Published online 18 January 2023

NAR Cancer, 2023, Vol. 5, No. 1 **1**
https://doi.org/10.1093/narcan/zcac040

Systemic structural analysis of alterations reveals a common structural basis of driver mutations in cancer

Tomer Meirson [1,*], David Bomze[2], Ora Schueler-Furman[3], Salomon M. Stemmer[1,2] and Gal Markel[1,4,*]

[1]Davidoff Cancer Center, Rabin Medical Center-Beilinson Hospital, Petah Tikva, 49100, Israel, [2]Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, 6997801, Israel, [3]Department of Microbiology and Molecular Genetics, Institute for Biomedical Research Israel-Canada, Faculty of Medicine, The Hebrew University of Jerusalem, Jerusalem, 9112001, Israel and [4]Department of Clinical Microbiology and Immunology, Sackler Faculty of Medicine, Tel Aviv University, Tel-Aviv, 6997801, Israel

Received June 21, 2022; Revised October 17, 2022; Editorial Decision November 15, 2022; Accepted December 04, 2022

ABSTRACT

A major effort in cancer research is to organize the complexities of the disease into fundamental traits. Despite conceptual progress in the last decades and the synthesis of hallmark features, no organizing principles governing cancer beyond cellular features exist. We analyzed experimentally determined structures harboring the most significant and prevalent driver missense mutations in human cancer, covering 73% (n = 168178) of the Catalog of Somatic Mutation in Cancer tumor samples (COSMIC). The results reveal that a single structural element—κ-helix (polyproline II helix)—lies at the core of driver point mutations, with significant enrichment in all major anatomical sites, suggesting that a small number of molecular traits are shared by most and perhaps all types of cancer. Thus, we uncovered the lowest possible level of organization at which carcinogenesis takes place at the protein level. This framework provides an initial scheme for a mechanistic understanding underlying the development of tumors and pinpoints key vulnerabilities.

INTRODUCTION

The dominant view of cancer research assumes that rules governing the transformation of normal cells into malignant cells share common traits. Thus, a major effort was put into organizing the complexities of the disease into a small number of molecular, biochemical and cellular traits. In the landmark, thought-provoking review, Hanahan and Weinberg integrated and dissected the deluge of information on cancer biology into general cell biology properties necessary to achieve the cancer phenotype (1). Not free of critique, nearly a quarter-century after defining the hallmarks of cancer, they remain the blueprints for understanding the core traits of cancer (2–5). This framework was synthesized as acquired capabilities and enabling characteristics on the cellular level that are regulated by cancer-associated signaling pathways (1,3). These hallmarks were central to the strategy applied since 1971 when the National Cancer Act was declared with the aim of reaching a mechanistic understanding of cancer to develop more effective treatments (6). Since then, technical improvements and advances in structure determination and sequencing techniques have revolutionized cancer research and its analytical depth. However, a mechanistic understanding of the structural components of molecular aberrations driving cancer remains elusive by any objective evaluation.

The focus on genomic aberrations stems from the diversity and acquired capabilities during carcinogenesis that depend largely on the succession of alterations that perturb the normal function of proteins. Most mutations do not confer a cancer hallmark or contribute directly to the tumor phenotype (7), whereas specific mutations in tumor suppressors and oncogenes confer a clear selective advantage to the cancer cells in the local tissue environment and enables their outgrowth (3). Driver mutations, defined by their ability to 'drive' tumor progression, are often categorized into two types: gain-of-function (GoF) and loss-of-function (LoF) (8).

The study of genomic aberrations, and mainly coding mutations, has been focused on the phenotypic consequences of distinct GoF and LoF mutations; whereas, common structural basis for mutations is lacking. To date, mutations are often classified as detrimental, damaging, intolerant or destabilizing (9–12) without a clear delineation of the contrasting functions (i.e. GoF versus LoF) or conflicting associations. For example, many mutations have minimal effect on protein stability or free energy, while others hyperstabilize the active conformation (13,14). The rationale for

*To whom correspondence should be addressed. Tel: +972 3 9377990; Fax: +972 3 9377902; Email: tomermrsn@gmail.com
Correspondence may also be addressed to Gal Markel. Email: markel@tauex.tau.ac.il

studying cancer genomics and proteomics was to uncover the lowest possible level of organization at which tumorigenesis occurs (5). However, no conceptualization beyond cellular physiology exists, indicating that dissecting the complexities of cancer biology further into molecular and biochemical traits is a key challenge.

MATERIALS AND METHODS

Cancer mutation dataset

Mutation data were obtained from the Catalog of Somatic Mutation in Cancer (COSMIC) repository v94 representing the largest and most comprehensive resource of human somatic mutations (15). Specifically, driver mutations were downloaded from the Cancer Gene Census, which contains mutations that have been causally implicated in cancer (16). We used genes classified in Tier 1 defined as such on the basis of documented and reproducible activity relevant to cancer and well-defined evidence of changing the activity of the gene product in a way that promotes oncogenic transformation (16). To gain further support for the significance of the mutations, we excluded variants not associated with pathogenic activity (minimal clinical significance below 4) as annotated by ClinVar (17) based on the American College of Medical Genetics and Genomics (ACMG)/Association for Molecular Pathology (AMP) recommendations (18). Also, to select the most important variants, we included prevalent mutations based on the following criteria: at least 10 unique mutated samples, and the ratio of mutated to tested tumor samples was at least 0.05%. Finally, we manually curated experimental evidence from the literature. We included all type of coding mutations, except nonsense mutations since their structural implications cannot be attributed with certainty to a single position. We identified 205 driver mutations in 45 driver genes at 133 unique positions, all of which were missense mutations. Further support of this set of driver genes was provided by recent studies that explored the sequence context of driver mutations and identified them as pathogenic (19–21).

Structure collection

Protein 3D structures determined by X-ray crystallography or electron microscopy with a resolution cut-off of <3.5 Å were download from the Protein Data Bank (PDB) database (22). Mapping between the PDB structure and the wild-type (WT) or mutated variant was performed using the R package MAPDB. If more than one structure was identified for a given variant, we selected the PDB with the lowest resolution and regular secondary structure assignment. PDB structures missing the residue of interest were excluded. Also, structures with irregular or phosphorylated residues near the variant were excluded.

PISCES dataset

A representative dataset of nonredundant high quality X-ray structures was downloaded from PISCES on 6 November 2020 (23). The structures were culled from the PDB at 20% sequence identity, with resolutions better than 1.6 Å and *R*-factors <0.25, yielding 3757 models.

Functional role assignment

The effects of mutations can be defined in several ways and described on several levels. Variants can lead to myriad of activities causing complex GoF or LoF associations that highly depend on the experimental system. Thus, the interpretation of the functional role can lead to conflicting conclusions, especially when characterizing the effects on signaling pathways. Also, classification of mutations based on biochemical assays, e.g. effect on phosphorylation, could be caused by the mutation directly or indirectly. Further, mutations can generate similar effects by destabilizing the inactive state or stabilizing the active state. Therefore, to characterize the most fundamental effects of the aberration and avoid ambiguous classification, we focused on evidence of the direct structural and functional implications and the main mode of action. Mutations associated with loss of functionality but gained important neomorphic capabilities were considered GoF mutations. For example, mutations leading to a new protein–protein interaction at the expense of another or shifted the substrate specificities (24). For P53 mutations lacking sufficient experimental evidence, we utilized data from the IARC TP53 database (https://p53.iarc.fr/) (25).

Secondary structure assignment

The initial assignment of secondary structures was performed using STRIDE (26). We used DSSP (27) to assign bends at irregular structures and expanded the turn assignments. The commonly used assignment programs do not assign κ-helix (alternative designation for polyproline II [PPII]). Therefore, to assign κ-helix we used the method, which was recently introduced (28,29).

We calculated the root-mean-square dihedral deviations (RMSdD) of the peptide backbone torsional angles ϕ and ψ as a measure of the average deviation from a reference κ-helix. The RMSdD of ϕ and ψ angles is given by

$$\mathrm{RMSdD}_{\phi} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\phi_i - \phi_r)^2} \quad (1)$$

$$\mathrm{RMSdD}_{\psi} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\psi_i - \psi_r)^2} \quad (2)$$

where N is the total number of residues with calculated torsional angles ϕ_i and ψ_i, and ϕ_r and ψ_r are the reference angles of $\phi = -78^\circ$, $\psi = 146^\circ$ (30). The mean RMSdD, which incorporates both torsional dihedral angles, is then given by

$$\mathrm{RMSdD} = \frac{\mathrm{RMSdD}_{\psi} + \mathrm{RMSdD}_{\psi}}{2} \quad (3)$$

To include short segments of κ-helix, at least two consecutive residues with mean RMSdD below the cutoff (ε) of 17 were defined as the criteria for the assignment (28,31,32).

Overall, we assigned the following regular secondary structures: π-helix, α-helix, β-strand, k-helix, 3_{10}-helix (Figure 1A), as well as nonregular secondary structures: bend, bridge, coil, β-turn, α-turn, π-turn or unclassified

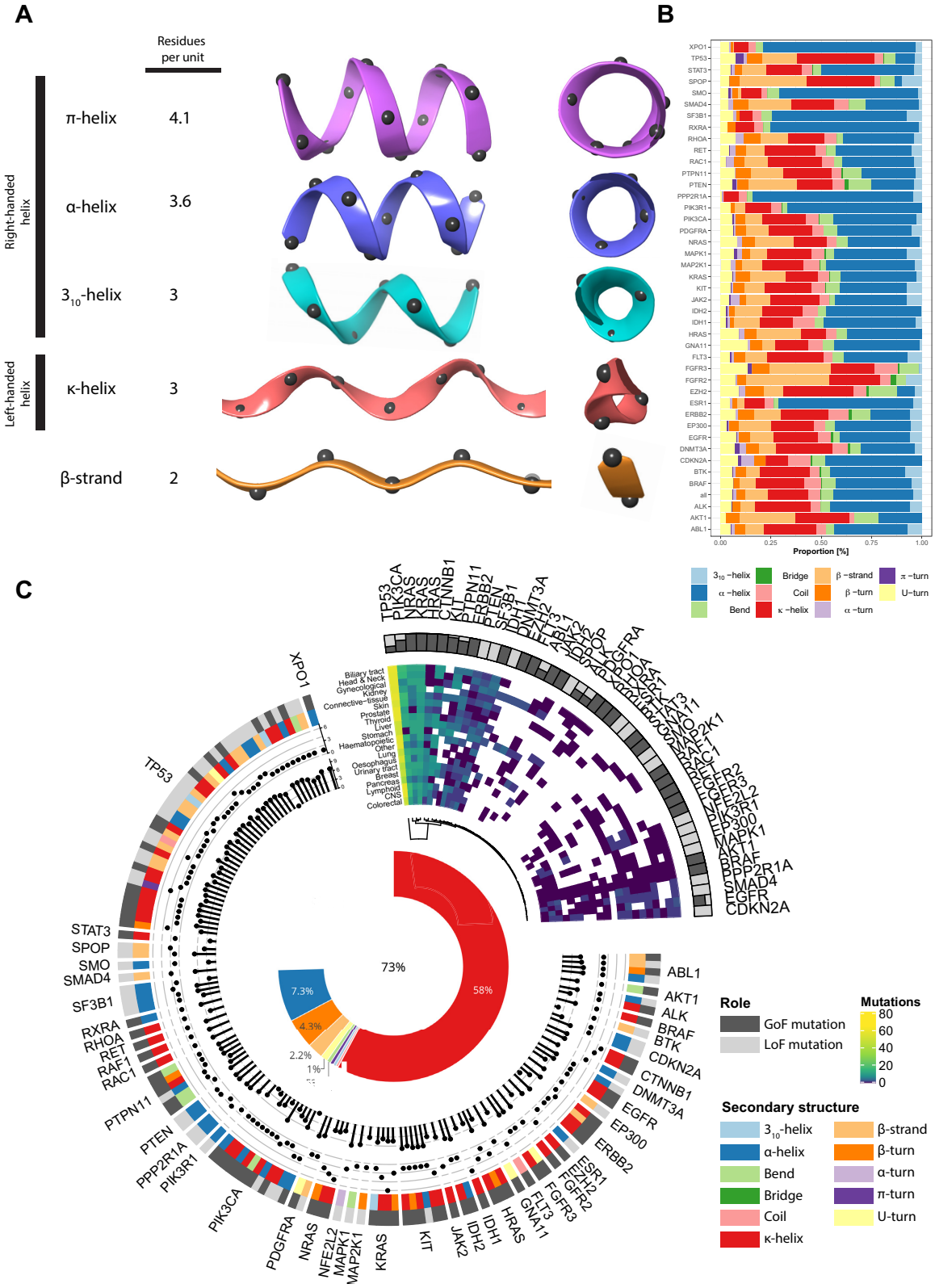

Figure 1. Overview of the structural landscape of missense driver mutations. (**A**) Illustration of the regular secondary structures in ribbon representation in two views. Carbons are represented as black spheres. (**B**) Distribution of the structural elements in the selected PDB models for each gene. (**C**) Circos plot representation of the most prevalent and significant missense driver mutations in human cancer covering 73% ($n = 168\,178$) of COSMIC tumor samples. Left: Track order: functional role (GoF and LoF are colored in dark and light gray, respectively) (i); secondary structure of the native variant (ii), number of variants at the position (iii) and SIFT score (iv). The central donut chart depicts the proportion of tumors with missense driver mutations targeting the secondary structures. Right: circos plot distribution of mutations in various cancers.

turn (U-turn). Of note, none of the residues in our dataset were assigned to π-helix.

Structural characterization

To investigate the major structural components forming the environment of the variants, we characterized the association with nearby neighboring residues. We stratified the nearby residues into flanking and interacting residues. Flanking residues were defined as the three residues N- and C-terminal to the variant (up to six in total). Variant interacting residues were determined if they formed any noncovalent interactions, including van der Waals using a distance cut-off of <5.0 Å. Structural analyses were performed with the Bio3D package (33) in R version 4.0.3. Flanking residues which also interacted with the variant remained with the original annotation.

Characterization of post-transcriptional modifications

To determine the relationship between the structural features of driver mutations and residues harboring post-transcriptional modifications (PTMs), we obtained data from PhosphoSitePlus (34). We considered nearby PTMs if the site was no more than three residues from the variant, flanking or interacting residues. The PTMs included phosphorylation, ubiquitination, methylation, acetylation, SUMOylation, *O*-GlcNAcylation and *O*-GalNAcylation.

Molecular visualization programs

The PDB structures were visualized using PyMOL 2.4.1 (35) and the academic version of Schrodinger Maestro v12.5 (36). Specifically, aligned PDB structures were visualized using PyMOL, and reassigned secondary structures were visualized using Maestro v12.5. κ-helices and 3_{10}-helices were represented as ribbons and tubes, respectively. Ribbons were drawn passing through alpha carbons.

Statistical methods

The association between the secondary structure of the variant and the abundance or the functional roles was estimated using the Chi-square test. Odds ratio estimates and 95% confidence intervals were calculated using the R package epiR. The Haldane–Anscombe correction was used when one of the cells had zero value. *P*-values are two-sided with a Bonferroni correction for multiple comparisons.

RESULTS

Here, we identified the most prominent mutations which drive different types of human cancer based on their significance and prevalence in the COSMIC database (15) (see Materials and Methods section). Each variant was curated and annotated as GoF or LoF based on the literature. Structural analysis was performed on low-resolution solved structures containing the native or mutated residue. We assigned the secondary structure of the selected PDB models using STRIDE (26) and DSSP (27). Since κ-helix, commonly known as polyproline type II helix, is not assigned by the widely used programs, a recently introduced method was used for this assignment (28,29). Briefly, we calculated the RMSdD of the peptide backbone torsional angles φ and ψ as a measure of the average deviation from a reference κ-helix (31,32).

We collected and curated the functional role of 205 (60% GoF and 40% LoF) driver mutations at 133 unique sites in 45 tumor suppressor and oncogenes from the COSMIC database and assigned the following secondary structures: π-helix, α-helix, β-strand, κ-helix and 3_{10}-helix (Figure 1A), as well as nonregular secondary structures: bend, bridge, coil, β-turn, α-turn, π-turn or unclassified turn (U-turn). Of note, none of the variants or structures contained residues assigned as π-helix (Figure 1B). Solved structures of mutated proteins were identified for 22% (45/205) of the dataset; therefore, we focused on the structural analysis of the WT variants. The distribution of the driver mutations across 45 genes and 20 anatomical sites is shown in Figure 1C and Supplementary Figure S1. This compilation of driver mutations covers 73% ($n = 168\,178$) of all sequenced tumors in COSMIC with missense mutations, which provides extensive coverage of the genomic landscape of driver mutations in human cancer (Supplementary Table S1). Over 96% of tumor samples contained only one driver mutation (Supplementary Figure S2). Notably, we found that a single structural element, κ-helix, covers 58% of the tumors in which at least one variant is assigned with this motif in the WT form. The next structures were far less prevalent with 7.3% and 4.3% for α-helix and β-turn, respectively.

Given the observation that κ-helix, as the target of driver mutations, makes up the bulk of tumor samples with 81% ($n = 135\,312$) of the dataset (Figure 2A–D), we further explored the structural composition of the mutations. The most represented genes were JAK2, KRAS, BRAF and TP53, with κ-helix being the most common structural element (Figure 2D). The distribution of structural motifs across different tumor types reveals a consistent pattern, with κ-helix being the most widespread assignment (Figures 2C and 3A). Exceptions include breast and urinary tract cancer with α-helix and U-turns, respectively, the most common conformations, followed by κ-helix. In breast cancer, this can be explained by a high rate of PIK3CA mutations, most of which occurred at α-helices (Figure 2D). Comparison between the WT and mutated forms demonstrated that the structural assignment was not affected by the aberration in most variants (69%), indicating that mutations tend to maintain the WT secondary structure (Figure 2E). Further, κ-helix persisted as the predominant structure in mutated form (Figure 2E, Supplementary Figures S6). To gain insights into the functional role of the structural motifs in tumor biology, we stratified the variants into GoF and LoF mutations. Figure 2F shows a distinct pattern where κ-helices dominated GoF mutations with 68%. In LoF, the most common structural motifs were α-helices (48%) and β-strands (28%).

We next examined whether the increased proportions of the structural components reflect their relative abundance or are over-represented in the dataset. Analysis of hypergeometric distribution (Figure 3B) revealed significant and exclusive enrichment of κ-helix in driver mutations (odds ratio (OR) = 2.2, 95% confidence interval (CI): 1.5–3.1,

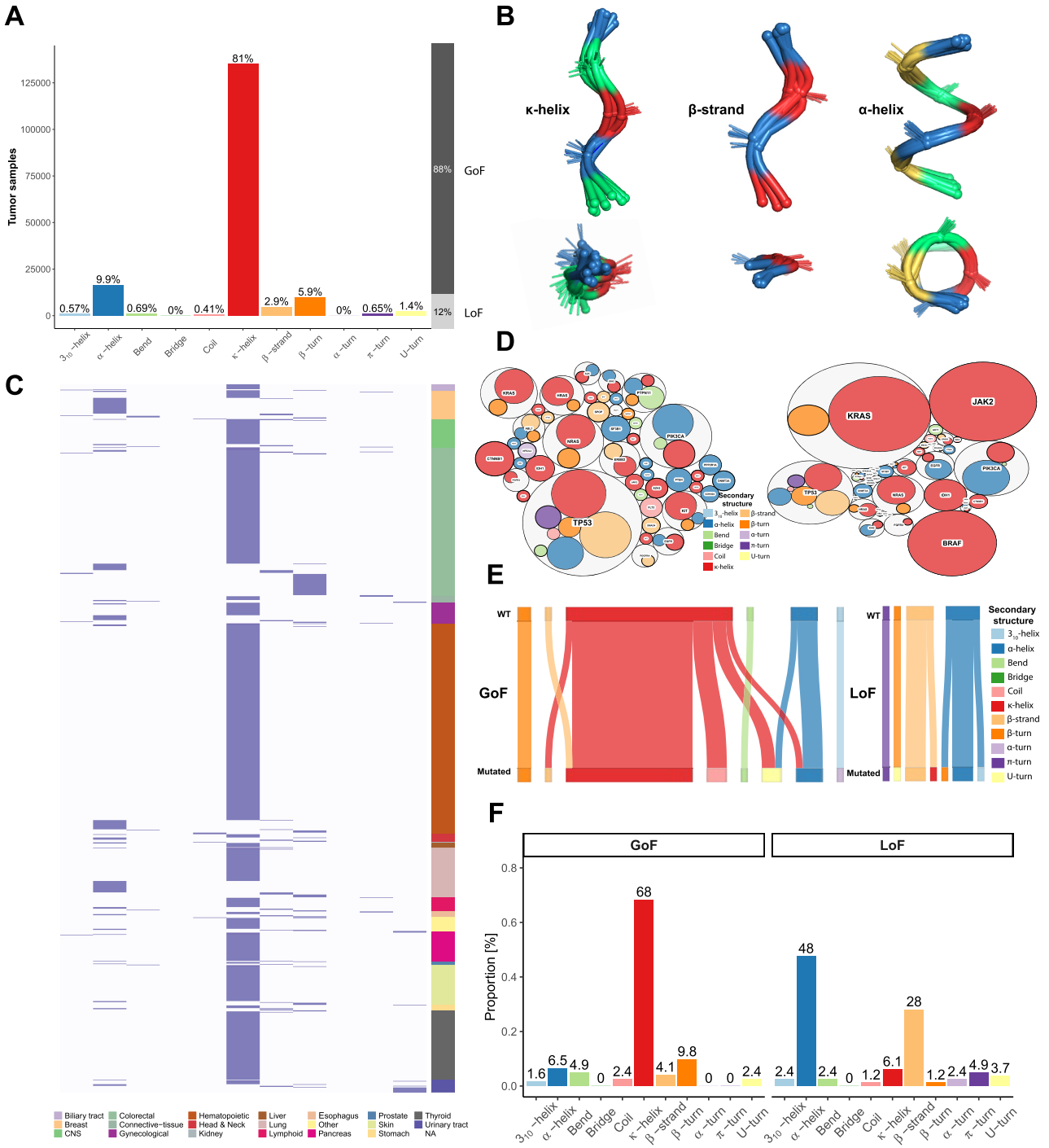

Figure 2. Structural characterization of missense driver mutations. (**A**) Proportion of tumors among 168 178 samples with identified missense driver mutations targeting the various secondary structures. Right bar, proportion of tumors with mutations annotated as gain-of-function (GoF) or loss-of-function (LoF). (**B**) Representative conformations targeted by mutations in tube representation with each residue colored in a different color. Shown are structurally aligned structures of κ-helices, β-strands and α-helices with at least three, two and four residues, respectively. (**C**) Distribution of missense driver targeting the secondary structures stratified by anatomical tumor site. (**D**) Per-gene prevalence of structural motifs based on the amount of driver mutations (left) or the amount of tumor samples harboring the mutations (right). (**E**) Association between the wild-type (WT) and mutated conformations for 45 variants (GoF: 34, LoF: 11). (**F**) Structural characterization of residues targeted by 205 variants stratified by functional role (GoF: 123, LoF: 82).

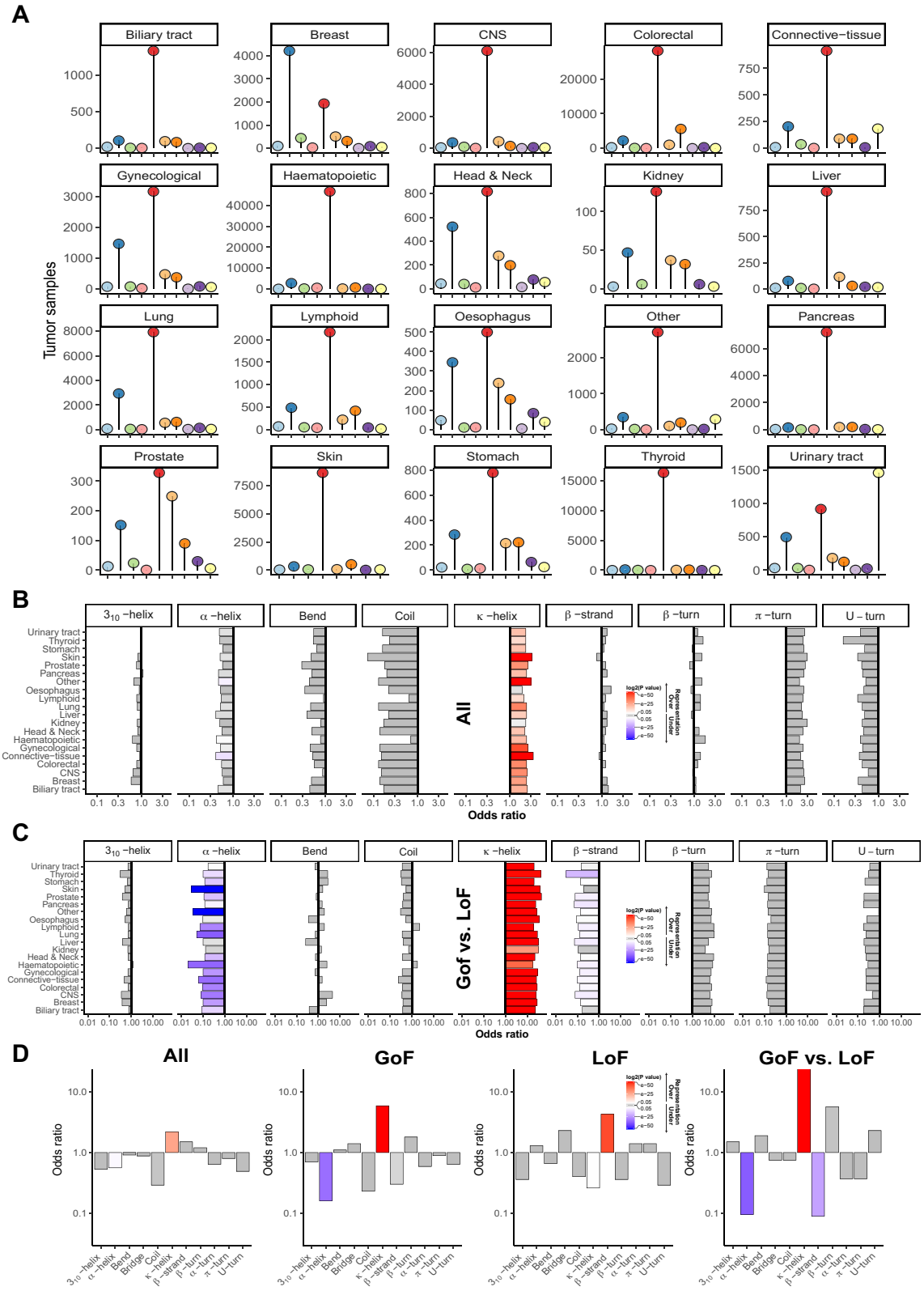

Figure 3. Enrichment of structural elements in driver mutations. (**A**) Summary of tumors with missense driver mutations targeting the secondary structures stratified by anatomical tumor site. (**B**) Enrichment of secondary structures of driver mutations in tumor samples ($n = 168\,178$ samples) and (**C**) among gain-of-function (GoF) versus los-of-function (LoF) mutations ($n = 133$ unique sites). (**D**) Enrichment of structural elements in driver mutations among all structures in the dataset (left) and over-representation in GoF, LoF and among GoF versus LoF mutations ($n = 133$ unique sites) (right panels). Odds ratios are shown on a logarithmic scale. Bonferroni-adjusted *P*-values for multiple testing are shown.

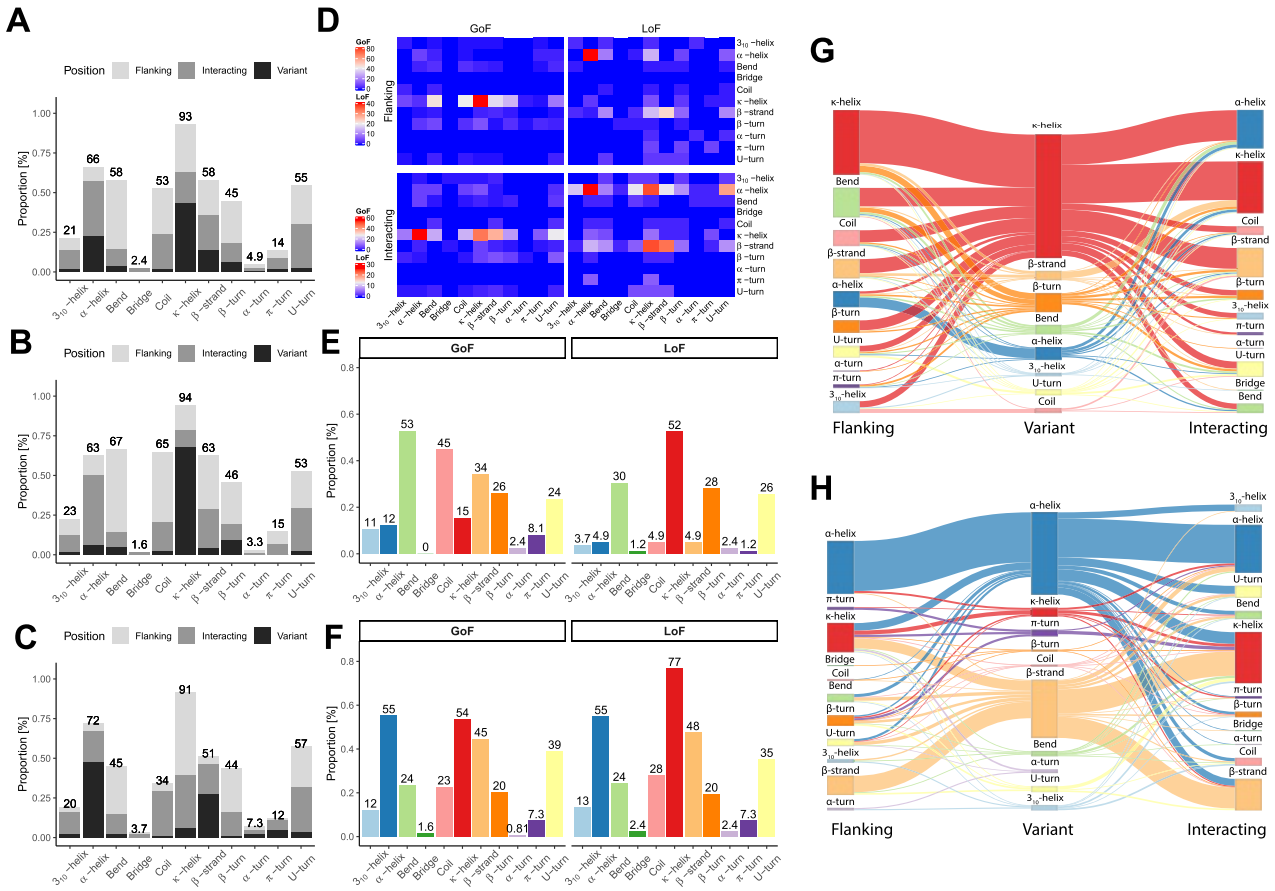

Figure 4. Structural characterization of the environment of missense driver mutations. Proportion of secondary structures of the variant, flanking or interacting residues of the wild-type (WT) form for (**A**) all (n = 205), (**B**) gain-of-function (GoF) (n = 123) and (**C**) loss-of-function (LoF) (n = 82) mutations. (**D**) Heatmap depicting the number of associations between the conformation of the WT variant (Y-axis) and the flanking or interacting residues (X-axis). Proportion of variants with at least one residue forming the conformation in the (**E**) flanking and (**F**) interacting residues. To focus on the unique tendency of conformations to flank the variant, proportions shown for elements different than that of the variant. Relationship between the conformations of the variants and corresponding interacting or flanking residues in (**G**) GoF and (**H**) LoF driver mutations in the WT form.

P = 6.5e-05). This is mostly attributable to κ-helix enrichment in GoF mutations (OR = 5.9, 95% CI: 3.7–9.4, P = 1.2e-15). In LoF mutations, β-strands were significantly enriched (OR = 4.3; 95% CI, 2.5–7.6, p = 2.3e-7), but not α-helices owing to the high abundance in proteins (Figure 1B). The enrichments improved when we used a large representative ensemble of sufficiently diverse protein structures with low resolutions, with significant enrichment in all anatomical sites (Supplementary Figure S3). The length distribution for the structural motifs is shown in Supplementary Figure S4. When evaluating the association between the functional role of the variants and the secondary structure regardless of their abundance, κ-helices were significantly enriched in GoF in all tumor types, including breast and urinary tract cancer (Figure 3C–D). In contrast, α-helices and β-strands were significantly enriched in LoF.

To investigate the major structural components forming the environment of the variants, we characterized the association with nearby neighboring residues: flanking (within three amino acids) and interacting residues (Figure 4, Supplementary Figures S5 and S6). We found that the primary motif in the local environment of the variant was κ-helix in both types of mutations (Figure 4A). In GoF and LoF, the κ-helical motif was implicated in 94% and 91% of the variants in the native state, respectively, albeit with contrasting differences (Figure 4B and C). High GoF proportions of κ-helix are mainly attributable to the variant structure, whereas high LoF proportions are attributable to the interacting and flanking residues (Figure 4D–H). Remarkably, κ-helix was implicated in over 99% of the tumor samples when considering both the variant and the local environment (Supplementary Figure S7). This observation indicates that nearly all driver mutations target κ-helix or affect it directly. The next frequent elements were bends (87%) which tend to flank κ-helices in GoF mutations (Figure 4D and G). We explored the relationship between the structural features of the driver mutations and PTMs. Figure 5A depicts the distribution of PTMs and the associated structural assignments, with phosphorylation being the most common modification, whereas no relevant glycosylation or SUMOylation sites were identified. Consistent with the assignment of the variants in the native state, the most abundant conformation of PTM sites were κ-helices with 29% of variants and 35% of GoF variants

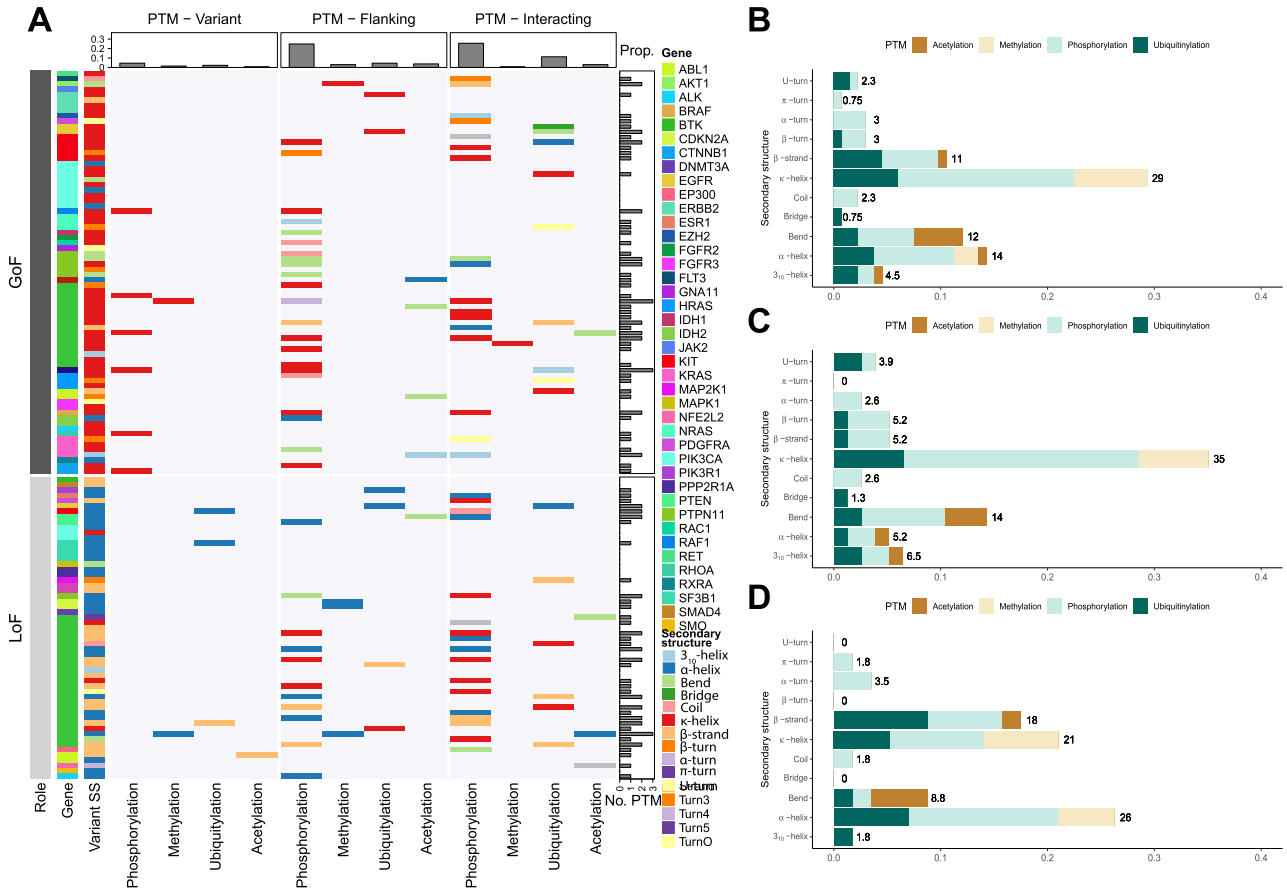

Figure 5. Structural characterization of the post-transcriptional modification (PTM) sites associated with missense driver mutations. (**A**) Conformational distribution of PTM sites associated with the variant and its surrounding. Summary of variants with PTM sites for (**B**) all, (**C**) gain-of-function (GoF) and (**D**) loss-of-function (LoF) driver mutations in the wild-type (WT) form. PTM sites were considered if they were positioned within three amino acids from a flanking or interacting residue.

(Figure 5B–D). Conversely, α-helices were slightly more represented in LoF variants (26%), followed by κ-helices (21%) and β-strands (18%).

DISCUSSION

In this study, we found that most driver mutations occur at a distinct structural motif and share the κ-helix conformation, which is also associated with the vast majority of tumor samples. Thus, the κ-helix conformation serves as the dominant form targeted by driver mutations, suggesting that it represents the lowest denominator in carcinogenesis. This observation provokes a number of questions. What is unique about the κ-helical conformation? Why is it enriched in cancer? How has it fallen under the radar?

κ-helix, commonly known as polyproline type II helix, is an elongated left-handed structure, with three residues per turn and a 3-fold rotational symmetry along the helical axis (29). Thus, the motif exhibits the least number of residues required to cover the full range of 360 degrees in the stereochemical space (Figures 1A and 2B). The extended character of κ-helix does not support regular patterns of intrachain hydrogen bonds, allowing for fast conformational changes (37). Since other noncovalent interactions can stabilize κ-helix, its unsatisfied hydrogen bonds are free to engage with the surrounding residues (37). Together, the flexibility and the availability of hydrogen bonds make the structure an ideal molecular switch or a binding motif. The structural element has been described as a 'functional block' as it is often linked to a specific function, including allosteric regulation and protein–protein interactions (28,38). Thus, it is not surprising that mutations occur at κ-helices that participate in the regulation and binding interfaces of oncogenes and tumor suppressors and alter their on/off states or typical interactions. Despite extensive research into the structural and functional properties of cancer mutations, κ-helix as a key player in driver mutations has been overlooked, which can be ascribed to several reasons. First, the conformation is not assigned by the widely used secondary structure assignment programs, including DSSP, which are employed in the PDB (32), and relatedly, lack any graphical representation in molecular visualization programs (28,29). Second, although prolines have a high propensity to form the conformation, hence the historical term 'polyproline', it has been widely criticized for its misleading association with mostly prolines, while proline-free

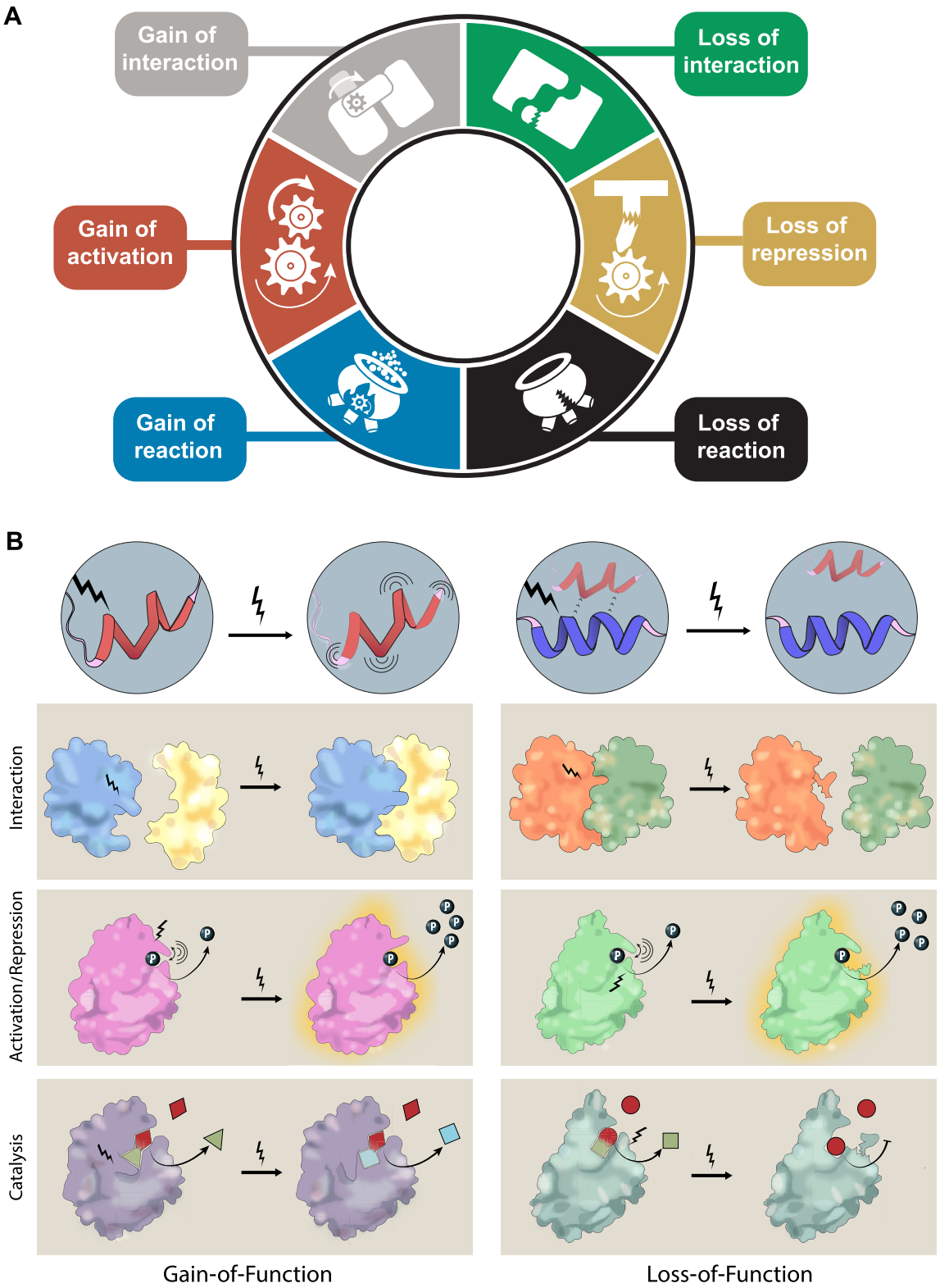

Figure 6. Proposed model of functional capabilities at the molecular level acquired by driver mutations (**A**). Illustrative example showing the effect of mutations targeting κ-helix or α-helix leading to gain-of-function (GoF) or loss-of-function (LoF) mutations (**B**).

structures are just as common or possibly more common in proteins (28,29,32,38).

Although our analyses show that κ-helix stands out among all secondary structures, driver mutations target different regions in GoF and LoF mutations. We found that GoF mutations tend to occupy κ-helices, which are associated with greater flexibility that may facilitate activating conformational changes and neomorphic binding affinities. In contrast, LoF mutations tend to target the structurally rigid α-helices and β-strands that are responsible for maintaining stability at allosteric regulation sites and binding interfaces, which are often mediated by nearby flexible regions such as κ-helix. Such observations are increasingly supported by in-depth structure–function analysis of driver mutations (39,40).

The binary categorization of driver mutations into GoF and LoF is useful but cannot accommodate the range of functions of mutations in cancer. We suggest that the vast catalog of driver mutations is a manifestation of six essential alterations in the molecular machinery governing cancer biology (Figure 6): gain of interaction, gain of activation, gain of reaction, loss of interaction, loss of repression and loss of reaction. These structural alterations—acquired during tumor development—represent mechanistic underpinning for enabling reprogramming of hallmark capabilities by cellular circuitry. Analogous to the synthesis by Hanahan and Weinberg, we propose that these six molecular hallmarks of cancer are shared in most or perhaps all types of human cancer.

Examples of oncogenic GoF driver mutations arising in κ-helices and involving these molecular hallmarks include: $P53^{R175H}$ disrupting the recruitment of MRN/ATM to DNA damage sites by physically interacting with Mre11, leading to ATM inactivation (41); $JAK2^{V617F}$ hyperstabilizing the stimulatory state using steric mechanism on JH1 domain and enhancing JAK2 activity (13); $IDH1^{R132}$ mutants gain neomorphic enzymatic activity by converting α-ketoglutarate to 2-hydroxyglutarate and modulating metabolic effects (42). In contrast, examples of LoF driver mutations in tumor suppressors arising in rigid elements, such as α-helix, include $DNMT3A^{R882H}$ that disrupts WT tetramers leading to reduced methyltransferase activity (43), whereas $PTEN^{R130G}$ losses phosphatase activity while maintaining dimerization with WT, which constrains its phosphatase activity in a dominant-negative manner (44).

Contrary to the paradigm that mutations produce oncogenes with dominant GoF and tumors suppressors with recessive LoF (1), both classes of cancer genes can elicit cancer phenotype by losing or gaining new functions. Also, the binary classification of genes into oncogenes and tumor suppressors is inaccurate for many genes and debatable for others (45–47). Notably, the history of p53 and the oscillating interpretations of its function serve as an example of how scientific paradigms evolve and influence perceptions about cancer-causing genes (45). The current classification of p53 still remains inconclusive. Similarly, the binary classification of driver versus passenger and GoF versus LoF mutations suffers from similar limitations, and it is difficult to unambiguously ascribe the functions of particular mutations to a single capability. Also, the functionally of cancer mutations may operate in concert or span a continuum spectrum of effects (39). For example, the GoF $JAK2^{V617F}$ mutation also has LoF features due to its impairment of JH2 catalytic activity, which further enhances JAK2 (JH1) activity due to loss of phosphorylation at negative regulatory sites (13). Thus, to avoid ambiguous classification, we focused on the direct and primary reported mode of action.

Although we focused on structures with the lowest resolutions, different structures might display various configurations as proteins are dynamic in nature and particularly at different states or environmental conditions. Other limitations of the study include the lack of representation of variants without solved structure, especially of rare mutations. Nonetheless, the current dataset spans the full range of tumor types and, based on the high coverage (73%) of COSMIC database, is implicated in the majority of human cancer.

Our results reveal that a single structural element lies at the core of missense driver mutations and suggest that a small number of molecular traits are shared by most and perhaps all types of cancer. Envisioning that cancer research is a logical science that could be reduced to machines and their parts, we uncover the lowest possible level of organization at which carcinogenesis takes place at the protein level. By providing a glimpse into the toolbox of cancer, we anticipate that a mechanistic understanding underlying the development of tumors and key vulnerabilities will be outlined in greater detail and clarity in the future.

DATA AVAILABILITY

The code for the structural assignment of k-helix is available at https://github.com/tomermeirson/Kappa-helix and https://doi.org/10.5281/zenodo.7343583.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Cancer Online.

FUNDING

Samueli Foundation grant for Integrative Immuno-Oncology (to G.M.); Israel Science Foundation.
Conflict of interest statement. None declared.

REFERENCES

1. Hanahan,D. and Weinberg,R.A.(2000). The hallmarks of cancer. *Cell*, **100**, 157–170.
2. Fouad,Y.A. and Aanei,C. (2017) Revisiting the hallmarks of cancer. *Am. J. Cancer Res.*, **7**, 1016.
3. Hanahan,D. and Weinberg,R.A. (2011) Hallmarks of cancer: the next generation. *Cell*, **144**, 646–674.
4. Meirson,T., Gil-Henn,H. and Samson,A.O. (2020) Invasion and metastasis: the elusive hallmark of cancer. *Oncogene*, **39**, 2024–2026.
5. Sonnenschein,C. and Soto,A.M. (2013) The aging of the 2000 and 2011 Hallmarks of Cancer reviews: a critique. *J. Biosci.*, **38**, 651–663.
6. Rauscher,F.J. Jr (1974) The National Cancer Program and the National Cancer Act of 1971. *Nat. Cancer Prog. Int. Cancer Res*. Vol. **3**.
7. Bozic,I., Antal,T., Ohtsuki,H., Carter,H., Kim,D., Chen,S., Karchin,R., Kinzler,K.W., Vogelstein,B. and Nowak,M.A. (2010) Accumulation of driver and passenger mutations during tumor progression. *Proc. Natl Acad. Sci. USA*, **107**, 18545–18550.

8. Li,Y., Zhang,Y., Li,X., Yi,S. and Xu,J. (2019) Gain-of-function mutations: an emerging advantage for cancer biology. *Trends Biochem. Sci*, **44**, 659–674.
9. 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.
10. Baugh,E.H., Simmons-Edler,R., Müller,C.L., Alford,R.F., Volfovsky,N., Lash,A.E. and Bonneau,R. (2016) Robust classification of protein variation using structural modelling and large-scale data integration. *Nucleic Acids Res.*, **44**, 2501–2513.
11. Kircher,M., Witten,D.M., Jain,P., O'roak,B.J., Cooper,G.M. and Shendure,J. (2014) A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.*, **46**, 310–315.
12. Ng,P.C. and Henikoff,S. (2003) SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res.*, **31**, 3812–3814.
13. Bandaranayake,R.M., Ungureanu,D., Shan,Y., Shaw,D.E., Silvennoinen,O. and Hubbard,S.R. (2012) Crystal structures of the JAK2 pseudokinase domain and the pathogenic mutant V617F. *Nat. Struct. Mol. Biol.*, **19**, 754–759.
14. Blanden,A.R., Yu,X., Blayney,A.J., Demas,C., Ha,J.-H., Liu,Y., Withers,T., Carpizo,D.R. and Loh,S.N. (2020) Zinc shapes the folding landscape of p53 and establishes a pathway for reactivating structurally diverse cancer mutants. *Elife*, **9**, e61487.
15. Tate,J.G., Bamford,S., Jubb,H.C., Sondka,Z., Beare,D.M., Bindal,N., Boutselakis,H., Cole,C.G., Creatore,C. and Dawson,E. (2019) COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res.*, **47**, D941–D947.
16. Sondka,Z., Bamford,S., Cole,C.G., Ward,S.A., Dunham,I. and Forbes,S.A. (2018) The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nat. Rev. Cancer*, **18**, 696–705.
17. Landrum,M.J., Lee,J.M., Benson,M., Brown,G.R., Chao,C., Chitipiralla,S., Gu,B., Hart,J., Hoffman,D. and Jang,W. (2018) ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.*, **46**, D1062–D1067.
18. Richards,S., Aziz,N., Bale,S., Bick,D., Das,S., Gastier-Foster,J., Grody,W.W., Hegde,M., Lyon,E. and Spector,E. (2015) Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.*, **17**, 405–423.
19. Banerjee,S., Raman,K. and Ravindran,B. (2021) Sequence neighborhoods enable reliable prediction of pathogenic mutations in cancer genomes. *Cancers*, **13**, 2366.
20. Dietlein,F., Weghorn,D., Taylor-Weiner,A., Richters,A., Reardon,B., Liu,D., Lander,E.S., Van Allen,E.M. and Sunyaev,S.R. (2020) Identification of cancer driver genes based on nucleotide context. *Nat. Genet.*, **52**, 208–218.
21. Agajanian,S., Oluyemi,O. and Verkhivker,G.M. (2019) Integration of random forest classifiers and deep convolutional neural networks for classification and biomolecular modeling of cancer driver mutations. *Front. Mol. Biosci.*, **6**, 44.
22. Sussman,J.L., Lin,D., Jiang,J., Manning,N.O., Prilusky,J., Ritter,O. and Abola,E.E. (1998) Protein Data Bank (PDB): database of three-dimensional structural information of biological macromolecules. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **54**, 1078–1084.
23. Wang,G. and Dunbrack Jr,R.L. (2003) PISCES: a protein sequence culling server. *Bioinformatics*, **19**, 1589–1591.
24. Songyang,Z. and Cantley,L.C. (1995) Recognition and specificity in protein tyrosine kinase-mediated signalling. *Trends Biochem. Sci*, **20**, 470–475.
25. Petitjean,A., Mathe,E., Kato,S., Ishioka,C., Tavtigian,S.V., Hainaut,P. and Olivier,M. (2007) Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database. *Hum. Mutat.*, **28**, 622–629.
26. Heinig,M. and Frishman,D. (2004) STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res.*, **32**, W500–W502.
27. Kabsch,W. and Sander,C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers: Original Res. Biomol.*, **22**, 2577–2637.
28. Meirson,T., Bomze,D., Markel,G. and Samson,A.O. (2020) κ-helix and the helical lock and key model: a pivotal way of looking at polyproline II. *Bioinformatics*, **36**, 3726–3732.
29. Meirson,T., Bomze,D., Kahlon,L., Gil-Henn,H. and Samson,A.O. (2020) A helical lock and key model of polyproline II conformation with SH3. *Bioinformatics*, **36**, 154–159.
30. Stapley,B.J. and Creamer,T.P. (1999) A survey of left-handed polyproline II helices. *Protein Sci.*, **8**, 587–595.
31. Meirson,T., Bomze,D. and Markel,G. (2021) Structural basis of SARS-CoV-2 spike protein induced by ACE2. *Bioinformatics*, **37**, 929–936.
32. Mansiaux,Y., Joseph,A.P., Gelly,J.-C. and de Brevern,A.G. (2011) Assignment of PolyProline II conformation and analysis of sequence–structure relationship. *PLoS One*, **6**, e18401.
33. Grant,B.J., Skjaerven,L. and Yao,X.Q. (2021) The Bio3D packages for structural bioinformatics. *Protein Sci.*, **30**, 20–30.
34. Hornbeck,P.V., Kornhauser,J.M., Tkachev,S., Zhang,B., Skrzypek,E., Murray,B., Latham,V. and Sullivan,M. (2012) PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Res.*, **40**, D261–D270.
35. DeLano,W.L. (2002) Pymol: an open-source molecular graphics tool. *CCP4 Newsl. Protein Crystallogr.*, **40**, 82–92.
36. Bell,J., Cao,Y., Gunn,J., Day,T., Gallicchio,E., Zhou,Z., Levy,R. and Farid,R. (2012) PrimeX and the Schrödinger computational chemistry suite of programs.
37. Cubellis,M., Caillez,F., Blundell,T. and Lovell,S. (2005) Properties of polyproline II, a secondary structure element implicated in protein–protein interactions. *Proteins: Struct., Funct., Bioinf.*, **58**, 880–892.
38. Adzhubei,A.A., Sternberg,M.J. and Makarov,A.A. (2013) Polyproline-II helix in proteins: structure and function. *J. Mol. Biol.*, **425**, 2100–2132.
39. Agajanian,S., Odeyemi,O., Bischoff,N., Ratra,S. and Verkhivker,G.M. (2018) Machine learning classification and structure–functional analysis of cancer mutations reveal unique dynamic and network signatures of driver sites in oncogenes and tumor suppressor genes. *J. Chem. Inf. Model.*, **58**, 2131–2150.
40. Alsulami,A.F., Torres,P.H., Moghul,I., Arif,S.M., Chaplin,A.K., Vedithi,S.C. and Blundell,T.L. (2021) COSMIC Cancer Gene Census 3D database: understanding the impacts of mutations on cancer targets. *Briefings Bioinf.*, **22**, bbab220.
41. Liu,D., Song,H. and Xu,Y. (2010) A common gain of function of p53 cancer mutants in inducing genetic instability. *Oncogene*, **29**, 949–956.
42. Reitman,Z.J., Parsons,D.W. and Yan,H. (2010) IDH1 and IDH2: not your typical oncogenes. *Cancer Cell*, **17**, 215–216.
43. Russler-Germain,D.A., Spencer,D.H., Young,M.A., Lamprecht,T.L., Miller,C.A., Fulton,R., Meyer,M.R., Erdmann-Gilmore,P., Townsend,R.R. and Wilson,R.K. (2014) The R882H DNMT3A mutation associated with AML dominantly inhibits wild-type DNMT3A by blocking its ability to form active tetramers. *Cancer Cell*, **25**, 442–454.
44. Papa,A., Wan,L., Bonora,M., Salmena,L., Song,M.S., Hobbs,R.M., Lunardi,A., Webster,K., Ng,C. and Newton,R.H. (2014) Cancer-associated PTEN mutants act in a dominant-negative manner to suppress PTEN protein function. *Cell*, **157**, 595–610.
45. Soussi,T. (2010) The history of p53: a perfect example of the drawbacks of scientific paradigms. *EMBO Rep.*, **11**, 822–826.
46. Cameron,E.R. and Neil,J.C. (2004) The Runx genes: lineage-specific oncogenes and tumor suppressors. *Oncogene*, **23**, 4308–4314.
47. Yang,L., Han,Y., Saiz,F.S. and Minden,M. (2007) A tumor suppressor and oncogene: the WT1 story. *Leukemia*, **21**, 868–876.