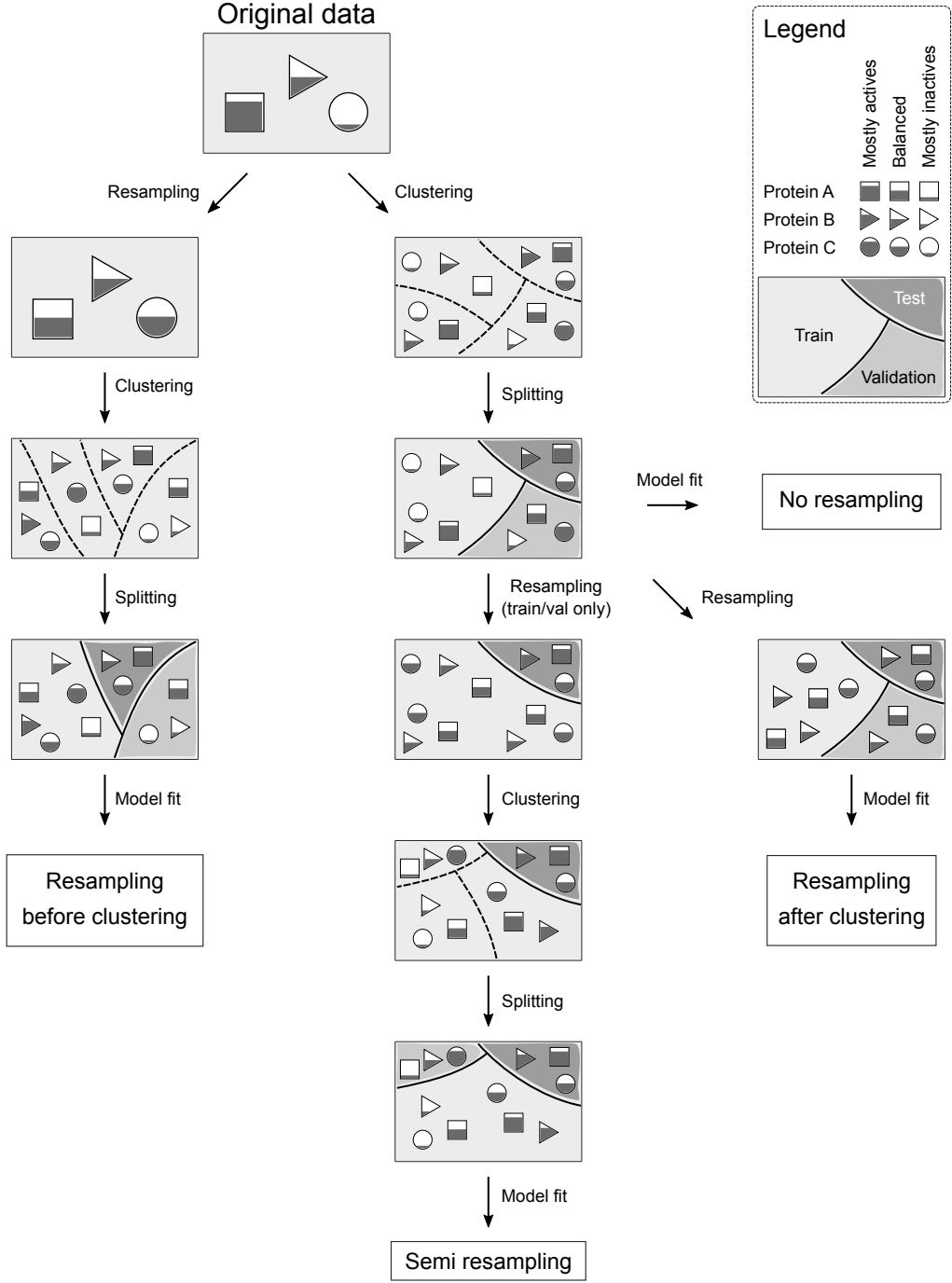
Original data
Resampling
Clustering
Legend
Mostly actives
Balanced
Mostly inactives
Protein A
Protein B
Protein C
Test
Train
Validation
Clustering
Splitting
Model fit
No resampling
Splitting
Resampling
(train/val only)
Resampling
Model fit
Clustering
Model fit
Resampling
before clustering
Resampling
after clustering
Splitting
Model fit
Semi resampling