Supporting Information

General purpose structure-based drug discovery neural network scorefunctions with human interpretable pharmacophore maps

BP Brown¹*, J Mendenhall², AR Geanes², J Meiler³, ⁴ *

1 Chemical and Physical Biology Program, Medical Scientist Training Program, Center for Structural Biology, Vanderbilt University, Nashville, TN 37232 USA.

2 Department of Chemistry, Center for Structural Biology, Vanderbilt University, Nashville, TN 37232 USA

3 Department of Chemistry, Center for Structural Biology, Vanderbilt University, Nashville, TN 37232 USA ; Departments of Pharmacology and Biomedical Informatics, Vanderbilt University, Nashville, TN 37212 USA.

4 Institute for Drug Discovery, Leipzig University Medical School, Leipzig, SAC 04103, Germany

*Co-correspondence: BPB (benjamin.p.brown17@gmail.com), JM (jens@meilerlab.org)
| Descriptor Name               | Description                                                                 | Value Type                                      |
|------------------------------|------------------------------------------------------------------------------|-------------------------------------------------|
| Atom_VCharge                 | Atom partial charge\(^1\)                                                   | Charge value (\(\text{e}\))                    |
| Atom_SigmaCharge             | Atom partial charge\(^2\)                                                   | Charge value (\(\text{e}\))                    |
| Atom_HBondDonorTernary       | Discriminates hydrogen bond donors (OH/NH), acceptors (O/N), and non-       | Returns 1 for hydrogen bond donors, -1 for hydrogen bond acceptors, and 0 otherwise |
|                              | hydrogen bonding atoms. Donors are considered strictly, such that donors    |                                                 |
|                              | cannot also be considered acceptors.                                        |                                                 |
| Atom_IsInAromaticRingTernary | Discriminates atoms that are in                                              | Returns 1 if the atom is part of an aromatic ring system and -1 otherwise |
|                              | aromatic\(^3\) rings from other atoms                                       |                                                 |
| Atom_IsInRingTernary         | Discriminates atoms that are in rings from other atoms                      | Returns 1 if the atom is part of a ring system and -1 otherwise |
|                              |                                                                              |                                                 |
| Atom_Polarizability          | Atom polarizability\(^4\)                                                   | Polarizability (\(\text{au}\))                 |
| Atom_HydrophobicTernary      | Discriminates hydrophobic\(^5\) atoms from non-hydrophobic atoms           | Returns 1 if the atom is hydrophobic and -1 otherwise |
| Atom_IsHTernary              | Discriminates heavy atoms from hydrogen atoms                               | Returns 1 if the atom is a hydrogen atom and -1 for a heavy atom |
| ENegNonC                     | Difference in atom electronegativity\(^6\) with carbon                     | Returns the difference in electronegativity between the selected element and the carbon. If the difference is less than 0 then returns 0. |
Table S2. BCL-AffinityNet feature decomposition. All features are of the BCL property type “3DAPairRealSpace” or “3DAPairRealSpaceAsym”. 3DAPairRealSpaceAsym differs from 3DAPairRealSpace in that the -+ and +- sign pairings are discriminated in the former but not the latter. This allows incorporation of protein-ligand directionality. The “Template” customizable implementation in the BCL descriptor framework defines the property type (e.g. 3DAPairRealSpace), sign pairings, distance bin range, and step size (example code object files are provided as Supporting Information). Multiple instances of the Template definition can be used with different protein and ligand base property types. This is the feature set with which BCL::Affinity was trained; however, users can customize the object file to train their own version of BCL-AffinityNet, generate descriptors for use with external software (e.g. WEKA, as was done for the Random Forest model in this manuscript), or generate traditional ligand-based QSAR models. In total, the model of BCL-AffinityNet trained in this manuscript contained 490 feature columns.

| Code Object File Template | Ligand | Protein | Sign Pairings | Range  | Step Size | Feature Count |
|---------------------------|--------|---------|---------------|--------|-----------|---------------|
| 3DAPairRS050              | Atom_Vcharge | Atom_Vcharge | -/+/-/+     | 0.0 - 7.0 | 0.5        | 42            |
|                           | Atom_SigmaCharge | Atom_SigmaCharge | -/+/-/+     | 0.0 - 7.0 | 0.5        | 42            |
|                           | Atom_HBondDonorTernary | Atom_HBondDonorTernary | -/+/-/+     | 0.0 - 7.0 | 0.5        | 42            |
|                           | Atom_IsInRingTernary | Atom_IsInRingTernary | -/+/-/+     | 0.0 - 7.0 | 0.5        | 42            |
|                           | Atom_IsInAromaticRingTernary | Atom_IsInAromaticRingTernary | -/+/-/+     | 0.0 - 7.0 | 0.5        | 42            |
|                           | Multiply(Atom_HydrophobicTernary,Atom_Polarizability) | Multiply(Atom_HydrophobicTernary,Atom_Polarizability) | -/+/-/+     | 0.0 - 7.0 | 0.5        | 42            |
|                           | Multiply(Atom_HydrophobicTernary,Atom_Polarizability) | Multiply(Atom_HydrophobicTernary,Atom_Polarizability) | -/+/-/+     | 0.0 - 7.0 | 0.5        | 42            |
|                           | IsHTernary | IsHTernary | -/+/-/+     | 0.0 - 7.0 | 0.5        | 42            |
| 3DAPairRSPP050             | ENegNonC | ENegNonC | ++           | 0.0 - 7.0 | 0.5        | 14            |
| 3DAPairRSAsym050           | Atom_Polarizability | Atom_Polarizability | ++           | 0.0 - 7.0 | 0.5        | 14            |
|                           | Multiply(Atom_HydrophobicTernary,Atom_Polarizability) | Multiply(Atom_HydrophobicTernary,Atom_Polarizability) | -/+/-/+     | 0.0 - 7.0 | 0.5        | 56            |
|                           | Multiply(Atom_HydrophobicTernary,Atom_Polarizability) | Multiply(Atom_HydrophobicTernary,Atom_Polarizability) | -/+/-/+     | 0.0 - 7.0 | 0.5        | 56            |
| 3DAPairRS-PP-PM-Asym050    | ENegNonC | ENegNonC | ++/-        | 0.0 - 7.0 | 0.5        | 28            |
|                           | Multiply(Atom_HydrophobicTernary,Atom_Polarizability) | Multiply(Atom_HydrophobicTernary,Atom_Polarizability) | ++/-        | 0.0 - 7.0 | 0.5        | 28            |
|                           | Atom_IsInAromaticRingTernary | Atom_IsInAromaticRingTernary | ++/-        | 0.0 - 7.0 | 0.5        | 28            |
| Total Feature Count       |        |        |              |        |           | 490           |
**Table S3. BCL-DockANNScore feature decomposition.** All features are of the BCL property type “3DAPairRealSpace”, “3DAPairRealSpaceAsym”, or 3DInterHBondCode. 3DAPairRealSpaceAsym differs from 3DAPairRealSpace in that the -+ and +- sign pairings are discriminated in the former but not the latter. Unlike 3DAPairRealSpace and 3DAPairRealSpaceAsym, 3DInterHBondCode is not modular and does not support generalization to alternative protein or ligand features. This is the feature set with which BCL-DockANNScore was trained; however, users can customize the object file to train their own version of BCL-DockANNScore, generate descriptors for use with external software, or generate traditional ligand-based QSAR models. In total, the model of BCL-DockANNScore trained in this manuscript contained 850 feature columns.

| Code Object File Template | Ligand | Protein | Sign Pairings | Range | Step Size | Feature Count |
|---------------------------|--------|---------|---------------|-------|-----------|---------------|
| **3DAPairRS050**          |        |         |               |       |           |               |
| Atom_Vcharge              | Atom_Vcharge |       | -/+;++/+     | 0.0 - 7.0 | 0.5        | 42            |
| Atom_SigmaCharge          | Atom_SigmaCharge |       | -/+;++/+     | 0.0 - 7.0 | 0.5        | 42            |
| Atom_HBondDonorTernary    | Atom_HBondDonorTernary |       | -/+;++/+     | 0.0 - 7.0 | 0.5        | 42            |
| Atom_isInRingTernary      | Atom_isInRingTernary |       | -/+;++/+     | 0.0 - 7.0 | 0.5        | 42            |
| Multiply(Atom_HydrophobicTernary,Atom_Polarizability) | Multiply(Atom_HydrophobicTernary,Atom_Polarizability) |       | -/+;++/+     | 0.0 - 7.0 | 0.5        | 42            |
| IsHTernary                | IsHTernary |       | -/+;++/+     | 0.0 - 7.0 | 0.5        | 42            |
| **3DAPairRSPP050**        |        |         |               |       |           |               |
| ENegNonC                  | ENegNonC |       | ++           | 0.0 - 7.0 | 0.5        | 14            |
| Atom_Polarizability       | Atom_Polarizability |       | ++           | 0.0 - 7.0 | 0.5        | 14            |
| **3DAPairRSAsym050**      |        |         |               |       |           |               |
| Atom_isInAromaticRingTernary | Multiply(Atom_HydrophobicTernary,Atom_Polarizability) |       | -/+;++/+     | 0.0 - 7.0 | 0.5        | 56            |
| Multiply(Atom_HydrophobicTernary,Atom_Polarizability) | Multiply(Atom_HydrophobicTernary,Atom_Polarizability) |       | -/+;++/+     | 0.0 - 7.0 | 0.5        | 56            |
| **3DAPairRS-PP-PM-Asym050** |        |         |               |       |           |               |
| ENegNonC                  | Multiply(Atom_HydrophobicTernary,Atom_Polarizability) |       | ++/+/-       | 0.0 - 7.0 | 0.5        | 28            |
| ENegNonC                  | Atom_isInAromaticRingTernary |       | ++/+/-       | 0.0 - 7.0 | 0.5        | 28            |
| **3DInterHBondCode**      |        |         |               |       |           |               |
| HBondDonorAcceptorPair    | ++     |         | 0.0 - 4.0 / 0.0 - 360.0 | 0.5 / 8.0 |           | 360           |
| **Total Feature Count**   |        |         |               |       |           | 850           |
Table S4. BCL-AffinityNet top 20 features by input sensitivity model consistency score.

| Feature ID                                                                 | Input Sensitivity Score |
|---------------------------------------------------------------------------|-------------------------|
| Partial(3DAPairRSAsym050(Multiply(Atom_HydrophobicTernary,Atom_Polarizability),Atom_IsInAromaticRingTernary),indices(55)) | 0.93                    |
| Partial(3DAPairRSAsym050(Multiply(Atom_HydrophobicTernary,Atom_Polarizability),Atom_IsInAromaticRingTernary),indices(51)) | 0.78                    |
| Partial(3DAPairRS050(Multiply(Atom_HydrophobicTernary,Atom_Polarizability)),indices(41)) | 0.78                    |
| Partial(3DAPairRS050(Atom_HBondDonorTernary),indices(34))                  | 0.77                    |
| Partial(3DAPairRS050(Atom_SigmaCharge),indices(37))                        | 0.75                    |
| Partial(3DAPairRS050(Atom_SigmaCharge),indices(41))                        | 0.74                    |
| Partial(3DAPairRS050(Atom_SigmaCharge),indices(29))                        | 0.73                    |
| Partial(3DAPairRS050(Multiply(Atom_HydrophobicTernary,Atom_Polarizability),Atom_IsInAromaticRingTernary),indices(19)) | 0.72                    |
| Partial(3DAPairRS050(Atom_HBondDonorTernary),indices(25))                  | 0.67                    |
| Partial(3DAPairRS050(Atom_SigmaCharge),indices(24))                        | 0.67                    |
| Partial(3DAPairRS050(Multiply(Atom_HydrophobicTernary,Atom_Polarizability),Atom_IsInAromaticRingTernary),indices(28)) | 0.67                    |
| Partial(3DAPairRS-PP-PM-Asym050(ENegNonC,Multiply(Atom_HydrophobicTernary,Atom_Polarizability)),indices(10)) | 0.65                    |
| Partial(3DAPairRS-PP-PM-Asym050(ENegNonC,Atom_IsInAromaticRingTernary),indices(40)) | 0.64                    |
| Partial(3DAPairRS050(Multiply(Atom_HydrophobicTernary,Atom_Polarizability),Atom_IsInAromaticRingTernary),indices(47)) | 0.62                    |
| Partial(3DAPairRS050(Atom_Polarizability),indices(13))                      | 0.61                    |
| Partial(3DAPairRS050(Atom_Vcharge),indices(34))                            | 0.60                    |
Figure S1. BCL-AffinityNet input sensitivity mixed sign symmetric feature decomposition aggregated by sign bin. Each plot represents the cross-validation model consistency score for each property in the “3DAPairRS050” template. Column values represent the average consistency score across all distance bins for the specified sign pairing. Here, Sign ID 0 represents the -/- sign pairing, Sign ID 1 represents the +/- sign pairing, and Sign ID 2 represents both of the undifferentiated -/+ and +/- sign pairings. Error bars indicate standard deviation across all distance bins.
Figure S2. BCL-AffinityNet input sensitivity mixed sign asymmetric feature decomposition aggregated by sign bin. Each plot represents the cross-validation model consistency score for each property in the “3DAPairRSAym050” template. Column values represent the average consistency score across all distance bins for the specified sign pairing. Here, Sign ID 0 represents the -/- sign pairing, Sign ID 1 represents the +/- sign pairing, Sign ID 2 represents the -/+ sign pairing, and Sign ID 3 represents the +/+ sign pairing. Error bars indicate standard deviation across all distance bins.
Figure S3. BCL-AffinityNet input sensitivity restricted mixed sign asymmetric feature decomposition aggregated by sign bin. Each plot represents the cross-validation model consistency score for each property in the “3DAPairRS-PP-PM-Asym050” template. Column values represent the average consistency score across all distance bins for the specified sign pairing. Here, Sign ID 0 represents the +/+ sign pairing and Sign ID 1 represents the +/- sign pairing. Error bars indicate standard deviation across all distance bins.
Figure S4. BCL-AffinityNet input sensitivity same sign symmetric feature decomposition aggregated by sign bin. Each plot represents the cross-validation model consistency score for each property in the “3DPairRSPP050” template. Column values represent the average consistency score across all distance bins for the specified sign pairing. Here, Sign ID 0 represents the ++ sign pairing. Error bars indicate standard deviation across all distance bins.
Table S5. BCL-DockANNscore top 20 features by input sensitivity model consistency score.

| Feature ID                                                                 | Input Sensitivity Score |
|---------------------------------------------------------------------------|-------------------------|
| Partial(3DAPairRS050(IsHTernary),indices(16))                           | 0.85                    |
| Partial(3DAPairRS-PP-PM-Asym050(ENegNonC,Atom_IsInAromaticRingTernary),indices(9)) | 0.84                    |
| Partial(3DAPairRS050(IsHTernary),indices(13))                           | 0.84                    |
| Partial(3DAPairRS-PP-PM-Asym050(ENegNonC,Atom_IsInAromaticRingTernary),indices(11)) | 0.84                    |
| Partial(3DAPairRS050(IsHTernary),indices(15))                           | 0.82                    |
| Partial(3DAPairRS050(IsHTernary),indices(19))                           | 0.82                    |
| Partial(3DAPairRS050(IsHTernary),indices(12))                           | 0.80                    |
| Partial(3DAPairRS050(IsHTernary),indices(21))                           | 0.78                    |
| Partial(3DAPairRS050(Atom_Vcharge),indices(10))                         | 0.76                    |
| Partial(3DAPairRS050(Atom_SigmaCharge),indices(35))                     | 0.76                    |
| Partial(3DAPairRS050(Atom_Vcharge),indices(34))                         | 0.76                    |
| Partial(3DAPairRS050(Atom_SigmaCharge),indices(34))                     | 0.75                    |
| Partial(3DAPairRS050(IsHTernary),indices(33))                           | 0.75                    |
| Partial(3DAPairRS050(Atom_IsInAromaticRingTernary),indices(6))          | 0.73                    |
| Partial(3DAPairRS050(Atom_IsInRingTernary),indices(11))                 | 0.73                    |
| Partial(3DAPairRS050(IsHTernary),indices(18))                           | 0.71                    |
| Partial(3DAPairRS-PP-PM-Asym050(ENegNonC,Atom_IsInAromaticRingTernary),indices(13)) | 0.71                    |
| Partial(3DAPairRS-PP-PM-Asym050(ENegNonC,Atom_IsInAromaticRingTernary),indices(7)) | 0.71                    |
| Partial(3DAPairRS050(IsHTernary),indices(22))                           | 0.69                    |
| Partial(3DAPairRS050(Atom_Vcharge),indices(20))                         | 0.67                    |
Figure S5. BCL-DockANNScore input sensitivity mixed sign symmetric feature decomposition aggregated by sign bin. Each plot represents the cross-validation model consistency score for each property in the “3DAPairRS050” template. Column values represent the average consistency score across all distance bins for the specified sign pairing. Here, Sign ID 0 represents the -/- sign pairing, Sign ID 1 represents the +/- sign pairing, and Sign ID 2 represents both of the undifferentiated +/- and +/- sign pairings. Error bars indicate standard deviation across all distance bins.
Figure S6. BCL-DockANNScore input sensitivity mixed sign asymmetric feature decomposition aggregated by sign bin. Each plot represents the cross-validation model consistency score for each property in the “3DAPairRSAym050” template. Column values represent the average consistency score across all distance bins for the specified sign pairing. Here, Sign ID 0 represents the -/- sign pairing, Sign ID 1 represents the +/+ sign pairing, Sign ID 2 represents the -/+ sign pairing, and Sign ID 3 represents the +/- sign pairing. Error bars indicate standard deviation across all distance bins.
Figure S7. BCL-DockANNScore input sensitivity restricted mixed sign asymmetric feature decomposition aggregated by sign bin. Each plot represents the cross-validation model consistency score for each property in the “3DAPairRS-PP-PM-Asym050” template. Column values represent the average consistency score across all distance bins for the specified sign pairing. Here, Sign ID 0 represents the +/- sign pairing and Sign ID 1 represents the +/- sign pairing. Error bars indicate standard deviation across all distance bins.
Figure S8. BCL-DockANNScore input sensitivity same sign symmetric feature decomposition aggregated by sign bin. Each plot represents the cross-validation model consistency score for each property in the “3DAPairRSPP050” template. Column values represent the average consistency score across all distance bins for the specified sign pairing. Here, Sign ID 0 represents the +/+ sign pairing. Error bars indicate standard deviation across all distance bins.
Table S6. BCL-AffinityNet comparison to KDEEP and RF-Score on the CASF-2016 coreset. Results reported as Pearson correlation coefficient (R).

| Descriptor Set   | Model Type          | Benchmark Test Set |                  |
|------------------|---------------------|--------------------|------------------|
|                  |                     | CASF-2016 coreset  | Pearson R        |
|                  |                     | (n=285)            |                  |
| BCL-AffinityNet  | DNN 2x512-32        | 0.84               |
| Molecular Weight | Scalar Property     | 0.50               |
| TPSA             | Scalar Property     | 0.20               |
| LogP             | Scalar Property     | 0.32               |
| Polarizability   | Scalar Property     | 0.51               |
| $\Delta_{Vina}RF_{20}$ | RF Docking Score | 0.82               |
| $\Delta_{Vina}RF_{20}$ | RF Docking Score | 0.73               |
| $\Delta_{Vina}XGB$ | Extreme Gradient Boosting Docking Score | 0.80               |
| KDEEP            | Grid-based CNN      | 0.82               |
| RF-Score         | RF                  | 0.80               |

a. As reported in Su et al. 2019<sup>7</sup>. Note that Su et al. contained 140 complexes in the training set that overlapped with the test set. RMSE not reported.

b. As reported in Lu et al. 2019<sup>8</sup>. Note that Lu et al. retrained $\Delta_{Vina}RF$ excluding test set complexes. RMSE not reported.

c. As reported in Jimenez et al., 2018<sup>9</sup>. Note that Jimenez et al. used the 58 target version of the coreset (290 protein-ligand pairs), while here all other metrics report statistics for the canonical 57 target version (285 protein-ligand pairs).
Figure S9. Evaluation of ligand and protein pocket property bias in the PDBbind v.2016 coreset. The 7568 sample training set derived from PDBbind v.2016 was used to build a Y-scramble version of BCL-AffinityNet (i.e. result labels were shuffled between training samples prior to training), a signed 3DA LB QSAR model, or a signed 3DA pocket-based QSAR model. Training was completed with five-fold random-split cross-validation. Columns and error bars represent the mean and standard deviation of NMAE (blue) or Pearson correlation coefficient (red) across either the five-fold random-split cross-validations (training) or five-fold random splits of the PDBbind v.2016 coreset (testing).
Figure S10. Performance evaluation on the LB AD test set. A total of 995 training samples were excluded from the initial training set of 7568 samples (see Methods for details). The remaining 6573 training samples were used to train BCL-AffinityNet (i.e. a single-task regression DNN with PLC features), a signed 3DA LB QSAR model, or a signed 3DA pocket-based QSAR model. Training was completed with five-fold random-split cross-validation. Columns and error bars represent the mean and standard deviation of NMAE (blue) or Pearson correlation coefficient (red) across either the five-fold random-split cross-validations (training) or five-fold random splits of the combined AD test set (testing).
Figure S11. Performance evaluation on the pocket AD test set. A total of 379 training samples were excluded from the initial training set of 7568 samples (see Methods for details). The remaining 7189 training samples were used to train BCL-AffinityNet (i.e. a single-task regression DNN with PLC features), a signed 3DA LB QSAR model, or a signed 3DA pocket-based QSAR model. Training was completed with five-fold random-split cross-validation. Columns and error bars represent the mean and standard deviation of NMAE (blue) or Pearson correlation coefficient (red) across either the five-fold random-split cross-validations (training) or five-fold random splits of the combined AD test set (testing).
Figure S12. Performance evaluation of WEKA random forest affinity prediction model trained with PLC descriptors. Models trained for evaluation on the PDBbind v.2016 coreset were trained with 7568 total samples. For models being evaluated on the Combined AD test set, a total of 1377 training samples were excluded from the initial training set of 7568 samples (see Methods for details on training splits). The remaining 6191 training samples were used to train AffinityRF (i.e. a random forest model trained with the same PLC descriptors as BCL-AffinityNet) using WEKA v.3.8.4. Columns and error bars represent the mean and standard deviation of Pearson correlation coefficient (red) across either five-fold random-split cross-validations (training) or five-fold random splits of the coreset or combined AD test set (testing).
Figure S13. **BCL-AffinityNet training cross-validation statistics and external test-set performance.** BCL-AffinityNet was trained with a normalized mean absolute error (NMAE; defined as the quotient of mean absolute error and mean absolute deviation) objective function and evaluated with both NMAE (blue columns) and Pearson correlation coefficient (red columns) on the PDBbind2016 coreset, CSAR NRC HiQ I set, and CSAR NRC HiQ II set. Each training set initially consists of PDBbind2016 refined set protein-ligand complexes plus complexes in the general set that have $K_d$ data available. All complexes in the PDBbind2016 coreset are excluded from all training sets. For each of the other external test sets, all redundant complexes between the training set and test were removed from the training set. Each of the above training sets is paired with one test set (the test set is the one directly adjacent on the right). Training columns are expressed as the average and standard deviation across five random split cross-validations. Test set columns are expressed as the average and standard deviation of five random splits of the test set.
Figure S14. BCL:DockANNScore training cross-validation statistics. BCL-DockANNScore was trained with an area under the curve (AUC) objective function and evaluated with AUC on five result labels: classification of a docked pose as being less than or equal to 1.0, 2.0, 3.0, 5.0, or 8.0 Å RMSD from the native pose. For each protein-ligand complex in the PDBbind2016 refined set, 250 decoys were generated with a standard RosettaLigand flexible docking protocol\textsuperscript{10}. These decoys were utilized in addition to the native protein-ligand complex provided in the dataset for training with five-fold cross-validation. Columns and error bars represent the average and standard deviation across five random split cross-validations.
Figure S15. Docking power evaluation of BCL-DockANNScore excluding native poses. Comparison of BCL-DockANNScore docking power to other methods from the CASF2016 benchmark by Su et al.\textsuperscript{7} when recovering the nearest non-native poses under 2.0 Å RMSD (A) within the top 3 poses, (B) within the top 2 poses, and (C) within the top 1 poses. Error bars indicate the 90% confidence interval. Green indicates BCL-DockANNScore.
Figure S16. Docking power evaluation binding funnel analysis of BCL-DockANNScore.
Figure S17. Training set five-fold random-split cross-validation of BCL-AffinityNet with variable PLC descriptor interaction shell distances. Columns and error bars represent the mean and standard deviation of NMAE (blue) or Pearson correlation coefficient (red) across the five-fold random-split cross-validations of the 7568 sample training set (see Methods for additional details).
Figure S18. Training set five-fold random-split cross-validation of BCL-AffinityNet with variable PLC descriptor smoothing parameters. Columns and error bars represent the mean and standard deviation of NMAE (blue) or Pearson correlation coefficient (red) across the five-fold random-split cross-validations of the 7568 sample training set (see Methods for additional details).
Figure S19. Training set five-fold random-split cross-validation of BCL-AffinityNet with variable model architectures. ANN indicates a single hidden layer was used. DNN indicates two hidden layers were used. Sigmoid and ReLU indicate the transfer function used for ANNs. All DNNs employed leaky rectifier transfer functions. The number following the transfer function indicates the number of neurons in the hidden layers. ANNs have one number and DNNs have two numbers corresponding to the respective hidden layer sizes. Columns and error bars represent the mean and standard deviation of NMAE (blue) or Pearson correlation coefficient (red) across the five-fold random-split cross-validations of the 7568 sample training set (see Methods for additional details).
Supplementary Methods

BCL-DockANNScore was trained using the PDBbind v.2016 refined set plus decoy poses generated with RosettaLigand. Briefly, a maximum of 100 conformers were generated for each small molecule ligand using the BCL conformer generator\textsuperscript{11, 12} with the following command:

\begin{verbatim}
bcl.exe molecule:ConformerGenerator \\
-ensemble_filenames <input_mols> \\
-conformers_single_file <output_confs> \\
top_models 100 \\
-max_iterations 2000 \\
-cluster \\
-conformation_comparer SymmetryRMSD 0.25
\end{verbatim}

Afterward, parameters files were generated with the Rosetta \texttt{molfile_to_params.py} script. All target protein PDB files were processed to remove water molecules, cofactors, and ligands with the \texttt{clean_pdb.py} script available in the Rosetta macromolecular modeling suite scripts directory. Docking was performed as described previously\textsuperscript{10, 13} to generate 250 decoys for each refined set complex. No restraints were implemented to enforce specific RMSD ranges. All RMSDs were computed in Rosetta with respect to the native pose provided for each refined set complex.

Additional PDB preparation for the BCL

All PDB files for training and testing are first processed with the Rosetta \texttt{clean_pdb.py} script available in the Rosetta macromolecular modeling suite scripts directory:

\begin{verbatim}
clean_pdb.py -allchains <protein.pdb> Z
\end{verbatim}

Proteins are then formatted appropriate for further processing in the BCL:

\begin{verbatim}
bcl.exe protein:PDBConvert protein_Z.pdb \\
-aaclass AAComplete \\
-min_sse_size 0 0 0 \\
bcl_pdb Full \\
-write_hydrogens \\
-output_prefix <prefix>
\end{verbatim}

*Note – the BCL does not recognize N- and C- terminal caps in PDB files (e.g. OXT); simply delete or convert to regular element type name (e.g. O).

Small molecule preparation

All molecules for training and testing were first processed in the BCL to assure correct atom type assignments:

\begin{verbatim}
bcl.exe molecule:Filter \\
-input_filenames <input.sdf> \\
-output_matched <output.sdf> \\
defined_atom_types \\
\end{verbatim}

Molecules were additionally processed to remove disconnected elements (e.g. salt ions):
bcl.exe molecule:Split
 -input_filenames <input.sdf> \
 -output <output.sdf> \
 -implementation Largest

**Descriptor generation and neural network model training**

Descriptors for the neural networks were generated with the BCL GenerateDataset application in the **descriptor** namespace:

bcl.exe descriptor:GenerateDataset \n -source "SdfFile(filename=<input_mols>)" \ # generate from SDF \n -output <output.bin> \ # output as BCL partial binary file \n -feature_labels <features.object> \ # contains desired features \n -result_labels <results.object> # contains desired results

The result label for BCL-AffinityNet was simply, "pKd". The result label for BCL-DockANNScore is available in the **Supporting Information**.

Datasets were randomized prior to model training:

bcl.exe descriptor:GenerateDataset \n -source "Randomize(Subset(filename=<output.bin>))" \n -output <output.rand.bin>

For additional details on the formatting of the feature and result object files, please look at the provided sample code object files.

Finally, we can train the model using the “launch.py” script provided with the BCL.

```
./bcl/scripts/machine_learning/launch.py -t cross_validation \n  --config-file <configuration_file> \n  --dataset <output.rand.bin> \n  --id my_model \n  --local
```

The “launch.py” script is a convenience wrapper for complex model training, cross-validation, and feature selection tasks. For a sample configuration file, see provided file “config.ini”.

**Training a K-means applicability domain model**

To train our applicability domain (AD) model, we will directly use the BCL **model:Train** application instead of the “launch.py” script.

bcl.exe model:Train "ApplicabilityDomainKohonen( \n  shuffle=0, \n  balance=0, \n  balance max repeats=1000000, \n  balance target ratio=1, \n  map dimensions(75), \n)"
Input sensitivity analysis

Input sensitivity analysis in the BCL can be achieved rapidly with a single command-line:

```
Bcl.exe descriptor:ScoreDataset \
-source "Subset(filename=<output.rand.bin>)" \
-score "InputSensitivityNeuralNetwork( \
storage=File(directory=<model_directory>,prefix=model), \
weights=(consistency=1))" \
-output <output_file> \
-scheduler PThread <n_threads>
```

Here, the dataset to be passed to the application is the dataset we previously built with the descriptor:GenerateDataset application. We specify that we will calculate the input sensitivity of a neural network with the score flag, and we provide as arguments the neural network to use and indicate that consistency is our final desired metric. Then we just specify an output file via output and the desired number of threads across which the application will distribute the task.

Building pharmacophore maps

All benchmark evaluations were performed with the default small molecule – receptor poses provided for each respective benchmark. Prior to pharmacophore map generation, protein-ligand complexes were minimized into the BCL-AffinityNet score function using a Monte Carlo – Metropolis local perturbation algorithm in the BCL cheminfo:MoleculeFit application:

```
bcl.exe cheminfo:MoleculeFit \
-input_filenames <input_mols.sdf> -random_seed -routine 3 \
-output_filename <output_mols.molfit.sdf> \
-pose_dependent_scoring_function \
"AffinityNet(report_dg=True, receptor filename=<my_receptor.pdb>)"
```
receptor_filename <my_receptor.pdb> \
-refinement_pose_sampling_iterations 100 \ 
-n_cycles 5

For ease of comparison, pharmacophore maps were generated with the <output_mols.molfit.sdf> molecules and mapped to the pre-minimized structures.

Relative pharmacophore maps were built with the following command-line:

bcl.exe cheminfo:MoleculeFeatures \
-input_filenames <input_mols.sdf> \
-model "AffinityNet(receptor filename=iGluR5_A_bcl.pdb,\ 
report_dg=True)" \
-output <output_datafile> \ 
-perturb_type Molecules \ 
-atom_score_distribution_type Rigorous \ 
-color_minmax -1.0 1.0

Absolute pharmacophore maps were built with the following command-line:

bcl.exe cheminfo:MoleculeFeatures \
-input_filenames <input_mols.sdf> \
-model "AffinityNet(receptor filename=iGluR5_A_bcl.pdb,\ 
report_dg=True)" \
-output <output_datafile> \ 
-perturb_type Atoms \ 
-color_minmax -1.0 1.0
Supplementary References

1. Gilson, M. K.; Gilson, H. S. R.; Potter, M. J., Fast Assignment of Accurate Partial Atomic Charges: An Electronegativity Equalization Method That Accounts for Alternate Resonance Forms. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1982-1997.
2. Gasteiger, J.; Marsili, M., Iterative Partial Equalization of Orbital Electronegativity - a Rapid Access to Atomic Charges. *Tetrahedron Lett.* **1980**, *36*, 3219-3228.
3. Aromaticity Detection in Marvin.
4. Miller, K. J., Additivity Methods in Molecular Polarizability. *J. Am. Chem. Soc.* **1990**, *112*, 8533-8542.
5. Chan, S. L.; Labute, P., Training a Scoring Function for the Alignment of Small Molecules. *J. Chem. Inf. Model.* **2010**, *50*, 1724-1735.
6. Pauling, L., The Nature of the Chemical Bond. Iv. The Energy of Single Bonds and the Relative Electronegativity of Atoms. *J. Am. Chem. Soc.* **1932**, *54*, 3570-3582.
7. Su, M.; Yang, Q.; Du, Y.; Feng, G.; Liu, Z.; Li, Y.; Wang, R., Comparative Assessment of Scoring Functions: The Casf-2016 Update. *J. Chem. Inf. Model.* **2019**, *59*, 895-913.
8. Lu, J.; Hou, X.; Wang, C.; Zhang, Y., Incorporating Explicit Water Molecules and Ligand Conformation Stability in Machine-Learning Scoring Functions. *J. Chem. Inf. Model.* **2019**, *59*, 4540-4549.
9. Jiménez, J.; Škalič, M.; Martínez-Rosell, G.; De Fabritiis, G., Kdeep: Protein–Ligand Absolute Binding Affinity Prediction Via 3d-Convolutional Neural Networks. *J. Chem. Inf. Model.* **2018**, *58*, 287-296.
10. DeLuca, S.; Khar, K.; Meiler, J., Fully Flexible Docking of Medium Sized Ligand Libraries with RosettaLigand. *PLoS One* **2015**, *10*, e0132508.
11. Jeffrey, M.; Benjamin, B.; Sandeepkumar, K.; Jens, M., Bcl::Conf – Improved Open-Source Knowledge-Based Conformation Sampling Using the Crystallographic Open Database. 2020.
12. Kothiwale, S.; Mendenhall, J. L.; Meiler, J., Bcl::Conf: Small Molecule Conformational Sampling Using a Knowledge Based Rotamer Library. *J. Cheminform.* **2015**, *7*, 47.
13. Brown, B. P.; Mendenhall, J.; Meiler, J., Bcl::Molalign: Three-Dimensional Small Molecule Alignment for Pharmacophore Mapping. *J. Chem. Inf. Model.* **2019**, *59*, 689-701.