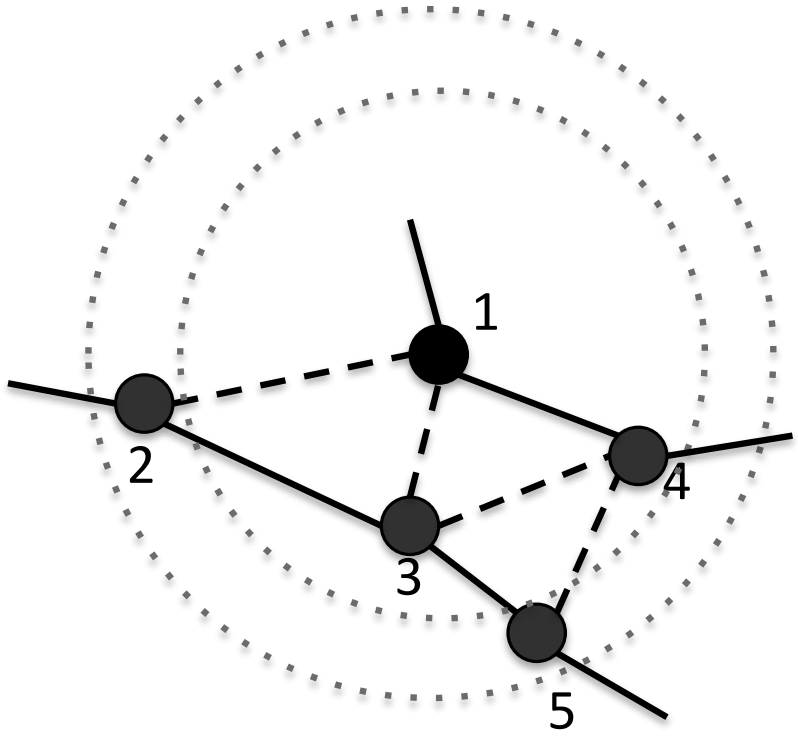
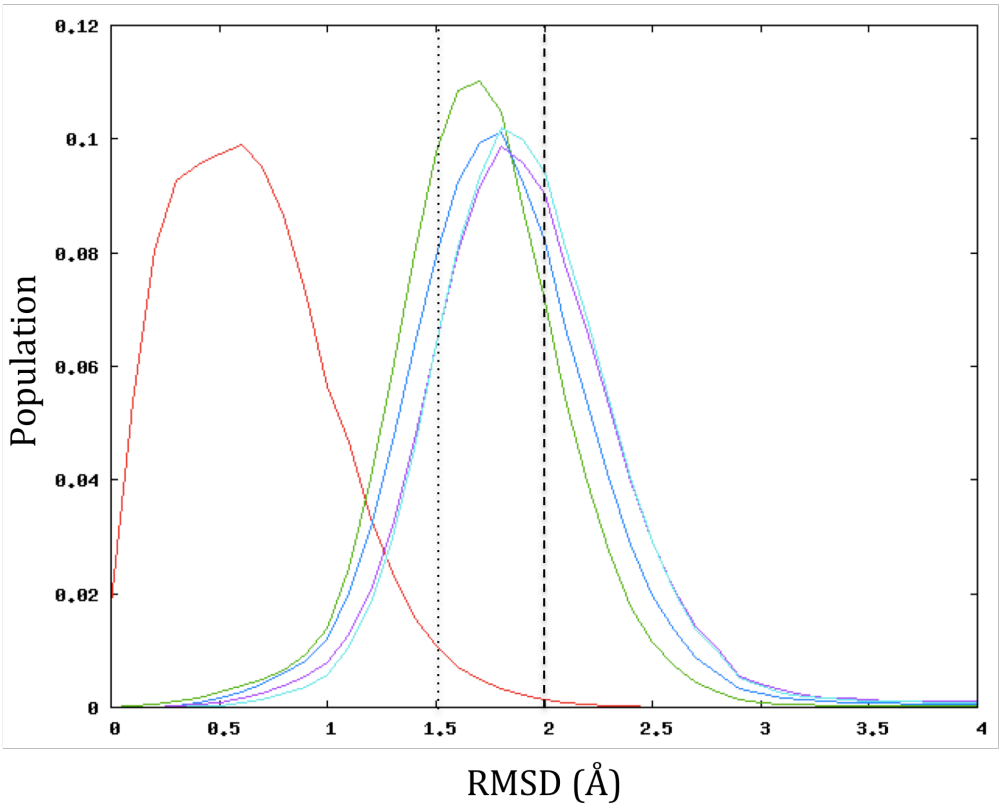
Supplemental Table and Figures

	TerMo[1]	Dom[2]	Fam[2]	Super[2]	Fold[2]	Class[2]	PDB[2]
coil	-- -- -- C4	32229	5809	4128	3122	409	148
	-- -- T1 C3	72849	9341	6359	4744	609	214
	-- E1 -- C3	61365	8055	5372	3985	542	207
	H1 -- -- C3	48912	8155	5546	4209	577	263
turn	-- -- T4 --	18634	4112	2944	2252	296	107
	-- -- T3 C1	65781	8449	5821	4345	560	220
	-- E1 T3 --	25494	5194	3706	2857	467	201
	H1 -- T3 --	31460	6585	4631	3564	570	241
extended							
	-- E4 -- --	541857	9877	6185	4462	503	214
	-- E3 -- C1	243119	11157	7261	5292	687	272
	-- E3 T1 --	116240	9625	6403	4753	677	259
	H1 E3 -- --	160050	8535	5702	4409	679	278
helix	H4 -- -- --	481119	12974	8618	6547	884	387
	H3 -- -- C1	200444	12341	8515	6552	1052	441
	H3 -- T1 --	216333	12710	8731	6696	1102	433
	H3 E1 -- --	142722	6803	4676	3754	599	268
mixed							
	-- -- T2 C2	95261	10412	6702	4965	591	244
	-- E2 -- C2	128678	10411	6895	5066	682	260
	H2 -- -- C2	94130	11024	7598	5796	960	376
	-- E2 T2 --	50346	7692	4987	3793	540	244
	H2 -- T2 --	80826	10406	7090	5489	897	370
	H2 E2 -- --	120143	7733	5257	4140	652	293
	-- E1 T1 C2	96856	10822	7120	5200	642	273
	H1 -- T1 C2	87672	11783	7935	5972	946	359
	H1 E1 -- C2	58232	8105	5452	4149	646	261
	-- E1 T2 C1	76329	9668	6542	4885	713	282
	H1 -- T2 C1	79652	11497	7799	5891	913	346
	H1 E1 T2 --	28795	6171	4310	3444	627	259
	-- E2 T1 C1	125608	11717	7781	5772	865	353
	H1 E2 -- C1	102568	9446	6418	4996	816	330
	H1 E2 T1 --	57732	8033	5519	4355	779	363
	H2 -- T1 C1	146711	13827	9426	7170	1124	453
	H2 E1 -- C1	93745	8612	5963	4646	918	379
	H2 E1 T1 --	62116	7665	5004	3994	698	320
	H1 E1 T1 C1	69183	9490	6471	5000	936	371

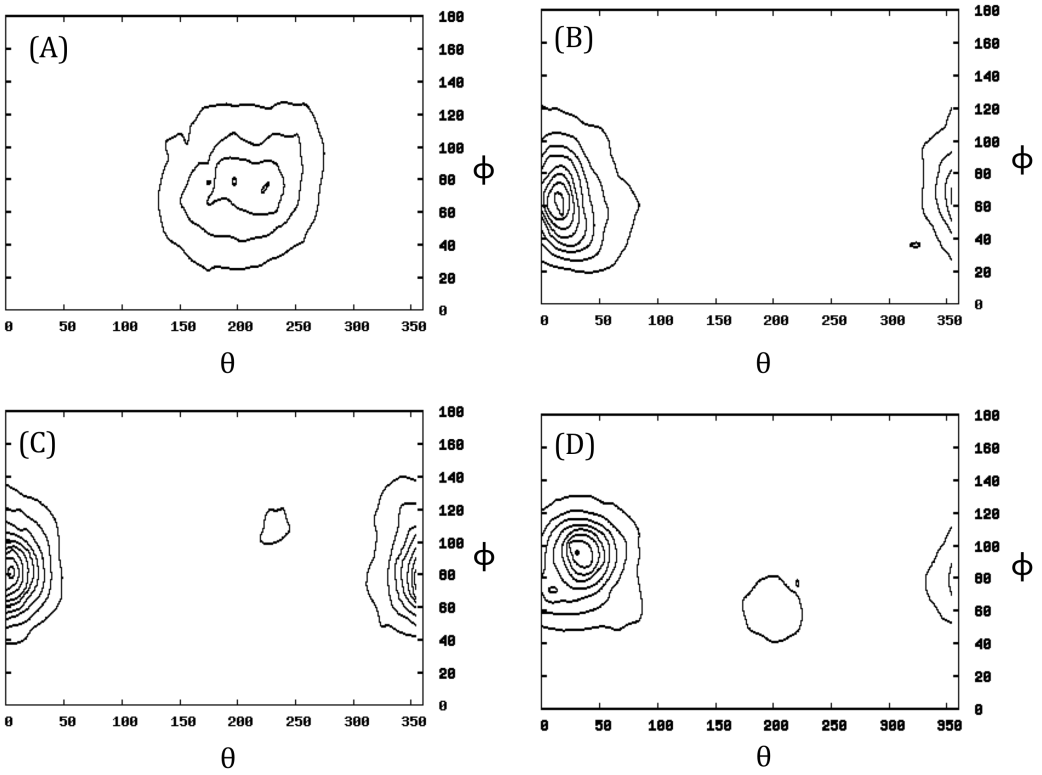
Supplemental Table S1. Results of clustering cliques: Number of clusters. [1]For TerMo definitions of secondary structure H: # of residues in helix, E: # of residues in extened sheet, T: # of residues in turn, C: # of residues in coil. [2]In SCOP, there are 15,273 total domains (Dom); 3,463 families (Fam); 1,776 superfamilies (Super); 1,086 folds (Fold); and 7 classes (Class).



Supplemental Figure S1. Differences between the Delaunay tessellation and a radial distance cutoff. The numbered dark balls represent residues. Solid black lines represent the backbone connections between the residues. Black dashed lines are the contacts defined by the Delaunay tessellation. The grey dotted lines represent two radial distance cutoffs from atom 1. The short cutoff misses the contact with atom 2. A longer contact introduces a spurious contact with atom 5. The sets of atoms 1-2-3, 1-3-4, and 3-4-5 represent three atom RPGs, respectively.

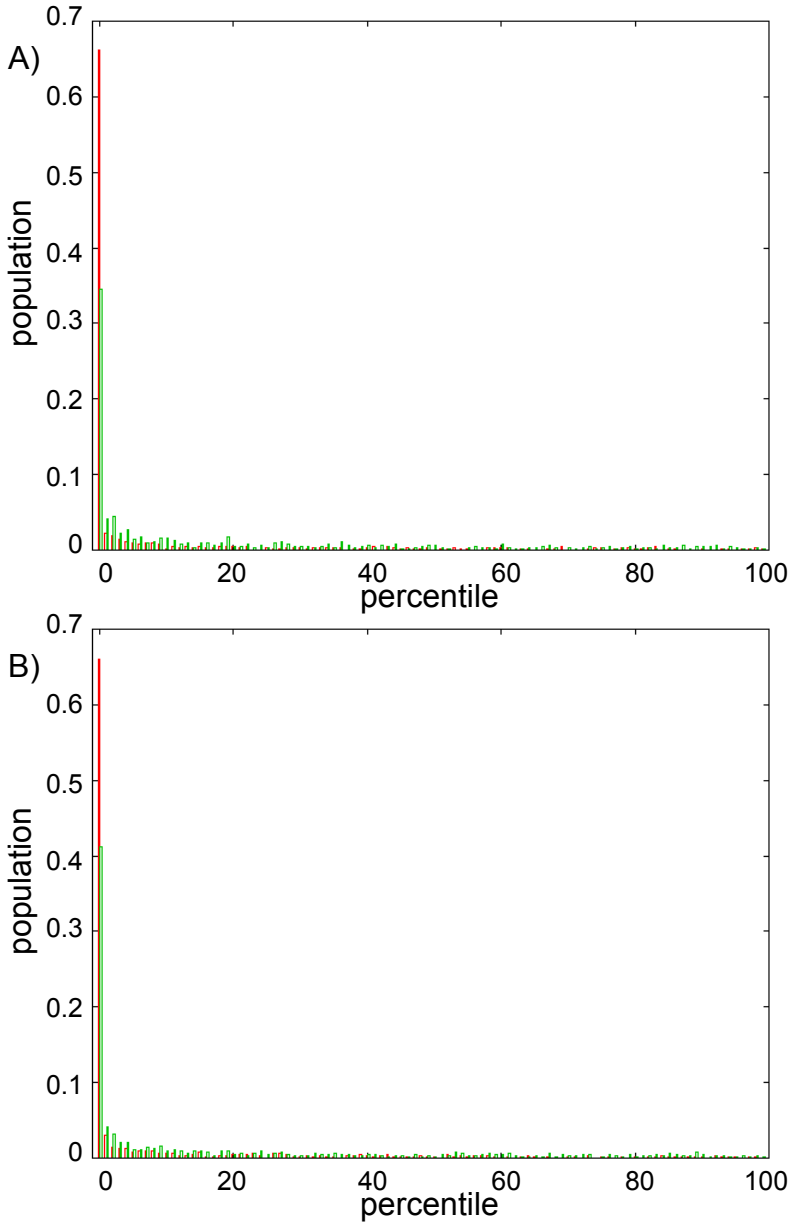


Supplemental Figure S2. Cα-centroid RMSD distributions for Family clique pairs with differing amounts of sequence alignment. Red: all four residues aligned. Green: three residues aligned. Blue: two residues aligned. Magenta: one residue aligned. Cyan: no residues aligned. The dotted and dashed lines show the short (1.5 Å) and long (2.0 Å) cutoffs used. 97.9% of completely aligned cliques have an RMSD < 1.5 Å and 99.8% have an RMSD < 2.0 Å. 18.9% of partially or completely unaligned cliques have an RMSD < 1.5 Å and 65.8% have an RMSD < 2.0 Å.

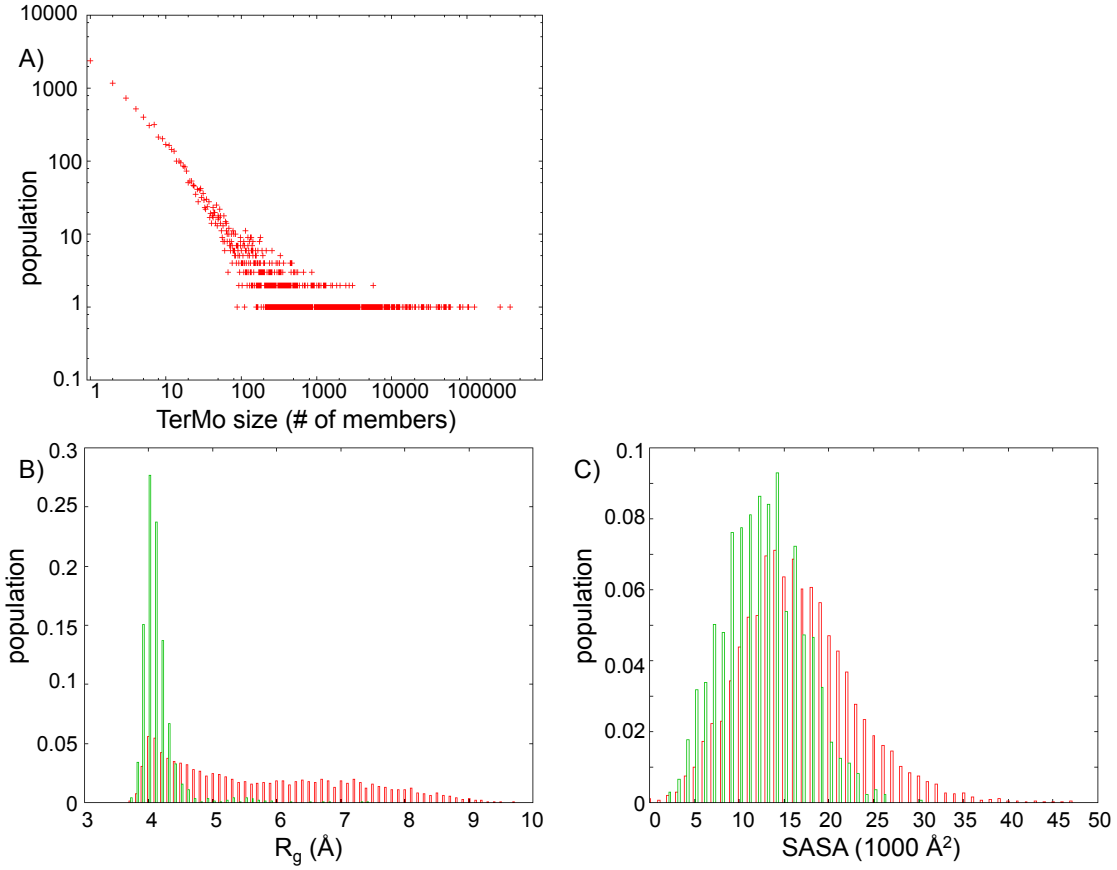


Supplemental Figure S3. Side-chain orientations, expressed as the angular coordinates (θ,ϕ) of the Cα-centroid vector, for each position in the mixed helix-sheet TerMo [H1 E3 - -] shown in Figure 2A of the main text. (A) Position 1, the helical residue. (B) Position 2, the first sheet residue. (C) Position 3, the second sheet residue. (D) Position 4, the third sheet residue.

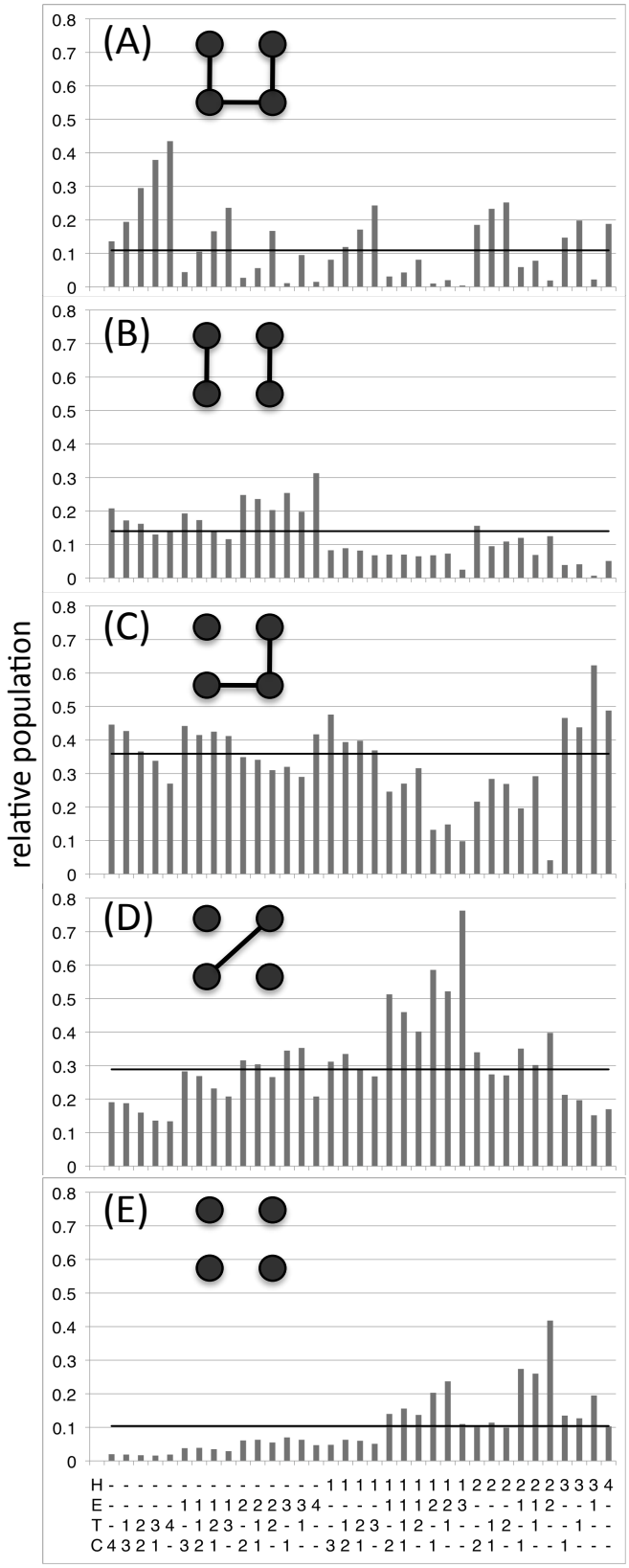
While a 2.0 Å RMSD cutoff allows a comparison the entire PDB, the resulting clusters do not always preserve side-chain orientations. For example, the RPGs in the helix-sheet TerMo [H1 E3 - -] were aligned to their representative structure and the orientations of their Cα-centroid vectors were calculated in polar coordinates (θ,ϕ). The dominant orientations of the sheet positions are parallel to one another but antiparallel to the helix position: essentially a helical residue pointing into three sheet residues. Small but significant orientation subpopulations are seen at positions 3 and/or 4, 2% for both positions, 10% at only 3, and 15% at only 4.



Supplemental Figure S4: Termo properties as percentiles of random clusters. The (A) radius of gyration (Rg) and (B) solvent accessible surface area (SASA) were calculated for all RPGs in all TerMos with at least 100 members. We then calculated the average and sample standard deviation of Rg and SASA for each TerMo and 1000 random clusters of the same size as each TerMo. We then determined the percentile of the distribution defined by our random clusters occupied by the true TerMo mean (red) and standard deviation (green). These percentiles are collected for all TerMos and shown in the histograms above. If the clustering were random, we would expect the above distributions to be uniform. As the distributions are strongly biased toward low percentiles (i.e. low probabilities of the real values being seen in random clusters), we conclude that our clustering is not random. The ensemble mean and standard deviation of Rg are 4.0 Å and 0.5 Å and those of SASA are 11000 Å2 and 9400 Å2 (see supplemental figure S5).



Supplemental Figure S5: TerMo properties. A complete listing of all TerMos, their populations, radii of gyration, and solvent accessible surface areas are available from the authors. A) Histogram of TerMo populations. Note that both axes are logarithmic. The average size of a TerMo cluster is 405 members and 98% of the RPG cliques are in TerMos of at least 100 or larger. B) Histogram of average TerMo radius of gyration (R_g). Red: all TerMos with a mean of 5.73 Å and deviation of 1.42 Å. Green: Termos with at least 100 members with a mean of 4.18 Å and deviation of 0.32 Å. Large Rg TerMos do not cluster well, so end up in low population TerMos. C) Histogram of average TerMo solvent accessible surface areas (SASA). Red: all TerMos with a mean of 16,928 $Å^2$ and a deviation of 6,338 $Å^2$. Green: Termos with at least 100 members with a mean of 12,975 $Å^2$ and a deviation of 4,471 $Å^2$. TerMos with high average SASA are more likely to be found on the surface of the protein.



Supplemental Figure S6. The distribution of cliques with different amounts of sequence locality as a function of secondary structural class. The secondary structural classes are labeled with four numbers representing the number of residues in helix (H), sheet (E), turn (T), and coil (C). For clarity, a "-" indicates that there are no residues in that secondary structure. If consecutive residues in a clique are separated by fewer than four sequence positions, those residues are considered local. The horizontal line is the fraction of all cliques with a given sequence locality. In the inset diagrams, the circles represent the residues in a clique. Circles connected by a line are local in sequence. (A) all local (4) cliques. (B) mixed local and non-local (2+2) cliques. (C) mixed (3+1) cliques. (D) mixed (1+2+1) cliques. (E) all non-local (1+1+1+1) cliques.