C. Lori, M. Petrosino, R. Chiaraluce, C. Tallant, S. Knapp, V. Consalvi, PLos One 11 (2016).
[18] M.M. Santoro, D.W. Bolen, Biochemistry 27 (1988) 8063–8068.
[19] S. Gianni, Y. Ivarsson, P. Jemth, M. Brunori, C. Travaglini-Allocatelli, Biophys. Chem. 128 (2007) 105–113.
[20] M.J. Parker, J. Spencer, A.R. Clarke, J. Mol. Biol. 253 (1995) 771–786.
[21] T. Kiefhaber, F.X. Schmid, J. Mol. Biol. 224 (1992) 231–240.
[22] I.E. Sanchez, T. Kiefhaber, J. Mol. Biol. 325 (2003) 367–376.
[23] M. Oliveberg, Acc. Chem. Res. 31 (1998) 765–772.
[24] C. Travaglini-Allocatelli, S. Gianni, M. Brunori, Trends Biochem. Sci. 29 (2004) 535–541.
[25] C. Travaglini-Allocatelli, S. Gianni, V.K. Dubey, A. Borgia, A. Di Matteo, D. Bonivento, F. Cutruzzola, K.L. Bren, M. Brunori, J. Biol. Chem. 280 (2005) 25729–25734.
[26] D.M. Korzhnev, T.L. Religa, W. Banachewicz, A.R. Fersht, L.E. Kay, Science 329 (2010) 1312–1316.
[27] K. Kamagata, K. Kuwajima, J. Mol. Biol. 357 (2006) 1647–1654.
[28] A.A. Nickson, J. Clarke, Methods 52 (2010) 38–50.
[29] V. Daggett, A.R. Fersht, Trends Biochem. Sci. 28 (2003) 18–25.
[30] S. Gianni, N.R. Guydosh, F. Khan, T.D. Caldas, U. Mayor, G.W.N. White, M.L. DeMarco, V. Daggett, A.R. Fersht, Proc. Natl. Acad. Sci. USA 100 (2003) 13286–13291.
[31] V. Munoz, L. Serrano, Nat. Struct. Biol. 1 (1994) 399–409.

International Journal of
Molecular Sciences

MDPI

Article

Single-Nucleotide Polymorphism of PPARγ, a Protein at the Crossroads of Physiological and Pathological Processes

Maria Petrosino [1,†], Laura Lori [1,†], Alessandra Pasquo [2], Clorinda Lori [1], Valerio Consalvi [1], Velia Minicozzi [3], Silvia Morante [3], Antonio Laghezza [4], Alessandra Giorgi [1], Davide Capelli [5] and Roberta Chiaraluce [1,*]

1 Department of Biochemical Sciences "A. Rossi Fanelli", Sapienza University of Rome, P.le A. Moro, 5, 00185 Rome, Italy; maria.petrosino@uniromad1.it (M.P.); laura.lori@uniroma1.it (L.L.); clorinda.lori@uniroma1.it (C.L.); valerio.consalvi@uniroma1.it (V.C.); alessandra.giorgi@uniroma1.it (A.G.)
2 SSPT-BIOAG-BIOTEC ENEA Casaccia ENEA, 00123 Rome, Italy; alessandra.pasquo@enea.it
3 Department of Physics, University of Rome Tor Vergata and INFN, Via della Ricerca Scientifica 1, 00133 Roma, Italy; Velia.Minicozzi@roma2.infn.it (V.M.); silvia.morante@roma2.infn.it (S.M.)
4 Dipartimento di Farmacia-Scienze del Farmaco, University of Bari, 70126 Bari, Italy; antonio.laghezza@uniba.it
5 Istituto di Cristallografia, Consiglio Nazionale delle Ricerche, Via Salaria Km. 29, 300, Monterotondo Stazione, 00015 Roma, Italy; davide.capelli80@gmail.com
* Correspondence: roberta.chiaraluce@uniroma1.it; Tel.: +39-064-9910-956; Fax; +39-064-440-062
† These authors contributed equally to this work.

Academic Editors: Emil Alexov and Stephen A. Bustin
Received: 31 October 2016; Accepted: 1 February 2017; Published: 10 February 2017

Abstract: Genome polymorphisms are responsible for phenotypic differences between humans and for individual susceptibility to genetic diseases and therapeutic responses. Non-synonymous single-nucleotide polymorphisms (nsSNPs) lead to protein variants with a change in the amino acid sequence that may affect the structure and/or function of the protein and may be utilized as efficient structural and functional markers of association to complex diseases. This study is focused on nsSNP variants of the ligand binding domain of PPARγ a nuclear receptor in the superfamily of ligand inducible transcription factors that play an important role in regulating lipid metabolism and in several processes ranging from cellular differentiation and development to carcinogenesis. Here we selected nine nsSNPs variants of the PPARγ ligand binding domain, V290M, R357A, R397C, F360L, P467L, Q286P, R288H, E324K, and E460K, expressed in cancer tissues and/or associated with partial lipodystrophy and insulin resistance. The effects of a single amino acid change on the thermodynamic stability of PPARγ, its spectral properties, and molecular dynamics have been investigated. The nsSNPs PPARγ variants show alteration of dynamics and tertiary contacts that impair the correct reciprocal positioning of helices 3 and 12, crucially important for PPARγ functioning.

Keywords: PPARγ; molecular dynamics; protein stability; single-nucleotide polymorphism

1. Introduction

This study is focused on some natural variants of the Peroxisome Proliferator-Activated Receptor γ (PPARγ), a nuclear receptor that belongs to the superfamily of ligand inducible transcription factors, involved in several biological processes and in the maintenance of cellular homeostasis [1]. Nuclear receptors are multi-domain transcription factors that bind to DNA and regulate the expression of genes. PPARs (α, β/δ and γ) form heterodimers with retinoid X receptor (RXR) and, in the

presence of a ligand, adopt an active conformation. Gene regulation by these receptors is related to the ligand-dependent recruitment of coactivators, which is necessary to create a complex that binds to Peroxisome Proliferator Response Elements (PPRE) [2,3]. PPARγ is composed of different functional domains: two activation functional domains, AF-1 and AF-2, and a ligand binding domain (LBD) connected to a DNA binding domain (DBD) by a hinge region (Figure 1).

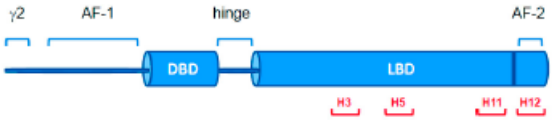


Figure 1. Schematic representation of nuclear receptor PPARγ. The ligand binding domain (LBD) is linked to the DNA binding domain (DBD) by a hinge. The residues involved in ligand binding are located in helix 3 (H3), helix 5 (H5), helix 11 (H11), and helix 12 (H12). Helices are numbered according to Nolte et al. [2]. PPARγ isoform 1 (UniProt ID P37231-2) is 28 residues shorter than PPARγ isoform 2 (UniProt ID P37231-1) at the N-terminus (γ2).

PPARγ, expressed in the adipose tissue [4,5], regulates adipocyte differentiation and insulin sensitization, playing a key role in the regulation of lipid metabolism in mature adipocytes and macrophages [6], with a direct impact on type 2 diabetes, dyslipidemia, atherosclerosis, and cardiovascular diseases [7,8]. In addition to the role in lipid metabolism, PPARγ has been reported to play a role in several processes related to cellular differentiation and development and to carcinogenesis [9]. Moreover, PPARγ has been implicated in inflammation [10] and is expressed in colon, breast, and prostate cancers [4,9,11,12]. As far as the role played by PPARγ in cancer, the association of loss of function variants with colon cancer [13] along with some evidence of inhibition of cell proliferation and induction of apoptosis suggest its potential anti-neoplastic effects [9]. The activation of PPARγ by agonist drugs [14] such as thiazolidinediones has been proposed as antineoplastic therapy [15]. However, it is not yet clear whether the use of PPARγ ligands as drugs could reduce the risk of cancer development [15].

Some rare missense mutations in PPARγ may cause profound phenotypic changes in affected individuals, contributing to the risk of dyslipidemia, type 2 diabetes [16], and colon cancer [8,13,17–23]. Indeed, a point mutation in the PPARγ ligand-binding domain (LBD) may alter structural interactions that are important for its stabilization, thus affecting ligand binding and the receptor transcriptional function. The molecular mechanism of most PPARγ mutations, related to lipodystrophy and insulin resistance, is not clear [23,24] and the structural reason for the decrease in functional activity of PPARγ variants has been identified in the case of F360L [25] and V290M [26]. These variants are nsSNPs, or missense variants, i.e., single-nucleotide variations occurring in the DNA coding region that lead to a polypeptide with a change in the amino acid sequence. The effect of nsSNPs has been related to changes in protein stability, protein–protein interactions, and protein functions [27,28]. Comparative analyses of phenotypically vs. thermodynamically characterized variations revealed that, on average, the variation types most involved in disease are also associated with a pronounced effect on protein stability [29–31]. However, the strength of this association is not sufficient to consider protein destabilization as the unique mechanistic cause explaining the onset of diseases [29], and the impact of nsSNPs on protein function can be unambiguously clarified only by thorough experimental analysis [29,32]. Computational studies predicted that around 30% of protein variants resulting from nsSNPs are less stable than the wild type [33]. Moreover, in silico studies have predicted the impact of nsSNPs on protein structure, stability, function, and interactions and have analyzed how these variations may affect disease susceptibility [34,35]. However, the experimental assessment of in vitro stability of common variants is required to determine the biophysical effects of mutations on protein structure and function [32,36,37].

PPARγ, at the crossroads of physiological and pathological processes such as metabolic control and adipogenesis, inflammation, apoptosis, and cancer, is particularly interesting for the study of the effects of nsSNPs on its structural stability, thermodynamic, and dynamic properties in solution. Several natural variants of PPARγ LBD, such as F360L, V290M, R357A, R397C, and P467L, have been associated with lipid metabolism disorders as well as cancer, e.g., Q286P, R288H [13]. More than 30 PPARγ natural variants are reported in COSMIC, a database designed to store and display somatic mutation information relating to human cancers [38,39].

In this study we selected PPARγ variants V290M, R357A, R397C, F360L, P467L, Q286P, R288H, E324K, and E460K, all located in the LBD, which are expressed in cancer tissues and/or associated with partial lipodystrophy and insulin resistance. In the selection of the variants, we focused on those mutations that were located on putatively critical positions in the structure and that may lead to alteration of the polarity of the residue, such as E324K, E460K, R357A, and R397C, or in the secondary structure propensity, as in the case of Q286P. In particular, the variants Q286P, R288H, V290M, E324K, E460K, and P467L, located in H3, H5, H11, and H12, are in close proximity of the residues involved in ligand binding (Figure 1 and S1).

We have investigated the effect of single amino acid substitution on the thermal and thermodynamic stability and the spectral properties of the above mentioned PPARγ variants by comparing experimental data with molecular dynamics (MD) simulations.

The alterations in protein stability and function may be driven by non-covalent interactions changes and modification of conformational dynamics of the variants. In most cases the stability of the expressed protein variants has been suggested to be responsible for the impact and/or consequences of the mutations on the pathological conditions or genetic susceptibility to diseases of the individuals. However, recent studies on natural protein variants in solution revealed that the perturbation of tertiary structure is not necessarily followed by changes in thermal and/or thermodynamic stability [40]. Indeed, the changes in side-chain flexibility of a mutated residue may lead to local variation in protein dynamics. Analysis of physico-chemical properties of natural variants may be helpful to reveal local structural changes that may not affect the overall folding of the structure, or may not be evident from the analysis of the variants crystal structure due to the conformational constraints the protein is subjected to in the crystal [40].

MD simulations are well suited to capture effects of point mutations on protein dynamics and detect any minor changes associated with an nsSNP. Detailed knowledge at the atomic level allows for an understanding of the structural and functional relationship upon mutation. In this study we use MD (in silico) analysis and (in vitro) thermodynamic studies to investigate the effect of nsSNPs of PPARγ natural variants.

2. Results

2.1. PPARγ Variants

In this study we focused on nine PPARγ variants (Q286P, R288H, V290M, E324K, R357A, F360L, R397C, E460K, and P467L) located in the LBD and associated with lipid metabolism disorders or to cancer (Figures 2 and S1). Four of these PPARγ variants, Q286P, R288H, E460K, and E324K have been found in cancer of the large intestine, lung, and endometrium, as reported in the COSMIC database (http://cancer.sanger.ac.uk/cosmic) [38]. The other five variants (V290M, R357A, F360L, P467L, and R397C) are related to alteration of metabolic control [23]. The variants Q286P, R288H, and V290M are located on helix 3 (H3), close to the ligand binding site, as is also the case with E324K, which is situated on helix 5 (H5). The variants R357A, E460K, and R397C are located in loops and F360L and P467L at the beginning of two small helices; the latter is the only variant in close proximity of one of the binding sites for the LXXLL helix of the coactivator [2]. The position of the mutated residues mapped onto the PPARγ structure is shown in Figures 2A and S1. The selected mutations encompass four surface exposed residues, R357A, R288H, E460K, and P467L, and five more buried

residues, Q286P, V290M, F360L, R397C, and E324K. Site-directed mutagenesis and available bacterial expression systems were used to produce recombinant proteins of the identified mutants [25] with the purpose of studying the consequences of the mutations on PPARγ spectral properties and thermal and thermodynamic stability. Introduction of these mutations resulted in soluble recombinant protein for Q286P, R288H, V290M, R357A, F360L, E460K, and P467L, whereas E324K and R397C could not be expressed in the soluble fraction even when different induction conditions were used. Mapping of these mutations onto the structure of PPARγ LBD revealed that E324 and R397 are both involved in one of the two salt bridges that play a pivotal role in the domain stabilization (Figure 2C,D).

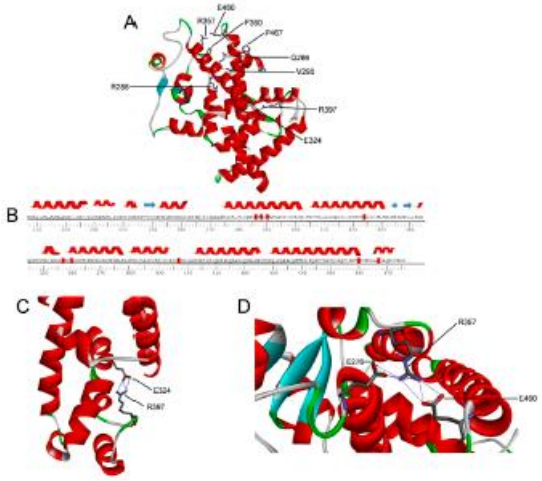


Figure 2. Amino acid sequence and structure of PPARγ LBD. (**A**) Structure of PPARγ LBD (PDB code: 1PRG) shown as a ribbon diagram; (**B**) secondary structural elements are shown at the top of the amino acid sequence. Mutated residues are highlighted in red; (**C**) Local environment of residues R397 and E324 involved in one salt bridge; (**D**) local environment of residue R357 engaging two salt bridges with residues E460 and E276.

2.2. Spectroscopic Characterization of PPARγ Wild Type and Variants

The near-UV circular dichroism (CD) spectrum of wild-type PPARγ, a protein lacking tryptophan residues, shows a strong positive contribution centered at around 263 nm, flanked by two positive shoulders at 270 nm and 258 nm, accompanied by fine structure features at 275–285 nm (Figure 3A). The near-UV CD spectra of F360L, P467L, and Q286P differ significantly from those of the wild type, either in intensity or in one of the positive shoulders that is blue-shifted to around 268 nm. In particular, the intensity of the near-UV CD spectrum of Q286P is significantly higher than that of the wild type and of F360L and P467L. V290M, R357A, R288H, and E460K display near-UV CD spectra closely similar to that of the wild type except for an overall decrease in the dichroic activity. Moreover, R357A and R288H show slight differences in the 270–280 nm region with respect to the wild type (Figure 3A).

The fluorescence emission spectra of the PPARγ wild type and variants are similar, but not identical. They all have a maximum emission wavelength around 308 nm, characteristic of tyrosine contribution (Figure 3B).

The far-UV CD spectra of PPARγ wild type and all the variants are typical of alpha helical proteins, showing local minima at around 208 and 222 nm and a zero intercept at around 200 nm. Interestingly, the wild type and variants show distinct contributions at 208 and 222 nm. The ratio of the molar ellipticity at 222 and at 208 nm ($[\Theta]_{222}/[\Theta]_{208}$) is 0.94 for the wild type and smaller for all the variants ranging from 0.86 for F360L, to 0.87 for R288H, 0.89 for R357A, 0.89 for E460K, 0.9 for P467L, 0.91 for Q296P, and 0.92 for V290M (Figure 3C). The 222/208 ellipticity ratio is indicative of interhelical contacts and has generally been used to distinguish between coiled coil helices and non-interacting helices (<0.9) [41,42]. The 222/208 ellipticity ratio below 0.9, observed for most of the variants, suggests different interhelical interactions and may indicate that the single amino acid substitutions induce significant changes of PPARγ structure in solution.

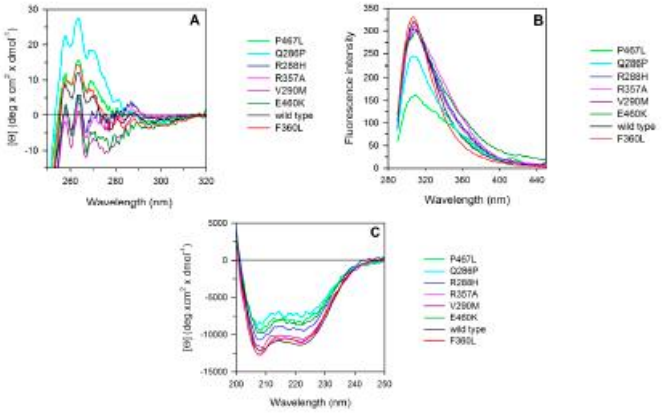


Figure 3. Spectroscopic properties of PPARγ wild type and variants. (**A**) Near-UV CD spectra were recorded in a 1 cm path-length quartz cuvette at 4.60 mg/mL protein concentration in 50 mM Tris–HCl pH 8.0 containing 0.20 M NaCl and 2.0 mM Dithiothreitol (DTT); (**B**) intrinsic fluorescence emission spectra were recorded at 0.1 mg/mL protein concentration (274 nm excitation wavelength) in 20 mM Tris–HCl pH 8.0 containing 0.1 M NaCl and 0.2 mM DTT; (**C**) far-UV CD spectra were recorded in a 0.1 cm path-length quartz cuvette at 0.2 mg/mL protein concentration in 20 mM Tris–HCl pH 8.0 containing 0.20 M NaCl and 0.2 mM DTT.

2.3. *Thermal Unfolding*

The thermal stability of PPARγ wild type and variants was investigated by continuously monitoring the ellipticity changes at 222 nm in the temperature range between 20 and 75 °C (Figure 4). The transition curves of PPARγ wild type and variants were compared by measuring the melting temperature (T_m) that corresponds to the midpoint of the denaturation process as calculated by plotting the first derivative of the molar ellipticity values as a function of temperature (Figure 4 inset). The temperature-induced ellipticity changes at 222 nm, where the main amplitude was observed, occur in an apparent cooperative transition with T_m values ranging from 50.0 to 44.0 °C (Table 1). A modest increase in T_m values is observed for the variants P467L and R288H; all the other variants show T_m values lower than that of the wild type (Table 1), with E460K showing a T_m value five degrees below that of the wild type. Notably, the differences in the amplitude observed for the thermal transitions of most of the variants (Figure 4B) may be attributed to the difference in the dichroic activity at 222 nm of their corresponding native states, as also indicated in the far-UV CD spectra reported

in Figure 3C. The ellipticity changes induced by temperature are paralleled by the increase of the photomultiplier tube voltage above 370 V (data not shown), suggesting that the protein aggregation follows temperature-induced unfolding. The observed transitions are irreversible, as indicated by the spectra measured at the end of the cooling phase that differ from those of the native proteins measured at the beginning of the thermal transitions. Furthermore, cuvette inspection at the end of the cooling phase revealed the presence of precipitate in all the samples.

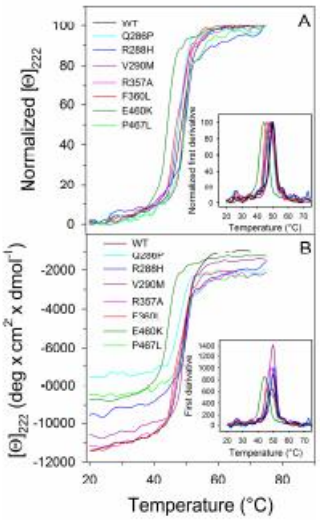


Figure 4. Thermal unfolding transition of PPARγ wild type and variants. Wild type and variants were heated from 20 to 75 °C in a 0.1 cm path-length quartz cuvette at 0.2 mg/mL protein in 20 mM Tris–HCl pH 8.0 containing 0.20 M NaCl and 0.2 mM DTT and the molar ellipticity at 222 nm ($[\Theta_{222}]$) was monitored continuously every 0.5 °C. (**A**) Normalized $[\Theta_{222}]$; (**B**) $[\Theta_{222}]$ before normalization. The insets show the first derivative of the same data as in (**A**,**B**).

Table 1. Melting temperatures and thermodynamic parameters for urea-induced unfolding equilibrium of PPARγ wild type and mutants measured by far-UV CD spectroscopy.

PPARγ	T_m (°C)	ΔG^{H_2O} (kcal/mol)	m (kcal/mol/M)	$[Urea]_{0.5}$ (M)
Wild type	49.5	3.37 ± 0.15	0.95 ± 0.05	3.16
Q286P	48.0	3.07 ± 0.15	0.84 ± 0.05	3.65
R288H	50.0	3.43 ± 0.12	0.94 ± 0.04	3.65
V290M	49.5	3.40 ± 0.10	0.89 ± 0.03	3.82
R357A	48.0	3.56 ± 0.10	1.00 ± 0.03	3.56
F360L	46.5	2.97 ± 0.10	0.83 ± 0.03	3.58
P467L	50.0	3.48 ± 0.11	0.93 ± 0.04	3.74
E460K	44.0	3.20 ± 0.16	1.07 ± 0.06	3.00

The temperature-induced changes were followed by monitoring the ellipticity at 222 nm. The T_m values were calculated by taking the first derivative of the ellipticity at 222 nm with respect to temperature. Urea-induced unfolding equilibrium data were measured at 10 °C in 20 mM Tris/HCl, pH 8.0, containing 0.2 M NaCl and 200 µM DTT by monitoring ellipticity at 222 nm $[\Theta_{222}]$. ΔG^{H_2O} and m values were obtained from Equation (3); $[Urea]_{0.5}$ was calculated from Equation (4). Data are reported as the mean ± SE of the fit.

2.4. Urea-Induced Equilibrium Unfolding Transitions

PPARγ wild type and variants reversibly unfold in urea at 10 °C in 20 mM TrisHCl, pH 8.0, containing 0.2 mM dithiothreitol (DTT) and 0.10 M NaCl. The effect of increasing urea concentrations (0–9 M) on the protein structure was analyzed by far-UV CD and fluorescence spectroscopy. Fluorescence and far-UV CD ellipticity changes during the unfolding transitions were monitored on the same samples. The ellipticity changes at 222 nm induced by urea show a sigmoidal dependence upon denaturant concentration, with an apparent two-state transition without any detectable intermediate (Figure 5A). The unfolding process is fully reversible upon dilution of the denaturant both for the wild type and variants with transition midpoints ranging from 3.82 (Table 1) to 3.00 M urea. The thermodynamic parameters relative to the apparent two-state equilibrium unfolding measured by far-UV CD have been fitted to a two-state model according to Equation (3) and do not indicate any significant difference between the variants and the wild type, except for F360L, which shows a less than 0.5 kcal/mol decrease of ΔG of unfolding (Table 1). Notably, the variant E460K also shows a 5.5 degree decrease in thermal stability, displaying a ΔG^{H_2O} value closely similar to that of the wild type. The values of m generally refer to the amount of protein surface area that becomes exposed to solvent upon unfolding [43]. Interestingly, all the m values determined for the PPARγ wild type and its variants, as measured by far-UV CD (Table 1), range between 0.83 and 1.07 kcal/mol/M, values four-fold lower than those predicted for a monomeric protein of 282 amino acid residues unfolded in urea [44,45]. Such low m values may be related to multi-state equilibrium unfolding, in line with the results obtained monitoring the unfolding process by intrinsic fluorescence (Figure 5B).

The fluorescence changes induced by increasing urea concentration for PPARγ wild type and all variants (recall that all these proteins lack tryptophan residues) are characterized by an increase in the fluorescence emission intensity and by a broadening of the emission spectra that remain centered at around 308 nm (Figure S2). These spectral changes, analyzed by monitoring the intensity averaged emission wavelength $\overline{\lambda}$, show a complex, non-two-state dependence upon increasing urea concentration for the wild type and for the variants Q286P, R288H, V290M, R357A, F360L, and P467L (Figure 5B). The data clearly indicate a three-state unfolding process and the population of a denaturation intermediate above 3.50 M (Figures 5B and S2), about the same urea concentration of the apparent denaturation midpoints observed by monitoring the ellipticity changes, with the exception of P467L, whose fluorescence intermediate becomes apparent above 2.0 M urea (Figures 5B and S2). The three-state transitions monitored by fluorescence, are not coincident with the two-state transitions monitored by far-UV CD and were fitted to a three-state unfolding process according to Equation (5), yielding the thermodynamic parameters reported in Table 2.

For the first transition, which represents the unfolding of the native to the intermediate state, $\Delta G^{H_2O}{}_{I\text{-}N}$ values of R288H and V290M are similar to those of the wild type, whereas those of P467L, F360L, R357A, and Q286P are significantly lower, suggesting a destabilization of the native state for these variants (Table 2). For the second transition, which represents the unfolding of the intermediate to the denatured state, the $\Delta G^{H_2O}{}_{U\text{-}I}$ values of R357A and F360L are higher than those of the wild type, suggesting that the intermediate state of the two variants is more stable. In the case of Q286P, R288H, V290M, and P467L, the $\Delta G^{H_2O}{}_{U\text{-}I}$ values are about half those of the wild type and suggest a less stable intermediate (Table 2). The differences in the values of ΔG^{H_2O}, from the native to the intermediate state and from the intermediate to the unfolded state, are mainly due to differences in m values, i.e., in the solvent exposed surface area upon unfolding. Notably, for the variants Q286P, R357A, F360L, and P467L, the m value from the native to the intermediate state is significantly lower than that of the wild type; in the transition from the intermediate to the unfolded state is observed a decrease of m value of Q286P, R288H, V290M, and P467L and an increase of m value of R357A and F360L. Taken together, these results indicate for all the variants a total ΔG^{H_2O} value, relative to the unfolding from the native to the denatured state, lower than that of the wild type and suggest a destabilization of the native state for Q286P, R357A, F360L, and P467L and a stabilization of the intermediate state of R357A and F360L. In the case of E460K, which shows a shallow unfolding transition, the changes of intensity

averaged emission wavelength $\overline{\lambda}$ at increasing urea concentration were fitted to a two-state unfolding process, according to Equation (3), yielding ΔG^{H_2O}, m and $[Urea]_{0.5}$ values of 1.81 ± 0.3 kcal/mol, 0.54 ± 0.08 kcal/mol/M and 3.37 M, respectively (Figure 5B, inset).

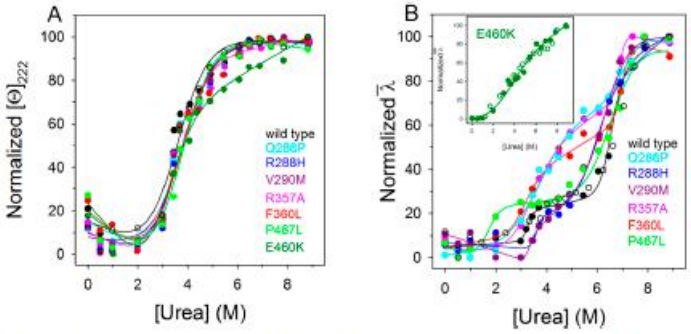


Figure 5. Urea-induced equilibrium unfolding of PPARγ wild type and variants. (**A**) Normalized molar ellipticity at 222 nm ($[\Theta]_{222}$) reported after removal of the high-frequency noise and the low-frequency random error by singular value decomposition algorithm (SVD). The continuous lines represent the nonlinear fitting of the normalized molar ellipticities at 222 nm to Equation (3); (**B**) Normalized intensity-averaged emission wavelength ($\overline{\lambda}$). The continuous lines represent the three-state fitting of the normalized intensity-averaged emission wavelength data to Equation (5). The inset in (**B**) shows the unfolding of E460K variant fitted according to Equation (3). The reversibility points (empty circles) are shown, for clarity, only for the wild type and for E460K and were not included in the nonlinear regression analysis. All the spectra were recorded at 10 °C, as described in Materials ad Methods.

Table 2. Thermodynamic parameters for urea-induced unfolding equilibrium of PPARγ wild type and mutants measured by fluorescence spectroscopy.

PPARγ	$m_{I\text{-}N}$ (kcal/mol/M)	$D50_{I\text{-}N}$ (M)	$\Delta G^{H_2O}{}_{I\text{-}N}$ (kcal/mol)	$m_{U\text{-}I}$ (kcal/mol/M)	$D50_{U\text{-}I}$ (M)	$\Delta G^{H_2O}{}_{U\text{-}I}$ (kcal/mol)
Wild type	5.30 ± 0.64	3.26 ± 0.09	17.27	2.47 ± 0.32	6.56 ± 0.04	16.20
Q286P	0.85 ± 0.10	3.73 ± 0.07	3.17	1.53 ± 0.18	7.00 ± 0.11	10.76
R288H	5.24 ± 0.52	3.57 ± 0.09	18.71	1.42 ± 0.17	5.89 ± 0.07	8.36
V290M	4.96 ± 0.64	3.32 ± 0.18	16.47	1.29 ± 0.14	6.41 ± 0.05	8.27
R357A	1.21 ± 0.16	3.31 ± 0.06	4.01	4.71 ± 0.50	6.88 ± 0.04	32.40
F360L	1.87 ± 0.22	3.09 ± 0.14	5.79	3.71 ± 0.48	6.86 ± 0.06	25.45
P467L	2.48 ± 0.24	1.66 ± 0.24	4.12	1.13 ± 0.16	6.58 ± 0.08	7.43

Urea-induced unfolding equilibrium data were obtained at 10 °C in 20 mM Tris/HCl, pH 8.0, containing 0.2 M NaCl and 200 μM DTT by measuring the fluorescence intensity averaged emission wavelength $\overline{\lambda}$. The free energy of unfolding from the native state to the intermediate ($\Delta G^{H_2O}{}_{I\text{-}N}$) and from the intermediate to the unfolded state ($\Delta G^{H_2O}{}_{U\text{-}I}$) were calculated from Equation (5). $D50_{I\text{-}N}$ and $m_{I\text{-}N}$ which are the midpoint and m value for the transition between native and intermediate state, respectively, and $D50_{U\text{-}I}$ and $m_{U\text{-}I}$ are the midpoint and m value for the transition between the intermediate and the unfolded state, respectively, were calculated from Equation (5). Data are reported ± SE of the fit.

2.5. Molecular Dynamics

MD simulations are invaluable in interpreting experimental data since they allow us to follow at the atom level the changes occurring in each of the mutant proteins. The basic data concerning the 10 (nine mutants plus the wild type) simulated systems are reported in Table 3. After 40 ns of equilibration, we followed the simulated systems for another 110 ns in the NVT *ensemble*. From the

collected configurations we computed the backbone root mean square deviations (r.m.s.d.) of each PPARγ variant with respect to the wild type starting structure (PDB ID: 1PRG) as a function of the simulation time. From this analysis we conclude that only the R357A and R397C r.m.s.d. are significantly different from the wild type r.m.s.d., meaning that these two variants are structurally the most distant ones from the PPARγ wild-type crystalline state. Moreover, a calculation of the gyration radius shows that R357A is the most compact system as it has the smallest gyration radius among all the mutants (Table 3), significantly smaller than that of the wild type. The other two interesting parameters that we found useful to monitor along the MD trajectory are the distances between H3 and helix 12 (H12) and between H12 and subportion 280–287 of PPARγ. In Table 3 (the last two columns) we report the mean value and the standard deviation of these two distances computed along the last 110 ns of the trajectory. In the Q286P variant, the H3–H12 distance is considerably smaller than in the wild type. In R288H and R357A we notice a large standard deviation due to the fact that H3–H12 distance oscillates. The F360L variant is the only one for which both the H3–H12 distance and the distance between H12 and the PPARγ 280–287 subportion are significantly larger (beyond errors) than in the wild type. This can be interpreted by saying that in the case of the F360L variant the strength of the inter-helical interactions is considerably reduced.

H3 appears to undergo secondary structural changes in four of the analyzed mutants, namely Q286P, R357A, F360L, and R397C. In Q286P, H3 assumes a 3-helix and turn secondary structure in the 277–287 region. In R357A it takes a coil structure in the 286–292 segment. In F360L it becomes turn and 3-helix in the 280–288 region, while in R397C its structure changes in a long segment 287–302 assuming a turn and a 3-helix secondary structure. In the wild type and in all the other variants, H3 stably remains in an α-helix structure. Compared to H3, the secondary structure of H12 is generally less stable. The reason is that H12 is at the C-terminal, hence it is located in a rather mobile region. H12 completely loses its α-helix secondary structure in favor of a turn structure only in R357A, F360L, and P467L.

Table 3. MD results for PPARγ wild type and the nine mutants.

System	Backbone r.m.s.d. (nm)	Gyration Radius (nm)	H3–H12 Distance (nm)	H12 Subportion (280–287) Distance (nm)
wild type	0.28 (0.02)	1.96 (0.01)	1.14 (0.05)	1.45 (0.05)
Q286P	0.26 (0.01)	1.96 (0.01)	1.00 (0.05)	1.42 (0.06)
R288H	0.26 (0.02)	1.95 (0.01)	1.2 (0.1)	1.46 (0.06)
V290M	0.28 (0.02)	1.95 (0.01)	1.15 (0.05)	1.40 (0.03)
E324K	0.28 (0.02)	1.94 (0.01)	1.05 (0.04)	1.39 (0.04)
R357A	0.39 (0.02)	1.92 (0.01)	1.3 (0.1)	1.5 (0.1)
F360L	0.28 (0.02)	1.98 (0.01)	1.44 (0.06)	1.60 (0.06)
R397C	0.38 (0.03)	1.98 (0.01)	1.26 (0.05)	1.45 (0.08
E460K	0.29 (0.02)	1.97 (0.01)	1.10 (0.06)	1.42 (0.05)
P467L	0.28 (0.02)	1.95 (0.02)	1.13 (0.06)	1.41 (0.05)

The r.m.s.d., the gyration radius, the H3–H12, and H12 subportion (280–287) distance mean values and (standard deviations) are computed on the last 110 ns of simulation.

The analysis of the Cα root mean square fluctuations (r.m.s.f.) per residue shows that the point mutations significantly alter the PPARγ mobility. In Figure 6 we show the r.m.s.f. of the three variants, E324K, R357A, and R397C, whose mobility is definitely higher than that of the wild-type protein. One notices that R397C is the PPARγ variant with the highest mobility and the largest number of involved residues.

By following the history of specific residues along the simulated trajectories, we can monitor the stability of some structurally important salt bridges. We have examined the history of the E259–R280, E324–R397, and E460–R357 salt-bridges along the MD history. The E259–R280 salt bridge is absent in the wild type, and is present only in the E324K, R357A, and R397C variants. The salt bridges E324–R397 and E460–R357 are always present and stable except in the variants where one of the amino

acids involved in the salt bridge is mutated (E324K and R397C for the first salt bridge; E460K and R357A for the second). It is worth noting that the R357–E460 distance is more stable and smaller in the Q286P variant (Figure S3A) than in all the other variants and in the wild type.

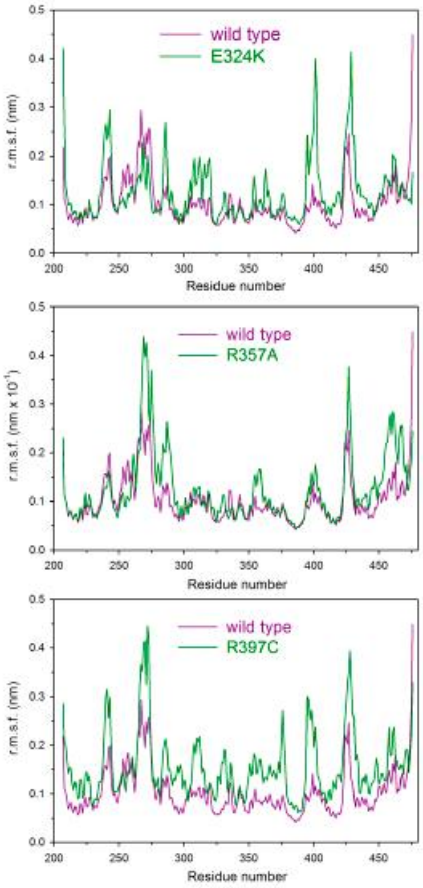


Figure 6. Mobility per residue of PPARγ wild type and E324K, R357A, and R397C variants. Cα root mean square fluctuations (r.m.s.f.) per residue for E324K, R357A, and R397C variants compared with the wild type r.m.s.f. On the *x*-axis is the residue number and on the *y*-axis is the mobility in nm.

The E276–R357 salt bridge is lost in the R357A variant (because of the point mutation), while it is quite stable in all the other cases, except the wild type and the E460K variant (Figure S3B). In Figure S3B we report the E276–R357 distance in the case of the E324K variant as an example of stability. Finally,

we monitored the distance between the R397 and R443 residues. We found that these two residues are always rather near except in the E324K and R397C variants. A possible explanation of such behavior is that the absence of the E324–R397 salt bridge in these two variants causes the residue R397 to move away from R443. The largest oscillations of R397–R443 distance are found in the Q286P variant even if the two residues remain closer than in E324K and R397C. In Figure S3C we report the time evolution of the R397–R443 distance in the case of the Q286P variant together with that of the wild type and the F360L variant. The latter is shown just to compare with a case where oscillations are small.

2.6. Transcription Activity

The transcription activity of F360L, R357A, P467L, and Q286P PPARγ variants was evaluated in comparison with wild-type PPARγ LBD in the presence of the full agonist rosiglitazone and LT175, a partial agonist that binds to a different region of PPARγ. For this purpose, GAL4–PPAR chimeric receptors were expressed in transiently transfected HepG2 cells according to a previously reported procedure [46]. As previously reported, the efficacy of both ligands remained basically unchanged towards F360L compared to the wild type, while the potency was significantly reduced [25]. A remarkable lowering in both efficacy and potency was shown for R357A and P467L (Figure S4 and Table S1). In particular, rosiglitazone displayed a remarkable lowering of potency; its EC_{50} value, in fact, was 7-fold higher against R357A and 18-fold higher against P467L compared to the wild type, whereas for LT175 this value turned out to be only about twice as high (Table S1). Singular behavior has been observed for the mutant Q286P, which was completely inactive and insensitive to both rosiglitazone and LT175 (Figure S5).

3. Discussion

In the post-genomic era, how human genetic and somatic variations are associated with diseases and how mechanisms form the basis of the relationship between genotype and phenotype are still open questions. Genetic polymorphism is responsible for phenotypic differences among humans and individual susceptibility to genetic disease and therapeutic responses. nsSNPs are of particular interest since the single-nucleotide variations occurring in the DNA coding region lead to a polypeptide with a change in the amino acid sequence that may affect the structure and/or function of the protein. The structural analysis of nsSNP protein variants may help in understanding the molecular basis of diseases and, since individuals carrying variants may respond differently to drugs, it may provide information for personalized drugs tailored to the individual variant. For complex diseases such as cancer and diabetes, SNPs may not function individually; rather, they work in coordination with other SNPs to manifest a disease condition. However, nsSNP variants may be utilized as efficient structural and functional markers of association with complex diseases.

Experimental [25,27,28,40] and computational [47–50] studies on several proteins related the effect of nsSNPs to the alteration of protein stability, protein–protein interactions, and protein functions. The interest in studying the effects of nsSNPs on structural stability and dynamic properties of PPARγ derives from the involvement of this nuclear receptor in a variety of biological processes such as adipocyte differentiation and insulin sensitization, as well as cellular differentiation and development and carcinogenesis [14]. Notably, PPARγ functions have been linked to several pathologies, ranging from metabolic disorders to cardiovascular disease, chronic inflammation, neurodegenerative disorders, and cancer [51,52]. PPARs ligands and other agents influencing PPAR signaling pathways have been shown to display chemopreventive potential by mediating tumor suppressive activities in a variety of human cancers and could represent novel targets to inhibit carcinogenesis and prevent tumor progression [53]. In addition, PPARγ agonists have recently been reported to lower the incidence of a number of neurological disorders [54]. All these functions are accomplished by binding PPARγ LBD to different ligands, which leads to conformational changes that promote the interaction with coactivator proteins in the nucleus [55].

PPARγ, a nuclear receptor that belongs to the ligand-dependent transcription factors, consists of a central DNA binding domain and a carboxy-terminal domain involved in ligand binding, dimerization, and transactivation. PPARγ adopts an active conformation that promotes transcription upon heterodimerization with RXR in the presence of a ligand. The ligand binding site is buried within the core of the LBD, which is folded into three layers of α-helices (Figures 2 and S1) [2].

Missense mutations in PPARγ LBD caused by nsSNPs may induce profound phenotypic changes in affected individuals, contributing to the risk of onset of various pathological states, like dyslipidemia, type 2 diabetes [16], and cancer [8,13,17–23]. The molecular mechanism that leads to the loss and/or alteration of PPARγ functions in nsSNP variants is not clear [23,24]. In this study we investigate the effect of the mutations on PPARγ LBD; to our knowledge, this is the first report that analyzes, in comparison with the wild type, nine PPARγ non-synonymous polymorphic variants of the LBD in terms of their spectroscopic properties in solution, thermodynamic and thermal stability, and molecular dynamics. The selection of the variants was focused on those mutations located in putatively critical positions, such as Q286P, R288H, V290M, E324K, E460K, and P467L, in close proximity to the residues involved in ligand binding (Figures 1 and S1). We also considered those non-conservative amino acid substitutions leading to alteration of the polarity of the residue, such as E324K, E460K, R357A, and R397C, or in the secondary structure propensity, as in the case of Q286P. All PPARγ variants were obtained as recombinant soluble proteins, with the exception of E324K and R397C, which could not be expressed in the soluble fraction even when different induction conditions were used. Interestingly, E324 and R397, located on H5 and on a loop, respectively, are both involved in one of the two salt bridges that may contribute to PPARγ stabilization (Figure 2C,D). The importance of the two salt bridges (Figure 2C,D) is also evident from the consequence of the mutation of the negatively charged E460 into a positively charged lysine at the end of H12, which breaks the salt bridge network formed by R357 and E276, both located on a loop. The importance of this salt bridge network, and of R357in particular, has already been described by the effect of its mutation into alanine on the global stabilization of the entire LBD [25]. As a matter of fact, E460K shows the lowest melting temperature, five degrees lower than that of the wild type, and a poorly cooperative urea-induced unfolding transition monitored by fluorescence changes, characterized by very low values of thermodynamic parameters (Figure 5B inset). Notably, the thermodynamic parameters, measured by monitoring the secondary structure changes by far-UV circular dichroism in the apparent two-state urea-induced unfolding transitions, are similar for all variants and only slightly different with respect to the wild type, with the exception of F360L, which shows ΔG^{H_2O}, m, and T_m values lower than those of the wild type (Table 1). These results suggest a similar overall secondary structure folding for all variants with respect to the wild type. On the other hand, the tertiary structure changes monitored by fluorescence reveal a complex non two-state process and significant differences among the natural variants. The analysis of the thermodynamic parameters obtained by fitting the fluorescence changes to a three-state unfolding reveals a decreased stability of the native state for all variants except for R288H and V290M. Interestingly, the variants P467L and Q286P show a destabilization of both the native and the intermediate state and are inactive. Both amino acid substitutions involve a proline residue and, in the case of Q286P, a residue located in the middle of H3; its functional relevance has been previously addressed in [56]. Tertiary structural variations between the wild type and variants are indicated by comparison of their near-UV CD spectra; in particular, amino acid substitutions in the variants F360L, P467L, and Q286P lead to changes in the overall protein tertiary arrangements, and minor tertiary changes are observed for all the other variants. Notably, all variants show a slight decrease in inter-helical interactions, as suggested by the decrease of 222/208 ellipticity ratio, more significant for F360L. These results, taken together, suggest a possible increase in the flexibility of the variants with respect to the wild type, as confirmed by molecular dynamics simulations. The most flexible variants (Figure 6) are E324K, R357A, and R397C, precisely the ones where the mutation affects a residue involved in one of the salt bridges that are supposed to contribute to the PPARγ LBD wild-type stabilization. Moreover, molecular dynamics simulations are able to confirm the presence

of small changes in the secondary structure of all the variants compared to the wild type and a more significant decrease of inter-helical interactions for the F360L variant (the last two columns of Table 3). The importance of inter-helical interactions and the correct reciprocal positioning of H3 and H12 has been previously reported as a crucial point for PPARγ function [57].

4. Materials and Methods

4.1. Plasmids and Site-Directed Mutagenesis

The LBD of PPARγ wild type (gene ID 5468, amino acids 174–477, expected molecular mass 34.5 kDa) and mutants were cloned in pET-28 plasmid for *Escherichia coli* expression as previously described [58]. The plasmid harboring the PPARγ wild-type gene was used to obtain mutant enzymes. The QuikChange Site-Directed Mutagenesis Kit (Stratagene, San Diego, CA, USA) was used to introduce the point mutations into the bacterial expression vector and into the vector expressing the chimeric receptor containing the yeast Gal4 DNA-binding domain fused to the wild-type PPARγ LBD used for the transcription activity assay [59]. The mutagenic synthetic oligonucleotides are shown in Table S2. Sequence analysis was performed to confirm the presence of the desired mutations and the absence of additional mutations.

4.2. Protein Preparation

PPARγ isoform 1 (UniProt ID P37231-2) wild type and mutants (Table S3) were expressed as N-terminally His-tagged proteins using a pET-28 vector and then purified as follows. *E. coli* KRX cells were transformed with the selected plasmid and were grown on an LB medium with 30 mg/mL kanamycin at 37 °C to an OD of 0.6. The culture was then induced with 5.0 mM rhamnose and further incubated at 18 °C for 20 h with vigorous shaking. Cells were collected by centrifugation and resuspended as a 20 mL culture in buffer A (20 mM Tris, 150 mM NaCl, 10% glycerol, 1 mM tris(2-carboxyethyl)phosphine–HCl (TCEP) pH 8.0) in the presence of protease inhibitors (Complete Mini EDTA-free; Roche Applied Science, Monza, Italy). The cells were sonicated and the soluble fraction was isolated by centrifugation (35,000× *g* for 45 min). The supernatant was applied to a Ni^{2+}–nitrilotriacetic acid column (GE Healthcare) and elution was performed with 0.25 M imidazole in buffer A. The pure fractions were concentrated to 2 mL using Millipore (Milano, Italy), concentrators and loaded onto a Superdex, 75 10/300, GE Healthcare (Milano, Italy), gel-filtration column on an ÄKTA FPLC system previously equilibrated with 50 mM Tris–HCl, 0.25 M NaCl, 2 mM DTT pH 8.0 at a flow rate of 1.0 mL/min. The eluates were collected and SDS–PAGE was used to test the purity of the protein. The proteins were identified by mass-spectrometric analysis. SDS–PAGE bands were cut from the gel and processed via tryptic proteolysis. The peptide mixtures were analyzed by a MALDI-ToF, AutoFlex II (Bruker Daltonics, Bremen, Germany) mass spectrometry instrument. Data were manually analyzed by a FlexAnalysis program (Bruker Daltonics) that revealed the expected site mutations according to a theoretical mass list of tryptic PPARγ peptides. The protein was then cleaved with thrombin protease (GE Healthcare (Milano, Italy); 10 U/mg) at room temperature for 2 h. The digested mixture was reloaded onto an Ni^{2+}–nitrilotriacetic acid column to remove the His tag and the undigested protein. The flowthrough was loaded onto a Q-Sepharose HP column (GE Healthcare) and eluted with a 0–500 mM gradient of NaCl in buffer B (20 mM Tris, 10% glycerol, 1 mM TCEP pH 8.0) with a BioLogic DuoFlow FPLC system (Bio-Rad Laboratories, Milano, Italy). Finally, the protein was purified by gel-filtration chromatography on a HiLoad Superdex 75 column (GE Healthcare) and eluted with buffer C (20 mM Tris, 1 mM TCEP, 0.5 mM EDTA pH 8.0). Protein quantification was determined according to OD_{280} measurement using the respective molar extinction coefficients ε of each protein, calculated according to [60].

4.3. Cell Culture and Transfections

Human hepatocellular liver carcinoma cell line HepG2 (Interlab Cell Line Collection, Genoa, Italy) was cultured in Minimum Essential Medium (MEM) containing 10% heat-inactivated fetal bovine serum, 100 U/mL of penicillin G, and 100 µg/mL of streptomycin sulfate at 37 °C in a humidified atmosphere of 5% CO_2 (250 ng). For transactivation assays, 1×10^5 cells per well were seeded in a 96-well plate and transfected after 24 h with K2 Transfection System (Biontex Laboratories GmbH, Munchen, Germany) according to the manufacturer's protocol using 0.20 µg/well of DNA. Cells were transfected with expression plasmids encoding the fusion protein Gal4–PPARγ–LBD (wild type, P467L, or Q286P mutant), pGal5TKpGL3, and pCMVβgal to normalize the transfection efficacy. Twenty-four hours after transfection, the medium was replaced with a fresh medium supplemented with rosiglitazone (ranging from 2 nM to 10 µM), LT175 (ranging from 100 nM to 10 µM) or DMSO 0.1%. After a further 24 h of incubation, cells were lysed and the luciferase activity in cell extracts was determined by a luminometer (VICTOR^3 V Multilabel Plate Reader, Perkin-Elmer, Monza, Italy) and normalized for β-galactosidase activity. Fold induction activity was calculated and plotted using GraphPad Prism 5.04 software (La Jolla, CA, USA). All transfection experiments were repeated at least twice with similar results. The results were expressed as mean ± SEM.

4.4. Spectroscopic Measurements

Intrinsic fluorescence emission spectra were recorded from 290 to 440 nm (274 nm excitation wavelength, 1 nm sampling interval), at 0.1 mg/mL protein concentration (3.25×10^{-2} AU at 280 nm) in 20 mM Tris–HCl pH 8.0 containing 0.1 M NaCl and 0.2 mM DTT with a LS50B spectrofluorimeter (Perkin-Elmer) using a 1.0 cm path length quartz cuvette. Far-UV (190–250 nm) CD spectra were monitored at a protein concentration of 200 µg/mL (6.50×10^{-2} AU at 280 nm) in 50 mM Tris-Cl pH 8.0, 0.2 mM DTT, 0.2 M NaCl, using a 0.1 cm path length quartz cuvette. Near-UV (250–320 nm) CD spectra were monitored at a protein concentration of 4.6 mg/mL (1.49 AU at 280 nm) in 50 mM Tris-Cl pH 8.0, 2.0 mM DTT, 0.2 M NaCl, in a 1.0 cm path length quartz cuvette. CD measurements were performed in a JASCO-815 spectropolarimeter (Jasco, Easton, MD, USA) and the results were expressed as the mean residue ellipticity ([Θ]), assuming a mean residue molecular mass of 110 per amino acid residue. All spectroscopic measurements were carried out at 10 °C.

4.5. Urea-Induced Equilibrium Unfolding

For equilibrium transition studies, PPARγ wild type and variants (final concentration ranging over 100–200 µg/mL) were incubated at 10 °C at increasing concentrations of urea (0−9 M) in 20 mM Tris/HCl, pH 8.0, in the presence of 0.2 M NaCl and 200 µM DTT. After 10 min, equilibrium was reached and intrinsic fluorescence emission and far-UV CD spectra (0.1-cm cuvette) were recorded in parallel at 10 °C. To test the reversibility of the unfolding, PPARγ wild type and variants were unfolded at 10 °C in 8.0 M urea at protein concentration ranging over 1.0–2.0 mg/mL in 20 mM Tris/HCl, pH 8.0, in the presence of 2 mM DTT and 0.2 M NaCl. After 10 min, refolding was started by 10-fold dilution of the unfolding mixture at 10 °C into solutions of the same buffer used for unfolding containing decreasing urea concentrations. The final protein concentration ranged over 100–200 µg/mL. After 24 h, intrinsic fluorescence emission and far-UV CD spectra were recorded at 10 °C. All denaturation experiments were performed in triplicate.

4.6. Thermal Denaturation Experiments

PPARγ wild type and variants (protein concentration ranging over 0.10–0.20 mg/mL) were heated from 20 to 75 °C in a 0.1 cm quartz cuvette with a heating rate of 1 degree × min^{-1} controlled by a Jasco programmable Peltier element (Jasco, Easton, MD, USA). The dichroic activity at 222 nm and the photomultiplier voltage (PMTV) were continuously monitored in parallel every 0.5 °C [61]. All the thermal scans were corrected for the solvent contribution at the different temperatures.

Melting temperature (T_m) values were calculated by taking the first derivative of the ellipticity at 222 nm with respect to temperature. All denaturation experiments were performed in triplicate.

4.7. Data Analysis

Far-UV CD spectra recorded as a function of urea concentration were analyzed by a singular value decomposition algorithm (SVD) using the software MATLAB (Math-Works, South Natick, MA, USA) to remove the high-frequency noise and the low-frequency random errors and determine the number of independent components in any given set of spectra, as described in [40].

The changes in intrinsic fluorescence emission spectra at increasing urea concentrations were quantified as the intensity-averaged emission wavelength, $\overline{\lambda}$, [62] calculated according to

$$\overline{\lambda} = \sum(I_i\lambda_i)/\sum(I_i) \quad (1)$$

where λ_i and I_i are the emission wavelength and its corresponding fluorescence intensity at that wavelength, respectively. This quantity is an integral measurement, negligibly influenced by the noise, which reflects changes in the shape and position of the emission spectrum.

Urea-induced equilibrium unfolding transitions monitored by far-UV CD ellipticity and intrinsic fluorescence emission changes were analyzed by fitting baseline and transition region data to a two-state linear extrapolation model [63] according to

$$\triangle G_{unfolding} = \Delta G^{H_2O} + m[\text{Urea}] = -RT\ln K_{unfolding} \quad (2)$$

where $\Delta G_{\text{unfolding}}$ is the free energy change for unfolding for a given denaturant concentration, ΔG^{H_2O} is the free energy change for unfolding in the absence of denaturant, m is a slope term that quantifies the change in $\Delta G_{\text{unfolding}}$ per unit concentration of denaturant, R is the gas constant, T is the temperature and $K_{\text{unfolding}}$ is the equilibrium constant for unfolding. The model expresses the signal as a function of denaturant concentration:

$$y_i = \frac{y_N + s_N[X]_i + (y_U + s_U[X]_i) \times \exp\left[(-\triangle G^{H_2O} - m[X]_i/RT\right]}{1 + \exp\left[\frac{(-\Delta G^{H_2O} - m[X]_i}{RT}\right]} \quad (3)$$

where y_i is the observed signal; y_U and y_N are the baseline intercepts for unfolded and native protein, respectively; s_U and s_N are the baseline slopes for the unfolded and native protein, respectively; $[X]_i$ is the denaturant concentration after the ith addition; ΔG^{H_2O} is the extrapolated free energy of unfolding in the absence of denaturant, and m is the slope of a $\Delta G_{\text{unfolding}}$ versus $[X]$ plot. The denaturant concentration at the midpoint of the transition, $[\text{Urea}]_{0.5}$, according to Equation (2), is calculated as:

$$[\text{Urea}]_{0.5} = \Delta G^{H_2O}/m \quad (4)$$

The denaturation curve obtained by plotting the fluorescence changes of the PPARγ wild type and variants induced by increasing urea concentrations was fitted to the following equation assuming a three-state model:

$$F = \frac{F_N + \exp\left(m_{I-N}\frac{[\text{urea}]-D50_{I-N}}{RT}\right) \times \left(F_I + F_U\exp\left(m_{U-I}\frac{[\text{urea}]-D50_{U-I}}{RT}\right)\right)}{1 + \exp\left(m_{I-N}\frac{[\text{urea}]-D50_{I-N}}{RT}\right) \times \left(1 + \exp\left(m_{I-N}\frac{[\text{urea}]-D50_{I-N}}{RT}\right)\right)} \quad (5)$$

where F is $\overline{\lambda}$, calculated according to Equation (1); m is a constant that is proportional to the increase in solvent-accessible surface area between the two states involved in the transition; $D50_{I\text{–}N}$ and $m_{I\text{–}N}$ are the midpoint and m value for the transition between N and I, respectively; and $D50_{U\text{–}I}$ and $m_{U\text{–}I}$ are the midpoint and m value for the transition between I and U, respectively [64]. The $\overline{\lambda}$ of the intermediate

state (I), F_I, is constant, whereas that of the folded state (N) and of the unfolded state (U), F_N and F_U, respectively, has a linear dependence on denaturant concentration:

$$F_N = a_N + b_N[\text{urea}] \quad (6)$$

$$F_U = a_U + b_U[\text{urea}] \quad (7)$$

where a_N and a_U are the baseline intercepts for N and U, and b_N and b_U are the baseline slopes for N and U, respectively. All unfolding transition data were fitted using Graphpad Prism 5.04 (La Jolla, CA, USA).

4.8. Molecular Dynamics Simulations

Molecular Dynamics (MD) simulations were performed with the GROMACS package [65–68]. The initial coordinates of the wild-type protein were taken from the crystal structure of the PPARγ receptor [2] (PDB ID: 1PRG). The coordinates of the nine variants were adapted from the wild-type coordinates by performing a point mutation. Each system was placed in a dodecahedral box of sufficiently large dimensions such that nearby images lay more than 10 Å away. The box was filled with water molecules and an appropriate number of counter-ions to make the whole system neutral. As in [25], an OPLS force field [69] was used to simulate PPARγ and all its variants.

The equilibration strategy adopted for the nine systems is quite standard and is explained in detail in [70,71]. The temperature was held fixed at 300 K using the v-rescale thermostat [72] with a coupling time of 0.1 ps. The single point charge (SPC) model was employed for water molecules. Periodic boundary conditions were used throughout the simulation. Coulomb interactions have been dealt with by a standard Particle Mesh Ewald algorithm [73]. A time step of 2 fs was used. A non-bond pair list cutoff of 1.0 nm was used. The list was updated every 10 steps.

Each one of the 10 systems was simulated for 120 ns in the NVT ensemble. The analysis of the numerical data obtained in the simulation was carried out by GROMACS and VMD [74] tools according to needs.

5. Conclusions

In conclusion, the nine nsSNP PPARγ variants associated with metabolic disorders and /or cancer show alterations in the dynamics and tertiary contacts that impair the correct reciprocal positioning of H3 and H12, crucially important for PPARγ functioning. These alterations may lead to changes in the interactions with ligands and influence the multiple biological functions of this nuclear receptor.

Supplementary Materials: Supplementary materials can be found at www.mdpi.com/1422-0067/18/2/361/s1.

Acknowledgments: This work was supported by Regione Lazio (Prot. FILAS-RU-2014-1020).

Author Contributions: Conceived and designed the experiments: Laura Lori, Maria Petrosino, Alessandra Pasquo, Clorinda Lori, Roberta Chiaraluce, Valerio Consalvi, Velia Minicozzi, Silvia Morante. Performed the experiments: Laura Lori, Maria Petrosino, Alessandra Pasquo, Clorinda Lori, Roberta Chiaraluce, Valerio Consalvi. Mass spectroscopy: Alessandra Giorgi; Cell Culture and Transfections: Antonio Laghezza; Data analysis: Maria Petrosino, Laura Lori, Alessandra Pasquo, Davide Capelli, Roberta Chiaraluce, Valerio Consalvi. Molecular dynamics: Velia Minicozzi, Silvia Morante. Wrote the paper: Valerio Consalvi, Roberta Chiaraluce, Velia Minicozzi, Silvia Morante.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

DTT	Dithiothreitol
LBD	Ligand-Binding Domain
MD	Molecular Dynamics
nsSNPs	Non-synonymous single-nucleotide polymorphisms
PPARγ	Peroxisome Proliferator-Activated Receptor γ

References

1. Sauer, S. Ligands for the nuclear peroxisome proliferator-activated receptor γ. *Trends Pharmacol. Sci.* **2015**, *36*, 688–704. [CrossRef] [PubMed]
2. Nolte, R.T.; Wisely, G.B.; Westin, S.; Cobb, J.E.; Lambert, M.H.; Kurokawa, R.; Rosenfeld, M.G.; Willson, T.M.; Glass, C.K.; Milburn, M.V. Ligand binding and co-activator assembly of the peroxisome proliferator-activated receptor-γ. *Nature* **1998**, *395*, 137–143. [PubMed]
3. Chandra, V.; Huang, P.; Hamuro, Y.; Raghuram, S.; Wang, Y.; Burris, T.P.; Rastinejad, F. Structure of the intact PPAR-γ-RXR-nuclear receptor complex on DNA. *Nature* **2008**, *456*, 350–356. [CrossRef] [PubMed]
4. Kersten, S.; Desvergne, B.; Wahli, W. Roles of PPARs in health and disease. *Nature* **2000**, *405*, 421–424. [PubMed]
5. Michalik, L.; Wahli, W. Involvement of PPAR nuclear receptors in tissue injury and wound repair. *J. Clin. Investig.* **2006**, *116*, 598–606. [CrossRef] [PubMed]
6. Anghel, S.I.; Wahli, W. Fat poetry: A kingdom for PPARγ. *Cell. Res.* **2007**, *17*, 486–511. [CrossRef] [PubMed]
7. Evans, R.M.; Barish, G.D.; Wang, Y.X. PPARs and the complex journey to obesity. *Nat. Med.* **2004**, *10*, 355–361. [CrossRef] [PubMed]
8. Semple, R.K.; Chatterjee, V.K.; O'Rahilly, S. PPARγ and human metabolic disease. *J. Clin. Investig.* **2006**, *116*, 581–589. [CrossRef] [PubMed]
9. Wang, T.; Xu, J.; Yu, X.; Yang, R.; Han, Z.C. Peroxisome proliferator-activated receptor γ in malignant diseases. *Crit. Rev. Oncol. Hematol.* **2006**, *58*, 1–14. [CrossRef] [PubMed]
10. Mandard, S.; Patsouris, D. Nuclear control of the inflammatory response in mammals by peroxisome proliferator-activated receptors. *PPAR Res.* **2013**, *2013*, 613864. [CrossRef] [PubMed]
11. Lehrke, M.; Lazar, M.A. The many faces of PPARγ. *Cell.* **2005**, *123*, 993–999. [CrossRef] [PubMed]
12. Tontonoz, P.; Spiegelman, B.M. Fat and beyond: The diverse biology of PPARγ. *Annu. Rev. Biochem.* **2008**, *77*, 289–312. [CrossRef] [PubMed]
13. Sarraf, P.; Mueller, E.; Smith, W.M.; Wright, H.M.; Kum, J.B.; Aaltonen, L.A.; de la Chapelle, A.; Spiegelman, B.M.; Eng, C. Loss-of-function mutations in PPARγ associated with human colon cancer. *Mol. Cell* **1999**, *3*, 799–804. [CrossRef]
14. Garcia-Vallvé, S.; Guasch, L.; Tomas-Hernández, S.; del Bas, J.M.; Ollendorff, V.; Arola, L.; Pujadas, G.; Mulero, M. Peroxisome proliferator-activated receptor γ (PPARγ) and ligand choreography: Newcomers take the stage. *J. Med. Chem.* **2015**, *58*, 5381–5394. [CrossRef] [PubMed]
15. Kim, J.H.; Song, J.; Park, K.W. The multifaceted factor peroxisome proliferator-activated receptor γ (PPARγ) in metabolism, immunity, and cancer. *Arch. Pharm. Res.* **2015**, *38*, 302–312. [CrossRef] [PubMed]
16. Chan, K.H.; Niu, T.; Ma, Y.; You, N.C.; Song, Y.; Sobel, E.M.; Hsu, Y.H.; Balasubramanian, R.; Qiao, Y.; Tinker, L.; et al. Common genetic variants in peroxisome proliferator-activated receptor-γ (PPARG) and type 2 diabetes risk among Women's Health Initiative postmenopausal women. *J. Clin. Endocrinol. Metab.* **2013**, *98*, E600–E604. [CrossRef] [PubMed]
17. Walkey, C.J.; Spiegelman, B.M. A functional peroxisome proliferator-activated receptor-γ ligand-binding domain is not required for adipogenesis. *J. Biol. Chem.* **2008**, *283*, 24290–24294. [CrossRef] [PubMed]
18. Barroso, I.; Gurnell, M.; Crowley, V.E.; Agostini, M.; Schwabe, J.W.; Soos, M.A.; Maslen, G.L.; Williams, T.D.; Lewis, H.; Schafer, A.J.; et al. Dominant negative mutations in human PPARγ associated with severe insulin resistance, diabetes mellitus and hypertension. *Nature* **1999**, *402*, 880–883. [PubMed]
19. Savage, D.B.; Tan, G.D.; Acerini, C.L.; Jebb, S.A.; Agostini, M.; Gurnell, M.; Williams, R.L.; Umpleby, A.M.; Thomas, E.L.; Bell, J.D.; et al. Human metabolic syndrome resulting from dominant-negative mutations in the nuclear receptor peroxisome proliferator-activated receptor-γ. *Diabetes* **2003**, *52*, 910–917. [CrossRef] [PubMed]
20. Meirhaeghe, A.; Amouyel, P. Impact of genetic variation of PPARγ in humans. *Mol. Genet. Metab.* **2004**, *83*, 93–102. [CrossRef] [PubMed]
21. Agostini, M.; Schoenmakers, E.; Mitchell, C.; Szatmari, I.; Savage, D.; Smith, A.; Rajanayagam, O.; Semple, R.; Luan, J.; Bath, L.; et al. Non-DNA binding, dominant-negative, human PPARγ mutations cause lipodystrophic insulin resistance. *Cell Metab.* **2006**, *4*, 303–311. [CrossRef] [PubMed]

22. Tan, G.D.; Savage, D.B.; Fielding, B.A.; Collins, J.; Hodson, L.; Humphreys, S.M.; O'Rahilly, S.; Chatterjee, K.; Frayn, K.N.; Karpe, F. Fatty acid metabolism in patients with PPARγ mutations. *J. Clin. Endocrinol. Metab.* **2008**, *93*, 4462–4470. [CrossRef] [PubMed]
23. Jeninga, E.H.; Gurnell, M.; Kalkhoven, E. Functional implications of genetic variation in human PPARγ. *Trends Endocrinol. Metab.* **2009**, *20*, 380–387. [CrossRef] [PubMed]
24. Visser, M.E.; Kropman, E.; Kranendonk, M.E.; Koppen, A.; Hamers, N.; Stroes, E.S.; Kalkhoven, E.; Monajemi, H. Characterisation of non-obese diabetic patients with marked insulin resistance identifies a novel familial partial lipodystrophy-associated PPARγ mutation (Y151C). *Diabetologia* **2011**, *54*, 1639–1644. [CrossRef] [PubMed]
25. Lori, C.; Pasquo, A.; Montanari, R.; Capelli, D.; Consalvi, V.; Chiaraluce, R.; Cervoni, L.; Loiodice, F.; Laghezza, A.; Aschi, M.; et al. Structural basis of the transactivation deficiency of the human PPARγ F360L mutant associated with familial partial lipodystrophy. *Acta. Crystallogr. D Biol. Crystallogr.* **2014**, *70*, 1965–1976. [CrossRef] [PubMed]
26. Puhl, A.; Webb, P.; Polikarpov, I. Structural dataset for the PPARγ V290M mutant. *Data Brief.* **2016**, *7*, 1430–1437. [CrossRef] [PubMed]
27. Pasquo, A.; Consalvi, V.; Knapp, S.; Alfano, I.; Ardini, M.; Stefanini, S.; Chiaraluce, R. Structural stability of human protein tyrosine phosphatase ϱ catalytic domaineffect of point mutations. *PLoS ONE* **2012**, *7*, e32555. [CrossRef] [PubMed]
28. Lori, C.; Lantella, A.; Pasquo, A.; Alexander, L.T.; Knapp, S.; Chiaraluce, R.; Consalvi, V. Effect of single amino acid substitution observed in cancer on Pim-1 kinase thermodynamic stability and structure. *PLoS ONE* **2013**, *8*, e64824. [CrossRef] [PubMed]
29. Casadio, R.; Vassura, M.; Tiwari, S.; Fariselli, P.; Martelli, P.L. Correlating disease-related mutations to their effect on protein stability: A large-scale analysis of the human proteome. *Hum. Mutat.* **2011**, *32*, 1161–1170. [CrossRef] [PubMed]
30. Thusberg, J.; Olatubosun, A.; Vihinen, M. Performance of mutation pathogenicity prediction methods on missense variants. *Hum. Mutat.* **2011**, *32*, 358–368. [CrossRef] [PubMed]
31. Khan, S.; Vihinen, M. Performance of protein stability predictors. *Hum. Mutat.* **2010**, *31*, 675–684. [CrossRef] [PubMed]
32. Kucukkal, T.G.; Petukh, M.; Li, L.; Alexov, E. Structural and physico-chemical effects of disease and non-disease nsSNPs on proteins. *Curr. Opin. Struct. Biol.* **2015**, *32*, 18–24. [CrossRef] [PubMed]
33. Allali-Hassani, A.; Wasney, G.A.; Chau, I.; Hong, B.S.; Senisterra, G.; Loppnau, P.; Shi, Z.; Moult, J.; Edwards, A.M.; Arrowsmith, C.H.; et al. A survey of proteins encoded by non-synonymous single nucleotide polymorphisms reveals a significant fraction with altered stability and activity. *Biochem. J.* **2009**, *424*, 15–26. [CrossRef] [PubMed]
34. Zhang, Z.; Miteva, M.A.; Wang, L.; Alexov, E. Analyzing effects of naturally occurring missense mutations. *Comput. Math. Methods Med.* **2012**, *2012*, 805827. [CrossRef] [PubMed]
35. Yates, C.M.; Sternberg, M.J. Proteins and domains vary in their tolerance of non-synonymous single nucleotide polymorphisms (nsSNPs). *J. Mol. Biol.* **2013**, *425*, 1274–1286. [CrossRef] [PubMed]
36. Ormond, K.E.; Wheeler, M.T.; Hudgins, L.; Klein, T.E.; Butte, A.J.; Altman, R.B.; Ashley, E.A.; Greely, H.T. Challenges in the clinical application of whole-genome sequencing. *Lancet* **2010**, *375*, 1749–1751. [CrossRef]
37. Kucukkal, T.G.; Yang, Y.; Chapman, S.C.; Cao, W.; Alexov, E. Computational and experimental approaches to reveal the effects of single nucleotide polymorphisms with respect to disease diagnostics. *Int. J. Mol. Sci.* **2014**, *15*, 9670–9717. [CrossRef] [PubMed]
38. Forbes, S.A.; Bindal, N.; Bamford, S.; Cole, C.; Kok, C.Y.; Beare, D.; Jia, M.; Shepherd, R.; Leung, K.; Menzies, A.; et al. COSMIC: Mining complete cancer genomes in the catalogue of somatic mutations in cancer. *Nucleic Acids Res.* **2011**, *39*, D945–D950. [CrossRef] [PubMed]
39. Zhao, M.; Ma, L.; Liu, Y.; Qu, H. Pedican: An online gene for pediatric cancers with literature evidence. *Sci. Rep.* **2015**, *5*, 11435. [CrossRef] [PubMed]
40. Lori, L.; Pasquo, A.; Lori, C.; Petrosino, M.; Chiaraluce, R.; Tallant, C.; Knapp, S.; Consalvi, V. Effect of BET missense mutations on bromodomain function, inhibitor binding and stability. *PLoS ONE.* **2016**, *11*, e0159180. [CrossRef] [PubMed]
41. Choy, N.; Raussens, V.; Narayanaswami, V. Inter-molecular coiled-coil formation in human apolipoprotein E C-terminal domain. *J. Mol. Biol.* **2003**, *334*, 527–539. [CrossRef] [PubMed]

42. Kiss, R.S.; Weers, P.M.; Narayanaswami, V.; Cohen, J.; Kay, C.M.; Ryan, R.O. Structure-guided protein engineering modulates helix bundle exchangeable apolipoprotein properties. *J. Biol. Chem.* **2003**, *278*, 21952–21959. [CrossRef] [PubMed]
43. Myers, J.K.; Pace, C.N.; Scholtz, J.M. Denaturant m values and heat capacity changes: Relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* **1995**, *4*, 2138–2148. [CrossRef] [PubMed]
44. Geierhaas, C.D.; Nickson, A.A.; Lindorff-Larsen, K.; Clarke, J.; Vendruscolo, M. BPPred: A computational tool to predict biophysical quantities of proteins. *Protein Sci.* **2007**, *16*, 125–134. [CrossRef] [PubMed]
45. Auton, M.; Holthauzen, L.M.; Bolen, D.W. Anatomy of energetic changes accompanying urea-induced protein denaturation. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 15317–15322. [CrossRef] [PubMed]
46. Pinelli, A.; Godio, C.; Laghezza, A.; Mitro, N.; Fracchiolla, G.; Tortorella, V.; Lavecchia, A.; Novellino, E.; Fruchart, J.C.; Staels, B.; et al. Synthesis, biological evaluation, and molecular modeling investigation of new chiral fibrates with PPARα and PPARγ agonist activity. *J. Med. Chem.* **2005**, *48*, 5509–5519. [CrossRef] [PubMed]
47. Fariselli, P.; Martelli, P.L.; Savojardo, C.; Casadio, R. INPS: Predicting the impact of non-synonymous variations on protein stability from sequence. *Bioinformatics* **2015**, *31*, 2816–2821. [CrossRef] [PubMed]
48. Bromberg, Y.; Rost, B. Correlating protein function and stability through the analysis of single amino acid substitutions. *BMC Bioinformatics* **2009**, *10*, S8. [CrossRef] [PubMed]
49. Fernald, G.H.; Capriotti, E.; Daneshjou, R.; Karczewski, K.J.; Altman, R.B. Bioinformatics challenges for personalized medicine. *Bioinformatics* **2011**, *27*, 1741–1748. [CrossRef]
50. Stefl, S.; Nishi, H.; Petukh, M.; Panchenko, A.R.; Alexov, E. molecular mechanisms of disease-causing missense mutations. *J. Mol. Biol.* **2013**, *425*, 3919–3936. [CrossRef] [PubMed]
51. Menendez-Gutierrez, M.P.; Roszer, T.; Ricote, M. Biology and therapeutic applications of peroxisome proliferator-activated receptors. *Curr. Top. Med. Chem.* **2012**, *12*, 548–584. [CrossRef] [PubMed]
52. Peters, J.M.; Shah, Y.M.; Gonzalez, F.J. The role of peroxisome proliferator-activated receptors in carcinogenesis and chemoprevention. *Nat. Rev. Cancer* **2012**, *12*, 181–195. [CrossRef] [PubMed]
53. Renaud, J.P.; Moras, D. Structural studies on nuclear receptors. *Cell. Mol. Life. Sci.* **2000**, *57*, 1748–1769. [CrossRef] [PubMed]
54. Chen, Y.C.; Wu, J.S.; Tsai, H.D.; Huang, C.Y.; Chen, J.J.; Sun, G.Y.; Lin, T.N. Peroxisome proliferator-activated receptor γ (PPAR-γ) and neurodegenerative disorders. *Mol. Neurobiol.* **2012**, *46*, 114–124. [CrossRef] [PubMed]
55. Berger, J.P.; Akiyama, T.E.; Meinke, P.T. PPARs: Therapeutic targets for metabolic disease. *Trends Pharmacol. Sci.* **2005**, *26*, 244–251. [CrossRef] [PubMed]
56. Pochetti, G.; Mitro, N.; Lavecchia, A.; Gilardi, F.; Besker, N.; Scotti, E.; Aschi, M.; Re, N.; Fracchiolla, G.; Laghezza, A.; et al. Structural insight into peroxisome proliferator-activated receptor γ binding of two ureidofibrate-like enantiomers by molecular dynamics, cofactor interaction analysis, and site-directed mutagenesis. *J. Med. Chem.* **2010**, *53*, 4354–4366. [CrossRef] [PubMed]
57. Kallenberger, B.C.; Love, J.D.; Chatterjee, V.K.; Schwabe, J.W. A dynamic mechanism of nuclear receptor activation and its perturbation in a human disease. *Nat. Struct. Biol.* **2003**, *10*, 136–140. [CrossRef] [PubMed]
58. Pochetti, G.; Godio, C.; Mitro, N.; Caruso, D.; Galmozzi, A.; Scurati, S.; Loiodice, F.; Fracchiolla, G.; Tortorella, P.; Laghezza, A.; et al. Insights into the mechanism of partial agonism: Crystal structures of the peroxisome proliferator-activated receptor γ ligand-binding domain in the complex with two enantiomeric ligands. *J. Biol. Chem.* **2007**, *282*, 17314–17324. [CrossRef] [PubMed]
59. Raspé, E.; Madsen, L.; Lefebvre, A.M.; Leitersdorf, I.; Gelman, L.; Peinado-Onsurbe, J.; Dallongeville, J.; Fruchart, J.C.; Berge, R.; Staels, B. Modulation of rat liver apolipoprotein gene expression and serum lipid levels by tetradecylthioacetic acid (TTA) via PPARalpha activation. *J. Lipid Res.* **1999**, *40*, 2099–2110. [PubMed]
60. Gill, S.C.; von Hippel, P.H. Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **1989**, *182*, 319–326. [CrossRef]
61. Benjwal, S.; Verma, S.; Rohm, K.H.; Gursky, O. Monitoring protein aggregation during thermal unfolding in circular dichroism experiments. *Protein Sci.* **2006**, *15*, 635–639. [CrossRef] [PubMed]
62. Royer, C.A.; Mann, C.J.; Matthews, C.R. Resolution of the fluorescence equilibrium unfolding profile of Trp aporepressor using single tryptophan mutants. *Protein Sci.* **1993**, *2*, 1844–1852. [CrossRef] [PubMed]

63. Santoro, M.M.; Bolen, D.W. Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl α-chymotrypsin using different denaturants. *Biochemistry* **1988**, *27*, 8063–8068. [CrossRef] [PubMed]
64. Rowling, P.J.; Cook, R.; Itzhaki, L.S. Toward classification of BRCA1 missense variants using a biophysical approach. *J. Biol. Chem.* **2010**, *285*, 20080–20087. [CrossRef] [PubMed]
65. Berendsen, H.J.C.; van der Spoel, D.; van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **1995**, *91*, 43–56. [CrossRef]
66. Lindahl, E.; Hess, B.; van der Spoel, D. Gromacs 3.0: A package for molecular simulation and trajectory analysis. *J. Mol. Model.* **2001**, *7*, 306–317. [CrossRef]
67. Van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.E.; Berendsen, H.J.C. Gromacs: Fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701–1718. [CrossRef] [PubMed]
68. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447. [CrossRef] [PubMed]
69. Jorgensen, W.L.; Maxwell, D.S.; Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236. [CrossRef]
70. Minicozzi, V.; Chiaraluce, R.; Consalvi, V.; Giordano, C.; Narcisi, C.; Punzi, P.; Rossi, G.C.; Morante, S. Computational and experimental studies on β-sheet breakers targeting Amyloid-β1–40 fibrils. *J. Biol. Chem.* **2014**, *289*, 11242–11252. [CrossRef] [PubMed]
71. Di Carlo, M.G.; Minicozzi, V.; Foderà, V.; Militello, V.; Vetri, V.; Morante, S.; Leone, M. Thioflavin-T templates amyloid-β(1–40) conformation and aggregation pathway. *Biophys. Chem.* **2015**, *206*, 1–11. [CrossRef] [PubMed]
72. Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101. [CrossRef] [PubMed]
73. Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An *N*·log (*N*) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092. [CrossRef]
74. Humphrey, W.; Dalke, A.; Schulten, K. VMD-Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38. [CrossRef]

Supplementary Materials: Single Nucleotide Polymorphism of PPARγ, A Protein at the Crossroads of Physiological and Pathological Processes

Maria Petrosino, Laura Lori, Alessandra Pasquo, Clorinda Lori, Valerio Consalvi, Velia Minicozzi, Silvia Morante, Antonio Laghezza, Alessandra Giorgi, Davide Capelli and Roberta Chiaraluce

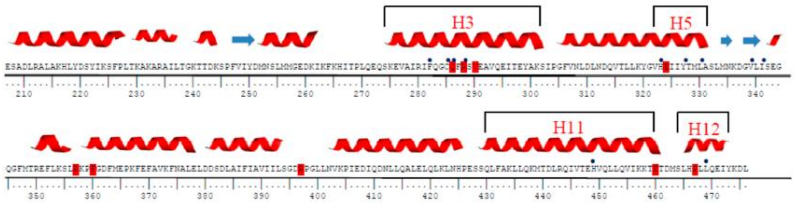


Figure S1. Amino acid sequence of PPARγ LBD. Secondary structure elements are shown at the top of the amino acid sequence. Mutated residues are highlighted in red. Blue dots indicate the residues involved in ligand binding (1PRG.pdb). Helices are numbered according to Nolte et al. [1].

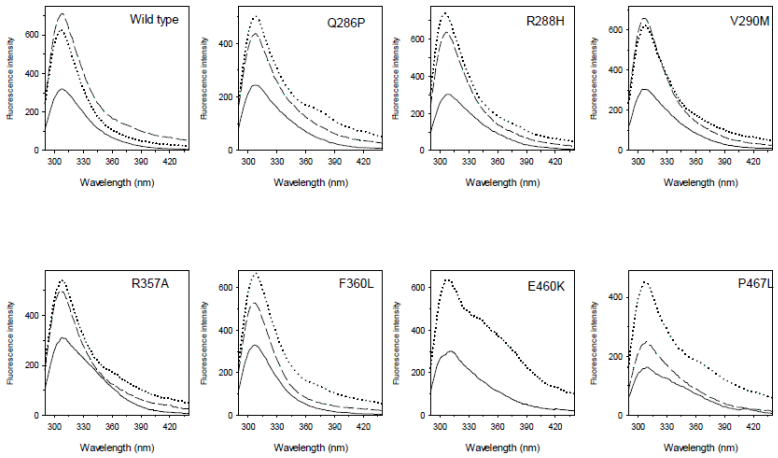


Figure S2. Intrinsic fluorescence emission spectra of PPARγ wild type and variants. Fluorescence spectra of PPARγ wild type and variants in 0 M (continuous lines), 8.0 M (dotted lines), 4.00 M (Q286P and F360L, dashed lines), and 4.07 M urea (wild type, R288H, V290M, and R357A, dashed lines) were recorded at 0.1 mg/mL protein concentration (274 nm excitation wavelength) in 20 mM Tris–HCl pH 8.0 containing 0.1 M NaCl and 0.2 mM DTT.

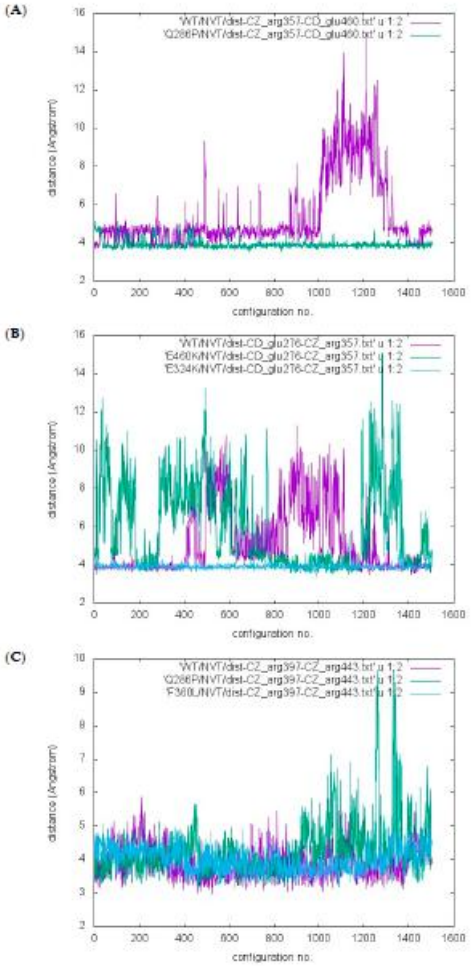


Figure S3. (**A**) Distance between R357 and E460. Distance between the carbon atom binding the two amino groups of arginine 357 lateral chain (CZ) and the carbon atom binding the carboxyl group of glutamic 460 lateral chain (CD), in the wild type and in the Q286P variant. On the x-axis is the configuration number; on the y-axis is the distance in Å; (**B**) Distance between E276 and R357. Distance between CD of glutamic 276 and CZ of arginine 357 in the wild type, in E460K, and in E324K variants. On the x-axis is the configuration number; on the y-axis is the distance in Å; (**C**) Distance between R397 and R443. Distance between CZ of arginine 397 and CZ of arginine 443 in the wild type, in Q286P, and in F360L variants. On the x-axis is the configuration number; on the y-axis is the distance in Å.

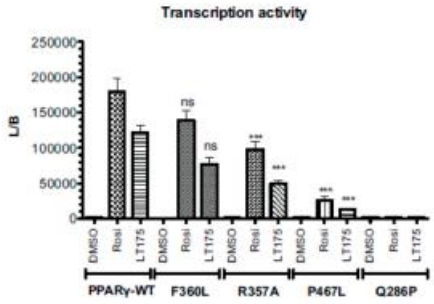


Figure S4. Transcription activity of wild type PPARγ and mutants in a PPARγ Gal4-based assay. Luciferase activity was normalized for difference in transfection efficiency by β-galactosidase activity and each point is the mean ± SEM of at least two experiments each performed in duplicate. Results are expressed as the ratio between luciferase activity and β-galactosidase activity (L/B). Differences between mutants and control (wild type with the same treatment) were significant (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; one-way ANOVA, Bonferroni test).

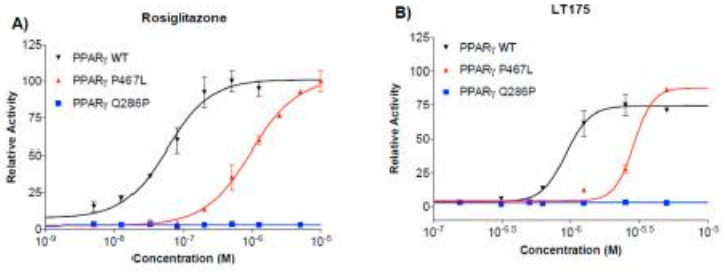


Figure S5. Transcription activity. Transcription activity of rosiglitazone (**A**); and LT175 (**B**) towards wild type PPARγ, PPARγ P467L, and PPARγ Q286P in a PPARγ Gal4-based assay. Results are expressed as a percentage of the highest efficacy obtained with rosiglitazone and each point is the mean ± SEM of two experiments each performed in duplicate.

Table S1. Transcription activity of PPARγ wild type and mutants in a PPARγ Gal4-based assay.

PPARγ	Rosiglitazone	LT175
PPARγ wild type	53 ± 17	1100 ± 180
PPARγ F360L	790 ± 70	2500 ± 500
PPARγ R357A	380 ± 70	1530 ± 270
PPARγ P467L	960 ± 180	2880 ± 290
PPARγ Q286P	-	-

Potency of rosiglitazone and LT175 towards wild type PPARγ and mutants in a PPARγ Gal4-based assay. Results are expressed as EC_{50} (nM) and are the mean ± SEM of two experiments each performed in duplicate.

Table S2. List of oligonucleotides used for site-directed mutagenesis.

PPARγ Variant	Oligonucleotide
Q286P	Forward: GCATCTTTCAGGGCTGCCCGTTTCGCTCCGTGGAG Reverse: CTCCACGGAGCGAAACGGGCAGCCCTGAAAGATGC
R288H	Forward: CAGGGCTGCCAGTTTCATTCCGTGGAGGCTGTGC Reverse: GCACAGCCTCCACGGAATGAAACTGGCAGCCCTG
V290M	Forward: GGGCTGCCAGTTTCGCTCCATGGAGGCTGTGCAGGAGATC Reverse: GATCTCCTGCACAGCCTCCATGGAGCGAAACTGGCAGCCC
R357A	Forward: GGAGTTTCTAAAGAGCCTGGCAAAGCCTTTTGGTG Reverse: CACCAAAAGGCTTTGCCAGGCTCTTTAGAAACTCC
F360L	Forward: GCCTGCGAAAGCCTCTGGGTGACTTTATGGAGCCC Reverse: GGGCTCCATAAAGTCACCCAGAGGCTTTCGCAGGC
P467L	Forward: GACATGAGTCTTCACCTGCTCCTGCAGGAG Reverse: CTCCTGCAGGAGCAGGTGAAGACTCATGTC
E460K	Forward: GCAGGTGATCAAGAAGACGAAGACAGACATGAGTCTTCACCCGC Reverse: GCGGGTGAAGACTCATGTCTGTCTTCGTCTTCTTGATCACCTGC
R397C	Forward: AGTGGAGACTGCCCAGGTTTGCT Reverse: AGCAAACCTGGGCAGTCTCCACT
E324K	Forward: CCTCAAATATGGAGTCCACAAGATCATTTACACAATGCTGGCC Reverse: GGCCAGCATTGTGTAAATGATCTTGTGGACTCCATATTTGAGG

Table S3. PPARγ variants.

PPARγ Variant	SNP ID	Reference
Q286P	rs121909242	[2]
R288H	rs28936407	[2]
V290M	rs72551362	[3]
E324K	rs530007199 COSM1037602	https://bioinfo.uth.edu/TSGene/gene_mutation.cgi?gene=5468#cosmic [4]; http://pedican.bioinfo-minzhao.org/gene_mutation.cgi?gene=5468 [5]
R357A	–	[6]
F360L	rs72551363	[7]
R397C	rs72551364	[6]
E460K	–	https://bioinfo.uth.edu/TSGene/gene_mutation.cgi?gene=5468#cosmic [4]; http://pedican.bioinfo-minzhao.org/gene_mutation.cgi?gene=5468 [5]
P467L	rs121909244	[6,8]

References

1. Nolte, R.T.; Wisely, G.B.; Westin, S.; Cobb, J.E.; Lambert, M.H.; Kurokawa, R.; Rosenfeld, M.G.; Willson, T.M.; Glass, C.K.; Milburn, M.V. Ligand binding and co-activator assembly of the peroxisome proliferator-activated receptor-γ. *Nature* **1998**, *395*, 137–143.
2. Sarraf, P.; Mueller, E.; Smith, W.M.; Wright, H.M.; Kum, J.B.; Aaltonen, L.A.; de la Chapelle, A.; Spiegelman, B.M.; Eng, C. Loss-of-function mutations in PPARγ associated with human colon cancer. *Mol. Cell* **1999**, *3*, 799–804.
3. Puhl, A.; Webb, P.; Polikarpov, I. Structural dataset for the PPARγ V290M mutant. *Data Brief.* **2016**, *7*, 1430–1437.
4. Tumor Suppressor Gene Database. Available online: https://bioinfo.uth.edu/TSGene/gene_mutation.cgi?gene=5468#cosmic (accessed on 31 October 2016).
5. Pediatric Cancer Gene Database. Available online: http://pedican.bioinfo-minzhao.org/gene_mutation.cgi?gene=5468 (accessed on 31 October 2016).
6. Jeninga, E.H.; Gurnell, M.; Kalkhoven, E. Functional implications of genetic variation in human PPARγ. *Trends Endocrinol. Metab.* **2009**, *20*, 380–387.

7. Lori, C.; Pasquo, A.; Montanari, R.; Capelli, D.; Consalvi, V.; Chiaraluce, R.; Cervoni, L.; Loiodice, F.; Laghezza, A.; Aschi, M.; et al. Structural basis of the transactivation deficiency of the human PPARγ F360L mutant associated with familial partial lipodystrophy. *Acta. Crystallogr. D Biol. Crystallogr.* **2014**, *70*, 1965–1976.
8. Kallenberger, B.C.; Love, J.D.; Chatterjee, V.K.; Schwabe, J.W. A dynamic mechanism of nuclear receptor activation and its perturbation in a human disease. *Nat. Struct. Biol.* **2003**, *10*, 136–140.

Biophysical Chemistry 229 (2017) 110–114

Contents lists available at ScienceDirect

Biophysical Chemistry

journal homepage: www.elsevier.com/locate/biophyschem

The effect of β-sheet breaker peptides on metal associated Amyloid-β peptide aggregation process

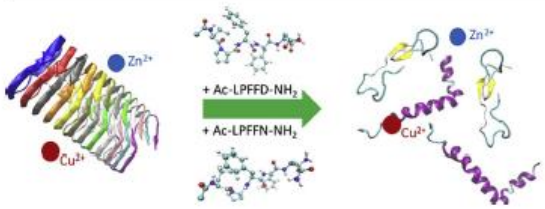
F. Stellato[a], Z. Fusco[b], R. Chiaraluce[c], V. Consalvi[c], S. Dinarelli[d], E. Placidi[a,d], M. Petrosino[c], G.C. Rossi[a,e], V. Minicozzi[a,*], S. Morante[a]

[a] Dipartimento di Fisica, Università di Roma Tor Vergata and INFN, Via della Ricerca Scientifica 1, 00133 Rome, Italy
[b] College of Engineering and Computer Science (CECS), Australian National University, Canberra, ACT 0200, Australia
[c] Dipartimento di Biochimica, Sapienza Università di Roma, Piazzale Aldo Moro 5, 00185 Rome, Italy
[d] Istituto di Struttura della Materia, CNR, via del Fosso del Cavaliere 100, 00133 Rome, Italy
[e] Centro Fermi - Museo Storico della Fisica e Centro Studi e Ricerche "Enrico Fermi", Rome, Italy

HIGHLIGHTS

- The effect of Cu(II)/Zn(II) ions and β-sheet breakers on $A\beta_{1-40}$ fibril formation
- CD spectroscopy and AFM monitor secondary structure changes and fibril formation
- No β-sheet structure transition and no fibrils in the presence of β-sheet breakers
- β-Sheet breakers peptides prevent fibril formation in the presence of metal ions.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Amyloid-β peptide
Fibrils
Atomic Force Microscopy
Circular Dichroism
Metal ions
Inhibitors

ABSTRACT

Far-UV Circular Dichroism experiments and Atomic Force Microscopy tomography are employed to assess the impact of β-sheet breakers on the $A\beta_{1-40}$ peptide aggregation process in the presence of Cu^{2+} or Zn^{2+} transition metals. In this work we focus on two specific 5-amino acids long β-sheet breakers, namely the LPFFD Soto peptide, already known in the literature, and the LPFFN peptide recently designed and studied by our team. We provide evidence that both β-sheet breakers are effective in reducing the $A\beta_{1-40}$ aggregation propensity, even in the presence of metal ions.

1. Introduction

Alzheimer's disease (AD) is an irreversible and progressive neurodegenerative disorder belonging to a large class of pathologies called amyloidoses [1].

The hallmark of all amyloidoses is the switching of endogenous proteins or peptides from their physiological soluble state to a pathological fibrillar insoluble one. The proteinaceous agent involved in the AD development is the Amyloid-β peptide (Aβ), a 39 to 43 amino acids long peptide that is produced by the proteolytic cleavage of a membrane protein called amyloid precursor protein (APP) [2].

The fingerprint of the disease in AD patient brains is the presence of self-associating fibrous protein aggregates which are deposited in the extracellular tissues in the form of characteristic cross β-sheet arrays [3].

* Corresponding author.
E-mail address: mi.nicozzi@roma2.infn.it (V. Minicozzi).

http://dx.doi.org/10.1016/j.bpc.2017.05.005
Received 5 April 2017; Received in revised form 9 May 2017; Accepted 11 May 2017
Available online 15 May 2017
0301-4622/

Mature fibrils are usually made by up to 6 protofilaments braided together to form a rope-like fiber. Fibers are 5–10 nm in diameter and up to a few microns in length [4]. Fibrils are unbranched structures, extremely stable and resistant to degradation by proteases and denaturants. It is a remarkable fact that the development of a cross β -sheet arrangement with the subsequent fibril formation occurs following an Aβ conformational change, in which the peptide partially loses its random coil or α-helical native structure in favour of a β-sheet secondary structure [5].

An attractive therapeutic strategy for AD could then be that of blocking the early steps of the misfolding and aggregation of the soluble Aβ peptides. It has been suggested that this goal can be achieved if the Aβ peptide is let to interact with short peptides, called β-sheet-breaker (BSBs) [6,7], as they have been shown to be able to inhibit the Aβ secondary structure modification that triggers the successive aggregation process and fibril formation.

In particular, it has been recently shown in ref. [8] that these peptides are able to block (or at least strongly delay) fibril formation when they come to interact with the Aβ hydrophobic region (the 17–21 segment). In this reference a careful spectroscopic analysis of the effect of different kinds of BSBs on the $A\beta_{1-40}$ secondary structure is presented. Experimental results were corroborated by classical Molecular Dynamics (MD) simulations. The focus was on two kinds of 5-amino acids long BSBs, namely Ac-LPFFD-NH_2 (also known as the Soto breaker [9]) and Ac-LPFFN-NH_2, a variation of the Soto breaker, purposely designed by the Authors to study the dependence upon the specific breaker amino acidic content of its ability to prevent $A\beta_{1-40}$ misfolding.

Experimental studies included in time Thioavin T (ThT) experiments, far-UV Circular Dichroism (far-UV CD) measurements and Electro-Spray-Ionization Mass Spectrometry (ESI-MS). ThT fluorescence provided experimental evidence of an enhanced (compared to the Soto breaker) ability of Ac-LPFFN-NH_2 to delay Aβ fibril formation. Far-UV CD measurements confirmed the strong stabilization effect of Ac-LPFFN-NH_2 on the $A\beta_{1-40}$ secondary structure, while ESI-MS data demonstrated that Ac-LPFFN-NH_2 preferentially interacts with the 17–21 hydrophobic $A\beta_{1-40}$ region. The general pattern that emerges from all these measurements was beautifully confirmed by MD simulations. The latter not only undoubtedly proved that both BSBs are able to interact with the Aβ peptide, reducing its residues mobility and interfering with possible secondary structure changes, but also confirmed the experimental finding that the BSB-Aβ interaction occurs precisely in the 17–21 hydrophobic $A\beta_{1-40}$ region.

Further recent studies [10,11,12,13,14,15,16,17] addressed the question of the role of metals, especially Cu^{2+} and Zn^{2+}, in Aβ misfolding and aggregation, as metals are found in high concentrations (~400 mM and ~1 mM, respectively) [18,19] in the Aβ plaques present in AD patient brains and recognized to affect the rate of fibril formation [20,21] as well as fibril morphology [22].

Moreover, Cu^{2+} and Zn^{2+} are seen to increase amyloid plaque formation and toxicity in AD model animals [23]. Therefore, the effectiveness of any inhibiting peptide should be tested in the presence of metal ions.

In the present work we report the results of far-UV CD spectroscopy and Atomic Force Microscopy (AFM) experiments that have been carried out with the specific purpose of assessing the degree of inhibiting power of the Ac-LPFFD-NH_2 and Ac-LPFFN-NH_2 BSBs on the $A\beta_{1-40}$ aggregation propensity when also Cu^{2+} and/or Zn^{2+} ions are present.

Our extensive far-UV CD study of the $A\beta_{1-40}$ secondary structure in a variety of physico-chemical conditions leads to the conclusion that the presence of metal ions does not significantly affect the ability of BSBs of inhibiting the Aβ peptide β-sheet secondary structure formation. AFM images nicely confirms the inhibiting power of both the BSBs we have investigated.

2. Materials and methods

2.1. Peptide preparation

$A\beta_{1-40}$ was purchased from the PolyPeptide Laboratories France SAS. Ac-LPFFD-NH_2 and Ac-LPFFN-NH_2 peptides were synthesized on Rink amide resin (65-mmol scale, 100 mg). Fmoc-amino acids-OH (Fmoc-AA-OH), diisopropylcarbodiimide (DIC), and hydroxybenzotriazole (HOBt) (3 eq, respectively) in *N*-methyl-2-pyrrolidinone are used for couplings. The coupling mixture was shaken overnight. After cleavage, the crude peptides were purified by reverse phase-HPLC on a Waters 600E liquid chromatography system by using a Waters Bondapac C-18 column (1.9 × 30 cm, 5 μm, 300 Å) for semi-preparative scale with elution at 8 ml/min by a linear gradient of 10–60% acetonitrile in 0.1% aqueous trifluoroacetic acid for 30 min. Peptide purity is measured to be > 97% by analytical HPLC (Waters μ-Bondapack C-18 column, 0.39 × 30 cm, 5 μm, 300 Å). The peptides were characterized on a Q-TOF MICRO spectrometer (Micromass, now Waters, Manchester, UK) equipped with an ESI source, in the positive ion mode, and data were analyzed using the MassLynx software.

Aliquots of lyophilized $A\beta_{1-40}$ were dissolved in 20 mM Tris-HCl, pH 7.5, at a 55 μM $A\beta_{1-40}$ final concentration. Ac-LPFFD-NH_2 as well as Ac-LPFFN-NH_2 were dissolved in ultra-high quality water, lyophilized, and stored at 253 K. BSBs are added in 10-fold molar concentration excess with respect to the peptide. Metals (in the form of chlorides, $CuCl_2$ and $ZnCl_2$) are dissolved at the slightly sub-stoichiometric concentration of 50 μM.

In order to have an exhaustive set of samples for assessing the influence of BSBs on the $A\beta_{1-40}$ metal ions induced secondary structure arrangement, we have prepared seven different samples as detailed in Table 1. In the first column we give the name of samples as it is used in the text. In column two and three the kind of metal and BSB present in solution is indicated.

2.2. Far-UV CD spectroscopy measurements

Far-UV CD measurements were performed with a Jasco J-720 spectropolarimeter in a 0.02 cm path length quartz cuvette thermostated at 293 K. The results are expressed in terms of the mean residue ellipticity [θ] assuming a mean residue weight of 110 Da per amino acid.

2.3. AFM measurements

AFM measurements were carried out using a NanoSurf Flex-AFM. Images were acquired in tapping mode by using BudgetSensor TAP190Al-G silicon nitride probes with nominal spring constant and resonant frequency of 49 N/m and 190 kHz, respectively. The acquisition parameters, such as gains, set-point and speed were adjusted to yield maximum image quality. Samples were prepared by drop-casting 3 μl of solution onto freshly cleaved mica surfaces. Before starting the

Table 1
Measured samples. In the first column we give the name of sample. In column two and three the kind of metal and BSB present in solution are indicated. Aβ concentration is always 55 μM, metal ions are added at a slightly sub-stochiometric concentration (50 μM) while the BSBs concentration is taken to be 10 times that of the Aβ peptide.

Sample	Metal ion (50 μM)	Breaker (550 μM)
Aβ	–	
Aβ + Cu	Cu^{2+}	
Aβ + Zn	Zn^{2+}	
Aβ + Cu + LPFFN	Cu^{2+}	LPFFN
Aβ + Cu + LPFFD	Cu^{2+}	LPFFD
Aβ + Zn + LPFFN	Zn^{2+}	LPFFN
Aβ + Zn + LPFFD	Zn^{2+}	LPFFD

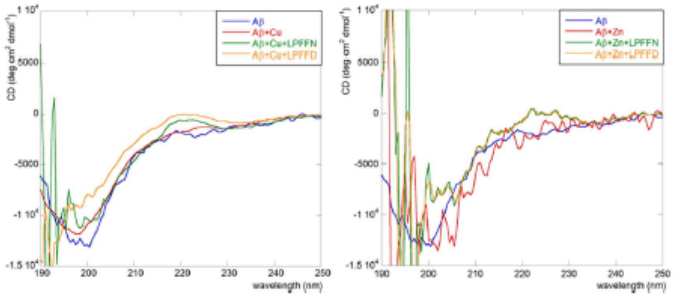


Fig. 1. Far-UV CD spectra at $t = 0$. In the presence of Cu^{2+} (left panel) and Zn^{2+} (right panel). (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

AFM measurements, each sample was allowed to completely dry in air for 15 min. All the AFM measurements were performed in air, at room temperature with a constant 30% relative humidity. AFM images were then processed using the Gwyddion 2.34 Data Processing Software [24] (www.gwyddion.net).

3. Results and discussion

3.1. Far-UV CD spectroscopy results

To monitor the $A\beta_{1-40}$ conformational changes possibly occurring in the various sample conditions under investigation (see Table 1), CD spectra were acquired immediately after sample preparation ($t = 0$) and after 24 h of continuous, mild stirring at 293 K. Stirring is necessary to promote secondary structure changes leading to fibrils formation.

Fig. 1 shows the CD spectra of the samples listed in Table 1 at $t = 0$. In the left panel we compare the CD spectra in the presence of Cu^{2+} (with and without BSBs) with the spectrum of the Aβ peptide alone. In the right panel we compare the CD spectra in the presence of Zn^{2+} (with and without BSBs) with the spectrum of the Aβ peptide alone. All of the spectra display a negative band with a minimum around 198 nm, which is typical of a random coil arrangement [25]. This finding confirms that at $t = 0$ the $A\beta_{1-40}$ secondary structure is not in a β-sheet configuration irrespective of its solution conditions.

Fig. 2 shows the corresponding (smoothed) spectra after 24 h of continuous, mild stirring at 293 K. The sample in the presence of Cu^{2+} (left panel) and in the absence of breakers display a visible increase of the negative molar ellipticity at ~218 nm and a positive peak at ~195 nm.

These two features, that are indicative of the conformational transition of the $A\beta_{1-40}$ peptide from random coil to a β-sheet secondary structure, are less pronounced in the presence of Zn^{2+} (right panel) thus suggesting that Zn^{2+} presence is more effective in limiting this Aβ peptide secondary structure change [26].

It is worth noticing that the spectra of the samples with and without BSB inhibitors substantially differ in the region between 190 nm and 200 nm (compare green and orange with red and blue data). This may be due to the relatively high BSB concentration that makes them to interact among themselves and form small aggregates of dimensions comparable to the beam wavelength. These aggregates fluctuate in the solution and the scattering events produced by the beam hitting them result in the peculiar features visible in the leftmost part of the spectrum.

In conclusion, the remarkable difference in the shape of the green and orange data compared to that of the red and blue data in Fig. 2 gives evidence for the fact that the BSBs ability in maintaining the $A\beta_{1-40}$ peptide in its native random coil structure [8] is not reduced in the presence of Zn^{2+} or Cu^{2+} metal ions.

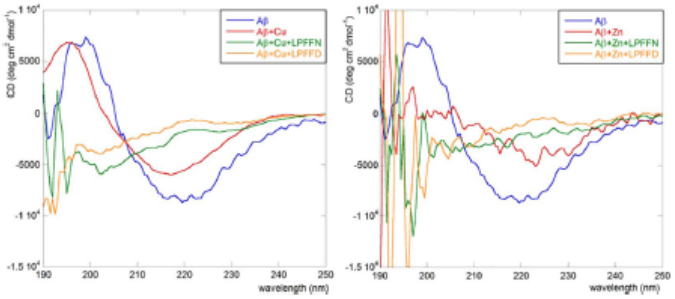


Fig. 2. Far-UV CD spectra after 24 h of continuous, mild stirring at room temperature. Left panel: samples in the presence of Cu^{2+}. Right panel: samples in the presence of Zn^{2+}. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

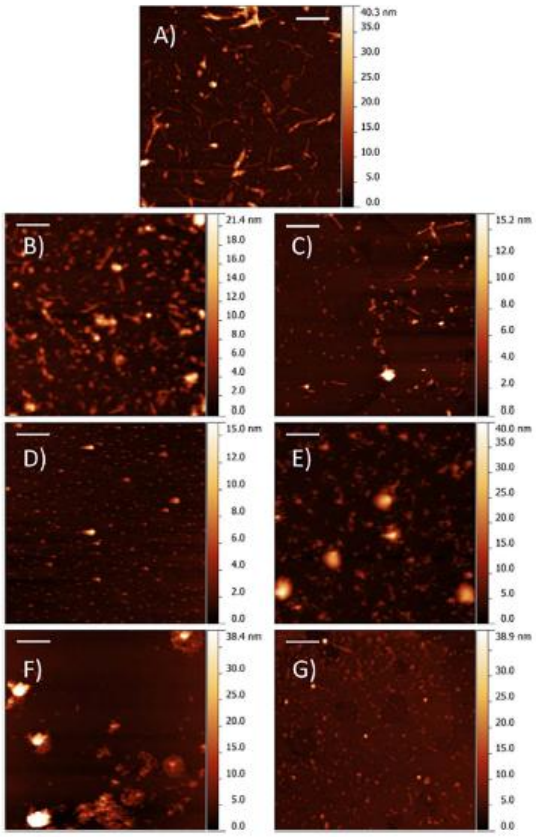


Fig. 3. Representative 6 × 6 μm² AFM images of: $A\beta_{1-40}$ alone (panel A), $A\beta_{1-40}$ in the presence of Cu^{2+} (panel B), $A\beta_{1-40}$ in the presence of Zn^{2+} (panel C), $A\beta_{1-40}$ in the presence of Cu^{2+} ions and LPFFN breakers (panel D), $A\beta_{1-40}$ in the presence of Zn^{2+} and LPFFN breakers (panel E), $A\beta_{1-40}$ in the presence of Cu^{2+} and LPFFD breakers (panel F), $A\beta_{1-40}$ in the presence of Zn^{2+} ions and LPFFD breakers (panel G). z –axis color scale has been optimized in each image to enhance visibility. The white scale bar is 1 μm. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.2. AFM results

Representative AFM topographies of the seven samples listed in Table 1 are shown in Fig. 3. Although a quantitative analysis of fibril formation is not possible due to the high inhomogeneity of the deposited sample on the mica substrate, the panorama of visible aggregates and/or fibrils is quite differentiated and provides us a fairly clear picture of the situation.

Indeed, in the AFM image of the sample containing the $A\beta_{1-40}$ peptide alone (Panel A) several aggregates with fibril-like structure and few fibrils bundles are visible. This is in agreement with the CD observed transition from coil to β-sheet secondary structure (compare blue data in Figs. 1 and 2) which is considered a crucial initial step toward fibril formation [6].

When metal ions, either Cu2 + (panel B) or Zn2 + (panel C), are present in solution, the number of fibrillar structures is visibly reduced, but fibrils are still present. The aggregate morphology is drastically modified, instead, when either LPFFN or LPFFD BSBs (panel D to G) are present. in this case fibrils are not any more visible and the predominant morphological structures are globular or amorphous aggregates with diameters ranging from few tens to few hundred nanometers. Again, absence of fibril aggregates is consistent with the CD spectra, where conformational transitions are not observed in the presence of BSBs (green and orange data display neither the typical increase of the negative molar ellipticity at about 218 nm nor the positive peak at about 195 nm).

4. Conclusions and outlooks

In a previous study [8] we have been able to establish the ability of the Ac-LPFFD-NH_2 and Ac-LPFFN-NH_2 BSBs in delaying Aβ peptide fibril formation.

Given the well-established, important role played by metal ions (especially Cu^{2+} and Zn^{2+}) in the process leading to misfolding and fibril formation [10,11,12,13,15,17,20], we have here investigated whether the effect of BSBs on the Aβ peptide aggregation process is modified in presence of metal ions.

To this purpose we have performed far-UV CD spectroscopy and AFM tapping mode measurements of $A\beta_{1-40}$ samples in the presence and in the absence of BSBs and/or metal ions.

CD data show that Ac-LPFFD-NH_2 and Ac-LPFFN-NH_2 BSBs keep their ability of "freezing" the native secondary structure of $A\beta_{1-40}$ peptides even when the latter are in the presence of metal ions. These results are nicely confirmed by AFM tomography. A careful inspection of AFM images reveals that fibrillar structures are present in all the samples prepared in the absence of BSBs, while only globular or amorphous aggregates are observed in samples containing BSBs. This is a high-valued information if BSBs will have to be ever used to prevent $A\beta_{1-40}$ fibrillation.

References

[1] R. Knowles, C. Wyatt, S.V. Buldyrev, I. Cruz, B. Urbanc, M. Hasselmo, B. Hyman, Plaque-induced neurite abnormalities: implications for disruption of neural networks in Alzheimer' disease, Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 5274–5279.

[2] S. Gandy, The role of cerebral amyloid beta accumulation in common forms of Alzheimer disease, J. Clin. Investig. 115 (2005) 1121–1129.

[3] L. Serpell, Alzheimer's amyloid fibrils. Structure and assembly, Biochim. Biophys. Acta 1502 (2000) 16–30.

[4] J. Jimenez, E. Nettleton, M. Bouchard, C. Robinson, C. Dobson, H. Saibil, The protofilament structure of insulin amyloid fibrils, Proc. Natl. Acad. Sci. U. S. A. (2002) 9196–9201.

[5] D. Selkoe, Folding proteins in fatal ways, Nature 426 (2003) 900–904.

[6] C. Soto, E. Castano, B. Frangione, N. Inestrosa, The α-helical to β-strand transition in the amino-terminal fragment of the amyloid β-peptide modulates amyloid formation, J. Biol. Chem. 270 (1995) 3063–3067.

[7] Q. Nie, X. Du, M. Geng, Small molecule inhibitors of amyloid-β peptide aggregation as a potential therapeutic strategy for Alzheimer disease, Acta Pharmacol. Sin. 32 (2011) 545–551.

[8] V. Minicozzi, R. Chiaraluce, V. Consalvi, C. Giordano, C. Narcisi, P. Punzi, G. Rossi, S. Morante, Computational and experimental studies on β-sheet breakers targeting a β_{1-40} fibrils, J. Biol. Chem. 289 (2014) 11242–11252.

[9] C. Adessi, M. Frossard, C. Boissard, S. Fraga, S. Bieler, T. Ruckle, F. Vilbois, S. Robinson, M. Mutter, W. Banks, C. Soto, Pharmacological profiles of peptide drug candidates for the treatment of Alzheimer disease, J. Biol. Chem. 278 (2003) 13905–13911.

[10] J. Dong, J. Shokes, R. Scott, D. Lynn, Modulating amyloid self-assembly and fibril morphology with Zn(II), J. Am. Chem. Soc. 128 (2006) 3540–3542.

[11] C. Ha, J. Ryu, C. Park, Metal ions differentially influence the aggregation and deposition of Alzheimer β-amyloid on a solid template, Biochemistry 46 (2007) 6118–6125.

[12] V. Minicozzi, F. Stellato, M. Comai, M. Dalla Serra, C. Potrich, W. Meyer-Klaucke, S. Morante, Identifying the minimal copper- and zinc-binding site sequence in amyloid-β peptides, J. Biol. Chem. 283 (2008) 10784–10792.

[13] V. Tougu, A. Tiiman, P. Palumaa, Binding of zinc (II) and copper (II) to the full-length Alzheimer amyloid β peptide, J. Neurochem. 104 (2008) 1249–1259.

[14] V. Minicozzi, S. Morante, G. Rossi, F. Stellato, N. Christian, K. Jansen, The role of metals in amyloid aggregation: experiments and ab initio simulations, Int. J. Quantum Chem. 108 (2008) 1992–2015.

[15] C. Sarell, S. Wilkinson, J. Viles, Substoichiometric levels of Cu^{2+} ions accelerate the kinetics of fiber formation and promote cell toxicity of amyloid- β from Alzheimer disease, J. Biol. Chem. 285 (2010) 41533–41540.

[16] S. Morante, G. Rossi, Metals in Alzheimer disease. A combined experimental and numerical approach, Adv. Alzheimer Res. 2 (2014) 100–147.

[17] E. De Santis, V. Minicozzi, O. Proux, G. Rossi, K. Silva, M. Lawless, F. Stellato, S. Saxena, S. Morante, Cu(II)-Zn(II) cross-modulation in amyloid-beta peptide binding: an x-ray absorption spectroscopy study, J. Phys. Chem. B 119 (2015) 15813–15820.

[18] M. Lovell, J. Robertson, W. Teesdale, J. Campbell, W. Markesbery, Copper, iron and zinc in Alzheimer disease senile plaques, J. Neurol. Sci. 158 (1998) 47–52.

[19] J. Ryu, K. Girigoswami, C. Ha, S. Ku, C. Park, Influence of multiple metal ions on β-amyloid aggregation and dissociation on a solid surface, Biochemistry 47 (2008) 5328–5335.

[20] A. Bush, W. Pettingell, G. Multhaup, M. d Paradis, J. Vonsattel, J. Gusella, K. Beyreuther, C. Masters, R. Tanzi, Rapid induction of Alzheimer Aβ amyloid formation by zinc, Science 265 (1994) 1464–1467.

[21] T. Burkoth, T. Benzinger, U. Voller, D. Morgan, D. Gregory, P. Thiyagarajan, R. Botto, S. Meredith, D. Lynn, Structure of the β-amyloid(10-35) fibril, J. Am. Chem. Soc. 122 (2000) 7883–7889.

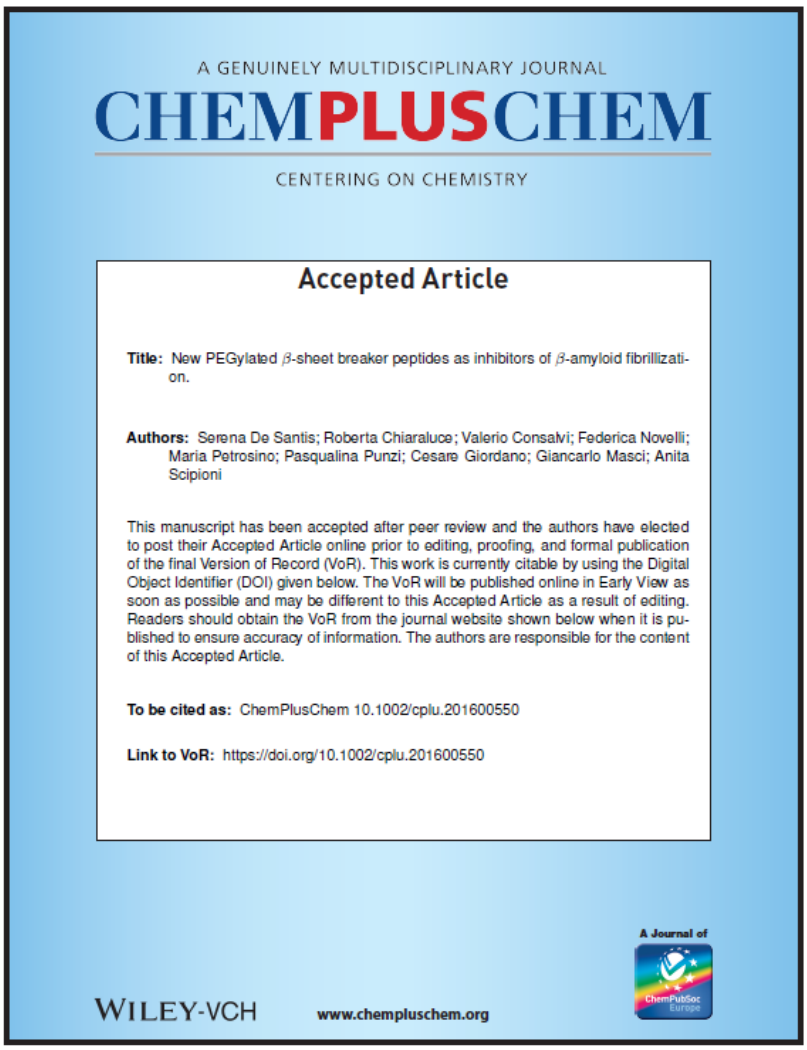
[22] D. Morgan, J. Dong, J. Jacob, K. Lu, R. Apkarian, P. Thiyagarajan, D.G. L., Metal switch for amyloid formation: insight into the structure of the nucleus, J. Am. Chem. Soc. 124 (2002) 12644–12645.

[23] J. Lee, T. Cole, R. Palmiter, S. Suh, J. Koh, Contribution by synaptic zinc to the gender-disparate plaque formation in human Swedish mutant APP transgenic mice, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 7705–7710.

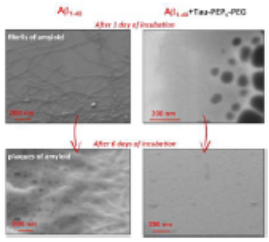
[24] D. Necas, P. Klapetek, Gwyddion: an open-source software for SPM data analysis, Cent. Eur. J. Phys. 10 (2012) 181–188.

[25] R. Woody, A. Dunker, Circular Dichroism and the Conformational Analysis of Biomolecules, Plenum Press, 1996.

[26] R. Banerjee, Effect of curcumin on the metal ion induced fibrillization of amyloid-β peptide, Spectrochim. Acta A 117 (2014) 798–800.


A GENUINELY MULTIDISCIPLINARY JOURNAL
CHEMPLUSCHEM
CENTERING ON CHEMISTRY
Accepted Article
Title: New PEGylated β-sheet breaker peptides as inhibitors of β-amyloid fibrillizati-on.
Authors: Serena De Santis; Roberta Chiaraluce; Valerio Consalvi; Federica Novelli; Maria Petrosino; Pasqualina Punzi; Cesare Giordano; Giancarlo Masci; Anita Scipioni
This manuscript has been accepted after peer review and the authors have elected to post their Accepted Article online prior to editing, proofing, and formal publication of the final Version of Record (VoR). This work is currently citable by using the Digital Object Identifier (DOI) given below. The VoR will be published online in Early View as soon as possible and may be different to this Accepted Article as a result of editing. Readers should obtain the VoR from the journal website shown below when it is published to ensure accuracy of information. The authors are responsible for the content of this Accepted Article.
To be cited as: ChemPlusChem 10.1002/cplu.201600550
Link to VoR: https://doi.org/10.1002/cplu.201600550
A Journal of
ChemPubSoc Europe
WILEY-VCH
www.chempluschem.org

PEGylated β-sheet breaker peptides: the PEGylated peptides H-Tau-Leu-Pro-Phe-Phe-Asp-NH_2 (Tau-PEP_5-PEG) and H-β-Ala-Leu-Pro-Phe-Phe-Asp-NH_2 self-assemble in nanoparticles capable of inhibiting β-amyloid fibrillization. PEGylation does not affect the anti-fibrillogenesis activity of the peptides *in vitro*. Importantly, electronic micrographs of β-amyloid incubated with the taurine-containing conjugate up to 6 days does not show fibrils proving the important role of the sulfonamide function against the amyloid aggregation.



PEGylated β-sheet breaker peptides as inhibitors of β-amyloid fibrillization.

Serena De Santis,[a] Roberta Chiaraluce,[b] Valerio Consalvi,[b] Federica Novelli,[a] Maria Petrosino,[b] Pasqualina Punzi,§[a] Fabio Sciubba,[a] Cesare Giordano,[c] Giancarlo Masci,[a] Anita Scipioni*[a]

Abstract: Three PEGylated β-sheet breakers peptides are designed as new inhibitors of β-amyloid fibrillization. The peptides Ac-Leu-Pro-Phe-Phe-Asp-NH_2, considered as the lead compound, and the hexamers in which taurine and β-alanine substitute the acetyl group, are conjugated to poly(ethylene glycol) obtaining self-assembling nanoparticles. The activity of the PEGylated peptides as inhibitors of amyloid fibrillization are tested *in vitro* using circular dichroism spectroscopy and scanning electron microscopy. The experimental results indicate that the PEGylation does not affect the ability of the β-sheet breakers peptides to inhibit fibrillogenesis *in vitro*. Moreover, microscopy images of β-amyloid incubated for 6 days with the taurine-containing conjugate suggest that this conjugate has the major anti-fibrillogenesis activity proving the important role of the sulfonamide function against the amyloid aggregation.

Introduction

Many degenerative pathologies, such as Parkinson's disease, type II diabetes, disorders caused by prions, Alzheimer's disease (AD), are related to amyloidosis, which are caused by the deposition of extracellular proteinaceous insoluble material with a reduced molecular weight.[1-2] This material accumulates in the form of amyloid aggregates that are deposited as plaques in the extracellular space, causing damage to organs involved and resulting in the onset of disease symptoms.[3,4] In particular, in AD, which is considered the most common form of dementia in adults,[5] the amyloid deposit consists of a 39-43 amino acid long peptide (Aβ), which in turn comes by cleavage of a larger amyloid precursor protein (APP).[6-8] Since at the origin of the fibril formation is the switching from α-helix to β-sheet of the amyloidogenic protein structure,[9] a rational pharmacological approach for preventing this process aims at developing compounds able to specifically stabilize the α-helical[10] and/or destabilize the β-sheet conformations.[11] In a recent past, the observation that short synthetic peptides, named β-sheet breaker peptides (BSBps), possess these specific capabilities, owing to a direct interaction with soluble oligomers or amyloid aggregates, addressed many scientific efforts with the hope of developing an effective therapeutic strategy.[12-15]

However, since the short half-life of peptides due to proteolytic degradation hinders their pharmacological use, many different strategies were adopted to reduce degradation and increase brain barrier penetration. In fact, unnatural or properly modified amino acids,[16-19] peptidomimetics[20-22] or retro-inverso peptides[23] with increased bioavailability were extensively studied. Furthermore, for this purpose, the potentiality of nanoparticles as β-sheet breaker (BSB) delivery systems was also exploited. Zhang *et al.* proposed poly(ethylene glycol)-poly(lactic acid) nanoparticles to load a novel BSBp (H102) and thus favoring its crossing through the blood brain barrier (BBB).[24] Similarly, Zheng *et al.* investigated the use of liposomes as H102 BSB delivery system.[25] As an alternative approach to protect nanoparticles from proteolysis *in vivo*, PEG has been extensively used due to its anti-biofouling properties and resistance to non-specific protein interactions and cell adhesion thus enhancing the bioactivity of the nanoparticles *in vivo*.[26-28] Thus, conjugation with poly(ethylene glycol) (PEG) has demonstrated to be an effective method to achieve increased stability and better pharmacokinetic properties of proteins, improving their therapeutic index and receiving market approval in several cases.[29] On the basis of such PEG properties and its wide use, Rocha and coworkers[30] conjugated PEG with the pentapeptide Ac-Leu-Pro-Phe-Phe-Asp-NH_2 (iAβ5p), which Adessi *et al.*[31] had shown to induce a relevant reduction in amyloid plaques in two different Alzheimer models and, for this reason, recognized as BSBp lead compound.[32] The authors demostrated that the PEGylation increases the stability of iAβ5p while retaining its capability of inhibiting amyloid Aβ fibrillogenesis.[30]

In our previous paper[33] we reported on the synthesis and fibrillogenesis inhibiting activity of new peptide derivatives related to the lead compound iAβ5p on which several strategic chemical modifications have been applied to improve its activity and metabolic stability. Adopting the $A\beta_{25-35}$ fragment as fibrillogenesis model, the activity of peptide derivatives H-Tau-Leu-Pro-Phe-Phe-Asp-NH_2 (Tau-PEP_5) and H-βAla-Leu-Pro-Phe-Phe-Asp-NH_2 (βAla-PEP_5), in which taurine or β-alanine were introduced in substitution of the acetyl group of iAβ5p, was tested *in vitro* by thioflavin T binding assay, circular dichroism (CD) spectroscopy, and scanning electron microscopy (SEM). Their ability to hinder the toxic effect of $A\beta_{25-35}$ *in vivo* was studied by monitoring the viability of human SH-SY5Y neuroblastoma cells and the prevention of superoxide anion radical release from BV2 microglial cells. Whilst βAla-PEP_5 was

[a] S. De Santis, F. Novelli, P. Punzi§, F. Sciubba, G. Masci, Prof. A. Scipioni
Dipartimento di Chimica Sapienza Università di Roma
P.le A. Moro,5 I-00185 Rome, Italy
E-mail: anita.scipioni@uniroma1.it
§ Current address: IRBM Science Park S.p.A., Via Pontina Km 30.600 - 00040 Pomezia (Roma), Italy

[b] R. Chiaraluce, Valerio Consalvi, M. Petrosino
Dipartimento di Scienze Biochimiche, Sapienza Università di Roma
P.le A. Moro,5 I-00185 Rome, Italy

[c] C. Giordano
Istituto di Biologia e Patologia Molecolari, CNR, Dipartimento di Chimica Sapienza Università di Roma
P.le A. Moro,5 I-00185 Rome, Italy

Supporting information for this article is given via a link at the end of the document.

shown to have a BSB activity comparable to that of iAβ5p, previous results suggested a remarkable role in the fibrillogenesis inhibitory activity of the sulphonamide function. Furthermore, Tau-PEP_5 shows a significant protective effect on cell viability, rescuing the cells from the toxicity exerted by the amyloid treatment. To improve their stability *in vivo*, we synthesized hybrid conjugates of the peptides Tau-PEP_5 and βAla-PEP_5 linking PEG (molecular weight 2000 Da, end-functionalized with an amine group) to the carboxylic group of the aspartic acid side chain (molecular structures in Figure 1). An analogous hybrid conjugate of iAβ5p was synthesized (Figure 1) to have a reference compound for the antifibrillogenis activity. The strategy to functionalize the aspartic acid side chain was chosen to preserve the integrity of the peptide backbone that is considered as the main responsible for the inhibition of Aβ fibrillogenesis.[10] Accordingly, exploiting a different conjugation site and PEG with lower molecular weight (2000 Da), we prepared a new PEGylated iAβ5p different from those investigated by Rocha *et al.* (3200 Da PEG linked to the *C*-terminal of aspartic acid and 5000 Da PEG to the *N*-terminal of the leucine.[30]

The aim of this work was the investigation of the self-assembling properties and the antifibrillogenesis activity of the new hybrid conjugates βAla-PEP_5-PEG, Tau-PEP_5-PEG and iAβ5p-PEG. Fluorescence spectroscopy and dynamic light scattering (DLS) were used to assess the aggregate features in water whilst the activity of the PEGylated BSBps as inhibitors of Aβ fibrillization was tested *in vitro* using CD spectroscopy and SEM.

Moreover, since it is widely accepted that a protein corona rapidly forms on the nanoparticle surface in a biological medium with important consequences for its transport,[34] the binding between the protein model bovine serum albumin (BSA) and the conjugated BSBps under aggregative conditions was investigated by steady-state fluorescence spectroscopy.

Figure 1 Molecular structures of the PEGylated peptides designed as possible β-sheet breakers: H-βAla-Leu-Pro-Phe-Phe-Asp(mPEG)-NH_2 (βAla-PEP_5-PEG), H-Tau-Leu-Pro-Phe-Phe-Asp(mPEG)-NH_2 (Tau-PEP_5-PEG) and Ac-Leu-Pro-Phe-Phe-Asp(mPEG)-NH_2 (iAβ5p). Conjugation with mPEG is obtained by linking mPEG-NH_2 with carboxylic group of the aspartic acid side chain.

Results and Discussion

Based on the proved activity as β-sheet breakers of the peptides Tau-PEP_5,[33] βAla-PEP_5[33] and iAβ5p,[32] the amphiphilic conjugates Tau-PEP_5-PEG, βAla-PEP_5-PEG and iAβ5p-PEG (molecular structures in Figure 1) were designed to self-assemble in core-shell nanoparticles with a high PEG surface density that should prevent the proteolysis of the peptide in biological environment. The BSBps, prepared by solid phase peptide synthesis, were conjugated in solution to a poly(ethylene glycol) end-functionalized with an amine group through the carboxylic group of the aspartic acid side chain. The purity of the conjugates was assessed by ESI-MS (Figures S1-S3) and monodimensional ^1H and ^{13}C NMR spectra and TOCSY (Total Correlation SpectroscopY) experiments as described in the Experimental Section. DOSY (Diffusion Ordered SpectroscopY) experiments were carried out on the PEGylated peptides to unambiguously demonstrate the PEG-peptide conjugation. In fact, DOSY techniques allow uniquely establishing if two molecules with significantly different molecular weight are present in solution as separate molecules or if they are linked, through their autodiffusion coefficients. The results are shown in Figure 2 where the DOSY spectra of βAla-PEP_5-PEG (panel A), Tau-PEP_5-PEG (panel B) and iAβ5p-PEG (panel C) are reported. The same value of the autodiffusion coefficient (highlighted by the solid lines) clearly shows that unreacted mPEG-NH_2 and BSBps are not present in the solution for all the conjugates. As an example of the sensitivity and selectivity of the DOSY technique, the DOSY spectrum of a mixture of Tau-PEP_5 and mPEG-NH_2 in 1/1 molar ratio is reported in Figure S4.

Before conjugation, the two moieties of the PEGylated peptides show different autodiffusion coefficients identified by the solid and dashed lines. In particular, the autodiffusion coefficients of the three conjugates are lower than those of the corresponding unreacted components as expected.

The self-assembling properties of the PEGylated BSBps in aqueous solution and their activity as inhibitors of $A\beta_{1-40}$ fibrillogenesis will be described in the following.

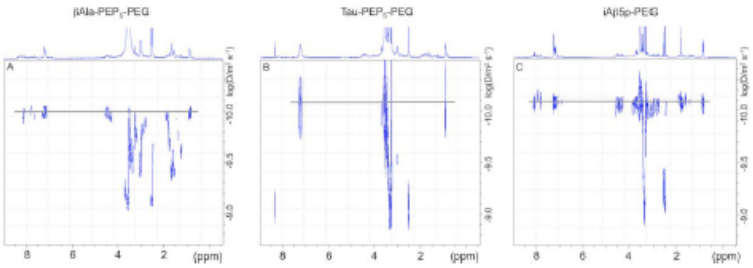


Figure 2. 2D DOSY experiments carried out in deuterated DMSO showing the auto-diffusion coefficients of βAla-PEP$_5$-PEG (panel A, D=9.99·10^{-10} m^2·s^{-1}), Tau-PEP$_5$-PEG (panel B, D=10.08·10^{-10} m^2·s^{-1}), and iAβ5p-PEG (panel C, D=10.08·10^{-10} m^2·s^{-1}). T = 298 K. The solid lines highlight the same values of the autodiffusion coefficient.

DLS measurements

As described in Materials and Methods, aqueous solutions of the conjugated BSBps were prepared by direct dissolution method at a concentration of 1 mg·ml^{-1}. Cumulant analysis of the autocorrelation function obtained at a scattering angle of 90° provides an average value of the hydrodynamic diameter, D_h, of about 210 nm for βAla-PEP$_5$-PEG and Tau-PEP$_5$-PEG and about 280 nm for iAβ5p-PEG. The corresponding normalized distribution functions of the average hydrodynamic diameter obtained by CONTIN analysis are reported in Figures 3 A-C. From these data iAβ5p-PEG nanoparticles (Figure 3C) appear to be slightly more polydisperse. DLS measurements were also carried out at different scattering angles, and the measured relaxation rates (Γ) were plotted against the square of the scattering vector *q* (data not reported). Straight lines passing through the origin were obtained indicating that the nanoparticles are spherical and slightly polydisperse.

The DLS measurements were repeated on solutions of βAla-PEP$_5$-PEG, Tau-PEP$_5$-PEG and iAβ5p-PEG dissolved in 2 ·10^{-2} mol·L^{-1} Tris·HCl, pH 7.4, the same experimental conditions adopted to monitor the inhibiting activity of amyloid Aβ$_{1-40}$ fibrillization by CD spectroscopy. Very similar hydrodynamic diameters and polydispersity are achieved for all the conjugated BSBps (data not shown).

Evaluation of the critical aggregation concentration (CAC) by fluorescence spectroscopy

The aggregation of amphiphilic molecules in aqueous media can be investigated by fluorescence spectroscopy using the pyrene as a probe. This method is based on the sensitivity of the pyrene emission spectrum to the polarity of the microenvironment sensed by the probe.[35-37] In particular, the CAC of the conjugates is determined by measuring the intensity ratio of the first over the third vibronic band ratio, I_3/I_1 (I_{384}/I_{373}), in 5·10^{-7} mol·L^{-1} aqueous solutions of pyrene at 298 K at conjugate concentration ranging from 2.6·10^{-6} to 3.2·10^{-4} mol·L^{-1} for βAla-PEP$_5$-PEG (Figure 4A), 6.0·10^{-6} to 3.7·10^{-4} mol·L^{-1} for Tau-PEP$_5$-PEG (Figure 4B) and 1.0·10^{-5} to 1.0·10^{-3} mol·L^{-1} for iAb5p-PEG (Figure 4C). Below the CAC, for low conjugate concentration, the value of I_3/I_1 ratio around 0.55 indicates that the probe is embedded in a hydrophilic environment. Increasing the conjugate concentration, the I_3/I_1 ratio increases suggesting the inclusion of the probe in a relatively hydrophobic microenvironment consequent to the aggregation of the hybrid peptide-polymer and the formation of the hydrophobic peptide core. This behavior, similar to that observed in equilibrium micelle formation, suggests that the PEGylated BSB peptides here investigated associate with a similar mechanism. The CAC is evaluated at the breaking point of the trend of I_3/I_1 versus the PEGylated concentration. A CAC of 2.0·10^{-6} mol·L^{-1} is obtained for βAla-PEP$_5$-PEG, 3.0·10^{-6} mol·L^{-1} for Tau-PEP$_5$-PEG and 3.0·10^{-4} mol·L^{-1} for iAβ5p-PEG. Such CAC values are not particularly low if compared to other peptide-PEG conjugates,[38] as expected considering the length of the hydrophobic peptide moiety. However, it could be a reasonable compromise between the resistance of the nanoparticles to the dilution, which should lead to increased circulation time, and the ability of the unimers to interact with the amyloid protein and act as fibrillogenesis inhibitors.

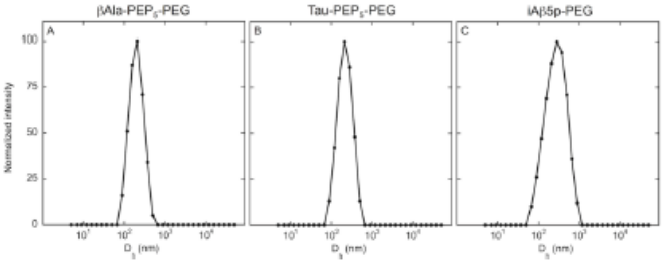


Figure 3. Normalized distribution function of the average hydrodynamic diameter as obtained by CONTIN analysis of DLS data at a scattering angle of 90°. Measurements were carried out on aqueous solution of βAla-PEP$_5$-PEG, Tau-PEP$_5$-PEG and iAβ5p-PEG at a concentration of 1 mg·mL^{-1}. T = 298 K.

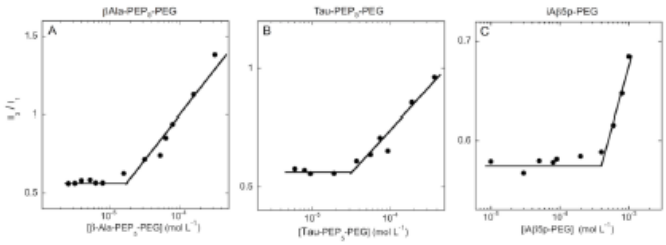


Figure 4. I_3/I_1 (I_{384}/I_{373}) ratio by fluorescence spectra of pyrene used as a probe as a function of βAla-PEP_5-PEG (panel A), Tau-PEP_5-PEG (panel B) and iAβ5p-PEG (panel C) concentration in aqueous solution. T = 298 K. The CAC is evaluated at the breaking point of the trends.

Evaluation by CD spectroscopy of *in vitro* inhibitory activity of $A\beta_{1-40}$ fibrillization.

CD spectroscopy was used to study *in vitro* the ability of the PEGylated βAla-PEP_5, Tau-PEP_5 and iAβ5p as inhibitors of $A\beta_{1-40}$ fibrillization.

When CD spectroscopy is adopted to evaluate the anti-fibrillogenesis activity of BSBps *in vitro*, the interference of CD spectra of the peptide moiety needs to be considered. In our case, however, the contributions of each conjugated peptide to the spectrum could be considered negligible as demonstrated by the normalized CD spectra reported in Figure S5.

The far UV CD spectral changes have been monitored as a function of time to follow the conformational changes of the synthetic $A\beta_{1-40}$ in solution in the absence or presence of 10 molar excess of βAla-PEP_5-PEG, Tau-PEP_5-PEG and iAβ5pPEG. Typically, lyophilized $A\beta_{1-40}$ $2.2 \cdot 10^{-5}$ $mol \cdot L^{-1}$ undergoes the far UV CD spectral changes reported in Figure S6.

The global changes in the spectral region 200–250 nm were analyzed by SVD (Single Value Decomposition) that indicates that only two spectral components contribute to the far-UV CD spectra. The most significant singular values are reported in Table S1. All the other singular values are well below 10 % of the largest singular value and progressively become negligible. The far UV CD changes of $A\beta_{1-40}$ observed over 27 h at pH 7.4 in $2 \cdot 10^{-2}$ $mol \cdot L^{-1}$ Tris·HCl can be summarized in two main events: an increase in molar ellipticity at around 220 nm, characteristic of *β*-sheet structure and an overall decrease in molar ellipticity mainly at around 200 nm, a typical turns-containing structure region (Figure S6).[39] Two main spectral components were identified after SVD of the far UV CD spectra of $A\beta_{1-40}$ alone monitored for 27 h. The changes of the first spectral component can be mainly referred to the appearance of *β*-sheet structure, as revealed by the increase of negative molar ellipticity at around 218 nm (Figure S7 A) accompanied by the disappearance of turns-containing structures as the decreasing of the second spectral component at around 200 nm suggests (Figure S7 B).

The inhibitory activity *in vitro* of βAla-PEP_5-PEG, Tau-PEP_5-PEG and iAβ5p-PEG was tested by monitoring their efficacy in preventing and/or interfering with the secondary structure conformational transition of $A\beta_{1-40}$ preliminary to the formation of fibrils and visible aggregates. The far UV CD spectra of $A\beta_{1-40}$ monitored in the presence of the conjugates over 27 h (Figures S8, S10 and S12) and analysed by SVD revealed two main spectral components, centred at around 203 and 218 nm, similarly to what observed for $A\beta_{1-40}$ alone with an inversion of the relative order of importance with respect to the control in the absence of inhibitors, as deduced from the respective S values of the first and second spectral component. The changes at 203 nm of the first spectral component (Figures S9 A and S11 A) and at 218 nm of the second spectral component (Figures S9 B and S11 B) in the presence of βAla-PEP_5-PEG and iAβ5p-PEG, respectively, are significantly reduced in amplitude in comparison with those of $A\beta_{1-40}$ alone. The time dependence of these spectral changes is shown in Figure 5 for $A\beta_{1-40}$ alone and in the presence of the conjugates. The changes of molar ellipticity at 203 nm of the first spectral component of $A\beta_{1-40}$ alone as a function of time shows an amplitude of about $1.2 \cdot 10^{4}$ $deg \cdot cm^{2} \cdot dmol^{-1}$. In the presence of βAla-PEP_5-PEG and iAβ5p-PEG, the amplitude of the changes at 203 nm is dramatically reduced with respect to that observed for $A\beta_{1-40}$ alone and decreases of about one order of magnitude (Figure 5 A). Similarly, the amplitudes of the changes at 218 nm of the second spectral component in the presence of βAla-PEP_5-PEG and iAβ5p-PEG appear to be decreased in comparison to those observed for $A\beta_{1-40}$ alone (Figure 5 B). These results suggest an interaction between the amyloid and βAla-PEP_5-PEG and iAβ5p-PEG that prevents the transition towards the β-sheet conformation of $A\beta_{1-40}$, at least along the observation period of about 30 hours. In the presence of Tau-PEP_5-PEG, the maximal amplitude for the changes of the first and second spectral components is observed over a different spectral range (Figure S13): at 213 nm for the first spectral component and at 203 and 218 nm for the second spectral component. The amplitudes for the spectral components at 213 and 203 nm are comparable to those of $A\beta_{1-40}$ alone even though sensitively reduced at 218 nm (Figures 13 A and B). These results may suggest the formation of β-like structures as a result of the interaction between $A\beta_{1-40}$ and the conjugate Tau-PEP_5-PEG that, however, do not evolve towards fibrillization. As a matter of fact, SEM analysis carried out on the amyloid in the presence of Tau-PEP_5-PEG demonstrates a strong inhibition of fibrillogenesis from such a conjugate up to 6 days of incubation, as will be shown in the next sub-section.

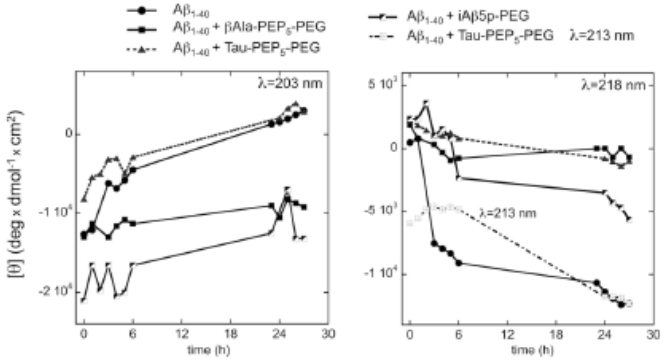


Figure 6. The changes of molar ellipticity of $2.2 \cdot 10^{-6}$ mol·L^{-1} $A\beta_{1\text{-}40}$ as a function of time for the spectral components measured at 203 nm (A) and 218 nm (B) were monitored in the presence of 10-fold excess of conjugates in $2 \cdot 10^{-5}$ mol·L^{-1} Tris-HCl, pH 7.4, T=293 K. The spectral component measured at 213 nm for $A\beta_{1\text{-}40}$ in the presence of Tau-PEP5-PEG is also reported in panel B. The spectral components were identified after SVD of the far UV CD spectra of $A\beta_{1\text{-}40}$ alone and in the presence of the conjugates. The lines are drawn as a guide to the eye.

Evaluation by scanning electron microscopy of *in vitro* inhibitory activity of $A\beta_{1\text{-}40}$ fibrillization

The effect of βAla-PEP$_5$-PEG, Tau-PEP$_5$-PEG and iAβ5p-PEG on $A\beta_{1\text{-}40}$ fibrillization was investigated also by SEM analysis.

In parallel with the experiments carried out with CD spectroscopy, samples of $2.2 \cdot 10^{-5}$ mol·L^{-1} solutions of $A\beta_{1\text{-}40}$ in the absence or in the presence of $2.2 \cdot 10^{-4}$ mol·L^{-1} PEGylated BSBps were also deposited on silicon support after 24 h of incubation as described in Materials and Methods.

SEM analysis shows that $A\beta_{1\text{-}40}$, reported as control, forms branched fibrils after 24 h of incubation at 293 K, as expected (Figure 6 A). The absence of fibrillar aggregates when $A\beta_{1\text{-}40}$ is incubated with the conjugates indicates that βAla-PEP$_5$-PEG (Figure 6 B), Tau-PEP$_5$-PEG (Figure 6 C) and iAβ5p-PEG (Figure 6 D) strongly inhibit fibrillogenesis.

The impact of the conjugates on $A\beta_{1\text{-}40}$ fibrillization after 6 days of incubation was shown on the SEM micrographs reported in Figure 7. SEM analysis shows that $A\beta_{1\text{-}40}$, reported as control, forms extended plaques (Figure 7A). The presence of the conjugates Tau-PEP$_5$-PEG (Figure 7C) strongly inhibits fibrillogenesis while rare fibrils are evident in the presence of iAβ5p-PEG (Figure 7D). The SEM image of $A\beta_{1\text{-}40}$ in the presence of βAla-PEP$_5$-PEG, instead, shows reduced amounts of fibrils demonstrating that such a conjugate only delays the aggregation.

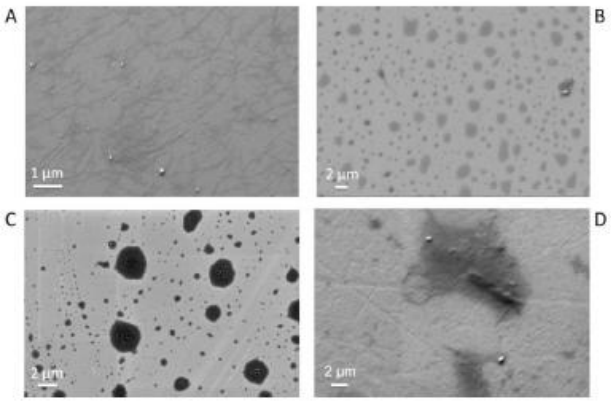


Figure 6 SEM micrographs of $2.2 \cdot 10^{-6}$ mol·L^{-1} $A\beta_{1-40}$ deposited from solutions of Tris HCl, pH 7.4, after an incubation of 24 h at T=293 K: alone (A); in the presence of βAla-PEP_8-PEG (B), Tau-PEP_8-PEG; (C) and IAβ5p-PEG (D). Molar ratio of PEGylated peptides/$A\beta_{1-40}$ = 10.

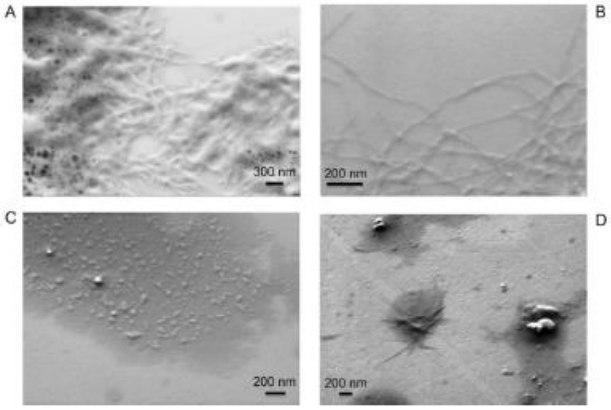


Figure 7 SEM micrographs of $2.2 \cdot 10^{-6}$ mol·L^{-1} $A\beta_{1-40}$ deposited from solutions of Tris HCl, pH 7.4, after an incubation of 6 days at T=293 K: alone (A); in the presence of βAla-PEP_8-PEG (B), Tau-PEP_8-PEG; (C) and IAβ5p-PEG (D). Molar ratio of PEGylated peptides/$A\beta_{1-40}$ = 10.

Evaluation of nanoparticles-BSA binding by steady-state fluorescence spectroscopy

The affinity of BSA for the conjugated peptides was investigated using static fluorescence quenching since this method allows evaluating the equilibrium binding constant between BSA and nanoparticles.[40,41] BSA, an abundant plasma protein and important as model protein, contains tryptophan residues, which allows it to absorb and fluoresce at characteristic wavelengths. Such spectroscopic features are exploited to investigate the

binding to proteins using fluorescence measurements. In fact, the local unfolding of BSA proteins into less compact structures can change the exposure of the tryptophan (Trp) residues to the aqueous environment leading to lower fluorescence intensity. Adopting Stern-Volmer analysis, it is possible to evaluate the binding constant of the association complex between the protein and possible ligands that perturb its tertiary structure (see the Experimental Section for details). The fluorescence spectra of BSA in the presence of increasing concentrations of the three PEGylated peptides were recorded. Stern-Volmer plots of Trp quenching measured at the

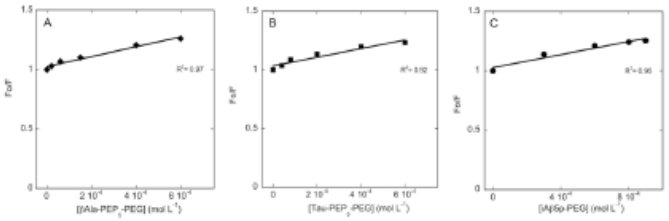


Figure 8 Stern-Volmer plots obtained for quenching of $7 \cdot 10^{-7}$ mol·L^{-1} BSA in Tris·HCl ($2 \cdot 10^{-2}$ mol·L^{-1}, pH 7.4) in the presence of βAla-PEP$_5$-PEG (A) and Tau-PEP$_5$-PEG (B); of $7 \cdot 10^{-6}$ mol·L^{-1} BSA in Tris·HCl ($2 \cdot 10^{-2}$ mol·L^{-1}, pH 7.4) in the presence of iASp-PEG (C). Measurements were carried out in aggregative conditions of the conjugates.

wavelength of 345 nm after adding increasing concentrations of the conjugates were shown in Figure 8 and the Stern-Volmer constants were determined for the three conjugates. As evident by the Stern-Volmer plots, the presence of the PEGylated peptides in aggregative conditions causes little quenching effects on the fluorescence of Trp of BSA. As a matter of fact, the Stern-Volmer constants, namely the binding constants of the complexes BSA/BSBp-PEG, are little than $5 \cdot 10^{2}$ mol^{-1}·L for all the conjugates (see the values reported in Table 1).

Table 1 *Stern-Volmer constants (K_{SV}) of the complexes BSA-PEGylated peptides in Tris-HCl, pH 7.4, T=298 K.*

BSA + BSBp-PEG	$K_{SV} \cdot 10^{-2}$ (mol^{-1}·L)
βAla-PEP$_5$-PEG	4.2
Tau-PEP$_5$-PEG	3.7
iAβ5p-PEG	2.8

Such values of the Stern-Volmer constants suggest that BSA undergoes little changes of the folded conformation and that the interactions with the conjugates are weak, at least in the conditions adopted during these experiments that are carried out in the presence of the PEGylated peptides at concentrations greater than their corresponding CAC.

Conclusions

The results reported in this paper show the ability of the conjugated BSB peptides to self-assemble in water solutions giving rise to PEGylated nanoparticles with plausible antibiofouling properties that could protect the peptide moiety, segregated in the nanoparticle core, from enzymatic proteolysis. Importantly, the experimental results indicate that after conjugation with PEG, the BSB peptides do not change their capability of inhibiting fibrillogenesis *in vitro*. Furthermore, SEM images of Aβ$_{1-40}$ incubated for 6 days with Tau-PEP$_5$-PEG suggest that such a conjugate has the major anti-fibrillogenesis activity. These results highlight the important role played by the sulfonamide function against the aggregation of the amyloid, in agreement with Giordano *et al.*[33] that showed the higher efficiency as BSB of Tau-PEP$_5$ compared to βAla-PEP$_5$ and iAβ5p. The improved efficiency of Tau-PEP$_5$-PEG could be due to the increased acidity of sulphonamide function that makes it a better hydrogen bond donor compared to the peptide amide moiety. The PEGylation of Tau-PEP$_5$ does not perturb sensitively the hydrogen bond pattern of the peptide region and the binding capability of Tau-PEP$_5$-PEG towards Aβ$_{1-40}$; as a consequence, the efficiency in preventing the aggregation of the amyloid is preserved.

Experimental Section

Materials

Aβ$_{1-40}$ was purchased from the Poly-Peptide Laboratories France SAS. Acetonitrile (MeCN, Chromasolv Plus, for HPLC), dichloromethane (DCM), *N*,*N*-dimethylformamide (DMF), 1-methyl-2-pyrrolidone (NMP), diethyl ether, hexane, trifluoroacetic acid (TFA), diisopropycarbodiimide (DIC), *N*,*N*-diisopropylethylamine (DIPEA), *N*,*N*,*N*',*N*'-tetramethyl-O-(benzotriazol-1-yl)uronium tetrafluoroborate (TBTU), thionyl cloride ($SOCl_2$), sodium azide (NaN_3), triethylamine (TEA), 4-(dimethylamino)pyridine (DMAP), *n*-butanol (*n*-BuOH), *N*-methylmorpholine (NMM) were purchased from Aldrich as peptide synthesis grade and used as provided. Palladium on activated carbon (10 wt% Pd/C) and spectrograde pyrene were from Fluka. DCM and DMF were dried over 4 Å molecular sieves and stored under argon. Piperidine (Fluka) was distilled before use. Rink amide (aminomethyl) polystyrene resin (200-400 mesh, 1% divinylbenzene, loading level 0.8-1.1 mmol·g^{-1}) was from Novabiochem. *N*-Fmoc-amino acids (Fmoc-L-Leu-OH,

Fmoc-L-Pro-OH), Fmoc-L-Phe-OH, Fmoc-L-Asp(OtBu)-OH, and Fmoc-βAla-OH) were from Aldrich.
Tris(hydroxymethyl)aminomethane hydrochloride (Tris·HCl, pH 7.4) was from Aldrich. Deuterated solvents for NMR spectroscopy were used as obtained from Aldrich. Methoxy poly(ethylene glycol) end-functionalized with an amine group ($mPEG_{45}$-NH_2) M_n = 2000, M_w/M_n = 1.03, Aldrich) was dried in a vacuum oven at 80 °C for 2 h before use. Thin Layer Chromatography (TLC) was performed on silica gel Merck 60 F254 plates.

Synthesis of the BSBps Fmoc-βAla-PEP_5, Fmoc-Tau-PEP_5 and iAβ5p.

iAβ5p and its analogues Fmoc-βAla-PEP_5 and Fmoc-Tau-PEP_5 were prepared manually by conventional solid-phase peptide synthesis (SPPS) on Rink-amide resin (65-mmol scale, 100 mg).[42] Fmoc-AA-OH, DIC and HOBt (3 eq, respectively) in NMP (5 mL) are used for couplings and the reaction mixtures were shaken mechanically for 3 h at room temperature. Fmoc-Tau-Cl, required for the preparation of the Tau-containing derivative Fmoc-Tau-PEP_5, was prepared according to Brouwer and co-workers[43] and coupled to the resin-bound protected pentapeptide in DCM in the presence of NMM (7 eq). The completion of the coupling reactions was checked by the ninhydrin-Kaiser test, and the removal of Fmoc was effected with a 20% piperidine solution in DMF (2 × 5 mL) for 15 min. For the cleavage from resin of the Fmoc protected peptide, a solution of aqueous 95% TFA (2 mL for 3 h) was used and the peptide was precipitated from the cleavage solution with peroxide-free dry diethyl ether/hexane 4/6 (v/v) at 0 °C. The precipitate was repeatedly washed with ether/hexane 4/6 (v/v), redissolved with water/dioxane 1/1 (v/v), and lyophilized. The crude peptide was purified by RP-HPLC on a Waters 600E liquid chromatography system by using a Waters μBondapak C-18 column (1.9 × 30 cm, 5 μm, 300 Å) for a semipreparative scale with elution at 8 mL min^{-1} by a linear gradient of 10–60% MeCN in 0.1% aqueous TFA in 30 min. The final peptide purity was >97% by analytical HPLC (Waters μBondapak C-18 column, (0.39 × 30 cm, 5 μm, 300 Å).
All the BSBps were identified by electrospray ionization mass (ESI-MS) spectrometry. ESI-MS positive spectra were reported in Figures S14-S16 with the relative assignments of the signals. Results of ESI-MS spectra of the peptides in methanol are listed below. Fmoc-Tau-PEP_5 and Fmoc-βAla-PEP_5 were further characterized by NMR spectroscopy. Monodimensional ^1H and ^{13}C NMR and TOCSY spectra of Fmoc-Tau-PEP_5 in DMSO-d_6 and Fmoc-βAla-PEP_5 in deuterated methanol allowed the assignment of most the signals (Table S2 and S3 of Supplementary Material).
Fmoc-Tau-PEP_5 Yield: 34 %
ESI-MS for $C_{50}H_{86}N_7O_{11}S$ (calculated 965.40), m/z (%): 988.17 (100) $[M+Na]^+$, 1004.16 (10) $[M+K]^+$, 1953.24 (6.3) $[2M+Na]^+$
Fmoc-βAla-PEP_5 Yield: 30 %
ESI-MS for $C_{51}H_{89}N_7O_{10}$ (calculated 929.43), m/z (%): 930.33 (14) $[M+H]^+$, 952.28 (100) $[M+Na]^+$, 968.29 (20) $[M+K]^+$.
iAβ5p Yield: 54 %
ESI-MS for $C_{35}H_{46}N_8O_8$ (calculated 678.28), m/z (%): 679.34 (100) $[M+H]^+$, 701.32 (17) $[M+Na]^+$.

Synthesis of Tau-PEP_5-PEG, βAla-PEP_5-PEG and iAβ5p-PEG. General Procedure

The PEGylation reactions were carried out under argon using 1 equivalent of each BSBp, 1.2 equivalents of TBTU, 1.2 equivalents of DMAP, 1.2 equivalents of TEA and 1.2 equivalents of $mPEG_{45}$-NH_2 in a volume of dry DMF to obtain a final concentration of $6.5 \cdot 10^{-3}$ mol L^{-1} in Fmoc-Tau-PEP_5 and Fmoc-βAla-PEP_5 whilst the PEGylation reaction of iAβ5p was carried out at a concentration of $1.8 \cdot 10^{-3}$ mol L^{-1}. The mixture was stirred for 48 h at room temperature and the reaction was monitored by TLC analysis (*n*-BuOH/AcOH/H_2O 6/2/2 v/v). The solution was partially concentrated under reduced pressure and exhaustively dialyzed against DMF for 2 days (Spectra/Por® 6 MWCO 1000 Da, SpectrumLabs). The dialyzed mixture was evaporated to dryness under reduced pressure to give a solid that was redissolved in water and passed through an Amberlite IR-120 (H) ion-exchange resin to eliminate the unreacted $mPEG_{45}$-NH_2. The fractions containing the expected product were pooled and lyophilized providing Fmoc-Tau-PEP_5-PEG, Fmoc-βAla-PEP_5-PEG and iAβ5p-PEG as white solids.
To remove the Fmoc protecting group, Fmoc-Tau-PEP_5-PEG and Fmoc-βAla-PEP_5-PEG were dissolved in 20% piperidine in DMSO and stirred for 1 h at room temperature. The conjugates were precipitated from the cleavage solution with peroxide-free dry diethyl ether at 0 °C. The precipitate was repeatedly washed with ether, redissolved with water/dioxane 1/1 and lyophilized.

Fmoc-Tau-PEP_5-PEG: Yield: 37 %
PEP_5-PEG: Yield: 51 %

Fmoc-βAla-PEP_5-PEG: Yield: 33 %
βAla-PEP_5-PEG: Yield: 49 %

iAβ5p-PEG: Yield: 86 %

The final conjugates were characterized by ESI-MS spectrometry and mono and bidimentional NMR spectroscopy. ESI-MS spectra of Fmoc-Tau-PEP_5-PEG, Fmoc-βAla-PEP_5-PEG and iAβ5p-PEG are shown in Figures S1-S3, respectively. NMR spectra were recorded by dissolving about 10 mg of each PEGylated peptides in 0.6 mL of DMSO-d_6. Monodimensional ^1H and ^{13}C NMR spectra and 2D TOCSY experiments allowed the assignment of most signals confirming the removal of the Fmoc protecting group and purity of the PEGylated BSBps. The 2D TOCSY spectra of Tau-PEP_5-PEG, βAla-PEP_5-PEG and iAβ5p-PEG are shown in Figures S17-S19, respectively; the corresponding signal assignments for each conjugate are reported in Table S4-S6.
DOSY experiments proved the integrity and purity of the three PEGylated BSBps (Figure 2). Such experimental results have been deeply discussed in the paragraph of Results and Discussion.

Preparation of the Nano-Aggregate Solutions

Aqueous solutions of PEGylated BSBps were prepared by direct dissolution method since such a method is prevalently adopted for relatively water-soluble conjugates where the molecular weight of the hydrophilic block exceeds that of the hydrophobic one.[44]
The conjugate is added to milli-Q water or 20 mM Tris·HCl solution (pH 7.4) and left under stirring overnight to allow the

self-assembly into nanoparticles. The resulting solution was sonicated in ice bath for about 60 minutes (UTA 18 sonicator, 280 W, FALC INSTRUMENTS) to prevent the formation of nanoparticles clusters and achieve stable dynamic light scattering measurements.

DLS measurements

DLS data were obtained with a Brookhaven Instruments Corp.BI-200SM goniometer equipped with a BI-9000AT digital correlator using a solid-state laser (125 mW, λ = 532 nm). Measurements of scattered light were made at a scattering angle θ of 90° and at temperature of 298.0±0.1 K (temperature control JULABO F10). Each experiment was repeated two or more times in a time range of 5-20 min. Cumulants analysis or CONTIN were used to fit the data. Solutions prepared as described in the previous section were filtered with 1 μm filters (Durapore, Millipore, USA) before DLS measurements.

Fluorescence Spectroscopy

Steady-state fluorescent spectra were measured using a Cary Eclipse (Varian) spectrofluorimeter equipped with a thermostated cell compartment and a SUPRASIL® quartz cell (10 × 4 mm). The fluorescence emission spectra (350-450 nm) of pyrene were measured using an excitation wavelength of 335 nm. The intensity ratio (I_3/I_1) of the third (I_3) and the first (I_1) vibronic bands of the emission spectrum of pyrene, at 384 and 373 nm, respectively, was used to monitor the formation of hydrophobic microdomains and evaluate the CAC.[35-37] A stock solution of pyrene ($5\cdot10^{-7}$ mol·L^{-1}) in water was prepared by adding 50 μL of a $5\cdot10^{-4}$ mol·L^{-1} pyrene solution in ethanol in a 50 mL volumetric flask, evaporating ethanol by a nitrogen flow and bringing to volume with milli-Q water. The solutions were kept at room temperature for 24 h to reach the solubilization equilibrium of pyrene in the aqueous phase.
For the determination of the CAC, the conjugates were dissolved in the aqueous $5\cdot10^{-7}$ mol·L^{-1} pyrene solution. Concentrations in the range $3.7\cdot10^{-4}$ -$6\cdot10^{-6}$ mol·L^{-1} for Tau-PEP$_5$-PEG, $3.2\cdot10^{-4}$ - $2.6\cdot10^{-6}$ mol·L^{-1} for βAla-PEP$_5$-PEG and $1.0\cdot10^{-3}$ -$1.0\cdot10^{-5}$ mol·L^{-1} for iAβ5p-PEG were analyzed.
The solutions were left under stirring overnight at about 37 °C before measurements in order to obtain a constant value of the emission intensity. Measurements were carried out in triplicates.
The evaluation of the binding constants between BSA and the conjugates were achieved by measuring the fluorescence quenching of the tryptophan residues of BSA. The emission of BSA was measured at a constant concentration in the presence of increasing PEGylated peptide concentrations in TRIS-HCl $2\cdot10^{-2}$ mol·L^{-1} at pH 7.4. The conjugates, at concentrations ≥ CAC to ensure the aggregative conditions, were incubated in protein solutions overnight at 4 °C in Teflon vials to avoid the absorption of the protein. Intrinsic fluorescence spectra were recorded between 300 and 450 nm with an excitation wavelength of 290 nm where the contributions of phenylalanine and tyrosine residues are negligible. Excitation and emission slits were 5 nm. The fluorescence quenching is described by the Stern-Volmer equation:[40,41]

$$\frac{F_0}{F} = 1 + K_{SV}[Q]$$

where F_0 and F are the fluorescence intensities of Trp residues of BSA in the absence and in the presence of the conjugates that act as quenchers (Q). K_{SV} is the Stern-Volmer quenching constant that becomes the binding constant if a static quenching is assumed. Stern-Volmer analysis, obtained reporting F_0/F as a function the increasing concentrations of the PEGylated peptides, provided the Stern-Volmer constants.
The concentration of iAβ5p-PEG was varied in the range $3\cdot10^{-4}$ - $9\cdot10^{-4}$ mol·L^{-1} while BSA was kept constant at $7\ 10^{-6}$ mol·L^{-1}. For βAla-PEP$_5$-PEG and Tau-PEP$_5$-PEG fluorescence measurements were carried out in $7\ 10^{-7}$ mol·L^{-1} BSA while the concentration range was between $2\cdot10^{-5}$ - $6\cdot10^{-4}$ mol·L^{-1} for βAla-PEP$_5$-PEG and $4\cdot10^{-5}$ - $6\cdot10^{-4}$ mol·L^{-1} for Tau- PEP$_5$-PEG. Fluoresceonce spectra were recorded at the temperature of 298 K.

CD Spectroscopy

CD spectra were performed in the far UV spectral region (250-190 nm) with a JASCO J 720 spectropolarimeter equipped with a Peltier device for temperature control to monitor the capability of the PEGylated BSBps of inhibiting the fibril formation of Aβ$_{1-40}$. The activity of inhibiting amyloid Aβ$_{1-40}$ fibrillogenesis was monitored recording CD spectra in a 0.1 cm path length quartz cell at 293 K. The results are expressed as the mean residue ellipticity assuming a mean weight of 110 per amino acid residue. Aliquots of lyophilized Aβ$_{1-40}$ were dissolved in $2\cdot10^{-2}$ mol·L^{-1} Tris·HCl, pH 7.4, at $2.2\cdot10^{-5}$ mol·L^{-1} final concentration in the absence or in the presence of $2.2\cdot10^{-4}$ mol·L^{-1} PEGylated BSBps. The far UV CD spectral changes were monitored, at increasing times up to 27 h. The CD spectra of the conjugated peptides recorded for control did not change in the same range of time. Far-UV CD spectra of Aβ$_{1-40}$ alone or in the presence of the conjugates were analyzed by the singular value decomposition algorithm (SVD)[45,46] using the software MATLAB (MathWorks, South Natick, MA). SVD is useful to find the number of independent components in a set of spectra and to remove the high frequency noise and the low-frequency random error. CD spectra in the 200-250-nm region (0.5 nm sampling interval) were placed in a rectangular matrix **A** of *n* columns, one column for each spectrum collected at different time. The **A** matrix is decomposed by SVD into the product of three matrices: **A**= **U** x **S** x **V**^{T} where **U** and **V** are orthogonal matrices and **S** is a diagonal matrix. The columns of **U** matrix contain the basis spectra, and the columns of the **V** matrix contain the time dependence of each basis spectrum. Both **U** and **V** columns are arranged in terms of their decreasing order of the relative weight of information, as indicated by the magnitude of the singular values in **S**. The diagonal **S** matrix contains the singular values that quantify the relative importance of each vector in **U** and **V**. An important feature of SVD analysis is that the signal-to-noise ratio is very high in the earliest columns of **U** and **V** and that the random noise is mainly accumulated in the latest **U** and **V** columns. The wavelength-averaged spectral changes as a function of time are represented by the columns of matrix **V**.

NMR Spectroscopy

One-dimensional NMR spectra were obtained on a Bruker AC 300 and a Bruker Avance 400 spectrometer (Bruker Spectrospin, Karlsruhe, Germany). The Bruker AC 300 operates at 300.13 and 75.47 MHz for ^{1}H and ^{13}C, respectively. The Bruker Avance 400 spectrometer operates at 9.4 T at 298 K by

using a 90° detection pulse and acquiring the FIDs into 64 kpoints; the spectral width was 6009.15 Hz and the relaxation delay was set to 9.5 s in order to achieve a 15 s total acquisition time to avoid relaxation effects. Monodimensional ^{13}C NMR spectra were performed by collecting 16000 scans for each spectrum at 298 K acquiring the FIDs into 32 kpoints, the spectral width was 25000 Hz and the relaxation delay was set to 2 s. Homonuclear ^{1}H - ^{1}H TOCSY experiments were performed in order to allow signal assignment. TOCSY experiments were acquired with spectral width of 6009.15 Hz in both dimensions, a data matrix of 8 K × 256 points, mixing time of 60 ms and relaxation delay of 2 s.

Pseudo-bidimensional ^{1}H homonuclear DOSY (Diffusion Ordered SpectroscopY) experiments were performed in order to assess the translational mobility of the molecules in solution. DOSY experiments were acquired at 298 K, spectral width of 15 ppm, 64 kpoints, acquisition time of 5.5 s, repetition time of 2 s, 16 scans with a diffusion time of 150 ms, a gradient length of 2.5 ms and 32 gradient points.

Chemical shifts were reported in parts per million (ppm) related to residual undeuterated solvent as an internal standard. The spectra were recorded dissolving about 10 mg of the samples in 0.6 mL of the suitable deuterated solvent.

ESI-MS Spectrometry

ESI-MS spectrometry was performed on a Q-Tof Micro spectrometer (Micromass, now Waters) equipped with an electron spray ionization (ESI) source, in the positive ion mode and data were analyzed using the MassLynx software (Waters). Mass spectra were measured in the range of *m/z* 200-2000.

SEM analysis

SEM micrographs were recorded with a ZEISS AURIGA 405 microscope (0.5-30 keV, 10^{-10} mbar). All the samples were prepared by deposition of 10 μL of the dispersions in water on a silicon support. Then, by lateral contact with a small piece of filter paper most of the solution was removed in order to leave a thin film of liquid on the support. The solvent was evaporated with a gentle flow of air. Finally, the sample was dried overnight under vacuum at room temperature. Samples of $2.2 \cdot 10^{-5}$ mol·L^{-1} solutions of $A\beta_{1-40}$ in the absence or in the presence of $2.2 \cdot 10^{-4}$ mol·L^{-1} PEGylated BSBps were deposited on silicon support after 24 h and 6 days of incubation to monitor the inhibition activity of the three conjugates towards $A\beta_{1-40}$ aggregation.

Keywords: β-sheet breakers peptides, hybrid peptide conjugates, PEGylation, self-assembly, nanoparticles

Acknowledgements: Thanks are due to CNIS Laboratory of Sapienza Università of Rome and Dr. Francesco Mura for providing SEM micrographs. Funding from Ateneo La Sapienza 2015 is also acknowledged.

[1] K.P. Kepp. *Chem. Rev.* **2012**, *112*, 5193-5239.
[2] Y. Takahashi, A. Ueno, H. Mihara, *ChemBioChem* **2002**, *3*, 637- 642.
[3] J. Greenwald, R. Riek. *Structure*, **2010**, *18*, 1244-1260.
[4] C. Soto. *J. Mol. Med.*, **1999**, 77, 412-418.
[5] A. Wimo, B. Winblad, H. Aguero-Torres, E. von Strauss. *Alzheimer Dis. Assoc. Dis.* **2003**, *17*, 63-67.
[6] R. Jakob-Roetne, H. Jacobsen. *Angew. Chem. Int. Ed.* **2009**, *48*, 3030-359.
[7] F. Chiti, C,M. Dobson. *Annu. Rev. Biochem.* **2006**, *75*, 333-366.
[8] V. John, L.H. Latimer, J.S. Tung, M.S. Dappen, In *Annu. Rep. Med. Chem. Academic press*, Athena Neurosciences, Inch. South San Francisco, California 94080, **1987**, 32, 11-20.
[9] C. Soto, E.M. Castano, B. Frangione, N.C. Inestrosa, *J. Biol. Chem.* **1995**, *270*, 3063-3067.
[10] C. Soto. *Mol. Med. Today* **1999**, *5*, 343-350.
[11] L.O. Tjernberg, J. Naslund, F. Lindqvist, J. Johansson, A.R. Karlstrom, J. Thyberg, L. Terenius, C. Nordstedt, *J. Biol. Chem.* **1996**, *271*, 29525-29528.
[12] C. Adessi, C. Soto. *Drug. Dev. Res.* **2002**, *56*, 184-193.
[13] A. Francioso, P. Punzi, A. Boffi, C. Lori, S. Martire, C. Giordano, M. D'Erme, L. Mosca. *Bioorg. Med. Chem.* **2015**, *23*, 1671-1683.
[14] L.D. Estrada, C. Soto *Curr. Pharm. Des.* **2006**, *12*, 2557-2567
[15] C. Giordano, A. Masi, A. Pizzini, A. Sansone,V. Consalvi, R. Chiaraluce, G. Lucente. *Eur. J. Med. Chem.* **2009**, *44*, 179-189.
[16] A. Botz, V. Gasparik, E. Deviller, A.R.F. Hoffmann, L. Callon, E. Chelain, O. Lequin, T. Brigaud, L. Khemtemourian. *Biopolymers (Peptide Science)*, **2016**, *104*, 601-610.
[17] S. Jagota, J. Rajadas. *Med. Chem. Res.* **2013**, *22*, 3991-4000.
[18] J. A. Loureiro, R. Crespo, H. Barner, P. M. Martins, F. A. Rocha, M. Coelho, M. C. Pereira, S. Rocha J. Mater. Chem. B 2014, 2, 2259-2264.
[19] C. Giordano, P. Punzi, C. Lori, R. Chiaraluce, V. Consalvi. *ChemPlusChem*, **2014**, *79*, 1036-1043.
[20] B. Dorgeret, L. Khemtemourian, I. Correia, J. -L. Soulier, O. Lequin, S. Ongeri. *Eur. J. Med. Chem.* **2011**, *46*, 5959-5969.
[21] J. Kaffy, D. Brinet, J.-L. Soulier, L. Khemtemourian, O. Lequin, M. Taverna, B. Crousse, S. Ongeri, *Eur. J. Med. Chem.* **2014**, *86*, 752–758.
[22] J. Kaffy, D. Brinet, J.-L. Soulier, I. Correia, N. Tonali, K.F. Fera, Y. Iacone, A.R.F. Hoffmann, L. Khemtémourian, B. Crousse, M. Taylor, D. Allsop, M. Taverna, O. Lequin, S. Ongeri. *J. Med Chem.* **2016**, *59*, 2025-2040.
[23] M.Taylor, S. Moore, J. Mayes, E. Parkin, M. Beeg, M. Canovi, M. Gobbi, D.M.A. Mann, D. Allsop. *Biochemistry* **2010**, *49*, 3261–3272.
[24] C. Zhang, X. Zheng, X Wan, X. Shao, Q. Liu, Z. Zhang, Q. Zhang. *J Control Release*. **2014**, *192*, 317–324.
[25] X. Zheng, X. Shao, C. Zhang, Y. Tan, Q. Liu, X. Wan, Q. Zhang, S. Xu, X. Jiang, *Pharm Res*, **2016**, *32*, 3837–3849.
[26] R. Murthy, C.E. Shell, M.A. Grunlan. *Biomaterials* **2008**, *30*, 2433-2439.
[27] J. Tan, J.L. Brash. *J. Biomed. Mater. Res.* **2009**, *90A*, 196–204.
[28] Q. Yu, Y. Zhang, H. Wang, J. Brash, H. Chen. *Acta Biomaterialia* **2011**, *7*, 1550-1557.
[29] R. Duncan. *Nat. Rev. Drug Discovery* **2003**, *2*, 347-360.
[30] S. Rocha, I. Cardoso, H. Börner, M.C. Pereira, M.J. Saraiva, M. Coelho. *Biochem. Biophys. Res. Commun.* **2008**, *360*, 397-401.
[31] B. Permanne, C. Adessi, G.P. Saborio, S. Fraga, M.-J. Frossard, I. Derwatcher, J. van Dorpe, W.A. Banks, F. van Leuven, C. Soto. *FASEB J.* **2002**, *16*, 860-862.
[32] C. Adessi, M.J. Frossard, C. Boissard, S. Fraga, S. Bieler, T. Ruckle, F. Vilbois, S.M. Robinson, M. Mutter, W.A. Banks, C. Soto. *J. Biol. Chem.* **2003**, *278*, 13905-13911
[33] C. Giordano, Sansone, A. Masi, A. Masci, L. Mosca, R. Chiaraluce, A. Pasquo, V. Consalvi. *Chem. Biol. Drug Des.* **2012**, *79*, 30-37.
[34] I. Lynch, K.A. Dawson. *Nano Today* **2008**, *3*, 40-47.
[35] K. Kalyanasundaram, J.K. Thomas. *J. Am. Chem. Soc.* **1977**,*99*, 2039-44.
[36] P. Lianos, M.L. Viriot, R. Zana. *J Phys Chem* **1984**, *88*, 1098–101.
[37] M. Wilhelm, C. Zhao, Y. Wang, R. Xu, M.A. Winnik, M.D. Croucher. *Macromolecules* **1991**, *24*, 1033-40.
[38] Y. Cheng, C. He, C. Xiao, J. Ding, X. Zhuang, Y. Huang, X. Chen. *Biomacromolecules* **2012**, *13*, 2053-2059.
[39] R.W. Woody, A.K. Dunker. In *Circular Dichroism and the Conformational Analysis of Biomolecules*. Plenum Press, *New York*, **1996**.

[40] A. Asadi, A.A. Saboury, A.A. Moosavi-Movahedi, A. Divsalar, M.N. Sarbolouki. *Int. J. Biol. Macromol.* **2008**, *43*, 262-270.
[41] J. Wu, C. Zhao, W. Lin, R. Hu, Q. Wang, H. Chen, L. Li, S. Chen, J. Zheng. *J. Mat. Chem B* **2014**, *2*, 2983-2992.
[42] E. Atherton, R. Sheppard in *Solid Phase Synthesis. A Practical Approach*. IRL Press, Oxford, **1989**.
[43] A.J. Brouwer, M.C.F. Monnee, R.M.J. Liskamp. *Synthesis* **2000**, *11*, 1579-1584.
[44] G. Floudas, P. Papadopoulos. *Macromolecules*, **2003**, *36*, 3673-83.
[45] E.R. Henry, J. Hofrichter. *Methods Enzymol.* **1992**, *210*, 129-192.
[46] W. C. Johnson Jr. *Methods Enzymol.* **1992**, *210*, 426-447.

PLOS ONE

RESEARCH ARTICLE

Effect of BET Missense Mutations on Bromodomain Function, Inhibitor Binding and Stability

Laura Lori[1], Alessandra Pasquo[2], Clorinda Lori[1], Maria Petrosino[1], Roberta Chiaraluce[1]*, Cynthia Tallant[3¤], Stefan Knapp[3¤], Valerio Consalvi[1]

1 Department of Biochemical Sciences "A. Rossi Fanelli", Sapienza University of Rome, Rome, Italy, **2** SSPT-BIOAG-BIOTEC ENEA Casaccia ENEA, Rome, Italy, **3** Nuffield Department of Clinical Medicine, Structural Genomics Consortium and Target Discovery Institute, University of Oxford, Oxford, United Kingdom

¤ Current address: Institute for Pharmaceutical Chemistry and Buchmann Institute for Life Sciences (BMLS), Johann Wolfgang Goethe-University, Frankfurt am Main, Germany
* roberta.chiaraluce@uniroma1.it

Citation: Lori L, Pasquo A, Lori C, Petrosino M, Chiaraluce R, Tallant C, et al. (2016) Effect of BET Missense Mutations on Bromodomain Function, Inhibitor Binding and Stability. PLoS ONE 11(7): e0159180. doi:10.1371/journal.pone.0159180

Editor: Fatah Kashanchi, George Mason University, UNITED STATES

Received: April 15, 2016

Accepted: June 28, 2016

Published: July 12, 2016

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by Regione Lazio (Prot. FILAS-RU-2014-1020).

Competing Interests: The authors have declared that no competing interests exist.

Abbreviations: Kac, acetylated Lysine residues; BRD, Bromodomain; BET, Bromo and Extra Terminal; nsSNP, nonsynonymous single nucleotide polymorphism; BLI, biolayer interference; ITC, isothermal titration calorimetry.

Abstract

Lysine acetylation is an important epigenetic mark regulating gene transcription and chromatin structure. Acetylated lysine residues are specifically recognized by bromodomains, small protein interaction modules that read these modification in a sequence and acetylation dependent way regulating the recruitment of transcriptional regulators and chromatin remodelling enzymes to acetylated sites in chromatin. Recent studies revealed that bromodomains are highly druggable protein interaction domains resulting in the development of a large number of bromodomain inhibitors. BET bromodomain inhibitors received a lot of attention in the oncology field resulting in the rapid translation of early BET bromodomain inhibitors into clinical studies. Here we investigated the effects of mutations present as polymorphism or found in cancer on BET bromodomain function and stability and the influence of these mutants on inhibitor binding. We found that most BET missense mutations localize to peripheral residues in the two terminal helices. Crystal structures showed that the three dimensional structure is not compromised by these mutations but mutations located in close proximity to the acetyl-lysine binding site modulate acetyl-lysine and inhibitor binding. Most mutations affect significantly protein stability and tertiary structure in solution, suggesting new interactions and an alternative network of protein-protein interconnection as a consequence of single amino acid substitution. To our knowledge this is the first report studying the effect of mutations on bromodomain function and inhibitor binding.

Introduction

Epigenetics has been defined as heritable changes in phenotype that are the consequence of changes in DNA sequence but are due to differences in the pattern of post-translational modification present in histone, other nuclear proteins and in DNA [1]. Changes in post-translational

modifications also called epigenetic marks is a principal mechanism regulating chromatin structure and gene transcription and dysregulation of epigenetic marks has been linked to the development of a large diversity of diseases. Acetylation of lysine residues (Kac) is one of the most frequently occurring post-translational modifications which controls a vast array of diverse cellular functions. Dysregulation of acetylation levels has been associated with the development of many diseases in particular to cancer and enzymes regulating acetylation have emerged as interesting targets for drug discovery [2–4]. Acetylation levels are reversibly maintained by a group of enzymes, the histone acetyl-transferases and histone deacetylases that "write" and "erase" acetylation marks on histones [5]. Acetylation sites in proteins are specifically recognized by small helical interaction modules called Bromodomains (BRDs).

The relevant importance of BRDs in drug design is highlighted in recent studies that report BRDs as a target site for the development of new cancer drugs [3, 6–9]. Inhibitors that specifically target the BET (Bromo and Extra Terminal) proteins selectively interfered with gene expression that mediated cellular growth and evasion of apoptosis in cancer [10–12]. The studies of these inhibitors have suggested that inhibition of BRDs may have several potential clinical applications [3, 13]. BET proteins (ubiquitously expressed BRD2, BRD3, BRD4 and testis-specific BRDT) belong to the subfamily II of BRDs, sharing a common architecture comprising two N-terminal BRDs, domain 1 and domain 2, that exhibit high level of sequence conservation as well as an extra terminal domain and a more divergent C-terminal recruitment domain.

Despite their low sequence identity, all BRDs share a conserved fold comprising a left-handed bundle of four alpha helices, connected with a characteristic hydrophobic cleft between two conserved loops [14]. This binding site specifically recognizes ε-aminoacetyl groups of nucleosomal histone. The first bromodomains of BETs have a preference binding to di-acetylated Kac present in histone H4. The binding mode of the acetyl-lysine interaction is highly conserved comprising an anchoring hydrogen bond to a conserved asparagine residue present in most BRDs as well as a water mediated hydrogen bond to a conserved tyrosine residue [15].

BRD4 and BRD2 have crucial roles in cell cycle control [16, 17]. BRD2 and BRD4 remain bound to mitotic chromatin [18]; this property has been suggested to be important for the maintenance of epigenetic memory during cell division [19, 20]. Constitutive over-expression of BRD2 in B-cells of transgenic mice results in the development of B-cell lymphoma and leukaemia [21]. Gene rearrangements of BRD3 and BRD4 with a testis specific protein called NUT (Nuclear protein in Testis) have been detected in aggressive carcinoma [22–24]. BRD3 expression is induced in activated lymphocytes and it is highly expressed in undifferentiated embryonic stem cells, whereas expression levels are reduced upon endothelial differentiation [25]. Down regulated expression or loss of BRD3 has been detected in biopsies of nasopharyngeal carcinomas [26] and altered expression levels have been found in bladder cancer [25].

Several mutations in BRDs have been identified in humans and they may play an important role in several diseases but the functional consequences of the recorded mutations has not been studied. These variants are nonsynonymous single nucleotide polymorphisms (nsSNPs), single nucleotide variations occurring in the coding region and leading to a polypeptide sequence with amino acid substitutions. A number of investigations have addressed the effect of nsSNPs on protein stability, protein-protein interactions and protein functions for several other protein families [27, 28]. Indeed, large-scale computational studies utilizing structural information indicate that many nsSNPs may affect protein stability by either increasing or decreasing protein stability. However, only a detailed experimental analysis can unequivocally reveal the effect of the missense mutation on protein function [29, 30].

Thus, a detailed characterization of BET bromodomain variants found in cancer will be important to evaluate the role of these sequence alteration affecting bromodomain function and potentially resistance to emerging treatment strategies. Indeed, recent reports already

identified mechanisms by which cancer cells may evade treatment by BET bromodomain inhibitors [31, 32]. We therefore report here the effect of BET bromodomain mutations found in the cancer patients and annotated in the COSMIC database (http://cancer.sanger.ac.uk/cosmic) [33] on bromodomain structure, the ability of BET bromodomains to bind acetylated target sequences as well as inhibitors and the consequences of these mutations on the conformational stability of the bromodomain fold.

Results

BET variants in cancer

Mutants that map to BET bromodomains were mined from the COSMIC database (http://cancer.sanger.ac.uk/cosmic) [33] and were mainly identified BET variants, present in cancer of the large intestine and lung or in hematopoietic malignancies. Most of the nsSNPS found in the COSMIC database were present in BRD2. We identified seven mutants in BRD2 in the sequence spanning both bromodomains (D160N, D160Y, Y153H, E140K, R100L, Q443H, R419W), one in the second bromodomain of BRD3 (2) (H395R) and two mutations in the first and second bromodomain of BRD4 (A89V, A420D) (Fig 1A). Most mutants were located on the two terminal helices αB and αC, some of them in the proximity of the acetyl-lysine binding site and in the loop regions (e.g. A89V in BRD4(1) and R100L in BRD2(1) are located in the ZA loop) but no sequence variants were found in the highly sequence diverse αA helices. The mutated residues present in the αB helix, namely E140K and Y153H in BRD2(1), R419W in BRD2(2), A420D in BRD4(2), do not alter the conserved sequence motif ϕxxDϕxxϕϕxNϕxxY (ϕ is a hydrophobic residue) that precedes the conserved arginine essential for Kac binding in

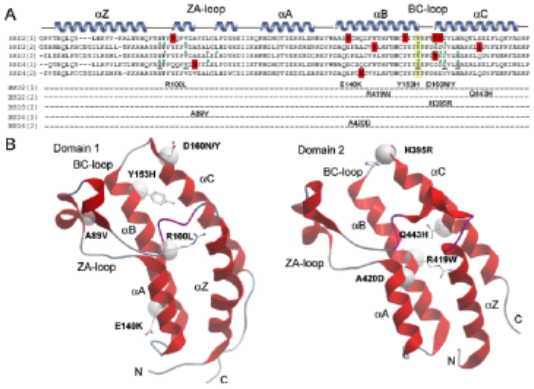


Fig 1. Alignment of BET bromodomain mutants. (A) Secondary structure elements are shown at the top of the sequence alignment. Mutated residues are highlighted in red and studied mutations are listed. The conserved asparagine (N391 in BRD3(2) numbering) is highlighted in yellow. The green dots represent the residues involved in binding with inhibitor JQ1 (PDB ID: 3ONI, 3S92, 3MXF). The residues underlined in blue are involved in PFI-1 binding (PDB ID: 4E96). (B) Location of the mutations. Shown are the first (left) and second bromodomain of BRD2. The mutated residues are highlighted in ball and stick and the position of Cα atoms are shown as a sphere. The main structural elements are labelled.

doi:10.1371/journal.pone.0159180.g001

BET proteins [14] but the tyrosine residue Y153, mutated in BRD2(1), is highly conserved in bromodomains suggesting a role in stabilizing the bromodomain fold. The residues D160 in BRD2(1) and H395 in BRD3(2), mutated in D160N, D160Y and H395R, respectively are located at the C-terminus of the BC loop, a region important for binding of Kac peptides and inhibitors (Fig 1A). The residue D161, located in the Kac binding region, was also mutated into D161N and D161Y to test the effect of this amino acid replacement in a region important for acetylated peptides binding. The location of the identified mutants mapped onto the structures of the first and second bromodomains is shown in Fig 1B [14]. All the mutations studied involve surface exposed residues and do not affect conserved amino acids, with the exception of Y153, a tyrosine residue present in all the sequence alignment in the family II BRDs (Fig 1). Hence, the consequences of the mutations were not obvious and we generated recombinant protein for each of the identified mutants using site directed mutagenesis and available bacterial expression systems [14].

Effects of BET mutants on structure

In order to obtain insight on changes in three dimensional structure as well as local interaction of the mutated sidechains we solved crystal structures of some of the generated mutants (Fig 2). All structures (BRD2(1)R100L, BRD2(1)Y153H, BRD2(1)D161Y, BRD2(2)Q443H, BRD3(2)H395R) were refined to high resolution maintaining favourable geometry (S1 Table). The three dimension structure of the mutants was found to be highly conserved showing only local structural alterations (Fig 2A). In the BRD2(1) mutant R100L the hydrogen bonds with backbone residues in helix αA were lost but this did not result in significant structural changes (Fig 2B). Mutations at BRD2(1) D161 as well as BRD2(2) Q443H (Fig 1) did also not result in any significant structural rearrangements in the refined crystal structures (Fig 2C and 2D).

Effects of BET mutants on conformation in solution

The conformation in solution of all mutants was studied spectroscopically using CD and fluorescence spectroscopy. The near-UV CD spectra of wild type bromodomains represent the spectral contributions of all aromatic residues and was characterized by a strong negative peak centred at around 280 nm and a positive one around 260 nm, accompanied, for BRD4(1) and BRD2(2), by fine structure features in the region of 290 nm (Fig 2E–2I). Notably, the significant differences observed in the near UV CD spectra of all the mutants indicated that their tertiary structure arrangements are different from that of the corresponding wild type proteins, as also suggested by differences observed in the intrinsic fluorescence spectra (S1 Fig). The intrinsic fluorescence spectra of most of the variants differed from the corresponding wild type proteins mainly in intensity, which was either significantly decreased or enhanced. As expected, based on differences in aromatic residue content, the maximal fluorescence emission wavelength of BRD2(2) R419W was red shifted with respect to that of the wild type protein (S1C Fig). These results pointed to local rearrangements of tertiary interactions for most of the BRDs variants in solution.

The far-UV CD spectrum of all mutants indicated some differences in secondary structure content probably due to differences in dynamic fluctuation in solution (Fig 3). The 222/208 ellipticity ratio, which is indicative of interhelical contacts present in helix bundle and coiled coil structures as expected in the conserved bromodomain fold, differed in the various wild type bromodomains and corresponded to 0.97, 0.98, 1.06, 1.16 and 2.15 in BRD2(1), BRD4(1), BRD4(2), BRD2(2) and BRD3(2), respectively. The 222/208 ellipticity ratio is generally used to distinguish between coiled coil helices ($\geq$ 1.0) and non-interacting helices (0.8–0.9) [34, 35]. This ratio was unchanged for all the variants, except for the BRD2(1) mutants Y153H and

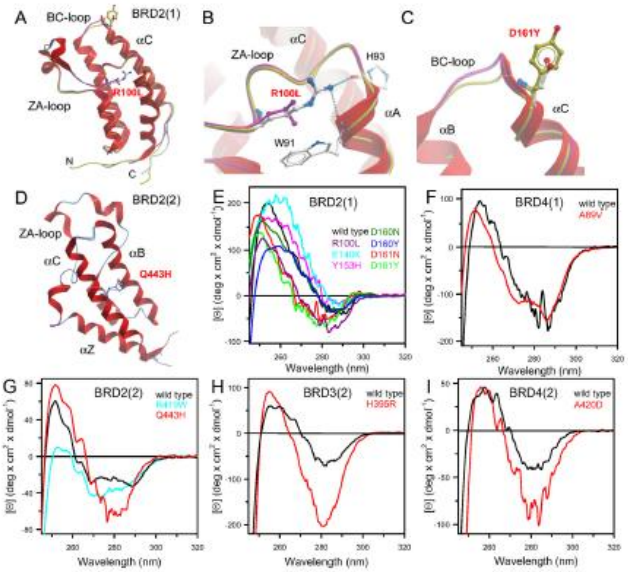


Fig 2. Structure of BET mutants and tertiary structure of mutants in solution. (A) Superimposition of wild type BRD2(1) shown as a ribbon diagram with the mutants BRD2(1) R100L and D161Y shown as protein worm in green and magenta, respectively. The mutated residues are shown in ball and stick representation and main structural elements are labelled. (B) Details of interactions formed by R100 in the wild type compared to the mutated residue. (C) Detailed view of BRD2(1) wild type and D161Y. (D) Superimposition of BRD2(2) shown as a ribbon diagram and the mutant BRD2(2) Q443H shown as protein worm in blue. (E) Comparison of the near UV CD spectra of wild type BRD2(1) and all generated mutants. (F) Comparison of the near UV CD spectra of wild type BRD4(1) and the mutants BRD4(1) A89V. (G) Comparison of the near UV CD spectra of wild type BRD2(2) and the mutants BRD2(2) R419W and Q443H. (H) Comparison of the near UV CD spectra of wild type BRD3(2) and the mutants BRD3(2) H395R. (I) Comparison of the near UV CD spectra of wild type BRD4(2) and the mutants BRD4(2) A420D. Near-UV CD spectra were recorded at 20°C in a 1.0-cm quartz cuvette in 20 mM Tris/HCl, pH 7.5 containing 0.20 M NaCl and 2.00 mM DTT, as described in Materials and Methods.

doi:10.1371/journal.pone.0159180.g002

E140K that show a 222/208 ellipticity ratio below 1.0, corresponding to 0.86 and 0.91, respectively (Fig 3A). The decrease in the 222/208 ellipticity ratio observed for these BRD2(1) variants suggested that a consequence of the amino acid substitutions was a perturbation of interhelical contacts that resulted in structural differences in solution, despite the fact that Y153 and E140 are not involved in any direct interactions with another helix and are solvent exposed (Fig 1B).

All in all we conclude that despite the high structural conservation of mutants in crystal structures the studied bromodomain mutants showed detectable changes of their spectroscopic properties in solution suggesting differences in three dimension structures and the dynamic properties of these mutants. Proteins are conformationally constrained in crystals and the

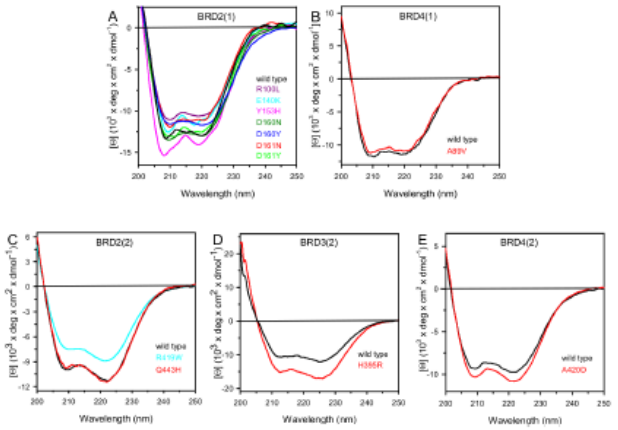


Fig 3. Far-UV CD spectra of wild type bromodomains and mutants. Far-UV CD spectra were recorded at 20°C in a 0.1-cm quartz cuvette in 20 mM Tris/HCl, pH 7.5 containing 0.20 M NaCl and 0.40 mM DTT, as described in Materials and Methods. Wild type spectra are shown as black solid lines and mutants are coloured as indicated in the figure.

doi:10.1371/journal.pone.0159180.g003

studied bromodomain mutations may therefore alter domain plasticity and structural properties that are only apparent in solution.

Effects of BET mutants on acetyl-lysine peptide and inhibitor binding

Some of the BET mutants identified in cancer, are located in close proximity to the acetyl-lysine binding site and we were therefore interested whether these mutants compromised the ability of the bromodomain to bind acetylated histone peptides as well as BET inhibitors. To address the first question we used biolayer interference (BLI), a technology that measures differences in refractive index in a similar way as the widely used surface plasmon resonance technology, and a library of biotin labelled acetylated histone peptides. Using this peptide library we measured the BLI response at 20 μM protein concentration on tips with immobilized acetylated and non-acetylated control peptides (Fig 4A). As expected, the first bromodomain of BRD2(1) interacted strongly with polyacetylated peptides of histone H4 [14]. This interaction was weakened by mutation in a region important for binding of Kac peptides (Fig 1A), such as D161 (Fig 4A). Second bromodomains in BETs show much weaker interaction with histones and other nuclear proteins that have been discussed as potential targets [36]. Indeed, we confirmed these reports and found only weak interaction with histones peptides that were present in our peptide library for BRD3(2) (Fig 4A).

For studying the effects of the mutations on binding of BET inhibitors, we used two structurally diverse panBET inhibitors: the thienodiazepine JQ1 [11] as well as the quinazolinone PFI-1 [14, 37, 38]. Binding of JQ1 was only modestly affected by all mutations (S2 Table and Fig 4D). Notably, the K_d for JQ1 was slightly increased for A420D, D161N and Y153H and slightly decreased for all the other variants. A different trend is observed for PFI-1: in this case

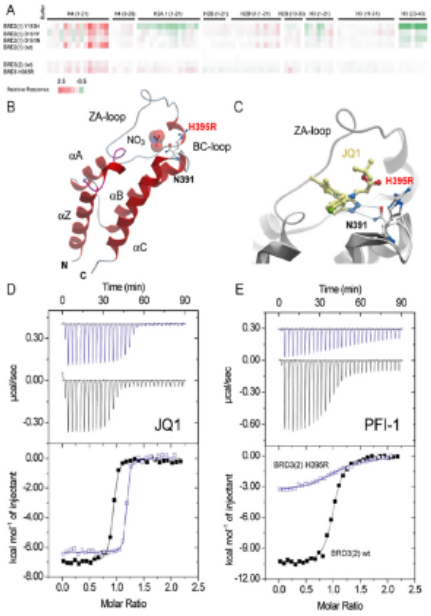


Fig 4. Binding of BET bromodomain mutants to acetylated histone peptides. (A) BLI study showing the interaction of some mutants with acetylated histone peptides. Shown is the maximum response at a protein concentration of 20 μM after subtraction of non-acetylated reference peptides using a colour code as indicated in the figure capture. (B) Structure of BRD3(2) mutant H395R. The mutated residues are highlighted. A nitrate molecule, present in the crystallization solution occupied the acetyl-lysine binding site in BRD3(2) H395R. The conserved asparagine (N391) formed canonical hydrogen bonds with the nitrated ion. (C) Superimposition of the wild type JQ1 complex with BRD3(2) H395R. (D) ITC experiments measuring the interaction of the panBET inhibitor JQ1 with wild type BRD3(2) (black curve) and BRD3(2) H395R (blue curve). Shown are raw titration heats (top panel) as well as normalized binding enthalpies as a function of ligand/protein ratio (lower panel). The best fit to a single binding site model is shown as solid lines. (E) ITC experiments showing the binding of the pan-BET inhibitor PFI-1 with BRD3(2) and the H395R mutant. A significant reduction in binding enthalpy and binding affinity is evident. Data on all ITC titrations are summarized in S2 Table.

doi:10.1371/journal.pone.0159180.g004

the K_d was slightly decreased for D161Y and A420D and was increased for all the other variants. The largest differences were observed for mutants of BRD3(2). H395R mutant and wild type protein with both the inhibitors JQ1 and PFI-1; this variant showed a 3-fold lower K_d for JQ1 and about 9-fold higher K_d for PFI-1 (Fig 4E). Comparison with the crystal structure of wild type BRD3(2) with JQ1 and the structure of BRD3(2) H395R suggested that the arginine residue introduced by this mutant may form a more favourable hydrogen bond with the carbonyl oxygen of the ester link (Fig 4B and 4C). All titration experiments are compiled in S2 Table.

Table 1. Melting temperatures and thermodynamic parameters for urea-induced unfolding equilibrium of BRDs wild type and mutants measured by far-UV CD and fluorescence spectroscopy.

	Tm (°C)	$\Delta G^{H}_{2}O$ (kcal/mol)		*m* (kcal/mol/M)		$[Urea]_{0.5}$ (M)	
		CD ($[\Theta]_{222}$)	Fluorescence	CD ($[\Theta]_{222}$)	Fluorescence	CD ($[\Theta]_{222}$)	Fluorescence
BRD2(1)							
Wild type	54.8	12.88 ± 0.93	12.91 ± 0.99	2.40 ± 0.17	2.23 ± 0.17	5.36	5.79
Y153H	47.8	6.39 ± 0.34	7.60 ± 0.61	1.49 ± 0.08	1.60 ± 0.13	4.30	4.75
E140K	44.0	9.40 ± 0.70	8.38 ± 0.63	2.16 ± 0.16	1.85 ± 0.14	4.35	4.54
R100L	49.0	7.66 ± 0.87	ΔG_1 = 8.20 ΔG_2 = 10.10	1.46 ± 0.16	m_1 = 3.16 ± 0.66 m_2 = 1.92 ± 0.14	5.42	2.47 ± 0.05 5.35 ± 0.02
D160N	49.0	8.78 ± 0.70	10.58 ± 0.82	2.03 ± 0.16	2.52 ± 0.19	4.32	4.20
D160Y	45.0	8.86 ± 0.59	8.71 ± 0.58	1.82 ± 0.12	1.80 ± 0.12	4.85	4.83
D161N	53.5	10.67 ± 0.91	9.89 ± 0.68	2.11 ± 0.33	2.09 ± 0.18	5.07	5.31
D161Y	50.0	9.50 ± 0.36	8.25 ± 0.47	1.71± 0.06	1.46 ± 0.08	5.55	5.63
BRD4(1)							
Wild type	52.9	8.82 ± 0.49	11.39 ± 0.97	1.27 ± 0.07	1.56 ± 0.13	6.96	7.32
A89V	53.0	9.81 ± 0.49	10.60 ± 0.83	1.46 ± 0.07	1.56 ± 0.12	6.70	6.80
BRD2(2)							
Wild type	53.0	7.97 ± 0.53	6.60 ± 0.39	1.50 ± 0.10	1.36 ± 0.08	5.29	4.86
Q443H	50.0	8.86 ±0.47	8.28 ± 0.40	1.77 ± 0.09	1.77 ± 0.09	5.01	4.68
R419W	45.0	7.95 ± 0.42	6.00 ± 0.36	1.72 ± 0.09	1.18 ± 0.07	4.63	5.10
BRD3(2)							
Wild type	51.6	7.14 ± 0.63	6.68 ± 0.52	1.41 ± 0.12	1.51 ± 0.11	5.06	4.44
H395R	49.0	9.17 ± 0.63	7.80 ± 0.55	2.02 ± 0.14	1.90 ± 0.13	4.53	4.10
BRD4(2)							
Wild type	54.0	8.66 ± 0.47	4.72 ± 0.23	1.59 ± 0.08	0.95 ± 0.05	5.43	4.95
A420D	51.0	9.10 ± 0.49	9.65 ± 0.53	1.75 ± 0.09	1.87 ± 0.10	5.20	5.15

The temperature-induced changes were followed by monitoring the ellipticity at 222 nm. The Tm values were calculated by taking the first derivative of the ellipticity at 222 nm with respect to temperature. Urea-induced unfolding equilibrium data were measured at 10°C in 20 mM Tris/HCl, pH 7.5, containing 0.2 M NaCl and 200 μM DTT by monitoring ellipticity at 222 nm [Θ_{222}] and intrinsic fluorescence emission. $\Delta G^{H}_{2}O$ and *m* values were obtained from Eq 3; $[Urea]_{0.5}$ was calculated from Eq 2. Intrinsic fluorescence emission data of BRD2(1) R100L were fitted to Eq 5. Data are reported as the mean ± SE of the fit.

doi:10.1371/journal.pone.0159180.t001

Effects of BET mutants on BRD stability

The thermal and thermodynamic stability of the wild type BRD2(1), BRD4(1) BRD2(2), BRD4 (2) and BRD3(2) were compared with the generated BET variants (Table 1).

It is noteworthy that all the natural variants studied, with the exception of BRD4(1) A89V, showed a significant decrease in the melting temperature (Tm) ranging from 1.3°C for BRD2 (1) D161N to 10.0°C for BRD2(1) E140K, when compared to the wild type proteins. The thermal stability of BRDs wild type and variants was investigated by continuously monitoring the ellipticity changes at 222 nm between 20 and 80°C. The observed thermal unfolding occured in an apparent two-state cooperative transition for all BRDs wild type and variants (S2 and S3 Figs). The midpoint of the unfolding, Tm, was calculated by plotting the first derivative of the molar ellipticity at 222 nm, where the main amplitude was observed, as a function of temperature (S2A and S2B Fig, inset; S3A–S3C Fig, inset). It is noteworthy that for all the BRDs variants, with the exception of BRD2(2) variants, the amplitudes of the ellipticity changes at 222 nm, i.e. the difference between the ellipticity measured at the end (80°C) and that at the beginning (20°C) of the thermal transition, were different from those measured for the wild type (S2C, S2D and S3D–S3F Figs). In particular, the amplitude of the ellipticity changes at 222 nm increased for all BRD2(1) variants, with the exception of D161N (S2C Fig), ranging from

1.3-fold for E140K, R100L, and D161Y, to 1.4-fold for D160N and D160Y and to 1.5-fold for Y153H. For the variants BRD3(2) H395R and BRD4(2) A420D a 1.6 and 1.2-fold increase of thermal transition amplitude is observed. These results point to a larger extent of loss of secondary structure upon thermal unfolding when compared to the wild type protein.

The lower Tm values and the higher loss of secondary structure elements upon thermal unfolding suggest that the point mutations induce a remarkable destabilization of the native state of BRDs. The temperature-induced ellipticity changes for all BRDs wild type and mutants were coincident with the heat-induced increase of the photomultiplier tube voltage (data not shown) suggesting that the temperature-induced unfolding is accompanied by protein aggregation [39]. Aggregation occurred also when thermal scans were performed at a lower heating rate with a low-temperature shifts of the apparent Tm; the differences between the apparent Tm of wild type and variants were the same as those measured at higher heating rate (data not shown). The observed transitions were irreversible as indicated by the spectra measured at the end of the cooling phase that differ from those of the native proteins measured at the beginning of the thermal transitions (data not shown).

The thermodynamic stability was studied by urea-induced equilibrium unfolding. BRD2(1), BRD4(1) BRD2(2), BRD4(2) and BRD3(2) wild type and variants reversibly unfold in urea at 10°C. The effect of increasing urea concentrations (0–8 M) on the structure of BRDs variants was analyzed by far-UV CD (S4A, S4B and S5A–S5C Figs) and fluorescence spectroscopy (S4C, S4D and S5D–S5F Figs) and compared to the effect exerted on the corresponding wild type. The same samples used to monitor the far-UV CD changes during the unfolding transition were used to monitor fluorescence emission changes, to allow a direct comparison of the data. The urea-induced changes in 222 nm ellipticity of all the BRDs wild type and mutants showed a sigmoidal dependence on denaturant concentration and follow an apparent two-state transition without any detectable intermediate (S4A, S4B and S5A–S5C Figs). Incubation of BRDs wild type and variants at increasing urea concentrations resulted in a progressive change of the intrinsic fluorescence emission intensity and a red-shift of the maximal emission wavelength from 345 nm, in the absence of denaturant, to about 366 nm, in 8 M urea (data not shown). Determination of the red-shift of the intrinsic fluorescence emission was obtained by calculating the intensity averaged emission wavelength, $\bar{\lambda}$, at increasing urea concentration (S4C, S4D, S4A, S4B and S5D–S5F Figs). This parameter is an integral measurement, negligibly influenced by the noise, and reflects changes in both the shape and the position of the emission spectrum. The urea-induced changes in $\bar{\lambda}$ of all the mutants are similar to that of the wild type proteins, show a sigmoidal dependence on urea concentration and follow an apparent two-state transition without any detectable intermediate, except for the variant of BRD2(1) R100L.

The thermodynamic parameters obtained from the analysis of the far-UV CD and fluorescence changes transitions are reported in Table 1. The difference between the free energy of urea-induced unfolding, ΔG^{H_2O}, of the variants and that of the wild type indicates a decrease in thermodynamic stability of 2–6 kcal/mol for BRD2(1) variants and an increase of about 2 kcal/mol for BRD3(2) mutant H395R. A minor increase in ΔG^{H_2O} (about 1 kcal/mol) was observed for all the other BRDs variants (Table 1). The decrease in ΔG values measured monitoring the far-UV CD changes of the BRD2(1) variants Y153H, R100L, D160Y and D161Y may be mainly referred to the lower m values, with respect to the wild type proteins. An increase in both ΔG and m values was observed only for the BRD3(2) mutant H395R. The changes in m values may indicate differences in the solvent exposed surface area upon unfolding between the variants and the wild type: decrease in m values is usually referred to a decrease in the solvent-exposed surface area upon unfolding. This is frequently ascribed to an increase in the compactness of the residual structure in the non-native state ensemble, rather than to an increase of the

accessible surface area of the native state [40–42]. A decrease in *m* value upon single mutation has been also referred, in some cases, to the population of a third intermediate state during chemical unfolding [43]; this is the case for R100L unfolding where an intermediate has been detected by monitoring the fluorescence changes (S4C Fig, inset). In 8.0 M urea, the ellipticity at 222 nm and the maximal fluorescence emission wavelength of the BRDs variants are comparable to those of the corresponding wild type proteins (data not shown) and did not indicate any increase in the structure of the non-native state to support the large decrease in *m* value observed for the BRD2(1) variants Y153H, R100L, D160Y and D161Y. In native conditions, the spectral properties of most of the variants point to tertiary structure arrangements different from those of the wild type and, for BRD2(1) E140K and Y153H variants (Fig 2), the 222/208 nm ratio < 1.0 (Fig 3A) may suggest a less compact native state.

The values of the thermodynamic parameters obtained from far-UV CD were comparable to those determined from fluorescence data for all the BRDs wild type, with the exception for BRD4(1) and BRD4(2) (Table 1); a non-coincidence of the thermodynamic parameters obtained by the two spectroscopic probes was also observed for the BRD2(2) variant R419W (Table 1). The lack of coincidence of the thermodynamic parameters determined using different spectral probes is generally considered as an indication that the unfolding does not follow a simple two-state process and suggests the presence of an undetected intermediate in the unfolding pathway. However, the *m* values obtained from the analysis of the far-UV CD and fluorescence unfolding transitions were within the range of those expected for a monomeric protein of about 115 amino acid residues denatured in urea [44]. Indeed, the presence of a folding intermediate was evident for the BRD2(1) R100L whose reversible urea unfolding transition monitored by the changes in fluorescence intensity averaged emission wavelength $\bar{\lambda}$ (S4C Fig, inset) was fitted as a three-state unfolding process, according to Eq 5 (Table 1).

Discussion

Bromodomains (BRDs) dysfunction has been linked to the development of several diseases [4]. In these regards, BRDs have recently emerged as interesting targets for the development of specific protein interaction inhibitors [3]. The object of this study were three of the four BET proteins BRD2, BRD3 and BRD4 and to our knowledge, this is the first spectroscopic characterization in solution of human BRDs variants found in cancer. Wild type and variants of BRD2(1), BRD4(1), BRD2(2), BRD3(2) and BRD4(2), were studied in order to investigate the effect of amino acid substitutions on their structure in solution and on their thermal and thermodynamic stability and interactions with peptides and inhibitors (Figs 1 and 2, S6 Fig). The structures of some of the variants that showed the most significant changes in inhibitor binding were determined by X-ray crystallography (Fig 2).

The single amino acid substitutions significantly affect the tertiary interactions of most of the BRDs variants studied, as indicated by the differences in their near UV CD fine structure and/or in the ellipticity amplitude, when compared to the corresponding wild type proteins. Due to the usually surface exposed location of the mutated residues, the consequences of the mutations were not obvious, however most of the amino acid substitutions in BRDs variants involve changes in the charge of solvent exposed residues, that may alter protein stability. Notably, the amino acid substitutions in most of the variants suggest an increase in side chain flexibility for the mutated residue [45] that may be responsible for local changes in protein dynamics. The changes in tertiary arrangements of most of the variants were also evident from their intrinsic fluorescence spectra that showed differences in relative intensity and/or in the maximum emission wavelength. Only the mutation of Ala89 to Val did not significantly affect the tertiary contacts of BRD4(1), as judged by a comparison of its spectral properties and of its

stability with those of the wild type. A loosening of the native tertiary structure was suggested also by the changes in the far UV CD spectra of some BRDs variants, particularly those of BRD2(1) Y153H and E140K for which the 222/208 ellipticity ratio below 1.0 points to a weakening of the interhelical contacts [34, 35]. Consequently the thermal and thermodynamic stability of these two variants were significantly decreased, as expected from the role of the conserved Y153 residue, located at the end of the αB helix, in stabilizing the bromodomain fold [14]. On the other hand, the substitution of the negatively charged glutamate with a positively charged lysine may induce local perturbations responsible for the observed destabilization, despite the lack of interaction of the solvent exposed side chain of E140 in the crystal structure of the wild type protein.

The thermal and thermodynamic stability of the BRD2(1) variants Y153H, E140K, R100I, D160N, D160Y and D161Y were remarkably lower than the corresponding wild type: the Tm values were at least 5 degrees below that of the wild type and the ΔG values are more than 3 kcal/mol lower than those of the wild type proteins (S3 Table). Amino acid substitutions in BRD2(1) variants involved changes in the charge of solvent exposed residues which would be expected to have a significant impact on protein stability. In general, the naturally occurring point mutations in BRD2(1) variants induced a remarkable destabilization of the native state, as suggested by the significant loss of secondary structure elements upon thermal unfolding and by the decrease in Tm and/or in ΔG values, e.g. in the mutant of the conserved Y153. However, for the other BRDs variants, differently from what observed for those of BRD2(1), the perturbation of tertiary contacts were not always accompanied by changes in both thermal and thermodynamic stability. The 8°C decrease in thermal stability of BRD2(2) R419W variant was not paralleled by a drop of its thermodynamic stability measured by urea induced-unfolding. Mutation of a charged arginine residue, solvent exposed, into an aromatic and more flexible tryptophan [45] may induce tertiary changes and lead to a protein with different hydrophobic interactions and, probably, more prone to thermal denaturation. On the other hand, for the BRD3(2) variant H395R the Tm is 2.6°C lower than that of the wild type, whereas the unfolding ΔG is even higher than that of the wild type, due to the higher m value of this variant (Table 1). Interestingly, both the variants BRD2(2) R419W and BRD3(2) H395R displayed significant differences in the overall far-UV CD ellipticity and in the tertiary arrangements, as indicated by the difference of their intrinsic fluorescence emission and near UV CD spectra with respect to the wild type, despite the fact that interhelical contacts in these variants were unaffected by the mutation, as suggested by the 222/208 ellipticity ratio similar to that of the wild type. In conclusion, the naturally occurring mutations in BRDs, caused by nsSNPs and found in cancer result in proteins with significantly altered physico-chemical properties, such as alteration of native tertiary contacts, loosening of the interhelical contacts or changes in thermal and/or thermodynamic stability [46]. In conclusion, despite the high structural conservation in crystal structures, the BRDs variants showed differences in spectral properties in solution that may suggest local structural changes and modifications of their dynamic properties.

The binding studies revealed that the BRD2(1) variants, namely Y153H, D161Y and D161N and BRD3(2) H395R showed significant differences compared to the wild-type proteins. The observed differences in inhibitor binding (S2 Table) may be referred to the fact that the mutated residues are located in close proximity of the BC loop, a region important for binding of Kac peptides and inhibitors (Fig 1A). Also the interactions with the acetylated histones peptides of H2A, H2B, H3 and H4, although preserved, showed some differences indicating that the mutations affect the binding interactions with the natural substrates. Interestingly, the superimposition of the X-ray structures of some of the variants with those of the wild type, revealed that the mutations did not cause overall misfolding of the structure (Figs 2 and 4). Minor structural changes were observed in BRD3(2) H395R where the amino acid substitution

resulted in a new hydrogen bond and affect interactions in the ligand binding site, as indicated by the different inhibitor binding affinity. Notably the remarkable changes in the interactions with histone recognition sequences and inhibitors, e.g. in the H395 mutant, were not accompanied by significant changes in the thermal and or thermodynamic stability of this variant (Fig 4 and S3 Table). A close relationship between the stability and the inhibitor binding data can be established only for Y153H which showed the largest differences in stability accompanied by a significant decrease in binding affinity for JQ1 and PFI-1.

Taken altogether, our results indicate that the mutants of BRDs found in cancer tissues did not alter significantly the overall folding. However, the significant alteration of the tertiary contacts observed in solution and the notable decrease in protein stability suggested an increase in conformational flexibility. All the mutated residues were solvent exposed, therefore they are not supposed to alter the global folding; however, a mutation of a residue on the surface may result in new and unknown interactions, thus the variants may acquire a new pattern of interactions and establish a novel and alternative network of protein-protein interconnection [47, 48]. This may be particularly important in BRDs that are physiologically embedded in multi-domain proteins and multi-subunit complexes.

Apart from PFI-1 binding to BRD3(2) H395R, the studied mutations are still inhibited by two diverse BET inhibitors and it is therefore not expected that the studied mutants will lead to drug resistance.

The results obtained from the study on BRDs nsSNPS may raise additional questions since BRDs are involved in the activation of oncogenes expression and most of the variants are less stable than the wild type, thus the variants might be expected to be less "deleterious". Considering the pivotal role of BET proteins in the regulation of the transcription of growth-promoting genes and cell cycle regulators, the phenotypic perturbations of BRDs variants may potentially lead to oncogene activation thus significantly affect the tumour development [4].

Materials and Methods

Plasmids and site-directed mutagenesis

The plasmids harboring the BRDs wild type genes [14] were used to obtain mutant enzymes. The wild type plasmids were subjected to site-directed mutagenesis using Quick Change Site-directed Mutagenesis Kit (Stratagene), combined with specific sense and antisense mutagenic oligonucleotides as shown in S4 Table. The presence of the desired mutations and the absence of unwanted additional mutations were confirmed by inserts sequencing.

Protein expression and purification

Wild type and mutant proteins were expressed in *E.coli*. Rosetta cells transformed with the selected plasmids were grown at 37°C in LB medium with Kan antibiotic to an $OD_{595} = 0.6$. Upon reached desired OD temperature was lowered up to 18°C, cultures were grown overnight after induction with 0.5 mM isopropyl-β-D-thiogalactopyranoside. Protein purification was carried out at 4°C modifying the protocol indicated by Filippakopoulos *et al.* [11]. Cells from one liter culture were collected by centrifugation, re-suspended in 50 ml of buffer A (50 mM HEPES, pH 7.5, 0.5 M NaCl, 5% Glycerol and 0.5 mM tris(2-carboxyethyl)phosphine–HCl (TCEP) containing a cocktail of EDTA-free protease inhibitors (Roche), sonicated in a Vibracell 75115 sonicator with 5 s boosts and 9 s pause, on ice. The sonicated cells were centrifuged and the supernatant, after an additional centrifugation at 15000 rpm, was applied to a 5 ml prepacked His Trap column (GE Healthcare) equilibrated in buffer A. The His- tagged fusion protein was eluted with 0.25 M imidazole in buffer A. The eluted protein was concentrated to a final volume of 2.5 ml on an Amicon concentrator Ultra-15 (Millipore) and then applied to a

PD-10 pre-packed column (GE Healthcare) to remove imidazole. The protein in buffer A was cleaved by recombinant His-tagged tobacco etch virus (TEV) protease (kindly provided by SK), overnight at 4°C. The digested mixture containing TEV protease, the His-tag and the cleaved protein was applied to a 5 ml pre-packed His Trap column (GE Healthcare) previously equilibrated in buffer A. The flow through containing the protein without His-tag was collected, and checked for purity and size by SDS–PAGE on a pre-casted NuPage 4–12% bis-Tris polyacrylamide gel (Invitrogen) (S7 and S8 Figs). Gels were stained with Coomassie blue R-250. The protein without His-tag was used for all structural and stability experiments. The mutant and wild type enzymes obtained were approximately 70 mg from one liter culture. Protein quantification was determined according to OD_{280} measurement using respective molar extinction coefficients ε of each protein, calculated according to Gill and Hippel [49].

Spectroscopic measurements

For intrinsic fluorescence emission measurements, the absorbance of the protein solutions at 280 nm was 0.10 AU, corresponding to a protein concentration ranging over 50.0–100 μg/mL. Intrinsic fluorescence emission measurements were carried out at 10°C with a LS50B spectrofluorimeter (Perkin-Elmer) using a 1.0 cm path length quartz cuvette. Fluorescence emission spectra were recorded from 300 to 450 nm (1 nm sampling interval), with the excitation wavelength set at 295 nm. For far-UV (190–250 nm) CD spectra, the absorbance of the protein solutions at 280 nm was 0.18 AU, corresponding to a protein concentration ranging over 100–190 μg/ml (0.4 mM DTT, 0.1 cm path length quartz cuvette) or 0.10 AU, corresponding to a protein concentration ranging over 50.0–100 μg/ml. (0.2 mM DTT, 0.2 cm path length quartz cuvette). For near-UV (250–320 nm) CD spectra, the absorbance of the protein solutions at 280 nm was 2.2 AU, corresponding to a protein concentration ranging over 1.25–2.28 mg/mL (2.0 mM DTT, 1.0 cm path length quartz cuvette). The results obtained from CD measurements were expressed as the mean residue ellipticity ([Θ]), assuming a mean residue molecular mass of 110 per amino acid residue. All spectroscopic measurements were carried out at 10°C in 20 mM Tris-HCl pH 7.5 containing 0.20 M NaCl.

Urea-induced equilibrium unfolding

For equilibrium transition studies, BRDs wild type and variants (final concentration ranging over 50.0–100 μg/ml) were incubated at 10°C at increasing concentrations of urea (0–8 M) in 20 mM Tris/HCl, pH 7.5, in the presence of 0.2 M NaCl and 200 μM DTT. After 10 min, equilibrium was reached and intrinsic fluorescence emission and far-UV CD spectra (0.2-cm cuvette) were recorded in parallel at 10°C. To test the reversibility of the unfolding, BRDs wild type and variants were unfolded at 10°C in 7.5 M urea at protein concentration ranging over 0.5–1.0 mg/ml in 20 mM Tris/HCl, pH 7.5, in the presence of 2 mM DTT and 0.2 M NaCl. After 10 min, refolding was started by 10-fold dilution of the unfolding mixture at 10°C into solutions of the same buffer used for unfolding containing decreasing urea concentrations. The final protein concentration ranged over 50.0–100 μg/mL. After 24 h, intrinsic fluorescence emission and far-UV CD spectra were recorded at 10°C. All denaturation experiments were performed in triplicate.

Thermal denaturation experiments

BRDs variants and wild type (protein concentration ranging over 0.10–0.20 mg/ml) were heated from 20°C to 80°C in a 0.1 cm quartz cuvette with a heating rate of 1 degree x min^{-1} controlled by a Jasco programmable Peltier element. The dichroic activity at 222 nm and the PMTV were continuously monitored in parallel every 0.5°C [39]. All the thermal scans were

corrected for the solvent contribution at the different temperatures. Melting temperature (Tm) values were calculated by taking the first derivative of the ellipticity at 222 nm with respect to temperature. All denaturation experiments were performed in triplicate.

Bio-layer Interferometry

In Bio-layer Interferometry (BLI) experiment the affinity between histones acetylated peptides and the proteins was measured. The proteins (20 μM) were dialyzed against the assay buffer (25 mM Hepes pH 7.5, 100 mM NaCl and 0.01% TWEEN). The biotinylated acetylated peptides of the H2A, H2B, H3 and H4 histones were immobilized onto strepdavidin biosensor (FòrteBio). All binding experiments were conducted at 25°C using an OctetRed 384 instrument (FòrteBio). Common cycles steps for analysis included 120 s of biosensor baseline equilibration step, associations in wells containing the free label protein for 240 s, and dissociations in buffer wells for 240 s. Reference subtraction was performed with the FòrteBio data analysis software to subtract the effect of baseline drift and the effect of nonspecific binding to biosensor tips without immobilized peptides.

Data analysis

Far-UV CD spectra recorded as a function of urea concentration were analyzed by a singular value decomposition algorithm (SVD) using the software MATLAB (Math-Works, South Natick, MA) to remove the high frequency noise and the low frequency random errors and determine the number of independent components in any given set of spectra. CD spectra in the 213–250 nm region were placed in a rectangular matrix A of n columns, one column for each spectrum collected at each time. The A matrix is decomposed by SVD into the product of three matrices: $A = U^*S^*V^T$, where U and V are orthogonal matrices and S is a diagonal matrix. The U matrix columns contain the basis spectra and the V matrix columns contain the urea dependence of each basis spectrum. Both U and V columns are arranged in terms of decreasing order of the relative weight of information, as indicated by the magnitude of the singular values in S. The diagonal S matrix contains the singular values that quantify the relative importance of each vector in U and V. The signal-to-noise ratio is very high in the earliest columns of U and V while the random noise is mainly accumulated in the latest U and V columns. The wavelength averaged spectral changes induced by increasing denaturant concentrations are represented by the columns of matrix V; hence, the plot of the columns of V versus the denaturant concentrations provides information about the observed transition.

The changes in intrinsic fluorescence emission spectra at increasing urea concentrations were quantified as the intensity-averaged emission wavelength, $\bar{\lambda}$, [50] calculated according to

$$\bar{\lambda} = \sum(I_i\lambda_i)/\sum(I_i) \quad (1)$$

where λ_i and I_i are the emission wavelength and its corresponding fluorescence intensity at that wavelength, respectively. This quantity is an integral measurement, negligibly influenced by the noise, which reflects changes in the shape and position of the emission spectrum.

Urea-induced equilibrium unfolding transitions monitored by far-UV CD ellipticity and intrinsic fluorescence emission changes were analysed by fitting baseline and transition region data to a two-state linear extrapolation model [51] according to

$$\Delta G_{unfolding} = \Delta G^{H_2O} + m[\text{Urea}] = -RTlnK_{unfolding} \quad (2)$$

where $\Delta G_{unfolding}$ is the free energy change for unfolding for a given denaturant concentration, ΔG^{H_2O} the free energy change for unfolding in the absence of denaturant and m a slope term

which quantifies the change in $\Delta G_{unfolding}$ per unit concentration of denaturant, R the gas constant, T the temperature and $K_{unfolding}$ the equilibrium constant for unfolding. The model expresses the signal as a function of denaturant concentration:

$$y_i = \frac{y_N + s_N[X]_i + (y_U + s_U[X]_i) * \exp[(-\Delta G^{H_2O} - m[X]_i)/RT]}{1 + \exp\left[\frac{(-\Delta G^{H_2O} - m[X]_i)}{RT}\right]} \quad (3)$$

where y_i is the observed signal, y_U and y_N are the baseline intercepts for unfolded and native protein, s_U and s_N are the baseline slopes for the unfolded and native protein, $[X]_i$ the denaturant concentration after the ith addition, ΔG^{H_2O} the extrapolated free energy of unfolding in the absence of denaturant, m the slope of a $\Delta G_{unfolding}$ versus [X] plot. The denaturant concentration at the midpoint of the transition, $[Urea]_{0.5}$, according to Eq 2, is calculated as:

$$[Urea]_{0.5} = \Delta G^{H_2O}/m \quad (4)$$

The denaturation curve obtained by plotting the fluorescence changes of the BRD2(1) variant R100L induced by increasing urea concentrations was fitted to the following equation assuming a three-state model:

$$F = \frac{F_N + \exp\left(m_{I-N}\frac{[urea]-D50_{I-N}}{RT}\right) \bullet \left(F_I + F_U \exp\left(m_{U-I}\frac{[urea]-D50_{U-I}}{RT}\right)\right)}{1 + \exp\left(m_{I-N}\frac{[urea]-D50_{I-N}}{RT}\right) \bullet \left(1 + \exp\left(m_{U-I}\frac{[urea]-D50_{U-I}}{RT}\right)\right)} \quad (5)$$

where F is $\bar{\lambda}$, calculated according to Eq 1, m is a constant that is proportional to the increase in solvent-accessible surface area between the two states involved in the transition, $D50_{I\text{-}N}$ and $m_{I\text{-}N}$ are the midpoint and m value for the transition between N and I, respectively, and $D50_{U\text{-}I}$ and $m_{U\text{-}I}$ are the midpoint and m value for the transition between I and U, respectively [52]. The $\bar{\lambda}$ of the intermediate state (I), F_I, is constant whereas that of the folded state (N) and of the unfolded state (U), F_N and F_U, respectively, has a linear dependence on denaturant concentration

$$F_N = a_N + b_N[urea] \quad (6)$$

$$F_U = a_U + b_U[urea] \quad (7)$$

where a_N and a_U are the baseline intercepts for N and U, b_N and b_U are the baseline slopes for N and U, respectively. All unfolding transition data were fitted by using Graphpad Prism 5.04.

Isothermal Titration Calorimetry

Isothermal titration calorimetry (ITC) experiments were performed at 20°C using a MicroCal VP-ITC calorimeter. BRDs proteins were extensively dialyzed with an Amicon Ultrafiltration device against the assay buffer (50 mM Hepes, pH 7.5, 100 mM NaCl, 0.5 mM TCEP). The assay buffer was also used to dilute the inhibitors 50 mM DMSO stock solutions to the experiment final concentration (15 μM). In all the experiments the inhibitors JQ1 or PFI-1 were placed in the sample cell under continuous stirring while the protein solution (ranging concentration between 200 and 250 μM) was loaded into the syringe injector. The titrations curves implied 30 injections of 6 μL at 180 s intervals. The thermodynamic data were processed with Origin 7.0 software provided by MicroCal.

Protein crystallization and structure determination

Aliquots of the purified proteins were set up for crystallization using a mosquito® crystallization robot (TTP Labtech). Coarse screens were typically setup onto Greiner 3-well plates using three different drop ratios of precipitant to protein per condition (100+50 nL, 75+75 nL and 50+100 nL). All crystallizations were carried out using the sitting drop vapour diffusion method at 4°C and were crystallized as described [14]. Crystals were cryo-protected using the well solution supplemented with additional ethylene glycol and were flash frozen in liquid nitrogen. Data were collected at diamond beamline I04 at a wavelength of 1.0121 Å. Indexing and integration was carried out using MOSFLM [53] and scaling was performed with SCALA [54]. Initial phases were calculated by molecular replacement with PHASER [55] using available structures of wild type proteins [14]. Initial models were built by ARP/wARP [56] and building was completed manually with COOT [57]. Refinement was carried out in REFMAC5 [58]. Thermal motions were analyzed using TLSMD [59] and hydrogen atoms were included in late refinement cycles.

Supporting Information

S1 Fig. Intrinsic fluorescence emission spectra of wild type bromodomains and mutants. Intrinsic fluorescence emission spectra were recorded at 20°C in 20 mM Tris/HCl, pH 7.5 containing 0.20 M NaCl and 200 μM DTT (295 nm excitation excitation wavelength) as described in Materials and Methods. The absorbance at 280 nm was 0.10 AU for all the protein solutions. Wild type spectra are shown as black solid lines and mutants are coloured as indicated in the figure.
(TIF)

S2 Fig. Thermal unfolding transition of BRD2(1) and BRD4(1) wild type and variants studied by CD spectroscopy. Wild type and variants were heated from 20°C to 80°C in a 0.1-cm quartz cuvette at 0.2 mg/ml in 20 mM Tris/HCl, pH 7.5 containing 0.20 M NaCl and 0.40 mM DTT and the molar ellipticity at 222 nm ($[\Theta_{222}]$) was monitored continuously every 0.5°C. Normalized $[\Theta_{222}]$ of BRD2(1) (A), BRD4 (1) (B); the insets show the first derivative of the same data as in (A) and in (B). (C) and (D) $[\Theta_{222}]$ before normalization.
(TIF)

S3 Fig. Thermal unfolding transition of BRD2(2), BRD3(2) and BRD4(2) wild type and variants studied by CD spectroscopy. Wild type and variants were heated from 20°C to 80°C in a 0.1-cm quartz cuvette at 0.2 mg/ml in 20 mM Tris/HCl, pH 7.5 containing 0.20 M NaCl and 0.40 mM DTT and the molar ellipticity at 222 nm ($[\Theta_{222}]$) was monitored continuously every 0.5°C. Normalized $[\Theta_{222}]$ of BRD2(2) (A), BRD3 (2) (B), BRD4(2) (C); the insets show the first derivative of the same data as in (A), (B) and in (C). (D), (E) and (F) $[\Theta_{222}]$ before normalization.
(TIF)

S4 Fig. Urea-induced equilibrium unfolding of BRD2(1) and BRD4(1) wild type and variants. (A) and (B) Normalized molar ellipticity at 222 nm ($[\Theta]_{222}$) reported after removal of the high-frequency noise and the low-frequency random error by SVD; (C) and (D) Normalized intensity-averaged emission wavelength ($\bar{\lambda}$). The continuous lines represent the nonlinear fitting of the normalized molar ellipticities at 222 nm and of the normalized intensity-averaged emission wavelength data to Eq 3, at increasing denaturant concentrations, calculated as described in Materials and Methods. The inset in (C) shows the three-state unfolding of BRD2(1) R100L variant fitted according to Eq 5. The reversibility points (empty circles) are shown, for clarity, only for the wild type and for R100L and were not included in the nonlinear regression analysis. All the spectra were recorded at 10°C as described in Materials and Methods.
(TIF)

S5 Fig. Urea-induced equilibrium unfolding of BRD2(2), BRD3(2) and BRD4(2) wild type and variants. (A), (B) and (C) Normalized molar ellipticity at 222 nm ($[\Theta]_{222}$) reported after removal of the high-frequency noise and the low-frequency random error by SVD; (D), (E) and (F) normalized intensity-averaged emission wavelength ($\bar{\lambda}$). The continuous lines represent the nonlinear fitting of the normalized molar ellipticities at 222 nm and of the normalized intensity-averaged emission wavelength data to Eq 3, at increasing denaturant concentrations, calculated as described in Materials and Methods. The reversibility points (empty circles) are shown, for clarity, only for the wild type and were not included in the nonlinear regression analysis. All the spectra were recorded at 10°C as described in Materials and Methods.
(TIF)

S6 Fig. Details of the environment of residues variants in BRD2. Shown is a structural overview (top left) and details of interactions with neighbouring residues within a radius of 6 Å. The mutated residues are shown in ball and stick representation and main structural elements are labeled.
(PDF)

S7 Fig. SDS-PAGE analysis of BRD2(1) and BRD4(1) wild type and variants. Lane 1, protein molecular mass markers; lane 2, wild type protein with His-tag; lane 3, wild type protein after overnight treatment with TEV protease; lane 4, purified wild type protein without His-tag. (A) BRD2(1) lane 5–11, purified variants without His-tag (15 kDa); (B) BRD4(1) lane 5, purified variant without His-tag (15 kDa). All the proteins were cleaved by TEV protease overnight at 4°C and purified on a His Trap column. The flow through containing the purified proteins without His-tag was collected and analyzed by SDS–PAGE. Gels were stained with Coomassie blue R-250.
(PDF)

S8 Fig. SDS-PAGE analysis of BRD2(2), BRD3(2) and BRD4(2) wild type and variants. Lane 1, protein molecular mass markers; lane 2, wild type protein with His-tag; lane 3, wild type protein after overnight treatment with TEV protease; lane 4, purified wild type protein without His-tag. (A) BRD2(2) lane 5 and 6, purified variants without His-tag (13 kDa); (B) BRD3(2) lane 5, purified variant without His-tag (13 kDa); (C) BRD4(2) lane 5, purified variant without His-tag (15 kDa). All the proteins were cleaved by TEV protease overnight at 4°C and purified on a His Trap column. The flow through containing the purified proteins without His-tag was collected and analyzed by SDS–PAGE. Gels were stained with Coomassie blue R-250.
(PDF)

S1 Table. Data collection and refinement statistics.
(PDF)

S2 Table. Dissociation constants and thermodynamic parameters from isothermal titration calorimetry assays. In all cases proteins were titrated into the ligand solution (reverse titration).
(PDF)

S3 Table. Structural location and difference in melting temperature and in free energy of urea-induced unfolding of BRDs wild type and mutants.
(PDF)

S4 Table. List of oligonucleotides used for site-directed mutagenesis.
(PDF)

Author Contributions

Conceived and designed the experiments: LL AP CL RC SK VC. Performed the experiments: LL AP CL CT. Analyzed the data: LL AP CL RC CT MP SK VC. Contributed reagents/materials/analysis tools: RC SK VC. Wrote the paper: LL AP RC SK VC.

References

1. Holliday R. The inheritance of epigenetic defects. Science. 1987; 238: 163–170 PMID: 3310230
2. Dawson MA, Kouzarides T, Huntly BJ. Targeting epigenetic readers in cancer. New England Journal of Medicine. 2012; 367: 647–657 doi: 10.1056/NEJMra1112635 PMID: 22894577
3. Filippakopoulos P, Knapp S. Targeting bromodomains: epigenetic readers of lysine acetylation. Nature Reviews Drug Discovery. 2014; 13: 337–356 doi: 10.1038/nrd4286 PMID: 24751816
4. Muller S, Filippakopoulos P, Knapp S. Bromodomains as therapeutic targets. Expert Reviews in Molecular Medicine. 2011; 13 e29 doi: 10.1017/S1462399411001992 PMID: 21933453
5. Dawson MA, Kouzarides T. Cancer epigenetics: from mechanism to therapy. Cell. 2012; 150: 12–27 doi: 10.1016/j.cell.2012.06.013 PMID: 22770212
6. Hewings DS, Rooney TP, Jennings LE, Hay DA, Schofield CJ, Brennan PE, et al. Progress in the development and application of small molecule inhibitors of bromodomain-acetyl-lysine interactions. Journal of Medicinal Chemistry. 2012; 55:9393–9413. doi: 10.1021/jm300915b PMID: 22924434
7. Kumar R, Li DQ, Müller, Knapp S. Epigenomic regulation of oncogenesis by chromatin remodeling. Oncogene. 2016 Jan 25. doi: 10.1038/onc.2015.513 [Epub ahead of print]
8. Martin LJ, Koegl M, Bader G, Cockcroft XL, Fedorov O, Fiegen D et al. Structure-based design of an in vivo active selective BRD9 inhibitor. Journal of Medicinal Chemistry. 2016; 59: 4462–4475. doi: 10.1021/acs.jmedchem.5b01865 PMID: 26914985
9. Sutherell CL, Tallant C, Monteiro OP, Yapp C, Fuchs JE, Fedorov O, et al. Identification and development of 2,3-Dihydropyrrolo[1,2-a]quinazolin-5(1H)-one inhibitors targeting bromodomains within the switch/sucrose nonfermenting complex. Journal of Medicinal Chemistry. 2016; 59: 5095–5101. doi: 10.1021/acs.jmedchem.5b01997 PMID: 27119626
10. Delmore JE, Issa GC, Lemieux ME, Rahl PB, Shi J, Jacobs HM, et al. BET Bromodomain Inhibition as a Therapeutic Strategy to Target c-Myc. Cell. 2011; 146: 904–917 doi: 10.1016/j.cell.2011.08.017 PMID: 21889194
11. Filippakopoulos P, Qi J, Picaud S, Shen Y, Smith WB, Fedorov O, et al. Selective inhibition of BET bromodomains. Nature. 2010; 468: 1067–1073 doi: 10.1038/nature09504 PMID: 20871596
12. Zuber J, Shi J, Wang E, Rappaport AR, Herrmann H, Sison EA, et al. RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. Nature. 2011; 478: 524–528 doi: 10.1038/nature10334 PMID: 21814200
13. Jung M, Gelato KA, Fernandez-Montalvan A, Siegel S, Haendler B. Targeting BET bromodomains for cancer treatment. Epigenomics. 2015; 7: 487–501 doi: 10.2217/epi.14.91 PMID: 26077433
14. Filippakopoulos P, Picaud S, Mangos M, Keates T, Lambert JP, Barsyte-Lovejoy D, et al. Histone recognition and large-scale structural analysis of the human bromodomain family. Cell. 2012; 149: 214–231 doi: 10.1016/j.cell.2012.02.013 PMID: 22464331
15. Owen DJ, Ornaghi P, Yang JC, Lowe N, Evans PR, Ballario P, et al.The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase gcn5p. The Embo Journal. 2000; 19: 6141–6149 PMID: 11080160
16. Dey A, Chitsaz F, Abbasi A, Misteli T, Ozato K. The double bromodomain protein Brd4 binds to acetylated chromatin during interphase and mitosis. Proceedings of the National Academy of Sciences of the United States of America. 2003; 100: 8758–8763 PMID: 12840145
17. Sinha A, Faller DV, Denis GV. Bromodomain analysis of Brd2-dependent transcriptional activation of cyclin A. Biochemical Journal. 2005; 387: 257–269 PMID: 15548137
18. LeRoy G, Rickards B, Flint SJ. The double bromodomain proteins Brd2 and Brd3 couple histone acetylation to transcription. Molecular Biology of the Cell. 2008; 30: 51–60
19. Nishiyama A, Dey A, Miyazaki J, Ozato K. Brd4 is required for recovery from antimicrotubule drug-induced mitotic arrest: preservation of acetylated chromatin. Molecular Biology of the Cell. 2006; 17: 814–823 PMID: 16339075
20. Yang Z, He N, Zhou Q. Brd4 recruits P-TEFb to chromosomes at late mitosis to promote G1 gene expression and cell cycle progression. Molecular Biology of the Cell. 2008; 28: 967–976

21. Belkina AC, Blanton WP, Nikolajczyk BS, Denis GV. The double bromodomain protein Brd2 promotes B cell expansion and mitogenesis. Journal of Leukocyte Biology. 2014; 95: 451–460 doi: 10.1189/jlb.1112588 PMID: 24319289

22. French CA. Pathogenesis of NUT midline carcinoma. Annual Review of Pathology. 2012; 7: 247–265 doi: 10.1146/annurev-pathol-011811-132438 PMID: 22017582

23. French CA, Kutok JL, Faquin WC, Toretsky JA, Antonescu CR, Griffin CA, et al. Midline carcinoma of children and young adults with NUT rearrangement. Journal of Clinical Oncology. 2004; 22: 4135–4139 PMID: 15483023

24. French CA, Miyoshi I, Aster JC, Kubonishi I, Kroll TG, Dal C, et al. BRD4 bromodomain gene rearrangement in aggressive carcinoma with translocation t(15;19). American Journal of Pathology. 2001; 159: 1987–1992 PMID: 11733348

25. Ishii H, Mimori K, Mori M, Vecchione A. Differentially expressed genes in endothelial differentiation. DNA and Cell Biology. 2005; 24: 432–437 PMID: 16008511

26. Zhou M, Peng C, Nie XM, Zhang BC, Zhu SG, Yu Y, et al. Expression of BRD7-interacting proteins BRD2 and BRD3 in nasopharyngeal carcinoma tissues. Ai Zheng 2003; 22: 123–127 PMID: 12600283

27. Lori C, Lantella A, Pasquo A, Alexander LT, Knapp S, Chiaraluce R, et al. Effect of single amino acid substitution observed in cancer on Pim-1 kinase thermodynamic stability and structure. PLoS One 2013; 8 e64824 doi: 10.1371/journal.pone.0064824 PMID: 23755147

28. Pasquo A, Consalvi V, Knapp S, Alfano I, Ardini M, Stefanini S, et al. Structural stability of human protein tyrosine phosphatase rho catalytic domain: effect of point mutations PLoS One 2012; 7 e32555 doi: 10.1371/journal.pone.0032555 PMID: 22389709

29. Casadio R, Vassura M, Tiwari S, Fariselli P, Luigi Martelli P. Correlating disease-related mutations to their effect on protein stability: a large-scale analysis of the human proteome. Human Mutation. 2011; 32: 1161–1170 doi: 10.1002/humu.21555 PMID: 21853506

30. Kucukkal TG, Petukh M, Li L, Alexov E. Structural and physico-chemical effects of disease and non-disease nsSNPs on proteins. Current Opinion in Structural Biology. 2015; 32: 18–24 doi: 10.1016/j.sbi.2015.01.003 PMID: 25658850

31. Fong CY, Gilan O, Lam EY, Rubin AF, Ftouni S, Tyler D, et al. BET inhibitor resistance emerges from leukaemia stem cells. Nature. 2015; 525: 538–542 doi: 10.1038/nature14888 PMID: 26367796

32. Rathert P, Roth M, Neumann T, Muerdter F, Roe JS, Muhar M, et al. Transcriptional plasticity promotes primary and acquired resistance to BET inhibition. Nature. 2015; 525: 543–547 doi: 10.1038/nature14898 PMID: 26367798

33. Forbes SA, Bindal N, Bamford S, Cole C, Kok CY, Beare D et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Research 2011; 39 D945–950 doi: 10.1093/nar/gkq929 PMID: 20952405

34. Choy N, Raussens V, Narayanaswami V. Inter-molecular coiled-coil formation in human apolipoprotein E C-terminal domain. Jornal of Molecular Biology. 2003; 334: 527–539

35. Kiss RS, Weers PM, Narayanaswami V, Cohen J, Kay CM, Ryan RO. Structure-guided protein engineering modulates helix bundle exchangeable apolipoprotein properties. Journal of Biological Chemistry. 2003; 278: 21952–21959 PMID: 12684504

36. Lamonica JM, Deng W, Kadauke S, Campbell AE, Gamsjaeger R, Wang H, et al. Bromodomain protein Brd3 associates with acetylated GATA1 to promote its chromatin occupancy at erythroid target genes. Proceedings of the National Academy of Sciences of the United States of America. 2011; 108: E159–168 doi: 10.1073/pnas.1102140108 PMID: 21536911

37. Fish PV, Filippakopoulos P, Bish G, Brennan PE, Bunnage ME, Cook AS, et al. Identification of a chemical probe for bromo and extra C-terminal bromodomain inhibition through optimization of a fragment-derived hit. Journal of Medicinal Chemistry. 2012; 55: 9831–9837 doi: 10.1021/jm3010515 PMID: 23095041

38. Picaud S, Da Costa D, Thanasopoulou A, Filippakopoulos P, Fish PV, Philpott M, et al. PFI-1 a highly selective protein interaction inhibitor targeting BET Bromodomains. Cancer Research. 2013; 73: 3336–3346 doi: 10.1158/0008-5472.CAN-12-3292 PMID: 23576556

39. Benjwal S, Verma S, Rohm KH, Gursky O. Monitoring protein aggregation during thermal unfolding in circular dichroism experiments. Protein Science. 2006; 15: 635–639 PMID: 16452626

40. Pradeep L, Udgaonkar JB. Effect of salt on the urea-unfolded form of barstar probed by m value measurements. Biochemistry. 2004; 43: 11393–11402 PMID: 15350126

41. Shortle D. Staphylococcal nuclease: a showcase of m-value effects. Advances in Protein Chemistry. 1995; 46: 217–247 PMID: 7771319

42. Wrabl J, Shortle D. A model of the changes in denatured state structure underlying m value effects in staphylococcal nuclease. Nature Structural Biology. 1999; 6: 876–883 PMID: 10467101

43. Spudich G, Marqusee S. A change in the apparent m value reveals a populated intermediate under equilibrium conditions in Escherichia coli ribonuclease HI. Biochemistry. 2000; 39: 11677–11683 PMID: 10995235

44. Geierhaas CD, Nickson AA, Lindorff-Larsen K, Clarke J, Vendruscolo M. BPPred: a Web-based computational tool for predicting biophysical parameters of proteins Protein Science. 2007; 16: 125–134 PMID: 17123959

45. Fuchs JE, Waldner BJ, Huber RG, von Grafenstein S, Kramer C, Liedl KR. Independent Metrics for Protein Backbone and Side-Chain Flexibility: Time Scales and Effects of Ligand Binding. Journal of Chemical Theory and Computation. 2015; 11: 851–860. doi: 10.1021/ct500633u PMID: 26579739

46. Gaboriau DC, Rowling PJ, Morrison CG, Itzhaki LS. Protein stability versus function: effects of destabilizing missense mutations on BRCA1 DNA repair activity. Biochemical Journal. 2015; 466: 613–624 doi: 10.1042/BJ20141077 PMID: 25748678

47. Gao M, Zhou H, Skolnick J. Insights into disease-associated mutations in the human proteome through protein structural analysis. Structure. 2015; 23: 1362–1369 doi: 10.1016/j.str.2015.03.028 PMID: 26027735

48. Engin HB, Kreisberg JF, Carter H. Structure-based analysis reveals cancer missense mutations target protein interaction interfaces. PLoS One. 2016 Apr 4; 11(4).

49. Gill SC, von Hippel PH. Calculation of protein extinction coefficients from amino acid sequence data. Analytical Biochemistry. 1989; 182: 319–326 PMID: 2610349

50. Royer CA, Mann CJ, Matthews CR. Resolution of the fluorescence equilibrium unfolding profile of trp aporepressor using single tryptophan mutants. Protein Science. 1993; 2: 1844–1852 PMID: 8268795

51. Santoro MM, Bolen DW. Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. Biochemistry 1988; 27: 8063–8068 PMID: 3233195

52. Rowling PJ, Cook R, Itzhaki LS. Toward classification of BRCA1 missense variants using a biophysical approach. Journal of Biological Chemistry 2010; 285: 20080–20087 doi: 10.1074/jbc.M109.088922 PMID: 20378548

53. Leslie AGW, Powell H. 2007; MOSFLM (Cambridge: MRC Laboratory of Molecular Biology)

54. Evans P 2007; SCALA—scale together multiple observations of reflections (Cambridge: MRC Laboratory of Molecular Biology)

55. McCoy AJ, Grosse-Kunstleve RW, Storoni LC, Read RJ. Likelihood-enhanced fast translation functions. Acta Crystallographica Section D Biological Crystallography. 2005; 61: 458–464

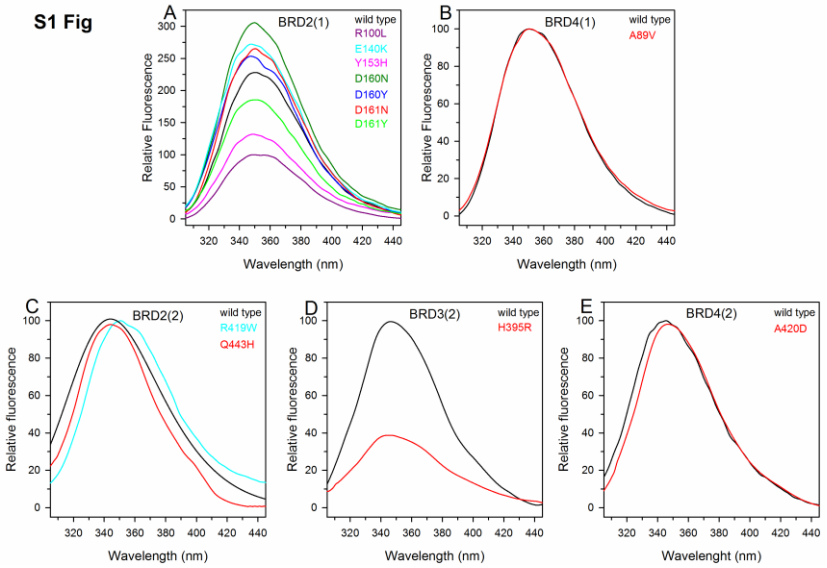
56. Perrakis A, Morris R, Lamzin VS. Automated protein model building combined with iterative structure refinement. Nature Structural Biology. 1999; 6: 458–463 PMID: 10331874

57. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallographica Section D Biological Crystallography 2004; 60: 2126–2132

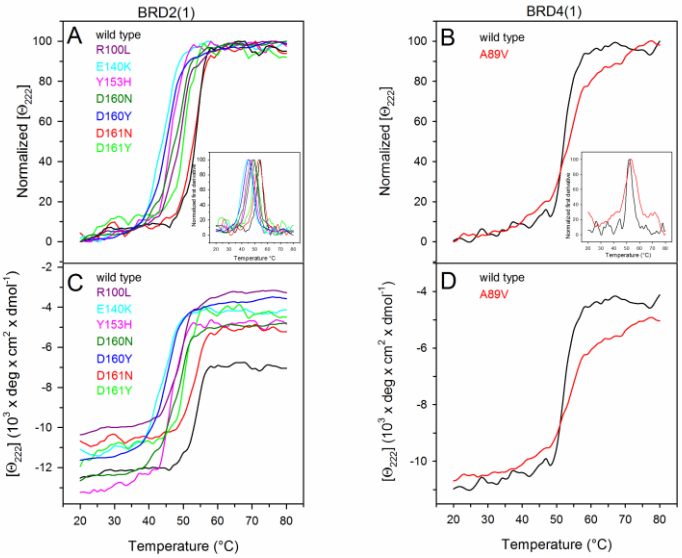
58. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallographica Section D Biological Crystallography. 1997; 53: 240–255

59. Painter J, Merritt EA. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. Acta Crystallographica Section D Biological Crystallography. 2006; 62: 439–450

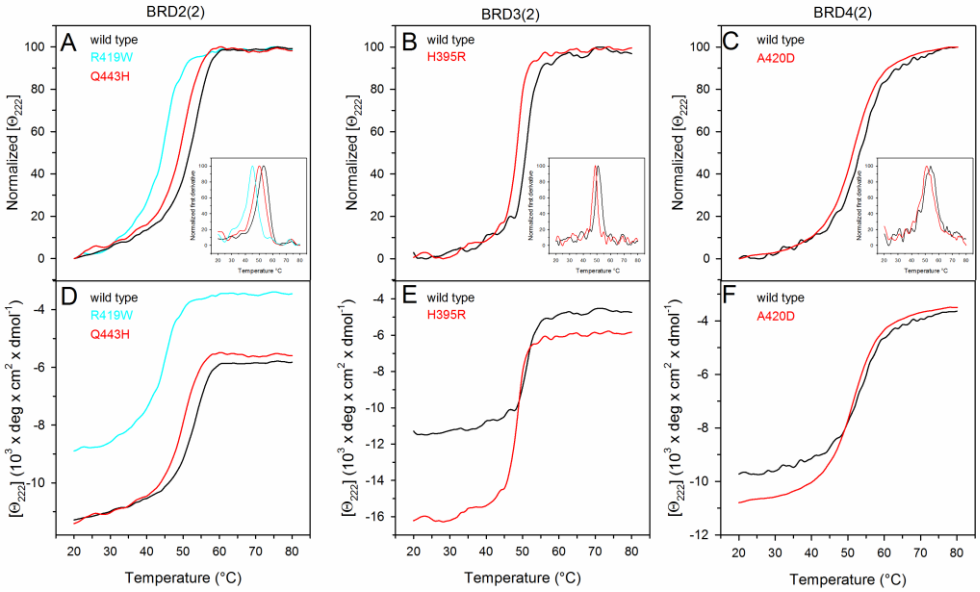
S1 Fig


A
BRD2(1)
wild type
R100L
E140K
Y153H
D160N
D160Y
D161N
D161Y
Relative Fluorescence
Wavelength (nm)
B
BRD4(1)
wild type
A89V
Relative Fluorescence
Wavelength (nm)
C
BRD2(2)
wild type
R419W
Q443H
Relative fluorescence
Wavelength (nm)
D
BRD3(2)
wild type
H395R
Relative fluorescence
Wavelength (nm)
E
BRD4(2)
wild type
A420D
Relative fluorescence
Wavelenght (nm)

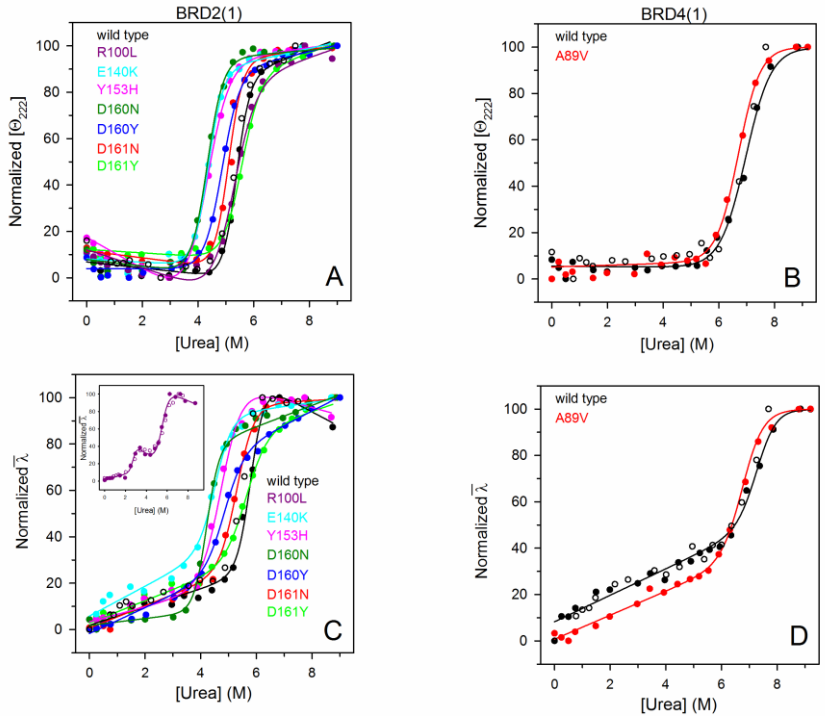
S2 Fig


BRD2(1)
BRD4(1)
A
wild type
R100L
E140K
Y153H
D160N
D160Y
D161N
D161Y
Normalized [Θ222]
B
wild type
A89V
Normalized [Θ222]
C
wild type
R100L
E140K
Y153H
D160N
D160Y
D161N
D161Y
[Θ222] (10³ x deg x cm² x dmol⁻¹)
Temperature (°C)
D
wild type
A89V
[Θ222] (10³ x deg x cm² x dmol⁻¹)
Temperature (°C)

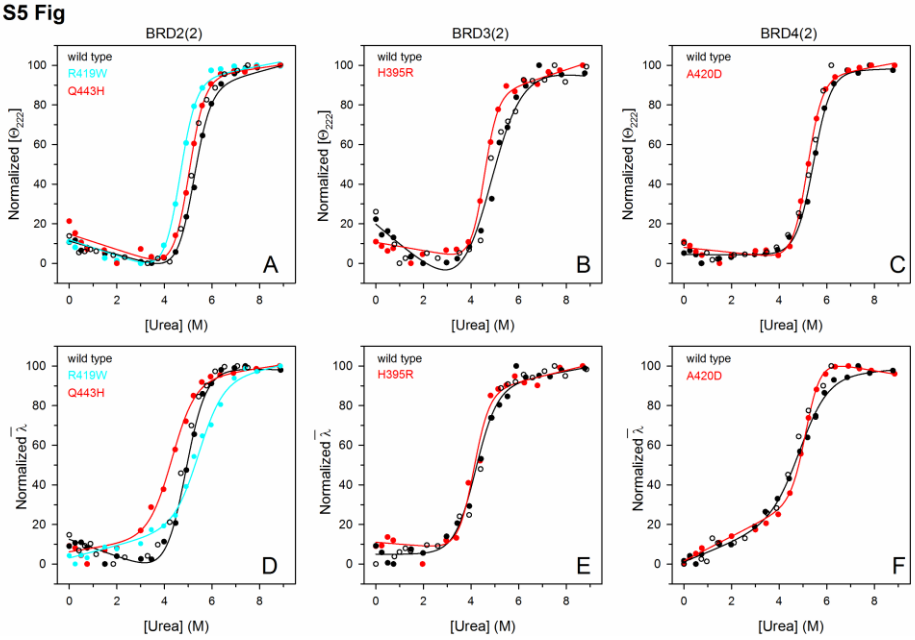
S3 Fig


BRD2(2)
BRD3(2)
BRD4(2)
A
wild type
R419W
Q443H
B
wild type
H395R
C
wild type
A420D
D
wild type
R419W
Q443H
E
wild type
H395R
F
wild type
A420D
Normalized [Θ222]
[Θ223] (10³ x deg x cm² x dmol⁻¹)
Temperature (°C)

S4 Fig


BRD2(1)
BRD4(1)
wild type
R100L
E140K
Y153H
D160N
D160Y
D161N
D161Y
A89V
A
B
C
D
Normalized [Θ222]
Normalized λ̄
[Urea] (M)

S5 Fig



S1Table. Data collection and refinement statistics

PDB ID	5HEL	5HEM	5HEN	5HFQ	5HFR
Protein / mutant	BRD2(1) / Y153H	BRD2(1) / D161Y	BRD2(1) / R100L	BRD2(2) / Q443H	BRD3(2) / H395R
Space group	$P\,4_3\,2_1\,2$	$P\,4_3\,2_1\,2$	$C\,1\,2\,1$	$P\,2\,2_1\,2_1$	$P\,2_1\,2\,2_1$
Cell dimensions:					
a, b, c (Å)	47.74 47.74 125.79	68.73 68.73 113.36	115.38 55.76 68.45	32.02 52.81 72.31	64.83 92.55 102.95
α, β, γ (deg)	90.00 90.00 90.00	90.00 90.00 90.00	90.00 94.76 90.00	90.00 90.00 90.00	90.00 90.00 90.00
Resolution* (Å)	29.75 (1.45)	29.67 (1.65)	29.44 (1.79)	29.28 (1.40)	29.55 (1.70)
Unique observations*	26847 (3793)	33564 (4733)	40528 (2236)	24828 (1130)	68935 (3165)
Completeness* (%)	99.9 (99.2)	99.7 (99.4)	99.5 (93.5)	99.7 (95.8)	99.3 (87.4)
Redundancy*	13.7 (12.9)	6.0 (4.3)	5.9 (5.5)	12.3 (7.8)	12.7 (10.3)
Rmerge*	0.044 (0.336)	0.097 (0.373)	0.063 (0.790)	0.040 (0.063)	0.234 (2.483)
I/σI*	34.2 (9.8)	13.8 (3.8)	14.2 (2.0)	48.5 (22.8)	13.0 (2.7)
Refinement					
Resolution (Å)	1.45	1.65	1.79	1.40	1.70
R_{work} / R_{free} (%)	19.09 / 20.90	16.67 (19.89)	18.08 (22.27)	16.68 (18.53)	19.37 (22.09)
Number of atoms					
(protein/other/water)	989 / 12 / 10	1982 / 29 / 214	2713 / 12 / 148	920 / 0 / 169	3634 / 20 / 402
B-factors (Å^2)					
(protein/other/water)	18.04/26.93/29.27	15.52/22.23/24.35	35.84/39.32/38.13	9.68/0/22.08	19.38/32.81/28.94
r.m.s.d bonds (Å)	0.012	0.017	0.021	0.009	0.013
r.m.s.d angles (°)	1.479	1.664	1.902	1.306	1.486
Ramachandran					
Favoured (%)	100.00	100.00	99.38	100.00	98.61
Allowed (%)	0.00	0.00	0.62	0.00	1.39
Disallowed (%)	0.00	0.00	0.00	0.00	0.00

* Values in parentheses correspond to the highest resolution shell

S2 Table. Dissociation constants and thermodynamic parameters from isothermal titration calorimetry assays. In all cases proteins were titrated into the ligand solution (reverse titration).

Comp	Protein	Mutant	K_A (10^6 M^{-1})	K_D (nM)	N	ΔH (kcal/mol)	$T\Delta S$ (kcal/mol) *	ΔG (kcal/mol)
JQ1	BRD2(1)	wt	20.9 ± 0.7	47.8 ± 1.6	0.875 ± 0.001	-8.826 ± 0.017	0.976	-9.802
	BRD2(1)	D161N	14.3 ± 0.8	69.9 ± 3.9	0.978 ± 0.002	-6.877 ± 0.025	2.673	-9.550
	BRD2(1)	D161Y	57.8 ± 3.0	17.3 ± 0.9	1.100 ± 0.001	-7.356 ± 0.016	2.997	-10.353
	BRD2(1)	Y153H	18.0 ± 0.8	55.5 ± 2.6	0.899 ± 0.002	-9.689 ± 0.031	0.044	-20.09
	BRD2(1)	E140K	26.0 ± 0.9	38.5 ± 1.4	0.933 ± 0.009	-7.574 ± 0.015	2.330	-9.904
	BRD2(1)	R100L	35.8 ± 1.7	27.9 ± 1.3	1.090 ± 0.001	-7.918 ± 0.017	2.175	-10.093
	BRD2(2)	wt	19.7 ± 1.4	50.8 ± 3.7	0.908 ± 0.002	-4.368 ± 0.019	5.331	-9.699
	BRD2(2)	Q443H	42.4 ± 3.8	23.6 ± 2.1	0.913 ± 0.002	-5.169 ± 0.002	4.985	-10.154
	BRD2(2)	R419W	33.4 ± 2.4	29.9 ± 2.2	1.100 ± 0.0019	-3.274 ± 0.011	6.714	-9.988
	BRD3(2)	wt	42.7 ± 2.4	23.4 ± 1.3	0.902 ± 0.001	-6.930 ± 0.017	3.256	-10.186
	BRD3(2)	H395R	142.8 ± 8.2	7.0 ± 0.4	1.100 ± 0.001	-6.606 ± 0.011	4.265	-10.871
	BRD4(1)	wt	55.5 ± 3.7	18.0 ± 1.2	0.912 ± 0.001	-12.020 ± 0.033	-1.605	-10.415
	BRD4(1)	A89V	97.1 ± 8.5	10.3 ± 0.9	0.969 ± 0.001	-11.140 ± 0.033	-0.423	-10.717
	BRD4(2)	wt	55.2 ± 4.6	18.1 ± 1.5	0.931 ± 0.001	-5.844 ± 0.019	4.466	-10.31
	BRD4(2)	A420D	22.5 ± 1.9	44.4 ± 3.9	0.990 ± 0.003	-6.530 ± 0.035	3.285	-9.815

(continued on next page)

S2 Table *(continued)*

Comp	Protein	Mutant	K_A (10^6 M^{-1})	K_D (nM)	N	ΔH (kcal/mol)	$T\Delta S$ (kcal/mol) *	ΔG (kcal/mol)
PFI-1	BRD2(1)	wt	12.4 ± 0.4	80.6 ± 2.9	0.924 ± 0.001	-19.230 ± 0.048	-9.538	-9.692
	BRD2(1)	D161N	4.90 ± 0.1	204.1 ± 6.6	0.843 ± 0.002	-13.670 ± 0.042	-4.609	-9.061
	BRD2(1)	D161Y	16.6 ± 0.7	60.24 ± 2.6	1.050 ± 0.001	-17.000 ± 0.044	-7.201	-9.799
	BRD2(1)	Y153H	7.4 ± 0.3	135.3 ± 4.9	1.020 ± 0.002	-17.400 ± 0.051	-8.039	-9.361
	BRD2(1)	E140K	10.7 ± 0.5	93.5 ± 4.4	1.060 ± 0.002	-14.390 ± 0.048	-4.870	-9.520
	BRD2(1)	R100L	9.2 ± 0.3	108.3 ± 3.4	0.950 ± 0.001	-14.790 ± 0.036	-5.359	-9.431
	BRD2(2)	wt	5.0 ± 0.2	198.8 ± 6.6	0.985 ± 0.002	-8.998 ± 0.027	-0.008	-8.990
	BRD2(2)	Q443H	4.7 ± 0.2	213.7 ± 7.6	0.978 ± 0.002	-9.901 ± 0.036	-0.936	-8.965
	BRD2(2)	R419W	4.2 ± 0.2	238.7 ± 10.7	1.100 ± 0.003	-6.318 ± 0.025	2.521	-8.839
	BRD3(2)	wt	6.2 ± 0.2	161.0 ± 5.9	0.976 ± 0.002	-10.540 ± 0.034	-1.403	-9.137
	BRD3(2)	H395R	0.70 ± 0.03	1386.0 ± 65.0	1.030 ± 0.007	-3.552 ± 0.032	4.235	-7.787
	BRD4(1)	wt	17.3 ± 0.8	57.8 ± 2.6	0.964 ± 0.002	-19.750 ± 0.056	-9.883	-9.867
	BRD4(1)	A89V	8.5 ± 0.2	117.1 ± 2.8	0.918 ± 0.001	-18.340 ± 0.038	-8.875	-9.465
	BRD4(2)	wt	2.8 ± 0.1	362.3 ± 17.1	1.010 ± 0.004	-10.130 ± 0.062	-1.464	-8.666
	BRD4(2)	A420D	3.4 ± 0.2	297.6 ± 14.5	0.974 ± 0.004	-10.260 ± 0.060	-1.484	-8.776

* Assay temperature = 15^0 C, 288.15 K

S3 Table. Structural location and difference in melting temperature and in free energy of urea-induced unfolding of BRDs wild type and mutants

	Position of the mutation	ΔTm (°C)	$\Delta\Delta G^{H_2O}$ (kcal/mol)	
			CD ($[\Theta]_{222}$)	Fluorescence
BRD2(1)				
Y153H	αB helix	-7.0	-6.49 ± 0.64	-5.31 ± 0.82
E140K	αB helix	-10.8	-3.48 ± 0.82	-4.53 ± 0.81
R100L	ZA-loop	-5.8	-5.22 ± 0.90	-
D160N	BC-loop	-5.8	-4.10 ± 0.82	-2.33 ± 0.91
D160Y	BC-loop	-9.8	-4.02 ± 0.76	-4.20 ± 0.79
D161N	BC-loop	-1.3	-2.21 ± 0.92	-3.02 ± 0.84
D161Y	BC-loop	-4.8	-3.38 ± 0.65	-4.66 ± 0.97
BRD4(1)				
A89V	ZA-loop	0.1	0.99 ± 0.50	-0.79 ± 0.90
BRD2(2)				
Q443H	αC helix	-3.0	0.89 ± 0.50	1.68 ± 0.40
R419W	αB helix	-8.0	0.02 ± 0.48	-0.60 ± 0.38
BRD3(2)				
H395R	BC-loop	-2.6	2.03 ± 0.63	1.12 ± 0.54
BRD4(2)				
A420D	αB helix	-3.0	0.44 ± 0.48	4.93 ± 0.38

Common acronyms and abbreviations

AD	Alzheimer's disease
ALS	Amyotrophic lateral sclerosis
ART	ADP-rybosyl transferase
ARTD	Diphteria-toxin-like ADP-rybosyl transferase
ATP	Adenosine triphosphate
BER/SSBR	Base excision repair/single strand break repair
BLI	BioLayer Interferometry
1,3-BPG	1,3-Bisphosphoglycerate
BRD	Bromodomain
CAT	C-terminal catalytic domain
CD	Circular dichroism
COSMIC	Catalogue Of Somatic Mutation In Cancer
dbSNP	Single Nucleotide Polymorphism Database
ΔG	Gibbs free energy
DTT	Dithiothreitol
$\mathbf{E_a^{a\ddagger}}$	Activation energy
EDTA	Ethylenediaminetetraacetic acid
EMT	Epithelial–mesenchymal transition
ERK1/2	Extracellular signal–regulated kinases 1/2
FBS	Fetal Seurm Bovine
FRDA	Friedreich Ataxia
FXN	Frataxin
GOF	Gain of function
GWAS	Genome-wide association studies
HapMap	The International HapMap Project

HD	Huntington’s disease
hFXN	human Frataxin
HGV	Human Genome Variation
HIF-1α	Hypoxia-inducible factor 1-alpha
HIF-2α	Hypoxia-inducible factor 2-alpha
IPTG	Isopropil-β-D-1-tiogalattopiranoside
ISC	Iron-sulfur cluster
Isd11	Nfs1 interacting protein
Isu	Iron-sulfur cluster assembly scaffold protein
ITC	Isothermal titration calorimetry
LBD	Ligand Binding Domain
LOF	Loss of function
MD	**M**olecular dynamics
MATLAB	Matrix Laboratory
MEM	Minimal Essential Medium
MgADP	Adenosine 5′-diphosphate magnesium salt
MgATP	Adenosine 5′-triphosphate magnesium salt
NaCl	Sodium chloride
NCI	National Cancer Institute
Nfs1	Cysteine desulfurase
NHEJ	Nonhomologous end-joining
NHGRI	National Human Genome Research Institute
nsSNV	Non synonymous Single Nucleotide Variant
NTR	N-terminal region
OD_{600}	Optical density
OMIM	Online Mendelian Inheritance in Man
PARP	Poly (ADP-ribose) polymerase

PD	Parkinson's disease
PDB	Protein Data Bank
PDHK1	Pyruvate dehydrogenase (acetyl-transferring)] kinase isozyme 1
3-PG	3-Phosphoglycerate
PGK1	Phosphoglycerate kinase 1
Phleo	Phleomycin
Pim-1	Serine/threonine-protein kinase pim-1
PPARγ	Peroxisome proliferator-activated receptor gamma
PTPρ	Protein tyrosine phosphatase ρ
ROS	Reactive oxygen species
SF	Stopped-flow
SNP	Single Nucleotide Polymorphism
SNV	Single Nucleotide Variant
SVD	Singular Value Decomposition
TARGET	Therapeutically Applicable Research to Generate Effective Treatments
TCGA	The Cancer Genome Atlas
TCEP	Tris(2-carboxyethyl)phosphine hydrochloride
TGFβ	Transforming growth factor beta
T_m	Melting Temperature
TMZ	Temozolomide
Tris/HCl	Tris(hydroxymethyl)aminomethane hydrochloride
UniProt	Universal Protein
WGR	Tryptophan-glycine-arginine

References

A haplotype map of the human genome. International HapMap Consortium. Nature; 437:1299–320, 2005.

Adhikari J, West GM, Fitzgerald MC. Global analysis of protein folding thermodynamics for disease state characterization. J Proteome Res.;14(5):2287-97. doi: 10.1021/acs.jproteome.5b00057, 2015.

Adinolfi S, Trifuoggi M, Politou AS, Martin S and Pastore A. A structural approach to understanding the iron-binding properties of phylogenetically different frataxins. Hum. Mol. Genet. 11, 1865–1877, 2002.

Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Current Protocols in Human Genetics.; (SUPPL.76), 2013.

Akhavan S, Miteva MA, Villoutreix BO et al. “A critical role for Gly25 in the B chain of human thrombin”. Journal of Thrombosis and Haemostasis, vol. 3, no. 1, pp. 139–145, 2005.

Akue´-Ge´du R, Rossignol E, Azzaro S, Knapp S, Filippakopoulos P, et al. Synthesis, kinase inhibitory potencies, and in vitro antiproliferative evaluation of new Pim kinase inhibitors. J Med Chem 52: 6369–6381, 2009.

Altmeyer M, Messner S, Hassa PO, Fey M and Hottiger MO. Molecular mechanism of poly(ADP-ribosyl)ation by PARP1 and identification of lysine residues as ADP-ribose acceptor sites. Nucleic Acids Res., 37, 3723–3738, 2009.

Altshuler D, Daly MJ, Lander ES. Genetic Mapping in Human Disease. Science: 881–8, 2008.

Anizon F, Shtil AA, Danilenko VN, Moreau P Fighting tumor cell survival: advances in the design and evaluation of Pim inhibitors. Curr Med Chem. 17: 4114–4133, 2010.

Ame JC, Rolli V, Schreiber V, Niedergang C, Apiou F, Decker P, Muller S, Hoger T, Menissier-de Murcia J and de Murcia G. PARP-2, A novel mammalian DNA damage-dependent poly(ADP-ribose) polymerase. J. Biol. Chem., 274, 17860–17868, 1999.

Ame JC, Spenlehauer C and de Murcia G. The PARP superfamily. Bioessays, 26, 882–893, 2004.

Aravind L and Koonin EV. SAP - a putative DNA-binding motif involved in chromosomal organization. Trends Biochem. Sci., 25, 112–114, 2000.

Arodź T, Płonka PM. Effects of point mutations on protein structure are nonexponentially distributed. Proteins, Structure, Function and Bioinformatics.; 80(7):1780–90, 2012.

Bandyopadhyay S, Chandramouli K and Johnson MK. (2008) Ironsulfur cluster biosynthesis. Biochem. Soc. Trans. 36, 1112–1119, 2008.

Barroso I, Gurnell M, Crowley VE, Agostini M, Schwabe JW, Soos MA, Maslen GL, Williams TD, Lewis H, Schafer AJ, et al. Dominant negative mutations in human PPAR associated with severe insulin resistance, diabetes mellitus and hypertension. Nature, 402, 880–883, 1999.

Beck C, Boehler C, Guirouilh Barbat J, Bonnet ME, Illuzzi G, Ronde P, Gauthier LR, Magroun N, Rajendran A, Lopez BS, Scully R, Boussin

FD, Schreiber V, Dantzer F. PARP3 affects the relative contribution of homologous recombination and nonhomologous end-joining pathways. Nucleic Acids Res. 2014 May;42(9):5616-32. doi: 10.1093/nar/gku174, 2014.

Bencze KZ, Kondapalli KC, Cook JD, McMahon S, Millan-Pacheco C, Pastor N and Stemmler TL. The structure and function of frataxin. Crit. Rev. Biochem. Mol. Biol., 41, 269–291, 2006.

Bencze KZ, Yoon T, Bradley PB, Cowan JA and Stemmler TL. Human frataxin: iron structure and ferrochelatase binding surface. Chem. Commun. 14, 1798–1800, 2007.

Bendl J, Stourac J, Salanda O, Pavelka A, Wieben ED, Zendulka J, et al. PredictSNP: Robust and Accurate Consensus Classifier for Prediction of Disease-Related Mutations. PLoS Comput Biol.; 10:e1003440. doi: 10.1371/journal.pcbi.1003440 PMID: 24453961, 2014.

Beutler E. PGK deficiency. Br. J. Haematol. 136 3–11, 2007.

Bertout JA, Patel SAS, Simon MC. The impact of O2 availability on human cancer. Nat Rev Cancer 8(12): 967–975, 2008.

Bhattacharya R, Rose PW, Burley SK, Prlić A. Impact of genetic variation on three dimensional structure and function of proteins. PLoSONE12(3):e0171355, 2017.

Bird TD, Turner JL, Sumi SM, Bierman EL. Abnormal function of endocrine pancreas and anterior pituitary in Friedreich's ataxia. Studies in a family. Ann Intern Med.;88(4):478–81, 1978.

Boehler C, Dantzer F: PARP-3, a DNA-dependent PARP with emerging roles in double-strand break repair and mitotic progression. Cell Cycle., 10: 1023-1024. 10.4161/cc.10.7.15169, 2011.

Bonetti D, Toto A, Giri R, Morrone A, Sanfelice D, Pastore A, Temussi P, Gianni S, Brunori M. The kinetics of folding of frataxin. Phys Chem Chem Phys.;16(14):6391-7. doi: 10.1039/c3cp54055c, 2014.

Boroughs L.K., DeBerardinis R.J., Metabolic pathways promoting cancer cell survival and growth. Nat. Cell Biol. 17, 351–359, 2015.

Boukouris AE, Zervopoulos SD, Michelakis ED. Metabolic Enzymes Moonlighting in the Nucleus: Metabolic Regulation of Gene Transcription. Trends Biochem Sci.;41(8):712-30. doi: 10.1016/j.tibs.2016.05.013, 2016.

Bourke T, Keane D. Friedreich's Ataxia: a review from a cardiology perspective. Ir J Med Sci.;180(4):799–805, 2011.

Brown DK, Bishop ÖT. Role of Structural Bioinformatics in Drug Discovery by Computational SNP Analysis: A Proposed Protocol for Analyzing Variation at the Protein Level. Global Heart, 2017.

Bucciantini M, Calloni G, Chiti F, Formigli L, Nosi D, Dobson CM, and Stefani M. Prefibrillar amyloid protein aggregates share common features of cytotoxicity. J Biol Chem 279: 3137482, 2004.

Burrell RA, McGranahan N, Bartek J, Swanton C. The causes and consequences of genetic heterogeneity in cancer evolution. Nature 501(7467):338–45.10.1038/nature12625, 2013.

Busi MV, Maliandi MV, Valdez H, Clemente M, Zabaleta EJ, Araya A, Gomez-Casati DF. Deficiency of Arabidopsis thaliana frataxin alters activity of mitochondrial Fe-S proteins and induces oxidative stress. Plant J. 48 (6) 873e882, 2006.

Cao R, Shi Y, Chen S, Ma Y, Chen J, Yang J, Chen G and Shi T. dbSAP: single amino-acid polymorphism database for protein variation detection, 2017.

Calmels N, Seznec H, Villa P, Reutenauer L, Hibert M, Haiech J et al. Limitations in a frataxin knockdown cell model for Friedreich ataxia in a high-throughput drug screen. BMC Neurol.; 9: 46, 2009.

Campuzano V, Montermini L, Moltò MD, et al. Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. Science.;271(5254):1423–7, 1996.

Capriotti E, Fariselli P, Calabrese R, and Casadio R. "Predicting protein stability changes from sequences using support vector machines." Bioinformatics, vol. 21, supplement 2, pp. ii54–ii58, 2005.

Capriotti E, Fariselli P and Casadio R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res. 33, W306–W310, 2005.

Cardoso JG, Andersen MR, Herrgård MJ & Sonnenschein N. Analysis of genetic variation and potential applications in genome-scale metabolic modeling. Frontiers in bioengineering and biotechnology, 3, 2015.

Casadio R, Vassura M, Tiwari S, Fariselli P, and Martelli PL. Correlating disease-related mutations to their effect on protein stability: A large-scale analysis of the human proteome. Hum. Mutat., 2, 1161–1170, 2011.

Chae YK, Anker JF, Carneiro BA, Chandra S, Kaplan J, Kalyan A, et al. Genomic landscape of DNA repair genes in cancer. Oncotarget 7(17):23312–21.10.18632/oncotarget.8196, 2016.

Chasman D, Adams RM. Predicting the functional consequences of non-synonymous single nucleotide polymorphisms: structure-based assessment of amino acid variation. J Mol Biol 307: 683–706, 2001

Chen OS, Hemenway S, Kaplan J, Inhibition of Fe-S cluster biosynthesis decreases mitochondrial iron export: evidence that Yfh1p affects Fe-S cluster synthesis. Proc. Natl. Acad. Sci. U. S. A 99 (19) 12321e12326, 2002.

Chiarelli LR, Morera SM, Bianchi P, Fermo E, Zanella A, et al. Molecular Insights on Pathogenic Effects of Mutations Causing Phosphoglycerate Kinase Deficiency. PLoS ONE 7(2): e32065. doi:10.1371/journal.pone.0032065, 2012.

Chiti F, Webster P, Taddei N, Clark A, Stefani M, Ramponi G, and Dobson CM. Designing conditions for in vitro formation of amyloid protofilaments and fibrils. Proc Natl Acad Sci U S A 96: 35904, 1999.

Chiti F, Stefani M, Taddei N, Ramponi G, and Dobson CM. Rationalization of the effects of mutations on peptide and protein aggregation rates. Nature 424: 8058, 2003.

Chiti F, Dobson CM. Amyloid formation by globular proteins under native conditions. Nat Chem Biol 5:15–22, 2009.

Chung CC, Chanock SJ. Current status of genome-wide association studies in cancer. Hum. Genet., 130: 59-78, 2011.

Choy N, Raussens V, Narayanaswami V. Inter-molecular coiled-coil formation in human apolipoprotein E C-terminal domain. J. Mol. Biol. 334:527–539. doi: 10.1016/j.jmb.2003.09.059, 2003.

Collins FS, Guyer MS & Chakravarti A. Variations on a theme: cataloging human DNA sequence variation. Science 278, 1580−1581, 1997.

Condo' I, Ventura N, Malisan F, Tomassini B, Testi R. A pool of extramitochondrial frataxin that promotes cell survival. J. Biol. Chem. 281: 16750–16756, 2006.

Condo' I, Ventura N, Malisan F, Rufini A, Tomassini B, Testi R. In vivo maturation of human frataxin. Hum. Mol. Genet. 16, 1534-1540, 2007.

Cook JD, Bencze KZ, Jankovic AD, Crater AK, Busch CN, Bradley PB, Stemmler AJ, Spaller MR and Stemmler TL. Monomeric yeast frataxin is an iron binding protein. Biochemistry 45, 7767–7777, 2006.

Correira A, Adinolfi S, Pastore A and Gomes C. Conformational stability of human frataxin and effect of Friedreich's ataxia related mutations on protein folding. Biochem. J. 398, 605-611, 2006.

Correira A, Pastore C, Adinolfi S, Pastore A and Gomes C. Implication of Friedreich's ataxia mutations on frataxin structure and dynamics. FEBS J. 275, 3680-3690, 2008.

Cossée M, Durr A, Schmitt M. et al. Friedreich's ataxia: point mutations and clinical presentation of compound heterozygotes. Ann. Neurol. 45, 200–206, 1999.

Dawicki-McKenna JM, Langelier MF, DeNizio JE, Riccio AA, Cao CD, Karch KR, McCauley M, Steffen JD, Black BE, Pascal JM. PARP-1 activation requires local unfolding of an autoinhibitory domain. Mol. Cell. 60:755–768, 2015.

D'Amours D, Desnoyers S, D'Silva I and Poirier GG. Poly(ADP-ribosyl)ation reactions in the regulation of nuclear functions. Biochem. J., 342, 249–268, 1999.

de Cristofaro R, Carotti A, Akhavan S et al. The natural mutation by deletion of Lys9 in the thrombin A-chain affects the PKa value of catalytic residues, the overall enzyme's stability and conformational transitions linked to Na^+ binding. The FEBS Journal, vol. 273, no. 1, pp. 159–169, 2006.

De Michele G, Perrone F, Filla A, et al. Age of onset, sex, and cardiomyopathy as predictors of disability and survival in Friedreich's disease: a retrospective study on 119 patients. Neurology.;47(5):1260–4, 1996.

Dhe-Paganon S, Shigeta R, Chi YI, Ristow M and Shoelson SE. Crystal structure of human frataxin. J. Biol. Chem., 275, 30753–30756, 2000.

Didonna A, Isobe N, Caillier SJ, Li KH, Burlingame AL, Hauser SL, Baranzini SE, Patsopoulos NA and Oksenberg JR. A non-synonymous single-nucleotide polymorphism associated with multiple sclerosis risk affects the EVI5 interactome. Human Molecular Genetics, Vol. 24, No. 24, 2015.

Dixit A, Yi L, Gowthaman R, Torkamani A, Schork NJ, Verkhivker GM. Sequence and structure signatures of cancer mutation hotspots in protein kinases. PLoS One.; 4(10):e7485, 2009.

Dobson CM. Protein folding and misfolding. Nature, vol. 426, no. 6968, pp. 884–890, 2003.

Dürr A, Cossee M, Agid Y, et al. Clinical and genetic abnormalities in patients with Friedreich's ataxia. N Engl J Med.;335(16):1169–75, 1996.

Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr.;60(Pt 12 Pt 1):2126–2132, 2004.

Evans DAP, Manley KA, McKusick VA. Genetic control of isoniazid metabolism in man. Br Med J.;2:485-490, 1960.

Evans PR and Murshudov GN. How good are my data and what is the resolution? Acta Cryst.;D69, 1204–1214, 2013.

Fandrich M, Fletcher MA, and Dobson CM. Amyloid fibrils from muscle myoglobin. Nature 410: 1656, 2001.

Fermo E, Bianchi P, Chiarelli LR, Maggi M, Mandara GM, Vercellati C, Marcello AP, Barcellini W, Cortelezzi A, Valentini G, and Zanella A. A new variant of phosphoglycerate kinase deficiency (p.I371K) with multiple tissue involvement: Molecular and functional characterization. Mol. Genet. Metab. 106, 455−461, 2012.

Filippakopoulos P and Knapp S Targeting bromodomains: epigenetic readers of lysine acetylation. Nature Reviews Drug Discovery AOP, published online 22 April 2014.

Finocchiaro G, Baio G, Micossi P, Pozza G, di Donato S. Glucose metabolism alterations in Friedreich's ataxia. Neurology.;38(8):1292–6, 1998.

Fiorillo A, Petrosino M, Ilari A, Pasquo A, Cipollone A, Maggi M, Chiaraluce R, Consalvi V. The phosphoglycerate kinase 1 variants found in carcinoma cells display different catalytic activity and conformational stability compared to the native enzyme. PLoS One.; 13(7):e0199191. doi: 10.1371/journal.pone.0199191, 2018.

Folker J, Murdoch B, Cahill L, Delatycki M, Corben L, Vogel A. Dysarthria in Friedreich's ataxia: a perceptual analysis. Folia Phoniatr Logop.;62(3):97–103, 2010.

Forbes SA, Tang G, Bindal N, Bamford S, Dawson E, Cole C, Kok CY, Jia M, Ewing R, Menzies A, Teague JW, Stratton MR, and Futreal PA. COSMIC (the Catalogue of Somatic Mutations in Cancer): a resource to investigate acquired mutations in human cancer. Nucleic Acids Res. 38: D6527, 2010.

Forbes SA, Bindal N, Bamford S, Cole C, Kok CY, Beare D, Jia M, Shepherd R, Leung K, Menzies A, Teague JW, Campbell PJ, Stratton MR, Futreal PA.COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Res.;39(Database issue): D945-50. doi: 10.1093/nar/gkq929, 2011.

Frazer KA, Ballinger DG, Cox DR, et al. A second generation human haplotype map of over 3.1 million SNPs. Nature; 449:851–61, 2007.

Friedreich N. Über degenerative Atrophie der spinalen Hinterstränge (On degenerative atrophy of the spinal dorsal columns). Virchows Arch Pathol Anat Physiol Klin Med.;26:391–419, 1863.

Gallois-Montbrun S, Faraj A, Seclaman E, Sommadossi JP, Deville-Bonne D, et al. Broad specificity of human phosphoglycerate kinase for antiviral nucleoside analogs. Biochem Pharmacol 68: 1749–1756, 2004.

Gao M, Zhou H, Skolnick J. Insights into Disease-Associated Mutations in the Human Proteome through Protein Structural Analysis. Structure 23, 1362–1369, 2015.

García-Closas M, Malats N, Real FX, Welch R, Kogevinas M, Chatterjee N, Pfeiffer R, Silverman D, Dosemeci M, Tardon A, Serra C, Carrato A, García-Closas R, Castano-Vinyals G, Chanock S, Yeager M, Rothman N. Genetic variation in the nucleotide excision repair pathway and bladder cancer risk. Cancer Epidemiol Biomarkers Prev 15:536–542, 2006a.

Gavande NS, VanderVere-Carozza PS, Hinshaw HD, Jalal SI, Sears CR, Pawelczak KS, et al. DNA repair targeted therapy: the past or future of cancer treatment? Pharmacol. Ther. 160:65–83.10.1016/j.pharmthera.2016.02.003, 2016.

Gerber J, Muhlenhoff U and Lill R. An interaction between frataxin and Isu1/Nfs1 that is cruciale for Fe/S cluster synthesis on Isu1. EMBO Rep., 4, 906-911, 2003.

Gfeller D, Ernst A, Jarvik N, Sidhu SS, Bader GD. Prediction and experimental characterization of nsSNPs altering human PDZ-binding motifs. PloS ONE. 9 e94507, 2014.

Gibson TJ, Koonin EV, Musco G, Pastore A and Bork A. Friedreich's ataxia protein: phylogenetic evidence for mitochondrial dysfunction. Trends Neurosci., 19, 465–468, 1996.

Gondeau C, Chaloin L, Lallemand P, Roy B, Pe´rigaud C, et al. Molecular basis for the lack of enantioselectivity of human 3-phosphoglycerate kinase. Nucleic Acids Res. 36: 3620–3629, 2008.

Gong S, Blundell TL. Structural and functional restraints on the occurrence of single amino acid variations in human proteins. PLoS One;5:e9186, 2010.

Gonzaga-Jauregui C, Lupski JR, & Gibbs RA. Human genome sequencing in health and disease. Annual review of medicine, 63, 35-6, 2012.

Goode EL, Ulrich CM, Potter JD. Polymorphisms in DNA repair genes and associations with cancer risk. Cancer Epidemiol Biomarkers Prev 11:1513–1530, 2002.

Gottesfeld JM. Small molecules affecting transcription in Friedreich ataxia. Pharmacol Ther.; 116: 236–248, 2007.

Groenendyk J, Sreenivasaiah PK, Kim do H, Agellon LB, Michalak M. Biology of endoplasmic reticulum stress in the heart. Circ Res 107:1185–1197, 2010.

Guccini I, Serio D, Condò I, Rufini A, Tomassini B, Mangiola A, Maira G, Anile C, Fina D, Pallone F, Mongiardi MP, Levi A, Ventura N, Testi R, Malisan F. Frataxin participates to the hypoxia-induced response in tumors. Cell Death Dis.; 2(2): e123. doi: 10.1038/cddis.2011.5, 2011.

Guijarro JI, Sunde M, Jones JA, Campbell ID, and Dobson CM. Amyloid fibril formation by an SH3 domain. Proc Natl Acad Sci U S A 95: 42248, 1998.

Hakmé A, Wong H, Dantzer F, Schreiber V: The expanding field of poly (ADP-ribosyl) ation reactions. “Protein modifications: beyond the usual suspects” review series. EMBO Rep. 9: 1094-1100. 10.1038/embor.2008.191, 2008.

Hamosh A, Scott AF, Amberger JS, Bocchini CA, and McKusick VA. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Res. 33: D5147, 2005.

Hanahan D, Weinberg RA. The hallmarks of cancer. Cell 100: 57–70, 2000.

Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 144: 646–674, 2010.

Harding AE. Friedreich's ataxia: a clinical and genetic study of 90 families with an analysis of early diagnostic criteria and intrafamilial clustering of clinical features. Brain.;104(3):589–620, 1981.

Harris DA and True HL. New insights into prion structure and toxicity. Neuron 50:3537, 2006.

Hottiger MO, Hassa PO, Luscher B, Schuler H and Koch-Nolte F. Toward a unified nomenclature for mammalian ADP-ribosyltransferases. Trends Biochem. Sci., 35, 208–219, 2010.

Hua et al 2014

Huynen MA, Snel B, Bork P and Gibson TJ. The phylogenetic distribution of frataxin indicates a role in iron-sulfur cluster protein assembly. Hum. Mol. Genet. 10, 2463-2468, 2001.

Hwang, TL, Liang Y, Chien KY & Yu JS. Overexpression and elevated serum levels of phosphoglycerate kinase 1 in pancreatic ductal adenocarcinoma. Proteomics 6(7), 2259–2272, 2006.

Hyeonju S, Hyundeok K, Hyun SK & Sangwoo K. Somatic mutation driven codon transition bias in human cancer. Scientific Reports | 7: 14204 | DOI:10.1038/s41598-017-14543-1;, 2017.

Ikejima M, Noguchi S, Yamashita R, Ogura T, Sugimura T, Gill DM and Miwa M. The zinc fingers of human poly(ADP-ribose) polymerase are differentially required for the recognition of DNA breaks and nicks and the consequent enzyme activation. Other structures recognize intact DNA. J. Biol. Chem., 265, 21907–21913, 1990.

Janssens ACJ, & van Duijn CM. Genome-based prediction of common diseases: advances and prospects. Human molecular genetics, 17(R2), R166-R173, 2008.

Jindal HK, Vishwanatha JK. Functional identity of a primer recognition protein as phosphoglycerate kinase. J Biol Chem 265: 6540–6543, 1990.

Johnson AD, Zhang Y, Papp AC, Pinsonneault JK, Lim JE, Saffen D, ... & Sadée W. Polymorphisms affecting gene transcription and mRNA processing in pharmacogenetic candidate genes: detection through allelic expression imbalance in human target tissues. Pharmacogenetics and genomics, 18(9), 781, 2008.

Jones R, Ruas M, Gregory F et al."A CDKN2A mutation in familial melanoma that abrogates binding of p16 INK4a to CDK4 but not CDK6". Cancer Research, vol. 67, no. 19, pp. 9134–9141, 2007.

Jorde LB, Wooding SP. Genetic variation, classification and 'race'. Nat Genet.;36:S28–33, 2004.

Kabsch W. XDS. Acta Crystallogr D Biol Crystallogr. ;D66:125–132, 2010.

Kabbage M, Chahed K, Hamrita B, Guillier CL, Trimeche M, Remadi S, Hoebeke J, Chouchane L. Protein alterations in infiltrating ductal carcinomas of the breast as detected by nonequilibrium pH gradient electrophoresis and mass spectrometry. J Biomed Biotechnol: 564127, 2008.

Karchin R. Next generation tools for the annotation of human SNPs. Brief Bioinform 10: 3552, 2009.

Karchin R, Diekhans M, Kelly L et al. LS-SNP: largescale annotation of coding non-synonymous SNPs based on multiple information sources. Bioinformatics, vol. 21, no. 12, pp. 2814–2820, 2005

Karicheva O, Rodriguez-Vargas JM, Wadier N, Martin-Hernandez K, Vauchelles R, Magroun N, Tissier A, Schreiber V, Dantzer F. PARP3 controls TGFβ and ROS driven epithelial-to-mesenchymal transition and stemness by stimulating a TG2-Snail-E-cadherin axis. Oncotarget.; 7(39): 64109–64123. doi: 10.18632/oncotarget.11627, 2016.

Kersten S, Desvergne B, Wahli W. Roles of PPARs in health and disease. Nature, 405, 421–424, 2000.

Khan RH, Chaturvedi D, Mahalakshmi R. Methionine mutations of outer membrane protein X influence structural stability and beta-barrel unfolding. PLoS ONE, 8:e79351, 2013.

Khemtemourian L, Killian JA, Hoppener JW, and Engel MF. Recent insights in islet amyloid polypeptideinduced membrane disruption and its role in βcell death in type 2 diabetes mellitus. Exp Diabetes Res.: 421287, 2008.

Kiss RS, Weers PM, Narayanaswami V, Cohen J, Kay CM, Ryan RO. Structure-guided protein engineering modulates helix bundle exchangeable

apolipoprotein properties. J. Biol. Chem.; 278:21952–21959. doi: 10.1074/jbc.M302676200, 2003.

Kleine H, Poreba E, Lesniewicz K, Hassa PO, Hottiger MO, Litchfield DW, Shilton BH and Luscher B. Substrate-assisted catalysis by PARP10 limits its activity to mono-ADP-ribosylation. Mol. Cell, 32, 57–69, 2008.

Koutnikova H, Campunzano V and Koenig M. Maturation of wild-type and mutated frataxin by the mitochondrial processing peptidase. Hum. Mol. Genet. 7, 1485-1489, 1998.

Koukouritaki SB, Poch MT, Henderson MC et al. Identification and functional analysis of common human flavin-containing monooxygenase 3 genetic variants. Journal of Pharmacology and Experimental Therapeutics, vol. 320, no. 1, pp. 266–273, 2007.

Krishnan P, Gullen EA, Lam W, Dutschman GE, Grill SP, et al. Novel role of 3-phosphoglycerate kinase, a glycolytic enzyme, in the activation of L-nucleoside analogs, a new class of anticancer and antiviral agents. J Biol Chem 278: 36726–36732, 2003.

Kroncke BM, Vanoye CG, Meiler J, George AL Jr, Sanders CR. Personalized biochemistry and biophysics. Biochemistry; 54(16):2551-9. doi: 10.1021/acs.biochem.5b00189, 2015.

Kunz E, Rothammer S, Pausch H, Schwarzenbacher H,. Seefried FR, Matiasek K, Seichter D, Russ I, Fries R and Medugorac I. Confirmation of a non-synonymous SNP in PNPLA8 as a candidate causal mutation for Weaver syndrome in Brown Swiss cattle. Genet Sel Evol 48:21 DOI 10.1186/s12711-016-0201-5, 2016.

Labelle H, Tohmé S, Duhaime M, Allard P. Natural history of scoliosis inFriedreich's ataxia. J Bone Joint Surg Am.;68(4):564–72, 1986.

Labuda M, Poirier J and Pandolfo M. A missense mutation (W155R) in an American patient with Friedreich ataxia. Hum. Mutat. 13, 506–507, 1999.

Lander, E.S. The new genomics: global views of biology. Science 274, 536–539, 1996.

Langelier MF, Servent KM, Rogers EE and Pascal JM. A third zinc-binding domain of human poly(ADP-ribose) polymerase-1 coordinates DNA-dependent enzyme activation. J. Biol. Chem., 283, 4105–4114, 2008.

Langelier MF, Planck JL, Roy S and Pascal JM. Structural basis for DNA damage-dependent poly(ADP-ribosyl)ation by human PARP-1. Science, 336, 728–732, 2012.

Langelier MF and Pascal JM. PARP-1 mechanism for coupling DNA damage detection to poly(ADP-ribose) synthesis. Curr. Opin. Struct. Biol., 23, 134–143, 2013.

Lay AJ, Jiang XM, Kisker O, Flynn E, Underwood A, et al. Phosphoglycerate kinase acts in tumour angiogenesis as a disulphide reductase. Nature 408: 869–873, 2000.

Lee JE, Cho KE, Lee KE, Kim J, Bae YS. Nox4-mediated cell signaling regulates differentiation and survival of neural crest stem cells. Mol Cells.;37:907–11. doi: 10.14348/molcells.2014.0244, 2014.

Leger K, Bar D, Savic N, Santoro R and Hottiger MO. ARTD2 activity is stimulated by RNA. Nucleic Acids Res., 42, 5072–5082 2014.

Lehrke, M.; Lazar, M.A. The many faces of PPAR. Cell., 123, 993–999, 2005.

Levi F, Pasche C, La Vecchia C, Lucchini F, Franceschi S. Food groups and colorectal cancer risk. Br J Cancer; 79(7–8):1283–7, 1999.

Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, et al. The diploid genome sequence of an individual human. PLoS Biol.; 5:e254, 2007.

Li X, Jiang Y, Meisenhelder J, Yang W, Hawke DH, Zheng Y, Xia Y, Aldape K, He J, Hunter T, Wang L, Lu Z. Mitochondria-Translocated PGK1 Functions as a Protein Kinase to Coordinate Glycolysis and the TCA Cycle in Tumorigenesis. Mol Cell.; 61(5):705-19, 2016.

Li X, Tian X, Zhang B, Zhang Y, Chen J. Variation in dicer gene is associated with increased survival in T-cell lymphoma. PLoS One; 7(12): e51640. doi: 10.1371/journal.pone.0051640, 2012.

Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, et al. Environmental and heritable factors in the causation of cancer – analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med; 343(2):78–85, 2000.

Lindquist SL and Kelly JW. Chemical and biological approaches for adapting proteostasis to ameliorate protein misfolding and aggregation diseases: progress and prognosis. Cold Spring Harb. Perspect. Biol. 3, a004507, 2011.

Lori C, Lantella A, Pasquo A, Alexander LT, Knapp S, Chiaraluce R, Consalvi V. Effect of single amino acid substitution observed in cancer on Pim-1 kinase thermodynamic stability and structure. PLoS One.;8(6): e64824. doi: 10.1371/journal.pone.0064824, 2013.

Lori C, Pasquo A, Montanari R, Capelli D, Consalvi V, Chiaraluce R, Cervoni L, Loiodice F, Laghezza A, Aschi M et al. Structural basis of the transactivation deficiency of the human PPARγ F360L mutant associated with familial partial lipodystrophy. Acta. Crystallogr. D Biol. Crystallogr., 70, 1965–1976, 2014.

Lu Z, and Hunter T. Prolyl isomerase Pin1 in cancer. Cell Res. 24, 1033–1049, 2014.

Lunt SY, Vander Heiden MG. Aerobic glycolysis: Meeting the metabolic requirements of cell proliferation. Annu. Rev. Cell Dev. Biol. 27, 441–464 2011.

MacArthur DG, Manolio TA, Dimmock DP, Rehm HL, Shendure J, Abecasis GR, Adams DR, Altman RB, Antonarakis SE, Ashley EA, Barrett JC, Biesecker LG, Conrad DF, Cooper GM, Cox NJ, Daly MJ, Gerstein MB, Goldstein DB, Hirschhorn JN, Leal SM, Pennacchio LA, Stamatoyannopoulos JA, Sunyaev SR, Valle D, Voight BF, Winckler W, Gunter C. Guidelines for investigating causality of sequence variants in human disease. Nature.; 508(7497): 469–476, 2014.

Marmolino D, Acquaviva F, Pinelli M, Monticelli A, Castaldo I, Filla A et al. PPAR-gamma agonist Azelaoyl PAF increases frataxin protein and mRNA expression: new implications for the Friedreich's ataxia therapy. Cerebellum; 8: 98–103, 2009.

Martin-Hernandez K, Rodriguez-Vargas JM, Schreiber V, Dantzer F. Expanding functions of ADP-ribosylation in the maintenance of genome integrity. Semin. Cell Dev. Biol. 63:92–101, 2016.

McCarrey JR, Thomas K. Human testis-specific PGK gene lacks introns and possesses characteristics of a processed gene. Nature 326: 501–505, 1987.

McCarthy MI, MacArthur DG. Human disease genomics: from variants to biology. Genome biology, 18(1), 20, 2017.

Mooney S. Bioinformatics approaches and resources for single nucleotide polymorphism functional analysis. Brief Bioinform 6: 4456, 2005.

Mueller E, Drori S, Aiyer A, et al. Genetic analysis of adipogenesis through peroxisome proliferator-activated receptor g isoforms. J. Biol. Chem.; 277:41925 – 30, 2015.

Mueller E, Smith M, Sarraf P, et al. Effects of ligand activation of peroxisome proliferator-activated receptor g in human prostate cancer. Proc Natl Acad Sci U S A 2000; 97:10990 – 5 2000.

Mueller SC, Backes C, Kalinina OV, Meder B, Stöckel D, Lenhof HP, Meese E and Keller A. BALL-SNP: combining genetic and structural information to identify candidate non-synonymous single nucleotide polymorphisms. 2015.

Muhlenhoff U, Richhardt N, Ristow M, Kispal G, Lill R, The yeast frataxin homolog Yfh1p plays a specific role in the maturation of cellular Fe/S proteins. Hum. Mol. Genet. 11 (17) 2025e2036, 2002.

Müller S, Filippakopoulos P, Knapp S. Bromodomains as therapeutic targets. Expert Rev Mol Med. 13, e29, 2011.

Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr.;53(Pt 3):240–255. pmid:15299926, 1997.

Musco G, Stier G, Kolmerer B, Adinolfi S, Martin S, Frenkiel T, Gibson T and Pastore A. Towards a structural understanding of Friedreich's ataxia: the solution structure of frataxin Structure. Fold Des. 8, 695–707, 2000.

Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. Genome Research.; 11 863–74. https://doi.org/10.1101/gr.176601 PMID: 11337480, 2001.

Ng PC, Henikoff S. Accounting for human polymorphisms predicted to affect protein function. Genome Res 12: 436–446, 2002.

Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acid Research.; 31(13):3812–4. PMID: 12824425, 2003.

Ng PC and Henikoff S. Predicting the effects of amino acid substitutions on protein function. Annu. Rev. Genomics Hum. Genet. 7, 61–80, 2006.

Obaji E, Haikarainen T, Lehtio L. Characterization of the DNA dependent activation of human ARTD2/PARP2. Sci. Rep. 6:34487, 2016.

Ode H, Matsuyama S, Hata M et al. Computational characterization of structural role of the non-active site mutation M36I of human immunodeficiency virus type 1 protease. Journal of Molecular Biology, vol. 370, no. 3, pp. 598–607, 2007.

Oliveberg M, Tan YJ, Silow M, Fersht AR. The changing nature of the protein folding transition state: implications for the shape of the free-energy profile for folding. J Mol Biol.; 277(4):933-43, 1998.

Oliver AW, Ame JC, Roe SM, Good V, de Murcia G, Pearl LH. Crystal structure of the catalytic fragment of murine poly(ADP-ribose) polymerase-2. Nucleic Acids Res. 32:456–464, 2004.

Otto H, Reche PA, Bazan F, Dittmar K, Haag F, Koch-Nolte F. In silico characterization of the family of PARP-like poly(ADP-ribosyl)transferases (pARTs). BMC Genomics. 2005 Oct 4; 6:139. doi: 10.1186/1471-2164-6-139, 2005.

Ozbabacan SEA, Gursoy A, Keskin O, and Nussinov R. "Conformational ensembles, signal transduction and residue hot spots: application to drug discovery". Current Opinion in Drug Discovery and Development, vol. 13, no. 5, pp. 527–537, 2010.

Oktay Y, Dioum E, Matsuzaki S, Ding K, Yan LJ, Haller RG et al. Hypoxia-inducible factor 2alpha regulates expression of the mitochondrial aconitase chaperone protein frataxin. J. Biol. Chem.; 282: 11750–11756, 2007.

Pace CN, Hebert EJ, Shaw KL, Schell D, Both V, Krajcikova D, Sevcik J, Wilson KS, Dauter Z, Hartley RW and Grimsley GR. Conformational stability and thermodynamics of folding of ribonucleases Sa, Sa2 and Sa3. J. Mol. Biol. 279, 271–286, 1998.

Palmai Z, Chaloin L, Lionne C, Fidy J, Perahia D, et al. Substrate binding modifies the hinge bending characteristics of human 3-phosphoglycerate kinase: a molecular dynamics study. Proteins 77: 319–329, 2009.

Pascal JM, Ellenberger T. The rise and fall of poly(ADP-ribose): an enzymatic perspective. DNA Repair. 32:10–16, 2015.

Pasquo A, Consalvi V, Knapp S, Alfano I, Ardini M et al. Structural Stability of Human Protein Tyrosine Phosphatase ρ Catalytic Domain: Effect of Point Mutations. PLoS ONE 7(2): e32555. doi:10.1371/journal.pone.0032555, 2012.

Pastore A and Puccio H. Frataxin: a protein in search for a function. J. Neurochem. 126, 43-52; doi: 10.1111/jnc.12220, 2013.

Petukh M, Kucukkal TG, Alexov E. On Human Disease-Causing Amino Acid Variants: Statistical Study of Sequence and Structural Patterns. Hum Mutat 36:524–534, 2015.

Pires DE, Ascher DB, Blundell TL. mCSM: predicting the effects of mutations in proteins using graph-based signatures. Bioinformatics, 335-342, 2013.

Pradeep L, Udgaonkar JB. Effect of salt on the urea-unfolded form of barstar probed by m value measurements. Biochemistry; 43: 11393–11402, 2004.

Py B, Moreau PL. and Barras F. Fe-S clusters, fragile sentinels of the cell. Curr. Opin. Microbiol. 14, 218–223, 2011.

Qian X, Li X, Cai Q, Zhang C, Yu Q, Jiang Y, Lee JH, Hawke D, Wang Y, Xia Y, Zheng Y, Jiang BH, Liu DX, Jiang T, Lu Z. Correspondence information about the author Zhimin Lu Phosphoglycerate Kinase 1 Phosphorylates Beclin1 to Induce Autophagy. Molecular Cell, Volume 65, Issue 5, 917 - 931.e6, 2017.

Quan JJ, Song JN, Qu JQ. PARP3 interacts with FoxM1 to confer glioblastoma cell radioresistance. Tumour Biol.;36(11):8617-24. doi: 10.1007/s13277-015-3554-4, 2015.

Ramazzotti A, Vanmansart V and Foury F. Miyochondrial functional interactions between frataxin and Isu1p, the iron-sulfur cluster scaffold protein, in *Saccharomyces cerevisiae*. FEBS Lett. 557, 215-220, 2004.

Ramensky V, Bork P, and Sunyaev S. Human nonsynonymous SNPs: server and survey. Nucleic Acids Research, vol. 30, no. 17, pp. 3894–3900, 2002.

Richardson DR, Kalinowski DS, Lau S, Jansson PJ, Lovejoy DB. Cancer cell iron metabolism and the development of potent iron chelators as anti-tumour agents. Biochem Biophys Acta 1790:702–717. doi:10.1016/j.bbagen.2008. 04.003, 2009.

Richardson DR, Huang ML, Whitnall M, Becker EM, Ponka P, Rahmanto YS. The ins and outs of mitochondrial iron-loading: the metabolic defect in Friedreich's ataxia. J Mol Med; 88: 323–329, 2010.

Rignall TR, Baker JO, McCarter SL et al. "Effect of single active-site cleft mutation on product specificity in a thermostable bacterial cellulase". Applied Biochemistry and Biotechnology A, vol. 98-100, pp. 383–394, 2002.

Risch N. & Merikangas K. The future of genetic studies of complex human diseases. Science 273, 1516−1517. 1996.

Robinson PA. Protein stability and aggregation in Parkinson's disease. Biochem J 413: 113, 2008.

Rose PW, Prlić A, Bi C, Bluhm WF, Christie CH, Dutta S, Green RK, Goodsell DS, Westbrook JD, Woo J, Young J, Zardecki C, Berman HM, Bourne PE, Burley SK. The RCSB Protein Data Bank: views of structural biology for basic and applied research and education. Nucleic Acids Res. 2015 Jan;43(Database issue):D345-56. doi: 10.1093/nar/gku1214, 2015.

Rouleau M, McDonald D, Gagné P, Ouellet M, Droit A, Hunter JM, Dutertre S, Prigent C, Hendzel MJ, Poirier GG: PARP-3 associates with polycomb

group bodies and with components of the DNA damage repair machinery. J Cell Biochem., 100: 385-401. 10.1002/jcb.21051, 2007.

Rouleau M, Saxena V, Rodrigue A, Paquet ER, Gagnon A, Hendzel MJ, Masson JY, Ekker M, Poirier GG. A Key Role for Poly(ADP-Ribose) Polymerase 3 in Ectodermal Specification and Neural Crest Development. PLoS One.;6:e15834, 2011.

Royer CA, Mann CJ, Matthews CR. Resolution of the fluorescence equilibrium unfolding profile of trp aporepressor using single tryptophan mutants. Protein Science.; 2: 1844–1852, 1993.

Santoro MM, Bolen DW. Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. Biochemistry; 27: 8063–8068, 1998.

Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, et al. A map of human genome sequence variation containi. Nat. Genet., 2004.

Saris CJ, Domen J, Berns A. The pim-1 oncogene encodes two related protein-serine/threonine kinases by alternative initiation at AUG and CUG. EMBO J 10: 655–664, 1991.

Sauer, S. Ligands for the nuclear peroxisome proliferator-activated receptor . Trends Pharmacol. Sci., 36, 688–704, 2015.

Schulz TJ, Thierbach R, Voigt A, Drewes G, Mietzner B, Steinberg P, et al. Induction of oxidative metabolism by mitochondrial frataxin inhibits cancer growth: Otto Warburg revisited. J Biol Chem.; 281:977–981, 2006.

Schmucker S, Argentini M, Carelle-Calmes N, Martelli A and Puccio H. The in vivo mitochondrial two-step maturation of human frataxin. Hum. Mol. Genet. 17, 3521-3531, 2008.

Schmucker S, Martelli A, Colin F, Page A, Wattenhofer-Donze M, Reutenauer L and Puccio H. Mammalian frataxin: an essential function for cellular viability through an interaction with a preformed ISCU/NFS1/ISD11 iron-sulfur assembly complex. PLoS ONE 6, e16199, 2011.

Sherman MA, Fairbrother WJ, Mas MT. Characterization of the structure and properties of the His62-fAla and Arg38fAla mutants of yeast phosphoglycerate kinase: an investigation of the catalytic and activatory sites by site-directed mutagenesis and NMR. Prot Sci.; 1:752-760, 1992.

Shi Z, and Moult J. Structural and functional impact of cancer-related missense somatic mutations. J. Mol. Biol. 413, 495−512, 2011.

Shichijo S, Azuma K, Komatsu N, Ito M, Maeda Y, et al. Two proliferation-related proteins, TYMS and PGK1, could be new cytotoxic T lymphocyte-directed tumor-associated antigens of HLA-A2+ colon cancer. Clin Cancer Res 10: 5828–5836, 2004.

Shihab Hashem A, Rogers Mark F, Gough Julian, Mort Matthew, Cooper David N, Ian NM Day, Gaunt Tom R, Campbell Colin. An integrative approach to predicting the functional effects of non-coding and coding sequence variation. Bioinformatics, Volume 31, Issue 10, Pages 1536–1543, 2015.

Spiegelman BM. Fat and beyond: The diverse biology of PPAR. Annu. Rev. Biochem., 77, 289–312, 2008.

Stefl S, Nishi H, Petukh M, Panchenko AR and Alexov E. Molecular mechanisms of disease-causing missense mutations. J. Mol. Biol. 425, 3919−3936, 2013.

Stehling O, Elsasser HP, Bruckel B, Muhlenhoff U, Lill R. Iron-sulfur protein maturation in human cells: evidence for a function of frataxin. Hum Mol Genet.; 13:3007–3015, 2004.

Stemmler TL, Lesuisse E, Pain D, Dancis A. Frataxin and mitochondrial Fe-S cluster biogenesis. J. Biol. Chem.; 285: 26737–26743, 2010.

Stenson PD, Ball E, Howells K, Phillips A, Mort M, Cooper DN. Human Gene Mutation Database: towards a comprehensive central mutation database. J Med Genet.; 45(2):124-6. doi: 10.1136/jmg.2007.055210, 2008.

Stevanin G, Hahn V, Lohmann E et al. Mutation in the catalytic domain of protein kinase C γ and extension of the phenotype associated with spinocerebellar ataxia type 14. Archives of Neurology, vol. 61, no. 8, pp. 1242–1248, 2004.

Stone EA, Sidow A. Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. Genome Research.; 15(7):978–86, 2005.

Studer Romain A, Dessailly Benoit H , Orengo Christine A. Residue mutations and their impact on protein structure and function: detecting beneficial and pathogenic changes. Biochemical Journal, 449(3)581-594; DOI: 10.1042/BJ20121221, 2013.

Sun S, Liang X, Zhang X, Liu T, Shi Q, Song Y, Jiang Y, Wu H, Jiang Y, Lu X, Pang D. Phosphoglycerate kinase-1 is a predictor of poor survival and a

novel prognostic biomarker of chemoresistance to paclitaxel treatment in breast cancer. Br J Cancer.; 112(8):1332-9. doi: 10.1038/bjc.2015.114, 2015.

Takamiya O, Seta M, Tanaka K, and Ishida F. “Human factor VII deficiency caused by S339C mutation located adjacent to the specificity pocket of the catalytic domain”. Clinical and Laboratory Haematology, vol. 24, no. 4, pp. 233– 238, 2002.

Tanford C. Protein denaturation part C: Theoretical models for the mechanism of denaturation. Adv. Protein. Chem., 24:1{95, 1970.

Thierbach R, Schulz TJ, Isken F, Voigt A, Mietzner B, Drewes G, et al. Targeted disruption of hepatic frataxin expression causes impaired mitochondrial function, decreased life span and tumor growth in mice. Hum Mol Genet.; 14:3857–3864, 2005.

Thusberg J, Vihinen M. Pathogenic or not? And if so, then how? Studying the effects of missense mutations using bioinformatics methods. HumMutat 30:703–714, 2009.

Torti SV, Torti FM. Ironing out cancer. Can. Res. 71:1511–1514. doi:10.1158/0008-5472.CAN-10-3614, 2011.

Torti SV, Torti FM. Cellular iron metabolism in prognosis and therapy of breast cancer. Crit. Rev. Oncog. 18:435–448, 2013.

Tsai CI and Barondeau DP. Human frataxin is an allosteric switch that activates the Fe-S cluster biosynthetic complex. Biochemistry 49, 9132–4139, 2010.

Vagin A, Teplyakov A. MOLREP: an automated program for molecular replacement. J. Appl. Cryst.; 30: 1022–1025, 1997.

Valastyan Julie S. and Lindquist Susan. Mechanisms of protein folding diseases at a glance. Disease Models & Mechanisms 7, 9-14 doi:10.1242/dmm.013474, 2014.

Vas M, Varga A, Graczer E. Insight into the mechanism of domain movements and their role in enzyme function: example of 3-phosphoglycerate kinase. Curr Protein Pept Sci 11: 118–147, 2010.

Vendruscolo M, Zurdo J, MacPhee CE, Dobson CM. Protein folding and misfolding: a paradigm of self-assembly and regulation in complex biological systems. Phil. Trans. R. Soc. Lond. A.; 361: 1205–1222, 2003.

Wang Z. and Moult J. "Three-dimensional structural location and molecular functional effects of missense SNPs in the T cell receptor Vβ domain" Proteins, vol. 53, no. 3, pp. 748– 757, 2003.

Wang Z.and Moult J. "SNPs, protein structure, and disease" Human Mutation, vol. 17, no. 4, pp. 263–270, 2001

Wang J, Wang J, Dai J, Jung Y, Wei CL, Wang Y, Havens AM, Hogg PJ, Keller ET, Pienta KJ, Nor JE, Wang CY & Taichman RS. A glycolytic mechanism regulating an angiogenic switch in prostate cancer. Cancer Res. 67, 149-159, 2007.

Wang S, Awad KS, Elinoff JM, Dougherty EJ, Ferreyra GA, Wang JY, ... & Danner RL. G Protein-coupled Receptor 40 (GPR40) and Peroxisome Proliferator-activated Receptor γ (PPARγ) AN INTEGRATED TWO-RECEPTOR SIGNALING PATHWAY. Journal of Biological Chemistry, 290(32), 19544-19557, 2015.

Wang T, Xu J, Yu X, Yang R, Han ZC. Peroxisome proliferator-activated receptor in malignant diseases. Crit. Rev. Oncol. Hematol. 58, 1–14, 2006.

Wang X, Sun Y, Wong J, & Conklin DS. PPARγ maintains ERBB2-positive breast cancer stem cells. Oncogene, 32(49), 5512, 2013.

Wang Z, Shen D, Parsons DW, Bardelli A, Sager J, et al. Mutational analysis of the tyrosine phosphatome in colorectal cancers. Science 304: 1164–1166, 2004.

Wei H, Yu X. Functions of PARylation in DNA damage repair pathways. Genomics Proteomics Bioinformatics. 14:131–139, 2016.

Willard HF, Goss SJ, Holmes MT, Munroe DL. Regional localization of the phosphoglycerate kinase gene and pseudogene on the human X chromosome and assignment of a related DNA sequence to chromosome 19. Hum Genet 71: 138–143, 1985.

Wrabl J, Shortle D. A model of the changes in denatured state structure underlying m value effects in staphylococcal nuclease. Nature Structural Biology; 6: 876–883, 1999.

Wu Y, Wu Q, Zhang H, Chen C, Chen G, Yang H, Qin D, Fu H. Lack of genetic associations between PPAR-γ gene rs1801282 polymorphism and Alzheimer's disease in general population: A meta-analysis. Gene.; 563:120-4, 2015.

Xinjian Li et al. Mitochondria-translocated phosphoglycerate kinase 1 functions as a protein kinase to coordinate glycolysis and TCA cycle in tumorigenesis. Mol Cell.; 61(5): 705–719. doi:10.1016/j.molcel.2016.02.009, 2016.

Yamada Y, Banno Y, Yoshida H et al. “Catalytic inactivation of human phospholipase D2 by a naturally occurring Gly901Asp mutation”. Archives of Medical Research, vol. 37, no. 6, pp. 696–699, 2006.

Yankner BA and Lu T. Amyloid β-protein toxicity and the pathogenesis of Alzheimer disease. J. Biol. Chem. 284: 47559, 2009.

Yates CM, MJ Sternberg. The effects of non-synonymous single nucleotide polymorphisms (nsSNPs) on protein-protein interactions. J. Mol. Biol. 425 3949–3963, 2013.

Ye Y, Li Z and Godzik A. “Modeling and analyzing threedimensional structures of human disease proteins” Pacific Symposium on Biocomputing, pp. 439–450, 2006.

Yue P, Li Z and Moult J. Loss of Protein Structure Stability as a Major Causative Factor in Monogenic Disease. J. Mol. Biol. 353, 459–473, 2005.

Yue P, Moult J. Identification and Analysis of Deleterious Human SNPs. J Mol Biol. 356: 1263–1274. PMID: 16412461, 2006.

Zhang CZ, Spektor A, Cornils H, Francis JM, Jackson EK, Liu S3, Meyerson M, Pellman D. Chromothripsis from DNA damage in micronuclei. Nature 522, 179–184, https://doi.org/10.1038/nature14493, 2015.

Zhang H, Zheng W, Hua L, Wang Y, Li J, Bai H, Wang S, Du M, Ma X, Xu C, Li X, Gong B, Wang Y. Interaction between PPAR γ and SORL1 gene with Late-Onset Alzheimer's disease in Chinese Han Population. Oncotarget. doi: 10.18632/oncotarget.15691, 2017.

Zang XL, Han WQ, Yang FP, Ji KD, Wang JG, Gao PJ, He G and Wu SN. Association of a SNP in SLC35F3 Gene with the Risk of Hypertension in a

Chinese Han Population. Front Genet. 2016 Jun 20;7:108. doi: 10.3389/fgene.2016.00108. eCollection, 2016.

Zerrad L, Merli A, SchroÈder GF, Varga A, GraÂczer EÂ , Pernot P, et al. A spring-loaded release mechanism regulates domain movement and catalysis in phosphoglycerate kinase. J Biol Chem.; 286:14040±14048. https://doi.org/10.1074/jbc.M110.206813 PMID: 21349853, 2011.

Zhang Z, Norris J, Schwartz C, Alexov E. In silico and in vitro investigations of the mutability of disease-causing missense mutation sites in spermine synthase. PLoS One, 6:e20373, 2011.

Zhang Z, Teng S, Wang L, Schwartz CE and Alexov E. "Computational analysis of missense mutations causing Snyder-Robinson syndrome" Human Mutation, vol. 31, no. 9, pp. 1043–1049, 2010.

Zhang GY, Ahmed N, Riley C, Oliva K, Barker G, Quinn MA, Rice GE.bEnhanced expression of peroxisome proliferator-activated receptor gamma in epithelial ovarian carcinoma. Br. J. Cancer 92, 113–119, 2005.

Zhou T, Zhang Y, Macchiarulo A, Yang, Cellanetti M, Coto E, Xu P, Pellicciari R and Wang L. Novel Polymorphisms of Nuclear Receptor SHP Associated with Functional and Structural Changes THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 285, NO. 32, pp. 24871–24881, 2010.

Zimmermann MT, Urrutia R, Oliver GR, Blackburn PR, Cousin MA, Bozeck NJ, Klee EW. Molecular modeling and molecular dynamic simulation of the effects of variants in the TGFBR2 kinase domain as a paradigm for interpretation of variants obtained by next generation sequencing. PLoS ONE 12(2): e0170822. doi:10.1371/journal.pone.0170822, 2017.

Zippo A, De Robertis A, Serafini R, Oliviero S. PIM1-dependent phosphorylation of histone H3 at serine 10 is required for MYC-dependent transcriptional activation and oncogenic transformation. Nat Cell Biol 9: 932–944, 2007.