

Supplementary Table S2: Summary of the set of sequences with experimentally annotated phosphorylation sites. NR: non-redundant, here defined as having less than 40% sequence identity in the 25 residue-long fragment centered at S, T or Y.

Residue	Num. sequences	Num. sites	Num. non-sites	NR sites	NR non-sites
Serine	16,083	34,620	687,051	19,537	360,652
Threonine	7,228	7,867	201,785	4,612	121,149
Tyrosine	4,470	4,700	50,880	2,194	24,477