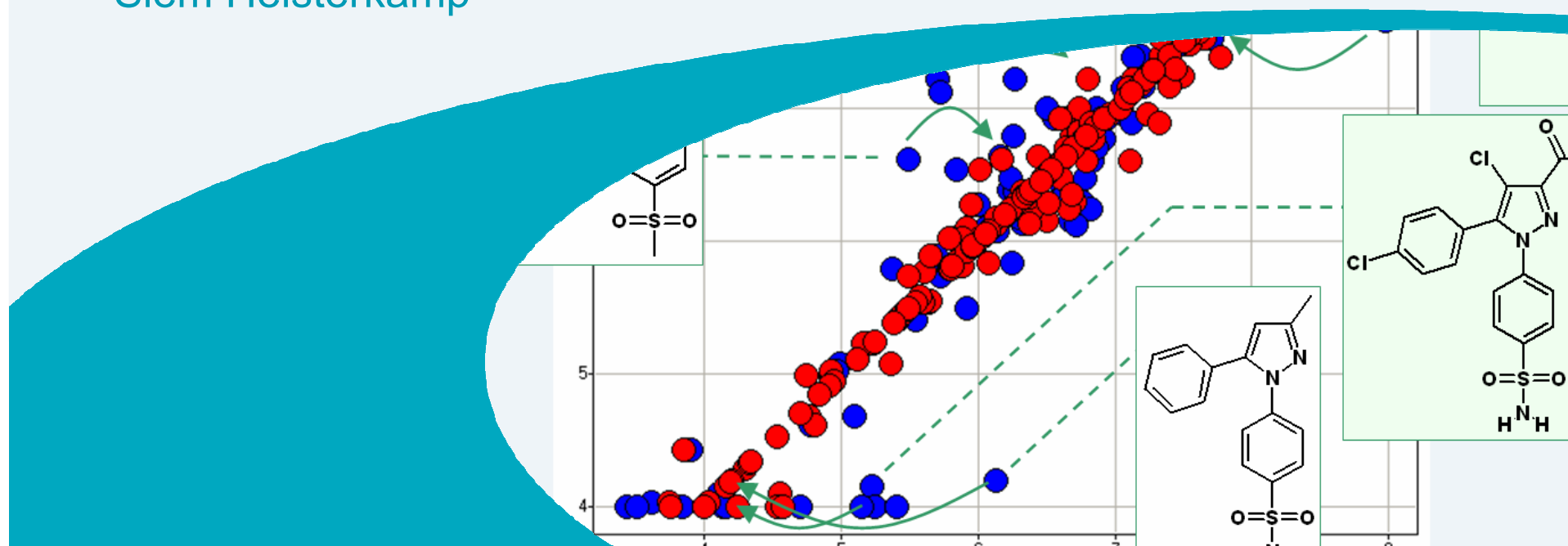


SeeSAR

an automated SAR analysis tool based on pairwise comparisons

Jos Lommerse
Markus Wagener
Siem Heisterkamp



Overview

- Introduction
- Generation of a SAR table
- Pairwise comparisons using interactive web-based tool
- Statistical analysis
- Conclusions



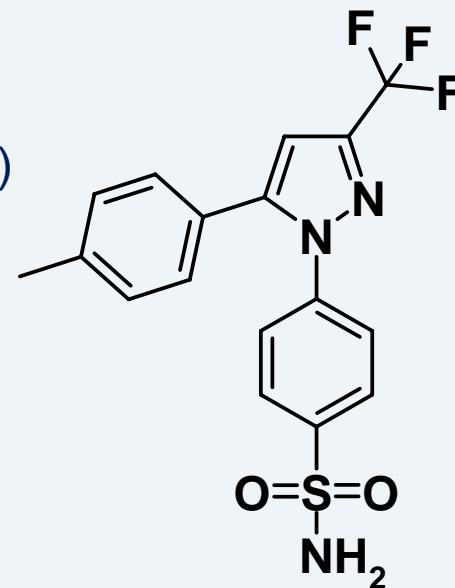
Introduction

How to quickly analyse internal and/or external data sets?

- Think as a medicinal chemist
- Analyse data as a *in silico* modeler

Example from literature data set:

- Cyclooxygenase-2 (COX-2) inhibitors
- Selective non-steroidal anti-inflammatory drug (NSAID)
- Diarylheterocycles
- J. Sutherland et al. J. Med. Chem. 2004, 47, 5541-5554 Chavette et al. J. Med. Chem. 2001, 44, 3223-3230.
- 210 compounds
- pIC50 range from 4 to 8.8

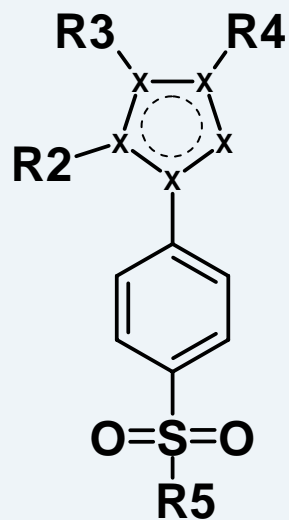


Celecoxib

Generation of a SAR table

Structure data-file

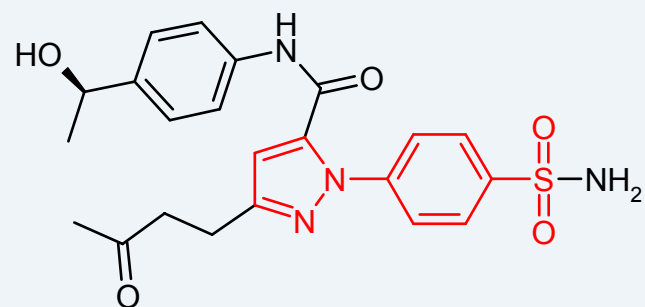
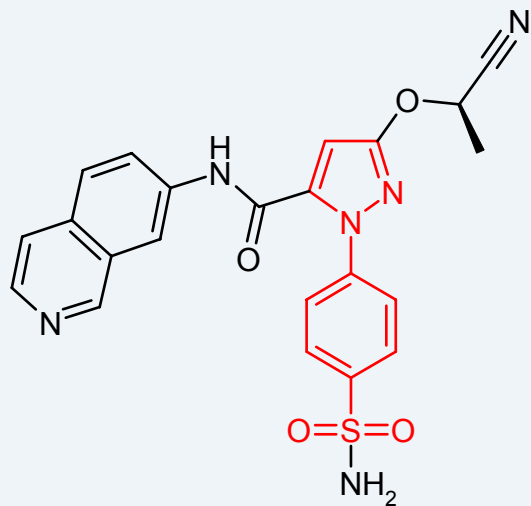
Description of core(s)



Compound	R1	R2	R3	...	Read-out

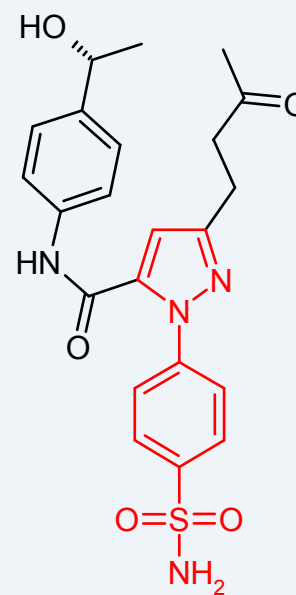
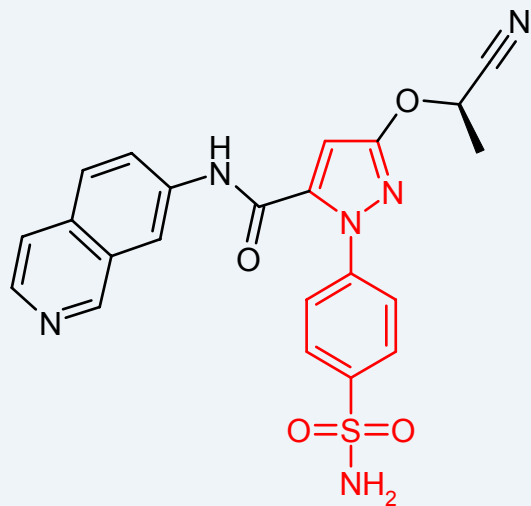
Alignment of Core and R-groups

1. Identification of the core



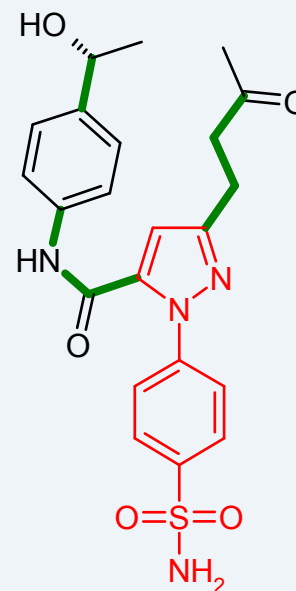
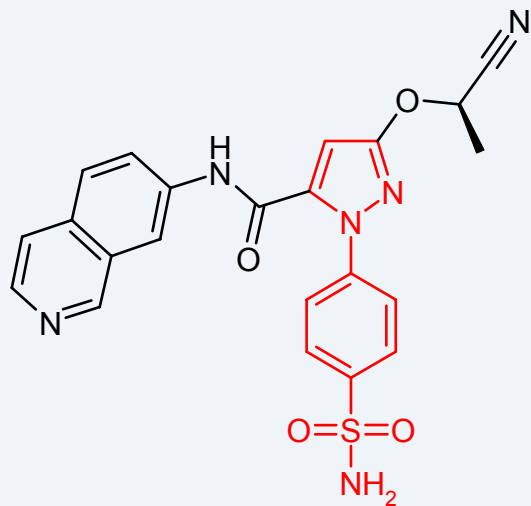
Alignment of Core and R-groups

2. Alignment of core



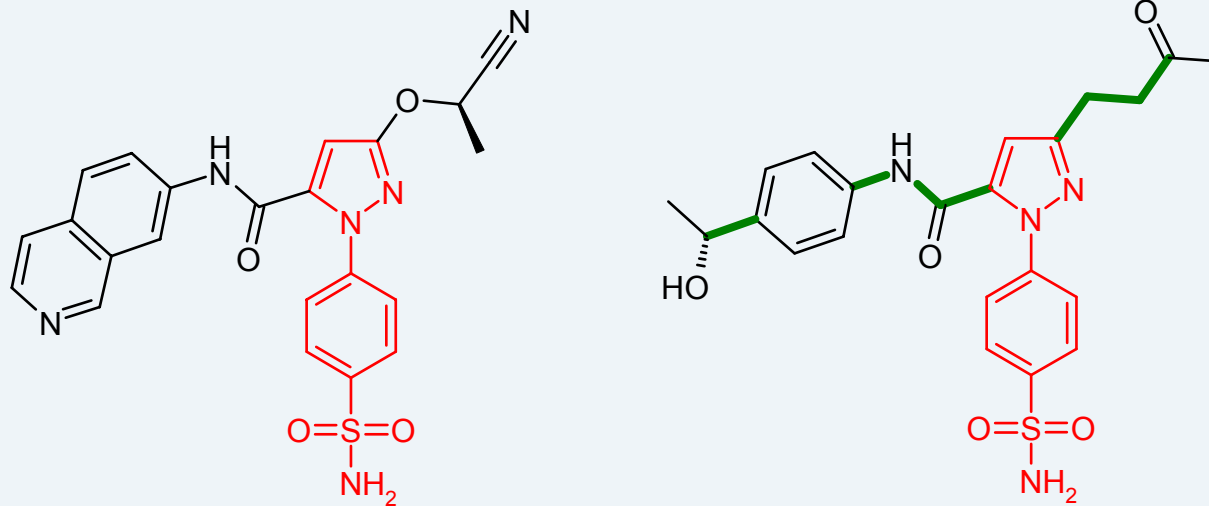
Alignment of Core and R-groups

3. Identification of rotatable bonds



Alignment of Core and R-groups

4. Optimization of overlap



Alignment of Core and R-groups

Scoring function inspired by SEAL:

$$score = \sum_i \sum_j w_{ij} \cdot \exp(-\alpha \cdot r_{ij}^2) - \sum_k \sum_l \frac{1}{r_{kl}^2}$$

α width of Gaussian

r_{ij} distance between atom i in structure 1
and atom j in structure 2

r_{kl} distance between atoms k and l in structure 2

w_{ij} weight for each distance interaction taking into account:

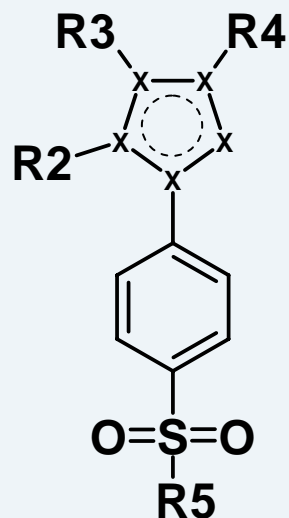
- heteroatoms
- core membership
- distance to core



SK Kearsley, GM Smith,
Tetrahedron Computer Methodology, 1990, 3, 615.

Web-based tool (1): Sar table

Core








X = C or N
R1 = core

Home		SAR R-Group Table					?			
	Struct.	R-1	R-2	R-3	R-4	R-5	Activity	Polar Surface	Molecular Weight	
7							5.79317	40.130001	371.425476	
8							5.41117	44.180000	359.414490	
9							5.84164	44.180000	427.412750	
10				No-Structure			6.61979	37.860001	346.830658	
11				No-Structure			6.95861	37.860001	400.802155	
12				No-Structure			6.92082	37.860001	366.357056	

Organon

Web-based tool (2): Substituent pair comparisons

Home		SeeSAR substituent data				?
		For: R-4 = fe9772d54d391154				
	Struct1.	Compound Pairs	Activity-Ref	Activity-Alt	SAR score (%)	
1 <input checked="" type="checkbox"/>		Reference	-	-	-	
8 <input checked="" type="checkbox"/>		142-64 10-54	4.10237 6.61979	4.02965 5.07831	-100.00	
3 <input checked="" type="checkbox"/>		169-107 10-49	4.20204 6.61979	6.69897 6.38722	0.00	
13 <input checked="" type="checkbox"/>		169-106 142-63 10-48	4.20204 4.10237 6.61979	6.88606 4.68403 6.21467	33.33	
7 <input checked="" type="checkbox"/>		142-55 10-11	4.10237 6.61979	5.77211 6.95861	100.00	

- Ranking of substituents to generate SAR
- Initial Score R_{init} for subst. X with activity A_x :

$$R_{init} = \frac{\sum_{x \neq i} (A_x < A_i) - \sum_{x \neq i} (A_x > A_i)}{\sum_{x \neq i} (A_x > A_i) + \sum_{x \neq i} (A_x = A_i) + \sum_{x \neq i} (A_x < A_i)}$$

















OR

$$R_{init} = \frac{\sum (\text{wins}) - \sum (\text{losses})}{\sum (\text{games})}$$



Web-based tool (3): Matrix of all substituent pairs

SeeSAR 2D-substituent matrix
R-4 Activity
MinDiff=0, MinFact=1

	Home	Overall Effect Corr2											
1 <input checked="" type="checkbox"/>		-100.0	***	-	-100.0 1 0 0 1	-100.0 3 0 0 3	-100.0 4 0 0 4	-	-100.0 2 0 0 2	-	-100.0 1 0 0 1	-	-
2 <input checked="" type="checkbox"/>		-100.0	-	***	-	-100.0 1 0 0 1	-	-100.0 1 0 0 1	-	-	-	-100.0 1 0 0 1	-
3 <input checked="" type="checkbox"/>		-86.2	100.0 0 0 1 1	-	***	100.0 0 0 1 1	-100.0 1 0 0 1	-	0.0 0 1 0 1	-	-100.0 1 0 0 1	-	-
4 <input checked="" type="checkbox"/>		-58.6	100.0 0 0 3 3	100.0 0 0 1 1	-100.0 1 0 0 1	***	0.0 1 0 1 2	100.0 0 0 1 1	-100.0 2 0 0 2	-	-100.0 1 0 0 1	100.0 0 0 1 1	-
5 <input checked="" type="checkbox"/>		-57.9	100.0 0 0 4 4	-	100.0 0 0 1 1	0.0 1 0 1 2	***	-	-33.3 2 0 1 3	-	0.0 1 0 1 2	-	-100.0 1 0 0 1

Web-based tool (4): Pairs of original compounds

Home SeeSAR compound comparisons For: Activity ?










Struct-1	Struct-2	Codes-1	Codes-2
		8.09691	7.82391
		6.35655	5.73755
		5.77211	4.68403
		6.95861	6.21467

Home SeeSAR compound comparisons For: Activity ?

Struct-1	Struct-2	Codes-1	Codes-2
		6.88606	4.20204
		4.68403	4.10237
		6.21467	6.61979

Web-based tool (5): Matrix of all substituent pairs

Home **SeeSAR 2D-substituent matrix**
R-4 Activity
MinDiff=0, MinFact=1

		Overall Effect Corr2				
1 <input checked="" type="checkbox"/>		-100.0	***	-	-100.0 1 0 0 1	-100.0 3 0 0 3
2 <input checked="" type="checkbox"/>		-100.0	-	***	-	-100.0 1 0 0 1
3 <input checked="" type="checkbox"/>		-86.2	100.0 0 0 1 1	-	***	100.0 0 0 1 1
4 <input checked="" type="checkbox"/>		-58.6	100.0 0 0 3 3	100.0 0 0 1 1	-100.0 1 0 0 1	***
5 <input checked="" type="checkbox"/>		-57.9	100.0 0 0 4 4	-	100.0 0 0 1 1	0.0 1 0 1 2

Corrected SeeSAR Score:

$$w_i^+ = (R_{init} + 1) / 2$$

$$w_i^- = 1 - w_i^+$$

$$R = \frac{\sum_{x \neq i} w_i^+ (A_x > A_i) - \sum_{x \neq i} w_i^- (A_x < w_i A_i)}{\sum_{x \neq i} w_i^+ (A_x > A_i) + \sum_{x \neq i} w_i^- (A_x < w_i A_i)}$$



Web-based tool (6): Generation of SAR

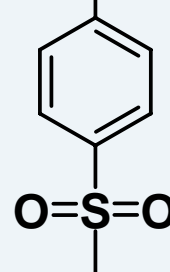
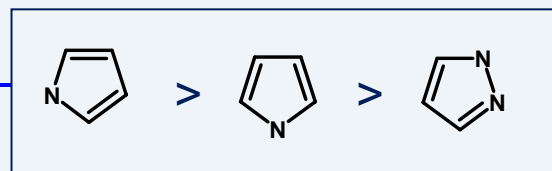
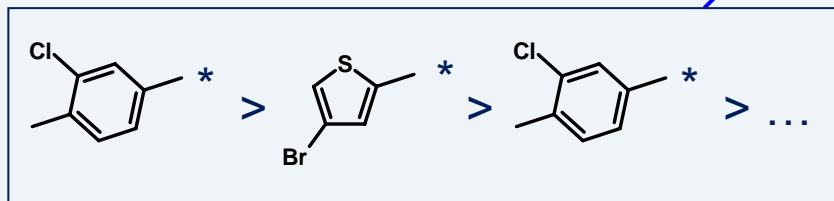
$\text{NO}_2 > \text{Cl} > \text{F} > \text{Me} \dots$

$\text{CF}_3 > \text{Br} > \text{Ph} > \text{CN} \dots$

R3 R4

R1

R2



R5

$\text{NH}_2 > \text{Me} > \text{NMe}_2 \sim \text{NHMe}$

SeeSAR analysis

- Quick overview of (large) data set of structurally related molecules
 - Easy interpretation by Medicinal Chemists (this IS how med chemists work!)
 - Some substituent dependencies can be discovered
 - Ranking substituents generates the SAR
-
- No estimate of importance of each substitution site
 - No estimate of contribution of each individual substituent
 - No interpretation if substituent does not occur in any pair
 - Substituent dependencies are not observed if the needed pair combinations are not available

Development of suitable statistical model to capture all (Q)SAR information



Regression analysis

Main effects model:

$$y = X_{\text{mean}} + X_{R1} + X_{R2} + X_{R3} + \dots + \varepsilon$$

(cf. Free-Wilson analysis)

Accounting for potential substituent dependencies:

$$y = X_{\text{mean}} + X_{R1} + X_{R2} + X_{R3} + \dots + X_{R1} \cdot X_{R2} + \varepsilon \quad (\text{only R1:R2 dependency})$$

$$y = X_{\text{mean}} + X_{R1} + X_{R2} + X_{R3} + \dots + X_{R1} \cdot X_{R3} + \varepsilon \quad (\text{only R1:R3 dependency})$$

etc.

$$y = X_{\text{mean}} + X_{R1} + X_{R2} + X_{R3} + \dots + X_{R1} \cdot X_{R2} + X_{R3} \cdot X_{R4} + \varepsilon$$

(both R1:R2 and R3:R4 dependencies)

etc.

> Select model which fits data best with fewest parameters, e.g. using Akaike Information Criterion (AIC).

AIC for various models

Model	df	AIC
cox2_main	152	463.7111
cox2_inter12	165	433.2666
cox2_inter13	155	416.4300
cox2_inter14	162	426.2110
cox2_inter15	155	463.5676
cox2_inter23	168	400.0482
cox2_inter24	170	427.2762
cox2_inter25	177	457.8914
cox2_inter34	169	169.1309
cox2_inter35	156	439.0895
cox2_inter45	161	446.1146

← Best model

Overall regression parameters for "cox2_inter34"

Response: pActivity

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Subst1	5	11.199	2.240	16.2955	1.149e-08
Subst2	103	201.807	1.959	14.2548	3.275e-15
Subst3	12	28.462	2.372	17.2564	5.282e-12
Subst4	27	55.013	2.038	14.8238	2.943e-13
Subst5	3	17.765	5.922	43.0826	1.868e-12
Subst3:Subst4	17	20.856	1.227	8.9258	1.026e-08
Residuals	39	5.360	0.137		

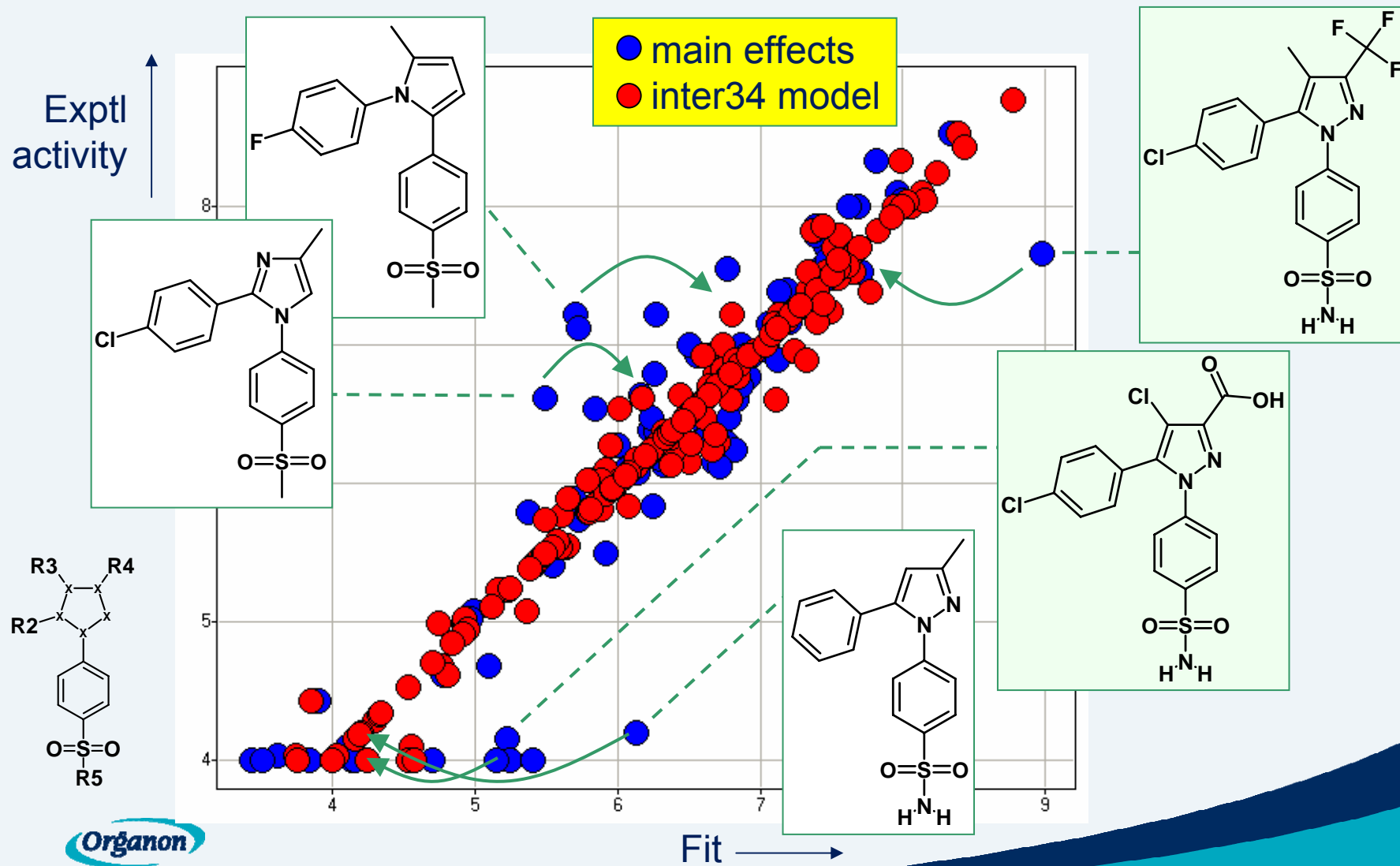
Number of different
substituents

"Average" contribution
of all substituents

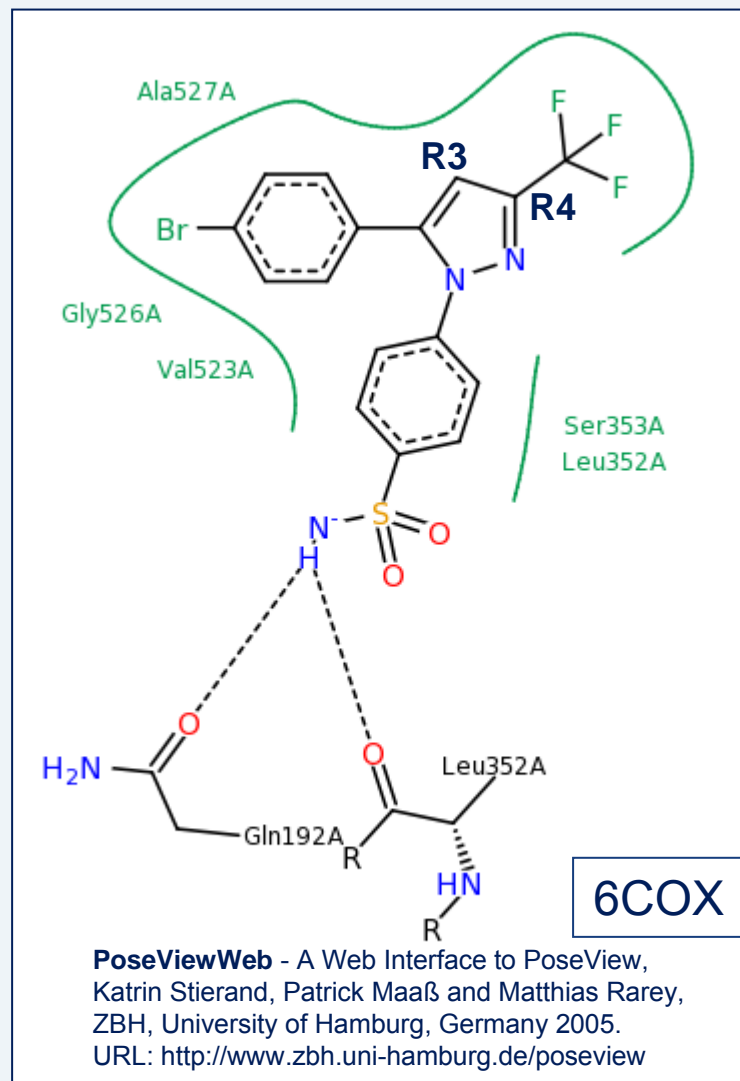
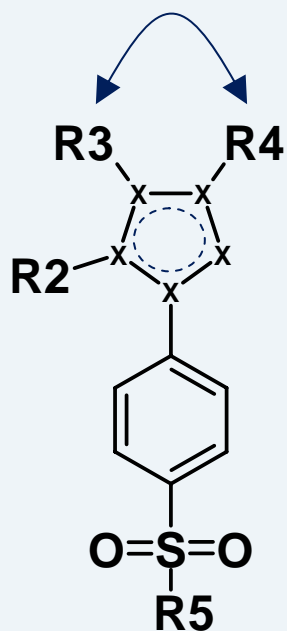
Significance level



Fit of "cox2_main" and "cox2_inter34" models

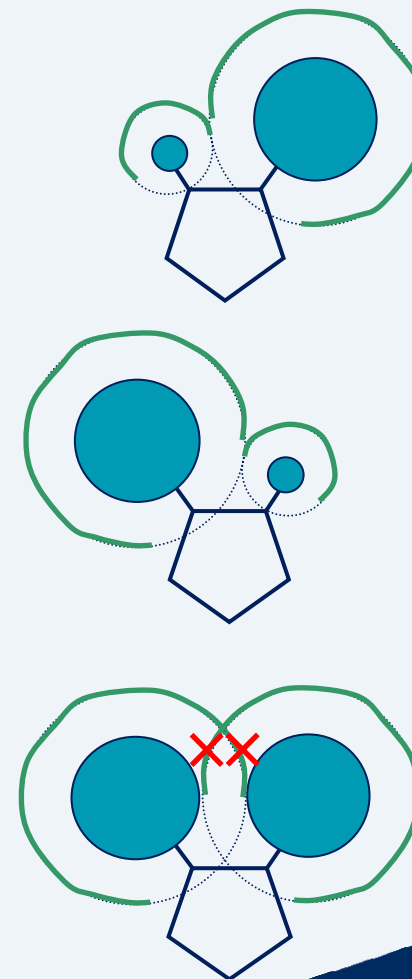


Dependency R3 en R4



PoseViewWeb - A Web Interface to PoseView,
 Katrin Stierand, Patrick Maaß and Matthias Rarey,
 ZBH, University of Hamburg, Germany 2005.
 URL: <http://www.zbh.uni-hamburg.de/poseview>

R3 R4



Conclusions

- SeeSAR: Combination of chemical intuition and model building resulting in more and better information for a given molecular data set
- Automated generation of SAR
- Novel compounds suggestions by combination of "strong" substituents and from insights of substituent dependencies

