

CHEMBIOCHEM

Supporting Information

© Copyright Wiley-VCH Verlag GmbH & Co. KGaA, 69451 Weinheim, 2007

Scaffold-Hopping by “Fuzzy” Pharmacophores and Application to RNA Targets

Yusuf Tanrikulu, Manuel Nietert, Ute Scheffer, Ewgenij Proschak, Kristina Grabowski, Petra Schneider, Markus Weidlich, Michael Karas, Michael Göbel, Gisbert Schneider

Defined by the Medicinal Chemistry Section of IUPAC,^[1] a pharmacophore is the “ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger (or to block) its biological response”. According to this concept, ligand-receptor interactions arise by individual functional group contributions. Since we do not know a priori which functional group actually contributes to the interaction they are termed “potential pharmacophore points” (PPPs). Ligand-receptor interactions take place in 3D space. Therefore 3D pharmacophore models represent the most intuitive choice. Noteworthy, in the absence of a receptor-relevant ligand conformation or conformation ensemble, quantitative structure-activity relationship (QSAR) studies that are based on 3D models can still be erroneous. In addition to the problem of conformer generation, an error-prone step in pharmacophore matching methods is the 3D-alignment of molecular features, that is, matching a screening molecule to a given pharmacophore model. To enable rapid database searching the explicit alignment step can be avoided by an alignment-free representation of pharmacophore patterns. One idea is to convert the spatial distribution of PPPs to a vector representation. Such vectors are referred to as “fingerprints”, “bitstrings”, “correlation vectors” (CV), or “spectra” depending on the type of information stored. The trick is to compare these reduced molecular representations instead of explicit 3D feature alignment thus formulating a pharmacophore search as a similarity search.^[2]

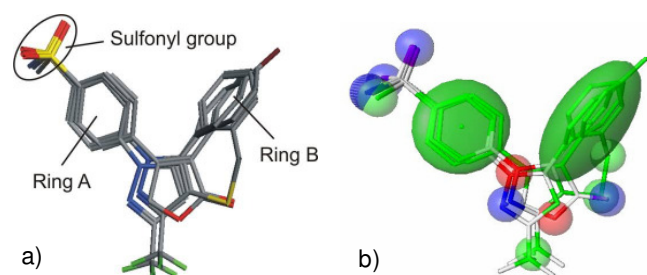


Figure S1. a) Molecular alignment of three COX-2 inhibitors (Rofecoxib, M5, SC-558). The common chemical groups, which are essential interactions for specific COX-2 inhibition, are labeled according to ref. [4]. b) LIQUID Pharmacophore model of a): the visualization shows the ellipsoidal PPP models of both ring systems, which are members of the maximum common substructure of the respective superposition.

In our software LIQUID, PPPs are modeled as trivariate Gaussian distributions and encoded as CV representations. The statistical spread of every PPP reflects the fuzziness of

the pharmacophore model: The probability of a certain interaction decreases with the increasing distance to the PPPs centroid. Three-dimensional visualization tools, e.g. PyMOL^[3], enable us to visually analyze generated pharmacophore models. An example of a LIQUID pharmacophore model is shown in Figure S1, derived from a molecular alignment of the COX-2 inhibitors Rofecoxib, M5, and SC-558.

LIQUID can be used to compute a pharmacophore model of either a single molecular conformation or a conformer ensemble. The first fundamental step is the atom type recognition, where potential pharmacophore features are assigned to the query's atoms. We consider three different interactions types: “lipophilic”, “hydrogen-bond donor” and “hydrogen-bond acceptor”. For example, oxygen atoms are always hydrogen-bond acceptor interaction points, whereas OH groups also possess a hydrogen-bond donor feature. An atom can represent none, one or maximal two (only hydrogen-bond donor + acceptor) of these features. Atoms lacking a pharmacophore feature are not considered for the model. Atom typing results in a transformation of the query molecule(s) into a three-dimensional disposition of interaction points. To gather a fuzzy approximation of this interaction field via Gaussian functions, single interaction points are clustered into PPPs. Every PPP represents a local maximum of the Gaussian distribution of interaction points it contains.

To determine the maxima of the interaction point distribution the cluster radius dependent local-feature-density (LFD) was introduced.^[5] It allows a quantification of common-type atoms in the spatial environment around an atom. Hereby the manually adjustable cluster radius is used to constrain the space around an atom where the maximum is determined. This also enables the user to generate pharmacophore models with varying fuzziness. The fuzzy pharmacophore model of all acetylpromazine docking modes described above was computed with a lipophilic cluster radius of 1.4 Å and 2.0 Å for hydrogen-bond donors and acceptors. The LFD of the k^{th} atom of pharmacophore type T is calculated by Eq. 1.

$$LFD(atom_k^T) = \sum_{i=1}^n \max\left\{0, 1 - \frac{D_2(atom_k^T, atom_i^T)}{r_c}\right\}, \quad (1)$$

where n is the total number of atoms of the pharmacophore type T in the query, D_2 is the Euclidean distance of two atoms, and r_c represents the cluster radius of the specific pharmacophore type. The closer a common-typed atom, the bigger is its impact to the considered atom's LFD. Clustering of the interaction points is done on the basis of the atoms' LFDs via a Union-Find strategy. The following pseudo-code illustrates how the algorithm works:

```

INIT: each atom is a singleton.
FOR each atom  $i$  of type  $T$ 
  FOR each atom  $j$  of type  $T$ 
    calculate Distance( $i, j$ )
    IF Distance  $\leq$  ClusterRadius  $r_c$  THEN
      FIND maxLFD(Cluster $_i$ )
      FIND maxLFD(Cluster $_j$ )
      IF maxLFD(Cluster $_i$ )  $\leq$  maxLFD(Cluster $_j$ ) THEN
        UNION Cluster $_j$  with Cluster $_i$ 

```

Initially, every atom represents a singleton. If a common-typed atom yielding a higher LFD inside the cluster radius of

the considered atom is found, both atoms get “united” into a cluster. Finally, each cluster of interactions points (feature-typed atoms) forms a PPP. One can observe that the number of final clusters depends on the adjusted cluster radius. The centroid of a PPP is calculated as geometric center of its clustered atoms.

We apply the principal component analysis (PCA)^[6] in order to compute the size and orientation of a PPP. The covariance matrix is built up from the Cartesian coordinates of the clustered atoms in relation to the PPPs centroid. The orientation of a PPP is given by the resulting principal components, because their directions span the data space according to the highest variances. Eigenvector approximation is done with the NIPALS algorithm.^[7] The corresponding Eigenvalues provide the distribution of the PPP in the direction of the Eigenvectors.

Having obtained the position, size and orientation of the PPPs, we encode the pharmacophore model as a correlation vector (Eq. 2).^[2,5] LIQUID computes a correlation-vector from the trivariate Gaussian functions, which are used to model the PPPs. Due to pairwise PPP correlation, we encounter six PPP pairs: “lipophilic- lipophilic”, “lipophilic-donor”, “lipophilic-acceptor”, “donor-donor”, “donor-acceptor” and “acceptor-acceptor”. Encoding yields an equally partitioned bin vector. In the case of PPP instances, the correlation vector reduces to scaled occurrence frequencies of atom pairs at distance intervals from 1 to 20 Å.^[2,8] Each bin (vector element) contains a correlated probability, which indicates the presence of a PPP pair at distance d :

$$CV_d^{A,B} = \frac{1}{\# \text{ pairs } (A, B)} \sum_i^A \sum_j^B \frac{1}{2} \{ \text{trivG}(\sigma_{1,2,3})_i \cdot \text{trivG}(\sigma_{1,2,3})_j \}, \quad (2)$$

for all PPPs i of pharmacophore type A and for all PPPs j of type B. trivG gives the trivariate Gaussian with standard deviations $\sigma_{1,2,3}$, and $\# \text{ pairs}(A, B)$ is the number of pairs of PPPs of type A and B. The result is a 120-dimensional correlation vector-based descriptor designed for fast virtual screening. In this study, the Euclidian distance was used to determine the similarity between vector representations of pairs of molecules.

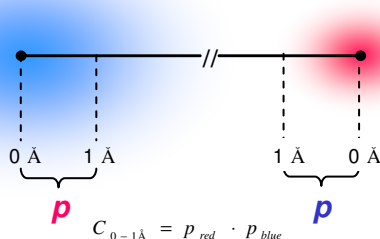


Figure S2. Correlation of probabilities in a certain distance bin

FRET Binding Assay. Fluorescence based binding assays were performed in a microplate reader (Safire²; Tecan; excitation wavelength 489 nm, emission wavelength 590 nm) in 96 well microplates (Corning 6860, black, non binding surface). TAR RNA (Biospring) and Tat peptide (Thermo

Electron Corporation) were both used at concentrations of 10 nM in a final volume of 100 μ L in TK buffer (50 mM Tris-HCl, 20 mM KCl, 0.01% Triton-X 100, pH 7.4) at 37 °C. Prior to titration, the RNA (100 nM in 5mM Tris-HCl, pH 7.4) was heated to 90 °C for 5 minutes and then immediately placed on ice for additional 2-5 minutes. The fluorescence of the blank (TK buffer only), pure peptide and of the Tat-TAR complex was determined first. Single-point measurements of potential inhibitors were carried out in triplicates at 50 or 100 μ M. Figure S3 shows the TAR model and labeled Tat peptide.

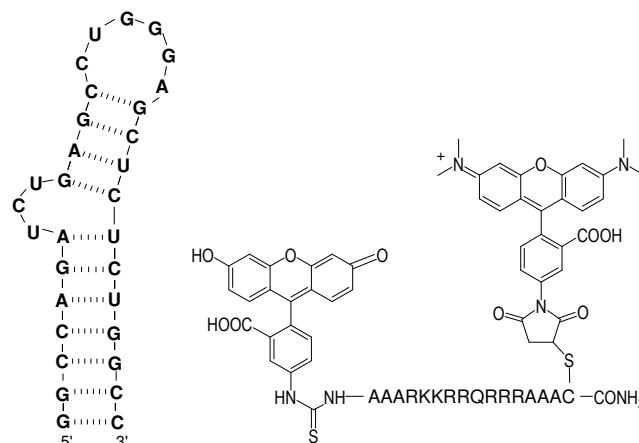


Figure S3. Structures of the HIV-1 TAR model and Tat peptide labeled with fluoresceine and rhodamine.

The values delivered by the Safire² spectrometer were transformed to obtain relative fluorescence intensity according to equation (3):

$$x_{ij}^* = \frac{x_{ik} - x_{k, \min}}{x_{k, \max} - x_{k, \min}}, \text{ where} \quad (3)$$

x_{ij}^* = scaled fluorescence intensity in [0,1]

x_{ik} = raw data value

$x_{k, \min}$ = minimum of the measured fluorescence activity (max. quenching value; max. of free peptide)

$x_{k, \max}$ = maximum of the measured fluorescence activity (max. fluorescence value; max. of bound peptide).

Original data from the FRET assay are available from file “Tanrikulu_FRET_primary_data.xls”.

Cell-free Transcription/Translation Assay. S30-extract of *E. coli* A19 was prepared as described previously.^[9] The reaction volume of 25 μ l contained 3.5 mM Tris-acetate (pH 8.2), 220 mM potassium acetate, 13 mM magnesium acetate, 100 mM HEPES-KOH (pH 8.0), 0.7 mM EDTA, 0.2 mM folic acid, 2 mM DTT, 2 % polyethylene glycol 8000, 1.2 mM ATP, 0.8 mM of the NTPs, 20 mM acetyl phosphate (lithium potassium salt), 20 mM phosphoenolpyruvate (trisodium salt), and 0.05 % sodium azide (components purchased from Sigma-Aldrich Chemie GmbH, Schnelldorf, Germany). 500 mg/L of *E. coli* tRNA mixture, 40 mg/L pyruvate kinase, 1 tablet/10 ml complete protease inhibitor (all Roche Diagnostics, Mannheim, Germany), 0.3 U/ μ l RNAsin RNase inhibitor (Promega GmbH, Mannheim, Germany), 3 U/ μ l T7-RNA-Polymerase (GE Healthcare Europe GmbH, Munich, Germany), and complete amino acid mixture (final concentration 0.75 – 1mM). 40–60 ng/ μ l of the plasmid pVEX2.3-GFP encoding GFP under the control of a T7-promoter and 30 % *E. coli* A19 S30 extract

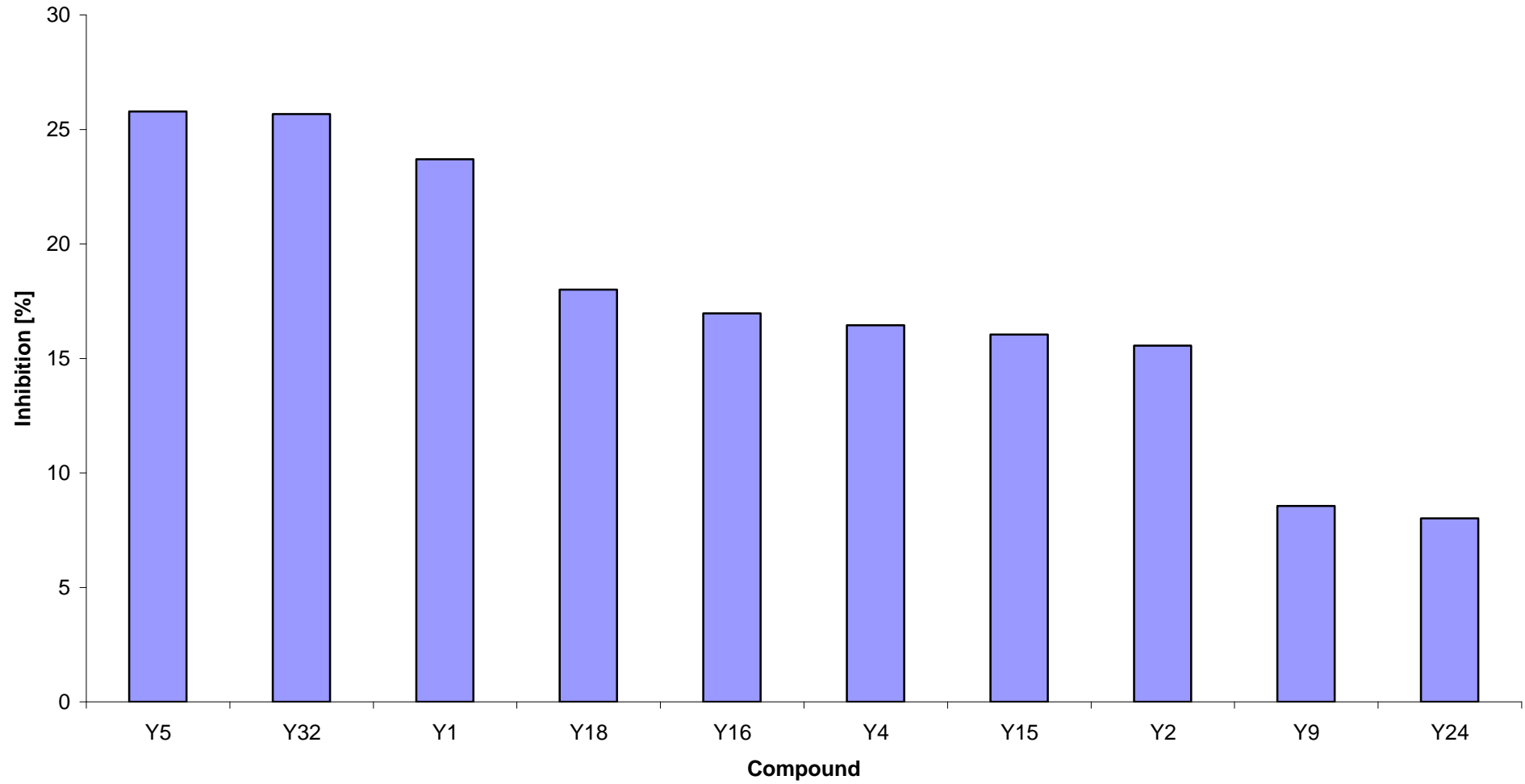
were added. Compounds were dissolved with 2% (v/v) dimethyl sulfoxide in the reaction mixture. Reaction mixtures were complemented with 100 μ M of potential inhibitor. Negative inhibition control reactions were complemented with water. CFTT-reactions were run for 5 h at 30°C. GFP was quantified (1:100 dilutions) by measuring the fluorescence with excitation/emission wavelengths of 395 and 509 nm, respectively using a Hitachi F-4500 Fluorescence Spectrophotometer. Single-point measurements of potential inhibitors were carried out in triplicates.

Original data from the cell-free transcription/translation assay are available from file "Tanrikulu_ivTT_primary_data.xls".

References

- [1] C.G. Wermuth, C.R. Gannelin, P. Lindberg, L.A. Mitschler, *Pure & Appl. Chem.* **1998**, 70, 1129-1143.
- [2] S. Renner, U. Fechner, G. Schneider, In: *Pharmacophores and Pharmacophore Searches* (T. Langer and E. Hoffmann), Wiley-VCH, Weinheim, **2005**, pp. 49-79.
- [3] W.L. DeLano, *The PyMOL Molecular Graphics System*, DeLano Scientific, San Carlos, CA, USA, **2002**.
- [4] A. Palomer, F. Cabre, J. Pascual, J. Campos, M. Trujillo, A. Entrena, M. Gallo, L. Garcia, D. Mauleon, A. Espinosa, *J. Med. Chem.* **2002**, 45, 1402-1411.
- [5] S. Renner, G. Schneider, *J. Med. Chem.* **2004**, 47, 4653-4664.
- [6] E. Jackson, *A User's Guide to Principal Components*, Wiley, New York, **1991**.
- [7] S. Wold, *Chemica Scripta* **1974**, 5, 97-106.
- [8] R.E. Carhart, D.H. Smith, R. Venkataraghavan R, *J. Chem. Inf. Comput. Sci.* **1985**, 25, 64-73.
- [9] C. Klammt *et al.* *Eur. J. Biochem.* **2004**, 271, 568-580.

in vitro TT inhibition assay

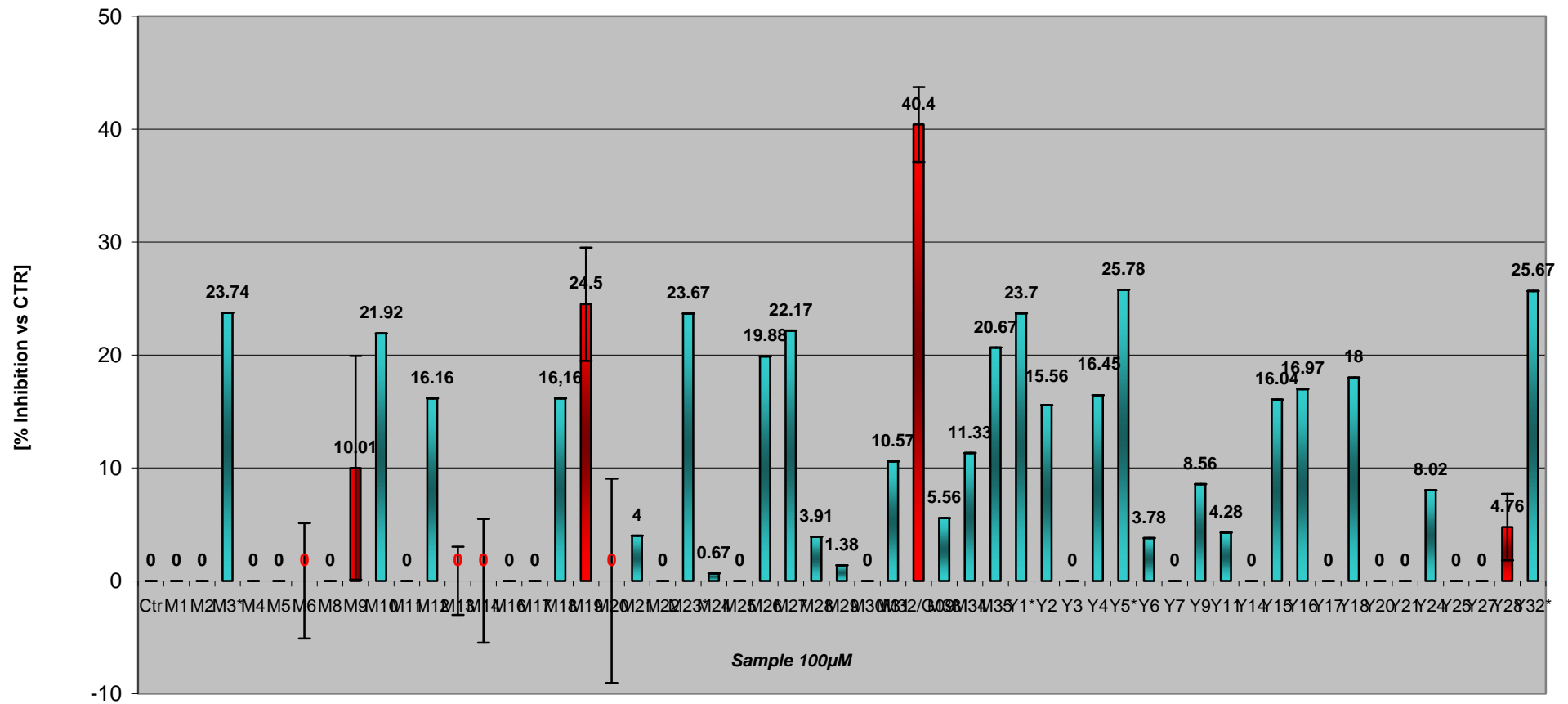


Y5	25.78
Y32	25.67
Y1	23.7
Y18	18
Y16	16.97
Y4	16.45
Y15	16.04
Y2	15.56
Y9	8.56
Y24	8.02

Manuel/Yusuf SPECS

		TT [%]	Hemmung [%]	STDEV/2
3fach	M32/G09	59.6	40.4	3.32
	Y5	74.22	25.78	0
	Y32	74.33	25.67	0
3fach	M19	75.5	24.5	5.02
	M3	76.26	23.74	0
	Y1	76.3	23.7	0
	M23	76.33	23.67	0
	M27	77.83	22.17	0
	M10	78.08	21.92	0
	M35	79.33	20.67	0
	M26	80.12	19.88	0
	Y18	82	18	0
	Y16	83.03	16.97	0
	Y4	83.55	16.45	0
	M12	83.84	16.16	0
	M18	83.84	16.16	0
	Y15	83.96	16.04	0
	Y2	84.44	15.56	0
	M34	88.67	11.33	0
	M31	89.43	10.57	0
3fach	M9	89.99	10.01	9.91
	Y9	91.44	8.56	0
	Y24	91.98	8.02	0
	M33	94.44	5.56	0
3fach	Y28	95.24	4.76	2.94
	Y11	95.72	4.28	0
	M21	96	4	0
	M28	96.09	3.91	0
	Y6	96.22	3.78	0
	M29	98.62	1.38	0
	M24	99.33	0.67	0
	Kontrolle	100	0	0
	M4	100.34	0	0
	M5	92.517	0	0
3fach	M6	142.32	0	5.11
	M8	117.35	0	0
	M11	153.5	0	0
3fach	M13	116.39	0	3.03
3fach	M14	103	0	5.49
	M16	113.64	0	0
	M17	133.3	0	0
3fach	M20	121.51	0	9.06
	M22	127.06	0	0
	M25	129.8	0	0
	M30	114.23	0	0
	Y3	136.4	0	0
	Y7	120.44	0	0
	Y14	120.8	0	0
	Y17	100	0	0
	Y20	135.78	0	0
	Y21	172.02	0	0
	Y25	103.67	0	0
	Y27	187.16	0	0

AK Schneider; Transcription/translation inhibition in-vitro vs control (September 2006)
 (red=triplicate experiment)



Overview

Laborindex	Weight	Vial Amount	Molmengen	mmol	10mM	μL in Epi
3	479.58	2.9	0.00604701	6.047008191	0.60470082	605
4	420.54	2.8	0.00665815	6.658153742	0.66581537	666
5	536.61	3.2	0.00596338	5.963384817	0.59633848	596
6	522.63	2.9	0.00554891	5.548911744	0.55489117	555
7	559.73	3.4	0.00607431	6.074313871	0.60743139	607
8	415.56	2.4	0.00577533	5.775325403	0.57753254	578
9	402.52	2.6	0.00645939	6.459386607	0.64593866	646
10	440.96	2.5	0.00566951	5.669512762	0.56695128	567
11	410.94	2.7	0.00657029	6.570286245	0.65702862	657
12	404.49	2.6	0.00642778	6.427783848	0.64277838	643
13	420.55	2.2	0.00523121	5.231207482	0.52312075	523
14	406.53	2.5	0.00614967	6.149668164	0.61496682	615
15	410.94	2.1	0.00511022	5.110222635	0.51102226	511

11	12
...	...
...	...
...	...
...	...
...	...

11	12
...	...
...	...
...	...
<hr/>	
...	...
...	...
...	...

11	12
...	...
...	...
...	...
...	...
...	
...	...
...	...

11	12
----	----

...

...

...

...

...

...

[Redacted]	
------------	--

...

...

...

...

11	12
...	...
28359	27082
10	10
...	...
...	...
...	...
...	...
...	...

11	12
...	...
...	...
...	...
...	...
[Yellow bar]	

SAFIRE II; Serial number: 12904200051; Firmware: V 1.35 08/2005 Safire2; XFLUOR4SAFIREII Version: V 4.62n

Date: 9.11.05

Time: 20:43

Measurement mode: Fluorescence
Excitation wavelength: 544 nm
Emission wavelength: 590 nm
Excitation bandwidth: 20 nm
Emission bandwidth: 20 nm
Gain (Manual): 122
Number of reads: 10
FlashMode: High sensitivity
Integration time: 100 µs
Lag time: 0 µs
Plate definition file: COS96fb.pdf
Part of the plate: E1 - E12
Z-Position (Manual): 7011 µm
Time between move and flash: 3 ms
Shake duration (Orbital Medium): 30 s
Shake settle time: 10 s
Target Temperature: 37 °C
Current Temperature: 27.9 °C

Rawdata (RFU) Temperature: 27.9 °C

<>	1	2	3	4	5	6	7	8	9	10
A
B
C
D
E	26729	27084	26426							
F	12	12	12							
G	50µM	50µM	50µM							
H							

11	12
...	...
...	...
...	...
28508	25547
12	12
50μM	50μM
...	...
...	...

11	12
...	...
...	...
...	...
...	...

SAFIRE II; Serial number: 12904200051; Firmware: V 1.35 08/2005 Safire2; XFLUOR4SAFIREII Version: V 4.62n

Date: 9.11.05

Time: 20:43

Measurement mode: Fluorescence
Excitation wavelength: 544 nm
Emission wavelength: 590 nm
Excitation bandwidth: 20 nm
Emission bandwidth: 20 nm
Gain (Manual): 122
Number of reads: 10
FlashMode: High sensitivity
Integration time: 100 µs
Lag time: 0 µs
Plate definition file: COS96fb.pdf
Part of the plate: E1 - E12
Z-Position (Manual): 7011 µm
Time between move and flash: 3 ms
Shake duration (Orbital Medium): 30 s
Shake settle time: 10 s
Target Temperature: 37 °C
Current Temperature: 27.9 °C

Rawdata (RFU)

Temperature:

27.9 °C

	1	2	3	4	5	6	7	8	9	10
<>
A
B
C
D
E										31937
F										13
G										50µM
H										sofort pellet

11	12
...	...
...	...
...	...
...	...
31902	30179
13	13
50µM	50µM
sofort pellet	sofort pellet

SAFIRE II; Serial number: 12904200051; Firmware: V 1.35 08/2005 Safire2; XFLUOR4SAFIREII Version: V 4.62n

Date: 9.11.05

Time: 20:43

Measurement mode: Fluorescence
Excitation wavelength: 544 nm
Emission wavelength: 590 nm
Excitation bandwidth: 20 nm
Emission bandwidth: 20 nm
Gain (Manual): 122
Number of reads: 10
FlashMode: High sensitivity
Integration time: 100 μ s
Lag time: 0 μ s
Plate definition file: COS96fb.pdf
Part of the plate: E1 - E12
Z-Position (Manual): 7011 μ m
Time between move and flash: 3 ms
Shake duration (Orbital Medium): 30 s
Shake settle time: 10 s
Target Temperature: 37 $^{\circ}$ C
Current Temperature: 27.9 $^{\circ}$ C

Rawdata (RFU)

Temperature:

27.9 $^{\circ}$ C

	1	2	3	4	5	6	7	8	9	10
<>
A
B
C
D
E	...			27201	27146	28045
F	...			14	14	14
G	...			50 μ M	50 μ M	50 μ M
H

11	12
...	...
...	...
...	...
...	...
[Yellow bar]	

11	12
...	...
...	...
27844	31452
15	15
Brocken	Brocken
drin	drin
50µM	50µM
...	...