Figure S1. Plots of Offset Frequency Functions for prefix/suffix of HCD spectra mentioned in the algorithm section. In the plot of the prefix OFF, the 3 most intense peaks are located at 1.0061, -26.9885 and -17.0041, and we can confidently infer their corresponding ion types as *b, a, and b-H$_2$O* (Table 2). In the plot of the suffix OFF, the peak located at 19.0159 is undoubtedly attributed to *y*-ions, and the next two peaks at 20.0189 and 21.0217 are mainly due to isotopic peaks of *y*-ions. The average distance between two adjacent peaks in this cluster is 1.0029, which is very close to the theoretical value, 1.0034. The offsets at 1.9905 and 1.0065 correspond to *y-NH$_3$* and *y-H$_2$O* ions. In addition, we also chose to consider the offset at 10.0106, which is due to the *y^{2+}* ions, in the pNovo algorithm. It appears in the triply charged spectra with a relatively higher frequency.

Table S1. Probabilities of common ion types appeared in HCD data, with the consideration of mass regions and the relations between these types of ions.

	total	low	medium	high
Prob (*a*)	0.200	0.464	0.097	0.051
Prob (*b*)	0.286	0.0.533	0.217	0.111
Prob(*y*)	0.668	0.661	0.783	0.539
Prob(y^0)	0.155	0.235	0.118	0.125
Prob(y^*)	0.177	0.273	0.156	0.111
Prob(y^{2+})	0.100	0.033	0.071	0.199
Prob(y^{2+})[#]	0.388	0.089	0.413	0.634
Prob($y|y^*$)	0.932	0.924	0.948	0.922
Prob($y|y^0$)	0.970	0.951	0.989	0.982
Prob($y^0|y$)	0.225	0.338	0.149	0.228
Prob($y^*|y$)	0.247	0.382	0.190	0.189
Prob(*b*|*a*)	0.739	0.691	0.877	0.867
Prob (*a*|*b*)	0.518	0.602	0.392	0.398
Prob (*b*|*y*)	0.328	0.159	0.254	0.649
Prob (*y*|*b*)	0.766	0.657	0.917	0.952
Prob($y|y^{2+}$)[#]	0.520	0.977	0.889	0.188

only triply charged spectra are considered.

Note: y^0 and $y*$ denote y-ions with a neutral loss of water and ammonia, respectively. Three regions, low, medium and high ones, are computed by evenly split the range between 0 and the value of the peptide molecular weight plus a Proton.

Table S2. Proteins used in the Experiments and their corresponding IDs in Swiss-Prot database (v.56.2)

Protein	ID in Swiss-Prot database
Myosin	Q28641
Glycogen phosphorylase	P00489
Serum albumin	P02769
Beta-galactosidase	P00722
Carbonic anhydrase	P00921
Trypsin inhibitor	P01070
Ovalbumin	P01012
Lysozyme	P00698

Table S3. Parameters of database search.

Item	Setting in pFind and Mascot
Database	Target-reversed strategy is used and the target database consists of the proteins in Table S2.
Enzyme	Trypsin
Maximum missed cleavage sites	2
Precursor tolerance	± 10 ppm
Fragment tolerance	± 0.01Da
Fixed Modifications	Carbamidomethylation (C)