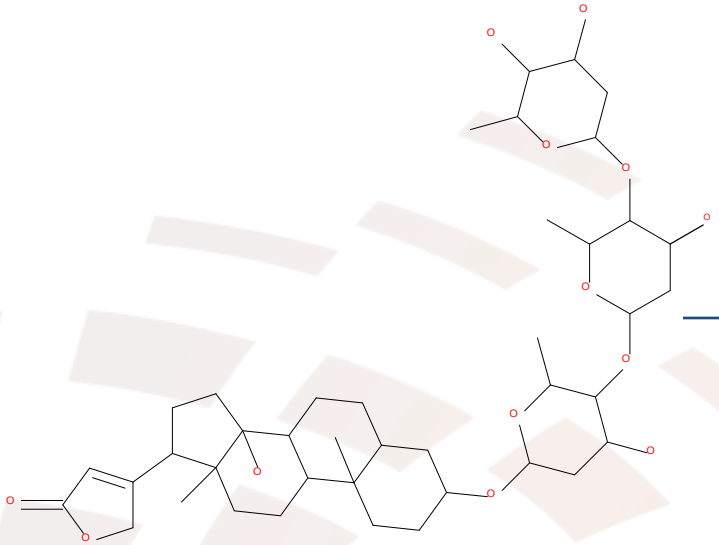
Conformal Prediction in Bioclipse

Ernst Ahlberg
Computational ADME & Safety

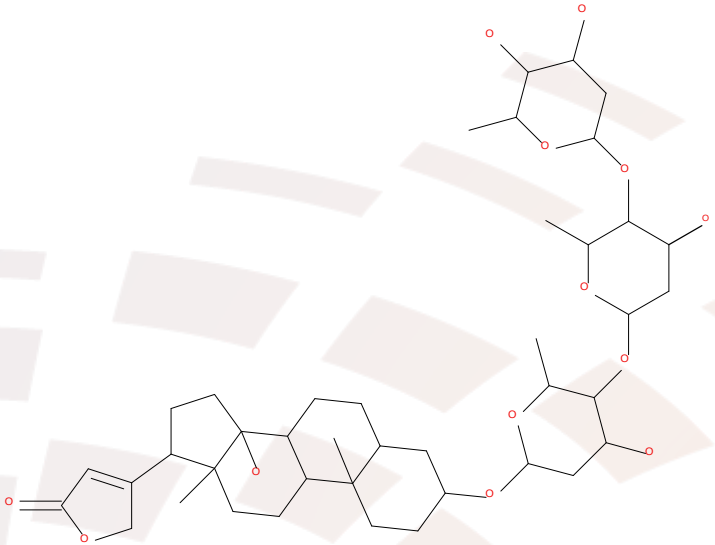
AstraZeneca

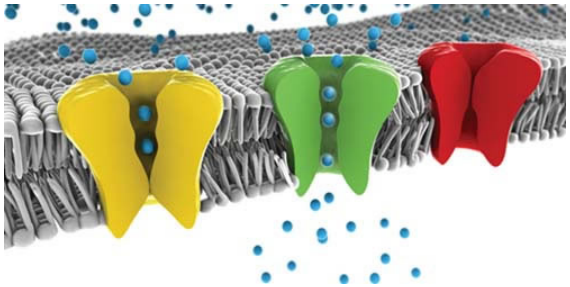
Computational Safety



AstraZeneca

Computational Safety





AstraZeneca

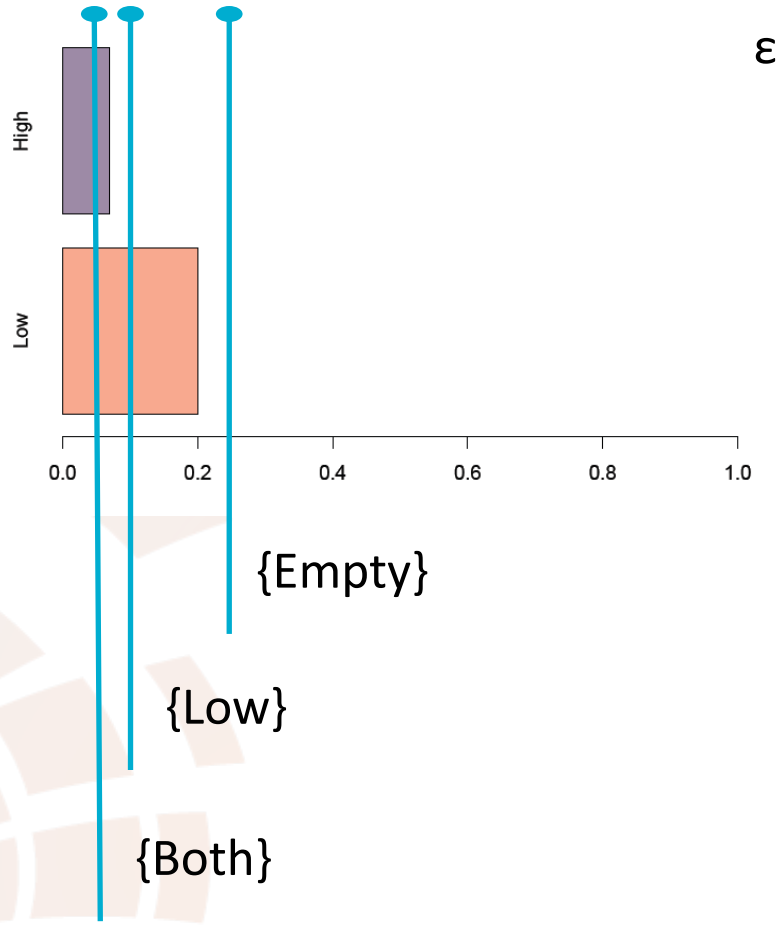
The Conformal Prediction Framework

- **Non Conformity measure**: a measure of how unusual an example looks relative to previous examples, described as a monotonically increasing function.
- **Prediction region**: A set Γ^{ε} that contains the true value with probability at least $1 - \varepsilon$.
- **Exchangeability**: Consider variables $z_1, \ldots, z_N$. Suppose that for any collection of N values, the N! different orderings are equally likely. Then we say that $z_1, \ldots, z_N$ are exchangeable.
- **Validity**: An Γ^{ε} is valid if it contains the truth 1-ε of the time.
- **Efficiency**: The smallest prediction region at a given ε produces the most informative prediction, ie $\Gamma_1^{\varepsilon} > \Gamma_2^{\varepsilon} \Rightarrow \Gamma_2^{\varepsilon}$ is more efficient than Γ_1^{ε}. Also, the prediction regions for different ε are nested such that we can obtain a small prediction region with a low level of confidence and with a higher level of confidence the prediction region will be expanded, containing the smaller prediction region. In short when $\varepsilon_1 \geq \varepsilon_2$ we have $\Gamma^{\varepsilon 1} \subseteq \Gamma^{\varepsilon 2}$.

Vladimir Vovk, Alex Gammerman, and Glenn Shafer. *Algorithmic Learning in a Random World*. Springer, New York, 2005

AstraZeneca

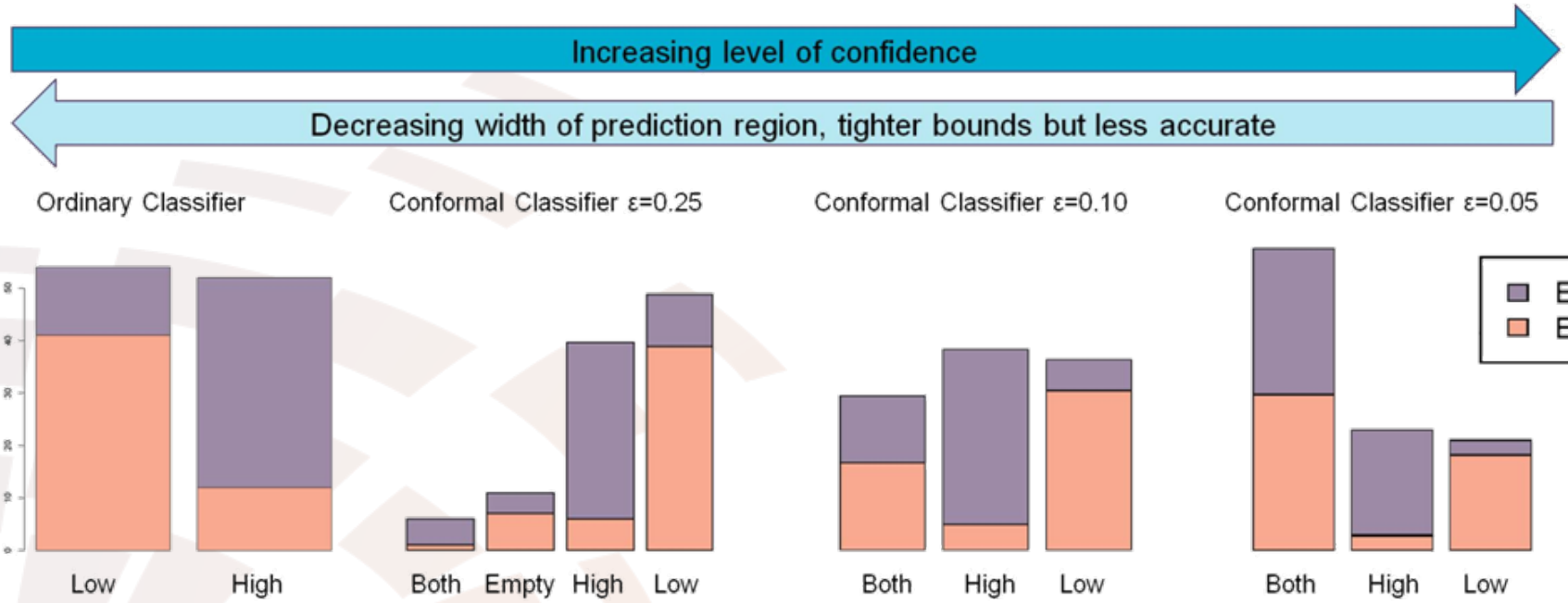
P-values for a single prediction



Prediction Regions at
ε 0.05, 0.10, 0.25

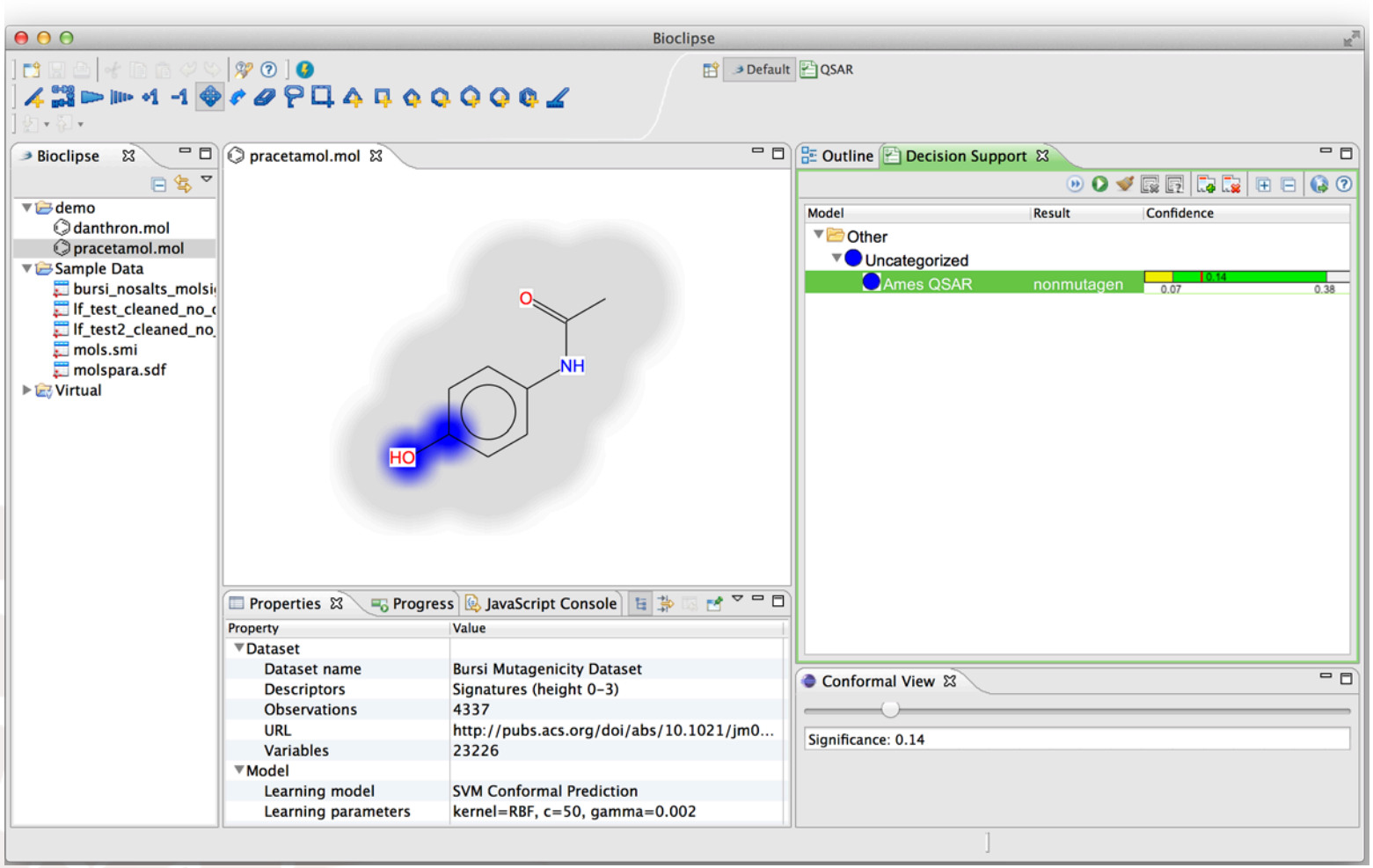
AstraZeneca

Compared to an ordinary classifier



AstraZeneca

Demo



AstraZeneca