Supplementary Material

G-LoSA: an efficient computational tool for local structure-centric biological studies and drug design

Hui Sun Lee and Wonpil Im

Type	Description
HD	Hydrogen bond donor
HA	Hydrogen bond acceptor
OH	Hydroxyl group
PC	Positively charged atom
NC	Negatively charged atom
AR	Aromatic ring
AL	Aliphatic hydrophobic group

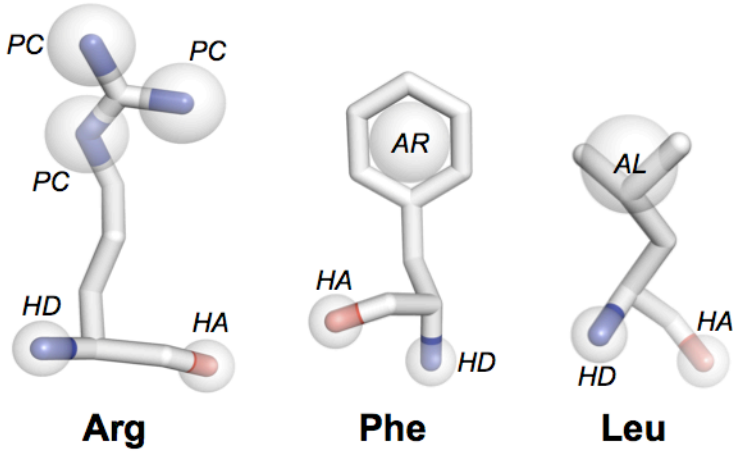


Figure S1. Definition of chemical features. For representative examples, the chemical feature points of three residues (Arg, Phe, and Leu) are illustrated as sphere representations.

Table S1. Definition of chemical features for each amino acid.

Amino acid	Chemical feature (PDB atom name)
Ala	HD (N), HA (O), AL (CB)
Arg	HD (N), HA (O), PC (NE), PC (NH1), PC (NH2)
Asn	HD (N), HA (O), HD (ND2), HA (OD1)
Asp	HD (N), HA (O), NC (OD1), NC (OD2)
Cys	HD (N), HA (O), AL (SG)
Gln	HD (N), HA (O), HD (NE2), HA (OE1)
Glu	HD (N), HA (O), NC (OE1), NC (OE2)
Gly	HD (N), HA (O)
His	HD (N), HA (O), PC (ND1), PC (NE2)
Ile	HD (N), HA (O), AL (CB, CG1, CG2, CD1)
Leu	HD (N), HA (O), AL (CG, CD1, CD2)
Lys	HD (N), HA (O), PC (NZ)
Met	HD (N), HA (O), AL (CE)
Phe	HD (N), HA (O), AR (CG, CD1, CD2, CE1, CE2, CZ)
Pro	HA (O), HA (N)
Ser	HD (N), HA (O), OH (OG)
Thr	HD (N), HA (O), OH (OG1), AL (CG2)
Trp	HD (N), HA (O), HD (NE1), AR (CD2, CE2, CE3, CZ2, CZ3, CH2), AR (CG, CD1, CD2, NE1, CE2)
Tyr	HD (N), HA (O), AR (CG, CD1, CD2, CE1, CE2, CZ)
Val	HD (N), HA (O), AL (CB, CG1, CG2)

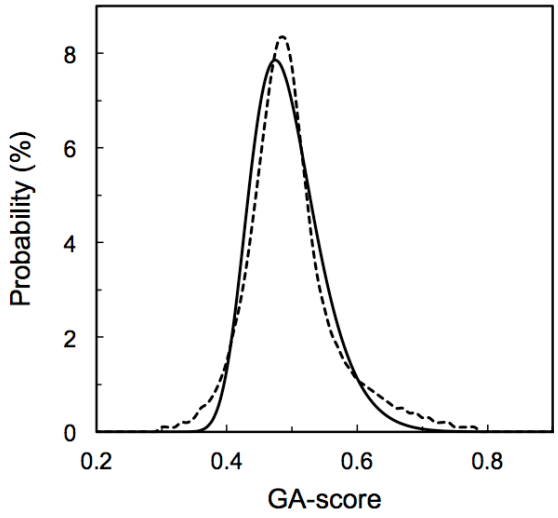


Figure S2. GA-score distribution from the random local structures. The dashed line is the observed probability density and the solid line is a direct fit using the Gumbel distribution.

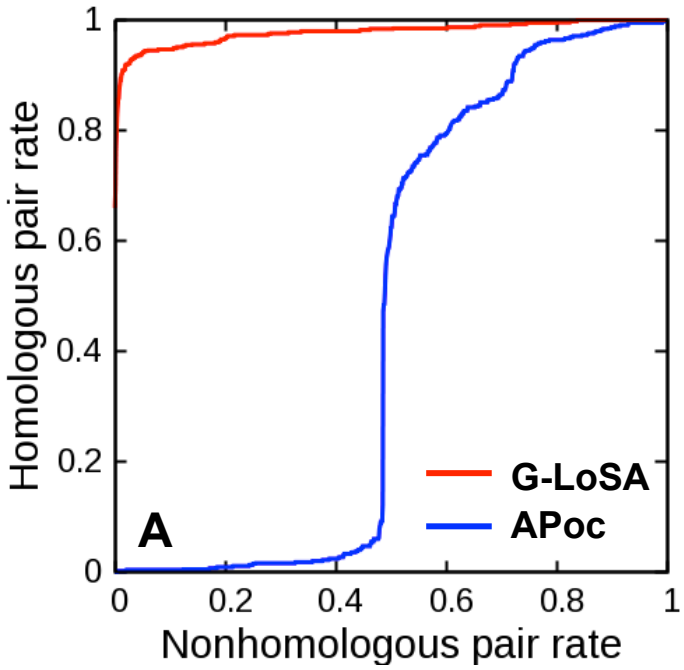
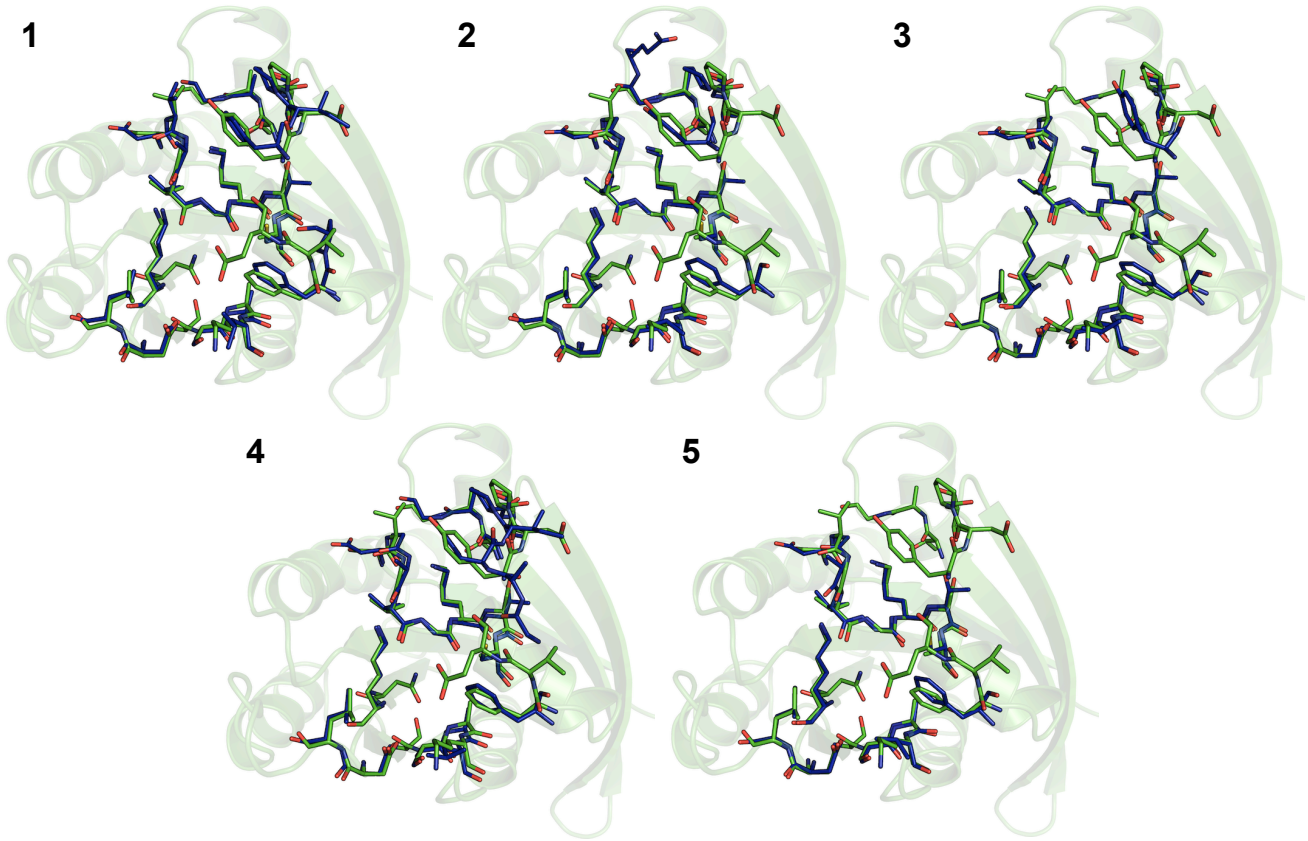
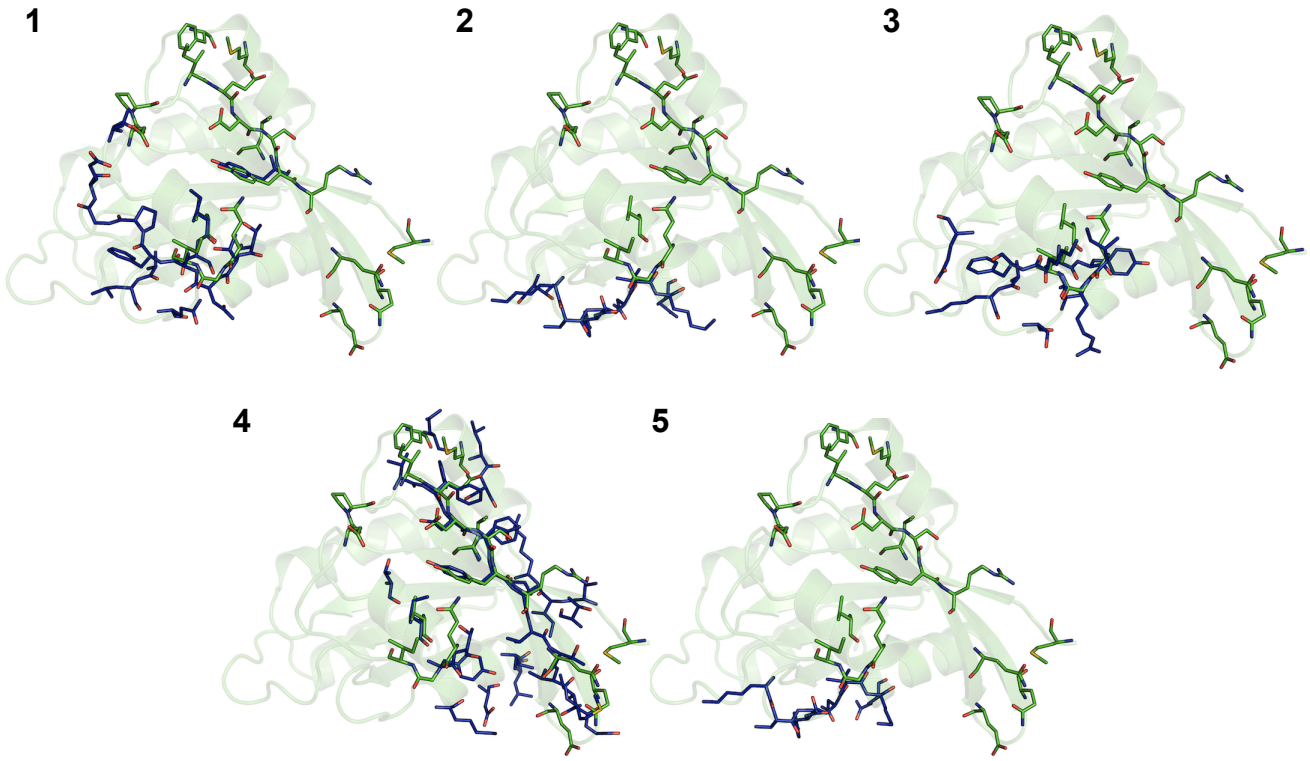


Figure S3. ROC plots of G-LoSA and APoc for Ca^{2+}-binding site benchmark set.



Template	GA-score (1)	TM-score (2)	Average of (1) and (2)	Recall value
1	0.890	0.898	0.894	0.846
2	0.882	0.897	0.889	0.808
3	0.897	0.891	0.889	0.654
4	0.875	0.901	0.888	0.846
5	0.893	0.882	0.887	0.577

Figure S4. The aligned structures of the top five ligand BS templates onto the target protein. The target protein and its native ligand BS (green) in cartoon and stick representation, respectively, are shown with the aligned templates (blue) in stick representation.



Template	GA-score (1)	TM-score (2)	Average of (1) and (2)	Recall value
1	0.722	0.900	0.811	0.368
2	0.766	0.854	0.810	0
3	0.732	0.888	0.810	0.158
4	0.727	0.889	0.808	0.579
5	0.769	0.847	0.808	0

Figure S5. The aligned structures of the top five protein BS templates onto the target protein. The target protein and its native protein BS (green) in cartoon and stick representation, respectively, are shown with the aligned templates (blue) in stick representation.

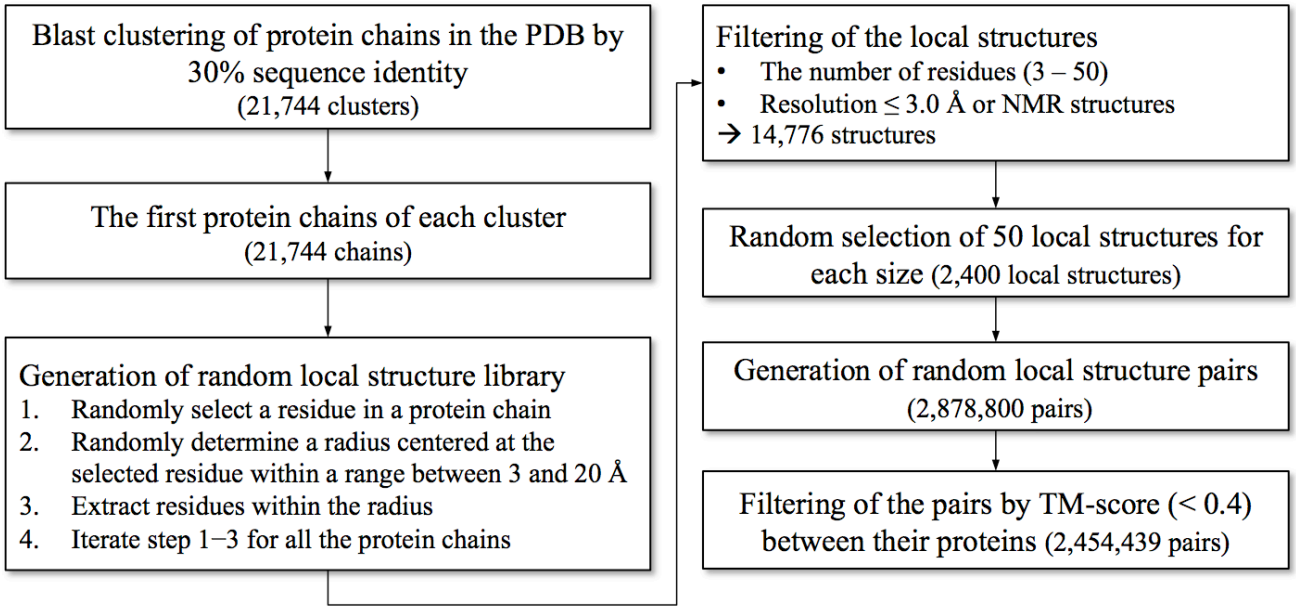


Figure S6. Schematic illustration of the overall procedure for the random local structure set preparation.