

SUPPORTING INFORMATION

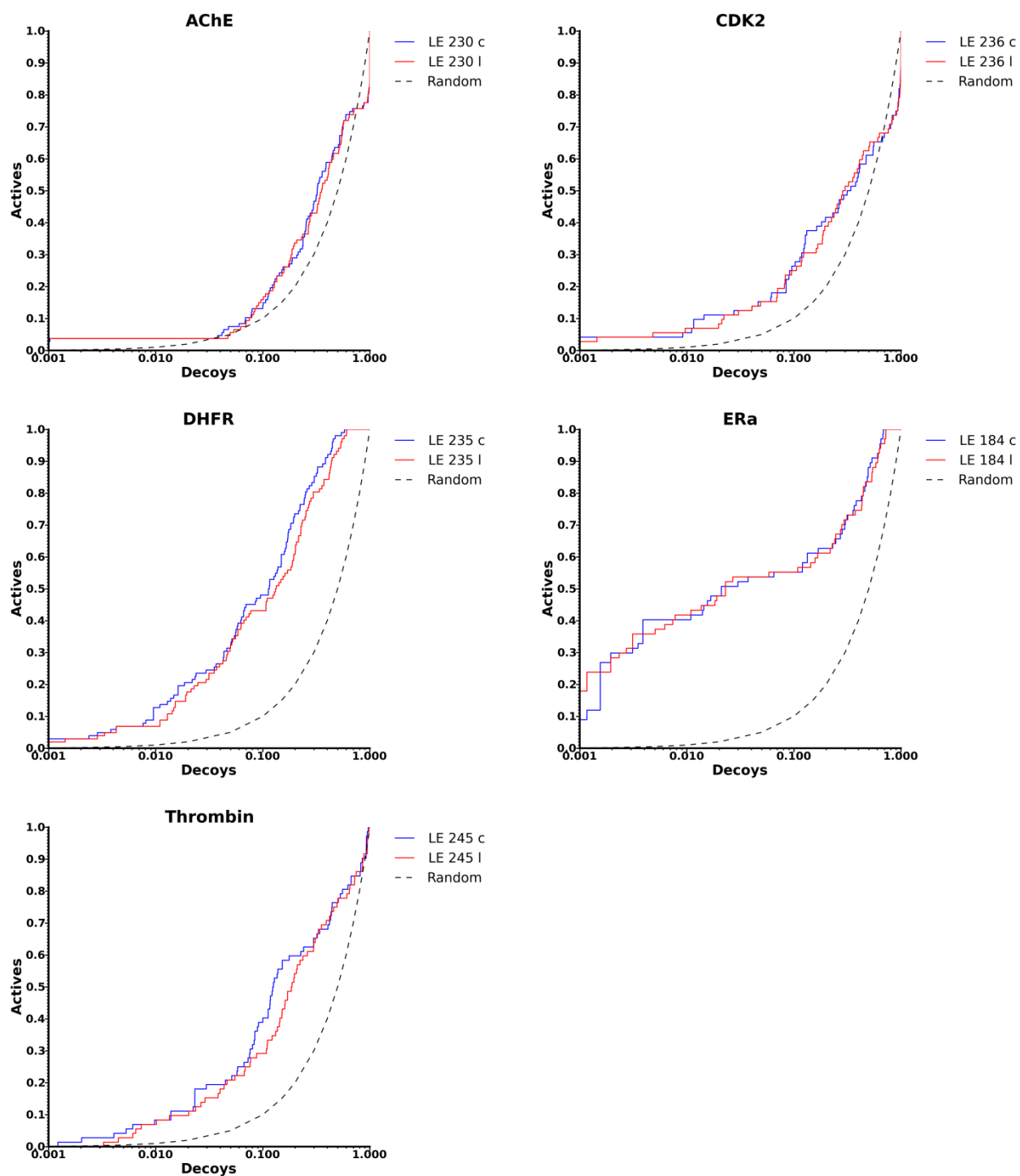


Figure S1. Docking enrichment plots for five protein systems using DUD. The fraction of actives ranked by docking score (y-axis) is plotted against the fraction of ranked decoys (x-axis; logarithmic scale). Results derived from docking to the Limoc ensemble (LE) using the average score of a poses cluster (blue, 'c') and overall lowest score (red, 'l'), are shown. Theoretical results from random selection (black dots) are shown for comparison.

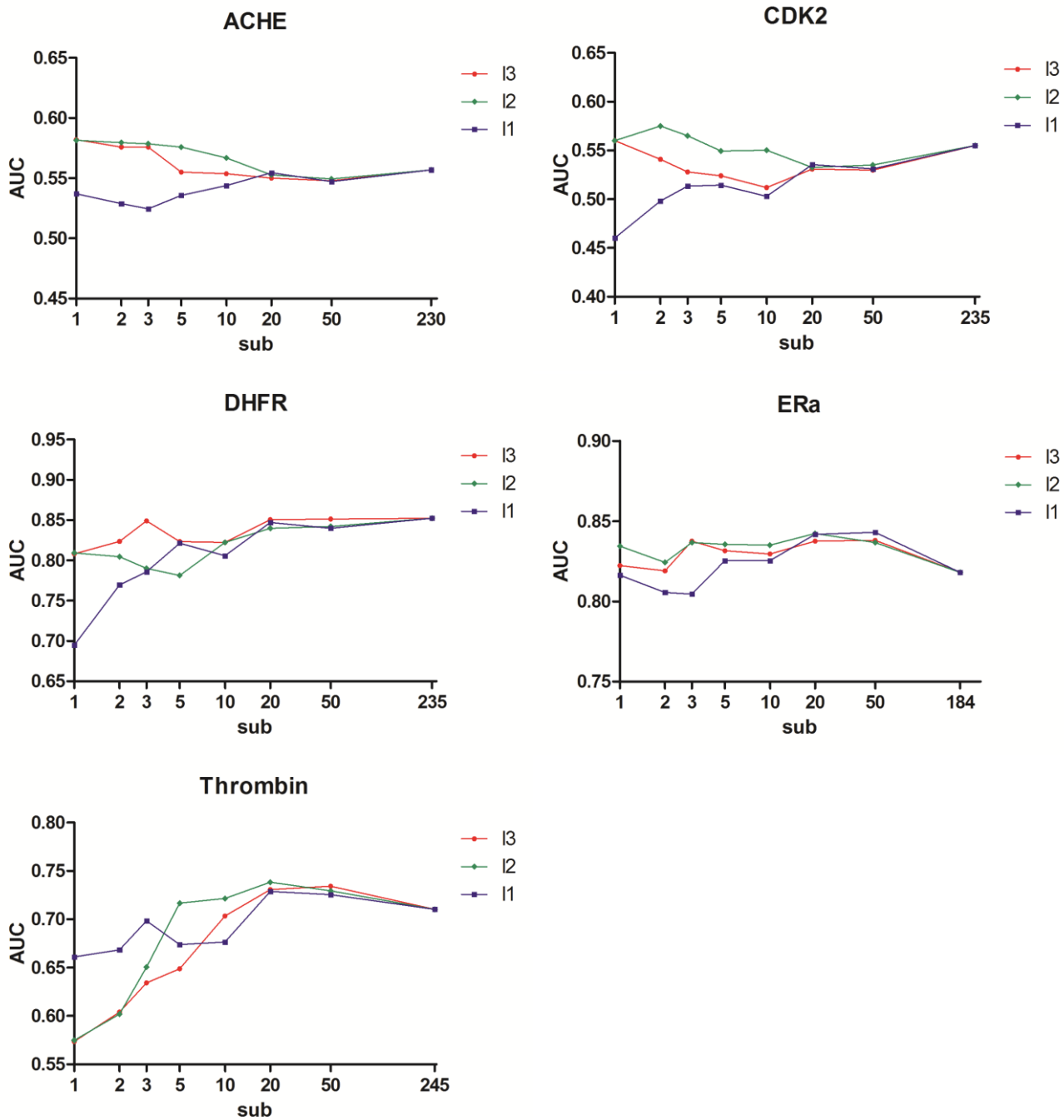


Figure S2. AUC as a function of the number of members in the sub-ensembles of protein structures. The results for frame selection by training using binding-pose reproduction of one (blue), two (green) and three (red) different co-crystallized ligands.

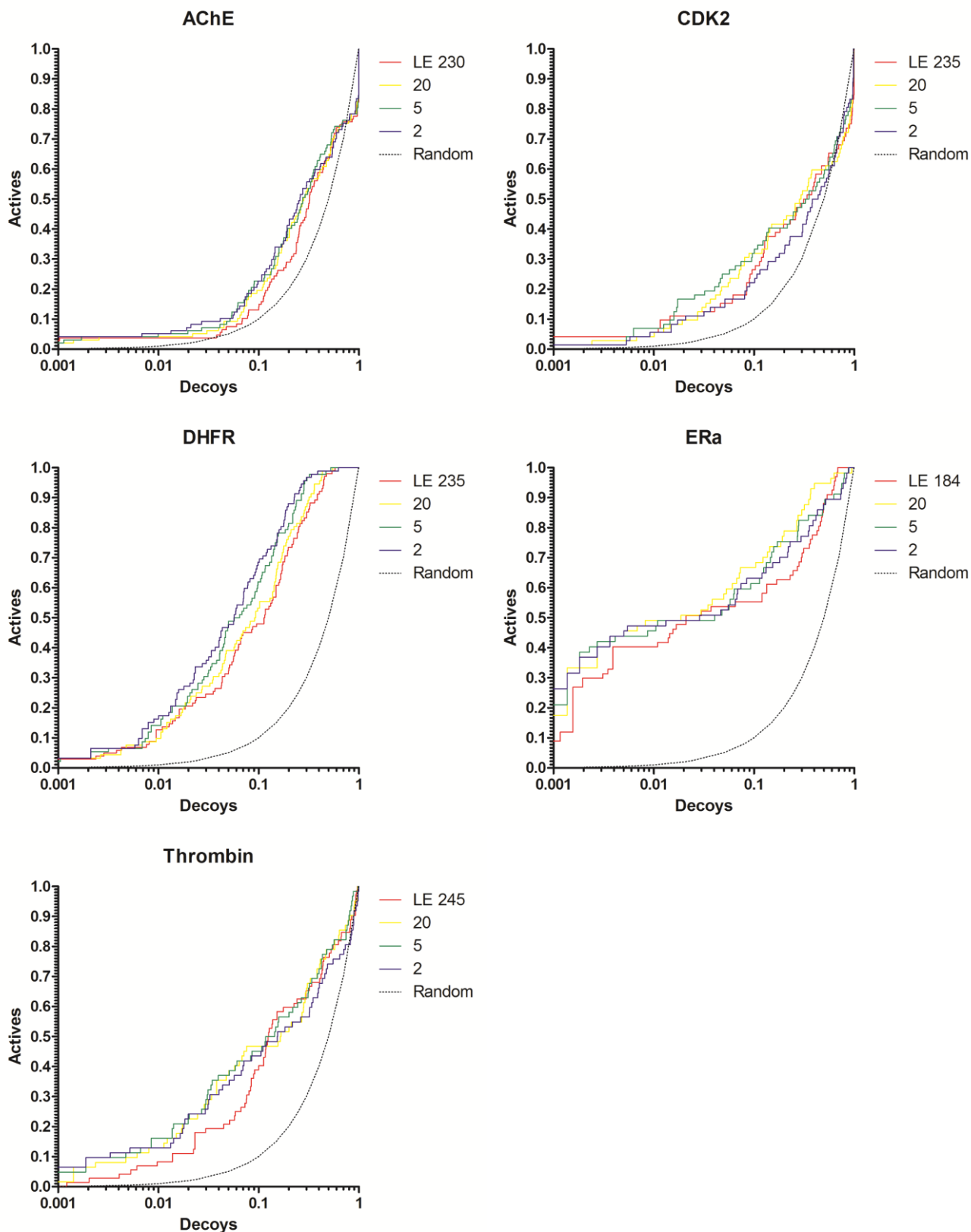


Figure S3. Docking enrichment plots for five protein systems using DUD. The fraction of actives ranked by docking score (y-axis) is plotted against the fraction of ranked decoys (x-axis; logarithmic scale). Results docking to the full Limoc EPS (red line) and sub-ensembles with 20 (yellow), 5 (green), and 2 (purple) members are shown. Members of the sub-ensembles were selected based on their potential to actives from decoys of a small sub-set of ligands from the DUD dataset. The results for using 10 actives are displayed.

Table S4. Slopes of linear regression models on quantitative indicators for enrichment quality using selection of EPS members based on binding pose predictions of one (l1) and two (l2) ligands. Negative regression (blue color) means increasing enrichment quality with reducing number of EPS members; positive regression in red.

| l2 | Thrombin | DHFR | ERa | CDK2 | ACHE |
|--------|----------|-------|--------|--------|--------|
| EF1 | -0.019 | 0.022 | 0.051 | -0.002 | 0.000 |
| EF10 | 0.044 | 0.062 | 0.000 | -0.045 | -0.025 |
| EF20 | 0.089 | 0.078 | -0.013 | -0.019 | -0.059 |
| AUC | 0.059 | 0.026 | -0.003 | -0.010 | -0.015 |
| LogAUC | 0.024 | 0.030 | 0.020 | -0.020 | -0.008 |

| l1 | Thrombin | DHFR | ERa | CDK2 | ACHE |
|--------|----------|-------|--------|-------|-------|
| EF1 | 0.070 | 0.030 | 0.008 | 0.005 | 0.000 |
| EF10 | 0.010 | 0.090 | 0.003 | 0.045 | 0.003 |
| EF20 | 0.074 | 0.123 | -0.014 | 0.033 | 0.018 |
| AUC | 0.025 | 0.056 | 0.010 | 0.032 | 0.124 |
| LogAUC | 0.012 | 0.051 | 0.007 | 0.025 | 0.007 |

Table S5. Results of multi-linear regression analysis for AChE, CDK2, DHFR, ERa and thrombin correlating $d_{ij}^<$ and $d_{ij}^>$ with the difference between $\overline{S}_{actives,p}$ and $\overline{S}_{decoys,p}$.

| Protein: AChE | $r^2 = 0.47$ | $C_0 = 0.04$ | |
|---------------------|--------------------|--------------------------|----------------|
| Atom i | Atom j | $d_{ij}^>$ or $d_{ij}^<$ | $c_{ij}^{>/<}$ |
| TRP84, C α | TYR121, C ζ | > | 0.12 |
| TRP84, C δ 2 | TYR127, C α | > | 0.07 |
| GLY119, C α | ILE441, C α | > | 0.25 |
| LEU127, C γ | PHE290, C γ | < | 0.14 |

| Protein: CDK2 | $r^2 = 0.27$ | $C_0 = 0.01$ | |
|-------------------|---------------------|--------------------------|----------------|
| Atom i | Atom j | $d_{ij}^>$ or $d_{ij}^<$ | $c_{ij}^{>/<}$ |
| ILE10, C α | LYS89, C ϵ | > | -0.06 |
| GLY11, C α | PHE82, C γ | > | 2.73 |
| VAL64, C α | LYS89, C α | > | 0.25 |
| VAL64, C α | ALA144, C α | > | -0.17 |
| PHE80, C γ | PHE82, C γ | < | 0.43 |
| ASP86, C γ | LUE134, C γ | > | 18.69 |
| LYS89, C α | ASP145, C α | < | 0.22 |

| Protein: DHFR | $r^2 = 0.45$ | $C_0 = -0.65$ | |
|---------------------|-------------------|--------------------------|----------------|
| Atom i | Atom j | $d_{ij}^>$ or $d_{ij}^<$ | $c_{ij}^{>/<}$ |
| ILE10, C γ 1 | GLU32, C α | > | 0.18 |
| ILE10, C γ 1 | PHE36, C α | > | 0.71 |
| LEU25, C α | LYS37, C α | > | 0.32 |
| GLU32, C α | ILE65, C α | > | 0.11 |
| SER34, O γ | PHE36, C γ | < | 0.59 |

| Protein: ERa | $r^2 = 0.74$ | $C_0 = -1.35$ | |
|----------------------|----------------------|--------------------------|----------------|
| Atom i | Atom j | $d_{ij}^>$ or $d_{ij}^<$ | $c_{ij}^{>/<}$ |
| LEU346, C α | LEU525, C α | < | 0.19 |
| GLU353, C δ | ARG394, C α | < | 0.26 |
| GLU353, C δ | LEU428, C α | > | 0.16 |
| MET 388, S δ | ARG394, N ϵ | > | 0.11 |
| ARG394, N ϵ | PHE404, C γ | > | 0.13 |
| PHE404, C α | HIS524, C α | < | 0.44 |

| Protein: Thrombin | $r^2 = 0.66$ | $C_0 = -0.35$ | |
|---------------------|----------------------|--------------------------|----------------|
| Atom i | Atom j | $d_{ij}^>$ or $d_{ij}^<$ | $c_{ij}^{> <}$ |
| HIS57, C_γ | GLU217, C_α | > | 1.76 |
| TYR60A, C_α | ILE174, $C_\gamma 1$ | > | 0.17 |
| TYR60A, C_ζ | LEU99, C_γ | < | 0.24 |
| TRP60D, $C\delta 2$ | SER195, O_γ | < | 0.34 |
| TRP60D, $C\delta 2$ | TRP215, C_α | > | 1.38 |
| LEU99, C_α | SER195, C_α | < | 0.45 |
| ALA190, C_β | GLY216, C_α | < | 0.34 |
| CYS191, S_γ | GLY216, C_α | < | 0.15 |
| GLU192, $C\delta$ | TRP215, C_α | < | 0.15 |