Supplementary Material for

# Hydrogen bond co-ordination in organic crystal structures: statistics, prediction and application

**Peter T. A. Galek[a]*, James A. Chisholm[b], Elna Pidcock[a] and Peter A. Wood[a]**

Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge, CB4 3EZ, United Kingdom, and [b]Optibrium Ltd., 7226 Cambridge Research Park, Beach Drive, Cambridge, Cambridgeshire, CB25 9TL

**Table S1.** Fitting statistics for logistic regression models to predict donor acceptor co-ordination derived in this work. The values represent how much better the optimised model equation (residual) fits the sample data, compared with the independent model (null). "D. O. F." is independent degrees of freedom per sample, "P value" is the probability that one could obtain a given result randomly assuming the null hypothesis is correct. The acceptance criterion was $P < 0.05$. "AIC" is Akaike's Information Criterion and is another estimate of the residual deviance associated with lost information as a result of applying the model as a predictor of the complete data. "Area under ROC (receiver operator characteristic) curve" is a measure of classification of predictions. It is desirable to achieve a value as close to unity as possible, whereas 0.5 indicates a random predictor prestigious.

| Model name | Residual deviance | D.O.F | Null deviance | D.O.F | log likelihood | P value (log likelihood test) | AIC | Area Under Roc curve |
|---|---|---|---|---|---|---|---|---|
| Br_Br_a_1 | 66.6885 | 574 | 66.6885 | 574 | -33.3442 | 0 | 68.6885 | 1 |
| Br_ion_Br_a_1 | 22.2611 | 119 | 28.1081 | 120 | -11.1306 | 0.00886732 | 26.2611 | 0.861582 |
| Br_ion_Br_a_2 | 47.2159 | 119 | 160.723 | 120 | -23.6079 | 8.43E-27 | 51.2159 | 0.971304 |
| Br_ion_Br_a_3 | 38.639 | 119 | 117.695 | 120 | -19.3195 | 3.06E-19 | 42.639 | 0.976043 |
| Br_ion_Br_a_4 | 26.5183 | 119 | 58.9206 | 120 | -13.2592 | 6.45E-09 | 30.5183 | 0.971239 |
| Cl_Cl_a_1 | 282.326 | 1751 | 295.455 | 1753 | -141.163 | 0.000704751 | 288.326 | 0.696132 |
| Cl_ion_Cl_a_1 | 2.51E-09 | 432 | 0 | 432 | -1.26E-09 | 0 | 2 | 1 |
| Cl_ion_Cl_a_2 | 153.58 | 431 | 470.395 | 432 | -76.7902 | 3.59E-71 | 157.58 | 0.968985 |
| Cl_ion_Cl_a_3 | 273.751 | 431 | 585.771 | 432 | -136.876 | 3.98E-70 | 277.751 | 0.943216 |
| Cl_ion_Cl_a_4 | 198.314 | 431 | 392.857 | 432 | -99.1568 | 1.63E-44 | 202.314 | 0.942884 |
| Cl_ion_Cl_a_5 | 60.2204 | 431 | 109.724 | 432 | -30.1102 | 1.01E-12 | 64.2204 | 0.939628 |
| I_I_a_1 | 8.08019 | 146 | 12.0012 | 148 | -4.0401 | 0.070395 | 14.0802 | 0.966216 |
| I_ion_I_a_1 | 41.1898 | 57 | 50.3971 | 58 | -20.5949 | 0.00131675 | 45.1898 | 0.782222 |
| I_ion_I_a_2 | 34.9963 | 57 | 46.8321 | 58 | -17.4982 | 0.000312031 | 38.9963 | 0.860294 |
| I_ion_I_a_3 | 3.42E-10 | 58 | 0 | 58 | -1.71E-10 | 0 | 2 | 1 |
| T1NH0_a_1 | 31.2866 | 48 | 50.9134 | 51 | -15.6433 | 9.67E-05 | 39.2866 | 0.902381 |
| T1NH0_a_2 | 15.5893 | 48 | 28.2037 | 51 | -7.79465 | 0.00258322 | 23.5893 | 0.921875 |
| T1OH0_a_1 | 791.218 | 733 | 882.139 | 738 | -395.609 | 2.08E-18 | 803.218 | 0.725646 |
| T1OH0_a_2 | 447.532 | 732 | 876.536 | 738 | -223.766 | 8.01E-90 | 461.532 | 0.931777 |
| T2NH0_N.2_a_1 | 1262.81 | 1507 | 1620.21 | 1513 | -631.405 | 1.96E-74 | 1276.81 | 0.822924 |
| T2NH0_N.2_a_2 | 77.3623 | 1510 | 89.24 | 1513 | -38.6811 | 0.00362295 | 85.3623 | 0.804342 |
| T2OH_O.3_a_1 | 335.013 | 340 | 376.044 | 344 | -167.507 | 1.26E-08 | 345.013 | 0.725963 |
| T2OH_O.3_a_2 | 56.1167 | 342 | 83.3969 | 344 | -28.0584 | 5.96E-07 | 62.1167 | 0.89881 |
| T2OH_O.3_d_1 | 110.961 | 340 | 230.827 | 344 | -55.4803 | 2.80E-25 | 120.961 | 0.860752 |
| T2OH_O.3_d_2 | 98.2693 | 341 | 123.403 | 344 | -49.1346 | 6.97E-06 | 106.269 | 0.836768 |
| T2OH0_a_1 | 337.047 | 616 | 429.306 | 621 | -168.524 | 1.09E-18 | 349.047 | 0.814743 |
| T3CS_S.2_a_1 | 600.738 | 535 | 743.963 | 539 | -300.369 | 2.84E-30 | 610.738 | 0.804428 |
| T3CS_S.2_a_2 | 252.064 | 534 | 380.879 | 539 | -126.032 | 2.07E-26 | 264.064 | 0.903727 |
| T3NH1_d_1 | 819.604 | 1351 | 1310.95 | 1355 | -409.802 | 2.49E-105 | 829.604 | 0.87555 |
| T3NH1_d_2 | 653.236 | 1349 | 806.332 | 1355 | -326.618 | 8.35E-31 | 667.236 | 0.815805 |
| T3NH2_a_1 | 91.8939 | 249 | 228.17 | 254 | -45.947 | 5.41E-28 | 103.894 | 0.785125 |
| T3NH2_d_1 | 56.8518 | 254 | 56.8518 | 254 | -28.4259 | 0 | 58.8518 | 1 |
| T3NH2_d_2 | 209.98 | 250 | 260.653 | 254 | -104.99 | 1.26E-10 | 219.98 | 0.798804 |
| T3NH2_d_3 | 190.001 | 250 | 211.197 | 254 | -95.0005 | 0.000132318 | 200.001 | 0.6758 |
| T4NH2_d_1 | 14.5676 | 257 | 32.7839 | 261 | -7.28378 | 0.000504397 | 24.5676 | 0.981982 |
| T4NH2_d_2 | 111.333 | 259 | 169.463 | 261 | -55.6666 | 1.19E-13 | 117.333 | 0.885756 |
| T4NH2_d_3 | 228.221 | 258 | 286.721 | 261 | -114.111 | 6.05E-13 | 236.221 | 0.796935 |
| T4NH2_d_4 | 36.6635 | 259 | 49.493 | 261 | -18.3318 | 0.000818619 | 42.6635 | 0.897276 |
| T4PO_O.2_a_1 | 70.1498 | 157 | 124.402 | 159 | -35.0749 | 8.28E-13 | 76.1498 | 0.90579 |
| T4PO_O.2_a_2 | 60.0967 | 157 | 160.129 | 159 | -30.0483 | 9.49E-23 | 66.0967 | 0.96582 |
| acyclic_n_N.2_a_1 | 106.294 | 398 | 332.706 | 404 | -53.1472 | 2.19E-46 | 120.294 | 0.978734 |
| acyclic_n_N.2_a_2 | 2.75E-06 | 403 | 14.0053 | 404 | -1.37E-06 | 9.70E-05 | 4 | 1 |
| al_oh_O.3_a_1 | 1948.74 | 2072 | 2604.59 | 2078 | -974.372 | 1.03E-138 | 1962.74 | 0.828376 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| al_oh_O.3_a_2 | 228.845 | 2074 | 338.92 | 2078 | -114.423 | 3.44E-23 | 238.845 | 0.907758 |
| al_oh_O.3_d_1 | 938.002 | 2073 | 1643.19 | 2078 | -469.001 | 1.85E-150 | 950.002 | 0.870978 |
| al_oh_O.3_d_2 | 546.692 | 2075 | 571.546 | 2078 | -273.346 | 7.97E-06 | 554.692 | 0.63995 |
| ar_n_N.ar_a_1 | 1452.03 | 1415 | 1967.86 | 1421 | -726.016 | 1.62E-108 | 1466.03 | 0.816017 |
| ar_n_N.ar_a_2 | 286.442 | 1415 | 371.599 | 1421 | -143.221 | 1.46E-16 | 300.442 | 0.876901 |
| ar_oh_O.3_a_1 | 1103.37 | 1459 | 1468.5 | 1463 | -551.686 | 4.72E-78 | 1113.37 | 0.853326 |
| ar_oh_O.3_a_2 | 131.87 | 1458 | 211.453 | 1463 | -65.935 | 4.94E-16 | 143.87 | 0.939889 |
| ar_oh_O.3_d_1 | 588.882 | 1460 | 2026.38 | 1463 | -294.441 | 1.0767E-311 | 596.882 | 0.970421 |
| ar_oh_O.3_d_2 | 287.546 | 1460 | 330.507 | 1463 | -143.773 | 1.23E-09 | 295.546 | 0.806458 |
| carbamoyl_N.am_d_1 | 39.1272 | 287 | 58.4125 | 289 | -19.5636 | 3.25E-05 | 45.1272 | 0.893192 |
| carbamoyl_N.am_d_2 | 198.048 | 285 | 290.233 | 289 | -99.0241 | 2.21E-19 | 208.048 | 0.856124 |
| carbamoyl_N.am_d_3 | 56.5818 | 285 | 73.2243 | 289 | -28.2909 | 0.00101225 | 66.5818 | 0.845301 |
| carbamoyl_O.2_a_1 | 79.8799 | 287 | 99.9358 | 289 | -39.94 | 2.21E-05 | 85.8799 | 0.858213 |
| carbamoyl_O.2_a_2 | 239.478 | 283 | 401.529 | 289 | -119.739 | 1.06E-32 | 253.478 | 0.888418 |
| carbamoyl_O.2_a_3 | 32.6189 | 284 | 58.4125 | 289 | -16.3094 | 4.37E-05 | 44.6189 | 0.97007 |
| carbonyl_O.2_a_1 | 3883.08 | 3606 | 4984.24 | 3612 | -1941.54 | 5.82E-235 | 3897.08 | 0.815624 |
| carbonyl_O.2_a_2 | 701.414 | 3606 | 1086.62 | 3612 | -350.707 | 2.09E-80 | 715.414 | 0.900188 |
| carboxylate_O.co2_a_1 | 416.463 | 1334 | 685.944 | 1339 | -208.232 | 1.79E-56 | 428.463 | 0.915544 |
| carboxylate_O.co2_a_2 | 1480.11 | 1336 | 1853.76 | 1339 | -740.053 | 5.60E-81 | 1488.11 | 0.789845 |
| carboxylate_O.co2_a_3 | 411.642 | 1335 | 537.894 | 1339 | -205.821 | 1.21E-26 | 421.642 | 0.846201 |
| cooh_O.2_a_1 | 773.26 | 788 | 1096.54 | 792 | -386.63 | 5.10E-69 | 783.26 | 0.852655 |
| cooh_O.2_a_2 | 308.047 | 787 | 404.802 | 792 | -154.024 | 1.24E-19 | 320.047 | 0.872601 |
| cooh_O.3_a_1 | 285.063 | 791 | 373.549 | 795 | -142.531 | 1.35E-18 | 295.063 | 0.833941 |
| cooh_O.3_d_1 | 64.5855 | 793 | 492.476 | 795 | -32.2927 | 6.08E-94 | 70.5855 | 0.979262 |
| cooh_O.3_d_2 | 159.794 | 792 | 179.479 | 795 | -79.8969 | 9.41E-05 | 167.794 | 0.763192 |
| cyclic_amide_N.am_d_1 | 279.069 | 1374 | 455.791 | 1378 | -139.534 | 1.86E-37 | 289.069 | 0.886709 |
| cyclic_amide_N.am_d_2 | 376.433 | 1373 | 436.414 | 1378 | -188.216 | 5.84E-12 | 388.433 | 0.772502 |
| cyclic_amide_O.2_a_1 | 1363.72 | 1488 | 1662.45 | 1491 | -681.862 | 9.35E-65 | 1371.72 | 0.788887 |
| cyclic_amide_O.2_a_2 | 783.738 | 1487 | 1212.47 | 1491 | -391.869 | 8.55E-92 | 793.738 | 0.898681 |
| cyclic_n_N.2_a_1 | 631.376 | 734 | 1015.54 | 740 | -315.688 | 3.50E-80 | 645.376 | 0.874271 |
| cyclic_n_N.2_a_2 | 128.043 | 738 | 161.95 | 740 | -64.0217 | 2.17E-08 | 134.043 | 0.85676 |
| cyclic_nn_N.2_a_1 | 183.392 | 176 | 253.03 | 182 | -91.6958 | 2.29E-13 | 197.392 | 0.82654 |
| cyclic_nn_N.2_a_2 | 1.06E-09 | 182 | 0 | 182 | -5.31E-10 | 0 | 2 | 1 |
| ether_O.3_a_1 | 1924.69 | 3705 | 2108.62 | 3711 | -962.347 | 2.43E-37 | 1938.69 | 0.749569 |
| nh3_N.4_d_1 | 1.24E-07 | 583 | 14.7449 | 585 | -6.22E-08 | 0.000314158 | 6 | 1 |
| nh3_N.4_d_2 | 25.1114 | 581 | 47.8688 | 585 | -12.5557 | 6.51E-05 | 35.1114 | 0.947595 |
| nh3_N.4_d_3 | 99.4905 | 582 | 117.074 | 585 | -49.7452 | 0.000254184 | 107.49 | 0.836963 |
| nh3_N.4_d_4 | 754.143 | 581 | 793.088 | 585 | -377.072 | 3.40E-08 | 764.143 | 0.636067 |
| nh3_N.4_d_5 | 219.002 | 579 | 265.125 | 585 | -109.501 | 1.28E-08 | 233.002 | 0.807623 |
| nitrile_N.1_a_1 | 624.196 | 495 | 676.138 | 497 | -312.098 | 2.63E-12 | 630.196 | 0.752096 |
| nitrile_N.1_a_2 | 66.8244 | 494 | 81.9693 | 497 | -33.4122 | 0.000798674 | 74.8244 | 0.848214 |
| nitro_O.2_a_1 | 1159.14 | 1329 | 1539.61 | 1333 | -579.569 | 2.29E-81 | 1169.14 | 0.821934 |
| nitro_O.2_a_2 | 248.734 | 1328 | 495.545 | 1333 | -124.367 | 1.31E-51 | 260.734 | 0.95212 |
| oxime_N.2_a_1 | 272.297 | 274 | 389.46 | 280 | -136.148 | 3.10E-23 | 286.297 | 0.866373 |
| oxime_O.3_a_1 | 124.569 | 276 | 154.313 | 280 | -62.2845 | 2.59E-06 | 134.569 | 0.753422 |
| oxime_O.3_d_1 | 1.63E-09 | 279 | 23.7666 | 280 | -8.15E-10 | 5.65E-07 | 4 | 1 |
| prim_amine_a_1 | 126.991 | 1159 | 400.081 | 1163 | -63.4954 | 3.42E-58 | 136.991 | 0.985196 |
| prim_amine_d_1 | 522.085 | 1160 | 618.635 | 1163 | -261.043 | 4.24E-21 | 530.085 | 0.809198 |
| prim_amine_d_2 | 1170.32 | 1158 | 1567.9 | 1163 | -585.159 | 4.89E-84 | 1182.32 | 0.823883 |
| prim_amine_d_3 | 533.372 | 1159 | 577.56 | 1163 | -266.686 | 2.81E-09 | 543.372 | 0.722639 |
| sec_amine_a_1 | 288.975 | 2114 | 687.75 | 2119 | -144.487 | 2.70E-84 | 300.975 | 0.966093 |
| sec_amine_d_1 | 2016.27 | 2114 | 2787.08 | 2119 | -1008.14 | 1.19E-164 | 2028.27 | 0.836787 |
| sec_amine_d_2 | 530.589 | 2115 | 635.289 | 2119 | -265.295 | 4.82E-22 | 540.589 | 0.816189 |
| sulfinyl_O.2_a_1 | 1.03E-07 | 31 | 32.7996 | 36 | -5.16E-08 | 1.88E-06 | 12 | 1 |
| sulfinyl_O.2_a_2 | 27.865 | 35 | 32.7996 | 36 | -13.9325 | 0.0152315 | 31.865 | 0.72043 |
| sulfonyl_O.2_a_1 | 1659.29 | 1552 | 2148.31 | 1557 | -829.647 | 9.33E-104 | 1671.29 | 0.804994 |
| sulfonyl_O.2_a_2 | 542.256 | 1553 | 923.04 | 1557 | -271.128 | 1.96E-81 | 552.256 | 0.915467 |
| tert_amine_N.3_a_1 | 507.52 | 576 | 750.648 | 580 | -253.76 | 9.75E-52 | 517.52 | 0.863894 |
| tert_amine_N.3_a_2 | 25.1526 | 578 | 37.5813 | 580 | -12.5763 | 0.00100027 | 31.1526 | 0.934833 |
| tert_ammonia_N.4_d_1 | 5.87981 | 228 | 171.402 | 233 | -2.9399 | 3.23E-34 | 17.8798 | 0.999653 |
| tert_ammonia_N.4_d_2 | 107.518 | 231 | 154.759 | 233 | -53.759 | 2.76E-11 | 113.518 | 0.864087 |
| thioether_S.3_a_1 | 144.323 | 898 | 144.323 | 898 | -72.1616 | 0 | 146.323 | 1 |
| transamide_N.am_d_1 | 1328.65 | 2237 | 2246.36 | 2243 | -664.327 | 2.78E-195 | 1342.65 | 0.886814 |
| transamide_N.am_d_2 | 306.515 | 2238 | 335.559 | 2243 | -153.258 | 1.03E-05 | 318.515 | 0.749887 |
| transamide_O.2_a_1 | 2001.01 | 2158 | 2800.02 | 2164 | -1000.51 | 1.25E-169 | 2015.01 | 0.85048 |
| transamide_O.2_a_2 | 901.113 | 2159 | 1378.83 | 2164 | -450.557 | 2.56E-101 | 913.113 | 0.88946 |
| water_O.3_a_1 | 524.097 | 501 | 555.031 | 502 | -262.049 | 1.38E-08 | 528.097 | 0.68727 |
| water_O.3_a_2 | 426.141 | 500 | 466.431 | 502 | -213.071 | 8.91E-10 | 432.141 | 0.701862 |
| water_O.3_d_1 | 60.3828 | 501 | 65.0741 | 502 | -30.1914 | 0.0176423 | 64.3828 | 0.594567 |
| water_O.3_d_2 | 220.591 | 501 | 232.818 | 502 | -110.296 | 0.000252439 | 224.591 | 0.681862 |
| water_O.3_d_3 | 299.315 | 501 | 303.102 | 502 | -149.657 | 0.0308631 | 303.315 | 0.59049 |