Human Molecular Genetics, 2019, Vol. 28, No. 1 1–15

doi: 10.1093/hmg/ddy323
Advance Access Publication Date: 12 September 2018
General Article

GENERAL ARTICLE

Insight into the specificity and severity of pathogenic mechanisms associated with missense mutations through experimental and structural perturbation analyses

Encarnación Medina-Carmona[1,2,‡], Isabel Betancor-Fernández[3,‡], Jaime Santos[4,‡], Noel Mesa-Torres[1], Silvia Grottelli[2], Cristina Batlle[4], Athi N. Naganathan[5], Elisa Oppici[6], Barbara Cellini[2], Salvador Ventura[4], Eduardo Salido[3], Angel L. Pey[†,1,*]

[1]Department of Physical Chemistry, University of Granada, Granada 18071, Spain, [2]Department of Experimental Medicine, University of Perugia, Piazzale Gambuli 1, Perugia 06132, [3]Centre for Biomedical Research on Rare Diseases, Hospital Universitario de Canarias, Tenerife, 38320, Spain, [4]Institut de Biotecnologia i de Biomedicina and Departament de Bioquímica i Biologia Molecular, Universitat Autónoma de Barcelona, Bellaterra 08193, Spain, [5]Department of Biotechnology, Bhupat & Jyoti Mehta School of Biosciences, Indian Institute of Technology Madras (IITM), Chennai 600036, India and [6]Department of Neurosciences, Biomedicine and Movement Sciences, Section of Biological Chemistry, University of Verona, Strada Le Grazie 8, Verona 37134, Italy

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

Abstract

Most pathogenic missense mutations cause specific molecular phenotypes through protein destabilization. However, how protein destabilization is manifested as a given molecular phenotype is not well understood. We develop here a structural and energetic approach to describe mutational effects on specific traits such as function, regulation, stability, subcellular targeting or aggregation propensity. This approach is tested using large-scale experimental and structural perturbation analyses in over thirty mutations in three different proteins (cancer-associated NQO1, transthyretin related with amyloidosis and AGT linked to primary hyperoxaluria type I) and comprising five very common pathogenic mechanisms (loss-of-function and gain-of-toxic function aggregation, enzyme inactivation, protein mistargeting and accelerated degradation). Our results revealed that the magnitude of destabilizing effects and, particularly, their propagation through the structure to promote disease-associated conformational states largely determine the severity and molecular mechanisms of disease-associated missense mutations. Modulation of the structural perturbation at a mutated site is also

[†]Angel L. Pey, http://orcid.org/0000–0001–7706-3243
[‡]Authors contributing equally.

Received: August 14, 2018. **Revised:** September 7, 2018. **Accepted:** September 9, 2018

shown to cause switches between different molecular phenotypes. When very common disease-associated missense mutations were investigated, we also found that they were not among the most deleterious possible missense mutations at those sites, and required additional contributions from codon bias and effects of CpG sites to explain their high frequency in patients. Our work sheds light on the molecular basis of pathogenic mechanisms and genotype–phenotype relationships, with implications for discriminating between pathogenic and neutral changes within human genome variability from whole genome sequencing studies.

Introduction

One of the bottlenecks in the whole genome sequencing era is the interpretation of the genetic variability in the human genome (1–3). High-throughput discrimination between pathogenic and neutral missense variants often relies in bioinformatic tools that do not explicitly account for the structural context of mutations (3–6). Consequently, computational biophysical studies show stronger predictive power particularly for some pathogenic traits such as intracellular stability (7,8). The structural destabilization required to cause loss-of-function due to protein aggregation, degradation and subcellular traffic defects seems to be quite small (a few kcal·mol^{-1}) (7–10), possibly reflecting the marginal thermodynamic stability of natural proteins *in vitro* (11) and *in cellulo* (8). In addition, the structural and energetic optimization of natural proteins explains the universal destabilizing effect of random missense mutations thus compromising protein function *in vivo* (11,12). Noteworthy, missense mutations along the entire protein structure, even those located far from catalytic centers and regulatory binding sites, often cause significant functional and regulatory perturbations likely through protein destabilization (13–24), supporting the notion that mutational effects propagate efficiently into the structure affecting distant functional sites (25,26). This propagation could be the basis of the multiple deleterious effects of artificial (human-made) and disease-associated missense mutations on different functional traits (*pleiotropy*) (16–23,27).

In contrast to the apparently universal effect of missense mutations on the stability and functional traits of proteins, small sets of missense mutations causing folding, misfolding and inactivation of human proteins constitute a remarkably large fraction of disease-causing alleles [see Supplementary Material, Table S1 for an illustrative set of diseases and (28)]. This implies the existence of some specific molecular mechanisms strongly *favoring* a frequent disease-associated mutation (amino acid X to Y) versus *random* mutations (X to Z, being Z any proteinogenic aminoacid $\neq$ Y), concomitantly shaping mutational disease spectra. Several mechanisms could be involved in this phenomenon, although their actual contribution is not well known (17,28). One class of mechanisms involves increased nucleotide mutation rates, and depends on the intrinsic stability and mutability of nucleotides (e.g. CpG sites) and DNA replication fidelity and repair efficiency (28,29). Another class involves alterations in protein structure, energetics and ultimately in function (17,24).

The effect of disease-associated mutations on organismal fitness (as the *reproductive success of genotypes*), pathogenicity (as the *severity of clinical manifestations*) and protein molecular properties (as the *perturbation of protein functionality in vivo*) are likely correlated, although the relationships between these effects can be complex and not universal in nature (17,28). It is plausible that the intrinsic mutability of those codons leading to frequent pathogenic mutations, combined with their specific effects at the molecular and biochemical levels, could shape, to a large extent, the mutational spectra of disease-associated alleles, explaining their high frequency (28). Importantly, the mechanisms by which a given amino acid substitution becomes a frequent mutation in terms of protein fold, mutated site and disease mechanism have been rarely addressed (8,17,30). Therefore, comprehensive computational, structural and experimental analyses on a range of specific diseases and pathogenic functional perturbations would certainly improve our understanding of the mechanisms shaping mutational spectra in human diseases (28). This in turn could pave the way to identify and optimize small molecule therapies to treat human genetic diseases (31) and to improve our predictive power on the pathogenicity of missense variants found in whole genome sequencing studies (1,2,8).

Mutational effects on protein stability are linked to pathogenic mechanisms (7,8,17,24,32,33), although how structural destabilization triggers a specific disease-associated mechanism is not well understood. In this work, we develop a proposal describing the relationships between structural and energetic contexts of mutations and their effects on relevant disease-associated conformational states leading to specific pathogenic mechanisms (sketched in Fig. 1). As we show in this work, this approach consistently explains experimental and computational perturbation analyses obtained for a large set of missense variants in three different disease-associated proteins displaying five of the most common pathogenic mechanisms: loss-of-function and gain-of-toxic function aggregation, enzyme inactivation, protein mistargeting and accelerated degradation.

We depicted relevant conformational states for a protein potentially associated with genetic disease in the context of a simple free energy profile (Fig. 1). For the sake of simplicity, we focused in the denaturation/unfolding process. This model clearly was an oversimplification (33–35) since it only considered a few relevant states along a single structural coordinate (versus multidimensional landscape representations) and it focused in the denaturation/unfolding process (note that the folding process is intrinsically more complex and interactive with the cellular protein folding machinery). However, its simplicity allowed it to display some key features of the relationships between mutation-induced protein destabilization and different pathogenic mechanisms investigated in this work. The native state (N) contained two domains (α and α/β) and existed in *equilibrium* with a partially unfolded state (M), which was assumed to be disease-relevant (i.e. for the protein to aggregate or mistarget, to be targeted for proteasomal degradation or represented an inactive conformation). The population of the M and N states was determined by the free energy difference between them (ΔG_{mis}), while the rate of conversion of N into M was determined by the activation barrier ($\Delta G^{\neq}$) with the TS (*transition state*). Thus, the value of ΔG_{mis} and $\Delta G^{\neq}$ is associated with the propensity to develop a given disease phenotype (35). The thermodynamic stability of the N state against global unfolding was determined by the energy difference with a *globally unfolded* state U (ΔG_{unf}). Note that mutational effects

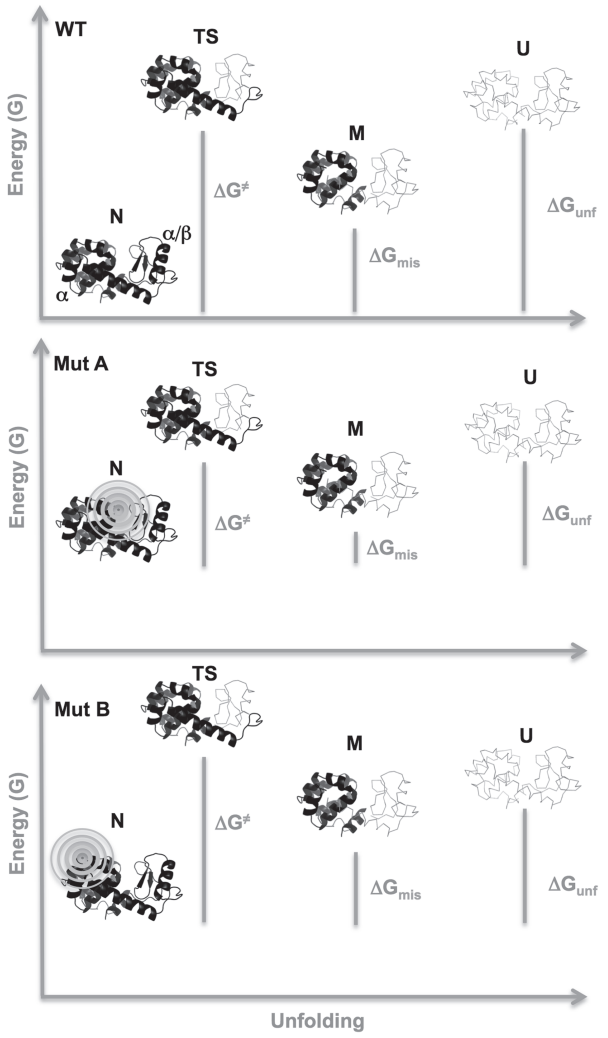

Figure 1. Stability (free energy; y-axis) profiles depicting the roles of destabilizing mutations in relevant states for pathogenic mechanisms as a function of the degree of unfolding (x-axis). Regions of the protein which were unstructured in non-native states (TS, M and U) were indicated as thin gray ribbons. Using the energy/folding landscape of the WT protein as a reference, we depicted two destabilizing mutations (Mut A and B) that locally affected different structural regions of the native state (N), thus decreasing the thermodynamic stability of N (ΔG_{unf}). However, propagation of the destabilizing effect of Mut A also reduced the thermodynamic (ΔG_{mis}) and kinetic ($\Delta G^{\neq}$) stabilities of N toward the partially unfolded state M by destabilizing regions structured in N and unstructured in the TS and M, thus becoming disease-causing. In contrast, Mut B did not destabilize N toward M because its effects did not propagate to unstructured regions in the TS and M, and thus, Mut A is not expected to be pathogenic. Concentric dark gray circles represented propagation of destabilizing effects from mutated sites.

on this variable can be estimated for instance by FoldX (or other suitable structure-based algorithms) reasonably assuming that destabilization at the mutated site fully translates into the thermodynamic stability (i.e. the unfolded state is *globally unfolded*) (36). In addition, since this type of calculation mostly considers the local effects on stability due to a mutation, we would expect a rough correlation between mutational effects on thermodynamic stability (referred to as $\Delta\Delta G$ in Figs 2–4) and the local destabilization caused around the mutated site. If we introduced two *similarly* destabilizing mutations (in terms of ΔG_{unf}), named mut A and B, which were located in structured regions of N, TS and M states, these two mutations should mainly affect ΔG_{unf} if their effects were local, and thus, they were not disease-associated, although FoldX analyses might reveal their destabilizing effects. A different scenario would be found if (i) mutations were located in regions unstructured in the disease-relevant TS or M states, which would promote pathogenic effects by decreasing ΔG_{mis} and/or $\Delta G^{\neq}$; (ii) propagation of the destabilizing effect [which dissipates exponentially through the structure with a half-dissipation radius of 5–10 Å (37) and Supplementary Material, Fig. S1] of mut A could affect unstructured regions in the TS and M states, leading to disease by reducing ΔG_{mis} and/or $\Delta G^{\neq}$ (Fig. 1). However, due to its different structural location, dissipation of destabilizing effects of mut B would not reach disease-relevant locations and thus would not be disease-causing (Fig. 1). In this context, modulation of the destabilizing effect at a given site (as we experimentally performed in this work) should reveal how destabilization of the native state ($\Delta\Delta G$) translates into changes in disease-relevant stabilities (i.e. in ΔG_{mis} and $\Delta G^{\neq}$) and gradually leads to different disease phenotypes. Additionally, information could be obtained regarding states relevant for different mechanisms (note that TS and M states may be structurally different for aggregation, mistargeting or degradation phenotypes for a given protein and disease, even if the structure of these states is preserved). Importantly, the effect of mutations in this context could not be simply derived from structural analyses of N, since the structure of TS or M was not known a priori, thus precluding a complete analysis of mutational effects on disease-relevant conformational states and mechanisms.

Results and Discussion

Structural and energetic determinants of pathogenic traits due to missense mutations

To investigate the link between structural and energetic perturbations due to missense mutations and particular pathogenic phenotypes according to the general scenario depicted in Figure 1, we selected three disease-associated proteins displaying a variety of common molecular pathogenic mechanisms. We focused on the underlying mechanisms by which mutational effects on the native structure could propagate beyond the mutated site, thus contributing to our understanding on how a single mutation might lead to more than one phenotypic trait and to what extent it happens. On these systems, we also introduced sets of random mutations by one to three nucleotide changes, named as *unnatural* mutations (since they are extremely rare in patients or global population; Table 1 and Supplementary Material, Tables S1 and S2) to vary the extent of structural and energetic perturbation at these sites. On the one hand, we evaluated the magnitude of the destabilization upon mutation by FoldX calculations (36,38) ($\Delta\Delta G$ in Figs 2–4) and carried out structural perturbation analysis (SPA). As mentioned above, FoldX calculations were also considered to provide estimates of the magnitude of mutational effects around the mutated site. In addition, SPA is a structure-based computational approach suitable for investigating the propagation of mutational effects from the mutated site (25,26,37). The reliability of SPA results depends on the quality of the structure used as template and thus, it is sensitive to small errors and uncertainties in these structural models. In addition, it implicitly considers that mutations do not largely affect protein structure, which is indeed true

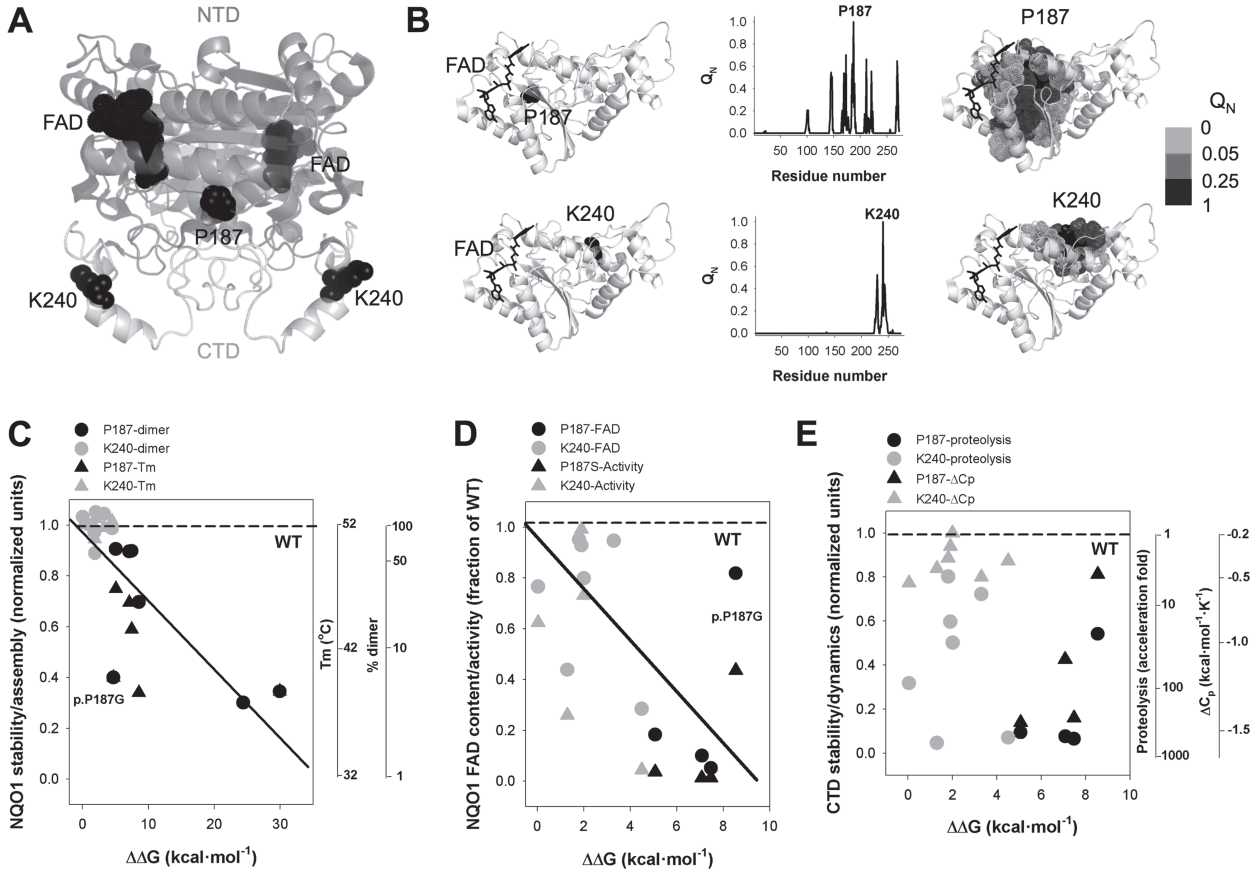

Figure 2. Perturbation analyses on cancer-associated NADP(H):quinone oxidoreductase 1 (NQO1). **(A)** Structure of NQO1 (2F1O) showing the N-(NTD) and C-terminal (CTD) domains, and the location of the FAD, P187 and K240 sites. **(B)** SPA analyses for P187 and K240 sites; Figures in the left showed the structural location of the sites in the NQO1 monomer. In the middle-right panels, SPA analyses were shown on the protein sequence and structure, respectively. Color scales indicated the extent of the propagation per residue, Q_N; **(C–E)** Relationships between local destabilization (ΔΔG, as determined by FoldX) and different molecular alterations: C, dimer assembly and stability; D, dimer functionality (FAD content and activity of purified proteins); E, stability/dynamics of the CTD. The x-axis represented the stability effects versus WT. The y-axis represented the extent of pathogenic effect characterized experimentally upon normalization to provide an operational scale related to the mutational effects on disease-relevant stabilities ($\Delta G^{\neq}$ and ΔG_{mis} in Fig. 1). Normalized scales in the left side provided values of 1 for the reference, WT (horizontal dashed lines) and values of 0 for extremely deleterious effects. Absolute scales for the different features can be found in the right side of y-axis. Thick solid lines are meant to guide the eye. Some data used in panels C–E were retrieved from (27) with permission.

for some of the natural mutations investigated here such as p.P187S in NADP(H):quinone oxidoreductase 1 (NQO1) and p.G170R in alanine:glyoxylate aminotransferase (AGT) (39–42). SPA was therefore used to investigate the propagation of mutational effects from the mutated site to functional regions of the protein *in silico*. Importantly, SPA was recently shown to reproduce the results of structural perturbations due to mutations, phosphorylation, protein–protein and protein–ligand interactions determined experimentally (37). On the other hand, the effects of these mutations on functional and stability traits were experimentally characterized to search for correlations with the results from structure-based perturbation analyses.

Cancer-associated NQO1: enzyme inactivation and accelerated proteasomal degradation

Human NADP(H):quinone oxidoreductase 1 (NQO1; GenBank ID: M81600) is a dimeric FAD-dependent enzyme catalyzing two-electron reduction of quinones, involved in activation of certain cancer pro-drugs and stabilization of cancer-associated transcription factors by physical interaction (43). The common polymorphism p.P187S (c.559C > T) is associated with cancer development through *in vivo* enzyme inactivation and enhanced proteasomal degradation of the enzyme by destabilization of the C-terminal domain (CTD) (43–46). P187 is fully buried in the structure and located far from the FAD binding site (Fig. 2A). The crystal structure of p.P187S in the holo-state (i.e. FAD-bound), and stabilized with the CTD binding inhibitor dicoumarol, has shown subtle local structural rearrangements (41). p.P187S affects different functional sites such as the FAD binding site and the CTD through long-range communication of structural and dynamic perturbations (45,47). Accordingly, SPA corfirmed that p.P187S affects the stability and packing of these two regions (Fig. 2B and Supplementary Material, Fig. S1). In contrast, the naturally occurring and rare cancer-associated p.K240Q (c.718A > C) somatic mutation (http://cancer.sanger.ac.uk/cosmic), placed in the CTD at a solvent-exposed location (Fig. 2A), locally perturbed the CTD (Fig. 2B and Supplementary Material, Fig. S1).

We have recently carried out an extensive biophysical characterization of six unnatural mutations at either P187 or K240 sites, as well as the cancer-associated p.P187S and p.K240Q mutations (Table 1), to determine whether modulation of local destabilization could affect different functional sites and to what extent (27). We focused on the propagation of mutational effects

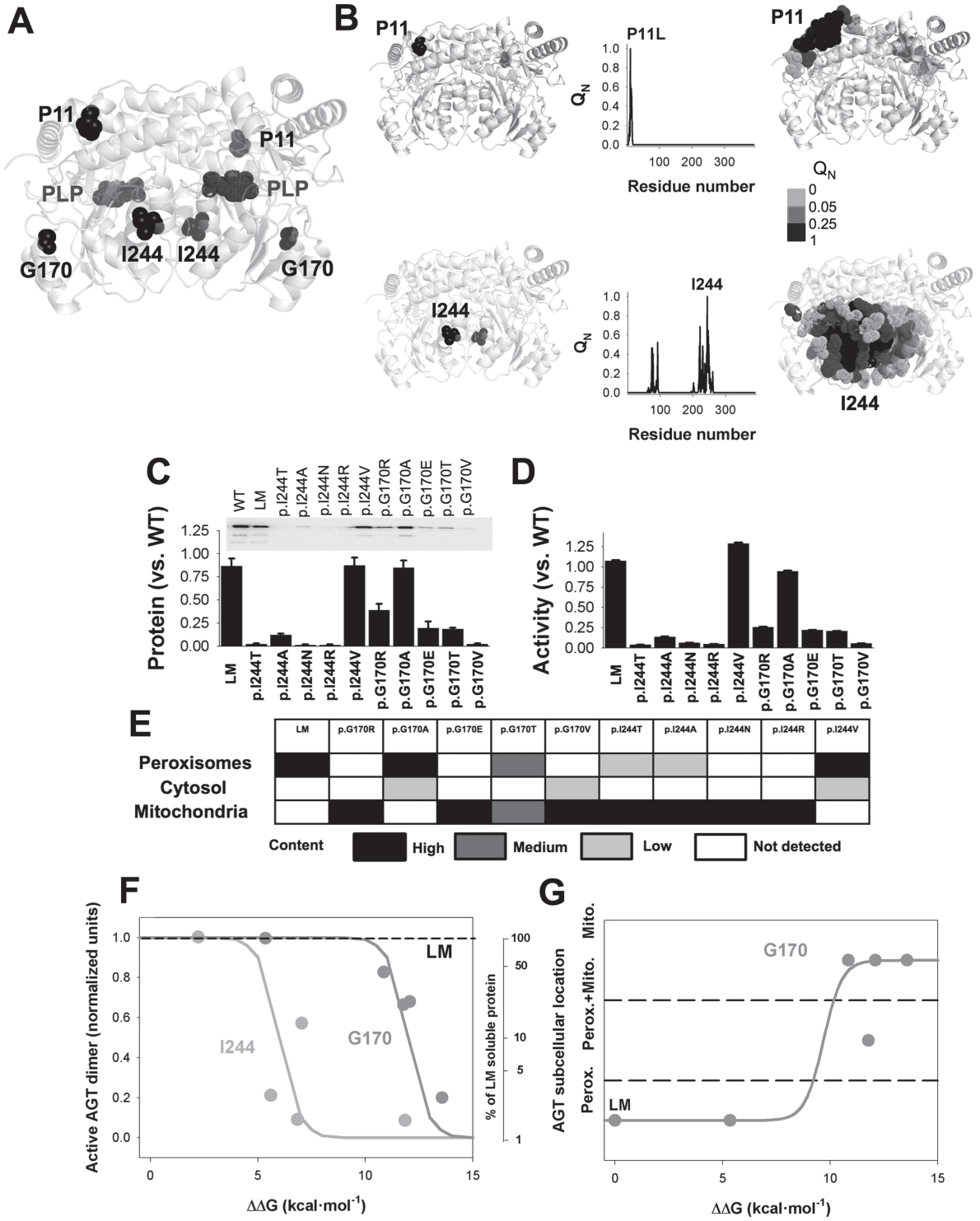

Figure 3. Perturbation analyses on AGT associated with aggregation and mistargeting in PH1. (**A**) Structure of AGT showing the location of the PLP molecule and the P11, G170 and I244 sites. (**B**) SPA analyses for P11 and I244 sites; Figures in the left showed the structural location of the sites in the AGT dimer. In the middle-left panels, SPA analyses were shown on the protein sequence and structure, respectively. Note that only one of the P11 and I244 residues in the dimer is oriented toward the reader. Color scales indicated the extent of the propagation per residue, Q_N; (**C** and **D**) Soluble AGT protein levels (C) and activity (D) upon expression in CHO-GO cells. Data were means ± S.D. from at least three independent measurements (normalized versus WT values). (**E**) Subcellular location of mutants at G170 and I244 sites. (**F** and **G**) Relationships between local destabilization (ΔΔG, as determined by FoldX) and protein yield as functional dimers (F) and subcellular location (G). The x-axis represented the local stability effects versus LM (*minor allele*). The y-axis represented the extent of pathogenic effect characterized experimentally upon normalization to provide an operational scale related to the mutational effects on disease-relevant stabilities ($\Delta G^{\neq}$ and ΔG_{mis} in Fig. 1). In panel F, normalized scales in the left side provided values of 1 for the reference, LM (horizontal dashed lines) and values of 0 for extremely deleterious effects. Absolute scales for the different features can be found in the right side of y-axis of panel F. Thick solid lines are meant to guide the eye.

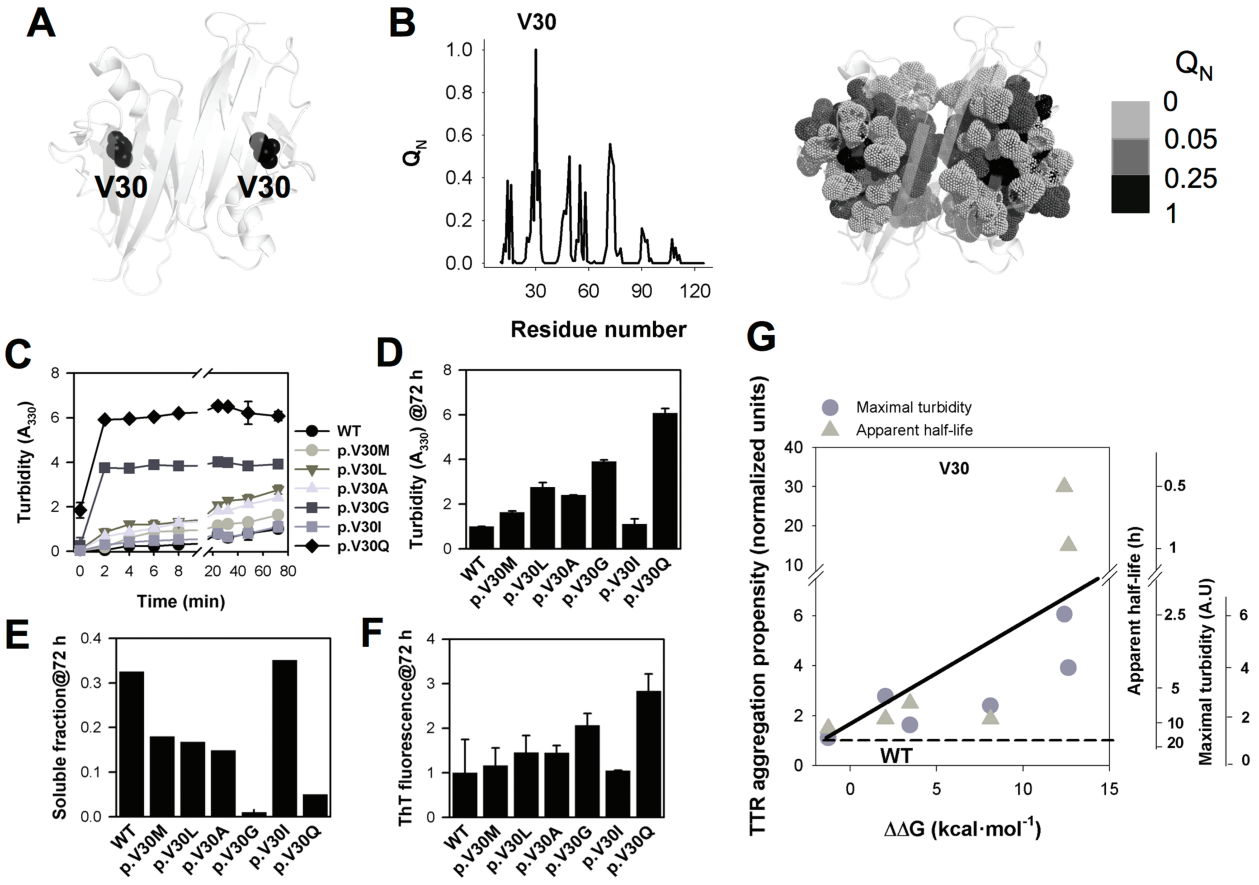

Figure 4. Perturbation analyses on amyloid-forming TTR. **(A)** Structure of TTR showing the location of the V30 site. **(B)** SPA analyses at V30 site. SPA analyses were shown on the protein sequence (left) and structure (right). Color scales indicated the extent of the propagation per residue, Q_N; **(C–F)** Aggregation of TTR variants at site V30. **(C)** Kinetics of aggregation monitored by turbidity measurements at 330 nm; (D–F) Aggregation (turbidity; D), fraction of soluble protein (E) and ThT fluorescence (F) after 72 h incubation. Data in panels C, D and F were mean ± s.d. from three replicates. **(G)** Relationships between local destabilization (ΔΔG, as determined by FoldX) and aggregation propensity. The x-axis represented the local stability effects versus WT. The y-axis represented the extent and kinetics of aggregation upon normalization to provide an operational scale related to the mutational effects on disease-relevant stabilities ($\Delta G^{\neq}$ and ΔG_{mis} in Fig. 1). In panel G, normalized scales in the left side of provided values of 1 for the reference, WT (horizontal dashed lines) and values larger than 1 for deleterious effects. Absolute scales for the different features can be found in the right side of y-axis of panel F. The thick solid line is meant to guide the eye.

to the monomer–monomer interface, affecting protein folding and dimer stability, the FAD binding site and the CTD (including the dicoumarol binding site). Thus, these analyses evaluated how these mutations affected the ability of the protein to fold, its stability and function. From these results, and according to the proposal depicted in Figure 1, we can gain insight into how destabilizing effects might partition between these different mechanisms (i.e. how changes in ΔG_{unf} translated into those in disease relevant ΔG_{mis} and $\Delta G^{\neq}$). We found simple relationships between ΔG_{unf} and the effects on dimer stability or activity (Fig. 2C and D and Supplementary Material, Fig. S2). Indeed, the thermal stability (T_m values), FAD content and specific activity of mutants at P187 and K240 sites decreased almost linearly with destabilization provided by FoldX, while the amount of stable dimer formed upon expression in *Escherichia coli* followed an exponential dependence on ΔΔG (Fig. 2C and D), with the exception of p.P187G which may cause increased conformational dynamics in addition to structural perturbations (27). The energetic graduation between conformational stability and activity indicated that a moderate destabilizing effect, even at a site away from the active site, propagated efficiently to the active site causing inactivation, while a larger structural destabilization severely compromised intracellular folding and stability as the main pathogenic trait. A similar interpretation could explain some degree of correlation observed between mild to moderate protein destabilization and catalytic perturbations due to recurrent mutations in other genetic diseases (17,22,24,30,48–50).

Another key pathogenic mechanism in loss-of-function diseases is accelerated proteasomal degradation of native or misfolded proteins (7,8,33,51). Proteasomal degradation is often determined by the presence of loosely folded and flexible regions in proteins (52–54). In NQO1, the stability and dynamics of the CTD are major determinants of accelerated degradation of p.P187S (associated with a reduction in the ΔG_{mis} and $\Delta G^{\neq}$ for degradation). Mutations at the P187 site affected similarly the stability and dynamics of the CTD (i.e. dicoumarol binding heat capacity, ΔC_p, and proteolysis rates) and ΔG_{unf} (Fig. 2E), suggesting that the larger the destabilizing effect at this site, the faster the intracellular degradation would be. However, mutations at the K240 site only affected CTD dynamics (i.e. proteolysis rates; Fig. 2E), and their effects on proteasomal degradation rates are expected to be milder that those of P187

Table 1. Disease-associated and unnatural mutations experimentally investigated by perturbation analyses

Protein	WT codon (amino acid)	Mutant codon (amino acid)[a]	Nature of the mutation	Minimal nucleotide substitutions
NQO1[b]	CCT (P187)	**T**CT (S187)	Frequent disease-associated	1
		GCT (A187)	Unnatural	1
		C**G**T (R187)	Unnatural	1
		ACT (T187)	Unnatural	1
		CT**T** (L187)	Unnatural	1
		GGT (G187)	Unnatural	2
		GAA (E187)	Unnatural	3
	AAA (K240)	**C**AA (Q240)	Rare disease-associated	1
		ATA (I240)	Unnatural	1
		GAA (E240)	Unnatural	1
		ACA (T240)	Unnatural	1
		CAC (H240)	Unnatural	2
		GCA (A240)	Unnatural	2
		GGA (G240)	Unnatural	2
AGT[c]	GGG (G170)	**A**GG (R170)	Frequent disease-associated	1
		GA**G** (E170)	Unnatural	1
		GT**G** (V170)	Unnatural	1
		GC**G** (A170)	Unnatural	1
		ACG (T170)	Unnatural	2
	ATC (I244)	**AC**C (T244)	Frequent disease-associated	1
		GTC (V244)	Unnatural	1
		AAC (N244)	Unnatural	1
		GCC (A244)	Unnatural	2
		CGC (R244)	Unnatural	2
TTR	GTG (V30)	**A**TG (M30)	Frequent disease-associated	1
		CTG (L30)	Rare disease-associated	1
		GGG (G30)	Rare disease-associated	1
		GC**G** (A30)	Rare disease-associated	1
		GA**G** (E30)	Unnatural	1
		CAG (Q30)	Unnatural	2
		ATA (I30)	Unnatural	2

[a]The nucleotides mutated are in bold.
[b]Characterized in (27).
[c]Mutations in AGT occur on the minor allele (i.e. in the presence of the P11L and I340M polymorphisms).

variants. Of course, while these relationships between local protein structure and dynamics and proteasomal degradation may hold for NQO1, they might not be general and possibly depend on specific structural and energetic features of different proteins (i.e. the properties of the N, M and TS states and how they are perturbed by mutations). Accordingly, good and simple overall correlations have been found between mutational effects on ΔG_{unf}, ΔG_{mis}, $\Delta G^{\neq}$ and proteasomal degradation rates in other genetic diseases (7,8).

AGT and primary hyperoxaluria type I: loss-of-function protein aggregation and subcellular mistargeting

Human alanine:glyoxylate amino transferase (AGT; GenBank ID: D13368) is a peroxisomal enzyme that forms active dimers (39). The N-terminal domain contains most of the active site residues, including the pyridoxal 5′-phosphate (PLP)-binding site and the dimerization interface, while the CTD is primarily involved in the peroxisomal targeting of the enzyme (39,55). AGT catalyzes the transamination of L-alanine and glyoxylate to form pyruvate and glycine, thus detoxifying the glyoxylate metabolite (56). Over 200 mutations in AGT have been identified in primary hyperoxaluria type I (PH1), a rare disease caused by deficient AGT activity (56,57). AGT exists in two polymorphic forms, the wild-type (WT; major allele) and the minor allele, which contains a largely destabilizing p.P11L polymorphism (c.32C > T) that predisposes toward disease-causing mutations (58), and a less relevant change p.I340M (c.1020A > G) (thus, the minor haplotype is also known and referred to as LM) (58). P11 is located at the N-terminal extended region far from the active site or dimerization interfaces (Fig. 3A). Most PH1-associated mutations cause either aggregation, particularly in the peroxisome or the cytosol (e.g. the frequent mutation p.I244T, c.731 T > C, in the LM, Supplementary Material, Table S1) or mistargeting to mitochondria, where the enzyme is metabolically useless (e.g. the frequent mutation p.G170R, c.508G > A, in the LM; Supplementary Material, Table S1). Both mechanisms are associated with low conformational stability and enhanced binding to molecular chaperones (10,58,59), which suggested that they may coexist in PH1 mutants (58–60). I244 and G170 are buried in the AGT structure (Fig. 3A), while p.G170R only caused minor changes in the structure (40). SPA analyses indicated that mutational effects at the I244 site would propagate to longer distances in the structure of AGT than those at the P11 site (Fig. 3B and Supplementary Material, Fig. S1; please note that these analyses cannot be performed at the G170 site).

We investigated two natural and eight unnatural mutations in AGT (Table 1). At the G170 site, we introduced mutations to A, V, T and E, of which those to V and T notably increased the

size of the side-chain, while that to E was *similar* to p.G170R (i.e. introduced an ionizable residue in the protein core). At the I244 site, we introduced mutations to V, N, A and R, of which that to A particularly decreased the size of the side-chain, that to R introduced a positive charge and that to N changed the polarity. We studied the folding, solubility, activity and subcellular localization of these AGT variants on the LM in transfected CHO-GO cells, a well-known model to study PH1 pathogenesis (58,60).

The mutations p.I244N and p.I244R had as dramatic effects on protein soluble levels and activity as those of p.I244T (at least 20-fold decrease), while p.I244A was milder and p.I244V had virtually no effect (Fig. 3C and D). Mutations at G170 were generally mild: p.G170R, p.G170E and p.G170 T decreased by 4-fold soluble protein and activity, while the most conservative p.G170A had virtually no effect and p.G170 V was the most deleterious (Fig. 3C and D). Regarding subcellular localization (Fig. 3E and Supplementary Material, Figs S3–S5), mutations at G170 showed a good correlation between soluble protein levels and subcellular location, with p.G170R, p.G170E and p.G170V being primarily mitochondrial, while p.G170T deviated from this trend since it was found in peroxisomes and mitochondria. Mutants at I244 were primarily found in mitochondria and to some extent in the cytosol or peroxisomes, although their main pathogenic effect seemed to be a decrease in folding efficiency (Fig. 3C and D). The more conservative variants p.G170A and p.I244V were primarily peroxisomal. Considering LM as a stability threshold for mutations to be crossed to affect proper folding and peroxisomal location (9,10), some plasticity was observed for these traits in the response to mutations, which was modulated by the structural location of the mutated site and the magnitude of the destabilizing effect. These results are consistent with the scenarios depicted in Figure 1.

Protein aggregation and trafficking defects can lead to loss-of-function diseases by simply decreasing the levels of active enzyme at the proper subcellular location (31,33). In PH1, mutation-induced destabilization of the native state may enhance aggregation and mitochondrial mistargeting as parallel and coexisting pathogenic mechanisms (10,58,59). There is some preference for some mutations to cause one or the other mechanism, with the recurrent mutations p.I244T and p.G170R (on LM) as the best examples. Destabilization by mutations at the I244 site (in terms of ΔG_{unf}) was more efficiently translated into intracellular aggregation than those of mutations at G170, and thus, into changes in ΔG_{mis} and $\Delta G^{\neq}$ for aggregation (Fig. 3F). Regarding mitochondrial mistargeting, changes in ΔG_{unf} of 10 kcal·mol^{-1} at the G170 site, which still allowed up to 50% of the protein to remain dimeric and active inside the cell, shifted the subcellular location of AGT from peroxisomal to mitochondrial (Fig. 3F and G). This behavior clearly supported a different response of intracellular aggregation and mistargeting traits due to destabilization at the G170 site (i.e. the destabilizing effect at this site propagates more efficiently into ΔG_{mis} and $\Delta G^{\neq}$ for mistargeting than for aggregation). Thus, the different capacity of mutational effects to be transmitted to relevant disease-associated states explained the different pathogenic mechanisms predominant in the frequent p.G170R and p.I244T mutations, the gradual shift in their pathogenic mechanisms found by fine-tuning the extent of native-state destabilization upon mutation (Figs 1 and 3F and G) and by ligands stabilizing the native state (61,62). To our knowledge, this has been the first structural and energetic insight into the mechanisms by which PH1-causing mutants preferentially led to intracellular aggregation or subcellular mistargeting.

Transthyretin and familial amyloidosis: gain-of-toxic-function aggregation

Human transthyretin (TTR; GenBank ID: M10605) is a paradigm for human degenerative diseases caused by protein misfolding or misassembly into structured protein aggregates (63). Structurally, TTR is a β-rich protein that normally assembles into tetramers, which are destabilized in aggregation-prone mutants associated with familial amyloidosis (63). p.V30M is the most prevalent disease mutation (64) while p.V30A, p.V30L and p.V30G are much less frequent (Table 1 and Supplementary Material, Table S1). p.V30A was characterized to be more aggressive than p.V30M (65), while the consequences of p.V30L are not well known (66). p.V30G is an extremely rare mutation leading to a very unstable protein (67). V30 is fully buried in the structure of TTR (Fig. 4A) and SPA analyses supported long-range communication of mutational effects at V30 site through the entire structure (Fig. 4B and Supplementary Material, Fig. S1).

We prepared four disease-associated (p.V30M, p.V30L, p.V30G and p.V30A) and three unnatural mutations (p.V30E, p.V30Q and p.V30I) (Table 1). Mutations to M, I and L were the most conservative, to G and A dramatically reduced the size of the side chain while mutations to E and Q changed the polarity or charge. p.V30E rendered very low expression levels as soluble protein precluding further experimental analysis and supporting a severe folding/stability effect.

Among disease-causing variants, p.V30M showed the slowest aggregation kinetics, which was faster in p.V30L and p.V30A, and remarkably accelerated by p.V30G. Among the unnatural mutations, aggregation kinetics of p.V30I was between those of WT and p.V30M, while p.V30Q aggregated extremely fast (Fig. 4C). The extent of aggregation showed the same trend found for aggregation kinetics (Fig. 4C and D) and mirrored the depletion of TTR tetramers (Fig. 4E). To identify the presence of amyloid species, we used Thioflavin T (Th-T) fluorescence and transmission electron microscopy (TEM). Results from Th-T fluorescence (Fig. 4F) correlated well with the kinetics and endpoint results from turbidimetry, while those from TEM showed that the amount of aggregates and their size agreed well with those obtained by other techniques (Supplementary Material, Fig. S6). Therefore, the chemical nature of the amino acid change at the V30 site largely determined the stability and aggregation propensity of TTR, revealing two unnatural mutations (p.V30E and p.V30Q) as the most aggressive.

Amyloidoses are associated with a wide variety of degenerative diseases, in which native state destabilization is one of the main features of disease-causing mutations that promote the population of aggregation-prone species (32,33,68). For TTR, the main determinant of toxic aggregation is the kinetic and thermodynamic destabilization of tetrameric TTR upon mutation, which increases the population of aggregation-prone monomers (32,69) consequently reducing relevant ΔG_{unf} , ΔG_{mis} and $\Delta G^{\neq}$ for aggregation (Fig. 1). Our results with mutants at the V30 site revealed a simple relationship between thermodynamic destabilization (effects on ΔG_{unf}) and aggregation propensity (effects on ΔG_{unf} and ΔG_{mis}) (Fig. 4G). An explanation for this behavior was that mutational effects at the V30 site propagated through the entire structure, affecting regions of the native state which were unstructured in the TS, M and U states. In accordance with this view, pharmacological chaperones binding to and stabilizing the tetramer (N state) prevented the aggregation of p.V30 M (70,71) just by increasing $\Delta G^{\neq}$, illustrating how long-range perturbations of the native tertiary and quaternary structure could be essential factors to understand and treat this pathology.

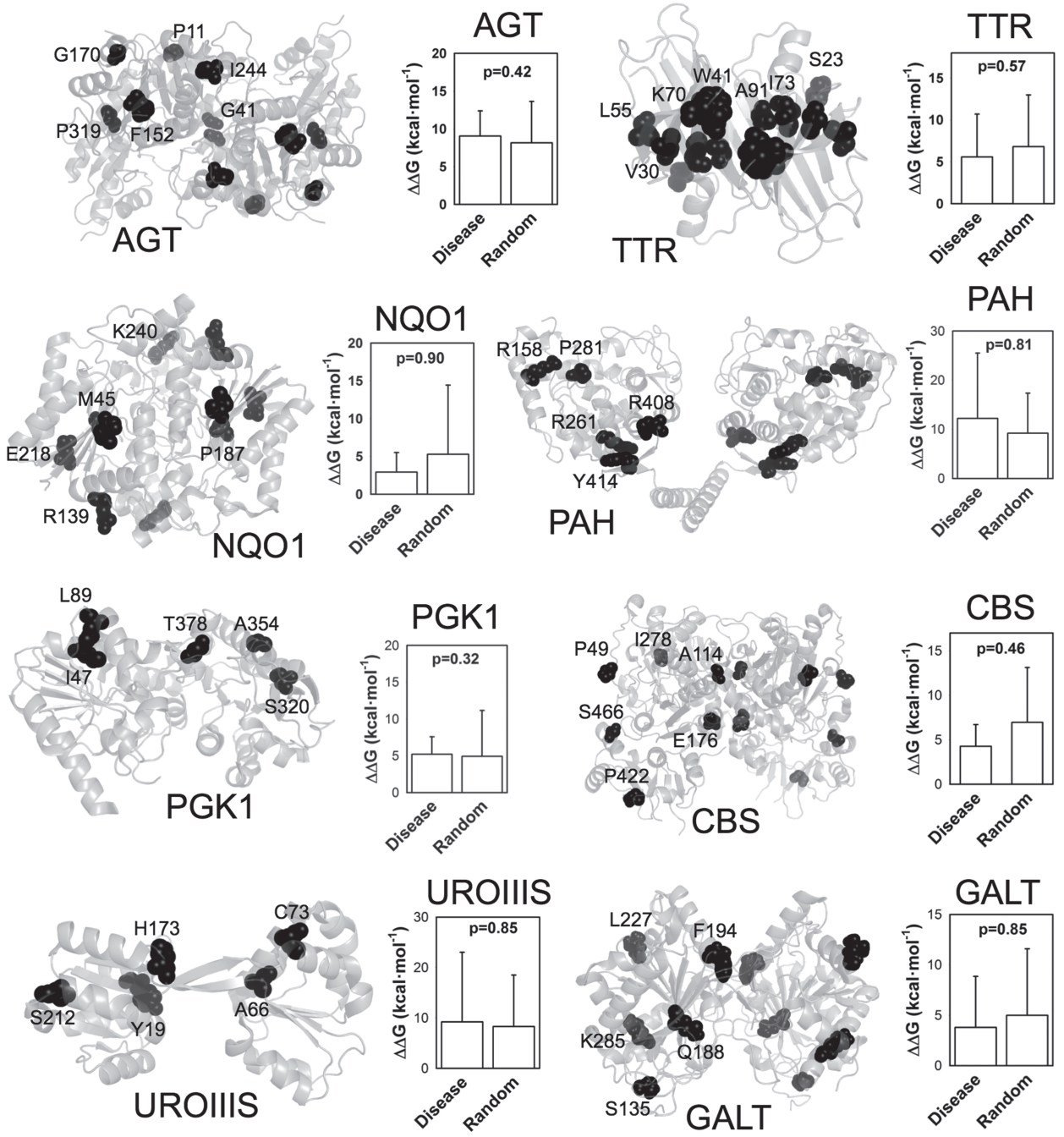

Figure 5. Native state destabilization by disease-associated alleles versus random mutations by FoldX. Disease-associated sites were shown as black sphere representation on the corresponding structure. Average destabilization (±s.d.) calculated for disease-associated alleles versus random mutations at these sites were shown in the plots. Statistical significance (*P*-values) was provided from a two-tailed Mann–Whitney test. AGT.- alanine:glyoxylate aminotransferase; TTR, transthyretin; NQO1, NADP(H):quinone oxidoreductase 1; PAH, phenylalanine hydroxylase; PGK1, phosphoglycerate kinase 1; CBS, cystathionine β-synthase; UROIIIS, uroporphyrinogen III synthase; GALT, galactose 1-phosphate uridylyltransferase.

On the mechanisms underlying frequent mutations in genetic diseases

The analyses described in this work showed that certain unnatural mutations at given disease-associated sites are more deleterious than frequent disease-causing mutations. This suggests that the predominance of certain frequent (recurrent) mutations in disease-associated mutational spectra (28) might not be due to their remarkably large destabilizing effect. To determine whether this could be a more general feature of frequent disease-associated mutations, we assessed the energetic impact of disease-associated versus *random* mutations on protein stability for 817 mutations in 43 disease-associated sites using computational tools (Fig. 5 and Supplementary Material, Fig. S7 and Table S1). These sites were found in eight proteins with different structural folds (all belonging to the

α/β class except TTR, which is an all β-protein) and forming in 75% of cases functional oligomers (dimers and/or tetramers) (Fig. 5). Sites were similarly distributed into different types of secondary structure elements (42% in helices, 26% in sheets and 32% in loops/coils; Supplementary Material, Table S1). Importantly, disease-associated mutations at these sites were associated with a variety of loss-of-function and gain-of-function mechanisms, in all cases linked to some extent with destabilization of protein structure and amenable to structure-based energetic analyses. In addition, the mutated sites fulfilled at least one of the following requirements (Supplementary Material, Table S1): (i) they were disease-associated through a highly predominant missense mutation (i.e. only p.C73R at the C73 site of UROIIIS); (ii) they were among the most common alleles associated with a given disease; (iii) they have been characterized experimentally (*in vitro*, *in cellulo* or *in vivo*) and shown to affect protein function, stability and/or misfolding tendency. Disease-associated missense mutations at these sites led to a wide range of changes in the chemical nature of the amino acid, affecting side chain polarity, charge and size and also displaying variable solvent exposure in protein structure (58% solvent exposed and 42% buried, using a 10% cut-off for accessibility; Supplementary Material, Table S1). Importantly, we found that the average destabilizing effect of disease-associated mutations did not significantly differ from those caused by random mutations at the same sites (Fig. 5). Accordingly, the effects of disease-associated mutations often clustered around the median destabilizing effect of all possible mutations at a given specific site, with only 28% above the median destabilizing effect (Supplementary Material, Fig. S7; considering an uncertainty of ∼1 kcal·mol^{-1} for FoldX analyses) (17). Therefore, disease-associated frequent mutations do not seem to cause a particularly large destabilization of the native fold.

To explain the high frequency of certain disease-associated mutations, in addition to sufficient energetic perturbation in an adequate conformational landscape, other genetic factors increasing mutation rates should also be considered. For instance, we may argue that disease-associated mutations require only a single nucleotide change (actually, all those experimentally analyzed in this work are due to single nucleotide substitutions; Table 1), although some of the most aggressive unnatural mutations can also be achieved through a single nucleotide substitution (p.P187R and p.P187L and p.K240E in NQO1; p.G170V in AGT; p.V30E in TTR; Table 1). In addition, 28% of these disease-associated frequent alleles compiled in Supplementary Material, Table S1 affected hypermutable CpG sites, which is close to the 30% estimated overall role of CpG sites in recurrent mutations (28,72–74). Moreover, frequent disease-causing alleles could also cause a codon bias leading to some *structural perturbations*, for instance by affecting native structures and cotranslational protein folding (75–77). Thus, recurrent mutations could alter protein folding/stability due to codon usage bias or tRNA abundance. Analyses of changes in codon usage (as a proxy for their effects on translation rates) (75) for this set of frequent mutations (Supplementary Material, Table S1), using statistics for codon frequency in human DNA (78), also revealed no significant bias (with average values of 1.8–1.9% for non-mutated and disease-associated codons).

Intriguingly, some unnatural mutations led to more severe phenotypes than those found in recurrent disease-associated mutations, thus raising the question about why these unnatural mutations are not frequent or present in patients. The practical absence of these unnatural mutations, neither identified in newborn screenings, patient-targeted or massive exome sequencing initiatives (see Supplementary Material, Tables S1 and S2) was unlikely associated with altered embryonic development or lethality due to extreme pathogenic effects (as inferred from phenotypic analyses of knock-out mice models) (79–82). Remarkably, these results supported that disease-asssociated frequent missense mutations might not be the absolute evil, they could just be deleterious enough, and in the *right* place and protein conformational ensemble, to eventually lead to a pathogenic state in combination with effects on other genetic factors.

Conclusions

We showed that disease-associated molecular mechanisms as well as the severity of phenotypes associated with missense mutations are linked to structural and energetic features of the mutated site and the energy landscape of the protein under study, in particular to the magnitude of the destabilizing effect at the mutated site and its propagation to relevant disease-associated structural spots. This provides a general framework to understand the molecular mechanisms of disease-associated frequent mutations, important to rationally develop mutation-specific therapies applicable to a large fraction of patients by targeting specific structural spots (31,83), and thus providing alternatives to the use of non-specific therapeutic agents such as proteasomal inhibitors (7,8). Overall, our work increases the understanding on genotype–phenotype relationships in disease from the molecular to the cellular levels, contributing to the translational development of pharmacological therapies for these diseases and the improvement of strategies to distinguish novel genetic variants found in whole-genome sequencing initiatives between potentially pathogenic and neutral.

Materials and Methods

Structure-energetics and perturbation analyses

Mutational effects on native state stability were evaluated using tridimensional structures and the FoldX energy field (36) as a plug-in of Yasara software (www.yasara.org/) (38). In this approach, the change in stability upon mutation was computed using an empirical effective energy function considering weighted contributions from van der Waals interactions, hydrogen bonds, electrostatics, solvation, side-chain and main-chain entropies (36). The following structures of human protein were retrieved from the Protein Data Bank database and used AGT (1H0C), NAD(P)H quinone oxidoreductase 1 (2F1O), transthyretin (5CN3), phenylalanine hydroxylase (2PAH), phosphoglycerate kinase 1 (2WZB), cystathionine β-synthase (1JBQ and 4COO), uroporphyrinogen III synthase (1JR2) and galactose 1-phosphate uridylyltransferase (5IN3). Upon energy-minimization of these WT structures, mutations were modelled allowing neighbouring residues to move, and their effect on unfolding free energy ($\Delta\Delta G$) was determined at 298.15 K, 100 mM ionic strength and pH 7 using an energy parameter *VdWDesign* of 2, which provides a ceiling of 5 kcal·mol^{-1} for the van der Waals clashes between two atoms. FoldX calculations were carried out for 53 disease-associated alleles in 43 positions and 764 random mutations at these positions (Supplementary Material, Fig. S7). Amino acid solvent accessibility was determined using the same set of crystallographic structures using GETAREA (84).

The extent of propagation of a mutational perturbation into the protein structure was estimated by the recently developed

structural perturbation analysis (SPA) (25,26,37), which takes just the protein structure and the identity of the mutation as input. The SPA can be condensed into two equations:

$$\Delta Q_{i,j} = Q_{i,j}^{WT} x_1 \left(1 - \frac{n_{mut}}{n_{WT}}\right) \quad (1)$$

$$\Delta Q_{j,k} = Q_{j,k}^{WT} x_2 \left(1 - \frac{n_{mut}}{n_{WT}}\right), \quad (2)$$

where $Q_{a,b}$ was the number of heavy atom contacts within a 6 Å cut-off between residues *a* and *b* (particularized in Eqs. 1 and 2 as i,j and j,k, respectively), ΔQ represented the loss of interactions (or the predicted perturbations) between neighbors in the first shell of residue i (defined by *j*) and between the first- and second-shell neighbors (*k*), *n* was the number of atoms of residue *i* in the WT (n_{WT}) or in the mutant (n_{Mut}), respectively. The values of x_1 and x_2 were fixed to 0.5 and 0.2, respectively; these values robustly predicted the change in stability induced by 375 mutations from 19 proteins (26). The extracted perturbations also followed an exponential dependence with distance (25) from the mutated site that was consistent with both all-atom molecular dynamics simulations (26) and experimental observations (37). The extent of propagation per residue (Q_N) was determined dividing the effect on that residue by the effect on the mutated site.

Expression and characterization of AGT variants in eukaryotic cells

For expression of AGT variants, the cDNA of AGT WT, LM, LM-p.G170R and p.I244T cloned into the pCI-neo plasmid (58) were used. Unnatural mutations at G170 and I244 sites (Table 1) were introduced using standard procedures (in these cases, codons were optimized for eukaryotic expression). For stable expression studies, chinese hamster ovary (CHO) cells expressing glycolate oxidase (GO; CHO-GO cells) were cultured at 37°C under O_2/CO_2 (19:1) atmosphere in Ham's F12 Glutamax medium (Invitrogen) supplemented with 10% (v/v) fetal bovine serum, 100 units·ml^{-1} penicillin and 100 µg·ml^{-1} streptomycin. To generate stably transfected cell clones expressing AGT variants, CHO-GO cells were transfected with Turbofect™ Transfection Reagent (Fermentas) according to the manufacturer's instructions. After 24 h, the selection agent G-418 at 1 mg·ml^{-1} was added to the medium and cells were allowed to grow for at least 4 weeks. For western blot analyses and activity measurements, cells were harvested and lysed in phosphate buffered saline (PBS) pH 7.2 (with protease inhibitors cocktail, Complete, Roche) by freezing/thawing cycles and immediately used.

For western blot analysis, 10 µg of each cell lysate was resolved using SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and transferred to nitrocellulose membranes. For AGT or GO detection the membrane was incubated with polyclonal rabbit anti-AGT serum (dilution 1:10000), or anti-GO serum (dilution 1:5000), kindly provided by Professor C.J. Danpure (University College London). As secondary antibodies, peroxidase-conjugated anti rabbit IgG was used (GE Healthcare; dilution 1:10000). Blotted proteins were detected and quantified with ECL® (Pierce Biotechnology, Rockford, IL), using a ChemiDoc XRS Imaging System (Bio-Rad, Hercules, CA) and the QuantityOne software.

The AGT enzymatic activity of CHO-GO cells was determined by incubating 100 µg of soluble lysate with 0.5 M L-alanine and 10 mM glyoxylate for 10 min at 37°C in the presence of 200 µM PLP in 0.1 M potassium phosphate buffer pH 7.4. Reactions were stopped by adding trichloroacetic acid 10% (v/v) and formed pyruvate was determined using a spectrophotometric assay described previously (85).

For immunofluorescence microscopy studies, ~3 × 10^5 cells CHO-GO cells expressing the analyzed species were seeded into each well of a 24 well plate, containing a 13 mm glass coverslip, and grown for 24 h at 37°C under O_2/CO_2 (19:1). Cells were fixed in 4% (w/v) paraformaldehyde for 20 min at room temperature (RT), permeabilized with 0.1% Triton X-100 in PBS for 5 min at RT, and then blocked in 3% bovine serum albumin 1% glycine in PBS. For the immunolabeling, rabbit polyclonal anti-human AGT and anti-peroxisomal GO were used as primary antibodies, and Alexa Fluor conjugated antibodies (Life technologies) were used as secondary antibodies (86). Nuclei were stained with DAPI and the coverslips were mounted over slides in AF1 medium (Dako). Mitochondria were stained in living cells by using Mitotracker Red (CMXRos version, Molecular Probes, Invitrogen) according to manufacturer instructions. Images were captured using a confocal laser-scanning fluorescence microscope Leica SP5 (Leica Microsystem, Manheim, Germany) at 63× magnification and analyzed using ImageJ. Images were processed using Adobe Photoshop.

Expression and purification of TTR protein variants in *E. coli*

Mutants at the V30 position (note that amino acid numbering in this protein excluded the 20 residue signal peptide) of human TTR were prepared by site-directed mutagenesis using the QuickChange Lightning kit (Agilent technologies, Santa Clara, California, USA), the pET-28a vector containing the full length cDNA for WT TTR as a template and codons were optimized for *E. coli* expression. All TTR proteins were expressed in *E. coli* BL21 (DE3) cells harboring the corresponding plasmid. Cultures in Luria-Bertani medium containing 50 µg mL^{-1} kanamycin were grown at 37°C to an OD at 600 nm of 0.6 and then induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at 37°C for 20 h. Harvested cells were resuspended in 20 mM Tris–HCl, 500 mM NaCl, pH 7.5 and lysed by sonication. Upon centrifugation, soluble extracts were fractionated by two cycles of ammonium sulfate precipitation, and TTR-containing fractions (precipitated between 50–90% ammonium sulfate) were dissolved in 25 mM Tris–HCl, pH 8 and dialyzed against the same buffer. Samples were then loaded onto a HiTrap Q HP anion exchange column (GE Healthcare Bio-Sciences, Uppsala, Sweden) equilibrated in 25 mM Tris–HCl, pH 8 and eluted with a linear 0–0.5 M NaCl gradient in the same solvent. TTR-enriched fractions were pooled, precipitated by 90% ammonium sulfate, resuspended in 25 mM Tris–HCl, 100 mM NaCl, pH 8 and further purified on the Superdex 75 HiLoad 26/60 prep grade column (GE Healthcare Bio-Sciences) equilibrated and eluted in 25 mM Tris–HCl, 100 mM NaCl, pH 8. Purity was confirmed by SDS-PAGE and protein concentration was determined spectrophotometrically ($\varepsilon_{(280)} = 77600\ M^{-1}\cdot cm^{-1}$).

Aggregation assays for TTR variants

TTR aggregation kinetics was monitored from initial soluble TTR species by measuring turbidity at 330 nm in a Cary-400 UV/Vis spectrophotometer (Varian Inc., USA). Samples containing 3.5 µM monomer in 200 mM sodium acetate, 100 mM KCl, pH 4.4 were incubated at 37°C without agitation and kinetics was monitored for 0–72 h. All subsequent experiments were carried

under the same conditions and incubations of 72 h, except Th-T binding experiments that were carried out with 0.35 μM of TTR. For Th-T fluorescence experiments, spectra were recorded with 25 μM of dye in a Cary Eclipse Fluorescence Spectrofluorometer (Agilent technologies) with an excitation wavelength of 440 nm and emission range from 460 to 600 nm (10 nm excitation and emission slits) and averaged from four accumulations. The intensity at 482 nm was used as a measure of amyloid formation. The results were normalized using the intensity of the WT TTR at final point. For sedimentation assays, TTR proteins were allowed to aggregate, and then samples were centrifuged at 20 000 g for 1 h, the supernatant was carefully recovered and its UV-spectra was registered. The percentage of the remaining TTR in the supernatant was calculated with respect to the initial concentration (0 h). For TEM analyses by negative staining, samples of TTR incubated for 72 h were placed onto carbon-coated copper grids and left for 5 min. The grids were then washed with distilled water, stained with 2% (*w*/*v*) uranyl acetate for 5 min., blotting with filter paper to remove excess of uranyl acetate and dried out before analysis. A TEM JEM-1400 (JEOL, Peabody, MA, USA) microscope was used operating at accelerating voltage of 120 kV. The more representative images of each grid were selected.

DNA sequence analyses

To investigate whether mutations were located at CpG dinucleotides, coding reference sequence of the gene of interest were manually scanned in both the sense and antisense strands for the three possible locations of the CpG dinucleotide in a codon: CGN, NCG and NNC-GNN (87). Codon bias/usage analyses were manually performed using recently reported data for codon usage statistics in human DNA (78).

Supplementary Material

Supplementary Material is available at *HMG* online.

Acknowledgements

E.M.C. was supported by a pre-doctoral fellowship from Junta de Andalucia and a short-term mobility fellowship from University of Granada. E.O. was supported by a fellowship from the Italian Ministery of Instruction University and Research (RBSI148BK3 to B.C.). S.V. has been granted an ICREA ACADEMIA award. A.N.N. is a Wellcome Trust/DBT Intermediate Fellow.

Conflict of Interest statement. None declared.

Funding

Spanish Ministry of Economy and Competitiveness (BIO2016-78310-R to S.V. and SAF2015-69796 to E.S.); Junta de Andalucia (P11-CTS-07187, to A.L.P.); the Oxalosis and Hyperoxaluria Foundation (OHF2017 to B.C.).

References

1. Salgado, D., Bellgard, M.I., Desvignes, J.P. and Beroud, C. (2016) How to identify pathogenic mutations among all those variations: variant annotation and filtration in the genome sequencing era. *Hum. Mutat.*, **37**, 1272–1282.
2. Chakravorty, S. and Hegde, M. (2018) Inferring the effect of genomic variation in the new era of genomics. *Hum. Mutat.*, **39**, 756–773.
3. Desvignes, J.P., Bartoli, M., Delague, V., Krahn, M., Miltgen, M., Beroud, C. and Salgado, D. (2018) VarAFT: a variant annotation and filtration system for human next generation sequencing data. *Nucleic Acids Res.*, **46**, W545–W553.
4. Sim, N.L., Kumar, P., Hu, J., Henikoff, S., Schneider, G. and Ng, P.C. (2012) SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res.*, **40**, W452–W457.
5. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S. and Sunyaev, S.R. (2010) A method and server for predicting damaging missense mutations. *Nat. Methods*, **7**, 248–249.
6. Chun, S. and Fay, J.C. (2009) Identification of deleterious mutations within three human genomes. *Genome Res.*, **19**, 1553–1561.
7. 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.
8. Nielsen, S.V., Stein, A., Dinitzen, A.B., Papaleo, E., Tatham, M.H., Poulsen, E.G., Kassem, M.M., Rasmussen, L.J., Lindorff-Larsen, K. and Hartmann-Petersen, R. (2017) Predicting the impact of Lynch syndrome-causing missense mutations from structural calculations. *PLoS Genet.*, **13**, e1006739.
9. Mesa-Torres, N., Salido, E. and Pey, A.L. (2014) The lower limits for protein stability and foldability in primary hyperoxaluria type I. *Biochim. Biophys. Acta*, **1844**, 2355–2365.
10. Mesa-Torres, N., Betancor-Fernández, I., Oppici, E., Cellini, B., Salido, E. and Pey, A.L. (2018) Evolutionary divergent suppressor mutations in conformational diseases. *Genes*, **9**, E352.
11. Soskine, M. and Tawfik, D.S. (2010) Mutational effects and the evolution of new protein functions. *Nat. Rev. Genet.*, **11**, 572–582.
12. 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.
13. Tang, Q. and Fenton, A.W. (2017) Whole-protein alanine-scanning mutagenesis of allostery: a large percentage of a protein can contribute to mechanism. *Hum. Mutat.*, **38**, 1132–1143.
14. Tzeng, S.R. and Kalodimos, C.G. (2012) Protein activity regulation by conformational entropy. *Nature*, **488**, 236–240.
15. Jack, B.R., Meyer, A.G., Echave, J. and Wilke, C.O. (2016) Functional sites induce long-range evolutionary constraints in enzymes. *PLoS Biol.*, **14**, e1002452.
16. 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.
17. Kiel, C., Benisty, H., Llorens-Rico, V. and Serrano, L. (2016) The yin-yang of kinase activation and unfolding explains the peculiarity of Val600 in the activation segment of BRAF. *Elife*, **5**, e12814.
18. Erlandsen, H., Pey, A.L., Gamez, A., Perez, B., Desviat, L.R., Aguado, C., Koch, R., Surendran, S., Tyring, S., Matalon, R. *et al.* (2004) Correction of kinetic and stability defects by tetrahydrobiopterin in phenylketonuria patients with certain phenylalanine hydroxylase mutations. *Proc. Natl. Acad. Sci. U. S. A.*, **101**, 16903–16908.

19. Pey, A.L., Perez, B., Desviat, L.R., Martinez, M.A., Aguado, C., Erlandsen, H., Gamez, A., Stevens, R.C., Thorolfsson, M., Ugarte, M. *et al.* (2004) Mechanisms underlying responsiveness to tetrahydrobiopterin in mild phenylketonuria mutations. *Hum. Mutat.*, **24**, 388–399.
20. Pey, A.L., Majtan, T., Sanchez-Ruiz, J.M. and Kraus, J.P. (2013) 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.
21. Pey, A.L., Mesa-Torres, N., Chiarelli, L.R. and Valentini, G. (2013) Structural and energetic basis of protein kinetic destabilization in human phosphoglycerate kinase 1 deficiency. *Biochemistry*, **52**, 1160–1170.
22. ben Bdira, F., Gonzalez, E., Pluta, P., Lain, A., Sanz-Parra, A., Falcon-Perez, J.M. 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.
23. Chiarelli, L.R., Morera, S.M., Bianchi, P., Fermo, E., Zanella, A., Galizzi, A. and Valentini, G. (2012) Molecular insights on pathogenic effects of mutations causing phosphoglycerate kinase deficiency. *PLoS One*, **7**, e32065.
24. 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.
25. Rajasekaran, N. and Naganathan, A.N. (2017) A self-consistent structural perturbation approach for determining the magnitude and extent of allosteric coupling in proteins. *Biochem. J.*, **474**, 2379–2388.
26. Rajasekaran, N., Suresh, S., Gopi, S., Raman, K. and Naganathan, A.N. (2017) A general mechanism for the propagation of mutational effects in proteins. *Biochemistry*, **56**, 294–305.
27. Pey, A.L. (2018) Biophysical and functional perturbation analyses at cancer-associated P187 and K240 sites of the multifunctional NADP(H):quinone oxidoreductase 1. *Int. J. Biol. Macromol*, **118**, 1912–1923.
28. Shendure, J. and Akey, J.M. (2015) The origins, determinants, and consequences of human mutations. *Science*, **349**, 1478–1483.
29. Baer, C.F., Miyamoto, M.M. and Denver, D.R. (2007) Mutation rate variation in multicellular eukaryotes: causes and consequences. *Nat. Rev. Genet.*, **8**, 619–631.
30. Fortian, A., Gonzalez, E., Castano, D., Falcon-Perez, J.M. and Millet, O. (2011) Intracellular rescue of the uroporphyrinogen III synthase activity in enzymes carrying the hotspot mutation C73R. *J. Biol. Chem.*, **286**, 13127–13133.
31. Tao, Y.X. and Conn, P.M. (2018) Pharmacoperones as novel therapeutics for diverse protein conformational diseases. *Physiol. Rev.*, **98**, 697–725.
32. Sekijima, Y., Wiseman, R.L., Matteson, J., Hammarstrom, P., Miller, S.R., Sawkar, A.R., Balch, W.E. and Kelly, J.W. (2005) The biological and chemical basis for tissue-selective amyloid disease. *Cell*, **121**, 73–85.
33. Hartl, F.U. (2017) Protein misfolding diseases. *Annu. Rev. Biochem.*, **86**, 21–26.
34. Cremades, N., Cohen, S.I., Deas, E., Abramov, A.Y., Chen, A.Y., Orte, A., Sandal, M., Clarke, R.W., Dunne, P., Aprile, F.A. *et al.* (2012) Direct observation of the interconversion of normal and toxic forms of alpha-synuclein. *Cell*, **149**, 1048–1059.
35. Balchin, D., Hayer-Hartl, M. and Hartl, F.U. (2016) In vivo aspects of protein folding and quality control. *Science*, **353**, aac4354.
36. Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F. and Serrano, L. (2005) The FoldX web server: an online force field. *Nucleic Acids Res.*, **33**, W382–W388.
37. Rajasekaran, N., Sekhar, A. and Naganathan, A.N. (2017) A universal pattern in the percolation and dissipation of protein structural perturbations. *J. Phys. Chem. Lett.*, **8**, 4779–4784.
38. Van Durme, J., Delgado, J., Stricher, F., Serrano, L., Schymkowitz, J. and Rousseau, F. (2011) A graphical interface for the FoldX forcefield. *Bioinformatics*, **27**, 1711–1712.
39. Zhang, X., Roe, S.M., Hou, Y., Bartlam, M., Rao, Z., Pearl, L.H. and Danpure, C.J. (2003) Crystal structure of alanine:glyoxylate aminotransferase and the relationship between genotype and enzymatic phenotype in primary hyperoxaluria type 1. *J. Mol. Biol.*, **331**, 643–652.
40. Djordjevic, S., Zhang, X., Bartlam, M., Ye, S., Rao, Z. and Danpure, C.J. (2010) Structural implications of a G170R mutation of alanine:glyoxylate aminotransferase that is associated with peroxisome-to-mitochondrion mistargeting. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.*, **66**, 233–236.
41. 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.
42. Asher, G., Dym, O., Tsvetkov, P., Adler, J. and Shaul, Y. (2006) The crystal structure of NAD(P)H quinone oxidoreductase 1 in complex with its potent inhibitor dicoumarol. *Biochemistry*, **45**, 6372–6378.
43. 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.
44. 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.
45. 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.
46. 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.
47. 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.
48. Lai, K. and Elsas, L.J. (2001) Structure-function analyses of a common mutation in blacks with transferase-deficiency galactosemia. *Mol. Genet. Metab.*, **74**, 264–272.
49. Lai, K., Willis, A.C. and Elsas, L.J. (1999) The biochemical role of glutamine 188 in human galactose-1-phosphate uridyltransferase. *J. Biol. Chem.*, **274**, 6559–6566.
50. Tongsook, C., Niederhauser, J., Kronegger, E., Straganz, G. and Macheroux, P. (2017) Leucine 208 in human histamine N-methyltransferase emerges as a hotspot for protein

stability rationalizing the role of the L208P variant in intellectual disability. *Biochim. Biophys. Acta*, **1863**, 188–199.
51. Sontag, E.M., Samant, R.S. and Frydman, J. (2017) Mechanisms and functions of spatial protein quality control. *Annu. Rev. Biochem.*, **86**, 97–122.
52. van der Lee, R., Lang, B., Kruse, K., Gsponer, J., Sanchez de Groot, N., Huynen, M.A., Matouschek, A., Fuxreiter, M. and Babu, M.M. (2014) Intrinsically disordered segments affect protein half-life in the cell and during evolution. *Cell Rep.*, **8**, 1832–1844.
53. 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.
54. Yu, H. and Matouschek, A. (2017) Recognition of client proteins by the proteasome. *Annu. Rev. Biophys.*, **46**, 149–173.
55. Mesa-Torres, N., Tomic, N., Albert, A., Salido, E. and Pey, A.L. (2015) Molecular recognition of PTS-1 cargo protein by Pex5p: implications for protein mistargeting in primary hyperoxaluria. *Biomolecules*, **5**, 121–141.
56. Salido, E., Pey, A.L., Rodriguez, R. and Lorenzo, V. (2012) Primary hyperoxalurias: disorders of glyoxylate detoxification. *Biochim. Biophys. Acta*, **1822**, 1453–1464.
57. Oppici, E., Dindo, M., Conter, C., Borri Voltattorni, C. and Cellini, B. (2018) Folding defects leading to primary hyperoxaluria. *Handb. Exp. Pharmacol.*, **245**, 313–343.
58. Mesa-Torres, N., Fabelo-Rosa, I., Riverol, D., Yunta, C., Albert, A., Salido, E. and Pey, A.L. (2013) The role of protein denaturation energetics and molecular chaperones in the aggregation and mistargeting of mutants causing primary hyperoxaluria type I. *PLoS One*, **8**, e71963.
59. Santana, A., Salido, E., Torres, A. and Shapiro, L.J. (2003) 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**, 7277–7282.
60. Fargue, S., Lewin, J., Rumsby, G. and Danpure, C.J. (2013) Four of the most common mutations in primary hyperoxaluria type 1 unmask the cryptic mitochondrial targeting sequence of alanine:glyoxylate aminotransferase encoded by the polymorphic minor allele. *J. Biol. Chem.*, **288**, 2475–2484.
61. Montioli, R., Oppici, E., Dindo, M., Roncador, A., Gotte, G., Cellini, B. and Borri Voltattorni, C. (2015) Misfolding caused by the pathogenic mutation G47R on the minor allele of alanine:glyoxylate aminotransferase and chaperoning activity of pyridoxine. *Biochim. Biophys. Acta*, **1854**, 1280–1289.
62. Oppici, E., Fargue, S., Reid, E.S., Mills, P.B., Clayton, P.T., Danpure, C.J. and Cellini, B. (2015) 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.
63. Eisele, Y.S., Monteiro, C., Fearns, C., Encalada, S.E., Wiseman, R.L., Powers, E.T. and Kelly, J.W. (2015) Targeting protein aggregation for the treatment of degenerative diseases. *Nat. Rev. Drug. Discov.*, **14**, 759–780.
64. Zhen, D.B., Swiecicki, P.L., Zeldenrust, S.R., Dispenzieri, A., Mauermann, M.L. and Gertz, M.A. (2015) Frequencies and geographic distributions of genetic mutations in transthyretin- and non-transthyretin-related familial amyloidosis. *Clin. Genet.*, **88**, 396–400.
65. Zhang, F., Hu, C., Dong, Y., Lin, M.S., Liu, J., Jiang, X., Ge, Y. and Guo, Y. (2013) The impact of V30A mutation on transthyretin protein structural stability and cytotoxicity against neuroblastoma cells. *Arch. Biochem. Biophys.*, **535**, 120–127.
66. Suhr, O.B., Andersen, O., Aronsson, T., Jonasson, J., Kalimo, H., Lundahl, C., Lundgren, H.E., Melberg, A., Nyberg, J., Olsson, M. *et al.* (2009) Report of five rare or previously unknown amyloidogenic transthyretin mutations disclosed in Sweden. *Amyloid*, **16**, 208–214.
67. Martin, S.E., Benson, M.D. and Hattab, E.M. (2014) The pathologic spectrum of oculoleptomeningeal amyloidosis with Val30Gly transthyretin gene mutation in a postmortem case. *Hum. Pathol.*, **45**, 1105–1108.
68. Buell, A.K., Dhulesia, A., Mossuto, M.F., Cremades, N., Kumita, J.R., Dumoulin, M., Welland, M.E., Knowles, T.P., Salvatella, X. and Dobson, C.M. (2011) Population of nonnative states of lysozyme variants drives amyloid fibril formation. *J. Am. Chem. Soc.*, **133**, 7737–7743.
69. Hammarstrom, P., Wiseman, R.L., Powers, E.T. and Kelly, J.W. (2003) Prevention of transthyretin amyloid disease by changing protein misfolding energetics. *Science*, **299**, 713–716.
70. Sant'Anna, R., Gallego, P., Robinson, L.Z., Pereira-Henriques, A., Ferreira, N., Pinheiro, F., Esperante, S., Pallares, I., Huertas, O., Rosario Almeida, M. *et al.* (2016) Repositioning tolcapone as a potent inhibitor of transthyretin amyloidogenesis and associated cellular toxicity. *Nat. Commun.*, **7**, 10787.
71. Bulawa, C.E., Connelly, S., Devit, M., Wang, L., Weigel, C., Fleming, J.A., Packman, J., Powers, E.T., Wiseman, R.L., Foss, T.R. *et al.* (2012) Tafamidis, a potent and selective transthyretin kinetic stabilizer that inhibits the amyloid cascade. *Proc. Natl. Acad. Sci. U. S. A.*, **109**, 9629–9634.
72. Robertson, K.D. (2005) DNA methylation and human disease. *Nat. Rev. Genet.*, **6**, 597–610.
73. Swami, M. (2010) Mutation: it's the CpG content that counts. *Nat. Rev. Genet.*, **11**, 456.
74. Antonarakis, S.E. (2006) *In eLS*. John Wiley and Sons.
75. Saunders, R. and Deane, C.M. (2010) Synonymous codon usage influences the local protein structure observed. *Nucleic Acids Res.*, **38**, 6719–6728.
76. O'Brien, E.P., Ciryam, P., Vendruscolo, M. and Dobson, C.M. (2014) Understanding the influence of codon translation rates on cotranslational protein folding. *Acc. Chem. Res.*, **47**, 1536–1544.
77. Pechmann, S. and Frydman, J. (2013) Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. *Nat. Struct. Mol. Biol.*, **20**, 237–243.
78. Athey, J., Alexaki, A., Osipova, E., Rostovtsev, A., Santana-Quintero, L.V., Katneni, U., Simonyan, V. and Kimchi-Sarfaty, C. (2017) A new and updated resource for codon usage tables. *BMC Bioinformatics*, **18**, 391.
79. Watanabe, M., Osada, J., Aratani, Y., Kluckman, K., Reddick, R., Malinow, M.R. and Maeda, N. (1995) Mice deficient in cystathionine beta-synthase: animal models for mild and severe homocyst(e)inemia. *Proc. Natl. Acad. Sci. U. S. A.*, **92**, 1585–1589.
80. Salido, E.C., Li, X.M., Lu, Y., Wang, X., Santana, A., Roy-Chowdhury, N., Torres, A., Shapiro, L.J. and Roy-Chowdhury, J. (2006) Alanine-glyoxylate aminotransferase-deficient mice, a model for primary hyperoxaluria that responds to adenoviral gene transfer. *Proc. Natl. Acad. Sci. U. S. A.*, **103**, 18249–18254.
81. Episkopou, V., Maeda, S., Nishiguchi, S., Shimada, K., Gaitanaris, G.A., Gottesman, M.E. and Robertson, E.J. (1993) Disruption of the transthyretin gene results in mice with depressed levels of plasma retinol and thyroid hormone. *Proc. Natl. Acad. Sci. U. S. A.*, **90**, 2375–2379.

82. Radjendirane, V., Joseph, P., Lee, Y.H., Kimura, S., Klein-Szanto, A.J., Gonzalez, F.J. and Jaiswal, A.K. (1998) Disruption of the DT diaphorase (NQO1) gene in mice leads to increased menadione toxicity. *J. Biol. Chem.*, **273**, 7382–7389.
83. Hole, M., Jorge-Finnigan, A., Underhaug, J., Teigen, K. and Martinez, A. (2016) Pharmacological chaperones that protect tetrahydrobiopterin dependent aromatic amino acid hydroxylases through different mechanisms. *Curr. Drug Targets*, **17**, 1515–1526.
84. Fraczkiewicz, R. and Braun, W. (1998) Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J. Comp. Chem.*, **19**, 319–333.
85. Cellini, B., Bertoldi, M., Montioli, R., Paiardini, A. and Borri Voltattorni, C. (2007) Human wild-type alanine: glyoxylate aminotransferase and its naturally occurring G82E variant: functional properties and physiological implications. *Biochem. J.*, **408**, 39–50.
86. Oppici, E., Roncador, A., Montioli, R., Bianconi, S. and Cellini, B. (2013) Gly161 mutations associated with primary hyperoxaluria type I induce the cytosolic aggregation and the intracellular degradation of the apo-form of alanine:glyoxylate aminotransferase. *Biochim. Biophys. Acta*, **1832**, 2277–2288.
87. Leroy, B., Anderson, M. and Soussi, T. (2014) TP53 mutations in human cancer: database reassessment and prospects for the next decade. *Hum. Mutat.*, **35**, 672–688.