

Table S1: Notation summary

$\underline{m}$	complete spectral dataset consisting of N MS/MS spectra
$\underline{m}_n$	the n -th MS/MS spectrum of the spectra dataset ($\underline{m}$)
K_n	number of m/z fragment peaks in the n -th MS/MS spectrum
m_{nk}	the k -th m/z fragment peak of the n -th MS/MS spectrum
$\underline{c}_n$	candidate list/set for the n -th MS/MS spectrum
C_n	size/length of candidate list/set $\underline{c}_n$
c_{nc}	c - th candidate of candidate list/set $\underline{c}_n$
f_{nck}	fragment-structure (of candidate c_{nc}) assigned to m/z fragment peak m_{nk}
$(\underline{m}_n, f_{nc})$	peak fragment assignment of candidate c_{nc} to MS/MS spectrum $\underline{m}_n$
$\perp$	placeholder indicating a non-annotated annotated m/z fragment peak
$\tilde{f}_k$	fingerprint of fragment-structure f_k
$\tilde{M}_{tr}$	reduced m/z fragment peak domain (for training MS/MS spectra)
$\tilde{F}_{tr}$	reduced fingerprint domain (for training MS/MS spectra)
$\mathcal{D}_{train}$	list of single peak fragment assignments
$\underline{m}_q$	query MS/MS spectrum
m_{qk}	the k -th m/z fragment peak of the query MS/MS spectrum
f_{qck}	fragment-structure (of candidate c_{qc}) assigned to m/z fragment peak m_{qk}
K_q	number of m/z fragment peaks in the query MS/MS spectrum
C_q	candidate list/set of query MS/MS spectrum $\underline{m}_1$
$\tilde{M}$	reduced m/z fragment peak domain (for training MS/MS spectra + query MS/MS spectrum)
$\tilde{F}$	reduced fingerprint domain (for training MS/MS spectra + query MS/MS spectrum)
$\theta_{mf} \sim$	probability of fragment fingerprint $\tilde{f}$ given m/z fragment peak m
θ_{mf}^{ML}	maximum likelihood estimator for $\theta_{mf} \sim$
$N_{mf} \sim$	absolute frequency of observing fragment fingerprint $\tilde{f}$ and m/z fragment peak m
$N_{m\perp}$	absolute frequency of observing a non-annotated m/z fragment peak m
$\pi_{mf} \sim$	hyper parameter for prior distribution
α, β	pseudo counts for prior distribution
θ_{mf}^{ML}	mean posterior estimator for $\theta_{mf} \sim$
l_{nkh}	m/z fragment loss between m/z fragment peaks m_{nk} and m_{nh}
f_{nchk}	fragment-substructure of f_{nck} excluding f_{nch} assigned to l_{nkh}
$\tilde{L}_{tr}$	reduced m/z fragment loss domain (for training MS/MS spectra)
$\tilde{F}_{tr}^L$	reduced loss fingerprint domain (for training MS/MS spectra)
$\mathcal{D}_{train}^L$	list of single loss fragment assignments
$\tilde{M}^L$	reduced m/z fragment loss domain (for training MS/MS spectra + query MS/MS spectrum)
$\tilde{F}^L$	reduced loss fingerprint domain (for training MS/MS spectra + query MS/MS spectrum)
$\psi_{mf} \sim$	hyper parameter for prior distribution of fragment loss model
α^L, β^L	pseudo counts for prior distribution of fragment loss model
$\phi_{lf} \sim$	probability of loss fragment fingerprint $\tilde{f}$ given m/z fragment peak m
ϕ_{lf}^{ML}	maximum likelihood estimator for $\psi_{lf} \sim$
N_{lf}^L	absolute frequency of observing fragment loss fingerprint $\tilde{f}$ and m/z fragment loss m
$S_{MetFrag}^c$	MetFrag score of a candidate c
S_{Peak}^c	statistical score evaluating fragment - m/z peak assignments of a candidate c
S_{Loss}^c	statistical score evaluating loss fragment - m/z loss assignments of a candidate c
$S_{RawPeak}^c$	non-normalized statistical score evaluating fragment - m/z peak assignments of a candidate c
$\omega_1, \omega_2, \omega_3$	single score weights
S_{Fin}^c	final/consensus score of a candidate c