SAPIENZA
UNIVERSITÀ DI ROMA

DOTTORATO DI RICERCA IN BIOCHIMICA

XXX CICLO

Single nucleotide variants in pathological tissues: effect on protein structure and stability

Dottoranda

Laura Lori

Docente guida
Prof. Roberta Chiaraluce

Coordinatore
Prof. Francesco Malatesta

Dicembre 2017

Index

Chapter 1

Single nucleotide polymorphisms and protein variants

Polymorphisms are generally defined as variations in DNA sequence that occur in at least 1% of the population. The vast majority of variants are single nucleotide polymorphisms, or SNPs. Because the human genome contains 3 billion nucleotides, and variations between individuals occur approximately once in each 300 base pair, approximately 10 million SNPs are expected to exist between any 2 genomes. Because only a fraction of these SNPs is likely to prove relevant to drug response, the ultimate goal will be to identify all functionally important variants.

Major pharmaceutical companies have responded to the growing emphasis on individualized therapy to improve drug efficacy and safety with large investments in pharmacogenomics research. The availability of SNP maps at high-resolution and DNA microarray, enables the basic research and the studies a clinical trials level to better identify disease-susceptibility genes for prognosis, drug discovery, and selection of therapy. If risk for a given disease is predicted to be high, as judged by the SNP pattern of a patient, preventive therapy and lifestyle adjustments (diet, exercise, etc.) may also be implemented.

A comprehensive SNP map will contain genetic variants relevant to drug transport, metabolism, and receptor interaction and, therefore, it needs to be considered in drug selection. Moreover, a comprehensive SNP map may also serve to alert the therapist when careful drug dosage monitoring is required. Stratifying patient populations using genome-wide SNP maps presents a major challenge to the pharmaceutical industry. The outcome from applying such an approach cannot be accurately gauged at present. Genetic heterogeneity appears to be a significant source of variability observed in the response to drugs. This variability means that information pertaining to interethnic and interindividual genetic differences can facilitate rational drug discovery in order to avoid or minimize the incidence of adverse events in clinical trials.

The success of this approach depends in large part on assembling an extensive, high-quality database of informative SNPs. Ultimately the vision of pharmacogenomics encompasses a genetic profile for each individual, containing sufficient information to select which drugs are most likely to be safe and effective in that person. The same insight will be of great importance in preventive health care, arguably the most desirable goal.

The dynamic complexity of the human genome, multigenic disease origins, and involvement of numerous genes in drug response impede the effective application of genome-wide SNP scanning in the clinic. Drug responses will most likely be associated with patterns of multiple polymorphically expressed traits, rather than single causative polymorphisms. Such patterns of genetic variants differ among distinct ethnic groups. This factor could obscure prediction of disease susceptibility and drug response across patient populations, and it points out to the need to

genetically stratify patients for clinical pharmacogenomic studies (Mancinelli et al., 2000).

SNPs may fall within coding as much as non-coding, or in the intergenic regions. SNPs within a coding sequence do not necessarily change the amino acid sequence of the protein that is translated from it, due to degeneracy of the genetic code. SNPs in the coding region are of two types, synonymous and nonsynonymous SNPs. Synonymous SNPs do not affect the protein sequence while nonsynonymous SNPs change the amino acid sequence of protein. The nonsynonymous SNPs are of two types: missense and nonsense. A missense mutation is a point mutation in which a single nucleotide change the resulting amino acid, while on the other hand the nonsense mutation corresponds to a premature stop codon leading to truncation of the resulting protein (Esch e al., 2015).

SNPs that are not in protein-coding regions may still affect gene splicing, transcription factor binding, messenger RNA degradation, or the sequence of non-coding RNA. SNPs located within the coding or regulatory regions of genes can cause qualitative and quantitative changes in gene expression, RNA splicing, protein translation, or gene function. As reported by The International Human Genome Sequencing Consortium 2001, only about 5% of the human genome codes for the production of proteins and most SNPs are found outside coding sequences. These variations may have effects in gene expression and regulation, by interrupting regulatory regions and affecting transcription factor binding. Millions of human SNPs have been identified so far, and these variants could be strongly correlated with phenotypic variations of traits/diseases (Cao et al, 2017). Genetic variation

can cause changes in phenotype if expression levels are altered or pre-mRNA splicing is affected.

Coding SNPs, especially nonsynonymous coding SNPs (nsSNPs, also referred to as missense mutations) are of particular interest, because the single amino acid substitution may affect the structure and/or function of the protein itself (Zhou et al, 2010), as reports in many studies (Chasman and Adams, 2001; Ng and Henikoff, 2002; Ramensky et al, 2002).

Structural studies have been implemented during the last decades since amino acid residue change has been found important for the protein 3-D structure (Altshuler et al, 2008; Botstein et al, 2003). Sequence changes at the amino acid level influence the shape, function, or binding properties of a given protein (Bhattacharya et al, 2017).

Non-synonymous SNPs (nsSNPs) cause changes in the amino acid residues and are important factors contributing to the functional diversity of the encoded proteins (Yates et al, 2003; Wang and Moult, 2001; Wang and Moult, 2003). Non-synonymous SNPs may also affect gene regulation for example by altering DNA and transcriptional binding factors and maintaining the structural integrity of cells and tissues. In addition, nsSNPs affect the functional roles of proteins involved in signal transduction of visual, hormonal, and other stimuli (Rajasekaran et al, 2007; Gfeller et al, 2014).

Point mutation of a particular residue can have different effect on protein structure and function (from no effect to complete disruption of its function). Amino acids with similar size and physico-chemical properties may have mild effect (e.g. substitution from leucine to valine). While changes in crucial area of protein can change the overall arrangement of

protein contacts. Important for example are the secondary structure elements destruction usually driven by substitution of proline residues. Many algorithms have been developed to predict the effect of amino acid changes in a protein structure. The consequences of missense mutations may be difficult to predict because of their diverseness and because a single amino acid change may lead to multiple effects.

Human diseases can be classified as monogenic or multifactorial (Janssens et al 2008). Monogenic diseases are caused by alterations in a single gene, and they segregate in families according to the traditional Mendelian principles of inheritance (Martorana et al, 2017). But, the vast majority of human diseases can be classified as multifactorial (Manolio et al, 2009). These conditions are also referred to as complex diseases, and they are responsible for most of the burden on our health care system (McCarthy, M. I., & MacArthur, D. G.,2017). Examples of these conditions include cardiovascular disease, cancer, type-2 diabetes, and a number of birth defects and psychiatric disorders (Avior et al, 2016). Notably about 40 SNPs have been shown to link with type-2 diabetes (McCarthy, M. I. 2010). By definition, complex diseases are caused by variation in many genes, and they may or may not be influenced by environment. The prediction of the pathogenicity of a nsSNP is based on the degree to which the function of the protein is impaired by the amino acid substitution; but it is further complicated by factors influencing the severity of the phenotype, such as the genetic background and the effect of environment (Stone and Sidow, 2005).

The study of the impact of mutations on protein structure, stability and function may provide a better understanding of the mechanisms of the diseases (Gao et al, 2015). Mutations in different types of functional sites will

typically have different consequences. About one quarter of the known missense SNPs in the human population are significantly deleterious to protein function in vivo (Yue and Moult, 2006). There is considerable evidence that the most common cause of protein loss-of-function relies on changing of the protein's thermodynamic stability. Most changes lead to protein folding equilibrium shifted towards the nonfunctional unfolded state, possibly coupled to irreversible aggregation and/or degradation by cell degradation's pathways (Yue et al, 2005; Casadio et al, 2011; Shi and Moult, 2011; Stefl et al, 2013; Petukh et al, 2015).

The most common cause of monogenic disease is a single base DNA variant resulting in an amino acid substitution and a high fraction of these substitutions appear to result in reduction of stability of the corresponding protein structure (Yue et al, 2005). Frequently, missense mutations are found to destabilize the corresponding protein (Stefl et al, 2013). Destabilizing mutations in the core of the protein lead to inactivation of many tumor suppressors in cancer. Protein stability is a key characteristic of a functional protein (Ramensky et al, 2002 Capriotti et al, 2005; Karchin et al, 2005; Ye et al, 2006; Zhang et al, 2011). SNPs can affect the stability of the protein by making it susceptible to proteolysis, by changing the thermal inactivation temperature or increases the susceptibility of the protein to degradation by lysosomal cysteine proteinase, leading to severe reduction in half-life of its (Bhattacharya et al, 2017). A change in the amino acid sequence can considerably affect protein stability (de Cristofaro et al, 2006; Koukouritaki et al, 2007; Ode et al, 2007) through perturbing conformational constraints and physicochemical parameters. Mutations involving Cys, Trp, or Pro residues are more likely to be disease associated. Cys mutations often involve breaking disulfide bridges or forming unwanted disulfide bonds; Trp

mutations usually significantly destabilize structures and mutations to Pro tend to disrupt helical structures (Gao et al, 2015). The degree of destabilization was found to be high for mutations that introduce drastic changes such as charged to neutral, relatively rigid to relatively flexible, or aromatic to aliphatic mutation types. For most of the cases, the destabilization was also accompanied by structural changes (Pasquo et al, 2012; Khan et al, 2013). This observation highlights the difficulties to predict the physiological relevance of mutations based on the effect delivered by a single methodology.

A single residue mutation can have effects on protein activity (interdicts access to the active site), aggregation (give rise to protein aggregation, which is a hallmark of some neurodegenerative diseases, e.g., Alzheimer's disease, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis and prion diseases), binding (can change the affinity of binding to partners, such as activators, repressors, or substrates) and assembly (change the quaternary structure) (Bhattacharya et al, 2017).

One of the fundamental problems in protein research is the understanding of protein stabilization by complex physical interactions. This problem has attracted the attention also in molecular medicine since it has been noticed that some diseases such as amyloidosis are likely to be related to changes in protein stability due to mutations occurring in the protein sequence. In the biomedical context SNPs are very important, as they are the most common source of variation in the human genome. Due to the redundant nature of the RNA triplet code that encodes proteins, many of these SNPs may not cause an amino acid change (synonymous mutations). However, when a SNP causes an amino acid change (a non-synonymous

mutation) there may be an effect on the structure and function of the encoded protein. The protein stability mechanism is studied by using site-directed mutagenesis followed by thermodynamic measurements and structure determination. Loss of function may lead to disease. It would therefore be extremely useful to be able to predict which mutations are likely to cause disease. Identifying those SNPs that promote susceptibility or protection to complex diseases will aid early diagnosis, prevention and help in finding treatments (Worth et al, 2011; Pires et al, 2013).

The analysis of large datasets of genetic variation can be performed with bioinformatics tools to filter the data (Alipoor et al, 2016; Ramensky et al, 2002, Shreya et al, 2015, Kalia et al, 2016). Computational methods such as SIFT (Ng and Henikoff, 2001; Ng and Henikoff, 2003) Polyphen-2 (Adzhubei et al, 2013), StructMAn (Gress et al, 2016) or MAPP (Stone and Sidow, 2005) classify SNPs according to negative, neutral, or positive effects on the structure or function of the protein. Knowledge of the 3D structure of a gene product is beneficial in predicting and understanding both function and role in disease. However, 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 SNPs on 3D protein structure going beyond the realm of prediction (Arodź and Płonka, 2012).

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). At the protein level, numerous tools have been developed to predict the impact of nonsynonymous SNPs on

protein function and stability (Yadav et al, 2014; Huang et al, 2015; Casadio et al, 2016). These tools usually fall into two categories. The first makes predictions based solely on the sequence of a protein, while the second takes into account structural information when making predictions (Brown, D. K., & Bishop, Ö. T. 2017). Determining which SNPs affect the phenotype would help the identification of the molecular mechanisms of disease and phenotypic variation (Hepp et al, 2015). There are many other tools to differentiate the deleterious or disease-associated SNPs occurring in a gene from the neutral or tolerated alterations, and these tools use approaches based on different features (Vendruscolo et al, 2003; Hepp et al, 2015). Methods based on the structural, physical and chemical properties of the wild type and mutant proteins are also available, and allow the identification of the SNPs 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). Studies using combination of different prediction tools have identified deleterious mutations in genes involved in different biological processes, including, for example, cancer (Hepp et al, 2015).

Polymorphisms may influence gene transcription, mRNA stability, and protein activity. Structural analysis of SNPs 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). SNPs may be also critical for personalized medicine, that is a medical model that proposes the

customization of healthcare with medical decisions, practices, and/or products being tailored to the individual patient. The use of genetic information has played a major role in some aspects of personalized medicine.

Chapter 2

Personalized medicine

SNPs are thought to be critical enablers in realizing the concept of personalized medicine and can be highly useful in diagnostics and therapeutics development (Markward, 2007). Personalized medicine (PM) is defined by the U.S. National Cancer Institute as “a form of medicine that uses information about a person's genes, proteins, and environment to prevent, diagnose, and treat disease”. The clinical applications of PM in cancer is broad, encompassing screening, diagnosis, prognosis, prediction of treatment efficacy, patient follow-up after surgery for early detection of recurrence, and the stratification of patients into cancer subgroup categories, allowing for individualized therapy. PM aims to eliminate the “one size fits all” model of medicine, which has centered on reaction to disease based on average responses to care (Diamandis et al., 2010).

Scientific advances since the completion of the Human Genome Project have confirmed that the genetic composition of individual humans has a significant role to play in predisposition to common diseases and therapeutic interventions. The translation of genetic and genomic data into the knowledge of patient care for prevention, diagnosis, prognosis and treatment has introduced a new paradigm for healthcare: PM. The traditional medicine model has relied on best practices emerging from large population studies and dictates a *one-size-fits-all* approach (Bennet et al., 2012). Although

synthesized evidence is essential to demonstrate the overall safety and efficacy of medical approaches, it falls short in explaining the individual variations that exist among patients. Recent advances in genome-wide association studies have revolutionized the practice of medicine, causing a shift to a *patient-centered* model (Chouchane et al., 2011) and offering tailored diagnostic and therapeutic strategies.

The field of medicine has seen many new advances in the last several decades. An enormous amount of new information has been gained about the human genome and the genetic variations between individuals (Diamandis et al., 2010). Added to this is the emergence of new technologies for global genomic analysis, including high-throughput sequencing, single nucleotide polymorphism (SNP) genotyping, and transcript profiling. The construction of haplotype maps of the genome (Frazer et al., 2007) is now allowing us to view DNA in a much bigger picture than ever before. This, coupled with enormous advances in computer systems and bioinformatics, has resulted in a revolutionary shift in medical care to the era of PM. The concept of PM is to use clinical, genomic, transcriptomic, proteomic, and other information sources to plot the optimal course for an individual in terms of disease risk assessment, prevention, treatment, or palliation. Thus, PM aims to eliminate the "one size fits all" model of medicine, which has mainly centered on reaction to a disease (treating the symptoms) based on average responses to care, by shifting the emphasis of patient care to cancer prevention and early intervention for high-risk individuals (Abrahams et al., 2005; Abrahams and Silver, 2009). Moreover, understanding the molecular profiles of individuals and how these can cause variations in disease susceptibility, symptoms, progression, and responses to treatment will lead to tailoring medical care to

fit each individual patient (Ginsburg and McCarthy, 2001; Allison 2008; Ginsburg and Willard, 2009).

Pharmacogenetics and pharmacogenomics deal with the genetic basis underlying variable drug response in individual patients. The traditional pharmacogenetic approach relies on studying sequence variations in candidate genes suspected of affecting drug response. On the other hand, pharmacogenomic studies encompass the sum of all genes, i.e., the genome. Numerous genes may play a role in drug response and toxicity, introducing a daunting level of complexity into the search for candidate genes. The high speed and specificity associated with genomic technologies enable the search for relevant genes and their variants to include the entire genome (Mancinelli et al., 2000).

Pharmacogenomic analysis can identify disease susceptibility genes representing potential new drug targets. All of this will lead to novel approaches in drug discovery, an individualized application of drug therapy, and new insights into disease prevention. pharmacogenomics may help focus effective therapy on smaller patient subpopulations which although demonstrating the same disease phenotype are characterized by distinct genetic profiles. Moreover, worldwide use of these drugs has revealed substantial interindividual differences in therapeutic response. Any given drug can be therapeutic in some individuals but less effective in others, and some individuals experience adverse drug effects whereas others are unaffected. Recognition of interindividual differences in drug response is an essential step towards optimizing therapy. Over the past decades, much evidence has emerged indicating that a substantial portion of variability in drug response is genetically determined, with age, nutrition, health status,

environmental exposure, and concurrent therapy playing important contributory roles.

Biologists have long accepted that the capacity of organisms to respond differently to their environment is genetically determined. Evidence of interindividual variations in the response to suxamethonium (succinylcholine), isoniazid, and debrisoquine was also scrutinized for a genetic connection. Clinical reports first surfaced in the late 1940s of peripheral neuropathy occurring in a substantial number of patients treated with the antituberculosis drug isoniazid (Hughes et al., 1954). These initial clinical observations were followed by the realization that slow metabolizers (acetylators), although frequencies varied, followed defined geographic and ethnic population distributions (Evans et al., 1960). We now know that the "slow acetylator phenotype" represents approximately 40% to 60% of Caucasians and results in slow clearance and the potential for associated toxicity from drugs such as isoniazid, procainamide, and phenelzine.

In cancer chemotherapy of acute lymphocytic leukemia, administration of drugs such as 6-mercaptopurine, 6-thioguanine, and azathioprine can cause severe hematologic toxicity or even death in patients possessing nonfunctional ("null") variants of thiopurine methyltransferase (TPMT). Each drug after it enters the body interacts with numerous proteins, such as carrier proteins, transporters, metabolizing enzymes, and multiple types of receptors (Sadee, 1998; Sadee 1999; Evans and Relling, 1999). These proteins determine drug absorption, distribution, excretion, target site of action, and pharmacological response. Moreover, drugs trigger downstream secondary events that may also vary among patients. As a result, multiple polymorphisms in many genes may affect drug response, requiring a

genome-wide search for the responsible genes. Profiling the expression pattern of genes in a target tissue reveals mechanisms of drug action in a genomic context, and it can serve to clarify interindividual differences in drug response that are downstream of immediate drug effects in the body.

"The Human Genome Project and advanced technology spin-offs emanating from it will have a profound impact on drug discovery, development, and therapy within the pharmaceutical industry" (Schachter, 1998). With the completion of the human genome project at the beginning of the 21st century, the biological sciences entered an unprecedented age of data generation, and made its first steps toward an era of personalized medicine (Brown, D. K., & Bishop, Ö. T. 2017). Since then, the world has embarked on a revolution in science and healthcare that is changing the way we live and carries the promise of individualizing clinical delivery to improve health, and prevent and cure human disease (Thomas et al, 2016). Innovative automated instrumentation, new analytical and informatics approaches, and novel strategies emerging from genome-based research will be essential for exploiting the massive primary sequence data. Genomic techniques are making it possible not only to identify tangible new gene targets for drug discovery efforts, but also to find associations between specific genetic markers and drug response in a patient population. Technical innovations such as DNA microarrays and microfluidic analytical devices are remodelling the biological sciences by enabling economy of scale for high-throughput DNA sequencing and gene mapping required for genomic research.

An evolving key element in genome-wide searches for genes relevant to disease and therapy is a comprehensive map of polymorphisms distributed

over the entire genome (Gonzaga-Jauregui et al, 2012, Tuxen; I. V et al, 2014; Karki et al, 2015).

Chapter 3

Defective protein variants

Any nsSNP can be deleterious either because it leads to disruption of a site that is directly involved in the function of a protein (e.g. a catalytic residue, a residue involved in ligand binding or a residue that forms a critical interaction with another protein), or because it causes destabilization of the protein structure. Both events can promote protein degradation or severe changing in protein function due to loss of the structural framework. The prediction of the consequences of nsSNPs, in order to discriminate neutral variants from those causative of a pathological phenotype, is a major research challenge as the rapid growth of genomic tools has produced a vast amount of information about genetic variation among individuals (Karchin 2009; Mooney 2005; Ng and Henikoff 2006; Cardoso et al, 2015). In order to have the greatest medical impact we must be able to separate genuine disease-causing or disease- associated genetic variants from the broader background of variants present in all human genomes that are rare, potentially functional but not actually pathogenic for the disease (MacArthur et al, 2014).

The importance of studying of protein folding and stability is manifested by the fact that protein misfolding is involved in many diseases (Dobson, 2003; Valastyan and Lindquist, 2014). Proteins consist of an elaborate arrangement of interior folds that collapse into a final thermodynamically stable structure and, for many proteins, only a modest

free- energy gain (Lindquist and Kelly, 2011) is associated with correct folding of a protein compared with its innumerable potential misfolded states.

The problem of how protein misfolding is promoting certain types of insoluble fibrillar aggregates, at the basis of several amyloid disorders, has been documented in vitro and a list of putative residue mutations causing protein aggregates is available (Chiti and Dobson, 2009; Naiki and Nagai, 2009; Uversky, 2008). Other human diseases different from amyloidosis have been proposed to be related to protein misfolding (Thusberg and Vihinen, 2009; Groenendyk et al, 2010).

Research on protein misfolding and aggregation is progressively gaining attention, largely because of its impact on the understanding of the molecular mechanisms underlying widespread pathologies involving amyloid formation such as Alzheimer's (Selkoe 1996) or Parkinson's (Trojanowski and Lee 1998) disease. Aggregation is a common undesirable event for polypeptide chains, involving the irreversible interaction of two or more denatured protein molecules leading to precipitation of protein. Missense mutations that trigger protein aggregation have been shown to be associated with an increasing number of pathologies (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).

Some genes for rare, heritable Mendelian disorders have been identified, in which variation of single gene is both necessary and sufficient to cause disease (Hamosh et al, 2005). On the other hand, complex polygenic diseases are much more complicated to be directly related to gene variation

since they are thought to be due to the combined effect of many different susceptibility such as genetic variants and environmental factors (multifactorial diseases). Small genetic changes may be responsible for medicinal drug delivery and disease susceptibility, which have already been verified in some diseases, such as diabetes mellitus and cancer (Chung and Chanock, 2011; Vieira et al, 2011).

Cancer

Cancer is a heterogeneous group of diseases characterized by uncontrolled growth of the cells. Cancers are generally classified by the type of cells or organ from which they originate. Since malignant growth can occur in virtually all locations of the body, there are over 100 different types of cancers. Cancer is an immensely complex and diverse disease; however, a set of characteristics are shared among almost all malignancies. Those characteristics, named hallmarks of cancer, are a unified set of capabilities that are acquired during tumorigenesis. The hallmarks of cancer are self-sufficiency in growth signals, insensitivity to growth-inhibitory signals, evasion of programmed cell death, limitless replicative potential, sustained angiogenesis, and tissue invasion and metastasis (Hanahan and Weinberg, 2000). The list has been further extended with emerging hallmarks such as deregulating cellular energetics and avoiding immune response. Additionally, enabling characteristics were proposed, which are tumor promoting inflammation, and genome instability and mutation (Hanahan and Weinberg, 2011). Cancer is an important health problem worldwide and it is one of the most frequent causes of death. Cancer is now becoming the first cause of

death in developed countries and the second in developing countries (Jemal et al, 2011).

Several types of cancer can be caused by environmental factors, which can lead to cancer by inducing DNA damage (Levi et al, 1999). Even though the exact onset and development of cancer have not been identified, it is determined that genetic susceptibility plays an important role on disease etiology (Lichtenstein et al, 2000). In cancer research, genetic aspects have become even more important for revealing the molecular basis of the disease that may have a predictive value in cancer development and may contribute to forecast the efficacy of treatment. Personalized therapies are under development, such as anticancer vaccination and viral gene therapy of cancer by adenovirus particles (Hall et al, 2010). Cancer resistance to drugs is of paramount importance when evaluating the response to chemotherapy; therefore, its relevance is unambiguous in clinical practice (Valera et al, 2004).

SNPs are the most common genetic causes of drug resistance (Cascorbi, 2006; Chou Dhuri and Klaassen, 2006). SNPs may lead to different network of protein interactions in signal transduction pathways (Effert and Volm, 2005). Resistance can also be caused by SNPs of membrane transporters (Cascorbi, 2006; Choudhuri and Klaassen, 2006). A detailed study of proteins associated with SNPs and involved in cancer may lead to a better understanding of the biological mechanism for cancer to improve prevention, early detection and treatment.

Chapter 4

Aim of the study

Mutations can frequently affect several protein biophysical parameters simultaneously and may or may not cause diseases (Gong et Blundell, 2010). There are many prediction studies on the effect of nsSNPs on protein stability. Computational analysis has predicted that around 30% of protein variants resulting from nsSNPs are less stable than the wild type, however experimental studies on the effect of mutations on protein structure, stability and function are still needed. Indeed, experimental studies have been carried out on very few proteins (Gao et al, 2015). These experimental stability studies, that require mutagenesis, protein expression and purification followed by thermal and chemical unfolding, are difficult to perform due to the cost and time needed for the entire process.

This study is focused on the comparison of the structural, biochemical and biophysical parameters between a wild-type protein and the single nucleotide variants found in pathological tissues. We have identified several nsSNPs related to pathological states in humans, such as cancer, and used site-directed mutagenesis to produce recombinant mutant proteins in *E.coli* cells. The variants carrying missense mutations, identified in pathological human tissues have been characterized and compared with the wild-type proteins in order to investigate whether the amino acid substitutions affected protein

conformational stability and interactions with ligand or inhibitors. We analyzed the effects induced by nsSNPs on protein structure and function, by means of structural stability, binding and activity studies. In particular, we selected variants from OMIM (Hamosh et al, 2005) and COSMIC (Forbes et al, 2010) databases for somatic mutations in cancer that store information about genetic variations and aim to collect mutations in all genes. Our study may provide information also on the rational basis for personalized medicine since SNPs may be linked to individual susceptibility either to develop disease or to drug response (Aneesh et al, 2009).

The proteins selected for this study, are:

- **Peroxisome proliferator-activated receptor** (PPARγ) a nuclear receptor in the superfamily of ligand inducible transcription factors. PPARγ plays crucial role within the cell, it is an important regulator of adipocyte differentiation and lipid metabolism. It promotes cell differentiation in hepatocytes, fibroblasts, myocytes, colon epithelium, and in adipocytes (Varga et al, 2011) role in lipid metabolism, PPARγ has been reported to be involved in several processes related to cellular differentiation and development and to carcinogenesis (Wang T et al,2006). Recent studies have indicated PPARγ related also to cardiovascular disease, chronic inflammation, neurodegenerative disorders and cancer (Menendez-Gutierrez MP et al, 2012). The genes activated by PPARγ stimulate lipid uptake in adipocytes and initiate adipogenesis (Anghel SI, et al. 2007).

  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. Our purpose is to study the effects of single amino acid substitution on PPARγ variants: thermodynamic stability, spectral properties, molecular dynamics and transcription activity have been investigated. The molecular dynamics simulated the movements and trajectories of all the atoms in these structures over a limited period of time. It can be used to explore the protein structure upon single amino acid substitution to obtain information about protein flexibility and dynamics (Brown, D. K., & Bishop, Ö. T. 2017).

In this project we also will investigate the effects of PPARγ wild-type and its variants in eukaryotic cell line in order to see if adipogenesis pathway and some cancer related markers are differently expressed. Detailed Western blot analysis will be performed after cell transfection either with PPARγ wild-type and variants in order to see the expression level of the proteins. For this study, we have selected those variants that showed the most significant differences in term of stability and structure in solution and in dynamics. Two interesting variants have been selected for studies in eukaryotic cell lines: E324K and R397C. Both residues, E324 and R397, are involved in one of the two salt bridges that may contribute to PPARγ stabilization.

Our purpose is to combine the structure and function relationship on natural occurring variants studied *in vitro*, with the results obtained in eukaryotic cell lines. The in vitro studies will let us explore the effect of the mutation on the protein structure and stability when compared with the wild type enzyme. On the other hand, the eukaryotic cell line studies will give us the chance to investigate how biochemical pathways are

modified in the presence of the variants and how they affect cell differentiation, cell death and apoptosis.

The analysis of the effect of PPARγ mutations on important pathways for the eukaryotic cells, such as cellular differentiation, apoptosis, propensity to tumorigenicity and adipogenesis (Farmer, S. R. 2005; Wang et al, 2013) may help in understanding the connection between mutation and disease. Indeed, the disruption of cellular pathways may result in the development of several diseases and pathological states (Subbarayan et al, 2004). nsSNPs can also result in altered levels of expression of proteins and mRNAs and may induce functional alteration of the biological properties of the proteins, so biochemical analysis of the cellular properties of polymorphisms are necessary to understand the impact of the mutations on the biological function of the proteins (Guo et al, 2016).

- **Bromodomains** (BRDs) are the only known protein recognition module that selectively targets ε-N-acetylation of lysines. Acetylation of lysine residues (Kac) is one of the most frequently occurring post-translational modifications which controls gene transcription and a vast array of diverse cellular functions. Deregulation of lysine acetylation levels has been associated with the development of many diseases, such as cancer, and enzymes regulating acetylation have emerged as interesting targets for drug discovery. In this study we selected some BRDs variants found in cancer tissues and we analyzed the effects of a single amino acid change on BRDs thermodynamic stability, spectral properties and the interaction with inhibitors JQ1 and PFI-1.

- **Phosphoglycerate kinase** (PGK), 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. The elevated levels of PGK1 protein, detected in the serum of patients affected by pancreatic cancer (Hwang et al., 2006; Sun et al., 2015) and in breast cancer tissues (Sun et al., 2015), suggest a plausible role of PGK1 as a cancer biomarker. In this study we selected the variants R38M, R65W, G166D, M189I, A199V, V216F and F241S associated to human carcinoma. We are generating recombinant protein in order to analyze the consequences of the mutations on PGK1 thermal and thermodynamic stability and the kinetic activity.

  A detailed investigation at the molecular and cellular level of natural protein variants may provide useful information for the development of new therapeutic strategies, particularly in the search of small molecules able to selectively interact with the variants, which is an essential preliminary step to personalized medicine.

Chapter 5

Peroxisome Proliferator Activated Receptor (PPARγ)

Peroxisome proliferator-activated receptors (PPARs) are members of the nuclear hormone receptor superfamily of ligand-activated transcription factors that regulate a large number of gene networks involved in cell metabolism, growth, differentiation, and inflammation (Picard et al, 2002; Lefebvre et al, 2006). There are three PPAR isoforms, α, β, and γ whose expression patterns vary among tissues. PPARs are a family of nuclear transcription factors that interact with PPAR response elements in controlling growth-regulatory gene expression (Vamecq et al, 1999). PPARs belong to the steroid hormone receptor superfamily and are involved in ligand-inducible lipid metabolism (Issemann et al, 1990). Prostanoids and their synthetic analogues and other long-chain fatty acids and their metabolites act as ligands for the three PPARs, α, β, and γ (Issemann et al, 1990; Tontonoz et al, 1994). The PPARγ selective ligand 15-deoxy- D12,14-prostaglandin J2 (15d-PGJ2) is produced in large amounts in the prostate gland via spontaneous conversion from prostaglandin D2 (Tokugawa et al, 1998). Another PPARγ ligand, 15-hydroxyeicosatetraenoic acid (15-HETE), is produced in normal epithelial cells (ECs) by 15-lipoxygenase-2 (15-LOX-2) (Shappell et al, 1999; Tang et

al, 2002). These PPARγ ligands, 15d-PGJ2 and 15-HETE, are likely to be involved in prostate carcinogenesis.

PPARγ is a human transcription factor whose activity is regulated by direct binding of lipid metabolites, and xenobiotics (Chawla et al, 2001). Agonist ligand binding (eicosanoids, prostaglandins and synthetic agonists), by promoting the stabilization of the PPARγ active conformation of the C-terminal activation function-2 helix (H12), triggers the recruitment of co-activator proteins that locally remodel chromatin and activate the cellular transcriptional machinery. The resulting protein complexes activate the transcription of specific target genes thus inducing intracellular signalling cascades that mediate the physiological effects of the ligands. In the absence of ligand, PPARγ has the potential to actively silence genes to which it is bound by recruiting transcriptional corepressor complexes containing nuclear receptor corepressor (N-CoR) or SMRT (silencing mediator of retinoid and thyroid receptors) containing histone deacetylase (HDAC) activity that represses the transcription of the target gene (Guan et al, 2005).

Nuclear receptor ligands bind in a cavity within a moderately conserved ligand binding domain (LBD) toward the C terminus of the nuclear receptor. A linker joins the LBD to the central DNA binding domain of the receptor. The LBDs of receptors have a conserved structure that consists of three layers of α-helices. The central layer is incomplete, leaving a cavity that can accommodate ligand. The size of this cavity varies according to the disposition of the helices in the outer layers. The ligand binding cavity in PPARγ is particularly large. Nuclear receptors activate transcription in response to ligand binding through the displacement of co-repressor proteins and the recruitment of coactivator proteins. The mechanism seems to involve

both a global stabilization of the LBD on ligand binding as well as a specific stabilization of the C-terminal helix 12 (Kallenberger et al, 2003). Notably, various partial agonists were shown to differentially stabilize distinct regions of the LBD (Bruning et al, 2007). The modulation of PPARγ activity may be ascribed to the modification of the conformation of the loop following helix 2' (Waku et al, 2009). Once bound to ligand, PPARγ binds coactivators in a helix 12– dependent fashion, resulting in a transcriptionally active complex. Given the importance of PPARγ as a transcription factor regulating key physiological processes and as a drug target for insulin sensitizers, many structures have been determined in complex with various synthetic agonists and with natural ligands (Cronet et al, 2001; Itoh et al, 2008; Montanari et al, 2008).

Ligand-activated PPARs form heterodimers with the 9-cis retinoic acid receptor (RXR-α) and bind to peroxisome proliferator response elements (PPRE) in the promoter regions of target genes to stimulate their expression (Bardot et al, 1993; Tontonoz et al, 1994). PPARγ activation can inhibit the activity of other pro-inflammatory transcription factors through transrepression mechanisms (Pascual et al, 2005). PPARγ is expressed in numerous cells including pulmonary vascular endothelial and smooth muscle cells where it plays a critical role in normal pulmonary vascular function (Lu et al, 2013; Green et al, 2012; Green et al, 2011).

There are two PPARγ isoforms, PPARγ 1 and PPARγ 2. Some cancers have been shown to express PPARγ (Segawa et al, 2002; Nwankwo et al, 2001¸ Mueller et al, 2000; Kubota et al, 1998) and prostate cancer expresses significantly more PPARγ than normal cells (Segawa et al, 2002; Nwankwo et al, 2001; Kubota et al, 1998). Investigators also have reported

the inhibitory effects of PPARγ agonists on prostate cancer cells (Kubota et al, 1998; Mueller et al, 2012; Butler et al, 2000; Shappell et al, 2001).

PPARs are implicated in tumor progression, cellular differentiation, and apoptosis. Hence, modulation of their function is considered as a potential target for cancer prevention and treatment. Notably, PPAR functions have been related to various pathologies, ranging from metabolic disorders to cardiovascular disease, chronic inflammation, neurodegenerative disorders and cancer (Menendez-Gutierrez et al, 2012; Peters et al, 2012). PPARs ligands and other agents, influencing PPAR signalling pathways, have been shown to display chemopreventive potential by mediating tumor suppressive activities in a variety of human cancers and could represent potential novel targets to inhibit carcinogenesis and prevent tumor progression (Renaud et al, 2000). In addition, PPARγ agonists have been recently reported to lower the incidence of a number of neurological disorders (Chen et al, 2012).

PPARγ is highly expressed in cancer tissues, such as gastric, colorectal, prostate and ovarian cancer (Subbarayan et al, 2004; Segawa et al., 2002; Zhang et al., 2005; Yao et al., 2010; Pang et al.,2015), and plays a crucial role in carcinogenesis (Han and Roman, 2007). In gliomas cells, PPARγ is downregulated, leading to a decrease in insulin sensitivity and an increase in neuroinflammation. Moreover, PPARγ contributes to regulate some key circadian genes. Abnormalities in the regulation of circadian rhythms and dysregulation in circadian clock genes are observed in gliomas. Circadian rhythms are dissipative structures, which play a key role in far-from- equilibrium thermodynamics through their interactions with WNT/beta- catenin pathway and PPARγ. In gliomas, metabolism,

thermodynamics, and circadian rhythms are tightly interrelated (Vallèe et al, 2017).

Some studies suggested that PPARγ had an anti-cancer effect through the inhibition of tumor growth, induction of cell differentiation and apoptosis (Shimada et al., 2002). It also affects tumorigenesis by modulating the expression of its target genes (Leung et al., 2004). However, the underlying mechanism of PPARγ on antitumor is not completely clear until now (Guo et al, 2016). Recently, some studies reported that PPARγ can regulate amyloidogenic pathways (D'Abramo et al, 2005; Sastre et al, 2006), they suggest that PPARγ may be a potential candidate gene for Alzheimer's disease. However, results on association between PPARγ and late-onset Alzheimer's disease were inconsistent yet (Helisalmi et al, 2008; Wu et al, 2015; Zhang et al, 2017). In addition, recent studies have indicated PPARγ agonists have anti- fibrogenic properties in organs besides the gut; however, their effects on human intestinal fibrosis are poorly understood (Koo et al, 2017).

The regulation of glucose metabolism and fatty acid storage is one of the most significant PPARγ activity. PPARγ is the target of many drugs used in the treatment of type 2 diabetes. The anti-diabetic thiazolidinediones (TZD) activate PPARγ and are efficient drugs against type 2 diabetes, despite the occurrence of several adverse effects that confirms the abundant pleiotropic actions of PPARγ (Menendez-Gutierrez et al, 2012). Notably, the marketed TZD class of antidiabetic agents binds to and activates PPARγ. These drugs enhance insulin sensitivity in target tissues and lower glucose and fatty acid levels in type 2 diabetic patients. However, despite their proven benefits, these drugs have been plagued by certain adverse effects: weight

gain, higher rate of bone fractures, fluid accumulation, and pulmonary edema. A detailed structural investigation of the PPARγ interaction with new compounds may lead to the development of new class of drugs targeting lipid and glucose metabolism regulation (Pochetti et al, 2010).

There is evidence that some rare missense mutations in PPARγ have profound phenotypic effects in affected individuals, contributing to the risk of dyslipidaemia and type 2 diabetes (Chan et al., 2013) and colon cancer (Barroso et al., 1999; Savage et al, 2003; Meirhaeghe & Amouyel, 2004; Agostini et al, 2006; Semple et al, 2006; Tan et al, 2008; Jeninga et al, 2009). Some of these mutations in PPARγ have been related to lypodistrophy and insulin resistance (Jeninga et al, 2009; Visser et al, 2011), although the molecular mechanism by which these mutations act is not clear. Many of these variants are nsSNPs, single nucleotide variations occurring in the coding region and leading to a polypeptide sequence with amino acid substitutions (Dixit et al, 2009). Expression and structural characterization of the PPARγ variants associated with energy metabolism disorders, carcinogenesis and tumour progression may increase our knowledge about the molecular interactions in PPARs related disease by studying the effects of these mutations on PPARγ conformational stability and binding interactions. We studied the effect of the human PPARγ F360L mutant associated with familial partial lipodystrophy (Al-Shali et al, 2004) on PPARγ thermal stability, binding interaction and structure. The structure solution of PPARγ F360L revealed the structural reason for its defective behaviour and for the significant decrease of functional activity and binding to synthetic ligands (Lori et al, 2014).

This study is focused on nsSNP variants of the ligand binding domain (LBD) 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, molecular dynamics and transcription activity have been investigated.

Chapter 6

Bromodomains

Bromodomains (BRDs) dysfunction has been linked to the onset and development of several diseases (Müller et al, 2011). BRDs are the only known protein recognition module that selectively targets ε-N-acetylation of lysines. Acetylation levels are reversibly maintained by a group of enzymes, the histone acetyl-transferases (HATs) and histone deacetylases (HDACs) that "write" and "erase" acetylation marks on histones. To date little is known of the "reading" process of Kac that involves BRDs. Kac is recognized by a central hydrophobic cavity and is anchored by a hydrogen bond with an asparagine residue present in most BRDs. However, the substrates (e.g. the acetylated sequences that are specifically recognized) of most BRDs are largely unknown. Proteins that contain BRDs have been implicated in the development of a large variety of diseases (Filippakopoulos and Knapp, 2014).

A comprehensive study of the BRDs variants associated with carcinogenesis and tumour progression may increase our knowledge about the molecular interactions in BRDs related disease by studying the effects of these mutations on BRD conformational stability and binding interactions. The aim of this study is the expression and structural characterization of the BET BRDs wild-type and variants carrying missense mutations and related to

pathological states in humans to investigate whether these mutations affect BRDs conformational stability and interactions with inhibitors.

The human proteome encodes over 40 proteins that contain more than 60 diverse BRDs (Filippakopoulos and Knapp, 2012). Despite their low sequence identity, all BRDs share a conserved fold comprising a left-handed bundle of four alpha helices, linked by diverse loop regions that contribute to substrate specificity. The conserved BRD fold contains a deep, largely hydrophobic acetyl lysine binding site, which represents an attractive pocket for the development of small, pharmaceutically active molecules.

The inhibitors that specifically target the BET proteins selectively interfered with gene expression that mediated cellular growth and evasion of apoptosis in cancer (Dawson et al, 2011; Delmoreet al, 2011; Zuber et al, 2011). BET proteins belong to the subfamily II of BRDs, sharing a common architecture comprising two N-terminal BRDs that exhibit high level of sequence conservation as well as an extra-terminal (ET) domain and a more divergent C-terminal recruitment domain. There are four proteins in this subfamily, BRD2, BRD3 (LeRoy et al, 2008), BRD4 (Ullah et al, 2008) and BRDT (Moriniere et al, 2009).

BRD2 contains two domains (BD1 and BD2), which are considered to be responsible for binding to acetylated chromatin. It is ubiquitously expressed in mammalian tissues.

Importantly, BET-BRDs stay associated with chromatin during mitosis. This unique feature of the BET proteins is utilized by papilloma and Kaposi's sarcoma-associated herpesviruses for tethering their genomes to the mitotic chromosome of their host, resulting in viral propagation during the

cell division. BRD2 has been shown to bind to the transcription factor E2F, but also co-activator TAFs, members of the SWI/SNF complex, HATs and HDACs (histone deacetylases) to regulate the expression of diverse genes (Nakamura et al, 2007). BRD2 can function either as a transcriptional co-activator or co-repressor, depending on the specific promoter and cellular context and has been shown to regulate the expression of cyclin A. Constitutive over-expression of BRD2 in the B-cells of transgenic mice results in the development of B-cell lymphoma and leukaemia. BRD3 contains two bromodomains (BD1 and BD2) and is ubiquitously expressed in human adult and fetal tissues, with highest expression in testis, ovary, placenta, uterus, and brain. BRD3 expression is induced in activated lymphocytes and it is highly expressed in undifferentiated embryonic stem (ES) cells whereas expression levels are reduced upon endothelial differentiation (Ishii et al, 2005). Bromodomains in BRD4 (BD1 and BD2) bind to specific acetylated lysines in histones H3 and H4 and BRD4 has been isolated in complex with the replication factor complex. Mouse embryos nullizygous for BRD4 are non-viable due to an early post-implantation defect, suggesting that BRD4 is required for fundamental cellular processes (Houzelstein et al, 2002).

In our study we selected three of the four BET proteins BRD2, BRD3 and BRD4, that belong to the subfamily II of BRDs. Wild type and variants of BRD2(1), BRD4(1), BRD2(2), BRD3(2) and BRD4(2) have been studied in order to investigate the effect of amino acid substitutions on their structure in solution and on their thermal and thermodynamic stability. The interaction with inhibitors JQ1 and PFI-1 was investigated by isothermal titration calorimetry experiments (ITC).

Chapter 7

Phosphoglycerate kinase

Phosphoglycerate kinase (PGK) (EC 2.7.2.3) is an essential 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. Human PGK can phosphorylate L-nucleoside analogues, which are used in antiviral and anticancer therapies (Krishnan et al, 2003; Gallois-Montbrun et al, 2004; Gondeau et al, 2008). PGK was also shown to participate in the DNA replication and repair in mammal cell nuclei (Jindal et al, 1990). Finally, extracellular PGK has been recently reported to exhibit thiol reductase activity on plasmin, leading to angiostatin formation, which inhibits angiogenesis and tumor growth (Lay et al, 2000; Shichijo et al, 2004). PGK is a typical hinge- bending monomeric enzyme containing 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).

Two human phosphoglycerate kinase isoenzymes, PGK1 and PGK2, have been so far identified, 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. PGK2, also known as

testis form, is unique to meiotic/postmeiotic spermatogenic cells. PGK1, in addition to its metabolic function, may play different roles. This enzyme may be secreted in the extracellular environment by tumor cells and act as a thiol reductase regulating angiogenesis (Wang et al., 2007). In addition, translocation of PGK1 to the nucleus is related to binding to alpha DNA polymerase (Boukouris et al., 2016). 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). Recently, the protein kinase activity of PGK1 has been related to initiation of autophagy (Qian et al., 2017).

Several somatic mutations of PGK1 have been 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). Different mutations of the PGK1 gene have been so far identified and the reasons for the phenotypic variability associated with mutations are still unknown (Chiarelli et al, 2012).

In our study we selected some PGK1 mutations 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. We selected seven PGK1 variants (R38M, R65W, G166D, M189I, A199V, V216F and F241S) associated to human carcinoma. We are generating recombinant protein for each of the identified mutants using site directed mutagenesis and available bacterial expression systems. Introduction of these mutations results in soluble recombinant protein and allows to study

the consequences of the mutations on PGK1 thermal and thermodynamic stability and the kinetic activity.

Chapter 8

Results and discussion

SNPs are the most frequent form of human genetic variation (Dunham et al, 2012) occurring, on average, approximately every 1200 base pairs (Stefl et al, 2013) nsSNPs occur in the DNA coding region and lead to an amino acid change in the protein, missense mutation, that is interesting in medical biology because it may affect protein function and lead to pathogenic phenotypes (Thusberg et al, 2009; Li et al, 2012; Sun et al, 2010). Furthermore, nsSNPs have been shown to be related to drugs sensitivity suggesting new therapeutic personalized strategies.

Several natural variants are associated with the development of diseases and have been catalogued in several databases such as OMIM (Hamosh et al, 2005), the Human Gene Mutation Database (Stenson et al, 2008) and COSMIC (Forbes et al, 2011).

In most of the cases the SNPs stability has been considered to be responsible of the mutations impact on the pathological conditions or on the genetic susceptibility to diseases of the individuals. The most common cause of protein loss of function is the destabilization of its native structure (Pasquo et al, 2012; Lori et al, 2013) and SNPs have been reported to affect protein folding and to produce changes in thermodynamic stability (Adhikari et al, 2015; Kroncke et al, 2015). Experimental studies on thermal and

thermodynamic stability of some natural protein variants expressed in cancer tissues revealed a decrease in thermal stability and an increase of protein flexibility (Pasquo et al, 2012; Lori et al, 2013; Rowling et al, 2010).

Understanding the functional consequences of nonsynonymous changes, and predicting potential causes, involves integration of information from multiple heterogeneous sources, including sequence, structural data and pathway relations between proteins, and the molecular basis of diseases. nsSNPs may influence an individual's susceptibility to disease -or response to drugs- through their impacts on a protein's structure and hence cause functional changes. The future of SNP analysis greatly lies in the development of personalized medicines that can facilitate the treatment of disorders induced by genomic variations. Personalized medicine is a medical model that proposes the customization of healthcare with medical decisions, practices, and/or products being tailored to the individual patient. The knowledge of SNPs will help in understanding pharmacokinetics or pharmacodynamics and how drugs act in individuals with different genetic variants. Diseases with different SNPs may become relevant pharmacogenomic targets for drug therapy.

PPARγ

In this study 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.

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. The selection of the variants was focused on those mutations located in putatively critical positions of the LBD, such as Q286P, R288H, V290M, E324K, E460K, and P467L, in close proximity to the residues involved in ligand binding. 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. 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. 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. Both amino acid substitutions involve a proline residue and, in the case of Q286P, a residue located in the middle of H3. 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 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 (Kallenberger et al, 2003).

The transcription activity of F360L, R357A, P467L, and Q286P PPAR γ variants was evaluated in comparison with wild-type PPAR γ 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 (Pinelli et al, 2005). As previously reported, the efficacy of both ligands remained basically unchanged towards F360L compared to the wild type, while the potency was significantly reduced (Lori et al, 2013). A remarkable lowering in both efficacy and potency was shown for R357A and P467L (Figure 1 and Table 1). Singular behavior has been observed for the mutant Q286P, which was completely inactive and insensitive to both rosiglitazone and LT175 (Figure 2).

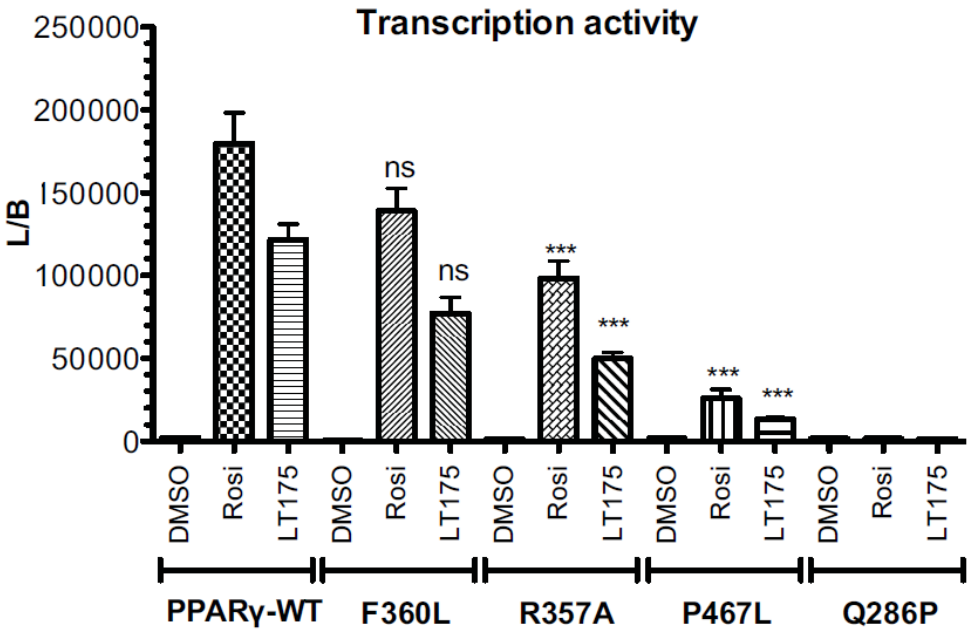


Figure 1 - Transcription activity of wild type PPARγ and mutants in a PPARγ Gal4-based assay.

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	-	-

Table 1 - Transcription activity of PPARγ wild type and mutants in a PPARγ Gal4-based assay

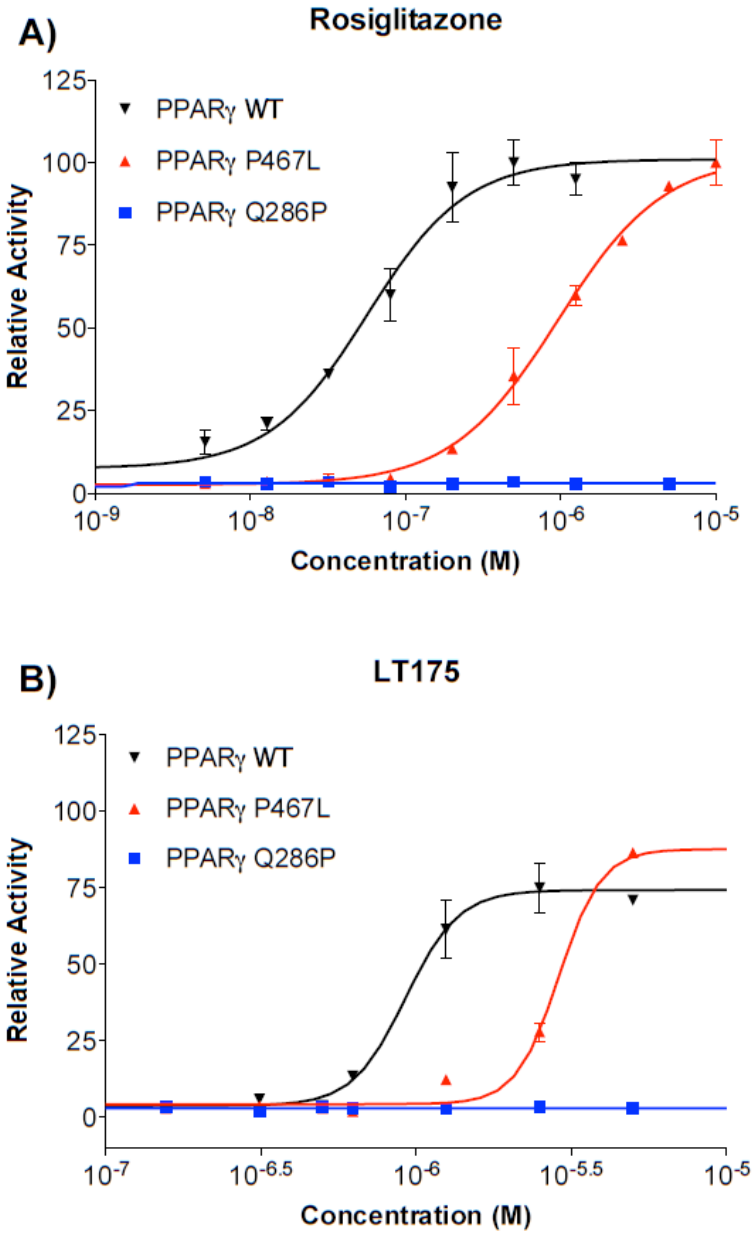


Figure 2 - Transcription activity. Transcription activity of rosiglitazone (A); and LT175 (B) towards wild type PPARγ, PPARγ P467L, and PPARγ Q286P in a PPARγ Gal4-based assay.

In this project we investigated the effects of PPARγ wild-type and the variants E324K and R397C in eukaryotic cell line, in order to see if adipogenesis pathway and some cancer related markers are differently expressed. The variants E324K and R397C are involved in one of the two salt bridges that may contribute to PPARγ stabilization and showed the most significant differences in term of stability and structural differences in solution and in dynamics. Interestingly, these two variants could not be expressed in the soluble fraction in *E. coli* even when different induction conditions were used. The experiments in eukaryotic cell line have been conducted in 3T3-L1 mouse embryonic fibroblast widely used for metabolism and obesity research (Wang et al, 2013). This cell line can differentiate into adipocites1 upon chemical induction. Human PPARγ C-HA tagged natural ORF mammalian expression plasmid and pCMV3 C-HA tagged empty plasmid were used to transfect low-passage 3T3-L1 preadipocytes (kindly provided by Dr. F. Fazi, Sapienza University, Rome). The 3T3-L1 semiconfluent cell cultures have been transfected with the mammalian transient expression vectors harboring wild-type, E324K and R397C PPARγ, respectively, and the pCMV-C-HA empty vector. The morphological differentiation was assessed under light microscopy. The differential expression of adipogenesis markers in transfected cells and relative controls have been detected by western blot analysis.

Bromodomains

In this study we selected three of the four BET proteins BRD2, BRD3 and BRD4, that belong to the subfamily II of BRDs. Wild type and variants of BRD2(1), BRD4(1), BRD2(2), BRD3(2) and BRD4(2) have been studied in order to investigate the effect of amino acid substitutions on their structure in solution and on their thermal and thermodynamic stability. Several mutations in BET (bromo and extra terminal) BRDs (BRD2, BRD3 and BRD4) have been identified in humans and they may play an important role in several diseases. BRDs are the only known protein recognition module that selectively targets ε-N-acetylation of lysines. Many of these BRDs mutations have been related to cancer. The single amino acid substitutions significantly affect the tertiary and secondary structures of most of the BRDs variants studied, as indicated by the differences in their near and far UV CD spectra, when compared to the corresponding wild type proteins. The thermal and thermodynamic stability in particular of the BRD2(1) variants Y153H, E140K, R100L, D160N, D160Y and D161Y are remarkably lower than the corresponding wild type. The mutations induced by nsSNPs and found in cancer may cause alteration of native tertiary contacts and changes in thermal and/or thermodynamic stability. The interaction with inhibitors JQ1 and PFI-1 was investigated by isothermal titration calorimetry experiments (ITC). The fitting of the binding isotherms with a one-site binding model gave the values of the Kd. The BRD2(1) variants, Y153H, D161Y and D161N and BRD3(2) H395R mutant show important differences in the Kd values, compared to wild-type.

Conclusion

In conclusion, the variants of the PPARγ show significant differences in terms of tertiary and secondary structures as suggested by near and far UV CD spectra. The nsSNP PPARγ variants associated to metabolic disorders and cancer show alteration of dynamics that impair the correct reciprocal positioning of structural elements crucially important for PPARγ functioning. The mutations tend to destabilize the PPARγ LBD and increase the flexibility of the variants suggesting a possible involvement in a different network of protein interactions. These alterations may lead to change in the interactions with ligands and influence the multiple biological functions of this nuclear receptor.

The single amino acid substitutions significantly affect the tertiary and secondary structures of most of the BRDs variants studied, as indicated by the differences in their near and far UV CD spectra, when compared to the corresponding wild type proteins. 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. This may be particularly important in BRDs that are physiologically embedded in multidomain proteins and multi-subunit complexes.

All in one, the results reported in this study point to alterations of tertiary contacts and change in protein flexibility and/or in stability for all the PPARγ and BRDS natural variants found in pathological tissues.

A detailed understanding of the changes of the investigated gene products at the molecular level to assess how genetic variations impact the protein folding, structure, function and interactions is required to develop new therapeutic strategies, particularly in the search of small molecules able to selectively interact with the variants, which is an essential preliminary step to personalized medicine, and help to identify new potential therapeutic targets. Precision medicine aims at classifying individuals into subpopulations that differ in their susceptibility to a particular disease. In addition, this approach may help in understanding the biology of the disease, in formulating the right diagnosis and prognosis and, at the last, it may provide information about the response to a specific pharmacological treatment. In conclusion, the structural analysis of protein variants expressed in pathological tissues may help in understanding the disease at the molecular level and, since individuals carrying variants may respond differently to drugs, it may provide information for personalized drugs tailored to the individual variant.

References

A haplotype map of the human genome. International HapMap Consortium. Nature 2005; 437:1299–320.

Abrahams E, Ginsburg GS, Silver M. The Personalized Medicine Coalition: goals and strategies. Am J Pharmacogenomics 2005;5: 345–55.

Abrahams E, Silver M. The case for personalized medicine. J Diabetes Sci Technol 2009; 3:680–4.

Adhikari J, West GM, Fitzgerald MC. Global analysis of protein folding thermodynamics for disease state characterization. J Proteome Res. 2015 May 1;14(5):2287-97. doi: 10.1021/acs.jproteome.5b00057. Epub 2015 Apr 9.

Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Current Protocols in Human Genetics. 2013;(SUPPL.76).

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 PPAR mutations cause lipodystrophic insulin resistance. Cell Metab. 2006, 4, 303–311

Alipoor B, Ghaedi H, Omrani MD, Bastami M, Meshkani R, Golmohammadi T A Bioinformatics Approach to Prioritize Single Nucleotide Polymorphisms in TLRs Signaling Pathway Genes IJMCM, Spring 2016 Vol 5, No 2

Allison M. Is personalized medicine finally arriving? Nat Biotechnol 2008; 26:509–17.

Al-Shali, K., Cao, H., Knoers, N., Hermus, A. R., Tack, C. J. & Hegele, R. A. (2004). A Single-Base Mutation in the Peroxisome Proliferator-Activated Receptor γ 4 Promoter Associated with Altered in Vitro Expression and Partial Lipodystrophy J. Clin. Endocrinol. Metab. 89, 5655–5660.

Altshuler D, Daly MJ, Lander ES. Genetic Mapping in Human Disease. Science. 2008: 881–8.

Aneesh, T. P., Sekhar, S., Jose, A., Chandran, L., & Zachariah, S. M. (2009). Pharmacogenomics: the right drug to the right person. Journal of clinical medicine research, 1(4), 191.

Anghel, S. I., & Wahli, W. (2007). Fat poetry: a kingdom for PPAR [gamma]. Cell research, 17(6), 486.

Arodź T, Płonka PM. Effects of point mutations on protein structure are nonexponentially distributed. Proteins, Structure, Function and Bioinformatics. 2012; 80(7):1780–90

Avior, Y., Sagi, I., & Benvenisty, N. (2016). Pluripotent stem cells in disease modelling and drug discovery. Nature reviews. Molecular cell biology, 17(3), 170.

Bardot O, Aldridge TC, Latruffe N, Green S. PPAR-RXR heterodimer activates a peroxisome proliferator response element upstream of the

bifunctional enzyme gene. Biochem Biophys Res Commun. 1993; 192:37–45. [PubMed: 8386511]

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 PPAR associated with severe insulin resistance, diabetes mellitus and hypertension. Nature 1999, 402, 880–883

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. 2014; 10:e1003440. doi: 10.1371/journal.pcbi.1003440 PMID: 24453961

Bennett C. C., Doub T. W., and Selove R., "EHRs connect research and practice: Where predictive modeling, artificial intelligence, and clinical decision support intersect," Health Policy and Technology, vol. 1, no. 2, pp. 105-114, Jun 2012.

Bhattacharya R, Rose PW, Burley SK, Prlić A (2017) Impact of genetic variation on three dimensional structure and function of proteins. PLoSONE12(3):e0171355

Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nature Genetics. 2003; 33 Suppl:228–37.

Boukouris AE(1), Zervopoulos SD(1), Michelakis ED(2). Trends Biochem Sci. 2016 Aug;41(8):712-30. doi: 10.1016/j.tibs.2016.05.013. Epub 2016 Jun

23.Metabolic Enzymes Moonlighting in the Nucleus: Metabolic Regulation of Gene Transcription.

Brown, D. K., & Bishop, Ö. T. (2017). Role of Structural Bioinformatics in Drug Discovery by Computational SNP Analysis: A Proposed Protocol for Analyzing Variation at the Protein Level. Global Heart.

Bruning JB, Chalmers MJ, Prasad S, Busby SA, Kamenecka TM, He Y, Nettles KW, Griffin PR. Partial agonists activate PPARgamma using a helix 12 independent mechanism. Structure. 2007 Oct;15(10):1258-71.

Bucciantini M, Calloni G, Chiti F, Formigli L, Nosi D, Dobson CM, and Stefani M (2004): Prefibrillar amyloid protein aggregates share common features of cytotoxicity. J Biol Chem 279: 3137482.

Butler R, Mitchell SH, Tindall DJ, Young CYF. Nonapoptotic cell death associated with S-phase arrest of prostate cancer cells via the peroxisome proliferator-activated receptor-g ligand, 15-deoxy-D- 12,14-prostaglandin J2. Cell Growth Differ 2000;11:49 – 61.

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

Capriotti E., Fariselli P., Calabrese R., and Casadio R. (2005), "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. (2005) I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res. 33, W306–W310

Cardoso, J. G., Andersen, M. R., Herrgård, M. J., & Sonnenschein, N. (2015). Analysis of genetic variation and potential applications in genome-scale metabolic modeling. Frontiers in bioengineering and biotechnology, 3.

Casadio R, Vassura M, Tiwari S, Fariselli P, and Martelli PL (2011) Correlating disease-related mutations to their effect on protein stability: A large-scale analysis of the human proteome. Hum. Mutat., 2, 1161–1170.

Cascorbi, I. Role of pharmacogenetics of ATP-binding cassette transporters in the pharmacokinetics of drugs. Pharmacol. Ther., 2006, 112: 457-473.

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- (PPARG) and type 2 diabetes risk amongWomen's Health Initiative postmenopausal women. J. Clin. Endocrinol. Metab. 2013, 98, E600–E604.

Chasman D, Adams RM. (2001) Predicting the functional consequences of non-synonymous single nucleotide polymorphisms: structure-based assessment of amino acid variation. J Mol Biol 307: 683–706, 2001

Chawla A, Repa JJ, Evans RM, Mangelsdorf DJ. Nuclear receptors and lipid physiology: opening the X-files. Science. 2001 Nov 30;294(5548):1866-70.

Chen YC, Wu JS, Tsai HD, Huang CY, Chen JJ, Sun GY, Lin TN. Peroxisome proliferator-activated receptor gamma (PPAR-γ) and neurodegenerative disorders. Mol Neurobiol. 2012 Aug;46(1):114-24

Chiarelli LR, Morera SM, Bianchi P, Fermo E, Zanella A, et al. (2012) Molecular Insights on Pathogenic Effects of Mutations Causing

Phosphoglycerate Kinase Deficiency. PLoS ONE 7(2): e32065. doi:10.1371/journal.pone.0032065

Chiti F, Dobson CM. 2009. Amyloid formation by globular proteins under native conditions. Nat Chem Biol 5:15–22.

Chiti F, Stefani M, Taddei N, Ramponi G, and Dobson CM (2003): Rationalization of the effects of mutations on peptide and protein aggregation rates. Nature 424: 8058.

Chiti F, Webster P, Taddei N, Clark A, Stefani M, Ramponi G, and Dobson CM (1999): Designing conditions for in vitro formation of amyloid protofilaments and fibrils. Proc Natl Acad Sci U S A 96: 35904.

Chouchane L., Mamtani R., Dallol A., and Sheikh J. I., "Personalized medicine: a patient-centered paradigm," J Transl Med, vol. 9, p. 206, 2011

Choudhuri, S., Klaassen, C.D. Structure, function, expression, genomic organization, and single nucleotide polymorphisms of human ABCB1 (MDR1), ABCC1 (MRP), and ABCG2 (BCRP) efflux transporters. Int. J. Toxicol., 2006, 25: 231-259.

Chung, C.C.; Chanock, S.J. Current status of genome-wide association studies in cancer. Hum. Genet., 2011, 130: 59-78.

Cronet P, Petersen JF, Folmer R, Blomberg N, Sjöblom K, Karlsson U, Lindstedt EL, Bamberg K. Structure of the PPARalpha and -gamma ligand binding domain in complex with AZ 242; ligand selectivity and agonist activation in the PPAR family. Structure. 2001 Aug;9(8):699-706.

d'Abramo C, Massone S, Zingg JM, Pizzuti A, Marambaud P, Dalla Piccola B, Azzi A, Marinari UM, Pronzato MA, Ricciarelli R. Role of peroxisome proliferator-activated receptor gamma in amyloid precursor protein processing and amyloid beta-mediated cell death. Biochem J. 2005; 391:693-8.

Dawson, M. A. Prinjha, R.K., Dittmann, A., Giotopoulos, G., Bantscheff, M., Chan, W.I., Robson, S.C., Chung, C.W., Hopf, C., Savitski, M.M., et al. et al. Inhibition of BET recruitment to chromatin as an effective treatment for MLL-fusion leukaemia. Nature 478, 529–533 (2011).

de Cristofaro R., Carotti A., Akhavan S. et al, (2006), "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.

Delmore, J.E., Issa, G.C., Lemieux, M.E., Rahl, P.B., Shi, J., Jacobs, H.M., Kastritis, E., Gilpatrick, T., Paranal, R.M., Qi, J., et al. (2011). BET bromodomain inhibition as a therapeutic strategy to target c-Myc. Cell 146, 904–917.

Diamandis M, White N.M.A. and Yousef G.M. Personalized Medicine: Marking a New Epoch in Cancer Patient Management Mol Cancer Res September 1 2010 (8) (9) 1175-1187; DOI: 10.1158/1541-7786.MCR-10-0264

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, 2015, Vol. 24, No. 24

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. 2009 Oct 16;4(10):e7485.

Dobson C. M. (2003), "Protein folding and misfolding," Nature, vol. 426, no. 6968, pp. 884–890, 2003.

Dunham I. et al. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012 Sep 6;489(7414):57-74. doi: 10.1038/nature11247.

Esch, E. W., Bahinski, A., & Huh, D. (2015). Organs-on-chips at the frontiers of drug discovery. Nature reviews. Drug discovery, 14(4), 248.

Evans DAP, Manley KA, McKusick VA. Genetic control of isoniazid metabolism in man. Br Med J. 1960;2:485-490.

Evans WE, Relling MV. Pharmacogenomics: translating functional genomics into rational therapeutics. Science. 1999;286:487-491.

Fandrich M, Fletcher MA, and Dobson CM (2001): Amyloid fibrils from muscle myoglobin. Nature 410: 1656.

Farmer, S. R. (2005). Regulation of PPAR [gamma] activity during adipogenesis. International journal of obesity, 29(S1), S13

Filippakopoulos, P., & Knapp, S. (2012). The bromodomain interaction module. *FEBS letters*, *586*(17), 2692-2704.

Filippakopoulos P and Knapp S Targeting bromodomains: epigenetic readers of lysine acetylation Nature Reviews Drug Discovery AOP, published online 22 April 2014

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. 2011 Jan;39(Database issue): D945-50. doi: 10.1093/nar/gkq929. Epub 2010 Oct 15

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 (2010): COSMIC (the Catalogue of Somatic Mutations in Cancer): a resource to investigate acquired mutations in human cancer. Nucleic Acids Res 38: D6527

Frazer KA, Ballinger DG, Cox DR, et al. A second generation human haplotype map of over 3.1 million SNPs. Nature 2007; 449:851–61.

Gallois-Montbrun S, Faraj A, Seclaman E, Sommadossi JP, Deville-Bonne D, et al. (2004) Broad specificity of human phosphoglycerate kinase for antiviral nucleoside analogs. Biochem Pharmacol 68: 1749–1756.

Gao M, Zhou H, Skolnick J (2015) Insights into Disease-Associated Mutations in the Human Proteome through Protein Structural Analysis Structure 23, 1362–1369

Gao M, Zhou H, Skolnick J (2015) Insights into Disease-Associated Mutations in the Human Proteome through Protein Structural Analysis Structure 23, 1362–1369

Gfeller D., A. Ernst, N. Jarvik, S.S. Sidhu, G.D. Bader, Prediction and experimental characterization of nsSNPs altering human PDZ-binding motifs. PloS ONE. 9 (2014) e94507.

Ginsburg GS, McCarthy JJ. Personalized medicine: revolutionizing drug discovery and patient care. Trends Biotechnol 2001; 19:491–6.

Ginsburg GS, Willard HF. Genomic and personalized medicine: foundations and applications. Transl Res 2009; 154:277–87.

Gondeau C, Chaloin L, Lallemand P, Roy B, Pe´rigaud C, et al. (2008) Molecular basis for the lack of enantioselectivity of human 3-phosphoglycerate kinase. Nucleic Acids Res 36: 3620–3629

Gong S, Blundell TL. Structural and functional restraints on the occurrence of single amino acid variations in human proteins. PLoS One 2010;5:e9186.

Gonzaga-Jauregui, C., Lupski, J. R., & Gibbs, R. A. (2012). Human genome sequencing in health and disease. Annual review of medicine, 63, 35-6

Green DE, Murphy TC, Kang B-Y, Kleinhenz J, Szyndralewiez C, Page P, Sutliff RL, Hart CM. The NOX4 inhibitor, GKT137831, attenuates hypoxia-induced pulmonary vascular cell proliferation. Am J Respir Cell Mol Biol. 2012; 47:718–726. [PubMed: 22904198]

Green DE, Sutliff RL, Hart CM. Is peroxisome proliferator-activated receptor gamma (PPARγ) a therapeutic target for the treatment of pulmonary hypertension? Pulmonary circulation. 2011; 1:33– 47. [PubMed: 21547012]

Gress A, Ramensky V, Büch J, Keller A and Kalinina OV. StructMAn: annotation of single-nucleotide polymorphisms in the structural context. Nucleic Acids Research, 2016, Vol. 44, Web Server issue

Groenendyk J, Sreenivasaiah PK, Kim do H, Agellon LB, Michalak M. 2010. Biology of endoplasmic reticulum stress in the heart. Circ Res 107:1185–1197.

Guan HP, Ishizuka T, Chui PC, Lehrke M, Lazar MA. Corepressors selectively control the transcriptional activity of PPARgamma in adipocytes. Genes Dev. 2005 Feb 15;19(4):453-61. Epub 2005 Jan 28.

Guijarro JI, Sunde M, Jones JA, Campbell ID, and Dobson CM (1998): Amyloid fibril formation by an SH3 domain. Proc Natl Acad Sci U S A 95: 42248.

Guo, F., Ren, X., Dong, Y., Hu, X., Xu, D., Zhou, H., ... & Zhao, Y. (2016). Constitutive expression of PPARγ inhibits proliferation and migration of gastric cancer cells and down-regulates Wnt/β-Catenin signaling pathway downstream target genes TERT and ENAH. Gene, 584(1), 31-37.

Hall, K.; Blair Zajdel, M.E.; Blair, G.E. Unity and diversity in the human adenoviruses: exploiting alternative entry pathways for gene therapy. Biochem. J., 2010, 431: 321-336.

Hamm JK, Park BH, Farmer SR: A role for C/EBPbeta in regulating peroxisome proliferator-activated receptor gamma activity during adipogenesis in 3T3-L1 preadipocytes. J Biol Chem2001, 276:4-71. 1846

Hamosh A, Scott AF, Amberger JS, Bocchini CA, and McKusick VA (2005): Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Res 33: D5147.

Han, S.W., Roman, J., 2007. Peroxisome proliferator-activated receptor γ: a novel target for cancer therapeutics? Anti-Cancer Drugs 18, 237–244

Hanahan D, Weinberg RA (2000) The hallmarks of cancer. Cell 100: 57–70.

Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. Cell 144: 646–674.

Harris DA and True HL (2006): New insights into prion structure and toxicity. Neuron 50:3537.

Helisalmi S, Tarvainen T, Vepsäläinen S, Koivisto AM, Hiltunen M, Soininen H. Lack of genetic association between PPARG gene polymorphisms and Finnish lateonset Alzheimer's disease. Neurosci Lett. 2008; 441:233-6.

Hepp D, Gonçalves GL, Freitas TROd (2015) Prediction of the Damage-Associated Non-Synonymous Single Nucleotide Polymorphisms in the Human MC1R Gene. PLoS ONE 10(3): e0121812. doi:10.1371/journal.pone.0121812

Houzelstein D, Bullock SL, Lynch DE, Grigorieva EF, Wilson VA, Beddington RS. Growth and early postimplantation defects in mice deficient for the bromodomain-containing protein Brd4. Mol Cell Biol. 2002 Jun;22(11):3794-802.

Huang W, Zhang H, Hao Y, Xu X, Zhai Y, Wang S, et al. (2016) A Non-Synonymous Single Nucleotide Polymorphism in the HJURP Gene Associated with Susceptibility to Hepatocellular Carcinoma among Chinese. PLoS ONE 11(2): e0148618. doi:10.1371/journal.pone.0148618

Hughes HB, Biehl JP, Jones AP, Schmidt, LH. Metabolism of isoniazid in man as related to the occurrence of peripheral neuritis. Am Rev Tuberculosis. 1954; 70:266-273.

Hwang, T.L., Liang, Y., Chien, K.Y. & Yu, J.S. Overexpression and elevated serum levels of phosphoglycerate kinase 1 in pancreatic ductal adenocarcinoma. Proteomics 6(7), 2259–2272 (2006).

Ishii H, Mimori K, Mori M, Vecchione A. Differentially expressed genes in endothelial differentiation. DNA Cell Biol. 2005 Jul;24(7):432-7

Issemann I, Green S. Activation of a member of the steroid hormone receptor superfamily by peroxisome proliferators. Nature 1990; 347:645 – 50.

Itoh T, Fairall L, Amin K, Inaba Y, Szanto A, Balint BL, Nagy L, Yamamoto K, Schwabe JW. Structural basis for the activation of PPARgamma by oxidized fatty acids. Nat Struct Mol Biol. 2008 Sep;15(9):924-31

Janssens, A. C. J., & van Duijn, C. M. (2008). Genome-based prediction of common diseases: advances and prospects. Human molecular genetics, 17(R2), R166-R173.

Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin 2011;61(2):69–90

Jeninga, E.H.; Gurnell, M.; Kalkhoven, E. Functional implications of genetic variation in human PPAR. Trends Endocrinol. Metab. 2009, 20, 380–387.

Jindal HK, Vishwanatha JK (1990) Functional identity of a primer recognition protein as phosphoglycerate kinase. J Biol Chem 265: 6540–6543.

Johnson, A. D., Zhang, Y., Papp, A. C., Pinsonneault, J. K., Lim, J. E., Saffen, D., ... & Sadée, W. (2008). 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

Kalia N, Sharma A, Kaur M, Kamboj SS and Singh J A comprehensive in silico analysis of non-synonymous and regulatory SNPs of human MBL2 gene SpringerPlus (2016) 5:811 DOI 10.1186/s40064-016-2543-4

Kallenberger BC, Love JD, Chatterjee VK, Schwabe JW. A dynamic mechanism of nuclear receptor activation and its perturbation in a human disease. Nat Struct Biol. 2003 Feb;10(2):136-40.

Karchin R (2009): Next generation tools for the annotation of human SNPs. Brief Bioinform 10: 3552.

Karchin R., Diekhans M., Kelly L. et al, (2005), “LS-SNP: largescale annotation of coding non-synonymous SNPs based on multiple information sources,” Bioinformatics, vol. 21, no. 12, pp. 2814–2820, 2005

Karki, R., Pandya, D., Elston, R. C., & Ferlini, C. (2015). Defining “mutation” and “polymorphism” in the era of personal genomics. BMC medical genomics, 8(1), 37.

Khan RH, Chaturvedi D, Mahalakshmi R: Methionine mutations of outer membrane protein X influence structural stability and beta-barrel unfolding. PLoS ONE 2013, 8:e79351.

Khemtemourian L, Killian JA, Hoppener JW, and Engel MF (2008): Recent insights in islet amyloid polypeptideinduced membrane disruption and its role in βcell death in type 2 diabetes mellitus. Exp Diabetes Res 2008: 421287.

Koo JB, Nam MO, Jung Y, Yoo J, Kim DH, Kim G, Shin SJ, Lee KM, Hahm KB, Kim JW, Hong SP, Lee KJ, Yoo JH. Anti-fibrogenic effect of PPAR-γ agonists in human intestinal myofibroblasts. BMC Gastroenterol. 2017 Jun 7;17(1):73. doi: 10.1186/s12876-017-0627-4.

Koukouritaki S. B., Poch M. T., Henderson M. C. et al, (2007), “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. (2003) 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.

Kroncke BM, Vanoye CG, Meiler J, George AL Jr, Sanders CR. Personalized biochemistry and biophysics. Biochemistry. 2015 Apr 28;54(16):2551-9. doi: 10.1021/acs.biochem.5b00189. Epub 2015 Apr 15.

Kubota T, Koshizuka K, Williamson E, et al. Ligand for peroxisome proliferator-activated receptor g (troglitazone) has potent antitumor effect against human prostate cancer both in vitro and in vivo. Cancer Res 1998; 58:3344 – 52.

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 (2016) 48:21 DOI 10.1186/s12711-016-0201-5

Lay AJ, Jiang XM, Kisker O, Flynn E, Underwood A, et al. (2000) Phosphoglycerate kinase acts in tumour angiogenesis as a disulphide reductase. Nature 408: 869–873

Lefebvre P, Chinetti G, Fruchart JC, Staels B. Sorting out the roles of PPAR alpha in energy metabolism and vascular homeostasis. J Clin Invest. 2006; 116:571–580. [PubMed: 16511589]

LeRoy, G., Rickards, B. and Flint, S.J. (2008) The double bromodomain proteins Brd2 and Brd3 couple histone acetylation to transcription. Mol. Cell 30, 51–60.

Leung, W.K., Bai, A.H.C., Chan, V.Y.W., Yu, J., Chan, M.W.Y., To, K.F., Wu, J.R., Chan, K.K., Fu, Y.G., Chan, F.K.L., Sung, J.J.Y., 2004. Effect of peroxisome proliferator activated receptor ligands on growth and gene expression profiles of gastric cancer cells. Gut 53, 331–338.

Levi F, Pasche C, La Vecchia C, Lucchini F, Franceschi S. Food groups and colorectal cancer risk. Br J Cancer 1999;79(7–8):1283–7.

Li X, Jiang Y, Meisenhelder J, Yang W, Hawke DH, Zheng Y, Xia Y, Aldape K, He J, Hunter T, Wang L, Lu Z. Mol Cell. 2016 Mar 3;61(5):705-19. Mitochondria-Translocated PGK1 Functions as a Protein Kinase to Coordinate Glycolysis and the TCA Cycle in Tumorigenesis.

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. 2012;7(12): e51640. doi: 10.1371/journal.pone.0051640. Epub 2012 Dec 10.

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 2000;343(2):78–85.

Lindquist, S. L. and Kelly, J. W. (2011). Chemical and biological approaches for adapting proteostasis to ameliorate protein misfolding and aggregation

diseases: progress and prognosis. Cold Spring Harb. Perspect. Biol. 3, a004507.

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. 2013 Jun 5;8(6): e64824. doi: 10.1371/journal.pone.0064824. Print 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 PPAR F360L mutant associated with familial partial lipodystrophy. Acta. Crystallogr. D Biol. Crystallogr. 2014, 70, 1965–1976

Lu X , Bijli KM., Ramirez A, Murphy TC, Kleinhenz J, and Hart CM Hypoxia downregulates PPARγ via an ERK 1/2-NF-κB-Nox4- dependent mechanism in human pulmonary artery smooth muscle cells Free Radic Biol Med. 2013 October; 63: 151–160. doi:10.1016/j.freeradbiomed.2013.05.013.

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. (2014) Guidelines for investigating causality of sequence variants in human disease. Nature. 2014 April 24; 508(7497): 469–476

Mancinelli L, Cronin M and Sadée W. Pharmacogenomics: The Promise of Personalized Medicine. AAPS PharmSci 2000; 2 (1) Article 4

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., ... & Cho, J. H. (2009). Finding the missing heritability of complex diseases. Nature, 461(7265), 747.

Markward, N. J. (2007) AMIA Annual Symposium Proceedings, Chicago, IL, November 10–14, 2007, p. 1041, American Medical Informatics Association

Martorana, D., Bonatti, F., Mozzoni, P., Vaglio, A., & Percesepe, A. (2017). Monogenic Autoinflammatory Diseases with Mendelian Inheritance: Genes, Mutations, and Genotype/Phenotype Correlations. Frontiers in Immunology, 8.

McCarrey JR, Thomas K (1987) Human testis-specific PGK gene lacks introns and possesses characteristics of a processed gene. Nature 326: 501–505

McCarthy, M. I. (2010). Genomics, type 2 diabetes, and obesity. New England Journal of Medicine, 363(24), 2339-2350.

McCarthy, M. I., & MacArthur, D. G. (2017). Human disease genomics: from variants to biology. Genome biology, 18(1), 20.

Meirhaeghe, A.; Amouyel, P. Impact of genetic variation of PPAR in humans. Mol. Genet. Metab. 2004, 83, 93–102.

Menendez-Gutierrez MP, Roszer T, Ricote M. Biology and therapeutic applications of peroxisome proliferator- activated receptors. Curr Top Med Chem. 2012;12(6):548-84

Montanari R, Saccoccia F, Scotti E, Crestani M, Godio C, Gilardi F, Loiodice F, Fracchiolla G, Laghezza A, Tortorella P, Lavecchia A, Novellino E, Mazza F, Aschi M, Pochetti G. Crystal structure of the peroxisome proliferator- activated receptor gamma (PPARgamma) ligand binding domain complexed with a novel partial agonist: a new region of the hydrophobic pocket could be exploited for drug design. J Med Chem. 2008 Dec 25;51(24):7768-76.

Mooney S (2005): Bioinformatics approaches and resources for single nucleotide polymorphism functional analysis. Brief Bioinform 6: 4456.

Moriniere, J. et al. (2009) Cooperative binding of two acetylation marks on a histone tail by a single bromodomain. Nature 461, 664–668

Mueller E, Drori S, Aiyer A, et al. Genetic analysis of adipogenesis through peroxisome proliferator-activated receptor g isoforms. J Biol Chem 2002; 77:41925 – 30

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.

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

Müller S, Filippakopoulos P, Knapp S (2011). Bromodomains as therapeutic targets. Expert Rev Mol Med. 13, e29.

Naiki H, Nagai Y (2009). Molecular pathogenesis of proteinmisfolding diseases: pathological molecular environments versus quality control systems against misfolded proteins. J Biochem 146:751–756.

Nakamura Y, Umehara T, Nakano K, Jang MK, Shirouzu M, Morita S, Uda-Tochio H, Hamana H, Terada T, Adachi N, Matsumoto T, Tanaka A, Horikoshi M, Ozato K, Padmanabhan B, Yokoyama S. Crystal structure of the human BRD2 bromodomain: insights 131 into dimerization and recognition of acetylated histone H4. J Biol Chem. 2007 Feb 9;282(6):4193-201

Ng PC, Henikoff S. (2002) Accounting for human polymorphisms predicted to affect protein function. Genome Res 12: 436–446, 2002.

Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. Genome Research. 2001; 11 863–74. https://doi.org/10.1101/gr.176601 PMID: 11337480

Ng PC, Henikoff S. Predicting the Effects of Amino Acid Substitutions on Protein Function. Annu Rev Genomics Hum Genet. 2006; 7: 61–80. PMID: 16824020

Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acid Research. 2003; 31(13):3812–4. PMID: 12824425

Ng, P. C. and Henikoff, S. (2006) Predicting the effects of amino acid substitutions on protein function. Annu. Rev. Genomics Hum. Genet. 7, 61–80

Nwankwo JO, Robbins ME. Peroxisome proliferator-activated receptor-g expression in human malignant and normal brain, breast and prostate-derived cells. Prostaglandins Leukot Essent Fatty Acids 2001; 64:241 – 5.

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.

Palmai Z, Chaloin L, Lionne C, Fidy J, Perahia D, et al. (2009) Substrate binding modifies the hinge bending characteristics of human 3-phosphoglycerate kinase: a molecular dynamics study. Proteins 77: 319–329

Pang, T., Kaufman, A., Choi, J., Gill, A., Drummond, M., Hugh, T., Samra, J., 2015. Peroxisome proliferator-activated receptor-alpha staining is associated with worse outcome in colorectal liver metastases. Mol. Clin. Oncol. 3, 308– 316.

Pascual G, Fong AL, Ogawa S, Gamliel A, Li AC, Perissi V, Rose DW, Willson TM, Rosenfeld MG, Glass CK. A SUMOylation-dependent pathway mediates transrepression of inflammatory response genes by PPAR-gamma. Nature. 2005; 437:759–763. [PubMed: 16127449]

Pasquo A, Consalvi V, Knapp S, Alfano I, Ardini M, et al. (2012) 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

Peters JM, Shah YM, Gonzalez FJ. The role of peroxisome proliferator-activated receptors in carcinogenesis and chemoprevention. Nat Rev Cancer. 2012 Feb 9;12(3):181-95

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.

Picard F, Auwerx J. PPAR(gamma) and glucose homeostasis. Annu Rev Nutr. 2002; 22:167–197. [PubMed: 12055342]

Pinelli, A., Godio, C., Laghezza, A., Mitro, N., Fracchiolla, G., Tortorella, V., ... & Crestani, M. (2005). Synthesis, biological evaluation, and molecular modeling investigation of new chiral fibrates with PPARα and PPARγ agonist activity. Journal of medicinal chemistry, 48(17), 5509-5519.

Pires, D. E., Ascher, D. B., & Blundell, T. L. (2013). mCSM: predicting the effects of mutations in proteins using graph-based signatures. Bioinformatics, 30(3), 335-342.

Pochetti G, Mitro N, Lavecchia A, Gilardi F, Besker N, Scotti E, Aschi M, Re N, Fracchiolla G, Laghezza A, Tortorella P, Montanari R, Novellino E, Mazza F, Crestani M, Loiodice F. Structural insight into peroxisome proliferator- activated receptor gamma binding of two ureidofibrate-like enantiomers by molecular dynamics, cofactor interaction analysis, and site-directed mutagenesis. J Med Chem. 2010 Jun 10;53(11):4354-66

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

Rajasekaran R., C. Sudandiradoss, C.G. Doss, R. Sethumadhavan, Identification and in silico analysis of functional SNPs of the BRCA1 gene. Genomics 90 (2007) 447–452.

Ramensky V, Bork P, Sunyaev S. Human non-synonymousSNPs: server and survey. Nucleic Acids Research. 2002; 30(17):3894–900. PMID: 12202775

Ramensky V., Bork P., and Sunyaev S., "Human nonsynonymous SNPs: server and survey," Nucleic Acids Research, vol. 30, no. 17, pp. 3894–3900, 2002.

Renaud JP, Moras D. Structural studies on nuclear receptors. Cell Mol Life Sci. 2000 Nov;57(12):1748-69

Robinson PA (2008): Protein stability and aggregation in Parkinson's disease. Biochem J 413: 113.

Rowling PJ, Cook R, Itzhaki LS. Toward classification of BRCA1 missense variants using a biophysical approach. J Biol Chem. 2010 Jun 25;285(26):20080-7. doi: 10.1074/jbc.M109.088922. Epub 2010 Apr 8.

Sadee W. Finding the right drug for the right patient. Drug Metab Drug Interact. 1997; 14:55-82. Pharm Res. 1998; 15:959-963. Sadee W. Pharmacogenomics. BMJ. 1999; 319:1286.

Sastre M, Dewachter I, Rossner S, Bogdanovic N, Rosen E, Borghgraef P, Evert BO, Dumitrescu-Ozimek L, Thal DR, Landreth G, Walter J, Klockgether T, van Leuven F, et al. Nonsteroidal anti-inflammatory drugs repress beta- secretase gene promoter activity by the activation of PPAR gamma. Proc Natl Acad Sci U S A. 2006; 103:443-8.

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-. Diabetes 2003, 52, 910–917

Schachter Schachter B. Pharming the Genome. Biomednet [serial online]. October 30, 1998;41.

Segawa Y, Yoshimura R, Hase T, et al. Expression of peroxisome proliferator- activated receptor (PPAR) in human prostate cancer. Prostate 2002; 51:108 –16.

Segawa, Y., Yoshimura, R., Hase, T., Nakatani, T., Wada, S., Kawahito, Y., Kishimoto, T., Sano, H., 2002. Expression of peroxisome proliferator-activated receptor (PPAR) in human prostate cancer. Prostate 51, 108–116.

Selkoe DJ (1996): Amyloid βprotein and the genetics of Alzheimer's disease. J Biol Chem 271: 182958. Trojanowski JQ and Lee VM (1998): Aggregation

of neurofilament and α-synuclein proteins in Lewy bodies: implications for the pathogenesis of Parkinson disease and Lewy body dementia. Arch Neurol 55: 1512

Semple, R. K., Chatterjee, V. K. & O'Rahilly, S. (2006). PPAR gamma and human metabolic disease. J. Clin. Invest. 116, 581–589

Shappell SB, Boeglin WE, Olson SJ, Kasper S, Brash AR. 15- Lipoxygenase-2 (15-LOX-2) is expressed in benign prostatic epithelium and reduced in prostate adenocarcinoma. Am J Pathol 1999; 155:235 – 45.

Shappell SB, Gupta RA, Manning S, et al. 15S-Hydroxyeicosatetraenoic acid activates peroxisome proliferator-activated receptor g and inhibits proliferation in PC3 prostate carcinoma cells. Cancer Res 2001; 61:497 – 503.

Shi, Z., and Moult, J. (2011) Structural and functional impact of cancer-related missense somatic mutations. J. Mol. Biol. 413, 495−512

Shichijo S, Azuma K, Komatsu N, Ito M, Maeda Y, et al. (2004) 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.

Shimada, T., Kojima, K., Yoshiura, K., Hiraishi, H., Terano, A., 2002. Characteristics of the peroxisome proliferator activated receptor γ ligand induced apoptosis in colon cancer cells. Gut 50, 658–664.

Shreya M. Patel, Prakash G. Koringa, Bhaskar B. Reddy, Neelam M. Nathani, Chaitanya G. Joshi In silico analysis of consequences of non-

synonymous SNPs of Slc11a2 gene in Indian bovines Genomics Data 5 (2015) 72–79

Stefl, S., Nishi, H., Petukh, M., Panchenko, A. R., & Alexov, E. (2013). Molecular mechanisms of disease-causing missense mutations. Journal of molecular biology, 425(21), 3919-3936.

Stefl, S., Nishi, H., Petukh, M., Panchenko, A. R., and Alexov, E. (2013) Molecular mechanisms of disease-causing missense mutations. J. Mol. Biol. 425, 3919−3936

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. 2008 Feb;45(2):124-6. doi: 10.1136/jmg.2007.055210.

Stone EA and Sidow A (2005) Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. Genome Res. 2005 Jul;15(7):978-86. Epub 2005 Jun 17.

Stone EA, Sidow A. Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. Genome Research. 2005; 15(7):978–86. A. Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. Genome Research. 2005; 15(7):978–86.

Subbarayan V, Sabichi AL, Kim J, Llansa N, Logothetis CJ, Lippman SM. and Menter DG. Differential Peroxisome Proliferator-Activated Receptor-γ

Isoform Expression and Agonist Effects in Normal and Malignant Prostate Cells Cancer Epidemiol Biomarkers Prev 2004; 13(11). November 2004

Subbarayan, V., Sabichi, A. L., Kim, J., Llansa, N., Logothetis, C. J., Lippman, S. M., & Menter, D. G. (2004). Differential peroxisome proliferator-activated receptor-γ isoform expression and agonist effects in normal and malignant prostate cells. Cancer Epidemiology and Prevention Biomarkers, 13(11), 1710-1716

Sun S, Liang X, Zhang X, Liu T, Shi Q, Song Y, Jiang Y, Wu H, Jiang Y, Lu X, Pang D. Br J Cancer. 2015 Apr 14;112(8):1332-9. doi: 10.1038/bjc.2015.114. Epub

Sun T, Lee GS, Oh WK, Pomerantz M, Yang M, Xie W, Freedman ML, Kantoff PW. Single-nucleotide polymorphisms in p53 pathway and aggressiveness of prostate cancer in a Caucasian population. Clin Cancer Res. 2010 Nov 1;16(21):5244-51. doi: 10.1158/1078-0432.CCR-10-1261. Epub 2010 Sep 20.

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 PPAR mutations. J. Clin. Endocrinol. Metab. 2008, 93, 4462–4470.

Tang S, Bhatia B, Maldonado CJ, et al. Evidence that arachidonate 15-lipoxygenase 2 is a negative cell cycle regulator in normal prostate epithelial cells. J Biol Chem 2002; 277:16189 – 201.

Thomas, S., Wolstencroft, K., de Bono, B., & Hunter, P. J. (2016). A physiome interoperability roadmap for personalized drug development. Interface focus, 6(2), 20150094.

Thusberg J, Vihinen M. 2009. Pathogenic or not? And if so, then how? Studying the effects of missense mutations using bioinformatics methods. HumMutat 30:703–714.

Tokugawa Y, Kunishige I, Kubota Y, et al. Lipocalin-type prostaglandin D synthase in human male reproductive organs and seminal plasma. Biol Reprod 1998; 58:600 – 7.

Tontonoz P, Hu E, Graves RA, Budavari AI, Spiegelman BM. mPPARg2: tissue-specific regulator of an adipocyte enhancer. Genes Dev 1994; 8:1224 – 34.

Tontonoz P, Hu E, Spiegelman BM. Stimulation of adipogenesis in fibroblasts by PPAR gamma 2, a lipid-activated transcription factor. Cell. 1994; 79:1147– 1156. [PubMed: 8001151]

Tuxen, I. V., Jønson, L., Santoni-Rugiu, E., Hasselby, J. P., Nielsen, F. C., & Lassen, U. (2014). Personalized oncology: genomic screening in phase 1. Apmis, 122(8), 723-733.

Ullah, M. et al. (2008) Molecular architecture of quartet MOZ/MORF histone acetyltransferase complexes. Mol. Cell. Biol. 28, 6828–6843

Uversky VN. 2008. Amyloidogenesis of natively unfolded proteins. Curr Alzheimer Res 5:260–287.

Valastyan Julie S. and Lindquist Susan (2014) Mechanisms of protein folding diseases at a glance Disease Models & Mechanisms (2014) 7, 9-14 doi:10.1242/dmm.013474

Valera, E.T.; Scrideli, C.A.; Queiroz, R.G.; Mori, B.M.; Tone, L.G. Multiple drug resistance protein (MDR-1), multidrug resistance-related protein (MRP) and lung resistance protein (LRP) gene expression in childhood acute lymphoblastic leukemia. Sao Paulo Med. J., 2004, 122: 166-171.

Vallée A, Lecarpentier Y, Guillevin R, Vallée JN.Thermodynamics in Gliomas: Interactions between the Canonical WNT/Beta-Catenin Pathway and PPAR Gamma. Front Physiol. 2017 May 30;8:352. doi: 10.3389/fphys.2017.00352. eCollection 2017.

Varga, T., Czimmerer, Z., & Nagy, L. (2011). PPARs are a unique set of fatty acid regulated transcription factors controlling both lipid metabolism and inflammation. Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease, 1812(8), 1007-1022.

Vas M, Varga A, Graczer E (2010) 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.

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. 2003; 361: 1205–1222.

Vieira, S.M.; Monteiro, M.B.; Marques, T.; Luna, A.M.; Fortes, M.A.; Nery, M.; Queiroz, M.; Dib, S.A.; Vendramini, M.F.; Azevedo, M.J.; Canani, L.H.;

Parisi, M.C.; Pavin, E.J.; Giannella-Neto, D.; Corrêa-Giannella, M.L. Association of genetic variants in the promoter region of genes encoding p22phox (CYBA) and glutamate cysteine ligase catalytic subunit (GCLC) and renal disease in patients with type 1 diabetes mellitus. BMC Med.Genet., 2011, 12: 129.

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 PPAR mutation (Y151C). Diabetologia 2011, 54, 1639–1644.

Waku T, Shiraki T, Oyama T, Fujimoto Y, Maebara K, Kamiya N, Jingami H, Morikawa K. Structural insight into PPARgamma activation through covalent modification with endogenous fatty acids. J Mol Biol. 2009 Jan 9;385(1):188-99.

Wang, T., Xu, J., Yu, X., Yang, R., & Han, Z. C. (2006). Peroxisome proliferator-activated receptor γ in malignant diseases. Critical reviews in oncology/hematology, 58(1), 1-14.

Wang Z. and Moult J. (2003), “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. (2001), “SNPs, protein structure, and disease,” Human Mutation, vol. 17, no. 4, pp. 263–270, 2001

Wang, J., Wang, J., Dai, J., Jung, Y., Wei, C.L., Wang, Y., Havens, A.M., Hogg, P.J., Keller, E.T., Pienta, K.J., Nor, J.E., Wang, C.Y. & Taichman, R.S. A glycolytic mechanism regulating an angiogenic switch in prostate cancer. Cancer Res. 67, 149-159 (2007).

Wang, L., Li, L., Ran, X., Long, M., Zhang, M., Tao, Y., ... & Ren, J. (2013). Lipopolysaccharides reduce adipogenesis in 3T3-L1 adipocytes through activation of NF-κB pathway and downregulation of AMPK expression. Cardiovascular toxicology, 13(4), 338-346.

Wang, X., Sun, Y., Wong, J., & Conklin, D. S. (2013). PPARγ maintains ERBb2-positive breast cancer stem cells. Oncogene, 32(49), 5512.

Willard HF, Goss SJ, Holmes MT, Munroe DL (1985) 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.

Worth, C. L., Preissner, R., & Blundell, T. L. (2011). SDM—a server for predicting effects of mutations on protein stability and malfunction. Nucleic acids research, 39(suppl_2), W215-W222.

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. 2015; 563:120-4.

XL Zang, WQ Han, FP Yang, KD Ji, JG Wang, PJ Gao, G He and SN Wu, Association of a SNP in SLC35F3 Gene with the Risk of Hypertension in a Chinese Han Population

Yadav P., Chatterjee A., Bhattacharjee A. Identification of deleterious nsSNP sinα, μ, π and θ class of GST family and their influence on protein structure Genomics Data 2 (2014) 66–72

Yankner BA and Lu T (2009): Amyloid β-protein toxicity and the pathogenesis of Alzheimer disease. J Biol Chem 284: 47559.

Yao, L., Liu, F., Sun, L., Wu, H., Guo, C., Liang, S., Liu, L., Liu, N., Han, Z., Zhang, H.,Wu, K., Fan, D., 2010. Upregulation of PPARgamma in tissue with gastric carcinoma. Hybridoma (Larchmt) 29, 341–343.

Yates C.M., M.J. Sternberg, The effects of non-synonymous single nucleotide polymorphisms (nsSNPs) on protein-protein interactions. J. Mol. Biol. 425 (2013) 3949–3963.

Ye Y., Li Z. and Godzik A., (2006), "Modeling and analyzing threedimensional structures of human disease proteins," Pacific Symposium on Biocomputing, pp. 439–450, 2006.

Yue P and Moult J Identification and Analysis of Deleterious Human SNPs J. Mol. Biol. (2006)

Yue P and Moult J Identification and Analysis of Deleterious Human SNPs J. Mol. Biol. (2006)

Yue P, Li1 Z and Moult1 J (2005) Loss of Protein Structure Stability as a Major Causative Factor in Monogenic Disease J. Mol. Biol. (2005) 353, 459–473

Yue P, Moult J. Identification and Analysis of Deleterious Human SNPs. J Mol Biol. 2006; 356: 1263–1274. PMID: 16412461

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. 2017 Feb 25. doi: 10.18632/oncotarget.15691.

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 2011, 6:e20373.

Zhang Z, Teng S, Wang L, Schwartz CE and Alexov E (2010), "Computational analysis of missense mutations causing Snyder-Robinson syndrome," Human Mutation, vol. 31, no. 9, pp. 1043–1049, 2010.

Zhang, G.Y., Ahmed, N., Riley, C., Oliva, K., Barker, G., Quinn, M.A., Rice, G.E., 2005. Enhanced expression of peroxisome proliferator-activated receptor gamma in epithelial ovarian carcinoma. Br. J. Cancer 92, 113–119.

Zhou T, Zhang Y, Macchiarulo A, Yang, Cellanetti M, Coto E, Xu P, Pellicciari R and Wang L (2010) Novel Polymorphisms of Nuclear Receptor SHP Associated with Functional and Structural Changes THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 285, NO. 32, pp. 24871–24881

Zhou, T., Zhang, Y., Macchiarulo, A., Yang, Z., Cellanetti, M., Coto, E., ... & Wang, L. (2010). Novel polymorphisms of nuclear receptor SHP associated with functional and structural changes. Journal of Biological Chemistry, 285(32), 24871-24881

Zuber, J. et al. RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. Nature 478, 524–528 (2011).

Appendix

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).

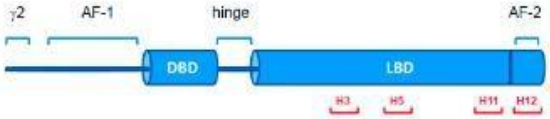


Figure 1. Schematic representation of nuclear receptor PPARγ. The ligand binding domain (LBD) is linked to the DNA binding domain (DBD) by a hinge. The residues involved in ligand binding are located in helix 3 (H3), helix 5 (H5), helix 11 (H11), and helix 12 (H12). Helices are numbered according to 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).

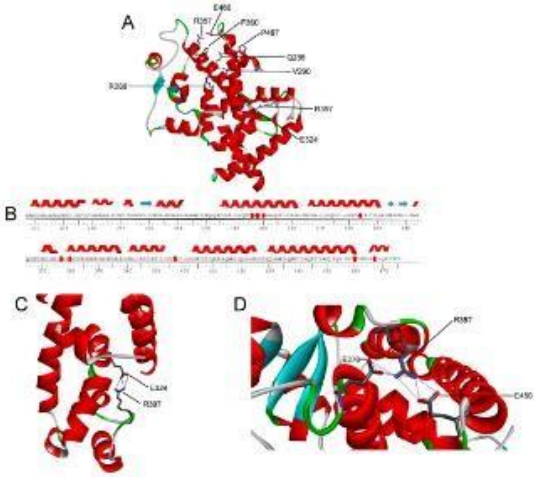


Figure 2. Amino acid sequence and structure of PPARγ LBD. (**A**) Structure of PPARγ LBD (PDB code: 1PRG) shown as a ribbon diagram; (**B**) secondary structural elements are shown at the top of the amino acid sequence. Mutated residues are highlighted in red; (**C**) Local environment of residues R397 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.

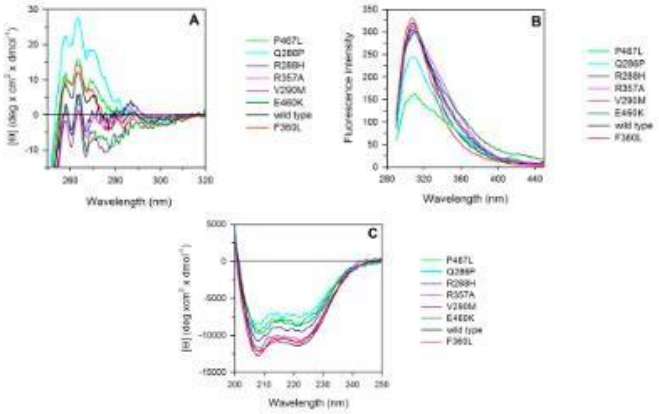


Figure 3. Spectroscopic properties of PPARγ wild type and variants. (**A**) Near-UV CD spectra were recorded in a 1 cm path-length quartz cuvette at 4.60 mg/mL protein concentration in 50 mM Tris–HCl pH 8.0 containing 0.20 M NaCl and 2.0 mM Dithiothreitol (DTT); (**B**) intrinsic fluorescence emission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

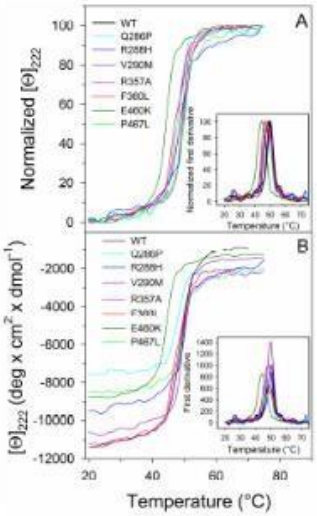


Figure 4. Thermal unfolding transition of PPARγ wild type and variants. Wild type and variants were heated from 20 to 75 °C in a 0.1 cm path-length quartz cuvette at 0.2 mg/mL protein in 20 mM Tris–HCl pH 8.0 containing 0.20 M NaCl and 0.2 mM DTT and the molar ellipticity at 222 nm ($[\Theta_{222}]$) 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 [Θ_{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).

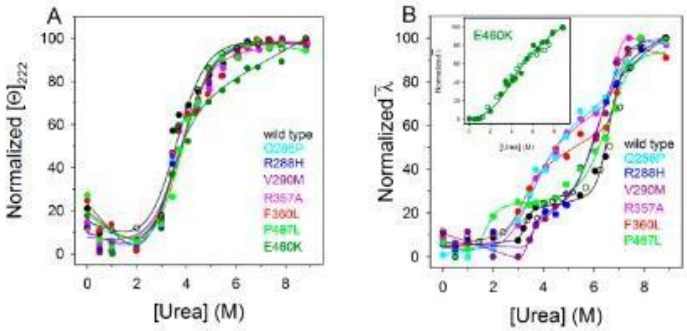


Figure 5. Urea-induced equilibrium unfolding of PPARγ wild type and variants. (**A**) Normalized molar ellipticity at 222 nm ($[\Theta]_{222}$) reported after removal of the high-frequency noise and the low-frequency random error by singular value decomposition algorithm (SVD). The continuous lines represent 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-N} (kcal/mol/M)	$D50_{I-N}$ (M)	$\Delta G^{H_2O}{}_{I-N}$ (kcal/mol)	m_{U-I} (kcal/mol/M)	$D50_{U-I}$ (M)	$\Delta G^{H_2O}{}_{U-I}$ (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 ($\Delta G^{H_2O}{}_{I-N}$) and from the intermediate to the unfolded state ($\Delta G^{H_2O}{}_{U-I}$) were calculated from Equation (5). $D50_{I-N}$ and m_{I-N} which are the midpoint and m value for the transition between native and intermediate state, respectively, and $D50_{U-I}$ and m_{U-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.

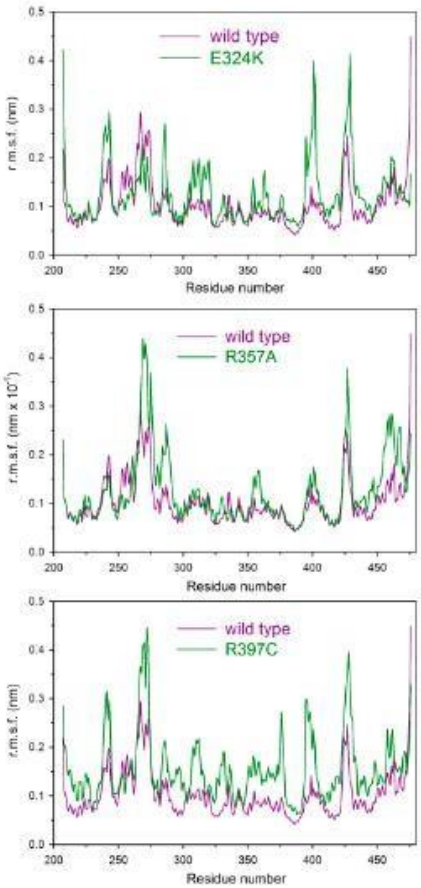


Figure 6. Mobility per residue of PPARγ wild type and E324K, R357A, and R397C variants. Cα root mean square fluctuations (r.m.s.f.) per residue for E324K, R357A, and R397C variants compared with the wild type r.m.s.f. On the *x*-axis is the residue number and on the *y*-axis is the mobility in nm.

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

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

where $\Delta G_{unfolding}$ 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 $\Delta G_{unfolding}$ per unit concentration of denaturant, R is the gas constant, T is the temperature and $K_{unfolding}$ is the equilibrium constant for unfolding. The model expresses the signal as a function of denaturant concentration:

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

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 $\Delta G_{unfolding}$ versus $[X]$ plot. The denaturant concentration at the midpoint of the transition, $[\text{Urea}]_{0.5}$, according to Equation (2), is calculated as:

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

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 domain:effect 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. Raspe´, 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

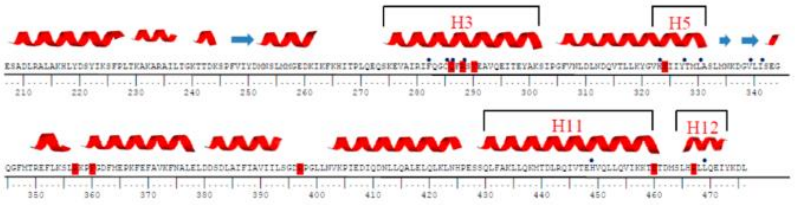


Figure S1. Amino acid sequence of PPARγ LBD. Secondary structure elements are shown at the top of the amino acid sequence. Mutated residues are highlighted in red. Blue dots indicate the residues involved in ligand binding (1PRG.pdb). Helices are numbered according to Nolte et al. [1].

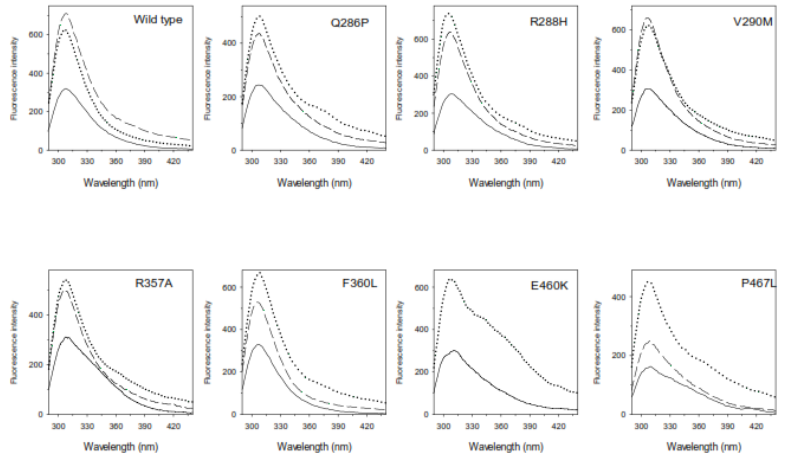


Figure S2. Intrinsic fluorescence emission spectra of PPARγ wild type and variants. Fluorescence spectra of PPARγ wild type and variants in 0 M (continuous lines), 8.0 M (dotted lines), 4.00 M (Q286P and F360L, dashed lines), and 4.07 M urea (wild type, R288H, V290M, and R357A, dashed lines) were recorded 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.

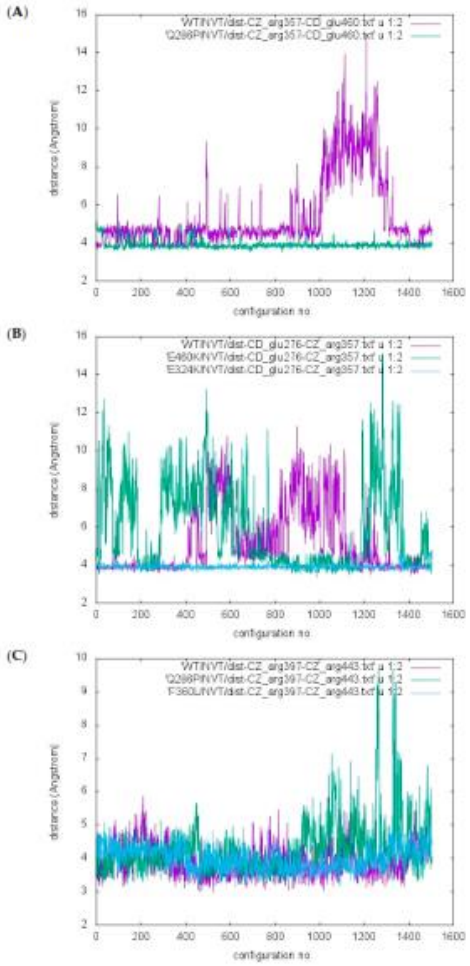


Figure S3. (**A**) Distance between R357 and E460. Distance between the carbon atom binding the two amino groups of arginine 357 lateral chain (CZ) and the carbon atom binding the carboxyl group of glutamic 460 lateral chain (CD), in the wild type and in the Q286P variant. On the *x*-axis is the configuration 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.

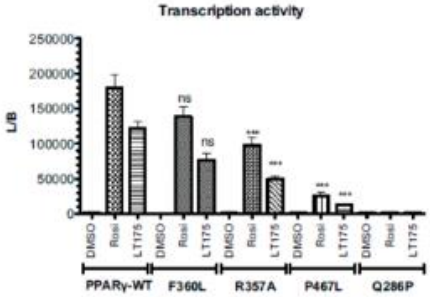


Figure S4. Transcription activity of wild type PPARγ and mutants in a PPARγ Gal4-based assay. Luciferase activity was normalized for difference in transfection efficiency by β-galactosidase activity and each point is the mean ± SEM of at least two experiments each performed in duplicate. Results are 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).

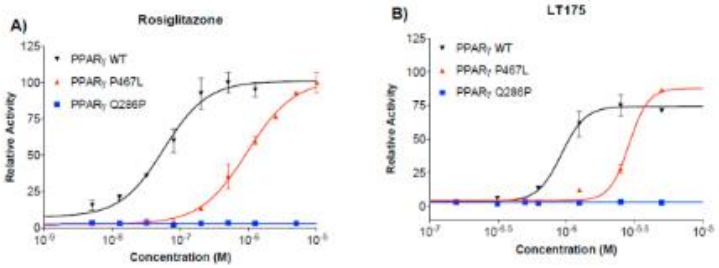


Figure S5. Transcription activity. Transcription activity of rosiglitazone (**A**); and LT175 (**B**) towards wild type PPARγ, PPARγ P467L, and PPARγ Q286P in a PPARγ Gal4-based assay. Results are expressed as a percentage of the highest efficacy obtained with rosiglitazone and each point is the mean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

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

OPEN ACCESS

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: Fafah Kashanchi, George Mason University, UNITED STATES

Received: April 15, 2016

Accepted: June 26, 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

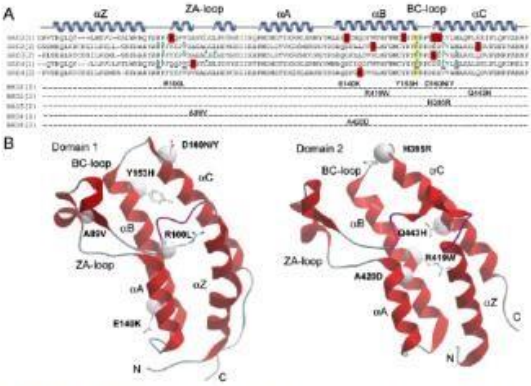


Fig 1. Alignment of BET bromodomain mutants. (A) Secondary structure elements are shown at the top of the sequence alignment. Mutated residues are highlighted in red and studied mutations are listed. The conserved asparagine (N391in BRD3(2) numbering) is highlighted in yellow. The green dots represent the residues involved in binding with inhibitor JQ1 (PDB ID: 3ONI, 3S92, 3MXF). The residues underlined in blue are involved in PFI-1binding (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

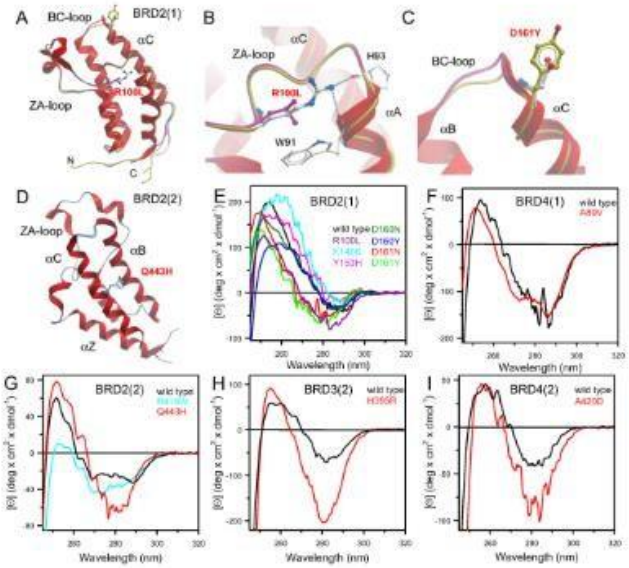


Fig 2. Structure of BET mutants and tertiary structure of mutants in solution. (A) Superimposition of wild type BRD2(1) shown as a ribbon diagram with the mutants BRD2(1) R100L and D161Y shown as protein worm in green and magenta, respectively. The mutated residues are shown in ball and stick representation 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

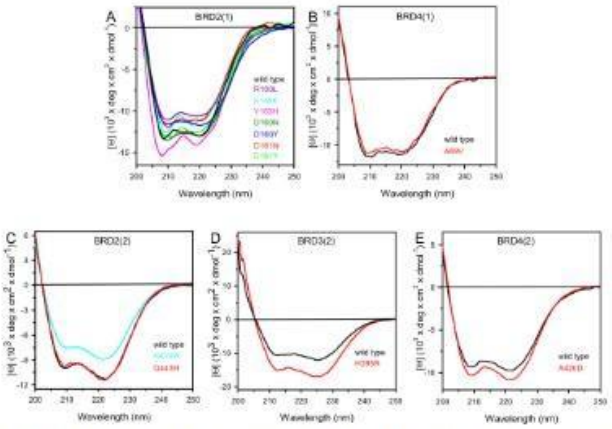


Fig 3. Far-UV CD spectra of wild type bromodomains and mutants. Far-UV CD spectra were recorded at 20°C in a 0.1-cm quartz cuvette in 20 mM Tris/HCl, pH 7.5 containing 0.20 M NaCl and 0.40 mM DTT, as described in Materials and Methods. Wild type spectra are shown as black solid lines and mutants are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

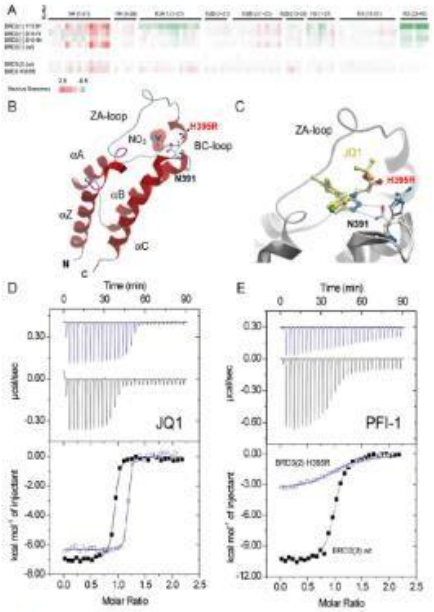


Fig 4. Binding of BET bromodomain mutants to acetylated histone peptides and inhibitors. (A) BLI study showing the interaction of some mutants with acetylated histone peptides. Shown is the maximum response at a protein concentration of 20 μM after subtraction of non-acetylated reference peptides using 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, R100L, 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 Vibra-cell 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 pre-packed 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: $A = U \cdot S \cdot V^T$, 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

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

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:

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

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:

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

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

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

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

where a_N and a_U are the baseline intercepts for N and U, 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

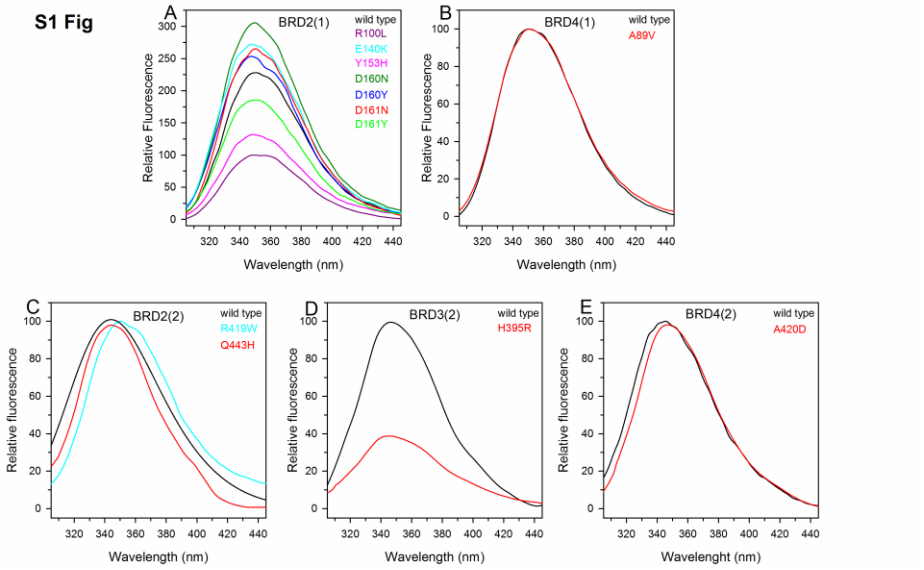
56. Perrakis A, Morris R, Lamzin VS. Automated protein model building combined with iterative structure refinement. Nature Structural Biology. 1999; 6: 458–463 PMID: 10331874

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

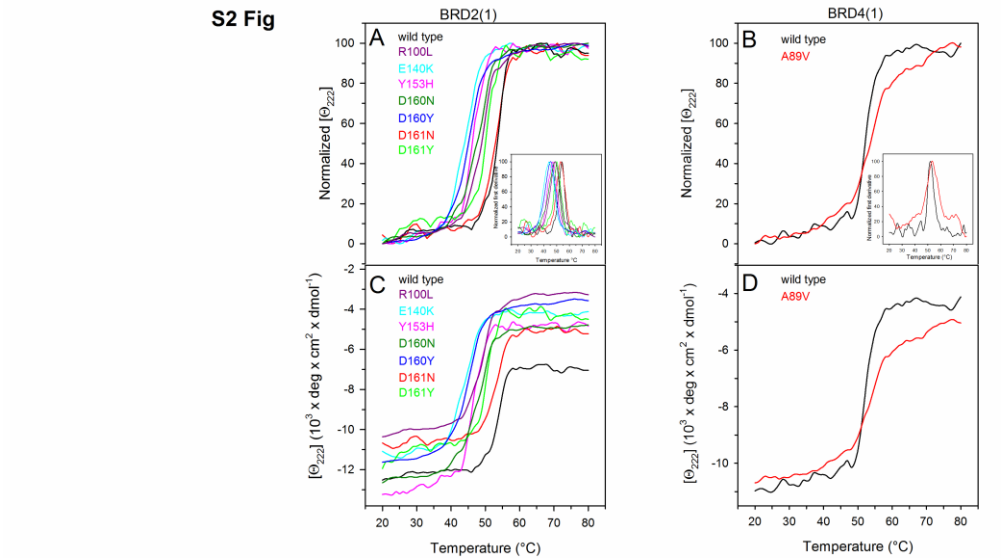


S1Table. Data collection and refinement statistics

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 Fig



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

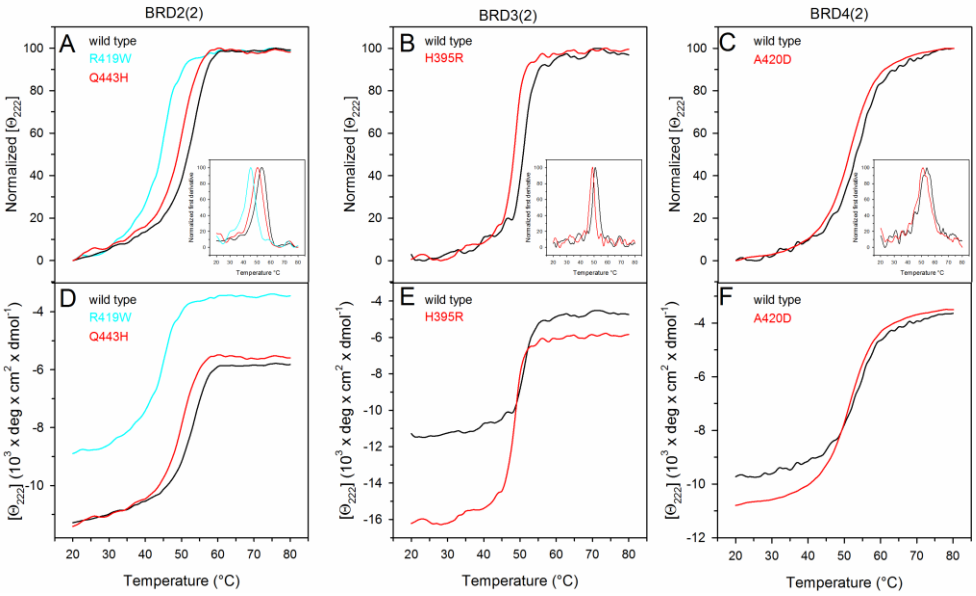
(continued on next page)

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

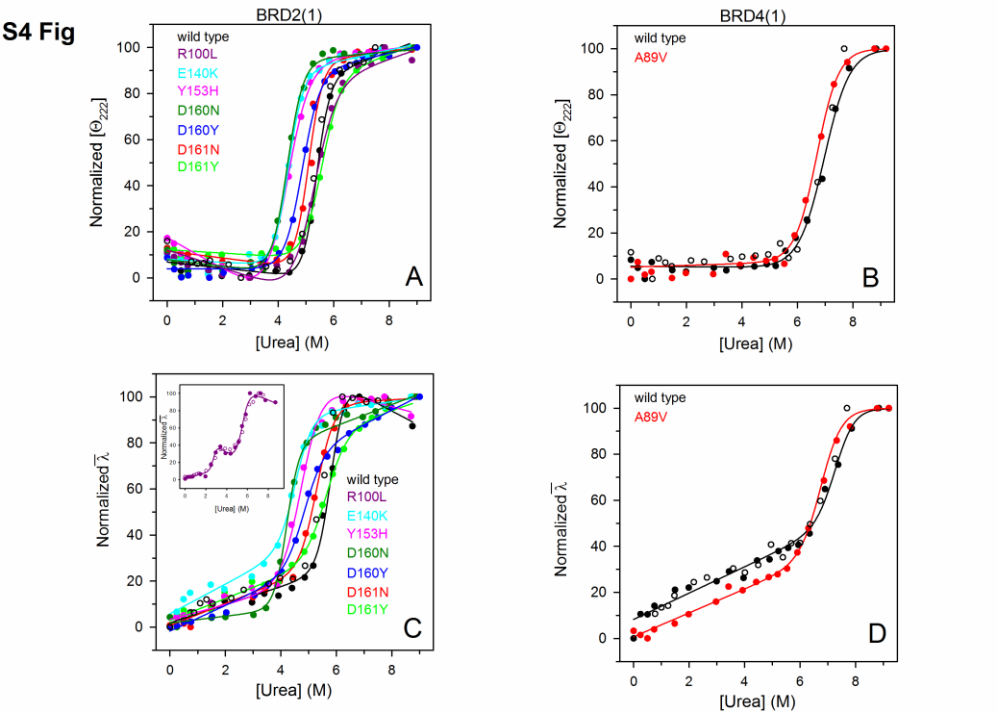
S3 Fig



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

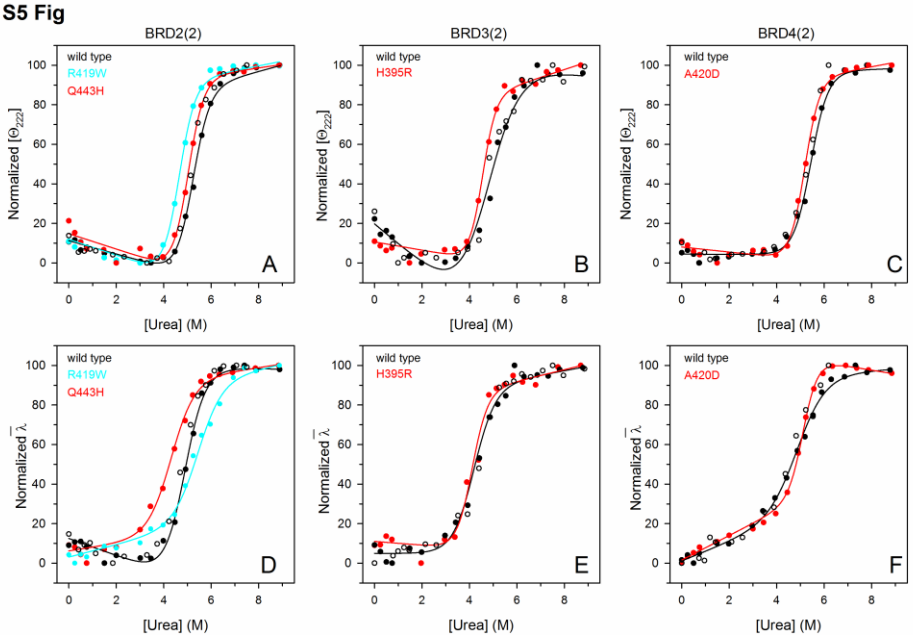
S4 Fig


BRD2(1)
wild type
R100L
E140K
Y153H
D160N
D160Y
D161N
D161Y
Normalized [Θ222]
[Urea] (M)
A
BRD4(1)
wild type
A89V
Normalized [Θ222]
[Urea] (M)
B
Normalized λ̄
wild type
R100L
E140K
Y153H
D160N
D160Y
D161N
D161Y
[Urea] (M)
C
wild type
A89V
Normalized λ̄
[Urea] (M)
D

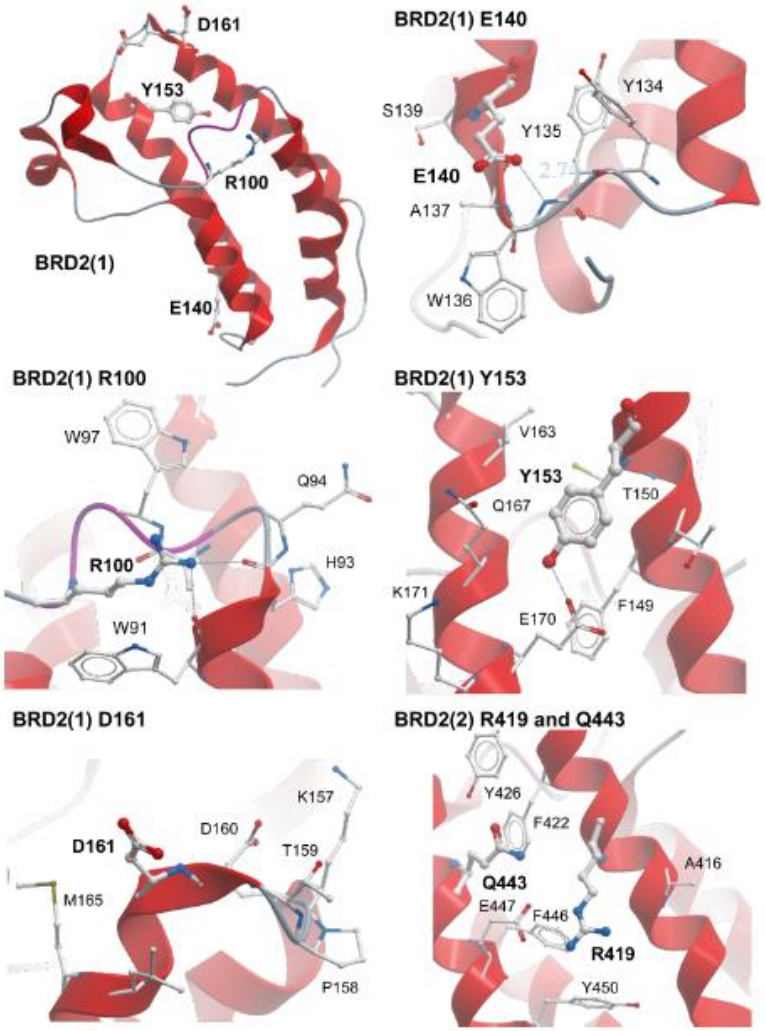
S4 Table. List of oligonucleotides used for site-directed mutagenesis

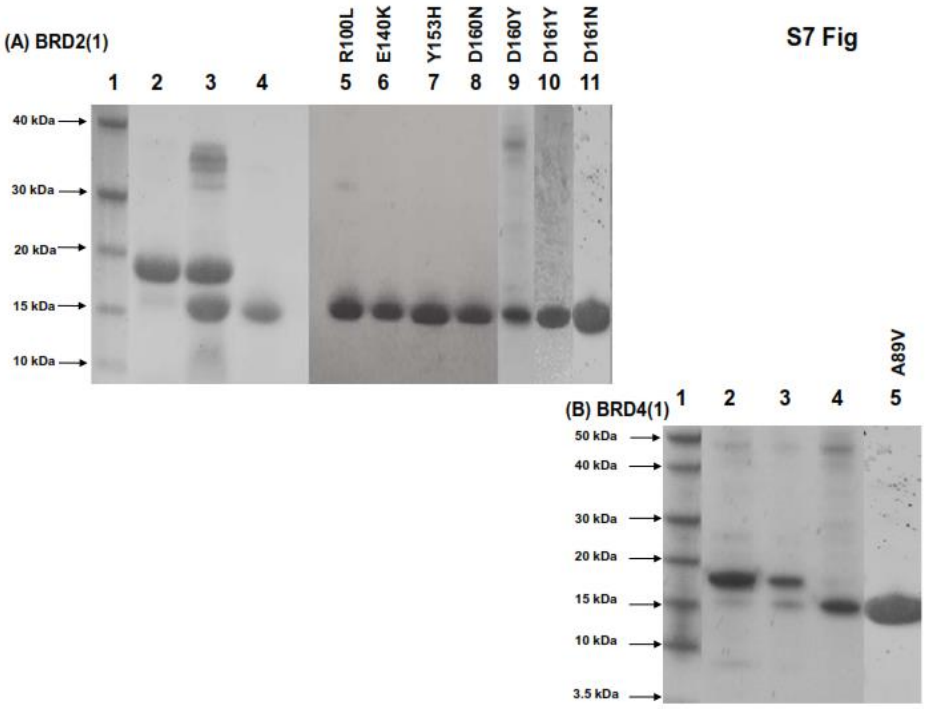
Mutant	Primer sequences (5' to 3')
BRD2(1)	
R100L FW	GTG GCC ATT TCT CCA GCC GGT GGA T
R100L REV	ATC CAC CGG CTG GAG AAA TGG CCA C
E140K FW	GGG CGG CCT CTA AAT GCA TGC AGG
E140K REV	CCT GCA TGC ATT TAG AGG CCG CCC
Y153H FW	GTT TAC GAA TTG TCA TAT CTA TAA CAA ACC G
Y153H REV	CGG TTT GTT ATA GAT ATG ACA ATT CGT AAA C
D160N FW	TAA CAA ACC GAC CAA TGA TAT CGT CTT GAT GG
D160N REV	CCA TCA AGA CGA TAT CAT TGG TCG GTT TGT TA
D160YFW	TAA CAA ACC GAC CTA TGA TAT CGT CTT GAT GG
D160Y REV	CCA TCA AGA CGA TAT CAT AGG TCG GTT TGT TA
D161N FW	CAA ACC GAC CGA TAA TAT CGT CTT GAT GG
D161N REV	CCA TCA AGA CGA TAT TAT CGG TCG GTT TG
D161YFW	CAA ACC GAC CGA TTA TAT CGT CTT GAT GG
D161YREV	CCA TCA AGA CGA TAT AAT CGG TCG GTT TG
BRD4(1)	
A89V FW	GCA GCC TGT GGA TGT CGT CAA GCT GAA CC
A89V REV	GGT TCA GCT TGA CGA CAT CCA CAG GCT GC
BRD2(2)	
R419W FW	GCG GCG GAT GTA TGG CTC ATG TTT TCG
R419W REV	CGA AAA CAT GAG CCA TAC ATC CGC CGC
Q443H FW	GGC ACG AAA GCT ACA CGA TGT ATT TGA G
Q443H REV	CTC AAA TAC ATC GTG TAG CTT TCG TGC C
BRD3(2)	
H395R FW	AAT CCC CCA GAC CGC GAG GTT GTG GCC AT
H395R REV	ATG GCC ACA ACC TCG CGG TCT GGG GGA
BRD4(2)	
A420D FW	CAG GAG TTT GGT GAT GAC GTC CGA TTG
A420D REV	CAA TCG GAC GTC ATC ACC AAA CTC CTG

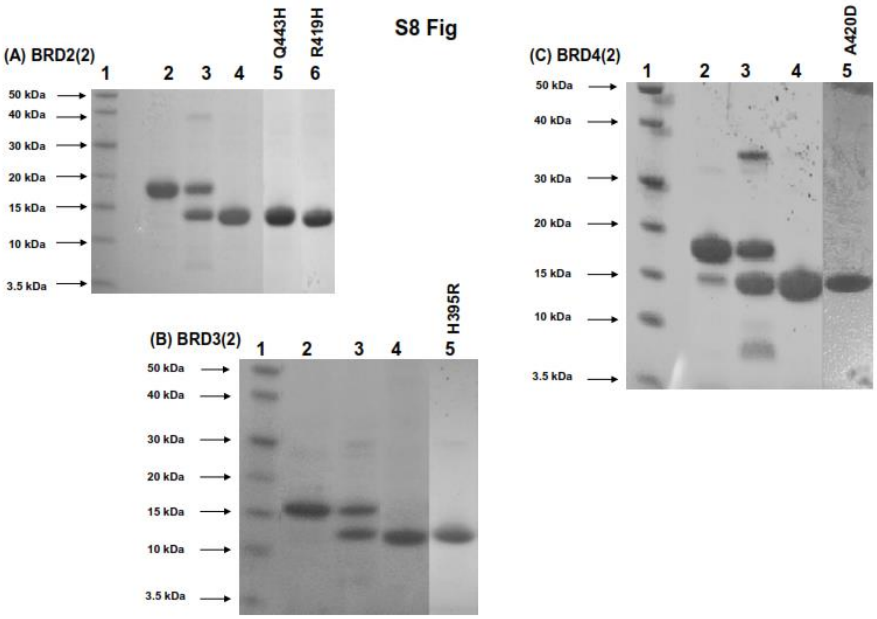
S5 Fig


BRD2(2)
BRD3(2)
BRD4(2)
wild type
R419W
Q443H
H395R
A420D
Normalized [Θ222]
Normalized λ̄
[Urea] (M)
A
B
C
D
E
F

S6 Fig


D161
Y153
R100
BRD2(1)
E140
BRD2(1) E140
S139
Y134
Y135
E140
A137
W136
BRD2(1) R100
W97
Q94
R100
H93
W91
BRD2(1) Y153
V163
Y153
Q167
T150
K171
E170
F149
BRD2(1) D161
K157
D160
D161
T159
M165
P158
BRD2(2) R419 and Q443
Y426
F422
A416
Q443
E447
F446
R419
Y450


S7 Fig
(A) BRD2(1)
1 2 3 4 5 6 7 8 9 10 11
R100L
E140K
Y153H
D160N
D160Y
D161Y
D161N
40 kDa
30 kDa
20 kDa
15 kDa
10 kDa
(B) BRD4(1)
1 2 3 4 5
A89V
50 kDa
40 kDa
30 kDa
20 kDa
15 kDa
10 kDa
3.5 kDa


S8 Fig
(A) BRD2(2)
1 2 3 4 5 6
Q443H
R419H
50 kDa
40 kDa
30 kDa
20 kDa
15 kDa
10 kDa
3.5 kDa
(B) BRD3(2)
1 2 3 4 5
H395R
(C) BRD4(2)
1 2 3 4 5
A420D

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 Roma, Italy
[b] SSPT-BIOAG-BIOTEC ENEA Casaccia ENEA, Roma, 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

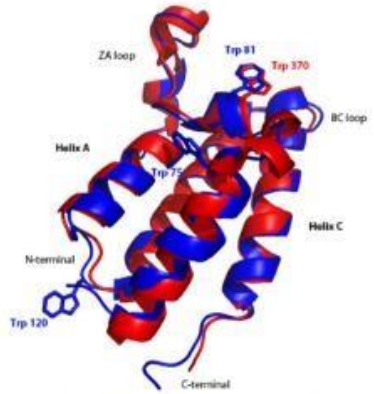


Fig. 1. Structural alignment of BRD2(2) (pdb id: 3oni), shown in red, and BRD4(1) (pdb id: 3uvx), shown in blue. The Trp residues are shown in sticks representations. BRD2(2) contains only one Trp (Trp370), structurally homologous to Trp81 of BRD4(1)); BRD4(1) presents two additional Trp residues (Trp75 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

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

where ΔG_{unf} is the free energy change for unfolding for a given denaturant concentration, $\Delta G_2^{H_2}$ 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:

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

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, $[\text{Urea}]_{0.5}$, according to Eq. (2), is calculated as:

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

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: $\ln k_0 = \ln k_0^{\circ} + m_0 (RT)^{-1}[\text{urea}]$, where k_0° and m_0 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.

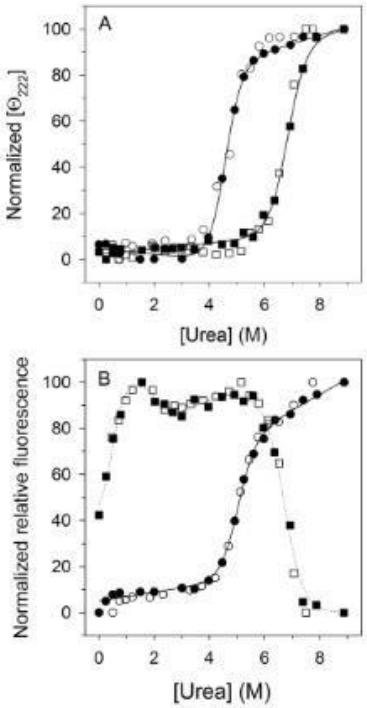


Fig. 2. Urea-induced equilibrium unfolding of BRD2(2) and BRD4(1). (A) Normalized molar ellipticity at 222 nm ($[\Theta]_{222}$) of BRD2(2) (circles) and BRD4(1) (squares) reported after removal of the high-frequency noise and the low-frequency random error by SVD. Continuous lines are the nonlinear 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 ($[\Theta]_{222}$)
$\Delta G^{H_2O}_2$ (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).

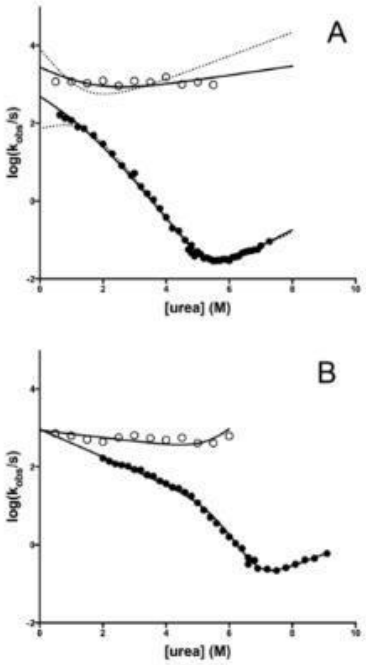


Fig. 3. Semilogarithmic plot (chevron plot) of the observed folding and unfolding rate constants as a function of urea concentration by SF (closed symbols) and T-jump (open symbols) experiments for BRD2(2) (panel A) and BRD4(1) (panel B). In panel A, the lines are the best fit to a three-state model 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_{IN} = 570 ± 211; m_{IN} = 0.58 ± 0.34; k_{NI} = 0.00010 ± 0.00008; m_{NI} = 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_{IN} = 875 ± 115; m_{IN} = 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, $k_1 \sim 100\ s^{-1}$ and $k_2 \sim 1\ s^{-1}$) (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

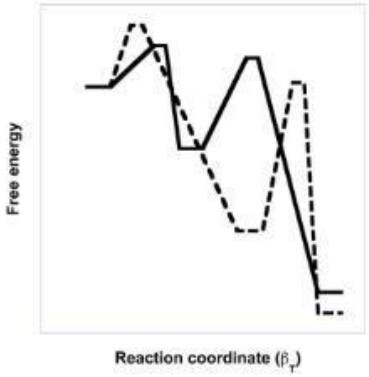


Fig. 4. Schematic energy diagrams for the folding of BRD2(2) (continuous line) and BRD4(1) (dashed line). The free energy values and the β_T values, were calculated from the kinetic parameters obtained by fitting the chevron plots of the two BRD domains to an on pathway intermediate (see the legend to 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): $\Delta G_{UI} = 1.14$ kcal mol^{-1}; $\Delta G_{IN} = 10.16$ kcal mol^{-1}; β_T(TS1) = 0.32; β_T(I) = 0.40; β_T(TS2) = 0.71. BRD4(1): $\Delta G_{UI} = 6.83$ kcal mol^{-1}; $\Delta G_{IN} = 14.80$ kcal mol^{-1}; β_T(TS1) = 0.22; β_T(I) = 0.75; β_T(TS2) = 0.80.

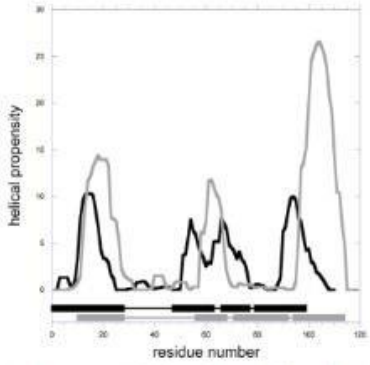


Fig. 5. Helical propensity of BRD2(2) (black line) and BRD4(1) (grey line). Helical propensity was calculated by the AGADIR program [31] (http://agadir.crg.es) Boxes represent the four helices, as determined from the X-ray structure of BRD2(2) (pdb id: 3oni; black boxes) and BRD4(1) (pdb id: 3uvx; grey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

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) ($\Delta\Delta G_{UI} = 5.7 \pm 0.8$ 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 ($\Delta\Delta G_{IN} = 4.3 \pm 1.3$ 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 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] B. 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. Cui, S.E. Ibaq, S. Rinaldo, J. Dogan, F. Cutruzzola, A. Engstrom, S. Gianni, P. Lundstrom, F. Jemth, Biochemistry 51 (2012) 8971–8979.
[9] S. Gianni, J. Dogan, F. 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. Kaplita, S. Chandrasekaran, M. Thompson, Biophys. Chem. 136 (2008) 7–12.
[12] S. Picaud, M. Strocchio, S. Terracciano, G. Lauro, J. Mondal, D.L. Daniels, R. Riccio, G. Bifulco, I. Bruno, P. Filippakopoulos, J. Med. Chem. 58 (2015) 2718–2736.
[13] A. Borgia, D. Borgia, C. Travaglini-Allocatelli, A. Di Matteo, M. Brunori, J. Biol. Chem. 281 (2006) 9351–9356.
[14] A.P. Capaldi, M.C.R. Shanny, C. Kleanthous, H. Roder, S.E. Radford, Nat. Struct. Biol. 8 (2001) 68–72.
[15] K. Teilum, 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. Paiardini, 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, E.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. Kawajima, 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.