phase according to Das and Pappu (56). This property is calculated using the fraction of positive and negative charges in a blob and is particularly relevant for IDPs. Nearly all structured proteins fall in the class 1 (weak polyampholyte) part of the Das–Pappu phase diagram. In contrast to structured proteins, IDPs can be found in all five Das–Pappu phases, including class 2 (Janus or boundary region), class 3 (strong polyampholyte), class 4 (negatively charged strong polyelectrolyte), and class 5 (positively charged strong polyelectrolyte).

The blob hydrophobicity class and charge class are fundamentally correlated; while blob charge class does not explicitly consider hydrophobicity, increasing the number of charged residues will reduce the average hydrophobicity of a blob. The extent of this correlation is shown in Fig. 4*A*, which breaks down the fraction of h- and p-blobs that fall in each Das–Pappu charge class. As expected, most h-blobs (90%) fall in class 1 (weak polyampholyte), followed by 9% in class 2 (Janus). The p-blobs are more evenly distributed across classes, with the highest fraction (42%) classified as strong polyelectrolytes. Fig. 4 *B* and *C* shows the SNP distributions for blobs with different hydrophobicity class and charge class, respectively. Since there are five charge classes and only three hydrophobicity classes, we hypothesized that in nonaggregating proteins, blob charge class would have a stronger association with disease than blob hydrophobicity class. Instead, we found that the strongest dSNP enrichment as a function of charge class (1.09-fold, $P < 10^{-58}$ for weak polyampholyte blobs) is comparable to or even slightly less than the strongest enrichment found for hydrophobicity class (1.15-fold, $P < 10^{-100}$ for h-blobs).

Furthermore, the charge-based and hydrophobicity-based classification schemes yield similar trends with protein aggregation: Fig. 4*B* shows that a given dSNP in an aggregating protein is just as likely to be found in an h-blob as if it were in a nonaggregating protein, and Fig. 4*C* shows an analogous result for dSNPs in various charge classes. However, nSNPs in aggregating proteins are found slightly less frequently in blobs classified as h-blobs or class 1 (globular, weak-polyampholyte) blobs than nSNPs in nonaggregating proteins. As a result, we do observe a small increase in overall dSNP enrichment for h-blobs/weak-polyampholyte blobs of aggregating proteins relative to nonaggregating proteins.

We then used the results from unconstrained-length blobulation to further stratify the dSNP enrichment in aggregating proteins by hydrophobicity threshold $H^\star$ and blob length. The enrichment calculations from Fig. 2*D* were partitioned between aggregating and nonaggregating proteins and are shown in Fig. 4*D*. The highly enriched (blue) band is shifted toward the origin for aggregating proteins, indicating that sensitivity to mutation is found in shorter and more weakly hydrophobic blobs. The differential enrichment values using whole-sequence blobulation in Fig. 4 *B* and *C* arise from collapsing the distributions in Fig. 4*D* along a single value of $H^\star$. This result suggests that lower-hydrophobicity thresholds may be appropriate for predicting disease risk in known-aggregating proteins and underscores the importance of a multidimensional analysis for distinguishing between different groups of proteins.

Disease-Associated SNPs Are Enriched for Mutations That Change Local Blob Characteristics and Overall Protein Blob Topology. Whole-sequence blobulation yields a series of h-blobs, connected by p- and s-blobs, which we term the "blobular topology." Such a topology is analogous to the classic protein topology of secondary structure elements, although the location of edges and number of elements may be distinct. A SNP can alter the blobular topology by moving a short stretch of contiguous residues above or below the minimum blob size, either forming a new small h-blob or dissolving an existing small h-blob, respectively. A SNP can also split a long h-blob by interrupting a long contiguous hydrophobic sequence or merge two smaller h-blobs into one long h-blob by removing such an interruption.

Here we tested whether the dSNPs were more likely to change the topology determined by whole-sequence blobulation ($H^\star = 0.4$, $L_{\min} = 4$). Fig. 5*A* displays the fraction of nSNPs and dSNPs that cause each type of topological change. In the background case we expect to see more formation than dissolution, since the

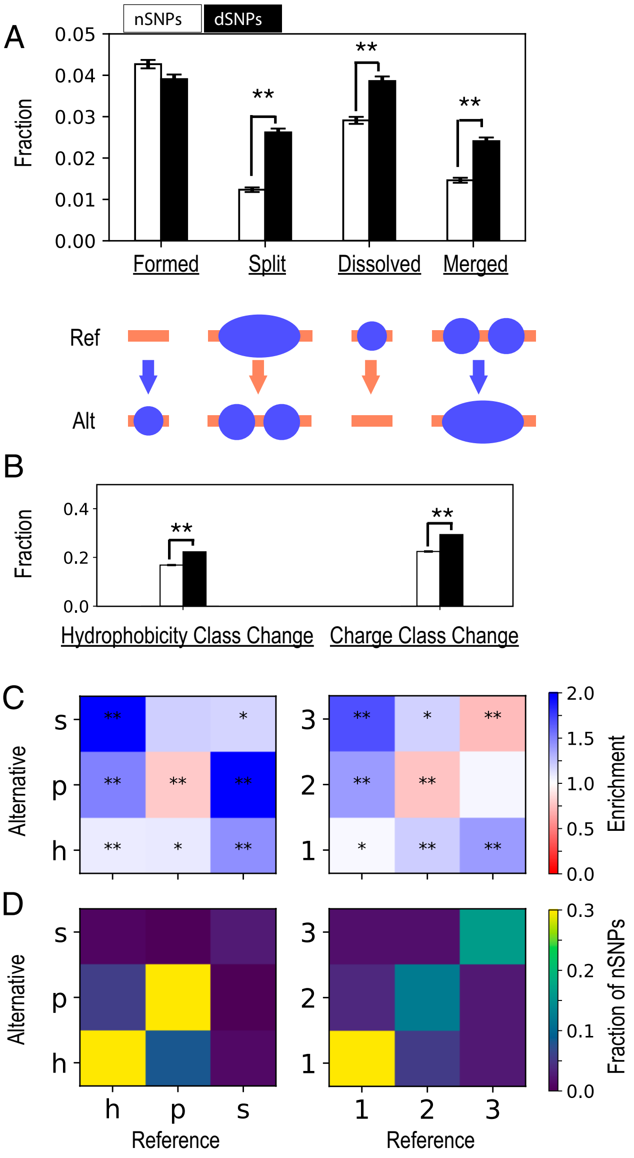

Fig. 5. Distribution of SNPs that change blob properties. (*A*) Fraction of nSNPs or dSNPs that change the blobular topology by either forming or dissolving an h-blob or by splitting one h-blob or merging two h-blobs. (*B*) Fractions of nSNPs or dSNPs that induce a change in hydrophobicity class or charge class. (*C*) The enrichment of dSNPs relative to nSNPs that induce a specific transition between the blob containing the reference allele (*x* axis) and the blob containing the alternative allele (*y* axis). This is shown for two blob properties, hydrophobicity class (*Left*) and charge class (*Right*), where the charge class categories are the same as in Fig. 4. (*D*) The overall proportion of nSNPs that induce each of the blobular topology transitions shown in *C*. The plot displays only charge classes 1 to 3 because fewer than 1.5% of SNPs involve charge class 4 or class 5. Significant enrichment or depletion in dSNPs is annotated with * ($P < 5 \times 10^{-3}$) or ** ($P < 5 \times 10^{-11}$) (binomial test). Errors bars in *A* and *B* represent one SE in the mean. All panels use whole sequence blobulation with $H^\star = 0.4$ and $L_{\min} = 4$.