inhibitory activity as well as *selectivity* with respect to the two cyclooxygenase isoforms.

Matched molecular pair analysis relates structural change to activity change on a qualitative basis. Three QSAR models were also generated for the COX isoforms. The first was a classification model for predicting which compounds would exhibit COX-2 IC_{50} below 10 μM. The second and third were regression models for predicting the actual inhibitory potency against COX-1 and COX-2.

These steps were followed by scaffold hopping into novel chemistry space and subsequent elaboration into virtual libraries for which a range of properties were predicted. Here, a battery of ADMET models from ADMET Predictor™[1] were used to predict physicochemical (solubility, logP, pK_a, etc.), pharmacokinetic (plasma protein binding, blood/plasma concentration ratio, fraction unbound in plasma, and volume of distribution), metabolic (high intrinsic clearance and CYP 3A4 inhibition), and toxicological (hERG inhibition, carcinogenicity, liver injury, etc.) properties for the compounds that were predicted to be sufficiently active. Compounds that violated ADMET property thresholds determined from a reference subset of the World Drug Index (WDI) were penalized. A list of synthesis candidates was then constructed by manual inspection of the 240 that remained in the library. Those that were successfully synthesized were then submitted for biological testing and limited ADMET property determination.

4 Computational Methods

4.1 Data Set

While much literature on cyclooxygenase inhibitors has appeared since 1990, we focused on the data reported in articles published by either Searle (which was eventually acquired by Pfizer) or Merck to provide a starting point for ligand-based design using a few hundred compounds. The actual molecular structures together with COX-1 and COX-2 IC_{50} values were extracted from several sources including the ChEMBL database (Gaulton et al. 2012) and a published compilation (Chavatte et al. 2001) as well as from the primary literature.

4.2 Extracting Data from ChEMBL

The ChEMBL database can be extremely helpful for creating knowledge databases because it makes structure, biological data, and journal references for a given target accessible from a single source. The steps used to extract our COX-2 data from

[1] ADMET Predictor™ is distributed by Simulations Plus, Inc., Lancaster CA, http://www.simulations-plus.com.