Supporting Information. A driving force for polypeptide and protein collapse

Antonello Merlino,[a] Nicola Pontillo,[a] Giuseppe Graziano[b*]

[a] Dipartimento di Scienze Chimiche, Università degli Studi di Napoli Federico II

Complesso Universitario di Monte Sant'Angelo, Via Cintia, 80126 Napoli, Italy

[b] Dipartimento di Scienze e Tecnologie, Università del Sannio

Via Port'Arsa 11, 82100 Benevento, Italy

Phone: +39/0824/305133; Fax: +39/0824/23013

E-mail: graziano@unisannio.it

The analysis of the composition of protein surface has been carried out using a dataset of 178 ultra-high resolution structures of non-redundant monomeric proteins deposited in the Protein Data Bank (www.rcsb.org[1]). The following pdb codes have been analyzed:

1AHO,1BRF,1BXO,1BYI,1C75,1CBN,1EA7,1F94,1G66,1GCI,1GQV,1GVT,1HJ9,1IC6,1IEE,1IQZ,1IUA,1IXH,1K4I,1K5C,1KWF,1L9L,1LUG,1M1Q,1MC2,1MJ5,1MNZ,1MUW,1MXT,1N9B,1NWZ,1OEW,1OK0,1P9G,1PJX,1PQ5,1PWM,1R6J,1RTQ,1SSX,1TG0,1TQG,1U2H,1UCS,1UFY,1UG6,1UNQ,1US0,1V0L,1VB0,1VBW,1VYR,1W0N,1X6X,1X8P,1XMK,1YLJ,1YWA,1Z8A,1ZUU,1ZZK,2BF6,2BF9,2BW4,2BZZ,2DDX,2E4T,2FDN,2FLA,2FMA,2FOU,2FVY,2FWH,2G6F,2JFR,2NRL,2O9S,2OV0,2PNE,2PPN,2PVB,2PWA,2QXW,2RH2,2UU8,2VB1,2VHK,2WFI,2XFR,2XJP,2XOD,2XOM,2XU3,2Y78,2ZPM,3A38,3AKQ,3C78,3D1P,3DHA,3DK9,3EA6,3F7L,3FSA,3G21,3GOE,3HGP,3IP0,3JU4,3KFF,3KLR,3M5Q,3O4P,3O5Q,3PUC,3PYP,3Q46,3Q8J,3QPA,3UI4,3VIG,3W5H,3WGE,3WVM,3X0I,3X2L,3X32,3ZOJ,3ZR8,3ZSJ,4ACJ,4AQO,4AYO,4BCT,4BM8,4BVM,4EA9,4EIC,4EKF,4F18,4G78,4GA2,4GCA,4HGU,4HNO,4HS1,4IAU,4IGS,4K8M,4KQP,4KXU,4LAU,4LFS,4MTU,4MZC,4NPD,4PNO,4QB3,4UYR,4WEE,4WKA,4WPK,4XDX,4ZC9,5A8C,5AKR,5AL6,5AVD,5D8V,5D9E,5DP2,5E9N,5EWO,5HB7,5HBS,5IG6,7A3H,8A3H.

These models have been analyzed using the Vadar server.[2] The structures have been solved by X-ray crystallography at a resolution higher than 0.99 Å, with an average resolution of 0.91 Å. The average secondary structure content of these proteins is: α-helix = 24.0 $\pm$ 20.3 %, β-sheet = 33.9 $\pm$ 18.4%, coil = 40.3 $\pm$ 11.6%. The analyzed total accessible surface areas range from 1816 to 35872 $Å^2$, with an average value of 9020 $\pm$ 4699 $Å^2$.

References

1. P.W. Rose, C. Bi, W.F. Bluhm, C.H. Christie, D. Dimitropoulos, S. Dutta, R.K. Green, D.S. Goodsell, A. Prlic, M. Quesada, G.B. Quinn, A.G. Ramos, J.D. Westbrook, J. Young, C. Zardecki, H.M. Berman, P.E. Bourne, The RCSB Protein Data Bank: new resources for research and education. *Nucleic Acids Res.*, 2013, **41** (Database issue), D475-82.

2. L. Willard, A. Ranjan, H. Zhang, H. Monzavi, R.F. Boyko, B.D. Sykes and D.S. Wishart, VADAR: a web server for quantitative evaluation of protein structure quality. *Nucleic Acids Res.*, 2003, **31**, 3316-3319.