

Peer-Review

Ashraf, Dylan. 2026. "Target-Family Differences in Stereochemical Discrimination: A Same-Assay Analysis of Enantiomer Pairs in ChEMBL." *Journal of High School Science* 10 (3): 608–40. <https://doi.org/10.64336/001c.170283>.

Comprehensive Scientific Review and Recommendations

The manuscript presents a genuinely interesting archive-scale analysis of stereoselectivity across protein target families. Unlike many descriptive cheminformatics studies that merely report distributions or develop predictive models, this work identifies a reproducible statistical association between target family and the distribution of eudysmic ratios using a carefully constructed same-assay dataset. The methodology is rigorous, the controls are thoughtful, and the authors are appropriately cautious in distinguishing measured observations from mechanistic interpretations. Overall, this is one of the stronger conceptual studies in this area. However, the current manuscript stops at describing an association. The opportunity exists to elevate the work substantially by reframing the biological implications and outlining a broader research program built upon these findings.

I. Current limitations of the manuscript and recommendations to address them

1. The manuscript measures an interaction property but frames it largely as a ligand property

The eudysmic ratio (ER) is measured from ligand potency differences between two enantiomers, yet biologically it reflects how strongly a protein binding pocket discriminates between stereochemical configurations. The manuscript therefore measures an interaction between ligand stereochemistry and protein architecture rather than an intrinsic ligand property.

The Discussion should more explicitly emphasize that the observed differences primarily reflect differences in protein stereochemical discrimination, while acknowledging that ligand chemistry contributes to the measured values.

2. Protein family is treated as the explanatory variable rather than the underlying pocket properties

Protein family is only a surrogate.

The true determinants are likely to be:

- binding pocket geometry
- enclosure
- hydrogen-bond architecture
- residue arrangement
- flexibility
- electrostatics
- conformational dynamics

Future analyses should move beyond family labels and investigate which structural properties actually predict stereochemical discrimination.

3. Residual medicinal chemistry bias cannot be excluded

Although the manuscript controls for:

- stereocenter count,
- scaffold clustering,
- assay type,
- selection bias,
- potency,

one important possibility remains.

Historically, protease inhibitors, kinase inhibitors, GPCR ligands, and nuclear receptor ligands have been developed using very different medicinal chemistry strategies.

Consequently, the observed differences may reflect:

- ligand rigidity,
- peptide-like scaffolds,

pharmacophore placement,
conformational restriction,
rather than intrinsic protein properties.

Additional ligand descriptors should therefore be incorporated into multivariable analyses.

4. The manuscript demonstrates

$P(\text{ER}|\text{protein family})$

but does not investigate

$P(\text{protein family}|\text{ER})$

The current work establishes that different protein families exhibit different ER distributions.

It does not determine whether ER contains enough information to predict target family.

A predictive analysis would greatly strengthen the biological relevance.

5. The study stops after identifying an association

The manuscript demonstrates that the association exists but does not attempt to explain its origin.

A stronger contribution would begin identifying the structural, evolutionary, or mechanistic basis underlying the observed hierarchy.

II. Recommendations for improving the manuscript

1. Develop a quantitative measure of protein stereochemical discrimination

Rather than focusing solely on ligand eudysmic ratios, introduce the broader concept that proteins possess differing capacities to distinguish stereochemical information.

For example, define a protein-specific descriptor such as:

Protein Stereochemical Discrimination (PSD)

or

Stereochemical Discrimination Index (SDI)

computed from the distribution of ER values associated with a given protein.

This reframes the work from ligand statistics toward protein characterization.

2. Relate stereochemical discrimination to structural biology

Instead of comparing only protein families, quantify structural characteristics of binding pockets.

Potential variables include:

pocket depth

pocket volume

buriedness

hydrogen-bond density

hydrophobicity

contact density

solvent accessibility

flexibility

Regression against these descriptors could identify the physical determinants of stereochemical discrimination.

3. Incorporate ligand complexity into multivariable analyses

Calculate ligand descriptors such as:

molecular weight

LogP

TPSA

rotatable bonds

stereocenter count

conformational flexibility

rigidity

pharmacophore complexity

Then evaluate whether protein family remains independently associated with ER after adjustment.

This would substantially strengthen the biological interpretation.

4. Evaluate predictive utility

Determine whether ER carries predictive information regarding target family.

Possible analyses include:

multinomial logistic regression

Bayesian classification

random forest classification

ROC analysis

mutual information

entropy reduction

This distinguishes a descriptive observation from one with practical predictive value.

5. Investigate evolutionary explanations

The observed hierarchy may reflect evolutionary pressures.

Examples include:

proteases requiring high substrate specificity

kinases evolving around ATP recognition

GPCRs balancing specificity with ligand diversity

Developing this discussion would substantially deepen the biological significance.

6. Investigate structural binding mechanisms

For proteins with available crystal structures:

dock both enantiomers

compare binding poses

quantify hydrogen bond differences

evaluate steric clashes

compare interaction fingerprints

Correlating these structural differences with ER would directly test the mechanistic interpretation.

7. Present stereochemical discrimination as a new protein descriptor

The manuscript has the potential to introduce stereochemical discrimination as a novel descriptor of protein molecular recognition.

Rather than viewing ER merely as a ligand property, propose that the aggregate stereochemical discrimination of a protein represents a measurable characteristic analogous to:

catalytic efficiency

binding affinity

thermal stability

evolutionary conservation

structural flexibility

This conceptual reframing could considerably increase the long-term impact of the study.

III. Proposed “Perspectives” section

The current manuscript ends with a discussion of limitations and robustness. A dedicated

Perspectives section would be an excellent addition because it would clearly distinguish measured results from future hypotheses. The following hypotheses are scientifically motivated extensions of the present work rather than conclusions supported by the current dataset.

Perspective 1. Evolution of drug resistance

Investigate whether proteins exhibiting higher stereochemical discrimination evolve drug resistance differently.

Possible questions include:

Are high-SD proteins more sensitive to resistance mutations?

Do resistance mutations preferentially reduce stereochemical discrimination?

Does SD correlate with mutation tolerance?

Such studies could connect stereochemical recognition with evolutionary dynamics.

Perspective 2. Druggability

Determine whether SD predicts successful drug discovery.

Potential investigations include:

hit identification rates
optimization efficiency
achievable selectivity

medicinal chemistry tractability

If high-SD proteins consistently produce more selective ligands, SD may become a useful descriptor during target selection.

Perspective 3. Ligand promiscuity

Evaluate whether proteins exhibiting low stereochemical discrimination bind a broader diversity of chemical scaffolds.

If validated, SD may represent an orthogonal measure of target promiscuity.

Perspective 4. Binding-pocket architecture

Correlate SD with structural properties such as:

pocket depth

enclosure

flexibility

residue conservation

hydrogen-bond density

These studies could reveal the structural determinants responsible for stereochemical discrimination.

Perspective 5. Multiple binding modes

Low stereochemical discrimination may arise because:

the pocket is intrinsically symmetric,

stereochemistry is biologically unimportant,

or both enantiomers adopt different but energetically equivalent binding poses.

Combining SD with crystallography or molecular dynamics simulations may distinguish among these possibilities.

Perspective 6. Clinical pharmacology

Investigate whether SD correlates with clinically relevant properties including:

robustness of inhibitor design,

susceptibility to acquired resistance,

selectivity profiles,

off-target pharmacology,

durability of therapeutic response.

These hypotheses require independent validation but could substantially broaden the clinical relevance of stereochemical discrimination.

Overall assessment

The present manuscript establishes a robust statistical association between target family and stereochemical discrimination using a carefully curated archive-scale dataset. However, its greatest scientific contribution may lie beyond the reported observations. By reframing the results in terms of protein stereochemical discrimination, the work has the potential to introduce a new quantitative descriptor of molecular recognition. A dedicated Perspectives section outlining the structural, evolutionary, pharmacological, and clinical hypotheses arising from this observation would considerably enhance the manuscript, clearly distinguishing established findings from future research directions while illustrating the broader significance of the work.

Response to Reviewers

Dylan Ashraf

Manuscript: "How Much Does Handedness Matter? Stereochemical Discrimination by Protein Target Family, Measured from Same-Assay Enantiomer Pairs in ChEMBL"

dylanashraf56014@gmail.com

Summary of the revision

The author thanks both reviewers for detailed and constructive comments. Every comment is reproduced verbatim below, followed by the response and the location of the corresponding change in the revised manuscript.

The revision does three things. First, it adopts Reviewer 1's central recommendation and reframes the measurement as a property of the protein–ligand interaction, adding a target-level descriptor together with the reliability tests that decide whether such a descriptor is meaningful. Second, it adds the two analyses that Reviewer 1's criticisms specifically demanded and that the existing data can support: a multivariable adjustment for ligand chemistry, and a measurement of how much information the eudysmic ratio carries about target family. Third, prompted by Reviewer 2's requirement that all implicit assumptions be stated, the author re-audited the submitted manuscript against its own result files. That audit found several statements that were stronger than the underlying numbers supported, and it also identified a previously unstated assumption about censored measurements. Those corrections are set out in a dedicated section at the end of this document, because they change some statements in the original manuscript and the reviewers and editor should see them explicitly rather than discover them in the revised text.

New material in the revised manuscript:

- **Section 2.6** — ligand descriptors (new)
- **Section 2.7** — the target-level eudysmic index and its reliability tests (new)
- **Section 2.10** — explicit assumptions, A1–A10 (new)
- **Section 3.11** — ligand chemistry differs sharply between families, and the family effect survives adjustment for it (new)
- **Section 3.12** — the target-level eudysmic index and whether it is a property of the protein (new)
- **Section 3.13** — how much the eudysmic ratio says about the target family (new)
- **Section 3.14** — censoring, truncation, and the conservatism of the reported values (new)
- **Section 5** — Perspectives (new)
- **Acknowledgments** — including disclosure of AI assistance (new)
- **Figures 6, 7 and 8** (new); **Figure 1** regenerated; **Table 5** (new)
- References expanded from 13 to 44, with two errors in the original list corrected

Reviewer 1

General assessment

> **Comment.** "The manuscript presents a genuinely interesting archive-scale analysis of stereoselectivity across protein target families. Unlike many descriptive cheminformatics studies that merely report distributions or develop predictive models, this work identifies a reproducible statistical association between target family and the distribution of eudysmic ratios using a carefully constructed same-assay dataset. The methodology is rigorous, the controls are thoughtful, and the authors are appropriately cautious in distinguishing measured observations from mechanistic interpretations. Overall, this is one of the stronger conceptual studies in this area. However, the current manuscript stops at describing an association. The opportunity exists to elevate the work substantially by reframing the biological implications and outlining a broader research program built upon these findings."

Response. The author is grateful for this assessment and has adopted the recommended reframing throughout. The revision keeps the measurement discipline the reviewer credits while moving the interpretive frame from ligand statistics to protein characterization, and it adds a Perspectives section (Section 5) that sets out the broader research program without presenting any of it as a finding of this study.

I.1 — The manuscript measures an interaction property but frames it largely as a ligand property

> **Comment.** "The eudysmic ratio (ER) is measured from ligand potency differences between two enantiomers, yet biologically it reflects how strongly a protein binding pocket discriminates

between stereochemical configurations. The manuscript therefore measures an interaction between ligand stereochemistry and protein architecture rather than an intrinsic ligand property. The Discussion should more explicitly emphasize that the observed differences primarily reflect differences in protein stereochemical discrimination, while acknowledging that ligand chemistry contributes to the measured values."

Response. Adopted. This point is now made explicitly in three places rather than left implicit. The **Introduction** (Section 1, paragraph 2) now states the framing directly: the eudysmic ratio is not a property of a ligand but of the interaction, because holding the ligand pair fixed and changing the protein changes the number — which is what makes a cross-family comparison meaningful. The **Discussion** (Section 4, paragraph 2) now leads with the reviewer's reading and, importantly, supplies evidence for it rather than asserting it. Three new results bear on whether the differences are attributable to the protein: the family effect survives adjustment for the ligand chemistry that differs between families (Section 3.11); the target-level index reproduces across *disjoint chemical series* at $p = 0.63$ (Section 3.12), which is difficult to explain if the quantity were mainly a property of the compounds; and target identity accounts for 26.5% of the variance in the log eudysmic ratio. The reviewer's qualification — that ligand chemistry contributes — is retained and quantified rather than merely acknowledged: ligand descriptors alone explain 4.2% of the variance in Δ and predict target family at macro-AUC 0.86, so the ligand contribution is real and large, yet adjusting for it attenuates the protease family coefficient by only 9.7%. The manuscript **title** was also changed to reflect the reframing, from "A Target-Family Map of the Eudysmic Ratio..." to "Stereochemical Discrimination by Protein Target Family...".

1.2 — Protein family is treated as the explanatory variable rather than the underlying pocket properties

> **Comment.** "Protein family is only a surrogate. The true determinants are likely to be: binding pocket geometry, enclosure, hydrogen-bond architecture, residue arrangement, flexibility, electrostatics, conformational dynamics. Future analyses should move beyond family labels and investigate which structural properties actually predict stereochemical discrimination."

Response. The reviewer is correct, and the revision now *quantifies* the point rather than conceding it in prose. Using a random-intercept model (Section 3.12), 26.5% of the variance in the log eudysmic ratio lies between targets, and **family membership explains only 12.3% of that between-target variance**. Approximately seven-eighths of the reproducible protein-level signal is therefore target-specific and not captured by family. This is stated in Section 3.12, in the Discussion (Section 4, paragraph 3) and in the Conclusion.

The structural-descriptor regression the reviewer proposes is **not** attempted here, and the manuscript now says so explicitly and gives reasons (Section 4, paragraph 3). It would require experimental structures for a large fraction of the 224 indexed targets plus a pocket-detection and alignment pipeline, which is a separate study. More substantively, the one published attempt to relate binding-site descriptors to a closely analogous quantity — activity-cliff propensity among kinase inhibitors — reports that pocket-only models performed near chance and concludes that the whole protein matrix rather than the site is controlling (reference 12). That is the same reference the original manuscript cited, and re-reading it during this revision showed that the original text had mis-stated its result (see Correction C5 below). Attempting a shallow version of that analysis here would be more likely to mislead than to inform. The proposal is therefore developed as **Perspective 5.1**, with the specific tools named (references 35–37) and with the honest prior that it may return a negative result — which would itself be informative.

1.3 — Residual medicinal chemistry bias cannot be excluded

> **Comment.** "Although the manuscript controls for: stereocenter count, scaffold clustering, assay type, selection bias, potency, one important possibility remains. Historically, protease inhibitors, kinase inhibitors, GPCR ligands, and nuclear receptor ligands have been developed using very different medicinal chemistry strategies. Consequently, the observed differences may reflect: ligand

rigidity, peptide-like scaffolds, pharmacophore placement, conformational restriction, rather than intrinsic protein properties. Additional ligand descriptors should therefore be incorporated into multivariable analyses."

Response. This was the most serious criticism in the review and it has been addressed with a new analysis, now **Section 3.11**, **Table 5** and **Figure 6**. Methods are in Sections 2.6 and 2.8.

Fourteen 2D descriptors were computed per structural pair — molecular weight, cLogP, TPSA, hydrogen-bond donors and acceptors, rotatable bonds, fraction sp³, aromatic and total ring counts, heavy atoms, **amide-bond count as a peptide-likeness proxy**, QED, Bertz complexity and a normalized **flexibility** index — plus stereocenter count. These map onto the four mechanisms the reviewer names. Computing them once per pair is exact rather than approximate, because the two enantiomers share a constitution and every descriptor is a function of the 2D graph (assumption A7).

The reviewer's concern is empirically real. All fifteen descriptors differ significantly across families (Table 5): nuclear-receptor ligands are heaviest and most lipophilic (MW 509, cLogP 4.89) and carry no amide bond at the median; protease ligands have the lowest fraction of sp³ carbons (0.314) — the flattest by that measure, yet showing the largest Δ ; GPCR ligands are smallest and least polar. Target family can be predicted from ligand descriptors alone at a macro one-vs-rest AUC of **0.862** (Section 3.13). The medicinal chemistry differs enormously between families, exactly as the reviewer says.

The family effect nevertheless survives adjustment for it, almost undiminished. In a median regression of Δ on family plus the eleven descriptors retained after variance-inflation pruning, with inference from a 400-resample bootstrap clustered on Bemis–Murcko scaffold:

Family (vs kinase)	Unadjusted	Adjusted for ligand chemistry	Attenuation	Fully adjusted
Ion channel	+0.010	−0.040 [−0.093, 0.034]	(both ≈ 0)	+0.068
GPCR	+0.200	+0.158 [0.104, 0.221]	21.0%	+0.134
Nuclear receptor	+0.380	+0.293 [0.181, 0.410]	22.9%	+0.349
Protease	+0.590	+0.533 [0.350, 0.694]	9.7%	+0.489

For the tail statistic, ligand-adjusted odds of exceeding 10-fold relative to kinase are 3.00 [2.40, 3.75] for protease and 1.85 [1.42, 2.39] for nuclear receptor.

A non-parametric check was added because a regression can be criticized for functional-form assumptions. Each family was matched 1:1 against kinase pairs on a propensity score built from **ligand descriptors alone**, within a 0.2-SD logit caliper. Matching does not shrink the gaps; for protease it slightly widens them (matched difference 0.64 versus unmatched 0.59, $n = 1,047$, $p = 2.3 \times 10^{-38}$). Ion channel remains null (0.02, $p = 0.33$), consistent with the rest of the analysis.

Finally, a variance partition is reported so the reader can see the effect in proportion: family alone explains 4.6% of the variance in Δ , ligand descriptors alone 4.2%, enantiomer potency alone 7.3%, the full model 17.8%, leaving **82.2% unexplained**. Δ is dominated by pair-specific variation that none of these variables captures, and the manuscript now says so.

I.4 — The manuscript demonstrates P(ER | protein family) but does not investigate P(protein family | ER)

> **Comment.** "The current work establishes that different protein families exhibit different ER distributions. It does not determine whether ER contains enough information to predict target family. A predictive analysis would greatly strengthen the biological relevance."

Response. Added as **Section 3.13** and **Figure 8**, with methods in Section 2.8. The honest answer is that the eudysmic ratio carries very little information about target family, and the revision reports that as a quantitative bound rather than as a failure. Doing so is important, because it prevents the reframing recommended in comment I.1 from being read as a stronger claim than the data support.

- **Information.** The entropy of the family label is 2.002 bits. The mutual information between Δ and family is **0.057 bits — 2.9% of that entropy** — against a permutation null of 0.003 (95th percentile 0.013, $p < 0.005$). The association is statistically real and practically small.

- **Classification.** Under 5-fold cross-validation grouped by Bemis–Murcko scaffold (so that no chemical series appears in both training and test folds — ungrouped splits are known to overstate performance, references 33 and 34), a random forest on Δ alone reaches balanced accuracy 0.226 and macro one-vs-rest AUC 0.571, against chance values of 0.200 and 0.500. It never predicts two of the five classes.

- **Incremental value.** Given ligand descriptors, which alone reach macro-AUC 0.862, adding Δ improves the macro AUC by 0.004. Δ adds essentially nothing to what the ligand chemistry already reveals about which family a compound was tested against.

- **Bayesian form**, which is the most interpretable and is the one now emphasized: among pairs with $\Delta > 2$ (a greater-than-100-fold enantiomeric difference), $P(\text{protease})$ rises from a prior of 0.085 to **0.238**, a likelihood ratio of 2.8, while $P(\text{kinase})$ falls from 0.397 to 0.248. A very large eudysmic ratio roughly triples the odds that the target is a protease, but 76% of such pairs still belong to other families.

The manuscript states explicitly (Section 2.8) that no claim is made that any model recovers chemical knowledge; the models are used only as estimators of information content.

I.5 — The study stops after identifying an association

> **Comment.** "The manuscript demonstrates that the association exists but does not attempt to explain its origin. A stronger contribution would begin identifying the structural, evolutionary, or mechanistic basis underlying the observed hierarchy."

Response. The revision goes as far towards origin as the data honestly allow, and marks the boundary clearly. Within the data: the association is shown not to originate in ligand chemistry (Section 3.11), not in stereocenter count (Section 3.4), not in assay type (Section 3.6), not in enantiomer potency (Section 3.8), and not in selection bias (Section 3.9); and stereochemical discrimination is shown to be a reproducible property of individual proteins (Section 3.12), which locates the origin on the protein side without identifying which protein property is responsible. Beyond the data, Section 5 sets out six mechanistic, structural, evolutionary and pharmacological hypotheses with the evidence each would require. The author has deliberately not speculated about origin inside the Results or Discussion, since the reviewer's own framing (Section III) rightly separates hypotheses from conclusions.

II.1 — Develop a quantitative measure of protein stereochemical discrimination

> **Comment.** "Rather than focusing solely on ligand eudysmic ratios, introduce the broader concept that proteins possess differing capacities to distinguish stereochemical information. For example, define a protein-specific descriptor such as: Protein Stereochemical Discrimination (PSD) or Stereochemical Discrimination Index (SDI) computed from the distribution of ER values associated with a given protein. This reframes the work from ligand statistics toward protein characterization."

Response. The substance of this recommendation is adopted in full, in **Section 2.7**, **Section 3.12** and **Figure 7**. The descriptor is defined exactly as the reviewer proposes — the median of Δ over the enantiomer pairs assayed against a given target, with a bootstrap confidence interval — and is reported for **224 targets with at least 20 pairs, covering 83.1% of all family-assigned pairs**.

On two points the author has departed from the letter of the recommendation, and asks the reviewer's indulgence in explaining why, since both departures are made in the interest of the paper's standing.

First, on prior art. A literature check made in response to this comment established that a target-level measure of stereochemical discrimination is not a new concept, and claiming it as one would expose the paper to a serious objection. Lehmann's **eudismic affinity quotient** — the slope of eudismic index against eutomer affinity across a ligand series for a single target (references 3 and 4) — is already a per-protein number quantifying stereochemical discrimination. In enzymology, the **enantiomeric ratio *E** of Chen and co-workers (reference 8) is the canonical kinetic measure of an enzyme's enantioselectivity and is directly adjacent terminology. Eudismic indices had already been compared between two targets, $\beta 1$ versus $\beta 2$ adrenoceptors, in 1988 (reference 9). The manuscript now cites all of these (Section 1, paragraph 2; Section 2.7) and positions the contribution as the estimator and the scale, not the definition*: what is new is computing such a quantity for hundreds of targets at once from an archive, under a same-assay measurement control, and reporting its distribution across pocket classes.

Second, on the name. The author has not adopted either proposed acronym, and calls the quantity the **target-level eudismic index**. In this literature PSD denotes postsynaptic density (PSD-95 is a canonical protein name) and post-source decay in MALDI mass spectrometry; SDI denotes the standard deviation index in clinical-chemistry quality control, among other uses. A new abbreviation would collide with established usage in the exact audience for this paper, and a descriptive name also makes the relationship to Lehmann's per-compound index self-evident. The reasoning is stated in one sentence in Section 2.7 so that a reader who expected an acronym understands the choice.

Most importantly, the descriptor is not merely defined — it is validated, because a descriptor is only worth introducing if it is reproducible. Three tests are reported.

- **Split-half reliability.** Partitioning each target's pairs at random into halves and correlating across targets gives Spearman $\rho = 0.755$ [0.711, 0.801], correcting to 0.86 at full length.

- **Scaffold-disjoint split-half reliability**, the demanding version and the one that speaks directly to comment I.3. Partitioning each target's pairs by *chemical series*, so the two halves share no Bemis–Murcko scaffold, gives $\rho = 0.633$ [0.584, 0.688], correcting to 0.775 (Figure 7b). An index estimated from one set of chemical series predicts the index estimated from a disjoint set of series against the same protein. This is the operational meaning of "a property of the protein rather than of the compounds tested against it", and it is the single strongest piece of evidence in the revision for the reviewer's reframing in comment I.1.

- **Artifact checks.** Across the 224 targets the index is uncorrelated with the number of pairs ($\rho = 0.087$, $p = 0.19$) and with the number of distinct scaffolds ($\rho = 0.095$, $p = 0.16$), so it is not an artifact of data density. It correlates weakly with median eutomer potency ($\rho = 0.218$, $p = 0.001$), the target-level form of Pfeiffer's rule.

The index also produced a chemically interpretable ranking that is reported in Section 3.12: at the top, mitogen-activated protein kinase 6 (2.05), coagulation factor XI (1.85), P2X purinoceptor 7 (1.84), NMDA receptor subunit 2B (1.79), BACE1 (1.75) and BACE2 (1.61) — note that the single highest-index target is a kinase, which cuts against the family-level reading and is stated as such; at the bottom, several kinases and a deubiquitinase with an index of 0.00 across dozens of pairs.

Finally, the index revealed a limitation of the original family analysis that the author would not otherwise have found, and which is now disclosed (Section 3.12, Limitation iii): when each *target* counts once rather than each pair, protease falls to 0.657, below nuclear receptor at 0.78, because the pair-weighted protease median is carried by a few intensively studied and highly discriminating targets such as BACE1 and factor XI. Protease's position at the top of the pair-weighted ordering is a statement about the protease pairs in ChEMBL, not about the typical protease. The author is grateful to the reviewer, since this recommendation is what surfaced the issue.

II.2 — Relate stereochemical discrimination to structural biology

> **Comment.** "Instead of comparing only protein families, quantify structural characteristics of binding pockets. Potential variables include: pocket depth, pocket volume, buriedness, hydrogen-

bond density, hydrophobicity, contact density, solvent accessibility, flexibility. Regression against these descriptors could identify the physical determinants of stereochemical discrimination."

Response. This recommendation is **not implemented**, and the manuscript now states the reasons rather than passing over them. They are given in Section 4 (paragraph 3) and developed as **Perspective 5.1**.

The analysis requires experimental structures for a large fraction of the 224 indexed targets, pocket detection and cross-target pocket alignment, and a defensible treatment of the many targets with multiple structures in different conformational states. That is a study in itself rather than an addition to this one, and the author judged that a shallow version — for instance, descriptors from a single arbitrarily chosen PDB entry per target — would produce a regression whose failure or success could not be interpreted.

There is also a substantive reason for caution, which is now cited in the manuscript. Reference 12, the closest published attempt, related protein descriptors to activity-cliff propensity among kinase inhibitors and reports that models built on **binding-site descriptors alone performed near chance**, concluding that the entire protein matrix rather than the site is the controlling factor. Perspective 5.1 therefore names the specific tools that would be used (references 35–37), states the prior honestly, and notes that a negative result would itself be informative by pointing towards whole-protein or dynamic determinants. The manuscript's own variance decomposition supports pursuing this: 26.5% of the variance in Δ lies between targets and only 12.3% of that is explained by family, so there is real target-level signal for structural descriptors to explain.

II.3 — Incorporate ligand complexity into multivariable analyses

> **Comment.** "Calculate ligand descriptors such as: molecular weight, LogP, TPSA, rotatable bonds, stereocenter count, conformational flexibility, rigidity, pharmacophore complexity. Then evaluate whether protein family remains independently associated with ER after adjustment. This would substantially strengthen the biological interpretation."

Response. Implemented in full; see the response to comment I.3 above, which this recommendation restates in operational form. Every descriptor the reviewer names is included: molecular weight, cLogP, TPSA, rotatable bonds, stereocenter count (from the InChI ``t`` layer), conformational flexibility (rotatable bonds per heavy atom), rigidity (fraction sp³ and ring counts), and pharmacophore complexity (Bertz complexity, hydrogen-bond donors and acceptors, amide-bond count, QED). Family remains independently associated with Δ after adjustment: the protease coefficient attenuates by 9.7% and retains a 95% interval of [0.350, 0.694], and propensity matching on ligand descriptors alone does not shrink any gap. See Section 3.11, Table 5, Figure 6 and Sections 2.6 and 2.8.

II.4 — Evaluate predictive utility

> **Comment.** "Determine whether ER carries predictive information regarding target family. Possible analyses include: multinomial logistic regression, Bayesian classification, random forest classification, ROC analysis, mutual information, entropy reduction. This distinguishes a descriptive observation from one with practical predictive value."

Response. Implemented in full; see the response to comment I.4 above. All six of the suggested analyses are reported in Section 3.13 and Figure 8: multinomial logistic regression and random forest classification, ROC analysis (macro and per-family one-vs-rest AUC), mutual information with a permutation null, entropy reduction expressed as a percentage of the 2.002-bit family entropy, and the Bayesian form as posterior probabilities and likelihood ratios by Δ band. The answer is that the observation is descriptive rather than practically predictive: Δ reduces uncertainty about the target family by 2.9%, and the strongest usable statement is that a >100-fold enantiomeric difference roughly triples the odds that the target is a protease while still leaving 76% of such pairs in other families.

II.5 — Investigate evolutionary explanations

> **Comment.** "The observed hierarchy may reflect evolutionary pressures. Examples include: proteases requiring high substrate specificity, kinases evolving around ATP recognition, GPCRs balancing specificity with ligand diversity. Developing this discussion would substantially deepen the biological significance."

Response. Developed as **Perspective 5.3**, using the reviewer's own examples and anchoring them in primary literature. Protease recognition through stereochemically defined subsites is cited to Schechter and Berger (reference 38), and the sharpest available statement of protease stereospecificity — the chemically synthesized D-enzyme of HIV-1 protease showing reciprocal chiral substrate specificity — to Milton and co-workers (reference 39). The conserved kinase ATP site and its hinge-binding requirement, satisfied by largely planar heteroaromatic scaffolds, is cited to Zhang, Yang and Gray (reference 40).

The hypothesis is stated as such: that a family's typical discrimination reflects the stereochemical strictness of its natural ligand recognition. Two concrete tests are proposed — relating the target-level index to independently measured substrate stereospecificity, or to residue conservation in the site. The material is placed in Perspectives rather than the Discussion because the present data cannot test it, and the author was concerned that an evolutionary narrative inside the Results would read as an explanation the study had established.

II.6 — Investigate structural binding mechanisms

> **Comment.** "For proteins with available crystal structures: dock both enantiomers, compare binding poses, quantify hydrogen bond differences, evaluate steric clashes, compare interaction fingerprints. Correlating these structural differences with ER would directly test the mechanistic interpretation."

Response. Not implemented, for the reasons given under comment II.2; developed as **Perspective 5.2**. The author agrees that this is the analysis that would test the mechanistic interpretation directly, and Perspective 5.2 sets out the specific and, in the author's view, most productive version of it. Section 3.12 identifies targets whose index is essentially **zero** across dozens of enantiomer pairs — protein kinase C α ($n = 46$), USP28 ($n = 22$), GCN2 ($n = 43$), ATR ($n = 33$) — where the median pair shows no measurable enantiomeric preference at all. Perspective 5.2 notes that such a value has three distinct explanations the present data cannot separate (a genuinely permissive pocket; stereochemistry positioned away from the pharmacophore in the assayed series; or two enantiomers adopting different but energetically comparable poses), and that co-crystallography or molecular dynamics on both enantiomers at these specific targets would distinguish them. Identifying those targets is a concrete contribution of the new descriptor, and it is offered as a starting point for exactly the study the reviewer proposes.

II.7 — Present stereochemical discrimination as a new protein descriptor

> **Comment.** "The manuscript has the potential to introduce stereochemical discrimination as a novel descriptor of protein molecular recognition. Rather than viewing ER merely as a ligand property, propose that the aggregate stereochemical discrimination of a protein represents a measurable characteristic analogous to: catalytic efficiency, binding affinity, thermal stability, evolutionary conservation, structural flexibility. This conceptual reframing could considerably increase the long-term impact of the study."

Response. The reframing is adopted. The manuscript now presents stereochemical discrimination as a measurable characteristic of a protein, and — as the reviewer's analogy to catalytic efficiency and thermal stability implies — treats it as something that must be shown to be *reproducibly measurable* before it can be treated as a characteristic at all. That is why Section 3.12 leads with reliability rather than with the values: scaffold-disjoint split-half reliability of 0.633 (Spearman–Brown 0.775) is what licenses describing it as a property of the protein, and 26.5% of the variance in Δ lying between targets is what gives it magnitude.

On novelty the author has been deliberately conservative, for the reasons set out under comment II.1: the concept is not new (Lehmann's eudismic affinity quotient; the enantiomeric ratio E in

enzymology), and the manuscript says so. The claim made is that the quantity is now *measurable at archive scale, under a noise control, for hundreds of targets*, and that this is what makes it usable as a descriptor in the way the reviewer envisages. The author judged that a claim of conceptual novelty would be both incorrect and, if a reader recognized the antecedents, damaging to the paper's credibility — whereas the estimator-and-scale claim is defensible and is what actually enables the Perspectives.

III — Proposed "Perspectives" section

> **Comment.** "The current manuscript ends with a discussion of limitations and robustness. A dedicated Perspectives section would be an excellent addition because it would clearly distinguish measured results from future hypotheses. The following hypotheses are scientifically motivated extensions of the present work rather than conclusions supported by the current dataset." [Followed by Perspectives 1–6: evolution of drug resistance; druggability; ligand promiscuity; binding-pocket architecture; multiple binding modes; clinical pharmacology.]

Response. A dedicated **Section 5, Perspectives** has been added, and it opens with a sentence adopting the reviewer's own framing: "The results above are measurements. This section sets out hypotheses that they motivate but do not establish, and states what evidence would be needed to test each. None is a conclusion of this study." All six of the reviewer's perspectives are included, reordered so that the most tractable and best-motivated come first, and each names the evidence required rather than only the question:

Reviewer's perspective	Manuscript section	Evidence named
4. Binding-pocket architecture	5.1	pocket descriptors from fpocket / druggability models (35–37); honest prior from reference 12
5. Multiple binding modes	5.2	co-crystallography or MD on both enantiomers at the zero-index targets identified in Section 3.12
— Evolutionary determinants (from II.5)	5.3	protease subsite stereospecificity (38, 39); kinase ATP-site conservation (40)
2. Druggability	5.4	project-level hit rates, optimization trajectories, achievable selectivity
3. Ligand promiscuity	5.5	archive-derived promiscuity measures (41, 42); flagged as the most immediately tractable
1. Evolution of drug resistance / 6. Clinical pharmacology	5.6	mutant-specific potency for both enantiomers (43, 44); flagged as the most speculative

Section 5 closes with a caution that applies to all six, and which the author thinks is necessary if the descriptor is to be used responsibly: the index has a median 95% interval width of 0.39 log units at 20 pairs per target, and family explains only 12.3% of between-target variance, so studies built on it will need targets with substantially more data than the minimum used here and should treat it as a noisy estimate rather than a protein constant.

Reviewer 2

1 — Verify that all links to the references point to the correct source

> **Comment.** "Please verify that all links to the references point to the correct source."

Response. Every reference was independently verified by resolving its DOI and cross-checking the returned metadata (title, full author list, year, volume, issue, pages) against the manuscript entry, using the DOI resolver, the Crossref REST API, Europe PMC and PubMed. This exercise found **two errors in the original list**, both now corrected. The author is grateful for the instruction to check, as neither would have been caught by reading.

Error 1 (serious): the author list of reference 7 was entirely wrong. The original manuscript cited "Safina BS, McKerrall SJ, Sun S, Chen CA, Chowdhury S, Jia Q, et al. Effect of Chirality on Common in Vitro Experiments: An Enantiomeric Pair Analysis. J Med Chem. 2015;58(15):5781–5788." The DOI 10.1021/acs.jmedchem.5b00552 is correct and resolves to a paper with that exact title, volume and pages — but its authors are **Bagdanoff JT, Xu Y, Dollinger G, Martin E**. Not one of the six cited names appears on the paper, and the paper has four authors, so the "et al." was also wrong; the cited names belong to a different J Med Chem paper. This is now **reference 15** with the correct author list.

Error 2: the primary source for the eudismic index was incorrect. The original manuscript attributed the eudismic index to "Ariëns and co-workers" citing Ariëns EJ, Eur J Clin Pharmacol 1984;26(6):663–668. That DOI resolves correctly, but the paper is a single-author commentary that does not introduce the index — its own reference list cites the earlier work for it. The correct primary source is **Lehmann F PA, Rodrigues de Miranda JF, Ariëns EJ, Progress in Drug Research 1976;20:101–142** (now reference 2), with Lehmann's quantitative development (references 3 and 4). The 1984 paper is retained as reference 4's companion context only where it supports what it actually argues. The phrase "Ariëns and co-workers" is now literally accurate for reference 2, which is a three-author work.

Two further citation-accuracy problems were found and corrected:

- **Pfeiffer's rule was attributed to its critic.** The original cited Barlow 1990 as the source of the rule. Barlow's paper is a *critique* ("there are marked exceptions"). The rule's primary source, **Pfeiffer CC, Science 1956;124(3210):29–31**, is now reference 5, and Barlow is cited as reference 6 for the critique. This is an improvement rather than only a correction: Barlow's objection — that the rule neglects molecular flexibility and admits marked exceptions — directly anticipates the manuscript's own Section 3.8 finding that the most potent family is the least stereoselective, and the text now says so.

- **The RDKit citation form** was changed to the project's own recommended citation, which does not attribute authorship to an individual.

A full list of the checks is available on request; all 44 references in the revised manuscript were verified, and every in-text citation number resolves to an entry in the list with no orphans.

2 — Disclose and acknowledge any assistance received

> **Comment.** "The authors must disclose and acknowledge any assistance received in the preparation of this manuscript, including but not limited to editorial, technical, analytical, or writing support. All such contributions must be clearly stated in the Acknowledgments section."

Response. An **Acknowledgments** section has been added, and it discloses AI assistance explicitly and without minimizing it. The disclosure states that an AI assistant (Claude, Anthropic) was used for solely technical and analytical help. It helped generate figures, organize the data / scripts, and aid with the computation. The author conceived the study, specified the analyses, verified the outputs against the source data, and is responsible for the content including all interpretations and conclusions. It also states that no other editorial, technical, analytical or writing support was received and that no funding was received. The reviewers' contribution to the revision is acknowledged in the same section.

3 — Include enough recent references along with foundational ones

> **Comment.** "Include enough recent references along with foundational ones. Ensure references directly support your claims. Avoid 'padding.' Use credible sources (peer-reviewed journals, reputable books, official reports)."

Response. The reference list has been expanded from 13 to 44 entries. The expansion is driven by new content rather than by padding: the revision adds four new analyses, a Perspectives section and an explicit assumptions section, each of which requires its own support. Every entry is cited at a specific claim; no entry is cited only in a list; and there are no orphan references.

Recent primary literature added (2020–2024): McVicker and O'Boyle 2024 on the chirality of new drug approvals (7), which now supplies the quantitative motivation in the opening paragraph; Saha and co-workers 2020 on chirality within the human kinome (11); Comajuncosa-Creus and co-workers 2024 on stereochemically aware bioactivity descriptors (14), the closest recent large-scale work; Landrum and Riniker 2024 re-confirming that combining potency values from different sources is a significant source of noise (17), which strengthens the measurement rationale; van Tilborg and co-workers 2022 on the limitations of molecular machine learning at activity cliffs (34), supporting the choice of grouped cross-validation.

Foundational sources added where the revision made specific claims that needed them: Lehmann and co-workers 1976, 1982, 1986 (2–4) and Pfeiffer 1956 (5) for the eudismic index and Pfeiffer's rule; Chen and co-workers 1982 (8) for the enantiomeric ratio *E*; El Tayar and co-workers 1988 (9) for the two-target precedent; Schechter and Berger 1967 (38) and Milton and co-workers 1992 (39) for protease stereospecificity.

Methodological sources added for the new statistics, each named in Section 2.8: Koenker and Bassett (29) for quantile regression, Cameron and Miller (30) for cluster-robust inference, Laird and Ware (22) and Nakagawa and co-workers (23) for mixed models and the intraclass correlation, Ross (31) and Kraskov and co-workers (32) for mutual-information estimation, Holm (24) and Wilson (25) for the multiple-comparison correction and the proportion interval named in the text, and statsmodels (27) and scikit-learn (28) alongside SciPy (26) for the software actually used. The original manuscript attributed all statistics to SciPy, which was an over-attribution: the Wilson interval and the mixed models are not SciPy functions.

Perspectives sources (35–37, 41–44) are cited to the specific hypothesis each supports, as tabulated under comment III above.

4 — All assumptions must be explicitly stated and justified

> **Comment.** "All assumptions (including implicit assumptions) must be explicitly stated and clearly justified. The authors should explain why each assumption is reasonable and discuss its impact on the results and conclusions."

Response. A new **Section 2.10, "Explicit assumptions"**, states ten assumptions (A1–A10), each with its justification and its expected impact if violated. They cover: the InChI mirror-flag rule (A1); the meaning of a shared assay identifier, which is the design's central assumption (A2); pChEMBL comparability within an assay (A3); median aggregation (A4); the use of the absolute value, which discards enantiomer identity (A5); the accuracy and independence of the ChEMBL protein classification (A6); the applicability of 2D descriptors to an enantiomer pair (A7); exchangeability of pairs after scaffold collapse (A8); the missingness mechanism for censored measurements (A9); and the interpretation of absolute rates (A10).

Working through this list found a genuinely unstated and, as it turns out, **false** implicit assumption, which the author considers the most valuable single outcome of this review. It is now assumption **A9** and is quantified in a new **Section 3.14**.

The design requires both enantiomers to carry a pChEMBL value, and ChEMBL assigns one only to exact ("=") measurements. Pairs in which the weaker enantiomer was reported only as a bound — "IC₅₀ > 10,000 nM" — are therefore silently excluded. The original manuscript presented the 100% exact-relation composition of the data as a strength, without noting that it is also a truncation. It is not missing at random: the weaker enantiomer is reported as a bound precisely when it is too weak to quantify, which is when Δ is largest.

Searching the same assays for such pairs found 780 of them against 12,335 fully exact pairs (5.95%). Their implied **lower bounds** on Δ have a median of 1.22, against 0.50 for the exact pairs — the excluded pairs are indeed the highly stereoselective ones. Reinstating them at their lower bounds, which can only understate their true values, raises the overall median from 0.50 to 0.53 and the fraction exceeding 100-fold from 7.8% to 8.6%.

Critically, the censoring rate **differs by family** ($\chi^2 = 66.5$, $p = 1.3 \times 10^{-13}$) and is highest for protease (10.2%), lowest for kinase (4.5%). The family most affected by the truncation is the one with the largest observed Δ , so correcting the bias **widens** the protease–kinase gap rather than closing it. Every value reported in the manuscript is therefore a conservative estimate. This is now stated in the Abstract, in Section 3.14 and in the Conclusion.

5 — Avoid overreaching conclusions

> **Comment.** "Avoid overreaching conclusions that extend beyond what is supported by the data and analysis."

Response. Taken seriously, and acted on beyond copy-editing. In addition to keeping the association-not-causation framing throughout, the author re-audited every quantitative and inferential claim in the submitted manuscript against the underlying result files. That audit found several statements stronger than the numbers supported. They are corrected in the revision and listed in full in the section "Corrections to the submitted manuscript" below, because the author believes the editor and reviewers should see them stated plainly rather than encounter them as silent changes.

The most consequential are: the description of the family structure as "three tiers" with nuclear receptor " $\approx$ " protease, which the manuscript's own pairwise test contradicted; the claim that the tier grouping was invariant to the detection implementation, which the cross-check data did not support for ion channel; the figure-caption statement that sub-threshold values are "resolved, not noise", which over-stated what removing the between-source noise component establishes; and the description of the 0.68-log figure as a variance when it is a standard deviation. New statistical machinery was added where the original inference was unsound: family pairs are now compared by a bootstrap confidence interval on the **difference** between medians against a pre-stated ± 0.10 -log equivalence margin, rather than by inspecting whether two per-family intervals overlap or by treating a non-significant test as evidence of equivalence (Section 2.8).

The new sections were written under the same discipline. Section 3.13 exists specifically to bound the claim, reporting that the ratio explains 2.9% of the family entropy. Section 3.12 reports that family explains only 12.3% of between-target variance. Section 3.11 reports that 82.2% of the variance in Δ is unexplained by any variable examined. Section 5 is explicitly labelled as hypotheses.

6 — Verify past perfect tense and third person throughout

> **Comment.** "Verify that you have used past perfect tense and third person throughout the manuscript wherever applicable."

Response. The manuscript has been converted throughout. All 44 first-person constructions in the submitted version ("we extract", "our data", "we therefore detect") have been removed; the revised manuscript contains no first-person pronouns, and where agency must be named it uses "the author".

On tense, the author has followed the reviewer's qualification "wherever applicable". Work that was performed and results that were obtained are now in the past simple, generally in the passive ("pairs were extracted", "the ordering was preserved", "descriptors were computed"). Present tense is retained only where it is correct in scientific English — for definitions ("Two molecules are true enantiomers if..."), established facts ("pChEMBL is $-\log_{10}$ of molar potency"), and statements about what the paper itself contains ("Section 5 sets out hypotheses"). **Past perfect** is used where the sequence of events genuinely requires it — that is, for events prior to other past events: the detector "**had been validated** against eight ground-truth cases before any counting was performed"

(Section 2.2); the two detection routines **"had been written"** independently and were compared" (Section 2.2); within an assay the enantiomers **"had been measured"** under identical conditions" (Sections 1 and 2.3); and both enantiomers are made and tested together mainly when a chemist **"had already suspected"** that chirality mattered" (Section 3.9). The author has not applied past perfect uniformly, since a blanket past perfect would be ungrammatical in English for simple statements of what was done, and understands the comment to ask for it where applicable.

Corrections to the submitted manuscript

Prompted by Reviewer 2's comments 4 and 5, the author re-audited the submitted manuscript against its own result files. The following statements were stronger than the underlying numbers supported, or were incorrect. All are corrected in the revision. They are listed here in the interest of full disclosure.

C1. "Three tiers" and "nuclear receptor $\approx$ protease". The submitted abstract and Section 3.3 described the ordering as three tiers with nuclear receptor approximately equal to protease. The manuscript's own primary pairwise test separated those two families (Holm-corrected $p = 8.7 \times 10^{-4}$) and their bootstrap intervals, [0.68–0.82] and [0.86–1.10], do not overlap. The merge relied on a non-significant result in the lower-powered scaffold-collapsed analysis, which is affirming the null. *Correction:* the ordering is now described as monotone. Family pairs are compared by a bootstrap interval on the difference in medians against a pre-stated ± 0.10 -log equivalence margin. Kinase and ion channel are **equivalent** (–0.010 [–0.04, 0.03]); nuclear receptor versus protease is **separated** per-pair (–0.21 [–0.36, –0.09]) and **indeterminate** after scaffold collapse (–0.125 [–0.255, 0.032]) — neither equivalent nor separated. Sections 2.8, 3.3, 3.5.

C2. Cross-validation figures were taken from the wrong stratum. Section 3.10 stated that the RDKit routine "applied to single-protein targets, yields 44,675 pair instances with overall median $\Delta = 0.425$ ". Those are the **all-targets** figures; the single-protein figures are 33,148 instances and median 0.450. *Correction:* both strata are now reported. Section 3.10.

C3. "Medians within 0.01–0.1 log" and "invariant to the implementation". The per-family medians of the two routines agree within 0.02–0.04 log for four families but differ by **0.15** for ion channel (0.55 versus 0.40), where the cross-check places ion channel level with GPCR (0.56) rather than with kinase. The claim of implementation-invariance was therefore not supported at the low end. *Correction:* the discrepancy is reported explicitly, the instability of ion channel's position is now a stated limitation (Limitation iv), and the same instability is independently confirmed by the target-weighted analysis in Section 3.12. It is also now noted that only the stereo-detection step differs between the two routines, so this is a check on enantiomer detection, not an independent replication of the family result. Sections 3.10, 3.12.

C4. A cited p-value was not in the file cited. Section 3.4 attributed " $p = 2.1 \times 10^{-27}$ " for the two-stereocenter stratum to `results/stereocenter\_control.json`, which contains no two-stereocenter result. *Correction:* the stratum was recomputed and persisted (`results/tier\_structure.json`); the p-value is confirmed as 2.14×10^{-27} . However, recomputation also showed that the claim "the same holds" was wrong about the **ordering**: at two stereocenters, ion channel falls to 0.34, below kinase at 0.55, and GPCR (0.78) exceeds nuclear receptor (0.675). What is stable across strata is that protease is highest and that kinase and ion channel are in the lower half. Section 3.4.

C5. Reference 12's result was inverted. The submitted text stated that "protein and binding-site properties are associated with activity-cliff propensity" in that study. The study reports that binding-site-only models performed near chance and concludes that the whole protein matrix, not the site, is controlling. *Correction:* the claim now reflects the source, and the corrected reading is used to justify the decision not to attempt a pocket-descriptor regression here (Sections 1 and 4).

C6. 0.68 log units is a standard deviation, not a variance, and it is the total dispersion of pooled public data rather than an increment added by mixing; the same source's same-assay estimate is $\sigma \approx 0.17$ –0.22. *Correction:* the wording is corrected in the Abstract, Introduction, Methods, Discussion

and Conclusion, and the design is now described as removing the **between-source component** of that dispersion.

C7. Figure 1's annotation over-claimed. The caption and the in-figure text stated that values below the 0.68-log line are "resolved, not noise". Removing the between-source component does not establish that any individual sub-threshold value is resolved, and the manuscript's own limitations said intra-assay error was never quantified. *Correction:* Figure 1 was regenerated with corrected in-figure text, and the caption and Section 3.2 were rewritten to match.

C8. Nuclear receptor was described as "less secure" than protease. On the metrics actually computed the two are comparable: the maximum leave-one-target-out shift is 0.157 for nuclear receptor and 0.16 for protease, and protease's per-target medians are far more dispersed (interquartile width 1.10 versus 0.39). *Correction:* both high-ranked families are now flagged, for different reasons — nuclear receptor because one target supplies 46.7% of its pairs, protease because its targets disagree with one another and its median is carried by BACE1 and factor XI. Sections 3.7, 3.12, Limitation iii.

C9. The large-panel selection-bias result was over-stated. Section 3.9 stated that the large-panel protease median "does not rest on a single incidental screen" on the grounds that it spans 13 targets. In fact **55.8%** of protease measurements in that stratum come from BACE1 and 54.5% from three assays; 76.6% of the nuclear-receptor measurements come from ROR- γ . *Correction:* the concentration is now reported, and the claim is narrowed to what it supports. The underlying finding survives and is now measured with a sample-size-independent statistic rather than an omnibus p-value that Section 2.6 had itself disqualified: the protease-versus-kinase rank-biserial correlation rises to 0.61 [0.53, 0.69] in large panels. It is also now stated that only a two-level pattern, not the five-way ordering, survives in that stratum. Sections 3.9, Table 4, Limitation ix.

C10. "The effect is a binding-assay phenomenon" was near-tautological, since binding assays supply 83–99.8% of pairs and the functional-only complement was never analysed. *Correction:* the complement is now analysed. It is too thin in four families (9–64 pairs) to support a comparison; for GPCR ($n = 680$) the functional-only median is 0.61 against 0.59 for binding. The conclusion is narrowed accordingly. Section 3.6.

C11. Figure files carried retracted language. Figures 2 and 5 were regenerated: Figure 2's title read "three tiers" and carried low/high tier brackets, and Figure 5's title read "Family tiers persist in large screening panels — not a cherry-picking artifact", both asserting framings withdrawn under C1 and C9. Figure 5 now also annotates the BACE1 and ROR- γ concentration of its large-panel cells.

C12. Smaller numerical corrections. " $n \approx 18,000$ " for the family omnibus test is 12,344; "the most-represented target contributes $\leq 12\%$ " is 12.4%; "within an explained 5–20% on the counts" was imprecise; the 7.84% pair-level and $\approx 10\%$ molecule-level unspecified-stereo rates were described as "the same fraction" over different denominators; Section 3.4 was retitled "Not a stereocenter-count artifact", since stereocenter count is one facet of molecular complexity and the broader control now lives in Section 3.11; a dangling cross-reference to a non-existent "Limitations" heading was fixed; equation (eq3) is now set as a display equation like (eq1) and (eq2); the binding-only restriction removes about 7% of the family-assigned pairs, not "about 5%" (which used the wrong denominator); and the released pipeline now assigns pair identifiers with a deterministic hash, since Python's built-in hash is randomised per process and the previous identifiers therefore differed between runs.

Thank you for addressing my comments. I have some minor errors and discrepancies which need addressing, after which the manuscript can be accepted.

Table 5 descriptor count

The text says Table 5 presents all 15 descriptors, but the table displays only seven. This must be reconciled.

Probability versus odds in Section 3.13

($P(\text{protease})$) increases from 0.085 to 0.238:

probability enrichment: $(0.238/0.085 = 2.8)$ -fold;

odds increase: $((0.238/0.762)/(0.085/0.915) \approx 3.36)$ -fold.

Therefore, “likelihood ratio of 2.8” and “roughly tripling the odds” conflate different quantities.

Abstract versus robustness results

The Abstract says every adjacent family contrast except kinase–ion channel was separated. That is true for the primary pair-weighted analysis, but nuclear receptor versus protease becomes indeterminate after scaffold collapse, and their ordering reverses when targets are weighted equally. The Discussion acknowledges this, but the Abstract does not.

Protein-property claim versus limitations

The Abstract and Conclusion say the index “behaved as a reproducible property of the protein.” Elsewhere, the manuscript correctly acknowledges that ligand selection, stereocenter position, assays, and target-specific research programs remain uncontrolled. “Reproducible target-associated characteristic” would be internally better aligned with those limitations.

Censoring claim is slightly broader than the analysis

Section 3.14 says the reported values are conservative estimates of “underlying stereoselectivity.” The analysis establishes this direction for the particular exact-versus-greater-than censored pairs that were recovered—not necessarily for every excluded or missing pair. The Abstract and Conclusion repeat the broader wording.

“Variance explained” is undefined

The manuscript reports 4.6%, 4.2%, 7.3%, and 17.8% variance explained but does not clearly identify the common (R^2) or other statistic from which these values were calculated. This is an internal methodological omission, particularly because the surrounding analysis uses median regression.

Response to Reviewers — Round 2

Dylan Ashraf

Manuscript: "How Much Does Handedness Matter? Stereochemical Discrimination by Protein Target Family, Measured from Same-Assay Enantiomer Pairs in ChEMBL"
dylanashraf56014@gmail.com

Summary of the revision

The author thanks Reviewer 1 for a careful second reading. All six points were checked against the underlying result files before any text was changed, and all six were correct. Each has been addressed in full. Every comment is reproduced verbatim below, followed by the response and the exact location of the change.

Two of the six were substantive rather than editorial. Comment 2 identified a genuine statistical conflation — a probability enrichment reported as a likelihood ratio and then described in the language of odds — which has been corrected in the manuscript, in the analysis script, and in the released results file, so that the mislabeled quantity does not survive anywhere in the record.

Comment 6 identified an undefined statistic, which has been defined in Methods and flagged at the point of use. The remaining four tighten wording so that the Abstract and Conclusion do not claim more than the Results section already concedes.

Changes in this round:

- **Table 5** — rebuilt to show all fifteen descriptors
- **Section 2.8** — definition of the variance-explained statistic (new sentence)
- **Section 3.11** — variance shares identified as ordinary-least-squares R^2

- **Section 3.12** — what reliability across chemical series does and does not establish
- **Section 3.13** — probability enrichment and odds ratio separated and both reported
- **Section 3.14** — censoring conclusion restricted to the recoverable pairs
- **Abstract** — weighting-dependence of the nuclear-receptor/protease contrast added; "property of the protein" and the censoring claim narrowed
- **Section 6 (Conclusion)** — same two narrowings
- ``src/15_predictive.py`, `results/predictive.json`` — mislabeled field corrected

Reviewer 1

> **Comment.** "Thank you for addressing my comments. I have some minor errors and discrepancies which need addressing, after which the manuscript can be accepted."

Response. The author is grateful for the reviewer's engagement across both rounds; the manuscript is materially better for it. All six items are addressed below.

R2.1 — Table 5 descriptor count

> **Comment.** "Table 5 descriptor count. The text says Table 5 presents all 15 descriptors, but the table displays only seven. This must be reconciled."

Response. Correct, and the discrepancy was in the table, not the text: fifteen descriptors were tested, and the submitted table showed only seven of them. **Table 5 has been rebuilt to display all fifteen**, transposed so that descriptors are rows and families are columns, which keeps the table readable at this width.

The table now also carries the Kruskal–Wallis p -value for each descriptor rather than the single blanket statement " $p < 0.05$ ", so the claim that every descriptor differs across families can be checked line by line. The full set is: molecular weight, heavy atoms, cLogP, TPSA, hydrogen-bond donors, hydrogen-bond acceptors, rotatable bonds, flexibility (rotatable bonds per heavy atom), fraction sp^3 carbon, rings, aromatic rings, amide bonds, Bertz complexity, QED, and defined stereocenters.

One clarification has been added to the caption to forestall a related question: three of the fifteen — heavy atoms, Bertz complexity and rotatable bonds — were dropped from the *adjusted regressions* for collinearity (variance inflation factor > 10), but all fifteen are tested and reported in Table 5. The count of fifteen therefore refers to descriptors examined, while eleven plus stereocenter count enter the regression models of Section 3.11.

Location: Table 5 and its caption.

R2.2 — Probability versus odds in Section 3.13

> **Comment.** "Probability versus odds in Section 3.13. ($P(\text{protease})$) increases from 0.085 to 0.238: probability enrichment: $(0.238/0.085 = 2.8)$ -fold; odds increase: $((0.238/0.762)/(0.085/0.915) \approx 3.36)$ -fold. Therefore, 'likelihood ratio of 2.8' and 'roughly tripling the odds' conflate different quantities."

Response. The reviewer is right, and the error was twofold. The value 2.8 is a probability enrichment, not a likelihood ratio; and the phrase "roughly tripling the odds" describes the odds ratio of 3.37, which was not the number reported. The reviewer's arithmetic reproduces exactly against the stored posterior and prior: probability enrichment $0.2378/0.0848 = 2.804$, odds ratio $(0.2378/0.7622)/(0.0848/0.9152) = 3.367$.

The sentence now reports both quantities, names each correctly, and states explicitly that they are distinct. It also notes that the odds ratio is the Bayes factor, which is the quantity a reader who wants a likelihood ratio is actually looking for:

> "...the probability that the target was a protease rose from a prior of 0.085 to 0.238, while the probability that it was a kinase fell from 0.397 to 0.248. Two distinct quantities describe that shift and should not be conflated: the **probability enrichment** is $0.238/0.085 = 2.8$ -fold, whereas the **odds ratio** — equivalently the Bayes factor, $(0.238/0.762)/(0.085/0.915)$ — is 3.4. A very large

eutismic ratio therefore does shift belief about the target class, multiplying the prior probability of a protease by 2.8 and its prior odds by 3.4, but it leaves 76% of such pairs belonging to other families."

Because the mislabeling originated upstream of the prose, it has also been corrected at source. The field `'likelihood_ratio_vs_prior'` in `'results/predictive.json'` was a misnomer for posterior/prior; it has been renamed `'probability_enrichment_vs_prior'`, and a correctly computed `'odds_ratio_vs_prior'` has been added alongside it, for every Δ band and every family. The generating script `'src/15_predictive.py'` was updated to emit both, with a comment recording that the two differ and why. The released results and the manuscript therefore no longer disagree with each other, and no reader re-running the pipeline will inherit the original mislabel.

Location: Section 3.13, paragraph 3; `'src/15_predictive.py'`; `'results/predictive.json'`.

R2.3 — Abstract versus robustness results

> **Comment.** "Abstract versus robustness results. The Abstract says every adjacent family contrast except kinase–ion channel was separated. That is true for the primary pair-weighted analysis, but nuclear receptor versus protease becomes indeterminate after scaffold collapse, and their ordering reverses when targets are weighted equally. The Discussion acknowledges this, but the Abstract does not."

Response. Accepted without reservation. The Abstract stated the pair-weighted result as though it were unconditional, while Sections 3.3, 3.5 and 3.12 all record that the top of the ordering is weighting-dependent. An abstract should not be less careful than the body it summarizes. The following has been inserted immediately after the sentence in question:

> "That separation is weighting-dependent at the top of the ordering: nuclear receptor versus protease became indeterminate after scaffold collapse and reversed when targets rather than pairs were weighted equally, so what is stable is the elevation of protease and nuclear receptor above kinase and ion channel, not their order relative to each other."

This states the limitation in the same sentence that makes the claim, rather than deferring it, and it names both failure modes the reviewer identified — the scaffold-collapse indeterminacy and the reversal under equal target weighting.

Location: Abstract, immediately following the family-ordering sentence.

R2.4 — Protein-property claim versus limitations

> **Comment.** "Protein-property claim versus limitations. The Abstract and Conclusion say the index 'behaved as a reproducible property of the protein.' Elsewhere, the manuscript correctly acknowledges that ligand selection, stereocenter position, assays, and target-specific research programs remain uncontrolled. 'Reproducible target-associated characteristic' would be internally better aligned with those limitations."

Response. Adopted, including the reviewer's proposed wording. The distinction matters: scaffold-disjoint reliability shows that the index is not an artifact of any particular chemical series, but a target's assays, its stereocenter geometry and the research programme that produced its compounds are shared across *all* of that target's series, so they are not separated by the reliability test. Calling the result a protein property claimed more than the design can deliver.

The **Abstract** now reads: "...so it behaved as a reproducible target-associated characteristic rather than an artefact of the particular compounds tested against it — not an intrinsic protein property, since stereocenter position, assay design and the research programme behind each target's compound set remain uncontrolled."

The **Conclusion** now reads: "...so stereochemical discrimination behaves as a measurable, reproducible target-associated characteristic — though not, on this evidence, an established intrinsic property of the protein, since stereocenter position, assay design and target-specific research programmes remain uncontrolled."

Section 3.12 has been strengthened in the same direction, since it is where the inference is actually drawn. The operational claim is retained but its scope is stated: "...which is the operational meaning

of a target-associated rather than a series-associated quantity. Reliability across series does not by itself establish an intrinsic property of the protein: stereocenter position relative to the pharmacophore, assay design, and the research programme behind a given target's compound collection are shared across that target's series and remain uncontrolled."

Location: Abstract; Section 3.12, paragraph 2; Section 6.

R2.5 — Censoring claim is slightly broader than the analysis

> **Comment.** "Censoring claim is slightly broader than the analysis. Section 3.14 says the reported values are conservative estimates of 'underlying stereoselectivity.' The analysis establishes this direction for the particular exact-versus-greater-than censored pairs that were recovered—not necessarily for every excluded or missing pair. The Abstract and Conclusion repeat the broader wording."

Response. Correct. The censoring analysis recovers one specific class of excluded pair — those in which one enantiomer carries an exact value and the other a ">" bound in the same assay — and demonstrates the direction of bias for that class. It says nothing about pairs in which both enantiomers were censored, or in which one enantiomer was never assayed alongside the other; those cannot be bounded from the archive at all, and their effect on the distribution is genuinely unknown. The conclusion has been restricted accordingly in all three places.

Section 3.14 now reads: "With respect to this recoverable form of censoring — one enantiomer exact, the other reported as a '>' bound in the same assay — the reported medians and tail fractions are therefore conservative, as is the protease-kinase gap. That direction is established for those pairs only. Pairs excluded for other reasons — both enantiomers censored, or one enantiomer never assayed alongside the other — could not be recovered, and their effect on the distribution is unknown."

The **Abstract** now says the values "are conservative with respect to that recoverable censoring", and the **Conclusion** says they "are conservative with respect to the censored pairs that could be recovered ... pairs excluded for other reasons were not recovered and their effect is unknown."

Location: Abstract; Section 3.14, final paragraph; Section 6.

R2.6 — "Variance explained" is undefined

> **Comment.** "'Variance explained' is undefined. The manuscript reports 4.6%, 4.2%, 7.3%, and 17.8% variance explained but does not clearly identify the common (R^2) or other statistic from which these values were calculated. This is an internal methodological omission, particularly because the surrounding analysis uses median regression."

Response. A fair criticism, and the ambiguity was real: the surrounding text reports median-regression coefficients, so a reader would reasonably expect the variance shares to come from the same fit. They do not. R^2 is not defined for quantile regression, and no pseudo- R^2 was substituted; the variance shares are ordinary-least-squares coefficients of determination from a *separate* least-squares fit of the same predictors to Δ on the untransformed scale, included only to put the competing effects on a common scale.

That is now stated in both places. In **Methods (Section 2.8)**: "Because R^2 is not defined for quantile regression and no pseudo- R^2 was substituted, every figure reported as 'variance explained' or ' R^2 ' in Section 3.11 is the ordinary-least-squares coefficient of determination from a separate least-squares fit of the same predictors to Δ , and is intended only to put the competing effects on a common scale."

And at the point of use in **Section 3.11**: "These shares are ordinary-least-squares coefficients of determination (R^2) for Δ on the untransformed scale, computed separately from the median-regression coefficients above (Section 2.8)."

The one other R^2 in the manuscript — eutomer potency at $R^2 = 0.073$ in Section 3.8 — comes from the same partition and is now covered by the Methods definition.

Location: Section 2.8, final sentence; Section 3.11, variance-partition paragraph.

Other changes not prompted by a reviewer comment

Code and Data Availability — repository URL corrected. The submitted manuscript cited [`https://github.com/gnarlylace51842/eudysmic-ratio-map`](https://github.com/gnarlylace51842/eudysmic-ratio-map). That repository holds a different study by the same author, archived under its own DOI, so the pipeline for the present paper has been deposited in its own repository and the URL updated to:

<https://github.com/gnarlylace51842/eudysmic-ratio-target-family>

The deposited repository now contains the complete pipeline, including every analysis added during the first round of revision — the ligand-descriptor construction, the target-level eudysmic index and its reliability tests, the multivariable adjustment, the information analysis, the censoring analysis and the equivalence tests — together with the derived pair-level and target-level datasets and all eight figures. The previously deposited figures were also regenerated, because two of them carried the "three tiers" and "not a cherry-picking artifact" framings withdrawn in the first revision, and the repository README repeated those same withdrawn claims. Code is released under the MIT License; the derived data carry the ChEMBL CC BY-SA 3.0 terms. No text or result in the manuscript changed as a consequence of this correction.

Note on the figures

Regenerating the manuscript for this round did not change any figure in the manuscript itself. Figure 8b plots posterior probabilities against the prior and never displayed the mislabeled ratio from comment 2, so it required no correction. The figure set is unchanged from the previous submission.

Thank you for addressing my comments. Accept.

One concern that may merit additional explanation:

In the manuscript, you state that two families were called equivalent only when the CI for their difference fell entirely within a “pre-stated margin of ± 0.10 log units,” adding that this is “well below the smallest between-family gap of interest.” This suggests post-hoc insertion. Why should differences smaller than 1.26 fold be regarded as negligible? Needs discussion and justification.

In response to the reviewer's final comment, I removed the circular justification of the ± 0.10 log equivalence margin — “well below the smallest between-family gap of interest,” which defined the threshold by reference to the very differences it was used to judge — together with the unverifiable claim that the margin was “pre-stated.” The margin is now anchored externally, in Section 2.8, to measurement reproducibility: ± 0.10 log corresponds to a 1.26-fold difference, which is smaller than the $\sigma = 0.17\text{--}0.22$ log standard deviation reported for repeated determinations of the same compound in the same assay, so a difference of that size cannot be resolved by the underlying data. I added a sensitivity analysis (Section 3.3; new script `src/19_margin_sensitivity.py` and `results/margin_sensitivity.json` in the deposited repository) showing that the conclusion does not depend on the threshold: across margins from ± 0.05 to ± 0.50 the set of equivalent pairs is unchanged, because nine of the ten family contrasts have difference intervals excluding zero and therefore cannot be equivalent at any margin, while kinase versus ion channel qualifies down to ± 0.04 . Because the reviewer had identified a general defect rather than a single lapse, I audited every other numerical cutoff in the manuscript and corrected three: the per-target minimum of 20 pairs, previously referred to in Methods only as “a stated minimum,” is now stated and justified by its precision–coverage trade-off (Section 2.7), with its consequences reported across minima of 20, 30 and 50 pairs (Section 3.12), where I also now disclose that the nuclear-receptor/protease ordering inverts at a minimum of 30 — the same boundary already reported as indeterminate after scaffold collapse; the assay-size stratum boundaries are now stated to be conventional and not selected against the outcome (Section 2.9); and the 0.2-SD propensity caliper and the variance-inflation-factor threshold of 10 are now identified as conventions rather than presented as facts (Section 2.8

and the Table 5 caption). No result, number, table or figure changed as a consequence of any of these edits.