

# Additional file

## Assessing the Calibration in Toxicological in Vitro Models with Conformal Prediction

Andrea Morger, Fredrik Svensson, Staffan Arvidsson McShane, Niharika Gauraha, Ulf Norinder, Ola Spjuth and Andrea Volkamer

Table S1: Number of compounds available per Tox21 dataset and endpoint before standardisation.

| endpoint          | Tox21Train |           | Tox21Test |           | Tox21Score |           |
|-------------------|------------|-----------|-----------|-----------|------------|-----------|
|                   | actives    | inactives | actives   | inactives | actives    | inactives |
| NR_AhR            | 950        | 7214      | 30        | 241       | 73         | 537       |
| NR_AR             | 380        | 8977      | 3         | 288       | 12         | 574       |
| NR_AR.LBD         | 303        | 8291      | 4         | 248       | 8          | 574       |
| NR_Aromatase      | 360        | 6861      | 18        | 196       | 39         | 489       |
| NR_ER             | 937        | 6756      | 27        | 237       | 51         | 465       |
| NR_ER.LBD         | 446        | 8302      | 10        | 276       | 20         | 580       |
| NR_PPAR- $\gamma$ | 222        | 7957      | 15        | 251       | 31         | 574       |
| SR_ARE            | 1097       | 6067      | 47        | 186       | 93         | 462       |
| SR_ATAD5          | 338        | 8748      | 25        | 246       | 38         | 584       |
| SR_HSE            | 428        | 7718      | 10        | 256       | 22         | 588       |
| SR_MMP            | 1142       | 6174      | 38        | 199       | 60         | 483       |
| SR_p53            | 537        | 8092      | 28        | 240       | 41         | 575       |

Table S2: Mean  $\pm$  standard deviation values over all twelve endpoints for observed error rate and efficiency at SL 0.2 for all experiments.

| nr. | name                  | error rate at SL 0.2 | efficiency at SL 0.2 |
|-----|-----------------------|----------------------|----------------------|
| 1   | <i>internal_CV</i>    | $0.17 \pm 0.01$      | $0.77 \pm 0.13$      |
| 2   | <i>pred_score</i>     | $0.31 \pm 0.12$      | $0.72 \pm 0.14$      |
| —   | <i>pred_test</i> *    | $0.26 \pm 0.11$      | $0.70 \pm 0.15$      |
| 3   | <i>pred_score_SCP</i> | $0.27 \pm 0.12$      | $0.73 \pm 0.13$      |
| 4   | <i>train_update</i>   | $0.23 \pm 0.06$      | $0.71 \pm 0.15$      |
| 5   | <i>cal_update</i>     | $0.21 \pm 0.05$      | $0.51 \pm 0.18$      |
| 6   | <i>cal_update.2</i>   | $0.18 \pm 0.01$      | $0.50 \pm 0.17$      |

\**pred\_test*: the CV-models from *internal\_cv* were used to make predictions on Tox21Test.

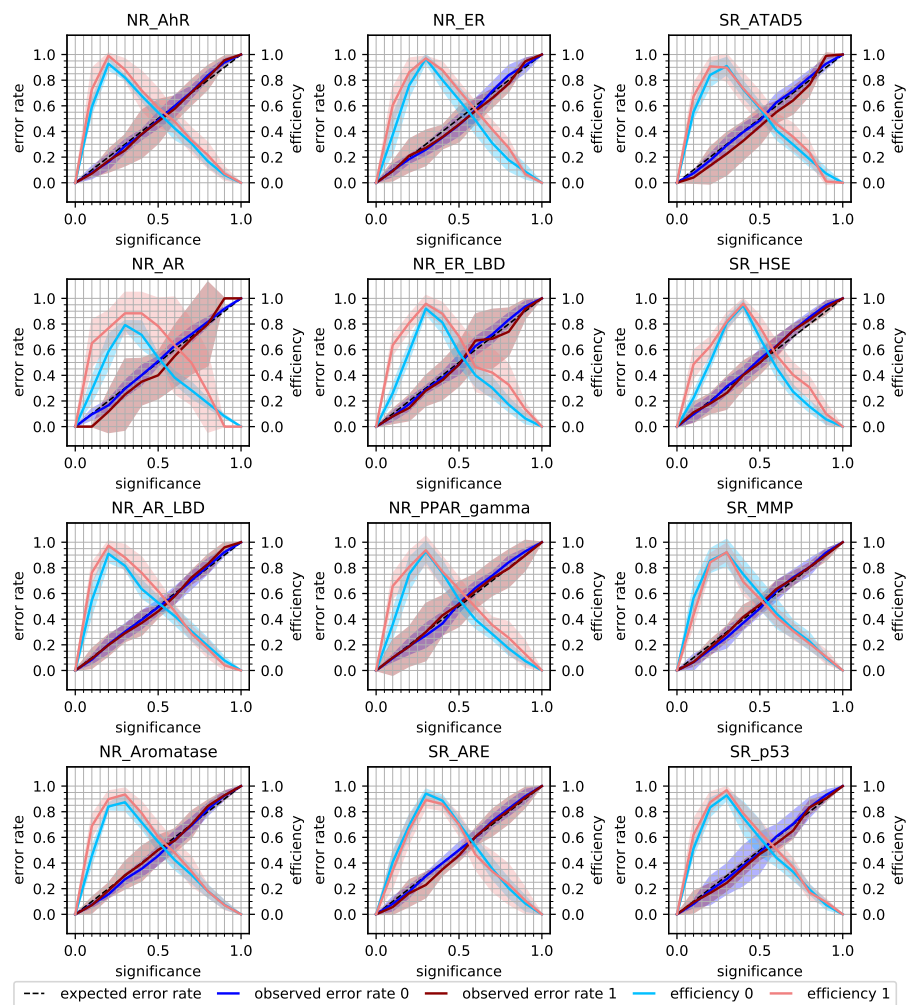


Figure S1: *1-internal-CV*: ACP models were trained and calibrated on Tox21Train and internally validated. CEPs for all twelve Tox21 endpoints are shown. Class 0: inactive compounds, class 1: active compounds. For a detailed explanation of all the components in the CEP, see Figure 2.

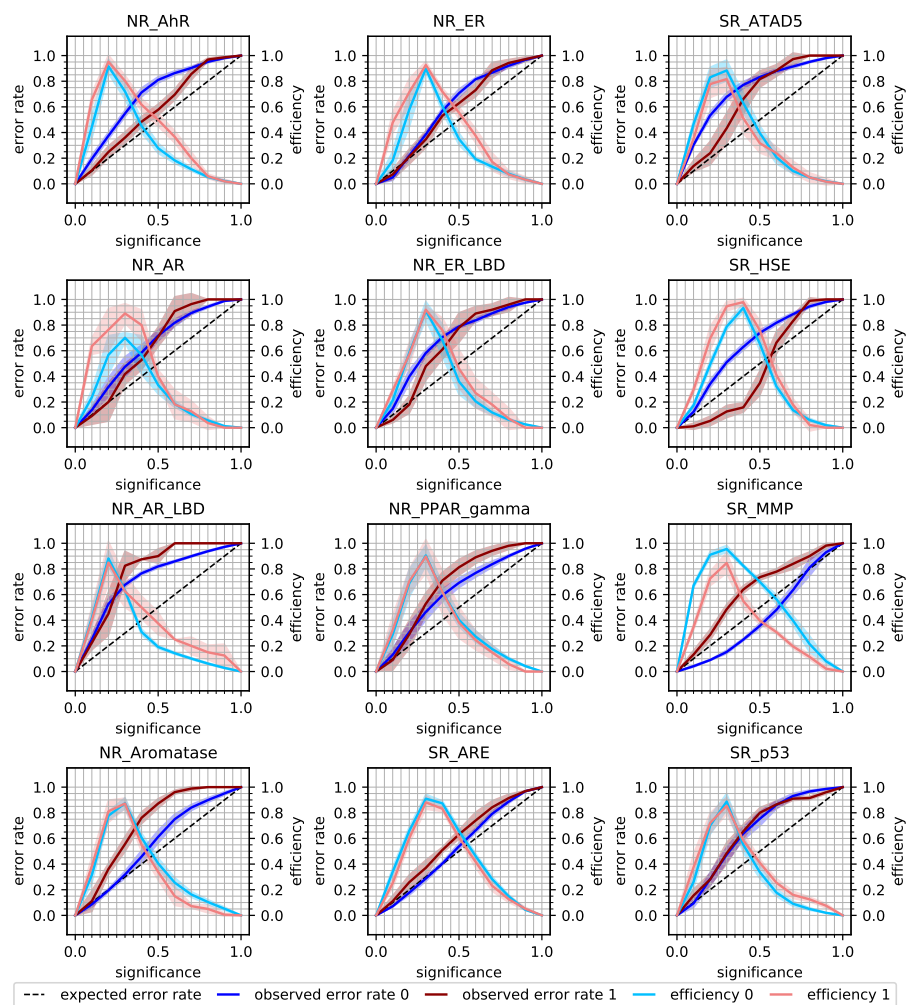


Figure S2: *2-pred\_score*: ACP models were trained and calibrated on Tox21Train and predictions were made for Tox21Score. CEPs for all twelve Tox21 endpoints are shown. Class 0: inactive compounds, class 1: active compounds. For a detailed explanation of all the components in the CEP, see Figure 2.

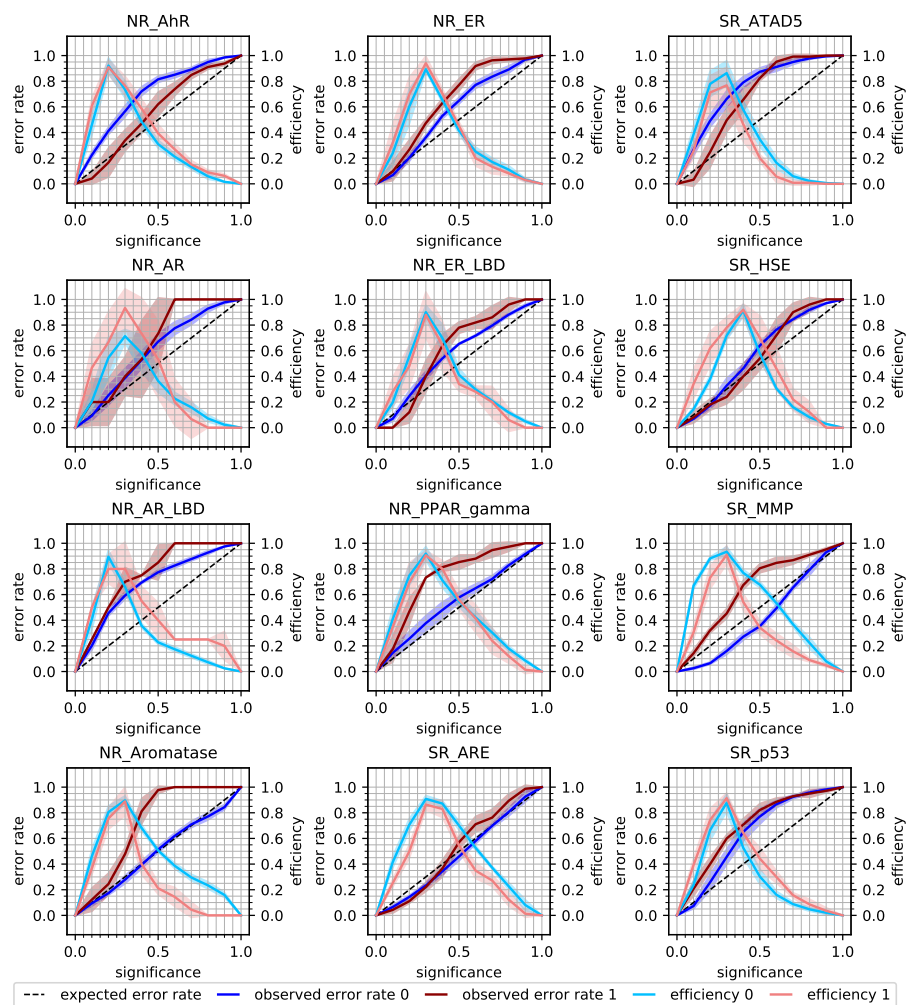


Figure S3: *pred.test*: ACP models were trained and calibrated on Tox21Train and predictions were made for Tox21Test. CEPs for all twelve Tox21 endpoints are shown. Class 0: inactive compounds, class 1: active compounds. For a detailed explanation of all the components in the CEP, see Figure 2.

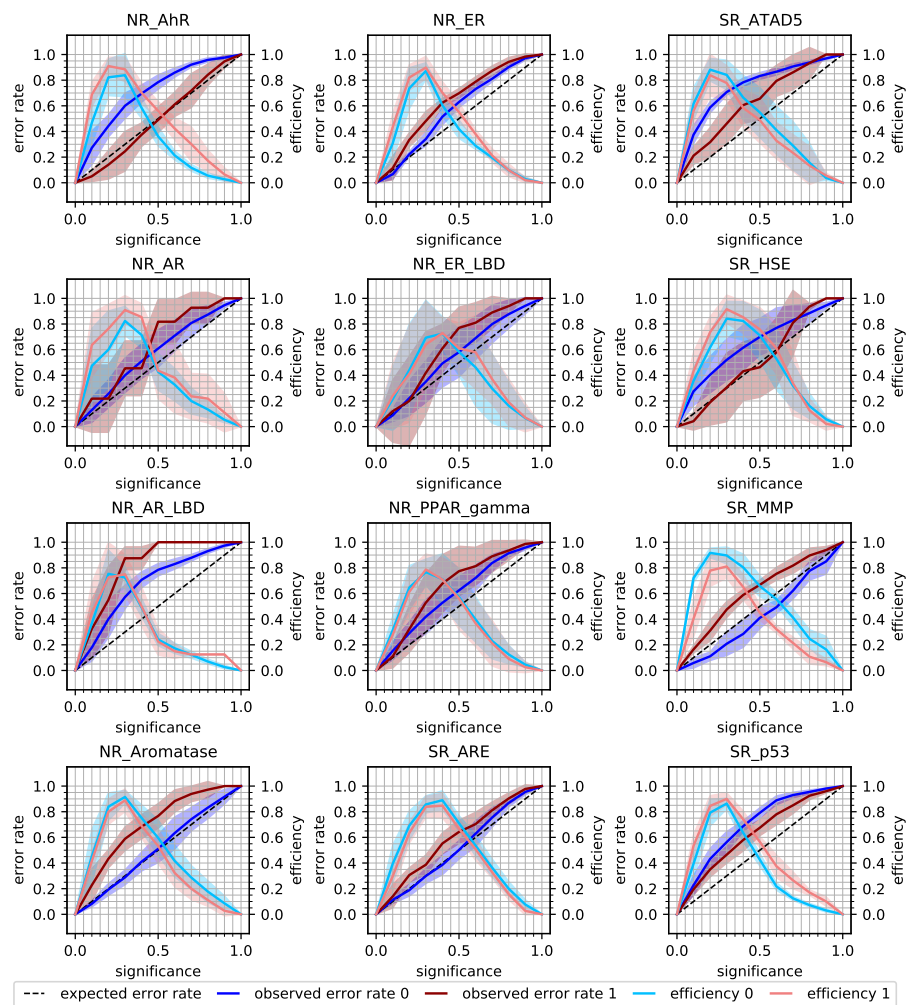


Figure S4: *3-pred\_score\_SCP*: SCP models were trained on Tox21Train and predictions made for Tox21Score. CEPs for all twelve Tox21 endpoints are shown. Class 0: inactive compounds, class 1: active compounds. For a detailed explanation of all the components in the CEP, see Figure 2.

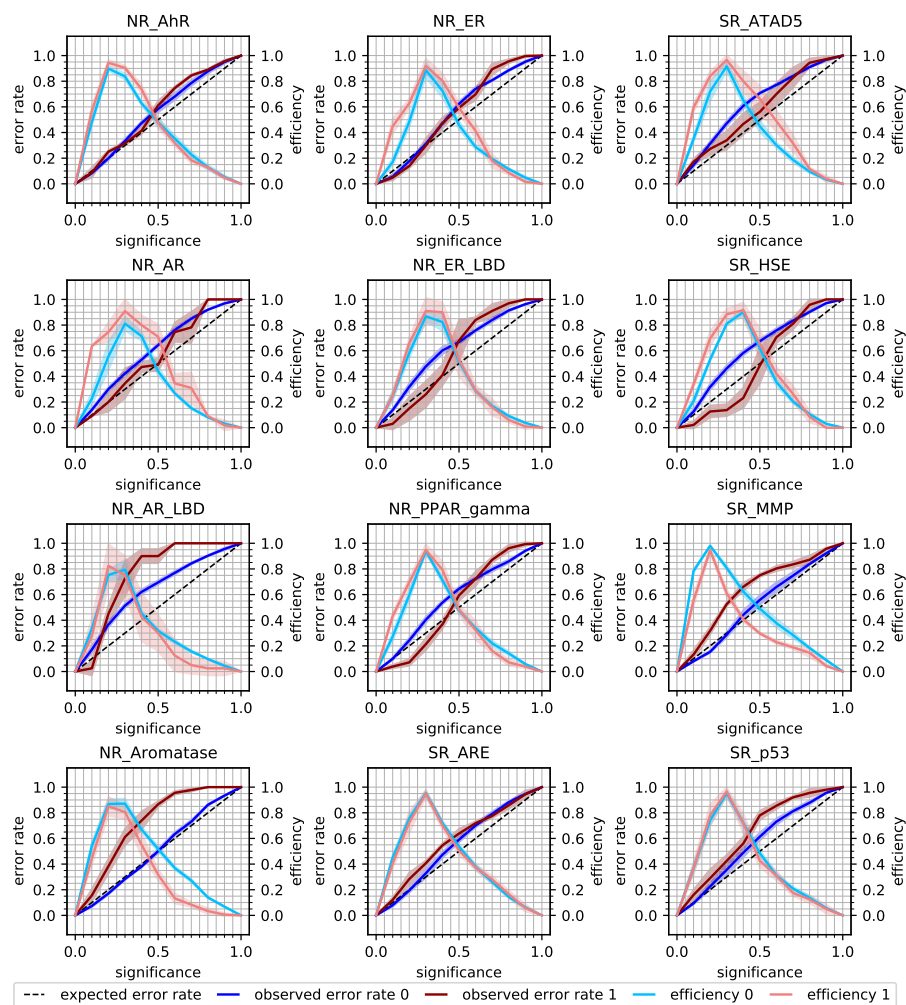


Figure S5: *4-train-update*: The training set from Tox21Train was updated with Tox21Test. An ACP model was retrained and predictions were made for Tox21Score. CEPs for all twelve Tox21 endpoints are shown. Class 0: inactive compounds, class 1: active compounds. For a detailed explanation of all the components in the CEP, see Figure 2.

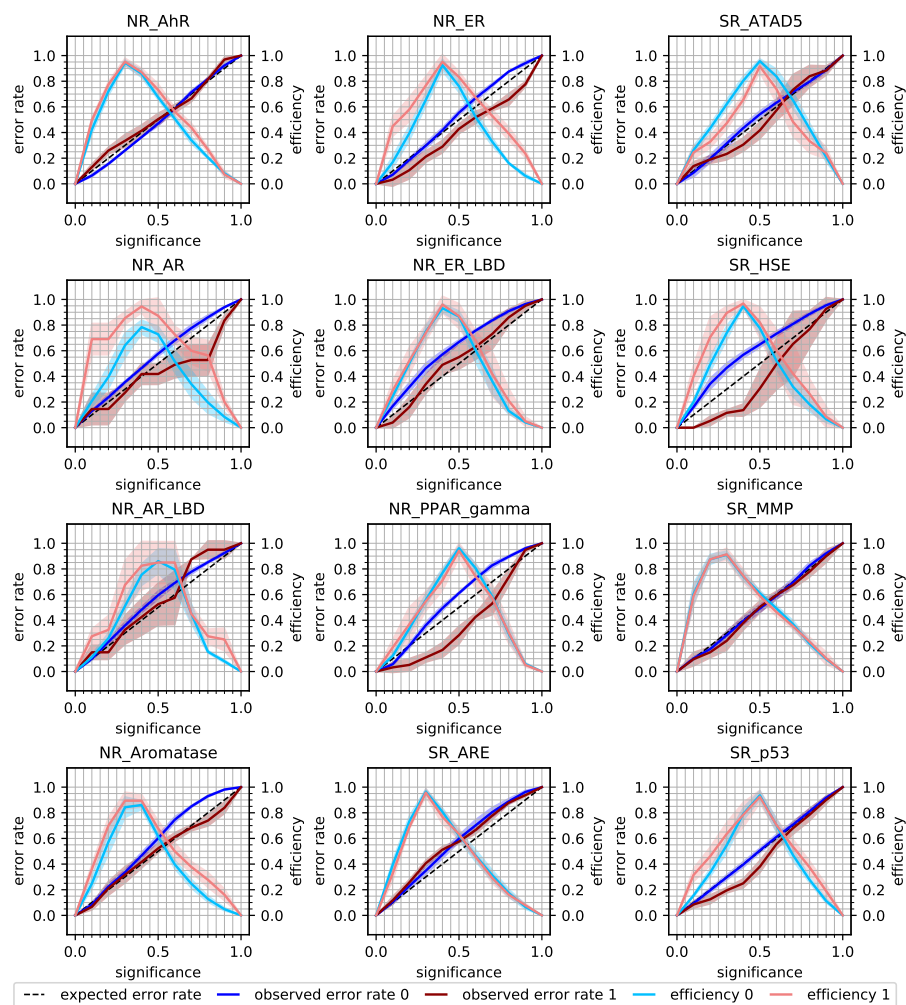


Figure S6: *5-cal\_update*: ACP models were trained on Tox21Train and calibrated on Tox21Test. Predictions were made for Tox21Score. CEPs for all twelve Tox21 endpoints are shown. Class 0: inactive compounds, class 1: active compounds. For a detailed explanation of all the components in the CEP, see Figure 2.

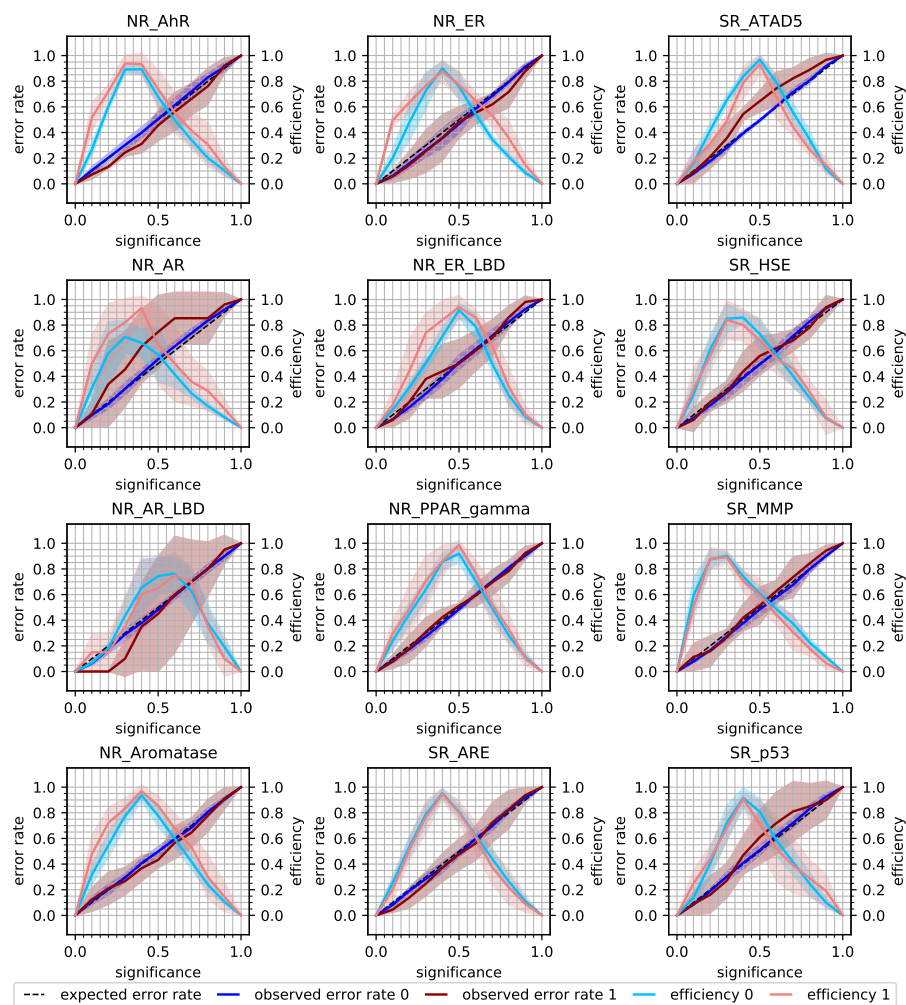


Figure S7: *6-cal\_update\_2*: ACP models were trained on Tox21Train and calibrated on 50% of Tox21Score. Predictions were made for the other 50% of Tox21Score. CEPs for all twelve Tox21 endpoints are shown. Class 0: inactive compounds, class 1: active compounds. For a detailed explanation of all the components in the CEP, see Figure 2.

Table S3: Mean RMSD values over all 12 endpoints, calculated for all compounds, as well as for active and inactive compounds, separately.

| nr. | name                  | all   | actives | inactives |
|-----|-----------------------|-------|---------|-----------|
| 1   | <i>internal_CV</i>    | 0.022 | 0.032   | 0.022     |
| 2   | <i>pred_score</i>     | 0.150 | 0.167   | 0.154     |
| —   | <i>pred_test*</i>     | 0.116 | 0.180   | 0.119     |
| 3   | <i>pred_score_SCP</i> | 0.121 | 0.147   | 0.124     |
| 4   | <i>train_update</i>   | 0.090 | 0.135   | 0.089     |
| 5   | <i>cal_update</i>     | 0.054 | 0.073   | 0.058     |
| 6   | <i>cal_update_2</i>   | 0.018 | 0.057   | 0.018     |

\**pred\_test*: the CV-models from *internal\_cv* were used to make predictions on Tox21Test.