OXFORD UNIVERSITY PRESS | Human Molecular Genetics

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

Journal:	*Human Molecular Genetics*
Manuscript ID	HMG-2017-TF-00293.R1
Manuscript Type:	2 General Article - UK Office
Date Submitted by the Author:	n/a
Complete List of Authors:	Medina-Carmona, Encarnación; University of Granada, Physical Chemistry Fuchs, Julian; Leopold-Franzens-University Innsbruck, Theoretical Chemistry Gaviria, Jose A.; Laboratorio de Estudios Cristalográficos, IACT (CSIC-UGR), Protein crystallization and crystallography Mesa-Torres, Noel; University of Granada, Physical Chemistry Neira, Jose Luis; Universidad Miguel Hernandez de Elche, Instituto de Biología Molecular y Celular Salido, Eduardo; University La Laguna, Pathology Palomino-Morales, Rogelio; University of Granada, Biochemistry and Molecular Biology I Burgos, Miguel; University of Granada, Genetics; University of Granada, Biomedical research center Timson, David ; University of Brighton, Pharmacy and Biomolecular Sciences Pey, Angel; University of Granada, Physical Chemistry
Key Words:	Conformational disease, Protein stability, Divergent evolution, Loss-of-function mechanism

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

Encarnación Medina-Carmona[a], Julian E. Fuchs[b,1], Jose A. Gavira[c], Noel Mesa-Torres[a], Jose L. Neira[d,e], Eduardo Salido[f], Rogelio Palomino-Morales[g], Miguel Burgos[h], David J. Timson[i] and Angel L. Pey[a,*].

[a] Department of Physical Chemistry, Faculty of Sciences, University of Granada, Granada (Spain).

[b] Institute of General, Inorganic and Theoretical Chemistry, Faculty of Chemistry and Pharmacy, University of Innsbruck, Innsbruck (Austria).

[c] Laboratorio de Estudios Cristalográficos, IACT (CSIC-UGR), Armilla, Granada, (Spain).

[d] Instituto de Biología Molecular y Celular, Universidad Miguel Hernández, Avda. del Ferrocarril s/n, 03202 Elche (Alicante), Spain.

[e] Instituto de Biocomputación y Física de los Sistemas Complejos (BIFI), 50009 Zaragoza, Spain.

[f] Centre for Biomedical Research on Rare Diseases (CIBERER), Hospital Universitario de Canarias, Tenerife (Spain).

[g] Department of Biochemistry and Molecular Biology I, Faculty of Sciences and Biomedical Research Center (CIBM), University of Granada, Granada (Spain).

[h] Department of Genetics and Institute of Biotechnology, Faculty of Sciences, and Biomedical Research Center (CIBM), University of Granada, Granada (Spain)

[i] School of Pharmacy and Biomolecular Sciences, The University of Brighton, Brighton (UK).

[1] Present address: Department of Medicinal Chemistry, Boehringer Ingelheim RCV GmbH & Co KG, Vienna, Austria.

[*] to whom correspondence should be addressed: angelpey@ugr.es. Department of Physical Chemistry, University of Granada, Av. Fuentenueva s/n, E-18071, Granada (Spain). Phone: +34-958243173. Fax: +34-958242747.

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 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 (SI text) 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 (Figure 1A,B and Table S1), consequently increasing their kinetic stability (i.e. slowing down irreversible inactivation, 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 (Figure 1C and D), which occurs between Ser72 and Val73 (18). The H80R mutation has no effect on the sensitivity of P187S towards proteolysis (Figure 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)) (Figure 1E, F and 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 (Figure 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; Figure 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 human enzyme (Figure 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 (Figure 2A-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 $kcal \cdot mol^{-1}$ for human H80R and rat NQO1 compared to the human WT enzyme (Figure 2C). Overall, Arg80 develops more stabilizing electrostatic interactions with its environment in rat than in human NQO1 (Figure 2C-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 (Figure 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 (Figure 2E-F), indicating that the Arg-to-His mutation has been

recently fixed along primate evolution. 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 (Figure 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 esentially 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 (Figure 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 (Figure 3, 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 (Figure 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 (Figure 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 (Figure S1A and S2A). In human NQO1, all six consensus mutations are individually stabilizing (Figure 1A), and only the adjacent A27V and A28E mutations are spatially close (Figure S1B), while the five consensus mutations on human AGT are individually stabilizing and their effects are additive (28), with four of them forming a intersubunit cluster of stabilizing interactions (Figure S2B). With a few exceptions, the consensus mutations do not cause large changes in the solvent accessibility (Figure S1C and S2C), and a significant fraction of them show an electrostatic origin for their stabilizing effects (Figure S1D and S2D). In rat NQO1, the consensus amino acids Arg80 and Glu28 combined cause a ≈3 kcal·mol^{-1} stabilization compared to His80 and Ala28 in the human enzyme, due to local optimization of electrostatic interactions (Figure S1D,E and Figure 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 kcal·mol^{-1} of electrostatic stabilization, besides some small

differences with hAGT-RHEAM such as the absence of Glu344 in the mouse enzyme (Figure S2 D,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 (Figure 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 (Figure 4B), causing a concomitant large increase in specific activity (Figure 4C) without perturbing other kinetic properties (e.g. the apparent affinities for NADH and DCPIP as substrate; Table S4 and Figure 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 (Figure 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 (Figure 4D and S5). Molecular dynamics (MD) simulations recapitulate the structural switch caused by H80R (Figure S6) further showing that this switch dynamically stabilizes the loop 57-66, particularly in the apo-state of P187S (Figure 4E,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 (Figure 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; Figure 4G and H). The levels of endogenous NQO1 P187S in Caco-2 cells (lower immunodetected bands in Figure 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 (Figure 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 (Figure 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) (Figure 4K and S7). Importantly, NQO1 activity measurements (Figures 4J and K) correlate well with the levels of transfected NQO1 proteins (Figure 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 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 (Figure 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 Figure 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 (Figure 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; Figure 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 (Figure 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 (Figure 5C-D). FAD titrations confirmed stronger binding of FAD to P187S in the presence of E247Q, particularly in combination with H80R (Figure 5E and S5), while very tight FAD binding was found for all variants of WT NQO1 (Figure 5F and 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 (Figure 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 epistastic 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 (Figure 1C-D), while the E247Q mutation stabilizes the CTD and the distal NTD, and importantly, enhances the local stabilization exerted by H80R on the NTD (Figure 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 SI) 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 ~0.6 $mg \cdot mL^{-1}$ and the absorption spectra registered in a HP 8453 spectrophotomoter (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: $\varepsilon_{280} = 47900 \cdot (1 + \frac{22000}{11300} \cdot A_{450})$, that considers the contribution from apo-NQO1 (ε_{280}= 47900 $M^{-1} \cdot cm^{-1}$, 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 (ε_{280}= 22000 and ε_{450}= 11300 $M^{-1}\cdot cm^{-1}$; (43)). This procedure yields estimates for the FAD content in purified samples by using a reference value of ε_{450}= 11300 $M^{-1}\cdot cm^{-1}$ 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 $0.5^{o}C\cdot min^{-1}$. 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 $2^{o}C\cdot min^{-1}$ 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 (Figure S5).

The interaction of NQO1 and SAMp73α was investigated by NMR spectroscopy. NMR data were acquired at 20 ºC on a Bruker Avance DRX-500 spectrometer equipped with a triple-resonance probe and z-gradients. Samples

containing NQO1 proteins (at ~300 μM, in monomer units) and ^{15}N-labelled SAMp73α (~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 analyzed 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 hours 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 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 provide 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 $U \cdot mL^{-1}$ penicillin and 100 $\mu g \cdot mL^{-1}$ 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 $\mu g \cdot mL^{-1}$ 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 analyzed 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^{o}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 $2 \cdot 10^4$ 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 $mg \cdot mL^{-1}$ of Alamar Blue solution (Sigma-Aldrich) was added to each well and incubated for 12 hours 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 ^oC as described in (49).

Acknowledgements

ALP thanks Prof. Jose Manuel Sanchez-Ruiz for support. JLN thanks C. Arrowsmith for the kind gift of the SAMp73 vector. We acknowledge the ESRF for provision of synchrotron radiation time at beam lines ID29 and ID30A-1, and the staff for their helpful support. This work was supported by the Spanish Ministry of Economy and Competitiveness (BIO 2015 66426-R to JMSR, CTQ 2015-64445-R to JLN, "Factoría Española de Cristalización", Consolider-Ingenio 2010 to JAG and SAF2015-69796 to ES), Junta de Andalucia (P11-CTS-07187 to ALP) and FEDER funds. EMC acknowledges a pre-doctoral fellowship from Junta de Andalucia.

Conflict of Interest

The authors declare no conflict of interest.

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, LLC. (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.

Legends to Figures

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 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 (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).

Figure 2. Overall conservation of the stabilizing effect of Arg80 across species and its divergence along primate evolution. A and 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 $kcal\cdot mol^{-1}$) 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 and F)

Conservation of the EER cluster among representative individual mammalian sequences (panel E) and a set of fifty-three mammalian sequences (see SI text) 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.

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 and 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 SI text); 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 aproximate times at which mutations have occurred according to ancestral sequence reconstruction.

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 5-6 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

and 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 and 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 Figure 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 and 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.

Abbreviations

AGT.- alanine:glyoxylate aminotransferase; CTD.- C-terminal domain; DCPIP.- 2,6-Dichlorophenolindophenol; FAD.- flavin-adenin dinucleotide; HEPES.- 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; MD.- molecular dynamics; NMR.- nuclear magnetic resonance; NQO1.- NAD(P)H:quinone oxidoreductase 1; NTD.- N-terminal domain.

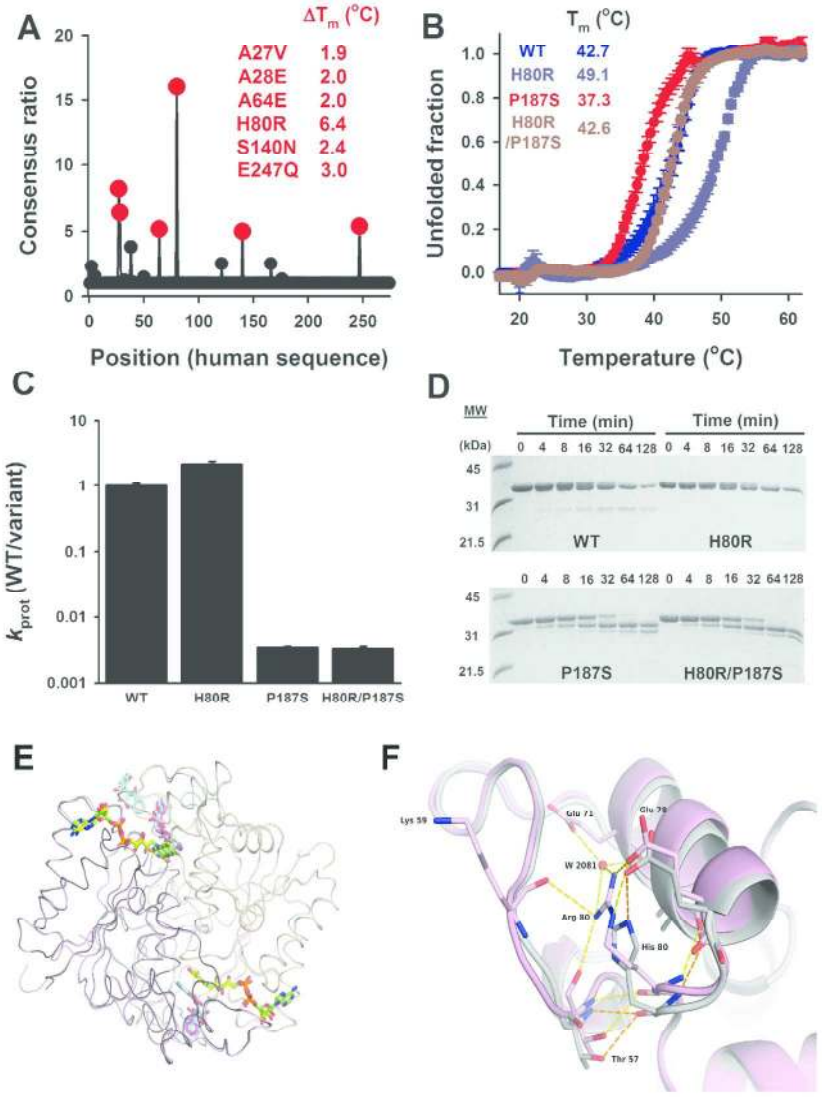

Figure 1

141x184mm (300 x 300 DPI)

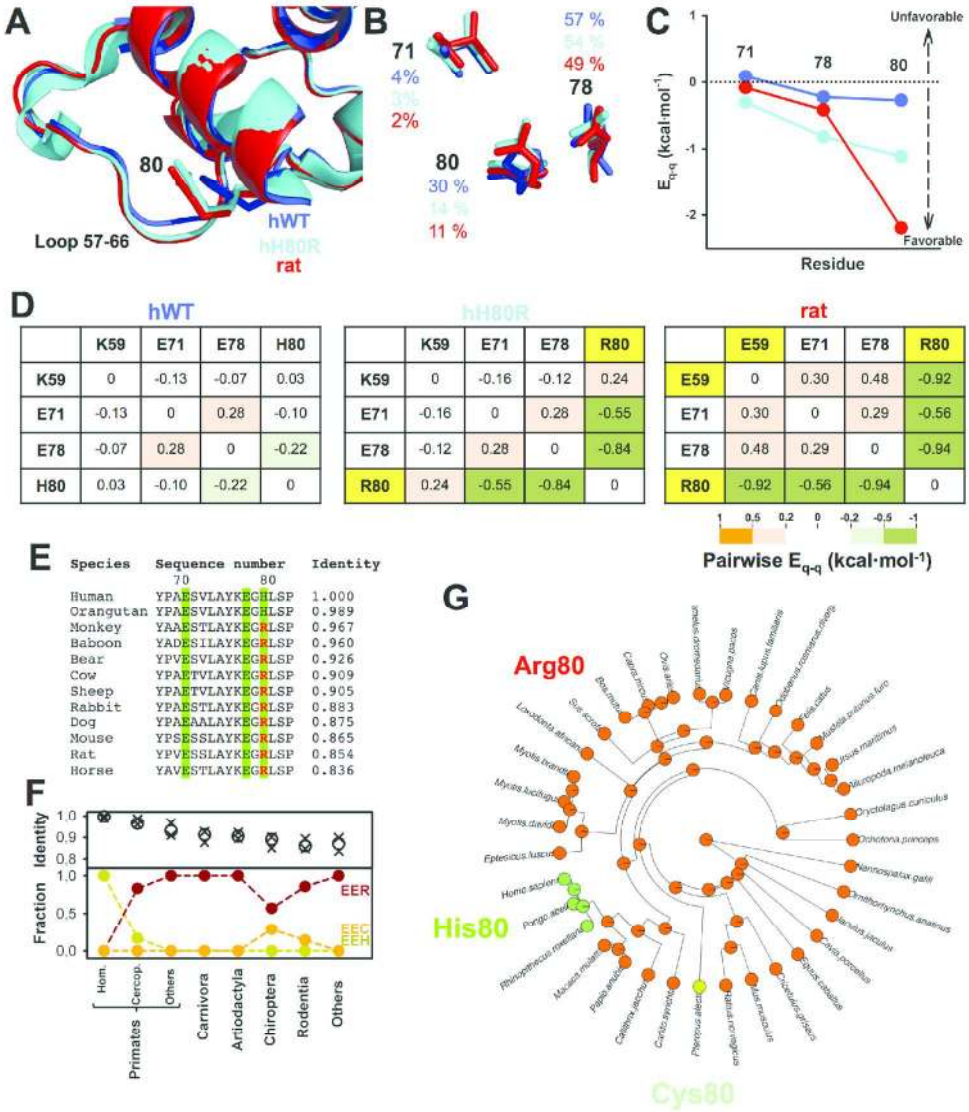

Figure 2

153x180mm (300 x 300 DPI)

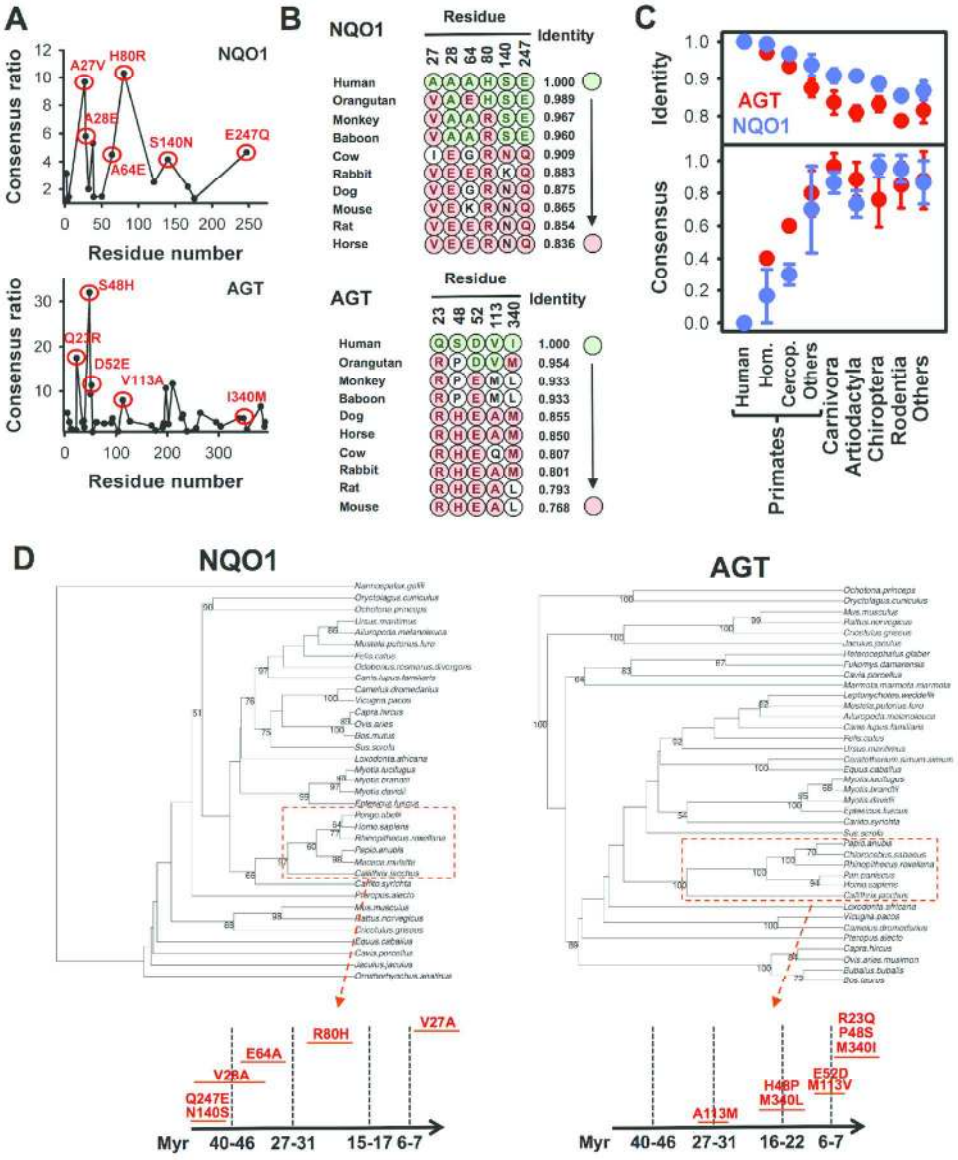

Figure 3

148x180mm (300 x 300 DPI)

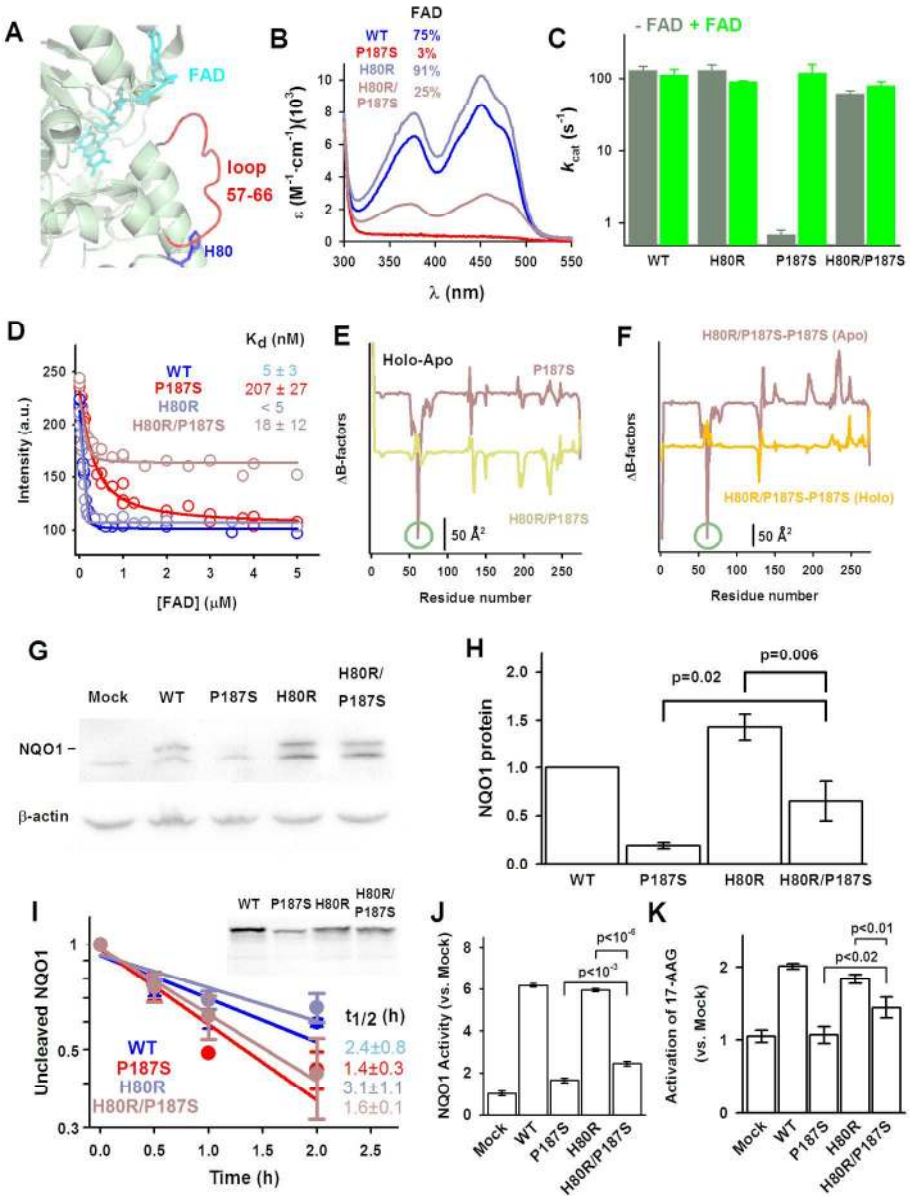

Figure 4

308x399mm (150 x 150 DPI)

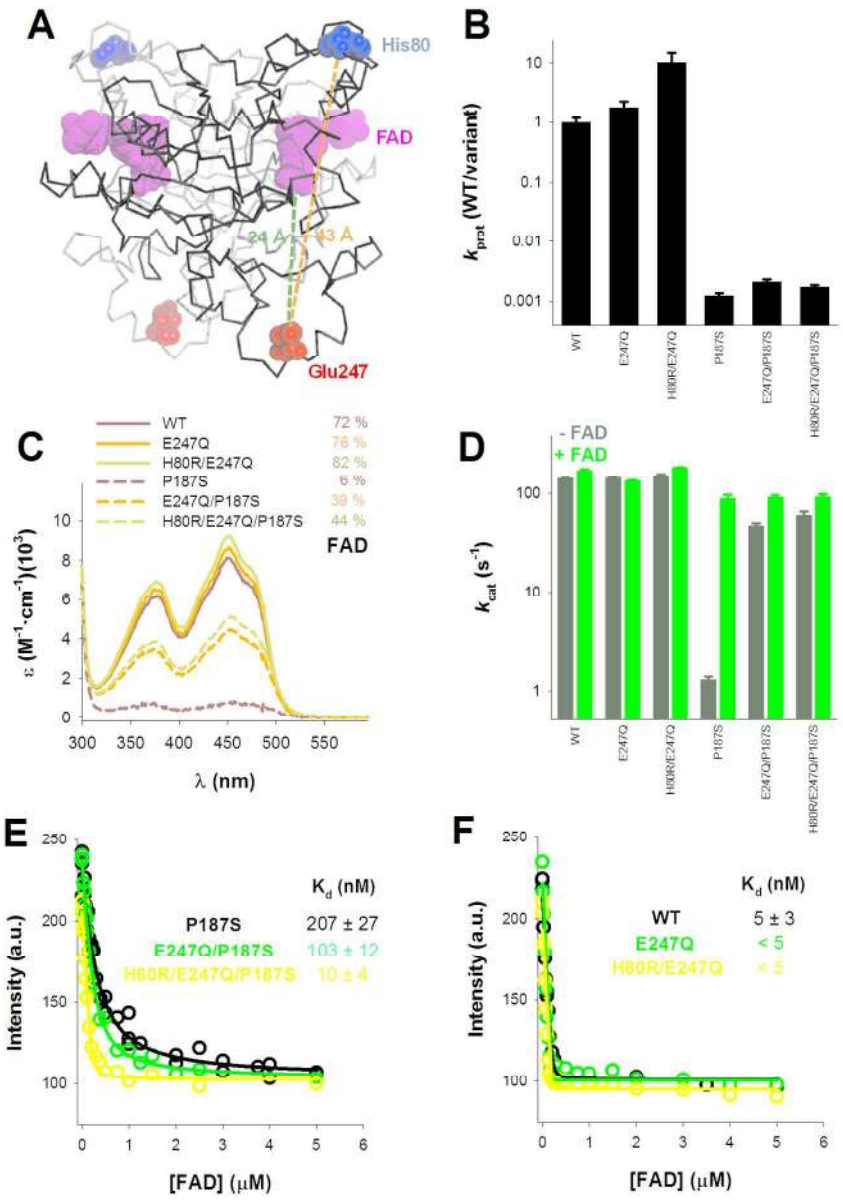

Figure 5

285x399mm (150 x 150 DPI)