

Combining Ensemble Learning with a Fragment-Based Topological Approach to Generate New Molecular Diversity in Drug Discovery: In Silico Design of Hsp90 Inhibitors

Alejandro Speck-Planche^{a*}

^aResearch Program on Biomedical Informatics (GRIB), Hospital del Mar Medical Research Institute (IMIM), 08003, Barcelona, Spain

CONTACT: Alejandro Speck-Planche (alejspivanovich@yahoo.es)

Phone: +34 678042076

Table S1. Local measures of the statistical quality and performance of the mt-QSAR-EL model.

\mathbf{b}_t	$\mathbf{N}_{\text{Total}}(\mathbf{b}_t)$	$\mathbf{N}_{\text{Active}}(\mathbf{b}_t)$	$\mathbf{N}_{\text{Inactive}}(\mathbf{b}_t)$	$[\mathbf{Sn}(\%)]_{\mathbf{b}_t}$	$[\mathbf{Sp}(\%)]_{\mathbf{b}_t}$	Series
Hsp90 α	486	252	234	85.32%	75.21%	Training Set
Hsp90 β	288	121	167	80.17%	92.81%	Training Set
Hsp90 α	161	83	78	89.16%	76.92%	Test Set
Hsp90 β	94	38	56	78.95%	87.50%	Test Set

The symbols in this table have the following meanings:

\mathbf{b}_t refers to the biological target (Hsp90 isoform). Therefore, Hsp90 α is the heat shock protein 90 alpha, while Hsp90 β is the heat shock protein 90 beta.

$\mathbf{N}_{\text{Total}}(\mathbf{b}_t)$ – total number of molecules depending on the target.

$\mathbf{N}_{\text{Active}}(\mathbf{b}_t)$ – number of active molecules depending on the target.

$\mathbf{N}_{\text{Inactive}}(\mathbf{b}_t)$ – number of inactive molecules depending on the target.

$[\mathbf{Sn}(\%)]_{\mathbf{b}_t}$ – Local sensitivity expressed as the percentage of molecules correctly classified as active depending on the target.

$[\mathbf{Sp}(\%)]_{\mathbf{b}_t}$ – Local specificity expressed as the percentage of molecules correctly classified as inactive depending on the target.