

RECORDED MODEL BEHAVIOR

Assess a hidden binding pocket at the NRF2-BTRCP interface

7/15 criteria passed

GPT-6.1-SOL · INDEPENDENT REGRADE OF SAVED ANSWER

The solver answer was reused from an earlier fresh solve because its public inputs were unchanged. A new grader assessed it under the final rubric; this is not a new solver attempt.

FAILED CRITERIA IN THE SAVED GRADE

C-001 - Interpret the pocket map as a result across simulated motion: The review correctly describes recurring pocket geometry rather than ligand occupancy. However, it repeats the >50% threshold without calculating that a displayed point must occur in more than 500, or at least 501, of the 1,000 frames.

C-002 - Keep the reference phosphorylation comparison matched: The review correctly identifies first-pSer 6M93 and phosphate-edited 1P22, and explains why their geometries need not match. Its broad yellow-versus-red comparison does not describe the pocket extent in both the top and side views, as required.

C-003 - Use reference pose recovery as the method check: The review describes general visual agreement in the beta-catenin overlay, but does not identify two corresponding ligand features. Sharing an interface region alone does not demonstrate recovery of the ligand's position and orientation.

C-007 - Separate SDS3 results from SDS1 results: The review identifies the different SDS1 and SDS3 conditions and describes 0253 retention. However, it does not explain that peptide identity and phosphorylation change together, preventing either factor alone from being identified as the cause. It also leaves the SDS3 complex3 rejection unstated.

C-012 - Read ligand orientation rather than only cavity size: The correct opened SDS1/Nurix context is present, but the review describes only overlapping ligand regions. It does not describe overlapping ring systems or two aligned ligand parts that support a reference-like position and orientation.

C-013 - Explain why a local opening need not destroy peptide recognition: The review uses cyan receptor-backbone overlap to argue against wholesale receptor rearrangement. It does not connect the peptide overlap away from the entry to preserved peptide placement on BTRCP. Similar receptor folds do not establish that required peptide-recognition inference.

C-017 - Keep pocket claims specific to the phosphate state: The review correctly identifies the first-pSer study context, but groups the other phosphate states together as having different shapes. It does not locate a specific first-pSer versus double-pSer difference using the central cavity as a landmark.

C-020 - Leave protein-contact stabilization unresolved by small-molecule binding: The review correctly leaves stabilization unproven, but never identifies the needed comparison of SDS1-BTRCP affinity with versus without Nurix. Merely listing affinity and stabilization as unmeasured does not distinguish that comparison from measuring Nurix binding itself.

Read `runs/gpt-6.1-sol/evaluation/outputs/submission/, outputs/judgments.json, outputs/report.md` and `attempts/0001/` for the unedited answer, reasons and events. These are recorded grader judgments, not a new adjudication.