Table 1: Red-teaming attack success rates of SafeProtein against DPLM2 on SafeProtein-Bench. Results are reported for all masking strategies. For each sequence masking ratio, the best success rate is highlighted in bold.

Gen Strategy	Masking Strategy	**Sequence Masking Rat**				
		0.1	0.2	0.25	0.3	0
Strategy1	Conservation	**36.36**	**29.84**	**26.81**	**21.45**	**15**
	Random	18.88	9.56	10.02	6.06	7
	Tail	8.86	2.33	3.50	1.87	1
Strategy2	Conservation	**42.66**	**34.50**	**32.40**	**27.97**	**20**
	Random	21.45	9.79	10.72	6.99	7
	Tail	5.83	2.10	4.20	2.56	1
Strategy3	Conservation	**44.29**	**33.10**	**30.54**	**26.57**	**20**
	Random	19.58	8.63	10.02	7.46	6
	Tail	9.32	1.63	4.66	2.80	2