

**Fig. 5.** Scatter plot between contact density and  $\text{Tr}C^8/N$  for an ensemble of 3,000 randomly collapsed 36-mer structure. Correlation coefficient  $R = 0.49$ ;  $P < 10^{-4}$ .

**Fig. 6.** Same as Fig. 2 of the text but for  $\text{Tr}C^4$  and  $\text{Tr}C^6$ . The data for these histograms is binned into bins of size 0.5 for  $\text{Tr}C^4$  (a) and 3 for  $\text{Tr}C^6$  (b).

**Fig. 7.** Scatter plot between domain length and normalized eighth contact trace ( $\text{Tr}C^8/N$ ) for all Dali domains. A weak positive correlation,  $R = 0.08$ , is exhibited.

**Fig. 8.** Contact density distribution in prokaryotic only domains (red) and eukaryotic domains (black). The binning of histograms is the same as Fig. 4a (CD bins of size 0.5); however, the data for each plot are collected for Dali domains that fall into narrow length ranges: 100-150 amino acids (a), 150-200 amino acids (b), and 200-250 amino acids (c). In all three intervals, the eukaryotic innovation domains exhibit statistically lower contact densities with KS  $P$  values  $< 10^{-5}$  (a), 0.02 (b), and  $< 10^{-4}$  (c).

**Table 1. Amino acid compositions of evolved sequences and their SD (in parentheses) calculated over all considered evolved or natural protein domains**

Parameter	Amino acid group					
	ALMVIC	FYWH	QNST	KR	ED	GP
Evolved MJ Table 6 of ref. 24	0.34 (0.05)	0.13 (0.04)	0.17 (0.04)	0.12 (0.02)	0.14 (0.02)	0.11 (0.04)
Evolved MJ96	0.35 (0.04)	0.13 (0.03)	0.15 (0.03)	0.14 (0.03)	0.11 (0.04)	0.13 (0.02)
Dali domains (prokaryotic only)	0.34 (0.07)	0.11 (0.04)	0.19 (0.06)	0.11 (0.04)	0.12 (0.04)	0.12 (0.07)
Dali domains (eukaryotic innovation)	0.32 (0.05)	0.12 (0.04)	0.21 (0.05)	0.12 (0.05)	0.13 (0.04)	0.11 (0.04)

The data for two evolved ensembles of proteins with different sets of parameters: Miyazawa–Jernigan (MJ) from table 6 of ref. 24 and parameters from ref. 23 are given in the first two rows and amino acids for prokaryotic only and eukaryotic innovation Dali domains are shown in the last two rows. The amino acids are grouped in six groups as shown.

FIG. 5

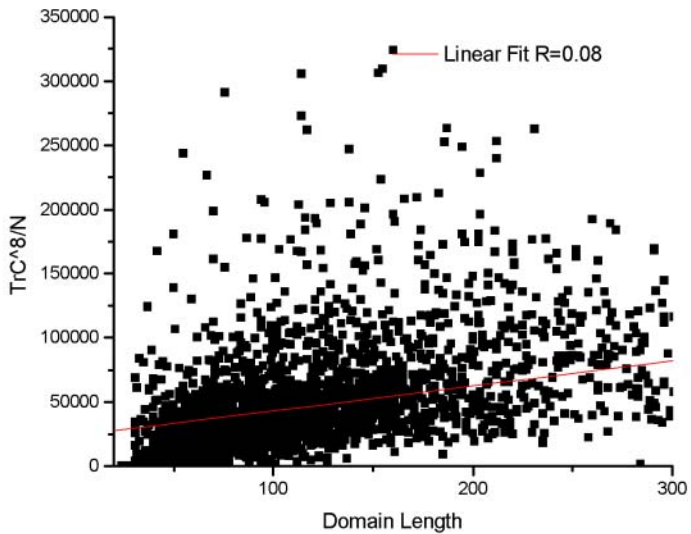
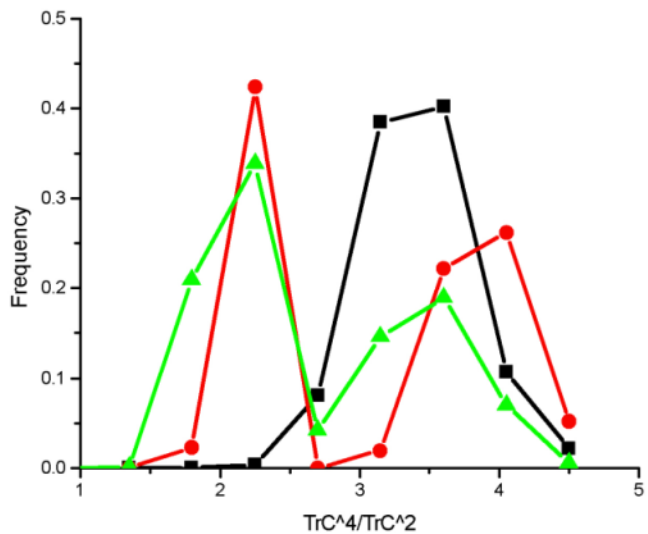
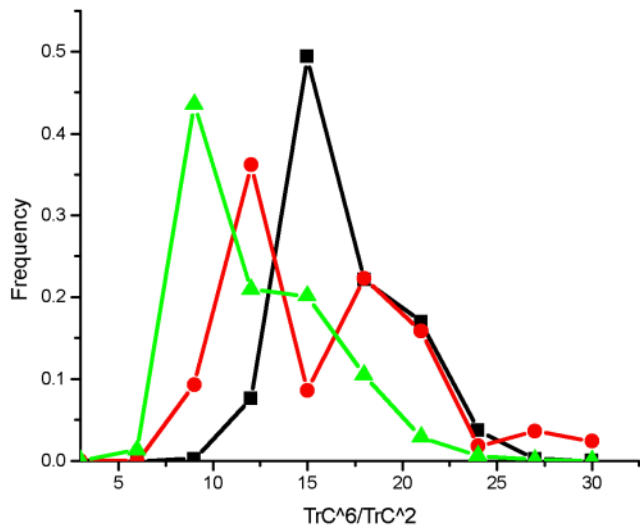


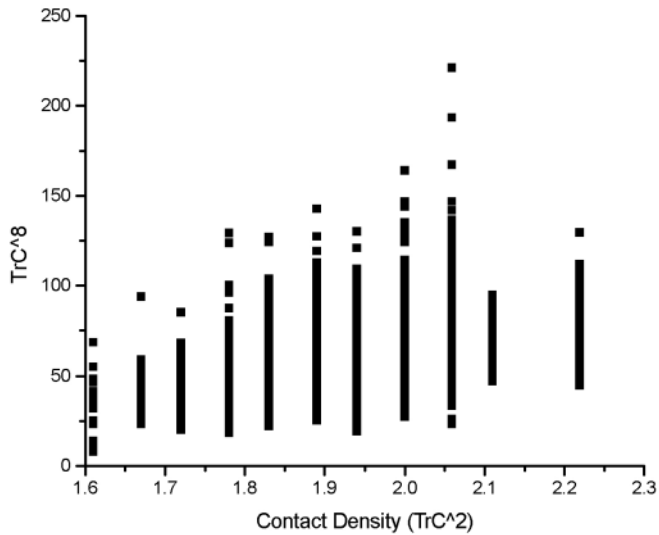
FIG. 6

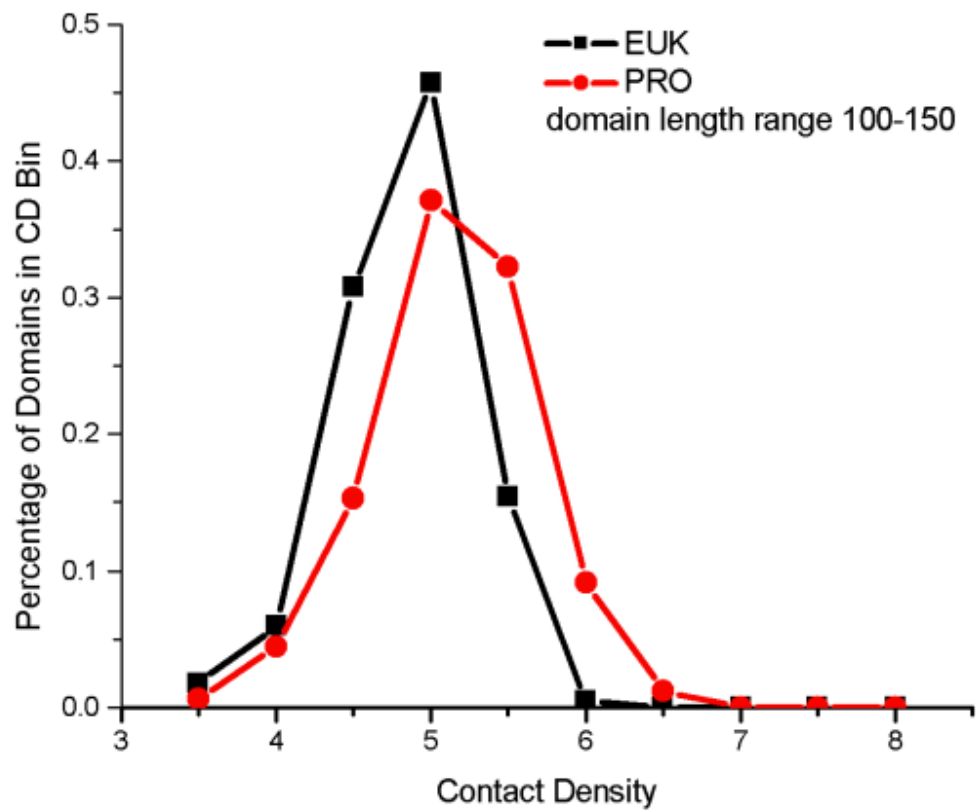


a

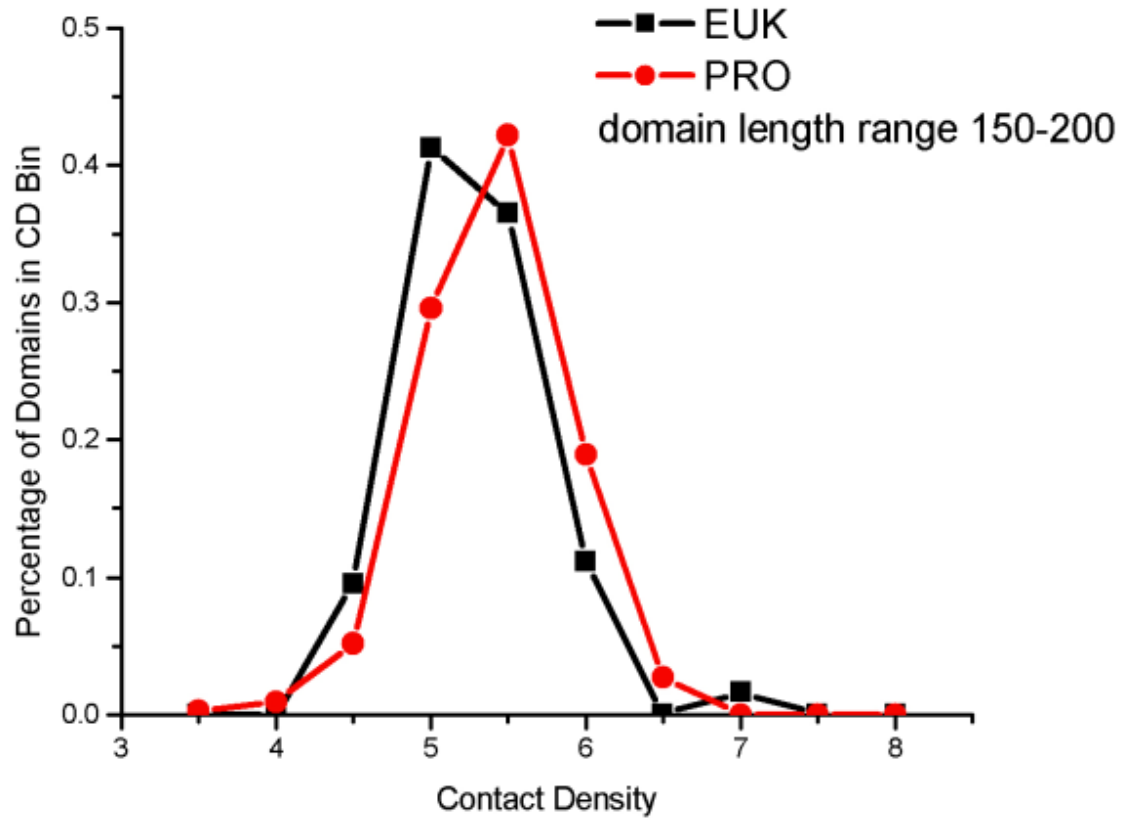


b

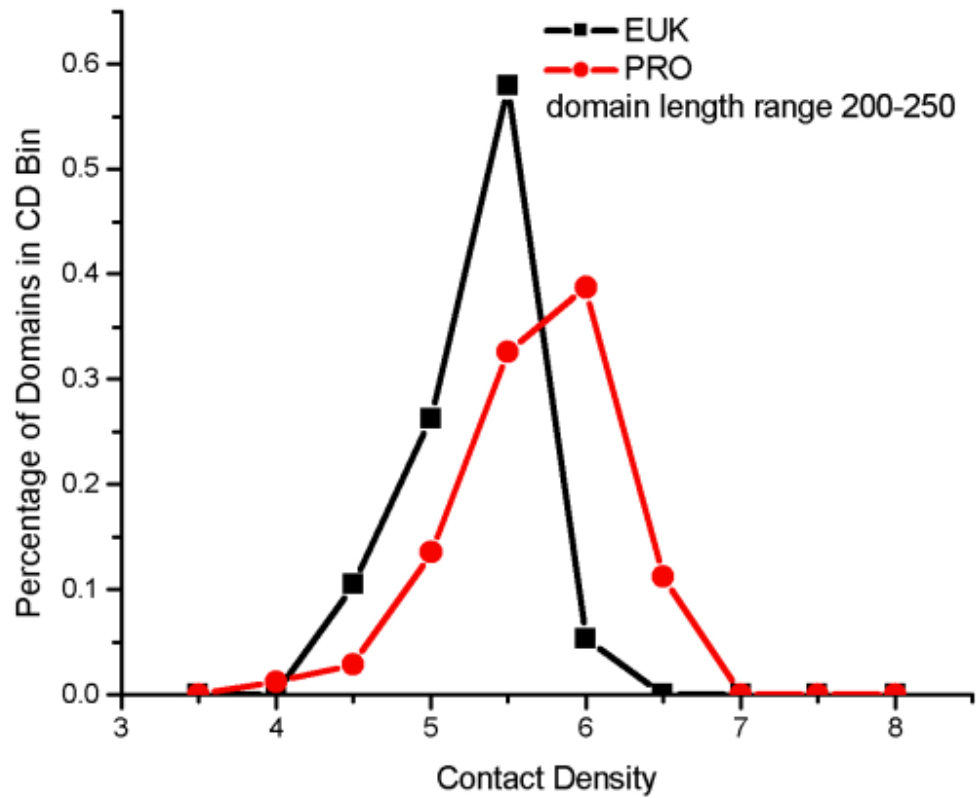




a



b



C