Supplemental material section S6 for

Prediction of LCMSMS properties of peptides from sequence by deep learning

Shenheng Guan1,2, Michael F. Moran2,3, and Bin Ma1

1 David R. Cheriton School of Computer Science, University of Waterloo, Waterloo, N2L 3G1, Canada

2 Program in Cell Biology and SPARC BioCentre, Hospital for Sick Children, 686 Bay St, Toronto, ON, M5G 0A4, Canada

3 Department of Molecular Genetics, University of Toronto, 686 Bay St, Toronto, ON, M5G 0A4, Canada

We provide the following information as the supplemental materials:

1. A Python function for encoding peptide sequence (with PSI format modifications)

one-hot-encode\_peptide\_ion.py

1. Training data for the three models (iRT/irt\_reg\_data.pickle, ChargeState/zfit\_one\_hot.pickle, and HCDspectrumPredict/ucsd\_hcd\_splib\_2ndhalf.pickle)
2. Trained models in \*.h5 format (iRT/irt\_reg\_data\_filtered\_bidirLSTM2\_masking\_model.h5, ChargeState/zfit\_bidirLSTM2\_masking\_model.h5, HCDspectrumPredict/ucsd\_hcd\_splib\_2ndhalf\_model.h5)
3. Predicted data (iRT/irt\_reg\_data\_filtered\_bidirLSTM2\_masking\_result.pickle, ChargeState/zfit\_bidirLSTM2\_masking\_result.pickle, and HCDspectrumPredict/ucsd\_hcd\_splib\_2ndhalf\_ions.pickle)
4. Codes to look at the results (iRT/plot\_irt.py,

ChargeState/plot\_zfit\_result.py,

and HCDspectrumPredict/plot\_predicted\_spectrum.py