Nature's Shortcut to Protein Folding

Fernando Bergasa-Caceres*[1,4]

Elisha Haas[2,5]

Herschel A. Rabitz[3,6]

1. Universidad Autonoma de Madrid, Cantoblanco 28049, Spain
2. The Goodman Faculty of life sciences, Bar-Ilan University, Ramat Gan 52900, Israel
3. Princeton University, NJ 08544, USA
4. Bergasa@Princeton.edu
5. Elisha.Haas@biu.ac.il
6. HRabitz@Princeton.edu

*Corresponding author

SCM existing predictions for primary contacts

In this supplementary section, the results obtained in previous work for the primary contacts of ten proteins within the SCM (references 48 and 96-99 in the main article) are summarized in condensed form. The best and next to best predicted contacts are listed in tables S1 and S2, and the best predicted contacts are represented on the crystal structures in figures S1 to S9 (the best predicted contact for adenylate kinase was presented in the main body of the article). The figures have been elaborated employing Protein Workshop[2].

1. Calculational method to identify initial non-local contacts from primary sequence information

The methodology employed in all cases to predict the location of the primary contacts of the 10 proteins was explained in full in ref. 48 of the main text. The methodology relies only on the primary sequence properties to determine the location of the primary contact. Because the amino acid side chains are significantly larger than the typical peptide bond length, it is expected that early contacts, nucleated by a loop defined by any two amino acids, will immediately involve segments including several amino acids. Thus, the typical early contact segment size was taken to be ~ 5 amino acids. Since the hydrophobic stabilization energy of the contact is determined by the hydrophobicity of the segments involved, hydrophobicity values h_k were obtained from the Fauchere-Pliska scale[1] and

were assigned to each residue. The N- and C-terminal residues carry a charge and their hydrophobicity should be much less than that assigned by the scale to amino acids within the chain. Thus, a value of zero was assigned to the hydrophobicity of the end residues. Then the hydrophobicity h_k of each residue was added over a segment contact window of five amino acids centered at residue *i*, resulting in a segment hydrophobicity $h_{i,5}$ (a value of ~ 0.5 is equivalent to a change in energy of *k*T (1)). In order to determine the highest propensity contact (i.e., the primary contact), the $h_{i,5}$ value of a segment centered at residue *i* was added to the $h_{j,5}$ value of a segment centered at residue *j*, located 65-85 amino acids apart along the sequence, to give a contact propensity $P_{ij} \sim (h_{i,5} + h_{j,5})$, a difference in propensity of ~ 0.45 reflect a difference in energy of ~*k*T.

2. Results

Table S1: Best predicted primary contact.

Protein	PDB structure	Primary contact	Propensity
Cythochrome *c*[1]	1HRC	9-13 on 94-98	10.1
Myoglobin[1]	1MBN	28-32 on 111-115	12.6
Ribonuclease A	1KF5	43-47 on 116-120	9
Barnase[2]	1BNR	13-17 on 93-97	10.7
α-Lactalbumin	1A4V	27-31 on 101-105	12.7
Hen lysozyme	1DPX	28-32 on 107-111	11.4
Leghemoglobin[1,3]	1LH1	43-47 on 109-113	11.4
β-Lactoglobulin	1BEB	19-23 on 103-107	13.1
Staphyloccocal nuclease	1STN	34-38 on 111-115	10.6
Adenylate kinase	4AKE	3-7 on 79-83	10.6

Table S2. Second best predicted primary contact.

PDB structure	Second best contact	Propensity
1HRC	9-13 on 81-85	9.3
1MNB	7-11 on 72-76	12.2
1KF5	26-30 on 106-110	8.7
1BNR	3-7 on 88-92	9.0
1A4V	51-55 on 116-120	9.7
1DPX	54-58 on 120-124	10.4
1LH1	65-69 on 136-140	11.2
1BEB	29-33 on 103-107	12.4
1STN	11-15 on 89-93	10.5
4AKE	105-109 on 178-182	10.5

1. For additional visual clarity the protein has been represented without the heme group present in the crystal structure

2. The result for barnase is different than that presented in ref.48, which was a printing error (a duplication of the result for myoglobin)

3. The experiments to which the SCM predictions were compared in ref.97 were carried out on apoleghemoglobin. The heme group in leghemoglobin is "wedged" between the two segments defining the best primary contact (see figure S7).

References

1. Fauchere, J. L.; Pliska, V. Hydrophobic parameters II of amino-acid side chains from the partitioning of N-acetyl-amino-acid amides *Eur. J. Med. Chem.* 1983, *18*, 369-75

2. Moreland, J.L.; Granada, A.; Buzko, O. V.; Zhang, Q.; P.E. Bourne, P. E. The molecular biology toolkit (MBT): A modular platform for developing molecular visualization applications *BMC Bioinformatics*, 2005, 6, 21

Figures

Figure S1: Primary contact for cythochrome *c*

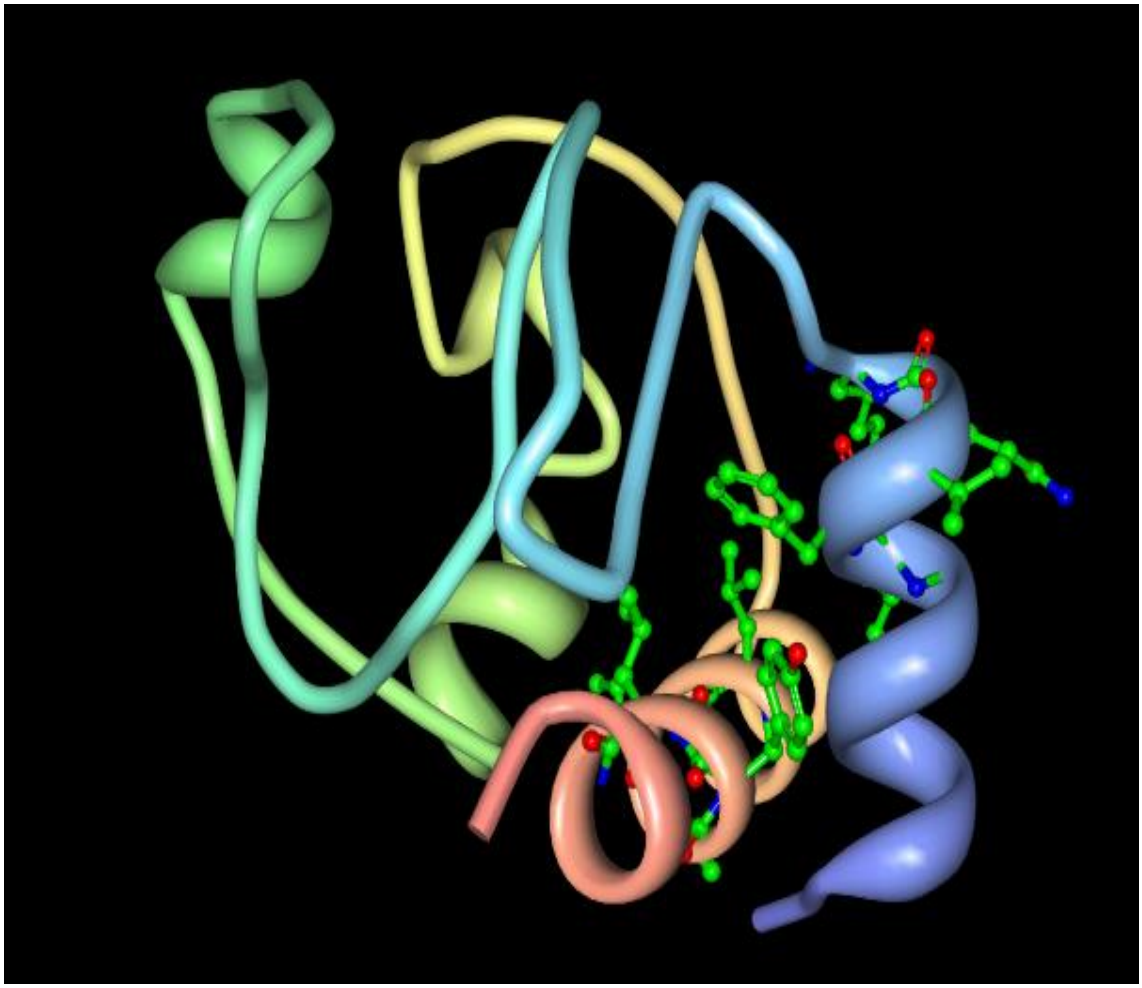


Figure S2: Primary contact for myoglobin

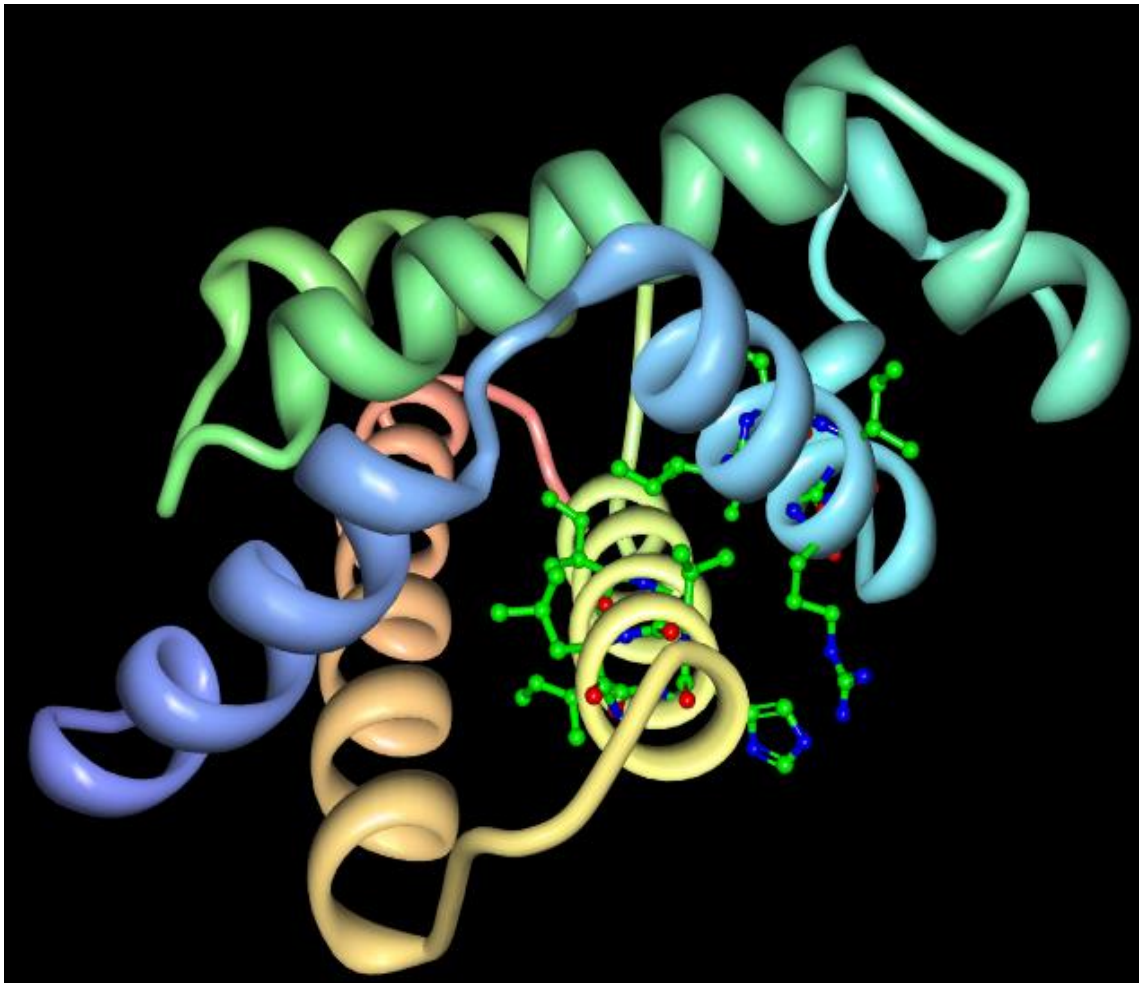


Figure S3: Primary contact for ribonuclease A

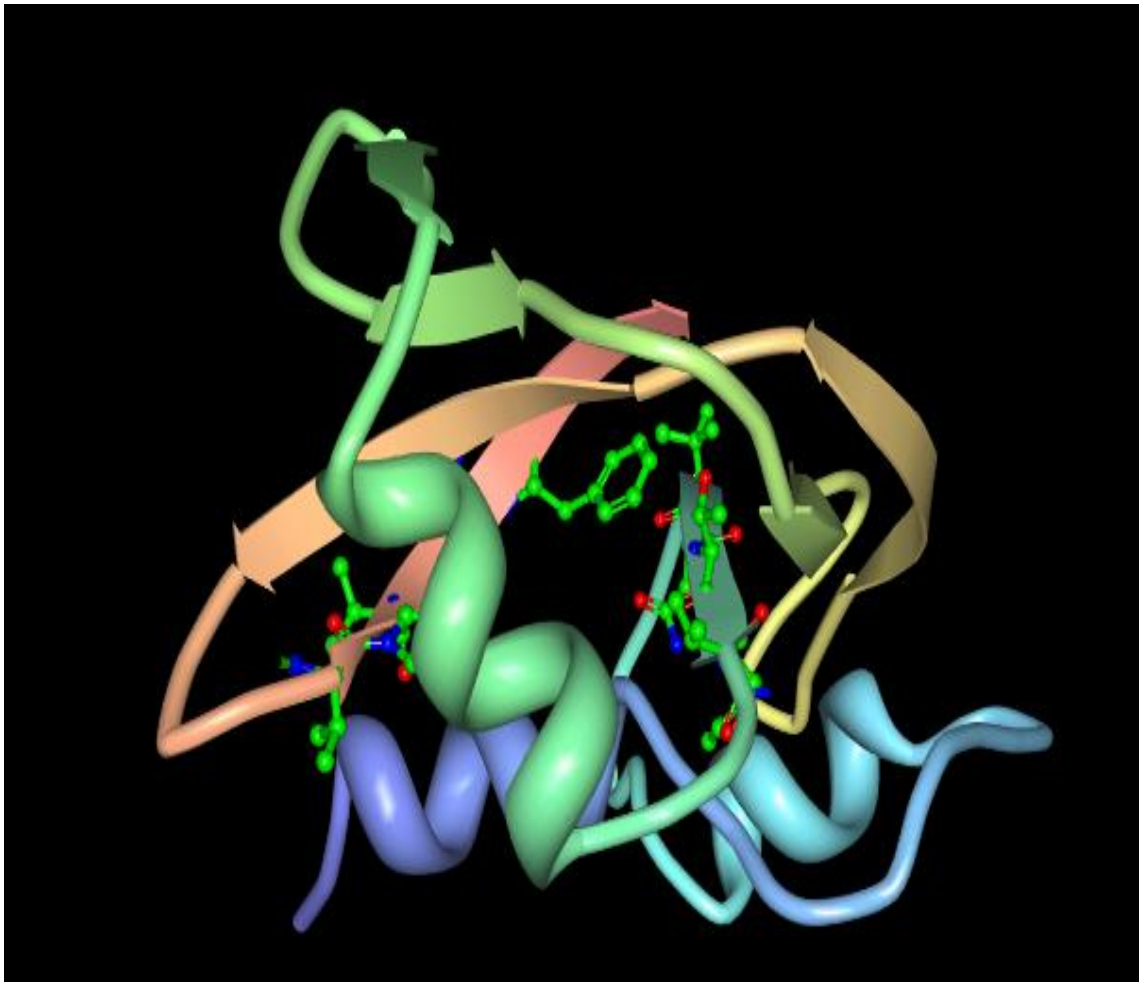


Figure S4: Primary contact for barnase

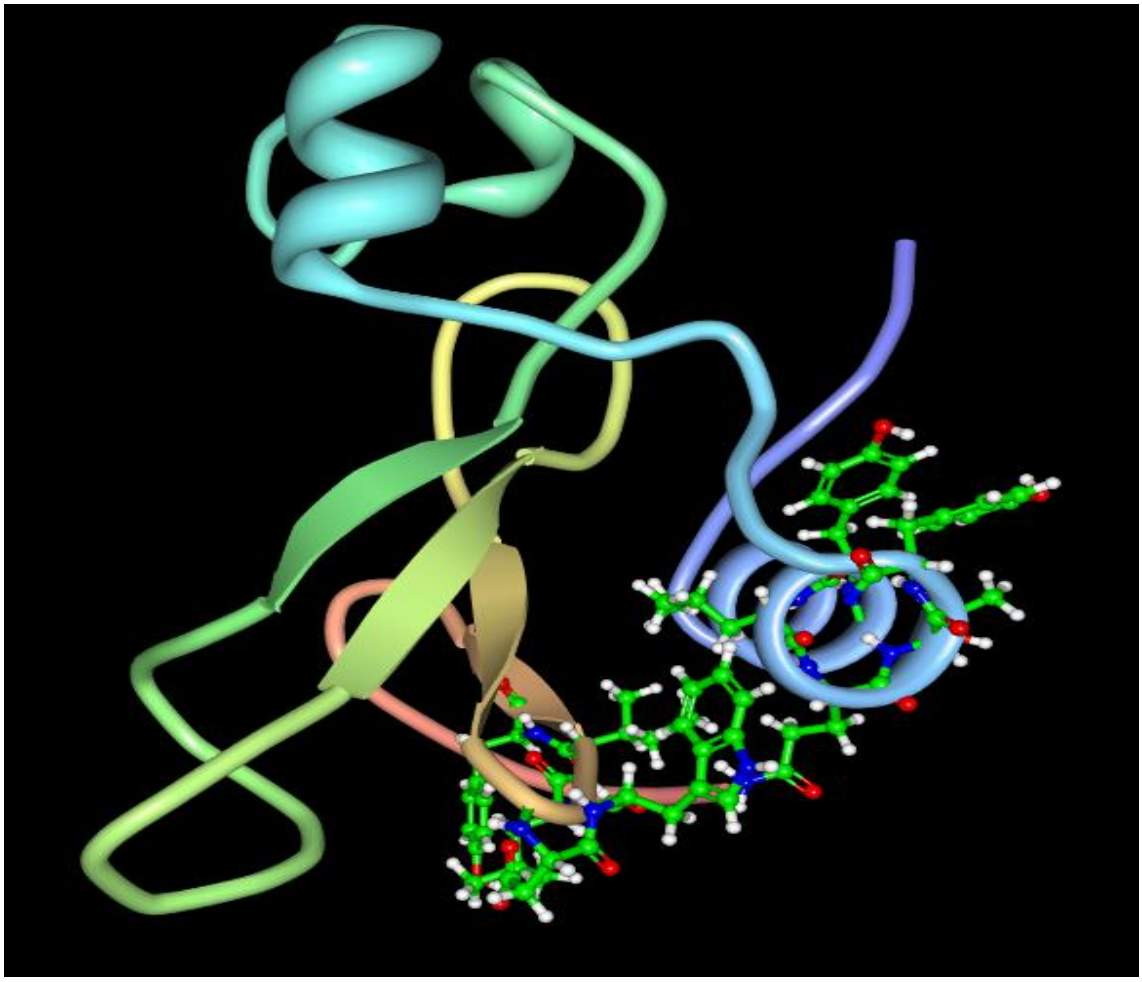


Figure S5. Primary contact for α-lactalbumin

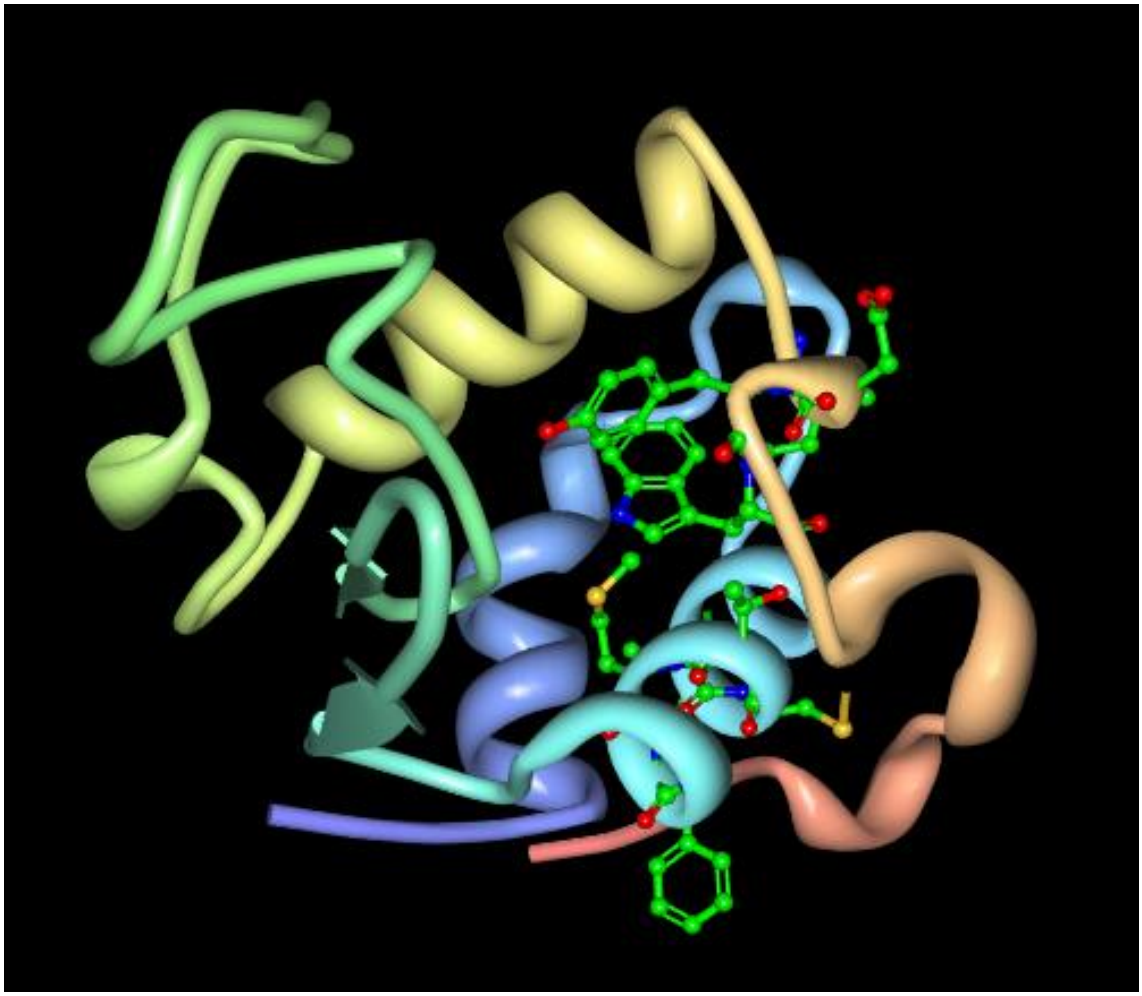


Figure S6. Primary contact for hen lysozyme

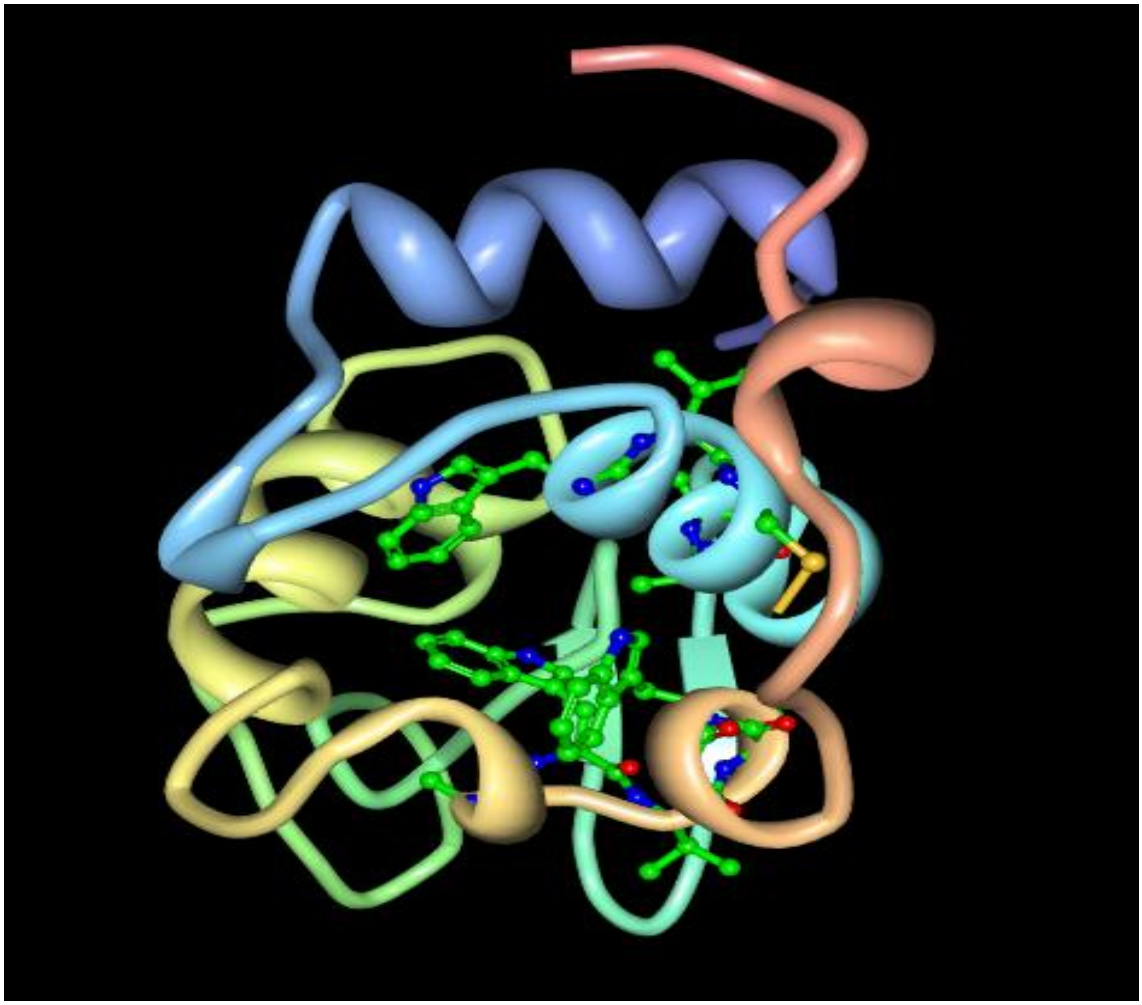
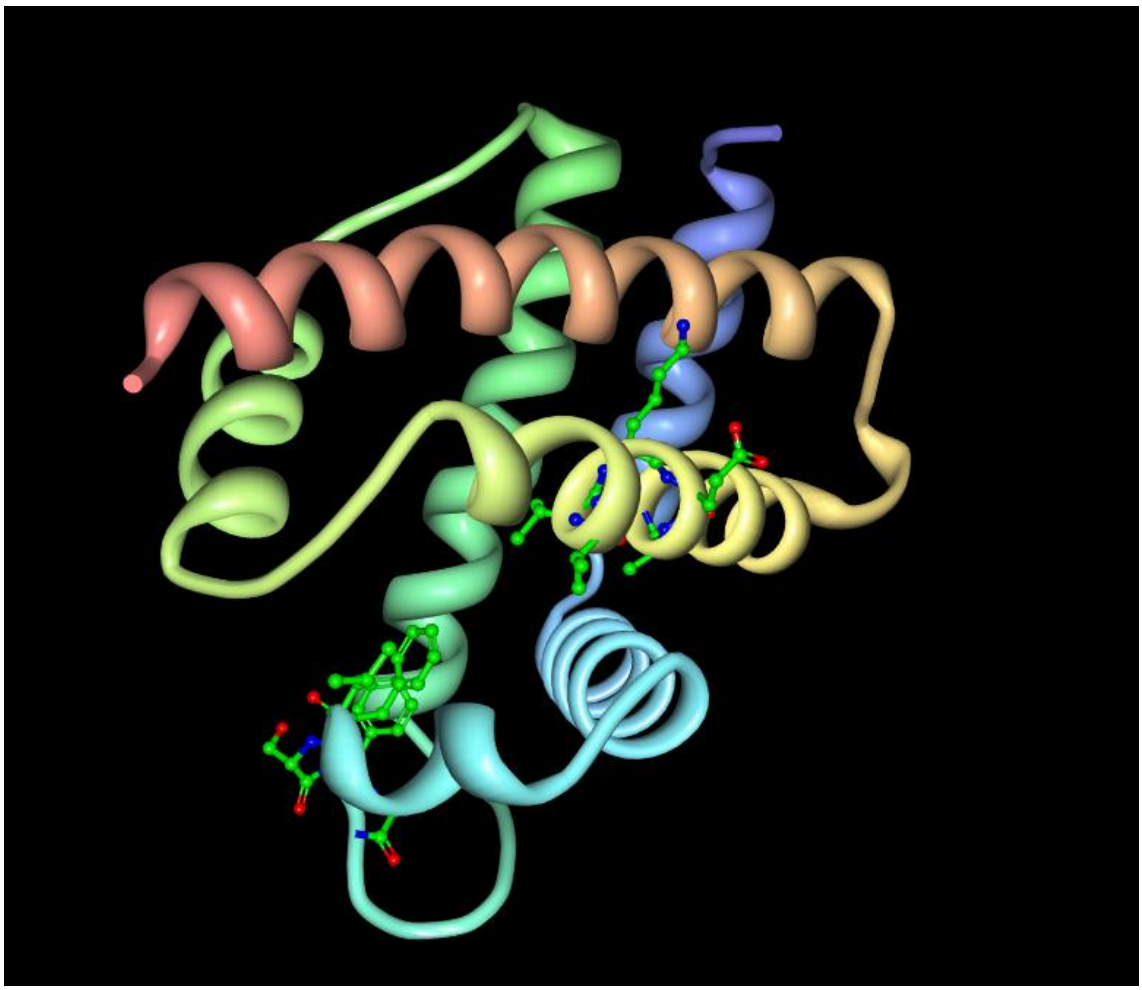


Figure S7. Primary contact for leghemoglobin: (a) not showing the heme group; (b) showing the heme group

(a)



(b)

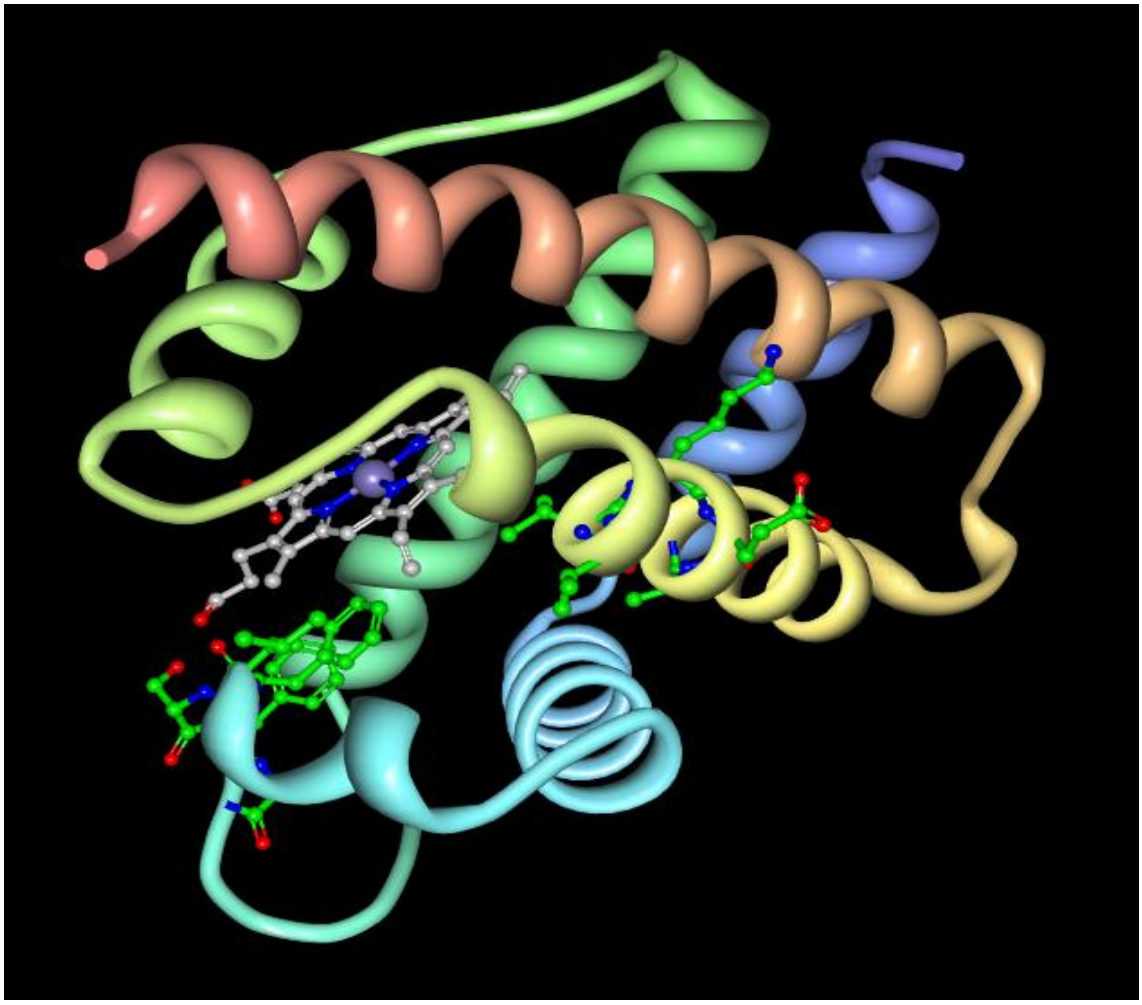


Figure S8. Primary contact for β-Lactoglobulin

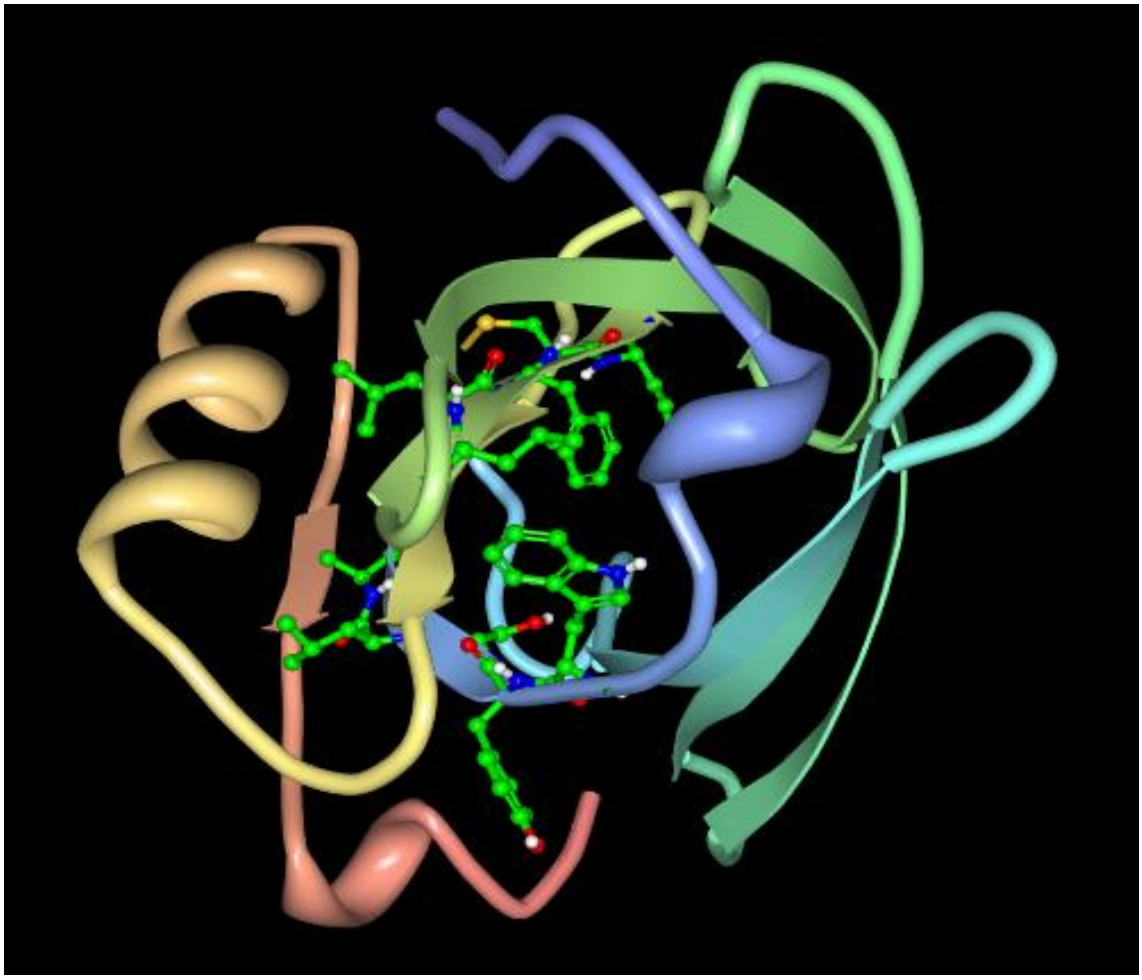


Figure S9. Primary contact for staphyloccocal nuclease