We have a database of (~18000) known sequences
For each sequence we predict the resulting chromatogram (peak heights and position and quadratic peak spacing correction), and then bin the resulting channels (A,C,G,T) (the binning width is the average peak-peak distance)
The resulting 18000 (sequences) *1800 (bins) *5 (ACGTN) matrix is stored in **predbinSqr**

For the experimental mixture we have 5 sequences in equal concentration: Photobacter leignatti, Vibrio Fischeri, Escherichia coli, staphylococcus epidermidis, Enterococcus faecalis

The measured chromatogram is stretched and shifted (to agree with the predicted chromatogram of the mix – this is cheating to save time since we known the bacteria in the mix. Otherwise we would need to scan all shifts/streaches in the surrounding). It is then binned and the resulting 1300 (bins – since the experimental sequence doesn't cover all the 16S rRNA sequence) * 4 (ACGT) matrix is stored in **mixnvhbin**

We then try to reconstruct mixnvhbin using the database in predbinSqr using GPSR
We run it on a part (100-800) of the measured mixtures since this is where the predicted and measured chromatograms agree. I used tau=~10000 to get a more or less correct number of large non-zeros entries.

To test if we succeeded we look for the known sequence names in the results by testing the names of the largest predicted frequencies using **RedName** which is a 18000 cell array containing the name of each sequence in predbinSqr.

A sample how to use everything is in **TryItAll**

The fuction **TryGPSRBinAlign** does the work of trying to solve the minimization problem.

Enjoy ☺