Table 1 Number of training and test compounds for each model

Model	Training pool compounds	Test set compounds	Total
COX-1 regression	172	42	214
COX-2 classification	490	54	544
COX-2 regression	384	43	427
COX-2/COX-1 selectivity	182	32	214

external test set that was withheld from the training process. Various statistics (root mean square error (RMSE), coefficient of determination, etc.) for the training and test sets are analyzed in order to select the best model. It is important that the statistics for the training and test sets are comparable in order to choose a model that is not overtrained. For example, if the training set RMSE is 0.3 and the test set RMSE is 1.0, then the model is likely to be overtrained. Once a model is selected, it can be used to predict activity or property values for new molecules.

ADMET Predictor version 6.5 was used to build a COX-1 regression model, COX-2 classification and regression models, and a COX-2/COX-1 selectivity model. The numbers of compounds in the training pools and test sets for each of the models are listed in Table 1. Compounds were categorized as positive (inhibitory) for the COX-2 classification model if their IC_{50} was 10 μM or below.

Some of the IC_{50} values were qualitative, e.g., COX-2 $IC_{50} > 10$ μM. These data were used to build the COX-2 classification model, but could not be used for the COX-2 regression model. Much of the COX-1 inhibition data was qualitative because researchers were typically looking for COX-2-selective molecules. Thus, fewer compounds were available for COX-1 models.

7 Combinatorial Elaboration and Fragment Assembly

The known COX-2-selective inhibitors in our data set contain three rings. These rings are labeled S (scaffold), A, and B in Fig. 8. The scaffold, ring S, is the most distinctive portion of cyclooxygenase inhibitors. There is some diversity in ring A, whereas ring B is always a phenyl ring with either a sulfonamide or methyl sulfone in the *para* position. We focused on generating relatively uncharacterized chemical series in our scaffold-hopping example.

Scaffold hopping was performed in two different ways – combinatorial elaboration and fragment assembly – in the design module of MedChem Studio. The former algorithm is similar to the one described by Stewart et al. (2006). Examples of transformations include reversing an amide, i.e., swapping positions of the amine and carbonyl, and replacing a carboxylic acid with a tetrazole ring. The "Combinatorial Transforms" option in MedChem Studio accesses approximately 120 potential molecular transformations developed in collaboration with medicinal chemists. Rofecoxib was used as the starting structure.

Fragment assembly generates novel molecules by replacing a designated portion of a known inhibitor with fragments from a database of drug fragments. The database of drug fragments was generated by breaking molecules from the World