## Angewandte Chemie

## Supporting Information

A False-Positive Screening Hit in Fragment-Based Lead Discovery: Watch out for the Red Herring<br>Jonathan Cramer ${ }^{+}$, Johannes Schiebel ${ }^{+}$, Tobias Wulsdorf, Kristof Grohe, Eszter Eva Najbauer, Frederik R. Ehrmann, Nedyalka Radeva, Nina Zitzer, Uwe Linne, Rasmus Linser, Andreas Heine, and Gerhard Klebe*

anie_201609824_sm_miscellaneous_information.pdf

## Table of contents

1. Supporting figures ..... 2
2. Supporting tables ..... 11
3. Experimental section ..... 12
3.1. General experimental information ..... 12
3.2. Ageing of compound $\mathbf{1}$ ..... 12
3.3. Photochemical activation of $\mathbf{1}$ ..... 13
3.4. Conversion of compound $\mathbf{5}$ to compound $\mathbf{4}$ ..... 13
3.5. Tandem MS analysis of compounds 2, 5, 4 and $\mathbf{6}$ ..... 13
3.6. Quantum chemical calculations ..... 15
3.7. NOESY and HSQC NMR experiments in the presence of endothiapepsin ..... 16
3.8. Crystallization, soaking and X-ray diffraction data collection ..... 17
3.9. Structure determination ..... 17
3.10. Estimation of standard deviations for different geometric parameters ..... 18
3.11. Queries in the Cambridge Structural Database (CSD) ..... 18
4. Experimental data ..... 19
Compound 1 ..... 19
Compound 2 ..... 19
Compound $\mathbf{3}$ (TFA salt) ..... 19
Compound 4 ..... 19
Compound 5 (TFA salt) ..... 19
Compound 6 (TFA salt) ..... 19
Compound 8 ..... 19
Compound 9 (TFA salt) ..... 19
5. Data from quantum chemical calculations ..... 20
6. Data from NMR, MS, and HPLC experiments ..... 27
6.1. NMR spectroscopy ..... 27
6.2. Mass spectrometry ..... 33
7. Supporting references ..... 49

## 1. Supporting figures



B


Figure S1: Monitoring the reaction of 1 by HPLC and NMR spectroscopy. (A) HPLC chromatogram of a sample of 1 after incubation for $0,24,72$, and 96 h at $50^{\circ} \mathrm{C}$ in methanol. (B) Time-resolved ${ }^{1} \mathrm{H}-\mathrm{NMR}$ spectrum of 1 in the presence of EP. Over the course of 12 h , spectra were recorded at 1 h intervals. For clarity, the figure only displays spectra every 3 hours, namely after 0 (blue), 3 (red), 6 (green), and 9 hours of incubation (violet). The major peaks are derived from 1, whereas additional peaks come up over time.




3


4

Figure S2: EP in complex with reaction products of $\mathbf{1}$. The $\mathrm{mF}_{\mathrm{o}}-\mathrm{DF}_{\mathrm{c}}$ electron density maps for bound ligands are depicted as gray meshes at the $3 \sigma$ level prior to the inclusion of each respective ligand into the model. (A) Interaction between EP and molecule 3. The depicted structure EP-3 is the result from a diffraction experiment with a crystal soaked with the HPLC isolate $\mathrm{m} / \mathrm{z}=192.3$. Electron density has been observed for four copies of compound $\mathbf{3}$ stacking on top of Phe291 in a regular arrangement. The density was sufficient to model the two molecules closest to Phe291 (the second molecule has only been partially modeled). The TFA molecule shown in two alternative conformations in orange stems from the HPLC purification and binds in a small pocket on the surface of EP. (B) Electron density for molecule 4 in the EP binding pocket after soaking of $\mathbf{5}$, the precursor of $\mathbf{4}$. While the isolated molecule $\mathbf{4}$ itself did not bind with sufficient occupancy to the enzyme to make it visible in the electron density, the presence of $\mathbf{4}$ was indicated by soaking experiments with the better water-soluble compound $\mathbf{5}$ that is easily converted into $\mathbf{4}$ (Scheme 1 in the main manuscript). For better orientation, the picture contains molecule $\mathbf{2}$ as a stick model in gray, derived from a superimposition of the EP-1-2 onto the EP-4 structure (gray surface representation). The bicyclic component of 2 that differentiates it from molecule $\mathbf{4}$ is depicted in transparent colors. Presumably due to its poor aqueous solubility, $\mathbf{4}$ does not sufficiently occupy the EP binding site to allow modelling it in the EP-4 structure. However, for several reasons, we are very confident that $\mathbf{4}$ binds to EP in a similar manner as the corresponding part of $\mathbf{2}$. In particular, the $\mathrm{mF}_{\mathrm{o}}-\mathrm{DF}_{\mathrm{c}}$ (gray mesh) electron density map indicates the presence of the substituted pyrrole ring of 4 , which is further supported by the observation of the alternative conformations of several residues arising from the induced-fit binding discussed in Fig. 4B. Presumably, the simultaneous presence of apo and 4 -bound EP molecules in the crystal also results in electron density next to the aldehyde group of $\mathbf{4}$ originating from a water molecule of the apo protein, which is represented by the red sphere derived from a superimposition of the apo structure (PDB-code 4Y5L) onto the EP-4 structure. For clarity, the $\mathrm{mF}_{0}-\mathrm{DF}_{\mathrm{c}}$ map is only shown around the pyrrole ring. Moreover, an isomorphous difference map (depicted in green) between the EP-4 and a ligand-free EP structure of an isomorphous crystal (PDB-code 5P71) provides some evidence for the presence of the tricyclic core of $\mathbf{4}$ (gray sticks). In contrast, no density could be identified for the atoms of the bicyclus additionally present in $\mathbf{2}$ (transparent gray sticks).


1


4


Figure S3: ${ }^{1} \mathrm{H}-{ }^{13} \mathrm{C}-\mathrm{HSQC}$ NMR spectrum collected from an EP sample incubated with 1 after an incubation time of three weeks at $7{ }^{\circ} \mathrm{C}$. Cross-peaks corresponding to $\mathbf{1}$ and $\mathbf{4}$ are annotated in the spectrum. Additionally observed signals originate from an unknown molecule, presumably generated from hydrolysis of the 3 -chloropyridazine moiety in $\mathbf{1}$. An aldehyde function is clearly present in $\mathbf{4}$ as indicated by a ${ }^{1} \mathrm{H}$ chemical shift of 9.53 ppm and a ${ }^{13} \mathrm{C}$ chemical shift of 186.6 ppm .


Figure S4: Comparison of geometric parameters for the proposed methylene bridge of molecule 2 derived from the crystallographic experiment (EP-1-2 structure) with results from CSD searches for comparable compounds. (A) Definition of the CSD query and investigated geometric parameters. In order to investigate whether molecule 4 and 1 are fused to compound 2 via a methylene bridge, the geometric parameters highlighted in blue, green, orange, magenta and red derived from the crystallographic experiment have been compared with the distributions of the same parameters resulting from a CSD query on comparable molecules as defined in the gray box. In the following panels, the mean values and standard deviations of the individual histogram peaks are given on top of each plot and compared to the crystallographically identified value for $\mathbf{2}$ including estimated standard deviation (purple, according to the SHELXL refinement performed as described in the experimental section). These analyses reveal that 2 likely harbors an intact methylene bridge. (B) Distribution of the distances between atoms 2 and 3 as defined in panel A. (C) Histogram for the distance between atoms 3 and 4. (D) Distribution of the methylene bridge angles between atoms 2,3 and 4 . (E) Histogram for the torsion angle between atoms 4, 3,

2 and 1. If less than $180^{\circ}$ rotation of the bond between atoms 4 and 3 in a clockwise fashion (indicated by the magenta arrow in panel A ) is required in order to create the eclipsed conformation with the bond between atoms 2 and 1, the torsion angle is considered positive (else negative) according to IUPAC. (F) Histogram for the torsion angle between atoms $2,3,4$ and 5 . If less than $180^{\circ}$ rotation of the bond between atoms 2 and 3 in a clockwise fashion (indicated by the red arrow in panel $A$ ) is required in order to create the eclipsed conformation with the bond between atoms 4 and 5 , the torsion angle is considered positive (else negative). All histograms have been generated using the statistical framework R.
A



X-ray

B







Figure S5: Comparison of geometric parameters for the proposed aldehyde group of molecule 2 derived from the crystallographic experiment (EP-1-2 structure) with results from CSD queries for comparable compounds. (A) Definition of the CSD queries and investigated geometric parameters. To decide whether 2 contains an aldehyde (blue boxes in panel A and transparent blue bars in panels B-D) or alcohol (red)
functionality, CSD queries were performed with pyrrole-3-carboxaldehydes and pyrrol-3-ylmethanols (left). Since the number of structures containing these motifs in the CSD was rather limited ( $\mathrm{N}=30$ and 4, respectively), similar queries have been performed on much more abundant benzaldehydes and benzylalcohols ( $\mathrm{N}=1204$ and 475, respectively). The three investigated parameters are defined in green, orange and magenta. For instance, the distance between the aldehyde/alcohol carbon and oxygen is highlighted in green. (B) Distribution of aldehyde (blue) vs. alcohol (red) bond lengths (between atoms 3 and 4). The mean values and standard deviations of the individual histogram peaks are given on top of each plot and compared to the crystallographically identified value for 2 including estimated standard deviation (purple, according to the SHELXL refinement performed as described in the experimental section). This comparison clearly indicates that compound 2 contains an aldehyde as supported by similar analyses of the associated angle and dihedral (panels $C$ and D). (C) Histogram for the angle between atoms 2, 3 and 4 (for the definition of these numbers see panel A). (D) Distribution of torsion angle values between atoms $1,2,3$ and 4 . If less than $180^{\circ}$ rotation of the bond between atoms 1 and 2 in a clockwise fashion (indicated by the magenta arrow in panel A) is required in order to create the eclipsed conformation with the bond between atoms 3 and 4, the torsion angle is considered positive (else negative) according to IUPAC. These figures have been prepared using the statistical program R.
A


B


C

D

E


Figure S6: Non-planar pyridazinium and pyridinium cations. (A) Comparison of the non-planarity of the pyridazinium heterocycle of molecule 2 as derived from the crystallographic experiment (EP-1-2 structure) with usual geometric features of pyridazinium ions. A CSD query was set up as defined in the gray box. The below histogram reports the distribution of distances between the carbon atom highlighted in red from the plane that is made up by the residual atoms of the heterocycle (out-of-plane deviation).

The mean values and standard deviation of the single histogram peak is given on top of the plot and compared to the crystallographically identified value including estimated standard deviation (purple, according to the SHELXL refinement performed as described in the experimental section). (B) Comparison of the non-planarity of the pyridazinium heterocycle of molecule 2 with usual geometric features of pyridinium ions. Since only seven pyridazinium moieties are currently present in the CSD fulfilling our search criteria (for details see also the experimental section), we additionally searched for much more frequent pyridinium ions $(\mathrm{N}=3107)$ and plotted the results in the same way as described for panel A. Both analyses underline that these heterocyclic cations are usually planar. However, some exceptions with significantly shifted carbon positions out of the ring plane do exist. We visually analyzed all small molecule crystal structures with $\mathrm{d}>0.08 \AA$ including their packing and found that this deviation can frequently be attributed to unusually twisted and strained molecular structures as found e.g. in helicenes. Two outliers, however, are very interesting with respect to the observed phenomenon (panels C and D). Both are host-guest complexes that bear similarities to the EP-1-2 protein-ligand complex (panel E). All complexes have two partially negatively charged oxygen atoms in close vicinity to the positively charged pyridinium/pyridazinium nitrogen and attached slightly pyramidalized carbon atom. The latter atom is highlighted in red along with the distance from the plane while important intermolecular distances are shown in orange. (C) A host-guest complex containing a significantly non-planar pyridinium ion (CSD-code XUBDAA). ${ }^{[1]}$ (D) A rotaxane containing a slightly distorted pyridinium ion (CSD-code XUXGAZ). ${ }^{[2]}$ (E) The high-resolution EP-1-2 crystal structure containing a non-planar pyridazinium moiety. The environment of the pyridazinium cation is highlighted. All histograms have been generated using the statistical framework R while structural figures have been prepared using Pymol.

## 2. Supporting tables

Table S1: X-ray diffraction data collection and refinement statistics.

| Structure | EP-2 | EP-1-2 | EP-3 | EP-4 |
| :---: | :---: | :---: | :---: | :---: |
| PDB code | 5LWR | 5LWS | 5LWT | 5LWU |
| Data collection and processing ${ }^{\text {a }}$ |  |  |  |  |
| Wavelength ( $\AA$ ) | 0.91841 | 0.91841 | 0.91841 | 0.91841 |
| Beamline | BESSY BL14.2 | BESSY BL14.1 | BESSY BL14.1 | BESSY BL14.1 |
| Detector | MARMOSAIC 225 | PILATUS 6M | PILATUS 6M | PILATUS 6M |
| Space group | $\mathrm{P} 2_{1}$ | $\mathrm{P} 2_{1}$ | $\mathrm{P} 2_{1}$ | $\mathrm{P} 2_{1}$ |
| Cell dimensions |  |  |  |  |
| a, b, c ( $\AA$ ) | 45.2, 72.8, 52.6 | 45.3, 73.1, 53.0 | 45.4, 73.1, 52.7 | 45.3, 73.3, 53.1 |
| $\alpha, \beta, \gamma\left({ }^{\circ}\right)$ | 90.0, 109.2, 90.0 | 90.0, 109.8, 90.0 | 90.0, 109.6, 90.0 | 90.0, 109.9, 90.0 |
| Resolution range ( $\AA$ ) | 42.7-1.25 (1.32-1.25) | 41.2-1.03 (1.09-1.03) | 42.7-1.07 (1.13-1.07) | 42.6-1.11 (1.18-1.11) |
| Wilson B factor ( $\AA^{2}$ ) | 10.5 | 9.3 | 9.0 | 10.3 |
| No. of unique reflections | 87290 (12864) | 151655 (23423) | 142071 (22805) | 128260 (20551) |
| Average redundancy | 4.0 (3.0) | 3.9 (3.8) | 3.7 (3.6) | 3.7 (3.6) |
| $\mathrm{R}_{\text {merge }}$ (\%) | 3.8 (31.8) | 5.5 (51.5) | 5.1 (48.5) | 4.7 (48.6) |
| Completeness (\%) | 98.1 (89.7) | 94.6 (90.6) | 99.5 (98.9) | 99.6 (99.0) |
| $<\mathrm{I} / \sigma(\mathrm{I})>$ | 20.7 (3.5) | 11.8 (2.3) | 13.1 (2.3) | 13.5 (2.3) |
| Refinement |  |  |  |  |
| Resolution range ( $\AA$ ) | 28.1-1.25 | 36.5-1.03 | 29.4-1.07 | 39.8-1.11 |
| No. of reflections (work / free) | 82903 / 4363 | 144059 / 7581 | 134943 / 7103 | 121835 / 6413 |
| $\mathrm{R}_{\text {cryst }}$ (\%) | 11.3 | 11.6 | 12.8 | 12.3 |
| $\mathrm{R}_{\text {free }}(\%)$ | 13.1 | 13.2 | 14.8 | 14.3 |
| No. of refined residues | 330 | 330 | 330 | 330 |
| No. of ligand atoms | 38 | 49 | 26 |  |
| No. of other ligand atoms ${ }^{\text {b }}$ | 44 | 47 | 43 | 42 |
| No. of water molecules | 329 | 350 | 312 | 326 |
| RMSD, bond lengths ( $\AA$ ) | 0.006 | 0.007 | 0.007 | 0.006 |
| RMSD, bond angles ( ${ }^{\circ}$ ) | 0.9 | 1.0 | 1.0 | 1.0 |
| Ramachandran plot (\%) ${ }^{\text {c }}$ |  |  |  |  |
| Most favored | 93.9 | 93.5 | 94.2 | 94.2 |
| Additionally allowed | 6.1 | 6.5 | 5.8 | 5.8 |
| Generously allowed | 0 | 0 | 0 | 0 |
| Disallowed | 0 | 0 | 0 | 0 |
| Average B factors $\left(\AA^{2}\right)^{\text {d }}$ |  |  |  |  |
| All protein atoms | 12.0 | 11.5 | 11.6 | 12.6 |
| Main chain | 11.0 | 10.5 | 10.7 | 11.7 |
| Side chain | 12.8 | 12.5 | 12.5 | 13.4 |
| Ligand atoms | 11.7 | 17.1 | 30.7 |  |
| Other ligand atoms ${ }^{\text {b }}$ | 21.6 | 23.2 | 25.3 | 25.8 |
| Waters | 29.1 | 26.1 | 24.8 | 27.7 |

[^0]
## 3. Experimental section

### 3.1. General experimental information

1-Chloro-5,6,7-trimethyl-6H-pyrrolo[3,4-d]pyridazine (1) was purchased as a hydrochloride salt from Enamine (UKR). ${ }^{1} \mathrm{H}$ and ${ }^{13} \mathrm{C}$ spectra were recorded on a JEOL ECX-400 or Bruker AV II 300 instrument. All chemical shift values are reported in ppm relative to the non-deuterated solvent signal. Trimethylsilyl propanoic acid was used as an external standard for ${ }^{13} \mathrm{C}$ NMR spectra in $\mathrm{D}_{2} \mathrm{O}$. To describe the multiplicity of the signal, the following abbreviations were used: $\mathrm{s}=$ singlet, $\mathrm{q}=$ quartet, $\mathrm{m}=$ multiplet. ESI-MS spectra were recorded on a Q-Trap 2000 system by Applied Biosystems. For high resolution ESI-MS, MS/MS and LC/MS a LTQ-FT Ultra (Thermo Fischer Scientific) or Orbitrap Velos Pro (Thermo Fisher Scientific) mass spectrometer were used. For HPLC chromatography a Shimadzu LC-20 system equipped with a diode array detector was used. Analytic separations were carried out with a MN Nucleodur 100-5 C18ec 4.6 $\times 250 \mathrm{~mm}$ column using a water-acetonitrile gradient with the addition of $0.1 \%$ TFA. For semipreparative separations a Waters XSelect CSH C18 $10 \times 250 \mathrm{~mm}$ column using a wateracetonitrile gradient was used. Preparative separations were carried out on a Varian PrepStar 218 instrument equipped with a MNagel Nucleodur 100-5 C18ec $32 \times 250 \mathrm{~mm}$ column, employing a water-acetonitrile gradient with the addition of $0.1 \%$ TFA.

### 3.2. Ageing of compound $\mathbf{1}$

To initiate the reaction cascade, a sample of $\mathbf{1}$ was placed in a clear glass flask or GC-vial, dissolved in methanol (Fisher, HPLC grade) and stirred at rt or $50^{\circ} \mathrm{C}$. Preliminary experiments were carried out in EP protein buffer ( 100 mM NaOAc pH 4.6 ), in the presence and absence of EP, respectively, as well as in water at various conditions. Since these experiments did not give any additional insight and are comparable to the results in methanol, the details are not reported here. For analytical HPLC and MS experiments, $1-5 \mathrm{mg}$ of $\mathbf{1}$ were dissolved and the resulting solution was used as is. The experiment was also carried out with exclusion of sunlight (brown glass vial, rt, 24 h ) and under protective gas (Argon, rt, 24 h ).

For a semi-preparative separation 20 mg of $\mathbf{1}(\mathrm{HCl}$ salt, 0.09 mmol$)$ were used, the solvent methanol was removed after the reaction ( 24 hours, $50{ }^{\circ} \mathrm{C}$ ) under reduced pressure and the resulting residue was taken up in $1 \mathrm{~mL} \mathrm{MeOH} / \mathrm{H}_{2} \mathrm{O}$ 1:9. The crude product was purified on a semi-preparative HPLC system equipped with a Waters XSelect CSH C18 (10 x 250 mm )
column using a water/acetonitrile gradient. After lyophilization, compound $4(4 \mathrm{mg}, 0.01 \mathrm{mmol}$, $13 \%$ ) was isolated as a light brown solid as the main product of the reaction. $12 \mathrm{mg}(0.05 \mathrm{mmol}$, $57 \%$ ) of the starting material could be recovered.

The preparative separation was carried out with 100 mg of $\mathbf{1}(\mathrm{HCl}$ salt, 0.43 mmol$)$. The solvent methanol was removed after the reaction ( 24 hours, $50^{\circ} \mathrm{C}$ ) under reduced pressure and the resulting residue was taken up in $2 \mathrm{~mL} \mathrm{MeOH} / \mathrm{H}_{2} \mathrm{O}$ 1:9. The crude product was purified on a preparative HPLC system using a water/acetonitrile gradient with the addition of $0.1 \%$ TFA. Fractions that contained a reaction product were collected and lyophilized. Compound 3 (TFA salt, $<1 \mathrm{mg},<3.29 \mu \mathrm{~mol},<0.76 \%$ ) was isolated as a light brown solid. Compound 5 (TFA salt, $31 \mathrm{mg}, 0.07 \mathrm{mmol}, 15 \%$ ) was isolated as a red oil as the main product of the reaction. Compound 6 (TFA salt, $<1 \mathrm{mg},<1.55 \mathrm{mmol},<0.4 \%$ ) was isolated as a red oil. $38 \mathrm{mg}(0.12 \mathrm{mmol}, 29 \%)$ of the starting material could be recovered.

### 3.3. Photochemical activation of $\mathbf{1}$

A sample of $\mathbf{1}(\mathrm{HCl}$ salt, $108 \mathrm{mg}, 0.47 \mathrm{mmol})$ was dissolved in 200 mL of methanol and placed in an immersion-type photochemical reaction apparatus equipped with a 150 W medium pressure Hg lamp (Hanau TQ 150). The solution was irradiated for 150 min and the solvent was subsequently removed under reduced pressure. The crude product was taken up in 2 mL $\mathrm{MeOH} / \mathrm{H}_{2} \mathrm{O} 1: 9$ and subjected to a preparative HPLC separation. After lyophilization, the TFA salt of compound $9(26 \mathrm{mg}, 0.07 \mathrm{mmol}, 15 \%)$ was isolated as a dark oil as the main product of the reaction.

### 3.4. Conversion of compound $\mathbf{5}$ to compound $\mathbf{4}$

A sample of $\mathbf{5}$ was placed in a screw cap vial and treated with $10 \%$ ammonia solution. After 24 h the sample was analyzed by HR-ESI-MS.

### 3.5. Tandem MS analysis of compounds 2, 5, 4 and 6

MS/MS experiments of compounds 2, 5, $\mathbf{4}$ and $\mathbf{6}$ were conducted on an Orbitrap Velos Pro (Thermo Fischer Scientific) mass spectrometer using an unpurified reaction mixture ( 24 hours, 50 ${ }^{\circ} \mathrm{C}$ ). Because of a superimposed signal, compound $\mathbf{2}$ could not be isolated in the ion trap. Instead it was analyzed by LC/MS/MS on a LTQ-FT Ultra (Thermo Fischer Scientific) instrument. Due
to the low concentration of $\mathbf{2}$, an ITMS2 configuration had to be used, which resulted in a reduced resolution of the spectra.





Scheme S1: Tandem MS fragmentation patterns observed for 2, 5, 4 and 6. Experimentally observed HRMS data is compared to calculated $\mathrm{m} / \mathrm{z}$ values of the postulated fragments.

### 3.6. Quantum chemical calculations



indole ([2',3']-pyrrolo)

monocyclic

Table S2: Overview of compounds investigated by quantum chemical calculations. Compounds I-X and P-X refer to bicyclic structures (left) while compounds B-X and Pyr-X refer to monocyclic structures (right).

| Compound | Scaffold | $\mathbf{Y}_{\mathbf{1}}$ | $\mathbf{Y}_{\mathbf{2}}$ | $\mathbf{R}_{\mathbf{1}}$ | $\mathbf{R}_{\mathbf{2}}$ |
| :---: | :---: | :---: | :---: | :---: | :---: |
| I-1 | isoindole | C | C | H | H |
| I-2 | isoindole | C | C | H | $\mathrm{CH}_{3}$ |
| $\mathbf{I - 3}$ | isoindole | C | C | Cl | $\mathrm{CH}_{3}$ |
| $\mathbf{I - 4}$ | isoindole | $\mathrm{C}-\mathrm{CH}_{3}$ | C | Cl | $\mathrm{CH}_{3}$ |
| P-1 | isoindole | N | N | H | H |
| P-2 | isoindole | N | N | H | $\mathrm{CH}_{3}$ |
| P-3 (1) | isoindole | N | N | Cl | $\mathrm{CH}_{3}$ |
| P-4 | isoindole | $N^{+}-\mathrm{CH}_{3}$ | N | Cl | $\mathrm{CH}_{3}$ |
| P-5 | indole | N | N | H | $\mathrm{CH}_{3}$ |
| P-6 | indole | N | N | Cl | $\mathrm{CH}_{3}$ |
| B-1 | monocyclic | C | C | H | - |
| B-2 | monocyclic | C | C | Cl | - |
| Pyr-1 | monocyclic | N | N | H | - |
| Pyr-2 | monocyclic | N | N | Cl | - |

Structures were initially build with $\mathrm{MOE}^{[5]}$ and optimized at the B3LYP/6-31+G(d) level using Gaussian09. ${ }^{[6]}$ No imaginary frequencies were found at the stationary points, indicating that the optimization converged to a true energy minimum. The optimized geometries were used as input
structures for single point SCF calculations using different basis sets ( $6-31 \mathrm{G}(\mathrm{d}), 6-31+\mathrm{G}(\mathrm{d})$ and $6-311+\mathrm{G}(\mathrm{d})$ ). The data trends discussed in the main text refer to calculations on the $6-311 \mathrm{G}(\mathrm{d})$ level of theory. The detailed results are described in Tables S3 to S16.

NICS calculation: Absolute NMR-shielding tensors were calculated using the GIAO method. ${ }^{[7]}$ The in-plane ring centers (NICS(0)) and out-of-plane ring centers (NICS(1.0)) were calculated based on the positions of carbon and nitrogen atoms only.

NBO Analysis: Wiberg bond indices (WBI) and natural atomic charges were calculated using NBO 3.1 ${ }^{[8]}$ as implemented in Gaussian09.

Electrophilicity index $\omega$ : The electrophilicity index $\omega$ was calculated according to Parr et al. ${ }^{[9]}$ as $\omega=\mu^{2} /(2 \eta)$. The chemical potential $\mu$ is defined as $\mu=0.5 \cdot\left(\mathrm{E}_{\mathrm{LUMO}}+\mathrm{E}_{\text {Номо }}\right)$ and the chemical hardness $\eta$ is defined as $\eta=0.5 \cdot\left(\mathrm{E}_{\text {LUмо }}-\mathrm{E}_{\text {Номо }}\right)$.

Example Gaussian input file:
$\%$ chk $=\mathrm{I}-1 \_6-311+\mathrm{Gd}$
\%nproc=2
\#B3LYP/6-311+G* SCF=tight Test freq nmr pop=nboread IOp(10/46=1) gfprint
remark line goes here
<<Coordinates>>
\$NBO RESONANCE NPA NBO NBOSUM BNDIDX E2PERT NLMO DIPOLE NRT PLOT \$END

### 3.7. NOESY and HSQC NMR experiments in the presence of endothiapepsin

NMR measurements were carried out with $300 \mu \mathrm{l}$ of protein/ligand solution ( $148 \mu \mathrm{M} \mathrm{EP}, 13 \mathrm{mM}$ compound 1) in a $50 \mathrm{mM} \mathrm{CD}_{3} \mathrm{CO}_{2} \mathrm{Na}$ buffer at a pD of 5.1 in a Shigemi tube. The time-resolved ${ }^{1} \mathrm{H}$ - and the $\left[{ }^{1} \mathrm{H}-{ }^{1} \mathrm{H}\right]$-NOESY spectra ( $\tau_{\mathrm{m}}=2 \mathrm{~s}$ ) were measured on a Bruker $800 \mathrm{MHz}{ }^{1} \mathrm{H}$ frequency spectrometer, equipped with a triple-resonance cryoprobe, at 298 K .12 consecutive ${ }^{1} \mathrm{H}$ - and 5 consecutive $\left[{ }^{1} \mathrm{H}-{ }^{1} \mathrm{H}\right]$-NOESY spectra were recorded every hour, starting from mixing the protein and the ligand. The $\left[{ }^{1} \mathrm{H}-{ }^{1} \mathrm{H}\right]$-NOESY spectra were recorded with $96152\left(t_{2}\right) * 272\left(t_{1}\right)$ complex points, $t_{1 \max }=14.2 \mathrm{~ms}, t_{2 \max }=3.0 \mathrm{~s}, 4$ scans, interscan delay 0.3 s . The HSQC spectra were measured three weeks after mixing the protein and the ligand, at a Bruker $600 \mathrm{MHz}{ }^{1} \mathrm{H}$ frequency spectrometer, equipped with a triple-resonance cryoprobe, at 298 K . The HSQC experiment was recorded as two separate experiments, focusing on the aliphatic and on the aromatic and carbonyl ${ }^{13} \mathrm{C}$ chemical shift range, respectively, each with 120 ppm spectral width in the indirect dimension, to ensure complete excitation of all ${ }^{13} \mathrm{C}$ nuclei. The phase-sensitive, echo/anti-echo-
edited HSQC spectra were recorded with $1024\left(t_{2}\right)^{*} 128\left(t_{1}\right)$ complex points, $t_{1 \max }=3.5 \mathrm{~ms}$, $t_{2 \text { max }}=53.2 \mathrm{~ms}$, 192 scans, and interscan delay of 2 s .

### 3.8. Crystallization, soaking and X-ray diffraction data collection

EP was extracted from Suparen samples, kindly supplied by DSM Food Specialties, as described previously. ${ }^{[10]}$ Subsequently, the protein was crystallized in its apo-form upon streak-seeding using the vapor diffusion method with a mother liquor composed of $100 \mathrm{mM} \mathrm{NH}_{4} \mathrm{OAc}, 100 \mathrm{mM}$ NaOAc $\mathrm{pH} 4.6,24-30 \%(\mathrm{w} / \mathrm{v})$ PEG 4000 at $17{ }^{\circ} \mathrm{C}{ }^{[11]}$ The originally purchased compound $\mathbf{1}$ has been soaked into EP crystals once at a concentration of 45 mM and thrice at 90 mM resulting in the structures EP-1-2 and EP-2, respectively. Following a similar approach, all soaking experiments with the compounds isolated via $\operatorname{HPLC}(\mathbf{1}, \mathbf{3}, \mathbf{4}, \mathbf{5}$ and $\mathbf{6})$ have been performed at two different concentrations of 90 and 250 mM . At 45 and 90 mM , soaking was performed for 48 h at $17^{\circ} \mathrm{C}$ in $70 \mathrm{mM} \mathrm{NH}_{4} \mathrm{OAc}, 70 \mathrm{mM}$ NaOAc $\mathrm{pH} 4.6,16-20 \%$ PEG 4000, $23 \%$ glycerol, $9 \%$ DMSO and 45 or 90 mM of the ligand, respectively. At the higher 90 mM ligand concentration, EP crystals have been transferred to this solution via a $1: 1$ mixture of this solution with the crystallization mother liquor. The same procedure was also applied at 250 mM where the soaking solution consisted of $65 \mathrm{mM} \mathrm{NH} 4 \mathrm{OAc}, 65 \mathrm{mM}$ NaOAc $\mathrm{pH} 4.6,14-18 \%$ PEG 4000, $10 \%$ glycerol, $25 \%$ DMSO and 250 mM ligand. In contrast to the 90 mM soaks, the duration of the experiment had been reduced to 24 h under the more harsh conditions at 250 mM . Finally, all crystals were directly flash-frozen in liquid nitrogen prior to data collection at the BESSY MX beamlines BL14.1 and BL14.2. ${ }^{[12]}$ Subsequently diffraction data have been processed using XDS. ${ }^{[13]}$

### 3.9. Structure determination

To be able to evaluate and compare the outcome of all diffraction experiments thoroughly, we used our automated refinement pipeline ${ }^{[14]}$ to determine structural models for all collected data sets via molecular replacement using Phaser ${ }^{[15]}$ and several Phenix ${ }^{[17]}$-based refinement steps. Following this approach, the data set with the highest quality and clearest ligand electron density has been chosen for each ligand-bound structure (EP-2: 45 mM, EP-1-2: $90 \mathrm{mM}, \mathrm{EP}-4: 90 \mathrm{mM}$, EP-3: 250 mM ). Subsequently, the associated models were manually further refined against structure factor amplitudes until convergence. Library files for all non-standard ligands were generated using the Grade web server application. ${ }^{[18]}$ For ligand 2, the obtained restraints have been modified in a way to allow the pyridazinium bridge head nitrogen and attached carbon to deviate from the ring plane as observed in the unrestrained SHELXL refinement described below. In order to obtain a reliable estimate of the real binding-site occupancy of 2, we first refined the occupancy of the system Asp33- $\mathrm{H}_{2} \mathrm{O}-\mathrm{Gly} 221-\mathrm{Thr} 222$ that appeared in two alternative conformations, one of which represents the apo-form and one the ligand-bound conformation
(Fig. 4B). After validation of the refined value via visual inspection of the anisotropic displacement parameters, the resulting occupancy for the latter conformational state was assigned also to ligand 2 ( $67 \%$ for EP-2 structure and $63 \%$ for EP-1-2). Finally, all resulting structures and structure factor amplitudes have been deposited in the PDB. Data collection and refinement statistics are given in Table S1. Figures displaying structural information have been prepared using Pymol. ${ }^{[19]}$

### 3.10. Estimation of standard deviations for different geometric parameters

To enable an estimation of standard deviations for geometric parameters, a least-squares BLOC refinement was performed against intensities using SHELXL. ${ }^{[20]}$ As a starting point, the final Phenix-refined EP-1-2 structure has been chosen and adapted in several modeling and refinement steps. For the subsequent determination of distances and angles within ligand 2, all of its atoms were refined in an unrestrained manner except for planarity restraints for the three ring systems excluding the carbon atom next to the pyridazinium nitrogen which was observed to be significantly out-of-plane (Fig. 3B). This observation was made via a similar refinement in which the planarity restraint for the bicyclic component of $\mathbf{2}$ was additionally omitted and the deviation of each atom of the pyridazinium ring from a least-squares plane made up by the residual five atoms of the ring was calculated along with estimated standard deviations. While the bridge head pyridazinium nitrogen was found to be located significantly below the plane $(-0.20 \pm 0.05 \AA$, corresponding to $-4.1 \sigma$ ) and the attached carbon atom accordingly above the plane $(0.23 \pm 0.05$ $\AA$, corresponding to $4.3 \sigma$ ), all other atoms did not deviate as significantly from their respective planes ( -0.11 to $0.06 \AA,-2.1$ to $1.1 \sigma$ ). In order to ensure that these results had not been biased by the ligand restraints used in the previous Phenix-based refinement steps, we repeated both analyses with starting structures that contained an alcohol CO bond length of $1.41 \AA$ instead of the aldehyde bond length of $1.22 \AA$ and an approximately planar pyridazinium ring, respectively. Importantly, the results did not deviate significantly from those obtained before.

### 3.11. Queries in the Cambridge Structural Database (CSD)

CSD searches have been performed using Conquest ${ }^{[21]}$ vs. 1.18 based on CSD vs. 5.37 as of November 2015 (with 2 updates) while the results were analyzed with Mercury ${ }^{[22]}$ vs. 3.8 and plotted via the statistical program R. ${ }^{[23]}$ The following adjustments have been used for all searches: 3D coordinates determined, only organics, $\mathrm{R} \leq 0.075$, not disordered, no errors and not polymeric. The investigated chemical structures and geometric parameters are defined in Figs. S4, S5 and S6.

## 4. Experimental data

Compound 1. ${ }^{1} \mathbf{H}$ NMR ( $400 \mathrm{MHz}, \mathrm{D}_{2} \mathrm{O}$ ) $\delta=2.78(\mathrm{~s}, 3 \mathrm{H}), 2.85(\mathrm{~s}, 3 \mathrm{H}), 3.97(\mathrm{~s}, 3 \mathrm{H}), 9.53(\mathrm{~s}$, $1 \mathrm{H}) .{ }^{13} \mathbf{C}$ NMR $\left(101 \mathrm{MHz}, \mathrm{D}_{2} \mathrm{O}\right) \delta=10.3,10.7,32.8,110.4,113.5,129.6,136.5,142.4,150.9$. HRMS (ESI+) calculated for $\mathrm{C}_{9} \mathrm{H}_{11} \mathrm{ClN}_{3}: 196.0636[\mathrm{M}+\mathrm{H}]^{+}$; found: 196.0636.

Compound 2. HRMS (ESI+) calculated for $\mathrm{C}_{27} \mathrm{H}_{29} \mathrm{ClN}_{9} \mathrm{O}: 530.2178$ [M] ${ }^{+}$; found: 530.2176.
Compound 3 (TFA salt). ${ }^{1} \mathbf{H}$ NMR ( $400 \mathrm{MHz}, \mathrm{D}_{2} \mathrm{O}$ ) $\delta=2.63(\mathrm{~s}, 3 \mathrm{H}), 2.65(\mathrm{~s}, 3 \mathrm{H}), 3.81(\mathrm{~s}, 3 \mathrm{H})$, $4.11(\mathrm{~s}, 3 \mathrm{H}), 9.15(\mathrm{~s}, 1 \mathrm{H})$. HRMS (ESI+) calculated for $\mathrm{C}_{10} \mathrm{H}_{14} \mathrm{~N}_{3} \mathrm{O}: 192.1131[\mathrm{M}+\mathrm{H}]^{+}$; found: 192.1128.

Compound $4 .{ }^{a}{ }^{\mathbf{1}} \mathbf{H}$ NMR ( $400 \mathrm{MHz}, \mathrm{MeOH}-d 4$ ) $\delta=2.26(\mathrm{~s}, 3 \mathrm{H}), 2.62(\mathrm{~s}, 3 \mathrm{H}), 2.63(\mathrm{~s}, 3 \mathrm{H}), 2.80$ (s, 3H), $3.63(\mathrm{~s}, 3 \mathrm{H}), 3.81(\mathrm{~s}, 3 \mathrm{H}), 8.52(\mathrm{~s}, 1 \mathrm{H}), 9.70(\mathrm{~s}, 1 \mathrm{H}) .{ }^{13} \mathbf{C}$ NMR ( $101 \mathrm{MHz}, \mathrm{MeOH}-d 4$ ) $\delta$ $=10.4,10.8,11.0,11.8,30.9,31.7,111.8,125.3,133.6,143.8,170.3$. HRMS (ESI + ) calculated for $\mathrm{C}_{18} \mathrm{H}_{21} \mathrm{~N}_{6} \mathrm{O}: 337.1771[\mathrm{M}+\mathrm{H}]^{+}$; found: 337.1767.

Compound 5 (TFA salt). ${ }^{1} \mathbf{H}$ NMR ( $400 \mathrm{MHz}, \mathrm{D}_{2} \mathrm{O}$ ) $\delta=2.45(\mathrm{~s}, 3 \mathrm{H}), 2.80(\mathrm{~s}, 3 \mathrm{H}), 2.83(\mathrm{~s}, 3 \mathrm{H})$, $2.87(\mathrm{~s}, 2 \mathrm{H}), 3.95(\mathrm{~s}, 3 \mathrm{H}), 3.97(\mathrm{~s}, 2 \mathrm{H}), 9.70(\mathrm{~s}, 1 \mathrm{H}) .{ }^{13} \mathbf{C}$ NMR $\left(101 \mathrm{MHz}, \mathrm{D}_{2} \mathrm{O}\right) \delta=10.6,10.8$, $10.9,12.3,33.4,105.6,109.4,115.6,116.2\left(\mathrm{q},{ }^{1} J_{C F}=288.7\right), 116.4,128.5,132.1,138.4,142.9$, 144.0, 144.9, 150.4, 151.4, 162.9 (q, ${ }^{2} J_{C F}=35.2$ ). HRMS (ESI+) calculated for $\mathrm{C}_{18} \mathrm{H}_{20} \mathrm{ClN}_{6}$ : $355.1438[\mathrm{M}]^{+}$; found: 355.1433.

Compound 6 (TFA salt). ${ }^{1} \mathbf{H}$ NMR ( $400 \mathrm{MHz}, \mathrm{D}_{2} \mathrm{O}$ ) $\delta=2.32(\mathrm{~s}, 3 \mathrm{H}), 2.54(\mathrm{~s}, 3 \mathrm{H}), 2.75(\mathrm{~s}, 3 \mathrm{H})$, $2.79(\mathrm{~s}, 3 \mathrm{H}), 2.87(\mathrm{~s}, 3 \mathrm{H}), 2.93(\mathrm{~s}, 3 \mathrm{H}), 3.89(\mathrm{~s}, 3 \mathrm{H}), 3.95(\mathrm{~s}, 3 \mathrm{H}), 4.04(\mathrm{~s}, 3 \mathrm{H}), 9.74(\mathrm{~s}, 1 \mathrm{H}), 10.15$ $(\mathrm{s}, 1 \mathrm{H}), 10.44(\mathrm{~s}, 1 \mathrm{H})$. HRMS (ESI+) calculated for $\mathrm{C}_{27} \mathrm{H}_{31} \mathrm{ClN}_{9} \mathrm{O}: 532.2335[\mathrm{M}]^{+}$; found: 532.2334 .

Compound 8. HRMS (ESI + ) calculated for $\mathrm{C}_{10} \mathrm{H}_{13} \mathrm{ClN}_{3} \mathrm{O}$ : $226.0742[\mathrm{M}+\mathrm{H}]^{+}$; found: 226.0730.
Compound 9 (TFA salt). ${ }^{1} \mathbf{H}$ NMR ( $400 \mathrm{MHz}, \mathrm{D}_{2} \mathrm{O}$ ) $\delta=3.44-3.49(\mathrm{~m}, 6 \mathrm{H}), 4.14(\mathrm{~s}, 3 \mathrm{H}), 5.07$ $(\mathrm{s}, 2 \mathrm{H}), 5.18(\mathrm{~s}, 2 \mathrm{H}), 9.84(\mathrm{~s}, 1 \mathrm{H}) .{ }^{13} \mathbf{C}$ NMR $\left(101 \mathrm{MHz}, \mathrm{D}_{2} \mathrm{O}\right) \delta=33.8,57.6,58.2,61.5,63.2$, 114.2, 115.6, $116.2\left(\mathrm{q},{ }^{1} J_{C F}=291.7\right), 128.0,135.1,143.1,150.9,162.8\left(\mathrm{q},{ }^{2} J_{C F}=43.1\right)$. HRMS (ESI+) calculated for $\mathrm{C}_{11} \mathrm{H}_{15} \mathrm{ClN}_{3} \mathrm{O}_{2}$ : $256.0847[\mathrm{M}+\mathrm{H}]^{+}$; found: 256.0833.
${ }^{a}$ Due to its poor solubility, compound $\mathbf{4}$ was dissolved in a mixture of MeOH-d4 and acetone-d6. The low concentration prohibited the observation of several signals of quaternary carbon atoms in the ${ }^{13} \mathrm{C}$ spectrum.

## 5. Data from quantum chemical calculations

Table S3: Natural atomic charges at atom position 4.

| Compound | $\underline{\mathbf{6 - 3 1 G}(\mathbf{d})}$ | $\underline{\mathbf{6 - 3 1}+\mathbf{G}(\mathbf{d})}$ | $\underline{\mathbf{6 - 3 1 1 + G}(\mathbf{d})}$ |
| :---: | :---: | :---: | :---: |
| $\mathbf{I}-\mathbf{1}$ | -0.21638 | -0.21759 | -0.18359 |
| $\mathbf{I}-2$ | -0.21592 | -0.21616 | -0.18410 |
| $\mathbf{I - 3}$ | -0.01476 | -0.01326 | -0.00804 |
| $\mathbf{I - 4}$ | -0.00645 | -0.00394 | 0.01021 |
| $\mathbf{P - 1}$ | 0.03557 | 0.03206 | 0.06757 |
| $\mathbf{P - 2}$ | 0.03546 | 0.03085 | 0.08457 |
| $\mathbf{P - 3 ~ ( 1 ) ~}$ | 0.22077 | 0.20894 | 0.22538 |
| $\mathbf{P - 4}$ | 0.23685 | 0.22813 | 0.24438 |
| $\mathbf{P - 5}$ | -0.03548 | -0.03769 | -0.00296 |
| $\mathbf{P - 6}$ | 0.15427 | 0.14564 | 0.16039 |
| $\mathbf{B - 1}$ | -0.23527 | -0.24218 | -0.20333 |
| $\mathbf{B - 2}$ | -0.23848 | -0.24463 | -0.20632 |
| $\mathbf{P y r - 1}$ | -0.01251 | -0.02041 | 0.02052 |
| $\mathbf{P y r - 2}$ | -0.01417 | -0.02107 | 0.01868 |

Table S4: Natural atomic charges at atom position 1.

| Compound | $\underline{\mathbf{6 - 3 1 G}(\mathbf{d})}$ | $\underline{\mathbf{6 - 3 1}+\mathbf{G}(\mathbf{d})}$ | $\underline{\mathbf{6 - 3 1 1 + G}(\mathbf{d})}$ |
| :---: | :---: | :---: | :---: |
| $\mathbf{I - 1}$ | -0.21638 | -0.21759 | -0.18359 |
| $\mathbf{I - 2}$ | -0.21594 | -0.21959 | -0.17120 |
| $\mathbf{I - 3}$ | -0.21468 | -0.21542 | -0.17048 |
| $\mathbf{I - 4}$ | -0.21542 | -0.21410 | -0.18760 |
| $\mathbf{P - 1}$ | 0.03439 | 0.03198 | 0.06752 |
| $\mathbf{P - 2}$ | 0.03546 | 0.03378 | 0.06459 |
| $\mathbf{P - 3 ~ ( 1 ) ~}$ | 0.03818 | 0.03664 | 0.07056 |
| $\mathbf{P - 4}$ | 0.12658 | 0.13490 | 0.16506 |
| $\mathbf{P - 5}$ | 0.01105 | 0.00766 | 0.04347 |
| $\mathbf{P - 6}$ | 0.01253 | 0.00908 | 0.04383 |
| $\mathbf{B - 1}$ | -0.23527 | -0.24218 | -0.20333 |
| $\mathbf{B - 2}$ | -0.04358 | -0.05584 | -0.04102 |
| $\mathbf{P y r - 1}$ | -0.01251 | -0.02041 | 0.02052 |
| $\mathbf{P y r - 2}$ | 0.15588 | 0.14349 | 0.16034 |

Table S5: WBI for atomic bond at position 2-3.

| Compound | 6-31G(d) | $\underline{6-31+G(d)}$ | 6-311+G(d) |
| :---: | :---: | :---: | :---: |
| I-1 | 1.2657 | 1.2663 | 1.2669 |
| I-2 | 1.2604 | 1.2603 | 1.2167 |
| I-3 | 1.2556 | 1.2542 | 1.2535 |
| I-4 | 1.2204 | 1.2198 | 1.2167 |
| P-1 | 1.2191 | 1.2227 | 1.2239 |
| P-2 | 1.2107 | 1.2129 | 1.2147 |
| P-3 (1) | 1.1936 | 1.1945 | 1.1935 |
| P-4 | 1.1269 | 1.1302 | 1.1264 |
| P-5 | 1.335 | 1.3374 | 1.3397 |
| P-6 | 1.3086 | 1.3095 | 1.3097 |
| B-1 | 1.4368 | 1.4371 | 1.439 |
| B-2 | 1.432 | 1.4324 | 1.4327 |
| Pyr-1 | 1.3976 | 1.4027 | 1.4047 |
| Pyr-2 | 1.3736 | 1.3765 | 1.3763 |

Table S6: WBI for atomic bond at position 3-4.

| Compound | $\mathbf{6 - 3 1 G}(\mathbf{d})$ | $\mathbf{6 - 3 1 + G ( d )}$ | $\mathbf{6 - 3 1 1 + \mathbf { G } ( \mathbf { d } )}$ |
| :---: | :---: | :---: | :---: |
| $\mathbf{I - 1}$ | 1.5988 | 1.5980 | 1.6005 |
| $\mathbf{I - 2}$ | 1.6013 | 1.6004 | 1.5679 |
| $\mathbf{I - 3}$ | 1.5624 | 1.5598 | 1.5623 |
| $\mathbf{I - 4}$ | 1.5758 | 1.5735 | 1.5679 |
| $\mathbf{P - 1}$ | 1.6445 | 1.6473 | 1.6468 |
| $\mathbf{P - 2}$ | 1.6492 | 1.6504 | 1.6436 |
| $\mathbf{P - 3 ~ ( 1 ) ~}$ | 1.6370 | 1.6387 | 1.6356 |
| $\mathbf{P - 4}$ | 1.5705 | 1.5672 | 1.5648 |
| $\mathbf{P - 5}$ | 1.5291 | 1.531 | 1.5299 |
| $\mathbf{P - 6}$ | 1.5408 | 1.5451 | 1.5437 |
| $\mathbf{B - 1}$ | 1.4368 | 1.4371 | 1.439 |
| $\mathbf{B - 2}$ | 1.4037 | 1.4027 | 1.4025 |
| $\mathbf{P y r - 1}$ | 1.4824 | 1.4866 | 1.4856 |
| $\mathbf{P y r - 2}$ | 1.4748 | 1.4782 | 1.4745 |

Table S7: $\operatorname{NICS}(0)$ values at $\operatorname{ring} \boldsymbol{A}$.

| Compound | 6-31G(d) ppm] $^{\text {d }}$ | $\underline{6-31+G(d) ~}[\mathrm{ppm}]$ | 6-311+G(d) 5 ppm$]$ |
| :---: | :---: | :---: | :---: |
| I-1 | -8.3573 | -6.7175 | -6.5780 |
| I-2 | -7.8801 | -6.3974 | -6.1650 |
| I-3 | -8.1807 | -7.3796 | -7.4937 |
| I-4 | -8.6765 | -7.3959 | -7.0781 |
| P-1 | -5.0951 | -4.4809 | -4.5637 |
| P-2 | -4.6257 | -4.1097 | -4.1701 |
| P-3 (1) | -5.6378 | -4.9806 | -4.9024 |
| P-4 | -6.2992 | -5.5065 | -5.2127 |
| P-5 | -8.9700 | -8.3265 | -8.3780 |
| P-6 | -8.9700 | -8.3265 | -8.3780 |
| B-1 | -9.6482 | -8.0216 | -7.9345 |
| B-2 | -10.2524 | -8.9027 | -9.9924 |
| Pyr-1 | -5.7556 | -5.0531 | -5.2485 |
| Pyr-2 | -6.4445 | -5.8008 | -5.8481 |

Table S8: $\operatorname{NICS}(0)$ values at ring $\boldsymbol{B}$.

| Compound | $\mathbf{6 - 3 1 G}(\mathbf{d})[\mathbf{p p m}]$ | $\mathbf{6 - 3 1 + \mathbf { G } ( \mathbf { d } ) \lceil \mathbf { p p m } ]}$ | $\underline{\mathbf{6 - 3 1 1}+\mathbf{G}(\mathbf{d})}$ [ppm] |
| :---: | :---: | :---: | :---: |
| $\mathbf{I - 1}$ | -17.4253 | -15.6495 | -15.3332 |
| $\mathbf{I - 2}$ | -16.8376 | -15.3714 | -15.2396 |
| $\mathbf{I - 3}$ | -17.2753 | -16.0619 | -15.8239 |
| $\mathbf{I - 4}$ | -16.9905 | -15.5229 | -15.2236 |
| $\mathbf{P - 1}$ | -17.9539 | -16.4588 | -16.0083 |
| $\mathbf{P - 2}$ | -17.3109 | -15.8880 | -15.6712 |
| $\mathbf{P - 3 ~ ( 1 ) ~}$ | -17.4879 | -16.4384 | -16.1498 |
| $\mathbf{P - 4}$ | -16.6222 | -15.4282 | -15.1428 |
| $\mathbf{P - 5}$ | -13.5300 | -12.7020 | -12.3425 |
| $\mathbf{P - 6}$ | -13.9308 | -13.1218 | -12.6420 |

Table S9: $\operatorname{NICS}_{\mathrm{ZZ}}(0)$ values at $\operatorname{ring} \boldsymbol{A}$.

| Compound | 6-31G(d) [ppm] | 6-31+G(d) $[\mathrm{ppm}]$ | $\underline{6-311+G(d) ~[p p m] ~}$ |
| :---: | :---: | :---: | :---: |
| I-1 | -5.6777 | -5.4037 | -6.0251 |
| I-2 | -3.1337 | -2.8154 | -3.4881 |
| I-3 | -29.4937 | -29.6613 | -30.4412 |
| I-4 | -2.5119 | -1.9326 | -2.4267 |
| P-1 | -2.4149 | -2.4149 | -3.5926 |
| P-2 | 0.1742 | 0.1742 | -0.9561 |
| P-3 (1) | 0.0075 | 0.0075 | -0.7008 |
| P-4 | 4.2965 | 4.6270 | 4.4654 |
| P-5 | -11.4265 | -11.9404 | -12.6243 |
| P-6 | -10.2845 | -10.4693 | -11.0410 |
| B-1 | -13.7989 | -13.1765 | -14.0643 |
| B-2 | -13.3587 | -12.6903 | -13.3565 |
| Pyr-1 | -10.4678 | -10.7302 | -11.5122 |
| Pyr-2 | -9.8905 | -9.8762 | -10.4523 |

Table S10: $\operatorname{NICS}_{Z Z}(0)$ values at ring $B$.

| Compound | $\mathbf{6 - 3 1 G} \mathbf{( d )}[\mathbf{p p m}]$ | $\mathbf{6 - 3 1 + \mathbf { G } ( \mathbf { d } ) \lceil \mathbf { p p m } ]}$ | $\mathbf{6 - 3 1 1 + \mathbf { G } ( \mathbf { d } ) [ \mathbf { p p m } ]}$ |
| :---: | :---: | :---: | :---: |
| $\mathbf{I - 1}$ | -16.6309 | -16.1241 | -16.6410 |
| $\mathbf{I - 2}$ | -13.8302 | -13.8624 | -14.5943 |
| $\mathbf{I - 3}$ | -13.8886 | -13.9385 | -14.6399 |
| $\mathbf{I - 4}$ | -12.6404 | -12.2745 | -12.9272 |
| $\mathbf{P - 1}$ | -16.7125 | -16.7125 | -16.8537 |
| $\mathbf{P - 2}$ | -13.7984 | -13.7984 | -14.5550 |
| $\mathbf{P - 3 ~ ( 1 ) ~}$ | -13.5122 | -13.5122 | -14.1549 |
| $\mathbf{P - 4}$ | -12.9942 | -12.6259 | -13.3833 |
| $\mathbf{P - 5}$ | -28.3137 | -28.2621 | -28.3954 |
| $\mathbf{P - 6}$ | -25.8995 | -25.6857 | -25.7820 |

Table S11: NICS(1.0) values at $\operatorname{ring} \boldsymbol{A}$.

| Compound | 6-31G(d) $[\mathrm{ppm}]$ | $\underline{\text { 6-31+G(d) }[\mathrm{ppm}]}$ | 6-311+G(d) 5 ppm$]$ |
| :---: | :---: | :---: | :---: |
| I-1 | -9.4985 | -8.2738 | -8.4818 |
| I-2 | -9.0535 | -7.9922 | -8.1357 |
| I-3 | -9.1508 | -8.0746 | -8.0797 |
| I-4 | -8.9200 | -7.8523 | -7.8267 |
| P-1 | -9.2444 | -8.3725 | -8.5583 |
| P-2 | -8.5047 | -7.7288 | -7.9358 |
| P-3 (1) | -8.2936 | -7.4413 | -7.6149 |
| P-4 | -7.2098 | -6.4063 | -6.4560 |
| P-5 | -11.7000 | -10.8731 | -11.0804 |
| P-6 | -11.1418 | -10.3154 | -10.5151 |
| B-1 | -11.1929 | -10.1364 | -10.1345 |
| B-2 | -11.0419 | -9.9924 | -10.0023 |
| Pyr-1 | -11.0397 | -10.2596 | -10.4739 |
| Pyr-2 | -10.6968 | -9.9392 | -10.0860 |

Table S12: NICS(1.0) values at ring $\boldsymbol{B}$.

| Compound | $\mathbf{6 - 3 1 G}(\mathbf{d})[\mathbf{p p m}]$ | $\mathbf{6 - 3 1 + \mathbf { G } ( \mathbf { d } ) \lceil \mathbf { p p m } ]}$ | $\underline{\mathbf{6 - 3 1 1}+\mathbf{G}(\mathbf{d})}$ [ppm] |
| :---: | :---: | :---: | :---: |
| $\mathbf{I - 1}$ | -13.8887 | -12.2812 | -12.3669 |
| $\mathbf{I}-\mathbf{2}$ | -13.6878 | -12.2126 | -12.4800 |
| $\mathbf{I - 3}$ | -13.7288 | -12.3437 | -12.3577 |
| $\mathbf{I - 4}$ | -13.4765 | -12.0651 | -12.1919 |
| $\mathbf{P - 1}$ | -14.1433 | -12.7693 | -12.8363 |
| $\mathbf{P - 2}$ | -13.8699 | -12.5762 | -12.9047 |
| $\mathbf{P - 3 ~ ( 1 ) ~}$ | -13.7434 | -12.5977 | -12.6247 |
| $\mathbf{P - 4}$ | -14.0630 | -12.8504 | -12.9812 |
| $\mathbf{P - 5}$ | -11.3494 | -10.4291 | -10.5042 |
| $\mathbf{P - 6}$ | -11.4074 | -10.4793 | -10.4817 |

Table S13: HOMO energies

| Compound | 6-31G(d) [eV] | 6-31+G(d) $[\mathrm{eV}]$ | 6-311+G(d) [eV] |
| :---: | :---: | :---: | :---: |
| I-1 | -4.72988 | -5.05424 | -5.09887 |
| I-2 | -4.37777 | -4.65396 | -4.69968 |
| I-3 | -4.64471 | -4.8942 | -4.93941 |
| I-4 | -4.58593 | -4.82294 | -4.86839 |
| P-1 | -5.78840 | -6.11440 | -6.15957 |
| P-2 | -5.33942 | -5.61997 | -5.66514 |
| P-3 (1) | -5.62486 | -5.88065 | -5.92555 |
| P-4 | -9.84780 | -9.98250 | -10.02440 |
| P-5 | -5.7751 | -6.1315 | -6.1645 |
| P-6 | -6.1476 | -6.3922 | -6.4390 |
| B-1 | -6.6940 | -6.9939 | -7.0385 |
| B-2 | -6.7016 | -6.9457 | -6.9822 |
| Pyr-1 | -6.3465 | -6.6997 | -6.7332 |
| Pyr-2 | -6.8390 | -7.1514 | -7.1835 |

Table S14: LUMO energies

| Compound | 6-31G(d) $\int \mathrm{eV}$ ] | $\underline{6-31+G(d) ~}[\mathrm{eV}]$ | $\underline{6-311+G(d)[e V]}$ |
| :---: | :---: | :---: | :---: |
| I-1 | -0.38612 | -0.78994 | -0.78994 |
| I-2 | -0.27946 | -0.63456 | -0.63456 |
| I-3 | -0.59592 | -0.91375 | -0.91375 |
| I-4 | -0.51211 | -0.81253 | -0.81253 |
| P-1 | -1.1428 | -1.50479 | -1.50479 |
| P-2 | -0.97117 | -1.28764 | -1.28764 |
| P-3 (1) | -1.25417 | -1.53227 | -1.53227 |
| P-4 | -5.92909 | -6.07902 | -6.07902 |
| P-5 | -0.6664 | -1.0275 | -1.0917 |
| P-6 | -0.9285 | -1.2251 | -1.2912 |
| B-1 | 0.0841 | -0.3951 | -0.4620 |
| B-2 | -0.3499 | -0.7761 | -0.8444 |
| Pyr-1 | -1.3910 | -1.7979 | -1.8618 |
| Pyr-2 | -1.8153 | -2.1742 | -2.2376 |

Table S15: HOMO-LUMO energy gaps

| Compound | $\underline{\mathbf{6 - 3 1 G}(\mathbf{d})[\mathbf{e V}]}$ | $\mathbf{6 - 3 1 + \mathbf { G } ( \mathbf { d } ) [ \mathbf { e V } ]}$ | $\underline{\mathbf{6 - 3 1 1}+\mathbf{G}(\mathbf{d})}[\mathbf{e V}]$ |
| :---: | :---: | :---: | :---: |
| $\mathbf{I - 1}$ | -4.34375 | -4.26429 | -4.30892 |
| $\mathbf{I - 2}$ | -4.09830 | -4.01939 | -4.06511 |
| $\mathbf{I - 3}$ | -4.04878 | -3.98048 | -4.02565 |
| $\mathbf{I - 4}$ | -4.07381 | -4.01041 | -4.05585 |
| $\mathbf{P - 1}$ | -4.64553 | -4.60961 | -4.65478 |
| $\mathbf{P - 2}$ | -4.36824 | -4.33232 | -4.37749 |
| $\mathbf{P - 3 ~ ( 1 ) ~}$ | -4.37069 | -4.34838 | -4.39328 |
| $\mathbf{P - 4}$ | -3.91871 | -3.90347 | -3.94538 |
| $\mathbf{P - 5}$ | -5.1087 | -5.1040 | -5.0727 |
| $\mathbf{P - 6}$ | -5.2191 | -5.1672 | -5.1479 |
| $\mathbf{B - 1}$ | -6.7781 | -6.5988 | -6.5765 |
| $\mathbf{B - 2}$ | -6.3517 | -6.1696 | -6.1378 |
| $\mathbf{P y r - 1}$ | -4.9555 | -4.9019 | -4.8714 |
| $\mathbf{P y r - 2}$ | -5.0238 | -4.9772 | -4.9459 |

Table S16: Electrophilicity index $\omega$.

| Compound | $\underline{\mathbf{6 - 3 1 G}(\mathbf{d})[\mathbf{e V}]}$ | $\underline{\mathbf{6 - 3 1}+\mathbf{G}(\mathbf{d})\lceil\mathbf{e V}]}$ | $\underline{\mathbf{6 - 3 1 1}+\mathbf{G}(\mathbf{d})}[\mathbf{e V}]$ |
| :---: | :---: | :---: | :---: |
| $\mathbf{I - 1}$ | 1.5064 | 2.0023 | 2.0120 |
| $\mathbf{I - 2}$ | 1.3231 | 1.7396 | 1.7499 |
| $\mathbf{I - 3}$ | 1.6958 | 2.1186 | 2.1276 |
| $\mathbf{I - 4}$ | 1.5949 | 1.9798 | 1.9893 |
| $\mathbf{P - 1}$ | 2.5853 | 3.1484 | 3.1550 |
| $\mathbf{P - 2}$ | 2.2791 | 2.7534 | 2.7608 |
| $\mathbf{P - 3 ~ ( 1 ) ~}$ | 2.7067 | 3.1593 | 3.1650 |
| $\mathbf{P - 4}$ | 15.8796 | 16.5220 | 16.4319 |
| $\mathbf{P - 5}$ | 2.0305 | 2.5104 | 2.5949 |
| $\mathbf{P - 6}$ | 2.3984 | 2.8073 | 2.9020 |
| $\mathbf{B - 1}$ | 1.6115 | 2.0685 | 2.1386 |
| $\mathbf{B - 2}$ | 1.9571 | 2.4161 | 2.4950 |
| $\mathbf{P y r - 1}$ | 3.0204 | 3.6827 | 3.7912 |
| $\mathbf{P y r - 2}$ | 3.7271 | 4.3682 | 4.4864 |

## 6. Data from NMR, MS, and HPLC experiments

### 6.1. NMR spectroscopy

${ }^{1} \mathrm{H}\left(400 \mathrm{MHz}, \mathrm{D}_{2} \mathrm{O}\right)$ and ${ }^{13} \mathrm{C}$ NMR ( $101 \mathrm{MHz}, \mathrm{D}_{2} \mathrm{O}$ ) of compound $\mathbf{1}$


${ }^{1} \mathrm{H}\left(400 \mathrm{MHz}, \mathrm{D}_{2} \mathrm{O}\right)$ NMR spectrum of compound $\mathbf{3}$

${ }^{1} \mathrm{H}(400 \mathrm{MHz}, \mathrm{MeOH}-d 4)$ and ${ }^{13} \mathrm{C}$ NMR ( 101 MHz , MeOH- $d 4$ ) spectra of compound 4


${ }^{1} \mathrm{H}\left(400 \mathrm{MHz}, \mathrm{D}_{2} \mathrm{O}\right)$ and ${ }^{13} \mathrm{C}$ NMR ( $101 \mathrm{MHz}, \mathrm{D}_{2} \mathrm{O}$ ) spectra of compound 5


${ }^{1} \mathrm{H}\left(400 \mathrm{MHz}, \mathrm{D}_{2} \mathrm{O}\right)$ NMR spectrum of compound $\mathbf{6}$

${ }^{1} \mathrm{H}\left(400 \mathrm{MHz}, \mathrm{D}_{2} \mathrm{O}\right)$ NMR spectrum of compound 9



### 6.2. Mass spectrometry

## ESI-MS and HR-MS spectra of compound 1

O:ILTQ-FTI...IAG Klebel151119_EM_007_KI
23.11.2015 13:56:44


O:ILTQ-FT ...IAG Klebel151119_EM_007_KI
23.11.2015 13:56:44


ESI-MS and HR-MS spectra of compound 2


## ESI-MS and HR-MS spectra of compound $\mathbf{3}$



O:ILTQ-FTI...IAG Klebel151119_EM_009_KI
11/23/2015 2:08:30 PM


## ESI-MS and HR-MS spectra of compound 4



O:ILTQ-FT1...IAG Klebel151207_EM_017_KI 11.12.2015 09:37:36


## ESI-MS and HR-MS spectra of compound 5



O:L2015\...UAG Klebel151119_EM_008_KI
23.11.2015 14:02:36


## ESI-MS and HR-MS spectra of compound 6



O:ILTQ-FT2015\...1151207_SY_018_KI
12/16/2015 2:25:33 PM


## ESI-MS and HR-MS spectra of compounds $\mathbf{8}$ and $\mathbf{9}$

O:IOrbitrapl0516|Klebel160513_SY_027_KI
5/18/2016 12:20:15 PM
160513_SY_027_KI \#46-54 RT: 0.65-0.76 AV: 9 SM: 7B NL: 9.33E7
F: FTMS + p NSI Full ms [100.00-600.00]


O:IOrbitrapl0516|Klebel160513 SY 027 KI
5/18/2016 12:20:15 PM



MS/MS Spectra of compound 2
160518_LC_028_KI 1 \#905-971 RT: 12.76-13.61 AV: 11 SM: 7B NL: 1.33E2


160518_LC_028_KI 1 \#966-994 RT: 13.64-13.98 AV: 5 NL: 4.67E2
F: TMS + CESI Full ms2 530.20@cid $35.00[145.00-2000.00]$


MS/MS Spectra of compound 4



160518_SY O26_Kl_2\#311-328 RT: 4.73-4.99 AV: 18 SM: 7B NL: 1.85E5
F: FTMS + p NSIW Full ms2 $337.20 @ c i d 35.00[90.00-600.00$ ]




160518_SY_026_KI_2\#259-275 RT: 3.87-4.14 AV: 17 SM: 7B NL: 7.45E4


MS/MS Spectrum of compound 5

160518_SY_026_KI_2\#74-107 RT: 1.11-1.59 AV: 34 SM: 7B NL: 4.78 E6
F:FTMS + $\begin{aligned} & \text { NSIFull ms } 2356.10 @ c i d 35.00 ~[95.00-600.00]\end{aligned}$
F: $\mathrm{FTM} \overline{\mathrm{S}}+\mathrm{p}$ - NSI Full ms 356.10@cid35.00 [95.00-600.00]

| 100 |
| :---: |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |


| $\mathrm{z}=1$ |
| :--- |



MS/MS Spectra of compound $\mathbf{6}$


160518_SY_026_KI_1\#858-967 RT: 7.23-8.47 AV: 107 SM: 7B NL: 5.11E3
F: FTMS + p NSIIFull ms2 $533.30 @ h c d 35.00[50.00-600.00]$




160518_SY_026_KI $1 \# 809-849$ RT: 6.64-7.08 AV: 39 SM: 7B NL: 1.33E1
F: FTMS + - NSIFuli ms2 $533.30 @ h c d 35.00$ [191.00-201.00]


ESI-MS and HR-MS spectra of the crude reaction mixture from the conversion of 5 to 4
O:ILTQ-FT....|AG Klebel151207_EM_016_KI $\quad 11.12 .2015$ 09:31:46

151207_EM_016_KI \#119 RT: 1.01 AV: 1 NL: 1.43E7
F: ITMS + c ESI Full ms [50.00-500.00]


O:ILTQ-FTI...IAG Klebel151207_EM_016_KI $\quad 11.12 .2015$ 09:31:46


## 7. Supporting references

[1] E. Biavardi, M. Favazza, A. Motta, I. L. Fragala, C. Massera, L. Prodi, M. Montalti, M. Melegari, G. G. Condorelli, E. Dalcanale, J. Am. Chem. Soc. 2009, 131, 7447-7455.
[2] Y. Gao, X.-Y. He, Z.-B. Wang, F. Zhang, Y.-Z. Li, H.-L. Chen, Supramol. Chem. 2009, 21, 699-706.
[3] R. A. Laskowski, M. W. MacArthur, D. S. Moss, J. M. Thornton, J. Appl. Crystallogr. 1993, 26, 283-291.
[4] G. J. Kleywegt, MOLEMAN - unpublished program, Uppsala University, Uppsala, Sweden, 1992-2004.
[5] Chemical Computing Group Inc., Molecular Operating Environment (MOE), 2013.08, 1010 Sherbooke St. West, Suite \#910, Montreal, QC, Canada, H3A 2R7, 2016.
[6] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, Gaussian 09, Revision C.01, Gaussian, Inc., Wallingford CT, 2010.
[7] P. v. R. Schleyer, C. Maerker, A. Dransfeld, H. Jiao, Hommes, Nicolaas J. R. van Eikema, J. Am. Chem. Soc. 1996, 118, 6317-6318.
[8] E. D. Glendening, A. E. Reed, J. E. Carpenter, and F. Weinhold, NBO Version 3.1.
[9] R. G. Parr, L. v. Szentpály, S. Liu, J. Am. Chem. Soc. 1999, 121, 1922-1924.
[10] H. Koster, T. Craan, S. Brass, C. Herhaus, M. Zentgraf, L. Neumann, A. Heine, G. Klebe, J. Med. Chem. 2011, 54, 7784-7796.
[11] a) J. Schiebel, N. Radeva, S. G. Krimmer, X. Wang, M. Stieler, F. R. Ehrmann, K. Fu, A. Metz, F. U. Huschmann, M. S. Weiss et al., ACS Chem. Biol. 2016, 1693-1701; b) N. Radeva, S. G. Krimmer, M. Stieler, K. Fu, X. Wang, F. R. Ehrmann, A. Metz, F. U. Huschmann, M. S. Weiss, U. Mueller et al., J. Med. Chem. 2016, 59, 7561-7575.
[12] U. Mueller, R. Förster, M. Hellmig, F. U. Huschmann, A. Kastner, P. Malecki, S. Pühringer, M. Röwer, K. Sparta, M. Steffien et al., Eur. Phys. J. Plus 2015, 130.
[13] W. Kabsch, Acta Crystallogr., Sect. D: Biol. Crystallogr. 2010, 66, 133-144.
[14] J. Schiebel, S. G. Krimmer, K. Rower, A. Knorlein, X. Wang, A. Y. Park, M. Stieler, F. R. Ehrmann, K. Fu, N. Radeva et al., Structure 2016, 24, 1398-1409.
[15] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, J. Appl. Crystallogr. 2007, 40, 658-674.
[16] P. D. Adams, P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve et al., Acta Crystallogr., Sect. D: Biol. Crystallogr. 2010, 66, 213-221.
[17] Smart, O.S., Womack, T.O., Sharff, A., Flensburg, C., Keller, P., Paciorek, W., Vonrhein, C., and Bricogne, G., Grade, Global Phasing Ltd., Cambridge, United Kingdom, 2011.
[18] The PyMOL Molecular Graphics System, Schrödinger, LLC.
[19] G. M. Sheldrick, Acta Crystallogr., Sect. C: Struct. Chem. 2015, 71, 3-8.
[20] I. J. Bruno, J. C. Cole, P. R. Edgington, M. Kessler, C. F. Macrae, P. McCabe, J. Pearson, R. Taylor, Acta Crystallogr., Sect. B: Struct. Sci. 2002, 58, 389-397.
[21] C. F. Macrae, I. J. Bruno, J. A. Chisholm, P. R. Edgington, P. McCabe, E. Pidcock, L. Rodriguez-Monge, R. Taylor, J. van de Streek, P. A. Wood, J. Appl. Crystallogr. (Journal of Applied Crystallography) 2008, 41, 466-470.
[22] R Development Core Team, R: A language and environment for statistical computing., R Foundation for Statistical Computing, Vienna, Austria, 2010.


[^0]:    ${ }^{\text {a }}$ Values in parenthesis refer to the highest resolution shell
    ${ }^{\mathrm{b}}$ Other ligands are glycerol, ethylene glycol, DMSO, acetate and TFA
    ${ }^{\text {c }}$ Calculated using PROCHECK ${ }^{[3]}$
    ${ }^{\mathrm{d}}$ Calculated using MOLEMAN ${ }^{[4]}$

