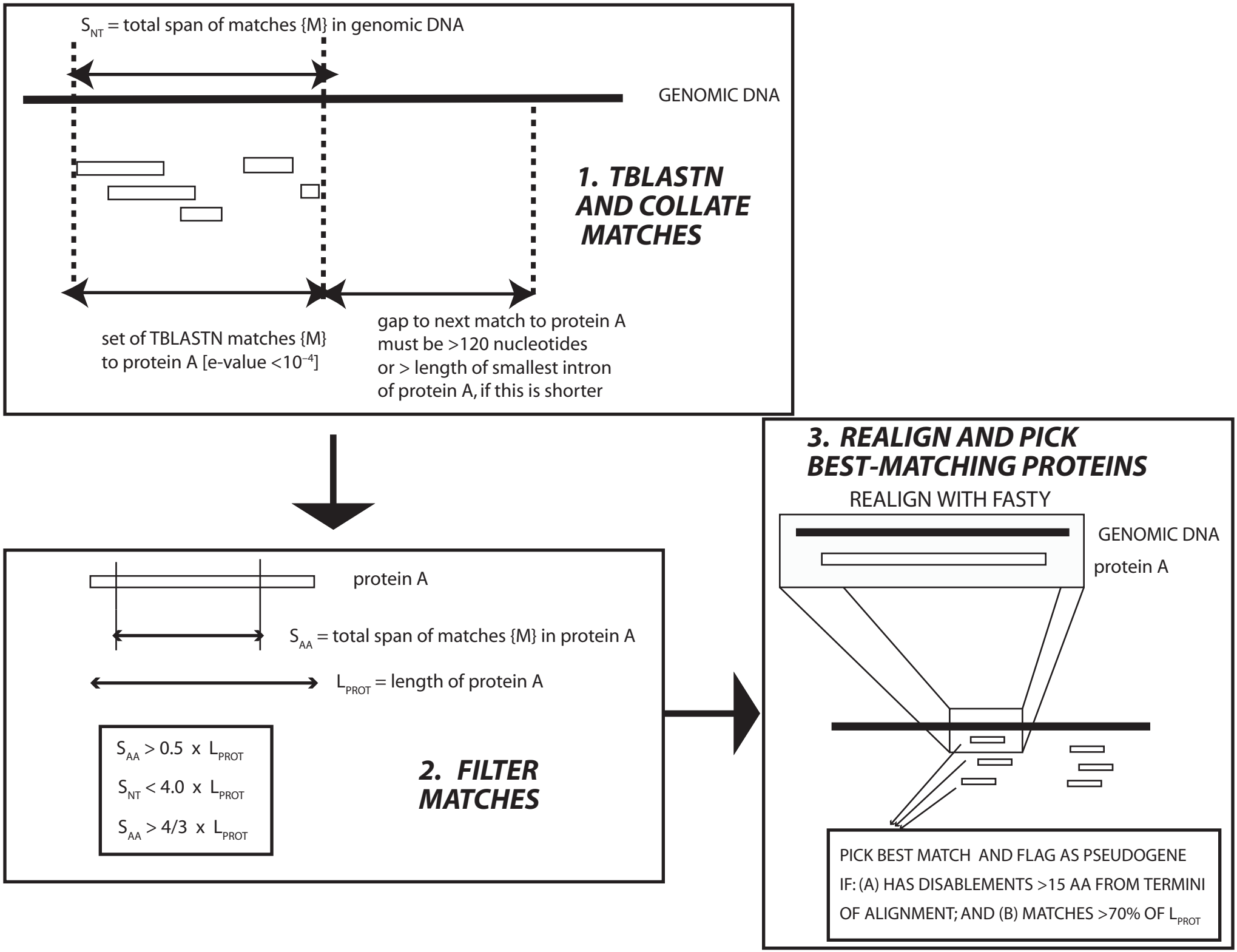
S_{NT} = total span of matches {M} in genomic DNA
GENOMIC DNA
1. TBLASTN AND COLLATE MATCHES
set of TBLASTN matches {M} to protein A [e-value <10^{-4}]
gap to next match to protein A must be >120 nucleotides or > length of smallest intron of protein A, if this is shorter
protein A
S_{AA} = total span of matches {M} in protein A
L_{PROT} = length of protein A
$S_{AA} > 0.5 \times L_{PROT}$
$S_{NT} < 4.0 \times L_{PROT}$
$S_{AA} > 4/3 \times L_{PROT}$
2. FILTER MATCHES
3. REALIGN AND PICK BEST-MATCHING PROTEINS
REALIGN WITH FASTY
GENOMIC DNA
protein A
PICK BEST MATCH AND FLAG AS PSEUDOGENE
IF: (A) HAS DISABLEMENTS >15 AA FROM TERMINI OF ALIGNMENT; AND (B) MATCHES >70% OF L_{PROT}