Supplement to "Optimized nonlinear gradients for reversed-phase liquid chromatography in shotgun proteomics"

Luminita Moruz,[†] Peter Pichler,[‡] Thomas Stranzl,[‡] Karl Mechtler,[‡,§] and Lukas Käll[*,¶,‖]

Science for Life Laboratory, Department of Biochemistry and Biophysics, Stockholm University, 17165 Solna, Sweden, Research Institute of Molecular Pathology, Bohr-Gasse 7, 1030 Vienna, Austria, and Science for Life Laboratory, School of Biotechnology, Royal Institute of Technology - KTH, 17165 Solna, Sweden

E-mail: lukas.kall@scilifelab.se

[*]To whom correspondence should be addressed

[†]Science for Life Laboratory, Department of Biochemistry and Biophysics, Stockholm University, 17165 Solna, Sweden

[‡]Research Institute of Molecular Pathology, Bohr-Gasse 7, 1030 Vienna, Austria

[¶]Science for Life Laboratory, School of Biotechnology, Royal Institute of Technology - KTH, 17165 Solna, Sweden

[§]Institute of Molecular Biotechnology, Vienna, Austria

[‖]Swedish e-Science Research Center, Royal Institute of Technology - KTH, 17165 Solna, Sweden

This document includes the following figures:

- Supplementary Figure S-1: Distribution of peptide identifications with linear gradients (2 hour runs)
- Supplementary Figure S-2: Retention time distributions with optimized gradients (2 hour runs)
- Supplementary Figure S-3: Peptides identified only with the MS1-optimized gradient (4 hour runs)
- Supplementary Figure S-4: Distribution of peptides identified using nonlinear gradients
- Supplementary Figure S-5: Peptides identified only with the *in silico*-optimized gradient (2 hour runs)
- Supplementary Figure S-6: Peptides identified only with the MS1-optimized gradient (2 hour runs)
- Supplementary Figure S-7: Chromatographic peak widths (2 hour runs)

Supplementary Figure S-1: Distribution of peptide identifications with linear gradients (2 hour runs)

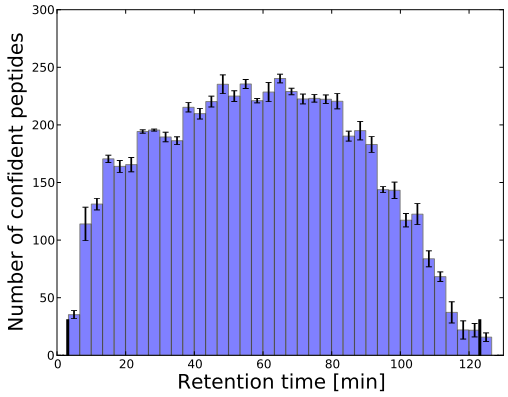

Figure S-1: **Uneven distribution of confident peptide identifications.** We display the average number of confident peptide identifications (FDR 1%) across the four replicates based on a 2 hour linear gradient as a function of retention time. The small segments on each bin indicate the standard deviation, while the two vertical black lines illustrate the start and end of the linear gradient.

Supplementary Figure S-2: Retention time distributions with optimized gradients (2 hour runs)

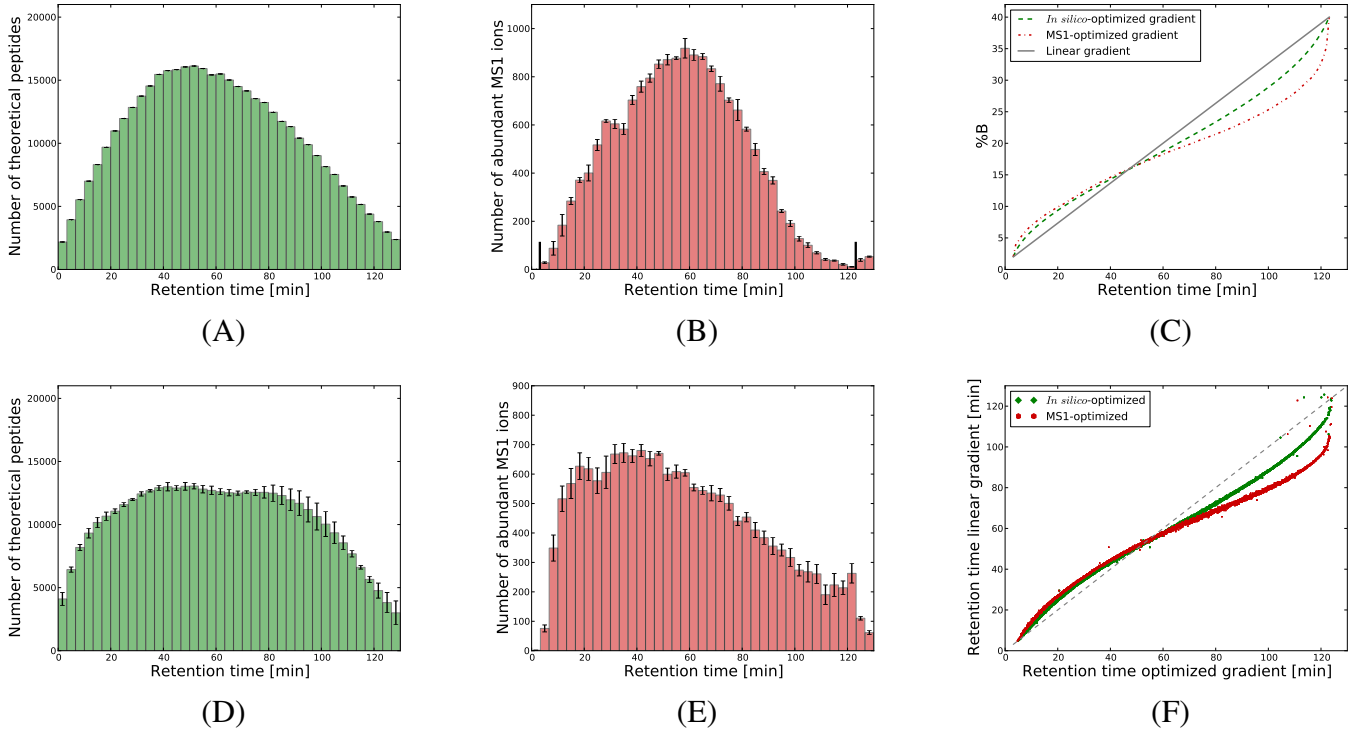

Figure S-2: **Nonlinear gradient functions.** In (A) we display the distribution of the predicted retention times for the theoretical peptides from an *in silico* digest of the human proteome when a linear gradient is used. (B) gives the average number of high-intensity MS1-features for the four replicates based on a linear gradient. In (C) we illustrate the *in silico*-optimized gradient designed to uniformize the distribution in (A), and one of the four MS1-optimized gradients calculated to even one of the distributions summarized in (B). (D) displays the average number of theoretical peptides as a function of predicted retention time when the *in silico*-optimized gradients were used. Similarly, (E) gives the average number of highly-abundant MS1-features yielded by the four replicates based on MS1-optimized gradients. The small segments on top of each bin give the standard deviation over the four replicates. In (F) we considered all the peptides identified at 1% FDR in both a run based on the linear gradient, and the corresponding runs based on the nonlinear gradients. We then plotted for each such peptide the retention time obtained with the linear gradient against the retention times in the runs based on the optimized gradients. All the figures correspond to 2 hour gradients.

Supplementary Figure S-3: Peptides identified only with the MS1-optimized gradient (4 hour runs)

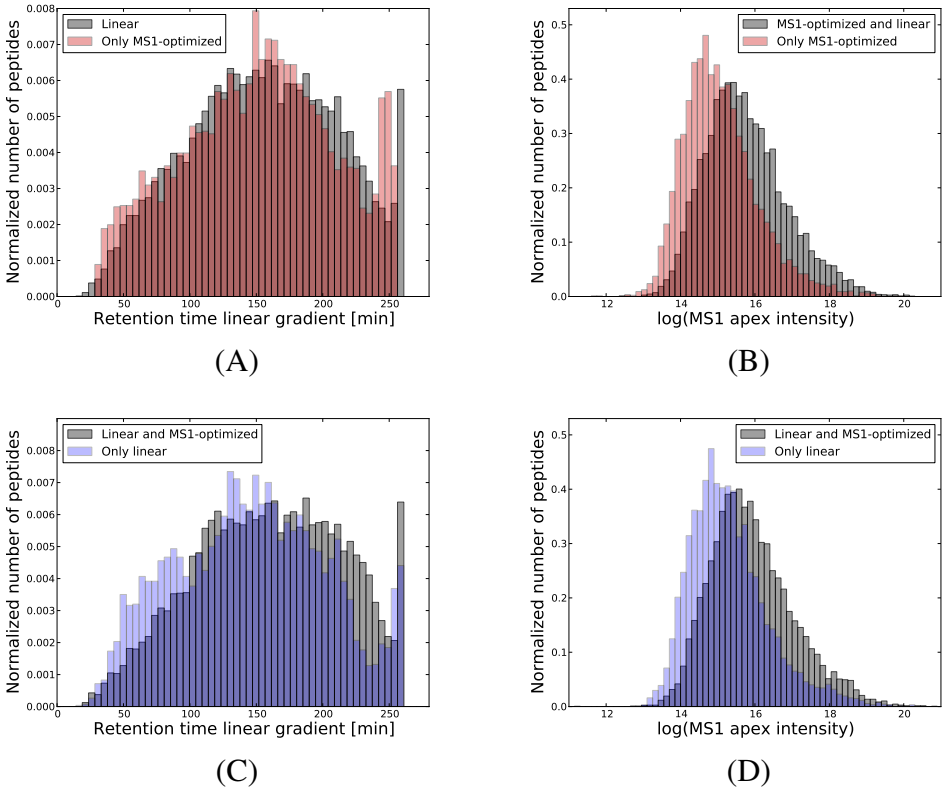

Figure S-3: **Peptides identified with only one type of gradient.** In (A) and (B) we give in red color the retention time distribution, and precursor intensities for the peptides identified at 1% FDR with the MS1-optimized, but not identified with the linear gradient. In (A) we give in gray the distribution of all the peptide identifications found with a linear gradient, while in (B) the gray distribution gives the precursor intensity of the common peptide identifications between the MS1-optimized gradient and the linear gradient. Figures (C) and (D) give similar representations for the peptides confidently identified with the linear gradient, but missed when using the MS1-optimized gradient. The red distribution in (A) gives the estimated retention time distribution according to a linear gradient. This was estimated by using a polynomial fit to relate the retention time of a peptide in a linear run with the retention time in the MS1-optimized run. This function was then used to translate the observed retention time of the peptides identified only with the MS1-optimized gradient, what would be the corresponding retention time when using a linear gradient.

Supplementary Figure S-4: Distribution of peptides identified using nonlinear gradients

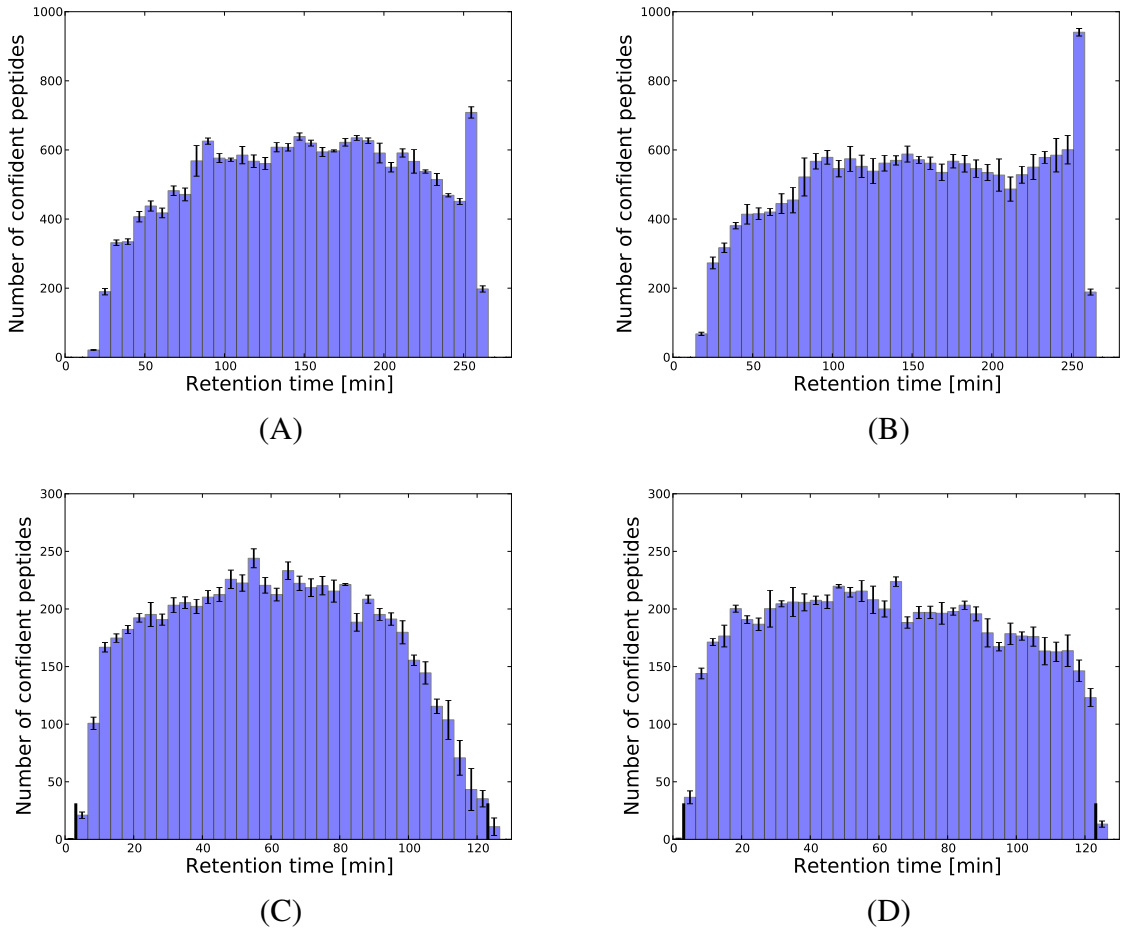

Figure S-4: **Distribution of confident peptide identifications.** (A) and (C) give the distribution of the peptides identified at 1% FDR using the *in silico*-optimized gradient for the 4 and 2 hour runs, respectively. Similarly, (B) and (D) display the distribution of confident peptide identifications when using the MS1-optimized gradients for 4 and 2 hour runs, respectively. The bars correspond to the average number of identifications, and the small black segments indicate the standard deviation across the four replicates run with each type of gradient.

Supplementary Figure S-5: Peptides identified only with the *in silico*-optimized gradient (2 hour runs)

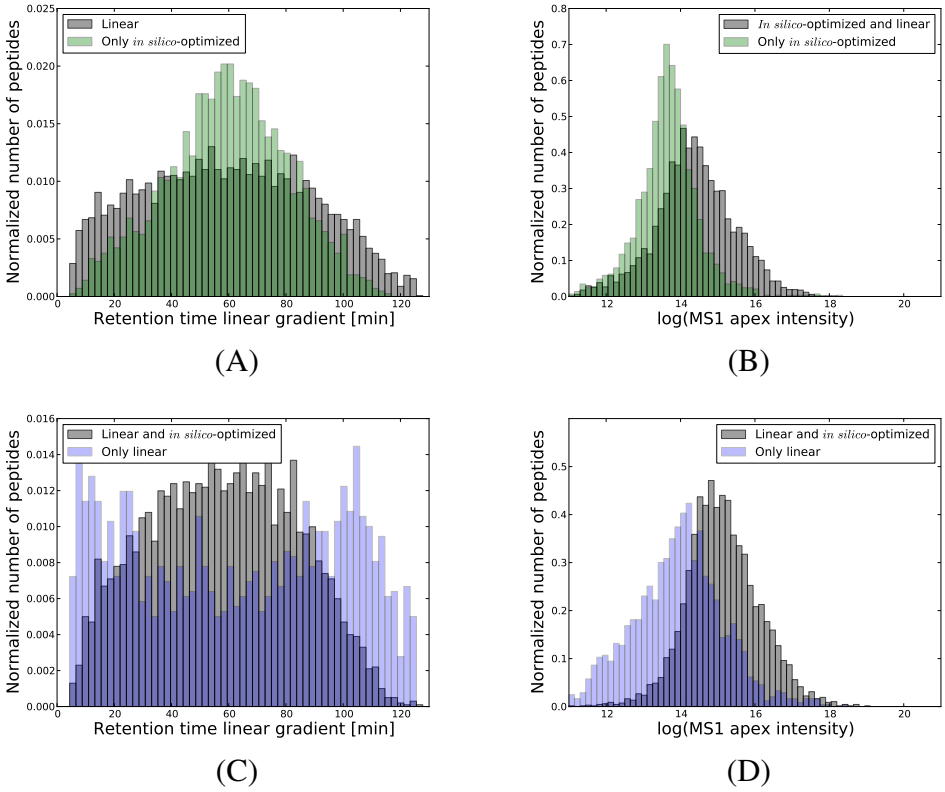

Figure S-5: **Peptides identified with only one type of gradient.** For one of the four replicates based on 2 hour gradient, we considered the peptides identified at 1% FDR when using the *in silico*-optimized gradient, but that were not identified with the linear gradient. In (A) we calculated the corresponding retention times that these peptides would have had if a linear gradient was used, and plotted the obtained distribution in green color. As a comparison, we give in gray the distribution of the confident peptides identified with the linear gradient. For the same peptides, (B) gives in green color the apex intensity of their precursor ions. In gray we display the precursor intensity of the common peptide identifications between the *in silico*-optimized and linear runs. (C) and (D) give similar representations for the peptides confidently identified with the linear gradient, but that were not present among the peptide identifications obtained with the *in silico*-optimized gradient.

Supplementary Figure S-6: Peptides identified only with the MS1-optimized gradient (2 hour runs)

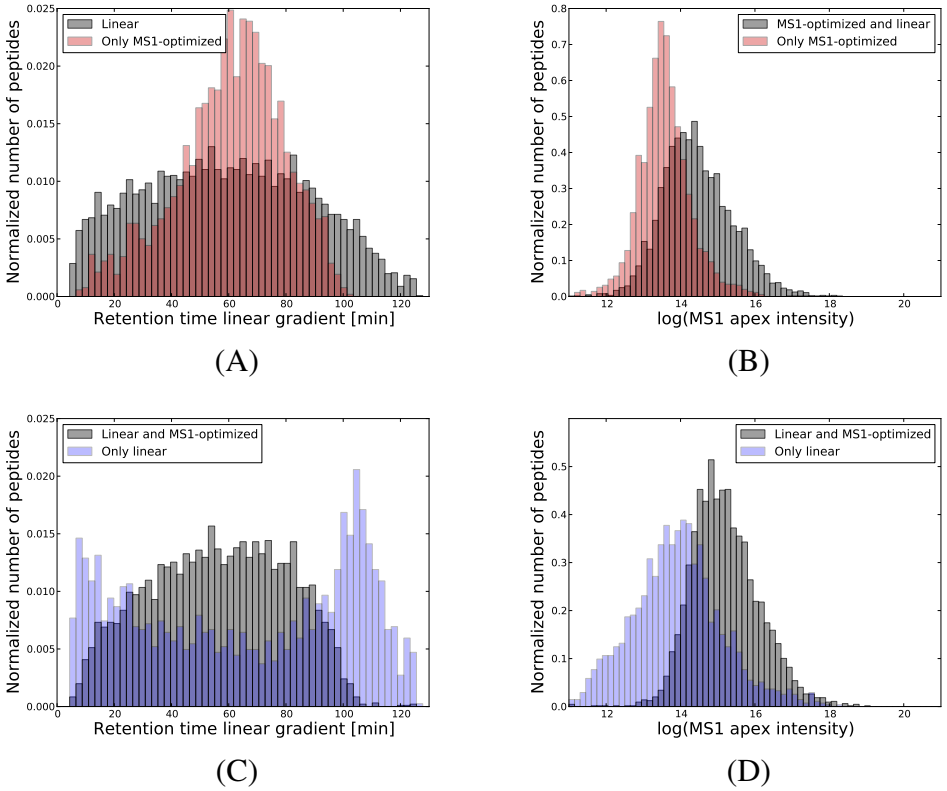

Figure S-6: **Peptides identified with only one type of gradient.** In (A) and (B) we give in red color the retention time distribution, and precursor intensities for the peptides identified at 1% FDR with the MS1-optimized, but not identified with the linear gradient. In (A) we give in gray the distribution of all the peptide identifications found with a linear gradient, while in (B) the gray distribution gives the precursor intensity of the common peptide identifications between the MS1-optimized gradient and the linear gradient. Figures (C) and (D) give similar representations for the peptides confidently identified with the linear gradient, but missed when using the MS1-optimized gradient. The red distribution in (A) gives the estimated retention time distribution according to a linear gradient. This was estimated by using a polynomial fit to relate the retention time of a peptide in a linear run with the retention time in the MS1-optimized run. This function was then used to translate the observed retention time of the peptides identified only with the MS1-optimized gradient, to the corresponding retention time when using a linear gradient.

Supplementary Figure S-7: Chromatographic peak widths (2 hour runs)

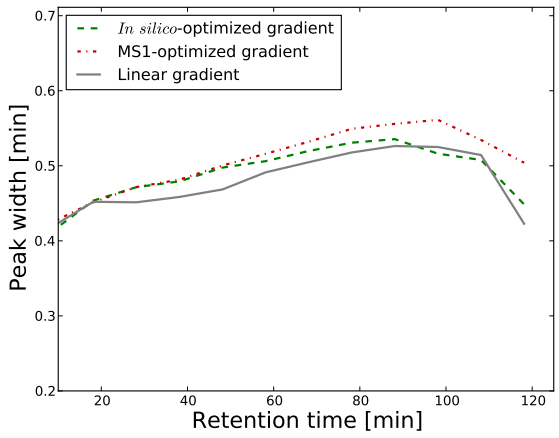

Figure S-7: **Chromatographic peak widths.** For each type of gradient, we display the estimated peak widths as a function of the retention time. The graphs corresponds to one of the four replicates.