Protein ensembles link genotype to phenotype

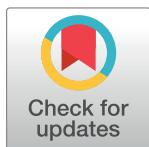**Ruth Nussinov**[1,2]*, **Chung-Jung Tsai**[1], **Hyunbum Jang**[1]

1 Cancer and Inflammation Program, Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research, National Cancer Institute at Frederick, Frederick, Maryland, United States of America, **2** Sackler Institute of Molecular Medicine, Department of Human Genetics and Molecular Medicine, Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel

* NussinoR@mail.nih.gov

Abstract

Classically, phenotype is what is observed, and genotype is the genetic makeup. Statistical studies aim to project phenotypic likelihoods of genotypic patterns. The traditional genotype-to-phenotype theory embraces the view that the encoded protein shape together with gene expression level largely determines the resulting phenotypic trait. Here, we point out that the molecular biology revolution at the turn of the century explained that the gene encodes not one but ensembles of conformations, which in turn spell all possible gene-associated phenotypes. The significance of a dynamic ensemble view is in understanding the linkage between genetic change and the gained observable physical or biochemical characteristics. Thus, despite the transformative shift in our understanding of the basis of protein structure and function, the literature still commonly relates to the classical genotype–phenotype paradigm. This is important because an ensemble view clarifies how even seemingly small genetic alterations can lead to pleiotropic traits in adaptive evolution and in disease, why cellular pathways can be modified in monogenic and polygenic traits, and how the environment may tweak protein function.

Introduction

The terms genotype and phenotype have been in use at least since the turn of the last century. Genotype has been defined as the genetic makeup of an organism or of a specific characteristic. Phenotype (from Greek phainein, meaning "to show," and typos, meaning "type") has been construed as the composite of the organism's observable characteristics or traits, such as morphology, development, biochemical, and physiological properties. Classically, the genotype of an organism has been described as the inherited genetic material coding for all processes in the organism's life. It provided some measurement of how an individual is specialized within a species based on its genomic sequence. By contrast, the phenotype referred to the observation that similar genotypes can differ in their expression under different environmental and developmental conditions. Typically, an individual's genotype relates to a particular gene of interest or to the combination of alleles that the individual organism or cell carries. To explain how the genotype determines the phenotype, population genetics [1] pointed out that (1) in real populations, phenotypic ratios are determined by the frequency of alleles in the population as well
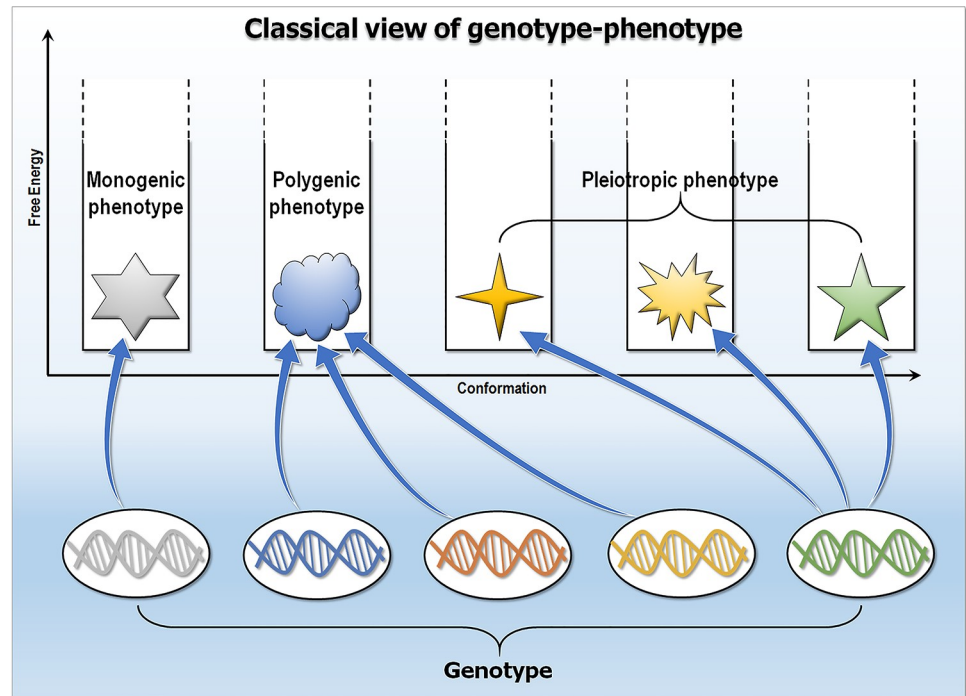
Fig 1. Classical view of genotype–phenotype. In this view, a protein or the gene product is considered to have one shape with a single function. Monogenic traits are expressed by single genes, whereas polygenic traits are affected by multiple genes. Seemingly unrelated phenotypic traits are pleiotropy that can be expressed by a single gene.

as by whether the alleles are in dominant or recessive form, (2) the number of phenotypes produced for a given trait depends on how many genes control that trait, and (3) there is no one-to-one mapping between genes and traits. Exactly what is a "trait" was not well defined.

The classical genotype–phenotype interpretation dates to a period when a protein, the gene product, was believed to exist in one shape with a single function (Fig 1). Evolution was perceived to optimize that shape for this function. Phenotype was considered as a visually observable property. Over a century later, with the understanding of the basis of protein structure and function having undergone a dramatic revolution, the genotype–phenotype paradigm remains unchanged. Scientific publications still commonly relate to it in terms of this weathered image. This view overlooks the fact that biomolecules exist as heterogeneous dynamic interconverting states with varying energies, and the multiple traits may mirror those protein states. The "second molecular biology revolution" [2], which imported newer concepts from physics and chemistry to molecular biology, such as the powerful idea of the free energy landscape [3], allows a new view of this genotype–phenotype dogma. Biomolecules should be thought of not as static single shapes but as statistical ensembles [4–7]. Here, we explain that structural ensembles—which allow proteins to fulfill their functions—link genotypes to phenotypes. Thus, within the broad cellular context, it is the network that controls transcription via gene regulation [8–30]; here, however, we relate to mutations that affect function at the lower, protein level.

Protein evolution in terms of biophysics

The evolution of proteins in terms of their conformational ensembles has not been overlooked [31, 32]. In a series of studies, the relationship of the protein's structure and conformational

dynamics to its function and thus its fitness has been explored, not through classical phylogenetic approaches, which largely overlook biophysical principles, but by evaluating how mutations impact protein structure, which already has marginal stability [33–35]. Mutations can also shift the equilibrium from an inactive, autoinhibited state to the active state, as for example observed in Raf and phosphoinositide 3-kinase (PI3K) [36]. A linkage between evolution and biophysics was also explored by changes in dynamic flexibility profiles [37], by protein interaction networks [38], by protein adaptation as observed by the functional impact of multiple mutations, by identifying key adaptive mutational solutions to the same selective pressure [39, 40], by extant fold-switching proteins [41], and by exploring the relationship between metastability, the fitness landscape, and sequence divergence [32]. Evolutionary selection has been explored in terms of the dynamics of structural evolution [42], and evidence for evolutionary selection in cotranslational folding was also found [43]. The stability of a viral protein was observed to correlate with its evolutionary dynamics [44], and the evolution of the biophysical fitness landscape of an RNA virus was explored as well [45]. Even a web app has recently been set up for such exploration [46]. Comprehensive analysis indicated that evolution conserves functional dynamic motions; clusters of conserved residues have a signature characteristic of protein domains, being spatially separated but individually compact [47]; and sequence and structurally conserved residues tend have a lower fluctuation than other residues [48]. Some studies emphasized the challenge of forecasting mutational outcomes, arguing that ensembles make evolution unpredictable [49].

Can the biophysical view of conformational ensembles link to classical evolutionary concepts, such as phenotypic plasticity, bet-hedging, canalization, and assimilation? Classically, a genotype's phenotype is viewed as relatively invariant, irrespective of different environments [50]. However, recently, a linkage was proposed via a "plasticity-first evolution" hypothesis that suggested how phenotypic plasticity may have facilitated macroevolutionary change [51]. Protein conformational diversity was suggested to correlate with evolutionary rate [52], and "phenotypic plasticity" through nongenetic heterogeneity was recently proposed to be driven by protein conformational dynamics [53]. The authors argued that mutations alter the relative probabilities of conformations, thus changing the effects of future mutations, resulting in uncertainty in the effect of each subsequent mutation and consequently prediction. Coincidentally, large-scale analysis has further shown that mutational effects on the conformations may also be small, even smaller than among proteins of identical sequences [54]. Finally, recent reviews described bridging the physical scales in evolutionary biology, from protein sequence space to fitness of organisms and populations [55], delineating the evolution of function from such a perspective [56].

The conformational space is vast, with the available X-crystal structures covering only a certain fraction, as recently elegantly documented for the Abl tyrosine kinase, for which molecular dynamics simulations and Markov state models identified a protein conformation apparently in a Lilly in-house structure of Abl with WHI-P15 but not in the Protein Data Bank (PDB) [57]. Even though here exploited for drug discovery, it is reasonable to expect that evolution has made use of such vast ensembles as well, adapting them in different ways [58, 59], including in thermostability [60, 61], diverse cellular environments [62], protein disorder and the switches between the ordered and the disordered state [63, 64], detailed linker histone sequence and posttranslational modification (PTM) [65] as well as catalysis of an (O-linked β-N-acetylglucosamine) O-GlcNAc PTM of nuclear and cytosolic protein [66], allosteric interaction networks and signaling pathways [67], and in higher-order organization [68]. This expectation not only can help in prediction of ligand binding [69] but also has inspired the proposition that accounting for conformational heterogeneity and dynamics would benefit protein design methods [70]. Along these lines, recent reviews underscored the role of

conformational dynamics in enzyme evolution toward new functions and suggested how conformational dynamism can be exploited in computational enzyme design protocols [71] and how a priori knowledge of an allosteric network could improve design through navigation of the design space [72]. Thus, biophysics is not concerned with quantitative association of gene loci nor with certain statistical measurements that are hallmarks of classical evolution; it views evolution at the basic conformational level and considers protein sequence and structural space and fitness of organisms and populations.

Statistical relationships between genotype and phenotype

Monogenic traits are affected by single genes, polygenic traits are affected by multiple genes, and pleiotropy occurs when one gene influences multiple, seemingly unrelated phenotypic traits [73–75] (Fig 1). Quantitative association of trait loci (QTL) aims to explain the genetic basis of variation in complex traits by linking phenotype data (trait measurements) to genotype data (e.g., single-nucleotide polymorphisms [SNPs]). However, quantifying traits is challenging; matters of contention include traits' definitions, interdependence, and selection. Statistics suggest that the frequency of pleiotropy is not high. Instead, the average phenotypic effect of a mutation on a trait increases with the number of traits that are affected by the mutation. Wagner and Zhang pointed out that pleiotropy may result from multiple molecular functions of a gene (type I pleiotropy) or from multiple outcomes of a single molecular function (type II pleiotropy) and that type II is the most common, which is why developing drugs that target only one particular phenotype of a pleiotropic gene may fail [74]. Here, we consider two types of scenarios (Fig 2). In the first, the mutations affect solely the protein (a "node"). This scenario applies only to monogenic phenotypes. In the second, the mutations create or break a protein–protein interaction (an "edge"). This scenario works by altering the cellular network. It can take place in gained monogenic or polygenic phenotypes. Since we focus on protein ensembles, we only consider mutations/SNPs in protein-coding regions.

Genome-wide association studies (GWAS) seek relationships between sites of common genome sequence variation and disease predisposition. They have already revealed the genetic basis of over 50 disease-susceptibility loci and provided insight into the allelic architecture of multifactorial traits [75], indicating that the approach can successfully identify common SNP variants with sufficiently large phenotypic change and illuminate relationships between changes in genome sequence and phenotypic variation. Challenges are not so much in detecting the statistical association signals but in relating them to the molecular mechanisms of phenotypic expression. There are 20,000 to 30,000 genes in the human body; there are millions of phenotypic traits [76]. Statistical studies such as those above can reveal relations and associations; they are unable to explain how SNPs, disease-related mutations, or mutations in adaptive evolution lead to phenotypic change. Since each record contains many variables, it is difficult to fully interpret the observed statistical biases [77]. This, however, becomes possible when we consider gene products as ensembles of dynamic, interconverting states. The linkage between genotype and phenotype can be understood in terms of the statistics of the ensemble.

The genotype encodes a conformational ensemble

In line with the rich biophysics literature, with some of the most recent cited above (as well as [34, 78]), and the long-standing awareness that multiple/promiscuous function drives transitions [35, 79], the genotype should be thought of not as determining a structure but instead as establishing a distinct conformational ensemble, which in turn specifies the phenotype (Fig 3). This ensemble embodies all states, including the orthosteric ligand-bound conformation, the activated (or inactivated) allosteric modulated states, posttranslationally modified states,
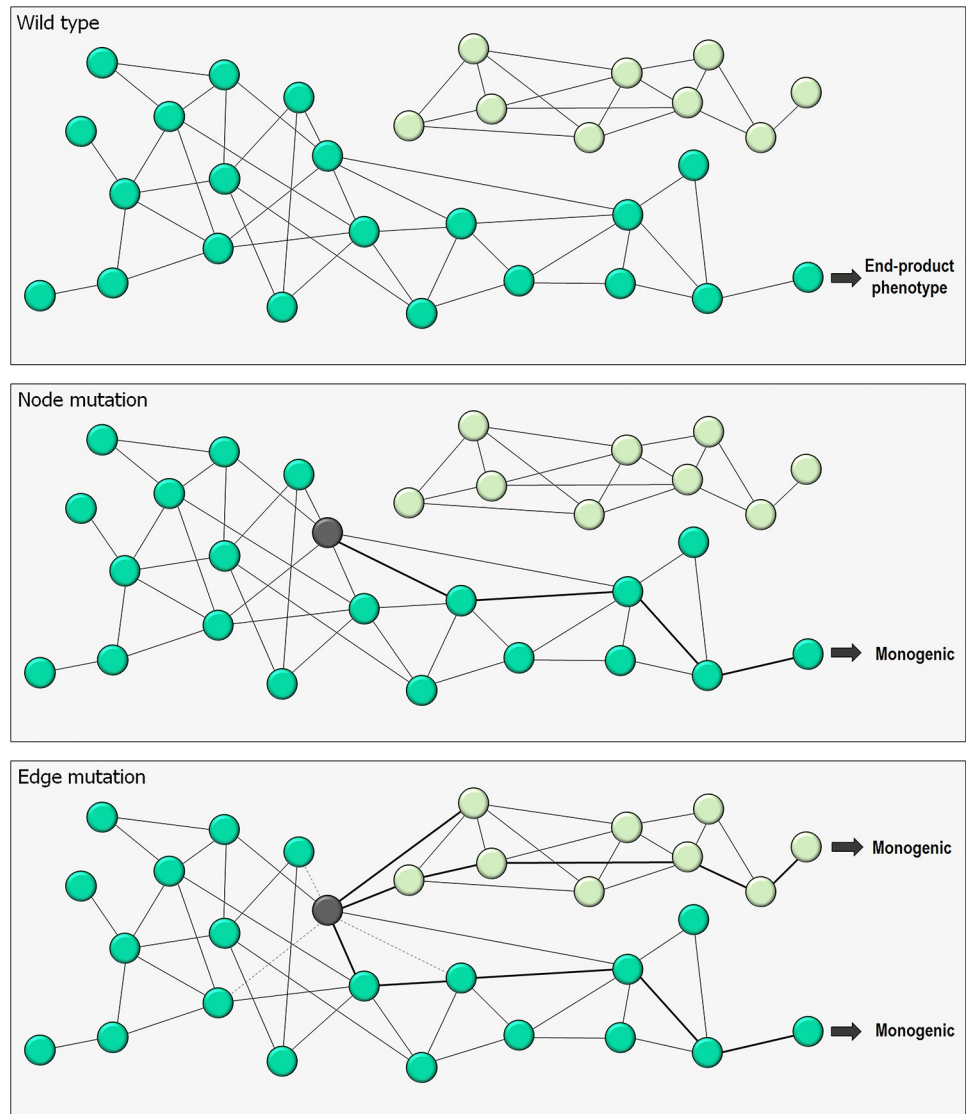
Fig 2. Network perturbations of genotype–phenotype. Mutations can shift phenotype traits generated from wild-type trait (upper panel), affecting solely the protein (a "node") or protein–protein interaction (an "edge"). The former and latter are node and edge mutations, respectively. The node mutation (middle panel) generates a monogenetic trait, whereas the edge mutation (bottom panel) creates or breaks the protein–protein interactions yielding monogenetic or polygenetic traits. An example shown for edge trait is polygenetic.

https://doi.org/10.1371/journal.pcbi.1006648.g002

transition states, and nonfunctional states serving as a reservoir for emerging functions [4–7]. The ensemble can be described by statistical mechanical laws, and its populations follow statistical distributions. Their specific distributions reflect the environment and conditions of the protein's milieu. The "environment and conditions" include both the physical surrounding, such as concentration, pH [80], presence of solutes or lipids, ions (where binding of sodium ions was shown to shift the population toward conformational states [81]) and covalent changes in the protein, such as mutations and their combinations, and phosphorylation, ubiquitylation, etc. A shift of the conformational ensemble that alters the highly populated state may define the phenotype. This thesis observes that evolution encoded all states and their populations in the genotype and optimized them, including rare states, for distinct phenotypes.
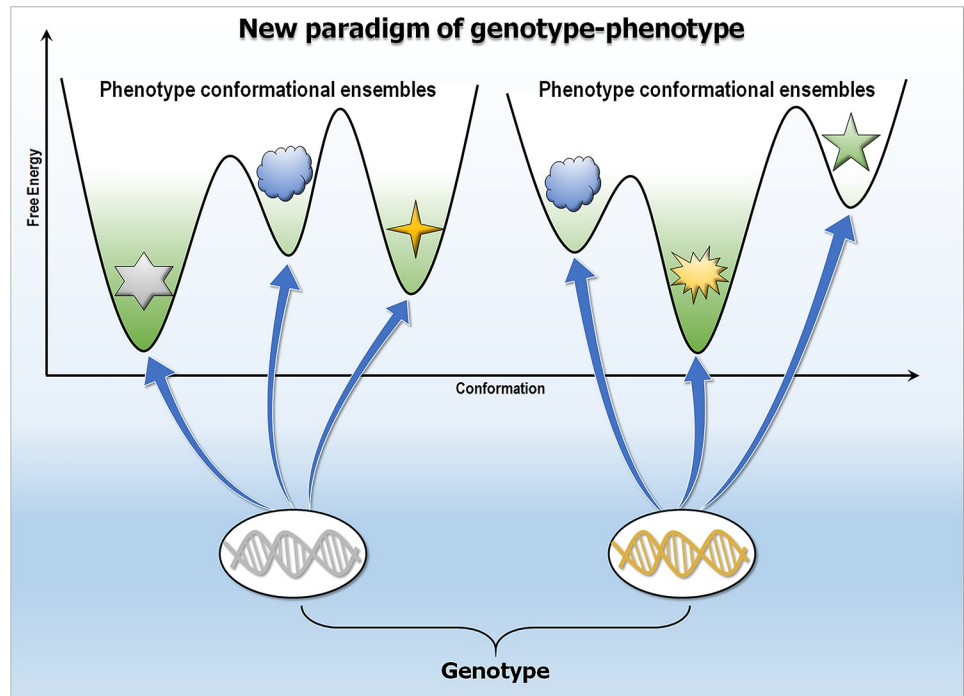
Fig 3. New paradigm of genotype–phenotype. In this view, genotype encodes a distinct conformational ensemble in all states. Populations determine the specific phenotype traits that link to genotype. Mutations shift the equilibrium of preexisting conformational ensembles altering phenotypes.

https://doi.org/10.1371/journal.pcbi.1006648.g003

The genotype—i.e., the genetic makeup of cells—and the phenotype have been in the lime-light for decades. The genotype and phenotype have been linked through expression of the specific encoded proteins. The linkage has been understood in terms of the specific three-dimensional structures that they obtain. However, the genetic makeup is expressed through ensembles of states and their interactions, and it is the distinct states in the ensemble that define the phenotype. A specific function is executed by a specific conformation, and a switch between the states, typically elicited by an allosteric event, can lead to a phenotypic change [82–86]. Examples include the G protein–coupled receptors (GPCRs), highly dynamic proteins that adopt multiple active states linked to distinct functional outcomes [87].

Structural ensembles link genotype to phenotype

Sickle cell anemia—a disorder that leads to atypical hemoglobin molecules that can distort red blood cells into a sickle, crescent shape resulting in anemia, repeated infections, and periodic episodes of pain—is one classical monogenic adaptive evolution example. It is a consequence of a single nucleotide/amino acid Glu → Val change. The conformational landscape of hemo-globin encompasses both states; however, the sickle cell conformation is visited more often in the mutant. A more recent example concerns melanocortin receptor 1 (MC1R), which plays a central role in regulation of animal coat color formation. Two nonsynonymous mutations in the *MC1R* gene previously associated with coat color in Chinese Minxian black-fur sheep are not present in the white coat (large-tailed Han) sheep [88]. A striking example concerns the coat color of deer mice [89]. Light coat color provides a selective advantage against visually hunting predators for mice that have recently colonized the light-colored soil of the Nebraska Sand Hills, in contrast to dark-colored deer mice inhabiting nearby dark soils. The phenotype

is composed of five color traits: dorsal hue, dorsal brightness, ventral color, dorsal–ventral boundary, and tail stripe. Each trait is statistically associated with a unique set of SNPs, with one exception, a serine deletion in exon 2 that was linked to both ventral color and tail stripe. Each color trait indicates selection. This example suggests (1) that the large-locus effect divides into small- to moderate-effect mutations and (2) that when one gene influences multiple, seemingly unrelated phenotypic traits, it is the individual mutations that bring the population closer to its phenotypic optimum. Thus, small minimally pleiotropic mutations occurring within a single gene may rapidly adapt an organism to multiple phenotypes. In the language of the free energy landscape, a set of mutations provides a complex conformational behavior; however, their combinations can more readily achieve specific optimal conformations. Similar strategies have been adopted by evolution through combinatorial sets of allosteric PTMs, which create recognition barcodes [90]. Each set of SNPs results in one type of coat color through distinct conformations.

A similar complex conformational behavior appears to be involved between the human MC1R pigmentation and skin cancer gene and youthful facial looks. Loss-of-function polymorphisms (multiple functions of a single gene) in human MC1R result in the yellow to reddish pheomelanin, which has a weaker ultraviolet (UV) shielding capacity. Recent analysis of over 8 million SNPs in 2,693 elderly Dutch Europeans indicated that MC1R SNP variants are most significantly associated with perceived facial age, with the homozygote MC1R risk haplotype looking up to 2 years older on average than noncarriers [91]. This association was independent of age, sex, skin color, and sun damage, such as wrinkling and pigmented spots, suggesting independence from melanin synthesis. The strongest associations with perceived facial age were for multiple SNPs; however, it was enhanced for four SNPs—variants rs1805005 (V60L), rs1805007 (R151C), rs1805008 (R160W), and rs1805009 (D294H)—known to be missense loss-of-function variants [92], causing phenotypes such as red hair color and pale skin [92, 93], and involved in age-related skin phenotypes such as pigmented spots [94]. Along similar lines, across a total of 2,459 patients at different ages and 53 families, the presence of a mutation in a sarcomere gene (cardiac myosin-binding protein C gene [*MYBPC3*] and myosin heavy chain 7 gene [*MYH7*]) in familial hypertrophic cardiomyopathy (HCM) has been associated with a number of phenotypic clinical features, including age at presentation, gender, family history of familial HCM and sudden cardiac death (SCD), and maximum left ventricular wall thickness (MLVWT) [95].

An additional combinatorial conformational barcodes example concerns *MYH9*-related disorders, a group of rare autosomal dominant platelet disorders caused by mutations in the *MYH9* gene encoding the nonmuscle myosin heavy chain II-A (NMMHC-IIA). Nonsyndromic forms are characterized by macrothrombocytopenia with giant platelets and leukocyte inclusion bodies; syndromic forms combine these hematological features with deafness and/or nephropathy and/or cataracts. A recent 8-year study of a large cohort of 109 patients from 37 sporadic cases and 39 unrelated families identified 43 genetic variants. Thirty-three of these were missense mutations. Distinct combinations of this heterogeneous mutational landscape resulted in specific disease phenotype [96].

Finally, a broad, systematic study of genotype–phenotype relationships mapped thousands of missense point mutations and in-frame insertions and deletions related to disease. Specific locations of distinct mutations of the same gene on the interface related to disease specificity. This was experimentally validated for the *MLH1*–*PMS2*, Wiskott–Aldrich syndrome protein (*WASP*)–cell division control protein 42 (*CDC42*), and tumor protein (*TP*) *63*–*TP73* interactions [97]. Mutations in protein–protein interfaces (edges) that change the conformational distributions can lead to a selection of a different binding partner, thereby leading to a phenotype change through altered interactions.

Single mutations

Above, sets of nonsense mutations combinatorially decide the preferred conformational states that provoke the phenotype. Below, we give examples of single-point mutations also acting by shifting the conformational landscape. Although in such cases it has been generally assumed that the phenotype is determined by a single ("driver") substitution, additional—albeit to date mostly unidentified—previous or subsequent mutational events may cooperate. Low-frequency presumed "passenger" mutations may act combinatorially with the driver in promoting distinct phenotypic expression [77, 98]. Somatic mutations evolve and accumulate over time and may be expected to affect the distributions of the conformational states, similar to the combinatorial examples above.

Plasma-membrane integrin αIIbβ3 is a major receptor in platelets during clotting. The L33P mutation in knockin mice reduces bleeding and clotting times and elevates the in vivo thrombosis phenotype, platelet attachment, and spreading onto fibrinogen [99]. Under unstimulated conditions, the mutation primes αIIbβ3 intracellular domains for outside-in signaling, increasing Src phosphorylation through talin interactions with the β3 cytoplasmic domain, leading to hypercoagulability and increased risk for coronary thrombosis and stroke. The mutation, resulting in a Pro32-Pro33 sequence, modifies the αIIbβ3 conformational equilibrium. The proline at position 33 was suggested to alter the flexibility in the β3 knee defined by the plexin-semaphorin-integrin (PSI), insulin-like growth factor 1 (IGF-1), and IGF-2 integrin αIIbβ3 extracellular domains, resulting in increased adhesion capacity of human platelet antigens (HPA)-1b platelets to fibrinogen [100]. Stormorken syndrome—a rare autosomal dominant disorder characterized by a phenotype that includes meiosis, thrombocytopenia/thrombocytopathy with bleeding time diathesis, intellectual disability, mild hypocalcemia, muscle fatigue, asplenia, and ichthyosis—provides a second example. The syndrome apparently results from a single gene defect, consistent with mendelian dominant inheritance. Stromal interaction molecule 1 (STIM1) protein mutation p.R304W is observed in patients but not in their unaffected family members. The STIM1 protein is an endoplasmic reticulum (ER) Ca^{2+} sensor. Data suggest that the STIM1 p.R304W mutation shifts the equilibrium toward an altered conformation of the inhibitory helix, unlocking the inhibitory state of STIM1. The mutation causes a gain of function increasing both resting Ca^{2+} levels and store-operated calcium entry [101]. An additional example of a mutation shifting the ensemble relates to the Niemann-Pick disease type C, a fatal neurodegenerative disease. Its major cause is mutations in the Niemann-Pick disease, type C1 (*NPC1*) gene, which encodes a late endosomal polytopic membrane protein required for intracellular cholesterol trafficking. One prevalent mutation (I1061T) causes a folding defect, which results in failure of endosomal localization of the protein, leading to loss-of-function phenotype [102].

Structural ensemble can link genotype to phenotype through phosphorylation

Hepatitis C virus (HCV) requires only 10 proteins for evading the immune system. Phosphorylation of the intrinsically disordered domain (IDD) of nonstructural protein 5A (NS5A), which is important for replication, changes its dynamics and represents a strategy to expand the viral proteome while limiting its coding capacity. The phosphorylation site is at Thr2332, near one of its polyproline-II motifs. Phosphorylation shifts the conformational ensemble of the NS5A IDD to a state that permitted detection of the polyproline motif by using ^{15}N-, ^{13}C-based multidimensional NMR spectroscopy. Mutating Thr2332 to alanine in HCV genotype 1b reduced the steady-state level of RNA by 10-fold; this change was lethal for genotype 2a. The lethal phenotype could be rescued by changing Thr2332 to glutamic acid, a phosphomimetic substitution.

The inability to produce pT2332-NS5A caused loss of integrity of the virus-induced membranous web/replication organelle [103]. The protein kinase A (PKA)-phosphorylated form of NS5A populates a conformation distinct from that of the unphosphorylated protein. The shifted distribution of the conformational ensemble encoded by the viral genome links HCV genotype to its phenotype, in this case via PTM of intrinsically disordered viral proteins.

The complexity of the genotype–Phenotype relationship

The association between genotype and phenotype is hard to understand. A structural view can help to illuminate the genotype/phenotype landscape. We distinguish between traits expressed by the protein itself (node traits) and by its interactions (edge traits) (Fig 2). Node traits are monogenic and can be considered to lie at the "end" of the pathway. The pigments in the deer mouse provide a good example. Edge traits can be monogenic or polygenic. Self-assembly of sickle cell hemoglobin—a monogenic adaptive evolution trait—creates new edges resulting in polymers. Cancer, a growth and proliferation disease, involves driver mutations in more than one protein; thus, polygenic traits involve the cellular network (Fig 4). For example, mutations in KRas, one of the most highly oncogenic proteins, are often coupled with mutations in Ras effectors [104] or with parallel pathways (Hippo/Yap1) [105, 106]. Oncogenic Ras does not rely on epidermal growth factor receptor (EGFR) signaling—that is, on an EGFR–Ras edge. Raf mutations can work by relieving the need for Ras-driven Raf side-by-side dimerization, which is required for activation [107, 108]. p53 mutations are similarly typically coupled with additional driver mutations in other proteins [109, 110]. Node mutations, such as the pigment examples above, affect only the node; edge mutations (e.g., the disease examples above), which
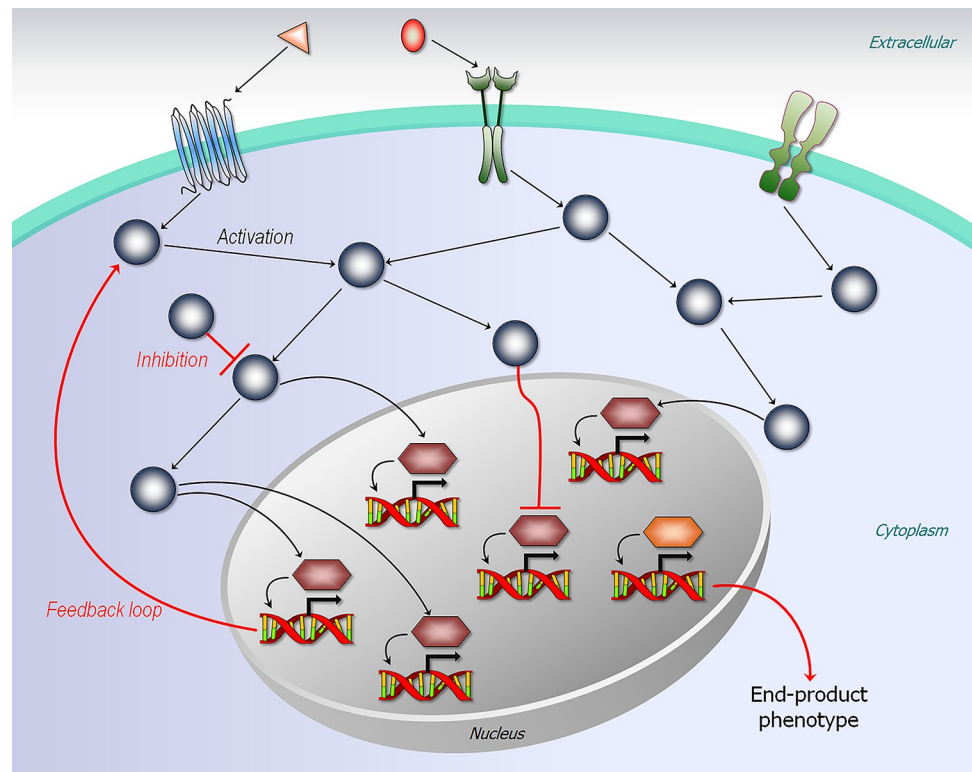

Fig 4. Cellular network. The network controls the transcriptional regulation for gene. Edge mutations can alter the cellular network, expressing the end-product phenotype.

https://doi.org/10.1371/journal.pcbi.1006648.g004

create or break interactions (Fig 2), alter the cellular network [111–116] (Fig 4). All traits—node, edge, and type I (multiple molecular functions of a gene) and the more common type II (multiple consequences of a single molecular function) pleiotropy [74]—can be understood in terms of statistics of protein ensembles (Fig 3). The shift in the populations of the sickle cell hemoglobin following the Glu → Val mutation favors polymerization, similar to the shift in Alzheimer amyloid β (Aβ) mutants, there resulting in amyloid formation [117, 118]. Likewise, edge formation/removal in oncogenic mutants arises from a redistribution of the conformational ensemble [119]. Because distinct mutations can shift the ensemble in different ways, the resulting edges can precipitate altered types of cancer by the same protein, as in the case of KRas driver mutations, which have different frequencies in distinct cancers [120]. This description holds both for adaptive evolution and for disease. The usefulness of a structural ensemble, node-versus-edge view can be seen in its drug targeting implications, as well as in prediction of phenotypes from genotypes.

Prediction of phenotype from genotype

Phenotypes may involve multiple traits that emerge from multiple nodes and edges, making it difficult to predict and relate specific nodes (proteins or genes) to specific traits [10, 13, 121]; however, networks of pairs of nodes and edges that drive phenotypes can be identified [122, 123]. To accomplish this goal, a network (or pathway)-based approach is adopted: instead of directly connecting genotype with phenotype, genotype variants are assembled into gene networks and subnetworks that are statistically connected to the phenotype are identified [10].

Cell signaling and phenotypic expression take place across time and space and are on length scales from nanometers to micrometers [124], which require consideration of how the genetic variation would affect the function in the cell. Recently, this has led to a strategy based on knowledge of cellular subsystems and their hierarchical organization as defined by the Gene Ontology (GO) or inferred from published datasets [10]. Genotype data are formulated hierarchically in terms of the consequences of the genetic variation at multiple cellular scales. The resulting "ontotype" is interpreted by logical rules. As an example, it was used to predict yeast cell growth phenotype of two new screens of double gene knockouts affecting DNA repair or nuclear lumen.

Observed conformational changes

To most clearly demonstrate our thesis, the examples should concomitantly (1) relate to signaling (i.e., pathways with nodes and edges), (2) present a clear visual phenotype (like the fur colors of deer mice or the aging of facial features), and (3) indicate a clear conformational change. We were unable to find such examples. The examples above imply conformational change; but they do not directly show it. Here, we present two examples indicating conformational change; however, they relate to pathogens—thus, no signaling pathway—and the phenotype they confer is disease. The first involves HIV-1 coreceptors (CCR5 and CXCR4) that are critical for virus tropism and pathogenesis (Fig 5). During infection, a mutation from a negatively to a positively charged residue at position 322 in the V3 loop of the HIV-1 envelope glycoprotein gp120 can accomplish a phenotype switch of R5 virus to an X4 virus, and this correlates with disease progression. The NMR structure of the V3 region of an R5 strain illustrates that positively charged and negatively charged residues at positions 304 and 322, respectively, oppose each other in the β-hairpin structure, resulting in stabilizing the R5 conformation. By contrast, in the X4 conformation, electrostatic repulsion between residues 304 and 322 induces a shift in the N-terminal strand, pointing to electrostatic interactions as modulating the conformation and thereby the phenotype switch [125]. The second example involves the classical
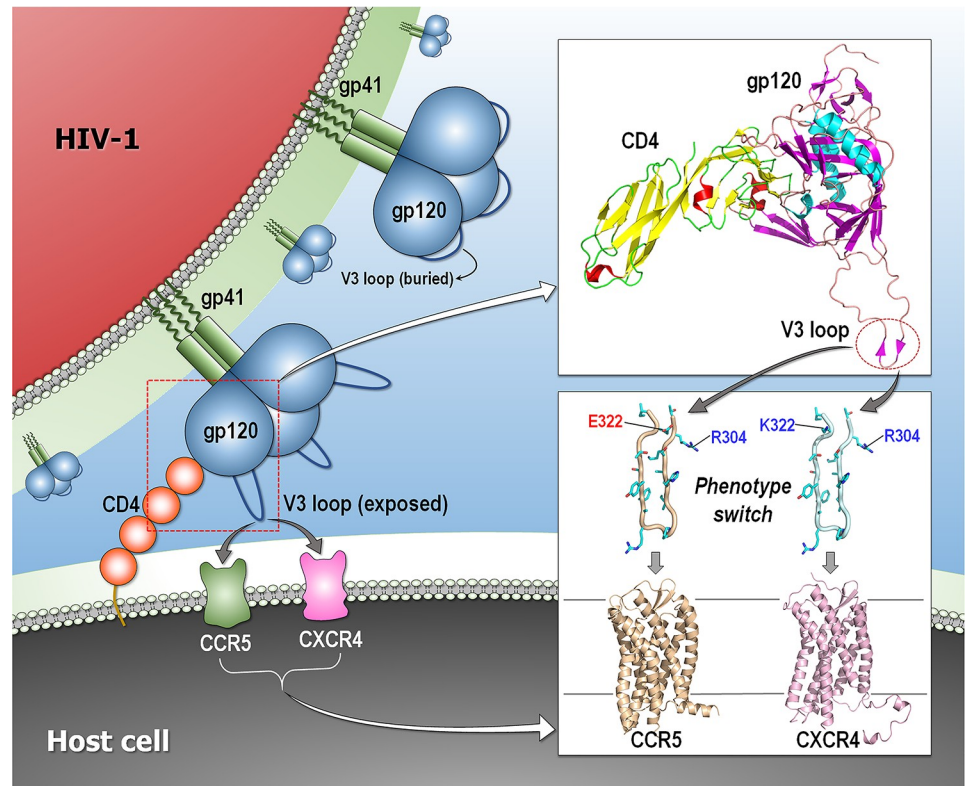
Fig 5. Phenotype switch of HIV-1 entry. Schematic diagram representing the initial process of the HIV-1 entry that led to fusion between the viral and the host cell membranes. The gp120 trimer undergoes a conformational change upon binding to cellular receptor CD4, exposing the variable loop V3. The V3 loop binds to the coreceptor (CCR5 and CXCR4), triggering the entry process. The wild-type V3 loop with positively and negatively charged residues at positions 304 and 322, respectively, recognizes CCR5. A phonotype switch by a mutation from a negatively to a positively charged residue at position 322 in the V3 loop alters coreceptor recognition to CXCR4. Modeled structures are the crystal structures of gp120 (PDB code: 24BC), CCR5 (PDB code: 4MBS), and CXCR4 (PDB code: 3ODU) and the NMR structure of V3 loop (PDB code: 2ESX). CCR5, C-C chemokine receptor type 5; CXCR4, C-X-C chemokine receptor type 4; PDB, Protein Data Bank.

phenomenon of yeast prion strain variants [126]. Infectious prion states—each of which has distinct conformation—cause distinguishable phenotypes. Even though the spectrum of conformations is identical to that of the noninfectious state, the relative populations differ. Solution NMR, amide hydrogen/deuterium (H/D) exchange, and mutagenesis of two strain conformations, termed Sc4 and Sc37, of the yeast Sup35 prion indicated an overlapping amyloid core composed of tightly packed β-sheets. This stable core is expanded in the Sc37 conformation, explaining why this strain has higher fiber stability, which impedes chaperone-mediated replication. The large conformational differences among prion strains provide a structural basis for their physiological phenotypic behavior.

Methods to associate ensembles and function

The fact that a protein-coding gene encodes not a single conformation but an ensemble of conformations is indisputable, as is the fact that this must play a role in the making of the phenotype(s). However, the questions of how to capture the multiple states and relate them to distinct functions are challenging. Until recently, X-ray crystallography could only capture a single conformation. Recently, time-resolved crystallography [127–134], diffuse X-ray scattering [135], and cryo-electron microscopy [136–138] have been shown capable of capturing

multiple conformations at high resolution along time. These can provide unparalleled insight into catalytic steps in enzyme reactions and binding/dissociations of proteins with their ligands, such as the dissociation of carbon monoxide from myoglobin. However, to date, high-resolution experimental methods linking specific macromolecular states to distinct functions are still challenging. To date, comparisons of available static crystal structures of a given protein complexed with different biological ligands relating to specific functions appear as the method of choice to attain this aim. Computations coupled with lower experimental resolutions have, however, made significant strides toward accomplishing this goal [139, 140], as well as other combined approaches or alone [141–144]. The heterogeneity of molecules is vast, the distinct mechanism of each molecule is often unknown, the sample size of each molecule is often very small, and often too these possess different PTMs. A related way to gauge the conformational diversity can be via a large-scale analysis of structures of proteins belonging to the same family [145]. The structural distribution observed in family members, which differ in sequence, is another reflection of a shift in the ensemble, as is of the proteins in different media [146]. Sequence divergence, media—water versus organic solvents—and molecular crowding are all changes in the environments, which redistribute the conformational ensemble. Conformational diversity of the native state modulates protein function [62, 147, 148] and may be revealed constructing and analyzing networks of evolutionarily coupled residues [149]. Other methods include mapping the conformational landscape of a dynamic enzyme by multi-temperature and X-ray free electron laser (XFEL) crystallography [150]; however, these methods are unable to relate specific conformers to distinct functions. Another recent approach focused on proteins that switch folds via remodeling of secondary structure in response to a few mutations (evolved fold switchers) or cellular stimuli (extant fold switchers), both allosteric events. Putative extant fold switchers with only one solved conformation were identified by incorrect secondary structure predictions and likely independent folding cooperativity, resulting in an estimate that 0.5%–4% of PDB proteins switch folds [41]. A method that identifies functionally interacting mutations in both extant and reconstructed ancestral sequences models pairwise functional dependencies and higher-order interactions that enable evolution of new protein functions. The results reveal that function-preserving mutation dependencies are frequently from structural contacts, whereas gain-of-function mutation dependencies are most commonly between residues distal in protein structure [39].

Conclusions: The significance of the conformational-level linkage

The "second molecular biology revolution" [2] calls for a new view of the genotype–phenotype dogma. The genotype does not encode only one state; it encodes ensembles of states. It is the ensembles that enable proteins to fulfill their functions; and thus, they are the ones that link genotype to phenotype. This new view of the genotype–phenotype association is not mere semantics. Its significance lies in deeper understanding of the connection of the disease phenotype with the genetic change and in providing the structural basis for disease-treating decisions [77]. Healthcare decisions are largely based on associating an observed disease phenotype with the genetic landscape, which is obtained through statistical analyses [151]. The conformational ensemble linkage can provide the mechanism of the mutation, point to cooperative "latent driver" mutations, and help in making predictions in precision medicine. Latent driver mutations behave as passengers; however, coupled with other emerging mutations, they may drive disease or phenotypic change [98]. Latent drivers are somatic mutations, emerging at any time during the individual life span. Eventually, the genetics of diseases is mediated by ensembles.

Classically, phenotype is what is observed; genotype is the genetic makeup serving as blueprint for protein expression and shape. Recently, the polycistronic nature of human genes has

also been shown to be critical to understanding the genotype–phenotype relationship, and the addition of alternative open reading frames in databases argued to help relate phenotypes and genotypes [152]. However, today we also understand that a genotype is expressed not by a single protein shape but by a large number of shapes, some of which defining altered phenotypes. Key factors include the conformational heterogeneity, the populations, and the environment. The old view of the genotype–phenotype association is unable to explain how different combinations of mutations in the same gene lead to altered phenotypic changes, as in the case of the deer mouse above. Current concepts explain that all conformations preexist, and an altered pattern of mutations shifts the equilibrium of the conformational ensemble, which now displays changed characteristics. We believe that it is time to broadly link the age-old phenotype/genotype view. We now know: their association is mediated by the statistical distribution of the ensemble of states.

So why is an ensemble view of a genotype–phenotype linkage important? Why update our perspective when current approaches have been working? First, it allows understanding of the physicochemical basis of observations. Second, the advent of the so-called precision medicine, in which treatments are envisioned to be tailored to a person's genetic profile, argues for a need to go beyond the multifactorial statistics. The free energy landscape of a protein can explain, and with time we hope quantify and predict [119, 153–155], how specific mutational combinations in an individual would alter the interactions of the protein and thus its cellular network and the prescribed drug regime. Conformational diversity of the native state can modulate protein function, with different ligands shifting the conformational equilibrium through their binding to highest-affinity conformers [147]. Conformational ensembles affect emerging functions and bear on enzyme and antibody promiscuity, signaling, protein–protein recognition, and preponderance of disease [156]. It is becoming increasingly clear that statistics based on combinations of multiple variables may not be enough to achieve this aim. Finally, third, creativity and innovation require in-depth understanding.

Acknowledgments

References

1. Lewontin RC (1985) Population genetics. Annu Rev Genet 19:81–102. https://doi.org/10.1146/annurev.ge.19.120185.000501 PMID: 3002243

2. Nussinov R, Wolynes PG (2014) A second molecular biology revolution? The energy landscapes of biomolecular function. Phys Chem Chem Phys 16(14):6321–6322. https://doi.org/10.1039/c4cp90027h PMID: 24608340

3. Frauenfelder H, Sligar SG, Wolynes PG (1991) The energy landscapes and motions of proteins. Science 254(5038):1598–1603. https://10.1126/science.1749933 PMID: 1749933

4. Tsai CJ, Kumar S, Ma B, Nussinov R (1999) Folding funnels, binding funnels, and protein function. Protein Sci 8(6):1181–1190. https://doi.org/10.1110/ps.8.6.1181 PMID: 10386868

5. Ma B, Kumar S, Tsai CJ, Nussinov R (1999) Folding funnels and binding mechanisms. Protein Eng 12 (9):713–720. https://10.1093/protein/12.9.713 PMID: 10506280

6. Tsai CJ, Ma B, Nussinov R (1999) Folding and binding cascades: shifts in energy landscapes. Proc Natl Acad Sci U S A 96(18):9970–9972. https://doi.org/10.1073/pnas.96.18.9970 PMID: 10468538

7. Kumar S, Ma B, Tsai CJ, Sinha N, Nussinov R (2000) Folding and binding cascades: dynamic landscapes and population shifts. Protein Sci 9(1):10–19. https://doi.org/10.1110/ps.9.1.10 PMID: 10739242

8. Pigliucci M (2010) Genotype-phenotype mapping and the end of the 'genes as blueprint' metaphor. Philos Trans R Soc Lond B Biol Sci 365(1540):557–566. https://doi.org/10.1098/rstb.2009.0241 PMID: 20083632

9. Dowell RD, Ryan O, Jansen A, Cheung D, Agarwala S, et al. (2010) Genotype to phenotype: a complex problem. Science 328(5977):469. https://doi.org/10.1126/science.1189015 PMID: 20413493

10. Yu MK, Kramer M, Dutkowski J, Srivas R, Licon K, et al. (2016) Translation of Genotype to Phenotype by a Hierarchy of Cell Subsystems. Cell Syst 2(2):77–88. https://doi.org/10.1016/j.cels.2016.02.003 PMID: 26949740

11. Sahni N, Yi S, Zhong Q, Jailkhani N, Charloteaux B, et al. (2013) Edgotype: a fundamental link between genotype and phenotype. Curr Opin Genet Dev 23(6):649–657. https://doi.org/10.1016/j.gde.2013.11.002 PMID: 24287335

12. Dao P, Kim YA, Wojtowicz D, Madan S, Sharan R, et al. (2017) BeWith: A Between-Within method to discover relationships between cancer modules via integrated analysis of mutual exclusivity, co-occurrence and functional interactions. PLoS Comput Biol 13(10):e1005695. https://doi.org/10.1371/journal.pcbi.1005695 PMID: 29023534

13. Kim YA, Cho DY, Przytycka TM (2016) Understanding Genotype-Phenotype Effects in Cancer via Network Approaches. PLoS Comput Biol 12(3):e1004747. https://doi.org/10.1371/journal.pcbi.1004747 PMID: 26963104

14. Magnusson R, Mariotti GP, Kopsen M, Lovfors W, Gawel DR, et al. (2017) LASSIM-A network inference toolbox for genome-wide mechanistic modeling. PLoS Comput Biol 13(6):e1005608. https://doi.org/10.1371/journal.pcbi.1005608 PMID: 28640810

15. Jang K, Kim K, Cho A, Lee I, Choi JK (2017) Network perturbation by recurrent regulatory variants in cancer. PLoS Comput Biol 13(3):e1005449. https://doi.org/10.1371/journal.pcbi.1005449 PMID: 28333928

16. Zhao C, Isenberg JS, Popel AS (2017) Transcriptional and Post-Transcriptional Regulation of Thrombospondin-1 Expression: A Computational Model. PLoS Comput Biol 13(1):e1005272. https://doi.org/10.1371/journal.pcbi.1005272 PMID: 28045898

17. Zarate YA, Smith-Hicks CL, Greene C, Abbott MA, Siu VM, et al. (2018) Natural history and genotype-phenotype correlations in 72 individuals with SATB2-associated syndrome. Am J Med Genet A 176 (4):925–935. https://doi.org/10.1002/ajmg.a.38630 PMID: 29436146

18. Sandor C, Beer NL, Webber C (2017) Diverse type 2 diabetes genetic risk factors functionally converge in a phenotype-focused gene network. PLoS Comput Biol 13(10):e1005816. https://doi.org/10.1371/journal.pcbi.1005816 PMID: 29059180

19. Orgogozo V, Morizot B, Martin A (2015) The differential view of genotype-phenotype relationships. Front Genet 6:179. https://doi.org/10.3389/fgene.2015.00179 PMID: 26042146

20. Fortuna MA, Zaman L, Ofria C, Wagner A (2017) The genotype-phenotype map of an evolving digital organism. PLoS Comput Biol 13(2):e1005414. https://doi.org/10.1371/journal.pcbi.1005414 PMID: 28241039

21. Ahnert SE (2017) Structural properties of genotype-phenotype maps. J R Soc Interface 14(132). https://doi.org/10.1098/rsif.2017.0275 PMID: 28679667

22. Wilson S, Filipp FV (2018) A network of epigenomic and transcriptional cooperation encompassing an epigenomic master regulator in cancer. NPJ Syst Biol Appl 4:24. https://doi.org/10.1038/s41540-018-0061-4 PMID: 29977600

23. Fernandez-Valverde SL, Aguilera F, Ramos-Diaz RA (2018) Inference of developmental gene regulatory networks beyond classical model systems: new approaches in the post-genomic era. Integr Comp Biol 58(4):640–653. https://doi.org/10.1093/icb/icy061 PMID: 29917089

24. Culyba MJ, Kubiak JM, Mo CY, Goulian M, Kohli RM (2018) Non-equilibrium repressor binding kinetics link DNA damage dose to transcriptional timing within the SOS gene network. PLoS Genet 14(6):e1007405. https://doi.org/10.1371/journal.pgen.1007405 PMID: 29856734

25. Mishra AK, Bernardo-Garcia FJ, Fritsch C, Humberg TH, Egger B, et al. (2018) Patterning mechanisms diversify neuroepithelial domains in the Drosophila optic placode. PLoS Genet 14(4):e1007353. https://doi.org/10.1371/journal.pgen.1007353 PMID: 29677185

26. Tan Z, Niu B, Tsang KY, Melhado IG, Ohba S, et al. (2018) Synergistic co-regulation and competition by a SOX9-GLI-FOXA phasic transcriptional network coordinate chondrocyte differentiation transitions. PLoS Genet 14(4):e1007346. https://doi.org/10.1371/journal.pgen.1007346 PMID: 29659575

27. Haque N, Ouda R, Chen C, Ozato K, Hogg JR (2018) ZFR coordinates crosstalk between RNA decay and transcription in innate immunity. Nat Commun 9(1):1145. https://doi.org/10.1038/s41467-018-03326-5 PMID: 29559679

28. Herrera-Delgado E, Perez-Carrasco R, Briscoe J, Sollich P (2018) Memory functions reveal structural properties of gene regulatory networks. PLoS Comput Biol 14(2):e1006003. https://doi.org/10.1371/journal.pcbi.1006003 PMID: 29470492

29. Meng Q, Li X, Zhu W, Yang L, Liang W, et al. (2017) Regulatory network and genetic interactions established by OsMADS34 in rice inflorescence and spikelet morphogenesis. J Integr Plant Biol 59 (9):693–707. https://doi.org/10.1111/jipb.12594 PMID: 28843032

30. Martin OC, Krzywicki A, Zagorski M (2016) Drivers of structural features in gene regulatory networks: From biophysical constraints to biological function. Phys Life Rev 17:124–158. https://doi.org/10.1016/j.plrev.2016.06.002 PMID: 27365153

31. Guseva E, Zuckermann RN, Dill KA (2017) Foldamer hypothesis for the growth and sequence differentiation of prebiotic polymers. Proc Natl Acad Sci U S A 114(36):E7460–E7468. https://doi.org/10.1073/pnas.1620179114 PMID: 28831002

32. Siltberg-Liberles J, Grahnen JA, Liberles DA (2011) The evolution of protein structures and structural ensembles under functional constraint. Genes (Basel) 2(4):748–762. https://doi.org/10.3390/genes2040748 PMID: 24710290

33. Sikosek T, Krobath H, Chan HS (2016) Theoretical Insights into the Biophysics of Protein Bi-stability and Evolutionary Switches. PLoS Comput Biol 12(6):e1004960. https://doi.org/10.1371/journal.pcbi.1004960 PMID: 27253392

34. Sikosek T, Chan HS (2014) Biophysics of protein evolution and evolutionary protein biophysics. J R Soc Interface 11(100):20140419. https://doi.org/10.1098/rsif.2014.0419 PMID: 25165599

35. Wroe R, Bornberg-Bauer E, Chan HS (2005) Comparing folding codes in simple heteropolymer models of protein evolutionary landscape: robustness of the superfunnel paradigm. Biophys J 88(1):118–131. https://doi.org/10.1529/biophysj.104.050369 PMID: 15501948

36. Nussinov R, Zhang M, Tsai CJ, Liao TJ, Fushman D, et al. (2018) Autoinhibition in Ras effectors Raf, PI3Kalpha, and RASSF5: a comprehensive review underscoring the challenges in pharmacological intervention. Biophys Rev 10(5):1263–1282. https://doi.org/10.1007/s12551-018-0461-0 PMID: 30269291

37. Modi T, Huihui J, Ghosh K, Ozkan SB (2018) Ancient thioredoxins evolved to modern-day stability-function requirement by altering native state ensemble. Philos Trans R Soc Lond B Biol Sci 373 (1749). https://doi.org/10.1098/rstb.2017.0184 PMID: 29735738

38. Kumar A, Butler BM, Kumar S, Ozkan SB (2015) Integration of structural dynamics and molecular evolution via protein interaction networks: a new era in genomic medicine. Curr Opin Struct Biol 35:135–142. https://doi.org/10.1016/j.sbi.2015.11.002 PMID: 26684487

39. Beleva Guthrie V, Masica DL, Fraser A, Federico J, Fan Y, et al. (2018) Network Analysis of Protein Adaptation: Modeling the Functional Impact of Multiple Mutations. Mol Biol Evol 35(6):1507–1519. https://doi.org/10.1093/molbev/msy036 PMID: 29522102

40. Lai JI, Verma D, Bailey-Kellogg C, Ackerman ME (2018) Towards conformational fidelity of a quaternary HIV-1 epitope: computational design and directed evolution of a minimal V1V2 antigen. Protein Eng Des Sel 31(4):121–133. https://doi.org/10.1093/protein/gzy010 PMID: 29897567

41. Porter LL, Looger LL (2018) Extant fold-switching proteins are widespread. Proc Natl Acad Sci U S A 115(23):5968–5973. https://doi.org/10.1073/pnas.1800168115 PMID: 29784778

42. Gilson AI, Marshall-Christensen A, Choi JM, Shakhnovich EI (2017) The Role of Evolutionary Selection in the Dynamics of Protein Structure Evolution. Biophys J 112(7):1350–1365. https://doi.org/10.1016/j.bpj.2017.02.029 PMID: 28402878

43. Jacobs WM, Shakhnovich EI (2017) Evidence of evolutionary selection for cotranslational folding. Proc Natl Acad Sci U S A 114(43):11434–11439. https://doi.org/10.1073/pnas.1705772114 PMID: 29073068

44. Klein EY, Blumenkrantz D, Serohijos A, Shakhnovich E, Choi JM, et al. (2018) Stability of the Influenza Virus Hemagglutinin Protein Correlates with Evolutionary Dynamics. mSphere 3(1):e00554. https://doi.org/10.1128/mSphereDirect.00554-17 PMID: 29299534

45. Rotem A, Serohijos AWR, Chang CB, Wolfe JT, Fischer AE, et al. (2018) Evolution on the biophysical fitness landscape of an RNA virus. Mol Biol Evol 35(10):2390–2400. https://doi.org/10.1093/molbev/msy131 PMID: 29955873

46. Razban RM, Gilson AI, Durfee N, Strobelt H, Dinkla K, et al. (2018) ProteomeVis: a web app for exploration of protein properties from structure to sequence evolution across organisms' proteomes. Bioinformatics 34(20):3557–3565. https://doi.org/10.1093/bioinformatics/bty370 PMID: 29741573

47. Granata D, Ponzoni L, Micheletti C, Carnevale V (2017) Patterns of coevolving amino acids unveil structural and dynamical domains. Proc Natl Acad Sci U S A 114(50):E10612–E10621. https://doi.org/10.1073/pnas.1712021114 PMID: 29183970

48. Dong Z, Zhou H, Tao P (2018) Combining protein sequence, structure, and dynamics: A novel approach for functional evolution analysis of PAS domain superfamily. Protein Sci 27(2):421–430. https://doi.org/10.1002/pro.3329 PMID: 29052279

49. Sailer ZR, Harms MJ (2017) Molecular ensembles make evolution unpredictable. Proc Natl Acad Sci U S A 114(45):11938–11943. https://doi.org/10.1073/pnas.1711927114 PMID: 29078365

50. Rees M, Jessica C, Metcalf E, Childs DZ (2010) Bet-hedging as an evolutionary game: the trade-off between egg size and number. Proc Biol Sci 277(1685):1149–1151. https://10.1098/rspb.2009.1541 https://doi.org/10.1098/rspb.2009.1541 PMID: 20018792

51. Levis NA, Pfennig DW (2018) Phenotypic plasticity, canalization, and the origins of novelty: evidence and mechanisms from amphibians. Semin Cell Dev Biol 88:80–90. https://doi.org/10.1016/j.semcdb.2018.01.012 PMID: 29408711

52. Javier Zea D, Miguel Monzon A, Fornasari MS, Marino-Buslje C, Parisi G (2013) Protein conformational diversity correlates with evolutionary rate. Mol Biol Evol 30(7):1500–1503. https://doi.org/10.1093/molbev/mst065 PMID: 23564939

53. Jolly MK, Kulkarni P, Weninger K, Orban J, Levine H (2018) Phenotypic Plasticity, Bet-Hedging, and Androgen Independence in Prostate Cancer: Role of Non-Genetic Heterogeneity. Front Oncol 8:50. https://doi.org/10.3389/fonc.2018.00050 PMID: 29560343

54. Shanthirabalan S, Chomilier J, Carpentier M (2018) Structural effects of point mutations in proteins. Proteins 86(8):853–867. https://doi.org/10.1002/prot.25499 PMID: 29569365

55. Bershtein S, Serohijos AW, Shakhnovich EI (2017) Bridging the physical scales in evolutionary biology: from protein sequence space to fitness of organisms and populations. Curr Opin Struct Biol 42:31–40. https://doi.org/10.1016/j.sbi.2016.10.013 PMID: 27810574

56. Tyzack JD, Furnham N, Sillitoe I, Orengo CM, Thornton JM (2017) Understanding enzyme function evolution from a computational perspective. Curr Opin Struct Biol 47:131–139. https://doi.org/10.1016/j.sbi.2017.08.003 PMID: 28892668

57. Meng Y, Gao C, Clawson DK, Atwell S, Russell M, et al. (2018) Predicting the Conformational Variability of Abl Tyrosine Kinase using Molecular Dynamics Simulations and Markov State Models. J Chem Theory Comput 14(5):2721–2732. https://doi.org/10.1021/acs.jctc.7b01170 PMID: 29474075

58. Razavi AM, Khelashvili G, Weinstein H (2018) How structural elements evolving from bacterial to human SLC6 transporters enabled new functional properties. BMC Biol 16(1):31. https://doi.org/10.1186/s12915-018-0495-6 PMID: 29540172

59. Murphy GS, Greisman JB, Hecht MH (2016) De Novo Proteins with Life-Sustaining Functions Are Structurally Dynamic. J Mol Biol 428(2 Pt A):399–411. https://doi.org/10.1016/j.jmb.2015.12.008 PMID: 26707197

60. Okafor CD, Pathak MC, Fagan CE, Bauer NC, Cole MF, et al. (2018) Structural and Dynamics Comparison of Thermostability in Ancient, Modern, and Consensus Elongation Factor Tus. Structure 26 (1):118–129. https://doi.org/10.1016/j.str.2017.11.018 PMID: 29276038

61. Jiang X, Chen G, Wang L (2016) Structural and dynamic evolution of the amphipathic N-terminus diversifies enzyme thermostability in the glycoside hydrolase family 12. Phys Chem Chem Phys 18 (31):21340–21350. https://doi.org/10.1039/c6cp02998a PMID: 27425569

62. Balu R, Mata JP, Knott R, Elvin CM, Hill AJ, et al. (2016) Effects of Crowding and Environment on the Evolution of Conformational Ensembles of the Multi-Stimuli-Responsive Intrinsically Disordered Protein, Rec1-Resilin: A Small-Angle Scattering Investigation. J Phys Chem B 120(27):6490–6503. https://doi.org/10.1021/acs.jpcb.6b02475 PMID: 27281267

63. Ahrens JB, Nunez-Castilla J, Siltberg-Liberles J (2017) Evolution of intrinsic disorder in eukaryotic proteins. Cell Mol Life Sci 74(17):3163–3174. https://doi.org/10.1007/s00018-017-2559-0 PMID: 28597295

64. Fuxreiter M (2018) Fuzziness in Protein Interactions-A Historical Perspective. J Mol Biol 430 (16):2278–2287. https://doi.org/10.1016/j.jmb.2018.02.015 PMID: 29477337

65. Ozturk MA, Cojocaru V, Wade RC (2018) Dependence of Chromatosome Structure on Linker Histone Sequence and Posttranslational Modification. Biophys J 114(10):2363–2375. https://doi.org/10.1016/j.bpj.2018.04.034 PMID: 29759374

66. Vaidyanathan K, Niranjan T, Selvan N, Teo CF, May M, et al. (2017) Identification and characterization of a missense mutation in the O-linked beta-N-acetylglucosamine (O-GlcNAc) transferase gene that segregates with X-linked intellectual disability. J Biol Chem 292(21):8948–8963. https://doi.org/10.1074/jbc.M116.771030 PMID: 28302723

67. Stetz G, Verkhivker GM (2017) Computational Analysis of Residue Interaction Networks and Coevolutionary Relationships in the Hsp70 Chaperones: A Community-Hopping Model of Allosteric Regulation

and Communication. PLoS Comput Biol 13(1):e1005299. https://doi.org/10.1371/journal.pcbi.1005299 PMID: 28095400

68. Shaytan AK, Landsman D, Panchenko AR (2015) Nucleosome adaptability conferred by sequence and structural variations in histone H2A-H2B dimers. Curr Opin Struct Biol 32:48–57. https://doi.org/10.1016/j.sbi.2015.02.004 PMID: 25731851

69. Guo Z, Thorarensen A, Che J, Xing L (2016) Target the More Druggable Protein States in a Highly Dynamic Protein—Protein Interaction System. J Chem Inf Model 56(1):35–45. https://doi.org/10.1021/acs.jcim.5b00503 PMID: 26650754

70. Johansson KE, Lindorff-Larsen K (2018) Structural heterogeneity and dynamics in protein evolution and design. Curr Opin Struct Biol 48:157–163. https://doi.org/10.1016/j.sbi.2018.01.010 PMID: 29413956

71. Maria-Solano MA, Serrano-Hervas E, Romero-Rivera A, Iglesias-Fernandez J, Osuna S (2018) Role of conformational dynamics in the evolution of novel enzyme function. Chem Commun (Camb) 54 (50):6622–6634. https://doi.org/10.1039/c8cc02426j PMID: 29780987

72. Raman S (2018) Systems Approaches to Understanding and Designing Allosteric Proteins. Biochemistry 57(4):376–382. https://doi.org/10.1021/acs.biochem.7b01094 PMID: 29235352

73. Solovieff N, Cotsapas C, Lee PH, Purcell SM, Smoller JW (2013) Pleiotropy in complex traits: challenges and strategies. Nat Rev Genet 14(7):483–495. https://doi.org/10.1038/nrg3461 PMID: 23752797

74. Wagner GP, Zhang J (2011) The pleiotropic structure of the genotype-phenotype map: the evolvability of complex organisms. Nat Rev Genet 12(3):204–213. https://doi.org/10.1038/nrg2949 PMID: 21331091

75. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, et al. (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet 9(5):356–369. https://doi.org/10.1038/nrg2344 PMID: 18398418

76. Tsai CJ, Ma BY, Nussinov R (2009) Protein-protein interaction networks: how can a hub protein bind so many different partners? Trends Biochem Sci 34(12):594–600. https://doi.org/10.1016/j.tibs.2009.07.007 PMID: 19837592

77. Nussinov R, Jang H, Tsai CJ (2014) The structural basis for cancer treatment decisions. Oncotarget 5 (17):7285–7302. https://doi.org/10.18632/oncotarget.2439 PMID: 25277176

78. James LC, Tawfik DS (2003) Conformational diversity and protein evolution—a 60-year-old hypothesis revisited. Trends Biochem Sci 28(7):361–368. https://doi.org/10.1016/S0968-0004(03)00135-X PMID: 12878003

79. Chen T, Vernazobres D, Yomo T, Bornberg-Bauer E, Chan HS (2010) Evolvability and single-genotype fluctuation in phenotypic properties: a simple heteropolymer model. Biophys J 98(11):2487–2496. https://doi.org/10.1016/j.bpj.2010.02.046 PMID: 20513392

80. Dong Y, Cao L, Tang H, Shi X, He Y (2017) Structure of Human M-type Phospholipase A2 Receptor Revealed by Cryo-Electron Microscopy. J Mol Biol 429(24):3825–3835. https://doi.org/10.1016/j.jmb.2017.10.019 PMID: 29079480

81. Xiao J, Melvin RL, Salsbury FR (2017) Mechanistic insights into thrombin's switch between "slow" and "fast" forms. Phys Chem Chem Phys 19(36):24522–24533. https://doi.org/10.1039/c7cp03671j PMID: 28849814

82. Gunasekaran K, Ma B, Nussinov R (2004) Is allostery an intrinsic property of all dynamic proteins? Proteins 57(3):433–443. https://doi.org/10.1002/prot.20232 PMID: 15382234

83. Nussinov R, Ma B, Tsai CJ, Csermely P (2013) Allosteric conformational barcodes direct signaling in the cell. Structure 21(9):1509–1521. https://doi.org/10.1016/j.str.2013.06.002 PMID: 24010710

84. del Sol A, Tsai CJ, Ma B, Nussinov R (2009) The origin of allosteric functional modulation: multiple pre-existing pathways. Structure 17(8):1042–1050. https://doi.org/10.1016/j.str.2009.06.008 PMID: 19679084

85. Tsai CJ, Nussinov R (2011) Gene-specific transcription activation via long-range allosteric shape-shifting. Biochem J 439(1):15–25. https://doi.org/10.1042/BJ20110972 PMID: 21916844

86. Nussinov R, Tsai CJ, Csermely P (2011) Allo-network drugs: harnessing allostery in cellular networks. Trends Pharmacol Sci 32(12):686–693. https://doi.org/10.1016/j.tips.2011.08.004 PMID: 21925743

87. Christopoulos A (2014) Advances in G protein-coupled receptor allostery: from function to structure. Mol Pharmacol 86(5):463–478. https://doi.org/10.1124/mol.114.094342 PMID: 25061106

88. Yang G-L, Fu D-L, Lang X, Wang Y-T, Cheng S-R, et al. (2013) Mutations in MC1R Gene Determine Black Coat Color Phenotype in Chinese Sheep. Sci World J 2013:675382. https://doi.org/10.1155/2013/675382 PMID: 24082855

89. Linnen CR, Poh Y-P, Peterson BK, Barrett RDH, Larson JG, et al. (2013) Adaptive Evolution of Multiple Traits Through Multiple Mutations at a Single Gene. Science 339(6125):1312–1316. https://doi.org/10.1126/science.1233213 PMID: 23493712

90. Nussinov R, Tsai CJ, Xin F, Radivojac P (2012) Allosteric post-translational modification codes. Trends Biochem Sci 37(10):447–455. https://doi.org/10.1016/j.tibs.2012.07.001 PMID: 22884395

91. Liu F, Hamer MA, Deelen J, Lall JS, Jacobs L, et al. (2016) The MC1R Gene and Youthful Looks. Curr Biol 26(9):1213–1220. https://doi.org/10.1016/j.cub.2016.03.008 PMID: 27133870

92. Valverde P, Healy E, Jackson I, Rees JL, Thody AJ (1995) Variants of the melanocyte-stimulating hormone receptor gene are associated with red hair and fair skin in humans. Nat Genet 11(3):328–330. https://doi.org/10.1038/ng1195-328 PMID: 7581459

93. Liu F, Struchalin MV, Duijn K, Hofman A, Uitterlinden AG, et al. (2011) Detecting low frequent loss-of-function alleles in genome wide association studies with red hair color as example. PLoS One 6(11): e28145. https://doi.org/10.1371/journal.pone.0028145 PMID: 22140526

94. Jacobs LC, Hamer MA, Gunn DA, Deelen J, Lall JS, et al. (2015) A Genome-Wide Association Study Identifies the Skin Color Genes IRF4, MC1R, ASIP, and BNC2 Influencing Facial Pigmented Spots. J Invest Dermatol 135(7):1735–1742. https://doi.org/10.1038/jid.2015.62 PMID: 25705849

95. Lopes LR, Rahman MS, Elliott PM (2013) A systematic review and meta-analysis of genotype-phenotype associations in patients with hypertrophic cardiomyopathy caused by sarcomeric protein mutations. Heart 99(24):1800–1811. https://doi.org/10.1136/heartjnl-2013-303939 PMID: 23674365

96. Saposnik B, Binard S, Fenneteau O, Nurden A, Nurden P, et al. (2014) Mutation spectrum and genotype-phenotype correlations in a large French cohort of MYH9-Related Disorders. Mol Genet Genomic Med 2(4):297–312. https://doi.org/10.1002/mgg3.68 PMID: 25077172

97. Wang X, Wei X, Thijssen B, Das J, Lipkin SM, et al. (2012) Three-dimensional reconstruction of protein networks provides insight into human genetic disease. Nat Biotechnol 30(2):159–164. https://doi.org/10.1038/nbt.2106 PMID: 22252508

98. Nussinov R, Tsai CJ (2015) 'Latent drivers' expand the cancer mutational landscape. Curr Opin Struct Biol 32(0):25–32. https://doi.org/10.1016/j.sbi.2015.01.004 PMID: 25661093

99. Oliver KH, Jessen T, Crawford EL, Chung CY, Sutcliffe JS, et al. (2014) Pro32Pro33 Mutations in the Integrin beta(3) PSI Domain Result in alpha IIb beta(3) Priming and Enhanced Adhesion: Reversal of the Hypercoagulability Phenotype by the Src Inhibitor SKI-606. Mol Pharmacol 85(6):921–931. https://doi.org/10.1124/mol.114.091736 PMID: 24695082

100. Jallu V, Poulain P, Fuchs PFJ, Kaplan C, de Brevern AG (2012) Modeling and Molecular Dynamics of HPA-1a and-1b Polymorphisms: Effects on the Structure of the beta 3 Subunit of the alpha IIb beta 3 Integrin. Plos One 7(11):e47304. https://doi.org/10.1371/journal.pone.0047304 PMID: 23155369

101. Morin G, Bruechle NO, Singh AR, Knopp C, Jedraszak G, et al. (2014) Gain-of-Function Mutation in STIM1 (P.R304W) Is Associated with Stormorken Syndrome. Hum Mutat 35(10):1221–1232. https://doi.org/10.1002/humu.22621 PMID: 25044882

102. Ohgane K, Karaki F, Noguchi-Yachide T, Dodo K, Hashimoto Y (2014) Structure-activity relationships of oxysterol-derived pharmacological chaperones for Niemann-Pick type C1 protein. Bioorganic Med Chem Lett 24(15):3480–3485. https://doi.org/10.1016/j.bmcl.2014.05.064 PMID: 24928400

103. Cordek DG, Croom-Perez TJ, Hwang J, Hargittai MRS, Subba-Reddy CV, et al. (2014) Expanding the Proteome of an RNA Virus by Phosphorylation of an Intrinsically Disordered Viral Protein. J Biol Chem 289(35):24397–24416. https://doi.org/10.1074/jbc.M114.589911 PMID: 25031324

104. Thevakumaran N, Lavoie H, Critton DA, Tebben A, Marinier A, et al. (2015) Crystal structure of a BRAF kinase domain monomer explains basis for allosteric regulation. Nat Struct Mol Biol 22(1):37–43. https://doi.org/10.1038/nsmb.2924 PMID: 25437913

105. Nussinov R, Tsai CJ, Jang H, Korcsmaros T, Csermely P (2016) Oncogenic KRAS signaling and YAP1/beta-catenin: Similar cell cycle control in tumor initiation. Semin Cell Dev Biol 58:79–85. https://doi.org/10.1016/j.semcdb.2016.04.001 PMID: 27058752

106. Nussinov R, Tsai CJ, Jang H (2017) A New View of Pathway-Driven Drug Resistance in Tumor Proliferation. Trends Pharmacol Sci 38(5):427–437. https://doi.org/10.1016/j.tips.2017.02.001 PMID: 28245913

107. Lavoie H, Therrien M (2011) CANCER A drug-resistant duo. Nature 480(7377):329–330. https://doi.org/10.1038/480329a PMID: 22170679

108. Poulikakos PI, Persaud Y, Janakiraman M, Kong XJ, Ng C, et al. (2011) RAF inhibitor resistance is mediated by dimerization of aberrantly spliced BRAF(V600E). Nature 480(7377):387–390. https://doi.org/10.1038/nature10662 PMID: 22113612

109. Abraham AG O'Neill E (2014) PI3K/Akt-mediated regulation of p53 in cancer. Biochem Soc Trans 42(4):798–803. https://doi.org/10.1042/BST20140070 PMID: 25109960

110. Meng X, Franklin DA, Dong J, Zhang Y (2014) MDM2-p53 Pathway in Hepatocellular Carcinoma. Cancer Res 74(24):7161–7167. https://doi.org/10.1158/0008-5472.CAN-14-1446 PMID: 25477334

111. Trinh HC, Kwon YK (2016) Edge-based sensitivity analysis of signaling networks by using Boolean dynamics. Bioinformatics 32(17):i763–i771. https://doi.org/10.1093/bioinformatics/btw464 PMID: 27587699

112. Nikolaev SI, Vetiska S, Bonilla X, Boudreau E, Jauhiainen S, et al. (2018) Somatic Activating KRAS Mutations in Arteriovenous Malformations of the Brain. N Engl J Med 378(3):250–261. https://doi.org/10.1056/NEJMoa1709449 PMID: 29298116

113. Zhong Q, Simonis N, Li QR, Charloteaux B, Heuze F, et al. (2009) Edgetic perturbation models of human inherited disorders. Mol Syst Biol 5:321. https://doi.org/10.1038/msb.2009.80 PMID: 19888216

114. AlQuraishi M, Koytiger G, Jenney A, MacBeath G, Sorger PK (2014) A multiscale statistical mechanical framework integrates biophysical and genomic data to assemble cancer networks. Nat Genet 46 (12):1363–1371. https://doi.org/10.1038/ng.3138 PMID: 25362484

115. Nishi H, Tyagi M, Teng S, Shoemaker BA, Hashimoto K, et al. (2013) Cancer missense mutations alter binding properties of proteins and their interaction networks. PLoS One 8(6):e66273. https://doi.org/10.1371/journal.pone.0066273 PMID: 23799087

116. Guven Maiorov E, Keskin O, Gursoy A, Nussinov R (2013) The structural network of inflammation and cancer: merits and challenges. Semin Cancer Biol 23(4):243–251. https://doi.org/10.1016/j.semcancer.2013.05.003 PMID: 23712403

117. Ma B, Nussinov R (2012) Selective Molecular Recognition in Amyloid Growth and Transmission and Cross-Species Barriers. J Mol Biol 421(2–3):172–184. https://doi.org/10.1016/j.jmb.2011.11.023 PMID: 22119878

118. Miller Y, Ma B, Nussinov R (2010) Polymorphism in Alzheimer Aβ Amyloid Organization Reflects Conformational Selection in a Rugged Energy Landscape. Chem Rev 110(8):4820–4838. https://doi.org/10.1021/cr900377t PMID: 20402519

119. Tsai CJ, Nussinov R (2014) The free energy landscape in translational science: how can somatic mutations result in constitutive oncogenic activation? Phys Chem Chem Phys 16(14):6332–6341. https://doi.org/10.1039/c3cp54253j PMID: 24445437

120. Prior IA, Lewis PD, Mattos C (2012) A comprehensive survey of Ras mutations in cancer. Cancer Res 72(10):2457–2467. https://doi.org/10.1158/0008-5472.CAN-11-2612 PMID: 22589270

121. Zuk O, Hechter E, Sunyaev SR, Lander ES (2012) The mystery of missing heritability: Genetic interactions create phantom heritability. Proc Natl Acad Sci U S A 109(4):1193–1198. https://doi.org/10.1073/pnas.1119675109 PMID: 22223662

122. Califano A, Butte AJ, Friend S, Ideker T, Schadt E (2012) Leveraging models of cell regulation and GWAS data in integrative network-based association studies. Nat Genet 44(8):841–847. https://doi.org/10.1038/ng.2355 PMID: 22836096

123. Greene CS, Krishnan A, Wong AK, Ricciotti E, Zelaya RA, et al. (2015) Understanding multicellular function and disease with human tissue-specific networks. Nat Genet 47(6):569–576. https://doi.org/10.1038/ng.3259 PMID: 25915600

124. Nussinov R (2013) The spatial structure of cell signaling systems. Phys Biol 10(4):045004. https://doi.org/10.1088/1478-3975/10/4/045004 PMID: 23913102

125. Rosen O, Sharon M, Quadt-Akabayov SR, Anglister J (2006) Molecular switch for alternative conformations of the HIV-1 V3 region: implications for phenotype conversion. Proc Natl Acad Sci U S A 103 (38):13950–13955. https://doi.org/10.1073/pnas.0606312103 PMID: 16966601

126. Toyama BH, Kelly MJ, Gross JD, Weissman JS (2007) The structural basis of yeast prion strain variants. Nature 449(7159):233–237. https://doi.org/10.1038/nature06108 PMID: 17767153

127. Olmos JL Jr., Pandey S, Martin-Garcia JM, Calvey G, Katz A, et al. (2018) Enzyme intermediates captured "on the fly" by mix-and-inject serial crystallography. BMC Biol 16(1):59. https://doi.org/10.1186/s12915-018-0524-5 PMID: 29848358

128. Tenboer J, Basu S, Zatsepin N, Pande K, Milathianaki D, et al. (2014) Time-resolved serial crystallography captures high-resolution intermediates of photoactive yellow protein. Science 346(6214):1242–1246. https://doi.org/10.1126/science.1259357 PMID: 25477465

129. Schmidt M, Srajer V, Henning R, Ihee H, Purwar N, et al. (2013) Protein energy landscapes determined by five-dimensional crystallography. Acta Crystallogr D Biol Crystallogr 69(Pt 12):2534–2542. https://doi.org/10.1107/S0907444913025997 PMID: 24311594

130. Rt Aranda, Levin EJ, Schotte F, Anfinrud PA, Phillips GN Jr. (2006) Time-dependent atomic coordinates for the dissociation of carbon monoxide from myoglobin. Acta Crystallogr D Biol Crystallogr 62 (Pt 7):776–783. https://doi.org/10.1107/S0907444906017318 PMID: 16790933

131. Schmidt M (2017) Time-Resolved Macromolecular Crystallography at Modern X-Ray Sources. Methods Mol Biol 1607:273–294. https://doi.org/10.1007/978-1-4939-7000-1_11 PMID: 28573577

132. Pawate AS, Srajer V, Schieferstein J, Guha S, Henning R, et al. (2015) Towards time-resolved serial crystallography in a microfluidic device. Acta Crystallogr F Struct Biol Commun 71(Pt 7):823–830. https://doi.org/10.1107/S2053230X15009061 PMID: 26144226

133. Moffat K (2014) Time-resolved crystallography and protein design: signalling photoreceptors and optogenetics. Philos Trans R Soc Lond B Biol Sci 369(1647):20130568. https://doi.org/10.1098/rstb.2013.0568 PMID: 24914168

134. Orville AM (2018) Entering an era of dynamic structural biology. BMC Biol 16(1):55. https://doi.org/10.1186/s12915-018-0533-4 PMID: 29848334

135. Wall ME, Wolff AM, Fraser JS (2018) Bringing diffuse X-ray scattering into focus. Curr Opin Struct Biol 50:109–116. https://doi.org/10.1016/j.sbi.2018.01.009 PMID: 29455056

136. Kaledhonkar S, Fu Z, White H, Frank J (2018) Time-Resolved Cryo-electron Microscopy Using a Microfluidic Chip. Methods Mol Biol 1764:59–71. https://doi.org/10.1007/978-1-4939-7759-8_4 PMID: 29605908

137. Komar AA (2018) Unraveling co-translational protein folding: Concepts and methods. Methods 137:71–81. https://doi.org/10.1016/j.ymeth.2017.11.007 PMID: 29221924

138. Fu Z, Kaledhonkar S, Borg A, Sun M, Chen B, et al. (2016) Key Intermediates in Ribosome Recycling Visualized by Time-Resolved Cryoelectron Microscopy. Structure 24(12):2092–2101. https://doi.org/10.1016/j.str.2016.09.014 PMID: 27818103

139. Li Y, Zhang Y, Grosseruschkamp F, Stephan S, Cui Q, et al. (2018) Specific Substates of Ras To Interact with GAPs and Effectors: Revealed by Theoretical Simulations and FTIR Experiments. J Phys Chem Lett 9(6):1312–1317. https://doi.org/10.1021/acs.jpclett.8b00342 PMID: 29488771

140. Ignatov M, Kazennov A, Kozakov D (2018) ClusPro FMFT-SAXS: Ultra-fast Filtering Using Small-Angle X-ray Scattering Data in Protein Docking. J Mol Biol 430(15):2249–2255. https://doi.org/10.1016/j.jmb.2018.03.010 PMID: 29626538

141. Mickert MJ, Gorris HH (2018) Transition-State Ensembles Navigate the Pathways of Enzyme Catalysis. J Phys Chem B 122(22):5809–5819. https://doi.org/10.1021/acs.jpcb.8b02297 PMID: 29742901

142. Romero-Rivera A, Garcia-Borras M, Osuna S (2017) Role of Conformational Dynamics in the Evolution of Retro-Aldolase Activity. ACS Catal 7(12):8524–8532. https://doi.org/10.1021/acscatal.7b02954 PMID: 29226011

143. Pandini A, Morcos F, Khan S (2016) The Gearbox of the Bacterial Flagellar Motor Switch. Structure 24(7):1209–1220. https://doi.org/10.1016/j.str.2016.05.012 PMID: 27345932

144. Sutto L, Marsili S, Valencia A, Gervasio FL (2015) From residue coevolution to protein conformational ensembles and functional dynamics. Proc Natl Acad Sci U S A 112(44):13567–13572. https://doi.org/10.1073/pnas.1508584112 PMID: 26487681

145. Zea DJ, Monzon AM, Parisi G, Marino-Buslje C (2018) How is structural divergence related to evolutionary information? Mol Phylogenet Evol 127:859–866. https://doi.org/10.1016/j.ympev.2018.06.033 PMID: 29953938

146. Rueda AJV, Monzon AM, Ardanaz SM, Iglesias LE, Parisi G (2018) Large scale analysis of protein conformational transitions from aqueous to non-aqueous media. BMC Bioinformatics 19(1):27. https://doi.org/10.1186/s12859-018-2044-2 PMID: 29382320

147. Saldano TE, Monzon AM, Parisi G, Fernandez-Alberti S (2016) Evolutionary Conserved Positions Define Protein Conformational Diversity. PLoS Comput Biol 12(3):e1004775. https://doi.org/10.1371/journal.pcbi.1004775 PMID: 27008419

148. Kalas V, Pinkner JS, Hannan TJ, Hibbing ME, Dodson KW, et al. (2017) Evolutionary fine-tuning of conformational ensembles in FimH during host-pathogen interactions. Sci Adv 3(2):e1601944. https://doi.org/10.1126/sciadv.1601944 PMID: 28246638

149. Jeon J, Nam HJ, Choi YS, Yang JS, Hwang J, et al. (2011) Molecular evolution of protein conformational changes revealed by a network of evolutionarily coupled residues. Mol Biol Evol 28(9):2675–2685. https://doi.org/10.1093/molbev/msr094 PMID: 21470969

150. Keedy DA, Kenner LR, Warkentin M, Woldeyes RA, Hopkins JB, et al. (2015) Mapping the conformational landscape of a dynamic enzyme by multitemperature and XFEL crystallography. Elife 4:e07574. https://doi.org/10.7554/eLife.07574 PMID: 26422513

151. Dancey JE, Bedard PL, Onetto N, Hudson TJ (2012) The Genetic Basis for Cancer Treatment Decisions. Cell 148(3):409–420. https://doi.org/10.1016/j.cell.2012.01.014 PMID: 22304912

152. Brunet MA, Levesque SA, Hunting DJ, Cohen AA, Roucou X (2018) Recognition of the polycistronic nature of human genes is critical to understanding the genotype-phenotype relationship. Genome Res 28(5):609–624. https://doi.org/10.1101/gr.230938.117 PMID: 29626081

153. Nussinov R, Tsai CJ, Ma B (2013) The Underappreciated Role of Allostery in the Cellular Network. Ann Rev Biophys 42:169–189. https://doi.org/10.1146/annurev-biophys-083012-130257 PMID: 23451894

154. Jenik M, Parra RG, Radusky LG, Turjanski A, Wolynes PG, et al. (2012) Protein frustratometer: a tool to localize energetic frustration in protein molecules. Nucleic Acids Res 40(W1):W348–W351. https://doi.org/10.1093/nar/gks447 PMID: 22645321

155. Nussinov R, Tsai CJ (2014) Free Energy Diagrams for Protein Function. Chem Biol 21:311–318. https://doi.org/10.1016/j.chembiol.2013.12.015 PMID: 24508196

156. Parisi G, Zea DJ, Monzon AM, Marino-Buslje C (2015) Conformational diversity and the emergence of sequence signatures during evolution. Curr Opin Struct Biol 32:58–65. https://doi.org/10.1016/j.sbi.2015.02.005 PMID: 25749052