

Table 7. Expected vs. actual frequency of the three-residue amyloidogenic motif VII surrounded by amyloid breakers in protein databases

| Motif* | $P_{VIIXX} (P_{XXVII})^\dagger$ | Nonredundant protein sequences (NR)* | | TrEMBL (Tr)* | | Swiss-Prot and TrEMBL (SP + Tr)* | | SwissProt (SP)* | |
|------------------------------|---------------------------------|--------------------------------------|----------------------------|----------------------------------|----------------------------|----------------------------------|----------------------------|----------------------------------|----------------------------|
| | | $O_{VIIXX} (O_{XXVII})^\ddagger$ | $E_{VIIXX} (E_{XXVII})^\S$ | $O_{VIIXX} (O_{XXVII})^\ddagger$ | $E_{VIIXX} (E_{XXVII})^\S$ | $O_{VIIXX} (O_{XXVII})^\ddagger$ | $E_{VIIXX} (E_{XXVII})^\S$ | $O_{VIIXX} (O_{XXVII})^\ddagger$ | $E_{VIIXX} (E_{XXVII})^\S$ |
| VII ++ | 3.699e ⁻⁶ | 492 | 470.07 | 401 | 469.28 | 489 | 469.44 | 88 | 115.63 |
| ++ VII | 3.699e ⁻⁶ | 1,022 | 470.07 | 794 | 469.28 | 1,006 | 469.44 | 213 | 115.63 |
| VII -- | 3.968e ⁻⁶ | 1,209 | 504.24 | 995 | 503.39 | 1,201 | 503.56 | 206 | 124.04 |
| -- VII | 3.968e ⁻⁶ | 1,739 | 504.24 | 1,552 | 503.39 | 1,744 | 503.56 | 193 | 124.04 |
| P VII | 1.298e ⁻⁵ | 3,471 | 1,648.8 | 2,877 | 1,646.0 | 3,395 | 1,646.6 | 517 | 405.59 |
| VII P | 1.298e ⁻⁵ | 3,192 | 1,648.8 | 2,686 | 1,646.0 | 3,193 | 1,646.6 | 475 | 405.59 |
| P + VII | 1.453e ⁻⁶ | 319 | 184.66 | 211 | 184.35 | 265 | 184.42 | 55 | 45.426 |
| VII + P | 1.453e ⁻⁶ | 244 | 184.66 | 179 | 184.35 | 226 | 184.42 | 48 | 45.426 |
| P - VII | 1.505e ⁻⁶ | 594 | 191.26 | 517 | 190.94 | 595 | 191.00 | 79 | 47.048 |
| VII - P | 1.505e ⁻⁶ | 272 | 191.26 | 199 | 190.94 | 259 | 191.00 | 61 | 47.048 |
| + P VII | 1.453e ⁻⁶ | 480 | 184.66 | 395 | 184.35 | 464 | 184.42 | 69 | 45.426 |
| VII P + | 1.453e ⁻⁶ | 300 | 184.66 | 241 | 184.35 | 299 | 184.42 | 58 | 45.426 |
| - P VII | 1.505e ⁻⁶ | 332 | 191.26 | 279 | 190.94 | 326 | 191.00 | 48 | 47.048 |
| VII P - | 1.505e ⁻⁶ | 336 | 191.26 | 265 | 190.94 | 336 | 191.00 | 72 | 47.048 |
| Mean difference [¶] | | 518.006 | | 378.516 | | 501.935 | | 37.256 | |
| df [¶] | | 13 | | 13 | | 13 | | 13 | |
| t value [¶] | | 3.2572 | | 3.0847 | | 3.2103 | | 3.3759 | |
| t probability [¶] | | 0.006242 | | 0.009455 | | 0.006831 | | 0.004968 | |

*Motif database scanning has been carried out by using PATTINPROT (1).

[†]Motif probabilities have been calculated as the product of the expected individual amino acid frequencies (see *Methods* in the main text).

[‡] $O_{VIIXX} (O_{XXVII})$ is the number of hits of an amyloid breaker motif that have been found on the database.

[§] $E_{VIIXX} (E_{XXVII})$ is the expected number of hits of a given amyloid breaker motif based on the average composition of proteins (see *Methods* in the main text) and the total number of VII hits found in database [$O_{VII} (NR) = 37,473$; $O_{VII} (Tr) = 37,410$; $O_{VII} (SP + Tr) = 37,423$; $O_{VII} (SP) = 369$].

[¶]Student's test of significance. VII is surrounded by amyloid breakers at a confidence level higher than 1%.

^{||}VII ++ is the only amyloid breaker motif found as expected from amino acid composition.

1. Combet, C., Blanchet, C., Geourjon, C. & Deleage, G. (2000) *Trends Biochem. Sci.* **25**, 147-150.