Supplementary Material

1). Primers used for generation of WT-CXCR4-GFP and Δ34-CXCR4-GFP constructs

Table S1: Primer sequences

Mutant (code)	Code	Primer	Sequence
CXCR4 aa 1-318	Δ34	Fwd	5'-GGGAAGCTTACCATGGAGGGG-3'
		Rev	5'-GGGGAATTCGGGTGAGTGCGTGCTGGG-3'
CXCR4 full length	WT	Fwd	5'-GGGAAGCTTACCATGGAGGGG-3'

2). Cell lines stably expressing empty vector, wildtype or C-terminally truncated CXCR4 receptors fused to GFP

Cells either stably expressing WT-CXCR4-GFP or Δ34-CXCR4-GFP were FACS-sorted to obtain single clones. Single clones were expanded to populations and then analysed by flow cytometry for total receptor levels and surface receptor expression (Fig.S1).
Surface receptor expression (Fig.S1B) was measured after staining cells with allophycocyanin (APC)-conjugated pan anti-CXCR4 (N-terminal extracellular epitope, eBioscience, 45min) or isotype-APC (eBioscience), followed by fixation of the cells (4% PFA, 15min) and analysis on a FACSCalibur flow cytometer (BD Bioscience, channel FL-4). As the cells were antibody stained before fixation, only those CXCR4 receptors that resided on the plasma membrane were stained. In the same run, we also measured the total receptor levels expressed. These were analysed by measuring the GFP fluorescence (surface and internal receptors) in channel FL-1. Following this analysis, we chose one clone-derived population for each, wildtype and mutant receptor expressing cells, which were subsequently used for all further experiments.

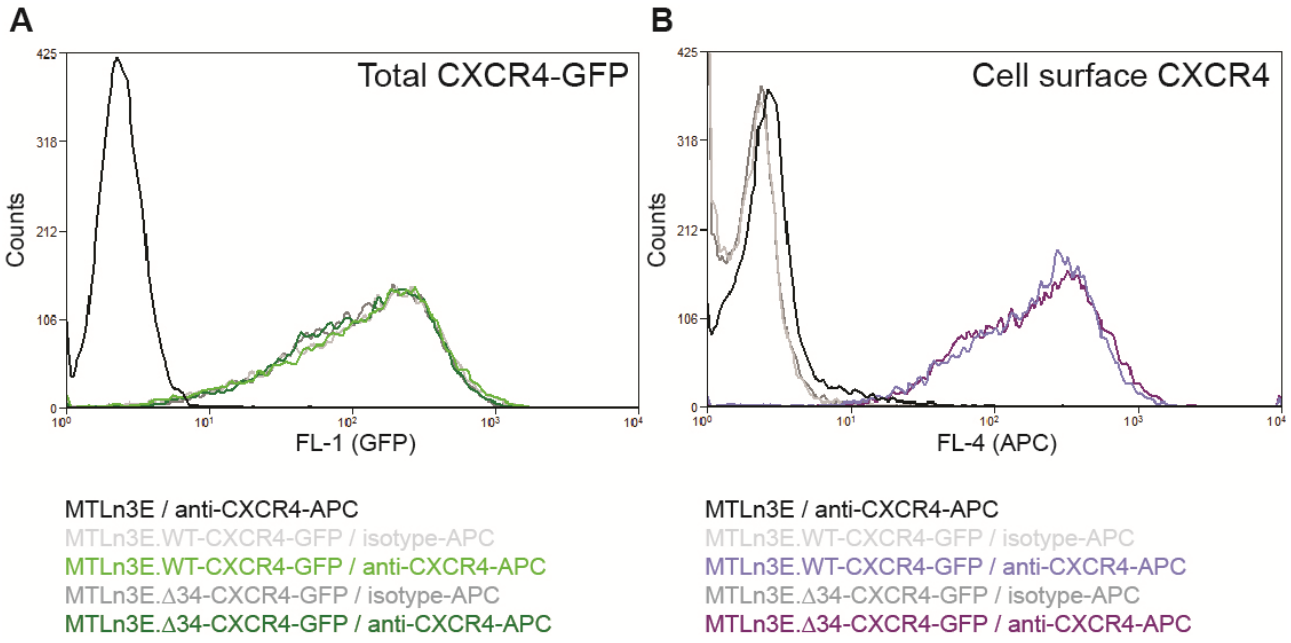

Figure S1: Total and surface receptor levels in cell lines stably expressing CXCR4 variants. (A) Total receptor expression was assessed by analysis of the GFP fluorescence (B) Surface receptor levels were assessed by antibody staining (pan anti-CXCR4-APC, N-terminal epitope) of live cells followed by cell fixation. Control experiments were performed using the corresponding isotype. Representative graphs of the flow cytometric analysis of the chosen clones are shown and each graph shows one out of the three experiments performed.

3). Permutation tests

A random permutation of the class assignments and subsequent cross-validation as described in the materials and methods section allows to judge the possibility that a particular Q^2 value occurs by chance. This method has been described in the literature as a reliable way to interpret the Q^2 values.[25] Figure S2-S4 (left) show the normalized histograms of the number of times that a particular Q^2 value is obtained in the 2000 × 4 different models, when the original class assignments are used (grey), as compared to randomly permutated class assignments (green). The amount of overlap between these two histograms reflects the quality of the predictions. Figure S2-S4 (right) show histograms for the number of components chosen for each of the models during repeated 4-fold cross-validation.

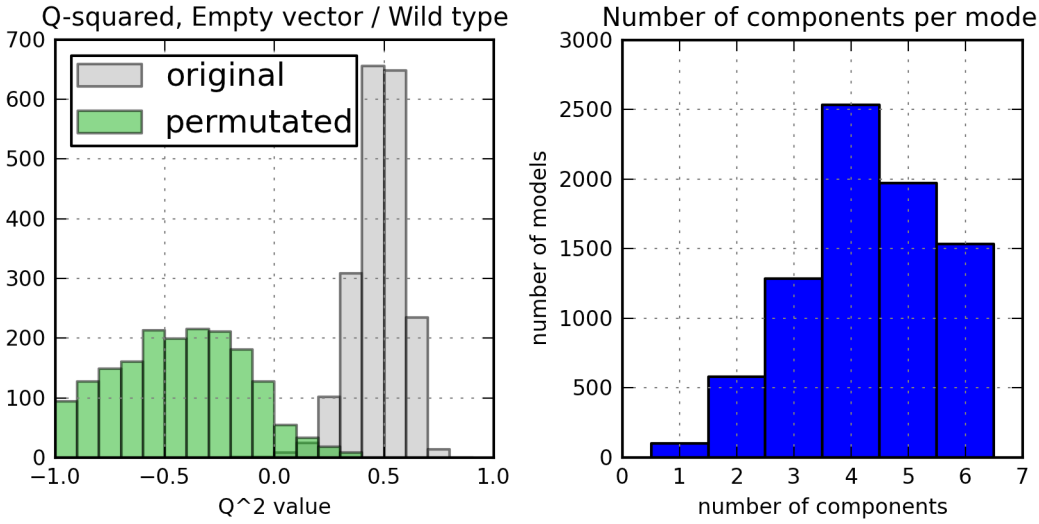

Figure S2: Empty vector vs. WT-CXCR4. Horizontal axis: Q^2 value or number of components, vertical axis: normalised or total number of models for which this value was obtained.

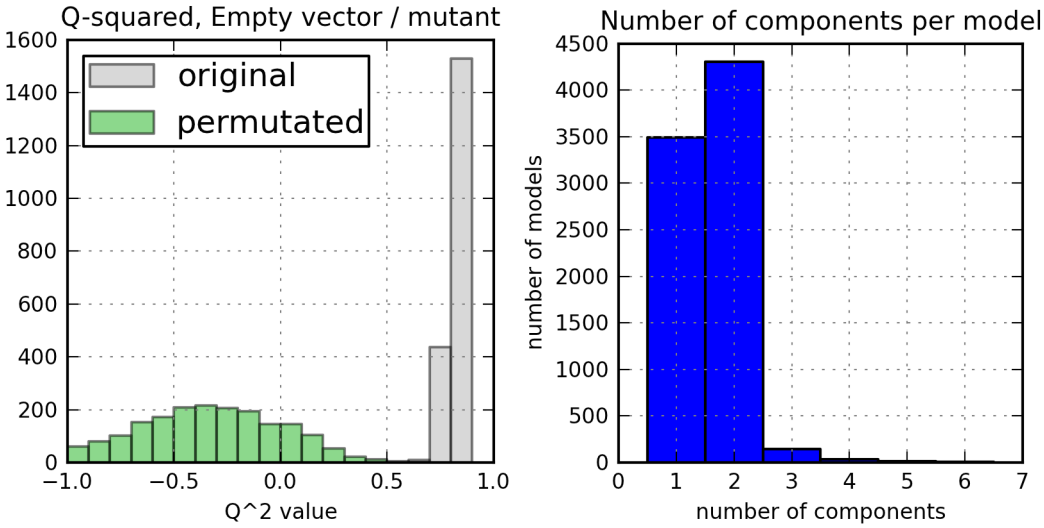

Figure S3: Empty vector vs. Δ34-CXCR4. Horizontal axis: Q^2 value or number of components, vertical axis: normalised or total number of models for which this value was obtained.

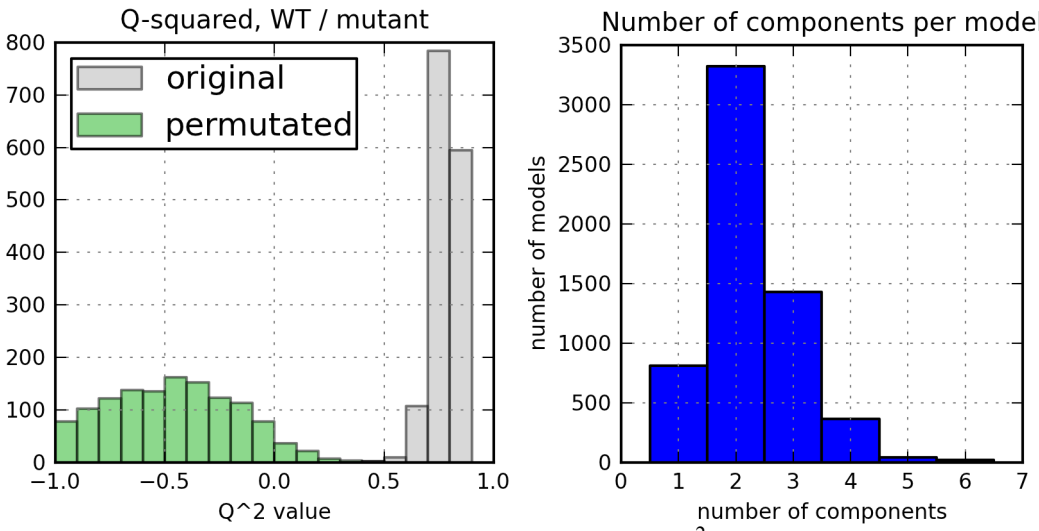

Figure S4: WT-CXCR4 vs. Δ34-CXCR4. Horizontal axis: Q^2 value or number of components, vertical axis: normalised or total number of models for which this value was obtained.

4). Assignments of NMR signals

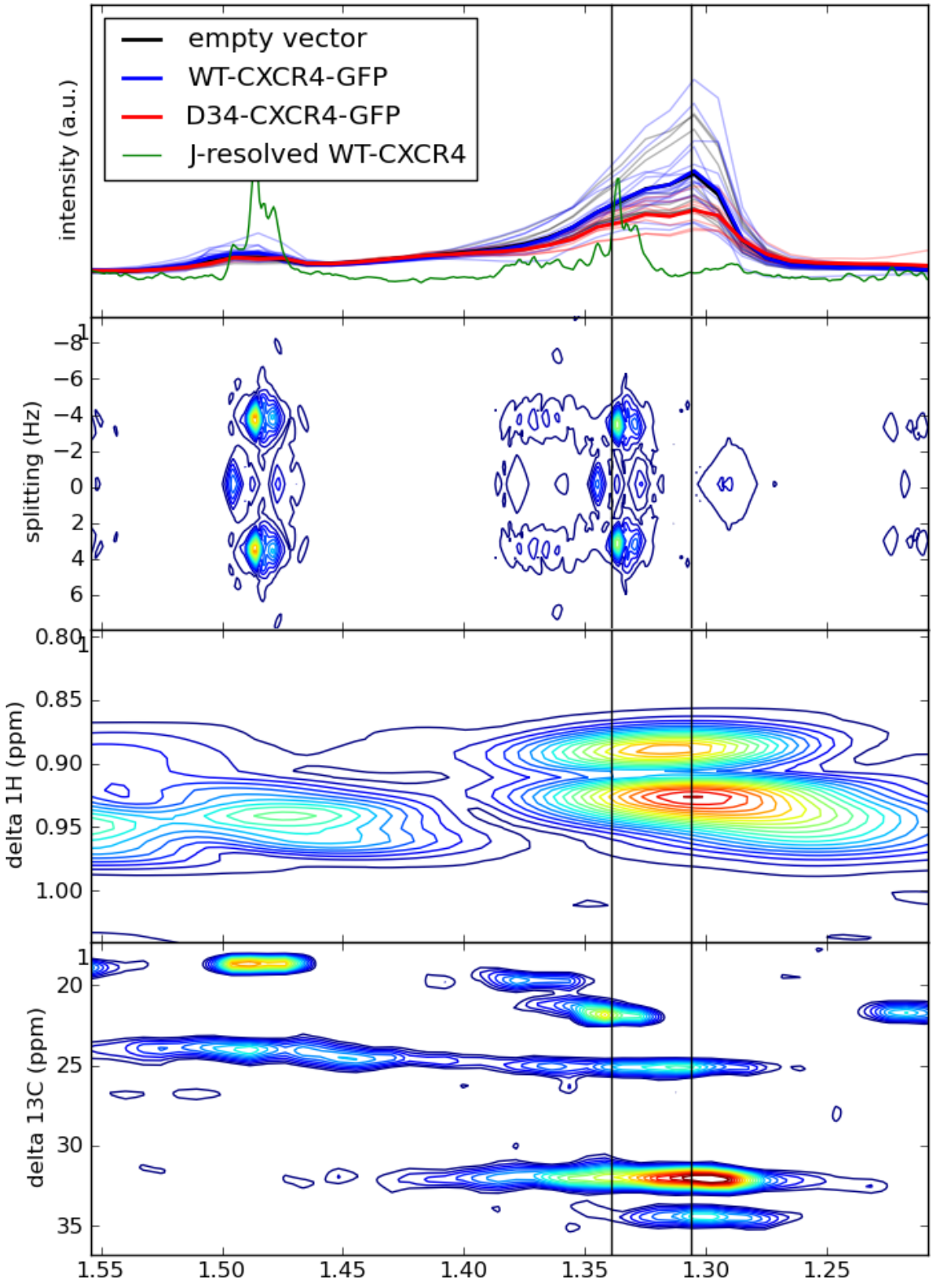

Figure S5: ^{13}C spectra were referenced to alanine at 1.48 ppm in the ^{1}H dimension and 18.70 ppm in the ^{13}C dimension. All ^{1}H spectra were referenced to TSP at 0 ppm. Fatty acids are assigned on the basis of their chemical shift at 1.33 ppm, the COSY cross-peak at 0.90 ppm, and ^{13}C chemical shifts at 25, 32 and 34 ppm. Also visible in this figure is lactate at 1.34 ppm, with a ^{13}C resonance at 22.7 ppm and a doublet in the J-resolved spectrum. The differences between the spectra of the different cell types occur in the broad component of the peak between 1.25 and 1.40 ppm, indicating lipids.

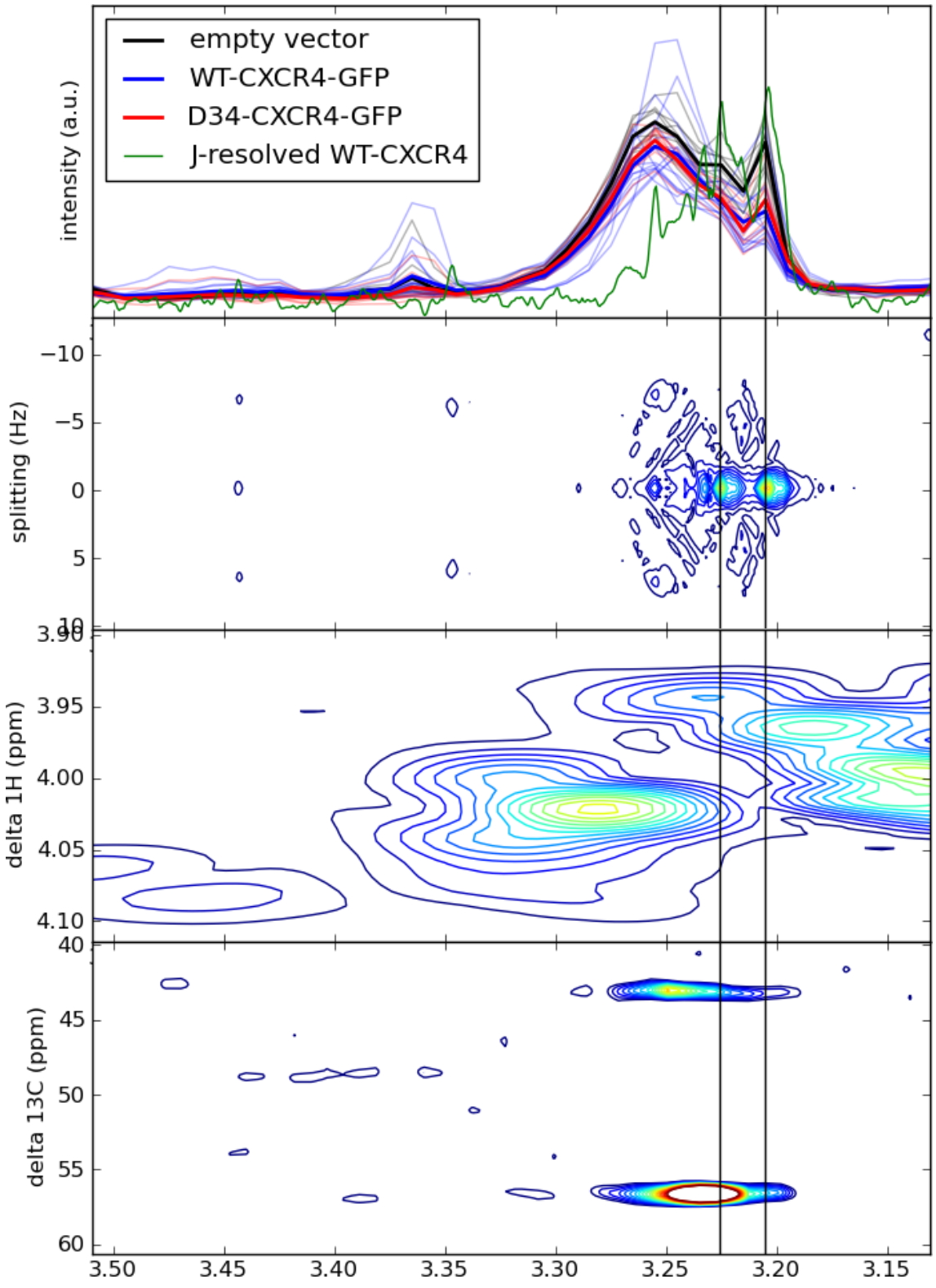

Figure S6: The peak at 3.21 ppm can be assigned to (total) choline on the basis of singlet at 3.21 ppm, and the HSQC cross-peak at a ^{13}C frequency of 56 ppm. The peak at 3.23 ppm with a ^{13}C cross-peak at 43 ppm is most likely phosphocholine.

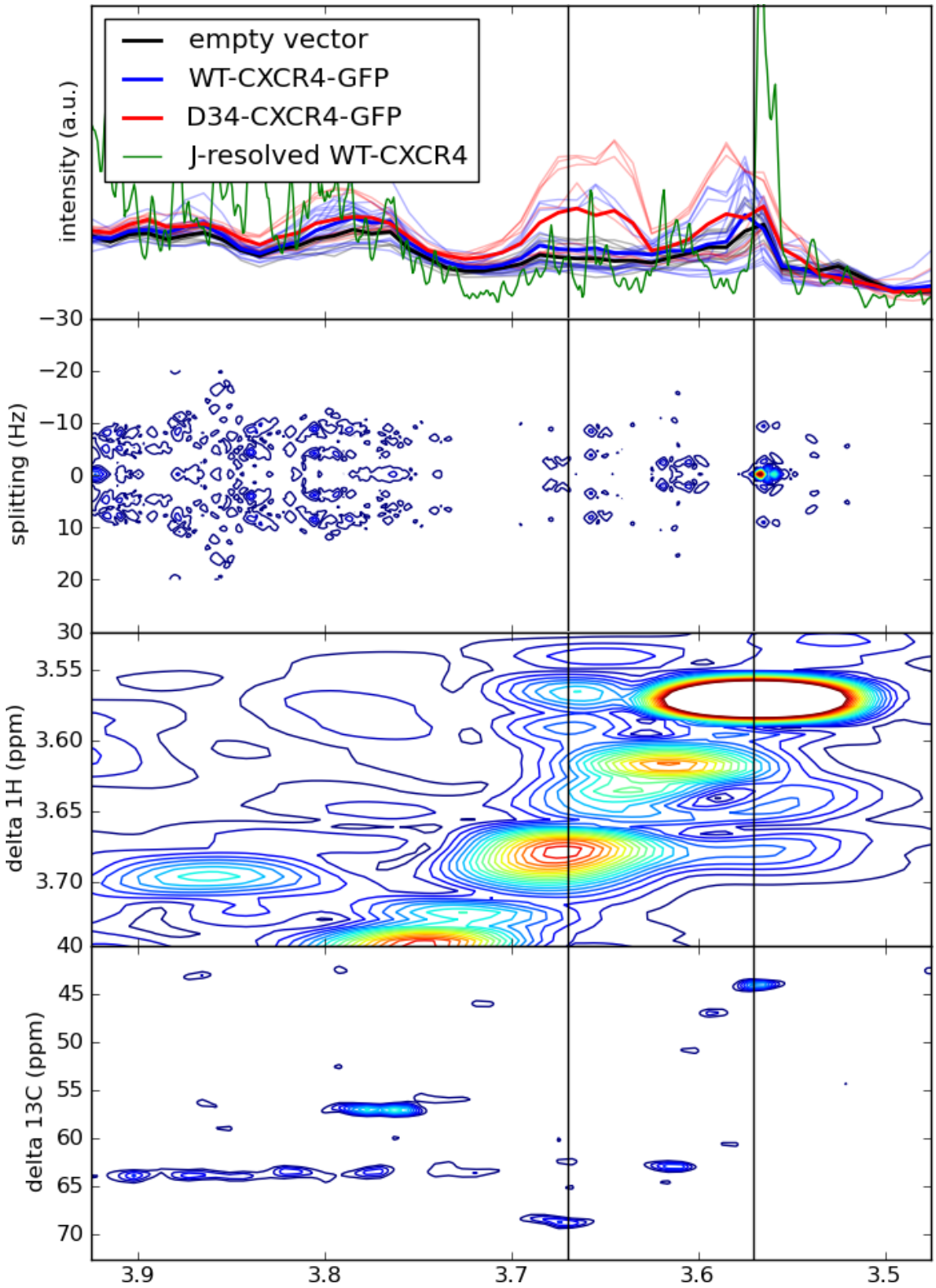

Figure S7: With the help of a STOCSY plot (S8), the resonance at 3.55 ppm is assigned to glycine based on the ^{1}H chemical shift and the ^{13}C cross-peak at 43 ppm. The peak at 3.66 ppm has a ^{13}C cross-peak at 68.58 ppm, possibly attributable to glycerophosphocholine (GPC). However the observation that these peaks are higher in the Δ34 cells does not correlate with the expected GPC peak at 3.23 ppm, which is lower in both the 4C and Δ34 cell types although this can be attributed to overlapping peaks in this region.

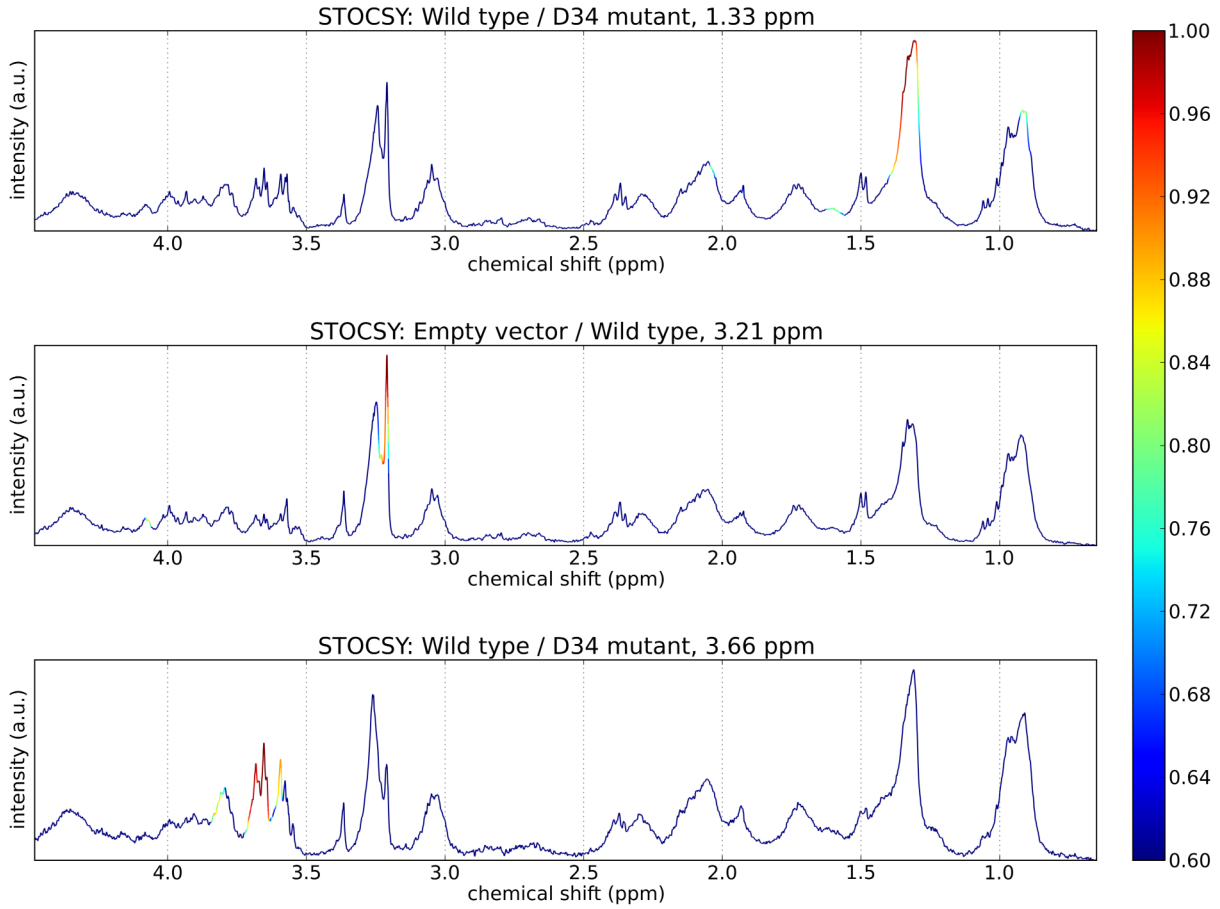

Figure S8: Statistical total correlation spectroscopy (STOCSY) plot of peaks assigned to lipids (top), choline (middle) and the unassigned peak at 3.66 ppm (bottom).