Supplementary File IV: Rationale for disabling "force re-do GS split"

To prevent overfitting in the reconstructed density map, cryo-EM employs the "gold-standard" refinement technique. The core principle of this approach involves equally dividing the dataset into two halves. These halves are kept independent during the refinement process. The resolution is determined by comparing two density maps derived from these half-sets, using the Fourier Shell Correlation (FSC) as a controlling metric.

To ensure independence—that is, to avoid any crosstalk between the two halves—CryoSieve selects particles within each half set independently, thereby upholding the complete separation required by the "gold-standard" refinement principle. However, within each half set, the particles, due to undergoing a sieving process, exhibit shared selection features and hence are not entirely independent from one another. Therefore, independence is maintained only between the two half sets. Disabling this option allows CryoSPARC to preserve the original gold-standard split specified in the star file. If not disabled, CryoSPARC may re-split the two halves randomly, exposing to the risk of overfitting.