

Christian Kramer

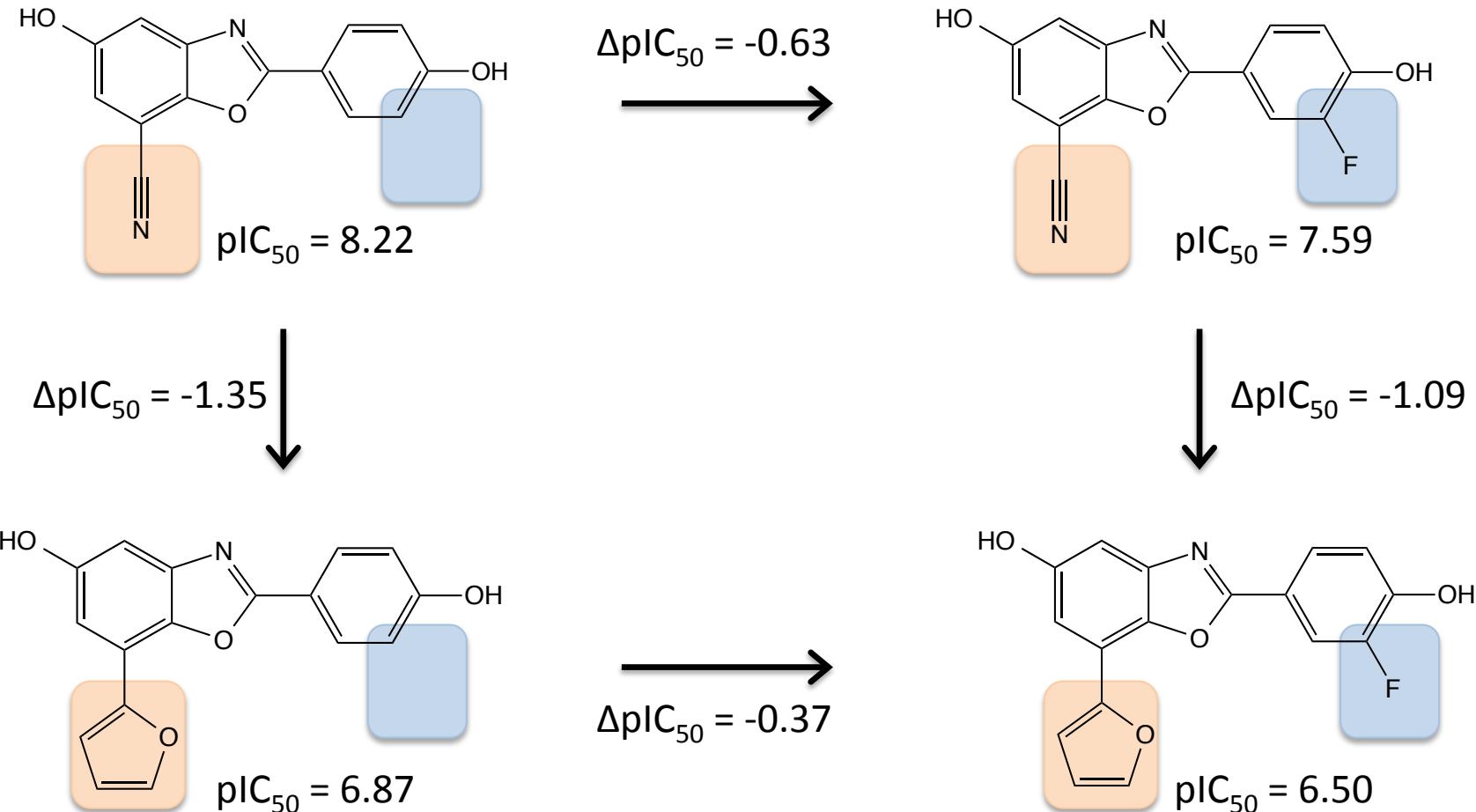
# Nonadditivity analysis with the RDKit

Christian Kramer  
University of Innsbruck

Christian Kramer

# What is Additivity?

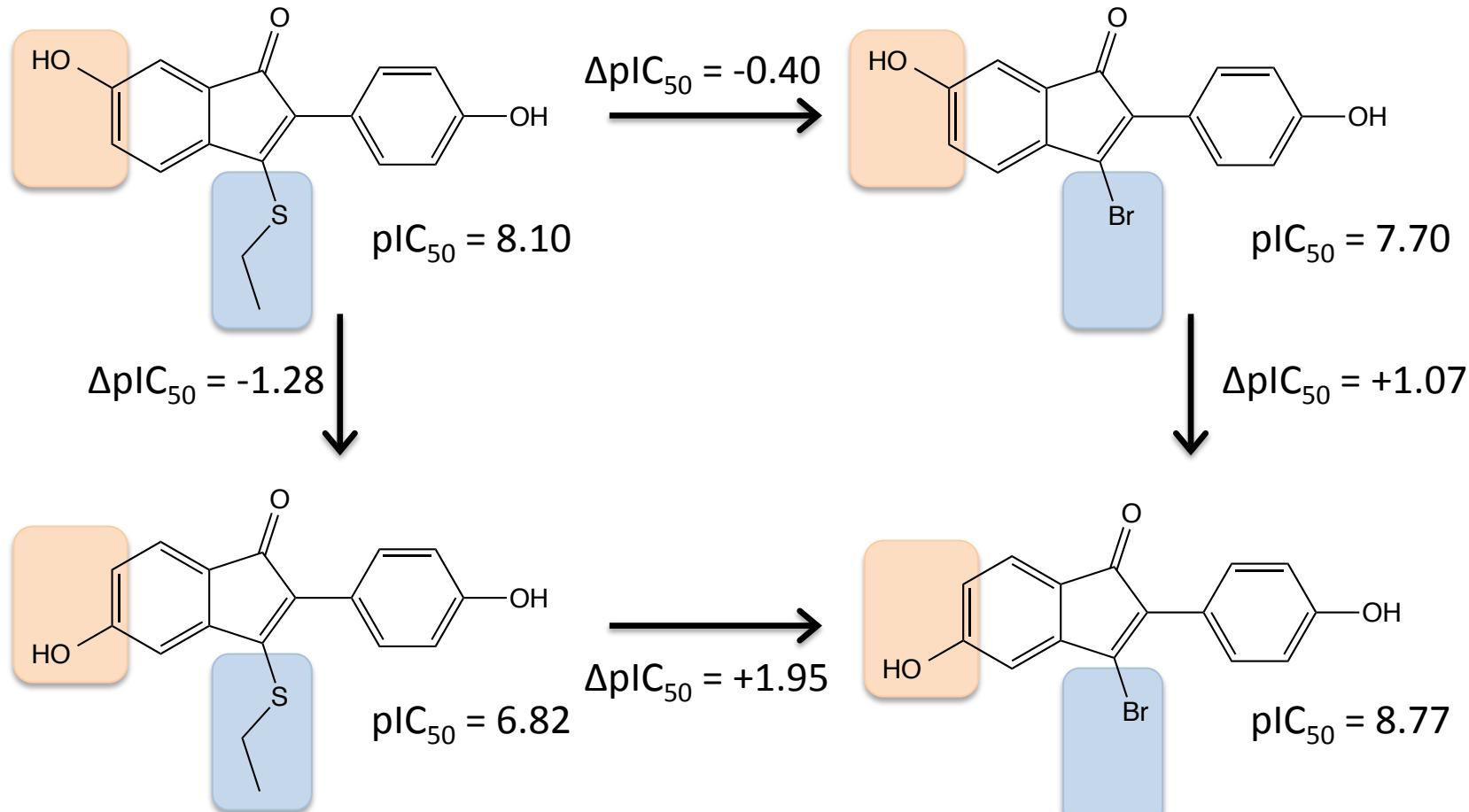
Target: Estrogen Receptor  $\beta$



Malamas, M.S. et al. Design and Synthesis of Aryl Diphenolic Azoles as Potent and Selective Estrogen Receptor- $\beta$  Ligands. *J. Med. Chem.* **2004**, 47, 5021-5040.

## What is Nonadditivity?

Target: Estrogen Receptor  $\beta$





Christian Kramer

What is Additivity?

### **Ligand-based design concept:**

The order of introducing substituents does not matter.

→ **Same substituent, same position:  
same free energy difference**

→ True if substituents do not interact (directly or indirectly)

### **CADD techniques assuming additivity:**

Free-Wilson Analysis

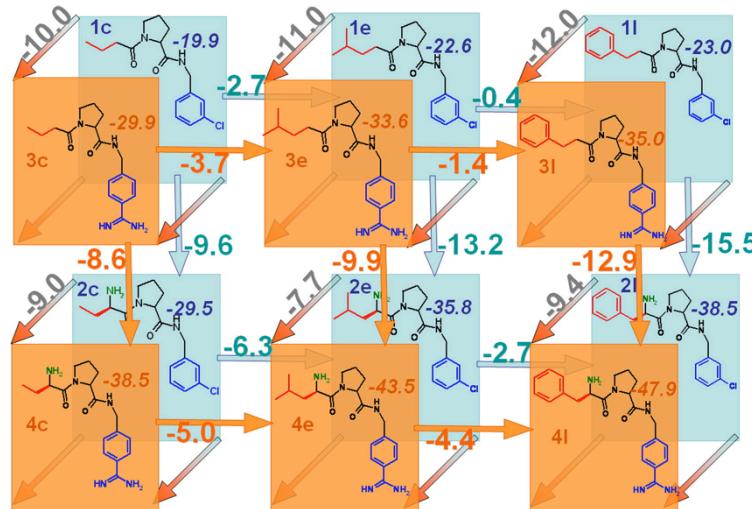
Linear QSAR models

Matched Molecular Pair Analysis

# Christian Kramer

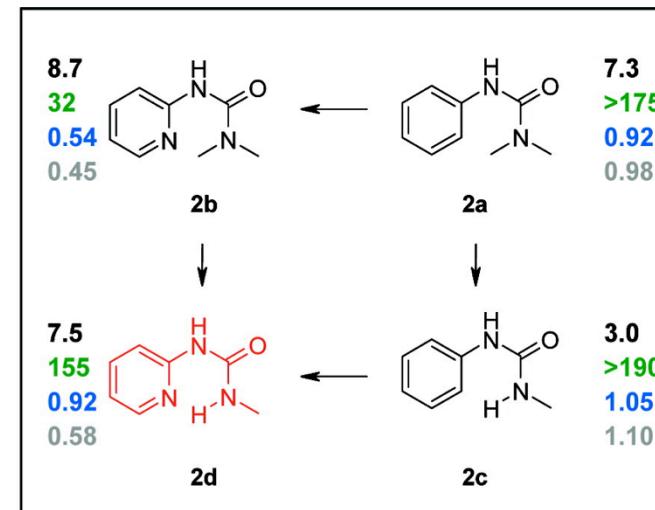
# Where does nonadditivity come from?

# Differences in residual mobility due to (missing) anchors



Baum, B.; Muley, L.; Smolinski, M.; Heine, A.; Hangauer, D.; Klebe, G. Non-additivity of Functional Group Contributions in Protein–Ligand Binding: A Comprehensive Study by Crystallography and Isothermal Titration Calorimetry. *J. Mol. Biol.* **2010**, *397*, 1042–1054.

## Intramolecular H-Bonding



Kuhn, B.; Mohr, P.; Stahl, M. Intramolecular Hydrogen Bonding in Medicinal Chemistry. *J. Med. Chem.* **2010**, 53, 2601-2611

- Water Solvation Shell
  - Complete Ligand rearrangement

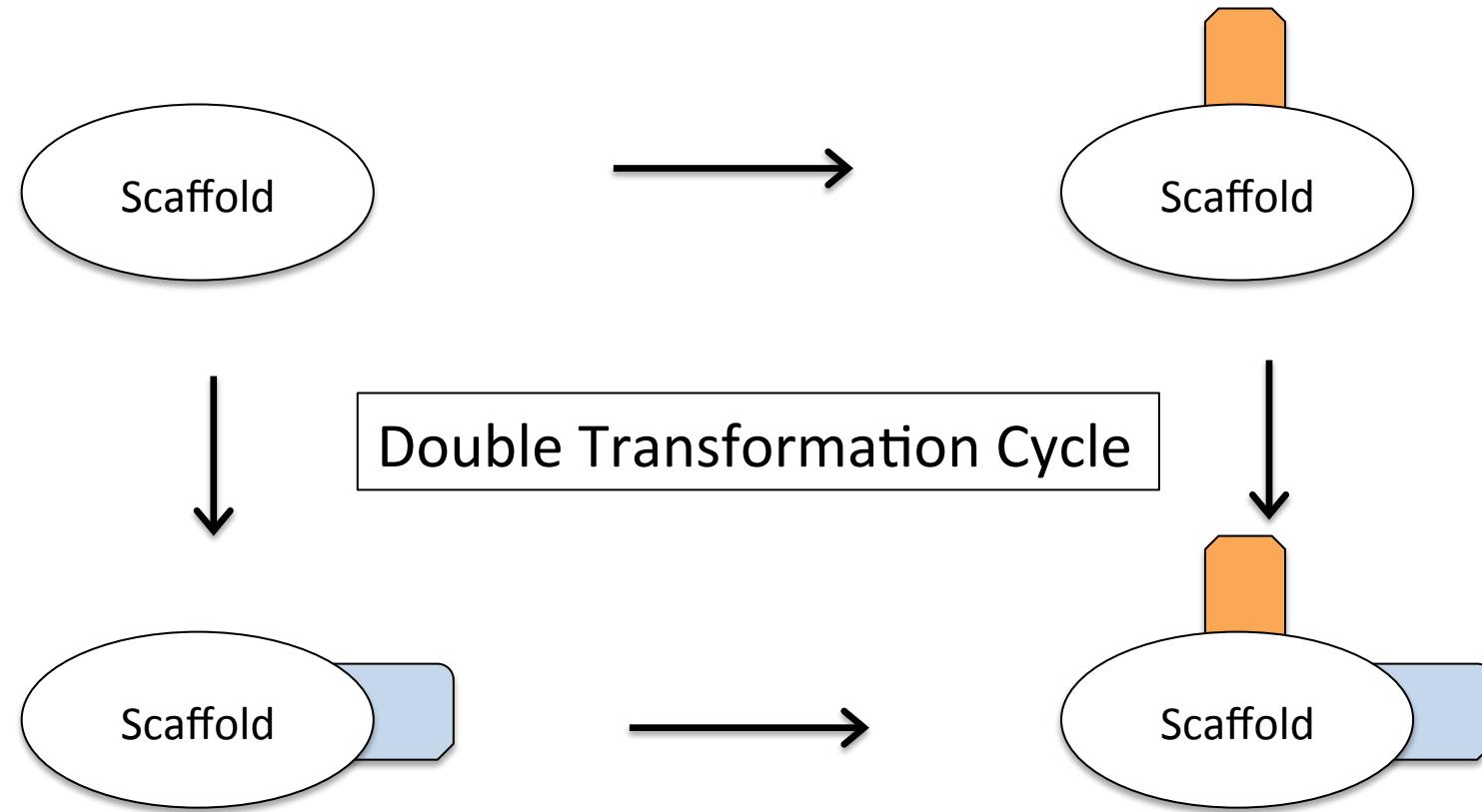
→ Nonadditivity is an important SAR feature

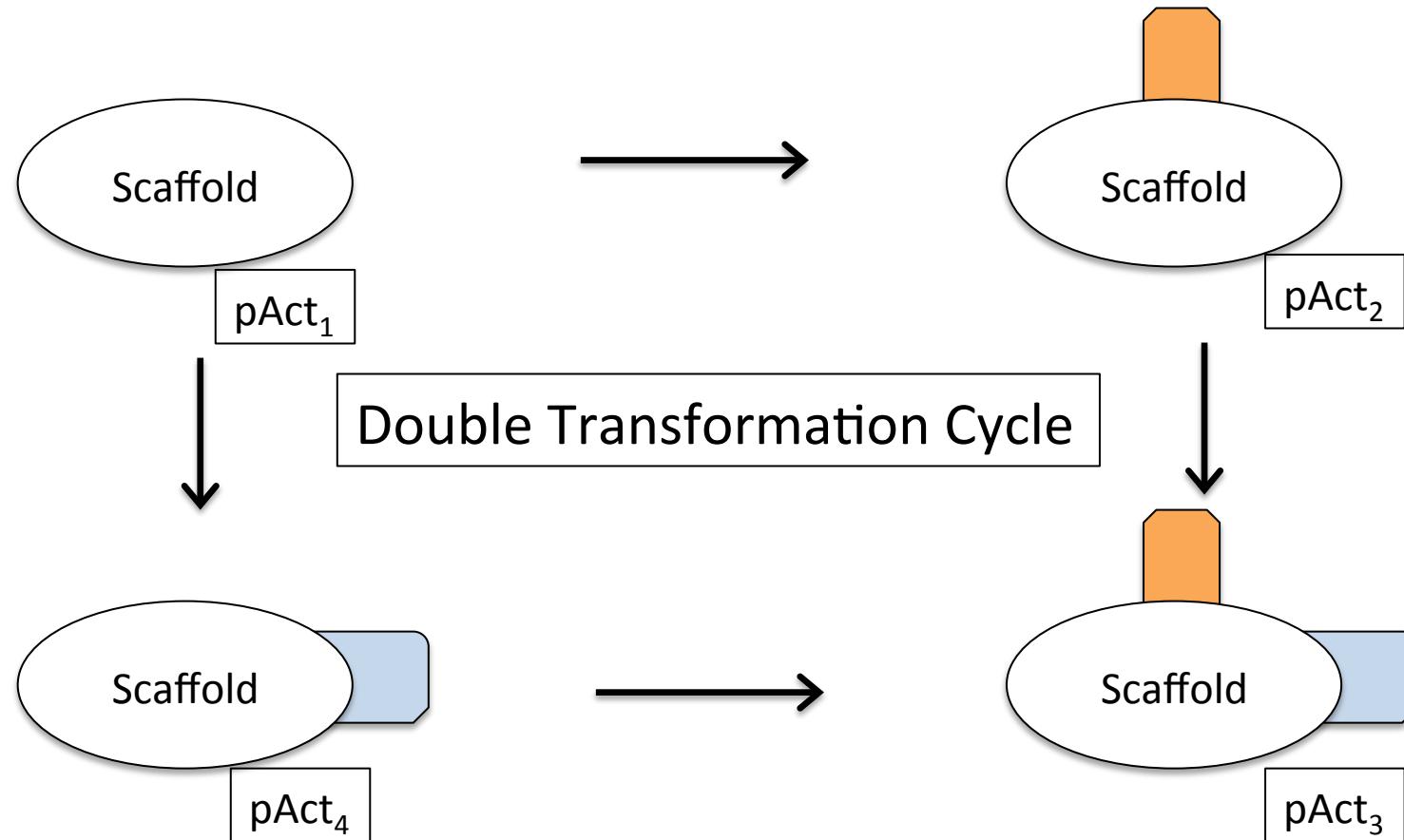
Christian Kramer

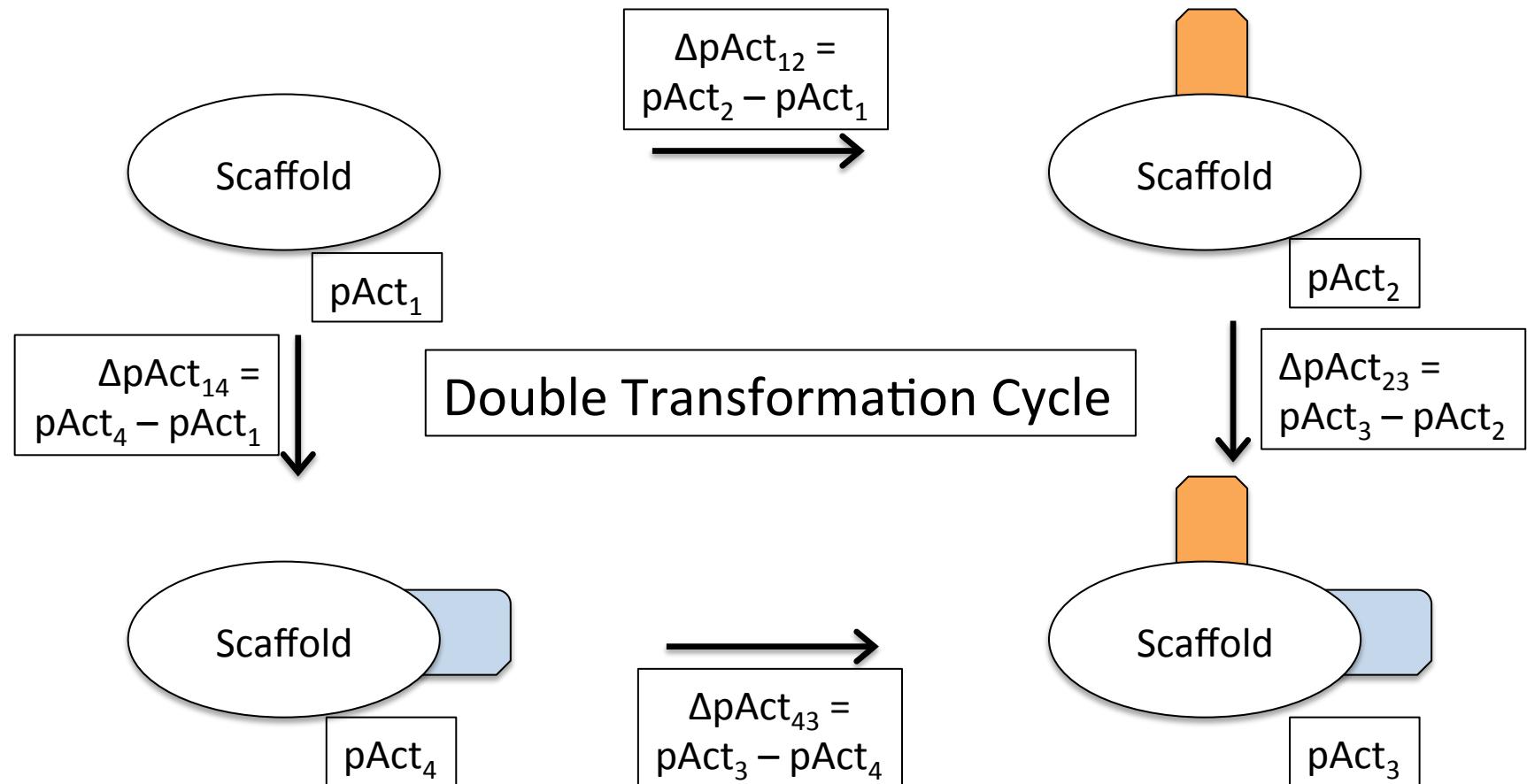
# Systematic Nonadditivity Analysis

Christian Kramer

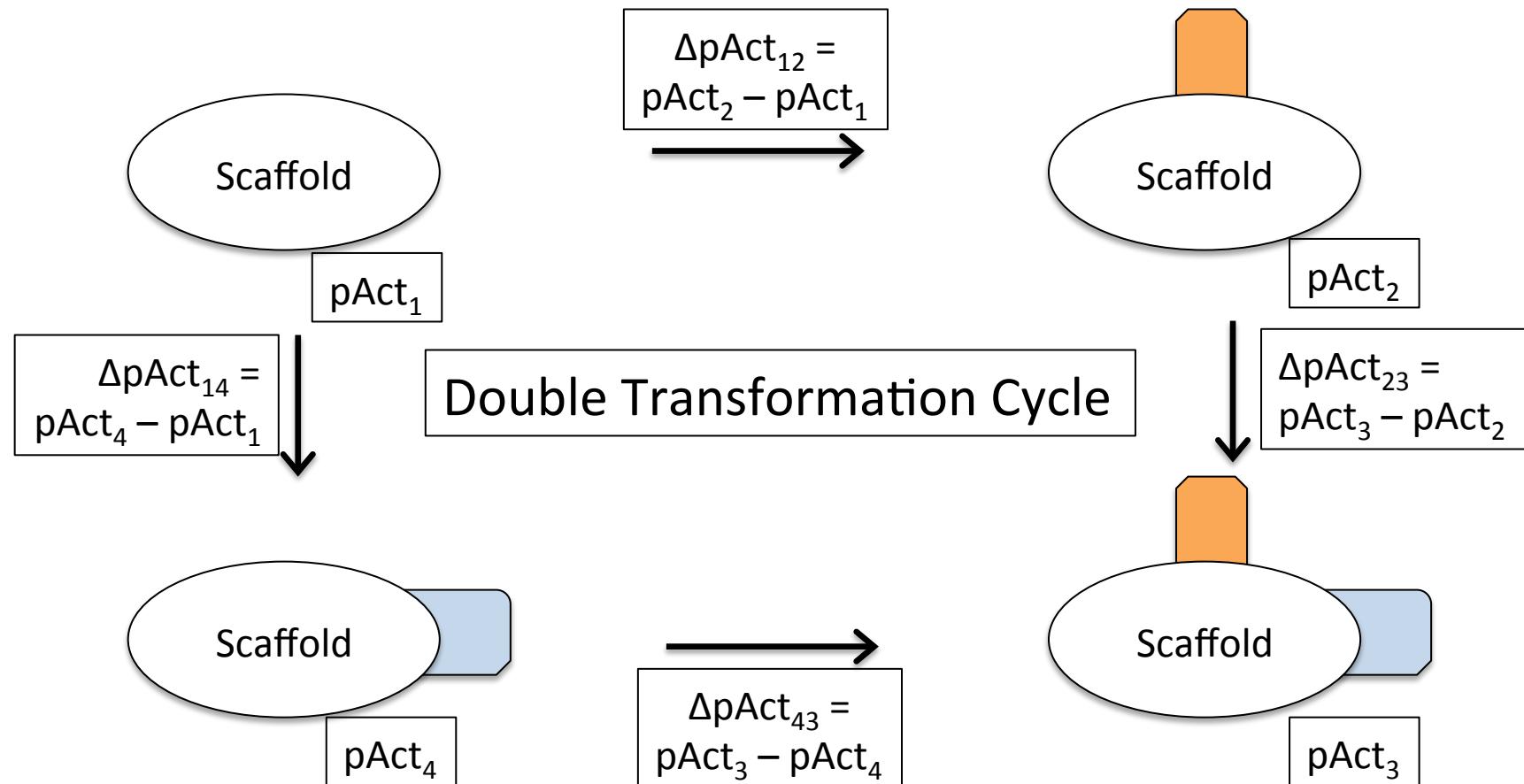
# Nonadditivity systematics





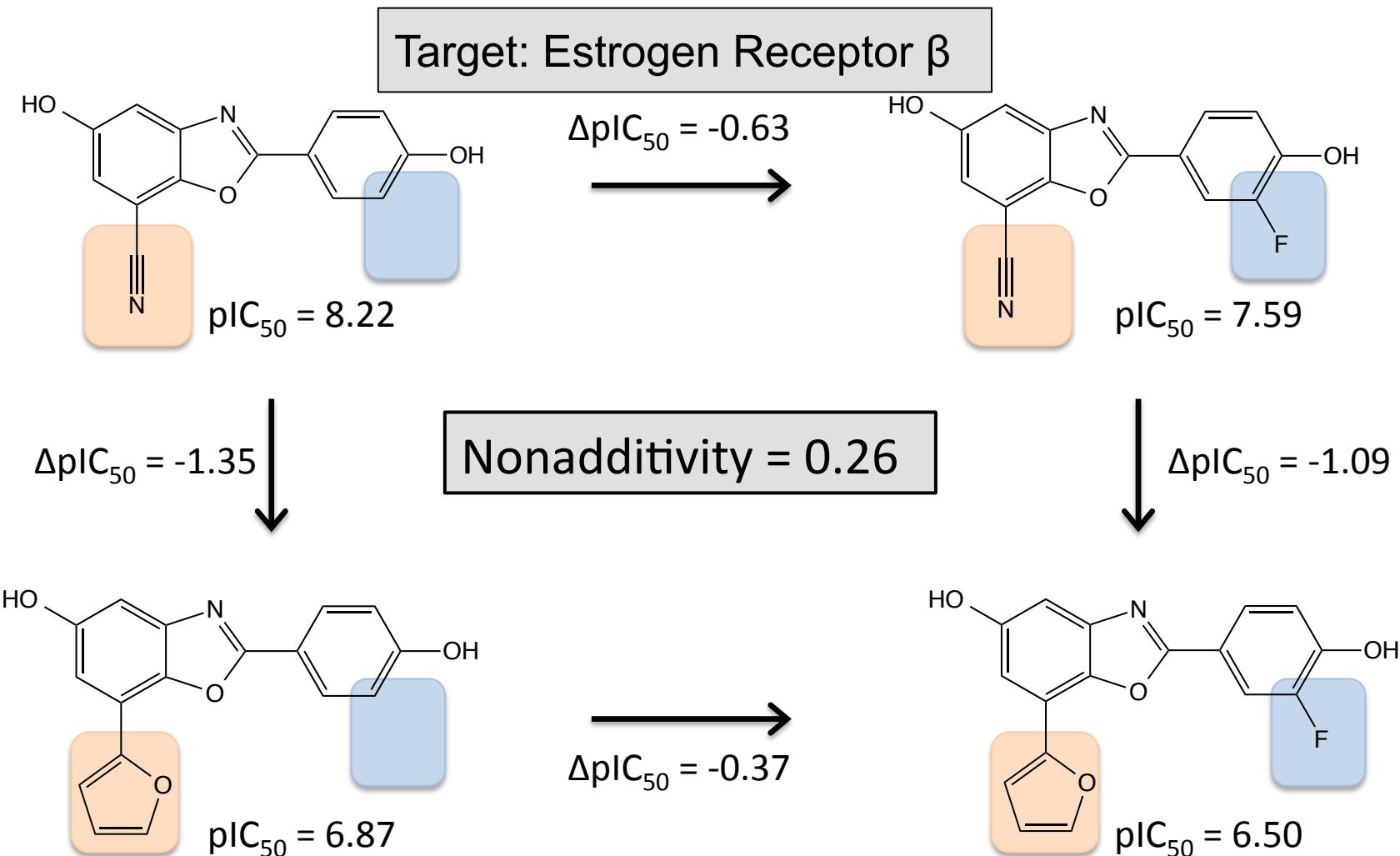


$$\text{Nonadditivity} = p\text{Act}_{43} - p\text{Act}_{12} = p\text{Act}_{23} - p\text{Act}_{14}$$



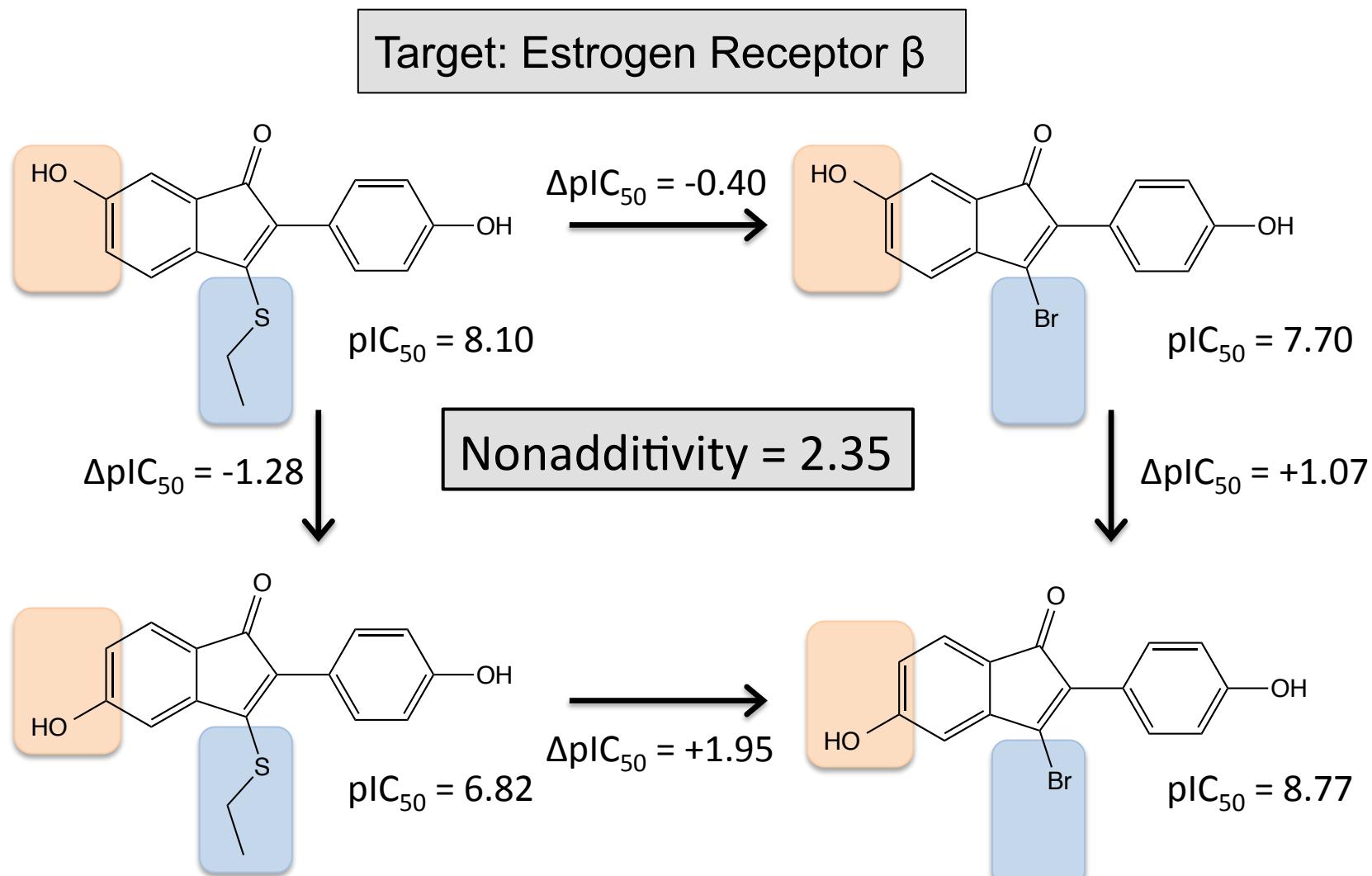
Christian Kramer

# What is Additivity?



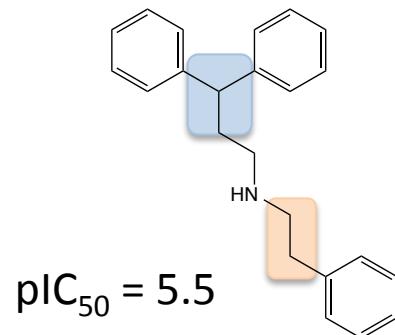
Malamas, M.S. et al. Design and Synthesis of Aryl Diphenolic Azoles as Potent and Selective Estrogen Receptor- $\beta$  Ligands. *J. Med. Chem.* **2004**, 47, 5021-5040.

# What is Nonadditivity?



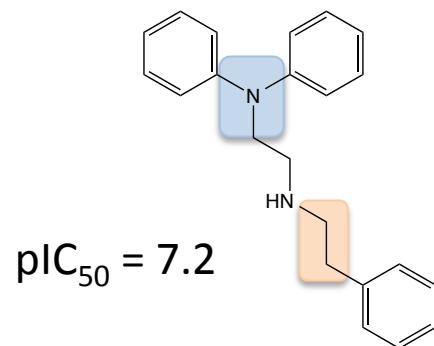
Christian Kramer

# Ligand-based Nonadditivity Analysis



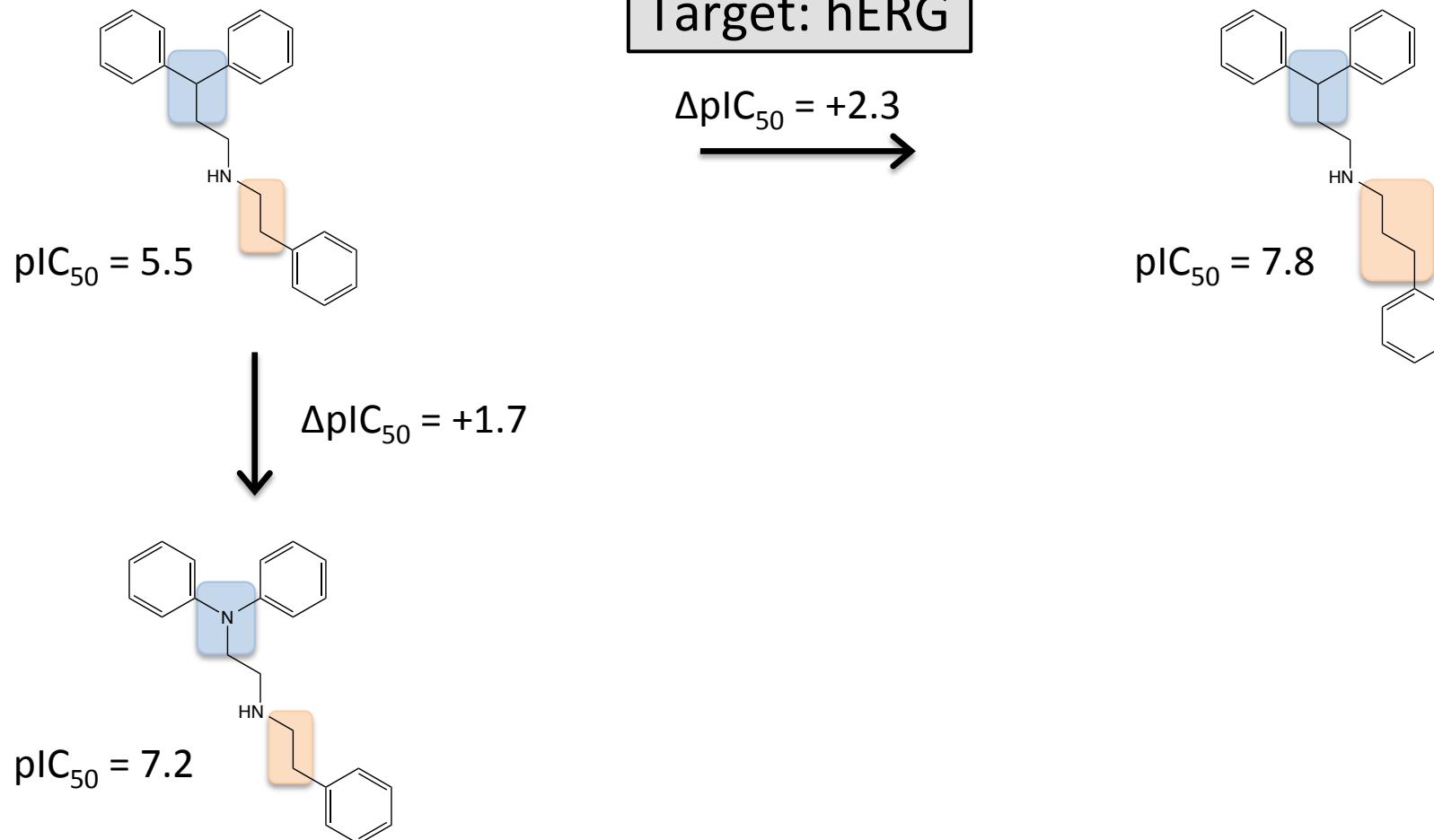
Target: hERG

$$\Delta\text{pIC}_{50} = +1.7$$

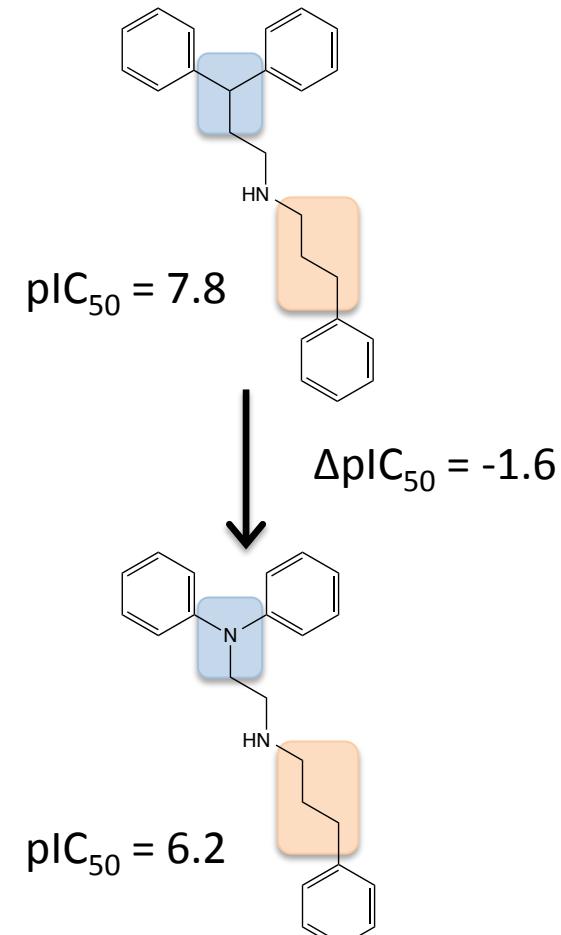
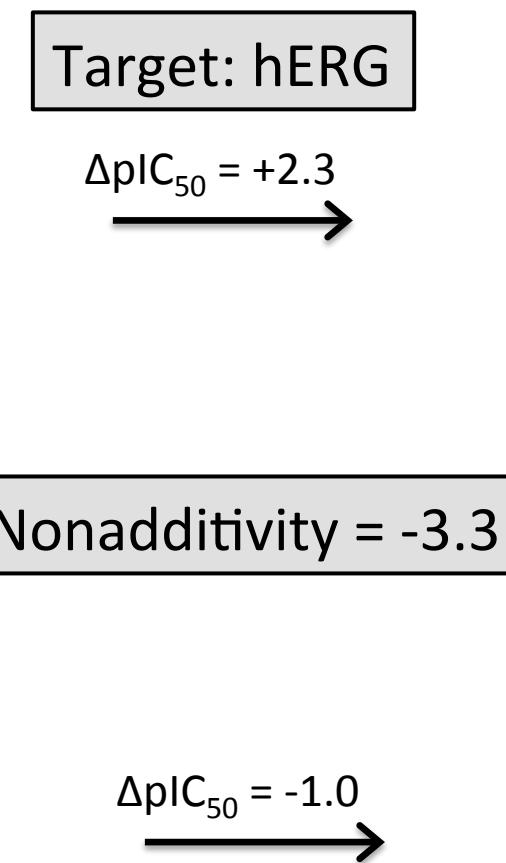
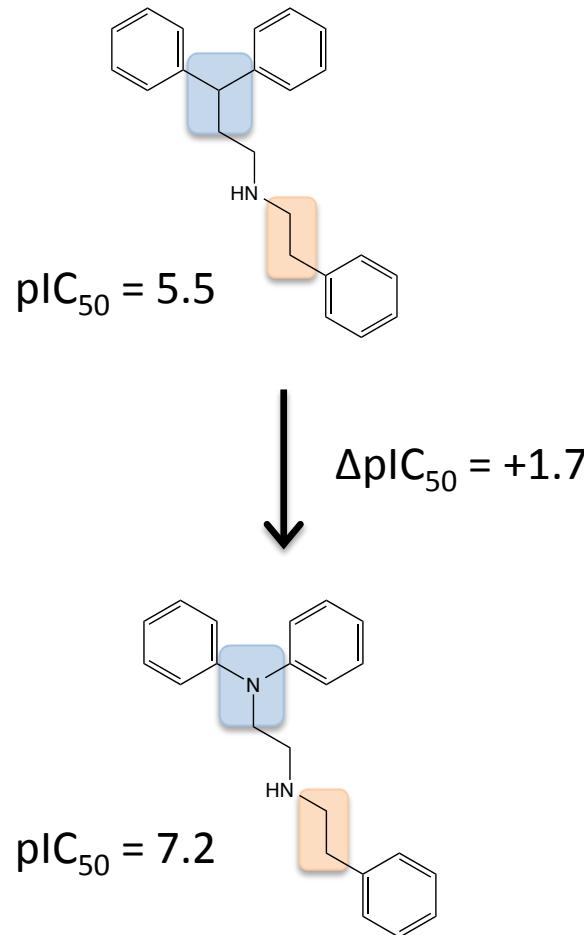


Christian Kramer

# Ligand-based Nonadditivity Analysis



# Ligand-based Nonadditivity Analysis



Christian Kramer

# Assembling Nonadditivity cycles - Technical Details



Christian Kramer

## Nonadditivity cycle creation - details

### Link between two structures

- Transformations from Jameed Hussain's RDKit MMPA implementation

### Cycle creation

- Build full neighbor dictionary (for every compound in set)
- Remove neighbor for unwanted transformations (e.g. linker exchange)
- Generate cycles with depth-first search
- Use a set of early stopping criteria to speed up search and eliminate double cycles

### Cycle interpretation

- Find Xray structures with similar ligands  
(automatic implementation possible for ChEMBL datasets)
- Compare nonadditivity distribution with expected noise distribution

## Cycle generation - Code

```

def get_circles(neighs):
    """
    Assemble circle information
    """

    neighs1 = copy.deepcopy(neighs)

    circles = []
    for k1 in neighs1.keys():
        neighs2 = copy.deepcopy(neighs1)
        for k2 in neighs1[k1]:
            if k2[0] in neighs1.keys():
                tr1_back = k2[1].split(">>")[1]+">>>"+k2[1].split(">>")[0]
                for k3 in neighs1[k2[0]]:
                    if k3[0] == k1: continue
                    if k3[0] in neighs1.keys():
                        if k1 in [i[0] for i in neighs[k3[0]]]: continue
                        tr2_back = k3[1].split(">>")[1]+">>>"+k3[1].split(">>")[0]
                        for k4 in neighs1[k3[0]]:
                            if not k4[1] == tr1_back: continue
                            if k4[0] == k2[0]: continue
                            if k4[0] in neighs2.keys():
                                for k5 in neighs1[k4[0]]:
                                    if not k5[1] == tr2_back: continue
                                    if k5[0] == k1: circles.append((k1,k2[0],k3[0],k4[0]))

```

Remove 1st compound from primary neighbor dictionary

Stop if 1<sup>st</sup> and 3<sup>rd</sup> compound are identical

Stop here if we build cycle of three

Stop if 1<sup>st</sup> and 3<sup>rd</sup> transformation are not inverse

Stop if 2<sup>nd</sup> and 4<sup>th</sup> compound are identical

Stop if 2<sup>nd</sup> and 4<sup>th</sup> transformation are not inverse

Remove 2<sup>nd</sup> compound from Secondary neighbor dictionary

Christian Kramer

# Nonadditivity and the Impact of Experimental Uncertainty

## What about Experimental Uncertainty?

Start:  $\text{Nonadditivity} = \text{pAct}_{43} - \text{pAct}_{12}$   
 $= \text{pAct}_3 - \text{pAct}_4 - \text{pAct}_2 + \text{pAct}_1$

with  $\text{pAct}_x = \text{pAact}_{x,true} + \epsilon_{x,uncertainty}$

and  $\epsilon_{x,uncertainty} \sim \mathcal{N}(\mu = 0, \sigma = \sigma_{uncertainty})$

gives  $\text{Nonadditivity} = \text{pAct}_{3,true} - \text{pAct}_{4,true} - \text{pAct}_{2,true} + \text{pAct}_{1,true} +$   
 $\epsilon_{3,uncertainty} - \epsilon_{4,uncertainty} - \epsilon_{2,uncertainty} + \epsilon_{1,uncertainty}$

since  $\sigma(\epsilon_1 + \epsilon_2 + \epsilon_3 + \epsilon_4) = \sqrt{\sigma_{uncertainty}^2 + \sigma_{uncertainty}^2 + \sigma_{uncertainty}^2 + \sigma_{uncertainty}^2}$

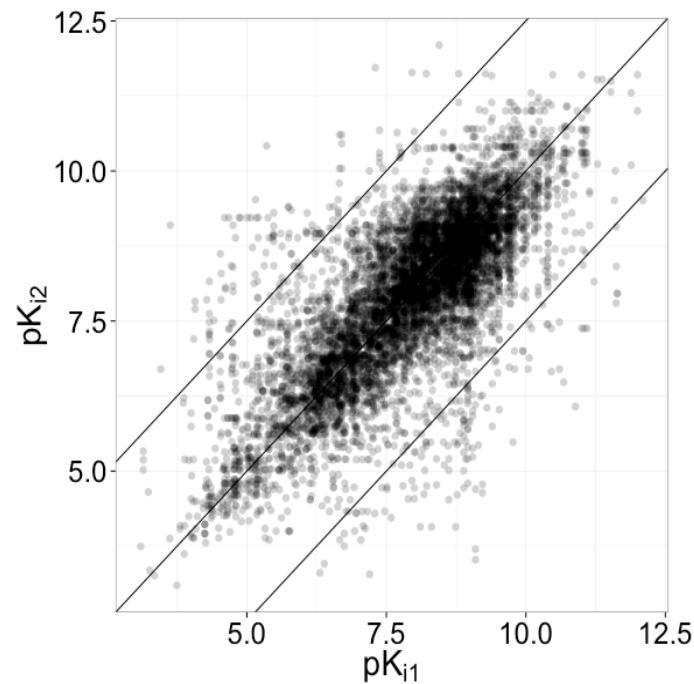
follows  $\epsilon_{\text{Nonadditivity}} \sim \mathcal{N}(\mu = 0, \sigma = 2 \cdot \sigma_{uncertainty})$

→ Apparent nonadditivity is strongly influenced by experimental uncertainty

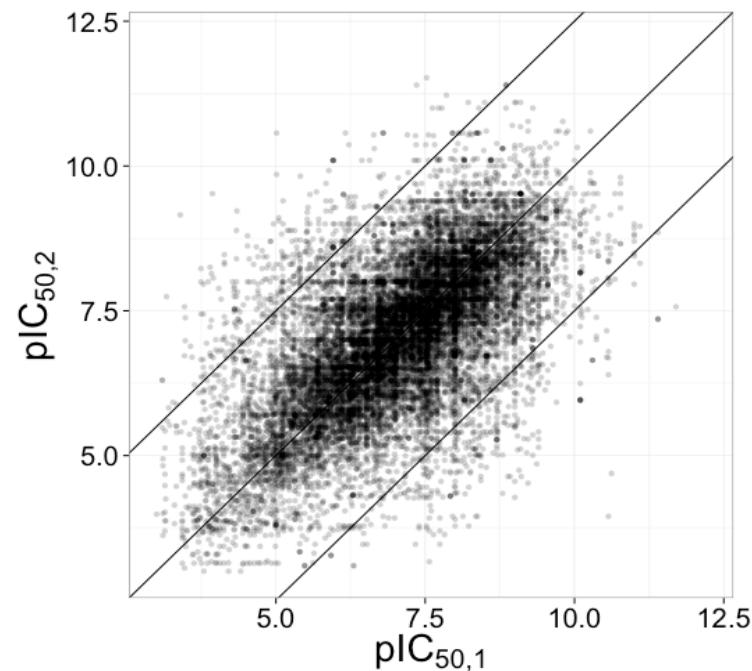
Christian Kramer

# The experimental uncertainty and variability of heterogeneous activity data

Heterogeneous  $pK_i$  data



Heterogeneous  $pIC_{50}$  data

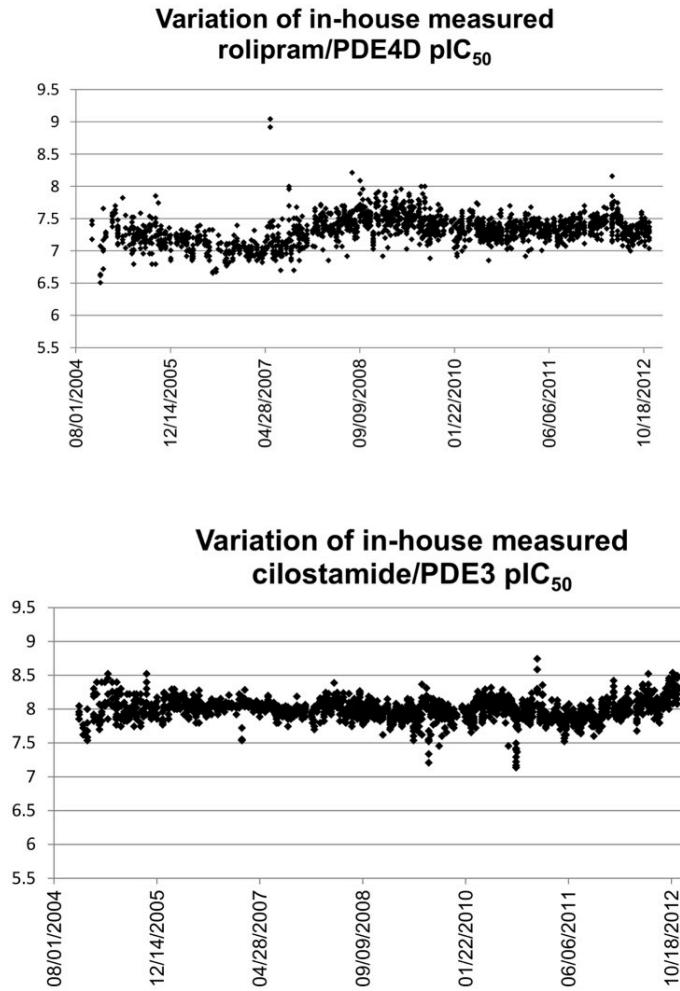


$$\begin{aligned}\sigma_{\text{hetero}, pK_i} \\ \sigma_{\text{hetero}, pIC50}\end{aligned}$$

$$\begin{aligned}= 0.54 \text{ log units} \\ = 0.69 \text{ log units}\end{aligned}$$

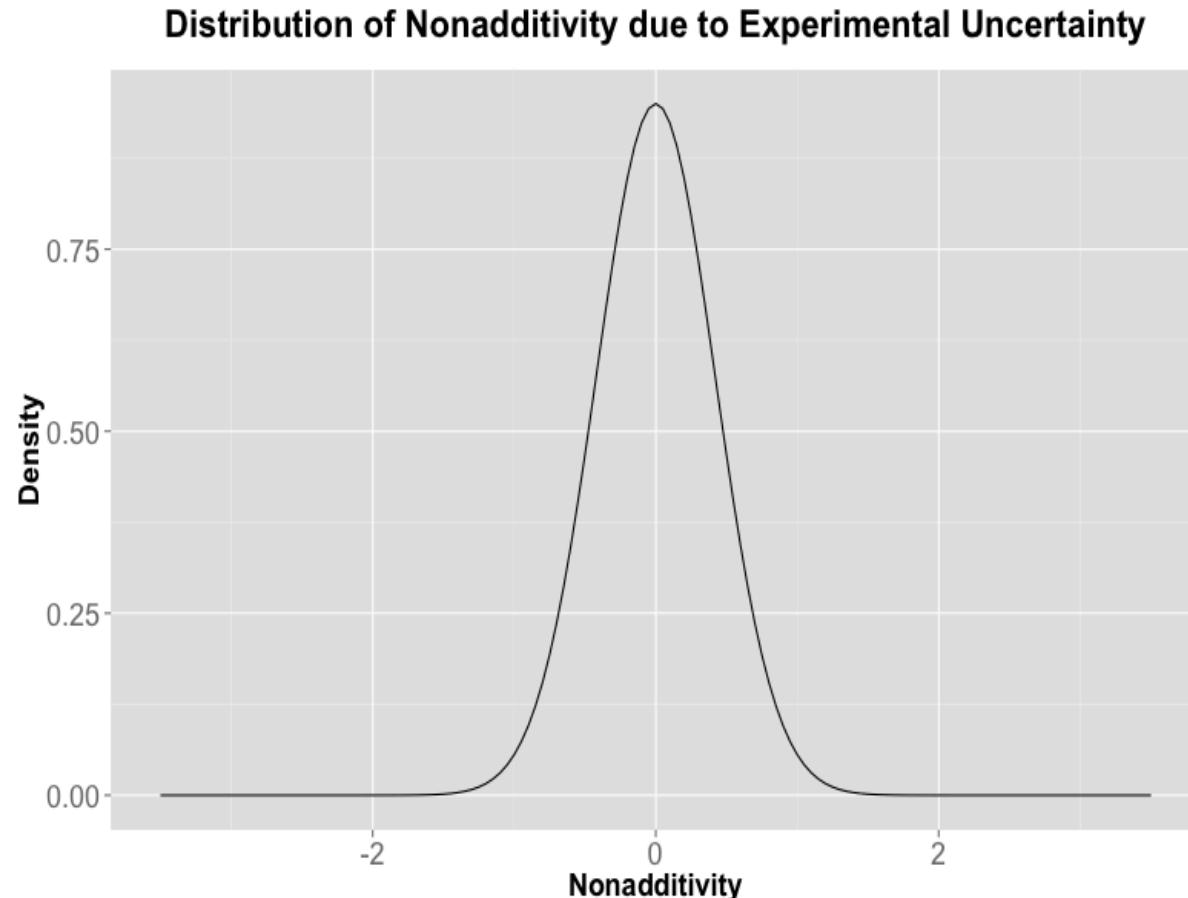
Christian Kramer

# Experimental uncertainty of homogeneous measurements



$\sigma_{\text{hetero, pKi}}$	= 0.54 log units
$\sigma_{\text{hetero, pIC50}}$	= 0.69 log units
$\sigma_{\text{homo, pIC50 \& pKi}}$	~ 0.20 log units

# Nonadditivity due to noise



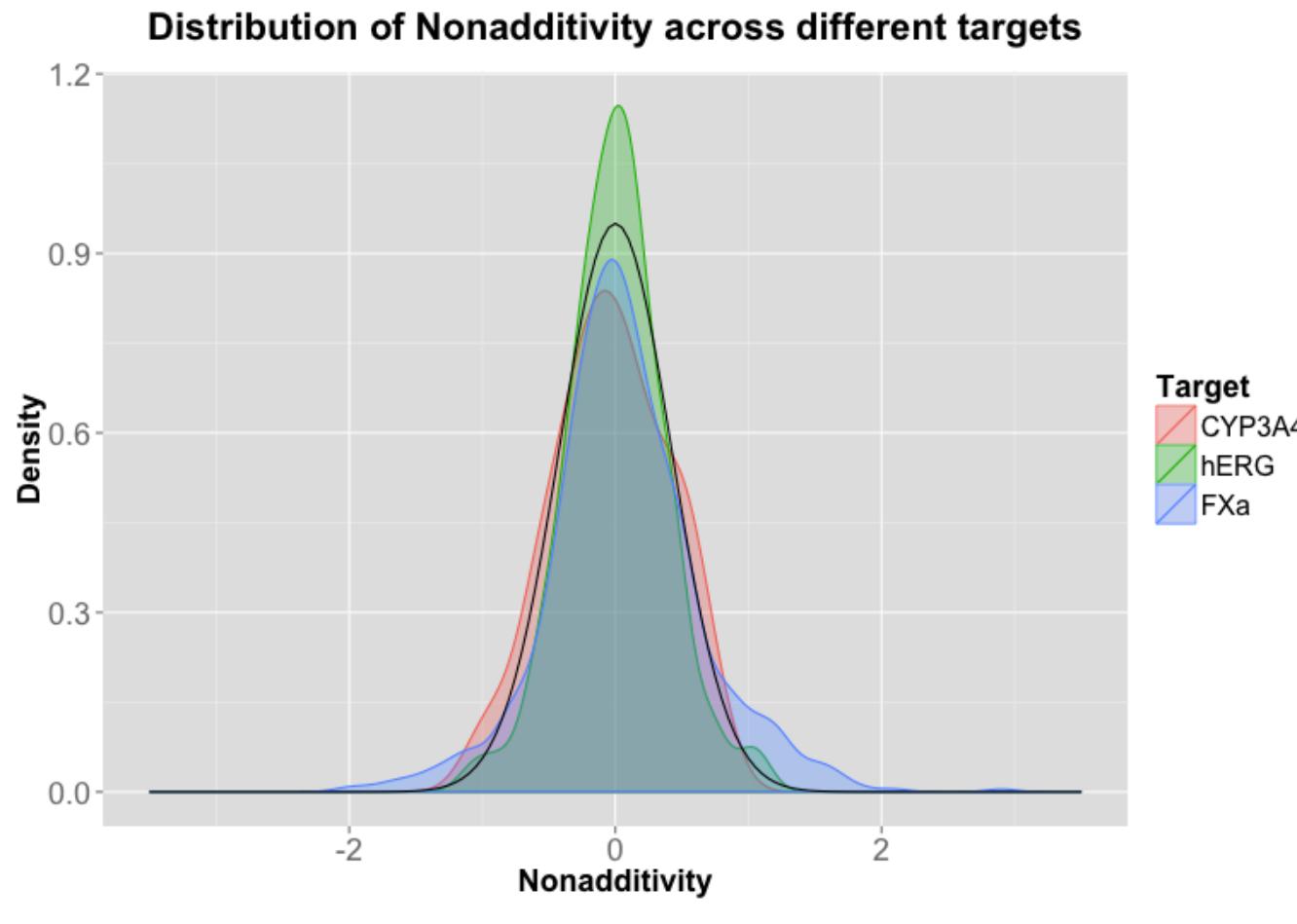
Line

Distribution from  
experimental  
uncertainty

$$\sigma = 0.21 \text{ log units}^*$$

This distribution is to be expected from perfectly additive datasets!

## Whole Dataset nonadditivity characterization



Data  
CHEMBL pK<sub>i</sub>

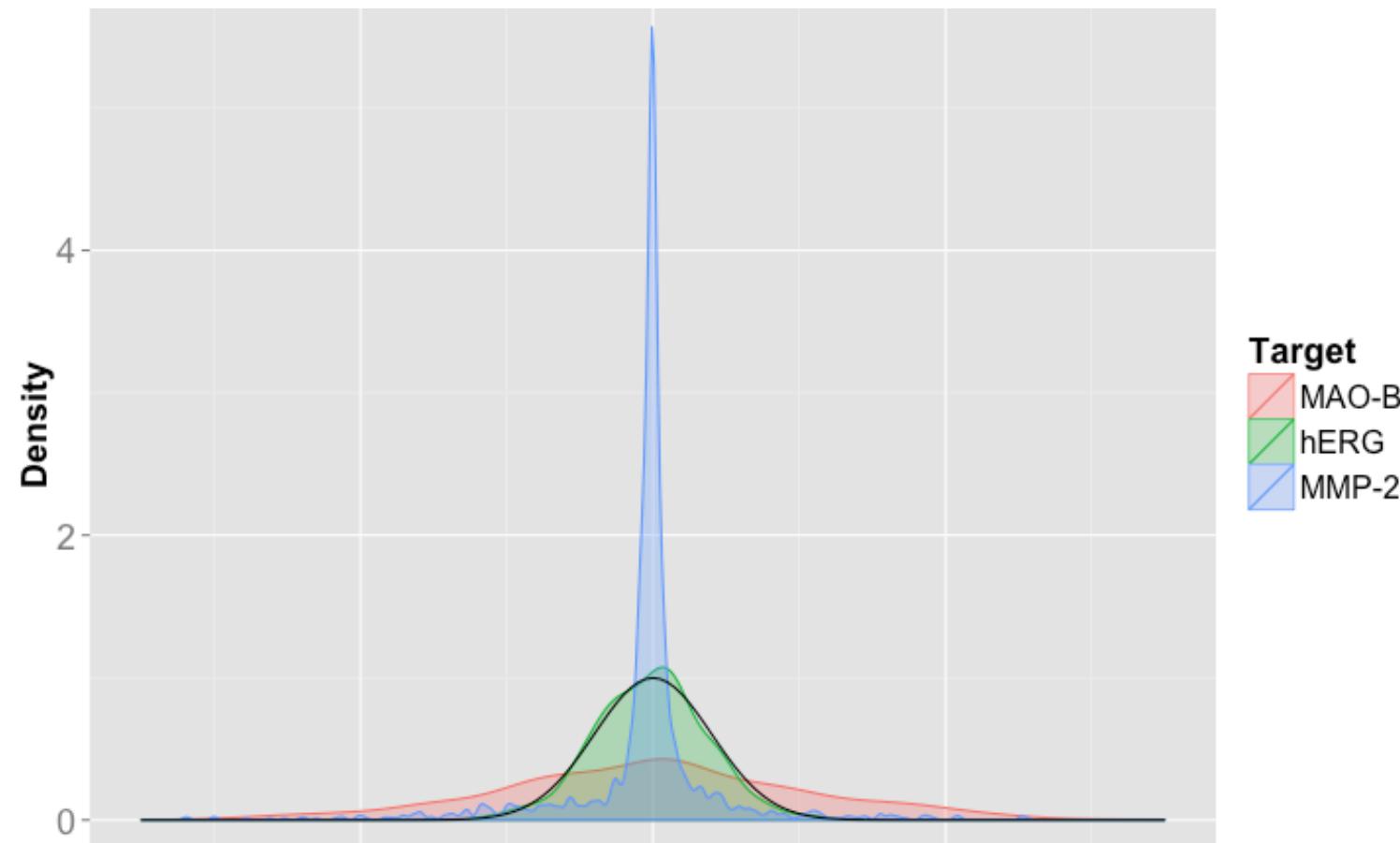
Code  
- RDKit  
- Jameed Hussain's  
MMPA Code

Filter  
- Same Assay Only  
- Selected linker  
substitutions only

Line  
Distribution from  
experimental  
uncertainty  
 $\sigma=0.21$  log units

# Christian Kramer

## Whole Dataset nonadditivity characterization – CHEMBL data

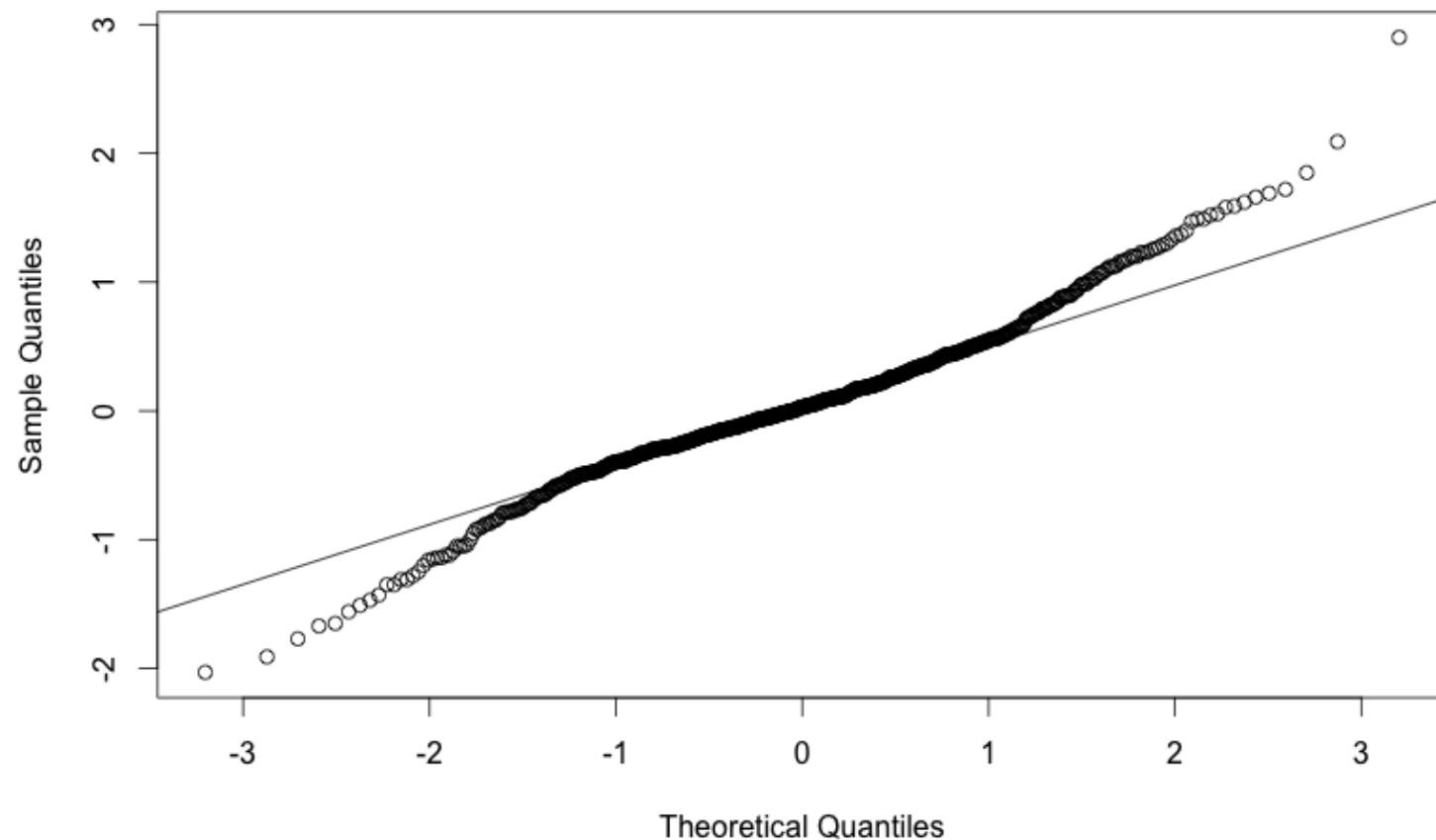


Distributions can be very different – Does this depend on the assay or the target?

Christian Kramer

# Visual Nonadditivity Distribution Analysis with QQ plots

**Factor Xa Ki Nonadditivity distribution**



Christian Kramer

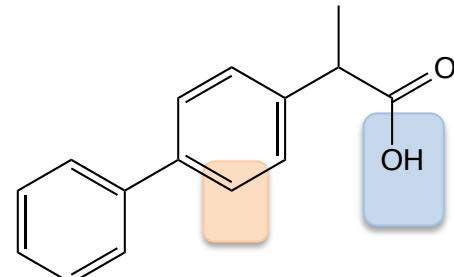
# SAR lessons from Nonadditivity Cycles

## Structural Reasons for Nonadditivity

- Extract all ligand datasets ( $K_i$  &  $IC_{50}$ ) from ChEMBL17 with more than 1000 ligands
  - Generate all double transformation cycles
  - Check PDB for explanatory structures:  
Same sequence, ligand similarity  $TC > 0.6$
  - Analyze all cycles (structure-based & original literature) with at least one ligand  $TC > 0.95$  and nonadditivity  $> 2.0$  (5 sigma rule, compare CERN particle physics rule)
- 80 cycles, 24 distinct scaffold series on 21 different targets

# scaffolds	Nonadditivity Reason
4	Covalent binders
3 (incl. one from above)	Data error
2	Insufficient Structural Data
1	Wrong Ligand in PDB structure
<b>15</b>	<b>Left for Analysis</b>

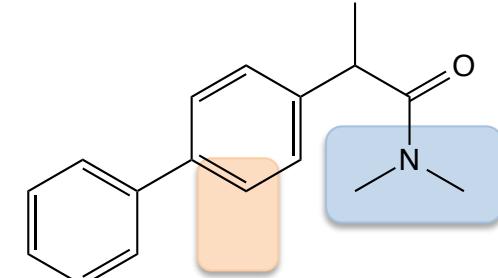
# Nonadditivity: Not so trivial reason



$\text{pIC}_{50} = 7.49$

Target: COX1

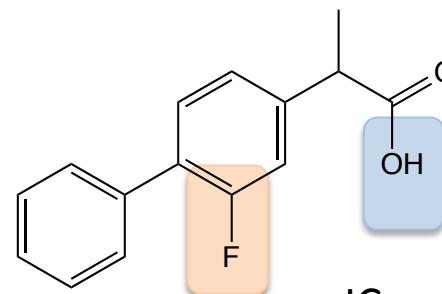
$$\Delta \text{pIC}_{50} = -2.17$$



$\text{pIC}_{50} = 5.32$

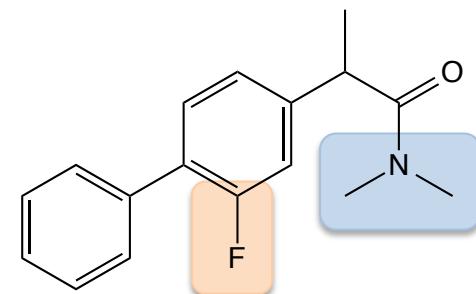
$$\Delta \text{pIC}_{50} = +0.43$$

Nonadditivity = 2.08



$\text{pIC}_{50} = 7.92$

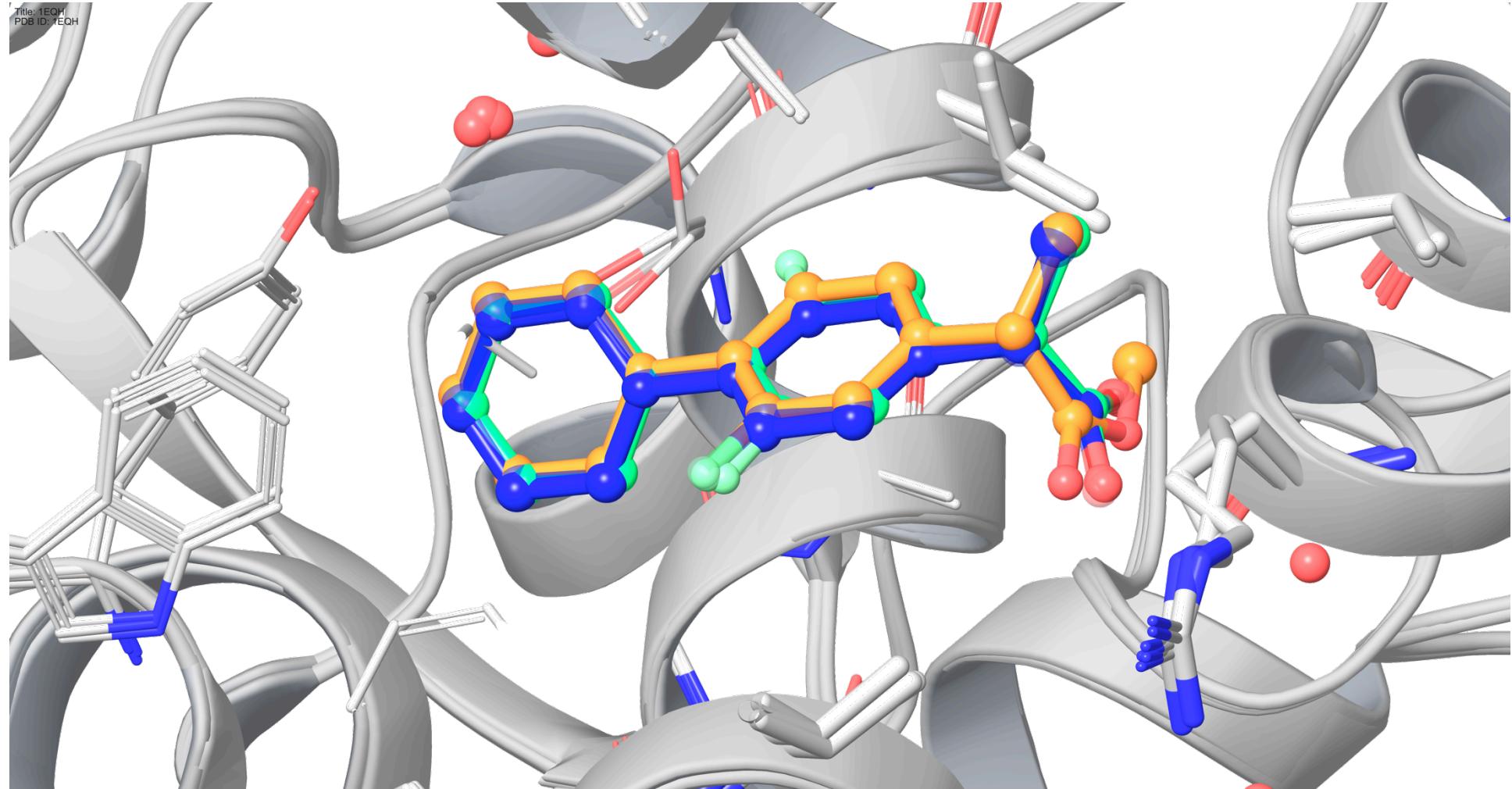
$$\Delta \text{pIC}_{50} = -4.25$$



$\text{pIC}_{50} = 3.67$

Christian Kramer

COX1: All Ligands are oriented the same way



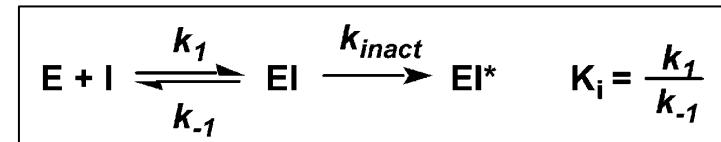
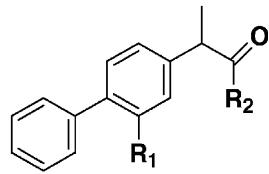
1EQH green; 2AYL: blue; 1HT5: orange  
31

# Explanation: Acids show tight binding behaviour

**Table 1.** Inhibition of cyclooxygenase-1 by flurbiprofen derivatives

Compd	R <sub>1</sub>	R <sub>2</sub>	IC <sub>50</sub> (μM) <sup>a</sup>	Time Dependent?
<b>1</b> (flurbiprofen)	F	OH	0.012±0.004	Yes
<b>2</b>	H	OH	0.032±0.002	Yes
<b>3</b>	F	OCH <sub>3</sub>	210±28	No
<b>4</b>	F	NH <sub>2</sub>	8.3±2	No
<b>5</b>	F	NH(CH <sub>3</sub> ) <sub>2</sub>	24±6	No
<b>6</b>	F	N(CH <sub>3</sub> ) <sub>2</sub>	215±7	No
<b>7</b>	H	OCH <sub>3</sub>	28.5±1	No
<b>8</b>	H	NH <sub>2</sub>	64.5±5	No
<b>9</b>	H	NH(CH <sub>3</sub> ) <sub>2</sub>	19.5±2	No
<b>10</b>	H	N(CH <sub>3</sub> ) <sub>2</sub>	4.8±1	No

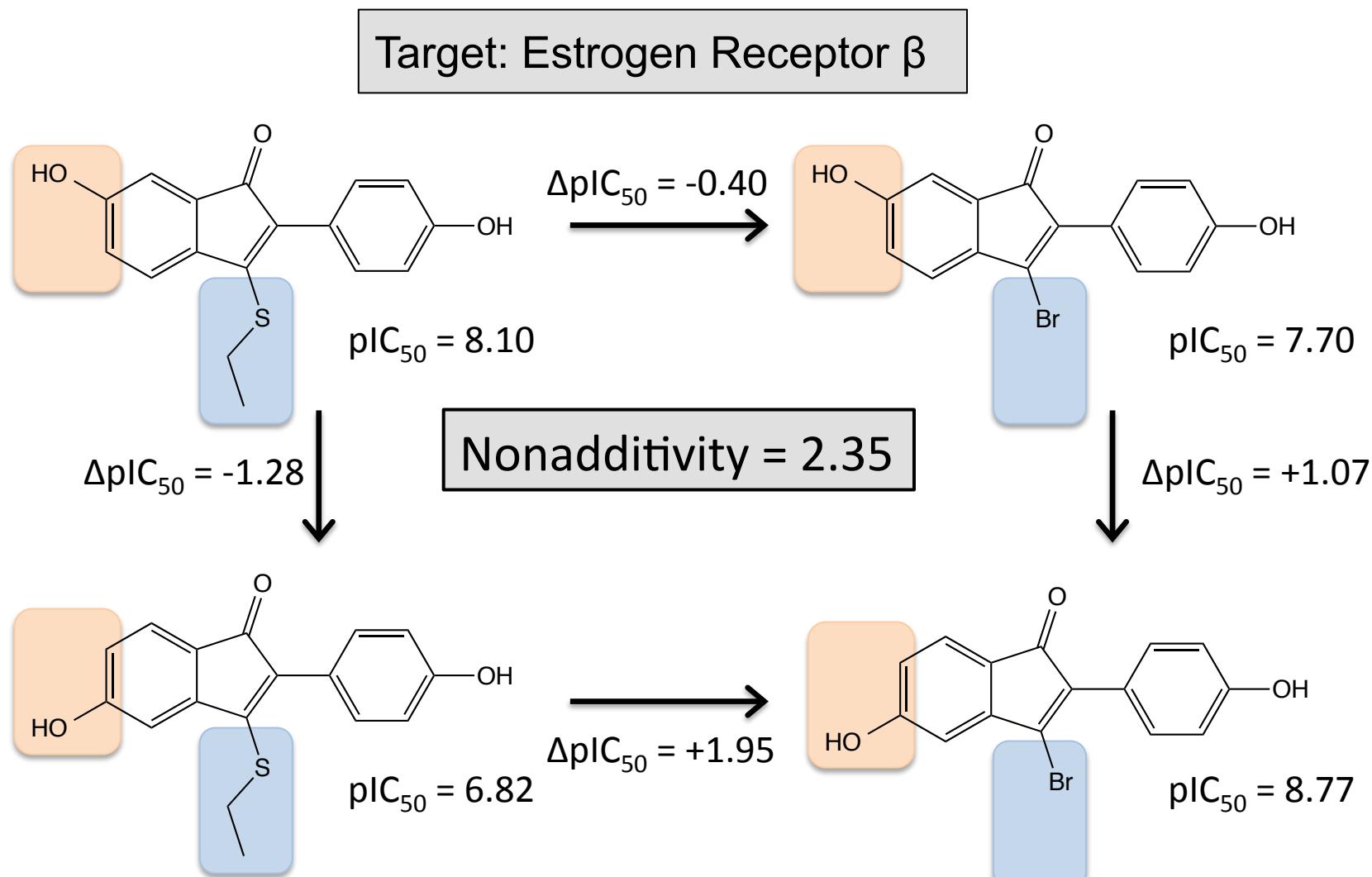
<sup>a</sup> Standard errors were estimated from two independent trials, each performed in triplicate.



**Table 2.** Kinetic constants for the inhibition of COX-1 by flurbiprofen (**1**) and its defluorinated analogue (**2**)

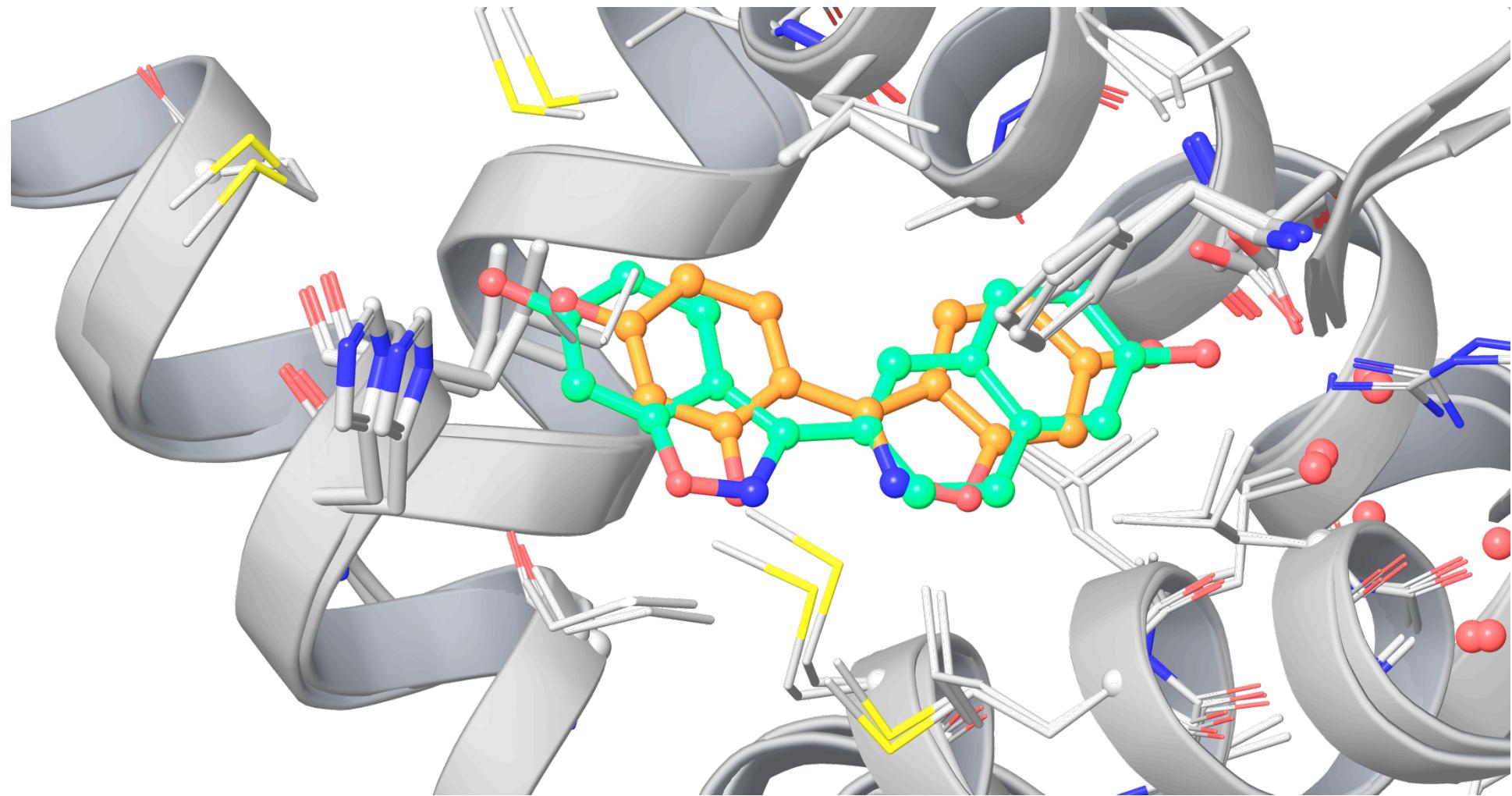
Compd	K <sub>i</sub> (μM) <sup>a</sup>	k <sub>inact</sub> (min <sup>-1</sup> )
<b>1</b>	1.0±0.5	1.0±0.4
<b>2</b>	0.13±0.03	0.19±0.02

## Example: Conformational Reorientation



Christian Kramer

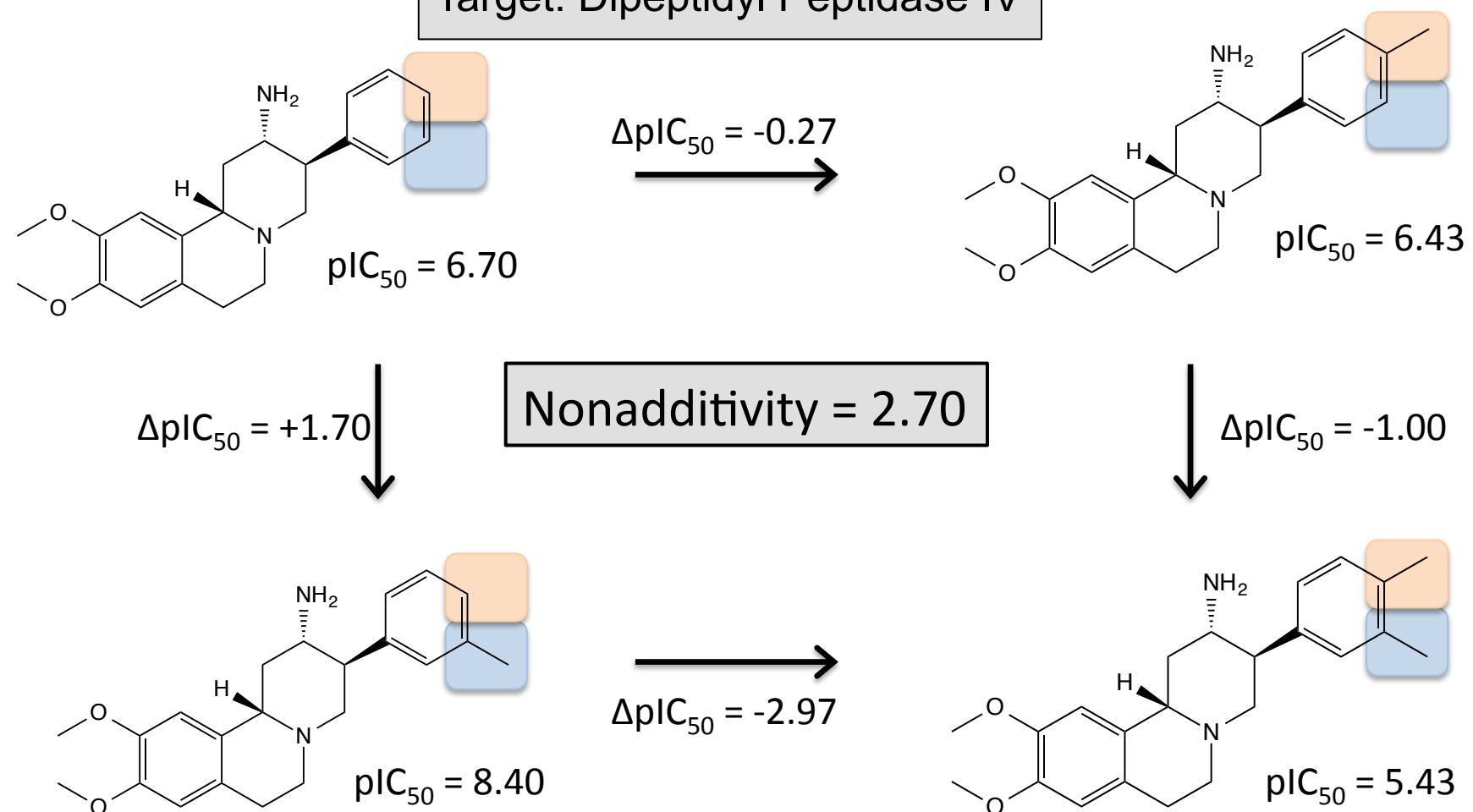
# Nonadditivity: Conformational Reorientation of Benzisoxazole derivatives



1U3S: green; 1U3Q: orange

Could you predict this with your docking/ scoring setup?

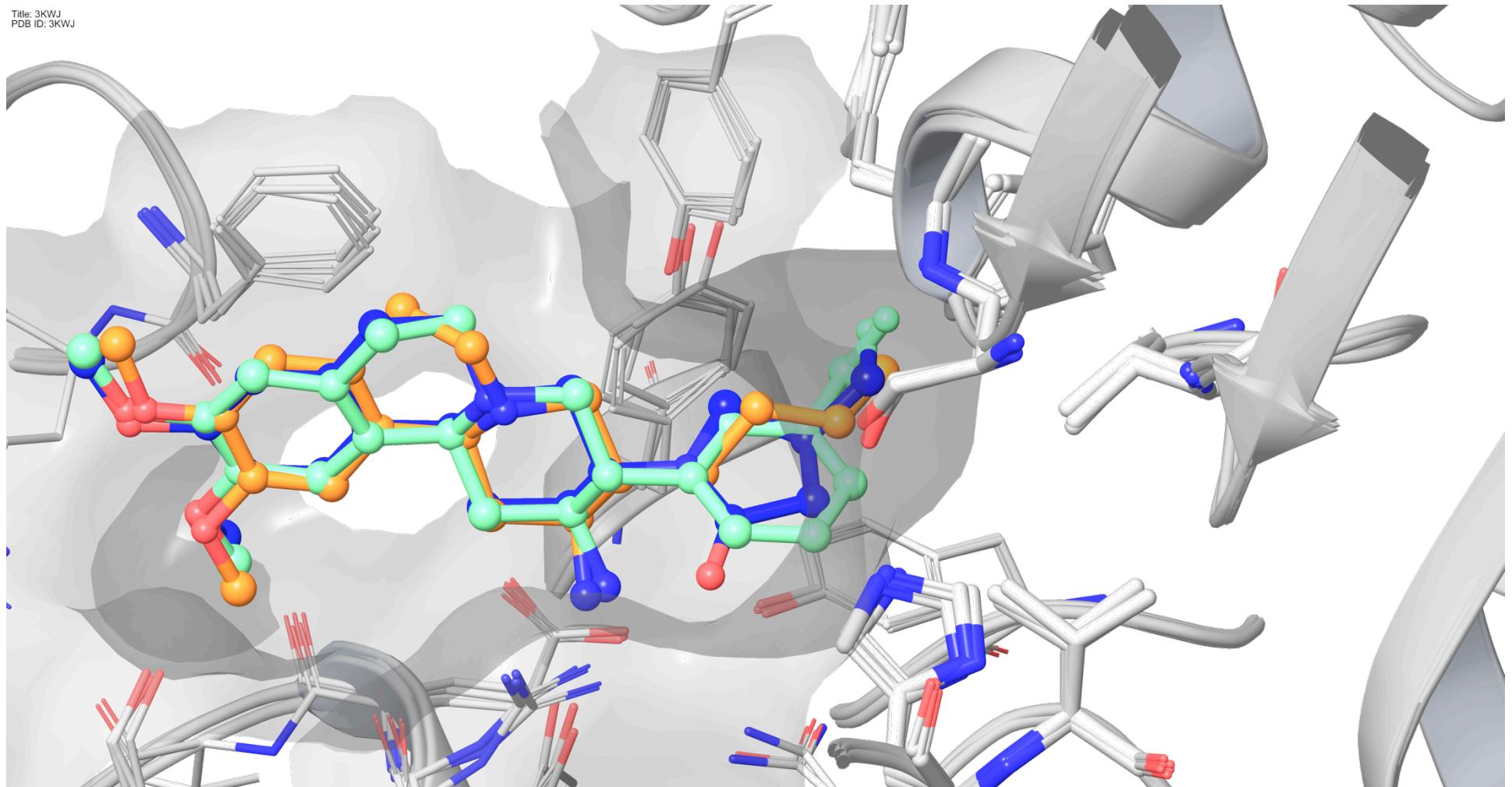
## Example: Substituent Interaction



Christian Kramer

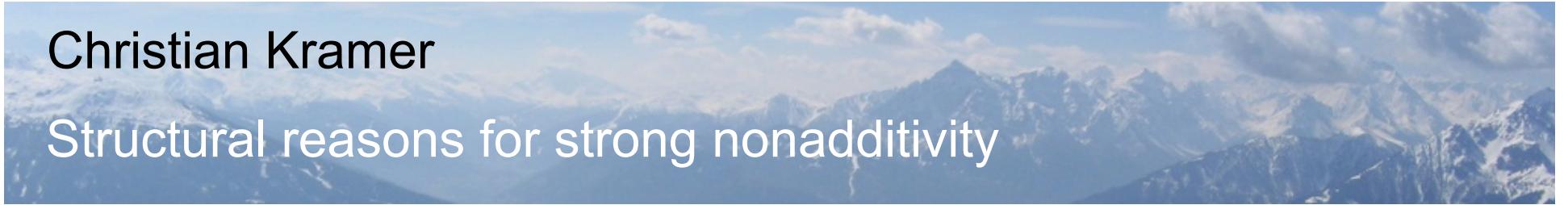
Nonadditivity: Pocket is too small for both Methyl groups

Title: 3KWJ  
PDB ID: 3KWJ



Could you predict this with your FEP setup?

3KWJ: green; 3KWF: blue; 3OC0: orange



Christian Kramer

Structural reasons for strong nonadditivity

### **Complete Rearrangement (Docking/Scoring cases)**

*DHFR, Thrombin, Estrogen β receptor, Factor Xa, LCK Kinase*

### **Substituent Interaction/ Concurrency (FEP cases)**

*CDK2, DPP IV, Thrombin, BACE, Glycogen Synthase Kinase*

### **Non-obvious cases**

*MMP13, CHK1, Glycogen Synthase Kinase, LCK Kinase, PI3 Kinase*

Nonadditivity is a **key SAR feature** that highlights systems where we can **learn about specific protein-ligand interactions**.

**Experimental Uncertainty** can lead to **apparent nonadditivity** and needs to be considered for the nonadditivity analysis.

Nonadditivity analysis provides important hints to **side-chain interactions, conformational clashes, and ligand rearrangements**.

Fully automated (incl. ChEMBL Binding) Nonadditivity Analysis code will be submitted to **RDKit Contrib.** → Code review?

# Liedl Group



Christian Kramer

Thank You

