

Erratum: How accurate and statistically robust are catalytic site predictions based on closeness centrality?

Eric Chea¹ and Dennis R. Livesay^{*2}

¹ Department of Biological Sciences, California State Polytechnic University, Pomona, CA, 91768 USA. ² Department of Computer Science and Bioinformatics Research Center, University of North Carolina at Charlotte, Charlotte, NC, 28223 USA.

In the original paper, there are errors in Table 2d. They are fixed here (highlighted in red). Fortunately, the absolute accuracies, improvements, etc. presented elsewhere (Table 2 and other tables/figures) were correct in the original manuscript. Meaning, the main results of the paper and the conclusions drawn there from are not affected. We sincerely apologize for any inconvenience this might have caused, and thank Zhang Tuo for alerting us of the errors.

Table 2. Evaluation of catalytic site predictions.¹

| Avg. # / PDB | Total accuracy ² | Per PDB accuracy ³ | p-value ⁴ | TP & FP rate ⁵ | TP:FP ratio | 1 correct per PDB ⁶ | 1 correct expect ⁷ |
|---|-----------------------------|-------------------------------|----------------------|---------------------------|-------------|--------------------------------|-------------------------------|
| (a.) Raw CC values (no filter) | | | | | | | |
| 1.3 | 6.0 | 2.7 (10.8) | 2.7E-09 | 2.1 / 0.4 | 6.0 | 7.6 | 1.1 |
| 2.4 | 6.8 | 4.2 (11.6) | 2.8E-22 | 4.9 / 0.7 | 7.2 | 15.0 | 2.0 |
| 3.6 | 6.5 | 4.5 (10.6) | 2.4E-30 | 7.0 / 1.0 | 6.9 | 19.9 | 3.1 |
| 4.6 | 6.3 | 4.7 (10.0) | 2.4E-37 | 8.8 / 1.3 | 6.9 | 23.4 | 3.9 |
| 5.7 | 6.3 | 4.9 (9.6) | 9.4E-47 | 11.0 / 1.6 | 6.9 | 27.6 | 4.8 |
| (b.) Solvent accessibility filter | | | | | | | |
| 1.1 | 14.2 | 7.5 (17.4) | 2.8E-42 | 5.3 / 0.3 | 18.7 | 15.9 | 1.0 |
| 2.2 | 13.0 | 9.2 (16.8) | 7.5E-72 | 9.7 / 0.6 | 16.9 | 25.4 | 1.9 |
| 3.3 | 11.1 | 8.7 (14.7) | 6.8E-82 | 12.2 / 0.9 | 14.2 | 29.3 | 2.9 |
| 4.4 | 10.8 | 8.9 (13.2) | 4.5E-103 | 15.8 / 1.2 | 13.7 | 36.7 | 3.9 |
| 5.4 | 10.4 | 8.9 (12.4) | 2.7E-120 | 18.8 / 1.4 | 13.2 | 41.3 | 4.8 |
| (c.) Residue identity filter | | | | | | | |
| 1.1 | 22.4 | 11.3 (21.0) | 3.8E-83 | 8.3 / 0.3 | 32.6 | 23.0 | 1.0 |
| 2.2 | 19.6 | 13.5 (19.8) | 8.8E-134 | 14.5 / 0.5 | 27.6 | 35.7 | 1.9 |
| 3.2 | 17.9 | 13.8 (18.2) | 0.0 | 19.2 / 0.8 | 24.7 | 42.8 | 2.8 |
| 4.3 | 17.6 | 14.3 (17.0) | 0.0 | 25.0 / 1.0 | 24.1 | 50.5 | 3.7 |
| 5.2 | 16.5 | 13.9 (15.3) | 0.0 | 29.3 / 1.3 | 22.4 | 56.2 | 4.7 |
| (d.) Combination filter (solvent accessibility + residue identity) | | | | | | | |
| 1.1 | 25.2 | 12.9 (21.8) | 6.8E-72 | 9.1 / 0.2 | 38.0 | 26.1 | 1.0 |
| 2.1 | 20.7 | 14.4 (20.7) | 2.4E-103 | 15.0 / 0.5 | 29.4 | 36.7 | 1.9 |
| 3.1 | 17.9 | 13.5 (17.1) | 6.5E-115 | 18.7 / 0.8 | 24.6 | 44.2 | 2.7 |
| 4.1 | 15.9 | 12.8 (14.6) | 2.6E-123 | 22.0 / 1.0 | 21.3 | 49.8 | 3.6 |
| 5.2 | 13.9 | 11.7 (13.1) | 8.1E-123 | 24.3 / 1.3 | 18.2 | 53.0 | 4.6 |

¹ Statistics describing the accuracy of the accessibility-filtered prediction on the SCOP superfamily dataset. ² Accuracy is defined as the percentage of correct catalytic residue predictions out of the total number of predictions for the entire collapsed dataset. In all cases, the random expectation is 0.9%. ³ Average value (and standard deviation) of accuracy calculated on a per protein basis. ⁴ The probability that the null hypothesis is correct calculated from the binomial distribution. ⁵ The true positive rate is the percent correct of the total number of catalytic residues within the CSA; similarly, the false positive rate is the percent incorrect predictions of the total number of noncatalytic residues. ⁶ The percent of proteins with at least one correct prediction. ⁷ The expected percent of proteins with at least one correct assuming a random model.