

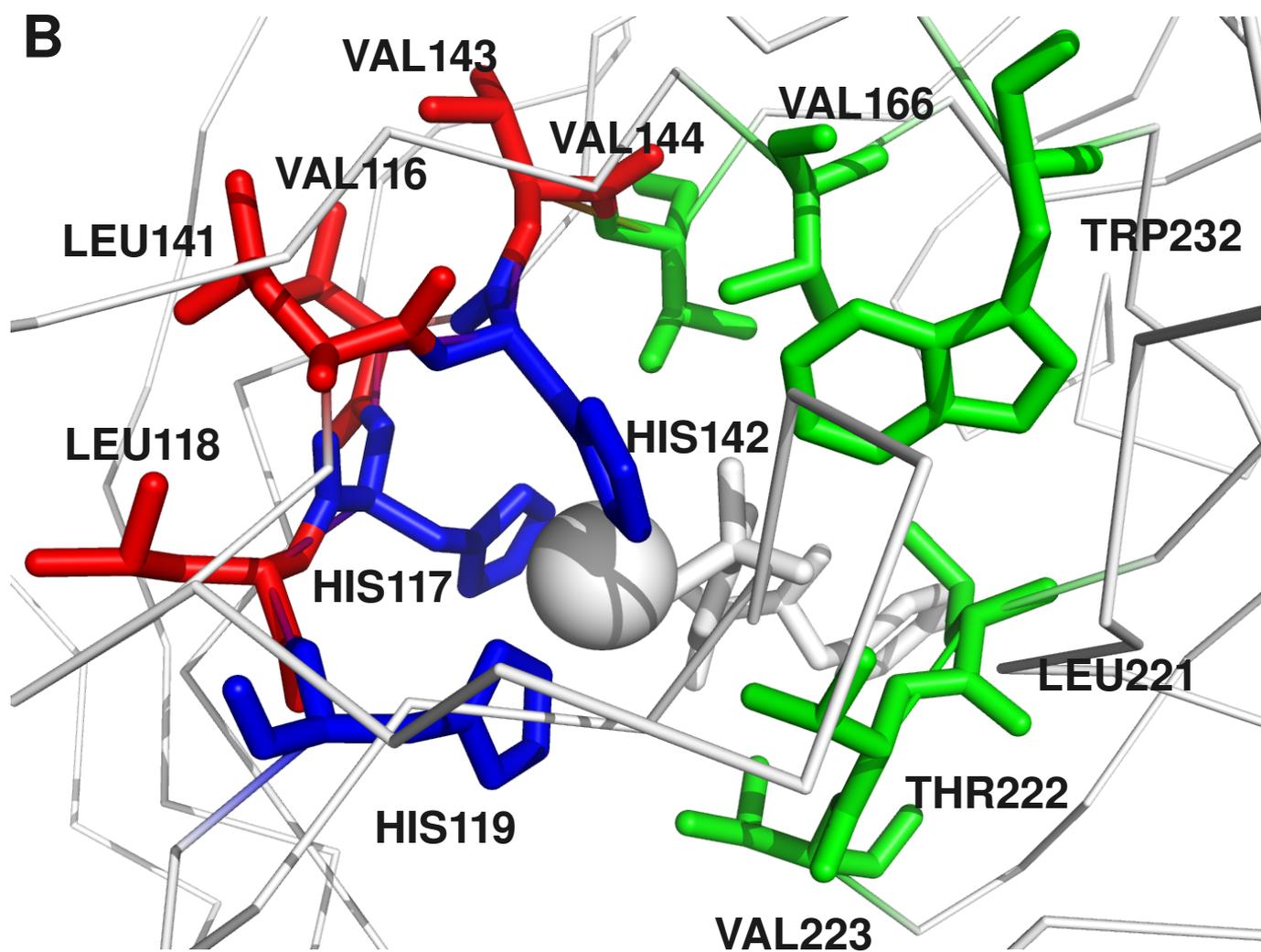
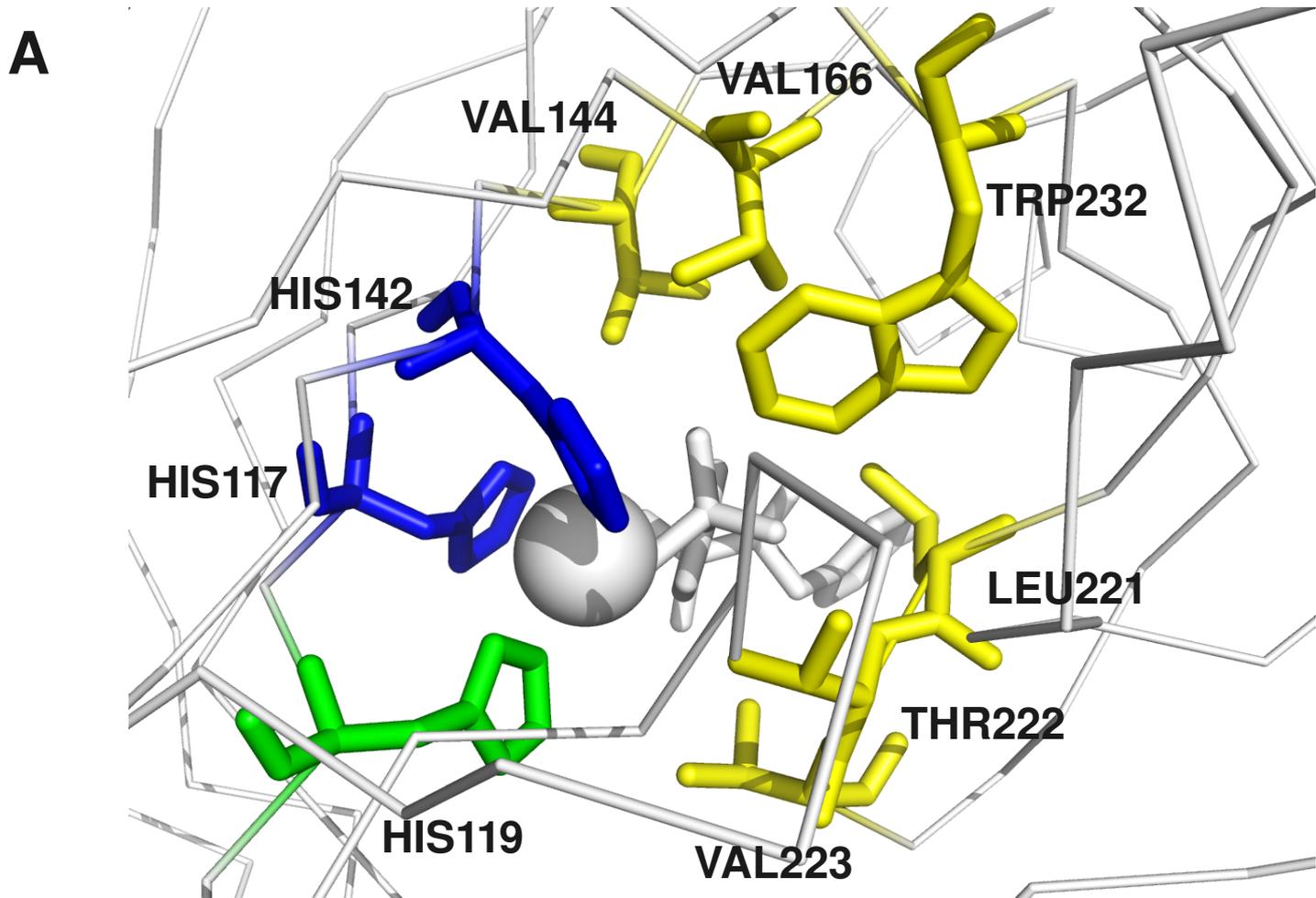
Supplementary Figure 2. Further examples of large differences occurring in the BDT and MCC scores assigned to predictions of binding site residues. Blue sticks represent the observed residues, red sticks represent incorrectly predicted residues, green sticks represent correctly predicted residues, yellow sticks represent neutral residues (ambiguous residues defined in CASP8 that were considered to be in the active site, or were considered to be in contact with an alternative ligand, but were nevertheless excluded from the observed subset) and the ligands are shown in white.

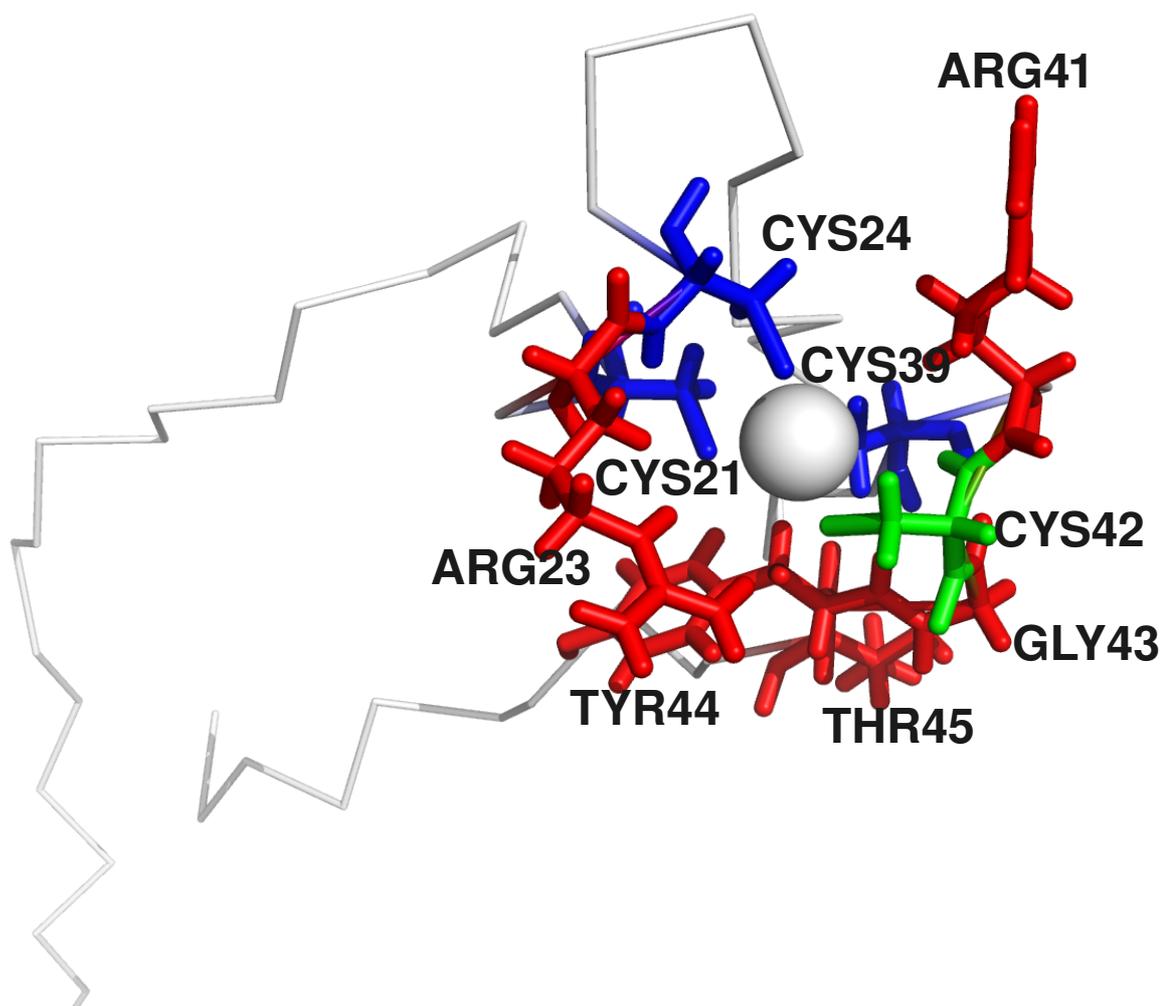
A) Ribbon diagram of CASP8 target T0426. The second prediction made by group FN450 is shown (T0426FN450_2). One residue (119) out of the 3 observed binding site residues (117, 119, 142) was correctly predicted. When the neutral residues are excluded from the calculations, this prediction results in an MCC score of 0.575 and a BDT score (with $d_o=3\text{\AA}$) of 0.333. If the neutral residues (114, 116, 221, 222, 223, 232) are included in the calculation as true negative residues (TNs) the MCC score is further increased to 0.705, however the BDT score remains at 0.333.

B) Ribbon diagram of CASP8 target T0426. This example shows a hypothetical prediction highlighting the difference between the MCC and BDT scores for the predicted residues close to the binding site, which were not included in the official observed subset. For the purposes of this hypothetical example the residues that were defined as neutral (114, 116, 221, 222, 223, 232) have been shown as green sticks and are predicted along with residues 116, 118, 141 and 143 (red sticks). If the “neutral” residues are considered to be correctly predicted then this results in an MCC score of 0.619 and BDT score (with $d_o=3\text{\AA}$) of 0.753. However, if the “neutral” residues are considered incorrect and are included in the MCC calculation as TNs, then the MCC score is drastically reduced to -0.00001289, whilst the BDT score is only reduced to 0.382. Finally, if the residues are considered to be truly neutral and are excluded from the MCC calculation, this still results in a low MCC score of -0.0134, whilst the BDT score remains at 0.382.

C) Ribbon diagram of CASP8 target T0480. The second prediction made by group FN057 is shown (T0480FN057_2). For this prediction, residue 42 was correctly predicted from the set of observed residues – 21, 24, 39, 42. However, group FN057 also over predicted residues 45, 43, 44, 42, 41 and 23. The MCC score for this prediction is 0.127 however the BDT score (with $d_o=3\text{\AA}$) is 0.436. Target T0480 has two disordered regions at the ends of the protein and if these regions are excluded from the MCC calculation then the MCC score is further reduced to 0.055 but the BDT score remains unchanged.

D) Predicted and observed residues for CASP8 target T0396. The ribbon backbone trace has been removed for clarity. The second prediction made by group FN407 is shown (T0396FN407_2). This prediction obtains a low MCC score of 0.007 and comparatively high BDT score (with $d_o=3\text{\AA}$) of 0.3833. Group FN407 had correctly predicted 6 residues 15, 75, 77, 78, 79 and 90 (green sticks) but had also incorrectly predicted 19 residues (red sticks) that were nevertheless close to the binding site. T0396 is an example of a relatively small 105 residue protein containing a large FAD binding site of 23 residues. In cases such as this it might be useful to set a more stringent distance threshold in order to differentiate more clearly between close predictions, thus if d_o is set to 1 then the BDT score for this prediction is reduced to 0.253.



C**D**