UNIVERSIDAD DE GRANADA

THE ROLE OF PROTEIN DYNAMICS IN LOSS-OF-FUNCTION DISEASE MECHANISMS: A NQO1 CANCER-ASSOCIATED POLYMORPHISM AS MODEL

(Programa Oficial de Doctorado en Química)

Encarnación Medina Carmona

Departamento de Química Física

Facultad de Ciencias

Universidad de Granada

INDEX

ABSTRACT

Mutations causing single amino acids replacements, also known as missense mutations, can noticeably affect diverse protein properties such as the ability to fold or the stability inside cells which often lead to disease. Specifically, disease-causing missense mutations that bring about a decreased functionality and/or stability produce the so-called loss-of-function diseases. There are numerous diseases associated with loss-of-function phenotype, involving from a large number of inherited monogenic diseases to more complex multifactorial diseases like cancer[1,2]. In particular, this doctoral thesis focuses on the study of a NAD(P)H: quinone-oxidoreductase 1 (NQO1) polymorphism, P187S, which has been associated with increased cancer risk[3]. P187S is a paradigm of loss-of-function mutation that causes significant defects on stability and function of NQO1. On the one hand, this polymorphism causes a severe reduction of enzymatic activity, possibly as a consequence of its low affinity for FAD, while its intracellular instability is linked to an enhanced proteasomal degradation[4–7]. Several researches have focused on P187S showing a X-ray structure cristallographic virtually identical compared to the structure of wild-type NQO1. In addition, Nuclear Magnetic Resonance (NMR) spectroscopy and proteolysis experiments confirmed the presence of partially unfolded states in solution although these results cannot totally explain the destructive effects of P187S *in vitro* and *in vivo*[5]. In this context, we suggest that the pathogenic mechanisms of this polymorphism could be deciphered with an in-depth analysis of the protein dynamics.

We used a combination of experimental and computational techniques to investigate the P187S effects on protein dynamics. Interestingly, the first section of our results showed that P187S seems to affect the local dynamics of two functionally and structurally distant sites of NQO1: the N-terminal domain (NTD), related to enzyme inactivation; and the C-terminal domain (CTD) which appears to play a key role in the low intracellular stability of P187S. In addition, the treatment of P187S with its natural ligand FAD may rescue *in vitro* activity of P187S but the combination with a second ligand (i.e. dicoumarol) that acts as stabilizer of the still dynamic CTD, is required to correct the degradation through this domain.

With the goal of evaluating the structural and functional role of CTD, this domain was exhaustively evaluated. Our results provided evidence for the existence of a multifunctional allosteric network between the CTD, the FAD binding site and the P187S site. For wild-type NQO1, we suggest that this network contributes to a very high affinity for FAD and dicoumarol, whereas P187S seems to disturb the network by dynamic and structural changes in the FAD binding site and the CTD affecting the binding affinity of P187S for both ligands.

In the third section, we tried to identify hot spots that would correct the NQO1 alterations caused by P187S by consensus approach[8,9]. Specifically, we found a suppressor mutation H80R which reactivates P187S by local and dynamic stabilization of FAD binding site. The deep characterization of the changes exerted by H80R suggested that this suppressor mutation mainly acts by shifting the conformational equilibrium of the apo-state towards more competent states for the binding of FAD. A second consensus mutation E247Q showed a stabilizing effect on the CTD and this

effect is propagated to the distal NTD supporting the existence of the previously proposed network. In addition, the combination of these two suppressor mutations strongly protected NQO1 from the loss-of-function caused by P187S.

In summary, the results of this doctoral thesis show that protein dynamics plays a fundamental role in the loss-of-function mechanisms associated with P187S, and its understanding is crucial to find alternatives that correct the deleterious effects of this polymorphism. We propose that our multi-disciplinary strategy could help to decipher complex mutational effects related to disease in many other loss-of-function genetic diseases, and to identify structural hot spot as targets for pharmacological correction.

RESUMEN

Las mutaciones que causan la substitución de un único aminoácido, también conocidas como mutaciones de cambio de sentido, pueden afectar notablemente diversas propiedades de las proteínas como por ejemplo su habilidad para plegarse, o su estabilidad dentro de las células y a menudo estos cambios conllevan a enfermedades. En particular, las enfermedades causadas por este tipo de mutaciones y que provocan una disminución de la funcionalidad y/o estabilidad de la proteína son comúnmente conocidas como enfermedades de pérdida de función. Existen numerosas enfermedades asociadas con el fenotipo de pérdida de función, incluyendo desde un gran número de enfermedades monogénicas hasta enfermedades multifactoriales más complejas como el cáncer[1,2]. Concretamente, esta tesis doctoral se centra en el estudio de un polimorfismo de la proteína NAD(P)H quinona oxidorreductasa 1 (NQO1), el polimorfismo P187S que ha sido asociado con un incremento del riesgo a padecer cáncer[3]. P187S es un paradigma de mutación de pérdida de función y causa defectos significativos en la estabilidad y función de NQO1. Por un lado, este polimorfismo provoca una reducción importante de la actividad enzimática, posiblemente como consecuencia de su baja afinidad por el ligando FAD, mientras su inestabilidad intracelular es relacionada con un aumento en la degradación proteasomal[4–7]. Varias investigaciones se han centrado en el estudio de P187S mostrando que este polimorfismo presenta una estructura cristalina virtualmente idéntica si la comparamos con la estructura de la proteína natural de NQO1. Además, estudios de Resonancia Magnética Nuclear (RMN) y experimentos de proteólisis confirmaron la

presencia de estados parcialmente desplegados en disolución aunque estos resultados no podían explicar los nocivos efectos de P187S en su totalidad. En este contexto, nosotros sugerimos que los mecanismos patogénicos de este polimorfismo podrían ser descifrados con un profundo análisis de la dinámica de la proteína.

Se usaron una combinación de técnicas experimentales y computacionales para investigar los efectos de P187S en la dinámica de la proteína. De modo interesante, la primera sección de nuestros resultados muestra que P187S parece afectar la dinámica local de dos sitios distantes funcionalmente y estructuralmente de NQO1: el dominio N-terminal (NTD) relacionado con la inactivación de la enzima, y el dominio C-terminal (CTD) el cual parece jugar un papel clave en la baja estabilidad intracelular de P187S. Además, el tratamiento de P187S con su ligando natural FAD podría rescatar su actividad *in vitro* aunque la combinación con un segundo ligando (p. ej. dicumarol), que actúe como estabilizador del todavía dinámico CTD, es necesaria para corregir su degradación a través de este dominio.

Con el objetivo de evaluar el papel funcional y estructural del CTD , este dominio fue exhaustivamente analizado en la segunda sección. Nuestros resultados aportan evidencias de la existencia de una red alostérica y multifuncional entre el CTD, el sitio de unión a FAD y el sitio para P187S. Nosotros sugerimos que para la proteína natural de NQO1 esta red contribuye con una alta afinidad por FAD y dicumarol. Sin embargo, la presencia de P187S parece alterar la red, mediante cambios estructurales y dinámicos en el sitio de unión a FAD y el CTD afectando la afinidad de unión de P187S por ambos ligandos.

En la tercera sección se intentaron encontrar puntos clave estructuralmente que corrigieran los defectos causados por P187S mediante la "técnica de consenso". En

concreto, se encontró una mutación de consenso H80R que reactiva P187S mediante la estabilización local del sitio de unión a FAD. La amplia caracterización de los cambios ejercidos por H80R sugirió que esta mutación actúa principalmente desplazando el equilibrio conformacional del estado apo de la proteína hacia estados más competentes para unir FAD. Una segunda mutación consenso E247Q muestra un efecto estabilizante sobre el CTD de P187S y el efecto se propaga al distante dominio NTD apoyando la existencia de la red previamente propuesta. Además la combinación de estas dos mutaciones supresoras protege fuertemente a NQO1 de la pérdida de función causada por P187S.

En general, los resultados de esta tesis doctoral muestran que la dinámica de proteínas juega un papel fundamental en los mecanismos de pérdida de función asociados con P187S y su comprensión es crucial para encontrar alternativas que corrijan los devastadores efectos de este polimorfismo. Nosotros proponemos que esta estrategia multidisciplinar podría ayudar a descifrar complejos efectos mutacionales asociados con enfermedad en otras muchas enfermedades de pérdida de función y además puede servir para identificar estructuralmente puntos clave que sirvan como dianas para la corrección farmacológica de estas enfermedades.

1. INTRODUCTION

1.1 MISSENSE POLYMORPHISMS AND THEIR EFFECTS

1.1.1 BACKGROUND

The human organism develops through tightly regulated mechanisms of cellular division, differentiation and death. During cellular division, the genetic content of the cell must be accurately replicated and therefore, all cells composing the whole human organism should contain exactly the same genetic information. Nevertheless, a significant number of changes can occur in the genetic sequence, and these changes are known as ***mutations.*** According to the cell type in which the mutations occur, they are classified as germinal or somatic mutations. Germinal mutations take place in the germ cell lineage (i.e. the sperm and egg) and can be transmitted to offspring, while the somatic mutations are acquired mutations, and are not heritable. Last ones occur when DNA is duplicated or they can be caused when DNA is damaged by environmental factors, including UV radiation, chemicals, and viruses. It is estimated that each human carries an average of 60 de novo mutations that arise during gametes formation in their parents, while estimated mutation rates in somatic cells are 4 to 25

times higher[10]. Together, these different types of mutations provide the raw material for evolution, but at the same time may place organisms at risk for disease[11,12].

In an evolutionary context, the changes in DNA sequence may are usually divided based on their "fitness" effects. Although the fitness is an abstract concept and a huge number of definitions have been offered, there is a common idea broadly accepted: the fitness is a measure of the capacity of an organism to survive and reproduce in the environment in which it is found[13,14]. Three categories can be considered when discussing the mutations effects on fitness[15,16]. First one includes mutations that are advantageous because they have positive effects on fitness and are essential to evolution. Second category encompasses, neutral mutations, which have little or no known effect on fitness. And finally, detrimental mutations which generally reduce organisms survival and/or fertility, and they are rarely fixed. Although, this simplistic classification serves as a first approximation to understand the mutations fate, it omits other traits that may also affect fitness, such as evolutionary context or living environment. In this way, the effects of mutation are context-dependent and for example, they may depend on the presence or absence of other mutations and/or their proximity on the same chromosome, or they may also depend on the living environment of the organism (a mutation can be neutral or deleterious in one environment but it can become beneficial if the context changes)[17,18].

Alternatively, changes in DNA sequences can be classified according to their pathogenicity, understanding pathogenicity as the property by which mutations confer risk for genetic disorders[10]. The National Human Genome Research Institute (NHGRI) defines genetic disorder as "a disease caused in whole or in part by a change in the

DNA sequence away from the normal sequence". When these disorders are caused by germinal mutations that only occur in the DNA sequence of a single gene, they are called **Mendelian disorders** (or monogenic disorders). On the contrary, the disorders that involve variations in multiple genes often coupled with environmental causes, are called multifactorial disorders. Although many genetic disorders can be categorized as multifactorial, research on monogenic diseases not only provide a priceless opportunity to learn about gene functions and for reconstructing normal and pathological pathways, but they would also contribute significantly to our understanding of the molecular genetic basis of common multifactorial diseases. There are more than 10000[*] known monogenic disorders, and different types of mutations can be the cause. The vast majority is associated with mutations that just involve a change on a single-base pair or a few base pairs and they are typically referred to as point mutations. Point mutations may consist on insertions or deletions of nucleotides in the DNA sequence shifting the way in which the sequence is read (frameshift mutations), or they can involve a substitution of one nucleotide for a different one. These last ones are known as nucleotide substitutions. Theoretically, if a nucleotide substitution exists in at least 1% of the total population, then it will be considered as **Single Nucleotide Polymorphisms** (SNP). SNPs are the most abundant variants in the human genome and they can be found in different parts of a gene, such as in coding or non-coding regions, with or without any effects on the gene product[19,20]. Special mention requires the SNPs that occur in the protein-coding part of the DNA sequence since they have the potential to change the protein codified. They can be categorized as: synonymous or silent, if the amino acid sequence is not altered (it is allowed for the

---

[*] The source which estimates this data is the World Health Organization.

redundancy in the genetic code); non synonymous or missense, where the change of a single base pair causes a change to a different amino acid in the resulting protein; or nonsense, if a codon that stands for an amino acid mutates to one of three stop codons in the gene code, bringing about an incomplete protein which is usually non-functional[21]. Traditionally, the focus has been non synonymous SNPs (nsSNPs), since many of them have been related to monogenic disease and they have also been associated with common multifactorial diseases such as heart disease, hypertension or cancer[22]. Recent evidences also suggest that synonymous SNPs (sSNPs) would result in aberrant mRNA splicing, or affect mRNA stability and thus, protein expression and enzymatic activity, leading to human diseases[23].

Of special interest for this thesis are the nsSNPs which will usually be refered to throughout this introduction as *missense polymorphisms*. They can directly affect diverse protein properties such as, protein folding and stability, protein function, protein–protein interactions, protein expression and subcellular localization[2,24–27]. Considering that most of the monogenic diseases are caused by missense polymorphisms, the analysis of their impacts on proteins increases our understanding of the relationships between protein structure and function, and it would allow us to decipher phenotype-genotype relations[28]. Numerous investigations about them have been undertaken, resulting in a wide variety of tools and databases such as, the Online Mendelian Inheritance in Man (OMIM) database, the Human Gene Mutation Database (HGMD), single nucleotide polymorphism database (dbSNP) or the Protein Mutant Database (PMD) among others[22].

A mutation can impact the function of a protein altering one or more of its properties such as structure, catalytic activity, or stability. From a biophysical point of view, understanding the mechanism by which a mutation impacts protein function has traditionally required specialized assays to measure these properties, for example, X-ray diffraction to study the structure or thermal denaturation to measure stability. Historically, missense polymorphisms have been studied paying close attention to understand how missense mutations influence folding, stability, solubility, activity and other structure-based properties, for example, through experimentally derived structures and homology models[29–31]. Nevertheless, and despite the wide available information, prediction of the effect and associated phenotypes of missense polymorphisms remains a challenge. Maybe, the principal cause is that the tools used for their analysis (i.e. thermal denaturation or X-ray crystallography) divide the collective effects of mutations into individual contributions from various structural and physicochemical features. However, many properties used to describe them are highly connected, making it difficult to separate them from each other[32,33]. Furthermore, it deserves to be mentioned that these collective effects at molecular levels must be linked to effects on protein function and stability *in vivo*[2,34].

1.1.2 PROTEIN FOLDING *in vitro* AND *in vivo*

Proteins are essential molecules in the most of cell processes, for instance, in signalling, catalysis, mobility, or recognition. There are proteins capable of performing their biological functions without achieving a stable structure, such as intrinsically

disorder proteins (IDPs) or proteins with intrinsically disordered regions. However, a large part of the proteins need to acquire a specific three-dimensional structure to carry out their function correctly[34–36]. Specifically, protein folding refers to the process by which a protein often folds in its right and functional conformation of low energy. This conformation is known as native state and is profoundly influenced by the sequence of amino acids. In particular, it has been shown that many proteins fold spontaneously *in vitro*, supporting the Anfinsen hypothesis for protein folding. It suggests that the linear sequence of the polypeptide chain contains all the necessary information to conduct the protein folding. This would imply a spontaneous process that would take place through a single step[37].

However, approximately 90% of all proteins in the cell are large proteins (>100 amino acids) and their folding require several steps (**Figure 1**)[35,38,39]. Although many aspects of protein folding *in vitro* are intrinsic to the biophysical properties of the protein itself, the folding process can be quite complex, requiring various intermediate steps along the way toward the native state. The intermediate steps are represented by various halfway conformations in a folding energy landscape[40].
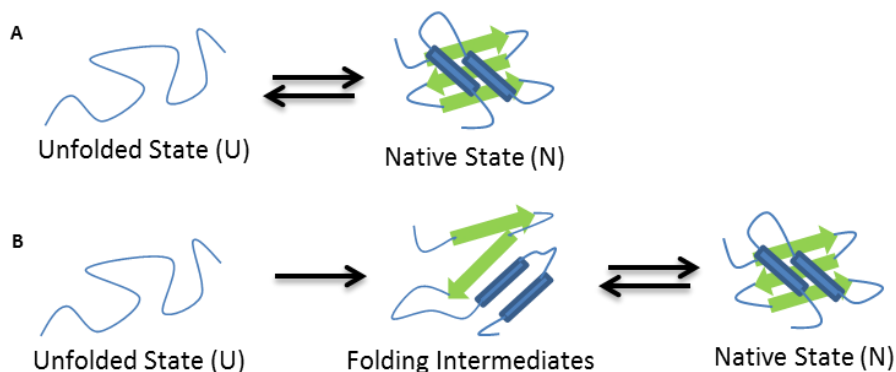

Figure 1: Protein folding process *in vitro* diagrams. A) Single-step process B) Folding process with intermediate steps.

The energy landscape theory of protein folding is a statistical description of a protein's potential energy surface[41]. The main idea that emerges from this theory is that the global folding landscape seems a funnel with traps in which the protein can transiently reside (significantly populated intermediates). Accordingly, several folding routes towards the native state can be followed[42]. Multiple factors in this dynamic process can result in misfolding. Specifically, a missense polymorphism can stabilize a non-native state in the process producing that this state is favoured over the native state. In addition, for many proteins the native state is not the state of least energy now that, they can form amyloid fibers or amorphous aggregates even more stable[34]. Given the link between correct folding and function, it is not surprising that these "misfolded" states may be associated with disease. However, the rugged potential surface navigated during folding *in vitro* is often modified *in vivo.* For instance, the folding process *in vivo* usually requires molecules to cross substantial kinetic energy barriers, and they avoid that intermediate states in folding process or misfolded states can be kinetically trapped in low energy wells before the native state is reached[38,40,43]. These molecules are present in all types of cells and cellular compartments and they are known as molecular chaperones (**Figure 2**)[42]. Some authors define molecular chaperone as "any protein that interacts, stabilizes or helps a non-native protein to acquire its native conformation, but is not present in the final functional structure" [38,44]. They are involved in different stages of protein folding that include from cotranslational folding, where they assists emerging chains from the ribosome, to later stages of the folding process, for instance cooperating on refolding processes of non-native states [38,39,43].
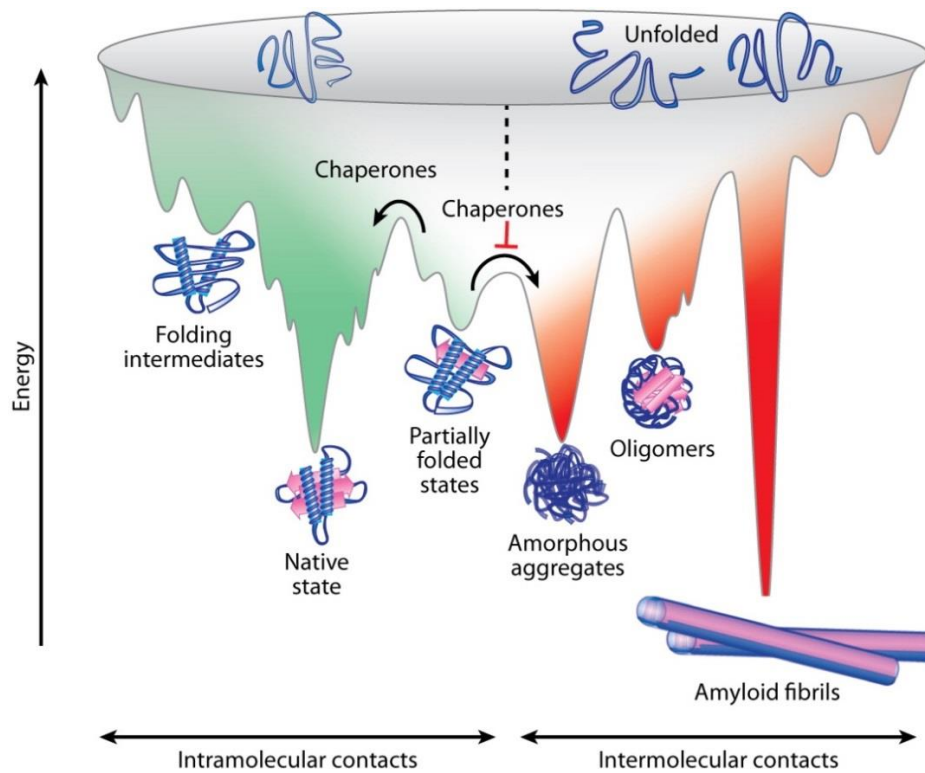
Figure 2. Energy landscape scheme of protein folding and aggregation. The green surface shows the multitude of conformations 'funneling' to the native state by intramolecular contacts. Red area shows the conformations moving toward amorphous aggregates or amyloid fibrils via intermolecular contacts. Both parts of the energy surface overlap Chaperones assist non-native states in overcoming free energy barriers and promoting folding to the native state. (Reprinted, with permission of Annual Review of Biochemistry, from Kim Y.E., 2013) [42].

Molecular chaperones do not raise the rate of individual steps in folding process; rather, they increase the efficiency of the overall process by reducing the probability of competing reactions as aggregation[39]. Folding efficiency *in vivo* depends on a balance between the physical properties of a protein like ***thermodynamic and kinetic stability***, and the interaction of the different protein states (partially folded or misfolded states)

with the ***protein homeostasis network*** throughout folding process[2]. At this juncture, it is important to remember that although some fundamental principles of protein folding have been established *in vitro* and they can be widespread, many details of the folding process *in vivo* depend on the particular environment where it takes place [39,43–45].

1.1.2.1 THERMODYNAMIC AND KINETIC STABILITY:

The intracellular levels of proteins in native state are influenced by their thermodynamic and kinetic stabilities[2,46]. Traditionally, the two-state equilibrium model has been used to introduce the concept of thermodynamic stability in a simple manner[2,35,47]. It describes well denaturation of many small proteins or single-domain proteins and it can be represented by the following model:

$$N \overset{K_U}{\leftrightarrow} U \tag{1}$$

where the protein "exist" only in two different macroscopic states, N for native state and U for unfolded state (or ensemble of unfolded states). This model implies a single step without significantly populated intermediate states. In this context, the thermodynamic stability is determined by the equilibrium unfolding constant (K_U), which is related to the standard unfolding free energy change (ΔG), (difference between the unfolded and native states) by the Lewis equation:

$$\Delta G(T) = G_U(T) - G_N(T) = -R \cdot T \cdot \ln(K_U) \tag{2}$$

The ΔG is dependent on temperature and at physiological temperatures (~37°C) it is usually a small number as a result of the enthalpy-entropy compensation considering that ΔG can be also expressed as:

$$\Delta G(T) = \Delta H(T) - T\Delta S(T) \tag{3}$$

These two temperature-dependent parameters, enthalpy (ΔH) and entropy (ΔS) changes, together with the denaturation heat capacity change (ΔCp) provide a whole description of thermodynamic stability.

Thus, thermodynamic stability determines unfolded protein fraction at equilibrium which often leads a constant value lower than one ($K_U < 1$) at physiological temperatures. It indicates that the unfolding equilibrium is shifted towards the functional native state leading to a thermodynamically stable situation which changes with temperature.

However, it must also be considered that the unfolding equilibrium is dynamic and displays kinetic features. The functional native state of the protein should be maintained, at least during a physiologically relevant time-scale, to talk about kinetic stability. The concept of kinetic stability supposes that in a folding/unfolding reaction there is a kinetic barrier that separates the native state from the non-functional forms, even if the native state is not the most thermodynamically stable[2,34,47]. The inactive forms may include unfolded and partially unfolded states, as well as irreversibly denatured states which are very common both *in vitro* and *in vivo.* In particular, the unfolded states are prone to establish nonnative intermolecular contacts with other adjacent macromolecules in consequence of the exposure of hydrophobic patches and

this usually leads to uncontrolled assembly reactions and subsequent aggregation[48,49]. These processes are associated with a high number of human diseases such as amyotrophic lateral sclerosis, Alzheimer's or Parkinson's diseases where thermodynamic equilibrium may not be applicable. Accordingly, kinetic stability can be described in terms of rate for irreversible denaturation and it is often interpreted by a simple Lumry–Eyring model which describes well the denaturation of many proteins.

$$N \overset{K_U}{\leftrightarrow} U \overset{k_{IRR}}{\longrightarrow} F \tag{4}$$
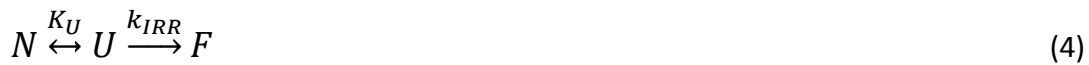
If transition state theory is applied, the rate for irreversible denaturation can be related to the free energy transition by the Eyring equation:

$$k_{IRR} = k_0 \cdot exp^{\left(-\Delta G^{\neq}/RT\right)} \tag{5}$$

where k_o is the preexponential factor (the hypothetical value of the rate constant when the activation free energy equals zero) and $\Delta G^{\neq}$ is the activation free energy[50]. At physiological temperature, if $\Delta G^{\neq}$ is high, the rate constant for irreversible denaturation will be low, and the protein will remain in the native state for long periods of time. This situation is referred as kinetically stable. Kinetic stability is very sensitive to mutational effects since small changes in $\Delta G^{\neq}$ can involve important changes in the time-scale in which the protein is functional and so, it is widely associated with diseases[2,47].

The relations between the thermodynamic and kinetic stabilities of a protein range from a perfect correlation to absolute independence between the two properties[47]. For instance, supposing a simple Lumry–Eyring scenario (4) where the irreversible step is very fast and unfolding step (N→U) is rate-limiting, the overall denaturation rate (k)

is determined by unfolding kinetics, K_U (1). In this situation changes on thermodynamic stability such as mutational effects, may or may not translate into kinetic stability including the extreme situation in which thermodynamic and kinetic stabilities can be completely decoupled[51]. Alternatively, if the U→F step is rate-limiting, the overall denaturation rate (k) will be the product of the unfolding equilibrium constant (1) times the rate constant for the U→F step, k_{IRR} (5) and under this situation the effects on thermodynamic stability will immediately affect kinetic stability.

1.1.2.2 PROTEIN HOMEOSTASIS NETWORK

As previously mentioned, between the *in vitro* and *in vivo* protein folding processes there are noticeable differences since many details of *in vivo* folding depend on the particular cellular environment, which makes it even more complex. The cellular environment is highly crowded (concentrations of macromolecules can reach 400 mg/mL) limiting the motions of polypeptide chains and favoring compact non-native states[42]. Furthermore, in eukaryotic cells, protein folding must occur in several distinct compartments with very different chemical composition, which can also result in different protein-folding problems that each cell must prevent and address[34]. Fortunately, the cells have a central network responsible for the generation and upkeep of the native state of the proteins and it is known as protein homeostasis network (or proteostasis network)[52]. The complex protein homeostasis network consists of numerous biological pathways which control protein synthesis, folding, trafficking and degradation and respond to specific forms to cellular stress[2,40,42,52]. The protein homeostasis network of mammalian cells involve approximately 1.300 different proteins implicated in protein synthesis (~400), conformational maintenance

(∼300), and degradation (∼700), although many of them take part in more than one pathway[42]. Of particular importance in this context are the previously introduced molecular chaperones[38,39,44–46].They are not only involved in the protein folding process of the proteins, but also in various aspects of proteome maintenance including macromolecular complex assembly, intracellular protein transport or assistance in protein degradation (**Figure 3**). Interestingly, there is a family of molecular chaperones that is key in the protein homeostasis network, the heat shock proteins (HSPs) family.

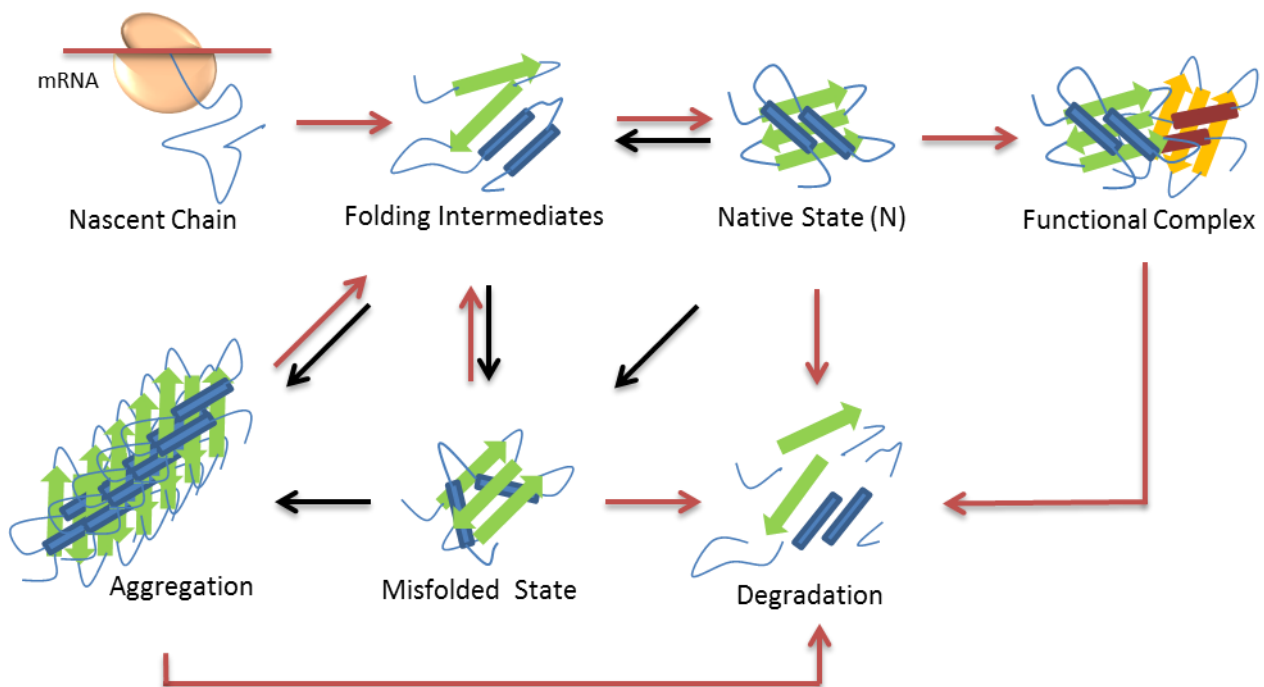**Figure 3.** Simple overview of different functions of molecular chaperones in protein homeostasis network. The red arrows show the steps where molecular chaperones take part.

These molecular chaperones, whose basic function is to selectively recognize non-native protein conformations, display a wide range of specificities for certain amino acid sequences and/or structural characteristics, although in most cases are

hydrophobic areas that are not exposed in the native protein conformation[53]. HSPs participate broadly in protein folding and refolding processes as well as show important implications in the processes of degradation.

In particular, a central feature in the arrangement of protein homeostasis network is the close interrelation between the molecular chaperones and different protein degradation pathways. The chaperones help to direct proteins whose functionality cannot be restored such as nonfunctional, misfolded or aggregated proteins toward degradation pathways and the balance between cellular protein folding and degradation is often referred to as protein quality control system (QCS)[40,54].

Proteins can be degraded either individually or on a large scale by several highly evolutionary conserved mechanisms[55,56]. Specifically, in the cytosol and nucleus of eukaryotic cells many degradation processes are carried out by the ubiquitin-dependent proteasome (26S proteasome). However, there are some proteins which are degraded by the proteasome without being ubiquitinated, although the mechanisms associated are well understood yet[57]. This type of degradation is performed by 20S proteasome and it relies on the inherent structural disorder of the protein degraded. Therefore, IDPs (or proteins with naturally disordered regions) as well as, proteins that contain unstructured regions due to oxidation, mutation or aging are susceptible to 20S degradation. The function of the 20S proteasome can be modulated by interactions with regulatory proteins. And interesting example of this regulation is given by NQO1, due to the fact that this protein physically binds to the 20S proteasome and protects proteins with unstructured regions (or IDPs) from degradation *(See section 2.2.4)*[57–59].

In addition, cells have an alternative strategy to maintain a balanced protein homeostasis when the mechanisms of refolding and proteosomal degradation are inhibited (or affected in another way). In this particular case, the misfolded protein can be sequestered as inclusion bodies which are degraded by selective autophagy, a degradation pathway interceded by lysosomes[55,56].

1.1.3 IMPACT OF MISSENSE POLYMORPHISMS AT MOLECULAR LEVEL

In this section we classify the different effects that a missense polymorphism can produce at molecular level in a protein (**Figure 4**), without forgetting that these effects can be interrelated [32,60].
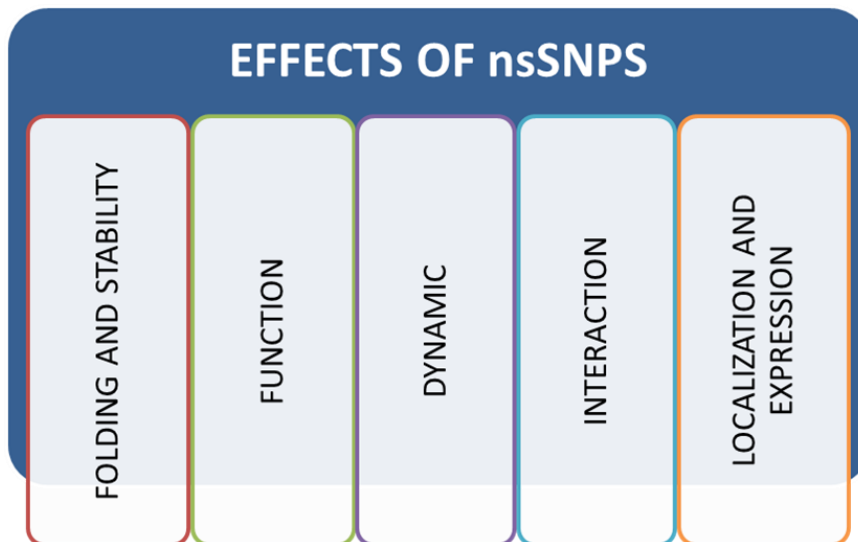

Figure 4.. Possible effects of missense polymorphisms at molecular levels in proteins.

1.1.3.1 EFFECTS ON PROTEIN FOLDING AND STABILITY

Multiple factors can alter the folding process resulting in destabilization of the correct fold and/or stabilization of a misfolded state. In particular, the missense polymorphisms can cause changes in many of the chemical and physical properties of a protein which can be associated with changes in its structure.

On the one hand, the missense polymorphisms can affect the protein structure by changes in thermodynamic stability and typically, the change in the folding free energy between wild type protein and the mutant protein ($\Delta\Delta G$) is used to quantify the magnitude of a mutation's effect on stability, since mutations can change both energy landscape and number of accessible conformations in both folded and unfolded states. Several comprehensive studies investigated the $\Delta\Delta G$ effects of missense mutations in different proteins showing that most mutations affect the protein stability in some way [29,61]. It has been observed that most globular proteins are marginally stable from a thermodynamic perspective, with a ΔG usually in the range of a −2 to −15 kcal/mol[2,62]. In order to put these values in context, the contribution of a single hydrogen bond to protein stability averages about 1 kcal/mol, showing that it can be easily affected[27]. Since, the presence of nsSNPs affect to the ΔG of both folded and unfolded states, they are usually classified as stabilizing or destabilizing depending on the final balance. Although the fraction of mutations classified like stabilizing or destabilizing varies according to the protein and the nature of these substitutions, suggesting that the distribution of $\Delta\Delta G$ effects might be unique for each protein, the most of missense polymorphisms that have been found associated with diseases are destabilizing and they lead to protein misfolding[61]. Protein misfolding can cause diseases known as

conformational diseases[39,64]. The degree of destabilization is usually elevated for mutations that introduce drastic changes of physicochemical properties of mutated site such as, breakage of salt bridges, alteration of interaction network or disruption of hydrogen bonds. Moreover, on average, mutations in core residues are much more destabilizing than mutations on the surface[61]. It has been shown that buried residues conserved throughout evolution are usually important parts of the hydrophobic core and any variation of theses might destabilize the fold. Nevertheless, the conserved residues that are exposed on the protein surface are usually involved in protein-protein interactions and the effects of their mutation will be seen later *(See section 1.1.3.4)*[46].

Special attention must be paid to classify the mutations in terms of their degree of stabilization/destabilization based on ΔΔG, since the protein kinetic stability must also be considered[2,47]. The missense polymorphisms can produce protein kinetic destabilization, even without altering the thermodynamic stability, which probably alters the interaction with protein homeostasis network producing effects on protein function[2,47].

1.1.3.2 EFFECTS ON PROTEIN FUNCTION

As indicated above, many pathological conditions are fundamentally rooted in protein misfolding. Given that protein function and structure are intrinsically linked, the most of mutations appear to indirectly affect activity via effects on protein structure or stability. Therefore, treating stability and activity as independently varying parameters is inherently confusing[65]. For instance, a missense polymorphism that decreases the protein stability, implying a reduction in the concentration of active protein, will cause

loss of function simply as a consequence of the latter[2]. In fact, the loss of activity be considerably greater if, for example, the destabilizing mutation also affected the specific activity of the folded protein[32,33,66]. Proteins perform different functions that are mediated by the various residues that comprise the functional sites and depending on the functional sites where the mutations occur there will be different consequences. For instance, some of the most dramatic effects occur in the actives sites of enzymes and the binding pockets of receptors[67]. In particular, the active sites are formed by a few crucial residues for the reaction (catalytic residues) with a larger number of surrounding residues which are important for ensuring correct attachment of the substrates and cofactors to the active site (binding site residues). Catalytic residues are highly conserved in evolution and it has been suggested that mutations in these residues are often very detrimental, while residues responsible for binding the substrate are not as vital to the catalytic function of the enzyme and can change through evolution, sometimes allowing the enzyme to accommodate new substrates[31,67,68]. Alternatively, the biochemical reaction is generally very sensitive to the accurate geometry of the active sites, so, any change that alters the active sites (observing or not a conformational change), for instance a mutation, will also affect the biochemical reaction, either completely abolishing the activity of the mutated protein or altering its kinetics[32,69].

Traditionally, the effects of disease mutations on protein function are categorized into two basic types[34,40,52]. On the one hand, effects related to *loss-of-function* of the protein. Numerous diseases are associated with loss-of-function phenotype, which is characterized by a decrease in the protein activity and can even cause a complete loss

of the activity[2,70]. These diseases are typically caused by inherited mutations which may lead to inefficient folding and excessive degradation, as it happens in cystic fibrosis, Gaucher disease and phenylketonuria or they may also produce an incorrect positioning of proteins in compartments where they are metabolically inefficient like in primary hyperoxaluria [2,40,52,71]. On the other hand, other effects imply mutations that confer altered functionality of a protein or increased protein activity and are associated with gain-of-function phenotype. Gain-of-function is sometimes linked with accumulation of toxic aggregates and diseases as known as amyotrophic lateral sclerosis, Parkinson, Alzheimer's and Huntington's diseases are related to this phenotype[34,40,52].

1.1.3.3 EFFECTS ON PROTEIN DYNAMICS

It is important to emphasize that proteins are dynamic molecules that require significant motion to carry out their different functions[60,65,72–74]. Some of the most remarkable examples of dynamics in function include protein folding and unfolding, ligand recognition and binding, protein-protein interaction and processes as complicated as transport through membranes, cell replication, transcription and translation, and assembly and disassembly of protein complexes[72]. Thus, **_protein dynamics_**, both local and global, play an important role in protein function and the proteins may be considered as networks of continuous motions that reflect local flexibility and a propensity for global structural plasticity[73]. Although sometimes the term dynamic and flexibility is used indiscriminately, some authors suggest the following difference between both terms: dynamics is used for intrinsic molecular motions, while flexibility is used to express the ability of a protein to adapt its structure

when it binds a ligand or in response to changes in the environment[75]. So, a protein can be flexible as a consequence of its dynamics, yet its dynamics does not automatically result in flexibility.

The synergy between structure and flexibility is critical to the protein functionality and the specific recognition of binding partners, such as small molecules, other proteins or nucleic acids, is essential for the many of the protein functions[72]. Any interaction between a protein and another molecule requires the protein to be able to adopt different conformations along the pathway of the biochemical reaction and the flexibility of the active site is considered as a requirement for the acceleration of the enzymatic reaction[73,76]. In particular, a protein must be stable enough to preserve it native three-dimensional structure, but flexible enough as to allow sufficient substrate binding, chemical reaction and product release. Often an increase of flexibility may lead to a concomitant rise of activity and loss of stability, leading to a balance between stability and function/flexibility[77].

The flexibility of a protein may result in either subtle changes, for instance when a few amino acid side chains of an enzyme move to bind a small substrate, or in more noticeable changes as when the folding of certain proteins (or protein domains) is promoted by the presence of the appropriate ligand[75]. Internal motions of a protein have different amplitudes and can be classified according to different timescales which span a large hierarchy of timescales, from picoseconds to milliseconds or even longer (**Figure 5**). The fastest events are small-amplitude motions (atoms moving only a
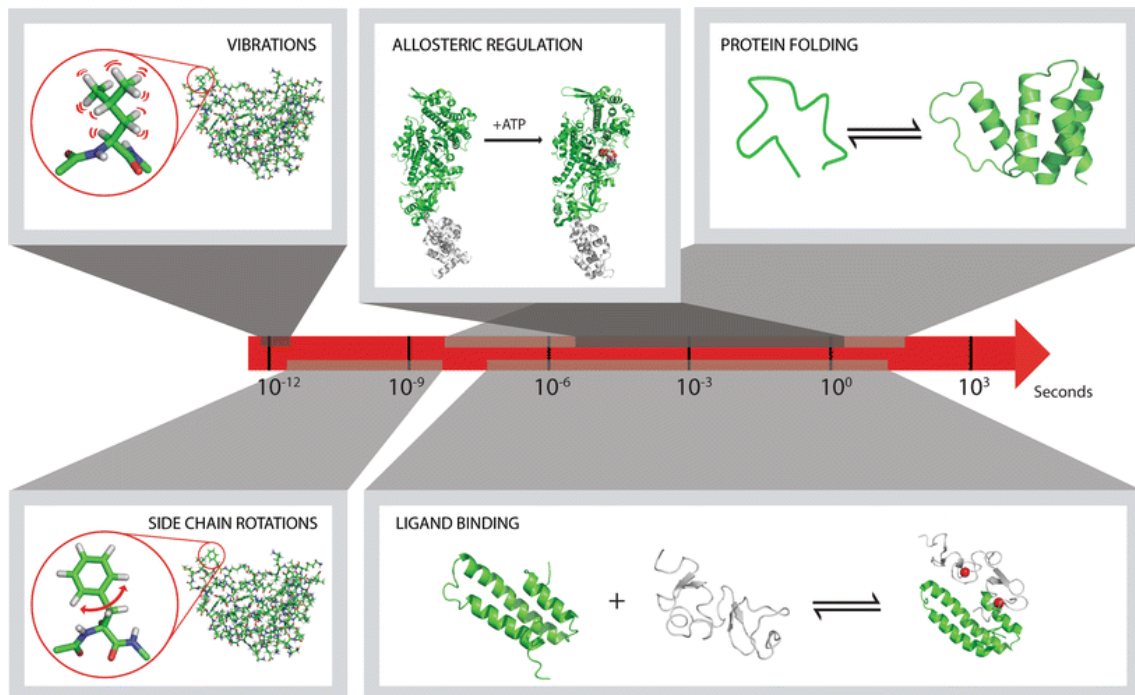
Figure 5. Schematic drawing of different protein dynamics process according to timescale in which they take place. . (Reprinted, with permission of Springer Nature , from Teilum K., 2009) [75].

fraction of an Ångström) which affect the protein structure from a more local standpoint. They include oscillations and vibrations of covalent bonds and fast side chain rotations on the picosecond to nanosecond timescale. At the other end of the timescale, large-amplitude motions can be found, such as ligand binding and protein folding/unfolding which may happen with time constants of hours[73,75,76].

Several studies have revealed changes of dynamics upon missense polymorphisms with important functional consequences[60,78,79]. In particular, the nsSNPs can affect the dynamics of the entire molecule or just a small region, and these changes in protein dynamics can affect the protein activity in several ways: firstly, they can inactivate or activate a protein, altering the concentration of active enzyme; secondly, they can affect the reaction parameters such as Michaelis-Menten constant, and finally they can

change the transition state of enzyme-substrate complex affecting parameters such as activation energy or pre-exponential factor[74].

A separate remark in protein dynamics should be done to allostery[60,78]. Allostery is a crucial process to living cells whereby a perturbation by an effector at one site of the protein leads to a functional change at another distant site. Allostery is a property of the conformational ensemble as a perturbation at any site in the structure leads to a shift in the distribution of the conformational states across the entire population. Therefore, allosteric structural and/or dynamic perturbations do not create new conformational states but rather, they change the relative distributions of the states within the ensemble[78].

Numerous researches have focused on the study of allosteric effects at the protein level, finding that an important condition for allostery seems to be the presence of a significant fraction of residues with low structural stability and high flexibility [75,80–84]. However, it is essential to study the allostery from a more global point of view to understand their full biological relevance. In other words, it is necessary to consider the allosteric effects not only in the protein, but in the pathway or pathways in which it participates and since these cellular pathways are interconnected, on the whole cell (i.e. a comprehensive view of the impact of an inactive or partially active protein can only be achieved by connecting molecular causes to system outcomes)[85,86]. Allostery plays a fundamental role in many of the processes in the living cell, which reveals its relevance in disease[78]. "Allosteric mutations" can cause disease through different mechanisms, for instance abolishing or creating sites for allosteric posttranslational modifications or changing the relative stabilities of the protein conformational states.

These mechanisms may produce protein deregulation, trapping the protein in an active or inactive conformation and producing diverse effects depending on how the function is associated with protein dynamics and the degree of change introduced to that functional protein motion[60,78].

1.1.3.4 EFFECTS ON PROTEIN INTERACTIONS

As previously commented, the proteins do not act in isolation in the cell but they are part of a global network of interactions with other proteins, nucleic acids, and other molecules, which have essential roles in regulatory processes, cellular functions, and signalling cascades[32,60,87–89]. Owing to their important roles in all biological processes, it is essential to understand how these interactions contribute to human disease. Clearly, a missense polymorphism which causes, for instance, protein aggregation, will lead to the loss of all the interactions of that protein, but also other more specific alterations of the macromolecular interactions can be related to missense polymorphisms.

There are two important factors in macromolecular interactions that can be affected by the presence of missense polymorphisms[90,91]. One of them is the specificity of the interaction, which plays a crucial role in recognizing unambiguous binding partners among a plenty of other molecules, and is in part determined by structural and physicochemical properties of interfaces of interaction[60,88]. Besides of specificity, the binding affinity is another important factor to consider in regards to protein interactions. The binding affinity between macromolecules to form a complex can be measured by the binding free energy ($\Delta G_{binding}$), which represents free-energy change of the complex-formation[60,88].

In particular, protein–protein interfaces (PPIs) are quite different from protein–ligand binding interfaces. Protein-ligand binding interfaces are often smaller and have a well-defined cavity for hosting small-molecule binders, whereas the PPIs are usually formed by two comparatively large surfaces with complementarity to each other, both in shape and in chemical nature (amino-acid side chains at the binding interface form preferentially stabilizing interactions of an electrostatic nature), but they are relatively flat [90,92,93]. Even then, high affinity and specificity can be achieved in PPIs since there are key residues that contribute a large amount to binding, and are called hot spots[32,66,87,93]. Hot spot is defined as a residue which replacement by an alanine implies a notable decrease in the free energy of binding ($\Delta\Delta G_{binding}$> 1.5 kcal/mol) and systematic analysis of hot spots has shown a nonrandom composition, preferring branched hydrophobic and aromatic amino acids such as Leu, Ile, Val, Phe, Tyr, Met, and Trp[90,94,95]. Because of the importance of these hot-spot residues to the interface, its mutation (i.e. changed by a charged or polar residue) can lead to loss or impairment of the interaction. In addition, there are also other ways in which an nsSNP would affect protein interactions for instance involving post-translational modifications, leading to novel interactions (i.e. in amyloidosis) and/or switch of function, or causing intrinsic disorders[87].

1.1.3.5 EFFECTS ON PROTEIN EXPRESSION AND LOCALIZATION

Cellular functions depend largely on the content and functionality of the proteins within the cells. There are numerous mechanisms which regulate the cellular functions varying from modulating gene expression to post-translational modifications[96,97]. In particular, gene expression in eukaryotes involves a wide range of mechanisms from

RNA transcription to subcellular localization, including processing of mRNA, translation into protein for protein-encoding genes, protein folding, and post-translational modification[96]. Gene expression is controlled at all steps in the process although much of the regulation is achieved at the level of transcription initiation when promoter and other specific regulatory sequences of genes are bound by transcriptional regulatory proteins. The amount of protein expressed is determined not only by the mentioned processes of the cell but also, in some cases, by inherited genetic determinants. Specifically, numerous SNPs in gene regulatory regions have been associated with variation in enzyme levels and human diseases[32,96].

Alternatively, an important mechanism, by which efficient functional regulation is achieved, is the control of protein localization. Many proteins have functions that are dependent on specific subcellular environments or they may perform the same functions in multiple subcellular compartments[32,97]. Transporting a protein to the correct compartment allows it to form the necessary interactions with its biological partners and take part in the corresponding biological networks. This transport can be mediated by signal peptides, which can also be masked or modified by carrier proteins that recognize a particular pattern of post-translational modifications, being part of an important regulatory mechanism[98]. The presence of missense polymorphisms in a signal peptide can impair its function and thus, direct the protein to a wrong subcellular compartment, which will also have deleterious effects on the other proteins that function there[32,98]

1. INTRODUCTION

1.2 NAD(P)H:quinone oxidoreductase I

NAD(P)H: quinone oxidoreductase-1 (NQO1; EC 1.6.5.2), also known as DT Diaphorase, was detected accidentally in 1958 by Ernster and Navazio in the soluble fraction of rat liver homogenates, during studies of NAD and NADP dehydrogenases[99]. They named it DT Diaphorase because of its ability to use both NADH and NADPH as cofactors in its obligate two-electron reduction of quinones substrates[100].

1.2.1 LOCATION, STRUCTURE AND REACTION MECHANISM

The human NQO1 (NQO1) enzyme is encoded by the *NQO1* gene (also referred as DIA4), which is maps to chromosomal location 16q22.1[101,102]. NQO1 is a flavoenzyme broadly distributed in human tissues, with highest levels in epithelial and endothelial cells, especially of the kidney and gastrointestinal tract[103]. In addition, this enzyme is highly expressed in some cancer cells[102–105]. Intracellularly, NQO1 is in essence a cytosolic enzyme, although it has also been localized in smaller amounts in mitochondria, endoplasmic reticulum and nucleus[101–103].

Structural studies have shown that NQO1 is a homodimeric enzyme with 273 residues per monomer, each with an approximated molecular mass of 31 KDa[105]. Each subunit

contains one molecule of flavin adenine dinucleotide (FAD), which is tightly bound. Two domains form each monomer, a N-terminal domain (residues 1-220), which is a large catalytic domain, and a C-terminal domain (residues 221-273), which is involved in dimerization and forms part of the binding site for the adenosine portion of the NAD[106]. The catalytic domain comprises a five-stranded parallel β-sheet in the middle of five α-helix, whereas the smaller C-terminal domain is composed of an antiparallel hairpin motif, an α-helix and several loops[107,108]. There are two functional active sites per dimer, each located at the interface between the two subunits. FAD molecules, attached non-covalently, form one of walls of each active site while the rest of the boundaries are formed by residues from both monomers (**Figure 6**)[105,107].
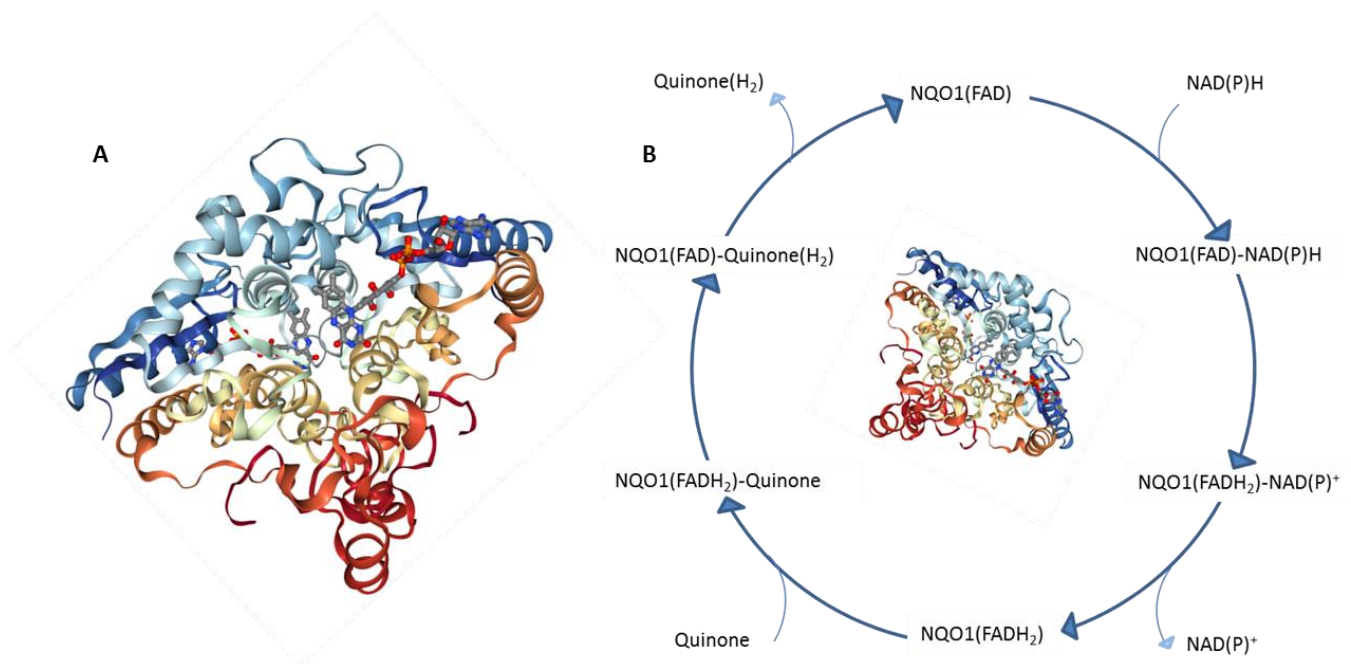

Figure 6.A) Crystal structure of native human NQO1 at 1.7 A resolution (PDB accession code: 1D4A) The non-covalently bound FAD is shown at both sites[107]. **B)** Ping-pong mechanism of NQO1.

The active sites are essentially large hydrophobic pockets with three defined binding spaces, one for the adenine and ribose parts of NAD(P)H, another for FAD molecule and one last for nicotinamide moiety of NAD(P)H or the electron acceptor molecule[105]. NQO1 active sites have the ability to accommodate a broad range of substrates such as, quinones, vitamin K1 and dichloroindophenol (DCPIP)[107].

The reaction proceeds via a substituted enzyme mechanism, also called "ping-pong", and the two-electron reduction occurs in two sequential steps (**Figure 6**)[100]. Firstly, NAD(P)H reduces the FAD cofactor bound in the active site of NQO1, forming NAD(P)$^+$ and FADH$_2$. Subsequently, NAD(P)$^+$ is released and the enzyme can consequently accept the second substrate, often a quinone, which is reduced by FADH$_2$ and forming again FAD. This regenerates the initial form of the enzyme, enabling a further round of catalysis. This reaction is strongly and competitively inhibited, with respect to NAD(P)H, by dicoumarol, a hydroxycoumarin with strong anticoagulant activity, which binds in the active site, presumably overlapping with the NAD(P)H binding site[100,105,106].

1.2.2 FUNCTIONS OF NQO1

NQO1 is essentially known for its antioxidant and detoxifying capacities, although it is a versatile protein showing multiple roles within the cells[105,109]. Some of the potential functions that have been described for NQO1 are summarized in **Figure 7**.
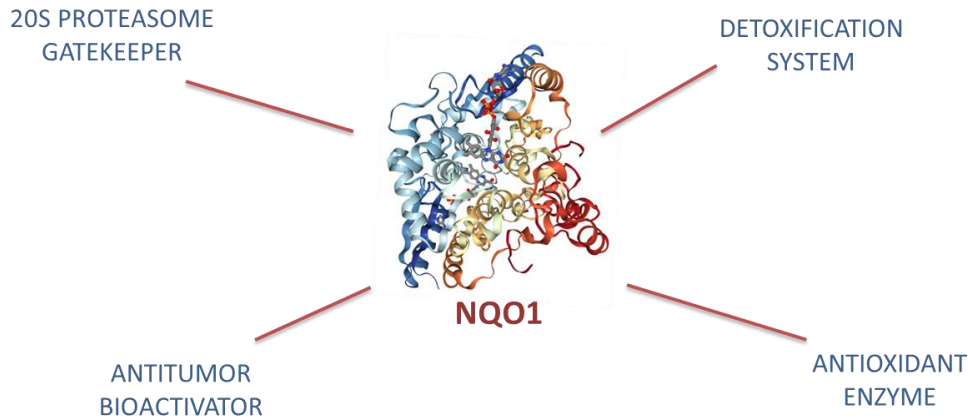
Figure 7. Schematic summarization of NQO1 functions.

1.2.2.1 NQO1 AS DETOXIFICATION SYSTEM

A considerable body of literature exists supporting the role of NQO1 as a detoxification system and this NQO1 role is attached to its catalytic mechanism: the obligatory two-electron reductions[103,110–113]. NQO1 prevents the formation of damaging semiquinones by obligate two-electron reductions of exogenous and endogenous quinones, quinoneimines, nitroaromatic compounds and azo dyes, to stable hydroquinones[105,108,114]. Semiquinones are very unstable substances which quickly react with molecular oxygen at physiological pH to form superoxide free radicals[111]. In addition, NQO1 acts as a Phase II detoxification enzyme, since the hydroquinone products of the NQO1 reaction are metabolites more water soluble and therefore, more easily excreted[102,112]. These reductions lessen quinone levels, minimizing the opportunities to participate in reactions that could lead to sulfhydryl depletion[108,113]. A separate remark should be made for the NQO1's ability in the protection against

benzene toxicity. Humans are exposed to benzene via contact with cigarette smoke, gasoline emissions, or products of incomplete combustion and NQO1 is capable of reducing benzoquinones to hydroquinone and catechol, resulting in detoxification[113,115,116].

1.2.2.2 NQO1 AS AN ANTIOXIDANT ENZYME

On the other hand, numerous reports have suggested that NQO1 may play an antioxidant role, possibly by the reduction of certain endogenous quinones[102,105,109,113]. NQO1 helps maintain some endogenous antioxidants in their reduced and active forms, and thus, NQO1 may help protect cellular membranes against oxidative damage[102]. In particular, NQO1 can reduce ubiquinone and vitamin E quinone devoided of antioxidant potential, to ubiquinol and Vitamin E hydroquinone respectively, both compounds with reported antioxidant capacities[109]. In addition, NQO1 can directly scavenge free superoxide by reaction with its cofactor, producing hydrogen peroxide as end product of the reaction[117]. Although NQO1 scavenges superoxide less efficiently than superoxide dismutase (SOD), this NQO1 property would be especially important in tissues with low SOD expression such as cardiovascular cells[117,118].

1.2.2.3 NQO1 AS ACTIVATOR OF CANCER PRODRUGS

This function of NQO1 is of special interest since, as previously mentioned, NQO1 is over-expressed in many cancerous tissues, compared to normal tissues. Bioactivation of antitumor quinones by NQO1 have been widely demonstrated, and different types of quinones such as, mitomycin C (MMC), indoloquinone E09, and

aziridinylbenzoquinones (AZQ) among others, have been characterized as quinones bioactivated by this enzyme[119–121].

1.2.2.4 NQO1 AS A GATEKEEPER OF 20S PROTEASOME

In addition, other protective roles can be associated with NQO1 beyond its enzymatic activity. In particular, an important role of NQO1 is its ability to physically interact with other proteins and there are numerous examples where such interactions are relevant for cellular function. NQO1 binds physically to 20S proteasome and protects some IDPs from degradation such as, p33, p53, p63, p73, c-Fos, C/EBP, PGC-1 or Ornithine decarboxylase (ODC). It seems that the regulatory role of NQO1 is conserved throughout evolution[57,58,122]. Apart from its association with the proteasome, NQO1 has been shown in several instances to bind the IDP substrates in an NADH-dependent manner, even competing with the 20S proteasome for substrate interaction, although the exact molecular mechanism underlying NQO1's ability to inhibit substrate degradation remains uncertain[90–93]. It should be highlighted the NQO1 functions as p53 protein stabilizer because tumor suppressor p53 plays an important role in protection of genome against internal and external stresses preventing the cell transformation and tumor formation. Under normal conditions, p53 is a short-lived protein that is regulated mainly by changes in its stability. However, in response to a range of stress signals, p53 is stabilized and activated, leading to cell cycle arrest, senescence, or apoptosis[123]. Furthermore, it is also important to mention the interplay between NQO1 and the 20S proteasome since, on the one hand NQO1 can prevent the proteolytic activity of the proteasome, while on the other hand, the proteasome regulates NQO1 levels since when NQO1 is in apo state (NQO1 in absence of FAD

binding) the protein is largely unstructured and it becomes a substrate of the 20S proteasome[57,58].

1.2.3 POLYMORPHISMS IN NQO1

There are more than 250 SNPs in the NQO1 gene according to the dbSNP database (http://www.ncbi.nlm.nih.gov/SNP), including the two most prevalent variants in the human population, NQO1*2 and NQO1*3. NQO1*2 is the polymorphism most frequent and it is a missense polymorphism which replaces the cytosine at position 609 of the cDNA for a thymine (c.C609T, rs1800566). This change codes for a proline to serine change in the residue 187 of the human protein (P187S). The NQO1*3 is less frequent and it represents a cytosine for thymine change at position 465 of the NQO1 cDNA (c.C465T, rs1131341). This change results in an arginine-to-tryptophan amino acid substitution at residue 139 (R139W) in the protein[112,113,115]. In this thesis NQO1*2 will be referred to as P187S and NQO1*3 as R139W.

1.2.3.1 THE P187S POLYMORPHISM

The majority of investigation on NQO1 polymorphisms have been focused on P187S now that this polymorphism is the most noticeable, both in terms of frequency and phenotypic consequences[113]. The P187S polymorphism has been found in wide variety of cell lines, many of them tumour cell lines commonly used in research. Its global allelic frequency in the human population is estimated to be 0.29, being as high as 0.5 in some Asian populations (http://www.ensembl.org/index.html). There is a clear relation between this polymorphism and increased cancer risk and there is a wide

variety of cancer types associated with P187S such as, gastrointestinal, liver, colorectal, lung, breast and thyroid among others[3,124–127]. In addition, P187S has also been related to increased susceptibility to poisoning by benzene and other carcinogenic chemicals[115,116,128]. Other functions of NQO1 are also affected by the presence of this polymorphism, for example, showing reduced efficiency as activator of cancer prodrugs, or affecting multiple protective roles associated with interaction of NQO1 with other proteins[129,130].

P187S is a typical example of mutation associated with loss-of-function and intracellular instability of the enzyme. This polymorphism was reported in 1980 by Edwards et al.,but it was not characterized for the first time until 1992[130,131]. Currently, it is well-known that P187S shows a reduction or even a total loss of the enzymatic activity in the homozygous mutant. It seems to be related to a strong decrease, 10-100 fold, of the FAD binding affinity[4,5,132]. In addition, this polymorphism gives rise to reduced intracellular stability of the protein undergoing a fast degradation by 20S/26S proteasome[6,7]. X-ray crystallographic analysis of P187S showed a structure virtually identical to the wild-type NQO1 structure, although Nuclear Magnetic Resonance (NMR) spectroscopy and proteolysis experiments confirmed the presence of partially unfolded states in solution and a particularly high flexibility of its C-terminus[5]. However, these results cannot totally explain the destructive effects of P187S *in vitro* and *in vivo*, and protein dynamics analysis in depth should be carry out (*as we do in Publication 1 and Publication 2*).

1.2.3.2 THE R139W POLYMORPHISM

This polymorphism was characterized by Pan and colleagues about 1995 while they were researching the MMC resistance of several types of cells lines[133,134]. R139W SNP is also associated with increased cancer risk, although it is less well studied than P187S SNP. It has an estimated global allele frequency of 0.02, reaching to 0.07 in Spanish Iberian population (http://www.ensembl.org)[4]. R139W SNP results of an enhanced alternative splicing events that can lead to a truncated mRNA with a deletion of exon 4, and therefore, to decrease the expression of the active full-length enzyme. Skipping of exon 4 quantitatively varies among cancer sublines and it removes residues 102–139 critical for quinone-binding site of NQO1 which carries out a very unstable and inactive protein[105,134,135]. In addition, R139W has been associated with reduced sensitivity towards some cancer prodrugs, such as MMC, in some cell lines[133,134].Even though R139W is traditionally related to childhood acute lymphoblastic leukaemia and studies to characterize this polymorphism have been performed, its involvement in disease development was not well determined so far=.

2. OBJECTIVES

The main objectives of this doctoral thesis are:

1. Analysing the relation between protein dynamics, stability and functionality of two NQO1 polymorphisms and understanding their loss-of-function mechanisms.

2. Determining the effects of P187S polymorphism on the C-terminal domain of NQO1 and its link with the deleterious effects associated with this polymorphism.

3. Studying the effects of P187S polymorphism on N-terminal domain, in particular on FAD binding site, and its rescue by suppressor mutations.

2. OBJETIVOS

Los objetivos principales de esta tesis doctoral son:

1. Analizar la relación entre la dinámica, la estabilidad y la funcionalidad de dos polimorfismos de NQO1 y entender sus mecanismos de pérdida de función.

2. Determinar los efectos del polimorfismo P187S en el dominio C-terminal de NQO1 y su relación con los efectos deletéreos asociados con este polimorfismo.

3. Estudiar los efectos del polimorfismo P187S en el dominio N-terminal, en particular en el sitio de unión a FAD, y su rescate mediante mutaciones supresoras.

3.   RESULTS

This chapter presents the results obtained throughout the development of this doctoral thesis. They are classified in different sections according to the proposed objectives.

OBJECTIVE 1

• PUBLICATION 1

- **E. Medina-Carmona**, R.J. Palomino-Morales, J.E. Fuchs, P.G. Esperanza, M.T. Noel, E. Salido, D.J. Timson, A.L. Pey, *Conformational dynamics is key to understanding loss-of-function of NQO1 cancer-associated polymorphisms and its correction by pharmacological ligands*, Sci. Rep. 6 (2016) 1–13. doi:10.1038/srep20331.
- JCR Impact Factor (2016): **4.259**
- Category name /Quartile in Category: **Multidisciplinary Sciences / Q1**

OBJECTIVE 2

• PUBLICATION 2

- **E. Medina-Carmona**, J.L. Neira, E. Salido, J.E. Fuchs, R. Palomino-Morales, D.J. Timson, A.L. Pey, *Site-to-site interdomain communication may mediate different loss-of-function mechanisms in a cancer-associated NQO1 polymorphism*, Sci. Rep. 7 (2017) 1–18. doi:10.1038/srep44532.
- JCR Impact Factor (2016): **4.259**
- Category name / Quartile in Category: **Multidisciplinary Sciences / Q1**

OBJECTIVE 3

• PUBLICATION 3

- **E. Medina-Carmona**, J.E. Fuchs, J.A. Gavira, N. Mesa-Torres, J.L. Neira, E. Salido, R. Palomino-Morales, M. Burgos, D.J. Timson, A.L. Pey, *Enhanced vulnerability of human proteins towards disease-associated inactivation through divergent evolution,* Hum. Mol. Genet. 26 (2017) 3531–3544. doi:10.1093/hmg/ddx238.
- JCR Impact Factor (2016): **5.340**
- Category Name / Quartile in Category: **Genetics & Heredity / Q1**

• PUBLICATION 4

- I.G. Muñoz, B. Morel, **E. Medina-Carmona**, A.L. Pey, *A mechanism for cancer-associated inactivation of NQO1 due to P187S and its reactivation by the consensus mutation H80R*, FEBS Lett. 591 (2017) 2826–2835. doi:10.1002/1873-3468.12772.
- JCR Impact Factor (2016): **3.623**
- Category Name / Quartile in Category: **Biophysics / Q1**

PUBLICATION 1

Conformational dynamics is key to understanding loss-of-function of NQO1 cancer-associated polymorphisms and its correction by pharmacological ligands.

SCIENTIFIC REPORTS

OPEN

Conformational dynamics is key to understanding loss-of-function of NQO1 cancer-associated polymorphisms and its correction by pharmacological ligands

Medina-Carmona Encarnación[1,*], Rogelio J. Palomino-Morales[2,*], Julian E. Fuchs[3,*,†], Padín-Gonzalez Esperanza[1], Mesa-Torres Noel[1], Eduardo Salido[4], David J. Timson[5,6] & Angel L. Pey[1]

Protein dynamics is essential to understand protein function and stability, even though is rarely investigated as the origin of loss-of-function due to genetic variations. Here, we use biochemical, biophysical, cell and computational biology tools to study two loss-of-function and cancer-associated polymorphisms (p.R139W and p.P187S) in human NAD(P)H quinone oxidoreductase 1 (NQO1), a FAD-dependent enzyme which activates cancer pro-drugs and stabilizes several oncosuppressors. We show that p.P187S strongly destabilizes the NQO1 dimer *in vitro* and increases the flexibility of the C-terminal domain, while a combination of FAD and the inhibitor dicoumarol overcome these alterations. Additionally, changes in global stability due to polymorphisms and ligand binding are linked to the dynamics of the dimer interface, whereas the low activity and affinity for FAD in p.P187S is caused by increased fluctuations at the FAD binding site. Importantly, NQO1 steady-state protein levels in cell cultures correlate primarily with the dynamics of the C-terminal domain, supporting a directional preference in NQO1 proteasomal degradation and the use of ligands binding to this domain to stabilize p.P187S *in vivo*. In conclusion, protein dynamics are fundamental to understanding loss-of-function in p.P187S, and to develop new pharmacological therapies to rescue this function.

The study of NAD(P)H quinone oxidoreductase 1 (NQO1, EC 1.6.5.2) polymorphisms is particularly interesting due to its enhanced expression in several types of cancer[1–3]. NQO1 is a dimeric, two-domain FAD-dependent enzyme (Fig. 1A) which catalyses the two electron reduction of quinones and related substrates through an enzyme-substituted mechanism in which NAD(P)H enters the active site, reduces the FAD and exits as the oxidised form, allowing the subsequent substrate binding and reduction by the $FADH_2$[4]. Primarily, NQO1 avoids the formation of reactive semiquinones, maintains antioxidants such as α-tocopherol and ubiquinone in their reduced state, and also activates some anticancer bioreductive drugs (e.g. mitomycin C (MMC) and a MMC analogue, EO9)[4,5]. Additionally, NQO1 interacts with tumour suppressors such as p53 and p73 and stabilizes them towards proteasomal degradation[6–9], while its interaction with the 20S proteasome prevents the degradation of a plethora of proteins with intrinsically disordered regions, including several cell cycle regulators, tumor suppressors and apoptotic proteins[10].

Two single nucleotide polymorphisms in NQO1 were originally isolated from cancer cell lines, namely rs1800566/c.C609T/p.P187S and rs1131341/c.C465T/p.R139W[11–13]. The frequency of the p.P187S allele is much

[1]Department of Physical Chemistry, Faculty of Sciences, University of Granada, Granada, Spain. [2]Department of Biochemistry and Molecular Biology I, Faculty of Sciences, University of Granada, Granada, Spain. [3]Centre for Molecular Informatics, Department of Chemistry, University of Cambridge, Cambridge, UK. [4]Hospital Universitario de Canarias, Tenerife, Spain. [5]School of Biological Sciences, Queen's University Belfast, Belfast, UK. [6]School of Pharmacy and Biomolecular Sciences, The University of Brighton, Brighton, UK. *These authors contributed equally to this work. †Present address: Institute of General, Inorganic and Theoretical Chemistry, Faculty of Chemistry and Pharmacy, University of Innsbruck, Innsbruck, Austria. Correspondence and requests for materials should be addressed to A.L.P. (email: angelpey@ugr.es)
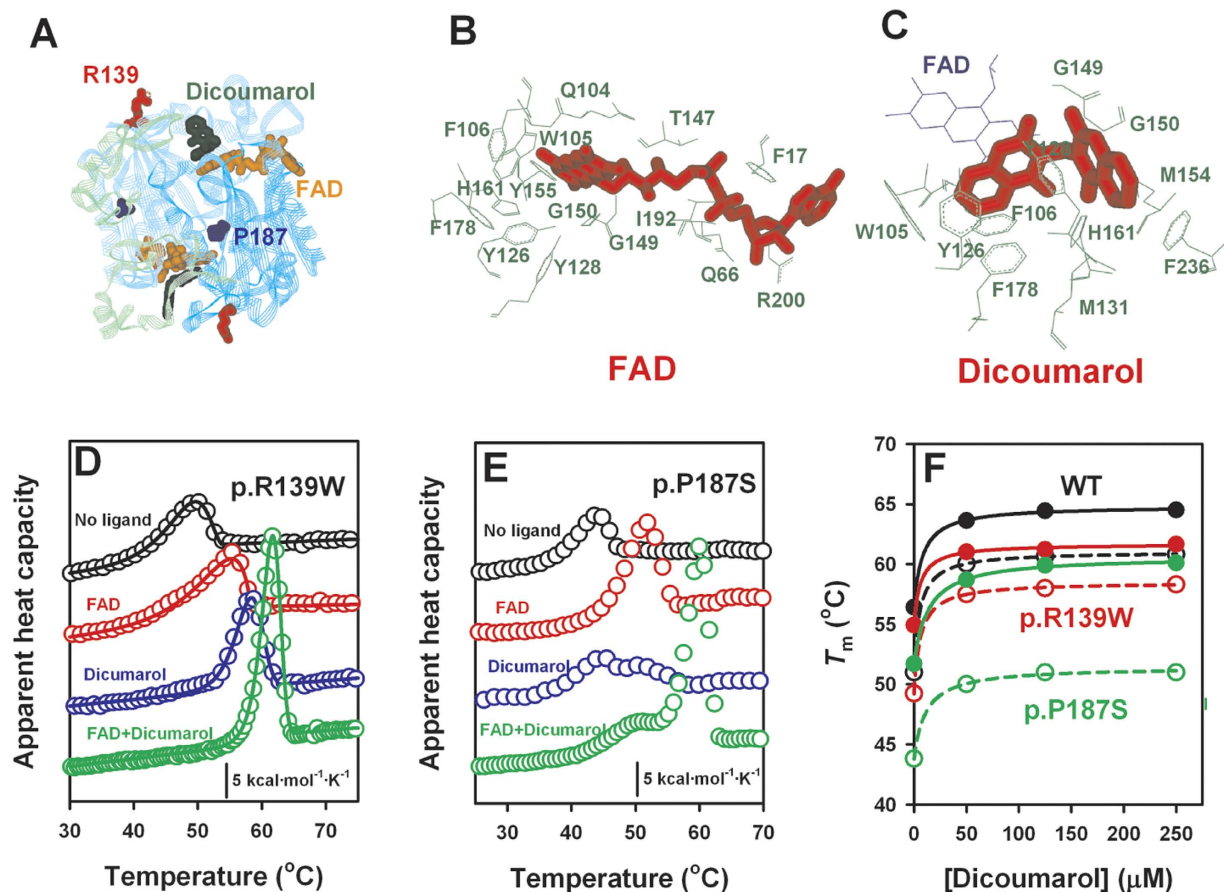
Figure 1. Dicoumarol and FAD mediated stabilization of NQO1 enzymes. (A) Structural location of Arg139 (red) and Pro187 (blue) and the ligands FAD (orange) and dicoumarol (dark green) based on the ternary complex determined by X-ray diffraction (PDB: 2F1O[35]). The N-terminal domain (residues 2–217) and C-terminal domain (residues 218–274) are displayed in cyan and light green, respectively; **(B,C)** Binding sites of FAD (B; PDB: 1D4A[36]) and dicoumarol (C, PDB: 2F1O[35]), as determined by X-ray diffraction. **(D,E)** DSC profiles of R139W and P187S enzymes in the absence and presence of $100\,\mu M$ FAD and $250\,\mu M$ dicoumarol. Lines in panel **(D)** are fits to a two-state kinetic model. **(F)** Dicoumarol concentration dependence of T_m values for NQO1 enzymes in the absence (open symbols) and presence (closed symbols) of $100\,\mu M$ FAD. For WT and p.R139W, T_m values are derived from fits to a two-state kinetic model, while for p.P187S correspond to the maximum of the heat capacity values for the high-temperature transition.

higher than of p.R139W in human population (about 30% vs. 1–2%; www.ensemble.org;)[14]. A recent comprehensive meta-analysis using over 21000 cases and 25000 controls have supported an association of the p.P187S polymorphism in homozygosis and an overall increased cancer risk[15]. Although p.R139W is rarely found in homozygosis, this polymorphism in heterozygosis has been associated with increased risk of developing acute lymphoblastic leukemia in children[16]. To our knowledge, no exhaustive association studies of these polymorphisms and the response to different chemotherapies have been described so far.

p.P187S is known to strongly reduce cellular NQO1 protein levels and activity[13,17], thus diminishing its efficiency in the reductive activation of cancer drugs such as MMC. The intracellular instability of p.P187S is caused by enhanced fast degradation by the 20S/26S proteasome[18,19] while its low activity is in part associated to its low affinity for FAD[18,20,21]. The p.R139W polymorphism is less well studied, showing WT activity levels towards using dichlorophenol indophenol (DCPIP) as substrate[20] but reduced activity towards MMC[11]. This polymorphism is thought to primarily reduce the sensitivity towards MMC in cell lines due to alternative splicing causing skipping of exon 4, and therefore, decreasing the expression of the active full-length enzyme[11,22]. Even though p.P187S does not affect the overall conformation of NQO1[20], a higher population of apo-NQO1 due to its defect in FAD binding may shift the folding equilibrium towards partially denatured states, which are much more sensitive to proteasomal degradation[18,20,21]. Therefore, boosting intracellular levels of FAD may protect p.P187S towards degradation[18,20,21]. Remarkably, X-ray crystallographic analysis of p.P187S has revealed no significant structural changes, while nuclear magnetic resonance (NMR) spectroscopy and proteolysis experiments supported partial unfolding and enhanced flexibility of p.P187S in solution[21].

Protein dynamics is a key factor in many aspects of protein function, such as enzyme catalysis[23], allosteric regulation[24] and protein degradation[25]. In the case of NQO1, polymorphisms may affect local dynamics therefore

NQO1 (Ligand)	K_{d1}	K_{d2}
WT (FAD)	~1 nM	60 nM
p.R139W (FAD)	~3 nM	70 nM
p.P187S (FAD)	400 nM	*N.app.*
WT (Dicoumarol)	122 ± 37 nM	*N.app.*
p.R139W (Dicoumarol)	56 ± 18 nM	*N.app.*
p.P187S (Dicoumarol)	11 ± 3 μM	*N.app.*

Table 1. Binding affinity of NQO1 variants for FAD and dicoumarol determined by ITC. Dissociation constants correspond to fittings to a two-sequential binding sites or to a single-set of independent binding sites (*N. app.* indicates a single set of sites model). FAD binding constants are the average of two independent titrations (from[20]) and dicoumarol binding constants are best-fits values ± fitting errors from a single titration (Fig. S1).

leading to changes in protein stability and functionality, and also, natural ligands may overcome these dynamic effects and be used to rescue NQO1 function. Thus, we explore here the relation between protein dynamics, stability and functionality of NQO1 polymorphisms using a combination of experimental and computational procedures, including biophysical binding and stability studies, analyses of the local flexibility by proteolysis and mass spectrometry, molecular dynamic simulations and expression studies in cultured cells. The results provided here are critical to understanding NQO1 loss-of-function in cancer patients and for the discovery of new ligands aimed at rescuing NQO1 function by shielding the polymorphic variants from proteasomal degradation.

Results

FAD and dicoumarol additively stabilize NQO1 enzymes towards thermal denaturation.
Arg139 and Pro187 are located in solvent exposed loops within the N-terminal catalytic domain (Fig. 1A). These two residues are not directly involved in the binding sites of FAD and/or NADH/dicoumarol (Fig. 1A–C), even though p.P187S strongly decreases NQO1 activity (k_{cat} of 0.5–1 s^{-1} for p.P187S *vs.* 100–150 s^{-1} for WT, in three different protein batches) and affinity for FAD[13,17,20,21] (see Table 1). Nevertheless, addition of FAD causes a strong kinetic and thermal stabilization of all NQO1 enzymes[18,20] (Fig. 1D–F). The p.P187S polymorphism also decreases the binding affinity for dicoumarol by 100-fold compared to WT and p.R139W (Table 1 and Fig. S1). As also seen with FAD, dicoumarol is able to enhance the thermal and kinetic stability of all NQO1 enzymes, and its stabilizing effect is additive to that of FAD (Fig. 1D–F). Therefore, two ligands interacting with different structural regions of NQO1 (FAD primarily with the N-terminal domain, while dicoumarol interacts with residues at both N-terminal and C-terminal domains, Fig. 1B,C) may enhance the global stability of NQO1 polymorphisms *in vitro*.

Proteolysis supports local changes in NQO1 protein dynamics by cancer-associated polymorphisms and their modulation upon ligand binding.
To investigate the potential effects of NQO1 polymorphisms on protein dynamics, possibly leading to reduced conformational stability and impaired catalytic function, we have performed proteolysis studies with thermolysin. First, we have determined the sensitivity of NQO1 variants towards degradation using different concentrations of protease and the effect of added FAD (Fig. S2), since as purified, WT and p.R139W contain significant amounts of FAD bound, while P187S is essentially an apo-protein[18,20]. As purified, WT and p.R139W show similar resistance towards proteolysis by thermolysin, while p.P187S is about 100-fold more sensitive. Addition of FAD has little or no effect in the sensitivity of native NQO1 enzymes towards degradation, even though some changes in the proteolysis patterns are observed (Fig. S2). These results agree well with those of a recent study performed with WT and p.P187S using trypsin[21]. Addition of dicoumarol has little effect on the extreme sensitivity of p.P187S towards degradation, while simultaneous addition of FAD and dicoumarol strongly protects p.P187S (Fig. S2).

To characterize the proteolysis patterns of NQO1 enzymes in the absence and the presence of FAD and dicoumarol, we have performed kinetic analyses by SDS-PAGE in combination with HPLC/ESI-MS and N-terminal sequencing of selected proteolysis products. WT and p.R139W are degraded at similar rates, although degradation of p.R139W is slightly faster, and the proteolysis rates of these two variants are not affected by the addition of FAD (Figs 2A,B and 3A). For these two variants, we observed the accumulation of two partially proteolyzed forms: i) one with a molecular mass of ~31.5 kDa (Table S1; note that our full-length construct is ~32.6 kDa) and a N-terminal sequence (Seq A, site WT-1; Fig. 3B) consistent with the cleavage in the far N-terminal sequence and the release of the his-tag (forming a species with a theoretical mass of 31555 Da). Therefore, this cleavage event essentially leads to the formation of a native full-length NQO1 enzyme; ii) a second form with a size of ~22.8 kDa (from HPLC/ESI-MS analyses, see Table S1) and the N-terminal sequence (Seq B, site WT-2; Fig. 3B), consistent with cleavage between Ser72-Val73 and that would generate a form with a theoretical mass of 22825 Da (in agreement with its experimentally determined mass; Table S1). Therefore, the main proteolysis product of WT and p.R139W by thermolysin is formed upon cleavage at the N-terminal region of NQO1 (named $I_{22.8}$) and contains residues 73–274. This cleavage site differs from those for trypsin and chymotrypsin previously reported in WT NQO1 (occurring at residues 230–240[26]), probably due to different protease specificities.

Proteolysis of p.P187S by thermolysin is very fast and shows a different pattern, with the accumulation of two partially proteolyzed forms (Fig. 2A): a transiently populated form with a mass of 29.9 kDa that leads to the formation of a 28.2 kDa product that is strongly stabilized by FAD (Fig. 2B). This proteolysis pattern also resembles that of p.P187S using trypsin[21]. The 28.2 kDa product has an N-terminal sequence (Seq C in Fig. 3B)
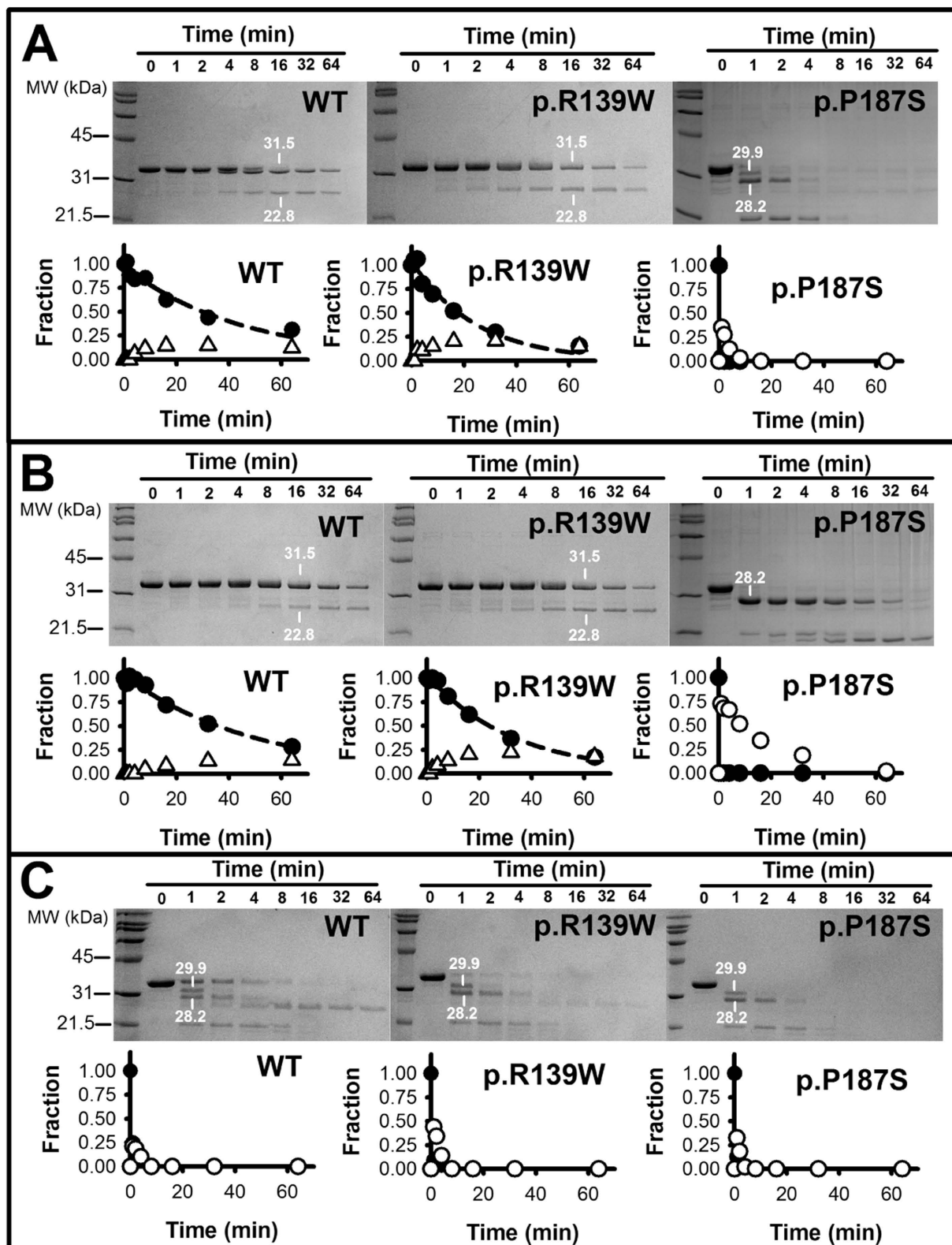
Figure 2. Proteolysis of NQO1 enzymes by thermolysin. Different panels display different experimental conditions: panel (**A**) show data for NQO1 enzymes as purified; panel (**B**) shows NQO1 enzymes in the presence of $100\,\mu M$ FAD; panel (**C**), apo-NQO1 enzymes. In each panel, SDS-PAGE gels are shown at the upper part, while the time-dependence of native bands (closed circles) and selected intermediates (open triangles for $I_{22.8}$ and open circles for $I_{28.2}$) are shown. Samples identified by HPLC/ESI-MS are indicated in the gels together with their molecular weight determined by this technique.
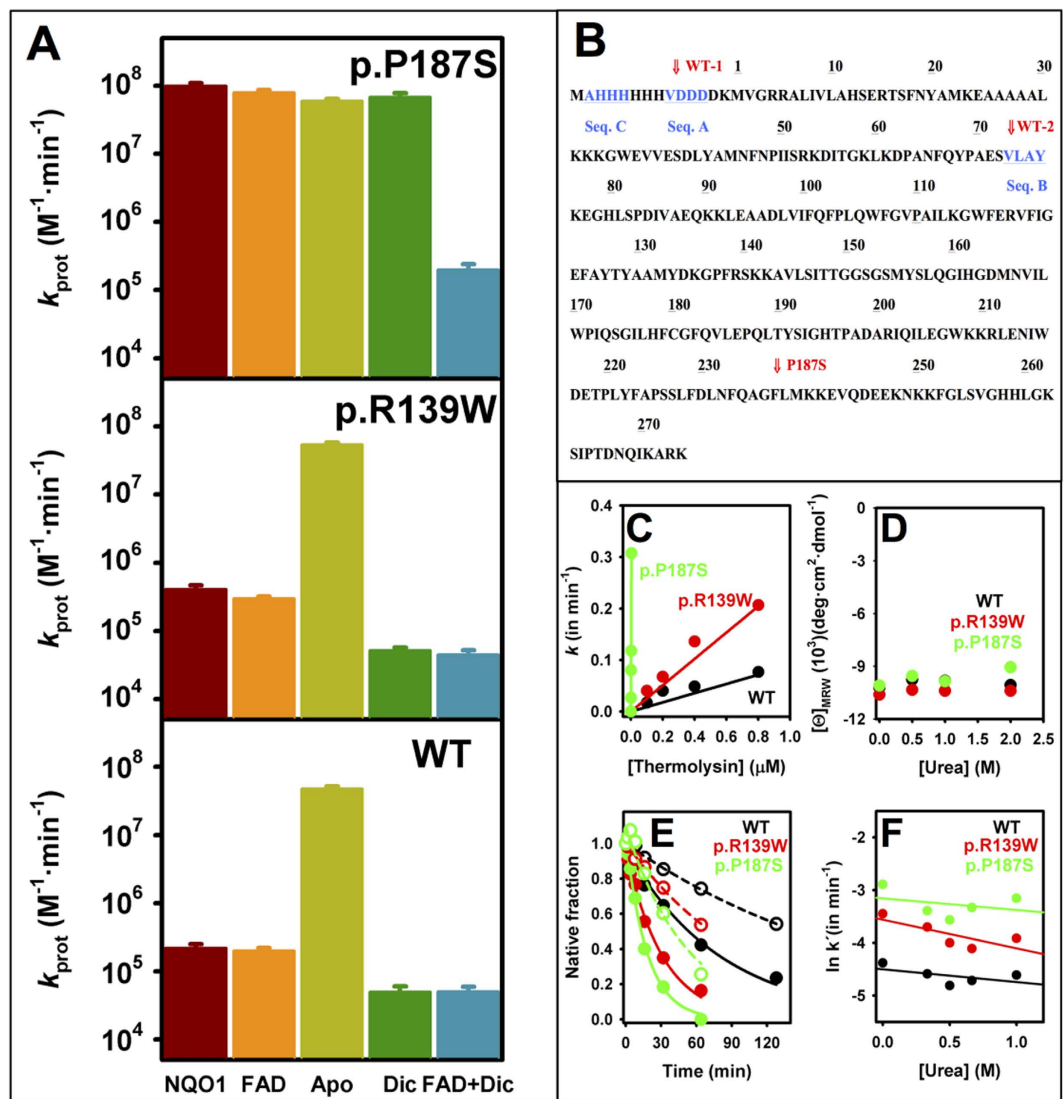
Figure 3. Proteolysis probes the local dynamics at N-terminal and C-terminal domains of NQO1 and the effects of polymorphisms and ligand binding. (**A**) Proteolysis rate constants of different NQO1 enzymes under different experimental conditions: NQO1 (as purified), FAD (in FAD excess), Apo (FAD withdrawn), Dic (in dicoumarol excess) and FAD + Dic (in FAD and dicoumarol excess). (**B**) Cleavage sites mapped onto the sequence of the WT NQO1 construct used in our study. Numbering corresponding to the NQO1 sequence excluding the N-terminal tag. Seqs (A–C) are N-terminal sequences determined for selected proteolysis products (highlighted in blue and underlined). The primary cleavage sites are labeled in red. See the main text for additional details. (**C–F**) proteolysis probes the local dynamics of NQO1 proteins. (**C**) The linear response of proteolysis rate constants with protease concentration shows that the proteolysis step is rate-limiting under these conditions; (**D**) Far-UV CD signals at 220 nm after 4 h incubation at selected urea concentrations (protein concentration 6 μM in monomer); (**E**) Proteolysis of NQO1 enzymes as holo-proteins after incubation with 1 M urea for at least two hours (open symbols) or without urea (closed symbols). Thermolysin concentrations were 0.1 μM (WT and p.R139W) and 1 nM (p.P187S). (**F**) Dependence of the proteolysis rate constants at different urea concentrations after correction by the urea effect on thermolysin activity reported by[28] (k'). The equilibrium m values are -0.14 ± 0.12 (WT), -0.30 ± 0.15 (p.R139W) and -0.12 ± 0.21 (p.P187S), in kcal·mol^{-1}·M^{-1}.

consistent with cleavage between residues 1 and 2 of our NQO1 construct, and thus, thermolysin is also cleaving at its C-terminal region. Since thermolysin cleaves at the N-terminus of hydrophobic residues[27], the cleavage site is either between Gly235-Phe236 (theoretical mass, 28087 Da) or Phe236-Leu237 (theoretical mass, 28234 Da) which we refer to as the P187S site (Fig. 3B) and the intermediate formed as I$_{28.2}$. Therefore, FAD binding seems to strongly affect NQO1 dynamics, but in the case of p.P187S, the far C-terminal region is still highly dynamic and sensitive to proteolysis (in agreement with[21]). Dicoumarol binding strongly protects (by ~400-fold; Fig. 3A) p.P187S towards degradation in the presence of FAD (Figs S2 and S3), supporting the hypothesis that its binding overcomes the dynamical effects of this polymorphism when FAD is also bound. In contrast, dicoumarol binding

modestly protects WT and p.R139W (~5-fold), and this stabilization does not depend on the addition of a FAD excess (Fig. 3A). We also noted that, a dicoumarol specific effect on the C-terminal dynamics of p.P187S could contribute to the low affinity of this variant for dicoumarol (Table 1).

Since p.P187S contains low levels of FAD as purified due to its low binding affinity[20,21] (and Table 1), we evaluated the proteolysis kinetic patterns of the three apo-proteins (Fig. 2C). Interestingly, all three NQO1 enzymes show a similar behaviour, being extremely sensitive towards cleavage, and thus they appear to be very dynamic ensembles (Figs 2C and 3A). In the case of WT and p.R139W, removal of FAD leads to a 200-fold increase in the sensitivity towards thermolysin (Fig. 3A), which qualitatively agrees with previous reports on WT NQO1 using trypsin[21,26]. Analyses of the proteolysis reactions of apo-NQO1 enzymes at short times revealed a very heterogeneous mixture of proteolysis products (Fig. 2C), but we also detected high levels of the two proteolysis products corresponding to those found in p.P187S, with masses of 29.9 and 28.2 kDa (Table S1). According to these data, and previous proteolysis experiments on apo WT using trypsin[21,26], proteolysis of apo-enzymes support a highly flexible C-terminal domain.

The cleavage site between Ser72-Val73 (site WT-2) is located in a helix and displays average B-factors and low solvent accessibility based on the crystal structures of either WT or p.P187S NQO1 (Fig. S4). In principle, this cleavage site would not be accessible in the native state, and thus, it would require a local unfolding event to become susceptible to proteolysis. In all structures analyzed, the cleavage site for p.P187S (Gly-235-Leu237; site P187S) is found at a more solvent-exposed region with high B-factors (Fig. S4), suggesting that this region would be highly dynamic in the native state. However, caution is required when using the X-ray crystal structure to derive dynamic information on these cleavage sites, and instead, we used dynamic information obtained from molecular dynamic simulations (see next section).

To determine whether cleavage at sites WT-2 and P187S occur through an unfolding event or due to a local fluctuation, we have measured the kinetics of proteolysis in the presence of low urea concentrations[28]. First, we confirmed that, under the experimental conditions, the proteolysis step is rate-limiting for all NQO1 enzymes (Fig. 3C). Then, we selected low urea concentrations (0–1 M) that do not cause significant unfolding based on Far-UV CD measurements (Fig. 3D). These conditions are thus reflecting proteolysis of NQO1 enzymes under near-native conditions. In this scenario, if significant unfolding is required to reach the cleavable state, the proteolysis rate constant will strongly depend on urea concentration (i.e. urea will speed up proteolysis), and the urea concentration dependence of the proteolysis rate constant will yield the equilibrium m value between the native and cleavable states, which is expected to be 3.1 kcal·mol^{-1}·M^{-1} for the complete unfolding of the NQO1 monomer[29]. Conversely, if a local fluctuation leads to the cleavable state, then the proteolysis rate constants will be essentially independent of urea concentration[28]. In our case, we found that proteolysis was somewhat slowed down by the presence of urea (Fig. 3E), due to the modest inhibition of thermolysin by urea[28]. When corrected for this effect (k' in Fig. 3F), equilibrium m values are almost zero (Fig. 3F), supporting the local fluctuation mechanism for the proteolysis of all three holo-NQO1 enzymes.

Effect of polymorphisms and ligand binding on NQO1 global stability from MD simulations.

We have then studied whether the changes in thermal/kinetic stability of NQO1 due to polymorphisms and ligand binding correlate with changes in the average dynamics of NQO1 using molecular dynamics (MD) simulations. In these simulations, the polymorphisms and/or the ligands have no significant effects on the overall secondary structure content (α-helix and β-sheet; Table S2) in agreement with previous experimental spectroscopic analyses[20]. Furthermore, we do not observe major differences in sampled conformations as the maximum RMSD between all systems is below 2 Å.

Three parameters that refer to the conformational dynamics were analyzed: B-factors (Fig. 4A), dihedral entropies and total vibrational entropies from normal mode analyses (Fig. S5). Average B-factors for the apo-proteins showed a good correlation with their effect on thermal/kinetic stability, with larger average values as the stability decreases. These effects were also observed for dihedral and vibrational entropies, but to a lower extent. Addition of FAD causes to all three proteins a remarkable decrease in the average B-factors (Fig. 4A), consistent with the strong stabilization exerted by FAD in all three variants[20] (and Fig. 1). Again, the same trend is observed for the dihedral and vibrational (Fig. S5) entropies, but the results are less pronounced. Addition of dicoumarol also reduces the flexibility of apo- and holo-forms of WT and p.P187S (Figs 4A and S5), consistent with its global stabilizing effect on both variants.

Since the NQO1 dimer dissociates prior to the rate-limiting step of thermal denaturation[20], the stability and flexibility of the monomer-monomer interface must strongly affect the kinetic stability of NQO1. Indeed, the destabilizing effect of polymorphisms and stabilizing effect of FAD binding are associated with large changes in the flexibility of the monomer-monomer interface (Fig. 4B). These changes in flexibility may affect the optimal interactions at the dimer interface, thus contributing to the kinetic stability of NQO1 enzymes.

Correlation between changes in local dynamics from proteolysis and MD simulations.

As native state local fluctuations govern NQO1 proteolysis kinetics, we have used our MD simulations to correlate proteolysis experiments by thermolysin (and trypsin[21]) with changes in flexibility at the residue level. In the case of NQO1, proteolysis kinetics at a certain site depends on the local dynamics of the site as well as on the intrinsic preference of the protease to cleave at that sequence. For these analyses, we have considered a window of about 10 residues around the cleavage site, since the conformation (and dynamics) of 10–12 residues around the cleavage site determines efficient protease binding and cleavage[30]. The flexibility of all three variants around the Ser72-Val73 cleavage site decreases upon FAD binding, whereas only WT and p.R139W also showed a similar decrease in the Gly235-Leu237 cleavage site (Fig. 4C–E). These results are consistent with fast cleavage of holo-p.P187S at the P187S site (Fig. 3A,B; please see[21] for similar results using trypsin). Similar analyses performed using dihedral entropies showed that FAD binding causes a modest decrease in flexibility at the Ser72-Val73, but no
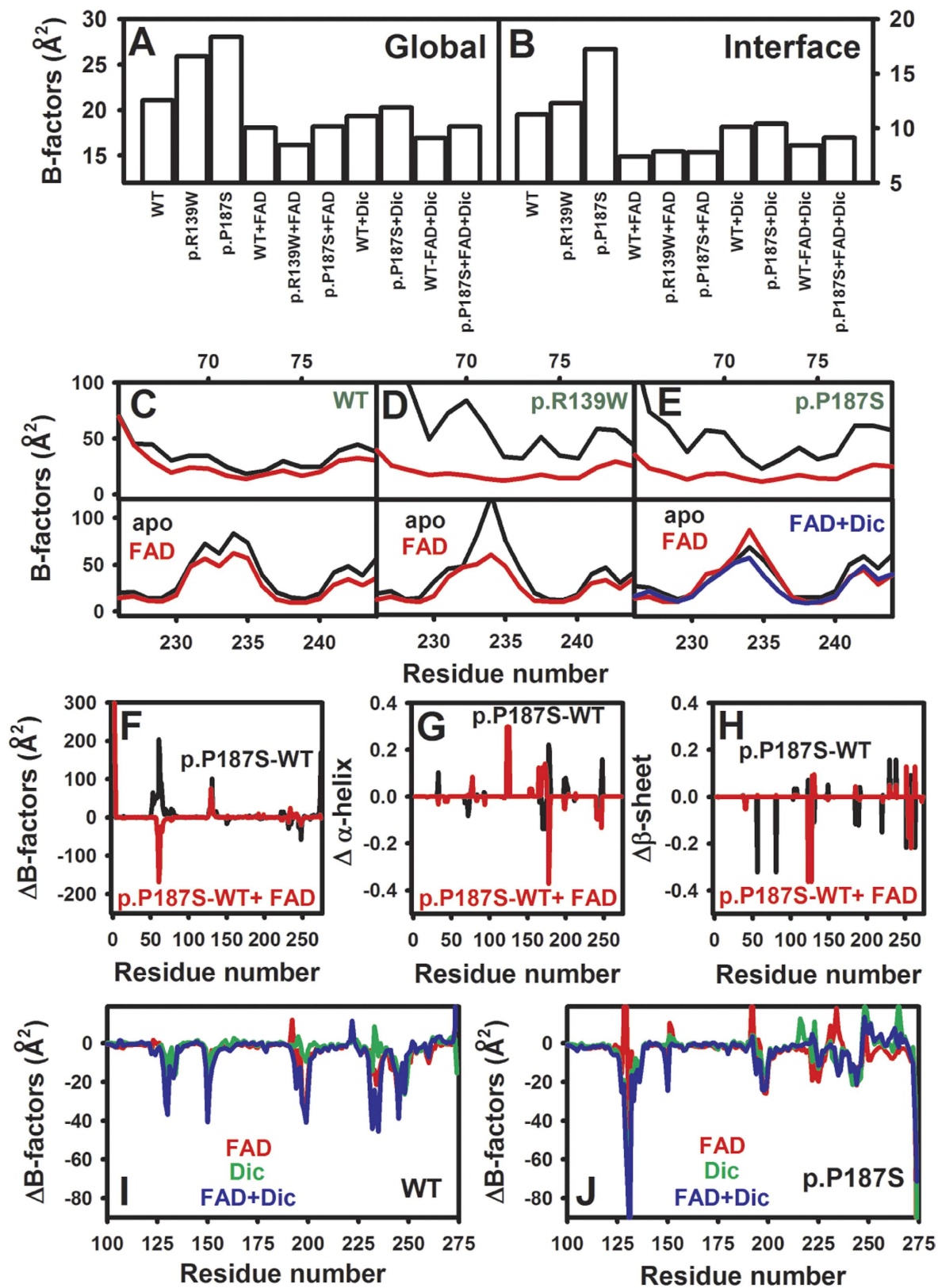
Figure 4. MD simulations support the role of global and local dynamics on the effects of polymorphisms on stability, catalytic function and ligand binding. (**A,B**) Effects on the global (**A**) and dimer interface (**B**) dynamics (B-factors); (**C–E**) Effects of NQO1 polymorphisms, FAD and FAD + dicoumarol binding on the dynamics at the residue level at the vicinity of the N-terminal and C-terminal primary cleavage sites of the NQO1 enzyme. (**F–H**) Difference in dynamics (**F**), α-helix (**G**) and β-sheet (**H**) propensities between p.P187S and WT NQO1 without bound ligands and with FAD bound at the residue level. (**I,J**) Effect of ligand binding (FAD, dicoumarol and FAD + Dicoumarol) on the local dynamics of WT (**I**) and p.P187S (**J**) at the residue level.

clear effect at the Gly235-Leu237 site (Fig. S5). Simultaneous binding of FAD and dicoumarol caused a decrease in flexibility at the Gly235-Leu237 cleavage site of p.P187S (Fig. 4E), supporting the conclusion that this variant is protected towards proteolysis in the presence of both ligands.

Changes in local dynamics at the FAD and NADH/Dicoumarol binding sites explain the catalytic impairment of p.P187S.
p.P187S binds FAD and dicoumarol (a competitive inhibitor of NADH) with ~10–100 fold lower affinity than the WT enzyme[20,21] (Table 1). To explain these alterations in binding affinity at the structural and dynamic levels, we have compared the flexibility (B-factors) and secondary structure propensities of p.P187S and WT NQO1 from our MD trajectories (Fig. 4F–J). It is worth noting that changes in either the ligand-free or the ligand-bound conformational ensembles are expected to affect the ligand binding affinity.

The p.P187S polymorphism seems to affect the flexibility of different regions important for FAD binding (Figs. 1B and 4F). As apo-protein, this polymorphism increases the flexibility of the region 58–67, which is compensated by FAD binding in p.P187S, and thus causes an energy (entropic) penalty to fix the right interaction of Gln66, Tyr67 and Pro68 with the FAD molecule. A significant increase in flexibility is also found for p.P187S, either as apo- and holo-enzyme, in the region 127–134 including Tyr126 and Tyr128 which are part of the FAD binding site, also possibly affecting the proper interaction with FAD. In the FAD-bound state of p.P187S, we also observed local changes in secondary structure propensities at several regions involved in the FAD binding site (residues 123–127, 170–179 and 241–248; Fig. 4G,H) which may also destabilize the FAD-NQO1 complex in this variant by distorting the FAD binding site.

Binding of dicoumarol has also significant effects on the local dynamics of NQO1 WT and p.P187S (Fig. 4I,J) altering the dynamics of three regions forming the dicoumarol binding site (Fig. 1C), and notably, these changes differ between WT and p.P187S (Fig. 4I,J). Residues 127–136, 149–155 and 230–238 involve regions of high flexibility in the apo-state of WT, which show much lower flexibility upon ligand binding, especially in the ternary (FAD + dicoumarol) complex (Fig. 4I). It is also important to note that in our MD simulations, dicoumarol does not adopt a well-defined binding pose in the absence of FAD, but binding becomes much stronger when FAD is also present. The most noticeable change in p.P187S is found in the region 127–136 (Fig. 4J), which is much more flexible in the apo-state than that for WT. Therefore, this region becomes much more *conformationally restrained* in the ternary complex, and thus, entropically penalizes dicoumarol binding to p.P187S explaining its lower affinity for this inhibitor.

Rescue of p.P187S in cultured cells requires stabilization of its C-terminal domain.
The dynamic alterations caused by p.P187S might be important to understand the low activity and stability of this polymorphism *in vivo*. Since small ligands such as FAD and dicoumarol correct these dynamic alterations *in vitro*, they could be used pharmacologically to rescue p.P187S. We have thus evaluated the effect of riboflavin and dicoumarol in NQO1 protein levels using cell lines endogenously expressing different NQO1 variants: HeLa (homozygotes for WT), HCT 116 (heterozygotes for p.R139W) and Caco-2 cells (homozygotes for p.P187S).

The steady-state levels of p.P187S in Caco-2 cells are enhanced 4-5 fold in the presence of dicoumarol, while riboflavin alone (a precursor of FAD) has a much lower effect (Fig. 5). Remarkably, these two ligands have small effects on the protein levels of WT (HeLa cells) and p.R139W (HCT 116 cells) (Fig. 5). Since the stability of p.P187S towards proteasomal degradation correlate well with its protein levels in cells, without significant effects on mRNA levels[19], these results suggest that dicoumarol binding is protecting p.P187S towards proteasomal degradation. Therefore, our results may imply that the intracellular stability of p.P187S strongly depends on the dynamics of its C-terminal domain, which could be therapeutically targeted by ligands binding to its C-terminal domain (such as dicoumarol). Moreover, since intracellular levels of NQO1 are determined to a large extent by the rate of degradation by the 26 S and/or 20 S proteasome[18,19], our results also suggest that the proteasome degrades p.P187S preferentially through its flexible C-terminal end.

Discussion

The effects of pathogenic genetic variations on protein folding, stability and function are often interpreted using high-resolution crystal structures. However, alterations in protein dynamics are not straightforwardly inferred from this type of structural approach. Particularly, the effect of the p.P187S polymorphism on the crystal structure of NQO1 have been shown to be marginal[21], and therefore, little insight into the pathogenic mechanisms leading to increased risk of cancer can be extracted from these *static* analyses. By contrast, the present study shows that the pathogenic effects of p.P187S can be studied in-depth using a combination of experimental and computational approaches aimed at correlating changes in protein dynamics, stability and functionality *in vitro* and inside cells. Indeed, the low activity and stability of p.P187S appears to have a dynamic origin, caused by a significant increase in local flexibility at the active site, particularly at the FAD and NADH/substrate binding sites and at the dimer interface. While riboflavin/FAD supplementation may rescue the activity of p.P187S by shifting the equilibrium towards the functional holo-state, the intracellular stability of this polymorphism seems to require a ligand (such as dicoumarol) that overcomes the dynamic alterations at the C-terminal domain, at least in our cell model systems. In the case of p.R139W, dynamic alterations are much weaker thus leading to only modest changes in activity and stability[11,20,22], and therefore, the main molecular pathogenic mechanism for this polymorphism is likely to be the skipping of exon 4 (residues 102–139 of the full-length protein) that destroys the binding sites of FAD and the substrate (Fig. 1B,C) rendering an inactive and unstable protein[11].

A recent study has shown that apo-NQO1 proteins are highly susceptible to proteolysis by trypsin, and in the case of p.P187S, this high susceptibility is not corrected upon FAD binding[21]. These results, in combination with those from NMR spectroscopy, have supported that p.P187S in the apo-state is highly dynamic, while in the holo-state its C-terminal domain remains flexible, which may compromise its catalytic efficiency[21]. Importantly, our present work allows to map changes in protein dynamics at the residue level due to p.P187S and ligand
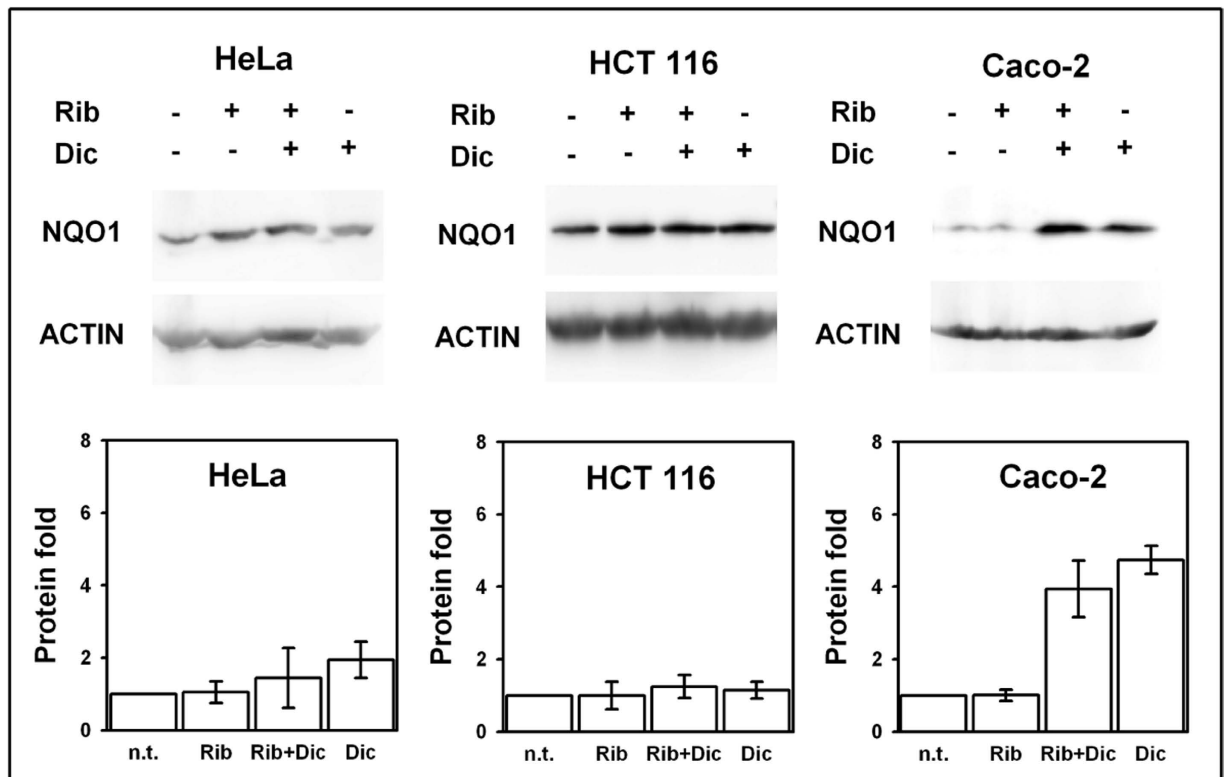
Figure 5. Pharmacological rescue of p.P187S. Effect of riboflavin (Rb) and/or dicoumarol (Dic) on the steady-state NQO1 protein levels in HeLa cells (expressing WT NQO1), HCT 116 cells (heterozygous for p.R139W) and Caco-2 cells (homozygous for p.P187S). Representative western-blots for NQO1 and β-actin in soluble extracts upon treatment with or without riboflavin and/or dicoumarol. The plots show the protein NQO1 levels corrected for β-actin levels and normalized using those from untreated cells. Data are mean ± S.E.M. from three independent experiments.

binding, and also to link them with alterations in protein stability (specially at the monomer:monomer interface and the C-terminal domain) and activity. Indeed, p.P187S affects the dynamics of both the C-terminal (as previously shown by[21]) and N-terminal domains, particularly at the FAD and NADH/dicoumarol binding sites, and these dynamic alterations in both domains contribute to its poor catalytic performance and binding affinity for these ligands. We also observe that FAD binding is not sufficient to correct the dynamic alterations at the C-terminal domain *in vitro* (as also shown by[21]) and in eukaryotic cells, and that dicoumarol binding is required for its correction. Our experiments thus support that the dynamics of the C-terminal domain, and its modulations by ligands, determine to a large extent the intracellular stability of p.P187S.

Our results have important implications for the search of pharmacological ligands correcting NQO1 intracellular stability and function, as new therapeutic approaches for those cancer patients bearing the p.P187S polymorphism, similar to those recently proposed to rescue unstable cancer-associated mutants of p53[31,32]. Protein degradation by the 20 S proteasome may proceed by either the N- and C-terminus, as well as from internal sequences, and depends on the protein substrate and the local stability of the initiation site recognized by the proteasome[25]. Similarly, degradation by the 26 S proteasome requires the recognition of a polyubiquitin tag and a disordered initiation site, often in the form of a dynamic tail of about 30 amino acids[33]. Therefore, local (rather than global) unfolding rates and/or thermodynamic stabilities may determine the directional preference of proteasomal to degrade proteins (as recently proposed by[34]). Our MD simulations show that the C-terminal end of NQO1 is more flexible than the N-terminal (Fig. S6), which may imply that degradation of the WT protein proceeds preferentially through its C-terminal. Accordingly, the enhanced proteasomal degradation of p.P187S would be linked to its dramatic effect on the C-terminal dynamics. Therefore, in the search for novel pharmacological chaperones to rescue p.P187S, we should take into account their effects on the stability and dynamics of the C-terminal domain.

In conclusion, we provide a comprehensive picture of the structural and dynamic consequences of NQO1 polymorphisms, linking changes in protein dynamics with impaired NQO1 function and FAD binding, particularly for p.P187S, and leading to increased cancer susceptibility (and to lower responses to some chemotherapeutic strategies). Moreover, alterations in protein dynamics can be modulated by small ligands, therefore setting a molecular framework to search for new pharmacological ligands to treat patients bearing these polymorphisms. Additionally, our approach could be used to unravel the pathogenic mechanisms underlying other loss-of-function genetic diseases associated to increased protein turnover, and notably, to dissect the complex

effects of mutations and polymorphisms on the conformational and dynamic features of native proteins ultimately causing catalytic and stability defects.

Materials and Methods

Protein expression and purification. NQO1 enzymes were expressed in and purified from *Escherichia coli* as N-terminal hexahis-tagged proteins as described[20]. Apo-proteins were obtained upon treatment with potassium bromide as described[20]. NQO1 enzymes were stored in HEPES-KOH 50 mM pH 7.4 at −80 °C and their concentration measured spectrophotometrically as described[20]. Three independent purifications of each NQO1 were used thorough this study, and control experiments were performed to confirm that the results are batch-independent.

Differential scanning calorimetry (DSC). DSC studies were performed on a capillary VP-DSC differential scanning calorimeter (Malvern Instruments) with a cell volume of 0.135 mL. Scans were performed at a 1–3 °C·min^{-1} in a temperature range of 5–80 °C using 5–30 μM NQO1 (monomer) in 50 mM HEPES-KOH, pH 7.4. In some experiments, FAD and/or dicoumarol (both from Sigma-Aldrich) were added to a final concentration up to 0.5 mM. Stock solutions of dicoumarol 10 mM were prepared in 0.1 M NaOH.

Activity measurements. NQO1 activity was measured in 50 mM HEPES-KOH, pH 7.4. A reaction mixture containing recombinant NQO1 and 0.5 mM NADH was incubated at 30 °C for 5 min in a 1 cm path length quartz cuvettes in a thermostatized Agilent 8453 diode-array spectrophotometer. The reaction was triggered by the addition of DCPIP 75 μM as the electron acceptor. Initial reaction rates were determined from changes in A_{600nm} resulting from the reduction of DCPIP and corrected for the non-enzymatic reaction. The NQO1 concentration used varied depending on the variant to ensure linearity over time and protein concentration: 1–2 nM for WT and 25–50 nM for p.P187S. The specific activity was calculated using $\varepsilon_{600\,nm} = 21000\,M^{-1}\cdot cm^{-1}$ for DCPIP.

Isothermal titration calorimetry (ITC). Experiments were performed in a VP ITC$_{200}$ microcalorimeter (Malvern Instruments) with a cell volume of 0.205 mL. Cell samples contained 5 μM NQO1 enzymes (dimer) in HEPES-KOH 50 mM pH 7.4 and were titrated using 0.25 mM dicoumarol by performing 30–50 injections of 0.8–1.2 μL each 180–240 s. Data were corrected for heats of dilution by performing dicoumarol titrations into buffer. After peak integration, binding isotherms were analyzed using a one-type of independent sites model using the software provided by the manufacturer.

Proteolysis by thermolysin. Thermolysin from *Bacillus thermoproteolyticus rokko* was purchased from Sigma-Aldrich, dissolved in HEPES-KOH 50 mM CaCl$_2$ 10 mM pH 7.4, stored at −80 °C and its concentration measured spectrophotometrically using $\varepsilon_{280\,nm} = 66086\,M^{-1}\cdot cm^{-1}$. Proteolysis was performed at 25 °C in HEPES-KOH 50 mM CaCl$_2$ 10 mM pH 7.4 using 20 μM in NQO1 (monomer), and the reaction was initiated by the addition of thermolysin (at a final concentration of 10 pM-1 μM) in a final volume of 200 μL. Proteolysis reactions were performed using the NQO1 proteins as purified or as apo-proteins. In some cases, FAD and/ or dicoumarol were added to a final concentration of 100 μM. At different times, 20 μL of the reaction mixture were withdrawn and proteolysis quenched by addition of 5 μL of EDTA 100 mM pH 8, mixed with Laemmli´s buffer and denatured at 95 °C for 5 min. Samples were resolved in 12% acrylamide SDS-PAGE gels, stained with Coomassie Blue and densitometered using ImageJ (http://rsbweb.nih.gov/ij/). After normalization of the intensities using that of the control sample in the absence of thermolysin, the Intensity (*I*) vs. Time (*t*) profiles were analyzed using a single exponential function:

$$I = I_0 \cdot exp^{-k \cdot t}$$

where k is the proteolysis rate constant and I_0 the initial intensity.

In some cases, NQO1 samples were incubated with urea (0–1 M) for 4 hours at 25 °C and then proteolysis rates constants were determined as described above. In this case, the observed proteolysis rate constants (k) were corrected for the corresponding inhibitory effect of urea on thermolysin at a given urea concentration reported by[28]. These corrected constants (k') were used to determine equilibrium m values from the slope of a ln k' vs. [urea] as previously described[28].

Limited proteolysis studied by mass spectrometry and N-terminal sequencing. In selected conditions, proteolysis reactions were performed as described above in a volume of 500 μL. After quenching with EDTA, samples were concentrated to 100 μL and submitted to two concentration/dilution cycles with water to a final volume of 100 μL using Vivaspin 10000 Da cut-off filters. These samples were analyzed by high performance liquid chromatography/electrospray-ionization mass spectrometry (HPLC/ESI-MS) at the high resolution mass spectrometry service at the Centro de Instrumentación Científica, University of Granada. HPLC/ESI-MS was performed in a Acquity UPLC system (Waters) using a gradient of water/formic acid (0.1%) and acetonitrile/formic acid (0.1%) using a Acquity UPLC BEH300 C4 column (2.1 × 50 mm; Waters) coupled to a QTOF Synapt62 HDMS (Waters).

For N-terminal sequencing, 100 μL of selected proteolysis mixtures were denatured in Laemmli´s buffer and run in a 12% SDS-PAGE followed by electrotransference to a polyvinylidene difluoride (PVDF) membrane. The membrane was stained with Coomassie blue G-250 and selected bands were cut, destained and equilibrated in water for N-terminal sequencing by the Edman´s method (performed at the service of Protein Chemistry, Centro de Investigaciones Biológicas, Madrid, Spain).

Circular dichroism spectroscopy.   The effect of low urea concentrations (0–2 M) on the conformation of holo-NQO1 enzymes was monitored by circular dichroism (CD) spectroscopy in a Jasco J-710 spectropolarimeter and thermostatized at 25 °C using a Peltier element. Protein samples (6 μM NQO1 monomer) were prepared in HEPES-KOH 50 mM pH 7.4 in the presence of 0–2 M urea, incubated at 25 °C for 4–6 h and their Far-UV CD spectra acquired using 1 mm path length cuvettes. Four scans were measured for each sample in a 210–260 nm range at 100 nm·min^{-1} scan rate and averaged. The corresponding blanks without NQO1 were measured similarly and subtracted.

Structural analyses based on the NQO1 crystal structures.   Structural calculations were performed using the structures 2F1O (for WT NQO1 in the presence of dicoumarol)[35], 1DXO (for WT NQO1 in the presence of duroquinone)[36] and 4CF6 (for p.P187S in the presence of Cybacron blue)[21]. Average B-factors were determined for the different protein chains described in the structures and the averages are presented. Solvent accessible surface areas (SASA) of the main-chain and side-chains were calculated using the Shrake-Rupley algorithm[37] with a radius of 1.4 Å for the solvent probe and the Chothia set for the protein atoms[38] and a -X-Gly-X-model for the unfolded state using a home-built software kindly provided by Prof. Jose M. Sánchez-Ruíz (Dept. of Physical Chemistry, University of Granada).

Molecular dynamics simulations.   We set up ten independent molecular dynamics simulations of wild-type and mutated NQO1 as apo systems, binary complexes with FAD or dicoumarol as well ternary complexes. Systems were prepared in dimeric state on basis of crystal structures of the complex with FAD (PDB: 1D4A[36]) and the ternary complex (PDB: 2F1O[35]). Systems were prepared for simulation using protonate3d and mutated in MOE[39].

Simulations were performed using the GPU implementation of pmemd in Amber12[40]. Protein residues were parametrized using the Amber forcefield 99SB-ILDN[41]. FAD and dicoumarol were parametrized within the Generalized Amber Force Field using additional bonding parameters for FADH[42]. Partial charges for ligands were derived as described earlier[43]. After employing an extensive equilibration protocol[44], unrestrained systems were sampled for 100 ns in NpT ensemble and 5,000 equal-spaced snapshots were saved to trajectory.

After ensuring thermodynamic and structural stability of simulations, we analyzed resulting conformational ensembles using cpptraj from AmberTools[45]. We analyzed secondary structure elements and B-factors of Cα atoms (derived from root mean squared fluctuations) after a global alignment as a metric for flexibility. Total vibrational entropies were calculated via a normal mode analysis of the covariance matrix of Cα atoms[46].

Furthermore, we extracted backbone dihedral angles and calculated dihedral entropies based on resulting distributions after employing parameter-free kernel density estimation[47]. Integration over probability densities yields a thermodynamics entropy depicting a metric for local flexibility of the backbone[48]. Dihedral entropies over all three backbone torsions were summed to give a total dihedral entropy per residue. All values are presented as arithmetic average over two dimer sub-units. Interface residues were defined as all residues with an atom closer than 3 Å to the adjacent sub-unit.

Cell culture and immunoblotting.   HeLa, HCT 116 and Caco-2 cell lines were grown in RPMI-1640 with ultraglutamine, supplemented with 10% of FBS (Hyclone) and 1× Antibiotic Antimycotic Solution (Sigma-Aldrich). Cells were incubated with riboflavin 10 μM and/or dicoumarol 100 μM (both from Sigma-Aldrich) for 24 h. Cells were then collected and lysed in RIPA buffer (50 mM Tris-HCl, 150 mM NaCl, 0.1% Triton X-100, 0.1% sodium dodecyl sulphate, 1 mM sodium orthovanadate, 1 mM NaF pH 8) with protease inhibitors (COMPLETE, from Roche) and lysed by freezing-thawing cycles and centrifuged at 18000 g for 15 min. Soluble extracts were collected and the amount of total protein determined by the BCA method (Pierce) using bovine serum albumin as a standard. Approximately 100 μg of total protein were denatured in Laemmli´s buffer at 95 °C for 5 min and loaded into 12% polyacrylamide gels and separated. After transferring to polyvinylidene difluoride membranes (GE Healthcare), immunoblotting was carried out using primary monoclonal antibodies anti-NQO1 and anti-β-actin (Santa Cruz Biotechnology) from mouse at 1:500 and 1:10000 dilutions, respectively. As secondary antibody, we used goat anti-mouse IgG-HRP (Santa Cruz Biotechnology) at 1:1000 dilution. Chemiluminiscent detection of bands was carried out using Clarity™ Western ECL Blotting Substrate (Biorad) and analyzed using a luminescent image analyzer LAS-4000 mini (Fujifilm).

References

1. Belinsky, M. & Jaiswal, A. K. NAD(P)H:quinone oxidoreductase1 (DT-diaphorase) expression in normal and tumor tissues. *Cancer Metastasis Rev* **12,** 103–117 (1993).
2. Buranrat, B. *et al.* NQO1 expression correlates with cholangiocarcinoma prognosis. *Asian Pac J Cancer Prev* **13** Suppl, 131–136 (2012).
3. Garate, M., Wani, A. A. & Li, G. The NAD(P)H:Quinone Oxidoreductase 1 induces cell cycle progression and proliferation of melanoma cells. *Free Radic Biol Med* **48,** 1601–1609, doi: 10.1016/j.freeradbiomed.2010.03.003 (2010).
4. Colucci, M. A., Moody, C. J. & Couch, G. D. Natural and synthetic quinones and their reduction by the quinone reductase enzyme NQO1: from synthetic organic chemistry to compounds with anticancer potential. *Org Biomol Chem* **6,** 637–656, doi: 10.1039/b715270a (2008).
5. Bailey, S. M. *et al.* Involvement of DT-diaphorase (EC 1.6.99.2) in the DNA cross-linking and sequence selectivity of the bioreductive anti-tumour agent EO9. *Br J Cancer* **76,** 1596–1603 (1997).
6. Asher, G., Bercovich, Z., Tsvetkov, P., Shaul, Y. & Kahana, C. 20 S proteasomal degradation of ornithine decarboxylase is regulated by NQO1. *Mol Cell* **17,** 645–655, doi: 10.1016/j.molcel.2005.01.020 (2005).
7. Asher, G., Lotem, J., Cohen, B., Sachs, L. & Shaul, Y. Regulation of p53 stability and p53-dependent apoptosis by NADH quinone oxidoreductase 1. *Proc Natl Acad Sci USA* **98,** 1188–1193, doi: 10.1073/pnas.021558898 (2001).
8. Asher, G., Lotem, J., Kama, R., Sachs, L. & Shaul, Y. NQO1 stabilizes p53 through a distinct pathway. *Proc Natl Acad Sci USA* **99,** 3099–3104, doi: 10.1073/pnas.052706799 (2002).

9. Asher, G., Tsvetkov, P., Kahana, C. & Shaul, Y. A mechanism of ubiquitin-independent proteasomal degradation of the tumor suppressors p53 and p73. *Genes Dev* **19,** 316–321, doi: 10.1101/gad.319905 (2005).
10. Ben-Nissan, G. & Sharon, M. Regulating the 20 S proteasome ubiquitin-independent degradation pathway. *Biomolecules* **4,** 862–884, doi: 10.3390/biom4030862 (2014).
11. Pan, S. S., Forrest, G. L., Akman, S. A. & Hu, L. T. NAD(P)H:quinone oxidoreductase expression and mitomycin C resistance developed by human colon cancer HCT 116 cells. *Cancer Res* **55,** 330–335 (1995).
12. Traver, R. D. *et al.* NAD(P)H:quinone oxidoreductase gene expression in human colon carcinoma cells: characterization of a mutation which modulates DT-diaphorase activity and mitomycin sensitivity. *Cancer Res* **52,** 797–802 (1992).
13. Traver, R. D. *et al.* Characterization of a polymorphism in NAD(P)H: quinone oxidoreductase (DT-diaphorase). *Br J Cancer* **75,** 69–75 (1997).
14. Flicek, P. *et al.* Ensembl 2014. *Nucleic Acids Res* **42,** D749–755, doi: 10.1093/nar/gkt1196 (2014).
15. Lajin, B. & Alachkar, A. The NQO1 polymorphism C609T (Pro187Ser) and cancer susceptibility: a comprehensive meta-analysis. *Br J Cancer* **109,** 1325–1337, doi: 10.1038/bjc.2013.357 (2013).
16. Eguchi-Ishimae, M. *et al.* The association of a distinctive allele of NAD(P)H:quinone oxidoreductase with pediatric acute lymphoblastic leukemias with MLL fusion genes in Japan. *Haematologica* **90,** 1511–1515 (2005).
17. Siegel, D., McGuinness, S. M., Winski, S. L. & Ross, D. Genotype-phenotype relationships in studies of a polymorphism in NAD(P)H:quinone oxidoreductase 1. *Pharmacogenetics* **9,** 113–121 (1999).
18. Moscovitz, O. *et al.* A mutually inhibitory feedback loop between the 20 S proteasome and its regulator, NQO1. *Mol Cell* **47,** 76–86, doi: 10.1016/j.molcel.2012.05.049 (2012).
19. Siegel, D. *et al.* Rapid polyubiquitination and proteasomal degradation of a mutant form of NAD(P)H:quinone oxidoreductase 1. *Mol Pharmacol* **59,** 263–268 (2001).
20. Pey, A. L., Megarity, C. F. & Timson, D. J. FAD binding overcomes defects in activity and stability displayed by cancer-associated variants of human NQO1. *Biochim Biophys Acta* **1842,** 2163–2173, doi: 10.1016/j.bbadis.2014.08.011 (2014).
21. Lienhart, W. D. *et al.* Collapse of the native structure caused by a single amino acid exchange in human NAD(P)H:quinone oxidoreductase(1.). *FEBS J* **281,** 4691–4704, doi: 10.1111/febs.12975 (2014).
22. Pan, S. S., Han, Y., Farabaugh, P. & Xia, H. Implication of alternative splicing for expression of a variant NAD(P)H:quinone oxidoreductase-1 with a single nucleotide polymorphism at 465 C > T. *Pharmacogenetics* **12,** 479–488 (2002).
23. Fraser, J. S. *et al.* Hidden alternative structures of proline isomerase essential for catalysis. *Nature* **462,** 669–673, doi: 10.1038/nature08615 (2009).
24. Tzeng, S. R. & Kalodimos, C. G. Dynamic activation of an allosteric regulatory protein. *Nature* **462,** 368–372, doi: 10.1038/nature08560 (2009).
25. Berko, D. *et al.* The direction of protein entry into the proteasome determines the variety of products and depends on the force needed to unfold its two termini. *Mol Cell* **48,** 601–611, doi: 10.1016/j.molcel.2012.08.029 (2012).
26. Chen, S., Deng, P. S., Bailey, J. M. & Swiderek, K. M. A two-domain structure for the two subunits of NAD(P)H:quinone acceptor oxidoreductase. *Protein Sci* **3,** 51–57, doi: 10.1002/pro.5560030107 (1994).
27. Fontana, A., Zambonin, M., De Filippis, V., Bosco, M. & Polverino de Laureto, P. Limited proteolysis of cytochrome c in trifluoroethanol. *FEBS Lett* **362,** 266–270 (1995).
28. Park, C. & Marqusee, S. Probing the high energy states in proteins by proteolysis. *J Mol Biol* **343,** 1467–1476, doi: S0022-2836(04)01085-X [pii] 10.1016/j.jmb.2004.08.085 (2004).
29. 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* **4,** 2138–2148, doi: 10.1002/pro.5560041020 (1995).
30. Hubbard, S. J., Eisenmenger, F. & Thornton, J. M. Modeling studies of the change in conformation required for cleavage of limited proteolytic sites. *Protein Sci* **3,** 757–768, doi: 10.1002/pro.5560030505 (1994).
31. Liu, X. *et al.* Small molecule induced reactivation of mutant p53 in cancer cells. *Nucleic Acids Res* **41,** 6034–6044, doi: 10.1093/nar/gkt305 (2013).
32. Wilcken, R. *et al.* Halogen-enriched fragment libraries as leads for drug rescue of mutant p53. *J Am Chem Soc* **134,** 6810–6818, doi: 10.1021/ja301056a (2012).
33. Inobe, T. & Matouschek, A. Paradigms of protein degradation by the proteasome. *Curr Opin Struct Biol* **24,** 156–164, doi: 10.1016/j.sbi.2014.02.002 (2014).
34. Rodriguez-Larrea, D. & Bayley, H. Protein co-translocational unfolding depends on the direction of pulling. *Nat Commun* **5,** 4841, doi: 10.1038/ncomms5841 (2014).
35. Asher, G., Dym, O., Tsvetkov, P., Adler, J. & Shaul, Y. The crystal structure of NAD(P)H quinone oxidoreductase 1 in complex with its potent inhibitor dicoumarol. *Biochemistry* **45,** 6372–6378, doi: 10.1021/bi0600087 (2006).
36. Faig, M. *et al.* Structures of recombinant human and mouse NAD(P)H:quinone oxidoreductases: species comparison and structural changes with substrate binding and release. *Proc Natl Acad Sci USA* **97,** 3177–3182, doi: 10.1073/pnas.050585797 (2000).
37. Shrake, A. & Rupley, J. A. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J Mol Biol* **79,** 351–371 (1973).
38. Chothia, C. The nature of the accessible and buried surfaces in proteins. *J Mol Biol* **105,** 1–12 (1976).
39. Labute, P. Protonate3D: assignment of ionization states and hydrogen coordinates to macromolecular structures. *Proteins* **75,** 187–205, doi: 10.1002/prot.22234 (2009).
40. AMBER 12 (University of California, San Francisco, 2012).
41. Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78,** 1950–1958, doi: 10.1002/prot.22711 (2010).
42. Antony, J., Medvedev, D. M. & Stuchebrukhov, A. A. Theoretical Study of Electron Transfer between the Photolyase Catalytic Cofactor FADH- and DNA Thymine Dimer. *J.Am.Chem.Soc.* **122,** 1057–1065 (2000).
43. Fuchs, J. E. *et al.* Dynamic regulation of phenylalanine hydroxylase by simulated redox manipulation. *PLoS One* **7,** e53005, doi: 10.1371/journal.pone.0053005 (2012).
44. Wallnoefer, H. G., Handschuh, S., Liedl, K. R. & Fox, T. Stabilizing of a globular protein by a highly complex water network: a molecular dynamics simulation study on factor Xa. *J Phys Chem B* **114,** 7405–7412, doi: 10.1021/jp101654g (2010).
45. Roe, D. R. & Cheatham, I. T.E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **9,** 3084–3095 (2013).
46. Fuchs, J. E., Fuchs, D. & Liedl, K. R. Dynamic regulation of phenylalanine hydroxylase. *Pteridines* **25,** 33–39 (2014).
47. Huber, R. G. *et al.* Entropy from state probabilities: hydration entropy of cations. *J Phys Chem B* **117,** 6466–6472, doi: 10.1021/jp311418q (2013).
48. Huber, R. G., Eibl, C. & Fuchs, J. E. Intrinsic flexibility of NLRP pyrin domains is a key factor in their conformational dynamics, fold stability, and dimerization. *Protein Sci* **24,** 174–181, doi: 10.1002/pro.2601 (2015).

Acknowledgements

Author Contributions

Study design and supervision: A.L.P. Development of methodology: E.M.C., R.P.M., J.E.F, E.S. and A.L.P. Protein expression and purification: E.M.C., E.P.G. and A.L.P. Biophysical studies: A.L.P. Proteolysis experiments: E.M.C., E.P.G., N.M.T. and A.L.P. Experiments in cell cultures: E.M.C and R.P.M. Molecular dynamic simulations: J.E.F. Data analysis and interpretation: R.P.M., J.E.F. and A.L.P. Drafting of the manuscript: J.E.F., D.J.T. and A.L.P. Revision of the manuscript: A.L.P. with contribution from the rest of the authors

Additional Information

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

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

How to cite this article: Encarnación, M.-C. *et al.* Conformational dynamics is key to understanding loss-of-function of NQO1 cancer-associated polymorphisms and its correction by pharmacological ligands. *Sci. Rep.* **6**, 20331; doi: 10.1038/srep20331 (2016).

PUBLICATION 2

Site-to-site interdomain communication may mediate different loss-of-function mechanisms in a cancer-associated NQO1 polymorphism.

SCIENTIFIC REPORTS

Site-to-site interdomain communication may mediate different loss-of-function mechanisms in a cancer-associated NQO1 polymorphism

Encarnación Medina-Carmona[1], Jose L. Neira[2,3], Eduardo Salido[4], Julian E. Fuchs[5], Rogelio Palomino-Morales[6], David J. Timson[7] & Angel L. Pey[1]

Disease associated genetic variations often cause intracellular enzyme inactivation, dysregulation and instability. However, allosteric communication of mutational effects to distant functional sites leading to loss-of-function remains poorly understood. We characterize here interdomain site-to-site communication by which a common cancer-associated single nucleotide polymorphism (c.C609T/p. P187S) reduces the activity and stability *in vivo* of NAD(P)H:quinone oxidoreductase 1 (NQO1). NQO1 is a FAD-dependent, two-domain multifunctional stress protein acting as a Phase II enzyme, activating cancer pro-drugs and stabilizing p53 and p73α oncosuppressors. We show that p.P187S causes structural and dynamic changes communicated to functional sites far from the mutated site, affecting the FAD binding site located at the N-terminal domain (NTD) and accelerating proteasomal degradation through dynamic effects on the C-terminal domain (CTD). Structural protein:protein interaction studies reveal that the cancer-associated polymorphism does not abolish the interaction with p73α, indicating that oncosuppressor destabilization largely mirrors the low intracellular stability of p.P187S. In conclusion, we show how a single disease associated amino acid change may allosterically perturb several functional sites in an oligomeric and multidomain protein. These results have important implications for the understanding of loss-of-function genetic diseases and the identification of novel structural hot spots as targets for pharmacological intervention.

Native proteins are heterogeneous ensembles sampling a wide variety of microscopic conformations[1]. The relative population of these states is determined by their intrinsic stabilities (i.e. *free energies*) as well as their rates of interconversion (i.e. *kinetic barriers*)[1,2]. Large-scale and slow collective motions (in the ms to s time scales) determine the transitions between a small number of low energy states that correspond to different *conformational states* (with intrinsically different structure and energetic balance)[1–4]. Within these low energy states, small-scale fluctuations rapidly occur, such as backbone and side-chain dynamics, which are the main source of *conformational dynamics*[1,3,4]. A deep structural, thermodynamic and kinetic understanding of collective and local protein motions is essential to understand protein function, regulation, degradation and evolution[4–10].

Protein allostery is the phenomenon that allows communication between two distant sites in a protein (e.g. two binding sites for different ligands or between a ligand binding and a mutated/post-translational modification

[1]Department of Physical Chemistry, Faculty of Sciences, University of Granada, Av. Fuentenueva s/n, 18071, Granada, Spain. [2]Instituto de Biología Molecular y Celular, Universidad Miguel Hernández, Avda. del Ferrocarril s/n, 03202, Elche, Alicante, Spain. [3]Instituto de Biocomputación y Física de los Sistemas Complejos (BIFI), 50009, Zaragoza, Spain. [4]Hospital Universitario de Canarias, Centre for Biomedical Research on Rare Diseases (CIBERER), Tenerife, Spain. [5]Institute of General, Inorganic and Theoretical Chemistry, Faculty of Chemistry and Pharmacy, University of Innsbruck, Innsbruck, Austria. [6]Department of Biochemistry and Molecular Biology I, Faculty of Sciences, University of Granada, Av. Fuentenueva s/n, 18071, Granada, Spain. [7]School of Pharmacy and Biomolecular Sciences, The University of Brighton, Brighton, UK. Correspondence and requests for materials should be addressed to A.L.P. (email: angelpey@ugr.es)

site)[11]. Ligand binding remodels protein conformational ensembles by stabilizing binding competent states from such a conformationally heterogenous pre-existing equilibrium. The consequent population-shift of microstates is *macroscopically* observed as a *conformational change*[11–14]. Accordingly, structural and dynamic aspects of these conformational changes have significant impact on protein function and regulation (e.g. on ligand binding affinity and energetics)[2–4,12,15]. A combination of functional, structural and thermodynamic analyses is a powerful approach to identify allosteric networks and pathways in proteins and to distinguish between different allosteric mechanisms (e.g. *induced-fit* vs. *conformational selection*)[4,12,14,16].

Disease-associated inherited changes in protein sequence (i.e. mutations and polymorphisms) affect protein activity, regulation and stability[17–22]. However, allosteric site-to-site communication underlying loss-of-function genetic variations has rarely been investigated in multi-domain oligomeric proteins[23–25]. Here, we hypothesize that the cancer-associated single nucleotide polymorphism p.P187S (rs1800566/c.C609T) in the NADP(H):quinone oxidoreductase 1 (NQO1; EC 1.6.5.2) mediates enzyme loss-of-function through long-range site-to-site effects originating at the p.P187S site and communicated to distant functional sites through a hypothetical *allosteric interaction network* (Fig. 1). NQO1 is a two-domain FAD-dependent enzyme (Fig. 1A,B) involved in the two-electron reduction of quinones, important for Phase II detoxification reactions, superoxide scavenging and activation of certain cancer pro-drugs[26]. The N-terminal domain (NTD) binds FAD and the smaller C-terminal domain (CTD, residues 225–274) is structurally involved in binding of NADH, substrates and competitive inhibitors (e.g. dicoumarol) (Fig. 1B,C). NQO1 also interacts with and stabilizes cancer-associated transcription factors such as p53 and p73α[27–30], even though the structural basis of this intracellular interaction is unknown. Although p.P187S strongly reduces NQO1 levels and activity *in vivo*, crystallographic analyses of holo-p.P187S have shown no significant effect on functional or structural sites such as FAD binding site, the CTD or the monomer:monomer interface[31]. However, molecular dynamics (MD) simulations, biophysical experiments and expression analyses have supported the hypothesis that p.P187S inactivates and destabilizes NQO1 through dynamic changes in the apo-state, in particular on two functionally and structurally distant sites (Fig. 1D): the FAD binding site, associated with enzyme inactivation; and the CTD, associated with enhanced proteasomal degradation of p.P187S[22,24,31].

In this work, we investigate how the p.P187S polymorphism affects the FAD binding site and the CTD of NQO1 through long-range communication of structural and dynamic perturbations (Fig. 1D), as well as their potential effects on the binding site of oncosuppressors. To do so, we characterize wild-type (WT) and p.P187S NQO1 in their full-length and C-terminal truncated (Δ50-NQO1) forms using a multi-disciplinary approach. Overall, we show how a single amino acid substitution may cause protein loss-of-function by simultaneously affecting multiple functional sites in a complex oligomeric and multidomain protein with metabolic and regulatory roles.

Results

The CTD of NQO1 plays an important role in the conformational equilibrium of the apo-state.

Structural analyses[32] have shown that the CTD forms part of NQO1 dimer interface (Fig. 1A,B), adding up to 35% of the total buried surface at this interface (960 Å^2 of 2700 Å^2 per monomer; calculated using cpptraj[33]). Calculations using the PISA server (http://www.ebi.ac.uk/pdbe/pisa/[34]) reveal that at least one residue of the CTD is involved in 40% of the hydrogen bonds and half of the salt bridges at the dimer interface. In addition, the CTD forms part of the dicoumarol binding site (Phe232 and Phe236; Fig. 1C), but it does not seem essential for FAD binding or stable folding of the NTD (Figure S1[31]). While Pro187 is not close to the FAD or dicoumarol binding sites, its shortest distance to the CTD is about 4 Å (with Gln268 and Ile269) (Fig. 1B,D). Therefore, it is unclear whether deletion of the CTD affect the dimerization, activity and stability of NQO1, as well as the structural and dynamic effects caused by p.P187S on the FAD binding site and the CTD (Fig. 1C,D and ref. 24).

NQO1 WT and p.P187S in their full-length and C-terminal deleted (lacking the last 50 residues at the C-terminus, Δ50-NQO1) versions were expressed in and purified from *E. coli*. To investigate their oligomeric state, their hydrodynamic behavior was analyzed by size-exclusion chromatography (SEC) and dynamic light scattering (DLS) (Figure S2A and S2B). Our results show that all NQO1 variants behave as dimers in solution even at concentrations ≤1 μM (note that dimers of NQO1 and Δ50-NQO1 are expected to have a molecular weight of 65.2 and 53.9 kDa, respectively). The 1-1 echo pulse sequence of NQO1 WT and P187S mutant were also used to measure T_2 relaxation times of the most down-field isolated amide protons of both spectra, yielding a molecular weight of approximately 78 kDa (at 20 μM concentration in monomer units; see Supplementary information). We attempted to detect dimer dissociation and to determine their dissociation constants by isothermal titration calorimetry (ITC; Figure S2C). However, dilution of NQO1 dimer caused little or no heat change compared to the blank (i.e. buffer) dilution heats, thus supporting that dissociation of dimers in the low micromolar range initially used during titrations occurs to little or negligible extent. This result is important considering that the amount of surface exposed to the solvent upon dimer dissociation (expected to be about 5400 Å^2; see above) would give a remarkable heat release, in the 10–15 kcal·mol^{-1} range at 25 °C, using the previously published structure-energetics correlations[35].

Remarkably, NQO1 proteins as purified show different amounts of FAD bound, with WT having the highest content, Δ50-WT showing intermediate levels (about 4-fold lower than for WT), and p.P187S and Δ50-p.P187S displaying negligible levels of bound FAD, as judged from their corresponding absorption near-UV/visible spectra (Fig. 2A). To investigate their overall conformation, we have prepared apo-proteins and then added precise concentrations of ligands to achieve different ligation states (holo-NQO1, with FAD, and holo-NQO1+ dicoumarol, with an excess of FAD and dicoumarol). The far-UV circular dichroism (CD) spectra of full-length and Δ50 apo-proteins show a high content of ordered secondary structure (Fig. 2B–E). Binding of FAD alters the secondary structure content, increasing the ellipticity at 222 nm (a hallmark for α-helical content) by 20–25% in all cases except for p.P187S in which this increase is of 10% (Fig. 2F,G; note that in all cases, a statistical comparison of a given variants as holo- vs. apo-protein yields a p value < 0.05 using a t-test). Binding of dicoumarol, that interacts
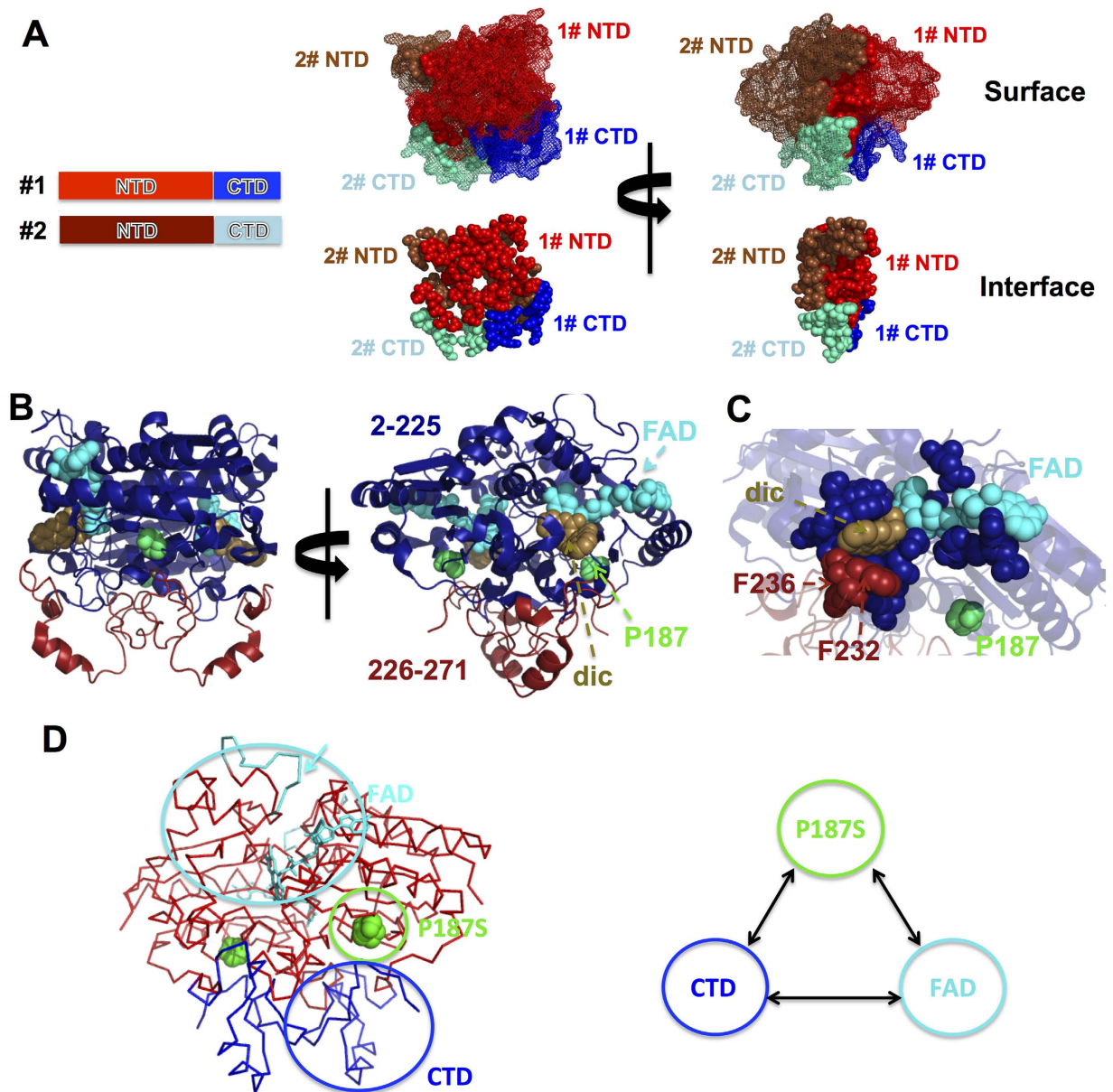
Figure 1. NQO1 structure, dimerization and functional definition of the hypothetical allosteric interaction network. (**A**) Two views of the NQO1 dimer showing the protein surface and the dimerization interface (as calculated using the PISA server[34]); the N-terminal domains (NTD) are displayed in red (monomer #1) and brown (monomer #2) while the C-terminal domains (CTD) are displayed in blue (monomer #1) and cyan (monomer #2); (**B**) Two views of the NQO1 dimers displaying secondary structures of NTD (residues 2-225, in blue) and the CTD (residues 226–271, in red). FAD, dicoumarol (dic) and P187 are shown in cyan, light brown and green, respectively. (**C**) Close-up view of the FAD and dicoumarol binding sites, with the residues belonging to the NTD (blue) and CTD (red) highlighted. (**D**) Schematic representation of the hypothesized allosteric interaction network involving the p.P187S and FAD binding sites and the CTD. In the FAD binding site (left panel), the highly dynamic loop 57–66 in the apo-state of p.P187S is indicated with a cyan arrow to highlight its far location from the CTD and p.P187S sites. Structures were displayed with Pymol[60] using the NQO1 structure (PDB:2F1O).

with both NTDs and CTDs in full-length NQO1, leads to further increase of 16% in holo-p.P187S ($p < 0.05$) but not in holo-WT, while it causes only small increments (6–9%) in Δ50 variants (Fig. 2F,G). Since apo-NQO1 intrinsically populates partially unfolded conformations[22,24,31,36], these results suggest that binding of FAD and dicoumarol to full-length and Δ50 variants leads to a shift in the conformational equilibrium of NQO1 towards more compact, ligand binding competent conformations. Accordingly, DLS experiments showed a decrease in the hydrodynamic radii of WT and p.P187S when FAD or FAD + dicoumarol are added, respectively (Fig. 2H), further supporting a more compact average dimer conformation in the presence of these ligands. In the case of
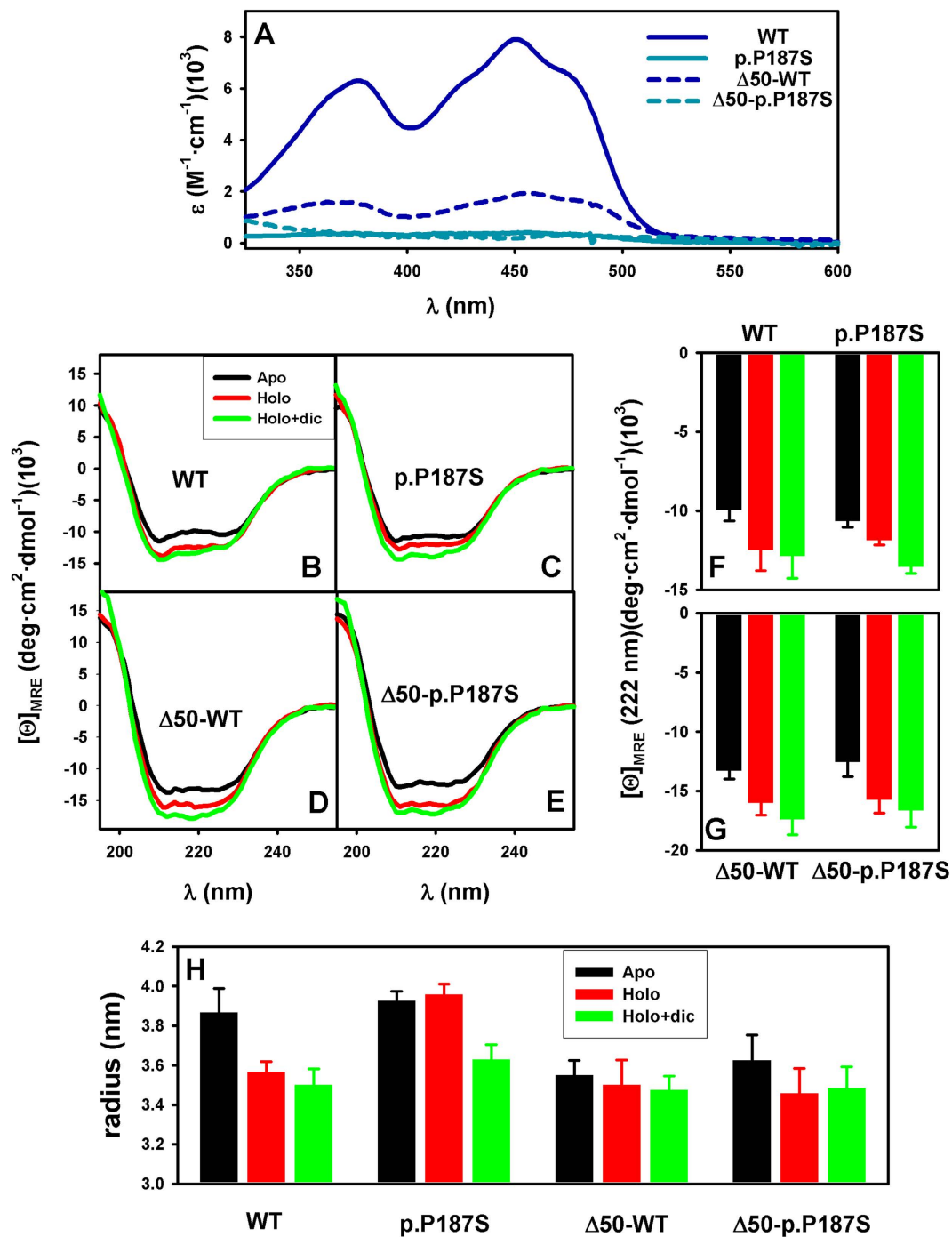
Figure 2. Conformational analysis of NQO1 variants. (**A**) FAD bound to the proteins as purified from Near-UV/visible absorption spectra. (**B**–**E**) Far-UV CD spectra of NQO1 variants as apo-proteins, in the presence of FAD (Holo) or FAD + dicoumarol (Holo + dic); (**F**,**G**) values of mean residue ellipticity ($[\Theta]_{MRE}$) at 222 nm from spectra shown in panels (**B**–**E**,**H**) hydrodynamic radius for NQO1 variants in the different ligation states. Data in panels A–G are the average from three independent experiments, while in panel H are from 3 to 6 independent experiments. Values in panels F, G and H are mean ± s.d. from replicates. In all cases, two to three different protein purifications were used, and the temperature was 25 °C.

the Δ50 variants, binding of FAD and/or dicoumarol led to much smaller effects on the hydrodynamic radius (Fig. 2H) indicating an important role of the CTD in the conformational equilibrium of apo-NQO1.

The dynamic CTD of p.P187S neither contributes to its low *in vitro* kinetic stability nor to ligand mediated stabilization of NQO1 dimers. p.P187S strongly decreases the *in vitro* thermal and kinetic stability of NQO1 by accelerating dimer unfolding and dissociation[22,24]. To investigate the role of the CTD on its
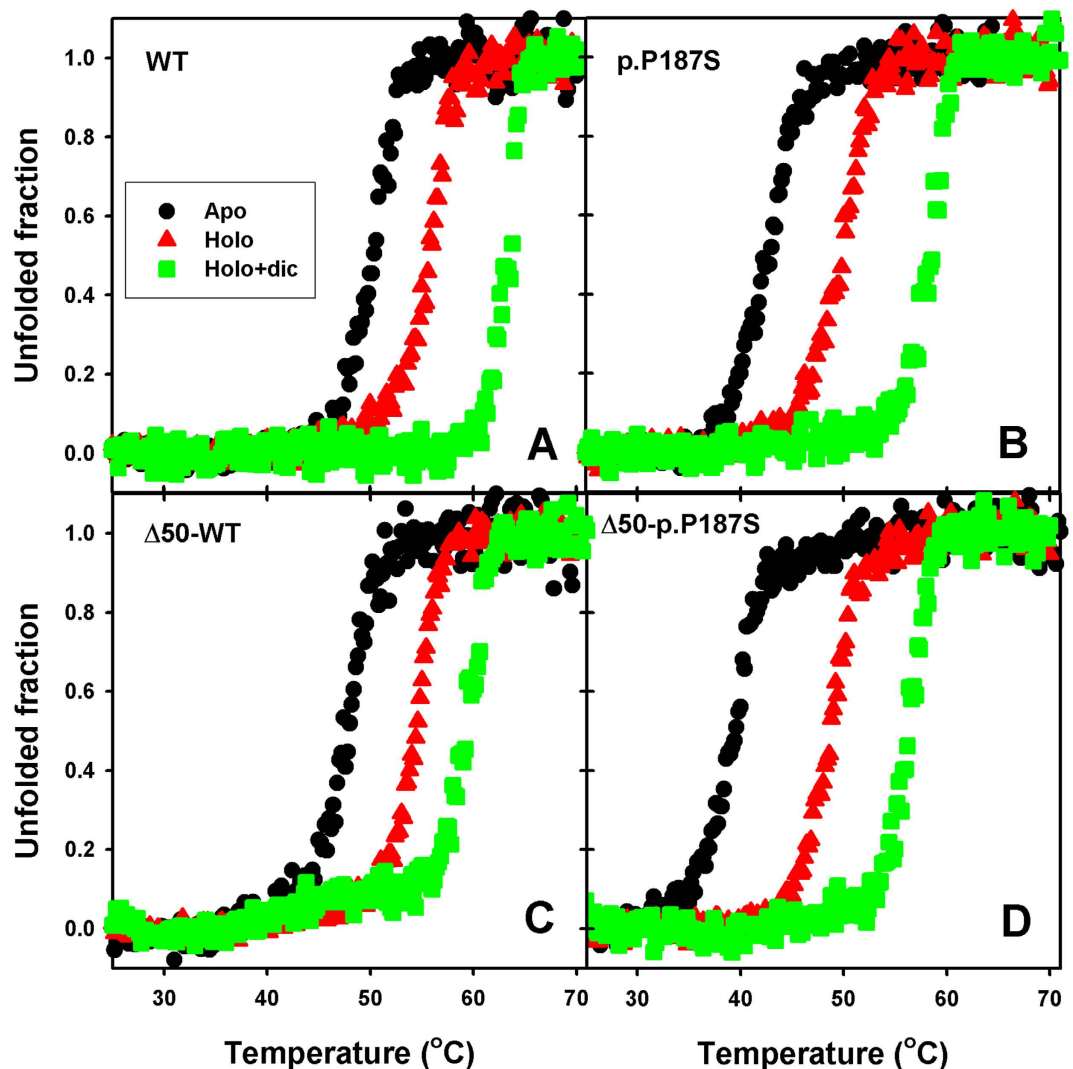
Figure 3. Thermal stability of NQO1 variants. Experiments were performed using NQO1 variants in different ligation states. Raw denaturation curves were normalized using linear pre- and post-transition baseline for sake of comparison.

Variant	T_m (°C)		
	Apo	**Holo**	**Holo + dic**
WT	50.4	55.7	63.7
p.P187S	42.6	49.9	58.7
Δ50-WT	47.9	54.5	59.3
Δ50-p.P187S	39.5	48.7	56.3

Table 1. Thermal stability of NQO1 and Δ50 variants. The apparent melting temperatures (T_m) are determined for proteins as apo-proteins, and in the presence of FAD or FAD + Dicoumarol.

low kinetic stability, we have performed thermal denaturation studies monitored by CD spectroscopy (Fig. 3). Removal of CTD causes a small 3 °C destabilization on the WT and p.P187S backgrounds (Fig. 3 and Table 1). A large excess of FAD (holo-proteins) causes a remarkable stabilization, ranging from 5 to 10 °C (Fig. 3 and Table 1), and an excess of FAD and dicoumarol have a roughly additive effect on stability (from 12–17 °C). These results imply that removal of CTD modestly destabilizes NQO1, and it is not the cause the low *in vitro* kinetic stability of p.P187S.

Partial proteolysis of full-length NQO1 proteins provides quantitative information on the effects of polymorphisms and ligand binding on local dynamics[24]. Partial proteolysis of Δ50 variants as apo-proteins showed a rapid decay of the native protein (Fig. 4A,B), and the appearance of 3–4 proteolysis products with molecular sizes of 16–19 kDa (Table S1) based on HPLC/ESI-MS. FAD binding causes a remarkable 10-fold increase in
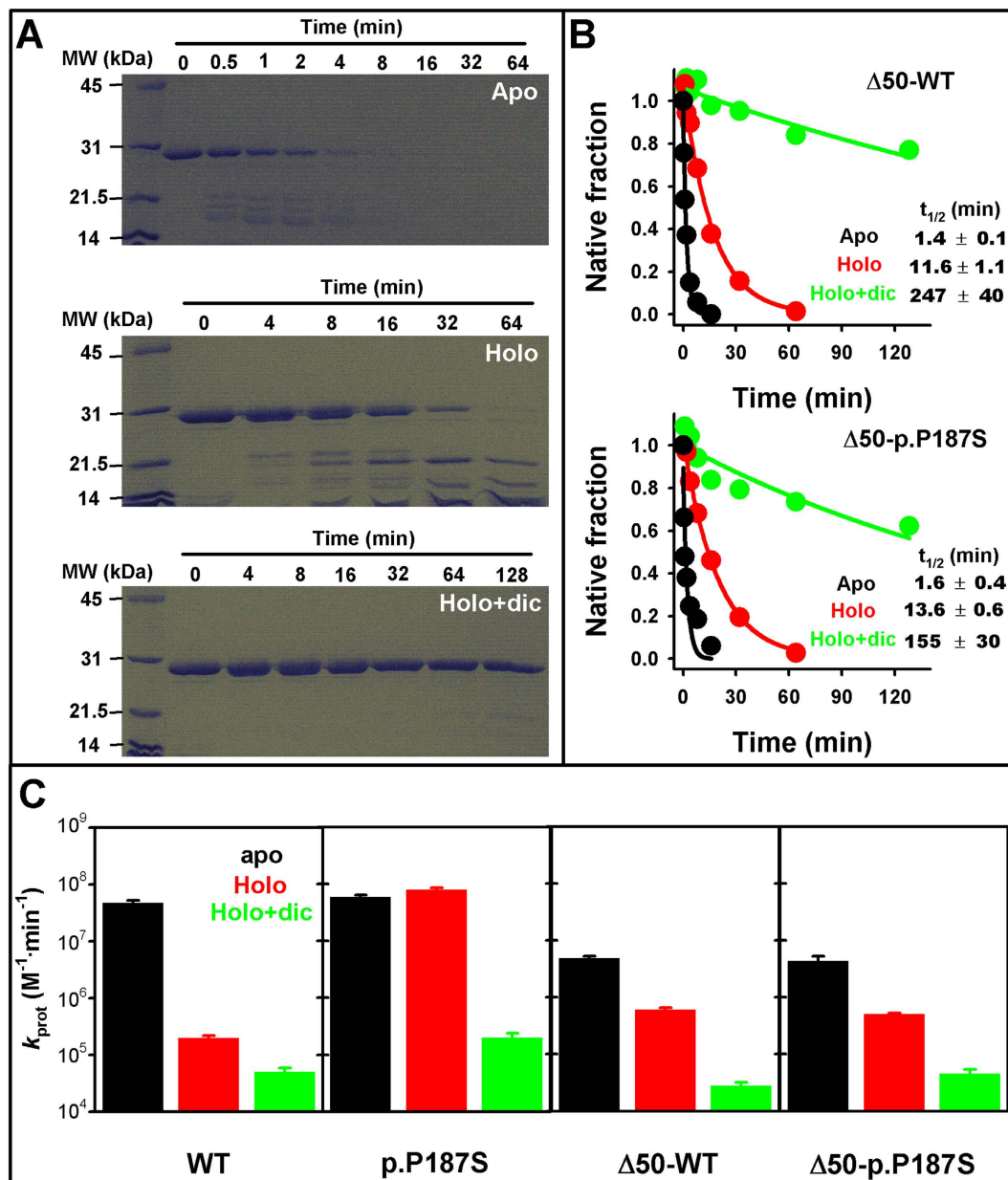
Figure 4. Partial proteolysis of Δ50-NQO1 variants. (**A**) representative SDS-PAGE gels of degradation kinetics of Δ50-WT in different ligation states (apo, holo and holo-holo + dic). The regions corresponding to molecular sizes from 45 to 14 kDa bands are shown for sake of clarity (see Figure S3 for additional pictures of gels); (**B**) Time-course of the degradation of the native band of Δ50 variants. Experiments performed at 25 °C in the presence of 0.1 μM of thermolysin. (**C**) proteolysis rate constants for full-length and Δ50-NQO1 variants in different ligation states (data for full-length variants are from ref. 24).

the resistance towards degradation (Fig. 4A,B) and the accumulation of 18 and 15 kDa partially proteolyzed species (Table S1). Binding of dicoumarol to holo-proteins leads to a 10–20 fold additional stabilization of the native Δ50 protein (Fig. 4A,B). Peptide fingerprinting of digested samples by MS/MS revealed that cleavage is initiated in the C-terminal part of Δ50-NQO1, involving several active cleavage sites in the region 150–170 (particularly dynamic in the apo-state of Δ50-NQO1, see our MD simulations below) leading to the formation of a stable intermediate comprising residues 1–143 (Figs 4A and S3). Our comparative analyses of proteolysis kinetics supports our previous proposal of a very dynamic CTD in the apo-state of NQO1, that remains dynamic in the holo-p.P187S unless dicoumarol is bound (Fig. 4C). We must note that an absolute comparison of local dynamics between NQO1 and Δ50 variants is not possible because experimental proteolysis rate constants are proportional to the sequence-dependent intrinsic proteolysis rate constants for primary cleavage sites under the conditions used.
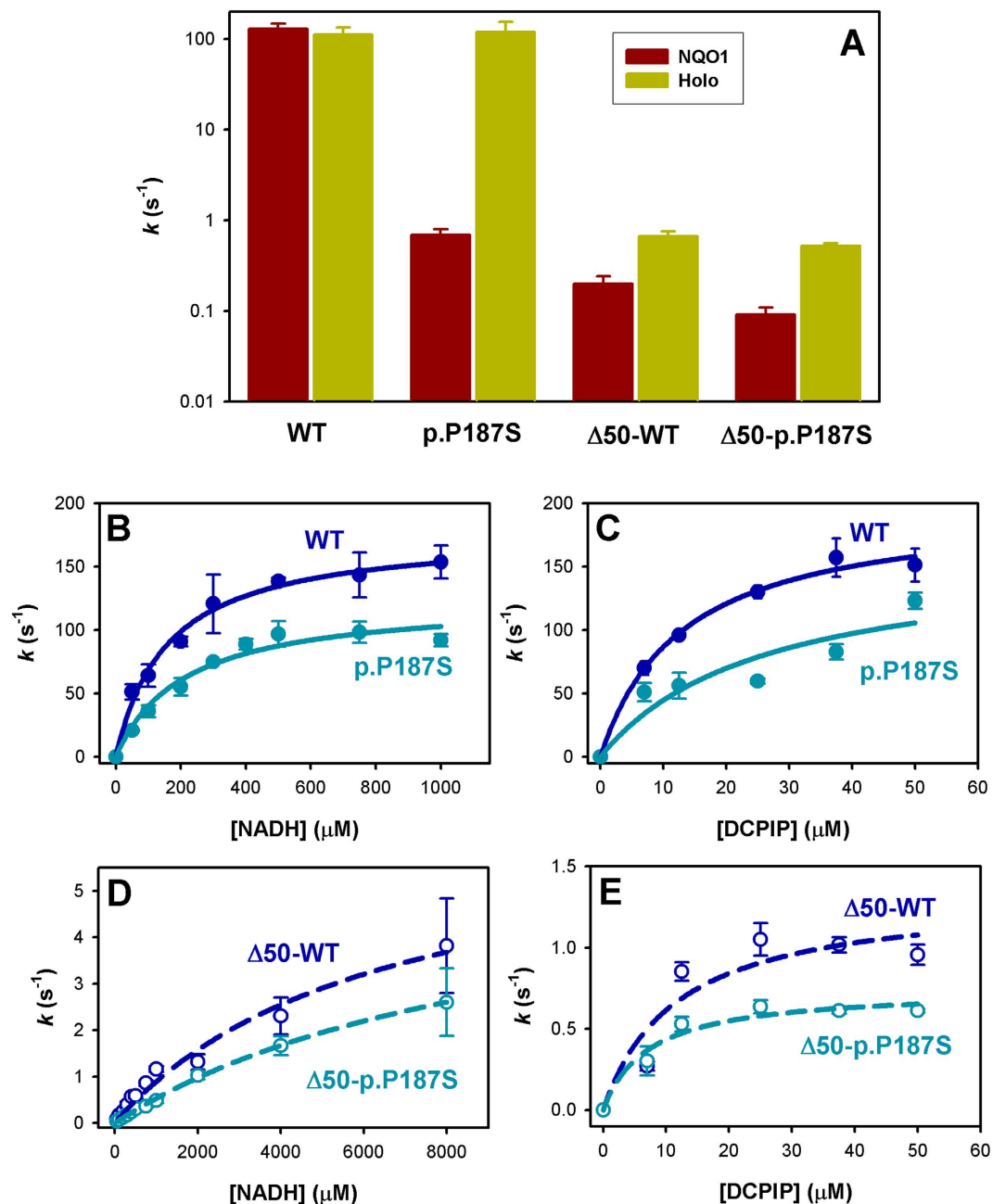
Figure 5. Enzyme kinetic analyses of NQO1 variants. (**A**) Activity measurements in the absence (NQO1) or the presence (Holo) of a FAD excess, using 70 μM DCPIP and 1 mM NADH. Please, note the logarithmic scale of the y-axis. (**B**) Activity dependence on NADH (**B,D**) and DCPIP (**C,E**) concentrations for WT/p.P187S (**B,C**) and Δ50-WT/Δ50-p.P187S (**D,E**). Lines are fits to the Michaelis-Menten equation.

The CTD of p.P187S undergoes a large conformational rearrangement upon dicoumarol binding and contributes to long-range dynamic effects affecting FAD binding. The low intracellular specific activity of p.P187S is largely associated with defective FAD binding (Fig. 5A and ref. 24). Removal of the CTD causes a large decrease in specific activity, that arises from a reduction in catalytic performance rather than in FAD content, since saturation with FAD only causes a modest increase in activity (Fig. 5A). The enzyme kinetic parameters for NADH and the substrate DCPIP (2,6-dichlorophenolindophenol) of holo-enzymes have been determined (Fig. 5B–E), revealing little effect of p.P187S on catalytic performance, while removal of the CTD leads to a 30-to-100-fold reduction in specific activity, and importantly, to a decrease in the apparent affinity for NADH by 30-to-50-fold (Fig. 5B–E and Table 2). These results confirm that p.P187S and deletion of the CTD cause NQO1 inactivation through different mechanisms: the former by reducing FAD binding affinity and the latter by strongly decreasing catalytic performance.

At this stage, we hypothesized that the CTD may particularly affect the catalytic performance of p.P187S *in vivo* due to long-range communication with FAD binding sites (Fig. 1D). NQO1 WT displays 10–100 higher

NQO1 variant	NADH		DCPIP	
	k_{cat} (s^{-1})	K_M (µM)	k_{cat} (s^{-1})	K_M (µM)
WT	177±7	158±21	199±11	13±2
p.P187S	124±10	210±48	160±58	26±20
Δ50-WT	6.6±1.0	6422±1668	1.3±0.3	12±8
Δ50-p.P187S	5.8±0.3	9985±797	0.74±0.07	7±3

Table 2. Enzyme kinetic parameters for holo NQO1 and Δ50 variants.

affinity for FAD than that of p.P187S and some degree of negative cooperativity[22], making a robust comparison of their binding energetics difficult. Nevertheless, removal of the CTD abolishes the negative cooperativity of NQO1 WT (as seen by the good description of their binding isotherms using a non-cooperative binding model; Figure S4A) and also the deleterious effect of p.P187S on FAD binding, with WT and p.P187S truncated forms showing similarly high binding affinities and thermodynamic signatures (Fig. 6A,B). Notably, removal of the CTD increases the affinity of p.P187S by 4-fold due to a modest favourable entropic contribution to binding (Fig. 6C). These results reveal a contribution of the CTD to FAD binding affinity, playing opposite roles in WT and p.P187S, probably through long-range site-to-site communication between the p.P187S and *functional* CTD and FAD binding sites. This is particularly interesting because removal of CTD has little or no structural and dynamic effects on the FAD binding mode of holo-p.P187S[31] (Figure S1 and the next section), thus indicating that communication between the p.P187S and functional sites mainly operates on the conformational equilibrium of the apo-state of p.P187S. Furthermore (see the next section), the entropic origin of the lower FAD binding affinity of p.P187S supports that it primarily arises from changes in protein conformational dynamics (i.e. *conformational entropy*, see Fig. 6C and the next section).

As purified, p.P187S binds dicoumarol with 90-fold lower affinity than NQO1 WT (Fig. 6D). This is mostly due to a much lower favourable enthalpic contribution, largely compensated by a more favourable entropic change (Fig. 6E). This large difference in affinity is likely explained by the much lower FAD content of p.P187S as purified. According to earlier crystallographic analyses and MD simulations[24,32], dicoumarol binding must be favoured by bound FAD (i.e. in the holo- vs. the apo-state), primarily through stronger (enthalpy-driven) interactions that would be counterbalanced by a large decrease in protein dynamics (i.e. loss of conformational entropy). Consistently, binding of dicoumarol to the holo-enzymes (i.e. saturated with FAD) shows 5-fold (WT) and 75-fold (p.P187S) higher affinities than those found for the proteins as purified (Figs 6F and S4B). These results show the important structural and energetic roles of FAD in the efficient binding of dicoumarol, and also indicate differences in dicoumarol binding to the holo-forms of p.P187S and WT. Indeed, binding of dicoumarol to holo-p.P187S shows a much larger favourable enthalpic contribution (about -21 kcal·mol^{-1}) which is compensated by a large decrease in entropy (Fig. 6G). A plausible explanation for this could be a large loss of solvation/conformational entropy counterbalanced by the structural reorganization of the protein, particularly at the highly dynamic CTD of p.P187S. In fact, simple structure-energetic correlations for protein folding[35] provides an enthalpy change at 25 °C of $\sim$0.21 kcal·mol^{-1}·(residue)$^{-1}$, consistent with a much larger conformational rearrangement (roughly a hundred residues) in p.P187S upon dicoumarol binding. According to this, dicoumarol binding to holo-Δ50-WT and holo-Δ50-p.P187S is very similar (Figure S4C), confirming that dicoumarol binding to the full-length holo-p.P187S is associated with large structural and dynamic rearrangements at its CTD (Fig. 6G,H).

Molecular dynamics (MD) simulations support the role of the CTD in the communication between p.P187S and FAD binding sites.

Molecular dynamics (MD) simulations have shown that p.P187S affects fast internal protein dynamics through long-range effects affecting the FAD binding site and the CTD[24]. To investigate whether CTD removal could intefere with the dynamic communication between p.P187S and the FAD binding sites, we performed MD simulations on Δ50-NQO1 variants, and analyzed their backbone dynamics using two relevant parameters[24]: B-factors (Fig. 7) and dihedral entropies (Figure S5).

Removal of the CTD increases the average flexibility of NQO1, particularly at the dimer interface, consistent with its modest *in vitro* kinetic destabilization, while thermal stabilization upon FAD and dicoumarol binding correlates well with ligand mediated reduction of global dynamics (Figs 7A,B and S5A-B). Interestingly, the decreased binding affinity for FAD in Δ50-NQO1 WT seems to originate from larger conformational fluctuations in the FAD binding site of the apo-state (at residues 57–66, 125–136 and 150–166; Figs 7C and S5C). Importantly, we observe that the CTD is essential for p.P187S to enhance local dynamics in the FAD binding site of the apo-state, particularly at the 57–66 loop (Fig. 7D). Therefore, our MD simulations explain at atomic resolution how removal of the CTD significantly enhances the binding affinity of p.P187S by a modest reduction of binding (conformational) entropy. Therefore, the CTD has opposite effects on the FAD binding site dynamics of WT and p.P187S in the apo-state (Fig. 7C,D), in agreement with our experimental observations (ref. 22 and Fig. 6A–C). Accordingly, CTD truncated forms of WT and p.P187S show similar FAD binding energetic and dynamic signatures (Figs 6B, 7E and S5E).

Regarding dicoumarol binding, holo-p.P187S binds the inhibitor with two-fold lower affinity than the truncated version due to a large entropic penalty counterbalanced by a large favorable enthalpic contribution (Fig. 6F–H). This can be explained *via* a large conformational restriction (i.e. a decrease in conformational entropy) of the region 231–236 upon binding (that includes Phe232 and Phe236, which directly interact with dicoumarol; Figs 1C, 7G, S5G and ref. 24). Moreover, this large conformational restriction, in turn, is enthalpically
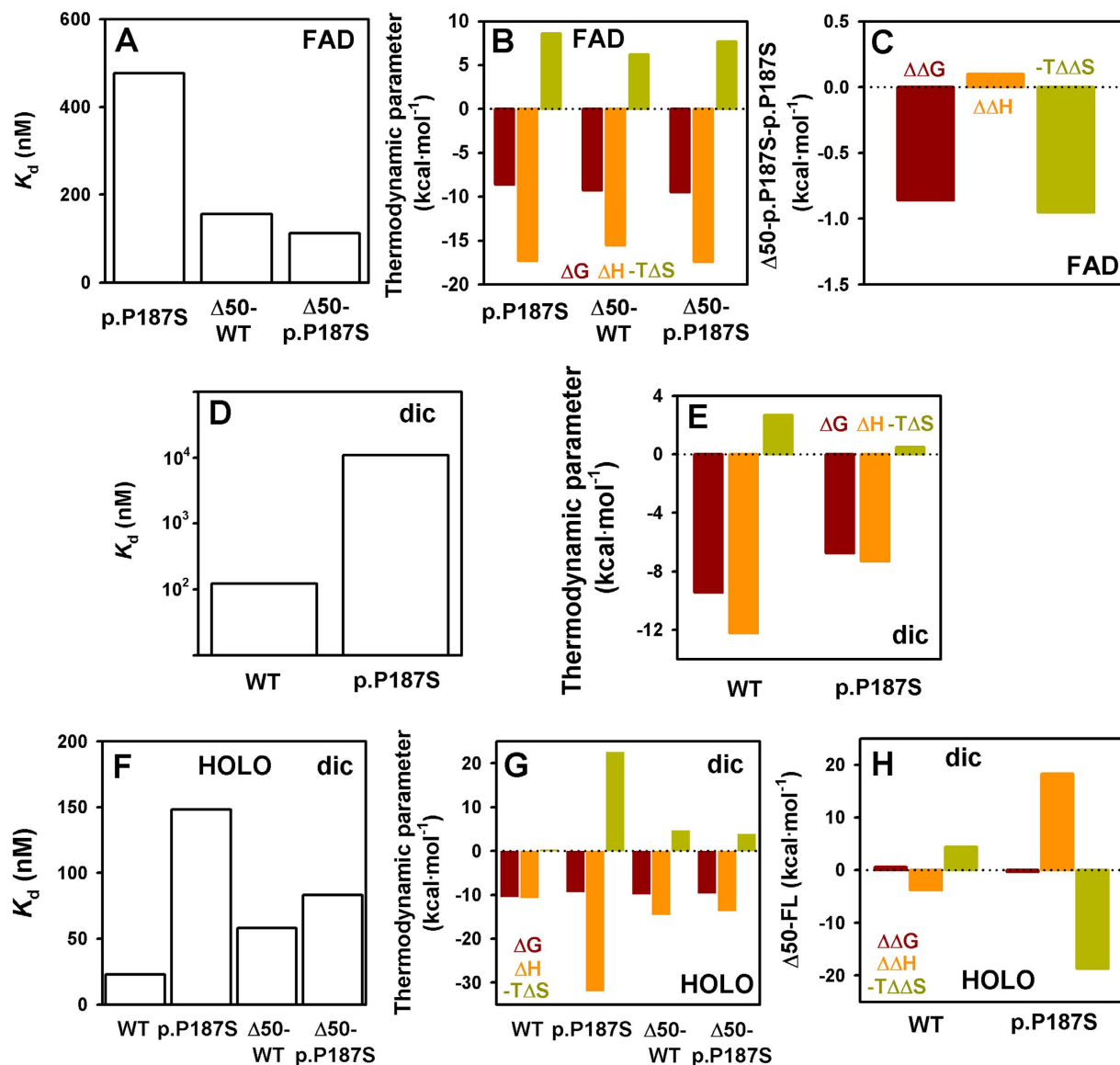
Figure 6. Thermodynamics of FAD and dicoumarol binding to NQO1 proteins. (**A**–**C**) Binding affinity (**A**) and thermodynamic profiles for FAD binding (**B**). In panel C, the effect of C-terminal withdrawal in thermodynamic binding parameters for p.P187S is shown. (**D,E**) Dicoumarol binding affinity (**D**) and thermodynamic binding profiles (**E**) for WT and p.P187S as purified. (**F**–**H**) Binding affinity (**F**) and thermodynamic profiles for dicoumarol binding (**G**) to holo-proteins. In panel H, the effects of C-terminal deletion in thermodynamic binding parameters are displayed.

compensated by the additional interactions established in the dicoumarol binding site at the CTD of p.P187S and the conformational change triggered upon binding.

Structural analyses of the interaction of NQO1 variants with the SAM domain of p73α show that an active NQO1 is not required for binding. Previous biochemical experiments have supported a role for the CTD in the interaction of NQO1 with oncosuppressors such as p73α. These interactions seem to be strengthened in the presence of NADH bound to NQO1 (even though this state is expected to be unstable, undergoing FAD reduction to $FADH_2$ and subsequent NAD^+ release), and may be inhibited by dicoumarol[30] but not by the mechanism-based inhibitor ES936 which also interacts with the CTD[32]. Truncation of the CTD (residues 209–274) or the presence of the intracellularly unstable p.P187S also seem to weaken these protein:protein interactions[29,37].

In addition to the transactivation domain, p73 also contains a sequence-specific DNA binding domain and an oligomerization region. Particularly, the p73α splicing variant has an extended C-terminus containing a sterile alpha motif (SAMp73) capable of repressing the function of p73 transactivation domain[38], and seemingly involved in the interaction with NQO1[30]. Therefore, we carried out NMR (HSQC) measurements on the interactions between NQO1 and SAMp73 using purified proteins. In the presence of NQO1 WT, the HSQC spectra of
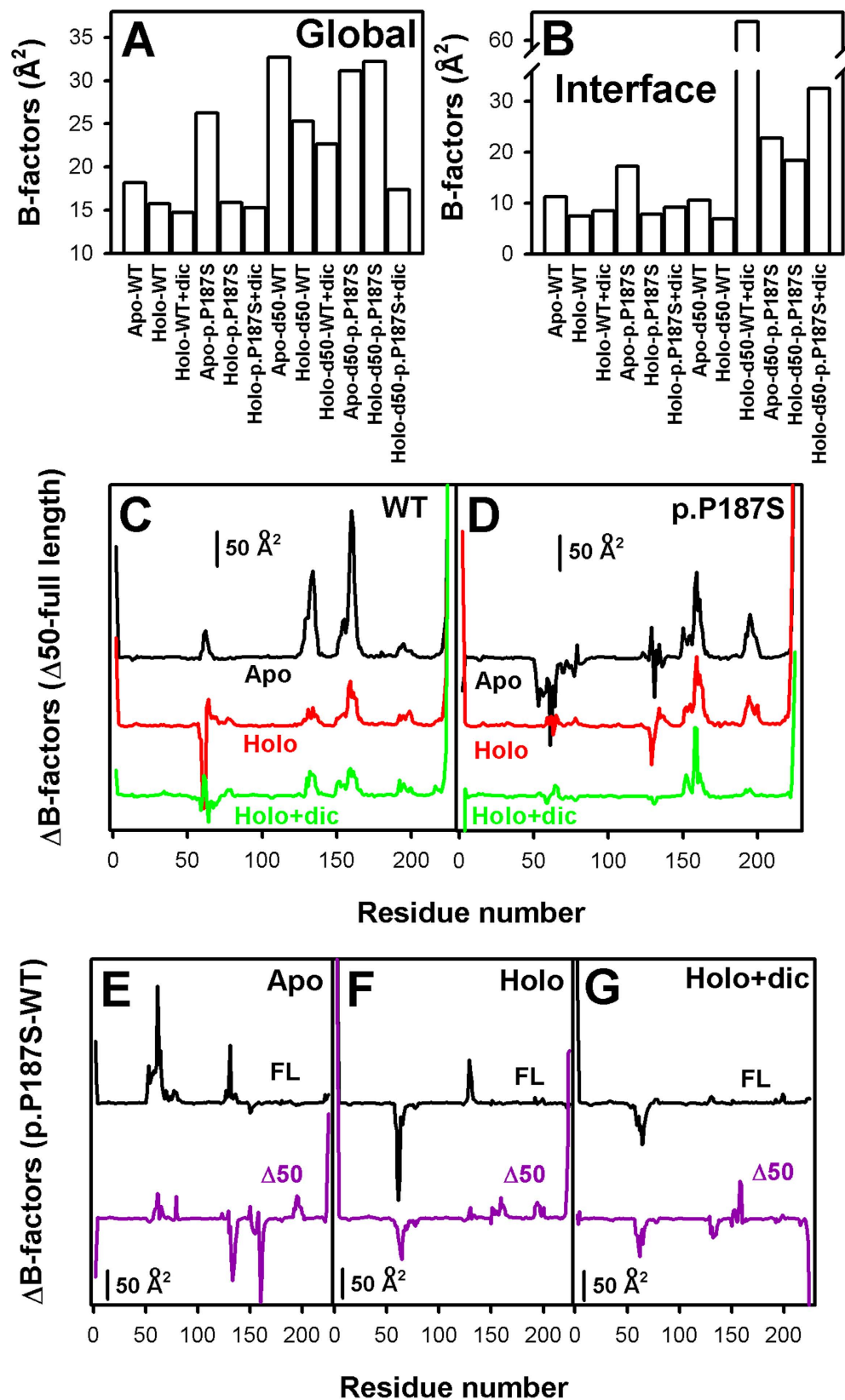
Figure 7. Dynamic effects (B-factors) of CTD truncation of NQO1 and ligand binding investigated by MD simulations. (**A**,**B**) Effects on the global (**A**) and dimer interface (**B**) dynamics of full-length and truncated NQO1 variants in different ligation states (apo, holo, and holo + dicoumarol); (**C**,**D**) Difference between C-terminal truncated and full length dynamics for WT (**C**) and p.P187S (**D**) in different ligation states; (**E**–**G**) Difference between p.P187S and WT as full-length and C-terminal truncated proteins in different ligation states ((**E**), apo; (**F**), holo; (**G**), holo + dicoumarol).

Figure 8. Structural analysis of the interaction between NQO1 and SAMp73. (**A**) Structure of SAMp73 (PDB: 1COK) showing the residues (in sticks) which were affected (either by signal broadening or by a CSP $\geq$ 0.01 ppm) (BRMB number 4413) by the presence of NQO1 WT as purified. The α-helices of SAMp73 are displayed with different colours (red for the fir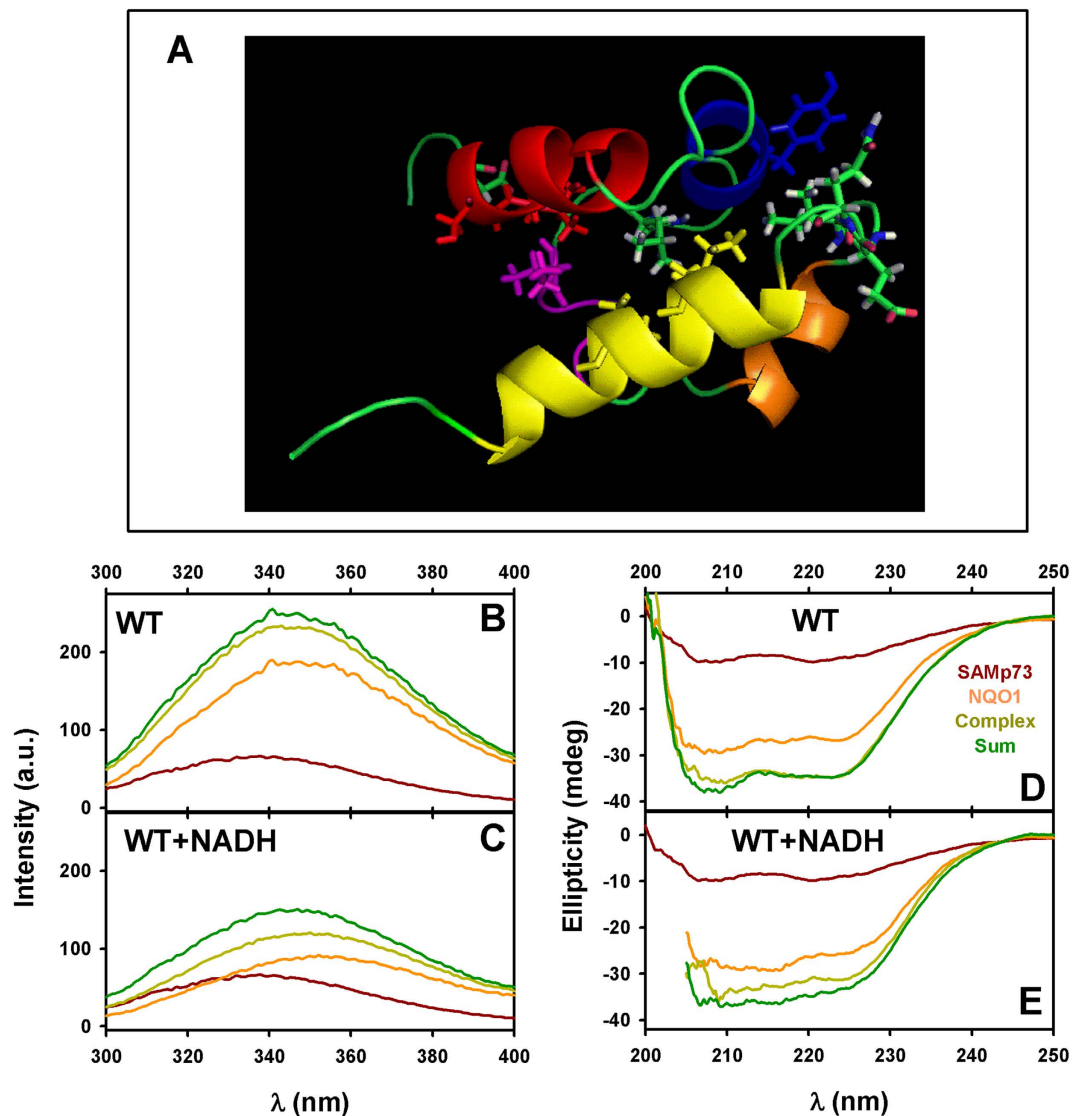st one; blue for the second; purple for the short 3_{10}-helix; gold for the fourth α-helix; and yellow for the fifth). The figure was produced with Pymol[60]. (**B,C**) Fluorescence spectra (upon excitation at 280 nm) of 2 μM SAMp73 or NQO1 WT, their sum, and the spectra of an equimolar mixture (2 μM of each protein in monomer units). (**D,E**) Far-UV CD spectra of 10 μM SAMp73 or NQO1 WT, their sum, and the spectra of an equimolar mixture (10 μM of each protein in monomer units). In (**B–E**), spectra in the absence or the presence of NADH are shown in different panels.

SAMp73 showed changes in the broadening or the chemical-shifts of the cross-peaks of a set of residues which are spatially close (Table S2), particularly in amino acids around Gly513 of SAMp73 and those involved in the N-terminal helix (Fig. 8A). Addition of NADH did not alter the set of residues affected by the presence of NQO1 WT, indicating that this interaction is not strictly NADH-dependent. Using fluorescence and far-UV CD spectroscopies (Fig. 8B–E), we also observed modest but reproducible structural changes upon NQO1:SAMp73 binding, and addition of NADH seemed to increase the extent of these structural changes. Attempts to provide quantitative data on the affinity for the interaction between NQO1 and SAMp73 by fluorescence and ITC were experimentally challenging, due to the apparent low affinity (possibly in the high micromolar range) and unstable calorimetric baselines (Figure S6 and data not shown). Since HSQC experiments showed that the spectrum of SAMp73 was not largely varied upon addition of NQO1, the changes observed are likely to occur in the environment of aromatic residues (Trp and Tyr) and the secondary structure of NQO1. Regarding SAMp73, Gly513 showed the largest changes in CSP (0.03) (Figure S7), while most of the affected cross-peaks showed broadening changes suggesting intermedium-slow kinetics for complex formation (within the NMR time scale), consistent with a low affinity interaction. Although the changes in chemical shifts observed were small, similar variations in
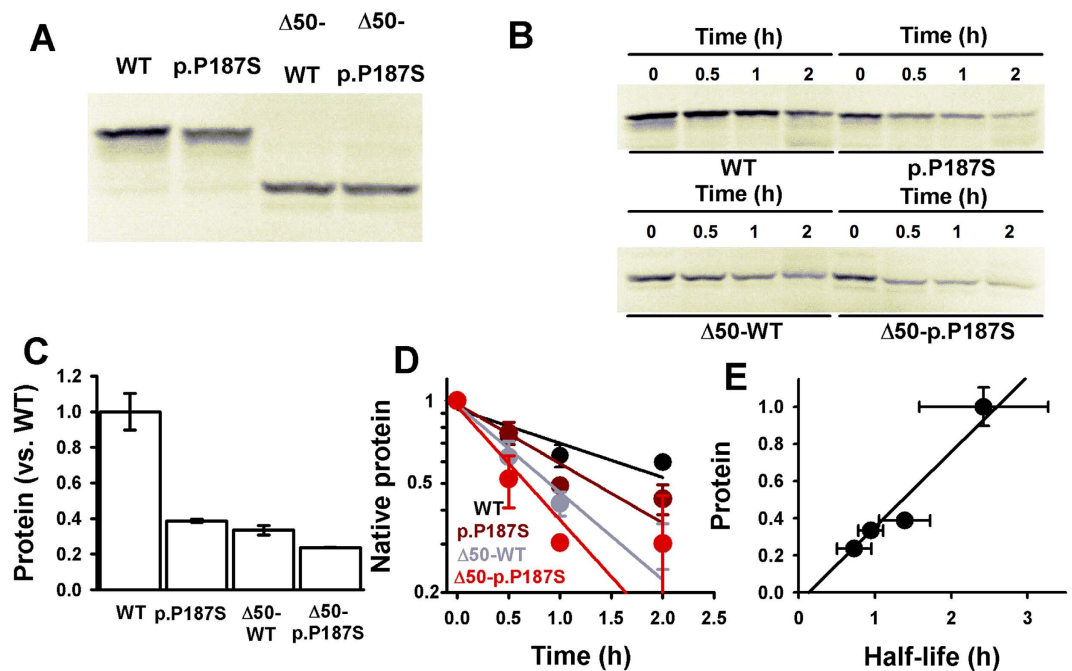
Figure 9. Pulse-chase experiments in rabbit reticulocytes. (**A**) Synthesis pulses of NQO1 variants. (**B**) Chase experiments. (**C**) Protein levels after synthesis pulses (normalized vs. NQO1 WT levels); (**D**) Semilogarithmic plots for chase kinetics; (**E**) correlation between protein levels after synthesis pulses and degradation half-lives. Data are mean ± s.d. from three independent experiments.

chemical shifts have been observed in other systems[39]. These results support that the interaction between NQO1 WT and SAMp73 is not strictly dependent on the presence of NADH, but instead, NADH binding and reduction of FAD might facilitate complex formation.

The interaction of p.P187S and Δ50-WT with SAMp73 caused similar broadening in a set of signals to those found for NQO1 WT (Table S2). Δ50-WT in the presence of NADH induced broadening of all cross-peaks, and even those of Tyr487 (Ile541), Asp490, Asn504 (Met539), Gly513, Ser516, Glu535, Leu545 and Gly551 disappeared completely (there is signal overlapping between Tyr487 and Ile541, and Asn504 and Met539). Thus, removal of the CTD, the presence of p.P187S or the absence of FAD bound (note that p.P187S is essentially an apo-protein as purified) do not abolish the interaction with SAMp73 but somewhat change the interaction mode. Interestingly, the binding mode of SAMp73 to NQO1 WT is also affected by the presence of dicoumarol, showing only small CSPs in a very reduced set of signals (Leu493, Gly513 and Trp542)(data not shown).

Overall, these biophysical analyses support that while different factors affecting the activity, structure and dynamics of NQO1 (including p.P187S, cofactors and inhibitors, or the CTD) may somewhat change the binding mode to SAMp73, this interaction is not prevented by any of the aforementioned factors. Moreover, our results also provide the first structural insights into the interaction between p73α and NQO1.

The CTD largely determines fast proteasomal degradation of p.P187S.
The highly dynamic CTD of p.P187S appears to act as an efficient initiation site for its proteasomal degradation[24]. Consistently, removal of the CTD had significant effects on protein levels and degradation rates upon expression in an eukaryotic expression cell free system (Fig. 9). In pulse synthesis, p.P187S showed 2.4-fold lower protein levels compared with WT as full-length proteins, while removal of the CTD (i.e. in Δ50 variants) caused a much larger decrease in WT protein levels than in p.P187S (3-fold vs. 1.5-fold; Fig. 9A–C). From chase experiments, we also observed a strong correlation between protein levels and degradation rates after a pulse of synthesis, supporting that proteasomal degradation rates strongly determine protein steady-state levels (Fig. 9B–E). Accordingly, p.P187S was degraded 1.75-fold faster than NQO1 WT (Fig. 9B,D), and removal of the CTD accelerated degradation of WT by 2.5-fold, but only that of p.P187S by 1.3-fold (Fig. 9B,D). These results show that removal of the CTD destabilizes NQO1 WT to a larger extent than the polymorphism, and therefore, the highly flexible CTD of p.P187S accelerates NQO1 proteasomal degradation. The enhanced degradation of Δ50-NQO1 is likely to be associated to its higher flexibility compared to that of full-length enzyme, as supported by our MD simulations (Fig. 7).

Discussion
In this work, we show that the CTD plays key functional and regulatory roles in NQO1, and remarkably, how it is crucial for the manifestation of the multiple deleterious effects of the cancer-associated p.P187S polymorphism. As expected from previous crystallographic evidence, the CTD is important for NQO1 activity, providing proper orientation of the coenzymes (FAD and NADH) and the substrate. Unexpectedly, we provide evidence for the existence of long-range communication between the p.P187S site and functional sites such as the CTD and the
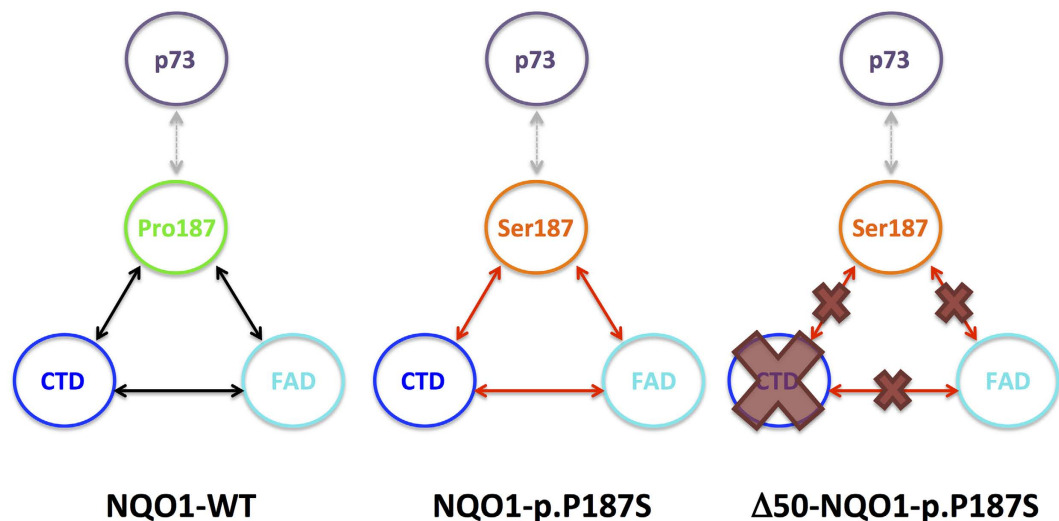
Figure 10. Schematic representation of the long-range communication between functional sites in NQO1 WT and p.P187S (*allosteric interaction network*). In NQO1 WT, the allosteric network between the P187 site, FAD binding site and CTD may contribute to the high-affinity cooperative binding of FAD and consequent structural and dynamic effects on the CTD upon binding. In NQO1 p.P187S, the S187 site affects FAD binding by changing the dynamics of its binding site and preventing structural and dynamic changes on CTD upon FAD binding. Removal of the CTD abolishes the dynamic effects of the S187 site on the FAD binding site. In all cases, alterations in this allosteric network seem to have a weak effect on the interaction site with p73α.

FAD binding sites, that we therefore refer to as the *allosteric interaction network* (Fig. 10). For NQO1 WT, we propose that this allosteric network contributes to its very high affinity for FAD as well as binding negative cooperativity, which are largely perturbed upon removal of the CTD. In addition, p.P187S may disturb this allosteric network promoting dynamic and structural alterations in the FAD binding site and the CTD and consequently reducing the binding affinity for FAD and dicoumarol. Perturbed allosteric communication between FAD binding sites and the CTD is further supported by the strong effect of CTD withdrawal in p.P187S, that eliminates the effects of the polymorphism on FAD binding site dynamics of the apo-state thus increasing its FAD binding affinity. Interestingly, other functional sites (such as the binding site for p73α) may not be strongly connected to this allosteric interaction network involving the FAD binding site, the CTD and the p.P187S site (Fig. 10).

FAD and dicoumarol cause different conformational changes upon binding to WT and p.P187S. FAD binding to NQO1 WT causes larger changes in hydrodynamic volume, secondary structure and intrinsic dynamics than those in p.P187S, while dicoumarol binding has larger effects on the conformation of holo-p.P187S (Figs 2,6 and 7, and ref. 24). These different conformational changes are mostly associated with the more dynamic and less structured CTD in the holo-state of p.P187S. Accordingly, structural calculations applied to dicoumarol binding energetics supported a much larger conformational change in the CTD for p.P187S (energetically resembling *folding* of a small protein). This scenario is consistent with a pre-existing conformational equilibrium in which apo-NQO1 mainly populates states non-competent for FAD binding, and that the presence of FAD shifts the conformational equilibrium towards less dynamic and more structured FAD binding competent states (this population-shift is macroscopically observed as a *conformational change*). For dicoumarol binding, the conformational change caused in p.P187S is larger than that of WT, due to the flexible and partially unstructured CTD of the polymorphism in the unbound holo-state, that must undergo large changes in structure and dynamics to reach the binding competent state (thus penalizing its binding). Importantly, we also provide evidence that p.P187S affects locally functional dynamics of the unbound conformational ensembles, particularly at the FAD binding of the apo-state (relevant for FAD interaction) and the CTD of the holo-state (important for dicoumarol binding). Unfortunately, the available data do not allow distinguish between *induced-fit* or *conformational-selection* mechanisms for FAD and dicoumarol binding to NQO1, that distinction would depend on whether binding competent states are significantly populated in the conformational equilibrium of apo-NQO1[12].

Proteasomal degradation of proteins is central to understanding the effect of genetic variations in loss-of-function inherited diseases[20]. In this context, dynamic or unstructured regions are key to drive recognition and efficient degradation by 20S and 26S proteasomes[8,9]. NQO1 is an excellent model to investigate proteasomal protein degradation, with particularly interesting roles of genetic variations, ligand binding and protein stability and dynamics in the modulation of degradation rates[8,22,24,36,40]. In the present work, we show that the enhanced dynamics of the CTD is key to determining the proteasomal degradation of NQO1, and furthermore, that once the CTD is removed (for instance by proteasomal degradation of this domain), the remaining protein is much more flexible and efficiently degraded (Figs 7 and 9). This result supports a step-wise directional degradation of NQO1 through its C-terminal end. More generally, our work highlights the idea of investigating mutational effects on protein dynamics as well as on global stability to assess the role of enhanced proteasomal degradation in loss-of-function genetic diseases.

To further understand the roles of NQO1 in human physiology and pathology, we must know the molecular and structural basis of its ability to interact with cell cycle regulators such as p53 and p73α and its inhibitory effect on the 20S proteasome[26,36]. Here, we present novel structural insights on the weak NQO1:oncosuppressor interaction. Importantly, we show that neither the CTD nor an active NQO1 is strictly required to interact with p73α, even though different ligands bound to NQO1 may modulate the intrinsically low affinity of this interaction, consistent with previous biochemical studies on these protein:protein interactions[29,30,37,41]. Interestingly enough, our studies show that the region of SAMp73 involved in binding to NQO1 is the same implicated in binding to other molecules at SAM domains: the so-called middle-loop-end-helix polypeptide-patch, with the fifth α-helix and the short 3_{10} helix as critical elements of secondary structure[42–45]. We must finally note that the available evidences support that intracellular destabilization of these oncosuppressors by p.P187S just mirrors its own intracellular instability rather than its inability to engage in these protein:protein interactions[24,29,36,40].

In conclusion, the present work lends support to the existence of complex allosteric communication between functional sites in a two-domain oligomeric protein with a remarkable functional chemistry and how a disease-associated single amino acid change may affect different functional sites through perturbation of allosteric networks. Due to the high prevalence of missense mutations affecting protein stability, activity and regulation[19,21,46–48], our mutational strategy combined with functional, structural, energetic and dynamic studies could help to decipher complex disease-associated mutational effects in many other loss-of-function genetic diseases and to identify structural targets for therapeutic correction.

Materials and Methods

Protein expression and purification.
The introduction of full-length NQO1 variants into the pET46 Ek/LIC vector have been recently described[22]. Site-directed mutagenesis was carried out using the QuikChange kit (Stratagene) to introduce a codon stop at the 225th residue of the coding sequence of NQO1 and to produce Δ50-NQO1 variants. All coding sequences were verified by DNA sequencing. Competent *E. coli* BL21 (DE3) cells were transformed with these plasmids and used to express NQO1 and Δ50-NQO1 proteins following previously described procedures[22]. Samples were buffer exchanged to 50 mM HEPES-KOH pH 7.4 and stored at −80 °C after flash-freezing in liquid nitrogen. Protein concentrations were measured spectrophotometrically using $\varepsilon_{280nm} = 47900\,M^{-1}\cdot cm^{-1}$. Correction for pre-bound FAD and preparation of apo-proteins were carried out as described[22]. NQO1 samples in different ligation states are designated as follows: apo-NQO1 (no ligand bound), NQO1 (as purified, partially saturated with FAD[22]), holo-NQO1 (saturated with FAD) and holo-NQO1 + dic (saturated with FAD and dicoumarol). FAD and dicoumarol stock solutions were prepared as previously described[22,24].

SAMp73 was produced and purified as described[49]. For 15N-SAMp73 expression, cells were grown in M9 minimal media, with 1 gr per liter of $^{15}NH_4Cl$, supplemented with vitamins, and purified as the unlabelled protein[49].

Isothermal titration calorimetry (ITC).
Experiments were performed in a ITC_{200} microcalorimeter (Malvern) with a cell volume of 0.205 mL. Titrations were performed in 50 mM HEPES–KOH pH 7.4. FAD titrations were carried out using 3.5–5 μM Δ50-apo-NQO1 variants (in dimer units) in the cell and 125 μM FAD in the titration syringe by performing 20–25 injections of 1.5 μL each 180 s. Dicoumarol titrations were carried out using 5 μM NQO1 or Δ50-NQO1 holo-variants (in monomer) supplemented with FAD (10 μM), and 100 μM dicoumarol + 10 μM FAD in the syringe by performing 20–25 injections of 1.5 μL each 180 s. Data were corrected for heats of dilution by performing titrations into buffer. After peak integration, binding isotherms were fitted using a model for identical and independent type of sites using the software provided by the manufacturer. Thermodynamic parameters are presented as the mean from two independent titrations.

Far-UV circular dichroism (CD) spectroscopy.
Circular dichroism (CD) was measured in Jasco J-710 or J-815 spectropolarimeters thermostatized using a Peltier element. NQO1 CD spectra were determined in 50 mM K-phosphate pH 7.4 at 25 °C, typically in a 195–260 nm range and using 50–100 nm/min scan rates. Samples of NQO1 proteins were prepared at 4 μM concentration in monomer, and 6 scans were acquired and averaged. In some cases, 10 μM FAD and 10 μM FAD + 10 μM dicoumarol were added. Experiments in the presence of SAMp73 and WT NQO1 were carried out at a equimolar concentration (10 μM in protein monomer) of each protein at 25 °C in 25 mM phosphate buffer pH 6.9. In all cases, appropriate blanks in the absence of proteins were acquired and subtracted from those of protein samples.

For thermal denaturation studies, NQO1 samples (4 μM in protein monomer) were equilibrated for 10 min at 20 °C in the absence or presence of ligands (50 μM FAD and 50 μM FAD + 50 μM dicoumarol) and thermal scans were performed up to 70–80 °C at a 2 °C/min scan rate. For sake of comparison, denaturation curves were normalized using pre- and post-transition baselines, and the half-denaturation temperature (T_m) was determined as the temperature at which half of the native signal was lost.

Dynamic light scattering.
Dynamic light scattering (DLS) was carried out in a DynaPro MSX instrument (Wyatt) using 1.5 mm path length cuvettes and 10 μM protein (in monomer units) in 50 mM HEPES-KOH pH 7.4 at 25 °C. 30 measurements were acquired for each NQO1 protein, in the absence or presence of ligands (20 μM FAD and 20 μM FAD + 20 μM dicoumarol) in three to six independent replicates, averaged and used to determine the hydrodynamic radius and polydispersity using the average autocorrelation function and assuming a spherical shape.

Fluorescence experiments.
Fluorescence spectra were acquired in a Varian Cary Eclipse spectrofluorimeter (Agilent, USA) interfaced with a Peltier unit. The excitation wavelengths were either 280 or 295 nm. The emission fluorescence was collected between 300–400 nm. Excitation and emission slits were 5 nm. Spectra were corrected by subtracting the corresponding blanks. Experiments were acquired with 2 μM SAMp73 or

NQO1-WT (in protein monomer), or a equimolar mixture of them, in 25 mM phosphate buffer pH 6.9 at 25 °C. NADH was added at a final concentration of 1 mM.

For titration experiments, a concentration of 4.2 μM NQO1-WT (in monomer) was used. Increasing concentrations of SAMp73 (from 0 to 8 μM) were added. The rest of the experimental set was the same as above.

Activity measurements. NQO1 activity was measured in 50 mM HEPES-KOH pH 7.4. A reaction mixture containing recombinant NQO1 and 0.5 mM NADH was incubated at 30 °C for 5 min in 1-cm path length quartz cuvettes in a thermostatized Agilent 8453 diode array spectrophotometer. The reaction was triggered by the addition of 70 μM DCPIP (2,6-dichlorophenolindophenol) as the electron acceptor. Initial reaction rates were determined from changes in A_{600nm} resulting from the reduction of DCPIP and corrected for the non-enzymatic reaction. To determine the kinetic parameters, NQO1 enzymes were incubated with a FAD excess (1 μM final concentration), NADH was kept constant at 1 mM and DCPIP was varied from 7–50 μM or, DCPIP was mantained at 50 μM and NADH was varied from 0.05–8 mM. Blanks in the absence of proteins were determined and subtracted from the reaction with NQO1. The NQO1 concentration used varied depending on the variant to ensure linearity over time and protein concentration: 1 nM (WT with and without added FAD, p.P187S with FAD), 25–50 nM (p.P187S without added FAD), 200–600 nM (Δ50-WT and Δ50-p.P187S, with and without added FAD). The specific activity was calculated using a $\varepsilon_{600nm} = 21000$ $M^{-1} \cdot cm^{-1}$ for DCPIP. k_{cat} and K_M values were determined using the Michaelis-Menten equation.

Proteolysis by thermolysin. Thermolysin from *Bacillus thermoproteolyticus rokko* was purchased from Sigma-Aldrich, disolved in 50 mM HEPES-KOH 10 mM CaCl$_2$ pH 7.4, stored at −80 °C and its concentration measured spectrophotometrically using a $\varepsilon_{280nm} = 66086$ $M^{-1} \cdot cm^{-1}$. Proteolysis was performed at 25 °C in 50 mM HEPES-KOH 10 mM CaCl$_2$ pH 7.4 using 20 μM Δ50-NQO1 (in monomer units), and the reaction was initiated by the addition of thermolysin at a final concentration of 0.1 μM. Proteolysis reactions were performed using Δ50-NQO1 as apo-proteins, upon addition of 100 μM FAD or 100 μM FAD and dicoumarol. Samples were withdrawn at different times, quenched with EDTA (20 mM final concentration) and denatured using Laemmli´s buffer at 95 °C for 5 min. Samples were resolved in 12% acrylamide SDS-PAGE gels, stained with Coomassie Blue and densitometered using ImageJ (http://rsbweb.nih.gov/ij/). After normalization of the intensities using that of the sample lacking thermolysin, the decay of the native band was analyzed using a single exponential function that yields the proteolysis rate constant k.

Mass spectrometry. To analyze the proteolysis products of Δ50-WT, the apo-protein (10 μM in monomer units) was incubated at 25 °C in 50 mM HEPES-KOH 10 mM CaCl$_2$ pH 7.4 with 100 nM thermolysin at three different conditions: (i) no ligand and 1 min reaction; (ii) FAD 100 μM, 16 min; (iii) FAD and dicoumarol (100 μM each) for 4 hours. The reaction volume was 150 μL. The reaction was stopped by addition of EDTA 20 mM.

125 μL of the final solution were exchanged to water using VIVAspin 500 filters (10 kDa) by two dilution-concentration cycles and submitted for High performance liquid chromatography/electrospray ionization mass spectrometry (HPLC/ESI-MS) to the High-resolution mass spectrometry unit, Centro de Instrumentación Científica (University of Granada). HPLC/ESI-MS was performed in a Acquity UPLC system (Waters), using a water/formic acid, acetonitrile/formic acid (each at 0.1%) gradient in a Acquity UPLC® BEH300 C4 column (2.1 × 50 mm, Waters) coupled with a Q-TOF Synapt62 HDMS (Waters).

10 μL of the final solution were mixed with an equal volume of Laemmli's buffer (x2), denatured at 95 °C and loaded into a SDS-PAGE gel (12% acrylamide). Two bands (#1 and #2) were manually excised from gels and submitted for analyses to the Proteomics Unit, Complutense University of Madrid, a member of ProteoRed network. Samples were in-gel reduced, alkylated and digested with trypsin, and after digestion, 1 μl of the supernatant was spotted onto a MALDI target plate, allowed to air-dry, and mixed with α-cyano-4-hydroxy-cinnamic acid matrix (Sigma) in 50% acetonitrile, and allowed again to air-dry at room temperature. MALDI-TOF MS analyses were performed in a 4800 Plus Proteomics Analyzer MALDI-TOF/TOF mass spectrometer (Applied Biosystems, MDS Sciex, Toronto, Canada). The MALDI-TOF/TOF operated in positive reflector mode with an accelerating voltage of 20000 V. All mass spectra were calibrated internally using peptides from the auto digestion of trypsin. Peptides of interest were subjected to MS/MS sequencing analyses using the 4800 plus Proteomics Analyzer. From MS spectra, suitable precursors were selected to be fragmented by CID (Collision Induced Disociation with atmospheric gas) using a 1 KV ion reflector operate method and a precursor mass window of ±4 Da. The Swissprot database (http://www.uniprot.org/help/uniprotkb) was used for protein identification without taxonomy restriction, and a home-made database with the sequence of recombinant Δ50-WT NQO1 was searched using MASCOT 2.3 (www.matrixscience.com) through the Global Protein Server v 3.6 from ABSciex, using the following parameters: enzyme, semitrypsin; carbamidomethyl cystein as fixed modification and oxidized methionine as variable modification; peptide mass tolerance, 50 ppm (PMFSearch) – 100 ppm (Combined search); up to 2 missed trypsin cleavage site; MS-MS fragments tolerance, 0.3 Da. The parameters for the combined search (Peptide mass fingerprint plus MS-MS spectra) were the same as described above. In any protein identification, the probability scores were greater than the score fixed by Mascot as significant with a p value < 0.05.

Molecular dynamics simulations. We performed molecular dynamics simulations of Δ50-NQO1 WT and p.P187S in the apo-state and in complex with FAD and/or dicoumarol. The simulations of full-length NQO1 used as comparison are described in ref. 24. Systems were prepared in dimeric state on basis of crystal structures of the complex with FAD (PDB: 1D4A[50]) and the ternary complex (PDB: 2F1O[32]). Proteins were protonated for simulation using protonate3d and mutated as well as truncated in MOE[51].

Simulations were performed using the GPU implementation of pmemd in Amber12[52]. Protein residues were parametrized using the Amber forcefield 99SB-ILDN[53]. FAD and dicumarol parameters were taken from an

earlier study[24]. After employing an extensive equilibration protocol[54], unrestrained systems were sampled for 100 ns in NpT ensemble and 5,000 equal-spaced snapshots were saved to trajectory.

After ensuring thermodynamic and structural stability of simulations, we analyzed resulting conformational ensembles using cpptraj from AmberTools[33]. We analyzed B-factors of Cα atoms (derived from root mean squared fluctuations along the simulations) after a single alignment to all Cα atoms as a metric for global flexibility. Additionally, we extracted dihedral angles of the protein backbone and calculated dihedral entropies based on resulting distributions after employing parameter-free kernel density estimation[55]. Integration over probability densities yields a thermodynamic entropy depicting a metric for local intrinsic flexibility of the protein backbone[56]. Independent dihedral entropies over all three backbone torsion angles were summed to yield a total dihedral entropy per residue. All values are presented as arithmetic average over two dimer sub-units. Interface residues were defined as all residues with at least one atom closer than 3 Å to the adjacent sub-unit.

NMR spectroscopy. The NMR data were acquired at 20 °C on a Bruker Avance DRX-500 spectrometer equipped with a triple-resonance probe and z-gradients. Proteins and ligands were prepared in 50 mM phosphate buffer pH 6.9. The ^{15}N-labelled SAMp73 concentration was in all cases ~200 μM, NQO1 variants (as purified) were prepared at ~300 μM. When necessary, ligands were added to a final concentration of 3 mM (dicoumarol) or 8 mM (FAD or NADH). The 2D ^1H-^{15}N HSQC (heteronuclear single-quantum coherence) experiments[57] were acquired in the following conditions: (i) isolated ^{15}N-SAMp73; (ii) ^{15}N-SAMp73 with NQO1 WT; (iii) ^{15}N-SAMp73 with NQO1 WT + NADH; (iv) ^{15}N-SAMp73 with Δ50-NQO1 WT; (v) ^{15}N-SAMp73 with Δ50-NQO1 WT + FAD; (vi) ^{15}N-SAMp73 with Δ50-NQO1 WT + FAD + NADH; (vii) ^{15}N-SAMp73 with NQO1 P187S; (viii) ^{15}N-SAMp73 with NQO1 P187S + FAD; (ix) ^{15}N-SAMp73 with NQO1 P187S + FAD + NADH; and, (x) ^{15}N-SAMp73 with NQO1 WT + dicoumarol (with stoichiometric amounts of dicoumarol). Samples containing Δ50-NQO1 led to precipitation during the experiments; experiments in the presence of dicoumarol also led to precipitation. Control experiments were acquired with ^{15}N-SAMp73 with FAD and ^{15}N-SAMp73 with NADH; no differences in either chemical shifts or signal broadening were observed when compared with the spectrum of isolated ^{15}N-SAMp73 (BMRB number 4413)[58].

Spectra were acquired in the phase sensitive mode. Frequency discrimination in the indirect dimensions was achieved by using the echo/antiecho-TPPI method. The spectra were acquired with 2 K complex points in the ^1H dimension, 128 complex points in the ^{15}N dimension, and 200 scans. The carrier of the ^1H dimension was set at the water frequency, and that of ^{15}N at 120 ppm. The spectral widths used were 10 and 35 ppm in the ^1H and ^{15}N dimensions, respectively. Water was suppressed with the WATERGATE sequence[59]. Data were zero-filled to double the number of original points in both dimensions, apodized with shifted squared sine-bell functions in the two dimensions and Fourier transformed with the program TopSpin 1.3 (Bruker). The assignment of the spectrum of SAMp73 under our conditions was carried out by comparison of the chemical shifts of the signals, with that previously published[58].

To analyse the differences between the 2D ^1H-^{15}N HSQC spectra of isolated SAMp73 and of the complexes, the chemical shift perturbation (CSP) was used. CSP was calculated as: $CSP = \sqrt{(\Delta\delta_H)^2 + 0.1(\Delta\delta_N)^2}$, where $\Delta\delta_H$ is the difference in chemical shift of the amide protons of isolated SAMp73 and that of SAMp73 in the corresponding complex; and $\Delta\delta_N$ is the difference in chemical shift of ^{15}N resonances of isolated SAMp73 and that of SAMp73 in the complex. Only variations in residues with CSP ≥ 0.01 ppm were considered significant. Broadening of signals was measured in comparison with the internal reference of Ala530. Changes in cross-peak intensity were considered significant when variations between the spectrum of isolated SAMp73 and that in the presence of NQO1 species were larger than 10%.

Expression studies in a cell-free system. Expression in a rabbit reticulocyte cell-free system (TnT system, Promega) was performed at 30 °C for 30 min using ^{35}S-Met and NOQ1 cDNA variants subcloned in pCIneo plasmids (Promega; 1 μg plasmid/reaction). Protein synthesis was stopped with 100 μg/ml cycloheximide, and kept at 30 °C for various chasing times. Aliquots were denatured in Laemmli's buffer and analyzed by SDS-PAGE and fluorography. The autoradiograms were scanned in a Molecular Imager FX (Bio-Rad) and the software LabImager (Bio-Rad) was used to measure the band volume adjusted to background. All experiments were performed three independent times and data are presented as mean ± s.d.

References

1. Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450,** 964–972 (2007).
2. Hilser, V. J., Garcia-Moreno, E. B., Oas, T. G., Kapp, G. & Whitten, S. T. A statistical thermodynamic model of the protein ensemble. *Chem Rev* **106,** 1545–1558 (2006).
3. Frederick, K. K., Marlow, M. S., Valentine, K. G. & Wand, A. J. Conformational entropy in molecular recognition by proteins. *Nature* **448,** 325–329 (2007).
4. Tzeng, S. R. & Kalodimos, C. G. Protein activity regulation by conformational entropy. *Nature* **488,** 236–240 (2012).
5. Henzler-Wildman, K. A. *et al.* A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* **450,** 913–916 (2007).
6. Kerns, S. J. *et al.* The energy landscape of adenylate kinase during catalysis. *Nat Struct Mol Biol* **22,** 124–131 (2015).
7. Tzeng, S. R. & Kalodimos, C. G. Dynamic activation of an allosteric regulatory protein. *Nature* **462,** 368–372 (2009).
8. Inobe, T. & Matouschek, A. Paradigms of protein degradation by the proteasome. *Curr Opin Struct Biol* **24,** 156–164 (2014).
9. Guharoy, M., Bhowmick, P., Sallam, M. & Tompa, P. Tripartite degrons confer diversity and specificity on regulated protein degradation in the ubiquitin-proteasome system. *Nat Commun* **7,** 10239 (2016).
10. Wilson, C. *et al.* Kinase dynamics. Using ancient protein kinases to unravel a modern cancer drug's mechanism. *Science* **347,** 882–886 (2015).
11. Motlagh, H. N., Wrabl, J. O., Li, J. & Hilser, V. J. The ensemble nature of allostery. *Nature* **508,** 331–339 (2014).
12. Vega, S., Abian, O. & Velazquez-Campoy, A. On the link between conformational changes, ligand binding and heat capacity. *Biochim Biophys Acta* **1860,** 868–878 (2016).

13. Nussinov, R. & Tsai, C. J. Allostery without a conformational change? Revisiting the paradigm. *Curr Opin Struct Biol* **30,** 17–24 (2015).
14. Tsai, C. J. & Nussinov, R. A unified view of "how allostery works". *PLoS Comput Biol* **10,** e1003394 (2014).
15. Hilser, V. J., Gomez, J. & Freire, E. The enthalpy change in protein folding and binding: refinement of parameters for structure-based calculations. *Proteins* **26,** 123–133 (1996).
16. Liu, T., Whitten, S. T. & Hilser, V. J. Functional residues serve a dominant role in mediating the cooperativity of the protein ensemble. *Proc Natl Acad Sci USA* **104,** 4347–4352 (2007).
17. Pey, A. L., Majtan, T., Sanchez-Ruiz, J. M. & Kraus, J. P. Human cystathionine beta-synthase (CBS) contains two classes of binding sites for S-adenosylmethionine (SAM): complex regulation of CBS activity and stability by SAM. *Biochem J* **449,** 109–121 (2013).
18. Pey, A. L., Padin-Gonzalez, E., Mesa-Torres, N. & Timson, D. J. The metastability of human UDP-galactose 4'-epimerase (GALE) is increased by variants associated with type III galactosemia but decreased by substrate and cofactor binding. *Arch Biochem Biophys* **562,** 103–114 (2014).
19. Muntau, A. C., Leandro, J., Staudigl, M., Mayer, F. & Gersting, S. W. Innovative strategies to treat protein misfolding in inborn errors of metabolism: pharmacological chaperones and proteostasis regulators. *J Inherit Metab Dis* **37,** 505–523 (2014).
20. Gregersen, N., Bross, P., Vang, S. & Christensen, J. H. Protein misfolding and human disease. *Annu Rev Genomics Hum Genet* **7,** 103–124 (2006).
21. Gomes, C. M. Protein misfolding in disease and small molecule therapies. *Curr Top Med Chem* **12,** 2460–2469 (2012).
22. Pey, A. L., Megarity, C. F. & Timson, D. J. FAD binding overcomes defects in activity and stability displayed by cancer-associated variants of human NQO1. *Biochim Biophys Acta* **1842,** 2163–2173 (2014).
23. Tsai, C. J. & Nussinov, R. The free energy landscape in translational science: how can somatic mutations result in constitutive oncogenic activation? *Phys Chem Chem Phys* **16,** 6332–6341 (2014).
24. Medina-Carmona, E. *et al.* Conformational dynamics is key to understanding loss-of-function of NQO1 cancer-associated polymorphisms and its correction by pharmacological ligands. *Scientific Reports* **6,** 20331 (2016).
25. Fuchs, J. E., Muñoz, I. G., Timson, D. J. & Pey, A. L. Experimental and computational evidence on conformational fluctuations as a source of catalytic defects in genetic diseases *RSC Adv.* **6,** 58604 (2016).
26. Pey, A. L., Megarity, C. F., Medina-Carmona, E. & Timson, D. J. Natural small molecules as stabilizers and activators of cancer-associated NQO1 polymorphisms. *Curr Drug Targets* **17,** 1506–1514 (2016).
27. Asher, G., Bercovich, Z., Tsvetkov, P., Shaul, Y. & Kahana, C. 20S proteasomal degradation of ornithine decarboxylase is regulated by NQO1. *Mol Cell* **17,** 645–655 (2005).
28. Asher, G., Lotem, J., Cohen, B., Sachs, L. & Shaul, Y. Regulation of p53 stability and p53-dependent apoptosis by NADH quinone oxidoreductase 1. *Proc Natl Acad Sci USA* **98,** 1188–1193 (2001).
29. Asher, G., Lotem, J., Kama, R., Sachs, L. & Shaul, Y. NQO1 stabilizes p53 through a distinct pathway. *Proc Natl Acad Sci USA* **99,** 3099–3104 (2002).
30. Asher, G., Tsvetkov, P., Kahana, C. & Shaul, Y. A mechanism of ubiquitin-independent proteasomal degradation of the tumor suppressors p53 and p73. *Genes Dev* **19,** 316–321 (2005).
31. Lienhart, W. D. *et al.* Collapse of the native structure caused by a single amino acid exchange in human NAD(P)H:quinone oxidoreductase(1). *FEBS J* **281,** 4691–4704 (2014).
32. Asher, G., Dym, O., Tsvetkov, P., Adler, J. & Shaul, Y. The crystal structure of NAD(P)H quinone oxidoreductase 1 in complex with its potent inhibitor dicoumarol. *Biochemistry* **45,** 6372–6378 (2006).
33. Roe, D. R. & Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **9,** 3084–3095 (2013).
34. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372,** 774–797 (2007).
35. Robertson, A. D. & Murphy, K. P. Protein Structure and the Energetics of Protein Stability. *Chem Rev* **97,** 1251–1268 (1997).
36. Moscovitz, O. *et al.* A mutually inhibitory feedback loop between the 20S proteasome and its regulator, NQO1. *Mol Cell* **47,** 76–86 (2012).
37. Xu, J. & Jaiswal, A. K. NAD(P)H:quinone oxidoreductase 1 (NQO1) competes with 20S proteasome for binding with C/EBPalpha leading to its stabilization and protection against radiation-induced myeloproliferative disease. *J Biol Chem* **287,** 41608–41618 (2012).
38. Liu, G. & Chen, X. The C-terminal sterile alpha motif and the extreme C terminus regulate the transcriptional activity of the alpha isoform of p73. *J Biol Chem* **280,** 20111–20119 (2005).
39. Fielding, L. NMR methods for the determination of protein-ligand dissociation constants. *Prog. Nuc. Magn. Reson.* **51,** 219–242 (2007).
40. Siegel, D. *et al.* Rapid polyubiquitination and proteasomal degradation of a mutant form of NAD(P)H:quinone oxidoreductase 1. *Mol Pharmacol* **59,** 263–268 (2001).
41. Anwar, A. *et al.* Interaction of human NAD(P)H:quinone oxidoreductase 1 (NQO1) with the tumor suppressor protein p53 in cells and cell-free systems. *J Biol Chem* **278,** 10368–10373 (2003).
42. Barrera, F. N., Poveda, J. A., Gonzalez-Ros, J. M. & Neira, J. L. Binding of the C-terminal sterile alpha motif (SAM) domain of human p73 to lipid membranes. *J Biol Chem* **278,** 46878–46885 (2003).
43. Leone, M., Cellitti, J. & Pellecchia, M. NMR studies of a heterotypic Sam-Sam domain association: the interaction between the lipid phosphatase Ship2 and the EphA2 receptor. *Biochemistry* **47,** 12721–12728 (2008).
44. Mercurio, F. A. *et al.* Solution structure of the first Sam domain of Odin and binding studies with the EphA2 receptor. *Biochemistry* **51,** 2136–2145 (2012).
45. Lee, H. J. *et al.* NMR structure of a heterodimeric SAM:SAM complex: characterization and manipulation of EphA2 binding reveal new cellular functions of SHIP2. *Structure* **20,** 41–55 (2012).
46. Pey, A. L. Protein homeostasis disorders of key enzymes of amino acids metabolism: mutation-induced protein kinetic destabilization and new therapeutic strategies. *Amino Acids* **45,** 1331–1341 (2013).
47. Pey, A. L., Maggi, M. & Valentini, G. Insights into human phosphoglycerate kinase 1 deficiency as a conformational disease from biochemical, biophysical, and *in vitro* expression analyses. *J Inherit Metab Dis* **37,** 909–916 (2014).
48. Gaboriau, D. C., Rowling, P. J., Morrison, C. G. & Itzhaki, L. S. Protein stability versus function: effects of destabilizing missense mutations on BRCA1 DNA repair activity. *Biochem J* **466,** 613–624 (2015).
49. Barrera, F. N., Garzon, M. T., Gomez, J. & Neira, J. L. Equilibrium unfolding of the C-terminal SAM domain of p73. *Biochemistry* **41,** 5743–5753 (2002).
50. Faig, M. *et al.* Structures of recombinant human and mouse NAD(P)H:quinone oxidoreductases: species comparison and structural changes with substrate binding and release. *Proc Natl Acad Sci USA* **97,** 3177–3182 (2000).
51. Labute, P. Protonate3D: assignment of ionization states and hydrogen coordinates to macromolecular structures. *Proteins* **75,** 187–205 (2009).
52. AMBER 12 (University of California, San Francisco, 2012).
53. Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78,** 1950–1958 (2010).
54. Wallnoefer, H. G., Handschuh, S., Liedl, K. R. & Fox, T. Stabilizing of a globular protein by a highly complex water network: a molecular dynamics simulation study on factor Xa. *J Phys Chem B* **114,** 7405–7412 (2010).

55. Huber, R. G. *et al.* Entropy from state probabilities: hydration entropy of cations. *J Phys Chem B* **117,** 6466–6472 (2013).
56. Huber, R. G., Eibl, C. & Fuchs, J. E. Intrinsic flexibility of NLRP pyrin domains is a key factor in their conformational dynamics, fold stability, and dimerization. *Protein Sci* **24,** 174–181 (2015).
57. Bodenhausen, G. & Ruben, D. Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Lett.* **69,** 185–189 (1980).
58. Chi, S. W., Ayed, A. & Arrowsmith, C. H. Solution structure of a conserved C-terminal domain of p73 with structural homology to the SAM domain. *EMBO J* **18,** 4438–4445 (1999).
59. Piotto, M., Saudek, V. & Sklenar, V. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J Biomol NMR* **2,** 661–665 (1992).
60. DeLano, W. L. The PyMOL Molecular Graphics System. *DeLano Scientific LLC, Palo Alto, CA, USA*. http://www.pymol.org (2002)

Acknowledgements

Author Contributions

Study design and supervision: A.L.P.; Development of methodology: E.M.-C., J.L.-N., R.P.-M., J.E.F., E.S., and A.L.P.; Protein expression and purification: E.M.-C., J.L.-N. and A.L.P.; Biophysical studies: J.L.-N. and A.L.P.; Proteolysis experiments: A.L.P.; Experiments in cell cultures and cell-free systems: E.M.-C., E.S. and R.P.M.; Molecular dynamic simulations: J.E.F.; Data analysis and interpretation: E.M.-C., E.S., J.L.-N., R.P.-M., J.E.F., D.J.T. and A.L.P.; Drafting and revision of the manuscript: A.L.P. All authors revised and approved the final version of the manuscript.

Additional Information

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

Competing Interests: The authors declare no competing financial interests.

How to cite this article: Medina-Carmona, E. *et al.* Site-to-site interdomain communication may mediate different loss-of-function mechanisms in a cancer-associated NQO1 polymorphism. *Sci. Rep.* **7**, 44532; doi: 10.1038/srep44532 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

PUBLICATION 3

Enhanced vulnerability of human proteins towards disease-associated inactivation through divergent evolution.

OXFORD

ORIGINAL ARTICLE

Enhanced vulnerability of human proteins towards disease-associated inactivation through divergent evolution

Encarnación Medina-Carmona[1], Julian E. Fuchs[2,†], Jose A. Gavira[3], Noel Mesa-Torres[1], Jose L. Neira[4,5], Eduardo Salido[6], Rogelio Palomino-Morales[7], Miguel Burgos[8], David J. Timson[9] and Angel L. Pey[1,*]

[1]Department of Physical Chemistry, Faculty of Sciences, University of Granada, Granada, Spain, [2]Institute of General, Inorganic and Theoretical Chemistry, Faculty of Chemistry and Pharmacy, University of Innsbruck, Innsbruck, Austria, [3]Laboratorio de Estudios Cristalográficos, IACT (CSIC-UGR), Armilla, Granada, Spain, [4]Instituto de Biología Molecular y Celular, Universidad Miguel Hernández, Avda. del Ferrocarril s/n, 03202 Elche, Alicante, Spain, [5]Instituto de Biocomputación y Física de los Sistemas Complejos (BIFI), 50009 Zaragoza, Spain, [6]Centre for Biomedical Research on Rare Diseases (CIBERER), Hospital Universitario de Canarias, Tenerife, Spain, [7]Department of Biochemistry and Molecular Biology I, Faculty of Sciences and Biomedical Research Center (CIBM), University of Granada, Granada (Spain), [8]Department of Genetics and Institute of Biotechnology, Faculty of Sciences, and Biomedical Research Center (CIBM), University of Granada, Granada, Spain and [9]School of Pharmacy and Biomolecular Sciences, The University of Brighton, Brighton, UK

*To whom correspondence should be addressed at: Department of Physical Chemistry, University of Granada, Av. Fuentenueva s/n, E-18071, Granada, Spain. Tel: +34 958243173; Fax: +34 958242747; Email: angelpey@ugr.es

Abstract

Human proteins are vulnerable towards disease-associated single amino acid replacements affecting protein stability and function. Interestingly, a few studies have shown that consensus amino acids from mammals or vertebrates can enhance protein stability when incorporated into human proteins. Here, we investigate yet unexplored relationships between the high vulnerability of human proteins towards disease-associated inactivation and recent evolutionary site-specific divergence of stabilizing amino acids. Using phylogenetic, structural and experimental analyses, we show that divergence from the consensus amino acids at several sites during mammalian evolution has caused local protein destabilization in two human proteins linked to disease: cancer-associated NQO1 and alanine:glyoxylate aminotransferase, mutated in primary hyperoxaluria type I. We demonstrate that a single consensus mutation (H80R) acts as a disease suppressor on the most common cancer-associated polymorphism in NQO1 (P187S). The H80R mutation reactivates P187S by enhancing FAD binding affinity through local and dynamic stabilization of its binding site. Furthermore, we show how a second suppressor mutation (E247Q) cooperates with H80R in protecting the P187S polymorphism towards inactivation through long-range allosteric communication within the structural ensemble of the protein. Our results support that recent divergence of consensus amino acids may have occurred with neutral effects on many functional and regulatory traits of wild-type human proteins. However, divergence at

3531

certain sites may have increased the propensity of some human proteins towards inactivation due to disease-associated mutations and polymorphisms. Consensus mutations also emerge as a potential strategy to identify structural hot-spots in proteins as targets for pharmacological rescue in loss-of-function genetic diseases.

Introduction

Through evolution, mutations causing single amino acid replacements may occur and fixate causing beneficial functional and stability changes (adaptive mutations), while others may not show any apparent change on protein functionality (neutral mutations) or affect secondary (promiscuous) functions, while largely deleterious mutations are more rarely fixed (1,2). Random mutations are often destabilizing, particularly when they affect the protein core (3), and may reduce protein stability below a certain threshold impairing functionality and organism fitness, which in principle often prevents their fixation (2). Fixation of destabilizing mutations can be tolerated if other stabilizing mutations occur before or afterwards (2,4). In this context, stability-activity trade-offs can be biologically buffered by different mechanisms, such as control of protein expression and modulation of protein misfolding by molecular chaperones (2,5). Importantly, mutations often show pleiotropic effects, manifesting different effects on primary and secondary functions as well as on protein stability (2,4). Mutations can also contribute to the evolution of protein function and the acquisition of regulatory mechanisms involving protein allostery through modulation of the dynamics of protein conformational ensembles (1,6–9).

The ability of proteins to fold into a native and stable structure intracellularly is often challenged by mutations, thus leading to disease (10,11). During evolution, mammals have gathered a vast network of ∼1300 different proteins involved in protein biogenesis, folding, trafficking, degradation and maintenance of the proteome named the *protein homeostasis network* (10,11). Mutations may critically affect the ability of proteins to fold and remain stable, causing loss of enzyme function and metabolic defects commonly due to accelerated proteasomal degradation (12). Global destabilization of the protein structure due to single amino acid replacements may play an important role in this loss-of-function mechanism (13–15), although recent studies support the idea that local stability and dynamics of protein structures may drive their directional degradation via the proteasome (15–19). Accordingly, small molecules aimed at locally or globally stabilizing protein structure constitute a promising new battery of pharmacological tools to treat these diseases (12,20–22). Mistargeting is another consequence of some mutations (23,24), and the compartmentalization of metabolic pathways along evolution (25,26) seems directly related to such mutations.

Proteins are often stabilized upon replacement of a rare amino acid at a given position by the most common one (i.e. the *consensus amino acid*) found in an alignment of homologous sequences (27–32). These *back-to-consensus* or *consensus mutations* increase protein stability primarily by local optimization of the interactions in the environment of the mutated site, and in some cases, by decreasing local flexibility (28,33,34). It has been argued that consensus mutations are reversions to the ancestral amino acid, and thus, their stabilizing effect reflects to some extent the thermophilic nature of ancestral proteins (35–37), although the universality of this relationship is unclear (38). Consensus mutations can act as global intragenic suppressors by providing a stability boost towards destabilizing

mutations (39). Since mammalian and vertebrate consensus amino acids can be strongly stabilizing when incorporated to human proteins (28,32), it is intriguing that during the evolutionary history of mammals divergence from consensus amino acids may have yielded some human proteins more prone to loss-of-function and disease-causing mutations.

To investigate the evolutionary divergence of consensus amino acids as potential intragenic suppressors of human disease, we have chosen to study the most common cancer-associated polymorphism (rs1800566/c.C609T/p.P187S) in NAD(P)H:quinone oxidoreductase 1 (NQO1) that causes a Pro187 to Ser187 amino acid replacement (40). This system is particularly suitable due to the multifunctional nature of this enzyme (allowing us to investigate specific effects on different individual functions) as well as the well-characterized mechanisms by which the polymorphism affects protein function and stability *in vivo* (18,41–45). NQO1 is a homodimeric FAD-dependent enzyme involved in the activation of some anti-cancer pro-drugs (41,46), which also interacts with and stabilizes transcription factors associated with cancer and cell-cycle regulation, such as p53, p73α and HIF-1α (47,48). While wild-type (WT) NQO1 is kinetically stable at physiological temperature *in vitro* (with a denaturation half-life of 9 h), the P187S polymorphism accelerates denaturation by a factor of 60 (43). More importantly, the P187S polymorphism largely abolishes NQO1 activity *in vivo* by affecting FAD binding and promotes its proteasomal degradation, through dynamic destabilization of two distant sites: the FAD binding site in the N-terminal domain (NTD, residues 1–224) and the small C-terminal domain (CTD, residues 225–274), respectively (18,42,44,49). As a consequence, binding of FAD to WT NQO1 protects it towards degradation *in vivo* while binding of the inhibitor dicoumarol is required to stabilize P187S, supporting a different directionality in their proteasomal degradation (i.e. initiated at the NTD vs. the CTD, respectively) and their specific blockage by ligands affecting protein local dynamics upon binding (18,44,45,49).

Results

Stabilizing consensus amino acids have been replaced during recent divergent evolution of two disease-associated human proteins

To identify consensus mutations for NQO1, we have used an alignment of mammalian sequences, including those with high (over 80%) identity to that of human NQO1 from eighteen different orders overall (Supplementary Materials) and a large representation of those from primates (15.3%), rodents (19.2%), artiodactyla (15.3%), carnivora (11.5%) and chiroptera (9.6%). Among these consensus mutations, H80R causes the most remarkable stabilizing effect either in WT or P187S NQO1 (Fig. 1A and B and Supplementary Material, Table S1), consequently increasing their kinetic stability (i.e. slowing down irreversible inactivation, Supplementary Material, Table S2). Importantly, H80R causes a local dynamic stabilization of the N-terminal domain (NTD) of WT NQO1, increasing by 2-fold its resistance towards partial proteolysis by thermolysin (Fig. 1C and D), which
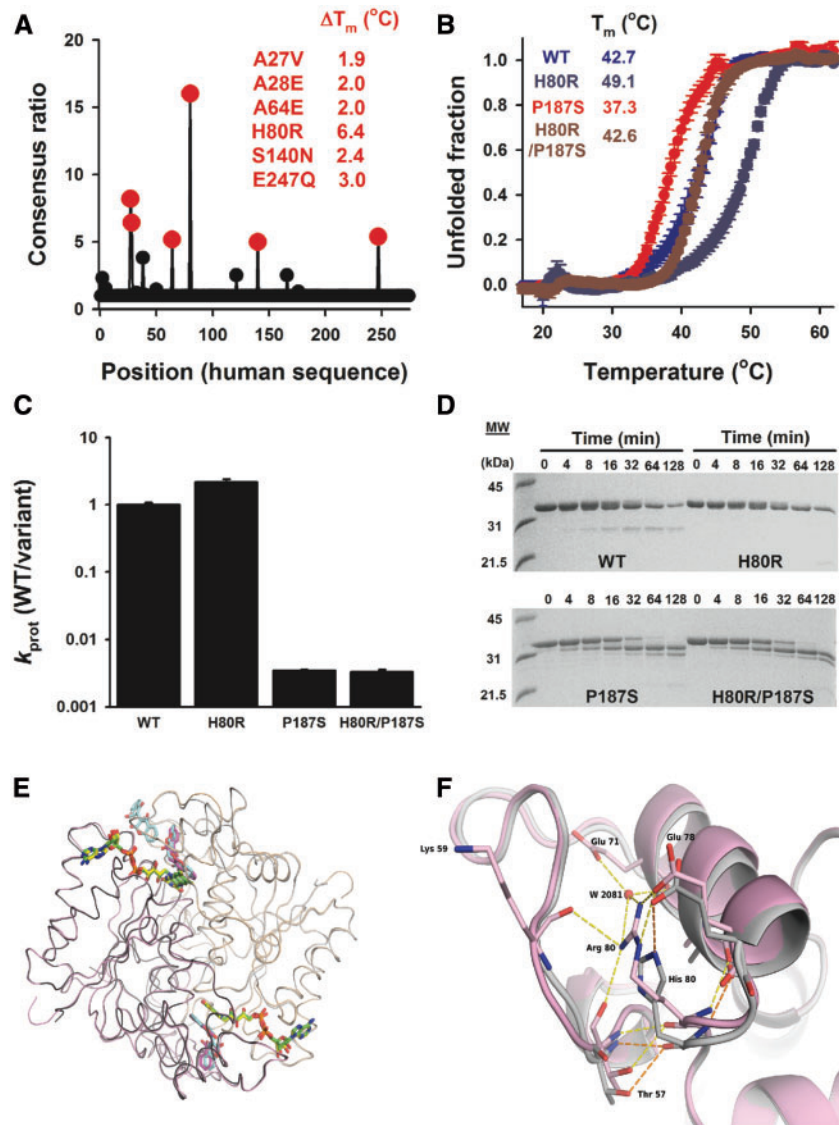
Figure 1. Structural basis of the global and local stabilization of human NQO1 by the consensus H80R mutation *in vitro*. (**A**) Consensus mutations derived from mammalian NQO1 sequences stabilize NQO1; the consensus ratios and changes in denaturation temperature vs. WT (ΔT_m; from $T_{0.5}$ values in Supplementary Material, Table S1) are shown; (**B**) Thermal stabilization of WT and P187S by the H80R mutation; data are means. from at least five experiments. T_m from $T_{0.5}$ values are shown (Supplementary Material, Table S1); C and D) Normalized proteolysis rates (**C**) and degradation patterns (**D**) for NQO1 enzymes showing local stabilization of the NTD by H80R. Normalized proteolysis rates are calculated from ratios of k_{prot} (in $M^{-1} \cdot min^{-1}$) for WT/variants, and thus, values higher/lower than one reflect quantitatively the stabilizing/destabilizing effect, respectively; (**E** and **F**) Structural comparison of H80R (PDB 5FUQ) and WT (PDB 2F1O) shows the preservation of the NQO1 overall fold (E) and the local stabilization exerted in the environment of the mutated site (F; WT in grey and H80R in pink).

occurs between Ser72 and Val73 (18). The H80R mutation has no effect on the sensitivity of P187S towards proteolysis (Fig. 1C and D), showing that the local stabilization of H80R on the NTD does not propagate to the cleavage site of P187S found at its CTD (18). To provide structural insight for the stabilizing effect of H80R, we have solved the structure of H80R in the presence of FAD and dicoumarol (a competitive inhibitor of NADH; (41)) (Fig. 1E and F, Supplementary Material, Table S3). The presence of the H80R mutation does not affect the overall fold of NQO1, with a RMSD of 0.42 Å (H80R, vs. WT, PDB ID: 2F1O, using PDBeFold (50)), showing two molecules of FAD and dicoumarol per dimer, and a third molecule of the inhibitor is found at the dimer interface (Fig. 1E). Interestingly, Arg80 undergoes a transition to a less solvent exposed environment than that of His80 (14% vs. 30%), and establishes a salt bridge with Glu78 (at 2.85 Å)

and other electrostatic interactions mainly involving Lys59, and a water molecule (W88) hydrogen-bonded to Glu71 (not present in the WT enzyme; Fig. 1F). Therefore, this cluster of electrostatic interactions mostly involving Glu71, Glu78 and Arg80 (hereafter named the 'EER' cluster) appears to be critical for the local stabilization exerted by H80R. Importantly, the EER cluster seems to stabilize the adjacent loop 57–66, which is highly dynamic in the P187S polymorphism and involved in its lower affinity for FAD (18).

Arg80 is found in most of the closely related mammalian sequences, suggesting that its stabilizing effect has disappeared recently in an evolutionary time scale. Specifically, this implies that the stabilizing effect of the consensus state at this position (i.e. Arg80) was present in mammalian ancestors of NQO1 and recently diverged to the non-consensus state found in the
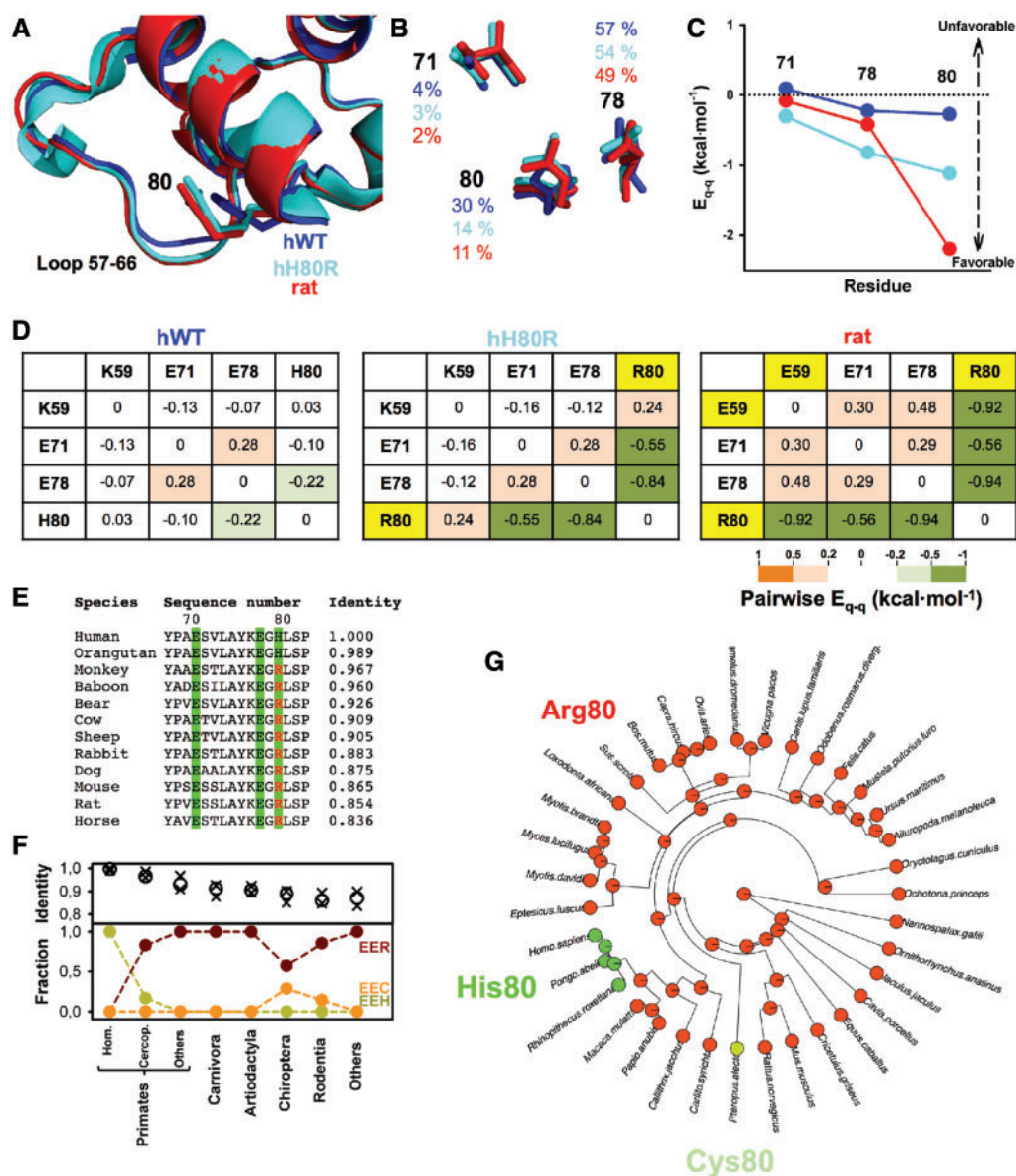
Figure 2. Overall conservation of the stabilizing effect of Arg80 across species and its divergence along primate evolution. (**A**, **B**) Structural overlay of human NQO1 WT (hWT, blue, PDB 2F1O) and H80R (hH80R, cyan, PDB 2FUQ) and rat NQO1 (red, PDB 1QRD), showing the environment of the 80 residue and the loop 57-66 (A) and the EER cluster (B; values indicate solvent accesibilities); (**C**) Total electrostatic interaction energies for individual residues of the EER cluster; (**D**) Pairwise electrostatic interaction energies (in kcal·mol^{-1}) between residue 59 and the EER cluster (residues in yellow are those diverging from human NQO1; the green and red color scale indicates the stabilizing/destabilizing nature of the interaction for visual aid); (**E**, **F**) Conservation of the EER cluster among representative individual mammalian sequences (panel E) and a set of fifty-three mammalian sequences (see Supplementary Materials) grouped by order/family as sequences diverge in identity (F, circles indicate mean values and crosses their corresponding ranges); (**G**) Ancestral sequence reconstruction for the Arg-to-His transition at position 80.

human enzyme (Fig. 2). Consequently, Arg80 is present in rat NQO1, and its available X-ray crystal structure allows direct comparison of the structural consequences of the His-to-Arg transition in these orthologues with high (85%) sequence identity. Structural superposition of human WT and H80R enzymes with rat NQO1 show that the local structural switch triggered by Arg80 in the human enzyme is found in the murine enzyme (Fig. 2A and B). In addition, the EER cluster is also structurally conserved in the murine protein, and Arg80 causes a similar local stabilization of the structure to that found in human H80R. The electrostatic local stabilization due to the EER cluster amounts to -1.8 and -2.3 kcal·mol^{-1} for human H80R and rat NQO1 compared to the human WT enzyme (Fig. 2C). Overall,

Arg80 develops more stabilizing electrostatic interactions with its environment in rat than in human NQO1 (Fig. 2C and D), mostly due to the presence of a non-conserved Glu59 in the rat enzyme which is a Lys in human NQO1 and most of the mammalian sequences. Nevertheless, a comparison of pairwise electrostatic interactions within the EER cluster (including Lys/Glu59), clearly shows that the local stabilizing role of the EER cluster found in human NQO1-H80R is present in the rat enzyme (Fig. 2D). This EER cluster contains the largely stabilizing Arg80 in most of mammalian sequences of different orders, including different primate families, but sharply shifts to His80 in the hominidae family (Fig. 2E and F), indicating that the Arg-to-His mutation has been recently fixed along primate evolution.
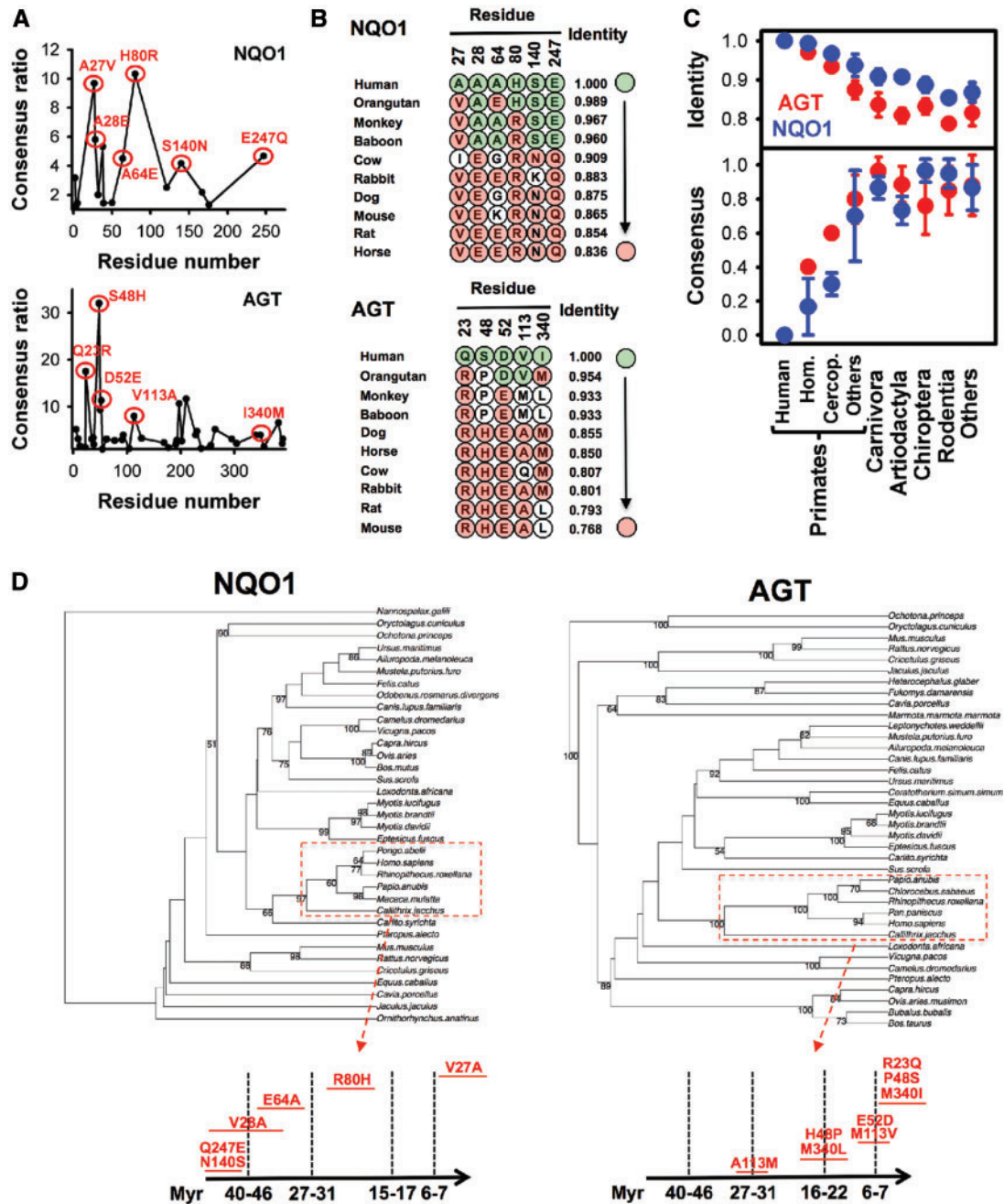
Figure 3. Recent evolutionary divergence of stabilizing consensus amino acids in NQO1 and AGT. (**A**) Consensus analyses performed using a similar set of mammalian sequences for NQO1 and AGT; (**B,C**) Divergence of the relevant set of consensus amino acids in highly similar mammalian sequences; in panel B, divergence is shown for individual sequences, in panel C as clustered in orders/families as the identity decreases (values are mean±s.d.; fifty-three sequences were used overall, see Supplementary Materials); (**D**) Phylogenetic trees based on sequences of NQO1 (left) and AGT (right). Numbers at some branches show the bootstrap values of 100 replicates. Only values above 50 are indicated. The lower panels show the approximate times at which mutations have occurred according to ancestral sequence reconstruction.

Ancestral sequence reconstruction indicates that the Arg-to-His transition likely occurred after divergence of Cercopithecoidea and Hominoidea superfamilies from a common ancestor about 30 Myr ago (Fig. 2G).

To test whether similar events to the recent Arg-to-His transition at position 80 in mammalian NQO1 may have occurred in other consensus amino acids as well as for other proteins, we have performed consensus analyses using essentially the same set of mammalian species for NQO1 with human alanine:glyoxylate aminotransferase 1 (hAGT). hAGT is associated,

due to inherited mutations and the polymorphism P11L, with the life-threating disease *primary hyperoxaluria type I* (OMIM #259900), in which genetic variations cause loss-of-function of hAGT activity notably due to protein mitochondrial mistargeting and peroxisomal aggregation (23,24). We have recently performed a detailed characterization of the effect of consensus mutations in hAGT, including quantitative stability and structural analyses for single- and multiple-consensus mutants (28). As shown in Figure 3A, consensus analyses reveal a similar set of stabilizing mutations to that previously found using

somewhat different groups of mammalian and eukaryotic sequences and experimentally validated for hAGT (28). Interestingly, murine orthologues of human NQO1 and hAGT (for which there are X-ray structures) display a large fraction of these consensus amino acids (Fig. 3B).

Detailed sequence, phylogenetic and structural analyses support the proposal that the fixation of a non-consensus (potentially disease-predisposing) amino acid in human NQO1 (i.e. His80) is not an unusual or isolated event (Fig. 3, Supplementary Materials, Figs. S1 and S2). Consensus amino acids are found in mammalian sequences with very high identity to the human orthologue, and divergence at these sites and consequent loss of stability seem to have occurred particularly along primate evolution (Fig. 3B and C). Indeed, ancestral sequence reconstruction analyses support that about 70–80% of the mammalian consensus amino acids analysed in NQO1 and AGT have been progressively replaced through the last 40 Myr (Fig. 3D).

Our interpretation on the loss of stability due to divergence from consensus amino acids implies that their stabilizing effects on human NQO1 (this work) and hAGT (28) are conserved between highly similar orthologue proteins, a plausible assumption considering that site specific amino acid preferences are often highly conserved along evolution (39,51). The available crystal structures for human and murine enzymes containing consensus amino acids at these sites provide further structural and energetic support for the interspecies conservation of the local stabilization exerted by consensus amino acids. Human and murine WT proteins, as well as human enzymes containing a single (NQO1-H80R) or multiple (hAGT-RHEAM) consensus mutations share the overall fold (Supplementary Materials, Figs S1A and S2A). In human NQO1, all six consensus mutations are individually stabilizing (Fig. 1A), and only the adjacent A27V and A28E mutations are spatially close (Supplementary Material, Fig. S1B), while the five consensus mutations on human AGT are individually stabilizing and their effects are additive (28), with four of them forming an intersubunit cluster of stabilizing interactions (Supplementary Material, Fig. S2B). With a few exceptions, the consensus mutations do not cause large changes in the solvent accessibility (Supplementary Materials, Figs S1C and S2C), and a significant fraction of them show an electrostatic origin for their stabilizing effects (Supplementary Materials, Figs S1D and S2D). In rat NQO1, the consensus amino acids Arg80 and Glu28 combined cause a $\approx 3\,\text{kcal}\cdot\text{mol}^{-1}$ stabilization compared to His80 and Ala28 in the human enzyme, due to local optimization of electrostatic interactions (Supplementary Material, Fig. S1D and E and Fig. 2). In mouse AGT, the consensus amino acids Arg23 and Glu52 create an electrostatic cluster of favourable interactions similar to that found in hAGT-RHEAM and including Arg175 and Arg333. This cluster amounts to $3–4\,\text{kcal}\cdot\text{mol}^{-1}$ of electrostatic stabilization, besides some small differences with hAGT-RHEAM such as the absence of Glu344 in the mouse enzyme (Supplementary Material, Fig. S2 D and E and (28)).

The consensus mutation H80R rescues P187S NQO1 activity intracellularly through local stabilization of the FAD binding site without affecting the interaction with p73α

To test whether mammalian consensus amino acids can act as suppressors of disease phenotype, we have further investigated the mutation H80R in the presence of the cancer-associated P187S NQO1 polymorphism. Notably, the local stabilization

exerted by H80R at the NTD affects the environment of the FAD binding site (Fig. 4A), whose destabilization by P187S in the apo-state (particularly at the loop 57–66) causes enzyme inactivation by reducing the affinity for FAD (18). Accordingly, H80R increases the content in FAD of P187S as purified (Fig. 4B), causing a concomitant large increase in specific activity (Fig. 4C) without perturbing other kinetic properties (e.g. the apparent affinities for NADH and DCPIP as substrate; Supplementary Materials, Table S4 and Fig. S3). Results from thermal denaturation experiments with an excess of FAD ruled out that the stabilizing effect of H80R on P187S is due to the presence of higher FAD levels in the proteins as purified (Supplementary Material, Fig. S4). Local stabilization of the FAD binding site of P187S by H80R also translates into a 10-fold higher binding affinity for FAD, with little consequence on the WT enzyme, as shown by fluorescence titrations (Fig. 4D, Supplementary Material, Fig. S5). Molecular dynamics (MD) simulations recapitulate the structural switch caused by H80R (Supplementary Material, Fig. S6) further showing that this switch dynamically stabilizes the loop 57–66, particularly in the apo-state of P187S (Fig. 4E and F).

The local stabilization exerted by H80R on the NTD should increase the FAD binding affinity of P187S, and consequently its activity inside cells, without affecting degradation rates or protein levels (which are largely influenced by the dynamic CTD; (18)). We have thus investigated the effects of H80R on the intracellular activity, protein levels and degradation rates of WT and P187S by combining experiments in stably transfected Caco-2 cells (showing low endogenous levels of NQO1 protein due to being homozygous for P187S; (18)) with pulse-chase experiments in a rabbit reticulocyte cell-free system (Fig. 4G–K). Western-blot analysis of transfected Caco-2 cells reveal a 3-fold increase in protein levels of P187S in the presence of the suppressor H80R mutation, while this effect is weak in WT NQO1 (1.4-fold increase; Fig. 4G and H). The levels of endogenous NQO1 P187S in Caco-2 cells (lower immunodetected bands in Fig. 4G) are somewhat higher when cells are transfected with NQO1 variants containing the H80R mutation, suggesting that the presence of this mutation favors hetero-oligomerization between exogenous (i.e. transfected) and endogenous P187S NQO1. Since degradation of NQO1 proteins upon expression in rabbit reticulocyte extracts is mediated by the proteasome (52), we have used this system to determine whether H80R affects proteasomal degradation rates. The obtained results show that H80R has little effect on the degradation rates of WT and P187S (Fig. 4I), suggesting that the effect of H80R on intracellular protein levels of P187S reflects an improved intracellular folding efficiency. Functional rescue of P187S by H80R is further supported by measurements of NQO1 activity in transfected cell extracts using DCPIP as substrate (Fig. 4J) and corroborated by the increased capacity of intact transfected cells to activate the drug 17-AAG (17-N-allylamino-17-demethoxygeldanamycin) into its cytotoxic form (53) (Fig. 4K, Supplementary Material, Fig. S7). Importantly, NQO1 activity measurements (Fig. 4J and K) correlate well with the levels of transfected NQO1 proteins (Fig. 4G and I), supporting that hetero-oligomerization of transfected proteins with endogenous P187S plays a minor role in the functional rescue exerted by H80R on P187S.

So far, we have shown that the consensus H80R mutation is capable of reactivating P187S NQO1 to a significant extent *in vitro* and in cells likely due to enhanced FAD binding affinity and folding efficiency. To determine whether this consensus mutation might affect further functions, such as their ability to interact with transcription factors (e.g. p73α), we have used NMR spectroscopy to investigate binding of NQO1 variants with the
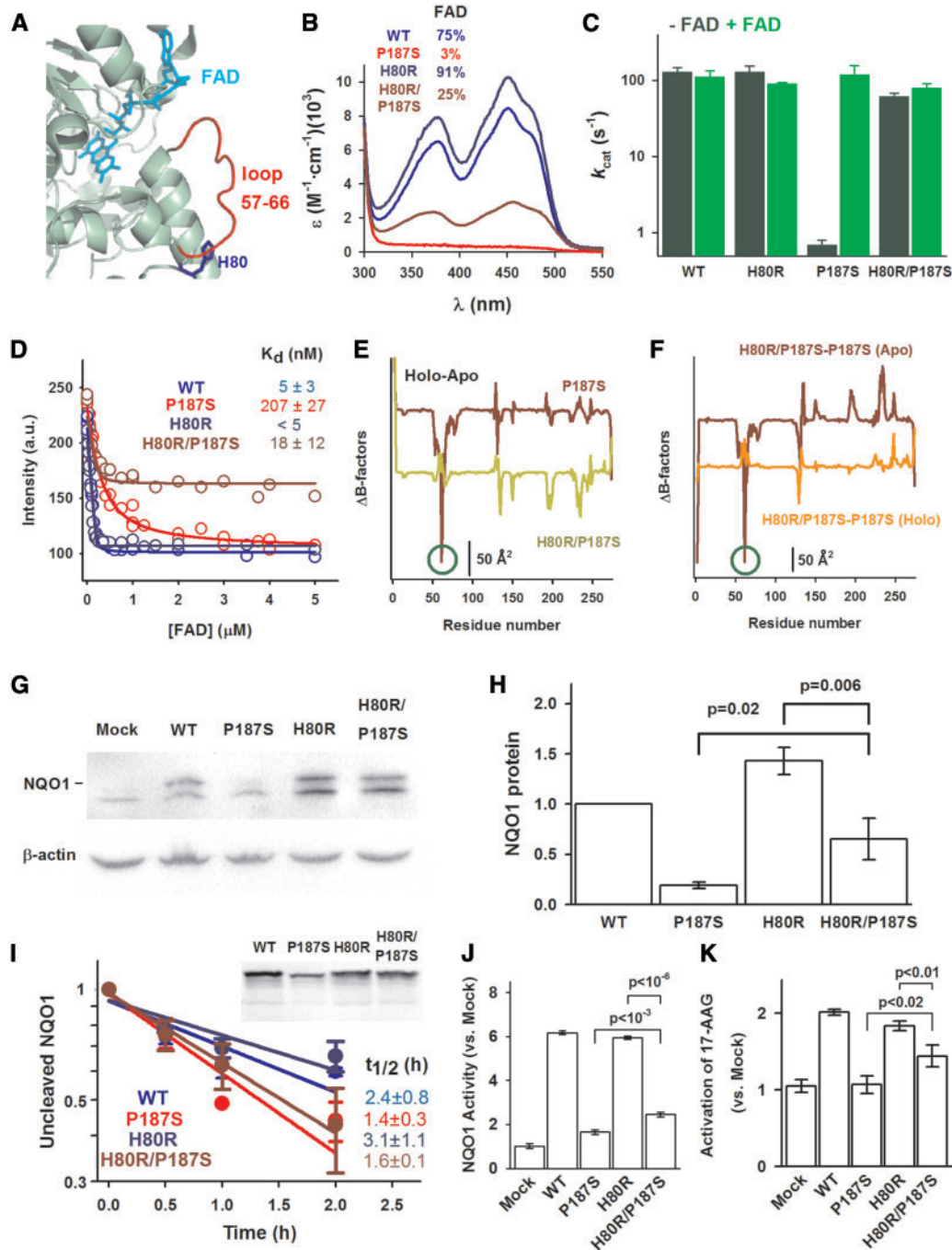
Figure 4. The H80R mutation protects cancer-associated P187S towards inactivation. (**A**) Structural location of the loop 57-66 and His80 relative to the FAD binding site (PDB: 2F10); (**B**) Absorption spectra of FAD bound to NQO1 enzymes as purified; data are the average from two independent purifications; the fractional degree of binding (in %) of FAD to NQO1 proteins is indicated; (**C**) Activity of purified NQO1 enzymes in the absence or presence of added FAD (mean±s.d. from 5 to 6 independent measurements); (**D**) FAD titrations of apo-NQO1 proteins monitored by fluorescence; Data are from two independent experimental series. Best-fit K_d values are displayed; (**E, F**) Dynamic changes at residue level from MD simulations showing specific effects of H80R on the backbone dynamics of apo-P187S (highlighted by green circles). (**G, H**) representative western-blots of Caco-2 cells stably transfected with pCINEO (mock, no insert), pCINEO-NQO1 WT, H80R, P187S, H80R/P187S (panel G). The position of transfected (exogenous) NQO1 in the blots is indicated. In panel H, densitometric analyses of exogenous NQO1 protein normalized using β-actin content (the signal of WT is set to 1 in each experimental series; n = 3; mean±S.D.). Statistical differences are provided by an unpaired t-test; (**I**) chase experiments after a pulse synthesis in rabbit reticulocyte extracts. Fits to a single exponential provide the half-lives for the corresponding NQO1 variants; mean±S.D. from three independent experiments; The inset displays a representative gel for a pulse of synthesis to show that proteins levels are comparable for all variants, although somewhat lower for P187S; (**J**) NQO1 activity in Caco-2 cell extracts transfected with different NQO1 variants and using DCPIP as substrate; (**K**) Capacity of different NQO1 variants to activate 17-AAG, determined as the maximal inhibition of cell viability and normalized using the value of Caco-2 cells transfected with pCINEO (mock)(see Supplementary Material, Fig. S7). Data in (**J** and **K**) are best-fit values±standard errors and statistical significance is provided by a t-test.
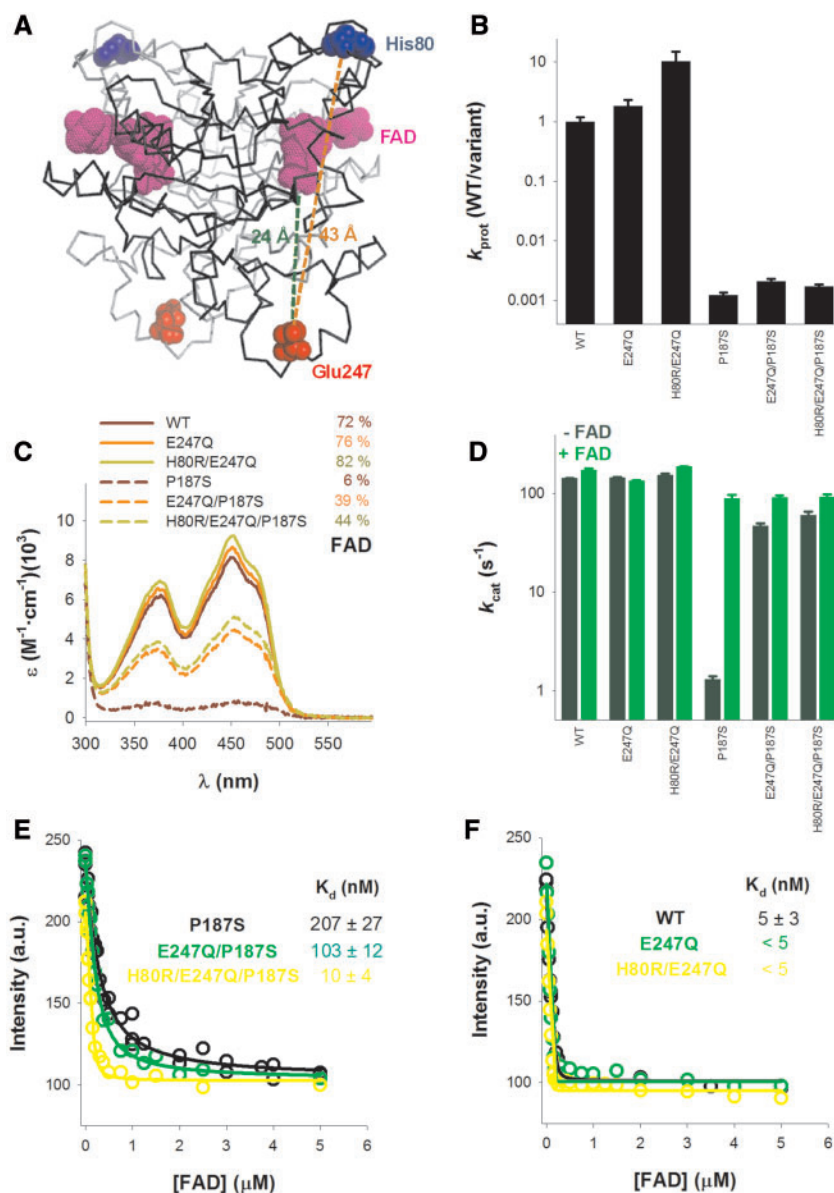
Figure 5. The consensus mutations E247Q and H80R cooperate to rescue P187S NQO1. (**A**) Location of Glu247 far from His80 and the FAD cofactor (PDB 2F1O); (**B**) Effect of the E247Q and H80R mutations on the normalized proteolysis rates by thermolysin; Normalized proteolysis rates were determined as described in Figure 1 and Materials and Methods; (**C**) Absorption spectra of FAD bound to NQO1 enzymes as purified; data are the average from two independent purifications; the fractional degree of binding (in %) of FAD to NQO1 proteins is indicated; (**D**) *In vitro* activity of purified NQO1 proteins in the absence or presence of FAD 1 μM. (**E, F**) Titrations of apo-NQO1 variants (E, P187S variants; F, WT variants) with FAD followed by intrinsic fluorescence. Data are from two independent experimental series. Best-fit K_d values are displayed.

C-terminal domain of p73α (SAMp73α) (49). Even though small quantitative differences are observed between NQO1 variants, these analyses support that the core of the residues involved in NQO1:SAMp73α interaction is not greatly affected by the H80R mutation (Supplementary Materials, Fig. S8 and Table S5).

E247Q cooperates with H80R to rescue NQO1 P187S through long-range epistatic interactions

During the evolution of proteins, the effects of a single mutation often depend on the presence of additional mutations (a type of *epistatic interactions*; (2,4)). In the context of this work, divergence from consensus amino acids at different sites in a short

evolutionary time period (see Fig. 3), may have cooperated to enhance the sensitivity of human NQO1 and AGT towards disease-associated inactivation through this type of interactions. This cooperation is further expected from the additive effects of consensus mutations on protein stability (28,30,54). To test this hypothesis, we have characterized the effects of the consensus NQO1 mutation E247Q, in the absence or presence of H80R (Fig. 5). These two sites are not spatially close (at ∼40 Å) and the Glu247 is also far from the FAD binding site (at ∼25 Å; Fig. 5A). We found that E247Q locally stabilizes the CTD of P187S, decreasing by 2-fold the sensitivity to partial proteolysis, and its effects propagate to the distal NTD enhancing the local stabilization exerted by H80R in the WT enzyme (Fig. 5B). This long-range communication of the stabilizing effect of E247Q

concomitantly increases the FAD content and specific activity of P187S, boosting the effects of H80R (Fig. 5C and D). FAD titrations confirmed stronger binding of FAD to P187S in the presence of E247Q, particularly in combination with H80R (Fig. 5E, Supplementary Material, Fig. S5), while very tight FAD binding was found for all variants of WT NQO1 (Fig. 5F, Supplementary Material, Fig. S5). As described earlier in this work for H80R, the presence of E247Q did not largely affect the core of residues in SAMp73α involved in the formation of its complex with NQO1 (Supplementary Materials, Fig. S8 and Table S5).

These results demonstrate that the additive stabilizing effects of consensus mutations can propagate to distal functional sites, in line with our recent findings on the long-range communication existing between the Pro187 site, the FAD binding site and the CTD, in particular when the CTD is withdrawn (49). Surprisingly, the epistatic effects described here for H80R and E247Q mutations support directionality in their propagation (in agreement with some ensemble descriptions of allosteric mutational effects; see (55)): the H80R mutation locally stabilizes the NTD, with no apparent effects on the stability or dynamics of the CTD (Fig. 1C and D), while the E247Q mutation stabilizes the CTD and the distal NTD, and importantly, enhances the local stabilization exerted by H80R on the NTD (Fig. 5B).

Discussion

We have provided evidence supporting that consensus amino acids among mammalian sequences have diverged over the last 50 Myr, decreasing the local stability of some human proteins and potentially predisposing them towards disease-associated, destabilizing and inactivating single amino acid changes. Due to the conservation of the stabilizing effect of consensus amino acids, at least in protein orthologues with highly similar sequences, their additive effects on stability, and the existence of long-range epistatic interactions between some consensus amino acids, it is intriguing to propose that some extant mammalian proteins (such as those from rodents) might be more robust towards disease-associated and destabilizing single amino acid replacements. Importantly, we demonstrate the potential of consensus amino acids to protect towards disease-associated single amino acid changes. This case also exemplifies well how deeply evolution can affect the protein conformational landscape through single mutations altering protein functionality *in vivo*, and particularly, the key role of long-range propagation of structural and dynamic effects to different functional sites. An interesting application of consensus mutations as disease suppressors could be the identification of druggable spots in proteins, i.e. to mimic the suppressor effect of the mutation using small pharmacological ligands.

From the perspective of natural selection, it is difficult to conceive how destabilizing mutations diverging from consensus amino acids are fixed along evolution, unless the mutated sites are under low selective pressure. In this case, mutations diverging from the consensus amino acid would be eventually fixed if they have little or no effect on protein function *in vivo* (or ultimately, in the reproductive capacity of the species), therefore constituting networks of neutral mutations (or *neutral networks*; (2)). Indeed, this *neutral network* scenario explains well the experimentally observed effects on human NQO1 and AGT proteins (this work and (28)). In the case of NQO1, a paradigm of multi-functional stress protein, mutation of the consensus Arg80 (to His) and Gln247 (to Glu) have little or no effect in many different molecular traits, such as its specific activity, FAD content, intracellular stability and folding efficiency, or interaction

with p73α. For human AGT, the simultaneous absence of up to five consensus amino acids only slightly affects its specific activity and has no impact on the intracellular folding, stability and peroxisomal import upon interaction with the peroxisomal receptor Pex5p (28). It can also be argued that the divergence of some consensus amino acids in these human proteins has been driven by natural selection, for instance by fine-tuning of some intracellular functions, unfortunately not yet identified by the functional analyses performed so far.

Our results with NQO1 have also important implications for understanding the stability and dynamics of the human flavo-proteome. A recent comprehensive proteomic analysis has shown that the stability of human flavo-proteins is strongly dependent on the bioavailability of the flavin cofactor, and NQO1 has emerged as a particularly sensitive case (45). Our results further highlight the importance of the apo-state ensemble, in particular its structure, stability and dynamics (18,42–45,49,56), in determining the cofactor binding affinity of human flavo-proteins, and consequently, their function and stability *in vivo*. This can be of particular relevance to understand the large number of metabolic diseases associated with flavo-proteins (linked to about 60% of all human flavo-proteins, based on a recent and exhaustive analysis (57)), and the significant fraction of them associated with riboflavin supplementation (57,58).

Materials and Methods

Consensus and phylogenetic analyses

The sequences of mammalian NQO1 and AGT enzymes were retrieved using BlastP and the human sequences as queries. This set of sequences (see Supplementary Materials) was used to identify consensus amino acids as those displaying the highest frequencies in the alignment. Therefore, the consensus ratio was defined as the number of sequences containing the most common amino acid in the alignment divided by the number of sequences containing the amino acid found in the human protein at the same site. In NQO1, the six consensus mutations displaying consensus ratios higher or equal to 5 were selected for further characterization.

Phylogenetic trees based on the sequences of the NQO1 and AGT proteins were made according to the UPGMA method with the phangorn package (59) from the gnu-R statistical software. Analysis of time divergence was performed according to (60,61).

Materials

Antibiotics, HEPES, cacodylate, KOH, DCPIP, FAD, NADH and thermolysin from *Bacillus thermoproteolyticus rokko* were purchased from Sigma Aldrich (Madrid, Spain). Chromatographic columns for protein purification were from GE Healthcare (Barcelona, Spain). Concentration devices (cut offs of 3–30 kDa) were from Millipore (Madrid, Spain) or Sartorius (Madrid, Spain). Other chemicals were purchased from standard suppliers.

Mutagenesis, expression and purification of NQO1 variants in *E. coli*

Site-directed mutagenesis was performed using the QuickChange lightening kit (Agilent Technologies, Madrid, Spain) by standard protocols and confirmed by sequencing. Expression in and purification from *E. coli* BL21 (DE3) was performed as described (43).

In vitro characterization of NQO1 proteins

NQO1 activity was measured by following the reduction of DCPIP as described (49). FAD content in purified NQO1 samples was estimated from UV-visible absorption spectra. Briefly, NQO1 proteins in K-HEPES 50 mM pH 7.4 were prepared $\sim$0.6 mg·mL^{-1} and the absorption spectra registered in a HP 8453 spectrophotometer (Agilent) at 25 °C. The corresponding spectra were converted into molar extinction units using a corrected extinction coefficient at 280 nm for each purified NQO1: $\epsilon_{280} = 47900 \cdot \left(1 + \frac{22000}{11300} \cdot A_{450}\right)$, that considers the contribution from apo-NQO1 ($\epsilon_{280} = 47900$ M^{-1}·cm^{-1}, derived from the primary sequence; (43)) and the FAD bound assuming that spectroscopic features of the cofactor do not largely depart from those in the free form ($\epsilon_{280} = 22000$ and $\epsilon_{450} = 11300$ M^{-1}·cm^{-1}; (43)). This procedure yields estimates for the FAD content in purified samples by using a reference value of $\epsilon_{450} = 11300$ M^{-1}·cm^{-1} for 100% saturation.

Thermal denaturation of NQO1 proteins was determined in K-HEPES 50 mM pH 7.4 at a protein concentration of 2 μM in the presence of the fluorescent probe SYPRO Orange (Life Technologies, Madrid, Spain). Thermal scans were registered in a IQ5 Real-time PCR detection system (Biorad, Madrid, Spain) by monitoring fluorescence upon excitation at 460 nm and emission at 510 nm at a scan rate of 0.5 °C·min^{-1}. The melting temperature was determined from the maximum of the first derivative (T_{der}) or as the half-denaturation temperature upon normalization of the unfolding curves ($T_{0.5}$) as described (21). Additionally, thermal stability of NQO1 proteins in the absence or presence of FAD (2 μM NQO1 monomer and 10 μM FAD) was determined by following intrinsic fluorescence (exc. 280 nm, em. 340 nm; slits 5 nm) in K-HEPES 50 mM pH 7.4. Thermal scans were registered using 0.3 cm path-length quartz cuvettes in a Cary-Eclipse spectrofluorimeter (Agilent Technologies), using 2 °C·min^{-1} scan rate and T_m values determined as described above ($T_{0.5}$ procedure). Data are reported as mean ± s.d. from three independent scans. Inactivation kinetics was determined by incubation of protein samples (40 μM) for different times at 42 °C, chilling on ice and measurement of residual activity as described (49).

Proteolysis kinetics by thermolysin were performed as recently described (18). Protease concentration was 100–200 nM (WT variants) or 0.5–1 nM (P187S variants). For each variant, two independent experiments were performed and data of intact NQO1 protein vs. time decays were used to determine first-order rate constants from exponential fits. These constants were divided by the thermolysin concentration used to yield second-order proteolysis rate constants (k_{prot}). For sake of comparison, ratios of k_{prot} value for WT and the corresponding variant are provided, which indicate a stabilizing effect when these are larger than one and a destabilizing effect when lower than one.

Fluorescence titrations were performed at 25 °C using 1 cm path-length cuvettes in a Cary Eclipse spectrofluorimeter (Agilent Technologies) and a final volume of 2 ml. FAD was prepared in K-HEPES 50 mM pH 7.4 (at a final concentration 0–5 μM) in 1.9 ml and incubated for 5 min in the cuvette while the fluorescence of these blanks (exc. 295 nm; em.: 340 nm; slits 5 nm) was registered. Then, 100 μl of apo-NQO1 (4 μM in monomer) was added, manually mixed, and fluorescence was measured for 10 min to ensure proper equilibration. The fluorescence intensity of the last 60 s was averaged and the signal of the blank

without protein was subtracted. Control experiments in the absence of FAD showed that apo-NQO1 did not significantly denature under these experimental conditions. For each protein variant, at least two different protein preparations and FAD stock solutions were used. K_d values were determined from fittings to 1:1 binding model using the following expression:

$$F = F_{apo} + \left(F_{holo} - F_{apo}\right) \cdot \left(\frac{[E] + [FAD] + K_d - \sqrt{([E] + [FAD] + K_d)^2 - 4 \cdot [E] \cdot [FAD]}}{2 \cdot [E]}\right)$$

where F is the fluorescence intensity of the sample at a given total FAD concentration ([FAD]), F_{holo} and F_{apo} are the intensities of the holo- and apo-proteins and $[E]$ is the total concentration of NQO1 (0.2 μM in monomer). In all cases, fittings using this model provided a good description of the experimental data within the experimental uncertainty (Supplementary Material, Fig. S5).

The interaction of NQO1 and SAMp73α was investigated by NMR spectroscopy. NMR data were acquired at 20 °C on a Bruker Avance DRX-500 spectrometer equipped with a triple-resonance probe and z-gradients. Samples containing NQO1 proteins (at $\sim$300 μM, in monomer units) and ^{15}N-labelled SAMp73α ($\sim$120 μM) were prepared in 50 mM phosphate buffer pH 6.9. The 2D ^{1}H-^{15}N HSQC (heteronuclear single-quantum coherence) experiments (62) were acquired in the phase sensitive mode. Experiments were acquired and analysed as described (49).

Crystallography

Crystallization trials were carried out via the hanging-drop vapour diffusion method using previously reported conditions (63) at pH 7–9 and crystal improvement was achieved by the capillary counter diffusion technique (64) set-up using the Domino Granada Crystallization Boxes® (Triana Science & Technology, Granada, Spain). Good quality diffracting crystals were obtained using 30% of PEG 3350, 200 mM sodium acetate, 100 mM sodium tricine pH 9.0, 1% DMSO and 10 mM dicoumarol, as precipitant agent. NQO1 H80R (70 mg·ml^{-1}) mixed with agarose (0.05% w/v) was used to fill capillaries of 0.2 mm inner diameter. Crystals were equilibrated for 24 h with the mother liquid supplemented with 15% (v/v) glycerol before being extracted from the capillaries and flash-cooled in liquid nitrogen. Data were collected at ID-29 beam-line of the European Synchrotron Radiation Facility (ESRF, Grenoble, France) from flashed cooled crystals. Data were indexed and integrated using XDS (65) and scaled with SCALA from the CCP4 suite (66). The structure was determined by molecular replacement using the structure of WT NQO1 (PDB: 1D4A) without waters, as search model. The molecular replacement solution was found using Phaser (67) locating the two monomers in the asymmetric unit. Structure refinement was done with phenix.refine (68) with cycles of manual building steps and ligand identification performed with Coot (69). Titration-Libration-Screw (TLS) was included in the last steps of refinement. Model quality was checked using MolProbity (70) implemented within the Phenix suite (68). Coordinates and structure factors have been deposited at the PDB with accession code 5FUQ. Figures were prepared with PyMOL (71). Details of data collection and processing, refinement statistics and quality

indicators of the final model are summarized in Supplementary Material, Table S3.

Electrostatic calculations

Calculation of the energy of charge–charge interactions ($E_{q\text{-}q}$) was performed using the solvent-accessibility-corrected Tanford–Kirkwood model (72). The input for these calculations were the atomic coordinates of: human NQO1 WT (2F1O) and H80R (2FUQ), rat NQO1 (1QRD), human AGT WT (1H0C) and RHEAM (4CBS), and mouse AGT (3KGX). We used an in-house software (kindly provided by Prof. Jose Manuel Sanchez-Ruiz, Department of Physical Chemistry, University of Granada) at pH 7.4 and 0.025 M (for NQO1) or 0.2 M (for AGT) ionic strength. The output of these analyses provides the energy of charge–charge interactions of a given ionizable residue with all or individual ionizable residues in the protein dimer, and are expressed per ionizable residue.

MD simulations

MD simulations were performed as recently described for WT and P187S NQO1 (18). The structures containing the mutation H80R were prepared using MOE (73) based on a crystal structure of the NQO1 complex with FAD (PDB: 1D4A) (63). The side-chain of Arg80 was modelled using two distinct conformations, solvent-exposed and partially buried. Nevertheless, after an extensive equilibration protocol (74) and a 100 ns sampling simulation, we observed a consistent orientation for Arg80 in the buried state after few ns. Analyses of MD trajectories were performed as recently described (18).

NQO1 expression in Caco-2 cells

Caco-2 cells were grown and maintained in Dulbecco's Modified Eagle Medium (Lonza, Barcelona, Spain) supplemented with 10% heat inactivated fetal bovine serum (HyClone, GE Healthcare, Barcelona, Spain), 100 U·ml^{-1} penicillin and 100 µg·ml^{-1} Streptomycin (Sigma Aldrich, Madrid, Spain) and cultured at 37 °C in a humidified incubator with 5% CO$_2$. Cells were transfected using Lipofectamine LTX with Plus Reagent (Thermo Fisher Scientific, Madrid, Spain) according to manufacturer's protocol and pCINeo plasmids containing the cDNA of NQO1 enzymes and selected using 400 µg·ml^{-1} of G418 (Sigma Aldrich) for a month.

For immunoblotting, cells were scrapped and lysed in RIPA buffer (50 mM Tris-HCl, 150 mM NaCl, 0.1% Triton X-100, 0.1% sodium dodecyl sulphate, 1 mM sodium orthovanadate, 1 mM NaF pH 8) with protease inhibitors (COMPLETE, from Roche, Spain). After centrifugation at 20000 g for 30 min, soluble extracts were collected and the amount of total protein determined by the BCA method (Pierce). Samples were denatured with Laemmli's buffer under reducing conditions, resolved using 12% SDS-PAGE and transferred to polyvinylidene difluoride membranes (GE Healthcare) using standard procedures. Immunoblotting was carried out using primary monoclonal antibodies anti-NQO1 and anti-β-actin (Santa Cruz Biotechnology) from mouse at 1:500 and 1:10000 dilutions, respectively. As a secondary antibody, we used chicken anti-mouse IgG-HRP (Santa Cruz Biotechnology) at 1:2000 dilution. Protein bands were visualized using luminol-based enhanced chemiluminiscence and images were acquired with ChemiDoc MP imaging system and analysed using Image Lab 5.0 software (both from BioRad Laboratories).

For NQO1 activity measurements, cells were collected in 50 mM K-HEPES pH 7.4 with protease inhibitors (COMPLETE, Roche) and lysed by freezing-thawing cycles. Soluble extracts were obtained upon centrifugation at 18000 g for 15 min at 4 °C and total protein content was determined by the BCA method (Pierce) and immediately used for measurements. Activity measurements were carried out as described (18,49) using DCPIP (75 µM) as electron acceptor and 0.5 mM NADH. Upon blank subtraction, the activity was normalized by the amount of total protein used in each measurement (50–130 µg).

For 17-AAG activation assays, cells were dispensed in 96 well plates at a density of 2·10^4 cells/well. After 24 h, cells were treated with 0–30 µM 17-AAG (Santa Cruz Biotechnology). After two days of incubation, 20 µL of 0.125 mg·ml^{-1} of Alamar Blue solution (Sigma-Aldrich) was added to each well and incubated for 12 h prior to absorbance measurement at 600 nm. Maximal inhibition of cell viability was determined from non-linear regression analysis of dose-response experiments.

NQO1 expression in a cell-free system

Pulse-chase experiments were performed using a TnT rabbit reticulocyte cell-free system (Promega, Madrid, Spain) at 30 °C as described in (49).

Supplementary Material

Supplementary Material is available at *HMG* online.

Acknowledgements

Funding

References

1. Tokuriki, N. and Tawfik, D.S. (2009) Protein dynamism and evolvability. *Science*, **324**, 203–207.
2. Soskine, M. and Tawfik, D.S. (2010) Mutational effects and the evolution of new protein functions. *Nat. Rev. Genet.*, **11**, 572–582.
3. Tokuriki, N., Stricher, F., Schymkowitz, J., Serrano, L. and Tawfik, D.S. (2007) The stability effects of protein mutations appear to be universally distributed. *J. Mol. Biol.*, **369**, 1318–1332.
4. Gong, L.I., Suchard, M.A. and Bloom, J.D. (2013) Stability-mediated epistasis constrains the evolution of an influenza protein. *Elife*, **2**, e00631.

5. Tokuriki, N. and Tawfik, D.S. (2009) Chaperonin overexpression promotes genetic variation and enzyme evolution. *Nature*, **459**, 668–673.

6. Wilson, C., Agafonov, R.V., Hoemberger, M., Kutter, S., Zorba, A., Halpin, J., Buosi, V., Otten, R., Waterman, D., Theobald, D.L. *et al.* (2015) Kinase dynamics. Using ancient protein kinases to unravel a modern cancer drug's mechanism. *Science*, **347**, 882–886.

7. Berezovsky, I.N., Guarnera, E., Zheng, Z., Eisenhaber, B. and Eisenhaber, F. (2016) Protein function machinery: from basic structural units to modulation of activity. *Curr. Opin. Struct. Biol.*, **42**, 67–74.

8. Guarnera, E. and Berezovsky, I.N. (2016) Allosteric sites: remote control in regulation of protein activity. *Curr. Opin. Struct. Biol.*, **37**, 1–8.

9. Motlagh, H.N., Wrabl, J.O., Li, J. and Hilser, V.J. (2014) The ensemble nature of allostery. *Nature*, **508**, 331–339.

10. Labbadia, J. and Morimoto, R.I. (2015) The biology of proteostasis in aging and disease. *Annu. Rev. Biochem.*, **84**, 435–464.

11. Kim, Y.E., Hipp, M.S., Bracher, A., Hayer-Hartl, M. and Hartl, F.U. (2013) Molecular chaperone functions in protein folding and proteostasis. *Annu. Rev. Biochem.*, **82**, 323–355.

12. Muntau, A.C., Leandro, J., Staudigl, M., Mayer, F. and Gersting, S.W. (2014) Innovative strategies to treat protein misfolding in inborn errors of metabolism: pharmacological chaperones and proteostasis regulators. *J. Inherit. Metab. Dis.*, **37**, 505–523.

13. McCorvie, T.J., Kopec, J., Pey, A.L., Fitzpatrick, F., Patel, D., Chalk, R., Streetha, L. and Yue, W.W. (2016) Molecular basis of classic galactosemia from the structure of human galactose 1-phosphate uridylyltransferase. *Hum. Mol. Genet.*, **25**, 2234–2244.

14. Pey, A.L., Stricher, F., Serrano, L. and Martinez, A. (2007) Predicted effects of missense mutations on native-state stability account for phenotypic outcome in phenylketonuria, a paradigm of misfolding diseases. *Am. J. Hum. Genet.*, **81**, 1006–1024.

15. Blouin, J.M., Bernardo-Seisdedos, G., Sasso, E., Esteve, J., Ged, C., Lalanne, M., Sanz-Parra, A., Urquiza, P., de Verneuil, H., Millet, O. *et al.* (2017) Missense UROS mutations causing congenital erythropoietic porphyria reduce UROS homeostasis that can be rescued by proteasome inhibition. *Hum. Mol. Genet.*, **26**, 1565–1576.

16. Guharoy, M., Bhowmick, P., Sallam, M. and Tompa, P. (2016) Tripartite degrons confer diversity and specificity on regulated protein degradation in the ubiquitin-proteasome system. *Nat. Commun.*, **7**, 10239.

17. Takahashi, K., Matouschek, A. and Inobe, T. (2015) Regulation of Proteasomal Degradation by Modulating Proteasomal Initiation Regions. *ACS Chem. Biol.*, **10**, 2537–2543.

18. Medina-Carmona, E., Palomino-Morales, R.J., Fuchs, J.E., Padín-Gonzalez, E., Mesa-Torres, N., Salido, E., Timson, D.J. and Pey, A.L. (2016) Conformational dynamics is key to understanding loss-of-function of NQO1 cancer-associated polymorphisms and its correction by pharmacological ligands. *Sci. Rep.*, **6**, 20331.

19. Berko, D., Tabachnick-Cherny, S., Shental-Bechor, D., Cascio, P., Mioletti, S., Levy, Y., Admon, A., Ziv, T., Tirosh, B., Goldberg, A.L. *et al.* (2012) The direction of protein entry into the proteasome determines the variety of products and depends on the force needed to unfold its two termini. *Mol. Cell*, **48**, 601–611.

20. Pey, A.L. and Martinez, A. (2007) Tetrahydrobiopterin for patients with phenylketonuria. *Lancet*, **370**, 462–463.

21. Pey, A.L., Ying, M., Cremades, N., Velazquez-Campoy, A., Scherer, T., Thony, B., Sancho, J. and Martinez, A. (2008) Identification of pharmacological chaperones as potential therapeutic agents to treat phenylketonuria. *J. Clin. Invest.*, **118**, 2858–2867.

22. Majtan, T., Pey, A.L., Ereño-Orbea, J., Martinez-Cruz, L.A. and Kraus, J.P. (2016) Targeting Cystathionine Beta-Synthase Misfolding in Homocystinuria by Small Ligands: State of the Art and Future Directions. *Curr. Drug Targets*, **17**, 1455–1470.

23. Salido, E., Pey, A.L., Rodriguez, R. and Lorenzo, V. (2012) Primary hyperoxalurias: Disorders of glyoxylate detoxification. *Biochim. Biophys. Acta*, **1822**, 1453–1464.

24. Oppici, E., Montioli, R. and Cellini, B. (2015) Liver peroxisomal alanine:glyoxylate aminotransferase and the effects of mutations associated with Primary Hyperoxaluria Type I: An overview. *Biochim. Biophys. Acta*, **1854**, 1212–1219.

25. Gabaldon, T. and Pittis, A.A. (2015) Origin and evolution of metabolic sub-cellular compartmentalization in eukaryotes. *Biochimie*, **119**, 262–268.

26. Martin, W. (2010) Evolutionary origins of metabolic compartmentalization in eukaryotes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, **365**, 847–855.

27. Serrano, L., Day, A.G. and Fersht, A.R. (1993) Step-wise mutation of barnase to binase. A procedure for engineering increased stability of proteins and an experimental analysis of the evolution of protein stability. *J. Mol. Biol.*, **233**, 305–312.

28. Mesa-Torres, N., Yunta, C., Fabelo-Rosa, I., Gonzalez-Rubio, J.M., Sanchez-Ruiz, J.M., Salido, E., Albert, A. and Pey, A.L. (2014) The consensus-based approach for gene/enzyme replacement therapies and crystallization strategies: the case of human alanine:glyoxylate aminotransferase. *Biochem. J.*, **462**, 453–463.

29. Pey, A.L., Rodriguez-Larrea, D., Bomke, S., Dammers, S., Godoy-Ruiz, R., Garcia-Mira, M.M. and Sanchez-Ruiz, J.M. (2008) Engineering proteins with tunable thermodynamic and kinetic stabilities. *Proteins*, **71**, 165–174.

30. Nikolova, P.V., Henckel, J., Lane, D.P. and Fersht, A.R. (1998) Semirational design of active tumor suppressor p53 DNA binding domain with enhanced stability. *Proc. Natl. Acad. Sci. U. S. A*, **95**, 14675–14680.

31. Steipe, B. (2004) Consensus-based engineering of protein stability: from intrabodies to thermostable enzymes. *Methods Enzymol.*, **388**, 176–186.

32. Trudeau, D.L., Kaltenbach, M. and Tawfik, D.S. (2016) On the Potential Origins of the High Stability of Reconstructed Ancestral Proteins. *Mol. Biol. Evol.*, **33**, 2633–2641.

33. Joerger, A.C., Allen, M.D. and Fersht, A.R. (2004) Crystal structure of a superstable mutant of human p53 core domain. Insights into the mechanism of rescuing oncogenic mutations. *J. Biol. Chem.*, **279**, 1291–1296.

34. Godoy-Ruiz, R., Perez-Jimenez, R., Ibarra-Molero, B. and Sanchez-Ruiz, J.M. (2005) A stability pattern of protein hydrophobic mutations that reflects evolutionary structural optimization. *Biophys. J.*, **89**, 3320–3331.

35. Risso, V.A., Gavira, J.A., Gaucher, E.A. and Sanchez-Ruiz, J.M. (2014) Phenotypic comparisons of consensus variants versus laboratory resurrections of Precambrian proteins. *Proteins*, **82**, 887–896.

36. Risso, V.A., Gavira, J.A., Mejia-Carmona, D.F., Gaucher, E.A. and Sanchez-Ruiz, J.M. (2013) Hyperstability and substrate promiscuity in laboratory resurrections of Precambrian beta-lactamases. *J. Am. Chem. Soc.*, **135**, 2899–2902.

37. Akanuma, S., Nakajima, Y., Yokobori, S., Kimura, M., Nemoto, N., Mase, T., Miyazono, K., Tanokura, M. and Yamagishi, A. (2013) Experimental evidence for the thermophilicity of ancestral life. *Proc. Natl. Acad. Sci. U. S. A*, **110**, 11067–11072.

38. Wheeler, L.C., Lim, S.A., Marqusee, S. and Harms, M.J. (2016) The thermostability and specificity of ancient proteins. *Curr. Opin. Struct. Biol.*, **38**, 37–43.

39. Risso, V.A., Manssour-Triedo, F., Delgado-Delgado, A., Arco, R., Barroso-delJesus, A., Ingles-Prieto, A., Godoy-Ruiz, R., Gavira, J.A., Gaucher, E.A., Ibarra-Molero, B. *et al.* (2015) Mutational studies on resurrected ancestral proteins reveal conservation of site-specific amino acid preferences throughout evolutionary history. *Mol. Biol. Evol.*, **32**, 440–455.

40. Lajin, B. and Alachkar, A. (2013) The NQO1 polymorphism C609T (Pro187Ser) and cancer susceptibility: a comprehensive meta-analysis. *Br. J. Cancer*, **109**, 1325–1337.

41. Pey, A.L., Megarity, C.F., Medina-Carmona, E. and Timson, D.J. (2016) Natural small molecules as stabilizers and activators of cancer-associated NQO1 polymorphisms. *Curr. Drug Targets*, **17**, 1506–1514.

42. Lienhart, W.D., Gudipati, V., Uhl, M.K., Binter, A., Pulido, S.A., Saf, R., Zangger, K., Gruber, K. and Macheroux, P. (2014) Collapse of the native structure caused by a single amino acid exchange in human NAD(P)H:quinone oxidoreductase 1. *FEBS. J.*, **281**, 4691–4704.

43. Pey, A.L., Megarity, C.F. and Timson, D.J. (2014) FAD binding overcomes defects in activity and stability displayed by cancer-associated variants of human NQO1. *Biochim. Biophys. Acta*, **1842**, 2163–2173.

44. Moscovitz, O., Tsvetkov, P., Hazan, N., Michaelevski, I., Keisar, H., Ben-Nissan, G., Shaul, Y. and Sharon, M. (2012) A mutually inhibitory feedback loop between the 20S proteasome and its regulator, NQO1. *Mol. Cell*, **47**, 76–86.

45. Martinez-Limon, A., Alriquet, M., Lang, W.H., Calloni, G., Wittig, I. and Vabulas, R.M. (2016) Recognition of enzymes lacking bound cofactor by protein quality control. *Proc. Natl. Acad. Sci. U. S. A*, **113**, 12156–12161.

46. Colucci, M.A., Moody, C.J. and Couch, G.D. (2008) Natural and synthetic quinones and their reduction by the quinone reductase enzyme NQO1: from synthetic organic chemistry to compounds with anticancer potential. *Org. Biomol. Chem.*, **6**, 637–656.

47. Asher, G., Tsvetkov, P., Kahana, C. and Shaul, Y. (2005) A mechanism of ubiquitin-independent proteasomal degradation of the tumor suppressors p53 and p73. *Genes Dev.*, **19**, 316–321.

48. Oh, E.T., Kim, J.W., Kim, J.M., Kim, S.J., Lee, J.S., Hong, S.S., Goodwin, J., Ruthenborg, R.J., Jung, M.G., Lee, H.J. *et al.* (2016) NQO1 inhibits proteasome-mediated degradation of HIF-1alpha. *Nat. Commun.*, **7**, 13593.

49. Medina-Carmona, E., Neira, J.L., Salido, E., Fuchs, J.E., Palomino-Morales, R., Timson, D.J. and Pey, A.L. (2017) Site-to-site interdomain communication may mediate different loss-of-function mechanisms in a cancer-associated NQO1 polymorphism. *Sci. Rep.*, **7**, 44352.

50. Krissinel, E. and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2256–2268.

51. Ashenberg, O., Gong, L.I. and Bloom, J.D. (2013) Mutational effects on stability are largely conserved during protein evolution. *Proc. Natl. Acad. Sci. U S A*, **110**, 21071–21076.

52. Siegel, D., Anwar, A., Winski, S.L., Kepa, J.K., Zolman, K.L. and Ross, D. (2001) Rapid polyubiquitination and proteasomal degradation of a mutant form of NAD(P)H:quinone oxidoreductase 1. *Mol. Pharmacol.*, **59**, 263–268.

53. Guo, W., Reigan, P., Siegel, D., Zirrolli, J., Gustafson, D. and Ross, D. (2005) Formation of 17-allylamino-demethoxygeldanamycin (17-AAG) hydroquinone by NAD(P)H:quinone oxidoreductase 1: role of 17-AAG hydroquinone in heat shock protein 90 inhibition. *Cancer Res.*, **65**, 10006–10015.

54. Rodriguez-Larrea, D., Perez-Jimenez, R., Sanchez-Romero, I., Delgado-Delgado, A., Fernandez, J.M. and Sanchez-Ruiz, J.M. (2010) Role of conservative mutations in protein multi-property adaptation. *Biochem. J.*, **429**, 243–249.

55. Liu, T., Whitten, S.T. and Hilser, V.J. (2007) Functional residues serve a dominant role in mediating the cooperativity of the protein ensemble. *Proc. Natl. Acad. Sci. U. S. A*, **104**, 4347–4352.

56. Lienhart, W.D., Strandback, E., Gudipati, V., Koch, K., Binter, A., Uhl, M.K., Rantasa, D.M., Bourgeois, B., Madl, T., Zangger, K. *et al.* (2017) Catalytic competence, structure and stability of the cancer-associated R139W variant of the human NAD(P)H:quinone oxidoreductase 1 (NQO1). *FEBS. J.*, **284**, 1233–1245.

57. Lienhart, W.D., Gudipati, V. and Macheroux, P. (2013) The human flavoproteome. *Arch. Biochem. Biophys.*, **535**, 150–162.

58. Henriques, B.J., Lucas, T.G. and Gomes, C.M. (2016) Therapeutic Approaches Using Riboflavin in Mitochondrial Energy Metabolism Disorders. *Curr. Drug Targets*, **17**, 1527–1534.

59. Schliep, K.P. (2011) phangorn: phylogenetic analysis in R. *Bioinformatics*, **27**, 592–593.

60. Hedges, S.B. and Kumar, S. (2009). *The timetree of life*. Oxford University Press, New York.

61. Hedges, S.B., Marin, J., Suleski, M., Paymer, M. and Kumar, S. (2015) Tree of life reveals clock-like speciation and diversification. *Mol. Biol. Evol.*, **32**, 835–845.

62. Bodenhausen, G. and Ruben, D. (1980) Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Lett.*, **69**, 185–189.

63. Faig, M., Bianchet, M.A., Talalay, P., Chen, S., Winski, S., Ross, D. and Amzel, L.M. (2000) Structures of recombinant human and mouse NAD(P)H:quinone oxidoreductases: species comparison and structural changes with substrate binding and release. *Proc. Natl. Acad. Sci. U. S. A*, **97**, 3177–3182.

64. Otalora, F., Gavira, J.A., Ng, J.D. and Garcia-Ruiz, J.M. (2009) Counterdiffusion methods applied to protein crystallization. *Prog. Biophys. Mol. Biol.*, **101**, 26–37.

65. Kabsch, W. (2010) Xds. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 125–132.

66. Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A. *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.*, **67**, 235–242.

67. Bunkoczi, G., Echols, N., McCoy, A.J., Oeffner, R.D., Adams, P.D. and Read, R.J. (2013) Phaser.MRage: automated molecular replacement. *Acta Crystallogr. D Biol. Crystallogr.*, **69**, 2276–2286.

68. Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 213–221.

69. Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 486–501.

70. Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S. and Richardson, D.C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 12–21.

71. Schrödinger, L.L.C. (2010) The PyMOL molecular graphic system.

72. Ibarra-Molero, B., Loladze, V.V., Makhatadze, G.I. and Sanchez-Ruiz, J.M. (1999) Thermal versus guanidine-induced unfolding of ubiquitin. An analysis in terms of the contributions from charge-charge interactions to protein stability. *Biochemistry*, **38**, 8138–8149.

73. C.C.G. Inc, (2013) Molecular Operating Environment, Montreal, QC, Canada.

74. Fuchs, J.E., Huber, R.G., von Grafenstein, S., Wallnoefer, H.G., Spitzer, G.M., Fuchs, D. and Liedl, K.R. (2012) Dynamic regulation of phenylalanine hydroxylase by simulated redox manipulation. *PLoS One*, **7**, e53005.

PUBLICATION 4

A mechanism for cancer-associated inactivation of NQO1 due to P187S and its reactivation by the consensus mutation H80R.

A mechanism for cancer-associated inactivation of NQO1 due to P187S and its reactivation by the consensus mutation H80R

Inés G. Muñoz[1], Bertrand Morel[2], Encarnación Medina-Carmona[2] and Angel L. Pey[2] (iD)

1 Crystallography and Protein Engineering Unit, Structural Biology Programme, Spanish National Cancer Research Centre (CNIO), Madrid, Spain
2 Department of Physical Chemistry, Faculty of Sciences, University of Granada, Spain

Correspondence

A. L. Pey, Department of Physical Chemistry, Faculty of Sciences, University of Granada. Av. Fuentenueva s/n, 18071 Granada, Spain
Fax: +34 958 242747
Tel: +34 958 243173
E-mail: angelpey@ugr.es

The cancer-associated P187S polymorphism in the NAD(P)H:quinone oxidoreductase 1 (NQO1) abolishes enzyme activity by strongly reducing FAD binding affinity. A single mammalian consensus mutation (H80R) protects P187S from inactivation. To provide mechanistic insight into these effects, we report here a detailed structural and thermodynamic characterization of FAD binding to these NQO1 variants. Our results show that H80R causes a population shift in the conformational ensemble of apo-P187S, remodelling the structure and dynamics of the FAD-binding site and reducing the energetic penalization arising from the equilibrium between apo- and holo-states. Our analyses illustrate how single amino acid changes can profoundly affect structural and mechanistic features of protein functional traits, with implications for our understanding of protein evolution and human disease.

Keywords: FAD binding thermodynamics; mutations in disease and evolution; structure-energetics relationships

Human NAD(P)H:quinone oxidoreductase 1 (NQO1) is a FAD-dependent, two-domain and dimeric multifunctional protein [1–3]. As an enzyme, it catalyses the two-electron reduction of quinones to form hydroquinones, acting as a detoxifyng and antioxidant enzyme, maintaining the reduced status of some vitamins and activating certain cancer prodrugs [1,2]. In addition, NQO1 interacts with other proteins, stabilizing transcription factors linked to cancer progression, such as p53, p73α and HIF-1α [4,5] and inhibiting the 20S proteasome further protecting these proteins from degradation [4,6]. Furthermore, its role in cancer is supported by its overexpression in certain types of tumours [7–9]. In the context of the flavo-proteome, NQO1 is also a paradigm of protein whose stability strongly depends on the bioavailability of the flavin precursor [10], possibly due to the highly dynamic and unstable nature of its cofactor-free (apo) form [11]. Therefore, a deep knowledge on FAD binding to human NQO1, particularly in structural and energetic terms, can provide clues to understand the protective role that flavin binding exerts in the human flavo-proteome.

A single nucleotide polymorphism in NQO1 (rs1800566/c.C609T/p.P187S) was originally found in cancer cell lines resistant to the antitumour quinone mytomicin C [12]. This polymorphism shows a high frequency in human population and is significantly associated with increased cancer risk [13]. The P187S polymorphism is known to abolish NQO1 activity by strongly decreasing FAD-binding affinity and accelerating protein turnover by the proteasome [3,6,11,14–16]. Intriguingly, the X-ray crystallographic structure

Abbreviations

ΔASA, change in accessible surface area; CD, circular dichroism; CTD, C-terminal domain; FTIR, Fourier transform infrared spectroscopy; HEPES, 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid; MD, molecular dynamic; NQO1, NAD(P)H:quinone oxidoreductase 1; NTD, N-terminal domain; SAXS, small-angle X-ray scattering.

of P187S in the FAD-bound (holo) state only showed subtle local rearrangements [3], which cannot explain the long-range communication of its effect to the N-terminal domain (NTD, where the FAD-binding site is placed) or the C-terminal domain (CTD) [11]. Accordingly, molecular dynamic (MD) simulations and proteolysis experiments have indicated that P187S affects the FAD binding affinity through local dynamic changes at its binding site in the apo-state, particularly in the loop 57–66, while increased protein turnover is associated with a highly flexible CTD and only prevented upon binding of the inhibitor dicoumarol [11]. Indeed, Pro187 and FAD-binding sites and the CTD constitute an allosteric network that communicates dynamic information through long distances in the protein conformation ensemble [15]. Very recently, we have identified a set of consensus amino acids that have diverged along recent evolutionary history of NQO1, yielding the human protein more susceptible towards inactivation by P187S [17]. Experimental characterization of a suppressor mutation (causing a His80Arg exchange) has shown that it strongly protects the polymorphism against inactivation through a structural switch in the apo-state of P187S that stabilizes the loop 57–66 close to the FAD-binding site [17].

In this work, we have investigated the mechanism of inactivation of NQO1 by P187S, as well as the rescue mechanism due to H80R, from the perspective of their structural and energetic consequences on FAD binding. We have particularly focused on the capacity of the apo-state to communicate the effects of single amino acid changes due to its highly dynamic nature. To do so, we have investigated the conformation of NQO1 in different ligation states using spectroscopies (Fourier transform infrared [FTIR] and circular dichroism [CD]), small-angle X-ray scattering (SAXS), calorimetric analyses of FAD binding and structure-energetics calculations. Our results highlight the importance of characterizing the effect of single amino acid exchanges in cofactor-free apo-states, in particular for those cases in which high-resolution structures are not available and often very challenging to be determined due to their intrinsically high instability.

Materials and methods

Mutagenesis, expression and purification of NQO1 variants in *Escherichia coli*

Site-directed mutagenesis was performed using the QuickChange lightening kit (Agilent Technologies, Madrid, Spain) by standard protocols and confirmed by DNA sequencing. Expression in and purification from *E. coli* BL21(DE3) was performed as described [14].

Isothermal titration calorimetry

Titrations with FAD by isothermal titration calorimetry (ITC) were performed by direct (FAD in syringe, apo-NQO1 in the cell) or reverse (apo-NQO1 in the syringe, FAD in the cell) titrations as described earlier [14,18] using an ITC_{200} microcalorimeter. Both type of titrations provide very similar results (Fig. 1 and Table 1). A model considering one set of independent sites was used to obtain thermodynamic parameters, including heat capacity changes from the experimentally observed linear temperature dependence of binding enthalpies.

To calculate structure-based values for binding enthalpies and heat capacity changes, we used the following parametrizations [19,20]:

$$\Delta H_{\text{X-RAY}} = -7.35 \times \Delta ASA_{\text{apolar}} + 31.06 \times \Delta ASA_{\text{polar}}$$

$$\Delta C_{p,\text{X-RAY}} = 0.45 \times \Delta ASA_{\text{apolar}} - 0.26 \times \Delta ASA_{\text{polar}}$$

where $\Delta ASA_{\text{apolar}}$ and $\Delta ASA_{\text{polar}}$ are the changes in apolar and polar solvent-accessible surface area (ΔASA, in Å^2), and $\Delta H_{\text{X-RAY}}$ and $\Delta C_{p,\text{X-RAY}}$ are expressed in cal·mol^{-1} and $\text{cal·mol}^{-1}\text{·K}^{-1}$ respectively. $\Delta ASA_{\text{apolar}}$ and $\Delta ASA_{\text{polar}}$ were estimated from NQO1 crystal structures in the FAD-bound and -unbound states using POPS (http://mathbio.nimr.mrc.ac.uk/wiki/POPS).

Small-angle X-ray scattering

Small-angle X-ray scattering measurements of NQO1 proteins in solution were performed at Diamond Light Source beamline B21 (Didcot, UK), using a BioSAXS robot for sample loading. Solutions of proteins in different ligation states at different concentrations were prepared in 20 mM Na-HEPES, 200 mM NaCl pH 7.4 and 1 mM TCEP at 277 K. Samples were exposed for 300 in 10 s acquisition blocks using a sample to detector distance of 3.9 m and X-ray wavelength of 1 Å. The SCATTER software package (www.bioisis.net) was used to analyse data, including buffer subtraction, scaling, merging and checking possible radiation damage of the samples. Guinier plots for all samples were linear at the low q range, indicative of monodisperse solutions with no presence of protein aggregation during data acquisition. Due to the monodispersity of samples, these data were suitable for structural analysis. R_g and D_{max} values were calculated with PRIMUS and GNOM, and shape estimation was carried out with DAMMIF/DAMAVER/DAMMIN. Real-space scattering profiles of atomic models were calculated using CRYSOL and the final *ab initio* models were superimposed with the high-resolution structure using the program SUPCOMB. The program OLIGOMER was used to calculate the ratio of the
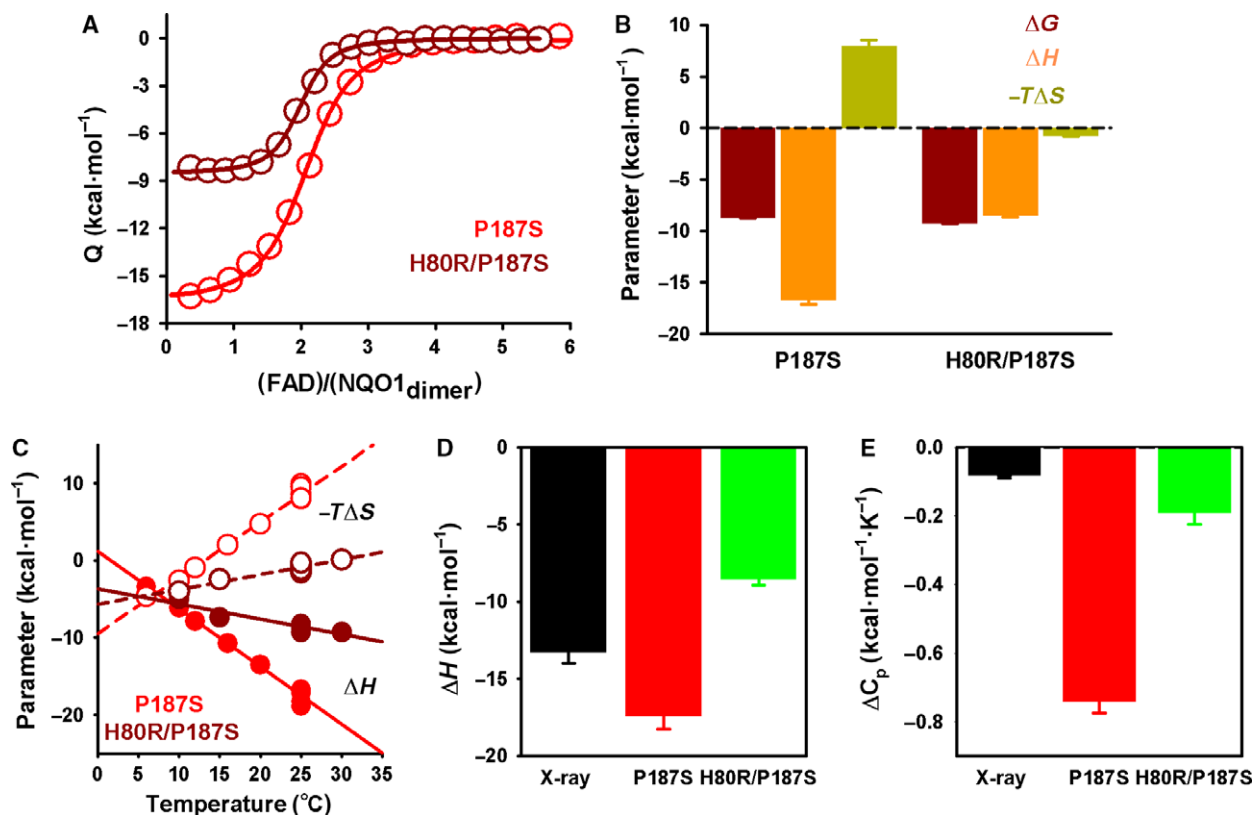
Fig. 1. Experimental and structure-based thermodynamic parameters for the binding of FAD to P187S and H80R/P187S. (A) Representative FAD titrations of apo-NQO1 proteins by ITC; (B) thermodynamic binding parameters for FAD at 25 °C ($n = 6$; mean ± SD); (C) temperature dependence of binding enthalpies and entropies; (D, E) comparison of experimental binding enthalpy and heat capacity (from linear fits shown in panel C) changes with those values derived from X-ray crystal structures (Fig. 2; mean ± SD from five different structures).

Table 1. FAD binding to P187S and H80R/P187S in K-HEPES and K-cacodylate (50 mM pH 7.4) buffers which have different ionization enthalpies (5.02 and −0.47 kcal·mol^{-1} respectively [27]).

NQO1	Buffer	K_d (nM)	ΔG (kcal·mol^{-1})	ΔH (kcal·mol^{-1})	$-T\Delta S$ (kcal·mol^{-1})
P187S	HEPES[a]	356 ± 63	−8.57 ± 0.30	−17.5 ± 0.8	+8.9 ± 0.8
	Cacodylate	420 ± 50	−8.69 ± 0.07	−17.3 ± 0.3	+8.6 ± 0.4
H80R/P187S	HEPES[a]	126 ± 50	−9.46 ± 0.26	−8.6 ± 0.4	−0.9 ± 0.4
	Cacodylate	198 ± 18	−9.14 ± 0.05	−8.1 ± 0.2	−1.0 ± 0.2

[a] Values are the average ± SD of six different titrations: three direct and three reverse titrations.

species present in the solutions. All the SAXS programs are included in the ATSAS programs package [21].

Fourier transform infrared spectroscopy

Samples of apo-NQO1 proteins were prepared at 100 μM (in protein subunit) in 50 mM K-HEPES pH 7.4, in the absence or presence of equimolar concentrations of FAD. FTIR spectra were recorded from 4000 to 900 cm^{-1} on a Bruker IFS 66 FTIR spectrophotometer (Bruker, Ettlingen, Germany) equipped with a BIO-ATR-II cell (Bruker) and thermostated at 25 °C by a water bath. Spectra were collected at 2 cm^{-1} resolution and averaged over 512 scans. Solvent

spectra were recorded under identical conditions and subtracted from the spectra of the proteins. Spectral contributions from residual water vapour were reduced using the atmospheric compensation filter built into the Bruker OPUS software (Bruker). The spectra were normalized by protein concentration.

Circular dichroism spectroscopy

Circular dichroism spectra were recorded in a Jasco J-710 (Jasco, Tokyo, Japan) at 25 °C in 20 mM K-phosphate pH 7.4 and using 1-mm path-length cuvettes. The 5 μM apo-NQO1 samples were prepared in the absence or presence of

25 μM dicoumarol and/or FAD. Five scans were recorded at 100 nm·min^{-1} and a bandwidth of 1 nm. In all cases, appropriate blanks in the absence of protein were recorded and subsequently subtracted.

Results and Discussion

Energetic dissection of the increase in FAD binding affinity of P187S due to the H80R mutation

Analysis of calorimetric titrations of P187S and H80R/P187S shows that the H80R mutation enhances by threefold the FAD binding affinity (Fig. 1A and Table 1). This affinity enhancement is associated with remarkable changes in the enthalpic and entropic contributions to binding (Fig. 1B). At 25 °C, FAD binding to P187S is much more favourable enthalpically [$\Delta\Delta H$ (H80R/P187S-P187S) = 8.9 ± 1.2 kcal·mol^{-1}] and strongly entropically penalized [$-T\Delta\Delta S$ (H80R/P187S-P187S) = −9.8 ± 1.2 kcal·mol^{-1}] (Table 1 and Fig. 1B). These results reveal a particular thermodynamic signature of H80R on FAD binding to P187S with structural implications: the local structural switch caused by H80R in the vicinity of the FAD-binding site may lead to a less favourable enthalpic contribution likely associated with local structural rearrangements, and entropically counterbalanced by a smaller loss of conformational entropy upon FAD binding due to the dynamic restriction of the FAD-binding site in the apo-state [17].

The temperature dependence of the binding enthalpy (ΔH) and entropy ($-T\Delta S$) are much larger for FAD binding to P187S (Fig. 1C), yielding an ~ 3.8-fold larger binding heat capacity (ΔC_p value; Fig. 1E). Since ionization effects can be ruled out (Table 1), binding ΔH and ΔC_p must arise from protein–ligand rigid body interactions that can be calculated using changes in accessible surface area (ΔASA) from crystal structures (Figs. 1D,E and 2) plus those arising from conformational changes upon FAD binding [19]. The calculated ΔASA$_{polar}$ and ΔASA$_{apolar}$ among different crystal structures of NQO1 are very similar, suggesting that the FAD binding mode is not largely affected by H80R and/or P187S in the holo-state (Fig. 2). Importantly, experimental ΔC_p values are larger than the values calculated from crystal structures of the FAD-bound, holo-state (Fig. 1E), confirming a departure from a lock-and-key binding mechanism and a remarkable conformational change upon FAD binding [11,15,18]. In this context, the different ΔH at 25 °C and ΔC_p for FAD binding between H80R/P187S and P187S (Fig. 1B,D,E) reflect the different magnitude of this conformational change at the structural level, and thus,

can be used to quantify the structural impact of H80R on the conformational ensemble of the apo-P187S [17].

As shown in Fig. 1E, the temperature-independent binding ΔC_p for FAD binding is much larger in P187S than the structure-based values (either for X-ray structures of WT, H80R or P187S), and it is substantially reduced by H80R. These results support two important conclusions [22,23]: (a) H80R causes a population shift in the conformational equilibrium of apo-P187S towards more compact (or structured) states; (b) FAD binding to either P187S or H80R/P187S complies with a mechanism in which FAD binding is strongly coupled with the conformational transition between binding competent and noncompetent states (i.e. an induced-fit mechanism), implying that FAD-binding competent states are not significantly populated in the conformational ensemble of the apo-state. Moreover, differences in binding ΔH at 25 °C and ΔC_p for FAD binding between H80R/P187S and P187S are reminiscent of the folding of a small protein domain: $\Delta\Delta H$ (at 25 °C) and $\Delta\Delta C_p$ are 8.9 ± 1.2 kcal·mol^{-1} and 0.55 ± 0.04 kcal·mol^{-1}, respectively, resembling the folding of an ~ 40-residue domain due to the presence of H80R (using well-known structure-energetic correlations [24]). These analyses strongly support that H80R affects locally the structure of apo-P187S, presumably in the vicinity of the FAD-binding site, as previously indicated by MD simulations and the X-ray crystal structure of holo-H80R [17].

In contrast, H80R facilitates FAD binding to P187S entropically (Fig. 1B and Table 1). These different entropic contributions can be rationalized from changes in: (a) translational and rotational entropy; (b) solvation entropy; and (c) conformational entropy [25]. The first term must not contribute significantly since FAD binding is a bimolecular interaction for both NQO1 variants. For the second term, we must consider ΔASA$_{polar}$ and ΔASA$_{apolar}$ associated with binding (i.e. the different structure of apo- and holo-states). Based on crystallographic analyses, both variants are expected to have fairly similar structures at the FAD-binding site as holo-proteins (Fig. 2), and thus, structural differences likely occur in the apo-state ensemble. In particular, the structural switch caused by H80R stabilizes the structure of the FAD-binding site of apo-P187S [17], thus contributing favourably to the binding entropy. Regarding the third term (conformational entropy), H80R restricts the dynamics of the FAD-binding site in the apo-state (particularly at the loop 57–66 [17]) thus reducing the entropy cost associated with FAD binding due to a favourable change in conformational entropy.

To further prove the structural and energetic consequences of the H80R mutation, we have characterized
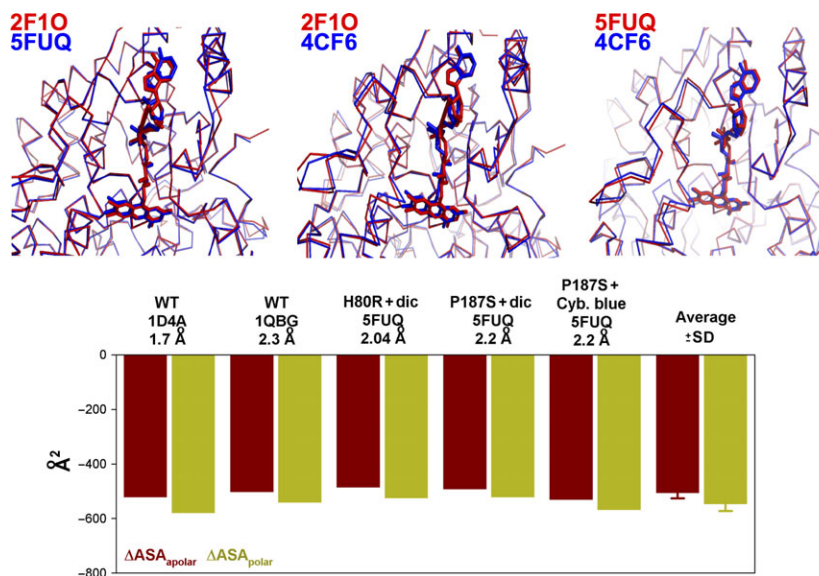
Fig. 2. Stuctural comparison of the FAD-binding site from different X-ray crystal structures of human NQO1. Upper panel: Pairwise structural overlay of FAD-binding sites from X-ray crystallography, using the structures corresponding to NQO1 enzymes with FAD and dicoumarol (WT, 2F1O; H80R, 5FUQ) or Cibacron blue (P187S; 4CF6). Lower panel: Changes in apolar (ΔASA_{apolar}) and polar (ΔASA_{polar}) accessible surface area due to FAD binding determined from the indicated X-ray crystal structures. Calculations were made using POPS (http://mathbio.nimr.mrc.ac.uk/wiki/POPS) and normalized per binding site.

the role of Glu71 and Glu78 by alanine scanning mutagenesis. These two glutamates constitute along with Arg80 a strong electrostatic cluster close to the FAD binding (the EER cluster [17]) that stabilizes the conformational switch triggered by H80R. As shown in Fig. 3A, Glu71 seems to be key to determining the affinity of P187S and H80R/P187S for FAD. A close inspection of binding ΔH and ΔC_p further supports our interpretation on the energetic and structural consequences of the H80R mutation (Fig. 3B,C). While E71A and E78A on P187S have little effect on binding heat capacity and enthalpies, these single mutants on H80R/P187S bring closer these variables to those of P187S, supporting that electrostatic interactions within the EER cluster are important to preserve the energetic (and likely, structural) differences in the FAD binding mechanism responsible for the increased FAD affinity of H80R/P187S (Fig. 3B,C). The important structural roles of Glu71 and Glu78 are further supported by the high instability displayed by double E71A/E78A mutants that precluded their large-scale purification and further experimental characterization.

FAD binding triggers global conformational changes while H80R locally affects the structure of the apo-state

So far, our results support that H80R corrects the FAD binding defect of P187S by causing a population shift in

the conformational ensemble of the apo-state towards more compact/structured states, and thus, energetically closer to those that are binding competent. Since FAD binding causes significant changes in the structure and dynamics of NQO1 (i.e. a conformational change [3,11,15]), this scenario is consistent with two limiting thermodynamic mechanisms: induced-fit (i.e. ligand binding is strongly coupled to the conformational change) or conformational selection (i.e. ligand binding is weakly coupled with the conformational change) [22].

To characterize changes in overall structure due to H80R for different ligation states, we have investigated the behaviour of NQO1 variants in solution using SAXS, CD and FTIR spectroscopies. The results of the SAXS experiments show that the pair-distribution functions $P(r)$ for NQO1 enzymes present single peaks, followed by a long extended tail at high r values (Fig. 4). The resulting *ab initio* three-dimensional shapes for different ligation states are shown in Fig. 5, and overlap well with models derived from crystallographic data (PDB codes: 1D4A, 2F1O, 4CET and 2FUQ). All apo-proteins showed a similarly elongated form (Fig. 5A,D,G). In agreement with a recent study using WT NQO1 [18], FAD binding causes a significant compaction of the solution ensemble, as shown by the decrease in R_g and D_{max} values in all three variants (Fig. 5B,E,H). Interestingly, binding of dicoumarol (a ligand that interacts with the CTD) leads to a slight contraction of holo-P187S and holo-H80R/
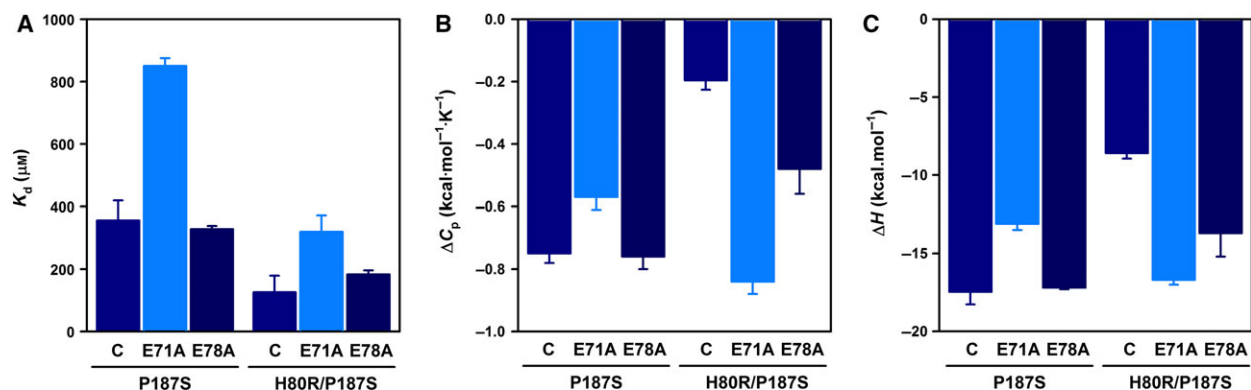
Fig. 3. Alanine scanning mutagenesis of Glu71 and Glu78 reveal their contributions to FAD-binding energetics and the rescue exerted by H80R. (A) Binding affinities at 25 °C; (B, C) binding heat capacity (B) and enthalpy (C; at 25 °C) changes. In panel (A) and (C), values are mean ± SD of $n = 6$ for controls (C) P187S and H80R/P187S, and $n = 3$ for alanine mutants. In panel (B), values are from the linear dependence of binding enthalpies on temperature.
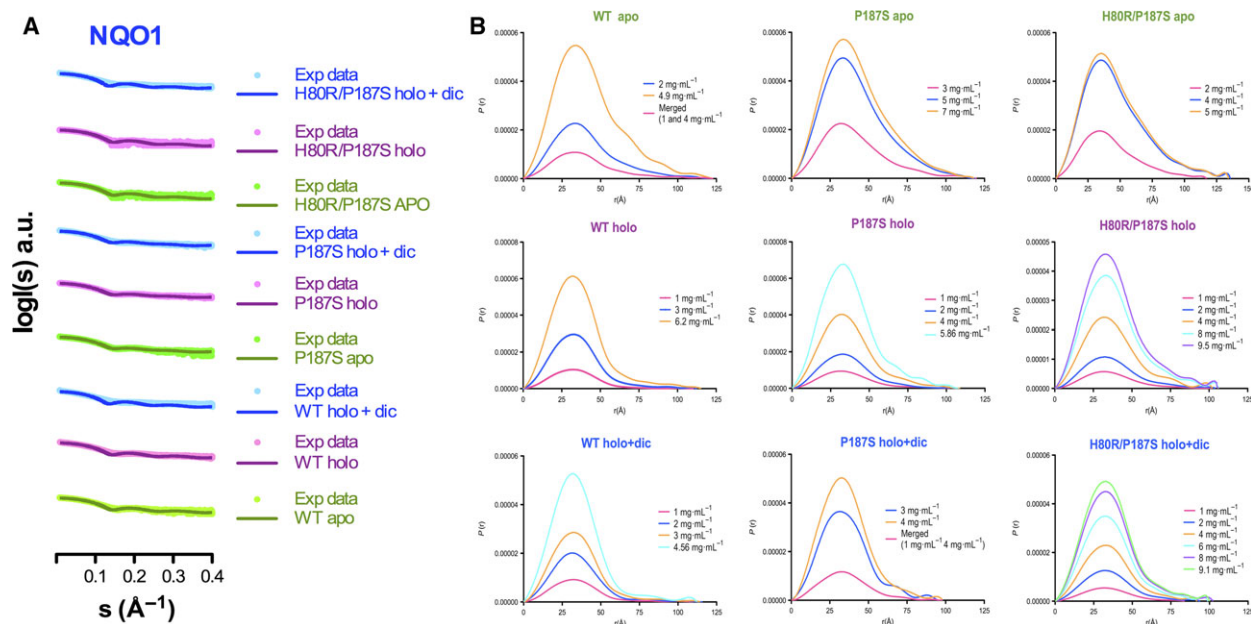

Fig. 4. SAXS analyses. (A) Scattering curves (dots) and theoretical scattering computed from the crystallographic models (smooth curves) in different ligation states. The data are vertically offset for clarity. a.u., arbitrary units. (B) $P(r)$ functions from SAXS data for NQO1 in different ligation states and at different protein concentrations.

P187S (Fig. 5C,F,I), which may reflect the compaction of the flexible CTD due to the P187S polymorphism [11,15].

We have also measured changes in secondary structure for different ligation states by FTIR and Far-UV CD (Fig. 6). The results suggest small differences in secondary structure between apo-WT and apo-P187S, particularly a decrease in the band intensities associated with α-helical content (i.e. bands centred at 208 and 222 nm, in CD, and at 1652 cm^{-1} in FTIR),

which seem to be corrected by the H80R mutation. Addition of FAD increases the content in ordered secondary structure, but to a lower extent in P187S and H80R/P187S, possibly reflecting a more disordered CTD due to P187S [3,11,15]. Consistently, dicoumarol binding further increases the ordered secondary structure content mostly in these two variants, supporting a larger conformational rearrangement in their CTD upon binding this inhibitor [11,15]. These results support our proposal on the local structural changes due
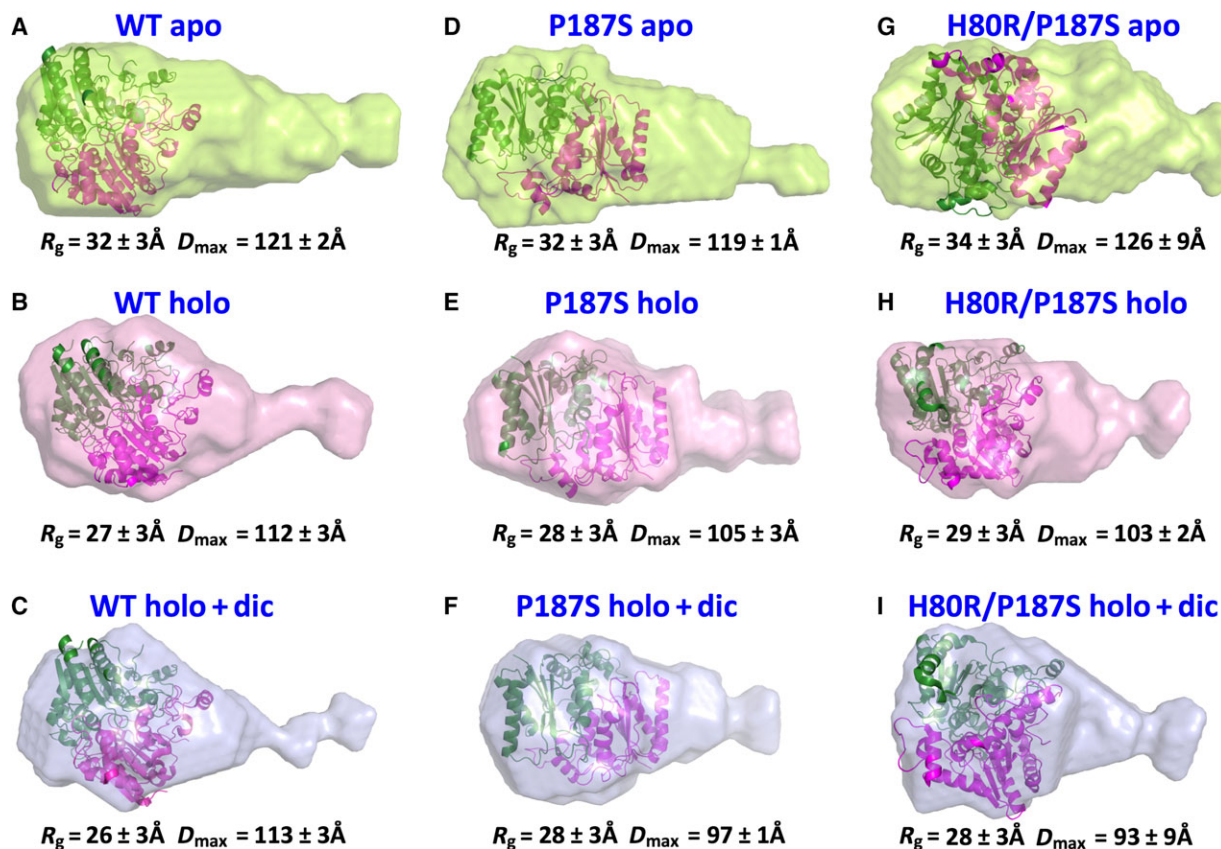
Fig. 5. Global conformational changes due to ligand binding and single amino acid replacements from SAXS. *Ab initio* shape reconstruction of NQO1 variants in different ligation states (apo, holo and holo + dic). The values of R_g and D_{max} are mean $\pm$ SD from measurements at different protein concentrations.

to H80R affecting locally the NTD, particularly in the apo-state.

Inactivation of NQO1 by P187S and reactivation by H80R can be unified by a simple thermodynamic and structural mechanism

Overall, we can integrate the structural, thermodynamic and dynamic information gathered to propose a mechanism of rescue of P187S by H80R in the context of an induced-fit mechanism (Fig. 7). In the absence of FAD, we assume the existence of a pre-existing equilibrium of conformations primarily populating noncompetent binding states, while the population of binding competent states is very low (i.e. K_{CONF}, defined as the ratio of concentrations of noncompetent and competent states, is very large; Fig. 7A [22]). Therefore, the apo-state is mostly populated by noncompetent binding states constituting a very flexible and dynamic ensemble (as seen by the large sensitivity of apo-proteins towards partial proteolysis, and also

supported by NMR, MD and SAXS analyses [3,11,17,18] and this work). In contrast, the binding competent states may structurally resemble the protein in the bound conformation (i.e. similar to that found in the X-ray crystal structures) and are marginally populated in the absence of FAD. In this scenario, binding of FAD provides the Gibbs free energy to shift the population from noncompetent to competent states, which is macroscopically observed as a conformational change [22]. Therefore, the observed affinity for FAD is related to the magnitude of the conformational transition (i.e. the value of K_{CONF}) and the intrinsic affinity for FAD (i.e. K_{FAD}) of the binding competent state (Fig. 7). Due to the minor structural effects of P187S and H80R on the holo-state, we can propose in the context of this mechanism, that the intrinsic affinity for FAD (K_{FAD}) is quite insensitive to these amino acid exchanges. Thus, the suppressor of H80R mainly operates on the pre-existing equilibrium in the absence of FAD by locally affecting the structure and dynamics of noncompetent binding states (i.e.
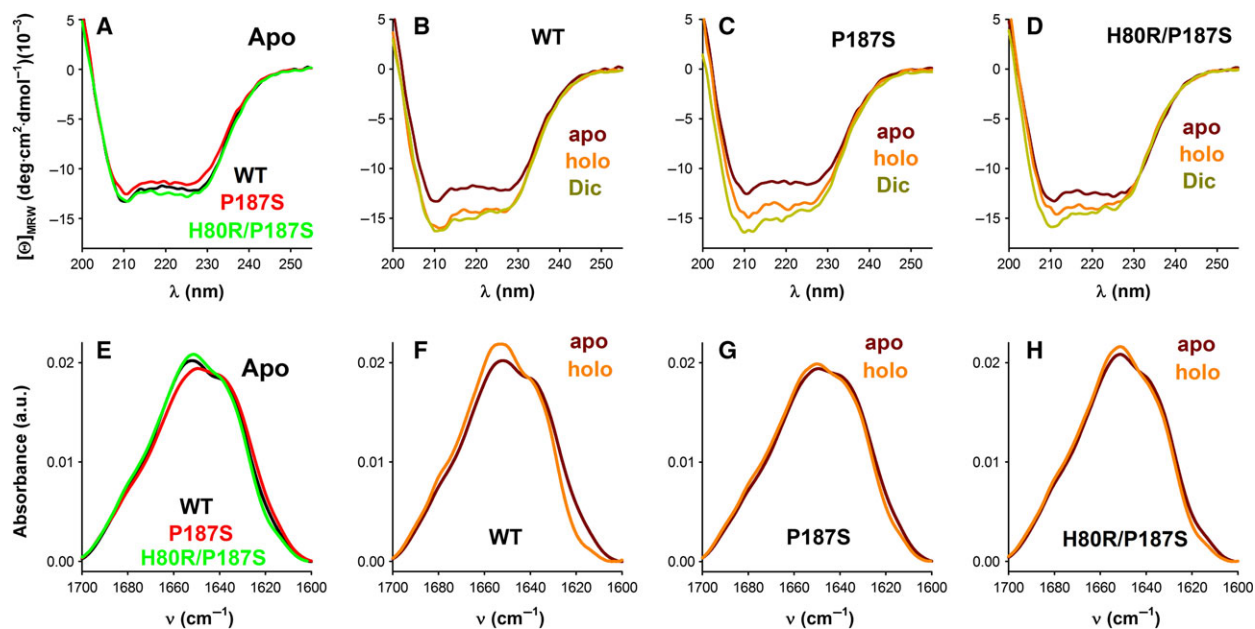
Fig. 6. ATR-FTIR and CD spectra reveal secondary structure changes in the conformational ensemble of apo- and holo-NQO1 by P187S and H80R. (A–D) Far-UV CD spectra of apo-proteins (A), and for WT (B), P187S (C) and H80R/P187S (D) in different ligation states (apo, holo and holo + dic). (E–H) FTIR spectra of apo-proteins (E), and for WT (F), P187S (G) and H80R/P187S (H) as apo- and holo-proteins. All the spectra are the average of two independent measurements.
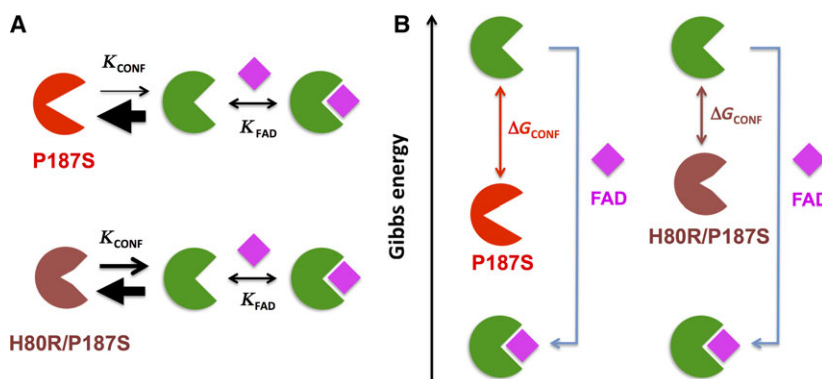

Fig. 7. A structural-thermodynamic mechanism to explain the effects of P187S on FAD binding and its rescue by H80R. (A) In the absence of FAD, the conformational ensemble of P187S is largely shifted towards binding noncompetent states (in red), while FAD-binding shift the population towards binding competent states (in green). The mutation H80R affects the conformational ensemble of the apo-state (in brown) shifting the equilibrium towards binding competent states in the absence of FAD (i.e. decreasing the value of K_{CONF}). (B) Gibbs free energy profile depicting the effect of H80R on the conformational ensemble of the apo-state, which decreases the energetic penalization due to the conformational change (ΔG_{CONF}) and concomitantly increases the overall FAD binding affinity. K_{CONF} stands for the equilibrium constant between noncompetent and competent states in the absence of ligand and K_{FAD} for the FAD (intrinsic) equilibrium association constant to competent states, while these two constants are related to the overall (apparent) affinity through the expression: $K_{FAD(app)} = K_{FAD}/1 + K_{CONF}$ [22].

K_{CONF}). It must be noted that in this scenario, the rescue by H80R can be explained by a modest threefold decrease in K_{CONF}, by bringing the noncompetent states closer to the competent states in terms of Gibbs free energy and thus decreasing the penalty to FAD binding due to the large conformational transition associated with binding (Fig. 7B).

Conclusions

Understanding the mechanisms by which disease-associated mutations affect protein stability and function, as well as the effects of second-site suppressor mutations may help to guide the development of novel therapeutic strategies aimed at treating inherited diseases

[17,26]. In this work, we provide a detailed mechanism that unifies the inactivation caused by P187S and its rescue by H80R in the context of a FAD binding mechanism in which ligand binding is strongly coupled to the conformational transition between binding competent and noncompetent states. In particular, P187S and H80R show opposite effects on the structure and dynamics of the FAD binding site in the apo-state, pointing towards a main role of fine-tuning structural features and conformational dynamics of the FAD-binding site to determine cofactor affinity. Therefore, a combination of structural and dynamic analysis using a wide variety of techniques (titration calorimetry, proteolysis, MD simulations, SAXS, FTIR and CD [17] and this work) support the existence of significant conformational plasticity in the apo-state of NQO1 in response to single amino acid exchanges. This integrated analysis completes and refines previous information obtained from high-resolution structural techniques such as X-ray crystallography or NMR spectroscopy [3,17,18]. We also provide here a detailed characterization of the suppressor effect of the consensus amino acid Arg80, which has recently diverged to His80 along primate evolution (about 30 million years ago), potentially predisposing human NQO1 towards cancer-associated inactivation [17]. Our work also supports that a detailed understanding of mutational effects on the structure and dynamics of protein conformational ensembles may help to provide a deeper knowledge on the mechanisms used by evolution to shape protein stability and function.

Acknowledgements

Author contributions

ALP designed the research study and wrote the initial draft. IGM, BM, EMC and ALP performed the research, revised the manuscript and approved the final version. IGM, BM and ALP contributed essential reagents or tools and analysed the data.

References

1 Pey AL, Megarity CF, Medina-Carmona E and Timson DJ (2016) Natural small molecules as stabilizers and activators of cancer-associated NQO1 polymorphisms. *Curr Drug Targets* **17**, 1506–1514.

2 Colucci MA, Moody CJ and Couch GD (2008) Natural and synthetic quinones and their reduction by the quinone reductase enzyme NQO1: from synthetic organic chemistry to compounds with anticancer potential. *Org Biomol Chem* **6**, 637–656.

3 Lienhart WD, Gudipati V, Uhl MK, Binter A, Pulido SA, Saf R, Zangger K, Gruber K and Macheroux P (2014) Collapse of the native structure caused by a single amino acid exchange in human NAD(P)H: quinone oxidoreductase(1.). *FEBS J* **281**, 4691–4704.

4 Asher G, Tsvetkov P, Kahana C and Shaul Y (2005) A mechanism of ubiquitin-independent proteasomal degradation of the tumor suppressors p53 and p73. *Genes Dev* **19**, 316–321.

5 Oh ET, Kim JW, Kim JM, Kim SJ, Lee JS, Hong SS, Goodwin J, Ruthenborg RJ, Jung MG, Lee HJ et al. (2016) NQO1 inhibits proteasome-mediated degradation of HIF-1alpha. *Nat Commun* **7**, 13593.

6 Moscovitz O, Tsvetkov P, Hazan N, Michaelevski I, Keisar H, Ben-Nissan G, Shaul Y and Sharon M (2012) A mutually inhibitory feedback loop between the 20S proteasome and its regulator, NQO1. *Mol Cell* **47**, 76–86.

7 Belinsky M and Jaiswal AK (1993) NAD(P)H:quinone oxidoreductase1 (DT-diaphorase) expression in normal and tumor tissues. *Cancer Metastasis Rev* **12**, 103–117.

8 Buranrat B, Chau-in S, Prawan A, Puapairoj A, Zeekpudsa P and Kukongviriyapan V (2012) NQO1 expression correlates with cholangiocarcinoma prognosis. *Asian Pac J Cancer Prev* **13** (Suppl), 131–136.

9 Garate M, Wani AA and Li G (2010) The NAD(P)H: quinone oxidoreductase 1 induces cell cycle progression and proliferation of melanoma cells. *Free Radic Biol Med* **48**, 1601–1609.

10 Martinez-Limon A, Alriquet M, Lang WH, Calloni G, Wittig I and Vabulas RM (2016) Recognition of enzymes lacking bound cofactor by protein quality control. *Proc Natl Acad Sci USA* **113**, 12156–12161.

11 Medina-Carmona E, Palomino-Morales RJ, Fuchs JE, Padín-Gonzalez E, Mesa-Torres N, Salido E, Timson DJ and Pey AL (2016) Conformational dynamics is key to understanding loss-of-function of NQO1 cancer-associated polymorphisms and its correction by pharmacological ligands. *Sci Rep* **6**, 20331.

12 Traver RD, Horikoshi T, Danenberg KD, Stadlbauer TH, Danenberg PV, Ross D and Gibson NW (1992) NAD(P)H:quinone oxidoreductase gene expression in

human colon carcinoma cells: characterization of a mutation which modulates DT-diaphorase activity and mitomycin sensitivity. *Cancer Res* **52**, 797–802.

13 Lajin B and Alachkar A (2013) The NQO1 polymorphism C609T (Pro187Ser) and cancer susceptibility: a comprehensive meta-analysis. *Br J Cancer* **109**, 1325–1337.

14 Pey AL, Megarity CF and Timson DJ (2014) FAD binding overcomes defects in activity and stability displayed by cancer-associated variants of human NQO1. *Biochim Biophys Acta* **1842**, 2163–2173.

15 Medina-Carmona E, Neira JL, Salido E, Fuchs JE, Palomino-Morales R, Timson DJ and Pey AL (2017) Site-to-site interdomain communication may mediate different loss-of-function mechanisms in a cancer-associated NQO1 polymorphism. *Sci Rep* **7**, 44352.

16 Siegel D, Anwar A, Winski SL, Kepa JK, Zolman KL and Ross D (2001) Rapid polyubiquitination and proteasomal degradation of a mutant form of NAD(P) H:quinone oxidoreductase 1. *Mol Pharmacol* **59**, 263–268.

17 Medina-Carmona E, Fuchs JE, Gavira JA, Mesa-Torres N, Neira JL, Salido E, Palomino-Morales R, Burgos M, Timson DJ and Pey AL (2017) Enhanced vulnerability of human proteins towards disease-associated inactivation through divergent evolution. *Hum Mol Genet*, https://doi.org/10.1093/hmg/ddx238.

18 Lienhart WD, Strandback E, Gudipati V, Koch K, Binter A, Uhl MK, Rantasa DM, Bourgeois B, Madl T, Zangger K *et al.* (2017) Catalytic competence, structure and stability of the cancer-associated R139W variant of the human NAD(P)H:quinone oxidoreductase 1 (NQO1). *FEBS J* **284**, 1233–1245.

19 Luque I and Freire E (2002) Structural parameterization of the binding enthalpy of small ligands. *Proteins* **49**, 181–190.

20 Gomez J, Hilser VJ, Xie D and Freire E (1995) The heat capacity of proteins. *Proteins* **22**, 404–412.

21 Petoukhov MV, Franke D, Shkumatov AV, Tria G, Kikhney AG, Gajda M, Gorba C, Mertens HD, Konarev PV and Svergun DI (2012) New developments in the program package for small-angle scattering data analysis. *J Appl Crystallogr* **45**, 342–350.

22 Vega S, Abian O and Velazquez-Campoy A (2016) On the link between conformational changes, ligand binding and heat capacity. *Biochim Biophys Acta* **1860**, 868–878.

23 Schrank TP, Bolen DW and Hilser VJ (2009) Rational modulation of conformational fluctuations in adenylate kinase reveals a local unfolding mechanism for allostery and functional adaptation in proteins. *Proc Natl Acad Sci USA* **106**, 16984–16989.

24 Robertson AD and Murphy KP (1997) Protein structure and the energetics of protein stability. *Chem Rev* **97**, 1251–1268.

25 D'Aquino JA, Gomez J, Hilser VJ, Lee KH, Amzel LM and Freire E (1996) The magnitude of the backbone conformational entropy change in protein folding. *Proteins* **25**, 143–156.

26 ben Bdira F, Gonzalez E, Pluta P, Lain A, Sanz-Parra A, Falcon-Perez JM and Millet O (2014) Tuning intracellular homeostasis of human uroporphyrinogen III synthase by enzyme engineering at a single hotspot of congenital erythropoietic porphyria. *Hum Mol Genet* **23**, 5805–5813.

27 Fukada H and Takahashi K (1998) Enthalpy and heat capacity changes for the proton dissociation of various buffer components in 0.1 M potassium chloride. *Proteins* **33**, 159–166.

4. METODOLOGY

The following table summarizes the methodology performed in this doctoral thesis indicating the publications (*see section 3*) where the assays are explained in detail.

METODOLOGY	REFERENCE IN SECTION 3 (Publication Number)
In silico Experiments	
Electrostatic Calculations	[1] [3]
Molecular Dinamic Simulation	[1] [2] [3]
Consensus and Phylogenetic Analyses	[3]
In vitro Experiments	
Site-Direct Mutagenesis	[2] [3] [4]
Protein Expression and Purification	[1] [2] [3] [4]
Activity Measurement	
• Specific Activity	[1] [2] [3]
• Inactivation Kinetics	[3]
• Kinetic Parameters Determination	[2]
Proteolysis Experiments	
• Proteolysis Kinetics	[1] [2] [3]
• Analysis of Proteolysis Products by Mass Spectrometry	[1] [2]
Differential Scanning Calorimetry (DSC)	[1]
Isothermal Titration Calorimetry (ITC)	[1] [2] [4]
Circular Dichroism Spectroscopy (CD)	[1] [2] [4]
Dynamic Light Scattering (DLS)	[2]

Fluorescence Experiments

- Thermal Denaturation Studies   [3]

- Thermal Stability Studies   [2] [3]

- Fluorescence Titrations   [2] [3]

Nuclear magnetic resonance spectroscopy (NMR)	[2] [3]
X-ray Crystallography	[3]
Small-Angle X-ray Scattering (SAXS)	[4]
Fourier Transform Infrared Spectroscopy (FTIR)	[4]
Expression Studies in Cell-Free System	[2] [3]
Cellular Experiments	
Mammalian Cells Cultures	[1] [3]
Stable Transfection of Cells	[3]
Immunoblotting	[1] [3]
Activity Measurement	[3]
Cell Viability Assays	[3]

5. DISCUSSION

Results of this thesis reveal that studying the alterations in protein dynamics and its relation with stability and functionality, *in vitro* and within cells, is crucial to understand in-depth the pathogenic mechanisms associated with loss-of-function genetic diseases. Frequently, the effects of pathogenic genetic mutations on protein stability and function have been interpreted using "static" high-resolution crystal structures. Nevertheless, they do not allow determining the global and local protein motions which are essential to protein function, as well as protein regulation, degradation or evolution[73,76,83,136–140]. In particular, our research focuses on the study of two NQO1 polymorphisms associated with enzyme loss-of-function, R139W and P187S. They are especially interesting due to they are overexpressed in certain types of tumours and have been related to a decreased efficiency in the bioactivation of some cancer prodrugs[3,124–128,130,141]. Our results have been divided into several sections with the aim of responding to each one of the objectives exposed and they are discussed from this point forward.

5.1 ALTERATIONS CAUSED BY P187S IN PROTEIN DYNAMICS MAY EXPLAIN THE PATHOGENIC EFFECTS ASSOCIATED WITH THIS POLYMORPHISM
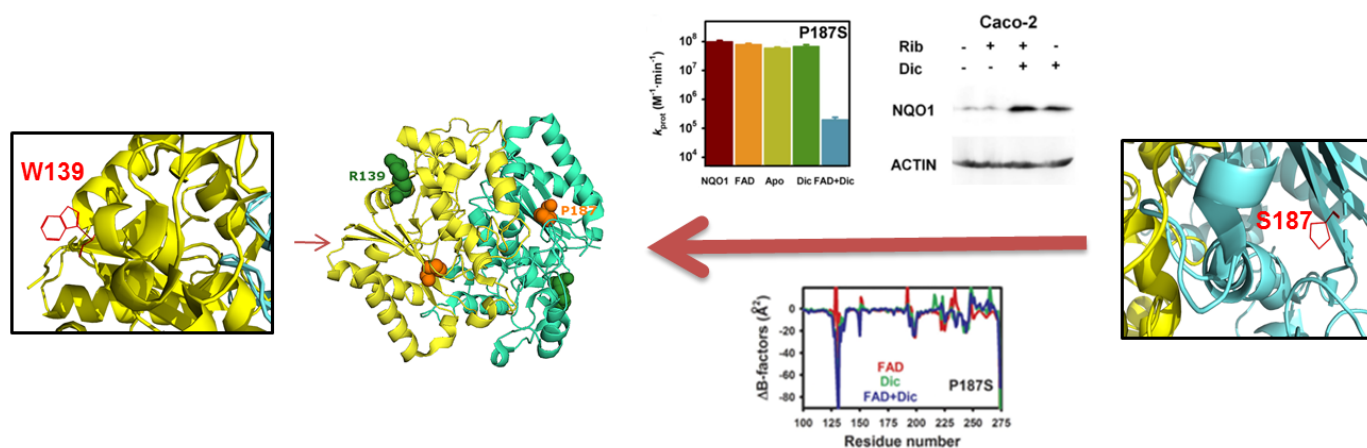
EFFECTS ON NQO1 DYNAMICS

To investigate the effects of NQO1 polymorphisms on protein dynamics we have used a combination of experimental and computational techniques such as, stability and binding studies, analyses of the local flexibility through proteolysis studies with thermolysin and mass spectrometry, molecular dynamic simulations and expression studies in eukaryotic cells.

On the one hand, although R139W has been linked to increased risk for cancer, its role in the disease development remained obscure and trying to decipher the molecular basis of its disease-associated mechanism meant a challenge for us. Our results about R139W showed that this polymorphism causes modest changes in protein dynamic which lead to only slight alterations in activity and stability in comparison to wild-type

NQO1. Specifically, although the NQO1 functionality in presence of R139W is virtually identical, a minor decrease in FAD binding affinity and kinetic stability have been found[4]. The functional changes seem to be caused by a slightly more dynamic active site, as our partial proteolysis experiments shown, while the perturbations in conformational stability are likely originated from dynamic alterations in the monomer- monomer interface suggested by our MD simulations (**Publication 1**). In light of these results, it seems that the main molecular mechanism associated with the pathogenicity of this polymorphism is probably the erroneous splicing of the premature mRNA that leads to the loss of exon 4, residues 102–139 of the full-length protein. It destroys the FAD and substrate binding site producing an inactive and unstable protein[134]. In accordance with our results, a subsequent study carried out a R139W characterization where X-ray crystallography and nuclear magnetic resonance (RMN) spectroscopy experiments were performed. Their results showed that this polymorphism is virtually identical to the wild-type protein, "structurally" speaking[142].

Nevertheless, although crystallographic analyses about P187S also revealed no significant structural changes, the alterations caused by P187S in NQO1 dynamics are much more striking and they can be directly linked to the pathogenic effects of P187S, such as inactivation of the enzyme and enhanced protein degradation by the proteosome[4–7]. Our results showed that even though Pro187 is not directly involved in the binding sites of FAD and NAD(P)H/dicoumarol, its replacement by a serine reduces both its binding affinity for FAD (related to a severe reduction of NQO1 activity) and dicoumarol compared to wild-type. As consequence of the difference in FAD binding

affinity, the NQO1 proteins show diverse amounts of FAD bound as purified, being P187S essentially an apo-protein.

Numerous genetic diseases have been associated with mutations which produce enzymes with reduced binding affinity for their ligands. Specifically, flavin-dependent human proteins appear as an exceptionally sensitive group and the supplementation with their ligands may increase the fraction of active holo-enzyme[143]. For this reason, different experiments in presence and absence of FAD and dicoumarol were performed by our group. The results showed that both the addition of FAD and dicoumarol cause a global stabilization *in vitro* of all NQO1 enzymes, and besides the dicoumarol stabilizing effects are additive to those of FAD **(Publication 1)**.

Considering that a considerable number of research have reflected that the intracellular levels of most flavoproteins are strongly dependent on flavin availability, a connection between flavin-dependent proteins stability and proteasomal degradation rates has been proposed[1,6,7,143,144]. This link may be explained through perturbations in protein dynamics, since minor structural deficits may be recognized by the proteasome for their degradations through dynamics[145]. With the goal of determining the potential effect of this polymorphism on protein dynamics, we have carried out proteolysis experiments with thermolysin. We observed that P187S, as purified, is much more sensitive to proteolysis than the wild-type protein and its high susceptibility is only protected by simultaneous addition of FAD and dicoumarol. To fully understand these results, we have characterized the proteolysis patterns of NQO1 proteins in the absence and the presence of these ligands. The apo-states of all NQO1 proteins are

highly dynamic, though only the holo-state of P187S presents a flexible C-terminal domain (CTD) which requires the dicoumarol binding for its correction.

Moreover, we have used molecular dynamics (MD) simulations to test our results *in vitro*. In particular, MD simulations did not show relevant effects on the overall secondary structure contents associated with P187S and/or the ligands binding and these results are also in accordance with spectroscopic analyses carried out previously[4]. However, our MD simulations showed large changes in the flexibility of monomer-monomer interface which may be associated both with the destabilizing effects of this polymorphism and with the stabilizing effects of FAD binding. Due to the NQO1 dimer dissociates prior to the rate-limiting step of thermal denaturation, these changes in flexibility of the dimer interface must strongly affect the kinetic stability of NQO1 enzymes[4]. In addition, our MD simulations showed that the CTD of NQO1 is more flexible that the N-terminal domain (NTD) suggesting that the NQO1 degradation proceeds preferentially by its CTD. It allows us to map changes at the residue level in protein dynamics caused by P187S or ligands binding **(Publication 1)**.

As previously mentioned, the dynamic changes due to P187S might be key to understand the low activity and stability *in vivo* of this polymorphism which seems to be susceptible for degradation by 20S proteasome[7]. For this reason, we have evaluated the effect of treatment with riboflavin (a precursor of FAD) and/or dicoumarol in NQO1 protein levels of different cells lines that express endogenously different NQO1 variants. Our results supported this hypothesis and showed increased P187S protein levels in presence of dicoumarol (while the treatment with riboflavin alone has no effect). It suggests that ligands that bind to the CTD (such as dicoumarol) would

protect P187S towards proteasomal degradation which preferentially occurs through of this end **(Publication 1)**. Moreover, our results were supported, either *in vitro* or *in vivo*, by an parallel and independent report[146]. They show the CTD of apo-NQO1 enzymes is targeted (at least *in vitro*) for degradation by the ubiquitin ligase C-terminal Hsp70-interacting protein (CHIP), and dicoumarol binding seems to produce the inhibition of ubiquitination and thus, increase the protein levels of P187S (or apo-enzymes)[146].

In this way, the results of this first section **(Publication 1)** show the two cancer-associated NQO1 polymorphisms have very different effects on protein dynamics since R139W seems to affect the NQO1 dynamic weakly, while P187S has important effects on protein dynamics which can be linked to deleterious effects of this polymorphism. **In particular, P187S appears to affect the local dynamics of two functionally and structurally distant sites: the NTD associated with enzyme inactivation, and the CTD associated with enhanced proteasomal degradation of P187S**[4,146]. Moreover, NQO1 symbolises an interesting example of a natively folded protein in which, its conformation, function, and stability are strongly modulated by binding to different natural ligands. Concretely, our results may have important implications to discover new pharmacological ligands to the P187S rescue *in vivo* which might allow the treatment of patients with this polymorphism.

5.2 P187S PERTURBS A MULTIFUNCTIONAL ALLOSTERIC NETWORK.
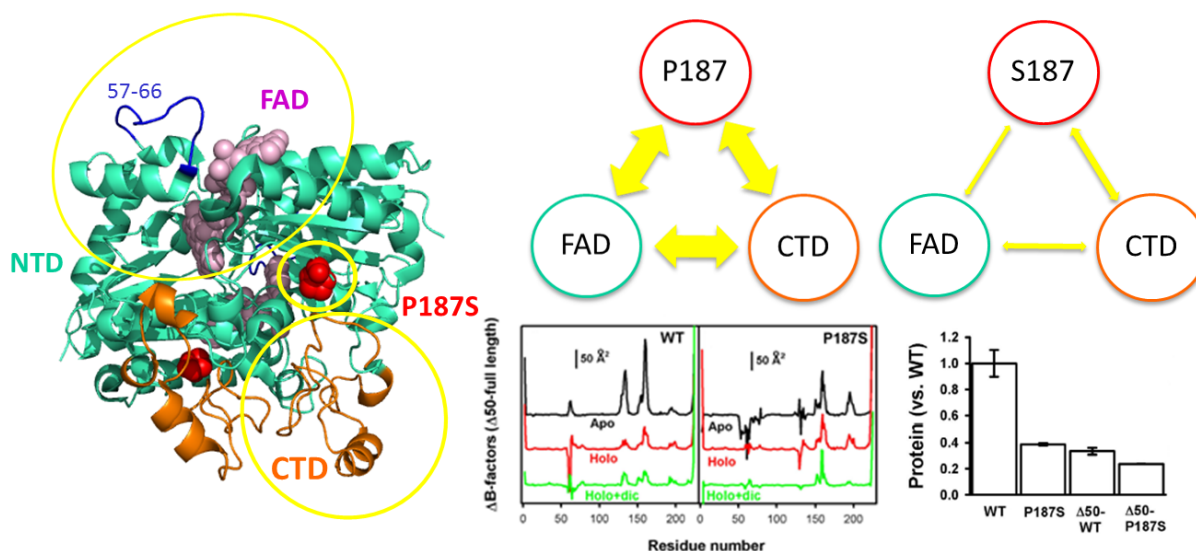
As we previously discussed, our results about P187S support the hypothesis that this polymorphism inactivates and destabilizes NQO1 through dynamic alterations which reach two functional and structurally distant sites of the NQO1, the FAD binding site and the CTD **(Publication 1)**. Crystallographic analyses have shown that the CTD is structurally involved in binding of NADH/substrates/competitive inhibitors (i.e. dicoumarol) and the dimer interface[106]. In addition, the CTD seems to play a key role in the low intracellular stability of the apo-NQO1 and to characterize the effects of P187S about the CTD may be crucial to understand in-depth the pathogenic mechanism associated with this polymorphism *in vivo*[146].

With the goal of studying the structural and functional role of CTD in NQO1 enzymes, we have analysed exhaustively wild-type and P187S variants, both in their full-length

form and removing the last 50 residues of the CTD (Δ50-WT and Δ50-P187S respectively).

In the first place, we have characterized the functional role of CTD observing that the lack of this domain strongly decreases the specific activity compared to wild-type NQO1. Given that the low intracellular specific activity of P187S is largely associated with defective FAD binding, and besides it is rescued by FAD addition, a similar behaviour would be foreseen in Δ50-variants. However, the saturation with FAD of these variants only causes modest increases in their specific activities suggesting that the inactivation is provoked by a different mechanism from that observed in P187S. Specifically, the low specific activity of Δ50-variants seems to lie in a strong reduction in catalytic performance which is caused by a decrease of the apparent affinity for NADH. Moreover, the removal of the CTD abolishes the detrimental effect of P187S on FAD (and dicoumarol) binding since the Δ50-variants show high binding affinities, similar in both variants. These results support our previous find which suggests that the loss-of-function associated with this polymorphism is explained by the existence of a long-range communication between the CTD, the FAD binding site and the P187S site **(Publication 1).** In particular, we provide evidence of the existence of this network which we name "allosteric interaction network". For wild-type, this allosteric network helps to its very high affinity for FAD, while P187S may disturb this network promoting dynamic and structural alterations in the FAD binding sites and the CTD, reducing its binding affinity for FAD (and dicoumarol) **(Publication 2).**

On the other hand, considering that P187S strongly reduces the *in vitro* kinetic stability of NQO1 affecting the dimer interface, we have evaluated the role of CTD in stability.

Our results showed that all NQO1 variants behave like dimers in solution and the CTD is not the cause of the kinetic destabilization connected with P187S[4]. The overall conformation of all NQO1 variants in diverse ligation states (apo-state, holo-state and holo+dicoumarol state) were analysed by spectroscopic techniques. An important role of CTD in the conformational equilibrium of apo-NQO1 is showed, suggesting a shift of this equilibrium towards competent conformations for ligand binding (more compact) in presence of FAD and/or dicoumarol. Interestingly, the binding to diverse ligands produces different conformational changes in wild-type and P187S. The FAD binding causes larger conformational changes in wild-type than in P187S now that, the CTD of holo-P187S remains as flexible and partially unfolded as in the apo-state. Nevertheless, the dicoumarol binding brings about the greatest effects on the conformation of holo-P187S. Dicoumarol binding to holo-P187S causes a noteworthy thermodynamic signature which energetically resembles the folding of a small protein. This latter may be associated with the large changes that the flexible and partially unstructured CTD of holo-P187S must undergo to reach binding competent states, more "rigid" **(Publication 1** and **Publication 2)**. This noticeable dynamic stabilization of the CTD in holo-P187S through dicoumarol binding may prevent its proteasomal degradation entailing a strong intracellular stabilization[146]. Interestingly, these results suggest that dicoumarol may operate as a pharmacological chaperone of P187S correcting the defects caused by this polymorphism and serving as basis for the design of new molecules which rescue this polymorphism *in vivo*[1,40,148]. It may seem counterintuitive to use an inhibitor to recover the stability and functionality in a loss-of-function disease. However, this strategy has been successfully used previously since the

inhibitory effect can be corrected *in vivo* by the natural substrate/cofactor binding at much higher concentrations intracellularly[1].

To investigate the contribution of CTD to the proteasomal degradation of NQO1 enzymes, experiments with Δ50-variants in eukaryotic cell free system have been performed analysing proteins levels and degradation rates. The results support a stepwise directional degradation of NQO1 through its C-terminal as previously we had already suggested, since when the CTD is removed, the remaining protein is much more flexible and efficiently degraded **(Publication 2)**. Our interpretation concurred with a other report where they suggest that the highly dynamic CTD of P187S accelerates NQO1 proteasomal degradation[146]. In addition, in this same report they show that the ubiquitination by CHIP in NQO1 is strongly dependent of CTD state and when this is removed, both wild-type and P187S, are degraded at comparable degradation rates.

Furthermore, considering that NQO1 is known to interact with and stabilize cancer-associated transcription factors such as p73α, we have investigated the role of CTD in the interaction with this cell cycle regulator. In particular, we have studied by NMR spectroscopy the interaction of NQO1 with the sterile alpha motif of p73α (SAMp73) which is thought to be involved in protein-protein interactions[149,150]. Interestingly, our results provide the first structural insights into the interaction between p73α and NQO1 and show that neither the CTD, nor the presence of P187S, nor the absence of bound FAD (active protein) is required in the interaction with p73. Still, different ligands bound to NQO1 may modulate the intrinsically weak interaction **(Publication 2)**. Our results are consistent with previous biochemical studies where a physical

interaction between NQO1 and p73, in an NADH-dependent manner, was suggested[151]. It is bear mentioning that the available evidences that supported intracellular destabilization of oncosupressors by P187S seem to derive from its own intracellular instability instead of its inability to interact with these proteins[6,7,129].

In conclusion, the results of this second section show the importance of the CTD in the multiple detrimental effects associated with P187S and lend evidences of a complex allosteric communication between different functional sites of NQO1 which is perturbed for this polymorphism. Since numerous genetic diseases are caused by missense mutations which alter different protein features (such as, protein stability, functionality and regulation), our multi-disciplinary strategy would help to decipher complex mutational effects associated with many other genetic diseases and identify structural hot spot as targets for therapeutic correction. In particular, in the context of the flavo-proteome our strategy can provide clues to decipher the dynamic flavin-dependence on flavoprotein levels *in vivo*[143,144].

5.3. SUPPRESSOR MUTATIONS MAY PROTECT NOQ1 AGAINST LOSS-OF-FUNCTION CAUSED BY P187S
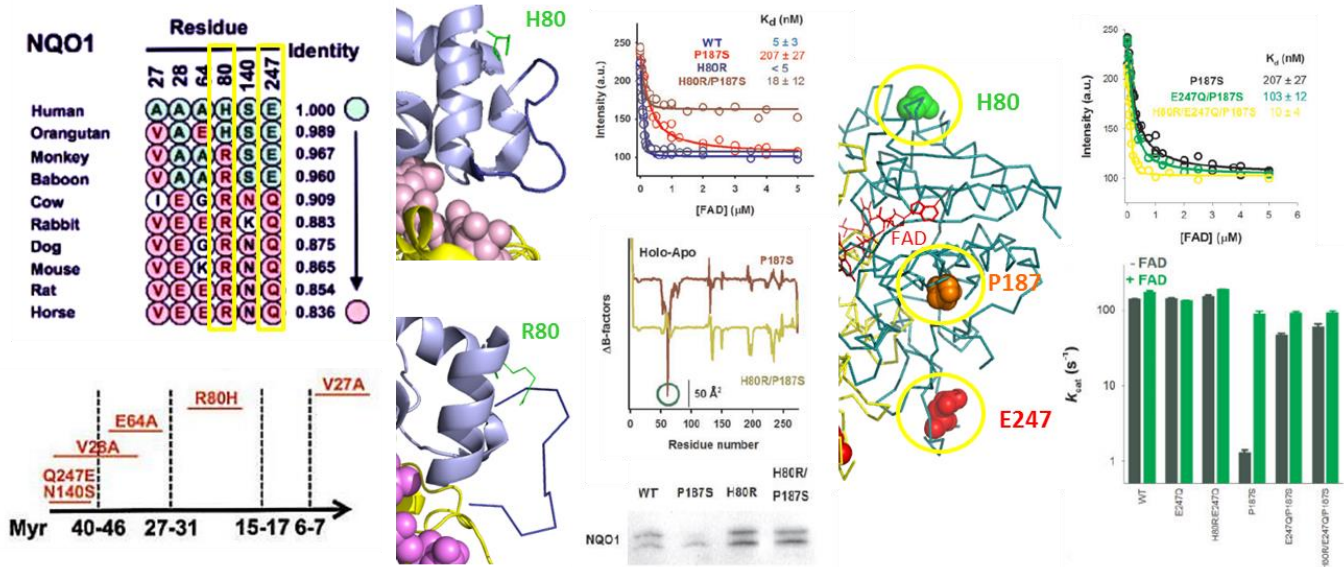
As a separate strategy to decipher and correct the pathologic defects caused by P187S, we used the consensus strategy. It is a semi-rational approach that uses statistical information contained in sequences alignments of proteins with high homology and it has been successfully applied to improve the thermodynamic and/or kinetic stability of several proteins and even sometimes, simultaneous protein activity increases have also been associated with consensus mutations[8,9,152–157]. On account of the sequence conservation, it might be expected that a considerable number of consensus mutations run into with identical amino acids, at the same position, in the ancestral proteins sequences[158]. Since a thermophilic character is traditionally associated with the ancestral life, the stabilizing effects of consensus mutations seem to reflect the ancestral protein hyperstability[158–160].

According to the Darwinian theory of evolution by natural selection, the mutations with beneficial effects on protein fitness are often fixated, while neutral or detrimental mutations are usually avoided by evolution[17,140]. From this point of view, it is difficult to comprehend how destabilizing mutations that differ from consensus amino acids can be fixed during evolution. However, Kimura's neutral theory suggests that the most of evolutionary changes at molecular level are caused by random fixation of selectively neutral or very nearly neutral mutations and thus, with no apparent effect on protein fitness[161]. Destabilizing mutations in themselves (i.e. some mutation which divergent from consensus) may be biologically compensated and buffered by different "trade-off" mechanisms, such as modulation of the protein misfolding by molecular chaperones or the existence of other stabilizing (consensus) mutations, causing a neutral effect in fitness[17,140,162–164]. In this way, a variation in genotype may have no apparent effect on phenotype at present, but can change the potential for future adaptations since, an adaptive event, as the acquisition of a new function, requires often of the simultaneous presence of two or more mutations[17,165]. On average, mutations that confer new functions are strongly destabilizing and their accumulation may give rise to proteins more prone to loss-of-function or other negative effects when the trade-off mechanisms do not exert enough compensatory effects. In these cases, the consensus mutations would act as global intragenic suppressors, rescuing detrimental mutations at several sites in a protein and avoiding their negative effects[65,163,166]. To conclude, the consensus strategy seems to be an excellent and simple choice to achieve the P187S rescue.

We have identified a set of six stabilizing mutations for NQO1 after aligning mammalian sequences with high identity. The H80R mutation caused the most remarkable stabilizing effect, both in wild-type and P187S. R80 has been found in most of the closely related mammalian, proposing that its stabilizing effect has disappeared recently (along primate evolution) in an evolutionary time scale **(Publication 3).**

The H80R causes an increase in the kinetic stability of the proteins, slowing down the irreversible inactivation of both wild-type and P187S, and proteolysis experiments by thermolysin showed that H80R causes a local dynamic stabilization of the NTD, but it has no effect in the CTD of the protein. With the aim of characterizing the H80R stabilizing effect, we have solved its structure in presence of FAD and dicoumarol (PDB accession code: 5FUQ). H80R mutation did not affect the overall fold of NQO1 although it causes a local structural switch (R80 changes a less solvent exposed environment than H80) which appears to be critical for its stabilization. Specifically, R80 forms a strong electrostatic cluster with two nearby glutamates ("EER" cluster) that apparently stabilizes the adjacent loop 57-66 which, according to our previous results, is highly dynamic in the apo-state of P187S and it is involved in its lower affinity for FAD **(Publication 2)**. Moreover, the structural switch caused by H80R and its stabilizing effect in the apo-state of P187S are also detected by our MD simulations. Since R80 seems to have diverged throughout primate evolution (30 Million years), current sequences of mammalians NQO1 still contain it. Consequently, R80 is present in rat NQO1 sequence and its available X-Ray crystal allows us to observe the same local structural switch caused by R80 together with the stabilizing EER cluster **(Publication 3).**

To see whether H80R may act as a suppressor of the detrimental effects caused by P187S, we have characterized this polymorphism in the presence of H80R. Accordingly, H80R causes a local stabilization of the FAD binding site of this polymorphism, which results in an increased FAD binding affinity and thus, in a significant enhancement of specific activity without perturbing other kinetic properties (i.e. apparent affinities for NADH or DCPIP). In addition, we have investigated in depth the structural and energetic consequences on FAD binding of R80 in P187S, evaluating its conformations in different ligation states. Our results showed that the structural and dynamic local changes caused by H80R in the FAD-binding site mainly act in the conformational ensemble of the apo-state, and shifting the equilibrium towards FAD binding competent states which, being more compact or structured, decrease the energetic penalization between apo- and holo-states present in the conformational equilibrium **(Publication 4)**. These results complete the information obtained by X-ray crystallography, NMR spectroscopy and MD simulations of H80R and suggest a mechanism of rescue of P187S by H80R in the context of an induced-fit mechanism that counterbalance the detrimental effects of P187S on the structure and dynamics of the FAD binding site in the apo-state **(Publication 3** and **Publication 4).**

So far, we have shown that the consensus H80R mutation is capable of reactivating P187S *in vitro* which should result in an increased activity within cells, without affecting degradation rates or protein levels (highly dependent on the stability and dynamic of CTD) **(Publication 1** and **Publication 2)**. To test it, we have investigated the effects of H80R on the intracellular activity, protein levels and degradation rates of wild-type and P187S by experiments in stably transfected Caco-2 cells and rabbit

reticulocyte cell-free system. On the one hand, the Western-blot analysis and measurement of NQO1 activity in transfected Caco-2 cells extracts show increased protein levels and the functional rescue of P187S in combination with H80R. While the experiments carry out in rabbit reticulocyte, where the NQO1 degradation is mediated by the proteasome, show little effect of H80R on the degradation rates, both P187S and wild-type[6]. This suggests that the effect of H80R on intracellular P187S protein levels reflects an improved intracellular folding efficiency **(Publication 3)**.

In addition, the presence of H80R could affect other important functions associated with NQO1 *in vivo* as its ability to interact with the transcription factor p73α[122,150,]**(Publication2)**. We have observed by NMR spectroscopy experiments that H80R has no large effects on interaction between NQO1 and SAMp73α **(Publication 3)**.

As previously mentioned, the effects of a mutation are context-dependent and they can be biologically altered by different mechanisms. In particular, the effects of a single mutation often depend on the presence of additional mutations (epistatic interactions) which play a key role in evolution[17,18,65,165]. Therefore, the effects of single amino acid mutations may be conditioned by the genetic background and their effects may be different depending on the sequential order in which the different mutations occur in the sequence. In the context of this thesis, the recent divergence from consensus amino acids, at different sites of the NQO1 sequence, may have contributed to lessen the tolerance of NQO1 to endure other destabilizing mutations like P187S. For this reason, we have characterized the effects of other consensus mutation for NQO1 in the absence or presence of H80R and P187S. The consensus mutation chosen is E247Q, a mutation structurally remote from H80R and FAD binding site. We expected that it

had some effect on the affected CTD of P187S which H80R is not able to stabilize (**Publication 1** and **Publication 2**). Interestingly, E247Q appears to locally stabilize the CTD of P187S and its stabilizing effects are spread to the distant NTD supporting the existence of allosteric communication between domains previously proposed (**Publication 2**). Specifically, the combination of these two mutations causes a significant increase in the FAD binding affinity of P187S and therefore, the subsequent increase in its specific activity. In addition, our NMR analyses show that the presence of Q247 seems to have no important effect on the residues involved in the interaction of NQO1 with SAMp73α (**Publication 3).**

Alternatively, we have observed that the recent fixation of non-consensus amino acids that potentially predispose towards the loss-of-function genetic diseases is not an isolated "incident" in NQO1. In particular, a similar situation has been found in the human alanine:glyoxylate aminotransferase 1 (AGT) protein, in which different genetic variations produce loss-of-function, by different mechanisms, of the enzyme causing primary hyperoxaluria type I (PH1)[71,168,169]. Previous works of our group showed that the combination of multiple consensus mutations in AGT (AGT-RHEAM) caused strong *in vitro* thermal and kinetic stabilization as well as an enhanced activity without affecting the foldability and peroxisomal targeting in mammalian cells[9]. Similar analyses to those carried out with NQO1 showed a recent divergence from consensus amino acids during mammalian evolution, in particular through primate evolution, which might decrease the local stability of AGT protein and predisposes it towards disease associated with destabilizing and/or inactivating single amino acids changes (**Publication 3)**.

Overall, the results of this third section provide evidences supporting that recently, in an evolutionary time scale, the propensity of some human proteins to undergo diseases caused by a single amino acid change has increased as consequence of the deflection of their consensus amino acids. In particular, our results demonstrate the potential of consensus amino acids to protect genetic diseases associated with a decrease in the local stability of the proteins which may affect the protein conformational landscape and alter its functionality *in vivo*. Perhaps, this consensus amino acids divergence in the human proteins analysed is nothing more that the result of the natural selection, for example fine-tuning some intracellular functions or even to facilitating the divergence to new functions unfortunately not yet identified for these proteins[18,140]. In any way, a detailed understanding of mutational effects on structure and dynamics of a protein is crucial, not only to see whether disease could be handled in different ways in mammals, but to provide a deeper knowledge on the mechanisms used by evolution to adapt the stability and function of the proteins. In addition, our results about NQO1 may be of great significance to understand the large number of metabolic diseases associated with flavo-proteins, highlighting the importance of the apo-state ensemble to determine their cofactor binding affinities[144].

6. CONCLUSIONS

Finally, after all the experimental work was done for the purposes of this study, we can draw the following conclusions:

1. The R139W polymorphism causes modest changes in protein dynamics that entail only weak alterations in activity and stability compared to wild-type. Thus, the main pathogenic mechanism for this polymorphism is likely the skipping of exon 4 which destroys the FAD and substrate binding site producing an inactive and unstable protein.

2. Both the low activity and stability of P187S appear to have a dynamic origin caused by a significant increase in local dynamics of two functionally and structurally distant sites: the N-terminal domain (NTD) associated with enzyme inactivation, and the C-terminal domain (CTD) associated with its enhanced proteasomal degradation.

3. The treatment with the natural ligand FAD may rescue the activity of P187S, but dicoumarol binding is required to correct the degradation through the flexible CTD in the holo-state.

4. We provide evidence for the existence of long-range communication between the CTD, the FAD binding site and the P187S site, which we refer to as allosteric interaction network. This network contributes to the high affinity for FAD and dicoumarol in wild-type, while the presence of P187S disrupts it causing a decrease in its binding affinities for both ligands.

5. The divergence, at several sites, from the consensus amino acids during primate evolution may have caused local protein destabilization in NQO1 (also in AGT) potentially predisposing it towards disease-associated single amino acid changes.

6. Our results show the potential of consensus amino acids to protect towards disease-associated single amino acid changes. In particular, the H80R suppressor mutation (NQO1 consensus mutation) reactivates P187S through local and dynamic stabilization of FAD binding site, which enhances its FAD binding affinity without having any effect on the CTD of the protein.

7. Structural and dynamic changes caused by H80R seem to mainly act in the conformational ensemble of the apo-state, shifting the equilibrium towards FAD binding competent states and decreasing the energetic penalization due to the conformational change.

8. A second suppressor mutation, E247Q, stabilizes the CTD of P187S and its effects are propagated to the distal NTD of the protein supporting the existence of the proposed long-range communication. Moreover, the stabilizing effects of E247Q are additive to those exerted by H80R strongly protecting P187S towards inactivation.

6. CONCLUSIONES

Para finalizar y tras todo el trabajo experimental llevado a cabo para los propósitos de este estudio, pueden extraerse las siguientes conclusiones:

1. El polimorfismo R139W causa cambios modestos en la dinámica de la proteína que implican solo alteraciones débiles en la actividad y estabilidad del polimorfismo comparado a la proteína natural. Por tanto, el principal mecanismo patogénico asociado con este polimorfismo parece ser la omisión del exón 4, la cual destruye el sitio de unión de FAD y los sustratos produciendo una proteína inactiva e inestable.

2. La baja actividad y estabilidad de P187S parecen tener un origen dinámico provocado por un aumento significativo en la dinámica local de dos sitios funcionales y estructuralmente distantes de la proteína: el dominio N-terminal (NTD) asociado con la inactivación de la enzima, y el dominio C-terminal (CTD) asociado con el aumento de la degradación proteasomal.

3. El tratamiento con FAD, ligando natural de la enzima, parece rescatar la actividad de P187S, sin embargo es necesaria la unión de dicumarol para corregir la degradación del aún flexible CTD en el estado holo.

4. Aportamos evidencias de la existencia de una comunicación de largo alcance entre el CTD, el sitio de unión a FAD y la posición P187S y a la que nos referimos como red de interacción alostérica. Esta red contribuye a la alta afinidad por FAD y dicumarol en la proteína natural, mientras la presencia de P187S la altera causando la disminución de su afinidad por ambos ligandos.

5. La divergencia, en varias posiciones de la secuencia, de aminoácidos consenso durante la evolución de los primates podría ser la causa de la desestabilización local de la proteína NQO1 (también para AGT) aumentando su predisposición a enfermedades causadas por el cambio de un único aminoácido.

6. Nuestros resultados muestran el potencial de los aminoácidos consenso para proteger enfermedades asociadas con el cambio de un único aminoácido. En particular, la mutación supresora H80R (mutación consenso de NQO1) reactiva P187S mediante una estabilización local y dinámica del sitio de unión de FAD, aumentando su afinidad por este ligando sin afectar al CTD de la proteína.

7. Los cambios estructurales y dinámicos causados por H80R parecen actuar principalmente en el conjunto conformacional del estado apo, desplazando el equilibrio hacia estados competentes para la unión de FAD y disminuyendo la penalización energética como consecuencia del cambio estructural que ocasiona.

8. Una segunda mutación supresora, E247Q, estabiliza el CTD de P187S y sus efectos se propagan al lejano NTD de la proteína, apoyando la existencia de la comunicación de largo alcance propuesta. Además los efectos estabilizantes de E247Q son aditivos a los ejercidos por H80R protegiendo considerablemente a P187S de inactivación.

7. FUTURE DIRECTIONS

Many human diseases are caused by missense mutations which appear as wild-type in other species. This phenomenon is commonly interpreted by the presence of compensatory residues ("genetic compensation") that mitigate their detrimental effects in the other species, being "neutral" mutations in the nonhuman orthologous[18,170–172]. In this sense, checking the effects of pathological mutations in these genetic backgrounds could be interesting.

Specifically, a large fraction of consensus amino acids were found in the actual sequences of murine orthologous of NQO1 and AGT after performing the mammalian sequence alignment. So, these backgrounds could help us assess the effects of disease-associated single amino acids mutation. The procedure to follow would be to introduce the mutation or mutations associated with disease in these backgrounds, and then, to characterize their structural and dynamic effects both *in vitro* and cellular experiment.

The fact that a reduction in the protein stability is the main mechanism by which the single amino acids variations cause diseases, our findings may be of great significance to understand the molecular basis of the involved diseases. In addition, our results may be important to understand the epistasis in protein evolution and to use animal models for human diseases.

Alternatively, our group is interested in evaluating whether disease-associated mutations are specific in the sequence space, in particular in relation to loss-of-function pathogenic mechanisms. Our curiosity is owing to that although random mutations are generally destabilizing and at least initially, non-conservative mutations

through a single nucleotide change should naturally occur, most missense disease-associated mutations lead to the preference of the amino acid present in the wild-type protein to only one or few among all possible, and the molecular basis of this mutation specificity is unknown[61].

In particular, evaluating the specificity of the pathogenic mechanisms in the sequence space we propose to test how different mutations of the same residue associated with disease, concretely in loss-of-function diseases, can influence the pathogenic mechanisms. For that, an in-depth characterization of a set of different amino acid changes is being performed on both P187S and the two most common mutations causing PH1, in addition to their expression in different cell systems. We hope to obtain relevant information that provides some light on the mutation specificity found in many missense diseases.

8. PERSONAL CONTRIBUTION

DIFFERENT MECHANISMS EXPLAIN THE LOSS-OF-FUNCTION CAUSED BY TWO NQO1 POLYMORPHISMS

- Site-Direct Mutagenesis
- Proteins Expression and Purification
- Activity Measurements
- Variants' Expression in Cell Cultures
- Inmunoblotting
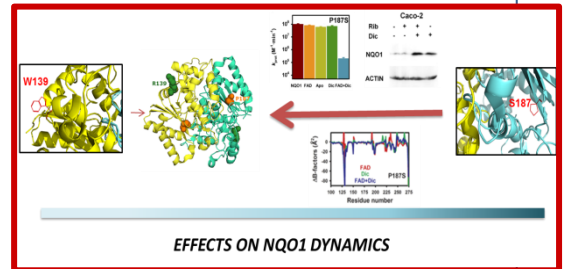
P187S DISRUPTS THE ALLOSTERIC INTERACTION NETWORK INVOLVED IN THE HIGH AFFINITY OF NQO1 FOR ITS LIGANDS

- Site-Direct Mutagenesis
- Proteins Expression and Purification
- Activity Measurements
- Mass Spectrometry
- Variants' Expression in Cell Cultures
- Inmunoblotting
- Activity Assay in Cell Extracts
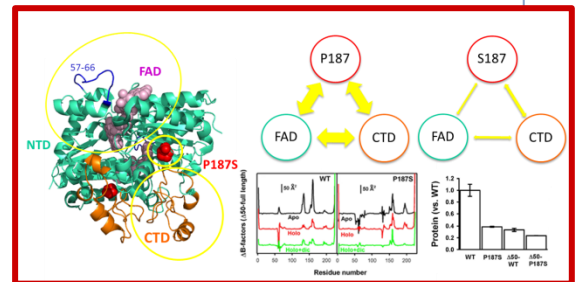
RECENT CONSENSUS AMINO ACIDS DIVERGENCE MAY CAUSE PROTEINS MORE PRONE TO LOSS-OF-FUNCTION DISEASES

- Site-Direct Mutagenesis
- Proteins Expression and Purification
- Activity Measurements
- Thermal Stability Studies
- Variants' Expression in Cell Cultures
- Inmunoblotting
- Cell Viability Assays
- Activity Assay in Cell Extracts
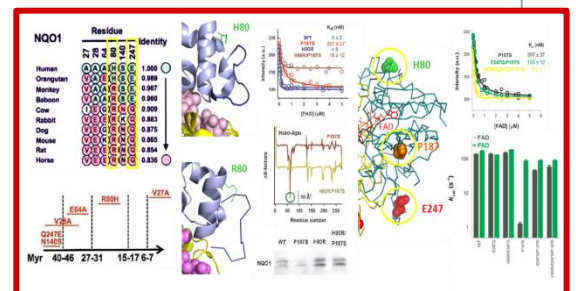
9. REFERENCES

1. Betancor-Fernandez, I., Timson, D. J., Salido, E. & Pey, A. L. Natural (and Unnatural) Small Molecules as Pharmacological Chaperones and Inhibitors in Cancer. in *Handbook of Experimental Pharmacology* (2017).

2. Pey, A. L. Protein homeostasis disorders of key enzymes of amino acids metabolism: Mutation-induced protein kinetic destabilization and new therapeutic strategies. *Amino Acids* **45,** 1331–1341 (2013).

3. Lajin, B. & Alachkar, A. The NQO1 polymorphism C609T (Pro187Ser) and cancer susceptibility: A comprehensive meta-analysis. *Br. J. Cancer* **109,** 1325–1337 (2013).

4. Pey, A. L., Megarity, C. F. & Timson, D. J. FAD binding overcomes defects in activity and stability displayed by cancer-associated variants of human NQO1. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1842,** 2163–2173 (2014).

5. Lienhart, W., Gudipati, V., Uhl, M. K., Binter, A. & Sergio, A. Collapse of the native structure caused by a single amino acid exchange in human NAD (P) H : quinone oxidoreductase. *FEBS J.* **281,** 4691–4704 (2014).

6. Siegel, D. *et al.* Rapid polyubiquitination and proteasomal degradation of a mutant form of NAD(P)H:quinone oxidoreductase 1. *Mol. Pharmacol.* **59,** 263–268 (2001).

7. Moscovitz, O. *et al.* A Mutually Inhibitory Feedback Loop between the 20S Proteasome and Its Regulator, NQO1. *Mol. Cell* **47,** 76–86 (2012).

8. Trudeau, D. L., Kaltenbach, M. & Tawfik, D. S. On the Potential Origins of the High Stability of Reconstructed Ancestral Proteins. *Mol. Biol. Evol.* **33,** 2633–2641 (2016).

9.   Mesa-Torres, N. *et al.* The consensus-based approach for gene/enzyme replacement therapies and crystallization strategies: the case of human alanine–glyoxylate aminotransferase. *Biochem. J.* **462,** 453–463 (2014).

10.   Shendure, J. & Akey, J. M. The origins, determinants, and consequences of human mutations. *Science (80-.).* **349,** 1478–1483 (2015).

11.   Martincorena, I. & Campbell, P. J. Somatic mutation in cancer and normal cells. *Science (80-.).* **349,** 1483–1489 (2015).

12.   Liu, B. *et al.* Mechanisms of mutagenesis: DNA replication in the presence of DNA damage. *Mutat Res Rev* **768,** 53–67 (2016).

13.   De Visser, J. A. G. M. & Krug, J. Empirical fitness landscapes and the predictability of evolution. *Nat. Rev. Genet.* **15,** 480–490 (2014).

14.   Orr, H. A. Fitness and its role in evolutionary genetics. *Nat. Rev. Genet.* **10,** 531–539 (2009).

15.   Loewe, L. & Hill, W. G. The population genetics of mutations: good, bad and indifferent. *Philos. Trans. R. Soc. B Biol. Sci.* **365,** 1153–1167 (2010).

16.   Eyre-Walker, A. & Keightley, P. D. The distribution of fitness effects of new mutations. *Nat. Rev. Genet.* **8,** 610–618 (2007).

17.   Soskine, M. & Tawfik, D. S. Mutational effects and the evolution of new protein functions. *Nat. Rev. Genet.* **11,** 572–582 (2010).

18.   Storz, J. F. Compensatory mutations and epistasis for protein function. *Curr. Opin. Struct. Biol.* **50,** 18–25 (2018).

19.   Brookes, A. J. The essence of SNPs. *Gene* **234,** 177–186 (1999).

20.   Kirk, B. W., Feinsod, M., Favis, R., Kliman, R. M. & Barany, F. Single nucleotide polymorphism seeking long term association with complex disease. *Nucleic Acids Res.* **30,** 3295–3311 (2002).

21.   de Souza, D. A. S. *et al.* Cystic fibrosis in Afro-Brazilians: XK haplotypes analysis

supports the European origin of p.F508del mutation. *Genetica* **145,** 19–25 (2017).

22. Wu, J. & Jiang, R. Prediction of Deleterious Nonsynonymous Single-Nucleotide Polymorphism for Human Diseases. *Sci. World J.* **2013,** (2013).

23. Sauna, Z. E. & Kimchi-Sarfaty, C. Understanding the contribution of synonymous mutations to human disease. *Nat. Rev. Genet.* **12,** 683–691 (2011).

24. Mesa-Torres, N., Salido, E. & Pey, A. L. The lower limits for protein stability and foldability in primary hyperoxaluria type i. *Biochim. Biophys. Acta - Proteins Proteomics* **1844,** 2355–2365 (2014).

25. Fraser-Pitt, D. & O'Neil, D. Cystic fibrosis – a multiorgan protein misfolding disease. *Futur. Sci. OA* **1,** fso.15.57 (2015).

26. Schuster-Böckler, B. & Bateman, A. Protein interactions in human genetic diseases. *Genome Biol.* **9,** R9 (2008).

27. Mesa-Torres, N., Tomic, N., Albert, A., Salido, E. & Pey, A. L. Molecular recognition of PTS-1 cargo proteins by Pex5P: Implications for protein mistargeting in primary hyperoxaluria. *Biomolecules* **5,** 121–141 (2012).

28. Yue, P., Li, Z. & Moult, J. Loss of protein structure stability as a major causative factor in monogenic disease. *J. Mol. Biol.* **353,** 459–473 (2005).

29. Wang, Z. & Moult, J. SNPs , Protein Structure , and Disease. *Hum. Mutat.* **17,** 263–270 (2001).

30. Ferrer-Costa, C., Orozco, M. & de la Cruz, X. Characterization of disease-associated single amino acid polymorphisms in terms of sequence and structure properties. *J. Mol. Biol.* **315,** 771–786 (2002).

31. Steward, R. E., MacArthur, M. W., Laskowski, R. A. & Thornton, J. M. Molecular basis of inherited diseases: A structural perspective. *Trends Genet.* **19,** 505–513 (2003).

32. Zhang, Z. *et al.* Analyzing Effects of Naturally Occurring Missense Mutations.

Comput. Math. Methods Med. **2012,** (2012).

33.   Araya, C. L. *et al.* A fundamental protein property, thermodynamic stability, revealed solely from large-scale measurements of protein function. *Proc. Natl. Acad. Sci.* **109,** 16858–16863 (2012).

34.   Valastyan, J. S. & Lindquist, S. Mechanisms of protein-folding diseases at a glance. *Dis. Model. Mech.* **7,** 9–14 (2014).

35.   Braselmann, E., Chaney, J. L. & Clark, P. L. Folding the proteome. *Trends Biochem. Sci.* **38,** 337–344 (2013).

36.   Oldfield, C. J. & Dunker, A. K. Intrinsically Disordered Proteins and Intrinsically Disordered Protein Regions. *Annu. Rev. Biochem.* **83,** 553–584 (2014).

37.   Anfinsen, C. B. Principles that Govern the Folding of Protein Chains. *Science (80-.).* **181,** 223–230 (1973).

38.   Vabulas, R., Raychaudhuri, S., Hayer-Hartl, M. & Ulrich Hartl, F. Protein Folding in the Cytoplasm and the Heat Shock Response. *Cold Spring Harb. Perspect. Biol.* **2,** (2010).

39.   Dobson, C. M. Protein folding and misfolding. *Nature* **426,** 884–890 (2003).

40.   Muntau, A. C., Leandro, J., Staudigl, M., Mayer, F. & Gersting, S. W. Innovative strategies to treat protein misfolding in inborn errors of metabolism: Pharmacological chaperones and proteostasis regulators. *J. Inherit. Metab. Dis.* **37,** 505–523 (2014).

41.   Onuchic, J. N., Luthey-Schulten, Z. & Wolynes, P. G. Theory of protein folding: the energy landscape perspective. *Annu. Rev. Phys. Chem.* **48,** 545–600 (1997).

42.   Kim, Y. E., Hipp, M. S., Bracher, A., Hayer-Hartl, M. & Ulrich Hartl, F. Molecular Chaperone Functions in Protein Folding and Proteostasis. *Annu. Rev. Biochem.* **82,** 323–355 (2013).

43.   Hartl, F. U., Bracher, A. & Hayer-Hartl, M. Molecular chaperones in protein folding and proteostasis. *Nature* **475,** 324–332 (2011).

44. Hartl, F. U. & Hayer-Hartl, M. Converging concepts of protein folding in vitro and in vivo. *Nat. Struct. Mol. Biol.* **16,** 574–581 (2009).

45. Hartl, F. U. Molecular chaperones in cellular protein folding. *Nature* **381,** 571–579 (1996).

46. Powers, E. T., Morimoto, R. I., Dillin, A., Kelly, J. W. & Balch, W. E. Biological and Chemical Approaches to Diseases of Proteostasis Deficiency. *Annu. Rev. Biochem.* **78,** 959–991 (2009).

47. Sanchez-Ruiz, J. M. Protein kinetic stability. *Biophys. Chem.* **148,** 1–15 (2010).

48. Pallarès, I. & Ventura, S. Understanding and predicting protein misfolding and aggregation: Insights from proteomics. *Proteomics* **16,** 2570–2581 (2016).

49. Balchin, D., Hayer-Hartl, M. & Hartl, F. U. In vivo aspects of protein folding and quality control. *Science (80-.).* **353,** (2016).

50. Costas, M. *et al.* Between-Species Variation in the Kinetic Stability of TIM Proteins Linked to Solvation-Barrier Free Energies. *J. Mol. Biol.* **385,** 924–937 (2009).

51. Jaswal, S. S., Sohl, J. L., Davis, J. H. & Agard, D. A. Energetic landscape of α-lytic protease optimizes longevity through kinetic stability. *Nature* **415,** 343–346 (2002).

52. Balch, W. E. & Morimoto, R. I. Adapting Proteostatis for Disease Intervention. *Science (80-.).* **319,** 916–919 (2008).

53. Buchner, J. Supervising the fold: functional principles of molecular chaperones. *FASEB J.* **10,** 10–19 (1996).

54. Labbadia, J. & Morimoto, R. I. The Biology of Proteostasis in Aging and Disease. *Annu. Rev. Biochem.* **84,** 435–464 (2015).

55. Shiber, A. & Ravid, T. Chaperoning proteins for destruction: Diverse roles of Hsp70 chaperones and their co-chaperones in targeting misfolded proteins to the proteasome. *Biomolecules* **4,** 704–724 (2014).

56. Sontag, E. M., Samant, R. S. & Frydman, J. Mechanisms and Functions of Spatial Protein Quality Control. **86,** 97–122 (2017).

57. Inobe, T. & Matouschek, A. Paradigms of protein degradation by the proteasome. *Curr. Opin. Struct. Biol.* **27,** 156–154 (2014).

58. Ben-Nissan, G. & Sharon, M. Regulating the 20S proteasome ubiquitin-independent degradation pathway. *Biomolecules* **4,** 862–884 (2014).

59. Jariel-Encontre, I., Bossis, G. & Piechaczyk, M. Ubiquitin-independent degradation of proteins by the proteasome. *Biochim. Biophys. Acta - Rev. Cancer* **1786,** 153–177 (2008).

60. 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.* **32,** 18–24 (2015).

61. Tokuriki, N., Stricher, F., Schymkowitz, J., Serrano, L. & Tawfik, D. S. The Stability Effects of Protein Mutations Appear to be Universally Distributed. *J. Mol. Biol.* **369,** 1318–1332 (2007).

62. Taverna, D. M. & Goldstein, R. A. Why Are Proteins Marginally Stable ? *Proteins* **109,** 105–109 (2002).

63. Pace, C. N. *et al.* Contribution of hydrogen bonds to protein stability. *Protein Sci.* **23,** 652–661 (2014).

64. Surguchev, A. & Surguchov, A. Conformational diseases: Looking into the eyes. *Brain Res. Bull.* **81,** 12–24 (2010).

65. DePristo, M. A., Weinreich, D. M. & Hartl, D. L. Missense meanderings in sequence space: A biophysical view of protein evolution. *Nat. Rev. Genet.* **6,** 678–687 (2005).

66. Pakula, A. A. & Sauer, R. T. Genetic Analysis of Protein Stability and Function. *Annu. Rev. Genet.* **23,** 289–310 (1989).

67. Studer, R. A., Dessailly, B. H. & Orengo, C. A. Residue mutations and their impact

on protein structure and function: detecting beneficial and pathogenic changes. *Biochem. J.* **449,** 581–594 (2013).

68. Porter, C. T. The Catalytic Site Atlas: a resource of catalytic sites and residues identified in enzymes using structural data. *Nucleic Acids Res.* **32,** 129D–133 (2004).

69. Camps, M., Herman, A., Loh, E. & Loeb, L. Genetic Constraints on Protein Evolution. *Crit. Rev. Biochem. Mol. Biol.* **42,** (2007).

70. Albers, C. A. *et al.* Compound inheritance of a low-frequency regulatory SNP and a rare null mutation in exon-junction complex subunit RBM8A causes TAR syndrome. *Nat. Genet.* **44,** 1–14 (2012).

71. Salido, E., Pey, A. L., Rodriguez, R. & Lorenzo, V. Primary hyperoxalurias: Disorders of glyoxylate detoxification. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1822,** 1453–1464 (2012).

72. Mittag, T., Kay, L. E. & Forman-Kaya, J. D. Protein dynamics and conformational disorder in molecular recognition. *J. Mol. Recognit.* **23,** 105–116 (2010).

73. Kokkinidis, M., Glykos, N. M. & Fadouloglou, V. E. *Protein flexibility and enzymatic catalysis*. *Advances in Protein Chemistry and Structural Biology* **87,** (2012).

74. Zhao, Q. On the indirect relationship between protein dynamics and enzyme activity. *Prog. Biophys. Mol. Biol.* **125,** 52–60 (2017).

75. Teilum, K., Olsen, J. G. & Kragelund, B. B. Functional aspects of protein flexibility. *Cell. Mol. Life Sci.* **66,** 2231–2247 (2009).

76. Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450,** 964–972 (2007).

77. Shoichet, B. K., Baase, W. A., Kuroki, R. & Matthews, B. W. A relationship between protein stability and protein function. *Proc. Natl. Acad. Sci.* **92,** 452–456 (1995).

78.   Nussinov, R. & Tsai, C.-J. Allostery in Disease and in Drug Discovery. *Cell* **153,** 293–305 (2013).

79.   Caron, N. S., Desmond, C. R., Xia, J. & Truant, R. Polyglutamine domain flexibility mediates the proximity between flanking sequences in huntingtin. *Proc. Natl. Acad. Sci.* **110,** 14610–14615 (2013).

80.   Luque, I. & Freire, E. Structural stability of binding sites: consequences for binding affinity and allosteric effects. *Proteins* **4,** 63–71 (2000).

81.   Ma, B., Tsai, C. J., Haliloğlu, T. & Nussinov, R. Dynamic allostery: Linkers are not merely flexible. *Structure* **19,** 907–917 (2011).

82.   Tsai, C. J. & Nussinov, R. A Unified View of 'How Allostery Works'. *PLoS Comput. Biol.* **10,** (2014).

83.   Tzeng, S. R. & Kalodimos, C. G. Dynamic activation of an allosteric regulatory protein. *Nature* **462,** 368–372 (2009).

84.   Tzeng, S. R. & Kalodimos, C. G. Protein dynamics and allostery: An NMR view. *Curr. Opin. Struct. Biol.* **21,** 62–67 (2011).

85.   Good, M. C., Zalatan, J. G. & Lim, W. A. Scaffold Proteins: Hubs for Controlling the Flow of Cellular Information. *Science (80-.).* **332,** 680–686 (2011).

86.   Nussinov, R., Tsai, C.-J. & Ma, B. The Underappreciated Role of Allostery in the Cellular Network. *Annu. Rev. Biophys.* **42,** 169–189 (2013).

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

88.   Li, M., Petukh, M., Alexov, E. & Panchenko, A. R. Predicting the Impact of Missense Mutations on Protein – Protein Binding A ffi nity. (2014).

89.   David, A. & Sternberg, M. J. E. The Contribution of Missense Mutations in Core and Rim Residues of Protein-Protein Interfaces to Human Disease. *J. Mol. Biol.* **427,** 2886–2898 (2015).

90. Kysilka, J. & Vondrášek, J. Towards a better understanding of the specificity of protein-protein interaction. *J. Mol. Recognit.* **25,** 604–615 (2012).

91. Janin, J. Principles of protein-protein recognition from structure to thermodynamics. *Biochimie* **77,** 497–505 (1995).

92. Bogan, A. A. & Thorn, K. S. Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* **280,** 1–9 (1998).

93. Li, Y., Liu, Z., Han, L., Li, C. & Wang, R. Mining the characteristic interaction patterns on protein-protein binding interfaces. *J. Chem. Inf. Model.* **53,** 2437–2447 (2013).

94. Moreira, I. S., Fernandes, P. A. & Ramos, M. J. Hot spots—A review of the protein–protein interface determinant amino-acid residues. *Proteins* **68,** 803–812 (2007).

95. Guo, W., Wisniewski, J. A. & Ji, H. Hot spot-based design of small-molecule inhibitors for protein-protein interactions. *Bioorganic Med. Chem. Lett.* **24,** 2546–2554 (2014).

96. Wang, X., Tomso, D. J., Liu, X. & Bell, D. A. Single nucleotide polymorphism in transcriptional regulatory regions and expression of environmentally responsive genes. *Toxicol. Appl. Pharmacol.* **207,** 84–90 (2005).

97. Bauer, N. C., Doetsch, P. W. & Corbett, A. H. Mechanisms Regulating Protein Localization. *Traffic* **16,** 1039–1061 (2015).

98. Jarjanazi, H., Savas, S., Pabalan, N., Dennis, J. W. & Ozcelik, H. Biological implications of SNPs in signal peptide domains of human proteins. *Proteins* **70,** 394–403 (2008).

99. Ernster, L., Danielson, L. & Ljunggren, M. DT diaphorase. I. Purification from the soluble fraction of rat-liver cytoplasm, and properties. *Biochim. Biophys. Acta* **58,** 171–188 (1962).

100. Hosoda, S, Nakamuru, W, Hayashi, K. Properties Diaphorase and Reaction

Mechanism from Rat Liver of DT. *October* **249,** 6416–6423 (1974).

101. Jaiswal, A. K., McBride, O. W., Adesnik, M. & Nebert, D. W. Human dioxin-inducible cytosolic NAD(P)H:menadione oxidoreductase. cDNA sequence and localization of gene to chromosome 16. *J. Biol. Chem.* **263,** 13572–13578 (1988).

102. Atia, A., Alrawaiq, N. & Abdullah, A. A review of NAD(P)H: Quinone oxidoreductase 1 (NQO1); A multifunctional antioxidant enzyme. *J. Appl. Pharm. Sci.* **4,** 118–122 (2014).

103. Danson, S., Ward, T. H., Butler, J. & Ranson, M. DT-diaphorase: A target for new anticancer drugs. *Cancer Treat. Rev.* **30,** 437–449 (2004).

104. Siegel, D. & Ross, D. Immunodetection of NAD(P)H:quinone oxidoreductase 1 (NQO1) in human tissues. *Free Radic. Biol. Med.* **29,** 246–253 (2000).

105. Pey, A. L., Megarity, C. & Medina-Carmona, Encarnación Timson, D. Natural Small Molecules as Stabilizers and Activators of Cancer-Associated NQO1 Polymorphisms. *Curr. Drug Targets* **17,** 1506–1514 (2016).

106. Asher, G., Dym, O., Tsvetkov, P., Adler, J. & Shaul, Y. The crystal structure of NAD(P)H quinone oxidoreductase 1 in complex with its potent inhibitor dicoumarol. *Biochemistry* **45,** 6372–6378 (2006).

107. Faig, M. *et al.* Structures of recombinant human and mouse NAD(P)H:quinone oxidoreductases: species comparison and structural changes with substrate binding and release. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 3177–82 (2000).

108. Li, R., Bianchet, M. A., Talalay, P. & Amzel, L. M. The three-dimensional structure of NAD(P)H:quinone reductase, a flavoprotein involved in cancer chemoprotection and chemotherapy: mechanism of the two-electron reduction. *Proc. Natl. Acad. Sci.* **92,** 8846–8850 (1995).

109. Ross, D. Quinone reductases multitasking in the metabolic world. *Drug Metab. Rev.* **36,** 639–654 (2004).

110. Lind, C., Hochstein, P. & Ernster, L. DT-diaphorase as a quinone reductase: A

cellular control device against semiquinone and superoxide radical formation. *Arch. Biochem. Biophys.* **216,** 178–185 (1982).

111. Munday, R. Activation and Detoxification of Naphthoquinones by NAD (P) H : Quinone Oxidoreductase. *Methods Enzymol.* **382,** 364–380 (2004).

112. Ross, D. *et al.* NAD(P)H:quinone oxidoreductase 1 (NQO1): Chemoprotection, bioactivation, gene regulation and genetic polymorphisms. *Chem. Biol. Interact.* **129,** 77–97 (2000).

113. Dinkova-Kostova, A. & Talalay, P. NAD(P)H:quinone acceptor oxidoreductase 1 (NQO1), a multifunctional antioxidant enzyme and exceptionally versatile cytoprotector. *Arch. Biochem. Biophys.* **501,** 116–123 (2010).

114. Talalay, P., De Long, M. J. & Prochaska, H. J. Identification of a common chemical signal regulating the induction of enzymes that protect against chemical carcinogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **85,** 8261–8265 (1988).

115. Ross, D. Functions and distribution of NQO1 in human bone marrow: Potential clues to benzene toxicity. *Chem. Biol. Interact.* **153–154,** 137–146 (2005).

116. Bauer, A. K. *et al.* Genetic Susceptibility to Benzene-induced Toxicity : Role of NADPH : Quinone Oxidoreductase-1 Genetic Susceptibility to Benzene-induced Toxicity : Role of NADPH : 929–935 (2003).

117. Siegel, D. *et al.* NAD(P)H : Quinone Oxidoreductase 1 : Role as a Superoxide Scavenger. *Mol. Pharmacol.* **65,** 1238–1247 (2004).

118. Zhu, H. *et al.* The highly expressed and inducible endogenous NAD(P)H:quinone oxidoreductase 1 in cardiovascular cells acts as a potential superoxide scavenger. *Cardiovasc. Toxicol.* **7,** 202–211 (2007).

119. Siegel, D., Gibson, N. W., Preusch, P. C. & Ross, D. Metabolism of mitomycin C by DT-diaphorase: role in mitomycin C- induced DNA damage and cytotoxicity in human colon carcinoma cells. *Cancer Res.* **50,** 7483–7489 (1990).

120. Phillips, R. M. Prospects for bioreductive drug development. *Expert Opin.*

Investig. Drugs **7,** 905–928 (1998).

121. Gutierrez, P. The role of a NAD(P)H oxidoreductase (DT-Diaphorase) in the bioactivation of quinone-containing antitumor agents: a review. *Free Radic. Biol. Med.* **29,** 263–275 (2000).

122. Sollner, S. *et al.* Quinone reductase acts as a redox switch of the 20S yeast proteasome. *EMBO Rep.* **10,** 65–70 (2009).

123. Meek, D. W. Regulation of the p53 response and its relationship to cancer. *Biochem. J.* **469,** 325–346 (2015).

124. Xu, L. L. *et al.* The NAD(P)H:quinone oxidoreductase 1 gene polymorphism and lung cancer: differential susceptibility based on smoking behavior. *Cancer Epidemiol. Biomarkers Prev.* **10,** 303–309 (2001).

125. Fagerholm, R. *et al.* NAD(P)H:quinone oxidoreductase 1 NQO1*2 genotype (P187S) is a strong prognostic and predictive factor in breast cancer. *Nat. Genet.* **40,** 844–853 (2008).

126. Lee, J. *et al.* NAD(P)H: Quinone Oxidoreductase 1 and NRH: Quinone Oxidoreductase 2 polymorphisms in papillary thyroid microcarcinoma: Correlation with phenotype. *Yonsei Med. J.* **54,** 1158–1167 (2013).

127. Peng, Q. *et al.* The NQO1 Pro187Ser polymorphism and breast cancer susceptibility: Evidence from an updated meta-analysis. *Diagn. Pathol.* **9,** (2014).

128. Rothman, N. *et al.* Benzene Poisoning , a Risk Factor for Hematological Malignancy , Is Associated with the NQO1 609 C → T Mutation and Rapid Fractional Excretion of Chlorzoxazone in Brief Benzene Poisoning , a Risk Factor for Hematological Malignancy , Is Associated with Mu. *Cancer Res.* 2839–2842 (1997).

129. Asher, G., Lotem, J., Kama, R., Sachs, L. & Shaul, Y. NQO1 stabilizes p53 through a distinct pathway. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 3099–3104 (2002).

130. Traver, R. D. *et al.* NAD (P) H : Quinone Oxidoreductase Gene Expression in

Human Colon Carcinoma Cells : Characterization of a Mutation Which Modulates DT-Diaphorase Activity and Mitomycin Sensitivity NAD (P) H : Quinone Oxidoreductase Gene Expression in Human Colon Carci. *Cancer Res.* **52,** 797–802 (1992).

131. Edwards, Y. H., Potter, J. & Hopkinson, D. A. Human FAD-Dependent NAD (P) H Diaphorase. **187,** 429–436 (1980).

132. Traver, R. D. *et al.* Characterization of a polymorphism in NAD(P)H: quinone oxidoreductase (DT-diaphorase). *Br. J. Cancer* **75,** 69–75 (1997).

133. Hu, L. T., Stamberg, J. & Pan, S. The NAD(P)H:quinone oxidoreductase locus in human colon carcinoma HCT 116 cells resistant to mitomycin C. *Cancer Res.* **56,** 5253–5259 (1996).

134. Pan, S. S., Forrest, G. L., Akman, S. A. & Hu, L. T. NAD(P)H:Quinone Oxidoreductase Expression and Mitomycin C Resistance Developed by Human Colon Cancer HCT 116 Cells. *Cancer Res.* **55,** 330–335 (1995).

135. Gasdaska, P. Y., Fisher, H. & Powis, G. An Alternatively Spliced Form of NQOj (DT-Diaphorase) Messenger RNA Lacking the Putative Quinone Substrate Binding Site Is Present in Human Normal and Tumor Tissues1 SW480 HT-29. *Cancer Res.* **55,** 2542–2547 (1995).

136. Henzler-Wildman, K. A. *et al.* A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* **450,** 913–916 (2007).

137. Tzeng, S. R. & Kalodimos, C. G. Protein activity regulation by conformational entropy. *Nature* **488,** 236–240 (2012).

138. Fraser, J. S., Clarkson, M. W., Degnan, S. C., Erion, R. & Alber, T. Hidden alternate structure of proline isomerase essential for catalysis. *Nature* **462,** 669–673 (2009).

139. Berko, D. *et al.* The Direction of Protein Entry into the Proteasome Determines the Variety of Products and Depends on the Force Needed to Unfold Its Two

Termini. *Mol. Cell* **48,** 601–611 (2012).

140. Tokuriki, N. & Tawfik, D. S. Protein Dynamism and Evolvability. *Science (80-.).* **324,** 203–208 (2009).

141. Eguch-Ishimae, M. *et al.* The association of a distinctive allele of NAD(P)H:quinone oxidoreductase with pediatric acute lymphoblastic leukemias with MLL fusion genes in Japan. *Haematologica* **90,** 1511–1515 (2005).

142. Lienhart, W.-D. *et al.* Catalytic competence, structure and stability of the cancer associated R139W variant of the human NAD(P)H:quinone oxidoreductase 1 (NQO1). *FEBS J.* **284,** 1233–1245 (2017).

143. Bruce N Ames, I. E.-S. and E. a S. High-dose vitamin therapy stimulates variant enzymes with\rdecreased coenzyme binding affinity (increased Km): relevance\rto genetic disease and polymorphisms. *Am J Clin Nutr* **75,** 616–658 (2002).

144. Lienhart, W. D., Gudipati, V. & MacHeroux, P. The human flavoproteome. *Arch. Biochem. Biophys.* **535,** 150–162 (2013).

145. Yu, H. & Matouschek, A. Recognition of Client Proteins by the Proteasome. *Annu. Rev. Biophys.* **46,** 149–173 (2017).

146. Martínez-Limón, A. *et al.* Recognition of enzymes lacking bound cofactor by protein quality control. *Proc. Natl. Acad. Sci.* **113,** 12156–12161 (2016).

147. Medina-Carmona, E. *et al.* Conformational dynamics is key to understanding loss-of-function of NQO1 cancer-associated polymorphisms and its correction by pharmacological ligands. *Sci. Rep.* **6,** 1–13 (2016).

148. Gámez, A. *et al.* Protein misfolding diseases: Prospects of pharmacological treatment. *Clin. Genet.* 450–458 (2017). doi:10.1111/cge.13088

149. Barrera, F. N., Poveda, J. A., González-Ros, J. M. & Neira, J. L. Binding of the C-terminal Sterile α Motif (SAM) Domain of Human p73 to Lipid Membranes. *J. Biol. Chem.* **278,** 46878–46885 (2003).

150. Liu, G. & Chen, X. The C-terminal sterile ?? motif and the extreme C terminus regulate the transcriptional activity of the ?? isoform of p73. *J. Biol. Chem.* **280,** 20111–20119 (2005).

151. Asher, G., Tsvetkov, P., Kahana, C. & Shaul, Y. A mechanism of ubiquitin-independent proteasomal degradation of the tumor suppressors p53 and p73. *Genes Dev.* **19,** 316–321 (2005).

152. Steipe, B., Schiller, B., Pluäckthun, A. & Steinbacher, S. Sequence statistics reliably predict stabilizing mutations in a protein domain. *Journal of Molecular Biology* **240,** 188–192 (1994).

153. Pey, A. *et al.* Engineering proteins with tunable thermodynamic and kinetic stabilities. *Proteins Struct. Funct. Genet.* **71,** 165–174 (2008).

154. Sullivan, B. J. *et al.* Stabilizing proteins from sequence statistics: The interplay of conservation and correlation in triosephosphate isomerase stability. *J. Mol. Biol.* **420,** 384–399 (2012).

155. Wijma, H. J., Floor, R. J. & Janssen, D. B. Structure- and sequence-analysis inspired engineering of proteins for enhanced thermostability. *Curr. Opin. Struct. Biol.* **23,** 588–594 (2013).

156. Lehmann, M., Pasamontes, L., Lassen, S. F. & Wyss, M. The consensus concept for thermostability engineering of proteins. *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* **1543,** 408–415 (2000).

157. Rodriguez-Larrea, D. *et al.* Role of conservative mutations in protein multi-property adaptation. *Biochem. J.* **429,** 243–249 (2010).

158. Risso, V. A., Gavira, J. A., Gaucher, E. A. & Sanchez-Ruiz, J. M. Phenotypic comparisons of consensus variants versus laboratory resurrections of Precambrian proteins. *Proteins Struct. Funct. Bioinforma.* **82,** 887–896 (2014).

159. Akanuma, S. *et al.* Experimental evidence for the thermophilicity of ancestral life. *Proc. Natl. Acad. Sci.* **110,** 11067–11072 (2013).

160. Risso, V. A., Gavira, J. A., Mejia-Carmona, D. F., Gaucher, E. A. & Sanchez-Ruiz, J. M. Hyperstability and substrate promiscuity in laboratory resurrections of precambrian β-lactamases. *J. Am. Chem. Soc.* **135,** 2899–2902 (2013).

161. Kimura, M. The neutral theory of molecular evolution: a review of recent evidence. *Japanese Journal of Genetics* **66,** 367–386 (1991).

162. Sikosek, T. & Chan, H. S. Biophysics of protein evolution and evolutionary protein biophysics. *J. R. Soc. Interface* **11,** 20140419–20140419 (2014).

163. Tokuriki, N. & Tawfik, D. S. Stability effects of mutations and protein evolvability. *Curr. Opin. Struct. Biol.* **19,** 596–604 (2009).

164. Socha, R. D. & Tokuriki, N. Modulating protein stability - Directed evolution strategies for improved protein function. *FEBS J.* **280,** 5582–5595 (2013).

165. Gong, L. I., Suchard, M. A. & Bloom, J. D. Stability-mediated epistasis constrains the evolution of an influenza protein. *Elife* **2013,** 1–19 (2013).

166. Risso, V. A. *et al.* Mutational studies on resurrected ancestral proteins reveal conservation of site-specific amino acid preferences throughout evolutionary history. *Mol. Biol. Evol.* **32,** 440–455 (2015).

167. Medina-Carmona, E. *et al.* Site-to-site interdomain communication may mediate different loss-of-function mechanisms in a cancer-associated NQO1 polymorphism. *Sci. Rep.* **7,** 1–18 (2017).

168. Santana, A., Salido, E., Torres, A. & Shapiro, L. J. Primary hyperoxaluria type 1 in the Canary Islands: a conformational disease due to I244T mutation in the P11L-containing alanine:glyoxylate aminotransferase. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 7277–82 (2003).

169. Oppici, E., Montioli, R. & Cellini, B. Liver peroxisomal alanine:glyoxylate aminotransferase and the effects of mutations associated with Primary Hyperoxaluria Type I: An overview. *Biochim. Biophys. Acta - Proteins Proteomics* **1854,** 1212–1219 (2015).

170. Xu, J. & Zhang, J. Why human disease-associated residues appear as the wild-type in other species: Genome-scale structural evidence for the compensation hypothesis. *Mol. Biol. Evol.* **31,** 1787–1792 (2014).

171. Gao, L. & Zhang, J. Why are some human disease-associated mutations fixed in mice? *Trends Genet.* **19,** 678–681 (2003).

172. Jordan, D. *et al.* Identification of cis-suppression of human disease mutations by comparative genomics. **524,** 225–229 (2015).

173. Oppici, E. *et al.* Pyridoxamine and pyridoxal are more effective than pyridoxine in rescuing folding-defective variants of human alanine: Glyoxylate aminotransferase causing primary hyperoxaluria type I. *Hum. Mol. Genet.* **24,** 5500–5511 (2015).