DB2: A database of monomeric globular proteins

Sankar Basu[1*] & Rahul Banerjee[1]

[1] Crystallography and Molecular Biology Devision; Saha Institute of Nuclear Physics, Kolkata - 700064

[*] nemo8130@gmail.com

The database has been used for the computation of shape (S_m) and electrostatic complementarity (E_m) of buried interior residues within proteins [1], and for the construction of the Complementarity Plot [1] [2] [3]

DB2 contains 400 high resolution crystal structures of monomeric globular proteins (i.e., each containing one polypeptide chain) culled at a resolution better than or equal to 2 Å, R-factor better than 20% and homologs removed at 30% sequence identity. This database is a subset of a larger database (DB1) discussed in detail in a previous report [4]. DB2 was assembled by removing proteins with deeply embedded prosthetic groups (e.g., cytochromes) and any missing atoms from DB1. DB2 (composed of 65 all-α, 70 all-β, 106 α|β, 124 α+β and 35 multidomain proteins). Sixty-two of these proteins were found to contain metal ions as an integral part of their structure.

PDB ID and chain identifier (underscored) for each polypeptide chain is given under its respective protein class. Hydrogen atoms were geometrically fixed to all structures by the program REDUCE [5]**.** In order to confirm the resonance states of histidine, polar hydrogens were reassigned by GROMACS (v.4) using its G43a1 force field [6]. 95% of the assignments were found to be identical in both programs.

all α:

1C02_B, 1C1K_A, 1EL4_A, 1ELK_B, 1EYH_A, 1EZJ_A (Ca^{+2}), 1G5N_A (Ca^{+2}), 1G8E_A, 1I8O_A, 1JKV_A (Ca^{+2}, Mn^{+3}), 1LJ9_A, 1MOJ_A (Fe^{+3}), 1MXR_B (Fe^{+3}, Hg^{+2}), 1OOH_B, 1OXJ_A, 1Q08_A (Zn^{+2}), 1Q5Z_A, 1QYZ_A, 1R1T_B, 1R5Z_B, 1R7J_A, 1SZH_A, 1TFJ_A, 1U2W_B (Zn^{+2}), 1U84_A, 1UCR_B, 1WPA_A, 1WXC_A, 1ZKR_A, 2A61_A, 2AIB_A, 2B8I_A, 2BA2_C, 2CWL_A, 2D48_A, 2D4X_A, 2DPO_A, 2E1N_B, 2ERB_A, 2F2B_A, 2FBQ_A, 2FD5_A, 2FMM_E, 2FP1_A, 2FU4_A, 2GXG_A, 2HRA_A, 2I3F_B, 2IC6_B, 2IGP_A, 2IMI_B, 2NZ7_B, 2O37_A, 2O70_F, 2OEB_A, 2OQG_C, 2PQR_B, 2QRW_K, 2QSA_A, 2RKN_A (Zn^{+2}), 3B9W_A, 3BS7_B, 3C1D_B, 3CHM_A, 3DHZ_B (Fe^{+2}).

all β:

1DJR_H, 1DMH_B (Fe^{+3}), 1EZG_B, 1F7D_A, 1I0R_B, 1I4U_A, 1IBY_B (Cu^{+2}), 1J71_A, 1K12_A

(Ca^{+2}), 1KJL_A, 1KT6_A, 1KZQ_B, 1LR5_C (Zn^{+2}), 1LSL_A, 1LXZ_A, 1LYQ_B, 1NTV_A, 1NYK_B, 1NYW_B, 1OU8_B, 1P3C_A, 1PXD_A, 1QB5_E, 1QXM_A, 1R77_B, 1R8O_A, 1ROC_A, 1RU4_A (Ca^{+2}), 1S1D_A (Ca^{+2}), 1SE0_A (Zn^{+2}), 1T2W_C, 1TH7_L, 1V8H_A, 1VM9_A, 1WD3_A, 1WLG_B, 1WLI_B, 1WS7_A, 1XU1_B, 1YFQ_A (Ca^{+2}), 1YOA_A, 1Z3T_A, 1ZVT_B, 2A6V_A, 2AYD_A (Zn^{+2}), 2D37_A, 2DCY_C, 2DP9_A, 2DUR_A (Ca^{+2}), 2EI5_A, 2ERF_A, 2ET1_A (Mn^{+2}), 2F6E_A, 2G9F_A, 2GGV_B, 2GH2_A (Fe^{+3}), 2HAL_A, 2ICC_A, 2IQY_A (Ca^{+2}), 2O8L_A, 2OFZ_A, 2OR7_B, 2PA7_A, 2PCX_A (Zn^{+2}), 2R5O_A, 2RJ2_A, 3BOV_A, 3BU1_A, 3D9X_C, 3F5R_A.

α | β:

1DJ0_A, 1EKG_A, 1EQ6_A, 1F1U_A, 1F46_B, 1F5M_B, 1G1T_A (Ca^{+2}), 1G2R_A, 1IQZ_A, 1J1G_A, 1J1Y_A, 1J27_A, 1J3W_B, 1JD5_A (Zn^{+2}), 1JYH_A, 1JYO_B, 1K2A_A, 1KJO_A (Zn^{+2}, Ca^{+2}), 1KQB_C, 1KUX_A, 1LN4_A, 1LO7_A, 1LQP_B, 1M4D_A, 1M4J_B, 1NE9_A, 1NWA_A, 1NWW_B, 1OZ9_A, 1PC4_A, 1PPV_B (Mg^{+2}, Mn^{+2}), 1R29_A, 1R45_A, 1RC9_A, 1RKI_A, 1RWZ_A, 1S5U_E, 1TA8_A, 1TKE_A, 1TP6_A, 1TR0_I, 1TU1_A, 1TUA_A, 1TUH_A, 1TUV_A, 1WRI_A, 1XBI_A, 1XMT_A, 1XPP_D, 1Y60_D, 1YAR_D, 1YGT_A, 1YPY_A, 1YWM_A, 1Z4R_A, 1ZDY_A, 1ZHQ_G, 1ZHX_A, 1ZT3_A, 2A15_A, 2AAL_A, 2ACY_A, 2B18_A, 2BBE_A, 2CVE_A, 2DT4_A, 2E11_B, 2E12_A, 2E8G_B, 2EA1_A, 2FHZ_A, 2FJR_B, 2FL4_A, 2FVV_A, 2FYG_A (Zn^{+2}), 2G64_A (Zn^{+2}), 2G8O_A, 2GEB_A (Ca^{+2}), 2GH5_B, 2GHT_A, 2GMN_B (Zn^{+2}), 2GU3_A, 2GWM_A, 2H2Z_A, 2I2Q_A, 2I7D_A, 2IF6_A, 2IG8_B, 2NR7_A, 2NTT_B, 2O28_A, 2OH5_A, 2P0W_A, 2P2O_B, 2P3H_A, 2PLQ_A, 2PU3_A, 2Q0L_B, 2QKP_B, 2Z51_A, 3BJK_C, 3BK8_A, 3C8I_B, 3COU_A, 3DA4_A, 3DHA_A (Zn^{+2}).

α + β:

1C7N_C, 1DGF_B, 1DJE_A, 1DQZ_B, 1EQC_A, 1EU8_A, 1F8M_B, 1G5T_A, 1G8A_A, 1I2A_A, 1I2K_A, 1I5G_A, 1I6W_B, 1I9C_B, 1IM5_A (Zn^{+2}), 1INL_C, 1IZC_A, 1JAY_B, 1JJF_A, 1JKE_C (Zn^{+2}), 1JX6_A (Ca^{+2}), 1K4M_B, 1K7C_A, 1LOK_A (Zn^{+2}), 1LUA_B, 1LWD_B, 1N3Y_A, 1NF9_A, 1O04_G, 1OOE_B, 1OZN_A, 1P6O_B (Zn^{+2}, Ca^{+2}), 1Q7L_A (Zn^{+2}), 1RKU_B, 1SG0_B (Zn^{+2}), 1SW5_B, 1T1V_B, 1TJY_A, 1TP9_B, 1TVP_A, 1UB3_B, 1UC7_A, 1UEH_B, 1UG6_A, 1UK8_A, 1V2X_A, 1V37_A, 1VFL_A (Zn^{+2}), 1WN5_B, 1WO8_A, 1WOU_A, 1WPN_A, 1WX1_B, 1XBY_A, 1XQ6_B, 1XS5_A, 1XX1_B, 1Y1P_A, 1Y1T_A, 1Y6V_A, 1YB6_A, 1YD0_A, 1YN9_B, 1YQZ_B, 1YZX_B, 1ZL0_A, 1ZOI_C, 1ZR6_A (Zn^{+2}), 1ZX0_C, 1ZZG_A, 1ZZW_B, 2A14_A, 2A1I_A, 2ACF_C, 2AEU_A, 2AGD_B, 2AP1_A (Zn^{+2}), 2AVD_B, 2B1K_A, 2B3F_B, 2B82_A, 2CYG_A, 2E5F_A, 2FB9_A, 2FLA_A, 2FQX_A, 2FY6_A, 2G84_A (Zn^{+2}), 2GSO_A (Zn^{+2}), 2H1C_A, 2H1V_A, 2HC9_A (Zn^{+2}), 2HL6_B, 2HOR_A, 2HPJ_A, 2HRC_B, 2HU9_A, 2HXU_A (Mg^{+2}), 2HY5_B, 2HZL_B, 2I49_A, 2I56_C (Zn^{+2}), 2NZX_C, 2OUI_B, 2P51_A (Mg^{+2}), 2PS1_A, 2QDX_A, 2QUL_B (Mn^{+2}), 2REM_C, 2RHJ_A, 2YW3_D, 2Z1Y_B, 2Z6R_B, 3B4U_B, 3B8X_B, 3BJE_B (Ca^{+2}), 3BOE_A, 3C9H_B, 3CQ5_B, 3D2H_A, 3DAF_A (Fe^{+2}), 3E96_B, 3ES7_B (Mg^{+2}), 3EVF_A.

Multidomain:

1DLJ_A, 1EEJ_B, 1F60_A, 1FP2_A, 1FW1_A, 1G8L_B, 1HM9_B (Ca^{+2}), 1HYO_B (Ca^{+2}), 1IG3_A,

1JAK_A, 1K7I_A (Zn^{+2}, Ca^{+2}), 1KTB_A, 1LBV_A, 1M1H_A, 1OWL_A, 1P1M_A, 1RA0_A (Fe^{+3}), 1RQP_C, 1SZN_A, 1T0F_A, 1T2D_A, 1T4B_B, 1U60_C, 1VBK_A, 1YU0_A (Ca^{+2}), 1Z6F_A, 1ZJC_A (Co^{+2}), 2CZ1_B, 2GA2_A (Mn^{+2}), 2GDQ_B, 2GGO_A, 2GZ1_B, 2GZJ_F (Zn^{+2}), 2PUZ_B (Fe^{+3}), 3ETJ_A (Mg^{+2}).

References:

1. Basu S, Bhattacharyya D, Banerjee R. Self-Complementarity within Proteins: Bridging the Gap between Binding and Folding. Biophys J. 2012;102: 2605–2614. doi:10.1016/j.bpj.2012.04.029

2. Basu S, Bhattacharyya D, Banerjee R. SARAMA: A Standalone Suite of Programs for the Complementarity Plot—A Graphical Structure Validation Tool for Proteins. J Bioinforma Intell Control. 2013;2: 321–323. doi:10.1166/jbic.2013.1059

3. Basu S, Bhattacharyya D, Banerjee R. Applications of complementarity plot in error detection and structure validation of proteins. Indian J Biochem Biophys. 2014;51: 188–200.

4. Basu S, Bhattacharyya D, Banerjee R. Mapping the distribution of packing topologies within protein interiors shows predominant preference for specific packing motifs. BMC Bioinformatics. 2011;12: 195. doi:10.1186/1471-2105-12-195

5. Reduce Software : Kinemage Website [Internet]. [cited 12 Jun 2016]. Available: http://kinemage.biochem.duke.edu/software/reduce.php

6. Molecular Dynamics Simulations - Gromacs [Internet]. [cited 12 Jun 2016]. Available: http://www.gromacs.org/Documentation/Terminology/Molecular_Dynamics_Simulations

Please cite Ref(1) if you use the database in your publication.