Table 5. Expected vs. actual frequency of amyloidogenic three-residue motifs in protein sequences

Motif*	P_X[†]	$R_{E(ADW)}$[‡] P_X/P_{ADW}	$R_{E(VIS)}$[‡] P_X/P_{VIS}	Nonredundant protein sequences*				TrEMBL*			
				$R_{O(ADW)}$[§]	$R_{O(VIS)}$[§]	$R_E/R_{O(ADW)}$[¶] ± 0.178[‖]	$R_E/R_{O(VIS)}$[¶] ± 0.167[‖]	$R_{O(ADW)}$[§]	$R_{O(VIS)}$[§]	$R_E/R_{O(ADW)}$[¶] ± 0.226[‖]	$R_E/R_{O(VIS)}$[¶] ± 0.220[‖]
VII	2.949e^{-4}	5.765	1.067	2.716	0.526	2.122	2.026	2.301	0.436	2.506	2.447
LEF	3.039e^{-4}	5.941	1.099	3.304	0.640	1.798	1.716	3.300	0.625	1.800	1.759
WEF	3.377e^{-5}	0.660	0.1222	0.524	0.102	1.259	1.202	0.366	0.069	1.802	1.761
LNF	1.770e^{-4}	3.458	0.640	4.130	0.801	0.837	0.799	4.127	0.782	0.838	0.819
VIV	3.318e^{-4}	6.486	1.200	4.933	0.955	1.315	1.256	4.968	0.952	1.306	1.2611
ADW	5.116e^{-5}	—	0.186	—	0.194	—	0.960	—	0.189	—	0.982
VIS	2.765e^{-4}	5.405	—	5.159	—	1.045	—	5.279	—	1.024	—

Motif*	P_X[†]	$R_{E(ADW)}$[‡] P_X/P_{ADW}	$R_{E(VIS)}$[‡] P_X/P_{VIS}	Swiss-Prot and TrEMBL*				Swiss-Prot*			
				$R_{O(ADW)}$[§]	$R_{O(VIS)}$[§]	$R_E/R_{O(ADW)}$[¶] ± 0.167[‖]	$R_E/R_{O(VIS)}$[¶] ± 0.161[‖]	$R_{O(ADW)}$[§]	$R_{O(VIS)}$[§]	$R_E/R_{O(ADW)}$[¶] ± 0.097[‖]	$R_E/R_{O(VIS)}$[¶] ± 0.115[‖]
VII	2.949e^{-4}	5.765	1.067	2.825	0.540	2.041	1.975	4.510	0.705	1.278	1.514
LEF	3.039e^{-4}	5.941	1.099	3.317	0.634	1.791	1.734	3.830	0.598	1.551	1.837
WEF	3.377e^{-5}	0.660	0.1222	0.520	0.099	1.270	1.229	0.453	0.071	1.457	1.726
LNF	1.770e^{-4}	3.458	0.640	4.181	0.799	0.827	0.801	3.347	0.523	1.033	1.224
VIV	3.318e^{-4}	6.486	1.200	4.989	0.955	1.300	1.256	5.809	0.904	1.117	1.327
ADW	5.116e^{-5}	—	0.186	—	0.191	—	0.974	—	0.156	—	1.191
VIS	2.765e^{-4}	5.405	—	5.231	—	1.033	—	6.402	—	0.844	—

*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).

[‡]R_E is the expected distribution of a given motif X with respect to a control amino acid sequence C (ADW or VIS): $R_{E(ADW)} = P_X/P_{ADW}$ and $R_{E(VIS)} = P_X/P_{VIS}$.

[§]R_O is the observed representation of a given three-residue motif X with respect to a control amino acid sequence (ADW or VIS) on the protein database: $R_{O(ADW)} = O_X/O_{ADW}$ and $R_{O(VIS)} = O_X/O_{VIS}$.

[¶]$R_E/R_O > 1$, X is underrepresented; $R_E/R_O \approx 1$, X appears as expected; $R_E/R_O < 1$, X is overrepresented.

[‖]± SE.

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