Homology modeling in a dynamical world

Alexander Miguel Monzon[1], Diego Javier Zea[2], Cristina Marino-Buslje[2, †, *], Gustavo Parisi[1, †, *]

[1] Departamento de Ciencia y Tecnología, Universidad Nacional de Quilmes, CONICET, Bernal, Argentina

[2] Structural Bioinformatics Unit, Fundación Instituto Leloir, CONICET, C1405BWE, Ciudad Autónoma de Buenos Aires, Buenos Aires, Argentina

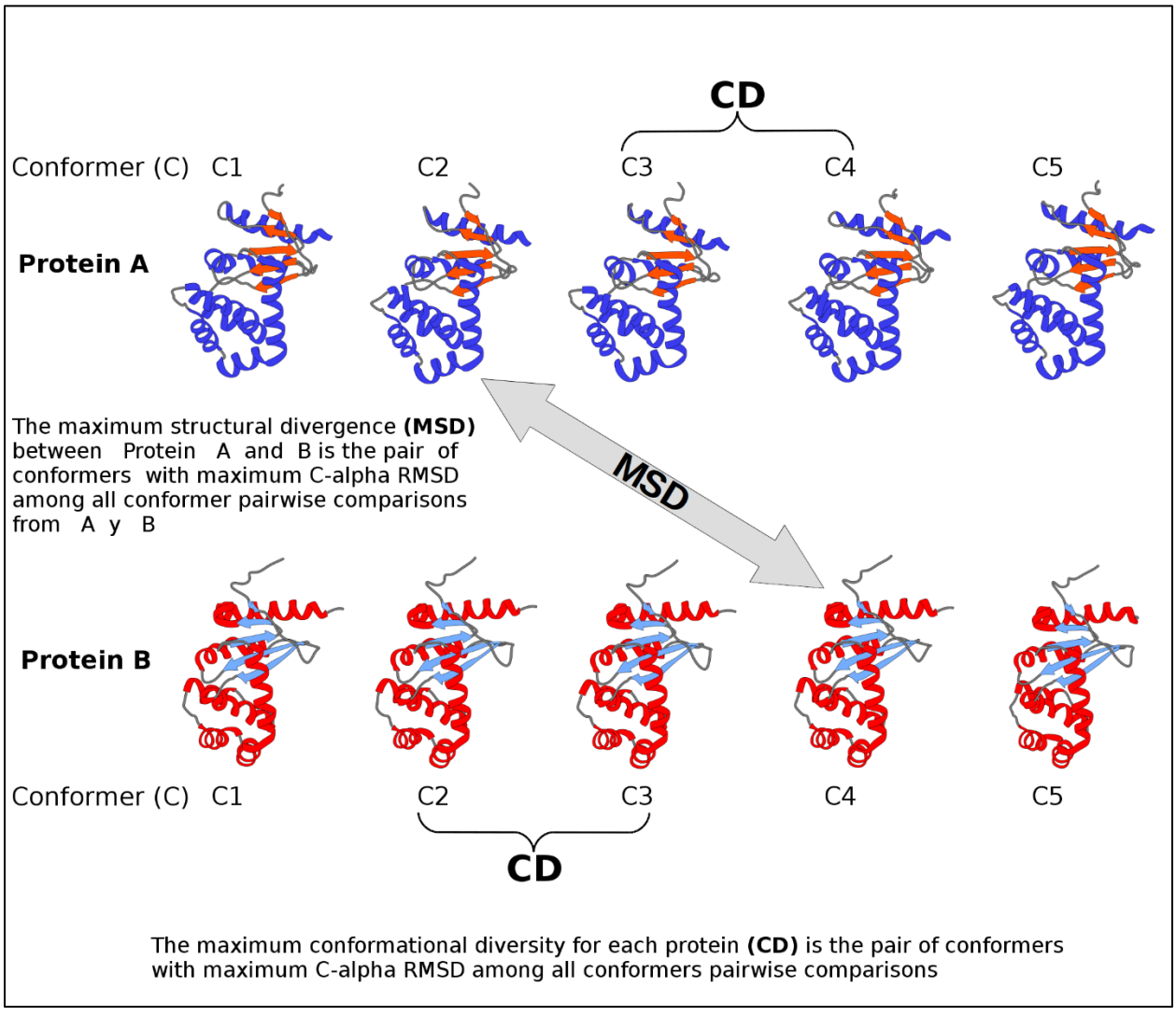
[†]these authors contributed equally to this work

*Correspondence to: gusparisi@gmail.com, Roque Sáez Peña 182, Bernal (B1876BXD) Argentina; cmb@leloir.org.ar, Av. Patricias Argentinas 435, Ciudad Autónoma de Buenos Aires (C1405BWE), Argentina.

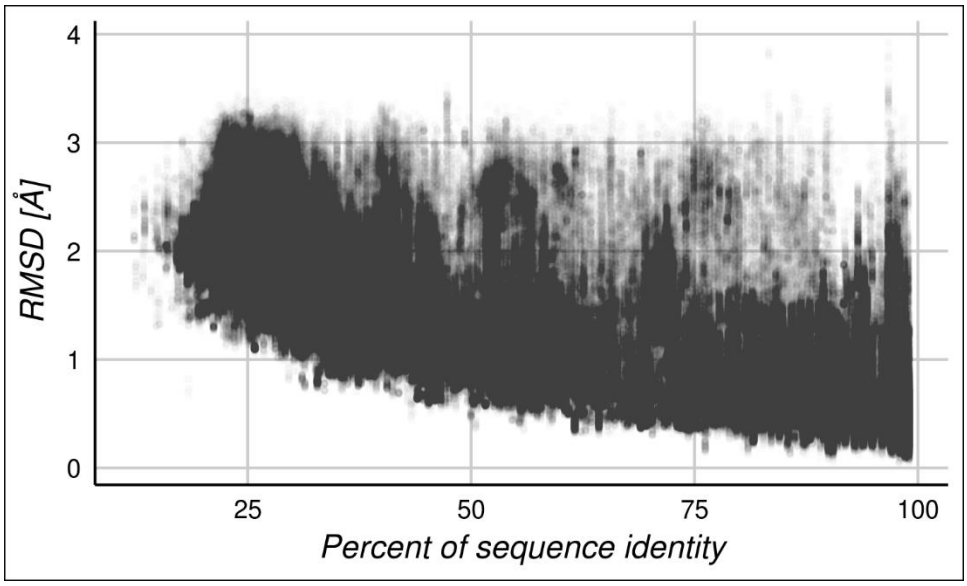
Supporting information

Structural measures	Average CD [Å]			
	(0,0.5]	*(0.5,1]*	*(1,1.5]*	*(1.5,4]*
RMSD	-0.83	-0.78	-0.55	-0.32
Fraction of unconserved SS	-0.75	-0.76	-0.74	-0.71
Fraction of unconserved RSA	-0.86	-0.87	-0.81	-0.80

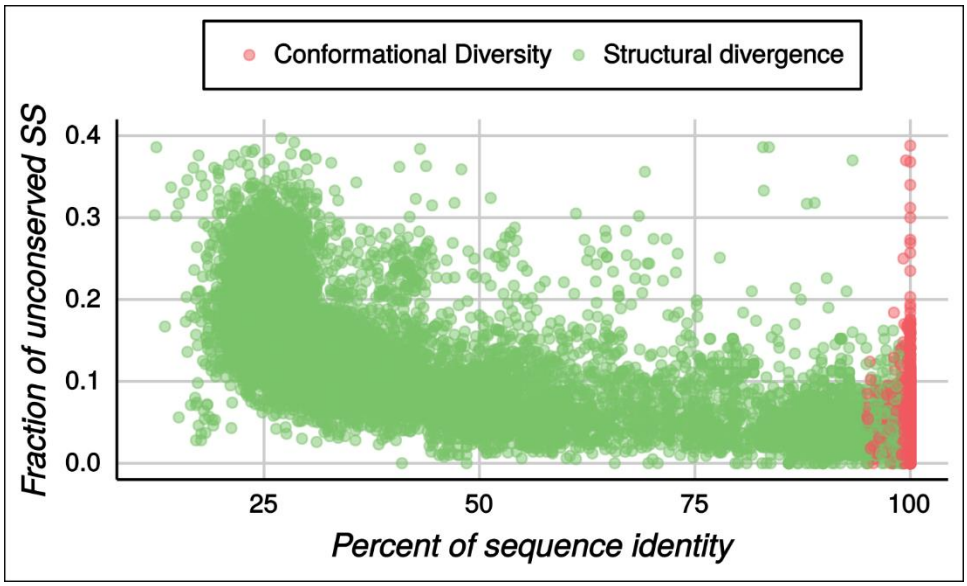
S1 Table. Spearman's rho coefficients calculated in bins of average CD between structural measures and percent of sequence identity.



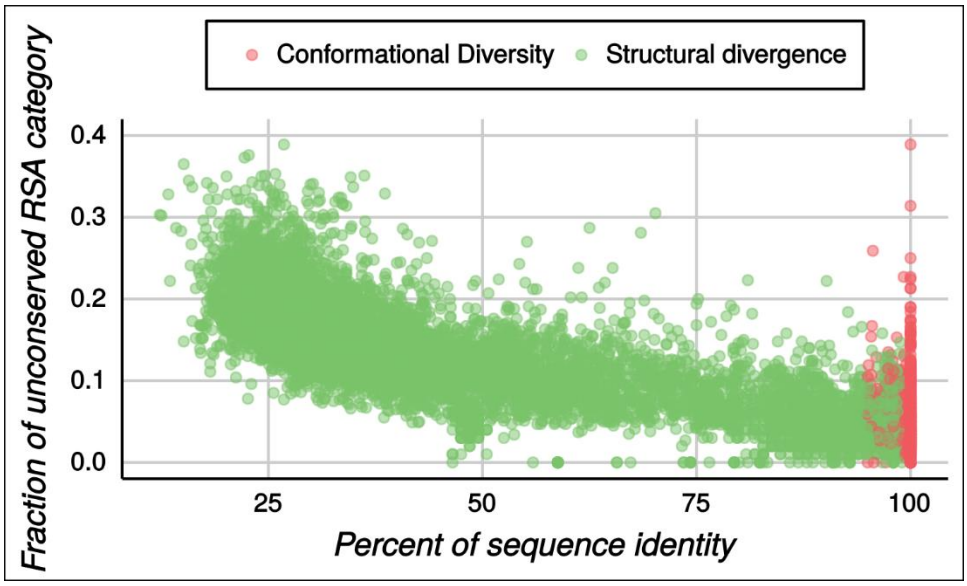
S1 Fig. Schematic representation of the derivation of MSD (Maximum Structural Divergence) and CD (conformational diversity)



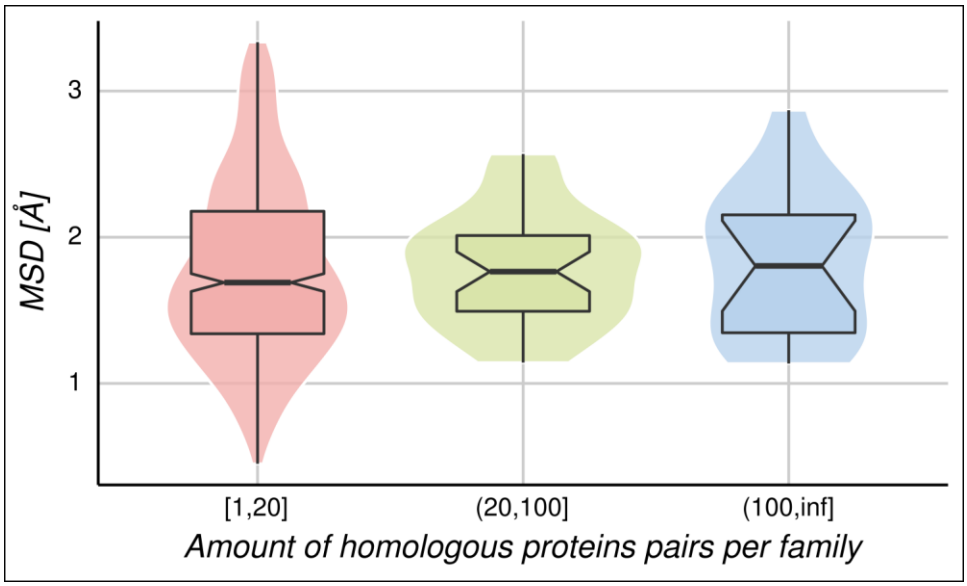
S2 Fig. RMSD vs percent of sequence identity. RMSD obtained from an all vs all comparison between two homologous proteins considering all their conformers. The figure contains about 3.5 million comparisons.



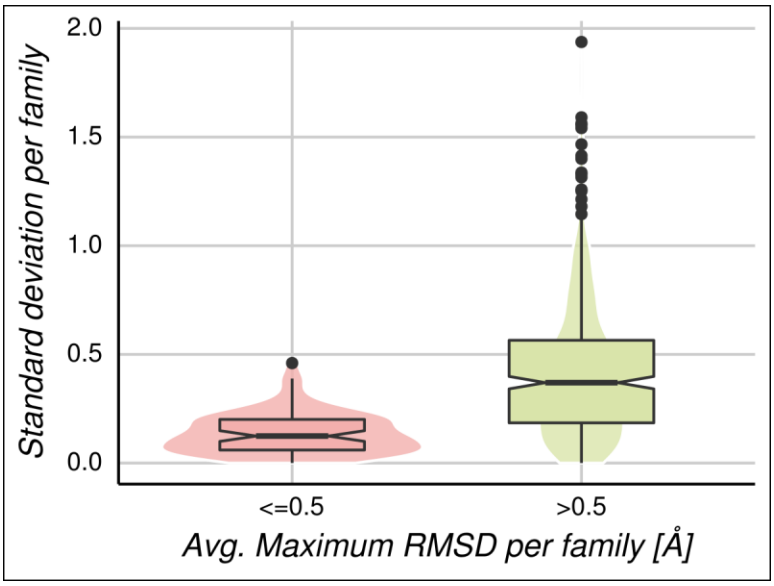
S3 Fig. Fraction of unconserved secondary structure (SS) (in pairs MSD and CD) versus sequence percent identity. Axis Y refer to the fraction of residues which change their secondary structure between structures from two homologous proteins (MSD), or from the same protein (CD). (A) Green dots: comparisons between homologous protein pairs. Red dots: comparison between conformers of the same protein.



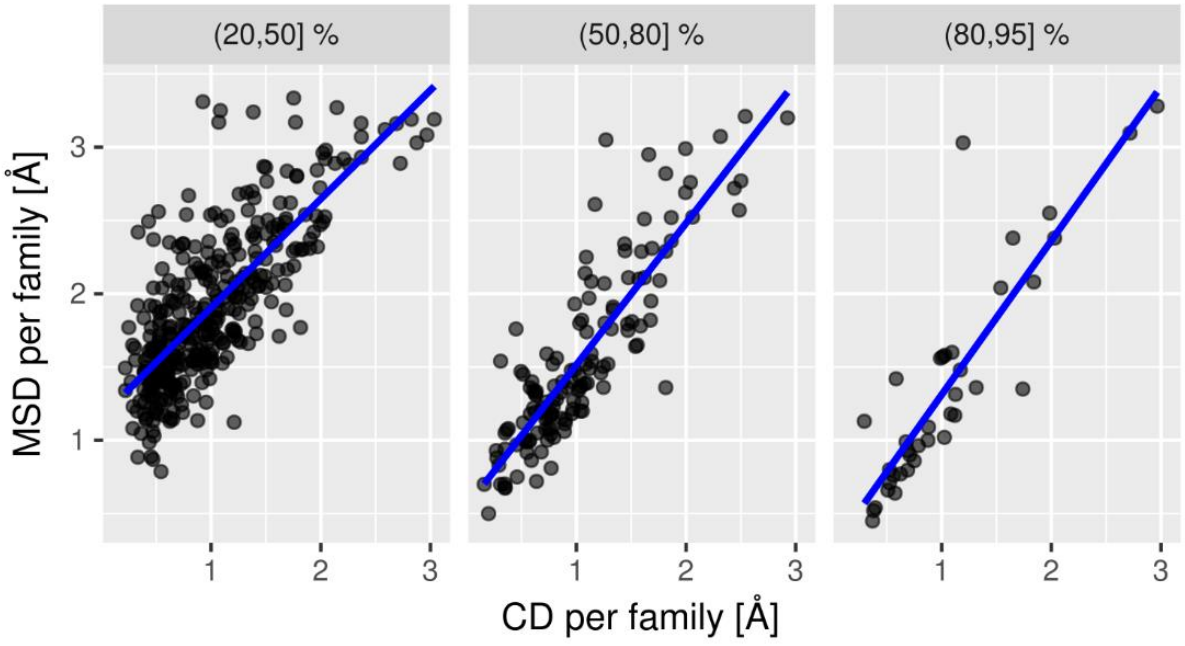
S4 Fig. Fraction of unconserved relative solvent accessibility category (RSA) (in pairs MSD and CD) versus sequence percent identity. Axis Y refer to the fraction of residues which change to buried/exposed between structures from two homologous proteins (MSD), or from the same protein (CD). Green dots: comparisons between homologous protein pairs. Red dots: comparison between conformers of the same protein.



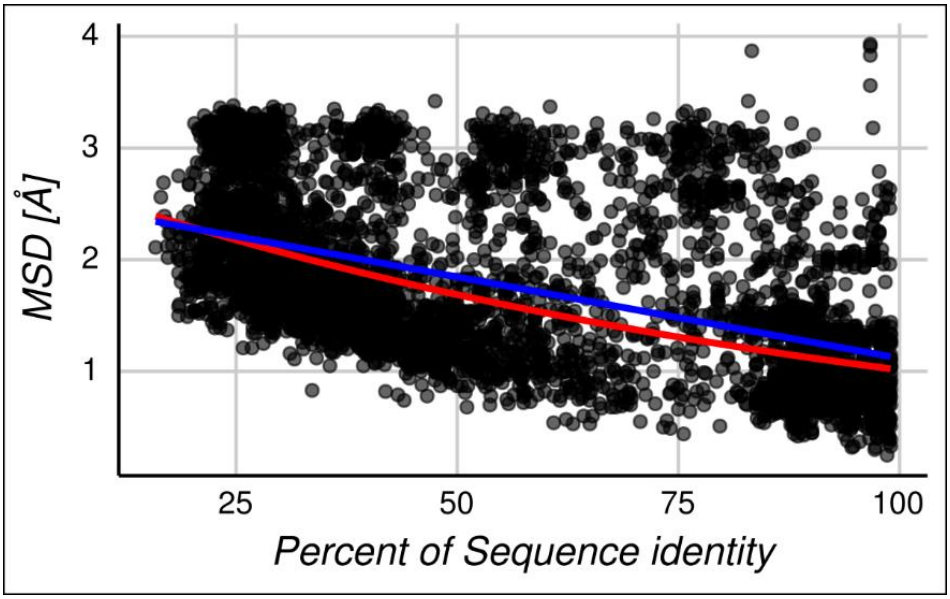
S5 Fig. Distributions of MSD in bins of homologous protein pairs number contained in each family. It is possible to see that the distribution of MSD is not influenced by how populated the family is.



S6 Fig. Distribution of the standard deviation of RMSD due to CD per family estimated over the CD for each protein by family. Standard deviations have been accumulated in two extreme cases, families with low (less than 0.5 Å of RMSD) and high CD (more than 0.5 Å of RMSD).



S7 Fig. Relationship between MSD and CD in bins of sequence identity percent. Each dot represents the average RMSD values for the MSD and the CD in a specific family. The number of families in each bin are 348, 138 and 38 with Pearson's correlation coefficients of 0.77, 0.87 and 0.88 respectively.



S8 Fig. Maximum structural divergence (MSD) versus percent of sequence identity for each homologous protein pair. The lineal (blue line) and exponential (red line) regressions are shown**.** All homologous protein pairs containing just ordered conformers. The linear and exponential fitted expressions are $RMSD = 2.57 - 0.015\,SEQID$ and $RMSD = e^{1.03}e^{-0.010\,SEQID}$, respectively.