A Scaffold-Split QSAR Model for Human EGFR Kinase Inhibition

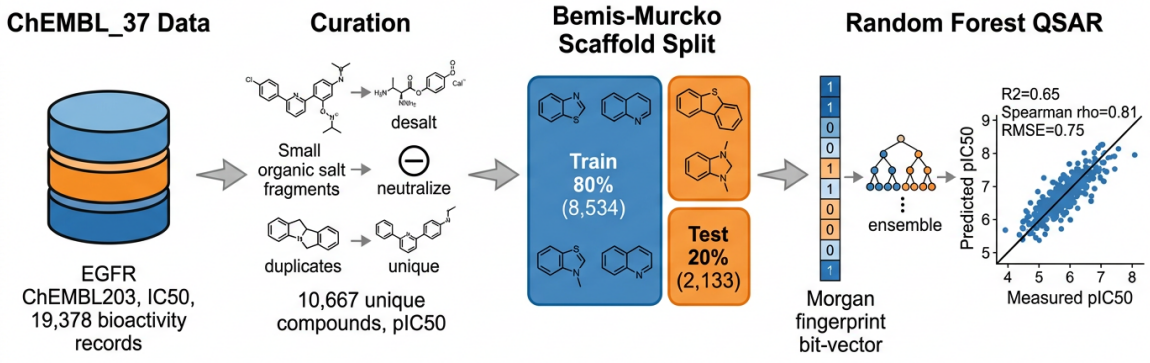
Data Curation, Bemis–Murcko Splitting, and Machine-Learning Evaluation on ChEMBL_37 Bioactivity Data (Target CHEMBL203)

K-Dense Web
contact@k-dense.ai

10 July 2026

Abstract

The human epidermal growth factor receptor (EGFR) is one of the most intensively pursued oncology drug targets, and quantitative structure–activity relationship (QSAR) models are widely used to prioritize its inhibitors. We assembled a bioactivity dataset for EGFR (ChEMBL target `CHEMBL203`) from the **ChEMBL_37** release (May 2026), restricting to single-protein IC_{50} assays with a defined `pChEMBL` value (19,378 activity records). Structures were desalted, neutralized, and tautomer-canonicalized with RDKit, and replicate measurements were aggregated by mean pIC_{50}, yielding **10,667 unique curated compounds** (pIC_{50} range 4.0–11.0). The dataset was partitioned with an 80/20 Bemis–Murcko scaffold split (seed 42; 8,534 train / 2,133 test) that enforces *zero* ring-scaffold overlap between partitions, so that held-out performance reflects extrapolation to novel chemotypes. Compounds were featurized as 2048-bit Morgan fingerprints (radius 2). A random forest and an XGBoost regressor were tuned by scaffold-grouped 5-fold cross-validation; the random forest was selected on lowest mean CV RMSE and retrained on the full training set. On the held-out test set the model achieved **RMSE** $= 0.746$, **MAE** $= 0.561$, $R^2 = 0.646$, and **Spearman** $\rho = 0.808$. The strong rank correlation, contrasted with a fitted predicted-vs-measured slope of 0.64 (regression-to-the-mean shrinkage of the extreme tails), indicates a model that is highly effective for *ranking and triage* even where absolute potency of out-of-domain, highly potent scaffolds is systematically under-predicted. We report the full per-compound test-set prediction table and the ten most potent held-out compounds with measured and predicted pIC_{50}.



Graphical abstract: end-to-end workflow from ChEMBL_37 retrieval and structure curation through Bemis–Murcko scaffold splitting to random-forest QSAR and held-out evaluation.

Contents

1 Introduction

1.1 EGFR as an anticancer drug target

The epidermal growth factor receptor (EGFR, also ErbB1/HER1) is a prototypical receptor tyrosine kinase whose ligand-induced dimerization and trans-autophosphorylation activate down-

stream RAS–MAPK and PI3K–AKT signaling that governs cell proliferation, survival, and migration [1]. Aberrant EGFR activity—through overexpression or activating kinase-domain mutations—is a validated oncogenic driver, most prominently in non-small-cell lung cancer (NSCLC). The 2004 discovery that somatic kinase-domain mutations (classically the exon 19 deletions and the L858R substitution) confer exquisite sensitivity to the small-molecule inhibitor gefitinib established EGFR as a cornerstone of precision oncology [2, 3]. Subsequent work defined the T790M "gatekeeper" mutation as a dominant mechanism of acquired resistance to first-generation inhibitors [4], motivating successive generations of tyrosine kinase inhibitors. Third-generation agents such as osimertinib, which target T790M-mutant disease and are now used first-line, have delivered substantial progression-free and overall survival gains [5, 6]. The continued clinical importance of EGFR, coupled with an unusually rich medicinal-chemistry record, makes it an excellent and well-populated target for data-driven modeling of inhibitor potency.

1.2 QSAR modeling and the purpose of this study

Quantitative structure–activity relationship (QSAR) modeling seeks a mathematical mapping from molecular structure to a measured biological endpoint, here the half-maximal inhibitory concentration (IC_{50}) expressed on the logarithmic pIC_{50} scale. QSAR models underpin virtual screening, hit triage, and lead optimization by allowing computational prioritization of compounds before synthesis and assay [7, 8]. Their reliability, however, is contingent on disciplined practice: rigorous chemical-structure curation, appropriate train/test partitioning, honest external validation, and an explicit statement of the applicability domain [9, 10]. A recurring failure mode is over-optimistic performance produced by random train/test splits, in which near-duplicate analogues of training compounds leak into the test set; splitting by chemical scaffold provides a substantially more realistic estimate of prospective, extrapolative performance [11, 12].

The purpose of this report is to document a complete, reproducible QSAR workflow for EGFR (CHEMBL203) end to end: (i) retrieval and filtering of bioactivity data from a stated ChEMBL release; (ii) structure standardization, de-duplication, and conversion to pIC_{50}; (iii) an 80/20 Bemis–Murcko scaffold split with a fixed random seed; (iv) fingerprint featurization and comparative model training; and (v) transparent evaluation on a genuinely held-out, scaffold-disjoint test set. We report the curated compound count, four complementary performance metrics (RMSE, MAE, R^2, Spearman ρ), the ten most potent held-out compounds with predicted and measured pIC_{50}, and the complete per-compound prediction table for the test set.

2 Methods

2.1 Data retrieval from ChEMBL

Bioactivity data were obtained from the **ChEMBL_37** release (database version reported by the ChEMBL web-services status endpoint; official release date 1 May 2026), which at query time contained 24,527,044 activities across 18,552 targets [13, 14]. The target `CHEMBL203` was confirmed as *Epidermal growth factor receptor*, organism *Homo sapiens*, target type `SINGLE PROTEIN`. Activities were queried with the constraints `target_chembl_id = CHEMBL203`, `standard_type = IC50`, and `pchembl_value__isnull = False` (i.e. a defined `pChEMBL` value). The server reported and returned **19,378** matching activity records, which were saved verbatim before any processing. The `pChEMBL` value—ChEMBL's harmonized negative-log molar potency—was adopted as the primary pIC_{50} label. As a quality control, we verified `pChEMBL` against an independent recomputation $9 - \log_{10}$(standard_value in nM) for

every record: all 19,378 activities were reported in nM, with mean absolute deviation 0.0022 (maximum 0.0055) and zero anomalies exceeding 0.5 log units, confirming internal consistency of the potency labels.

2.2 Structure standardization and curation

Chemical-structure curation followed established cheminformatics best practice [10, 15] and was implemented with RDKit 2026.03.3 (`rdMolStandardize`) [16]. Each unique input SMILES (10,734 distinct strings) was processed through a deterministic pipeline: (1) parse and sanitize; (2) **desalt** by retaining the largest covalent fragment (`LargestFragmentChooser`), removing counter-ions and solvates; (3) **neutralize** formal charges (`Uncharger`); (4) **canonicalize** the tautomer (`TautomerEnumerator.Canonicalize`); and (5) emit a canonical isomeric SMILES. All 10,734 unique structures were standardized successfully with no parse failures. Seventeen activity records carried missing or invalid structures and were dropped (19,378 → 19,361 records).

2.3 De-duplication and aggregation

Multiple activity records frequently correspond to the same compound (replicate assays, multiple literature reports). We de-duplicated on the standardized canonical SMILES and aggregated replicate pIC_{50} measurements by their arithmetic **mean**, which reduces assay noise while preserving an unbiased central estimate [15]. Of the resulting compounds, 4,163 had two or more measurements (one compound had as many as 247 records), and the per-compound standard deviation of pIC_{50} was retained as a data-quality annotation. Curation produced a final modeling set of **10,667 unique compounds**, with pIC_{50} mean 6.93, median 7.00, and range $[4.0, 11.0]$ (Figure 2a). The full stage-by-stage compound accounting is summarized in Figure 1 and Table 1.

2.4 Bemis–Murcko scaffold split

To estimate prospective performance on *novel chemotypes* rather than close analogues, the curated set was partitioned by molecular scaffold rather than randomly [11, 12]. The Bemis–Murcko framework of each compound—the union of its ring systems and the linkers connecting them, obtained after iterative removal of terminal side chains [17]—was computed with RDKit `MurckoScaffold.MurckoScaffoldSmiles` (`includeChirality=False`). Two acyclic compounds had no ring scaffold and were tagged `no_scaffold`. Compounds were grouped by scaffold; the sorted scaffold list was shuffled with a fixed seed (`random.Random(42)`) and whole scaffold groups were assigned to the test set until the 20% target was reached, with the remainder assigned to training. This guarantees that all compounds sharing a scaffold stay together in a single partition. The two `no_scaffold` compounds were split 80/20 at random (both fell to train). The procedure yielded **8,534 training** (80.00%) and **2,133 test** (20.00%) compounds, spanning 3,838 unique scaffolds in total (2,947 train, 891 test) with **zero ring-scaffold overlap** between partitions. The largest scaffold was the 4-anilinoquinazoline core (`c1ccc(Nc2ncnc3ccccc23)cc1`, 575 compounds), and 2,542 scaffolds were singletons. The train and test pIC_{50} distributions were well matched (train mean/median 6.95/7.03; test 6.86/6.93; Figure 2b).

2.5 Molecular featurization

Each compound was represented by a **2048-bit Morgan (circular) fingerprint of radius 2**, computed with the modern RDKit `rdFingerprintGenerator.GetMorganGenerator(radius=2,`

`fpSize=2048`) API and converted to a `uint8` bit vector [18, 19]. Morgan/ECFP-style fingerprints encode the local atomic environment around each atom out to the chosen radius by hashing substructure identifiers into a fixed-length bit string, and remain a strong, widely used baseline representation for bioactivity modeling [19]. All 8,534 training and 2,133 test compounds were featurized without failure (mean ≈62 bits set per training molecule). Identical featurization code and fingerprint specification were used for training and evaluation to guarantee consistency.

2.6 Model training, tuning, and selection

Two regression algorithms were compared: a **random forest** [20] and a gradient-boosted tree ensemble, **XGBoost** (`tree_method='hist'`) [21], both implemented in Python with scikit-learn 1.9.0 and xgboost 3.3.0 [22]. To avoid scaffold leakage during model selection, hyperparameters were tuned with **scaffold-grouped 5-fold cross-validation** (`GroupKFold` grouped by the Bemis–Murcko scaffold column), so that every validation fold holds out whole scaffolds; zero group leakage was verified across all folds. For each algorithm, `RandomizedSearchCV` (`n_iter`= 20, seed 42, `refit=neg_root_mean_squared_error`) explored the hyperparameter space, scoring RMSE, MAE, R^2, and Spearman ρ per fold. The model with the *lowest mean cross-validated RMSE* was selected and retrained on all 8,534 training compounds. The held-out test set played no role in model or hyperparameter selection and was used only for the final, one-shot evaluation reported below, in keeping with QSAR validation guidance [9, 7].

2.7 Evaluation metrics and reproducibility

Held-out performance was quantified with four complementary metrics: root-mean-square error (RMSE) and mean absolute error (MAE), both in pIC_{50} units; the coefficient of determination R^2; and the Spearman rank correlation ρ between predicted and measured pIC_{50}. RMSE and MAE measure absolute accuracy, R^2 measures the fraction of variance explained, and ρ measures rank agreement—the quantity most relevant to screening and triage, where correctly *ordering* candidates matters more than exact values [8]. A fixed seed (42) was used throughout (split, search, and model), and all package versions were recorded (Python 3.12.10, RDKit 2026.03.3, scikit-learn 1.9.0, xgboost 3.3.0, numpy 2.5.1). To confirm reproducibility, test-set fingerprints were regenerated from scratch at evaluation time using the fingerprint specification stored in the trained-model bundle; the recomputed metrics matched the model-selection-phase values to $|\Delta| = 0.0000$ across all four metrics (tolerance 0.01).

3 Results

3.1 Curated dataset

Curation of the 19,378 raw ChEMBL_37 IC_{50} records produced **10,667 unique EGFR compounds** with defined pIC_{50} labels. Table 1 traces the compound counts at each stage, and Figure 1 presents the same pipeline as a flow diagram. The curated potency distribution (Figure 2a) is approximately bell-shaped and mildly left-skewed, centered near pIC_{50} 7 (i.e. $\mathrm{IC}_{50} \approx 100$ nM) with a long tail of highly potent compounds extending to $\mathrm{pIC}_{50} > 10$ (sub-nanomolar), reflecting decades of lead optimization against this target.

Table 1: Compound accounting through the curation pipeline for EGFR (CHEMBL203), ChEMBL_37.

Stage	Count
Raw IC_{50} activity records (defined `pChEMBL`)	19,378
Unique input SMILES	10,734
Records dropped (missing/invalid structure)	−17
Records after structure filtering	19,361
Compounds with replicate measurements (mean-aggregated)	4,163
Unique curated compounds (final modeling set)	**10,667**
Training partition (80%)	8,534
Test partition (20%)	2,133

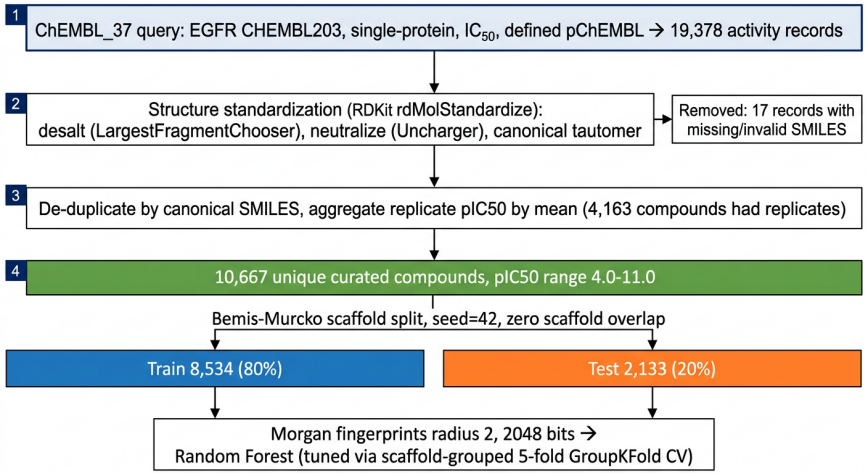
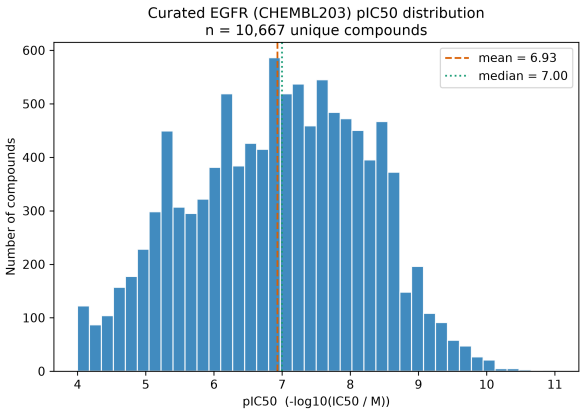
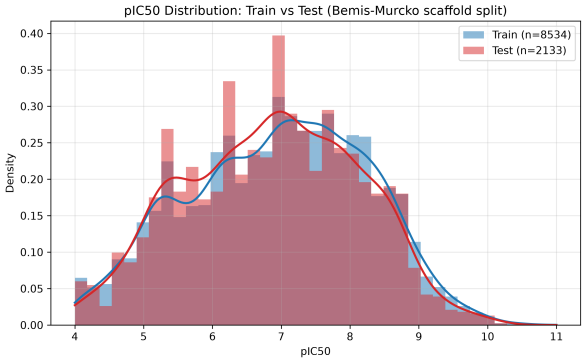


Figure 1: Data curation and modeling pipeline with per-stage compound counts. Raw ChEMBL_37 records are standardized (desalt, neutralize, canonical tautomer), de-duplicated by mean pIC_{50}, split 80/20 by Bemis–Murcko scaffold (seed 42; zero scaffold overlap), and modeled with 2048-bit Morgan fingerprints and a scaffold-CV-tuned random forest.



(a) Curated pIC_{50} distribution (10,667 compounds).



(b) Train vs. test pIC_{50} after scaffold split.

Figure 2: Potency distributions. (a) The curated dataset is centered near pIC_{50} 7 with a potent tail beyond pIC_{50} 10. (b) The 80/20 Bemis–Murcko scaffold split preserves closely matched pIC_{50} distributions across the (scaffold-disjoint) training and test partitions.

3.2 Model selection

Under scaffold-grouped 5-fold cross-validation, the random forest slightly but consistently outperformed XGBoost on the primary criterion of mean CV RMSE (Table 2, Figure 3). The random forest was therefore selected and retrained on the full training set with its best hyperparameters (`n_estimators`= 200, `max_features`= 0.3, `max_depth`=None, `min_samples_split`= 2, `min_samples_leaf`= 1). The narrow cross-validation standard deviations (±0.035 RMSE) indicate stable performance across scaffold folds.

Table 2: Scaffold-grouped 5-fold cross-validation performance (mean ± SD across folds) for the two tuned candidate models. The random forest was selected on lowest mean CV RMSE.

Model	CV RMSE	CV MAE	CV R^2	CV Spearman ρ
Random forest (selected)	$\mathbf{0.827 \pm 0.035}$	0.625 ± 0.033	0.592 ± 0.049	0.771 ± 0.039
XGBoost	0.838 ± 0.034	0.638 ± 0.033	0.581 ± 0.048	0.765 ± 0.041

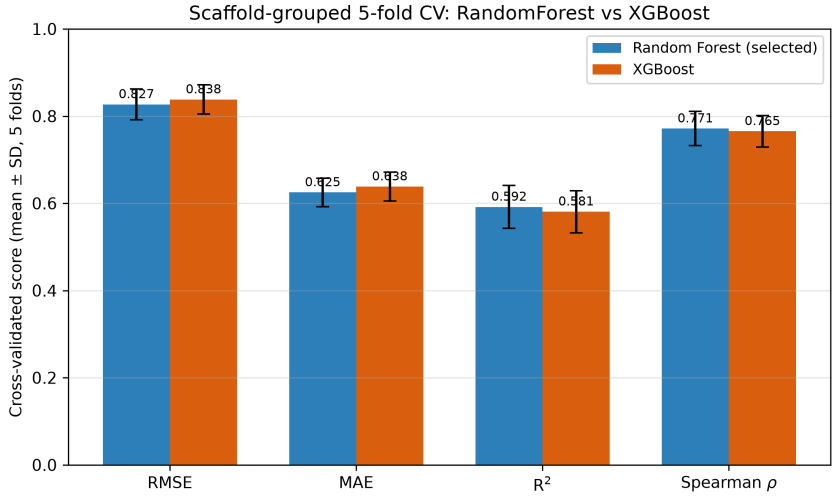


Figure 3: Scaffold-grouped 5-fold cross-validation comparison of the random forest and XGBoost candidates across all four metrics (mean ± SD). The random forest is marginally superior on RMSE, MAE, R^2, and Spearman ρ.

3.3 Held-out test performance

On the **2,133-compound scaffold-disjoint test set**, the selected random forest achieved the performance summarized in Table 3. These held-out values are actually slightly better than the cross-validation estimates—consistent with the final model benefiting from the full training set—and reproduced the model-selection-phase numbers exactly upon independent re-featurization.

Figure 4 shows the predicted-versus-measured scatter for the full test set. Points cluster around the ideal $y = x$ line across the whole potency range, and the fitted ordinary-least-squares trend line has slope 0.64 and intercept 2.50—the signature of *regression-to-the-mean shrinkage*, whereby the least potent compounds are somewhat over-predicted and the most potent compounds somewhat under-predicted (analyzed further in Section 4). Within ±0.5 pIC_{50} units of the measured value lie 55.9% of test compounds; within ±1.0 unit, 83.1%; and within ±1.5 units, 94.4%.

Table 3: Held-out test-set performance of the selected random-forest QSAR model (n = 2,133 scaffold-split compounds). RMSE and MAE are in pIC_{50} units.

Metric	**Value**
RMSE	0.746
MAE	0.561
R^2	0.646
Spearman ρ	0.808
Compounds evaluated	2,133
Unparseable SMILES	0
Residual mean $\pm$ SD	-0.027 ± 0.745
Residual range	$[-3.12, +3.02]$

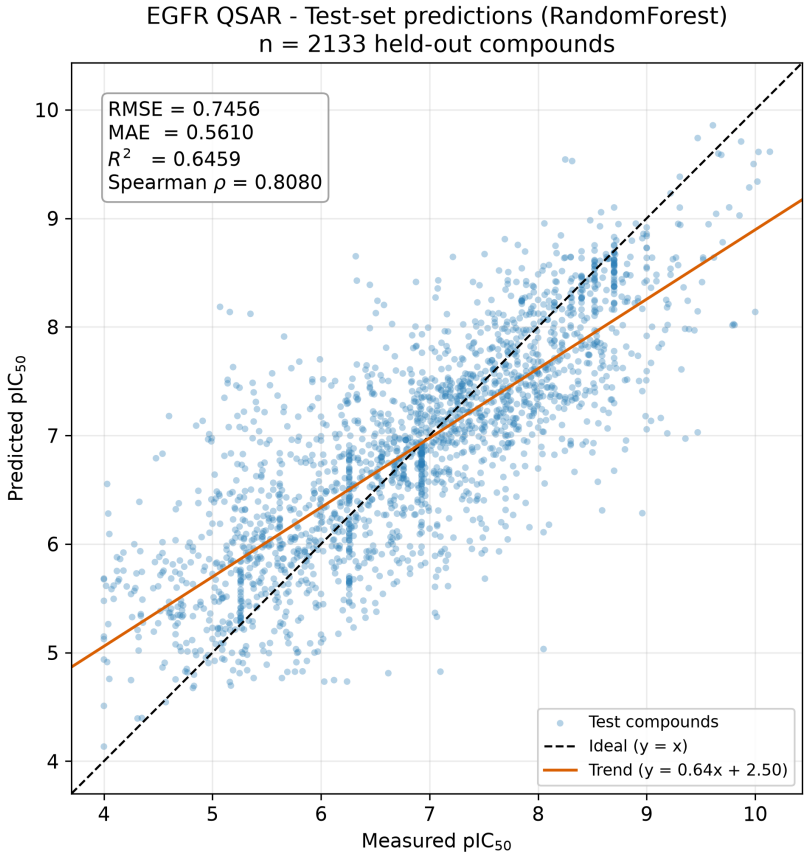


Figure 4: Predicted vs. measured pIC_{50} for the 2,133 held-out test compounds. The dashed line is the ideal $y = x$; the orange line is the ordinary-least-squares trend (slope 0.64). The inset reports RMSE = 0.7456, MAE = 0.5610, R^2 = 0.6459, and Spearman ρ = 0.8080. The compression of predictions toward the mean at the extremes is visible as the trend line crossing below $y = x$ at high potency.

3.4 Ten most potent held-out compounds

Table 4 lists the ten most potent test compounds by *measured* pIC_{50}, together with their predicted pIC_{50} and residual. All ten are sub-nanomolar ($pIC_{50} \geq 9.8$). The model recovers the extreme potency of this tail reasonably well: seven of the ten are predicted within $\sim$0.8 pIC_{50} units, and six share the same diarylaminopyrimidine scaffold (`c1cc(Nc2ccc3nccnc3c2)nc(Nc2ccc(N3CCC(N4CCC4)CC3)cc2)n1`), which is represented in the training set and hence well characterized. The two largest under-predictions—CHEMBL5196360 (residual +1.86) and CHEMBL5274692 (residual +1.79)—sit on distinct scaffolds and are discussed in Section 4.

Table 4: The ten most potent held-out (test-set) compounds ranked by measured pIC_{50}, with model-predicted pIC_{50} and residual (measured − predicted). All values on the pIC_{50} scale.

ChEMBL ID	Measured pIC_{50}	Predicted pIC_{50}	Residual
CHEMBL5186680	10.135	9.613	+0.522
CHEMBL5180388	10.030	9.612	+0.418
CHEMBL5393923	10.020	9.338	+0.682
CHEMBL5196360	10.000	8.138	+1.862
CHEMBL5399653	9.985	9.500	+0.485
CHEMBL5415311	9.945	9.284	+0.661
CHEMBL5193149	9.870	9.708	+0.162
CHEMBL5432965	9.855	9.026	+0.829
CHEMBL5274692	9.810	8.017	+1.793
CHEMBL5572504	9.800	8.025	+1.775

3.5 Full per-compound test-set table

The complete predicted-versus-measured table for all 2,133 held-out compounds is provided in Appendix A (Table 5), sorted by measured pIC_{50}. The identical data, including canonical SMILES and scaffold assignments, are available in machine-readable form as `results/test_predictions.csv`.

3.6 Residual diagnostics

Figure 5 characterizes the error structure. The residual distribution (right panel) is approximately symmetric and centered near zero (mean −0.027), confirming the absence of a global prediction bias. The residual-versus-measured plot (left panel) shows a clear positive trend (slope 0.36): residuals are negative (over-prediction) for weak compounds and positive (under-prediction) for potent compounds. Quantitatively, the mean residual rises from +0.35 for compounds with measured $pIC_{50} \geq 7$ to +0.54 for $pIC_{50} \geq 8$ and +0.88 for $pIC_{50} \geq 9$. This is the expected finite-data shrinkage of a tree-ensemble regressor rather than a coding artifact, and it directly explains the sub-unit slope in Figure 4.

4 Discussion

4.1 Overall model quality

The selected random forest explains roughly two-thirds of the held-out pIC_{50} variance ($R^2 = 0.646$) with a typical error of about 0.75 log units (RMSE) and a median-like absolute error of 0.56 log units (MAE)—i.e. predictions are on average within a factor of $\sim$3.6 in IC_{50}, and within

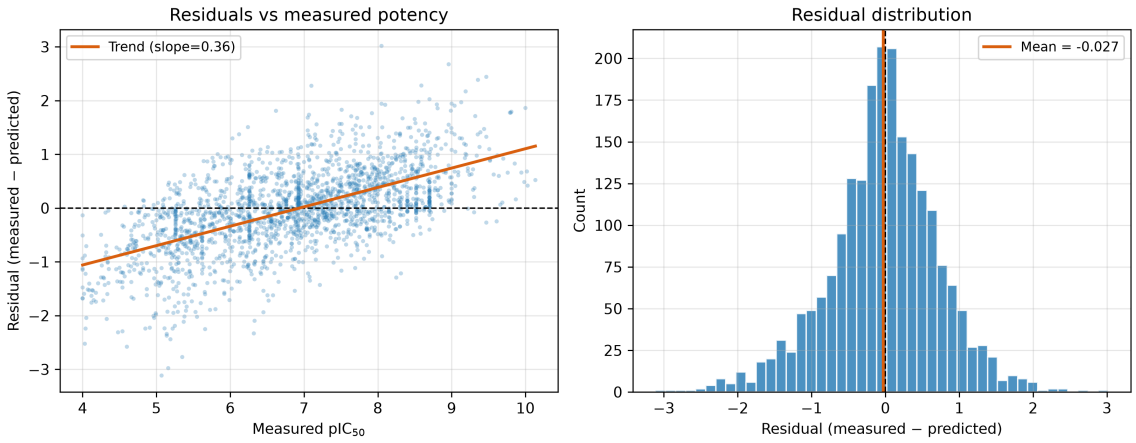


Figure 5: Residual diagnostics on the held-out test set. *Left:* residuals (measured − predicted) increase with measured potency (trend slope 0.36), the hallmark of regression-to-the-mean shrinkage. *Right:* the residual distribution is approximately symmetric and centered near zero (mean −0.027), indicating no systematic global bias.

a factor of ∼3 for the majority of compounds. Crucially, this was obtained under a *scaffold-disjoint* split, so the metrics reflect genuine extrapolation to chemotypes unseen in training rather than interpolation among near-duplicates. Random splits on the same data would almost certainly yield inflated scores, because analogues of test compounds would appear in training [11, 12]; the values here are therefore a conservative and honest estimate of prospective performance. The result is consistent with the well-documented ceiling on single-target Morgan-fingerprint QSAR under realistic splitting and with the broader QSAR-modeling literature [7, 8].

4.2 High rank correlation versus regression-to-the-mean shrinkage

The most striking feature of the results is the contrast between the high Spearman rank correlation ($\rho = 0.808$) and the shrunken regression slope (0.64). These two observations are complementary rather than contradictory. Spearman ρ depends only on the *ordering* of compounds, and a value of 0.808 means the model orders held-out compounds by potency very reliably. The sub-unit slope, by contrast, is a statement about *calibration* of absolute values: a random forest predicts by averaging the labels of training compounds in each leaf, so its outputs are bounded by the training-label range and are pulled toward the training mean, compressing the predicted dynamic range relative to the true one. This regression-to-the-mean effect is intrinsic to averaging estimators fitted on finite, noisy data and is amplified at the sparsely populated extremes of the pIC_{50} distribution. The residual analysis (Figure 5) makes the mechanism explicit: the shrinkage grows monotonically with potency, from +0.35 at $\mathrm{pIC}_{50} \geq 7$ to +0.88 at $\mathrm{pIC}_{50} \geq 9$.

The practical implication is favorable for the most common QSAR use cases. For virtual screening, hit triage, and library prioritization, what matters is whether the model ranks the truly potent compounds above the weak ones—precisely what a Spearman ρ of 0.808 certifies. Ranking-oriented metrics are increasingly recognized as more decision-relevant than absolute-error metrics for prospective screening [8]. Where absolute potency estimates are required (for example to set an assay concentration or to compare against a hard pIC_{50} threshold), the systematic under-prediction of the potent tail should be kept in mind: the model will tend to *underestimate* the potency of the best compounds, which is a conservative bias for triage but should not be read as a hard ceiling on true activity. Post-hoc calibration (e.g. isotonic or linear rescaling of predictions against a calibration set) could partially restore the dynamic range if absolute values are the priority.

4.3 Under-predicted potent compounds on novel scaffolds

The largest positive residuals concentrate among potent compounds whose scaffolds are poorly represented in (or absent from) the training set—the expected behavior when a model is asked to extrapolate beyond its applicability domain [23, 8]. Two examples from the top-ten table illustrate the pattern. CHEMBL5196360 (measured pIC_{50} 10.0, predicted 8.14; residual +1.86) is built on a benzo-fused imidazolone/quinazolinone-type framework distinct from the diarylaminopyrimidine series that dominates the potent training compounds, and CHEMBL5274692 (measured 9.81, predicted 8.02; residual +1.79) is a quinazoline bearing an acrylamide warhead and an unusual tetrahydrofuranyloxy substituent. Because the random forest cannot assign potency contributions to substructural motifs it has rarely encountered, it defaults toward the potency of the nearest well-populated neighborhoods and under-predicts these genuinely novel, highly potent chemotypes. More broadly across the whole test set, the most severe individual errors reflect the same phenomenon in both directions, including a small number of apparent activity cliffs—pairs of structurally similar molecules with large potency differences that are intrinsically difficult for any similarity-based model to resolve [24].

This behavior is not a defect to be "fixed" so much as a property to be *managed*. It argues for pairing the point prediction with an explicit applicability-domain assessment so that out-of-domain, potentially under-predicted compounds are flagged rather than silently down-ranked. A natural and inexpensive uncertainty signal is the variance of predictions across the individual trees of the random forest, which tends to be elevated precisely for compounds far from the training distribution; surfacing such flags would let a screening campaign treat high-uncertainty potent predictions as candidates for experimental follow-up rather than discarding them.

4.4 Limitations and future work

Several limitations temper the interpretation. First, IC_{50} values aggregated across heterogeneous assays, cell lines, and EGFR constructs (wild-type versus mutant) carry inter-laboratory variability of roughly 0.5–1.0 log units, which places a floor under the achievable RMSE; mean aggregation reduces but cannot eliminate this noise [15]. Second, the model uses a single 2D fingerprint representation; richer descriptors (physicochemical properties, 3D or pharmacophoric features) or modern graph-neural representations might extend the applicability domain, although scaffold-split gains from such methods are often modest [12]. Third, the model is target-agnostic to EGFR mutation status, whereas clinically the wild-type/mutant selectivity axis is central; a mutation-stratified or multi-task model would be a valuable extension. Finally, the honest but conservative scaffold split means these numbers should be read as a lower bound relative to interpolative deployment on chemistry close to the training set. Future work should add an explicit applicability-domain module and prediction-uncertainty estimates, explore calibration to recover dynamic range, and consider multi-task learning across the ErbB family and EGFR mutants.

4.5 Conclusion

We have documented a complete, reproducible QSAR workflow for human EGFR (CHEMBL203) on the ChEMBL_37 release: 19,378 IC_{50} records curated to **10,667 unique compounds**, split 80/20 by Bemis–Murcko scaffold (seed 42) into scaffold-disjoint training and test sets, featurized as 2048-bit radius-2 Morgan fingerprints, and modeled with a scaffold-CV-tuned random forest selected over XGBoost. On a genuinely held-out, scaffold-disjoint test set of 2,133 compounds the model attained **RMSE** 0.746**, MAE** 0.561**,** $R^2 = 0.646$**, and Spearman** $\rho = 0.808$. The high rank correlation makes the model well-suited to ranking and triage, while the regression-to-the-mean shrinkage of the potent tail—most

pronounced for highly potent compounds on novel scaffolds—defines the regime in which absolute-potency predictions should be treated with appropriate caution and, ideally, an explicit applicability-domain flag.

Data and Code Availability

Curated data, the trained model bundle, cross-validation results, per-compound predictions, and all figures are provided with this report. Key artifacts: `data/curated_egfr_bioactivity.csv` (10,667 compounds), `data/train_egfr.csv` / `data/test_egfr.csv` (scaffold split), `results/best_egfr_model.joblib` (random forest bundle with fingerprint spec and versions), `results/cv_results.json`, `results/test_predictions.csv` (full test-set table), and `results/top_10_potent_test.csv`. Software: Python 3.12.10, RDKit 2026.03.3, scikit-learn 1.9.0, xgboost 3.3.0, numpy 2.5.1. Random seed 42 throughout.

References

[1] Mark A. Lemmon and Joseph Schlessinger. Cell signaling by receptor tyrosine kinases. *Cell*, 141(7):1117–1134, 2010.

[2] Thomas J. Lynch, Daphne W. Bell, Raffaella Sordella, Sarada Gurubhagavatula, Ross A. Okimoto, Brian W. Brannigan, Patricia L. Harris, Sara M. Haserlat, Jeffrey G. Supko, Frank G. Haluska, David N. Louis, David C. Christiani, Jeff Settleman, and Daniel A. Haber. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *New England Journal of Medicine*, 350(21):2129–2139, 2004.

[3] J. Guillermo Paez, Pasi A. Jänne, Jeffrey C. Lee, Sean Tracy, Heidi Greulich, Stacey Gabriel, Paula Herman, Frederic J. Kaye, Neal Lindeman, Titus J. Boggon, Katsuhiko Naoki, Hidefumi Sasaki, Yoshitaka Fujii, Michael J. Eck, William R. Sellers, Bruce E. Johnson, and Matthew Meyerson. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science*, 304(5676):1497–1500, 2004.

[4] William Pao, Vincent A. Miller, Katerina A. Politi, Gregory J. Riely, Romel Somwar, Maureen F. Zakowski, Mark G. Kris, and Harold Varmus. Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain. *PLoS Medicine*, 2(3):e73, 2005.

[5] Tony S. Mok, Yi-Long Wu, Myung-Ju Ahn, Marina C. Garassino, Hye Ryun Kim, Suresh S. Ramalingam, Frances A. Shepherd, Yong He, Hiroaki Akamatsu, Willemijn S. M. E. Theelen, Chee Khoon Lee, Martin Sebastian, Alison Templeton, Helen Mann, Marcelo Marotti, Serban Ghiorghiu, and Vassiliki A. Papadimitrakopoulou. Osimertinib or platinum–pemetrexed in EGFR T790M-positive lung cancer. *New England Journal of Medicine*, 376(7):629–640, 2017.

[6] Jean-Charles Soria, Yuichiro Ohe, Johan Vansteenkiste, Thanyanan Reungwetwattana, Busyamas Chewaskulyong, Ki Hyeong Lee, Arunee Dechaphunkul, Fumio Imamura, Naoyuki Nogami, Takayasu Kurata, Isamu Okamoto, Caicun Zhou, Byoung Chul Cho, Ying Cheng, Eun Kyung Cho, Pei Jye Voon, David Planchard, Wu-Chou Su, Jhanelle E. Gray, Sang-We Lee, Rachel Hodge, Marcelo Marotti, Yuri Rukazenkov, and Suresh S. Ramalingam. Osimertinib in untreated EGFR-mutated advanced non-small-cell lung cancer. *New England Journal of Medicine*, 378(2):113–125, 2018.

[7] Artem Cherkasov, Eugene N. Muratov, Denis Fourches, Alexandre Varnek, Igor I. Baskin, Mark Cronin, John Dearden, Paola Gramatica, Yvonne C. Martin, Roberto Todeschini, Viviana Consonni, Victor E. Kuz'min, Richard Cramer, Romualdo Benigni, Chihae Yang, James Rathman, Lothar Terfloth, Johann Gasteiger, Ann Richard, and Alexander Tropsha. QSAR modeling: where have you been? where are you going to? *Journal of Medicinal Chemistry*, 57(12):4977–5010, 2014.

[8] Eugene N. Muratov, Jürgen Bajorath, Robert P. Sheridan, Igor V. Tetko, Dmitry Filimonov, Vladimir Poroikov, Tudor I. Oprea, Igor I. Baskin, Alexandre Varnek, Adrian Roitberg, Olexandr Isayev, Stefano Curtalolo, Denis Fourches, Yoram Cohen, Alan Aspuru-Guzik, David A. Winkler, Dimitris Agrafiotis, Artem Cherkasov, and Alexander Tropsha. QSAR without borders. *Chemical Society Reviews*, 49(11):3525–3564, 2020.

[9] Alexander Tropsha. Best practices for QSAR model development, validation, and exploitation. *Molecular Informatics*, 29(6-7):476–488, 2010.

[10] Denis Fourches, Eugene Muratov, and Alexander Tropsha. Trust, but verify: on the importance of chemical structure curation in cheminformatics and QSAR modeling research. *Journal of Chemical Information and Modeling*, 50(7):1189–1204, 2010.

[11] Robert P. Sheridan. Time-split cross-validation as a method for estimating the goodness of prospective prediction. *Journal of Chemical Information and Modeling*, 53(4):783–790, 2013.

[12] Zhenqin Wu, Bharath Ramsundar, Evan N. Feinberg, Joseph Gomes, Caleb Geniesse, Aneesh S. Pappu, Karl Leswing, and Vijay Pande. MoleculeNet: a benchmark for molecular machine learning. *Chemical Science*, 9(2):513–530, 2018.

[13] Barbara Zdrazil, Eloy Felix, Fiona Hunter, Emma J. Manners, James Blackshaw, Sybilla Corbett, Marleen de Veij, Harris Ioannidis, David Mendez Lopez, Juan F. Mosquera, María Paula Magariños, Nicolas Bosc, Ricardo Arcila, Tevfik Kizilören, Anna Gaulton, A. Patrícia Bento, Melissa F. Adasme, Peter Monecke, Gregory A. Landrum, and Andrew R. Leach. The ChEMBL database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. *Nucleic Acids Research*, 52(D1):D1180–D1192, 2024.

[14] Anna Gaulton, Louisa J. Bellis, A. Patrícia Bento, Jon Chambers, Mark Davies, Anne Hersey, Yvonne Light, Shaun McGlinchey, David Michalovich, Bissan Al-Lazikani, and John P. Overington. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Research*, 40(D1):D1100–D1107, 2012.

[15] Denis Fourches, Eugene Muratov, and Alexander Tropsha. Trust, but verify II: a practical guide to chemogenomics data curation. *Journal of Chemical Information and Modeling*, 56(7):1243–1252, 2016.

[16] Greg Landrum and The RDKit Contributors. RDKit: Open-source cheminformatics software. https://www.rdkit.org, 2024. Version 2026.03.3.

[17] Guy W. Bemis and Mark A. Murcko. The properties of known drugs. 1. molecular frameworks. *Journal of Medicinal Chemistry*, 39(15):2887–2893, 1996.

[18] H. L. Morgan. The generation of a unique machine description for chemical structures—a technique developed at chemical abstracts service. *Journal of Chemical Documentation*, 5(2):107–113, 1965.

[19] David Rogers and Mathew Hahn. Extended-connectivity fingerprints. *Journal of Chemical Information and Modeling*, 50(5):742–754, 2010.

[20] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[21] Tianqi Chen and Carlos Guestrin. XGBoost: a scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*, pages 785–794, 2016.

[22] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[23] Sabcho Dimitrov, Gergana Dimitrova, Todor Pavlov, Nadezhda Dimitrova, Grace Patlewicz, Jay Niemela, and Ovanes Mekenyan. A stepwise approach for defining the applicability domain of SAR and QSAR models. *Journal of Chemical Information and Modeling*, 45(4):839–849, 2005.

[24] Dagmar Stumpfe, Huabin Hu, and Jürgen Bajorath. Advances in exploring activity cliffs. *Journal of Computer-Aided Molecular Design*, 34(9):929–942, 2020.

A Full per-compound test-set prediction table

Table 5 lists all 2,133 held-out (scaffold-split) test compounds, sorted by measured pIC_{50} in descending order, with the random-forest predicted pIC_{50} and the residual (measured − predicted). Canonical SMILES and scaffold assignments for every row are available in `results/test_predictions.csv`.

Table 5: Full per-compound predicted-vs-measured table for the held-out test set (all 2,133 scaffold-split compounds), sorted by measured pIC_{50} (descending). Residual = measured − predicted.

#	ChEMBL ID	Measured	Predicted	Residual
1	CHEMBL5186680	10.135	9.613	+0.522
2	CHEMBL5180388	10.030	9.612	+0.418
3	CHEMBL5393923	10.020	9.338	+0.682
4	CHEMBL5196360	10.000	8.138	+1.862
5	CHEMBL5399653	9.985	9.500	+0.485
6	CHEMBL5415311	9.945	9.284	+0.661
7	CHEMBL5193149	9.870	9.708	+0.162
8	CHEMBL5432965	9.855	9.026	+0.829
9	CHEMBL5274692	9.810	8.017	+1.793
10	CHEMBL5572504	9.800	8.025	+1.775
11	CHEMBL6188493	9.790	8.014	+1.776
12	CHEMBL5396086	9.760	9.101	+0.659
13	CHEMBL5206902	9.700	8.780	+0.920
14	CHEMBL5434495	9.688	9.585	+0.103
15	CHEMBL4560430	9.660	9.596	+0.064
16	CHEMBL6142125	9.640	9.042	+0.598
17	CHEMBL5280967	9.620	8.639	+0.981
18	CHEMBL5203220	9.610	9.857	-0.247
19	CHEMBL2031298	9.570	8.945	+0.625
20	CHEMBL3357644	9.520	8.639	+0.881
21	CHEMBL5191510	9.520	8.620	+0.900
22	CHEMBL5825778	9.510	8.413	+1.097
23	CHEMBL5268781	9.470	7.029	+2.441
24	CHEMBL5204749	9.470	9.739	-0.269
25	CHEMBL3965262	9.470	8.143	+1.327
26	CHEMBL603266	9.440	7.968	+1.472
27	CHEMBL5271930	9.390	7.979	+1.411
28	CHEMBL5284924	9.378	8.241	+1.137
29	CHEMBL5220645	9.357	6.971	+2.385
30	CHEMBL5896727	9.353	7.983	+1.370
31	CHEMBL3906546	9.340	8.364	+0.976
32	CHEMBL516022	9.320	7.608	+1.712
33	CHEMBL5831921	9.310	8.939	+0.371
34	CHEMBL5415418	9.305	9.099	+0.206
35	CHEMBL5416489	9.305	9.383	-0.078
36	CHEMBL272865	9.290	8.569	+0.721
37	CHEMBL2031304	9.290	8.949	+0.341
38	CHEMBL5282641	9.270	8.639	+0.631
39	CHEMBL2029438	9.242	8.624	+0.618
40	CHEMBL4749862	9.220	9.231	-0.011
41	CHEMBL4741922	9.220	8.480	+0.740
42	CHEMBL5177272	9.220	8.535	+0.685
43	CHEMBL3890068	9.200	8.211	+0.989
44	CHEMBL5285052	9.193	8.018	+1.176
45	CHEMBL5180251	9.190	7.019	+2.171
46	CHEMBL5288888	9.188	8.180	+1.008
47	CHEMBL3954175	9.180	8.249	+0.931
48	CHEMBL6192014	9.170	7.972	+1.198

continued on next page

Table 5 – *continued from previous page*

#	ChEMBL ID	Measured	Predicted	Residual
49	CHEMBL474141	9.140	7.844	+1.296
50	CHEMBL5208907	9.130	7.161	+1.969
51	CHEMBL516373	9.110	7.810	+1.300
52	CHEMBL5281741	9.090	8.480	+0.610
53	CHEMBL5277739	9.090	8.421	+0.669
54	CHEMBL5177306	9.050	8.638	+0.412
55	CHEMBL5411886	9.050	8.001	+1.049
56	CHEMBL272187	9.040	8.360	+0.680
57	CHEMBL5182567	9.040	7.134	+1.906
58	CHEMBL6145524	9.040	7.668	+1.372
59	CHEMBL5964757	9.000	8.857	+0.143
60	CHEMBL5898177	9.000	8.726	+0.274
61	CHEMBL5786862	9.000	8.837	+0.163
62	CHEMBL5806859	9.000	8.916	+0.084
63	CHEMBL5873371	9.000	8.636	+0.364
64	CHEMBL5982314	9.000	8.426	+0.574
65	CHEMBL5839834	9.000	8.551	+0.449
66	CHEMBL4575583	9.000	7.057	+1.943
67	CHEMBL5973924	9.000	8.539	+0.461
68	CHEMBL6015239	9.000	8.729	+0.271
69	CHEMBL310853	9.000	7.917	+1.083
70	CHEMBL79215	9.000	7.592	+1.408
71	CHEMBL5755058	8.997	7.508	+1.489
72	CHEMBL5591885	8.990	7.178	+1.812
73	CHEMBL5280664	8.980	8.409	+0.571
74	CHEMBL5569244	8.970	8.857	+0.113
75	CHEMBL3623284	8.960	8.167	+0.793
76	CHEMBL3655337	8.960	8.817	+0.143
77	CHEMBL3960853	8.960	7.408	+1.552
78	CHEMBL4099329	8.960	6.284	+2.676
79	CHEMBL6143027	8.930	7.502	+1.428
80	CHEMBL3908905	8.920	8.235	+0.685
81	CHEMBL3663936	8.920	8.265	+0.655
82	CHEMBL6160966	8.920	7.734	+1.186
83	CHEMBL5566473	8.900	8.952	-0.052
84	CHEMBL3663935	8.890	8.219	+0.671
85	CHEMBL2029437	8.864	8.794	+0.070
86	CHEMBL6103055	8.860	7.754	+1.106
87	CHEMBL5178571	8.850	8.475	+0.375
88	CHEMBL5817731	8.850	8.645	+0.205
89	CHEMBL5905270	8.825	8.266	+0.559
90	CHEMBL489897	8.820	8.580	+0.240
91	CHEMBL3623283	8.820	7.515	+1.305
92	CHEMBL3623279	8.820	8.234	+0.586
93	CHEMBL6145609	8.815	8.484	+0.331
94	CHEMBL5594833	8.812	7.819	+0.993
95	CHEMBL5220529	8.810	6.953	+1.857
96	CHEMBL5283507	8.810	8.480	+0.330
97	CHEMBL5199341	8.800	6.872	+1.928
98	CHEMBL5846813	8.800	8.231	+0.569
99	CHEMBL6146710	8.800	7.509	+1.291
100	CHEMBL5266649	8.780	7.621	+1.159
101	CHEMBL6189765	8.780	8.016	+0.764
102	CHEMBL5281205	8.780	8.409	+0.371
103	CHEMBL516365	8.770	8.439	+0.331
104	CHEMBL5979002	8.770	8.279	+0.491
105	CHEMBL5176460	8.740	8.947	-0.207
106	CHEMBL4860398	8.720	8.797	-0.077
107	CHEMBL2048903	8.720	7.619	+1.101

continued on next page

Table 5 – *continued from previous page*

#	ChEMBL ID	Measured	Predicted	Residual
108	CHEMBL5985406	8.720	7.948	+0.772
109	CHEMBL205059	8.710	8.667	+0.043
110	CHEMBL3818563	8.705	7.663	+1.042
111	CHEMBL5755992	8.700	8.401	+0.299
112	CHEMBL5889876	8.700	8.376	+0.324
113	CHEMBL5997382	8.700	8.370	+0.330
114	CHEMBL6045539	8.700	8.582	+0.118
115	CHEMBL5906294	8.700	8.616	+0.084
116	CHEMBL5817509	8.700	8.567	+0.133
117	CHEMBL6022365	8.700	8.500	+0.200
118	CHEMBL5904771	8.700	8.304	+0.396
119	CHEMBL5860718	8.700	8.372	+0.328
120	CHEMBL5916523	8.700	8.509	+0.191
121	CHEMBL5772446	8.700	8.211	+0.489
122	CHEMBL5837504	8.700	8.616	+0.084
123	CHEMBL6017472	8.700	8.828	-0.128
124	CHEMBL5763901	8.700	8.462	+0.238
125	CHEMBL5990005	8.700	8.549	+0.151
126	CHEMBL492077	8.700	8.568	+0.132
127	CHEMBL6049695	8.700	7.777	+0.923
128	CHEMBL2419760	8.700	7.298	+1.402
129	CHEMBL121954	8.700	6.675	+2.025
130	CHEMBL5814763	8.700	8.159	+0.541
131	CHEMBL5987649	8.700	8.755	-0.055
132	CHEMBL5750351	8.700	8.222	+0.478
133	CHEMBL587723	8.700	8.505	+0.195
134	CHEMBL5962683	8.700	8.624	+0.076
135	CHEMBL5975352	8.700	8.574	+0.126
136	CHEMBL6033348	8.700	8.498	+0.202
137	CHEMBL5968820	8.700	8.435	+0.265
138	CHEMBL6023899	8.700	8.337	+0.363
139	CHEMBL5964384	8.700	8.218	+0.482
140	CHEMBL6013771	8.700	8.686	+0.014
141	CHEMBL6050257	8.700	8.677	+0.023
142	CHEMBL6007318	8.700	8.588	+0.112
143	CHEMBL5935427	8.700	8.561	+0.139
144	CHEMBL6050763	8.700	8.593	+0.107
145	CHEMBL5879075	8.700	8.371	+0.329
146	CHEMBL5916200	8.700	8.514	+0.186
147	CHEMBL5907670	8.700	8.414	+0.286
148	CHEMBL6039752	8.700	7.713	+0.987
149	CHEMBL5985452	8.700	8.611	+0.089
150	CHEMBL4780965	8.700	9.100	-0.400
151	CHEMBL5747811	8.700	8.685	+0.015
152	CHEMBL6008076	8.700	8.746	-0.046
153	CHEMBL6142200	8.680	7.442	+1.238
154	CHEMBL5436254	8.680	8.162	+0.518
155	CHEMBL5982062	8.680	7.986	+0.694
156	CHEMBL5406587	8.680	7.331	+1.349
157	CHEMBL2425736	8.670	7.330	+1.340
158	CHEMBL2031308	8.660	7.624	+1.036
159	CHEMBL312818	8.650	7.208	+1.442
160	CHEMBL5563672	8.650	8.629	+0.021
161	CHEMBL6142761	8.650	8.629	+0.021
162	CHEMBL4060996	8.640	7.065	+1.575
163	CHEMBL4130016	8.635	8.625	+0.010
164	CHEMBL6091887	8.630	7.546	+1.084
165	CHEMBL5969707	8.630	7.446	+1.184
166	CHEMBL4282543	8.630	8.339	+0.291

continued on next page

Table 5 – *continued from previous page*

#	ChEMBL ID	Measured	Predicted	Residual
167	CHEMBL271328	8.630	8.643	-0.013
168	CHEMBL5286636	8.625	8.670	-0.045
169	CHEMBL5780873	8.620	8.238	+0.382
170	CHEMBL5871844	8.610	8.521	+0.089
171	CHEMBL4060039	8.610	7.828	+0.782
172	CHEMBL296168	8.605	7.792	+0.813
173	CHEMBL5407489	8.600	7.644	+0.956
174	CHEMBL4460381	8.600	7.237	+1.363
175	CHEMBL5289304	8.595	8.453	+0.142
176	CHEMBL4099682	8.590	6.310	+2.280
177	CHEMBL6013902	8.590	7.510	+1.080
178	CHEMBL5934619	8.590	7.735	+0.855
179	CHEMBL2048912	8.590	7.681	+0.909
180	CHEMBL6144893	8.580	7.769	+0.811
181	CHEMBL5397801	8.580	9.056	-0.476
182	CHEMBL6146123	8.575	7.849	+0.726
183	CHEMBL6003270	8.575	7.621	+0.954
184	CHEMBL510597	8.570	8.609	-0.039
185	CHEMBL5425947	8.570	8.927	-0.357
186	CHEMBL3623285	8.570	8.167	+0.403
187	CHEMBL4862351	8.555	7.112	+1.443
188	CHEMBL1630112	8.550	7.815	+0.735
189	CHEMBL5940893	8.550	8.389	+0.161
190	CHEMBL4067871	8.550	6.573	+1.977
191	CHEMBL6163318	8.545	7.864	+0.681
192	CHEMBL3633146	8.540	7.960	+0.580
193	CHEMBL6191894	8.540	7.915	+0.625
194	CHEMBL2048906	8.540	7.863	+0.677
195	CHEMBL138940	8.533	8.134	+0.399
196	CHEMBL5858630	8.527	8.352	+0.174
197	CHEMBL5993712	8.523	7.399	+1.125
198	CHEMBL6041859	8.520	8.421	+0.099
199	CHEMBL6046237	8.520	8.615	-0.095
200	CHEMBL5770747	8.520	8.562	-0.042
201	CHEMBL5896195	8.520	8.531	-0.011
202	CHEMBL5755363	8.520	8.486	+0.034
203	CHEMBL5793602	8.520	8.479	+0.041
204	CHEMBL6013119	8.520	8.535	-0.015
205	CHEMBL5913125	8.520	8.597	-0.077
206	CHEMBL5762432	8.520	8.572	-0.052
207	CHEMBL5975926	8.520	8.439	+0.081
208	CHEMBL5812935	8.520	8.609	-0.089
209	CHEMBL5933788	8.520	7.828	+0.692
210	CHEMBL5824770	8.520	8.452	+0.068
211	CHEMBL474431	8.520	7.968	+0.552
212	CHEMBL5771859	8.520	8.102	+0.418
213	CHEMBL5988402	8.520	8.236	+0.284
214	CHEMBL6028869	8.520	8.184	+0.336
215	CHEMBL6039698	8.520	8.247	+0.273
216	CHEMBL5785533	8.520	8.751	-0.231
217	CHEMBL5932750	8.520	8.370	+0.150
218	CHEMBL5828286	8.520	8.431	+0.089
219	CHEMBL5913601	8.520	8.562	-0.042
220	CHEMBL5809880	8.520	8.154	+0.366
221	CHEMBL6044353	8.520	8.703	-0.183
222	CHEMBL1683967	8.520	7.572	+0.948
223	CHEMBL6015820	8.520	8.142	+0.378
224	CHEMBL2048796	8.520	7.442	+1.078
225	CHEMBL1934623	8.520	7.281	+1.239

continued on next page

Table 5 – *continued from previous page*

#	ChEMBL ID	Measured	Predicted	Residual
226	CHEMBL258940	8.520	7.846	+0.674
227	CHEMBL1243254	8.510	6.929	+1.581
228	CHEMBL5980501	8.510	7.837	+0.673
229	CHEMBL2048797	8.510	7.479	+1.031
230	CHEMBL5209046	8.500	6.844	+1.656
231	CHEMBL5087610	8.480	8.982	-0.502
232	CHEMBL5964366	8.480	8.307	+0.173
233	CHEMBL517127	8.480	7.878	+0.602
234	CHEMBL6023158	8.473	7.409	+1.065
235	CHEMBL6151439	8.470	7.740	+0.730
236	CHEMBL5178865	8.470	8.438	+0.032
237	CHEMBL6143935	8.470	7.744	+0.726
238	CHEMBL5412305	8.470	7.664	+0.806
239	CHEMBL5824924	8.460	7.744	+0.716
240	CHEMBL3932176	8.460	7.817	+0.643
241	CHEMBL3678960	8.450	7.129	+1.321
242	CHEMBL6133955	8.450	7.667	+0.783
243	CHEMBL3357652	8.450	8.379	+0.071
244	CHEMBL271408	8.450	8.081	+0.369
245	CHEMBL3902778	8.440	8.419	+0.021
246	CHEMBL472746	8.440	6.944	+1.496
247	CHEMBL6102777	8.440	7.417	+1.023
248	CHEMBL28418	8.435	7.836	+0.599
249	CHEMBL474341	8.430	7.371	+1.059
250	CHEMBL5748865	8.425	8.059	+0.366
251	CHEMBL5763517	8.410	8.337	+0.073
252	CHEMBL4862976	8.410	6.916	+1.494
253	CHEMBL136058	8.403	8.276	+0.127
254	CHEMBL473437	8.400	7.868	+0.532
255	CHEMBL6058591	8.400	8.386	+0.014
256	CHEMBL5772398	8.400	8.463	-0.063
257	CHEMBL5970015	8.400	8.265	+0.135
258	CHEMBL5837076	8.400	8.445	-0.045
259	CHEMBL6002527	8.400	8.697	-0.297
260	CHEMBL5854469	8.400	7.293	+1.107
261	CHEMBL5933251	8.400	8.261	+0.139
262	CHEMBL5964497	8.400	8.344	+0.056
263	CHEMBL6063486	8.400	8.160	+0.240
264	CHEMBL5968371	8.400	8.459	-0.059
265	CHEMBL6006209	8.400	8.268	+0.132
266	CHEMBL5777803	8.400	8.224	+0.176
267	CHEMBL5912163	8.400	7.830	+0.570
268	CHEMBL5858513	8.400	7.801	+0.599
269	CHEMBL5831214	8.400	8.104	+0.296
270	CHEMBL5985094	8.400	8.386	+0.014
271	CHEMBL6044391	8.400	8.633	-0.233
272	CHEMBL5796361	8.400	8.471	-0.071
273	CHEMBL5419895	8.390	7.994	+0.396
274	CHEMBL2048905	8.390	7.907	+0.483
275	CHEMBL4526992	8.390	8.319	+0.071
276	CHEMBL399373	8.380	7.265	+1.115
277	CHEMBL5593561	8.370	7.587	+0.783
278	CHEMBL5432141	8.370	7.473	+0.897
279	CHEMBL4752951	8.370	7.682	+0.688
280	CHEMBL50519	8.367	7.502	+0.865
281	CHEMBL5208454	8.365	7.389	+0.976
282	CHEMBL417478	8.360	7.941	+0.419
283	CHEMBL4105621	8.350	7.340	+1.010
284	CHEMBL3752363	8.345	8.473	-0.128

continued on next page

Table 5 – *continued from previous page*

#	ChEMBL ID	Measured	Predicted	Residual
285	CHEMBL204085	8.345	7.783	+0.562
286	CHEMBL3900264	8.340	7.342	+0.998
287	CHEMBL4087740	8.340	7.385	+0.955
288	CHEMBL53665	8.335	7.932	+0.403
289	CHEMBL121796	8.335	6.665	+1.670
290	CHEMBL5274514	8.327	8.126	+0.201
291	CHEMBL5950594	8.320	8.161	+0.159
292	CHEMBL4225166	8.320	7.590	+0.730
293	CHEMBL342828	8.317	8.340	-0.024
294	CHEMBL5631065	8.315	7.442	+0.873
295	CHEMBL5435775	8.313	9.528	-1.214
296	CHEMBL4227201	8.312	7.324	+0.988
297	CHEMBL5988665	8.310	8.227	+0.083
298	CHEMBL5870511	8.310	8.445	-0.135
299	CHEMBL5867757	8.310	8.269	+0.041
300	CHEMBL3655345	8.300	7.636	+0.664
301	CHEMBL444619	8.300	6.950	+1.350
302	CHEMBL5855225	8.300	8.419	-0.119
303	CHEMBL5921291	8.300	8.183	+0.117
304	CHEMBL5948070	8.300	8.270	+0.030
305	CHEMBL5866385	8.300	8.642	-0.342
306	CHEMBL5812367	8.300	8.645	-0.345
307	CHEMBL5988997	8.300	8.013	+0.287
308	CHEMBL5933810	8.300	8.308	-0.008
309	CHEMBL5767274	8.300	8.108	+0.192
310	CHEMBL5929024	8.300	7.355	+0.945
311	CHEMBL3962343	8.300	6.924	+1.376
312	CHEMBL3967836	8.300	6.911	+1.389
313	CHEMBL4435233	8.300	7.994	+0.306
314	CHEMBL3633142	8.292	8.178	+0.114
315	CHEMBL3930506	8.275	8.127	+0.148
316	CHEMBL3897723	8.270	7.302	+0.968
317	CHEMBL4278815	8.270	7.732	+0.538
318	CHEMBL2048909	8.260	7.919	+0.341
319	CHEMBL3903713	8.250	7.352	+0.898
320	CHEMBL5179357	8.250	9.542	-1.292
321	CHEMBL3357643	8.250	8.639	-0.389
322	CHEMBL3622676	8.250	7.923	+0.327
323	CHEMBL5870500	8.247	6.859	+1.388
324	CHEMBL4448107	8.247	7.334	+0.912
325	CHEMBL248393	8.240	7.246	+0.994
326	CHEMBL3921816	8.240	7.603	+0.637
327	CHEMBL5204270	8.240	8.634	-0.394
328	CHEMBL5284115	8.240	7.305	+0.935
329	CHEMBL4206716	8.228	7.971	+0.256
330	CHEMBL3959248	8.225	7.799	+0.426
331	CHEMBL473438	8.220	7.836	+0.384
332	CHEMBL471058	8.220	7.854	+0.366
333	CHEMBL5975399	8.220	8.063	+0.157
334	CHEMBL3963723	8.220	6.968	+1.252
335	CHEMBL5967931	8.220	8.164	+0.056
336	CHEMBL5862252	8.220	8.514	-0.294
337	CHEMBL3622648	8.220	7.731	+0.489
338	CHEMBL5196318	8.220	7.097	+1.123
339	CHEMBL4532084	8.210	8.843	-0.633
340	CHEMBL6149199	8.200	7.780	+0.420
341	CHEMBL6144949	8.200	7.699	+0.501
342	CHEMBL3966129	8.200	7.359	+0.841
343	CHEMBL3622666	8.190	8.278	-0.088

continued on next page

Table 5 – *continued from previous page*

#	ChEMBL ID	Measured	Predicted	Residual
344	CHEMBL3983586	8.190	7.362	+0.828
345	CHEMBL3655354	8.190	7.256	+0.934
346	CHEMBL4227091	8.187	7.576	+0.611
347	CHEMBL53156	8.184	7.683	+0.501
348	CHEMBL5786339	8.170	8.414	-0.244
349	CHEMBL5187605	8.170	8.679	-0.509
350	CHEMBL3926191	8.160	8.342	-0.182
351	CHEMBL3939771	8.160	8.419	-0.259
352	CHEMBL6102895	8.160	8.747	-0.587
353	CHEMBL3963403	8.160	7.379	+0.781
354	CHEMBL4225863	8.157	7.560	+0.597
355	CHEMBL2048908	8.150	7.872	+0.278
356	CHEMBL3416628	8.150	6.822	+1.328
357	CHEMBL2048907	8.150	7.823	+0.327
358	CHEMBL475446	8.150	7.935	+0.215
359	CHEMBL6015868	8.150	8.226	-0.076
360	CHEMBL5907571	8.150	8.155	-0.005
361	CHEMBL1934625	8.150	7.317	+0.833
362	CHEMBL5950707	8.150	7.712	+0.438
363	CHEMBL5870912	8.150	7.879	+0.271
364	CHEMBL311109	8.150	8.225	-0.075
365	CHEMBL5280471	8.150	7.436	+0.714
366	CHEMBL5836136	8.150	8.031	+0.119
367	CHEMBL5977553	8.150	8.492	-0.342
368	CHEMBL5901017	8.150	8.202	-0.052
369	CHEMBL5972316	8.150	8.189	-0.039
370	CHEMBL5842644	8.150	8.318	-0.168
371	CHEMBL6057006	8.140	7.991	+0.149
372	CHEMBL4228672	8.137	7.610	+0.526
373	CHEMBL4561606	8.135	7.168	+0.967
374	CHEMBL4215080	8.130	7.459	+0.671
375	CHEMBL4215076	8.127	7.670	+0.458
376	CHEMBL6134098	8.125	7.850	+0.275
377	CHEMBL2437471	8.123	7.442	+0.681
378	CHEMBL4079861	8.120	7.375	+0.745
379	CHEMBL6173924	8.120	7.951	+0.169
380	CHEMBL6041434	8.110	7.355	+0.755
381	CHEMBL5202945	8.110	7.877	+0.233
382	CHEMBL5993587	8.110	7.664	+0.446
383	CHEMBL5094288	8.100	7.536	+0.564
384	CHEMBL6060281	8.100	8.182	-0.082
385	CHEMBL5986956	8.100	8.571	-0.471
386	CHEMBL5786207	8.100	8.241	-0.141
387	CHEMBL4752026	8.100	6.178	+1.922
388	CHEMBL5591709	8.090	7.443	+0.647
389	CHEMBL3622649	8.090	8.068	+0.022
390	CHEMBL3818364	8.088	7.608	+0.480
391	CHEMBL2437467	8.087	7.652	+0.435
392	CHEMBL3233791	8.085	8.065	+0.020
393	CHEMBL5945719	8.083	8.322	-0.239
394	CHEMBL4097369	8.080	7.443	+0.637
395	CHEMBL3740144	8.080	7.514	+0.566
396	CHEMBL4060919	8.073	7.141	+0.932
397	CHEMBL4753389	8.070	7.625	+0.445
398	CHEMBL5177378	8.070	7.212	+0.858
399	CHEMBL4799276	8.070	7.603	+0.467
400	CHEMBL4457018	8.060	6.558	+1.502
401	CHEMBL5823169	8.060	7.628	+0.432
402	CHEMBL5266919	8.060	8.296	-0.236

continued on next page

Table 5 – *continued from previous page*

#	ChEMBL ID	Measured	Predicted	Residual
403	CHEMBL6120461	8.058	8.953	-0.896
404	CHEMBL5277831	8.050	6.109	+1.941
405	CHEMBL4280817	8.050	5.033	+3.017
406	CHEMBL506414	8.050	7.033	+1.017
407	CHEMBL6060509	8.050	8.120	-0.070
408	CHEMBL5994327	8.050	8.207	-0.157
409	CHEMBL5743124	8.050	8.071	-0.021
410	CHEMBL5758455	8.050	8.559	-0.509
411	CHEMBL6035061	8.045	7.362	+0.683
412	CHEMBL3639542	8.040	8.519	-0.479
413	CHEMBL5187169	8.040	7.035	+1.005
414	CHEMBL3956430	8.040	6.829	+1.211
415	CHEMBL2441563	8.035	7.491	+0.544
416	CHEMBL4083486	8.035	7.242	+0.793
417	CHEMBL2437488	8.030	7.679	+0.351
418	CHEMBL5205022	8.030	7.456	+0.574
419	CHEMBL5218754	8.030	6.895	+1.135
420	CHEMBL4581316	8.023	7.392	+0.632
421	CHEMBL3926558	8.020	8.047	-0.027
422	CHEMBL4445017	8.020	7.133	+0.887
423	CHEMBL4558324	8.015	7.354	+0.661
424	CHEMBL2437466	8.012	7.564	+0.448
425	CHEMBL5967585	8.000	8.484	-0.484
426	CHEMBL3965948	8.000	7.073	+0.927
427	CHEMBL5218990	8.000	6.695	+1.305
428	CHEMBL5844686	8.000	8.123	-0.123
429	CHEMBL5771679	8.000	7.575	+0.425
430	CHEMBL5821647	8.000	7.900	+0.100
431	CHEMBL5790308	8.000	8.249	-0.249
432	CHEMBL3655348	8.000	7.286	+0.714
433	CHEMBL251124	8.000	7.681	+0.319
434	CHEMBL5770894	8.000	7.743	+0.257
435	CHEMBL4565984	7.993	7.591	+0.403
436	CHEMBL5743902	7.990	7.963	+0.027
437	CHEMBL4065288	7.990	7.590	+0.400
438	CHEMBL4227210	7.990	7.568	+0.422
439	CHEMBL4095439	7.990	7.410	+0.580
440	CHEMBL4103194	7.990	7.271	+0.719
441	CHEMBL5178922	7.990	7.372	+0.618
442	CHEMBL4558227	7.990	7.338	+0.652
443	CHEMBL3950966	7.985	7.749	+0.236
444	CHEMBL3219127	7.974	7.464	+0.510
445	CHEMBL4473365	7.973	7.594	+0.379
446	CHEMBL5406792	7.970	7.861	+0.109
447	CHEMBL5288249	7.970	7.403	+0.567
448	CHEMBL1928315	7.970	7.630	+0.340
449	CHEMBL2437463	7.967	7.984	-0.017
450	CHEMBL3891846	7.965	7.442	+0.523
451	CHEMBL5433431	7.960	7.238	+0.722
452	CHEMBL6060963	7.960	7.807	+0.153
453	CHEMBL5832966	7.960	7.206	+0.754
454	CHEMBL5981434	7.960	8.028	-0.068
455	CHEMBL6043318	7.960	7.614	+0.346
456	CHEMBL4572315	7.960	8.369	-0.409
457	CHEMBL125920	7.960	7.223	+0.737
458	CHEMBL6047881	7.960	8.319	-0.359
459	CHEMBL5999776	7.960	8.123	-0.163
460	CHEMBL2437479	7.957	7.681	+0.276
461	CHEMBL3233779	7.955	8.190	-0.235

continued on next page

Table 5 – *continued from previous page*

#	ChEMBL ID	Measured	Predicted	Residual
462	CHEMBL2437465	7.952	7.482	+0.470
463	CHEMBL3357653	7.950	8.379	-0.429
464	CHEMBL5937036	7.950	7.759	+0.191
465	CHEMBL4458843	7.945	7.008	+0.937
466	CHEMBL4870265	7.945	7.180	+0.765
467	CHEMBL3982710	7.940	7.603	+0.337
468	CHEMBL4443062	7.938	8.168	-0.231
469	CHEMBL4520341	7.937	7.542	+0.394
470	CHEMBL6168400	7.933	7.535	+0.398
471	CHEMBL251314	7.930	7.265	+0.665
472	CHEMBL4449391	7.930	7.821	+0.109
473	CHEMBL3233790	7.930	8.138	-0.208
474	CHEMBL205798	7.920	7.107	+0.813
475	CHEMBL3680378	7.920	8.116	-0.196
476	CHEMBL5280584	7.920	6.155	+1.765
477	CHEMBL3671571	7.920	7.462	+0.458
478	CHEMBL399372	7.920	7.698	+0.222
479	CHEMBL480349	7.920	7.178	+0.742
480	CHEMBL5912978	7.920	7.971	-0.051
481	CHEMBL5966530	7.920	8.283	-0.363
482	CHEMBL256297	7.920	7.293	+0.627
483	CHEMBL5920247	7.905	7.365	+0.540
484	CHEMBL5594919	7.900	6.943	+0.957
485	CHEMBL4079501	7.890	7.554	+0.336
486	CHEMBL3114688	7.890	7.306	+0.584
487	CHEMBL6045905	7.890	7.459	+0.431
488	CHEMBL4531912	7.890	7.731	+0.159
489	CHEMBL5281748	7.890	8.421	-0.531
490	CHEMBL5961822	7.890	8.387	-0.497
491	CHEMBL5752067	7.890	8.110	-0.220
492	CHEMBL5803419	7.890	7.973	-0.083
493	CHEMBL6133741	7.890	7.592	+0.298
494	CHEMBL5414782	7.885	7.474	+0.411
495	CHEMBL1090357	7.885	7.643	+0.242
496	CHEMBL4534592	7.885	6.378	+1.507
497	CHEMBL428741	7.880	6.718	+1.162
498	CHEMBL6039588	7.880	8.270	-0.390
499	CHEMBL6151046	7.880	8.694	-0.814
500	CHEMBL5422592	7.875	7.312	+0.563
501	CHEMBL5430958	7.875	7.942	-0.067
502	CHEMBL4072299	7.873	7.332	+0.542
503	CHEMBL67003	7.870	6.832	+1.038
504	CHEMBL4094959	7.870	6.955	+0.915
505	CHEMBL3114700	7.870	7.280	+0.590
506	CHEMBL5437343	7.865	7.970	-0.105
507	CHEMBL5169547	7.860	7.377	+0.483
508	CHEMBL3910697	7.855	8.262	-0.407
509	CHEMBL4468924	7.850	7.208	+0.642
510	CHEMBL6051041	7.850	8.241	-0.391
511	CHEMBL6036007	7.850	8.244	-0.394
512	CHEMBL5779120	7.850	8.133	-0.283
513	CHEMBL5834209	7.850	8.540	-0.690
514	CHEMBL5973646	7.850	8.233	-0.383
515	CHEMBL5740900	7.850	8.213	-0.363
516	CHEMBL3233793	7.845	8.240	-0.395
517	CHEMBL3786453	7.845	7.540	+0.305
518	CHEMBL3233773	7.845	7.912	-0.067
519	CHEMBL4091276	7.845	7.529	+0.316
520	CHEMBL4227701	7.840	7.486	+0.354

continued on next page

Table 5 – *continued from previous page*

#	ChEMBL ID	Measured	Predicted	Residual
521	CHEMBL3233774	7.830	8.080	-0.250
522	CHEMBL5432280	7.825	7.418	+0.407
523	CHEMBL5270000	7.820	7.433	+0.387
524	CHEMBL1630116	7.820	7.968	-0.148
525	CHEMBL4280174	7.820	7.081	+0.739
526	CHEMBL514938	7.820	7.054	+0.766
527	CHEMBL5440481	7.820	7.550	+0.270
528	CHEMBL3939985	7.820	7.519	+0.301
529	CHEMBL4217992	7.815	7.838	-0.023
530	CHEMBL5976230	7.813	8.755	-0.942
531	CHEMBL5203543	7.810	7.681	+0.129
532	CHEMBL53796	7.810	7.525	+0.285
533	CHEMBL4467561	7.810	7.655	+0.155
534	CHEMBL5291434	7.805	7.289	+0.516
535	CHEMBL4853833	7.805	7.779	+0.026
536	CHEMBL5438568	7.800	6.650	+1.150
537	CHEMBL498133	7.800	7.732	+0.068
538	CHEMBL257815	7.800	7.159	+0.641
539	CHEMBL175557	7.800	7.058	+0.742
540	CHEMBL5876510	7.800	7.892	-0.092
541	CHEMBL5901440	7.800	8.414	-0.614
542	CHEMBL4444312	7.800	6.437	+1.363
543	CHEMBL5998521	7.795	7.991	-0.196
544	CHEMBL4869409	7.790	7.156	+0.634
545	CHEMBL3633143	7.789	7.767	+0.022
546	CHEMBL300217	7.782	7.581	+0.201
547	CHEMBL5085136	7.780	7.107	+0.673
548	CHEMBL31816	7.780	7.748	+0.032
549	CHEMBL3233778	7.780	8.037	-0.257
550	CHEMBL4476791	7.775	7.111	+0.664
551	CHEMBL6065436	7.770	7.819	-0.049
552	CHEMBL3948084	7.770	7.097	+0.673
553	CHEMBL3671561	7.770	7.518	+0.252
554	CHEMBL1683955	7.770	6.717	+1.053
555	CHEMBL338049	7.770	6.268	+1.502
556	CHEMBL4292351	7.770	7.160	+0.610
557	CHEMBL5928047	7.770	7.974	-0.204
558	CHEMBL5903269	7.767	8.066	-0.300
559	CHEMBL5083976	7.760	6.936	+0.824
560	CHEMBL5862522	7.757	7.561	+0.196
561	CHEMBL5282956	7.750	7.244	+0.506
562	CHEMBL4284071	7.750	7.008	+0.742
563	CHEMBL122182	7.750	6.926	+0.824
564	CHEMBL4129209	7.750	7.849	-0.099
565	CHEMBL4097883	7.750	7.324	+0.426
566	CHEMBL6060196	7.750	8.285	-0.535
567	CHEMBL291496	7.745	7.402	+0.343
568	CHEMBL4078914	7.745	7.235	+0.510
569	CHEMBL3919447	7.745	7.800	-0.055
570	CHEMBL4072048	7.745	7.455	+0.290
571	CHEMBL5953169	7.740	7.124	+0.616
572	CHEMBL3114686	7.735	7.316	+0.419
573	CHEMBL4104658	7.732	7.641	+0.092
574	CHEMBL207130	7.725	7.745	-0.020
575	CHEMBL5170653	7.725	6.928	+0.797
576	CHEMBL484270	7.725	8.505	-0.780
577	CHEMBL5203195	7.725	6.930	+0.795
578	CHEMBL3786343	7.721	7.440	+0.281
579	CHEMBL248114	7.720	6.893	+0.827

continued on next page

Table 5 – *continued from previous page*

#	ChEMBL ID	Measured	Predicted	Residual
580	CHEMBL5430071	7.720	7.366	+0.354
581	CHEMBL5201287	7.720	7.772	-0.052
582	CHEMBL6152563	7.720	7.772	-0.052
583	CHEMBL1683968	7.720	7.367	+0.353
584	CHEMBL5846336	7.720	8.396	-0.676
585	CHEMBL335648	7.720	6.043	+1.677
586	CHEMBL3982289	7.720	7.001	+0.719
587	CHEMBL5750469	7.720	7.957	-0.237
588	CHEMBL5284701	7.720	8.429	-0.709
589	CHEMBL4068047	7.720	7.425	+0.295
590	CHEMBL3233789	7.715	8.047	-0.332
591	CHEMBL3965087	7.715	7.286	+0.429
592	CHEMBL31570	7.715	7.824	-0.109
593	CHEMBL3233767	7.710	8.020	-0.310
594	CHEMBL4078120	7.710	7.368	+0.342
595	CHEMBL4089588	7.710	7.578	+0.132
596	CHEMBL4873098	7.710	7.858	-0.148
597	CHEMBL5952457	7.707	7.166	+0.542
598	CHEMBL5218459	7.700	6.314	+1.386
599	CHEMBL3360616	7.700	7.652	+0.048
600	CHEMBL175556	7.700	6.994	+0.706
601	CHEMBL490093	7.700	7.243	+0.457
602	CHEMBL5290340	7.700	6.505	+1.195
603	CHEMBL5933544	7.700	7.502	+0.198
604	CHEMBL4468893	7.700	8.344	-0.644
605	CHEMBL5408900	7.700	7.266	+0.434
606	CHEMBL4064361	7.700	6.420	+1.280
607	CHEMBL4078023	7.700	6.164	+1.536
608	CHEMBL3954684	7.700	6.890	+0.810
609	CHEMBL3956465	7.700	7.007	+0.693
610	CHEMBL3699612	7.700	7.658	+0.042
611	CHEMBL6039541	7.697	7.617	+0.079
612	CHEMBL3114701	7.695	7.081	+0.614
613	CHEMBL5954442	7.693	8.808	-1.115
614	CHEMBL5780557	7.690	7.985	-0.295
615	CHEMBL5426811	7.685	7.312	+0.373
616	CHEMBL243837	7.680	6.618	+1.062
617	CHEMBL3819553	7.680	7.646	+0.034
618	CHEMBL5882388	7.680	8.110	-0.430
619	CHEMBL2048910	7.680	7.839	-0.159
620	CHEMBL565714	7.680	7.211	+0.469
621	CHEMBL2048794	7.680	7.130	+0.550
622	CHEMBL2381516	7.677	7.332	+0.345
623	CHEMBL6006961	7.675	7.449	+0.226
624	CHEMBL4858714	7.675	7.430	+0.245
625	CHEMBL67057	7.675	6.789	+0.886
626	CHEMBL5198920	7.670	6.927	+0.743
627	CHEMBL4278886	7.670	7.355	+0.315
628	CHEMBL5900941	7.670	7.982	-0.312
629	CHEMBL4744973	7.660	7.407	+0.253
630	CHEMBL4752022	7.660	7.407	+0.253
631	CHEMBL5281270	7.660	6.201	+1.459
632	CHEMBL1934622	7.660	7.207	+0.453
633	CHEMBL5284179	7.660	7.044	+0.616
634	CHEMBL3765789	7.655	7.872	-0.217
635	CHEMBL6164231	7.650	7.176	+0.474
636	CHEMBL4861346	7.645	7.551	+0.094
637	CHEMBL3409483	7.640	6.863	+0.777
638	CHEMBL255438	7.640	7.736	-0.096

continued on next page

Table 5 – *continued from previous page*

#	ChEMBL ID	Measured	Predicted	Residual
639	CHEMBL5180965	7.640	8.520	-0.880
640	CHEMBL3416439	7.640	7.238	+0.402
641	CHEMBL5281266	7.640	7.501	+0.139
642	CHEMBL6034266	7.640	7.883	-0.243
643	CHEMBL5266823	7.640	5.932	+1.708
644	CHEMBL5286055	7.637	8.126	-0.489
645	CHEMBL3754803	7.630	7.794	-0.164
646	CHEMBL5875973	7.630	7.932	-0.302
647	CHEMBL6030424	7.630	7.397	+0.233
648	CHEMBL5962278	7.625	7.170	+0.455
649	CHEMBL5439143	7.620	6.474	+1.146
650	CHEMBL359630	7.620	7.054	+0.566
651	CHEMBL2425083	7.620	7.229	+0.391
652	CHEMBL3986363	7.618	7.178	+0.440
653	CHEMBL3910422	7.618	7.163	+0.455
654	CHEMBL4540442	7.610	7.485	+0.125
655	CHEMBL5183664	7.610	7.010	+0.600
656	CHEMBL6015235	7.610	6.823	+0.787
657	CHEMBL4855080	7.605	8.009	-0.404
658	CHEMBL5791264	7.600	7.609	-0.009
659	CHEMBL4437599	7.600	8.041	-0.441
660	CHEMBL5967623	7.600	7.024	+0.576
661	CHEMBL5772311	7.600	7.470	+0.130
662	CHEMBL6031205	7.600	7.110	+0.490
663	CHEMBL5969872	7.600	7.663	-0.063
664	CHEMBL5959647	7.600	7.378	+0.222
665	CHEMBL3233775	7.600	8.073	-0.473
666	CHEMBL5806097	7.600	7.500	+0.100
667	CHEMBL5180289	7.595	7.568	+0.027
668	CHEMBL6147474	7.590	6.824	+0.766
669	CHEMBL31656	7.590	7.615	-0.025
670	CHEMBL5286368	7.590	7.313	+0.277
671	CHEMBL3678959	7.587	7.608	-0.022
672	CHEMBL5840400	7.582	7.604	-0.021
673	CHEMBL6011498	7.580	8.120	-0.540
674	CHEMBL5943090	7.580	7.814	-0.234
675	CHEMBL3416629	7.580	6.726	+0.854
676	CHEMBL233325	7.580	6.883	+0.697
677	CHEMBL5437854	7.580	7.392	+0.188
678	CHEMBL4795878	7.575	7.693	-0.118
679	CHEMBL4786787	7.575	7.693	-0.118
680	CHEMBL5087931	7.575	7.805	-0.230
681	CHEMBL2419762	7.570	7.303	+0.267
682	CHEMBL500591	7.570	7.128	+0.442
683	CHEMBL4558356	7.570	8.248	-0.678
684	CHEMBL5403023	7.560	7.392	+0.168
685	CHEMBL5196024	7.560	6.880	+0.680
686	CHEMBL4161938	7.557	7.672	-0.116
687	CHEMBL3751976	7.555	7.796	-0.241
688	CHEMBL4064937	7.540	7.685	-0.145
689	CHEMBL5820578	7.540	8.130	-0.590
690	CHEMBL3233772	7.535	7.988	-0.453
691	CHEMBL4465103	7.530	7.225	+0.305
692	CHEMBL205966	7.520	7.308	+0.212
693	CHEMBL583218	7.520	7.438	+0.082
694	CHEMBL5189114	7.520	7.515	+0.005
695	CHEMBL111197	7.520	6.140	+1.380
696	CHEMBL4858409	7.515	7.593	-0.078
697	CHEMBL206782	7.510	7.191	+0.319

continued on next page

Table 5 – *continued from previous page*

#	ChEMBL ID	Measured	Predicted	Residual
698	CHEMBL4791249	7.510	7.079	+0.431
699	CHEMBL132348	7.510	5.944	+1.566
700	CHEMBL360958	7.510	7.120	+0.390
701	CHEMBL4871192	7.510	7.387	+0.123
702	CHEMBL1782692	7.510	5.681	+1.829
703	CHEMBL5981661	7.507	8.671	-1.164
704	CHEMBL5597916	7.507	6.934	+0.573
705	CHEMBL3741304	7.500	6.775	+0.725
706	CHEMBL6025269	7.500	7.036	+0.464
707	CHEMBL5284866	7.500	6.402	+1.098
708	CHEMBL3671577	7.500	7.497	+0.003
709	CHEMBL248321	7.500	7.099	+0.401
710	CHEMBL5275712	7.500	7.511	-0.011
711	CHEMBL3671572	7.500	7.429	+0.071
712	CHEMBL5517879	7.490	6.060	+1.430
713	CHEMBL5181684	7.490	6.969	+0.521
714	CHEMBL5901475	7.487	7.055	+0.432
715	CHEMBL1241487	7.480	5.913	+1.567
716	CHEMBL4868909	7.480	6.188	+1.292
717	CHEMBL256529	7.480	7.169	+0.311
718	CHEMBL204467	7.480	7.337	+0.143
719	CHEMBL4857353	7.480	7.066	+0.414
720	CHEMBL3085375	7.480	7.169	+0.311
721	CHEMBL392273	7.480	6.762	+0.718
722	CHEMBL5755313	7.478	6.559	+0.919
723	CHEMBL4877054	7.475	8.037	-0.562
724	CHEMBL5270334	7.470	7.566	-0.096
725	CHEMBL179766	7.470	7.179	+0.291
726	CHEMBL5271155	7.470	6.656	+0.814
727	CHEMBL5742497	7.470	7.949	-0.479
728	CHEMBL1683961	7.470	7.781	-0.311
729	CHEMBL54805	7.465	7.249	+0.216
730	CHEMBL388978	7.464	6.235	+1.229
731	CHEMBL355982	7.460	7.143	+0.317
732	CHEMBL5779959	7.460	8.384	-0.924
733	CHEMBL5283216	7.460	7.208	+0.252
734	CHEMBL5567575	7.453	6.991	+0.462
735	CHEMBL3986103	7.445	7.250	+0.195
736	CHEMBL2437487	7.443	6.444	+1.000
737	CHEMBL5613341	7.440	6.893	+0.547
738	CHEMBL2348417	7.440	7.397	+0.043
739	CHEMBL2048904	7.440	7.632	-0.192
740	CHEMBL5788794	7.440	8.226	-0.786
741	CHEMBL5273000	7.440	6.414	+1.026
742	CHEMBL4534719	7.438	6.549	+0.888
743	CHEMBL6160197	7.430	7.205	+0.225
744	CHEMBL5559947	7.430	7.350	+0.080
745	CHEMBL5994959	7.430	7.532	-0.102
746	CHEMBL5434067	7.430	7.749	-0.319
747	CHEMBL5859805	7.430	8.005	-0.575
748	CHEMBL5975268	7.430	6.400	+1.030
749	CHEMBL5806894	7.420	7.760	-0.340
750	CHEMBL4159404	7.420	7.069	+0.351
751	CHEMBL6052335	7.420	8.050	-0.630
752	CHEMBL5279794	7.410	7.449	-0.039
753	CHEMBL3753228	7.410	7.858	-0.448
754	CHEMBL4129287	7.403	7.849	-0.446
755	CHEMBL4102389	7.400	7.560	-0.160
756	CHEMBL346863	7.400	7.549	-0.149

continued on next page

Table 5 – *continued from previous page*

#	ChEMBL ID	Measured	Predicted	Residual
757	CHEMBL4094823	7.400	6.851	+0.549
758	CHEMBL4083642	7.400	6.222	+1.178
759	CHEMBL5286091	7.400	7.381	+0.019
760	CHEMBL5885640	7.400	8.414	-1.014
761	CHEMBL5825244	7.390	7.804	-0.414
762	CHEMBL5612021	7.390	6.890	+0.500
763	CHEMBL512391	7.390	7.024	+0.366
764	CHEMBL4797395	7.390	7.089	+0.301
765	CHEMBL5523623	7.390	5.955	+1.435
766	CHEMBL3953593	7.390	6.871	+0.519
767	CHEMBL5279933	7.390	6.585	+0.805
768	CHEMBL5276304	7.390	7.352	+0.038
769	CHEMBL5279491	7.380	7.524	-0.144
770	CHEMBL31630	7.380	7.284	+0.096
771	CHEMBL205235	7.380	7.228	+0.152
772	CHEMBL5188399	7.377	7.564	-0.187
773	CHEMBL597200	7.373	7.475	-0.102
774	CHEMBL4228120	7.370	7.134	+0.236
775	CHEMBL5967745	7.370	7.484	-0.114
776	CHEMBL5290929	7.370	7.061	+0.309
777	CHEMBL5269307	7.370	6.403	+0.967
778	CHEMBL340611	7.370	6.028	+1.342
779	CHEMBL5767529	7.367	7.213	+0.154
780	CHEMBL5946057	7.365	7.376	-0.011
781	CHEMBL3754564	7.365	7.821	-0.456
782	CHEMBL5976089	7.363	7.281	+0.082
783	CHEMBL4538707	7.360	7.911	-0.551
784	CHEMBL578044	7.360	7.410	-0.050
785	CHEMBL5205581	7.360	7.092	+0.268
786	CHEMBL2348416	7.360	7.220	+0.140
787	CHEMBL2437483	7.360	6.466	+0.894
788	CHEMBL5980602	7.357	7.123	+0.234
789	CHEMBL54788	7.355	7.495	-0.140
790	CHEMBL3671576	7.350	7.564	-0.214
791	CHEMBL2385960	7.350	7.217	+0.133
792	CHEMBL4214588	7.340	6.314	+1.026
793	CHEMBL6164475	7.340	6.484	+0.856
794	CHEMBL4280835	7.330	7.347	-0.017
795	CHEMBL255236	7.330	7.571	-0.241
796	CHEMBL2348418	7.330	7.296	+0.034
797	CHEMBL5268092	7.330	6.704	+0.626
798	CHEMBL4173696	7.330	6.592	+0.738
799	CHEMBL3938358	7.324	7.035	+0.289
800	CHEMBL3891187	7.324	6.999	+0.325
801	CHEMBL3954906	7.324	7.051	+0.273
802	CHEMBL3903641	7.324	6.943	+0.381
803	CHEMBL4291471	7.320	7.326	-0.006
804	CHEMBL402316	7.320	7.501	-0.181
805	CHEMBL5182231	7.315	6.848	+0.467
806	CHEMBL329867	7.315	6.444	+0.871
807	CHEMBL5556281	7.310	7.237	+0.073
808	CHEMBL1630114	7.310	7.457	-0.147
809	CHEMBL3752074	7.310	6.632	+0.678
810	CHEMBL525725	7.310	7.413	-0.103
811	CHEMBL4859542	7.310	8.023	-0.713
812	CHEMBL383246	7.310	7.699	-0.389
813	CHEMBL5772994	7.310	6.988	+0.322
814	CHEMBL204420	7.310	7.528	-0.218
815	CHEMBL5271718	7.310	7.422	-0.112

continued on next page

Table 5 – *continued from previous page*

#	ChEMBL ID	Measured	Predicted	Residual
816	CHEMBL332612	7.300	6.839	+0.461
817	CHEMBL504117	7.300	7.089	+0.211
818	CHEMBL3983201	7.300	7.471	-0.171
819	CHEMBL3944702	7.300	6.967	+0.333
820	CHEMBL5949708	7.300	6.413	+0.887
821	CHEMBL3963620	7.300	7.473	-0.173
822	CHEMBL381080	7.300	7.232	+0.068
823	CHEMBL2437458	7.298	7.710	-0.412
824	CHEMBL5956969	7.292	7.454	-0.161
825	CHEMBL3938794	7.290	7.365	-0.075
826	CHEMBL4856592	7.290	7.142	+0.148
827	CHEMBL597402	7.280	7.655	-0.375
828	CHEMBL3927914	7.280	7.451	-0.171
829	CHEMBL3905411	7.280	7.365	-0.085
830	CHEMBL3612593	7.280	6.557	+0.723
831	CHEMBL3416438	7.280	7.407	-0.127
832	CHEMBL6159859	7.280	7.199	+0.081
833	CHEMBL3977504	7.275	7.499	-0.224
834	CHEMBL5427952	7.270	7.363	-0.093
835	CHEMBL516487	7.270	7.057	+0.213
836	CHEMBL6036686	7.260	7.439	-0.179
837	CHEMBL3969056	7.260	7.304	-0.044
838	CHEMBL1933083	7.260	7.245	+0.015
839	CHEMBL1093234	7.260	7.168	+0.092
840	CHEMBL6013821	7.260	6.648	+0.612
841	CHEMBL4092966	7.260	7.286	-0.026
842	CHEMBL5193609	7.257	7.020	+0.237
843	CHEMBL6056421	7.253	7.337	-0.084
844	CHEMBL4095271	7.250	6.818	+0.432
845	CHEMBL5179214	7.250	6.933	+0.317
846	CHEMBL442938	7.250	6.531	+0.719
847	CHEMBL3671480	7.250	7.314	-0.064
848	CHEMBL5270335	7.240	7.061	+0.179
849	CHEMBL257814	7.240	7.360	-0.120
850	CHEMBL6030662	7.240	7.235	+0.005
851	CHEMBL4851795	7.240	7.176	+0.064
852	CHEMBL5945683	7.235	7.040	+0.195
853	CHEMBL4093954	7.235	6.917	+0.318
854	CHEMBL604914	7.230	7.702	-0.472
855	CHEMBL4284684	7.225	7.056	+0.169
856	CHEMBL4176629	7.220	6.059	+1.161
857	CHEMBL207247	7.220	7.108	+0.112
858	CHEMBL5779576	7.215	8.645	-1.430
859	CHEMBL3968839	7.210	7.120	+0.090
860	CHEMBL5286092	7.210	6.936	+0.274
861	CHEMBL6162629	7.205	7.480	-0.275
862	CHEMBL4099008	7.200	6.983	+0.217
863	CHEMBL1090356	7.200	7.035	+0.165
864	CHEMBL243629	7.200	6.592	+0.608
865	CHEMBL499534	7.200	7.231	-0.031
866	CHEMBL204625	7.200	7.342	-0.142
867	CHEMBL4531697	7.197	7.423	-0.226
868	CHEMBL6049383	7.194	7.138	+0.056
869	CHEMBL4228473	7.193	7.090	+0.103
870	CHEMBL4740108	7.190	6.797	+0.393
871	CHEMBL3954273	7.190	7.207	-0.017
872	CHEMBL3924500	7.190	7.398	-0.208
873	CHEMBL3896476	7.190	7.455	-0.265
874	CHEMBL5272404	7.190	7.357	-0.167

continued on next page

Table 5 – *continued from previous page*

#	ChEMBL ID	Measured	Predicted	Residual
875	CHEMBL92356	7.190	7.506	-0.316
876	CHEMBL137364	7.190	7.300	-0.110
877	CHEMBL4755651	7.180	6.682	+0.498
878	CHEMBL241919	7.180	6.664	+0.516
879	CHEMBL205776	7.180	7.011	+0.169
880	CHEMBL3754785	7.180	6.643	+0.537
881	CHEMBL320705	7.180	6.236	+0.944
882	CHEMBL4100655	7.177	6.933	+0.244
883	CHEMBL6033989	7.175	7.622	-0.447
884	CHEMBL6173880	7.175	6.950	+0.225
885	CHEMBL5822102	7.175	7.198	-0.023
886	CHEMBL1241676	7.170	5.901	+1.269
887	CHEMBL6026402	7.170	7.425	-0.255
888	CHEMBL6030528	7.170	7.532	-0.362
889	CHEMBL5746985	7.170	7.958	-0.788
890	CHEMBL5979481	7.170	7.516	-0.346
891	CHEMBL473436	7.170	7.782	-0.612
892	CHEMBL5564483	7.170	5.612	+1.558
893	CHEMBL4521214	7.160	8.118	-0.958
894	CHEMBL5201846	7.157	6.695	+0.462
895	CHEMBL4099713	7.157	7.822	-0.666
896	CHEMBL5572882	7.150	5.780	+1.370
897	CHEMBL5419002	7.150	7.393	-0.243
898	CHEMBL5964556	7.150	7.148	+0.002
899	CHEMBL5512476	7.150	5.846	+1.304
900	CHEMBL5266482	7.150	7.466	-0.316
901	CHEMBL5959742	7.147	6.884	+0.263
902	CHEMBL493428	7.147	7.380	-0.233
903	CHEMBL5199354	7.145	7.000	+0.145
904	CHEMBL3754213	7.145	7.755	-0.610
905	CHEMBL4438856	7.145	7.362	-0.217
906	CHEMBL4175782	7.145	7.360	-0.215
907	CHEMBL3233771	7.140	7.960	-0.820
908	CHEMBL4871569	7.140	5.955	+1.185
909	CHEMBL4750734	7.140	6.502	+0.638
910	CHEMBL3622622	7.140	7.107	+0.033
911	CHEMBL6001037	7.135	7.086	+0.049
912	CHEMBL4059772	7.130	6.663	+0.467
913	CHEMBL316127	7.130	6.962	+0.168
914	CHEMBL5563885	7.128	6.944	+0.184
915	CHEMBL2437482	7.127	6.557	+0.570
916	CHEMBL2437486	7.127	7.138	-0.012
917	CHEMBL264382	7.123	6.930	+0.194
918	CHEMBL1380839	7.120	7.462	-0.342
919	CHEMBL5564226	7.120	6.457	+0.663
920	CHEMBL461311	7.120	7.220	-0.100
921	CHEMBL5398132	7.115	7.358	-0.243
922	CHEMBL4569552	7.115	7.011	+0.104
923	CHEMBL4087570	7.115	6.783	+0.332
924	CHEMBL5589692	7.110	5.930	+1.180
925	CHEMBL4528630	7.110	7.611	-0.501
926	CHEMBL3977192	7.110	7.413	-0.303
927	CHEMBL3752365	7.110	7.854	-0.744
928	CHEMBL3936642	7.110	6.571	+0.539
929	CHEMBL4457447	7.110	6.978	+0.132
930	CHEMBL207246	7.105	7.420	-0.315
931	CHEMBL589588	7.100	4.825	+2.275
932	CHEMBL4636996	7.100	5.910	+1.190
933	CHEMBL379093	7.100	7.281	-0.181

continued on next page

Table 5 – *continued from previous page*

#	ChEMBL ID	Measured	Predicted	Residual
934	CHEMBL120979	7.100	6.668	+0.432
935	CHEMBL207410	7.100	7.148	-0.048
936	CHEMBL113070	7.100	6.068	+1.032
937	CHEMBL3416621	7.100	6.891	+0.209
938	CHEMBL5847006	7.099	7.083	+0.016
939	CHEMBL6020599	7.090	7.715	-0.625
940	CHEMBL5431658	7.085	7.443	-0.358
941	CHEMBL4105442	7.085	6.944	+0.141
942	CHEMBL4582195	7.080	6.643	+0.437
943	CHEMBL1093815	7.080	6.994	+0.086
944	CHEMBL5593262	7.080	6.309	+0.771
945	CHEMBL5273261	7.080	8.429	-1.349
946	CHEMBL1241581	7.080	5.760	+1.320
947	CHEMBL6013504	7.077	7.110	-0.033
948	CHEMBL4454929	7.077	7.433	-0.357
949	CHEMBL5089073	7.075	8.354	-1.279
950	CHEMBL3964380	7.070	7.205	-0.135
951	CHEMBL1242664	7.070	6.180	+0.890
952	CHEMBL248115	7.070	7.141	-0.071
953	CHEMBL327127	7.070	6.166	+0.904
954	CHEMBL1934630	7.070	7.229	-0.159
955	CHEMBL5594128	7.070	6.594	+0.476
956	CHEMBL205870	7.070	7.270	-0.200
957	CHEMBL5951162	7.067	5.330	+1.737
958	CHEMBL5199169	7.067	6.705	+0.362
959	CHEMBL5175316	7.060	6.643	+0.417
960	CHEMBL6000502	7.060	7.320	-0.260
961	CHEMBL3970399	7.060	6.959	+0.101
962	CHEMBL1242662	7.060	5.699	+1.361
963	CHEMBL2347972	7.055	6.826	+0.229
964	CHEMBL3233777	7.055	8.271	-1.216
965	CHEMBL6142537	7.050	7.772	-0.722
966	CHEMBL5276251	7.050	5.606	+1.444
967	CHEMBL1945448	7.050	5.366	+1.684
968	CHEMBL2435999	7.050	6.087	+0.963
969	CHEMBL6040102	7.050	7.507	-0.457
970	CHEMBL500217	7.050	6.568	+0.482
971	CHEMBL5194719	7.050	5.891	+1.159
972	CHEMBL2178368	7.040	7.171	-0.131
973	CHEMBL5966699	7.040	6.671	+0.369
974	CHEMBL3753208	7.040	6.934	+0.106
975	CHEMBL5561729	7.040	6.655	+0.385
976	CHEMBL5960825	7.037	7.156	-0.120
977	CHEMBL5874088	7.030	6.938	+0.092
978	CHEMBL3671558	7.030	7.339	-0.309
979	CHEMBL5760659	7.025	7.102	-0.077
980	CHEMBL4563980	7.020	7.223	-0.203
981	CHEMBL328977	7.020	7.525	-0.505
982	CHEMBL3754040	7.020	7.804	-0.784
983	CHEMBL481324	7.020	6.907	+0.113
984	CHEMBL56142	7.010	7.130	-0.120
985	CHEMBL5273791	7.010	7.452	-0.442
986	CHEMBL3931038	7.010	6.862	+0.148
987	CHEMBL2180203	7.003	7.321	-0.318
988	CHEMBL3133756	7.000	6.991	+0.009
989	CHEMBL3922922	7.000	7.241	-0.241
990	CHEMBL5273525	7.000	6.142	+0.858
991	CHEMBL3917105	7.000	7.046	-0.046
992	CHEMBL3775112	7.000	7.349	-0.349

continued on next page

Table 5 – *continued from previous page*

#	ChEMBL ID	Measured	Predicted	Residual
993	CHEMBL4207984	6.994	6.307	+0.687
994	CHEMBL5291481	6.990	6.337	+0.653
995	CHEMBL5290094	6.990	6.574	+0.416
996	CHEMBL4062509	6.987	7.356	-0.369
997	CHEMBL5983026	6.987	7.993	-1.007
998	CHEMBL4206312	6.975	7.175	-0.200
999	CHEMBL3947759	6.970	6.864	+0.106
1000	CHEMBL3416597	6.970	7.031	-0.061
1001	CHEMBL5272858	6.970	7.069	-0.099
1002	CHEMBL5275486	6.970	7.499	-0.529
1003	CHEMBL4067855	6.965	6.907	+0.058
1004	CHEMBL5970630	6.960	6.646	+0.314
1005	CHEMBL4780058	6.960	6.498	+0.462
1006	CHEMBL4166412	6.960	6.369	+0.591
1007	CHEMBL435809	6.960	5.826	+1.134
1008	CHEMBL5413223	6.960	6.444	+0.516
1009	CHEMBL208358	6.960	7.189	-0.229
1010	CHEMBL4863078	6.960	7.425	-0.465
1011	CHEMBL5870984	6.960	6.704	+0.256
1012	CHEMBL6063877	6.960	7.515	-0.555
1013	CHEMBL6159648	6.950	6.897	+0.053
1014	CHEMBL4286733	6.950	7.357	-0.407
1015	CHEMBL4446124	6.947	6.740	+0.207
1016	CHEMBL5542205	6.946	6.602	+0.344
1017	CHEMBL4278226	6.945	6.847	+0.098
1018	CHEMBL3775536	6.940	7.146	-0.206
1019	CHEMBL4642158	6.940	7.445	-0.505
1020	CHEMBL5962642	6.933	7.581	-0.648
1021	CHEMBL5766029	6.930	7.344	-0.414
1022	CHEMBL5967218	6.930	7.932	-1.002
1023	CHEMBL281543	6.930	7.762	-0.832
1024	CHEMBL6049268	6.927	6.727	+0.200
1025	CHEMBL5970361	6.927	6.770	+0.157
1026	CHEMBL5795317	6.927	6.552	+0.375
1027	CHEMBL5964547	6.927	6.363	+0.563
1028	CHEMBL5746419	6.927	6.744	+0.183
1029	CHEMBL5884157	6.927	6.734	+0.193
1030	CHEMBL5911586	6.927	6.535	+0.392
1031	CHEMBL5918405	6.927	6.689	+0.238
1032	CHEMBL6013487	6.927	6.818	+0.108
1033	CHEMBL6048496	6.927	6.464	+0.463
1034	CHEMBL5948133	6.927	6.579	+0.348
1035	CHEMBL6001660	6.927	6.835	+0.092
1036	CHEMBL5854211	6.927	6.731	+0.196
1037	CHEMBL6028598	6.927	6.456	+0.471
1038	CHEMBL5891594	6.927	6.841	+0.086
1039	CHEMBL5819792	6.927	6.176	+0.751
1040	CHEMBL5776776	6.927	6.799	+0.128
1041	CHEMBL5761824	6.927	6.825	+0.101
1042	CHEMBL5942546	6.927	6.764	+0.163
1043	CHEMBL5920414	6.927	6.833	+0.094
1044	CHEMBL6039812	6.927	6.910	+0.017
1045	CHEMBL6026968	6.927	6.829	+0.098
1046	CHEMBL5740925	6.927	6.909	+0.017
1047	CHEMBL5858538	6.927	6.570	+0.357
1048	CHEMBL5822838	6.927	5.976	+0.951
1049	CHEMBL6056251	6.927	6.841	+0.085
1050	CHEMBL5892354	6.927	6.877	+0.050
1051	CHEMBL5808700	6.927	6.508	+0.419

continued on next page

Table 5 – *continued from previous page*

#	ChEMBL ID	Measured	Predicted	Residual
1052	CHEMBL5746567	6.927	6.856	+0.071
1053	CHEMBL5791105	6.927	6.688	+0.239
1054	CHEMBL5977247	6.927	6.297	+0.629
1055	CHEMBL6041177	6.927	6.106	+0.820
1056	CHEMBL6003257	6.927	6.870	+0.056
1057	CHEMBL5950214	6.927	6.629	+0.298
1058	CHEMBL5820435	6.927	6.732	+0.195
1059	CHEMBL5913484	6.927	6.606	+0.321
1060	CHEMBL5932966	6.927	6.639	+0.288
1061	CHEMBL6016935	6.927	6.805	+0.121
1062	CHEMBL5778795	6.927	6.408	+0.519
1063	CHEMBL5798780	6.927	6.923	+0.004
1064	CHEMBL6044013	6.927	6.744	+0.182
1065	CHEMBL5984505	6.927	6.666	+0.261
1066	CHEMBL5781806	6.927	6.936	-0.009
1067	CHEMBL5954027	6.927	6.734	+0.192
1068	CHEMBL5741005	6.927	6.480	+0.447
1069	CHEMBL5759845	6.927	6.840	+0.086
1070	CHEMBL5836521	6.927	6.203	+0.724
1071	CHEMBL5825514	6.927	6.760	+0.167
1072	CHEMBL6026532	6.927	6.755	+0.171
1073	CHEMBL5846455	6.927	6.891	+0.035
1074	CHEMBL243839	6.925	6.550	+0.375
1075	CHEMBL4849337	6.925	7.098	-0.173
1076	CHEMBL5423938	6.925	7.274	-0.349
1077	CHEMBL5982869	6.922	7.353	-0.430
1078	CHEMBL4214264	6.920	6.947	-0.027
1079	CHEMBL5203711	6.920	5.920	+1.000
1080	CHEMBL4072986	6.920	6.836	+0.084
1081	CHEMBL2064403	6.920	6.161	+0.759
1082	CHEMBL1683964	6.920	7.049	-0.129
1083	CHEMBL5173862	6.920	5.804	+1.116
1084	CHEMBL1812572	6.920	7.244	-0.324
1085	CHEMBL4065503	6.920	7.201	-0.281
1086	CHEMBL508760	6.920	7.021	-0.101
1087	CHEMBL4227552	6.917	7.159	-0.242
1088	CHEMBL4082840	6.915	6.895	+0.020
1089	CHEMBL1907763	6.910	6.858	+0.052
1090	CHEMBL5757940	6.910	6.091	+0.819
1091	CHEMBL4752492	6.910	7.483	-0.573
1092	CHEMBL1907944	6.910	6.712	+0.198
1093	CHEMBL5277940	6.910	6.827	+0.083
1094	CHEMBL207037	6.910	7.261	-0.351
1095	CHEMBL5770880	6.905	7.415	-0.510
1096	CHEMBL26641	6.900	6.516	+0.384
1097	CHEMBL4073278	6.900	7.007	-0.107
1098	CHEMBL5403085	6.897	7.305	-0.408
1099	CHEMBL4873830	6.895	7.201	-0.306
1100	CHEMBL4226907	6.890	7.152	-0.262
1101	CHEMBL5284848	6.890	6.386	+0.504
1102	CHEMBL246072	6.890	6.541	+0.349
1103	CHEMBL234577	6.890	6.696	+0.194
1104	CHEMBL120319	6.890	6.733	+0.157
1105	CHEMBL4077608	6.890	7.118	-0.228
1106	CHEMBL398793	6.890	6.928	-0.038
1107	CHEMBL3612590	6.890	6.408	+0.482
1108	CHEMBL180227	6.890	6.744	+0.146
1109	CHEMBL5280555	6.888	7.381	-0.493
1110	CHEMBL5173986	6.885	7.137	-0.252

continued on next page

Table 5 – *continued from previous page*

#	ChEMBL ID	Measured	Predicted	Residual
1111	CHEMBL5810016	6.880	7.040	-0.160
1112	CHEMBL3753874	6.880	7.825	-0.945
1113	CHEMBL5747445	6.875	8.414	-1.539
1114	CHEMBL5836093	6.873	7.819	-0.946
1115	CHEMBL4459978	6.873	7.202	-0.329
1116	CHEMBL4646030	6.873	6.938	-0.065
1117	CHEMBL4849859	6.870	6.272	+0.598
1118	CHEMBL5907912	6.870	8.003	-1.133
1119	CHEMBL2180204	6.870	6.909	-0.039
1120	CHEMBL3622642	6.870	7.311	-0.441
1121	CHEMBL5087793	6.860	7.340	-0.480
1122	CHEMBL4073912	6.860	7.059	-0.199
1123	CHEMBL2086760	6.860	6.271	+0.589
1124	CHEMBL3775046	6.860	7.032	-0.172
1125	CHEMBL4463766	6.857	7.313	-0.456
1126	CHEMBL3752139	6.850	7.805	-0.955
1127	CHEMBL301612	6.845	5.913	+0.932
1128	CHEMBL69960	6.843	6.835	+0.008
1129	CHEMBL6014278	6.840	7.923	-1.083
1130	CHEMBL3087208	6.840	7.407	-0.567
1131	CHEMBL3297891	6.840	6.938	-0.098
1132	CHEMBL2093928	6.840	6.853	-0.013
1133	CHEMBL4099445	6.837	7.500	-0.663
1134	CHEMBL5630498	6.835	7.566	-0.731
1135	CHEMBL6026303	6.830	7.215	-0.385
1136	CHEMBL5290595	6.830	6.538	+0.292
1137	CHEMBL5523403	6.830	6.334	+0.496
1138	CHEMBL5276760	6.830	7.489	-0.659
1139	CHEMBL4091166	6.825	6.805	+0.020
1140	CHEMBL5209100	6.820	5.726	+1.094
1141	CHEMBL321193	6.820	5.979	+0.841
1142	CHEMBL3416437	6.820	6.981	-0.161
1143	CHEMBL4450061	6.820	7.256	-0.436
1144	CHEMBL2087354	6.820	7.609	-0.789
1145	CHEMBL391387	6.820	7.170	-0.350
1146	CHEMBL1641990	6.820	5.775	+1.045
1147	CHEMBL6147822	6.810	6.914	-0.104
1148	CHEMBL337027	6.800	5.848	+0.952
1149	CHEMBL3945555	6.800	6.510	+0.290
1150	CHEMBL5268062	6.800	6.561	+0.239
1151	CHEMBL5619771	6.800	6.552	+0.248
1152	CHEMBL5197952	6.800	6.972	-0.172
1153	CHEMBL3903707	6.794	6.877	-0.083
1154	CHEMBL5272533	6.785	6.595	+0.190
1155	CHEMBL5287295	6.785	6.868	-0.083
1156	CHEMBL5183327	6.780	7.243	-0.463
1157	CHEMBL6169985	6.780	6.831	-0.051
1158	CHEMBL5815306	6.780	6.958	-0.178
1159	CHEMBL5876067	6.780	6.958	-0.178
1160	CHEMBL5992872	6.780	6.958	-0.178
1161	CHEMBL5849328	6.780	6.958	-0.178
1162	CHEMBL3233770	6.775	6.994	-0.219
1163	CHEMBL3263964	6.770	6.318	+0.452
1164	CHEMBL5809369	6.760	6.639	+0.121
1165	CHEMBL6037302	6.760	6.598	+0.162
1166	CHEMBL5854828	6.760	6.609	+0.151
1167	CHEMBL5964110	6.760	6.684	+0.076
1168	CHEMBL5856323	6.760	6.335	+0.425
1169	CHEMBL24979	6.760	6.719	+0.041

continued on next page

Table 5 – *continued from previous page*

#	ChEMBL ID	Measured	Predicted	Residual
1170	CHEMBL5922408	6.760	6.639	+0.121
1171	CHEMBL5887560	6.760	6.689	+0.071
1172	CHEMBL5773640	6.760	6.820	-0.060
1173	CHEMBL1276179	6.760	6.264	+0.496
1174	CHEMBL1907764	6.760	6.712	+0.048
1175	CHEMBL5769228	6.760	6.628	+0.132
1176	CHEMBL4516287	6.753	7.278	-0.525
1177	CHEMBL4087243	6.753	7.571	-0.818
1178	CHEMBL4648924	6.750	7.072	-0.322
1179	CHEMBL552859	6.750	5.868	+0.882
1180	CHEMBL455433	6.750	6.282	+0.468
1181	CHEMBL2403108	6.747	6.605	+0.143
1182	CHEMBL3414594	6.740	6.789	-0.049
1183	CHEMBL3969628	6.740	7.123	-0.383
1184	CHEMBL5747102	6.738	6.432	+0.306
1185	CHEMBL4792279	6.730	6.564	+0.166
1186	CHEMBL3919271	6.725	6.964	-0.239
1187	CHEMBL247915	6.720	6.784	-0.064
1188	CHEMBL392774	6.720	6.763	-0.043
1189	CHEMBL247711	6.720	6.974	-0.254
1190	CHEMBL109631	6.720	5.825	+0.895
1191	CHEMBL1812573	6.720	6.071	+0.649
1192	CHEMBL453398	6.720	6.609	+0.111
1193	CHEMBL1090355	6.720	6.795	-0.075
1194	CHEMBL241918	6.715	6.529	+0.186
1195	CHEMBL5206298	6.703	6.857	-0.153
1196	CHEMBL6167047	6.700	6.705	-0.005
1197	CHEMBL2408829	6.700	6.652	+0.048
1198	CHEMBL393787	6.700	6.883	-0.183
1199	CHEMBL5867430	6.700	5.850	+0.850
1200	CHEMBL4475770	6.700	6.280	+0.420
1201	CHEMBL371863	6.700	6.062	+0.638
1202	CHEMBL322298	6.700	6.116	+0.584
1203	CHEMBL5195593	6.700	5.654	+1.046
1204	CHEMBL4291935	6.695	6.854	-0.159
1205	CHEMBL4289728	6.690	6.763	-0.073
1206	CHEMBL4746030	6.690	7.598	-0.908
1207	CHEMBL6167751	6.690	5.956	+0.734
1208	CHEMBL400413	6.680	6.712	-0.032
1209	CHEMBL332765	6.680	6.686	-0.006
1210	CHEMBL69964	6.680	6.991	-0.311
1211	CHEMBL4850856	6.680	6.355	+0.325
1212	CHEMBL178096	6.680	6.737	-0.057
1213	CHEMBL4777554	6.680	6.983	-0.303
1214	CHEMBL5757288	6.680	6.441	+0.239
1215	CHEMBL3924160	6.680	7.729	-1.049
1216	CHEMBL4575697	6.675	7.314	-0.639
1217	CHEMBL5789149	6.673	6.638	+0.035
1218	CHEMBL4642574	6.672	6.938	-0.265
1219	CHEMBL1095761	6.670	5.632	+1.038
1220	CHEMBL5619554	6.670	6.660	+0.010
1221	CHEMBL122411	6.670	6.757	-0.087
1222	CHEMBL3774396	6.670	6.729	-0.059
1223	CHEMBL4746023	6.670	6.406	+0.264
1224	CHEMBL360979	6.670	6.933	-0.263
1225	CHEMBL338935	6.660	5.919	+0.741
1226	CHEMBL50647	6.660	5.999	+0.661
1227	CHEMBL5286373	6.660	6.104	+0.556
1228	CHEMBL362611	6.660	7.044	-0.384

continued on next page

Table 5 – *continued from previous page*

#	ChEMBL ID	Measured	Predicted	Residual
1229	CHEMBL5900414	6.660	7.320	-0.660
1230	CHEMBL104153	6.660	6.049	+0.611
1231	CHEMBL5898708	6.660	6.565	+0.095
1232	CHEMBL5173070	6.658	6.620	+0.038
1233	CHEMBL3233780	6.655	8.027	-1.372
1234	CHEMBL5558740	6.650	6.332	+0.318
1235	CHEMBL4104717	6.650	6.586	+0.064
1236	CHEMBL4539566	6.645	6.021	+0.624
1237	CHEMBL4786701	6.645	7.142	-0.497
1238	CHEMBL5265927	6.640	6.373	+0.267
1239	CHEMBL4753913	6.640	6.397	+0.243
1240	CHEMBL5613826	6.640	5.735	+0.905
1241	CHEMBL1093235	6.640	6.877	-0.237
1242	CHEMBL219557	6.640	6.065	+0.575
1243	CHEMBL4756092	6.630	6.970	-0.340
1244	CHEMBL1242376	6.630	6.178	+0.452
1245	CHEMBL1958215	6.620	4.811	+1.809
1246	CHEMBL4164936	6.620	5.992	+0.628
1247	CHEMBL175621	6.620	7.096	-0.476
1248	CHEMBL3912361	6.620	7.043	-0.423
1249	CHEMBL138125	6.620	7.198	-0.578
1250	CHEMBL4751768	6.620	7.502	-0.882
1251	CHEMBL452980	6.620	6.746	-0.126
1252	CHEMBL153763	6.610	6.893	-0.283
1253	CHEMBL25425	6.610	6.833	-0.223
1254	CHEMBL113023	6.600	6.269	+0.331
1255	CHEMBL6058480	6.600	7.808	-1.208
1256	CHEMBL247101	6.600	7.110	-0.510
1257	CHEMBL5186969	6.600	6.918	-0.318
1258	CHEMBL5176588	6.590	7.131	-0.541
1259	CHEMBL338155	6.580	6.268	+0.312
1260	CHEMBL5810815	6.580	8.202	-1.622
1261	CHEMBL1683975	6.570	7.385	-0.815
1262	CHEMBL1242202	6.570	5.541	+1.029
1263	CHEMBL5195669	6.570	6.832	-0.262
1264	CHEMBL4069277	6.565	7.508	-0.943
1265	CHEMBL5813164	6.560	6.799	-0.239
1266	CHEMBL6000322	6.560	6.556	+0.004
1267	CHEMBL3894102	6.560	6.919	-0.359
1268	CHEMBL5892454	6.560	6.802	-0.242
1269	CHEMBL5962521	6.560	6.580	-0.020
1270	CHEMBL5836601	6.560	6.940	-0.380
1271	CHEMBL4535881	6.560	5.906	+0.654
1272	CHEMBL426051	6.560	7.040	-0.480
1273	CHEMBL6022727	6.560	6.532	+0.028
1274	CHEMBL6011191	6.560	6.757	-0.197
1275	CHEMBL5807906	6.560	6.849	-0.289
1276	CHEMBL5932064	6.560	6.766	-0.206
1277	CHEMBL4227281	6.555	6.431	+0.124
1278	CHEMBL4226965	6.553	6.803	-0.250
1279	CHEMBL3753461	6.550	7.203	-0.653
1280	CHEMBL5077665	6.550	7.555	-1.005
1281	CHEMBL362824	6.550	7.094	-0.544
1282	CHEMBL4213741	6.550	7.175	-0.625
1283	CHEMBL5793411	6.550	6.795	-0.245
1284	CHEMBL121260	6.540	6.729	-0.189
1285	CHEMBL3133751	6.540	6.469	+0.071
1286	CHEMBL5612514	6.540	5.705	+0.835
1287	CHEMBL5755984	6.535	6.877	-0.342

continued on next page

Table 5 – *continued from previous page*

#	ChEMBL ID	Measured	Predicted	Residual
1288	CHEMBL4086821	6.530	6.607	-0.077
1289	CHEMBL6000972	6.520	6.511	+0.009
1290	CHEMBL5747418	6.520	6.689	-0.169
1291	CHEMBL5773016	6.520	6.625	-0.105
1292	CHEMBL5930490	6.520	6.283	+0.237
1293	CHEMBL5820070	6.520	6.665	-0.145
1294	CHEMBL5953610	6.520	6.776	-0.256
1295	CHEMBL5741017	6.520	7.111	-0.591
1296	CHEMBL6036892	6.520	8.241	-1.721
1297	CHEMBL5893059	6.510	5.376	+1.134
1298	CHEMBL3403514	6.510	6.622	-0.112
1299	CHEMBL3233781	6.505	8.127	-1.622
1300	CHEMBL6188992	6.500	5.456	+1.044
1301	CHEMBL4277822	6.500	6.892	-0.392
1302	CHEMBL6174044	6.500	7.623	-1.123
1303	CHEMBL1242568	6.500	5.765	+0.735
1304	CHEMBL5815729	6.500	7.016	-0.516
1305	CHEMBL5740251	6.498	5.934	+0.564
1306	CHEMBL5437181	6.495	7.200	-0.705
1307	CHEMBL4647257	6.490	7.379	-0.889
1308	CHEMBL6055510	6.490	6.514	-0.024
1309	CHEMBL4583545	6.490	6.929	-0.439
1310	CHEMBL360211	6.490	6.994	-0.504
1311	CHEMBL5844317	6.480	7.221	-0.741
1312	CHEMBL324926	6.480	5.985	+0.495
1313	CHEMBL249511	6.480	7.153	-0.673
1314	CHEMBL129946	6.480	5.927	+0.553
1315	CHEMBL6024685	6.477	7.396	-0.919
1316	CHEMBL179310	6.470	6.546	-0.076
1317	CHEMBL5087785	6.470	7.563	-1.093
1318	CHEMBL5558842	6.470	6.659	-0.189
1319	CHEMBL5280292	6.470	6.410	+0.060
1320	CHEMBL6171608	6.470	6.411	+0.059
1321	CHEMBL6166776	6.470	7.385	-0.915
1322	CHEMBL5427116	6.460	6.740	-0.280
1323	CHEMBL2324871	6.460	5.869	+0.591
1324	CHEMBL265629	6.460	6.653	-0.193
1325	CHEMBL25610	6.460	6.826	-0.366
1326	CHEMBL5169328	6.450	7.481	-1.031
1327	CHEMBL3622659	6.450	7.989	-1.539
1328	CHEMBL5939208	6.450	8.387	-1.937
1329	CHEMBL6043477	6.450	6.103	+0.347
1330	CHEMBL5170278	6.450	7.434	-0.984
1331	CHEMBL5200760	6.450	7.446	-0.996
1332	CHEMBL6162368	6.450	6.526	-0.076
1333	CHEMBL5273342	6.450	6.394	+0.056
1334	CHEMBL2426277	6.443	5.937	+0.506
1335	CHEMBL6000385	6.440	6.355	+0.085
1336	CHEMBL4541573	6.440	6.212	+0.228
1337	CHEMBL552860	6.440	5.879	+0.561
1338	CHEMBL6061692	6.440	6.912	-0.472
1339	CHEMBL1242656	6.440	5.594	+0.846
1340	CHEMBL5570514	6.430	6.362	+0.068
1341	CHEMBL4092788	6.430	7.545	-1.115
1342	CHEMBL5077901	6.420	7.646	-1.226
1343	CHEMBL574059	6.420	5.932	+0.488
1344	CHEMBL1242293	6.420	5.876	+0.544
1345	CHEMBL1950289	6.420	6.684	-0.264
1346	CHEMBL243628	6.415	6.758	-0.343

continued on next page

Table 5 – *continued from previous page*

#	ChEMBL ID	Measured	Predicted	Residual
1347	CHEMBL1242853	6.410	6.085	+0.325
1348	CHEMBL1234815	6.410	5.321	+1.089
1349	CHEMBL574058	6.410	6.162	+0.248
1350	CHEMBL127367	6.410	7.607	-1.197
1351	CHEMBL4446992	6.400	5.824	+0.576
1352	CHEMBL522116	6.400	5.386	+1.014
1353	CHEMBL2426286	6.397	6.223	+0.174
1354	CHEMBL1830280	6.390	5.479	+0.911
1355	CHEMBL175933	6.390	7.049	-0.659
1356	CHEMBL3962032	6.380	7.062	-0.682
1357	CHEMBL3416620	6.380	6.667	-0.287
1358	CHEMBL102651	6.380	6.853	-0.473
1359	CHEMBL5185772	6.380	6.614	-0.234
1360	CHEMBL420059	6.380	6.967	-0.587
1361	CHEMBL4759405	6.370	6.320	+0.050
1362	CHEMBL3353404	6.365	6.374	-0.009
1363	CHEMBL3301625	6.360	6.746	-0.386
1364	CHEMBL1242201	6.360	5.686	+0.674
1365	CHEMBL6142419	6.360	6.610	-0.250
1366	CHEMBL4634817	6.350	5.797	+0.553
1367	CHEMBL101253	6.340	5.782	+0.558
1368	CHEMBL5173075	6.340	6.242	+0.098
1369	CHEMBL5171139	6.340	7.136	-0.796
1370	CHEMBL2385959	6.340	7.337	-0.997
1371	CHEMBL511478	6.340	7.114	-0.774
1372	CHEMBL3775897	6.340	6.793	-0.453
1373	CHEMBL3774808	6.340	6.776	-0.436
1374	CHEMBL6059204	6.337	7.328	-0.991
1375	CHEMBL5218821	6.335	6.394	-0.059
1376	CHEMBL6161385	6.330	7.654	-1.324
1377	CHEMBL5951478	6.330	8.426	-2.096
1378	CHEMBL5086054	6.330	7.380	-1.050
1379	CHEMBL5872048	6.323	7.352	-1.029
1380	CHEMBL5854207	6.320	6.539	-0.219
1381	CHEMBL6058583	6.320	6.249	+0.071
1382	CHEMBL111339	6.320	5.840	+0.480
1383	CHEMBL6020494	6.320	8.651	-2.331
1384	CHEMBL5749382	6.320	6.215	+0.105
1385	CHEMBL5905455	6.320	6.271	+0.049
1386	CHEMBL186580	6.310	7.136	-0.826
1387	CHEMBL5416668	6.310	6.164	+0.146
1388	CHEMBL2426288	6.302	6.552	-0.250
1389	CHEMBL4469900	6.300	6.510	-0.210
1390	CHEMBL396079	6.300	6.797	-0.497
1391	CHEMBL5275872	6.300	5.675	+0.625
1392	CHEMBL5082422	6.300	6.880	-0.580
1393	CHEMBL293584	6.300	5.231	+1.069
1394	CHEMBL1254523	6.300	6.075	+0.225
1395	CHEMBL3622641	6.290	7.522	-1.232
1396	CHEMBL325589	6.290	5.906	+0.384
1397	CHEMBL2070045	6.290	5.434	+0.856
1398	CHEMBL3930270	6.290	6.539	-0.249
1399	CHEMBL2324873	6.280	6.161	+0.119
1400	CHEMBL2324870	6.280	6.147	+0.133
1401	CHEMBL3917294	6.280	7.449	-1.169
1402	CHEMBL3900059	6.280	6.839	-0.559
1403	CHEMBL205652	6.270	6.438	-0.168
1404	CHEMBL1242385	6.270	5.996	+0.274
1405	CHEMBL307209	6.270	6.997	-0.727

continued on next page

Table 5 – *continued from previous page*

#	ChEMBL ID	Measured	Predicted	Residual
1406	CHEMBL5866027	6.260	6.573	-0.313
1407	CHEMBL6055007	6.260	5.534	+0.726
1408	CHEMBL6046464	6.260	6.845	-0.585
1409	CHEMBL6062263	6.260	6.659	-0.399
1410	CHEMBL5857633	6.260	6.461	-0.201
1411	CHEMBL6061137	6.260	6.056	+0.204
1412	CHEMBL5845905	6.260	6.504	-0.244
1413	CHEMBL5784829	6.260	6.185	+0.075
1414	CHEMBL5755010	6.260	6.407	-0.147
1415	CHEMBL5798581	6.260	6.093	+0.167
1416	CHEMBL5831631	6.260	6.137	+0.123
1417	CHEMBL6003806	6.260	6.396	-0.136
1418	CHEMBL4468148	6.260	6.428	-0.168
1419	CHEMBL5979639	6.260	5.657	+0.603
1420	CHEMBL5868212	6.260	5.641	+0.619
1421	CHEMBL5835276	6.260	5.937	+0.323
1422	CHEMBL5949435	6.260	5.683	+0.577
1423	CHEMBL5840304	6.260	6.077	+0.183
1424	CHEMBL5905715	6.260	5.626	+0.634
1425	CHEMBL5770897	6.260	6.101	+0.159
1426	CHEMBL6025324	6.260	6.814	-0.554
1427	CHEMBL5824026	6.260	6.038	+0.222
1428	CHEMBL5798489	6.260	6.545	-0.285
1429	CHEMBL5740635	6.260	5.929	+0.331
1430	CHEMBL5879924	6.260	5.880	+0.380
1431	CHEMBL5943719	6.260	6.598	-0.338
1432	CHEMBL5856597	6.260	6.612	-0.352
1433	CHEMBL5821638	6.260	5.952	+0.308
1434	CHEMBL5780633	6.260	6.776	-0.516
1435	CHEMBL5760493	6.260	6.033	+0.227
1436	CHEMBL5773851	6.260	6.704	-0.444
1437	CHEMBL5906808	6.260	5.641	+0.619
1438	CHEMBL5776801	6.260	5.302	+0.958
1439	CHEMBL5804118	6.260	5.473	+0.787
1440	CHEMBL5944879	6.260	5.702	+0.558
1441	CHEMBL5928380	6.260	6.425	-0.165
1442	CHEMBL5889296	6.260	5.590	+0.670
1443	CHEMBL6002504	6.260	5.684	+0.576
1444	CHEMBL5927148	6.260	5.650	+0.610
1445	CHEMBL6008308	6.260	5.658	+0.602
1446	CHEMBL5808792	6.260	6.137	+0.123
1447	CHEMBL5885178	6.260	6.183	+0.077
1448	CHEMBL5922580	6.260	6.371	-0.111
1449	CHEMBL5833583	6.260	6.620	-0.360
1450	CHEMBL5802687	6.260	5.959	+0.301
1451	CHEMBL5888462	6.260	6.694	-0.434
1452	CHEMBL5929629	6.260	6.771	-0.511
1453	CHEMBL5756489	6.260	6.814	-0.554
1454	CHEMBL6028388	6.260	6.908	-0.648
1455	CHEMBL5746447	6.260	5.461	+0.799
1456	CHEMBL5833997	6.260	6.684	-0.424
1457	CHEMBL5896750	6.260	6.508	-0.248
1458	CHEMBL5939963	6.260	6.667	-0.407
1459	CHEMBL6062944	6.260	6.225	+0.035
1460	CHEMBL5997917	6.260	6.758	-0.498
1461	CHEMBL5916856	6.260	6.508	-0.248
1462	CHEMBL5900800	6.260	6.861	-0.601
1463	CHEMBL3234736	6.260	6.242	+0.018
1464	CHEMBL5758784	6.260	6.784	-0.524

continued on next page

Table 5 – *continued from previous page*

#	ChEMBL ID	Measured	Predicted	Residual
1465	CHEMBL4634127	6.250	7.301	-1.051
1466	CHEMBL3622621	6.250	7.370	-1.120
1467	CHEMBL2437457	6.243	6.189	+0.054
1468	CHEMBL3234737	6.240	7.292	-1.052
1469	CHEMBL601040	6.240	4.733	+1.507
1470	CHEMBL362233	6.240	7.066	-0.826
1471	CHEMBL3263965	6.240	6.313	-0.073
1472	CHEMBL2424676	6.240	6.375	-0.135
1473	CHEMBL6006366	6.240	7.320	-1.080
1474	CHEMBL5942345	6.227	7.320	-1.094
1475	CHEMBL591897	6.220	6.607	-0.387
1476	CHEMBL6060780	6.220	6.036	+0.184
1477	CHEMBL5415565	6.220	7.124	-0.904
1478	CHEMBL555797	6.220	7.103	-0.883
1479	CHEMBL5892301	6.216	6.432	-0.216
1480	CHEMBL4793639	6.210	5.846	+0.364
1481	CHEMBL290106	6.210	5.485	+0.725
1482	CHEMBL4462528	6.210	6.828	-0.618
1483	CHEMBL2441273	6.200	6.115	+0.085
1484	CHEMBL5398727	6.200	5.390	+0.810
1485	CHEMBL4102288	6.200	6.515	-0.315
1486	CHEMBL1934621	6.190	7.188	-0.998
1487	CHEMBL1683956	6.190	6.807	-0.617
1488	CHEMBL3263373	6.190	6.385	-0.195
1489	CHEMBL4790553	6.180	6.229	-0.049
1490	CHEMBL251315	6.180	7.265	-1.085
1491	CHEMBL2325087	6.180	6.348	-0.168
1492	CHEMBL2179124	6.175	6.381	-0.206
1493	CHEMBL6007655	6.172	5.985	+0.187
1494	CHEMBL542887	6.170	5.957	+0.213
1495	CHEMBL111434	6.170	6.159	+0.011
1496	CHEMBL1645465	6.170	7.012	-0.842
1497	CHEMBL5945723	6.170	5.838	+0.332
1498	CHEMBL382041	6.170	6.802	-0.632
1499	CHEMBL2070048	6.160	5.173	+0.987
1500	CHEMBL4647325	6.160	5.657	+0.503
1501	CHEMBL195204	6.160	6.893	-0.733
1502	CHEMBL4873202	6.160	5.964	+0.196
1503	CHEMBL419022	6.160	6.110	+0.050
1504	CHEMBL5280288	6.150	6.240	-0.090
1505	CHEMBL5404368	6.150	6.554	-0.404
1506	CHEMBL4795888	6.150	5.866	+0.284
1507	CHEMBL298679	6.140	6.163	-0.023
1508	CHEMBL6025577	6.140	6.977	-0.837
1509	CHEMBL391738	6.140	5.984	+0.156
1510	CHEMBL5290879	6.140	6.563	-0.423
1511	CHEMBL5759916	6.137	5.791	+0.346
1512	CHEMBL91428	6.135	6.496	-0.361
1513	CHEMBL3916141	6.130	6.669	-0.539
1514	CHEMBL153409	6.130	7.502	-1.372
1515	CHEMBL149512	6.130	7.012	-0.882
1516	CHEMBL4649213	6.130	5.730	+0.400
1517	CHEMBL1242295	6.130	6.273	-0.143
1518	CHEMBL5850357	6.127	6.120	+0.006
1519	CHEMBL4591327	6.127	6.694	-0.567
1520	CHEMBL47940	6.125	6.289	-0.164
1521	CHEMBL5766497	6.120	6.753	-0.633
1522	CHEMBL4855312	6.120	6.487	-0.367
1523	CHEMBL1242289	6.120	5.780	+0.340

continued on next page

Table 5 – *continued from previous page*

#	ChEMBL ID	Measured	Predicted	Residual
1524	CHEMBL589589	6.120	4.747	+1.373
1525	CHEMBL1242199	6.120	6.041	+0.079
1526	CHEMBL5776043	6.120	6.276	-0.156
1527	CHEMBL5195321	6.113	6.156	-0.043
1528	CHEMBL6023442	6.110	6.036	+0.074
1529	CHEMBL5560711	6.100	6.525	-0.425
1530	CHEMBL22978	6.100	5.981	+0.119
1531	CHEMBL490569	6.100	6.239	-0.139
1532	CHEMBL5989746	6.100	5.831	+0.269
1533	CHEMBL6020779	6.095	7.460	-1.365
1534	CHEMBL5429694	6.080	6.206	-0.126
1535	CHEMBL4444231	6.060	6.668	-0.608
1536	CHEMBL4226151	6.050	7.096	-1.046
1537	CHEMBL5280575	6.050	6.486	-0.436
1538	CHEMBL2047029	6.050	6.484	-0.434
1539	CHEMBL2385976	6.050	6.345	-0.295
1540	CHEMBL2070201	6.040	6.466	-0.426
1541	CHEMBL4562744	6.030	6.130	-0.100
1542	CHEMBL1812571	6.030	6.335	-0.305
1543	CHEMBL5208369	6.030	6.753	-0.723
1544	CHEMBL589831	6.030	4.733	+1.297
1545	CHEMBL1645464	6.020	7.111	-1.091
1546	CHEMBL592211	6.020	5.934	+0.086
1547	CHEMBL5290503	6.020	6.823	-0.803
1548	CHEMBL4861796	6.020	5.446	+0.574
1549	CHEMBL3397276	6.020	7.517	-1.497
1550	CHEMBL1241490	6.020	5.126	+0.894
1551	CHEMBL1081849	6.010	6.203	-0.193
1552	CHEMBL436137	6.010	6.249	-0.239
1553	CHEMBL3897153	6.000	6.404	-0.404
1554	CHEMBL4097778	6.000	6.030	-0.030
1555	CHEMBL4527630	6.000	6.299	-0.299
1556	CHEMBL319244	6.000	5.540	+0.460
1557	CHEMBL3906473	6.000	6.195	-0.195
1558	CHEMBL2385977	6.000	6.507	-0.507
1559	CHEMBL5962946	6.000	6.347	-0.347
1560	CHEMBL3943489	6.000	6.954	-0.954
1561	CHEMBL1242754	6.000	5.166	+0.834
1562	CHEMBL184682	6.000	7.173	-1.173
1563	CHEMBL3969511	6.000	6.993	-0.993
1564	CHEMBL5179288	6.000	6.737	-0.737
1565	CHEMBL62176	6.000	5.093	+0.907
1566	CHEMBL5983716	5.995	6.419	-0.424
1567	CHEMBL6017443	5.990	6.259	-0.269
1568	CHEMBL4851555	5.990	6.365	-0.375
1569	CHEMBL1812560	5.990	7.701	-1.711
1570	CHEMBL2324872	5.980	6.142	-0.162
1571	CHEMBL5815780	5.980	6.648	-0.668
1572	CHEMBL5910943	5.980	6.578	-0.598
1573	CHEMBL954	5.980	5.436	+0.544
1574	CHEMBL3814940	5.973	7.345	-1.372
1575	CHEMBL590714	5.970	5.947	+0.023
1576	CHEMBL1958212	5.970	5.010	+0.960
1577	CHEMBL131020	5.960	6.142	-0.182
1578	CHEMBL1683949	5.960	6.273	-0.313
1579	CHEMBL3085380	5.960	6.965	-1.005
1580	CHEMBL131426	5.960	5.984	-0.024
1581	CHEMBL216048	5.955	5.417	+0.538
1582	CHEMBL4861493	5.950	5.455	+0.495

continued on next page

Table 5 – *continued from previous page*

#	ChEMBL ID	Measured	Predicted	Residual
1583	CHEMBL601828	5.950	5.596	+0.354
1584	CHEMBL2047026	5.950	6.022	-0.072
1585	CHEMBL5988212	5.942	6.036	-0.094
1586	CHEMBL1958216	5.940	5.000	+0.940
1587	CHEMBL2070200	5.940	6.468	-0.528
1588	CHEMBL5979972	5.940	6.470	-0.530
1589	CHEMBL4779681	5.935	6.553	-0.618
1590	CHEMBL2426290	5.933	6.495	-0.561
1591	CHEMBL308593	5.930	6.727	-0.797
1592	CHEMBL5401604	5.930	5.958	-0.028
1593	CHEMBL2442333	5.930	5.723	+0.207
1594	CHEMBL6011521	5.927	6.418	-0.491
1595	CHEMBL1254443	5.920	6.127	-0.207
1596	CHEMBL1812431	5.920	7.493	-1.573
1597	CHEMBL3233761	5.920	7.267	-1.347
1598	CHEMBL3397275	5.920	7.517	-1.597
1599	CHEMBL1945649	5.920	5.626	+0.294
1600	CHEMBL2426377	5.920	6.820	-0.900
1601	CHEMBL2442332	5.910	5.795	+0.115
1602	CHEMBL2426289	5.903	6.767	-0.864
1603	CHEMBL3604928	5.900	6.058	-0.158
1604	CHEMBL94019	5.900	7.244	-1.344
1605	CHEMBL1958221	5.900	5.085	+0.815
1606	CHEMBL335718	5.890	5.889	+0.001
1607	CHEMBL444337	5.890	5.987	-0.097
1608	CHEMBL1641992	5.890	5.574	+0.316
1609	CHEMBL5572677	5.890	6.357	-0.467
1610	CHEMBL111365	5.890	6.026	-0.136
1611	CHEMBL3612571	5.890	6.913	-1.023
1612	CHEMBL5760517	5.875	5.641	+0.234
1613	CHEMBL4868965	5.870	5.548	+0.322
1614	CHEMBL5935382	5.870	5.892	-0.022
1615	CHEMBL177688	5.870	5.841	+0.029
1616	CHEMBL1958022	5.870	4.923	+0.947
1617	CHEMBL5592551	5.860	5.846	+0.014
1618	CHEMBL205783	5.860	6.749	-0.889
1619	CHEMBL2442329	5.860	5.670	+0.190
1620	CHEMBL5977568	5.860	6.331	-0.471
1621	CHEMBL3416619	5.860	6.095	-0.235
1622	CHEMBL4447289	5.855	6.788	-0.933
1623	CHEMBL1242758	5.850	5.945	-0.095
1624	CHEMBL2047022	5.850	5.931	-0.081
1625	CHEMBL2047016	5.850	5.928	-0.078
1626	CHEMBL1240554	5.850	5.793	+0.057
1627	CHEMBL589830	5.830	4.827	+1.003
1628	CHEMBL544770	5.820	5.660	+0.160
1629	CHEMBL2424796	5.820	6.083	-0.263
1630	CHEMBL5179538	5.815	6.754	-0.939
1631	CHEMBL6039443	5.815	5.739	+0.076
1632	CHEMBL2316142	5.815	5.321	+0.494
1633	CHEMBL2047025	5.810	5.893	-0.083
1634	CHEMBL401930	5.800	6.280	-0.480
1635	CHEMBL78685	5.800	7.680	-1.880
1636	CHEMBL2437476	5.797	6.987	-1.191
1637	CHEMBL6043242	5.796	6.435	-0.639
1638	CHEMBL172333	5.795	5.937	-0.142
1639	CHEMBL5804483	5.795	5.630	+0.165
1640	CHEMBL3425667	5.790	6.148	-0.358
1641	CHEMBL5982923	5.790	5.953	-0.163

continued on next page

Table 5 – *continued from previous page*

#	ChEMBL ID	Measured	Predicted	Residual
1642	CHEMBL309276	5.790	5.006	+0.784
1643	CHEMBL398963	5.790	7.173	-1.383
1644	CHEMBL2316140	5.790	5.295	+0.495
1645	CHEMBL4073874	5.787	6.952	-1.166
1646	CHEMBL48614	5.780	5.877	-0.097
1647	CHEMBL1958214	5.780	4.998	+0.782
1648	CHEMBL6007759	5.780	6.165	-0.385
1649	CHEMBL499344	5.780	5.914	-0.134
1650	CHEMBL92006	5.775	6.786	-1.011
1651	CHEMBL336409	5.773	4.912	+0.862
1652	CHEMBL158275	5.770	5.032	+0.738
1653	CHEMBL3604930	5.765	6.060	-0.295
1654	CHEMBL3604935	5.760	6.064	-0.304
1655	CHEMBL4472571	5.760	6.403	-0.643
1656	CHEMBL592711	5.760	4.825	+0.935
1657	CHEMBL3604926	5.760	6.207	-0.447
1658	CHEMBL3604927	5.757	6.041	-0.284
1659	CHEMBL1242111	5.750	5.930	-0.180
1660	CHEMBL2047021	5.750	5.852	-0.102
1661	CHEMBL2437478	5.749	6.321	-0.572
1662	CHEMBL3604936	5.733	6.047	-0.314
1663	CHEMBL590526	5.730	5.801	-0.071
1664	CHEMBL5977533	5.730	6.961	-1.231
1665	CHEMBL5281446	5.727	7.620	-1.893
1666	CHEMBL1328065	5.720	6.063	-0.343
1667	CHEMBL4081727	5.720	6.024	-0.304
1668	CHEMBL338448	5.720	5.810	-0.090
1669	CHEMBL5903840	5.720	6.825	-1.105
1670	CHEMBL4780210	5.720	6.313	-0.593
1671	CHEMBL2316158	5.715	5.467	+0.248
1672	CHEMBL342792	5.715	5.165	+0.550
1673	CHEMBL4289083	5.710	6.335	-0.625
1674	CHEMBL5992914	5.710	5.694	+0.016
1675	CHEMBL592426	5.700	5.632	+0.068
1676	CHEMBL1958213	5.700	5.087	+0.613
1677	CHEMBL510730	5.700	5.562	+0.138
1678	CHEMBL1958209	5.690	4.822	+0.868
1679	CHEMBL5827368	5.690	7.084	-1.394
1680	CHEMBL2316157	5.690	5.142	+0.548
1681	CHEMBL1945447	5.680	5.376	+0.304
1682	CHEMBL3358006	5.680	6.430	-0.750
1683	CHEMBL356324	5.680	5.190	+0.490
1684	CHEMBL1095957	5.680	5.544	+0.136
1685	CHEMBL2442324	5.670	5.738	-0.068
1686	CHEMBL2316143	5.670	5.333	+0.337
1687	CHEMBL5870193	5.663	5.763	-0.100
1688	CHEMBL3233794	5.660	7.911	-2.251
1689	CHEMBL360209	5.660	7.262	-1.602
1690	CHEMBL3814021	5.655	7.216	-1.561
1691	CHEMBL3092320	5.655	5.866	-0.211
1692	CHEMBL592427	5.650	5.624	+0.026
1693	CHEMBL2047017	5.650	5.864	-0.214
1694	CHEMBL2047024	5.650	5.996	-0.346
1695	CHEMBL2316141	5.650	5.265	+0.385
1696	CHEMBL5822835	5.640	5.795	-0.155
1697	CHEMBL2426248	5.640	6.736	-1.096
1698	CHEMBL118000	5.640	6.496	-0.856
1699	CHEMBL1958034	5.640	4.766	+0.874
1700	CHEMBL3787112	5.640	5.998	-0.358

continued on next page

Table 5 – *continued from previous page*

#	ChEMBL ID	Measured	Predicted	Residual
1701	CHEMBL5825749	5.633	5.763	-0.130
1702	CHEMBL2047015	5.630	5.961	-0.331
1703	CHEMBL144791	5.627	5.277	+0.350
1704	CHEMBL5844933	5.620	6.458	-0.838
1705	CHEMBL130049	5.620	5.960	-0.340
1706	CHEMBL5981967	5.620	6.238	-0.618
1707	CHEMBL5964825	5.620	6.380	-0.760
1708	CHEMBL5860263	5.620	6.837	-1.217
1709	CHEMBL5938505	5.620	6.536	-0.916
1710	CHEMBL5826446	5.620	6.180	-0.560
1711	CHEMBL5795501	5.620	6.033	-0.413
1712	CHEMBL5761295	5.620	5.988	-0.368
1713	CHEMBL5779638	5.620	6.152	-0.532
1714	CHEMBL6056083	5.620	6.228	-0.608
1715	CHEMBL6017173	5.620	6.223	-0.603
1716	CHEMBL5827799	5.620	6.435	-0.815
1717	CHEMBL5836125	5.620	6.654	-1.034
1718	CHEMBL5840835	5.620	6.555	-0.935
1719	CHEMBL6000015	5.620	6.769	-1.149
1720	CHEMBL5995996	5.620	6.161	-0.541
1721	CHEMBL1958218	5.620	5.243	+0.377
1722	CHEMBL5777711	5.620	6.331	-0.711
1723	CHEMBL85021	5.610	5.464	+0.146
1724	CHEMBL5907728	5.608	6.453	-0.845
1725	CHEMBL2408824	5.607	6.061	-0.454
1726	CHEMBL101581	5.600	6.142	-0.542
1727	CHEMBL4745877	5.600	7.367	-1.767
1728	CHEMBL1095130	5.600	5.512	+0.088
1729	CHEMBL1097189	5.600	5.866	-0.266
1730	CHEMBL538798	5.600	5.823	-0.223
1731	CHEMBL4104427	5.600	5.800	-0.200
1732	CHEMBL3741069	5.590	5.415	+0.175
1733	CHEMBL3741655	5.590	5.552	+0.038
1734	CHEMBL2332119	5.580	5.486	+0.094
1735	CHEMBL1945648	5.580	5.631	-0.051
1736	CHEMBL130163	5.580	5.916	-0.336
1737	CHEMBL1241858	5.570	5.198	+0.372
1738	CHEMBL384067	5.565	5.681	-0.116
1739	CHEMBL5739918	5.560	6.501	-0.941
1740	CHEMBL6045276	5.560	6.221	-0.661
1741	CHEMBL5764214	5.560	6.564	-1.004
1742	CHEMBL5898434	5.560	6.304	-0.744
1743	CHEMBL1812569	5.550	6.238	-0.688
1744	CHEMBL182397	5.550	7.572	-2.022
1745	CHEMBL5969616	5.545	5.585	-0.040
1746	CHEMBL5797736	5.540	6.158	-0.618
1747	CHEMBL2333164	5.540	5.994	-0.454
1748	CHEMBL5739799	5.533	6.044	-0.511
1749	CHEMBL3628791	5.530	7.463	-1.933
1750	CHEMBL24944	5.530	5.088	+0.442
1751	CHEMBL3233763	5.525	6.957	-1.432
1752	CHEMBL1958023	5.520	4.995	+0.525
1753	CHEMBL4072410	5.520	6.445	-0.925
1754	CHEMBL2325109	5.520	5.978	-0.458
1755	CHEMBL5275784	5.520	6.213	-0.693
1756	CHEMBL608533	5.520	5.810	-0.290
1757	CHEMBL322395	5.520	6.237	-0.717
1758	CHEMBL1230609	5.517	5.990	-0.474
1759	CHEMBL5901840	5.510	6.269	-0.759

continued on next page

Table 5 – *continued from previous page*

#	ChEMBL ID	Measured	Predicted	Residual
1760	CHEMBL186747	5.500	6.923	-1.423
1761	CHEMBL1958037	5.500	4.889	+0.611
1762	CHEMBL204232	5.490	6.124	-0.634
1763	CHEMBL1016	5.490	6.337	-0.847
1764	CHEMBL1945446	5.490	5.541	-0.051
1765	CHEMBL6053847	5.483	5.584	-0.101
1766	CHEMBL2325110	5.480	5.988	-0.508
1767	CHEMBL1240553	5.480	5.754	-0.274
1768	CHEMBL2047020	5.480	5.996	-0.516
1769	CHEMBL461140	5.480	6.750	-1.270
1770	CHEMBL3133755	5.480	7.215	-1.735
1771	CHEMBL4084214	5.470	6.259	-0.789
1772	CHEMBL2424795	5.470	6.118	-0.648
1773	CHEMBL1958019	5.470	5.014	+0.456
1774	CHEMBL5860671	5.463	5.739	-0.277
1775	CHEMBL335931	5.460	5.970	-0.510
1776	CHEMBL1812570	5.460	6.098	-0.638
1777	CHEMBL1812564	5.460	6.070	-0.610
1778	CHEMBL2385987	5.460	6.678	-1.218
1779	CHEMBL132466	5.460	6.044	-0.584
1780	CHEMBL2385970	5.460	6.065	-0.605
1781	CHEMBL1958021	5.460	5.068	+0.392
1782	CHEMBL1762110	5.450	6.493	-1.043
1783	CHEMBL17157	5.450	5.685	-0.235
1784	CHEMBL83	5.450	5.536	-0.086
1785	CHEMBL3426225	5.450	6.618	-1.168
1786	CHEMBL4101735	5.450	6.633	-1.183
1787	CHEMBL2047018	5.450	5.867	-0.417
1788	CHEMBL5769479	5.445	6.730	-1.285
1789	CHEMBL1812561	5.440	6.188	-0.748
1790	CHEMBL6032965	5.440	6.939	-1.499
1791	CHEMBL4216765	5.440	5.642	-0.202
1792	CHEMBL132169	5.440	5.884	-0.444
1793	CHEMBL5273994	5.430	7.039	-1.609
1794	CHEMBL4204937	5.427	5.655	-0.228
1795	CHEMBL450786	5.420	5.861	-0.441
1796	CHEMBL332497	5.420	6.391	-0.971
1797	CHEMBL49120	5.410	6.061	-0.651
1798	CHEMBL3809489	5.410	6.251	-0.841
1799	CHEMBL5801676	5.410	5.881	-0.471
1800	CHEMBL541307	5.410	5.418	-0.008
1801	CHEMBL3740582	5.410	5.513	-0.103
1802	CHEMBL4102284	5.410	5.827	-0.417
1803	CHEMBL5286730	5.410	6.132	-0.722
1804	CHEMBL3234740	5.410	7.243	-1.833
1805	CHEMBL2316154	5.405	5.428	-0.023
1806	CHEMBL2437485	5.403	6.625	-1.222
1807	CHEMBL5174372	5.400	7.309	-1.909
1808	CHEMBL2424807	5.390	6.096	-0.706
1809	CHEMBL1641989	5.390	5.811	-0.421
1810	CHEMBL5419670	5.390	5.940	-0.550
1811	CHEMBL3233769	5.380	7.128	-1.748
1812	CHEMBL589809	5.380	5.628	-0.248
1813	CHEMBL2047019	5.380	5.960	-0.580
1814	CHEMBL5558817	5.380	5.590	-0.210
1815	CHEMBL4786003	5.380	6.045	-0.665
1816	CHEMBL1683971	5.380	7.397	-2.017
1817	CHEMBL1958039	5.380	4.766	+0.614
1818	CHEMBL1958024	5.370	5.241	+0.129

continued on next page

Table 5 – *continued from previous page*

#	ChEMBL ID	Measured	Predicted	Residual
1819	CHEMBL499345	5.370	5.635	-0.265
1820	CHEMBL1812434	5.370	7.485	-2.115
1821	CHEMBL1958217	5.370	5.288	+0.082
1822	CHEMBL3233762	5.370	6.974	-1.604
1823	CHEMBL5819273	5.363	6.034	-0.670
1824	CHEMBL2426378	5.355	6.501	-1.146
1825	CHEMBL106232	5.350	6.186	-0.836
1826	CHEMBL2385983	5.350	5.873	-0.523
1827	CHEMBL5760794	5.350	8.120	-2.770
1828	CHEMBL333553	5.345	5.779	-0.434
1829	CHEMBL418203	5.340	5.973	-0.633
1830	CHEMBL5287053	5.340	5.695	-0.355
1831	CHEMBL338944	5.340	5.861	-0.521
1832	CHEMBL213173	5.340	5.431	-0.091
1833	CHEMBL2316153	5.335	5.295	+0.040
1834	CHEMBL484109	5.330	5.923	-0.593
1835	CHEMBL603186	5.330	5.468	-0.138
1836	CHEMBL131590	5.330	5.884	-0.554
1837	CHEMBL1095465	5.330	5.714	-0.384
1838	CHEMBL1958028	5.320	4.838	+0.482
1839	CHEMBL2316159	5.320	4.981	+0.339
1840	CHEMBL484108	5.320	5.895	-0.575
1841	CHEMBL130208	5.320	5.846	-0.526
1842	CHEMBL4093642	5.310	5.919	-0.609
1843	CHEMBL1958020	5.310	5.059	+0.251
1844	CHEMBL3604929	5.310	6.092	-0.782
1845	CHEMBL4072392	5.300	6.488	-1.188
1846	CHEMBL4103121	5.300	6.315	-1.015
1847	CHEMBL361322	5.300	6.695	-1.395
1848	CHEMBL2385992	5.300	6.320	-1.020
1849	CHEMBL1241587	5.300	5.026	+0.274
1850	CHEMBL3233759	5.300	7.061	-1.761
1851	CHEMBL87084	5.300	6.213	-0.913
1852	CHEMBL6006814	5.290	6.044	-0.754
1853	CHEMBL2047028	5.290	5.965	-0.675
1854	CHEMBL5923228	5.290	5.213	+0.077
1855	CHEMBL113185	5.290	6.055	-0.765
1856	CHEMBL6020006	5.290	6.429	-1.139
1857	CHEMBL1812559	5.290	7.219	-1.929
1858	CHEMBL5945102	5.285	5.694	-0.409
1859	CHEMBL4863109	5.280	5.430	-0.150
1860	CHEMBL4868123	5.280	6.829	-1.549
1861	CHEMBL5804962	5.280	4.904	+0.376
1862	CHEMBL2029700	5.280	5.979	-0.699
1863	CHEMBL3263372	5.280	6.224	-0.944
1864	CHEMBL158805	5.280	4.990	+0.290
1865	CHEMBL113985	5.270	6.313	-1.043
1866	CHEMBL1958220	5.270	5.226	+0.044
1867	CHEMBL58	5.270	5.472	-0.202
1868	CHEMBL472743	5.270	6.729	-1.459
1869	CHEMBL1958031	5.270	5.003	+0.267
1870	CHEMBL5841526	5.260	5.383	-0.123
1871	CHEMBL5962503	5.260	5.580	-0.320
1872	CHEMBL3742060	5.260	5.470	-0.210
1873	CHEMBL6065282	5.260	5.416	-0.156
1874	CHEMBL6033572	5.260	6.062	-0.802
1875	CHEMBL6000404	5.260	5.554	-0.294
1876	CHEMBL5848379	5.260	5.322	-0.062
1877	CHEMBL5894758	5.260	5.316	-0.056

continued on next page

Table 5 – *continued from previous page*

#	ChEMBL ID	Measured	Predicted	Residual
1878	CHEMBL6007403	5.260	5.762	-0.502
1879	CHEMBL6026375	5.260	5.409	-0.149
1880	CHEMBL6055178	5.260	5.844	-0.584
1881	CHEMBL5977400	5.260	5.329	-0.069
1882	CHEMBL5863307	5.260	5.702	-0.442
1883	CHEMBL5955586	5.260	5.837	-0.577
1884	CHEMBL5943220	5.260	5.290	-0.030
1885	CHEMBL5964701	5.260	5.634	-0.374
1886	CHEMBL5743278	5.260	5.521	-0.261
1887	CHEMBL6013580	5.260	5.739	-0.479
1888	CHEMBL5899926	5.260	5.722	-0.462
1889	CHEMBL5939638	5.260	5.275	-0.015
1890	CHEMBL1958222	5.260	5.288	-0.028
1891	CHEMBL5800035	5.260	5.342	-0.082
1892	CHEMBL5997253	5.260	5.666	-0.406
1893	CHEMBL5788370	5.260	5.492	-0.232
1894	CHEMBL6012141	5.260	5.662	-0.402
1895	CHEMBL6032138	5.260	5.651	-0.391
1896	CHEMBL5792219	5.253	6.435	-1.182
1897	CHEMBL2316147	5.250	5.367	-0.117
1898	CHEMBL2029696	5.250	6.161	-0.911
1899	CHEMBL377730	5.250	5.752	-0.502
1900	CHEMBL186802	5.250	7.078	-1.828
1901	CHEMBL250315	5.250	7.440	-2.190
1902	CHEMBL1958210	5.250	4.811	+0.439
1903	CHEMBL4066118	5.240	4.914	+0.326
1904	CHEMBL4081698	5.235	6.403	-1.168
1905	CHEMBL5405436	5.230	6.101	-0.871
1906	CHEMBL4650321	5.230	7.180	-1.950
1907	CHEMBL5593611	5.230	6.736	-1.506
1908	CHEMBL1958219	5.220	5.297	-0.077
1909	CHEMBL2385982	5.220	5.762	-0.542
1910	CHEMBL2385988	5.220	6.586	-1.366
1911	CHEMBL1812556	5.220	7.461	-2.241
1912	CHEMBL4063701	5.210	6.322	-1.112
1913	CHEMBL5822544	5.210	6.057	-0.847
1914	CHEMBL2047027	5.210	6.131	-0.921
1915	CHEMBL2316150	5.210	5.212	-0.002
1916	CHEMBL4083467	5.205	6.335	-1.130
1917	CHEMBL3915306	5.190	6.010	-0.820
1918	CHEMBL335932	5.190	5.843	-0.653
1919	CHEMBL1958038	5.190	4.838	+0.352
1920	CHEMBL5201128	5.190	6.227	-1.037
1921	CHEMBL3233768	5.190	7.630	-2.440
1922	CHEMBL3133893	5.185	6.478	-1.293
1923	CHEMBL1958035	5.180	4.998	+0.182
1924	CHEMBL1812566	5.180	6.022	-0.842
1925	CHEMBL309765	5.170	5.338	-0.168
1926	CHEMBL5853228	5.170	5.824	-0.654
1927	CHEMBL3892960	5.170	6.835	-1.665
1928	CHEMBL5739782	5.170	5.853	-0.683
1929	CHEMBL84513	5.170	4.696	+0.474
1930	CHEMBL1241863	5.170	5.731	-0.561
1931	CHEMBL2316145	5.165	5.208	-0.043
1932	CHEMBL5868329	5.160	8.137	-2.977
1933	CHEMBL336410	5.160	4.729	+0.431
1934	CHEMBL57300	5.160	5.261	-0.101
1935	CHEMBL131715	5.160	5.910	-0.750
1936	CHEMBL4215703	5.160	6.464	-1.304

continued on next page

Table 5 – *continued from previous page*

#	ChEMBL ID	Measured	Predicted	Residual
1937	CHEMBL6016889	5.150	7.077	-1.927
1938	CHEMBL2316146	5.150	5.274	-0.124
1939	CHEMBL3085377	5.140	7.481	-2.341
1940	CHEMBL1812435	5.140	7.513	-2.373
1941	CHEMBL3586178	5.130	5.515	-0.385
1942	CHEMBL188386	5.130	5.948	-0.818
1943	CHEMBL6042351	5.130	5.874	-0.744
1944	CHEMBL336192	5.130	5.984	-0.854
1945	CHEMBL3287898	5.120	5.633	-0.513
1946	CHEMBL590525	5.120	5.217	-0.097
1947	CHEMBL6062321	5.110	6.391	-1.281
1948	CHEMBL604672	5.110	4.935	+0.175
1949	CHEMBL1812555	5.110	7.347	-2.237
1950	CHEMBL242349	5.110	6.812	-1.702
1951	CHEMBL89505	5.100	5.177	-0.077
1952	CHEMBL3221541	5.100	5.788	-0.688
1953	CHEMBL142148	5.100	5.505	-0.405
1954	CHEMBL1958211	5.100	5.085	+0.015
1955	CHEMBL2316144	5.095	5.285	-0.190
1956	CHEMBL1958033	5.090	4.945	+0.145
1957	CHEMBL1958025	5.090	5.027	+0.063
1958	CHEMBL337297	5.090	5.816	-0.726
1959	CHEMBL3586174	5.090	5.504	-0.414
1960	CHEMBL119156	5.085	5.493	-0.408
1961	CHEMBL5269346	5.080	5.537	-0.457
1962	CHEMBL1958029	5.080	5.087	-0.007
1963	CHEMBL3221546	5.070	5.784	-0.714
1964	CHEMBL1958036	5.070	5.226	-0.156
1965	CHEMBL5972723	5.070	8.185	-3.115
1966	CHEMBL2029699	5.070	6.186	-1.116
1967	CHEMBL2437474	5.065	6.446	-1.381
1968	CHEMBL1944921	5.060	5.560	-0.500
1969	CHEMBL543361	5.060	5.470	-0.410
1970	CHEMBL4560998	5.060	6.463	-1.403
1971	CHEMBL5863747	5.053	5.219	-0.165
1972	CHEMBL432941	5.050	5.171	-0.121
1973	CHEMBL1958223	5.050	5.227	-0.177
1974	CHEMBL2437470	5.040	6.483	-1.443
1975	CHEMBL1095464	5.040	5.829	-0.789
1976	CHEMBL5986062	5.030	6.419	-1.389
1977	CHEMBL3960786	5.030	5.672	-0.642
1978	CHEMBL137635	5.030	5.820	-0.790
1979	CHEMBL589068	5.030	5.079	-0.049
1980	CHEMBL487254	5.020	5.470	-0.450
1981	CHEMBL592710	5.020	4.924	+0.096
1982	CHEMBL89723	5.020	5.598	-0.578
1983	CHEMBL410659	5.020	6.639	-1.619
1984	CHEMBL2437475	5.015	6.385	-1.370
1985	CHEMBL3586179	5.010	5.509	-0.499
1986	CHEMBL358494	5.000	5.126	-0.126
1987	CHEMBL47173	5.000	5.228	-0.228
1988	CHEMBL247228	5.000	6.188	-1.188
1989	CHEMBL142898	5.000	5.724	-0.724
1990	CHEMBL2437484	4.994	6.976	-1.981
1991	CHEMBL116012	4.990	6.411	-1.421
1992	CHEMBL473141	4.990	6.946	-1.956
1993	CHEMBL1448	4.980	5.359	-0.379
1994	CHEMBL114073	4.973	6.104	-1.130
1995	CHEMBL425620	4.970	5.861	-0.891

continued on next page

Table 5 – *continued from previous page*

#	ChEMBL ID	Measured	Predicted	Residual
1996	CHEMBL5746124	4.970	7.012	-2.042
1997	CHEMBL514771	4.970	6.276	-1.306
1998	CHEMBL1958030	4.970	5.297	-0.327
1999	CHEMBL5792559	4.963	4.904	+0.060
2000	CHEMBL1958027	4.960	4.977	-0.017
2001	CHEMBL185327	4.960	6.965	-2.005
2002	CHEMBL5818193	4.947	5.208	-0.260
2003	CHEMBL589825	4.940	4.953	-0.013
2004	CHEMBL115519	4.930	5.791	-0.861
2005	CHEMBL1241675	4.920	5.755	-0.835
2006	CHEMBL1945445	4.920	5.406	-0.486
2007	CHEMBL5424516	4.910	7.135	-2.225
2008	CHEMBL337035	4.900	6.064	-1.164
2009	CHEMBL347659	4.890	5.019	-0.129
2010	CHEMBL1254364	4.890	5.440	-0.550
2011	CHEMBL1641994	4.890	5.467	-0.577
2012	CHEMBL472942	4.870	6.490	-1.620
2013	CHEMBL338449	4.870	5.893	-1.023
2014	CHEMBL113996	4.870	6.377	-1.507
2015	CHEMBL4872030	4.850	5.417	-0.567
2016	CHEMBL1958032	4.850	4.977	-0.127
2017	CHEMBL4878090	4.850	7.090	-2.240
2018	CHEMBL213616	4.840	5.011	-0.171
2019	CHEMBL83429	4.830	4.693	+0.137
2020	CHEMBL589120	4.830	4.797	+0.033
2021	CHEMBL428690	4.820	5.973	-1.153
2022	CHEMBL1945647	4.820	5.631	-0.811
2023	CHEMBL1240565	4.820	5.938	-1.118
2024	CHEMBL347160	4.820	4.674	+0.146
2025	CHEMBL2316148	4.820	5.336	-0.516
2026	CHEMBL56319	4.813	5.115	-0.302
2027	CHEMBL1945452	4.800	5.831	-1.031
2028	CHEMBL5896779	4.790	6.013	-1.223
2029	CHEMBL1947045	4.775	5.976	-1.201
2030	CHEMBL214863	4.770	4.818	-0.048
2031	CHEMBL1958026	4.770	5.010	-0.240
2032	CHEMBL5960969	4.763	4.850	-0.086
2033	CHEMBL1956885	4.760	5.640	-0.880
2034	CHEMBL589847	4.760	5.088	-0.328
2035	CHEMBL1241774	4.750	4.826	-0.076
2036	CHEMBL1945645	4.740	5.791	-1.051
2037	CHEMBL2316156	4.730	5.344	-0.614
2038	CHEMBL48760	4.730	5.434	-0.704
2039	CHEMBL4634317	4.720	5.667	-0.947
2040	CHEMBL147112	4.720	5.331	-0.611
2041	CHEMBL87746	4.720	5.390	-0.670
2042	CHEMBL5555046	4.710	5.447	-0.737
2043	CHEMBL86326	4.710	5.481	-0.771
2044	CHEMBL589832	4.710	5.015	-0.305
2045	CHEMBL1945453	4.700	5.848	-1.148
2046	CHEMBL1945444	4.700	5.343	-0.643
2047	CHEMBL1956886	4.700	5.727	-1.027
2048	CHEMBL5875009	4.700	5.052	-0.352
2049	CHEMBL1630578	4.690	6.375	-1.685
2050	CHEMBL2316152	4.680	5.347	-0.667
2051	CHEMBL1641991	4.680	5.423	-0.743
2052	CHEMBL1242661	4.680	5.270	-0.590
2053	CHEMBL1945646	4.670	5.599	-0.929
2054	CHEMBL2437477	4.662	6.321	-1.658

continued on next page

Table 5 – *continued from previous page*

#	ChEMBL ID	Measured	Predicted	Residual
2055	CHEMBL155607	4.660	4.753	-0.093
2056	CHEMBL1683952	4.660	6.035	-1.375
2057	CHEMBL347074	4.660	5.247	-0.587
2058	CHEMBL149201	4.660	5.169	-0.509
2059	CHEMBL600048	4.660	5.076	-0.416
2060	CHEMBL118109	4.657	5.743	-1.086
2061	CHEMBL148984	4.640	5.325	-0.685
2062	CHEMBL3353396	4.635	5.813	-1.178
2063	CHEMBL358728	4.620	5.327	-0.707
2064	CHEMBL154260	4.620	5.405	-0.785
2065	CHEMBL18797	4.620	5.683	-1.063
2066	CHEMBL1916951	4.610	5.370	-0.760
2067	CHEMBL5557733	4.600	5.468	-0.868
2068	CHEMBL438075	4.600	7.179	-2.579
2069	CHEMBL1916949	4.600	5.363	-0.763
2070	CHEMBL1945450	4.570	5.840	-1.270
2071	CHEMBL592240	4.570	4.544	+0.026
2072	CHEMBL1095445	4.570	5.410	-0.840
2073	CHEMBL592210	4.560	4.906	-0.346
2074	CHEMBL1641998	4.550	5.754	-1.204
2075	CHEMBL5918833	4.540	5.917	-1.377
2076	CHEMBL1945451	4.540	6.017	-1.477
2077	CHEMBL347854	4.540	4.863	-0.323
2078	CHEMBL1945644	4.540	5.739	-1.199
2079	CHEMBL187007	4.540	6.832	-2.292
2080	CHEMBL2312652	4.520	5.713	-1.193
2081	CHEMBL148514	4.520	5.383	-0.863
2082	CHEMBL1254363	4.510	5.881	-1.371
2083	CHEMBL3586175	4.500	5.611	-1.111
2084	CHEMBL146694	4.500	5.270	-0.770
2085	CHEMBL1094784	4.480	6.790	-2.310
2086	CHEMBL3353395	4.470	5.834	-1.364
2087	CHEMBL521155	4.460	5.193	-0.733
2088	CHEMBL4642842	4.455	5.593	-1.138
2089	CHEMBL314146	4.360	5.276	-0.916
2090	CHEMBL4640482	4.355	5.488	-1.133
2091	CHEMBL4102214	4.350	4.398	-0.048
2092	CHEMBL59099	4.350	5.744	-1.394
2093	CHEMBL1095463	4.340	5.339	-0.999
2094	CHEMBL363607	4.340	5.355	-1.015
2095	CHEMBL5200160	4.335	5.934	-1.599
2096	CHEMBL2437473	4.327	6.389	-2.062
2097	CHEMBL5633878	4.320	5.143	-0.823
2098	CHEMBL1910275	4.320	5.019	-0.699
2099	CHEMBL450622	4.310	5.209	-0.899
2100	CHEMBL3799345	4.310	4.394	-0.084
2101	CHEMBL57553	4.300	5.730	-1.430
2102	CHEMBL5555183	4.290	5.594	-1.304
2103	CHEMBL5532212	4.280	5.568	-1.288
2104	CHEMBL1254199	4.280	5.998	-1.718
2105	CHEMBL1094475	4.260	5.327	-1.067
2106	CHEMBL5560045	4.250	5.539	-1.289
2107	CHEMBL1087055	4.250	4.747	-0.497
2108	CHEMBL5559904	4.210	5.609	-1.399
2109	CHEMBL3353399	4.180	5.922	-1.742
2110	CHEMBL4553708	4.180	5.755	-1.575
2111	CHEMBL324439	4.150	5.862	-1.712
2112	CHEMBL2029697	4.140	5.675	-1.535
2113	CHEMBL5562728	4.130	5.656	-1.526

continued on next page

Table 5 – *continued from previous page*

#	ChEMBL ID	Measured	Predicted	Residual
2114	CHEMBL414013	4.120	5.752	-1.632
2115	CHEMBL3353400	4.120	5.878	-1.758
2116	CHEMBL5568100	4.090	5.617	-1.527
2117	CHEMBL485246	4.090	5.495	-1.405
2118	CHEMBL5567680	4.070	5.585	-1.515
2119	CHEMBL109259	4.050	6.093	-2.043
2120	CHEMBL504135	4.050	4.758	-0.708
2121	CHEMBL1094808	4.040	5.262	-1.222
2122	CHEMBL1947046	4.040	6.281	-2.241
2123	CHEMBL1087054	4.030	4.913	-0.883
2124	CHEMBL1910268	4.030	6.553	-2.523
2125	CHEMBL1095776	4.020	5.615	-1.595
2126	CHEMBL154911	4.000	5.124	-1.124
2127	CHEMBL57690	4.000	5.675	-1.675
2128	CHEMBL56081	4.000	5.682	-1.682
2129	CHEMBL1077107	4.000	4.936	-0.936
2130	CHEMBL3361128	4.000	5.477	-1.477
2131	CHEMBL312001	4.000	4.509	-0.509
2132	CHEMBL1242663	4.000	5.148	-1.148
2133	CHEMBL4075892	4.000	4.135	-0.135